UNDERSTANDING AND PREDICTING USER ENGAGEMENT BEHAVIOUR ON TWITTER WITH A COGNITIVE BEHAVIOURAL THEORY INSPIRED MACHINE LEARNING MODEL

By

VICTORIA YU

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial fulfillment of the requirements for the degree of

Master of Information Technology
In
Digital Media with a Specialization in Data Science
Carleton University
Ottawa, Ontario

Abstract

To help create more of an understanding on user engagement, this thesis proposes a model inspired by cognitive behaviour theory to explore user-to-user engagement on Twitter, in particular, factors that affect user engagement behaviour. Through case study and a novel model that combined machine learning and the cognitive-behaviour theory, this thesis examines tweets to determine how cognitive behaviour dynamics affect user engagement. The proposed model is evaluated on a dataset of tweets from 10 CEOs (n = 938). The results from the case study approach showed the content characteristics of these tweets such as sentiment analysis and subjectivity affectivity, but no clear idea on which factors affected engagement. The results from the proposed cognitive inspired machine learning model found that topics, emotions and cognitive perceptions play a role in affecting user-to-user engagement. The model predicts that higher engagement behaviours are motivated by topic and positive sentiments.

Acknowledgement

This thesis has been made possible by many individuals who helped me with forming the research aspects, narrowing the scope and perfecting the coding. First of all, I would like to thanks my thesis professor, Prof. Omair Shafiq for giving me the opportunity to work on this thesis. The design and implementation of this research has been quite challenging. Prof. Shafiq was guiding me through many challenges with patience and great empathy. During the extraordinary trying time of the past two years, Prof Shafiq guided me and helped me better understand many aspects of coding and researching within the discipline of computer and data science. I am very grateful for the weekly discussions. Without his support and timely advice, this research project would not have been advanced till the end. I have learned so much and developed self awareness about my strength and weakness. The experience of researching under the guidance of Prof Shafiq expanded my horizon in conducting research of digital media using data science tools.

It would be remised of me to not give my sincere thanks to Moneer Wardark and Nader Naseri Joojili for helping me with scraping Twitter data and fixing and running various codes for the model. I am thankful for you giving your personal time to answer many of the questions I had asked. Your dedication to help others and your superb technical competency are truly inspiring.

Finally, I would like to thank my family for their unconditional love and support throughout my years at Carleton. A special thank-you to my grandmother, Zhou Shufen for tirelessly cheering me on in my pursue of higher education.

Table of Contents

List of Figures

List of Tables

Chapter 1 - Introduction

Chapter 1 introduces key concepts that will be presented within this thesis, including what we will be considering for the concept of user engagement. Following this introduction, Chapter 1 will set out to have a brief discussion about the analytical tools and decision making models that are being tested to measure and predict user engagement on Twitter.

1.1 Context

The focus of this thesis is looking at the factors that affect user engagement behaviours on social media, in particular the factors that drive users to like or retweet on Twitter – one of the most powerful social media platforms in modern days.

Social media platforms are online spaces where people interact to enjoy themselves, to share information about the news or products, the weather, political views, or personal stories. Some popular social media platforms are Twitter, Facebook and Instagram. In recent years, social media platforms have become public spaces where heated debates take place. Activities on social media have become more prominent and increasingly drove by emotions. The effects of social media activities and events raised the importance of understanding the factors behind engagement behaviour. Despite its growing relevance, the academic research on understanding the motivation behind user engagement is limited in scope, limited in approach and often inconclusive [4]. It is particularly limited when it comes to understand what motivates users to engage with each other, and answering why do some tweets get more interactions than others?

One of the most popular social media platforms is Twitter – a microblogging and social network platform to disseminate information [1] where anyone can express thoughts and share anything they would like through messages known as tweets. Compared to other social media platforms, Twitter is considered to be the best social media site for research on social media and

user engagement [2] as there are a lot of resources for extracting and reading the data. Also, Tweets are text based and are short, making tweet data preferable for content analysis. In addition, Twitter is used globally and is known as the most effective ways to express thoughts and opinions. As a result, an enormous amount of data is also kept and is available for research to find out what users perceive from messages, react and response by certain engagement behaviours.

Twitter users are diverse and current research has recognized there is no one-size-fits-all user behavioral pattern [4][6][10][11][12][13][15]. However, there is a consensus among researchers that "retweets and likes are received by a small group of Twitter users". Research attempts have been made to classify users based on the number of followers and retweets, and by behaviour (frequency of use and content preferences) [1] [2] [45]. Overall, recent research found Twitter users are more educated and more likely to engage with new discoveries in science and are more interested to be informed and up to date on world events. A recent Pew research of Twitter users in the U.S. found similar characteristics among Twitter users, however, reported that "a large majority of tweets come from a small minority of tweeters" [45]. A 2021 research found "highly retweeted users" exceeded the influence of experts and professionals or mainframe journalists [45]. The dynamic of super user communities is observed by several recent research on spreading of false information during the Covid-19 pandemic [21][23][41]. They observed a small set of users who received disproportionate attention which is measured by the number of retweets and likes [21].

User engagement is a key concept in the research of digital social media. A comprehensive conceptualization of "user engagement" on social media is known to have the following facets: self-presentation by creating a self-identity on social media, action and participation through behaviours such as likes, retweets and replies on Twitter [1]. Recent research on evaluating the

behaviour of participation of users on social media is trending towards understanding the forces behind engagement of users on social media [1][2][5]. In particular, studies have examined the "determinants that drive virality" in the context of user-generated content including image content, memes, and videos [9]. With the content at the center of engagement, user engagement is used to describe "interaction with content" as a result of series cognitive and psychological processes such as thoughts, perception, attention, value, positive or negative feedback [5][6].

More recently, engagement behavioural such as likes and retweets on Twitter are becoming more practical for measuring user engagement [2] [7] [12]. The growing social media usage creates colossal amount of data that is informative about the cognitive reflections of the human mind. While machine learning tools are effective in investigating the factors that motivate millions and millions of user interactions, they are not able to unveil the cognitive content of user interactions, such as specific emotions or ideas relative to perceptions from stereotype to hate speech [3][4][5][21].

Effective online communication relies on the ability of content creators to turn users into active participants. Therefore, the more public the expression of interest with the content by the users, the higher level and impact the engagement is. The volume of likes, retweets and shares is typically an indicator of level of engagement. Considering motivation factors that drive this volume is thus useful for building models that better measure and enhance user engagement. Knowing what factors could increase user engagement, content creators could use the knowledge to fully utilize social media to spread truth, inform the public, and encourage discussions of social issues in a healthier and constructive manner.

1.2 Problem Statement

The main goal of the research is to attempt to expand our understanding of the factors that influence user-to-user engagement in social media. The way we have thought to expand this understanding is to tap into some of the cognitive factors such as how human thought process could influence engagement. These factors are less understood and deserve some spot light. We plan to use extensive feature engineering to detect cognitive features such as thoughts, perceptions and emotional responses to the content posted on social media.

The recent research efforts increasingly recognize the role of content characteristics, emotional response in influencing the level of user engagement. However, there are very limited conclusive findings to guide the research on how to approach evaluating user engagement through a cognitive lens. Defining emotional responses and measuring emotion response with data analytic tools could be difficult as well because machine learning tools are not always accurate, and most of the attempts to measure user engagement with social content are inconclusive [4]. Finally, the problem of having a dataset for this research is difficult. Due to privacy and security concerns, Twitter gave users control over account access, meaning it is not always easy to get into user accounts for public research purposes. In addition, most social media platforms like Twitter have design features that allow users to protect and limit interactions to only people they know or follow. As a result, access to user to user engagement dataset can be very limited. It also makes it hard to conduct research on the characteristics of content or measure levels of engagement in a comprehensive manner [2] [5]. Manual collection of tweets for each user is also time consuming and not practical. As a result, the size of tweet dataset feasible for this research is limited.

The thesis also aims to expand the understanding of user engagement on Twitter by focusing on individual user-to- individual user, and identify key actions for future research. This

is different from the traditional approach where user and organization level engagement was often the focus.

The above objectives are fulfilled by answering the following research questions:

- **RQ1:** What content characteristics can be detected from collected tweets?
- **RQ2:** Which content characteristics predicts positive affect on likes? On retweets?
- **RQ3:** What effects do emotion have on user engagement (likes and retweets)?
- **RQ4:** How do cognitive factors predict emotional effect on likes and retweets?

1.3 Overview of Proposed Solution

As part of the solution, this study proposed a novel cognitive-behaviour theory inspired machine learning model to improve measuring and predicting power. This research aims to focus on developing an extensive feature engineering criterion that will focus on content topic, lexical features, polarity contrast and sentiment and their correlation to user behaviours on Twitter. Our focus is on retweets and likes, which we define as "forms of engagement". Previous studies focus on limited features through text analysis and surveys while this research applies extensive feature engineering to detect cognitive features such as thoughts and perceptions (represented by topic and themes with sentence structures), emotional responses (positive, negative and neutral) and behaviours (represented by tweeting habits, frequency of tweets, time of tweets). The study evaluates the correlation of the input features to the level of engagement (likes + retweets). We collected tweet data through feature engineering architecture to predict the likelihood of likes or retweets.

User engagement has been conceptualized as measurable individual interaction with content created by other individuals online in the space of social media platform either it be Twitter, Facebook or Instagram and others [1][2][4]. The engagement behaviour is broadly

characterized as “liking” and “sharing” content in a post or tweets and measuring the number of sharing of original posts recommending its content to a wider audience. On Twitter, liking can be measured by the number of “likes” while sharing can be measured by the number of “retweets”. Liking and retweet are correlated but often modelled separately in social media studies [5][28]. This conceptualization of user engagement allows us to understand and measure a range of factors that might influence the level of engagement. For example, the number of likes on Twitter can be driven by the information the content provides by the value of “entertainment” and “excitement” embedded in the content, by content’s positive sentiment, the words, the time the tweet was posted, and level of influence of the content poster/creator [1][2][3][4][5]. In our proposed research model, retweets and likes are the focus although it is recognized that there are other forms of user engagement on Twitter such as replies and quote retweets that can also be quantitatively measured. We will pay particular attention to the factors that affect likes and retweets, especially the cognitive factors such as thoughts, perceptions and emotional responses.

Our proposed solution took into considerations that user engagement is not simply based on the unique features of a particular platform alone. It is, instead, driven by many complex factors that are reflections of the users themselves, especially how they make decisions [4]. In user to user engagement, individuals are responsible for the creation of the content and individual users chose to interact with the content having different motives, emotional and cognitive capacities. Theoretical models that predict user engagement have been slow coming due to these complexities. Most machine learning based models are not always able to perfectly detect sentiment, feelings and perceptions due to the complexity and “noises” of human behaviours [4][6][26].

The novelty of the proposal lies in its focus on role of human cognitive processes such as thoughts and perceptions in influencing the engagement behaviours of likes and retweets. The

massive amount of data left on Twitter opens opportunities to expand our understanding of the factors and integrate new elements to improve prediction models. This is because social media data can reflect the users' minds, their perceptions, and their cognitive biases. These deeper driving forces are expressed through language, concepts and emotions and is left on the social media platform. Social media has a strong cognitive component because they are mainly made of knowledge, that is, emotions and ideas, flowing from use-to-user along with social ties [6][26][28]. The angle focusing on cognitive dimensions of social media is only a recent phenomenon. In data science, prediction models focus on creating a clear connection between the outputs of a model and its input features. But language, feelings and emotion are often fuzzy, requiring tools for understanding the meaning of words and knowledge surrounding them. This was argued in more recent researches as achieved with an accessible and interpretable map of cognitive associations. In comparison, machine learning is powerful in tagging stances or sentiment patterns. However, the cognitive considerations are more interpretable. This is because the cognitive approach goes beyond tagging and listing but explain the meaningful cognitive associations between ideas as available to both text authors and readers [6]. The other theoretical framework is the "cognitive behaviour therapy" theory which is centred on the cognitive triad that recognize the connections among thoughts, feelings and behaviours.

1.4 Structure of Thesis

This thesis examines user-to-user engagement on Twitter with goals to expand the understanding of factors that influence user to user engagement on social media and to build cognitive model to predict the user engagement. To reach these ambitious goals, this research investigated the users as sender/creator of tweets and users behaviour as the receivers of the tweets. First, a case study approach is applied to understand the characteristics of the content created by a

group of influential users. A database was established consisting 10 datasets of tweets from 10 CEOs to extract various features to see which of them could have affected the level of user engagement on their posts. The features include volume, sentiment, lexical richness, language features, frequency of tweets and reactions of users to tweets (the number of likes and dislikes). Second, this research proposed a novel cognitive inspired machine learning model to predict behaviour. It is proposed to investigate how thoughts, perception and cognitive biases could affect emotions and the behaviour of liking and sharing. The proposed model is based on "cognitive behaviour" theory that suggests behaviours are results of internal processes of thoughts, perceptions and emotional reactions. The model is also validated with accuracy, recall and precision. The results from the case study and predictive model answer the research questions. Finally, future work to advance our understanding of the topic and improving the accuracy of the prediction model is explored.

Chapter 2 - Literature Review

In this chapter, a collection of research and studies was reviewed, compared and analyzed for understanding user engagement on social media network sites. The collection contains the most recent research published between 2018 and 2021 that studied user engagement with social media platforms from a variety of perspectives. During our search, the articles were not limited by whether they only used Twitter as a data source, nor was it limited by academic discipline and culture regions in order to collect as many relevant articles as possible. The literature review covers the objective, methodology, results, and the strengths and weaknesses of each research paper. Presenting the strengths and weaknesses reveals diverse perspective and research methods that shed light on the topic of user engagement on social media network sites. Following the review, a comparative analysis using tables is composed to illustrate key factors and themes already explored thus far on our subject matter. Finally, we include a brief discussion of research gaps emerged from the literature review and comparative analysis.

2.1 Literature Review

Existing research has provided definition of user engagement, introduced a variety of theories and tested scales to measure user engagement. Most recently the use of memes, pictures and videos has spurred research interests particularly in the area of communication strategies due to the potential of them to increase user engagement.

2.1.1 Conceptualization of user engagement

The first comprehensive research on user engagement in this collection is by L. Sigerson and C. Cheng, *Scales for Measuring user Engagement with Social Network Sites: A systematic review of psychometric properties* published in 2018 [1]. This research aims to conduct a

systematic review of the psychometric properties of various scales developed for measuring user engagements on social network sites. The researchers also attempt measuring 14 psychometric scales with five key criteria including: content validity, reliability, structural validity, associations with other variables and response processes. The conclusion of the research established the feasibility of looking at user engagement from psycho and social perspectives. At the same time, they conclude that further research is urgently needed to cover user engagement on social media networks.

The paper's most important strength [1] and contribution is the establishment of a good working definitions for "user engagement", referred as "a quality of user experience with technology" [1]. In this context, the "technology" refers to internet based platform that bears characteristics such as "challenge, aesthetic and sensory appeal, feedback, novelty, interactivity, perceived control and time, awareness, motivation, interest, and affect" [1]. Thus the outcomes of such user experience can be psychological and social. Another strength of this research is the 14 scales that provide tools for studying user engagement on social media sites. Among them, Psycho-social aspects of Facebook use (PSAFU), Gravitating Towards Facebook scale (GoToFB) measure psychological aspects of user engagement on Facebook. This research concludes that the distinctive theoretical frameworks show that options are available to use for data collection and to measure user engagement data. Finally this study recommended a step by step action plan when selecting a scale for measuring psychometric properties. This can be a useful strength for future researchers.

As for the disadvantages, this research focused almost entirely on Facebook with scales mostly assessed with Facebook users [1]. The narrow focus and limited understanding of the use of scales for other platforms such as Twitter. Therefore the scales used will have to be re-assessed

for Twitter users, which is not within the scope of this thesis. Nevertheless, it spurs interest in valid measurement tools.

A year after their study, Diaz-Faes et al conducted a study titled *Towards a Second Generation of Social Media Metrics* [2]. This research empirically examined the full range of Twitter user's behaviour within the Twitter platform. This research is said to have broken new ground for systematic analysis and characterization of social media users around science with a large-scale analysis of tweets. The researchers aim to build a reliable measurement source that can be utilized for the full range of Twitter. As a result, the large scale analysis established four latent dimensions that they called "Science Engagement", "Social Media Capital, "Social Media Activity" and "Science Focus" [2]. The activities of 1.3 million unique Twitter users were collected with the altmetic.com database which counted a total of 14 million tweets [2]. They concluded that although the proposed new social media metrics can be helpful, there is still a long way to go for potential use of the social media metrics in a research evaluation context. Empirical evidence that helps to disentangle what these metrics actually capture is missing and there is a dire need to define them [2].

The main advantage of this research is the highlighted profile description as an "appealing source" [2] for understanding the type of users. In the research, the authors used the "profile description" of the background and professional activity of users to complement their empirical findings. Another strength is that the research promotes the lens of "social media capital" to understand the meaning of the number of followers, the number of users followed and lists [2]. The social media capital metric illustrates the network of relationships established through user engagement with social media site. While the article itself states that it will work best for bio health

and health sciences and humanities, having social media metrics has the potential to be applied to other areas that might require the need for social media metrics.

However, the weakness of the research [2] relates to its database limitation in which only certain type of Twitter users is included. The typology of users around science may vary from cultural and social, and even from regional perspectives. Time frame limits the range of indicators expanding the potential of the metrics. For example, as society and science move forward, new indicators need to be considered to assess the social impact of science on the Twitter community. Finally, the latent dimensions focus entirely on user engagement to science and a focus on science could be a limitation that needs to be cross-validated. In a different setting beyond science, the validation will have to check the underlying structure using another sample.

Maria Teresa Borges-Tiago et al. looked into the factors of influencing user engagement in their 2018 study "*Exploring users' motivations to participate in viral communication on social media*" [3]. They looked into the factors that would influence a users' willingness to engage with viral content on social media. In particular, they explored different types of motivation that get users to interact with certain content. In this research, partial least squares structural equation modeling was used to examine data collected from Facebook users' survey with focus on a case study. The researchers also utilized "eWOM (electronic word of mouth) and viral-marketing-communication to focus on comment types, credibility and comparing the positive and negative content" [3]. Using tools such as the Bergman's model, Borges-Tiago et al. defined three different types of users based by how they interact with social media including "heavy user" the search-driven users and "social-driven users". They found that content with emotional tone and arousal level has significant impact on users' attitudes regarding sharing communication [3]. They also

reported the personality factors are most influential as to whether or not a user will engage with viral content.

The most relevant advantage of this research [3] is its focus on motivation factors of user engagement. It provides working definition of key concepts such as motivation and viral content. It is important to understand motivation from the user point of view, especially what catches their attention. Users use media for utility and as an information tool, both at work and in private" [3]. Another strength is the identification of three potential types of social media users; the heavy users, the search-driven users and the social-driven users. They observed that "heavy users" tend to "use a wide range of media frequently, using the most advanced facilities compared to the rest of the user population" [3]. Social-driven users were a "quite new and increasing user type because of the advent of social media applications" [3] and were defined by their tendency to use the Internet for convenience information and social motivation. This research has its own limitations. Particularly, the data is derived from user statistics from Facebook and is within marketing research. It is not clear if their findings apply to users on Twitter or other social media platform.

In terms of conceptualizing user engagement on Twitter, the meta-analysis conducted by Schreiner M., Fischer, T., and Riedl, R. [4] presented a review of 45 studies that looked at the impact of "content characteristics" on the level of engagement of Twitter users. Schreiner et al., reviewed the literature on the relationship between content, emotions and engagement. Because it is a meta-analysis, Schreiner, et al., did not establish any unique variables of their own. They identified several "appeal categories" that have positive effect on user engagement. They include "providing information", "providing entertainment", "leading to transactional actions", "opportunities" and "emotional such as content evoking fear, humor and romantic feelings" [4]. This review noted "high media richness" that include "pictures or videos" can have a positive

effect on user engagement behaviour. Emotional arousal seems to enhance engagement behaviour, including positive emotional response evoked by humor [4].

As a meta-analysis, it has a great advantage [4] of showing a wide range overview on a lot of the current researches. Even if it is described in a general sense, seeing all the traits and content characteristics helps to establish, in the best way, what the existing research landscape looks like. It offers a lot of different perspectives about the characteristics and content engagement.

The existing research has a couple of limitations [4]. Number one weakness is related to the nature of a meta-analysis. As such, it is hard to eliminate all publication bias. The 45 papers that were collected could have a bias that affects the results they found in the paper. Another weakness is that the researchers only consider immediate emotional responses which did not account for other affective constructs such as moods [4]. Lastly, the specific focus on content characteristics can be easily manipulated by the person creating the online content, and thus leaves further room for extension.

In *Perception and Practices of Social Engagement: A global perspective* [5], the researchers examined the conceptual, strategic, and practical dimensions of media engagement through a qualitative study of in-depth interviews with 74 experts [5]. This research went to a deep discussion about "user engagement" through the users themselves. The researchers recognized the multi-dimensional aspect of user engagement, and emphasized studying engagement presents a more complex qualitative nature. They found "promoting engagement" is a critical strategy to gain audience attention. It is a complex task as promoting engagement automatically has cognitive-affective-behavioral aspects, but is valued and practiced on a behavioral level most often [5].

This study highlights the cognitive aspect of user engagement, in particularly the users perceptions of engagement. The authors explore engagement through interviewing the experts

including CEOs of selected global companies who speak to what motivate them to engage users in the context of marketing. They found user engagement has cognitive, affective and behavioral aspects. Engagement has a quality component, something that goes beyond simply counting the likes and retweets, therefore content authors should try to put "soul" [5] in the content to appeal to users. Thus this study identified the cognitive association that is available to both users as receivers and users as the authors of the content. However, the entire study was conducted through interview with limited sample size (n=74). The engagement is strictly in the context of media marketing. Their findings have to be tested and quantified in other setting and with a bigger dataset.

At the end of 2021, Stella published their first comprehensive review called "*A Cognitive Network Science for Understanding Online Social Cognitions: A Brief Review*" [6]. The paper is about the cognitive components of social media and how investigations of language, semantics, and knowledge structure in social media can unveil conceptual associations between ideas as expressed on social media. The review outlined how cognitive approaches could enhance machine learning tools to explain user engagement behaviour. The author points out the power of machine learning tools, but they have not been able to inform on a variety of features, such as conceptual associations, interconnected emotions (fear in user retweets of Covid – 19) and perception about specific content and how it could influence behaviour. The author argues that the next generation tools suitable for processing and interpreting knowledge is the one that combines automatic tools and cognitive models. Machine learning and AI are powerful at spotting knowledge patterns and correlations often invisible to the human eye in a large dataset [6]. However, they often lack clear-cut cognitive grounding and interpretability. Cognitive networks can include emojis, hashtags and other feature of online language as much as they can contain useful information. Hashtags can provide information about the topic of messages thus allowing selection of messages around

specific topics. Cognitive interpretation also capture semantic prominence in text beyond word frequency. In summary, Stella provided cognitive tool for understanding and explaining the "information flow from user-to-user along with social ties" [6]. This information flow is used by users to either "build or express their own experience in ways that are not fully known yet"[6]. It points out social media are digitalizing massive amounts of users cognitions in terms of timelines and emotional content, but there has been little uptake using cognitive approach to look at user interaction on social media. The author points out that Big Data explorations of social media successfully investigated millions and millions of user interactions, but they did not unveil the cognitive content of such user to user interactions [6]. For example, the machine learning model is unable to tell whether ideas expressed are relative to aspects resemble stereotypical perceptions or hate speech. The review recommended integrating cognitive approaches such as using salience, perception and biases when explaining factors that influence engagement.

This review points to a crucial research direction by exploring cognitive phenomena related to salience, perception and biases when explaining what factors influence engagement [6]. The cognitive dimension of messages is key not only for investigating peoples' perception but also for studying salient features of users' minds like attitudes or personality traits to identify how individuals reason and behave on social media. through cognitive networks can greatly benefit several research lines studying cognition like semantic framing, stance detection and perception, emotional dynamics, and cognitive biases as outlined above.

2.1.2 Predicting user engagement – content characteristics, role of positive emotion and emerging features

While user engagement behaviour such as the decision of liking or sharing has been explained in a number of researches in recent years [7] [8] [9] [10] [11] [12] [13] [14] [15], predicting user engagement remains to be a challenge.

In a study about what make people more engaged on social media [10], researchers predicted social engagement from use of words. The study analyzed the role of word usage in Twitters replies and retweets by using psycholinguistics and correlating psycholinguistic category scores with replies and retweets. The researchers randomly sampled 17640 users then used the Twitter REST API to obtain their past 200 tweets. They made two categories called past_retweet_rate and past_response_rate and then used a rule-based method to identify any questions in the tweets [10]. Thanks to these rates and the information from them, they identified the questions users are more likely to respond to. In the end, they found that the predictive model of retweets from word uses have been able to identify users who are more willing to retweet a tweet [10]. The rate can be increased for users in virtual marketing, propagating emergency news and social/government/political campaigns. The regression analysis also confirms that this can be done with reasonably low mean error. The classification even manages to classify high/low responders/retweeters that can be done quite accurately. Although they did not look into ranking algorithms, they concluded that psycholinguistic analysis from word usage has significant correlations with social engagement behaviour in retweets and replies. In addition, the study found that a few psycholinguistic categories share common characteristics for responding and retweeting behaviour. Both models have shown a fair amount of accuracy. One way that the paper would like to improve is by extending their analytics to other social media platforms and verify how generally the findings can be used for social engagement. However, for the purposes of this thesis, the fact they focus on Twitter is very useful. Although another disadvantage that paper has is its time

period and whether or not quote retweets on Twitter will net similar results from the predictive models.

A prediction model is used to look at response and retweet based on the psycholinguistic category features to understand their predictive power. Using both regression analysis and a classification machine learning tool, the study concluded that word use is a strong predictor of retweet behaviour. People who use words such as “along” and “also” scored high and deemed as “likely to retweet”. People who use words such as “ask, call” and “contact” indicate those people are more likely to be more interactive and willing to retweet. Like the other researches [7][8] positive feelings are predictors of increasing level of response behaviour [10].

Several studies [9][12][13][14] [25] in our collection looked at how “thoughts” are expressed, and whether the format of expression, memes [12], videos, audios or photos [9] affect user engagement level. These researches tried to predict if topics and the usage of new format to express thoughts through social media can be used to see the changes in user engagement levels. The rising of memes soon caught the attention of Rieger, D., and Klimnt, C. [12]. In their study of memes in social media [12], they looked at the eudaimonic value internet memes have for users. Memes are increasingly prevalent on the Internet, especially on social media. The research found memes often stand out against the many other messages with positive hedonic qualities to which most users have already adapted to. Memes mixed with humor and hedonic aspects can increase the chance of eudemonic memes to receive relatively great attention and elaboration by users [12], which pave the way for more intense and diverse affective responses, including a heightened sense of inspiration, motivational intensions. Inspiration has been identified as an important motivational outcome of memes. The results reveal that memes evoke emotional responses from users.

The research on memes fills an important gap in understanding user engagement on social media. Therefore, it is considered a strength of the research [12] [13]. It covers the concept of user intention from the unique angle that has not been commonly considered in the existing body of literature. The casual user on the internet is more likely going to come across a meme of some form. Memes have evolved over the years from simple images to heavily edited videos or audio, and they have a major effect on the online culture. Memes are important pieces of the Internet and user interaction as they can easily gain a lot of traction.

Understanding motivations behind user preference is an important part of studying user engagement on social media. In *Is a Picture Worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement*, Li and Xie [9] took a deeper look at the traits of visual content on social media to determine what traits encourage people to engage with certain image posts. To do so, Li and Xie [9] analyzed the interaction that people had with photos of SUV models and then the interaction people had with photos of major US airlines. They found a "significant and robust positive mere presence effect of image content on user engagement in both airline and USV" and across Twitter and Instagram. This finding showed image content can be a powerful tool to improve user engagement.

The strength of this study [9] goes behind the scene of one of the most dynamic trend in social media – image-centric social media apps which became massively popular with smartphones and improved mobile internet-based technology. Visual content has taken a strong hold in social media leaving a gap for researchers to explore. As the first to explore the effect of image on user engagement popularity [9], this study fills the gap of studying non- text based content. The research asked if certain image characteristics attract more user engagement. If so what are the reasons? The researchers proposed 3 lens to show how to quantify the impact of image

content on social media– mere presence effect, image characteristics effect, image-text fit effect. In addition to fill the gap of studying image content, the researchers selected 3 data sets: 2 from Twitter and 1 from Instagram. For the 2 data sets from Twitter, researchers measured the engagement level of each tweet using the number of likes and retweets. Their findings shed light on the social media content effectiveness on Twitter especially the effectiveness of pictures used in tweets. For example, their analysis revealed using a picture increases user engagement, picture colorfulness and picture with a human face increase engagement. To overcome the challenge in mining image data, Li and Xie employed Google Cloud Vision API to extract information [9].

In *Integrated & Alone: Twitter Social Activism* [16], Simpson took a look at hashtags and whether using of it will trivialize social activism. With tweets scraped using Python on Twitter, 300,000 tweets were collected. However, since they also had to have a specific hashtag in it, such as #BlackLivesMatter, #TakeaKnee or #MeToo, the authors condensed tweets into a 2,000 tweet sample. Afterwards, these tweets were closely followed to see when they were used and how, especially to answer the question if they were used within the context of social activism. Simpson concluded the usage of hashtag facilitated the engagement of social activism. The research fills a gap in the current studies of Twitter. As hashtag is becoming an iconic feature of Twitter, its power should not be understated especially when user engagement is looked at under social movements.

One of the disadvantages is the narrow focus of the study. While it certainly teaches us about hashtag culture, it makes one wonder if hashtags can be effective in spreading information and attracting people's attention if they were not social movements. What is the difference between hashtags supporting a social movement and hashtags supporting something more dangerous? There might still have to be some validity and reliability done, or there are other factors that were not considered.

Although, Aldous et al [14] looked at the content from text based perspective only, their study analyzed and investigated the effects of the topics of social media posts on user engagement across five social media platforms. The researchers used different language features to input of their prediction models [14]. They evaluated user engagement with on-line content (n=3,163,373 postings from 53 new agencies) from Facebook, Instagram, Twitter, YouTube and Reddit. Researchers concluded user behaviour is different from platform to platform [14]. This conclusion is also confirmed by a big study by Voorveld et al. [25] that examined multiple platforms to test how they drive engagement with survey of 1,346 people aged 13 and older. They mapped out the user experiences with Facebook, YouTube, LinkedIn, Twitter, Google, Instagram, Pinterest and Snapchat.

In an emerging research on industrial applications of chatbots [7], researchers proposed a cognitive model for predicting emotions. Adikari et al., suggested extracting emotions from conversations, detect emotion transitions over time and predict real-time emotions and profile human participants based on their distinct emotional characteristics. The Chatobox model describes a generic cognitive process with "input" and "output" and flows through three layers: users of social media acquire 'information' from posts, to cognitive layer such as perception and emotional reaction, to responses layer which compile types of actions. The emotion prediction study was designed based on the second order Markov model. The forecasted emotion was assessed based on (1) correct emotion category as positive or negative and (2) correct emotion. The researchers found prediction accuracy increased when having more historical data as the emotion transitions are more accurately modeled. Taking a similar input and output approach, another study [8], looked at the role of perceived creativity and cognitive evaluation of thoughts in predicting engagement level on Instagram. Using structural equation modeling, researchers

confirmed that positive emotion is a strong predictor of user engagement while the perceived value of post is a weak predictor of the positive emotions triggered after viewing brands publications on Instagram.

The weaknesses for the above research are often related to the datasets. The study of memes [12] excluded the most advent consumers of memes on social media – the youth. The research was conducted with samples of adult users in Germany, thus may have skewed the level of responses without surveying how the youth respond to memes. Timing of viewing memes may also affect emotional responses. Participants estimate their emotional responses based on a meme they saw recently not when actually seeing one. Therefore the results may be screwed as emotional and motivation responses in retrospect is biased and unreliable based on memory not a real-time viewing of memes. For the study of the impact of using "picture" in content is limited [9] to firm/industry user generated content, for example media posts from two specific industries SUV and airlines. Thus the findings may not be applicable to "user to user" engagement. Similarly model limitation is evident in Aldous et al [14] in their study of influencing factors in social media platforms. They concluded the overall low similarities of content across social media platforms indicate that a model predicting user engagement on one platform is likely not to be transferable to another.

Finally, in *Text-based emotion detection: Advances, challenges, and opportunities* [24], the researchers took a deep dive in the subject and proposed models to extract emotions from text-based conversations to predict real-time emotions and profile human participants. That is based on the history of emotion states, the model could predict the next emotion state of the participants. This model is built based on Plutchik's emotion extraction process, which contains 8 emotions; anger, fear, sad, disgust, joy, anticipation, trust, surprise [24]. When applying this model,

researchers created a dictionary and train the machine to identify semantically similar words and capture the sentiment. User profiling based on the emotion detected to presume the emotional characteristics of participant. Is he showing negative emotion or positive ones? The prediction component of the model was able to predict the next state emotion with accuracies of 63% and 70%. Researchers found more historical data as the emotion transitions are more accurately modeled. In the research paper [24], researchers surveyed a variety of emotion detection models (ED) or sentiment analysis. This research conclude that detecting emotion from text-based messages such as tweets has its own challenges, especially text does not always let out clear cues of emotions. In particular, image, memes, emoji, gram errors and continuous evolution of new words as a result of language dynamics.

2.1.3 User Engagement - explained from the receiving users

Equally interesting are several researches who turn their focus to user profiles to see how traits and characteristics of receiving users can influence sharing behaviour and overall level of engagement. In *Understanding User Profiles on Social Media for Fake News Detection,* researchers examined how fake news is shared around social media. Shu et al. [22] investigated the correlation between user profile traits and fake news detection. They started the research by analyzing traits and characteristics that are commonly associated with users on Twitter who share fake news and those who don't share fake news. By constructing real-time data sets, Shu et al hope to see if a correlation between these two sets of variables exists. Profile features were collected. They include content related features as well as implicit profile features such as gender and age. The experimental results showed that there are specific users who are more likely to trust fake news than real news and that these users are more likely to reveal different features about themselves than those who trust real news.

A strength of this research is that it provides us with a framework in using feature engineering and studying for user profiles [22]. The research pointed out that explicit features with representative attributes such as the number of posts, the number of follower count, are available for all public users. They can be insightful for detecting the type of users who are more likely to trust fake news. Implicit features such as age, gender and personality are not readily available but are most used to depict user profiles to understand how user engaged with social media. Although not directly provided, researchers used "linear regression model" [22] with predictive lexica words and weights to find user data. The researchers found older people and female users are more likely to trust fake news. Using a "text-based personality prediction model", researchers found people with high extraversion and openness are more likely trust real news [22].

Despite its great potential, profiling in the research has its weakness [22]. First of all, very limited user profile features are included in the research, namely age, gender and personality leaving many others such as "political bias" and "user credibility" out of the research. This could potentially weakened the observation as bias and user credibility could be powerful traits to provide insights on how users engage on social media.

When the COVID-19 pandemic started, fake news is rampant resulting in misinformation on social media. "*Fake news and Covid-19: Modelling predictors*" is a timely study that looked into the factors contributing to fake news spreading on Covid-19. Apuke, O.D., and Omar, B. [23] examined if it was possible to make a model that predicts how fake news about Covid-19 was spread. By studying a sample of 385 in Nigeria, the researchers investigated the fake news phenomenon applying the Uses and Gratification framework. They found that *altruism* was the greatest motivator among all the variables that led to users spreading misinformation [23]. They

found entertainment had the lowest significant association with predicting the spread of false information about COVID-19.

The strength of this research is the prediction models established to figure out the motivations behind users information sharing behavior [23]. This research extended the common approach, applying Uses and Gratification theory by adding social impact theory to identify the needs and desires that an individual displays when using a particular media channel. One of the gratifications is "status-seeking" which refers to the user's gratification to attain status within their network. Previous work found "status-seeking" [23] be the strongest predictor of news sharing on social media. Under social impact theory, "parasocial interaction" refers to the degree or propensity of an individual to develop an emotional connection with a figure considered as a guide or a role model towards individuals they admire or respect. Views shared by public figures are held in high esteem. Parasocial interactions are positively associated with the sharing fake news on Covid -19 pandemic. The parasocial interaction model [23] has been the main approach to study user engagement with media. It has potential to be applied to analyze other behaviours. The framework can be useful to determine intrinsic motivations. The data analytical model Partial Least Square (PLS) was used to determine how motivation factors can affect the outcome of fake news sharing behaviour. The results of this research showed parasocial interactions predicted fake news sharing which is consistent with previous research findings. Different from previous study [22], however, this research found gender and age have no significant impact on fake news sharing. The biggest limitation of [23] this study is like many others lacks a more comprehensive inclusion of other factors such as peer influence, biases and other predictors.

Activism is studied again by Seong Jae Min [17] in "*From algorithmic disengagement to algorithmic activism: charting social media users' response to news filtering algorithm*s."[17]

However Min looked at activism from algorithms particularly by looking at how users react to algorithms and how it affects news getting filtered. The research surveyed 3441 Americans and researchers asked users to describe their experiences with algorithms. By looking at the results of the survey, the researchers concluded that human psychology and capacity determine meaningful tech use, and that those who have higher level of Internet skills and political efficacy are more likely to be characterized as "activists" [17].

Studying the algorithm and how it gets utilized in social media is very important. Algorithms are always around to track our interests to predict content a user is more likely than not going to interact with. Giving us different types of users and how they wind up responding to algorithms, including what they do with the news afterwards. It allows us to create a predictive model based on the type of users identified through algorithms [17].

The weakness of this research is that the categories of users are not comprehensive [17]. It is hard to truly say that one can safely and perfectly put categories of users into the same neatly fit descriptions. While this study recognizes human psychology influences engagement, it does not include many human factors such as age and gender to its predicting model. Therefore the categorization of users could be weakened.

2.1.4 Examining user behaviour and motivations in case studies of communities

In this fascinating study of inferring the unseen data from user conversations on social media, A.M Sadri et al., used their own metric called 'perplexity' to measure predictive capacity [15]. The main focus of the paper is for the researchers to take a look at user communities in social media networks and how the different interests apply to each other, and how people interact based on that. *Joint Inference of User Community and Interest Patterns in Social Media Networks* [15] concludes that users are linked not only by interest similarity [15] but also "affected by the

community ties between users". Using "pattern inference models such as Interest Pattern Model (IPM), researchers "predict user community" and observe the interaction topics and communities relate to two events visit by Bernie Sanders and a school festival within Purdue University. They conclude also that such prediction can be used to reach target audience and organize successful events.

One of the greatest advantages to this thesis is that it provides a several models for inferring user interest and the communities they belong to using large scale data from their interaction on social media [15]. They demonstrate it is possible to use these models to reveal data and information in the millions of online user conversations. Topic modelling methods such as Latent Dirichlet Allocation (LDA) is a widely used machine learning tool and the researchers successfully prescribed steps and procedures reaching out a conclusion. These models can be used in other topics. A disadvantage of the study is that the data that is offered here looks into very specific events that were only relevant to one University. As a result, there is a lack of representation and a sampling bias as it only applies to students who might attend that University. Researchers declared the challenge of scalability and issues with their algorithms for real world application given the size of the data [15].

Hutchinson's *Micro-platformization for digital activism on social media* [32] continued the research concerning user engagement of the digital activism, a community that is struggling with being visible on social media platforms. He argues visibility is a key concern for online activists to stay as a sustainable endeavour for the activists that engage in media content production on various social media website [32]. In this thesis, research data highlighting the commercial publishing strategies of social influencers on You Tube was presented in comparison to the lack of presence observed of activism. Hutchinson found social influencers with extraordinarily large

audiences dominate social websites. The influencers are mostly promoting commercial products for personal gain, whether it be monetary or social fame. The commercial influencers manipulate mechanics of platforms to increase visibility. The result of the research shows if social activists want to be sustainable and increase level of engagement they should adopt strategies used by commercial users who make sure using micro-platformization that connects content producers with users.

One advantage of this article is its insight in the power of content especially those that fall into the "entertainment category" [32]. It points out how commercial user community uses this category to stay visible and suggests activists should use an influencer to promote "morally valuable content" beyond self-branding [32]. However, the research [32] suffered from a lack of effective models for extracting data from YouTube. There may also be the chances that there is a bias as they seem to only focus on the videos of specific topics.

The case study of user engagement on social media in the context of University of Barcelona [29] with data from the University of Barcelona's Instagram provides relevant information regarding the complex use of social networks by students and the university as a way to generate engagement. They found overall the interaction of a university with its student users is limited. The topics and strategies greatly influence the responses. Many of the posts were studied, as well the types of engagement it was receiving by posts from the ones that were created quoting inspirational figures, to those posts letting students know about big events. Comments on the posts were closely examined. The researchers concluded that the insights observed could help higher education institution networks remain active and generate more participation in students. This, in turn, can help universities to develop stronger communication strategies to create better relationships with their stakeholders.

The focus on the impact of a university Instagram fills a research gap as the use of Instagram account by a higher education institute has not received a lot of research attention. In addition it provides a potential method that researchers could use to make Instagram a viable use for information. Instagram may not have as much net value compared to Twitter and Facebook. However, Instagram's influence is growing as noted by researchers [9] [14] [25]. As a weakness, though, this research is mostly based on one University's Instagram account, namely University of Barcelona. In addition, students' preference is changing. They may change the social network of reference as illustrated with the rising and declining popularity of Facebook. Finally, there might be the issue of "bot accounts" that do not belong to actual students. Random engagement from bot accounts skewed the level of engagement. Therefore the researchers called for more studies using a more significant number of higher education institutions and more social media platforms to get reliable results.

Brenna, G., et al focused on one of the most important parts about social media sharing behaviour [18]. In "*News Sharing User Behaviour on Twitter*", [18] the researchers look into the determinants of going viral or why news makes the rounds so quickly. With this in mind, researchers wish to observe the behaviour of users to determine what quality of online news articles motivates people to share the articles. The goal was to create pipelines to gather data that can enrich the dataset. This dataset focused on including news articles from major US news outlets and sharing activities with Twitter. At the end of it, they managed to implement their data collection pipeline and comprehensive dataset to provide an example in the context of political stance classification.

The strength of the article [18] is the use of navigational clickstream data. It allows researchers to show the usability that the sharing can be tracked. The user model also composed

of describing four different dimensions that helped them take a closer look at Twitter. Using social identity, user-generated content, news articles and the associated Twitter network as main dimensions can be useful in figuring out details about not only users, but small community clusters on Twitter.

One of the greatest weaknesses of this research is that the methods for the database are at the beginning of its development [18]. They require a lot more testing to be certain if they can truly tell us what sharing behaviour is like. In addition, we don't know how varied the applicability can be, or how reliable the results are. While it is interesting that it managed to sort users into their different political standings based on the data pipeline, it is questionable if the sorting method can be used for non-binary standing.

By taking a look at social networks focusing on the subject of dieting, Yasmine Probst and Qingcai Peng [19] observed the interaction between dietetics and consumers to determine how human behaviour is influenced when engaging on social networks. In *Social media in dietetics: Insights into use and user networks*, the researchers collected the data via surveys from 340 samples. They then used social media metrics on Twitter to determine influential dietetic networks. They utilized four measures of network centrality to find them (betweenness, eigenvector, closeness and degree). The findings of the paper suggest that some of the barriers for dietitians to engage higher level user engagement include not knowing where to start. They suggest to hire professional bodies to "upskill new users" as a way to increase engagement.

An important contribution from this research is that researchers managed to show how shared and similar topics draw more attention from each other [19]. Taking a look at how these topics develop user networks that interact with each other provides context in how it can be applied to other topics. One of the main weakness [18] with the subject was the response bias. There was

not a lot of turnout for responses of the survey in comparison, which can create some skewed results.

Aiming to understand everything that has to do with trending topics on Twitter, Tian Yang and Yilang Peng studied how much impact trending topics have on news media. In particular, they examined how trending topics can be digital gatekeepers for recommending news to users in *Importance of Trending Topics in the Gatekeeping of Social Media News Engageement: A Natural Experiment on Weibo* [20]. Most people would want to learn what type of events are happening in their vicinity or huge world events. The purpose of their research was to study the effect of digital keeping on user engagement with certain news topics. They focused on a Chinese social media site known as Weibo. They conducted an experimental case study when Weibo was shut down for a week. It was an external shut down event as a natural experiment to identify the effects of trending topics on user news engagement. They analyzed user engagement data from 36,239 posts. They found that trending topics increased user engagement with top news post and the impact of trending news is most noticeable for the least popular news outlets. Thus reducing the inequality among all news outlets (otherwise known as intra-outlet gatekeeping).

One of the greatest strengths [20] is the insight on the power of trending topics that is fairly understudied. It provides insights on how trending topics can increase user engagement level with posts. Trending topics are meaningful and can increase online social activism. It also gives us insight into how digital gatekeeping winds up working for people, when they want a specific topic to appear. It is certainly something people do not consider in their daily lives with how technology gets manipulated around us. The article explains that if a user utilized a trending topic, it can increase user engagement on their own posts.

The obvious weakness of [20] the research is its unique location and timing which is hard to replicate. It is a case study focused on a natural occurrence to a non-English social media platform used mostly by Chinese users. It is hard to duplicate same research. Nevertheless, trending topics are potentially strong influential factor for user engagement.

In *Stages of user engagement on Social Commerce Platforms: Analysis with the Navigational Clickstream Data,* Kumar, A., et al, [21] asked questions regarding social media engagement. In particular, what sort of effect does a users' social and commercial activity affect user engagement on a social commerce platform? They formed a conceptual framework to comprise the conceptualization, the definition of their performance metrics and then the factors affecting those metrics. They also utilized navigational clickstream data in order to keep note about where users were going. By collecting social activities, commercial activities, and giving definitions to user online activity, user characteristics, and situational factors, Kumar, A. et al determined that rank and distance were good predictors for user engagement and had a higher impact on decisions.

A big strength for this article [21] is the use of navigational clickstream data in the study. Navigational clickstream data is important for research in the context of social media. However, it also provides situational variables and a close look at the effect that these variables have. With this data, it allows us to see what action users take. Marketing utilizes this tool a lot in order to figure out what sort of items users are buying. Along with the kind of related products they decide to buy. A good example would be Amazon having offers of items that are "frequently bought together".

The weakness of the research [21] is its narrow focus on online shopping behaviour. In addition, the testing process is technical that may not be replicated easily. Finally, the model excluded other variables and fields that may affect the reliability and validity of the study.

Other case studies of activism [33] [34] in different geographic locations asked if social media offers potential motivator to call for actions to combat social injustice. Researchers found anger is a powerful motivator for prosocial engagement behaviour. Interestingly, they found unexpected that racial differences in perceptions of whether social media content involved racial discrimination did not translate into racial differences with respect to anger after exposure to such content.

2.2 Comparative Tables

Based on the above review, it is evident that user engagement is complex and multidimensional. There are also many challenges to measure level of user engagement. Although all research papers reviewed above have contributed to the understanding of the overall concept of user engagement, some are more relevant to the user to user engagement, which is the focus of this thesis. In this section, we summarized the most relevant concepts, measurement, and prediction models related to the purpose of this study on user to user engagement. The tables provide a glimpse of the current research landscape on this subject and highlight research gaps that need to be further explored.

Table 1 - Comparison of conceptualization of "user engagement"

Criteria	Research Paper [1][2]	Research Paper [3][5]	Research Paper [4]	Research Paper [6]
Research Focus	Define user engagement and systematic review of the psychometric properties of various scales for measuring user engagement on various social media sites	Define the cognitive, affective and behavioral aspects of user engagement from a qualitative approach.	Impact of content characteristics and emotion on user engagement	Define user engagement with cognitive input and out approach
Methodology	Empirical research and meta-analysis	Survey research and using Likert scale or analysis of user	As a mega study, they searched for the most recent	Cognitive science theories

		perception to understand predicting factors.	research papers that focused on the impact of content characteristics.	
Results	Found 14 scales to analyze user engagement from psycho and social aspects on social media sites.	Promoting engagement strategy has to consider the cognitive-affective-behavioral aspects of engagement.	Identified appeal categories that have positive effect on user engagement, including providing info, providing entertainment and transactional.	Identified using cognitive network theories to improve understanding of user engagement
Strength	Working definition of "user engagement"; highlighted psychological outcomes of user engagement on social platform. Demonstrate strength and weakness of psychometric properties to measure engagement.	Provides more comprehensive understanding of user engagement instead of forcing a correlation.	Provides insights on users on both ends of the information flow (user to user) Identified the mediating effect of emotional responses in the content-engagement relationship.	Provides direction on how to use cognitive behaviour model combined with machine learning tools in reconstructing how users semantically and emotionally frame events with contextual knowledge unavailable to machine learning.
Weaknesses	Measurement scales used cannot be applied to other platforms. Typology of users are science focused. Therefore prediction ability is limited.	Small data size and support of quantification.	Definition of engagement is narrow focussing on "interaction with content" only. Looked at emotional response only.	Data integration and individual variability.
Data Features & Sources	Big data with 1.3 million unique tweets. 45 studies with no unique variables.	Surveys including interviews of CEOs from big companies around the globe.	Sample of 45 publications	No dataset is used

Table 2 - Comparison of assessing the role of emerging features such as meme, image, hashtag in affecting user engagement

Criteria	Research Paper [7][10][11]	Research Paper [9][12][16]	Research paper [14][20]	Research Paper [19][9]
Research Focus	Cognitive approaches to investigate the role of languages, words, and emotions, and biases in influencing user engagement levels;	The use of "meme" [12] and "image" [9] and "hashtag" [16] to influence user engagement.	Understand how content design and trending topics affect user interact with different social platforms	Study user profile, personality and motivation to predict user behaviour – sharing fake news.
Method	Machine learning Markov chain 2	Google Cloud Vision API to extract image data. Developed model such as BVZINB to exam how meme/image affects likes and retweets.	Conduct a text-based topic analysis. Build prediction models for external posting behaviours.	Survey – Richter scale Twitter data to establish social media metrics that can be used to calculate and analyze social media interactions on Twitter.
Results	Word usage, language and emotion have positive impact on increasing level of user engagement.	Found motivations hedonic aspects of memes and pictures that motivate user engagement.	Found the most popular topic is "business". Predictive model has 85% precision low error rate.	Certain personalities are more likely to share fake news; and altruism was the biggest motivator to share fake news.
Strengths	Demonstrating promising predictive models for wide variety of social medial engagement behaviours.	Ground breaking by looking into the emotional responses of users to memes; effective prediction model.	Novel research on multi organizations; and present the analysis within a four-level framework of user engagement.	Provides prediction models to investigate the role of personalities in motivating sharing behaviour.
Weak-ness	Limited cognitive features were applied in modeling.	Exclusion of certain users and restricted analysis may underestimate the impact of memes and image.	Not all content features are considered such as images time and location	Omission of many other factors in testing and model including, peer influence, biases and user credibility.
Data Features	Random sample of Twitter users (n=17640) [10]	Contents characters for Twitter dataset	Big data n=3,163,373 posts from 53 news orgs	Datasets include news content and social context information.

		include for Instagram has text and image	across 5 platforms collected over an 8 month period.	

Table 3 - Various forms of user engagement examined in case study scenarios and micro community

Criteria	Research Paper [19][29]	Research Paper [32]	Research Paper [33][34]
Research Focus	Explore user engagement in the form of micro community on Twitter clusters such as dietetics [19] and education institutions [29] on Twitter.	Explore how the micro community – activism disengage or engage due to algorithms.	Case studies of micro community of online activism – explore how online activism stay visible on social media platforms that are dominated by commercial user communities.
Method	SPSS statistics software to analyze the data; Modelling: multiple linear regression model	Natural experiment design with engagement methods for news post as dependent variable Multi-level models	Ethnographically inspired qualitative and quantitative technics including interviews, site visits, with content analysis
Results	Found user engagement is limited due to narrow focused topics or interests in the setting of micro community.	It is difficult to differentiate between algorithm filters and actual gatekeeping	Commercial user engagement strategies such as using hashtag, image could increase user engagement with online activism.
Strengths	Provides potential promising platforms such as an SNS platform application performing interface (API) for collecting publicly available objective records on user activity.	Provides a formal definition of engagement in the context of marketing	Fill gaps of studying user engagement in the context of activism, and studying hashtag which is iconic feature of Twitter.
Weaknesses	Sample population is relatively limited to a smaller number which may skewed the results and limit the application to another community.	Definition of engagement is intended to apply specifically to marketing. Therefore, it may not be applicable for other setting.	Qualitative in nature. Did not elaborate on the design of the platform such as YouTube which is oriented to commercial activities and user entertainment, therefore limited for "online activism" regardless of strategies used to engage users.
Data Features	On line survey n=340	Information collected from the interviews	Qualitative analysis

			YouTube channel as dataset which is challenging to clean-up.

2.3 Gap Analysis

The results of the above literature review point to some gaps and challenges in conducting research on user to user engagement especially the investigation of the triangle relationship: thoughts to emotion to behaviours, repeat. For example, there is a lack of research on the type of interaction between a user as a content creator and a user as the receiver of the content [4]. These gaps illustrate the complexity of user-to-user engagement, especially the multi-dimensional factors that influence the level of engagement, from psychosocial, cognitive, and to the ambiguity of language. The research gaps point to the area where future research needs to focus on; in particular, the area that involves data analytics tools.

The existing research predominantly focuses on user engagement between the institution and general users with the goal to figure out how it impact the company's brand and sales. As a result, content characteristics are mostly about topics and functions for sales and marketing strategies [1] [2] [4]. Most studies, including those on the effect of new features, memes, hashtags, pictures, and video links, evaluate the use within the context of brand, marketing and sales. There is little insight on the content characteristics created by users/senders who are not representing a brand or company. On the receiving end, variables that can influence their likelihood of certain behaviours are often vast. Without some refining and scoping, it could be affecting credibility of the results. In addition, motivation factors to engage with non-marketing content are diverse and require a deeper dive to the "heart and soul" of users [11]. Researching the factors outside the marketing context has the potential to reveal more of the heart and soul type of factors.

In current research, there is a lack of research for a comprehensive theoretical framework applied to the understanding of behaviours on social media. While the conceptualization in existing literature acknowledged the multidimensional aspects of user engagement, the cognitive connection of "thoughts, feelings and behaviour" has not yet been applied to the digital media studies. The cognitive behaviour framework has inspired several applications [26] [27] [28] to investigate the emotional effects of content characteristics on engagement behaviour. As a recent study [6] pointed out, social media usage is creating massive datasets of digitalized human cognition such as their thoughts, emotions and behaviours. Some researchers adopted the stimulus response theory (S-O-R theory) which is an input and output theory to investigate the factors affecting user engagement. The S-O-R theory is similar to cognitive – feelings – behaviour which recognized the connection between thoughts, emotions and behaviours. A cognitive behaviour approach could also improve predictive power of machine learning models by detecting emotional responses to content characteristics, and predict potential user engagement either through liking or sharing. Currently, machine learning tools are not able to detect emotional reaction accurately due to the fuzzy nature of human language. For example, "sarcasm" is often used in tweets, and is often counted as a positive sentiment while it might actually mean to be negative. When limited words are available in lexicon used, the sentiment might be misclassified [4][24]. In the research [7] with chat bots, researchers found machine learning is not sophisticated to define positive and negative sentiments, therefore it could affect how we determine the role of sentiment (positive vs negative) in motivating user engagement.

Lastly, the most difficult gap for the existing research is the lack of datasets that are freely accessible for user to user engagement research, especially under the lens of thoughts—emotions—behaviours. Many researches in our literature review adapted to the lack of database

by traditional surveys and interviews, since those methods can gather more about what the user is thinking. However, those can still have their own problems. To the public eye, a tweet's popularity can be determined by its likes and retweets count. Replies can be turned off at the discretion of the original user. This information could contain more answers about engagement and what motivates it, but they are private to the original user. As a result, this research could not reach that level of detail and can only focus on likes and retweets.

Chapter 3 – Case Study Evaluation

Chapter 3 presents the framework followed by the case study evaluation, as well as the individual results from applying feature engineering codes on the tweets of each of the CEOs.

3.1 The Database

Given the problem with dataset availability for research, especially for ones that focus on content characteristics, our first course of action is to build a database for this research. This data base is established for the purpose of studying user-to-user engagement. For our database, we chose a group of 10 current and former CEOs of high tech firms in the US and Canada. The accounts are not directly associated with their respective firms, companies or brands therefore excluding group factor despite tweets that would discuss business ventures if they still work with them.

The decision to choose tech CEOs tweets as our study subject is based on the assumption that their text are more suitable for the study of data science and digital media together. They are both influencers on digital media platforms as well as influential users of various media platforms. Therefore, we believe that the tech CEOs and their large group of followers provide the best context to study user-to-user engagement.

The content of the tweets are the individual thoughts and perceptions. Their tweets are freely available and open for research. Examining this small group of highly engaged users is also effective getting better understanding what factors helped them to get the level of engagement.

The following table (Table 4) gives a basic account characteristics and demographic information. Due to how famous these people are, it is no surprise that they have large number of followers.

Table 4 - Selected CEOs & Basic Information

CEOs	Name of the Company	Number of Followers	# of Tweets Collected
Bill Gates	Microsoft	54,017,221	100
Elon Musk	Tesla/SpaceX	48,624,787	100
Jeff Bezos	Amazon	2,056,373	99
Corie Barry	Best Buy	3,285	100
Ginny Rometty	IBM	34,615	39
Lisa Su	AMD	93,142	100
Sundar Pichai	Google & Alphabet Inc.	3,569,192	100
Susan Wojcicki	YouTube	308,305	100
Tim Cook	Apple	12,680,857	100
Whitney Wolfe-Herd	Bumble	29,629	100

Given the gap in understanding the content characteristics, this study applies a layered approach starting with a "case study" model as the first step. The case study model was developed for the purpose of understanding the content characteristics of the tweets of the CEOs. We hypothesised that some relationship exist between characteristics such as sentiment, subjectivity, part of speech, time of day, frequency of user engagement behaviour. This is consistent with what we observed in existing studies [1] [4] [23] where researchers drew their inspiration from social psychology theory seeking to reflect a dynamic of changes in people as they form and share their thoughts.

3.2 Case Study Framework

The case study evaluation is based on data analysis through feature engineering to detect the characteristics of the content from the CEOs tweets they created. For our code, we utilized natural language processing and TextBlob as the package would allows us to detect and get values for factors such as sentiment analysis, sentiment polarity and subjective affectivity.

Sentiment Polarity. The polarity of a piece of text measures the extent to which the overall text's sentiment is negative or positive. Using the TextBlob package, the polarity of a text varies between -1 and 1; negative measures indicate negative sentiments, 0 indicates neutral sentiments, and positive measures indicate positive sentiments.

Subjectivity. The subjectivity of a piece of text is the extent to which it expresses an opinion. Subjectivity varies from 0 (neutral opinion) to 1 (strongest opinion).

Word Frequency Count. The number of words in a given tweet might also influence how it is shared; thus, we use word cloud to count this variable in our feature engineering.

Pattern of Sentences: Researchers have found well-structured sentences such as those with appropriate use of nouns, verbs and adjectives could affect how thoughts were expressed and may impact on how users interact with them [8][10][36].

Time Frequency: We assume the frequency of tweeting affects the level of engagement. The more frequency the higher of engagement.

Hour of Day: We surmised that the time of day when a tweet is posted might also have a bearing on its popularity.

Tweet Classification and User Mentions: We assume level of engagement is affected by the type of tweets, original, retweets.

To write our code, we used Jupyter Notebook to apply a tweepy API to scrape tweets from the Twitter accounts. For each CEO, we tried to collect about 100 tweets, although there were a few exceptions due to the nature of their accounts at the time, being Jeff Bezos and Ginni Rometty. The time period that the tweets covered were from February 2018 to March 8, 2021. After using tweepy to extract the tweets that we needed, data cleaning was performed on the tweets. Like all

other cases, data captured contain noisy and irrelevant content values. To prepare the data for processing, we performed the following procedures:

- removed symbols that frequently appear in the sentences including (), [].
- removed stop words in sentences and phrases
- removed extra rows and extra words

Evaluations from each of the individual case studies are presented below.

3.3 Case Study 1 – Bill Gates, CEO of Microsoft

A total of 100 tweets were collected from Bill Gates Twitter account. As the CEO of Microsoft, Bill Gates has roughly 50 million followers according to a November 2021 tracking analysis [61]. Our case study feature engineering unveiled the following characteristics.

3.3.1 Sentiment Polarity Analysis

The average sentiment result for Bill Gates' tweets in our data sample primarily lean to the positive. The overall sentiment for Bill Gates show that 77% of the tweets are positive, 12% are neutral and 11% are negative. It is interesting to see the negative and neutral are almost equivalent although their numbers are small compare to the positive ones. The positive sentiment appears to relate to Bill Gates frequent posts of trending topics such as major health events during the time period we examined. This is consistent to previous studies [20][35] which found trending topics leads to positive sentiment which then leads to higher levels of engagement.

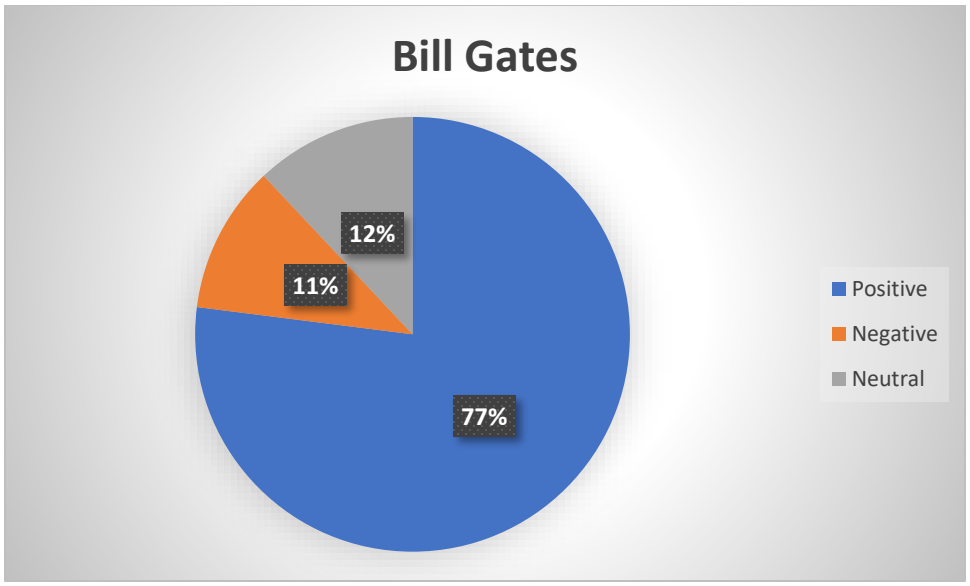

Figure 1 - Sentiment Analysis (Bill Gates)

When looking at the times of day Bill Gate posts, it appears he tends to posts more in the morning and afternoon, than he does in the evening. There is no shift of sentiment either twitting in the morning or afternoon or evening.

3.3.2 Subjective Affectivity Analysis

Based on the table and the scores that we have collected, the average of Bill Gates' subjective affectivity scores is 0.442**.**

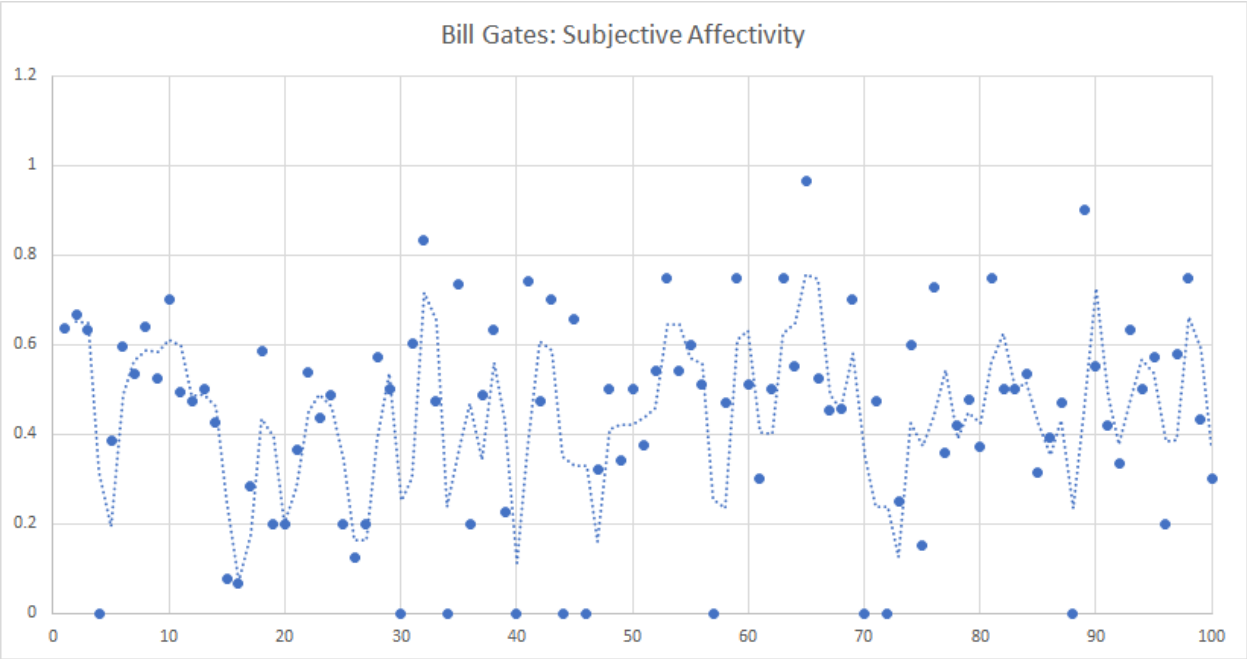

Figure 2 - Subjectivity Analysis (Bill Gates)

On average, the subjective affectivity is less than 1 (SA<1). The only tweet that comes the closest has the score of about 0.996. The tweet is about recommending a book, which would certainly have a very high subjective affectivity score. For those with 0 scores (SA = 0), they tend to be images that were posted without a caption. It is also noted tweets with adjectives are more subjective in comparison to those that have none.

3.3.3 Time Frequency

Bill Gates is known for his tremendous work in supporting vaccine development to fight COVID–19. Figure 3 indicates an increasing frequency of tweets in the period of October 27 2020 to March 8 2021 when vaccines against Covid-19 were advancing at a rapid speed. One tweet in particular praises India's government for being able to bring its scientists together to create a vaccine to fight COVID. This was the tweet that has the most likes and retweets among the 100 tweets we collected from Bill Gates.

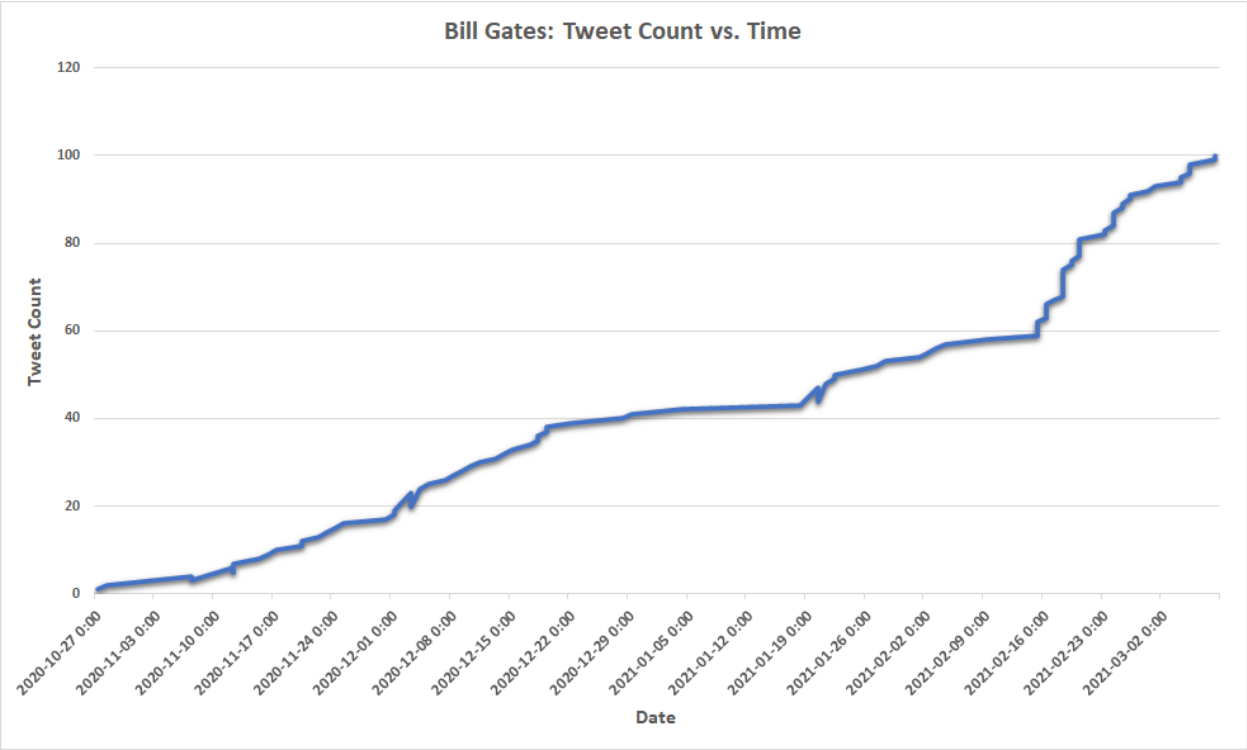

Figure 3 - Tweet Frequency by Time (Bill Gates)

The amount that he starts to post increases in the most recent months, also most likely related to the progress in advancing COVID-19 vaccine deployment around the world.

3.3.4 Tweet Classification Analysis

Most of the tweets that Bill Gates posts onto his Twitter are original. However, he does frequently retweet and sometimes comment on posts particularly with stories that are inspiring. Whenever Bill Gates retweets a post, it also tends to involve raising awareness whenever he appears for an interview.

3.3.5 Frequency of Tweets Analysis

Similarly to the time periods, the frequency of Bill Gates' tweets varies greatly. He posts fairly frequently, particularly when compared to some of the other CEOs that we had collected tweets from.

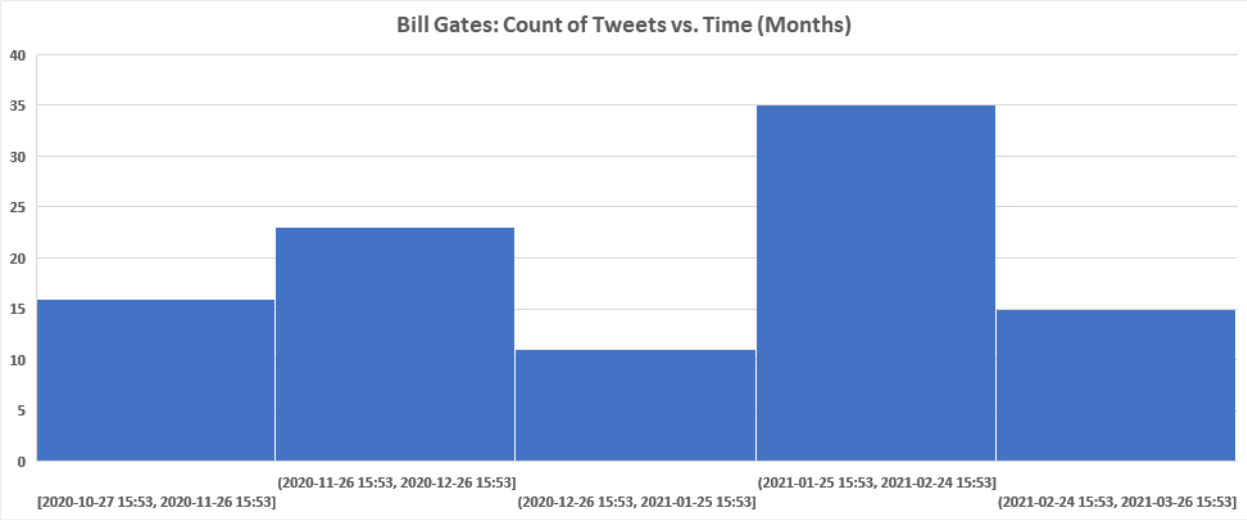

Figure 4 - Count of Tweets vs Time (Months) (Bill Gates)

As shown in Figure 4, the highest frequency time period when Bill Gates posted a lot, was between the time periods of January 25 2021 to February 24, 2021. This was during the time when Joe Biden was elected as the President of the United States.

Figure 5 - Bill Gates tweet on Jan 20, 2021 after Joe Biden won the presidency

3.3.6 Reaction Classification Analysis

In total, the number of interactions that tweets that Bill Gates receives is: 549,987 likes and 74,565 retweets. His most retweeted tweet has 20,494 retweets while his most liked has 87,923 likes.

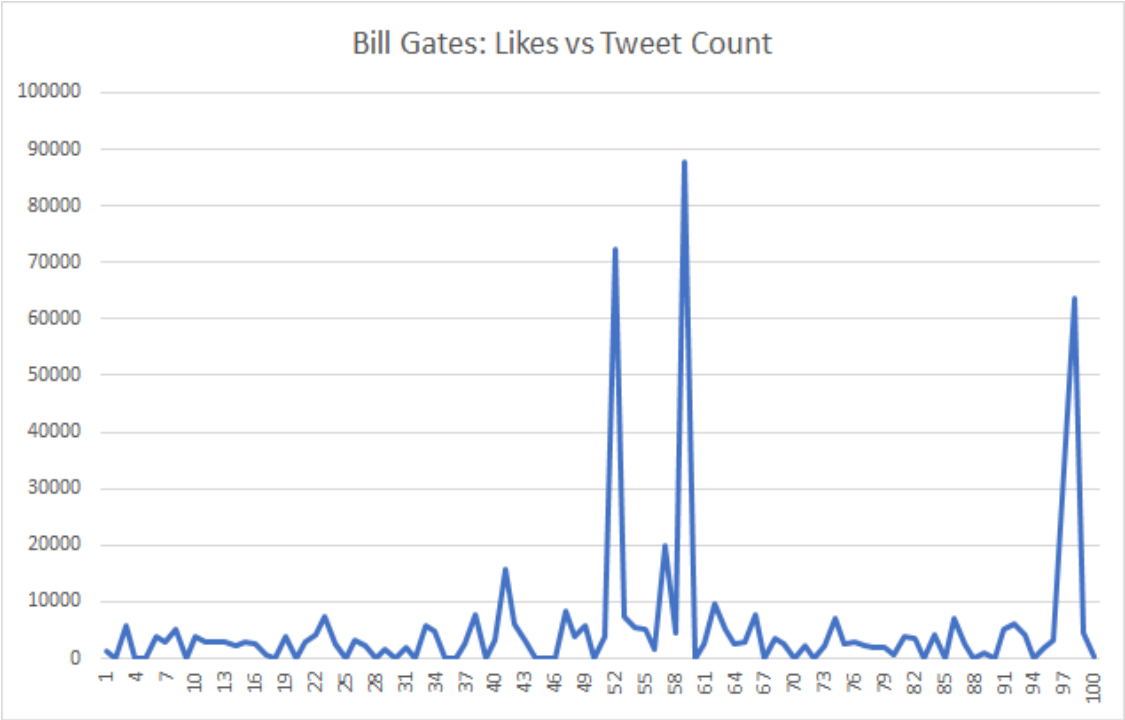

Figure 6 - Amount of Likes on Each Tweet (Bill Gates)

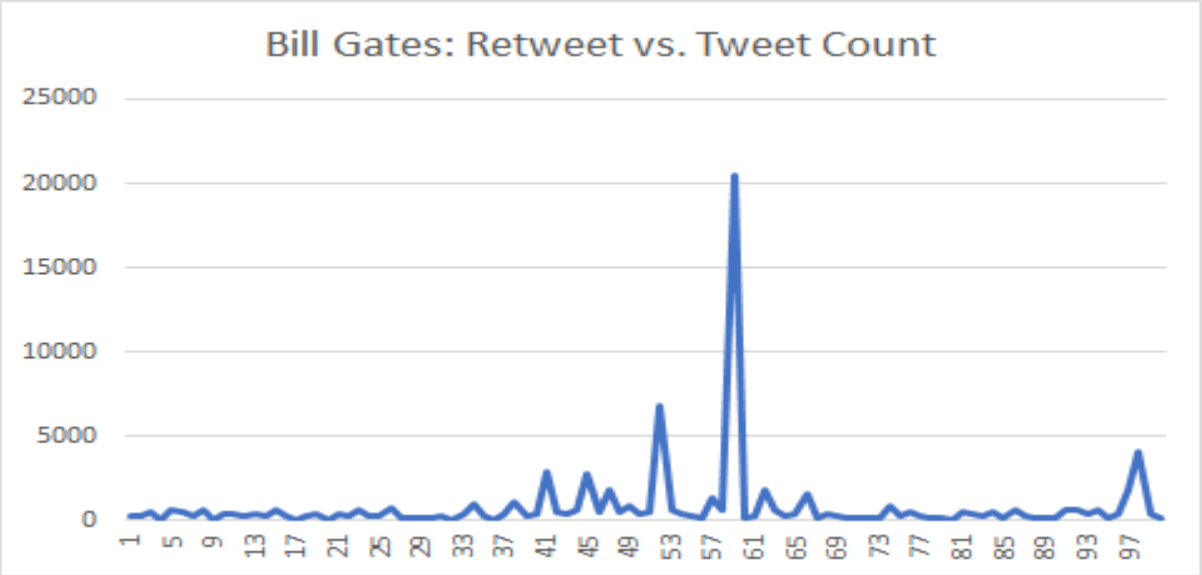

Figure 7 - Amount of Retweets on Each Tweet (Bill Gates)

These graphs also show the different spikes that come from the tweets that he posts. The tweet on Bill Gates' Twitter that is a part of our data that has the highest amount of likes and retweets is the one talking about COVID-19.

3.3.7 Word Count Analysis

The word cloud in illustration 1 shows the most frequent words among Bill Gates' tweets. They include words such as "climate change", "covid" "world", with "climate change" being his most used term.

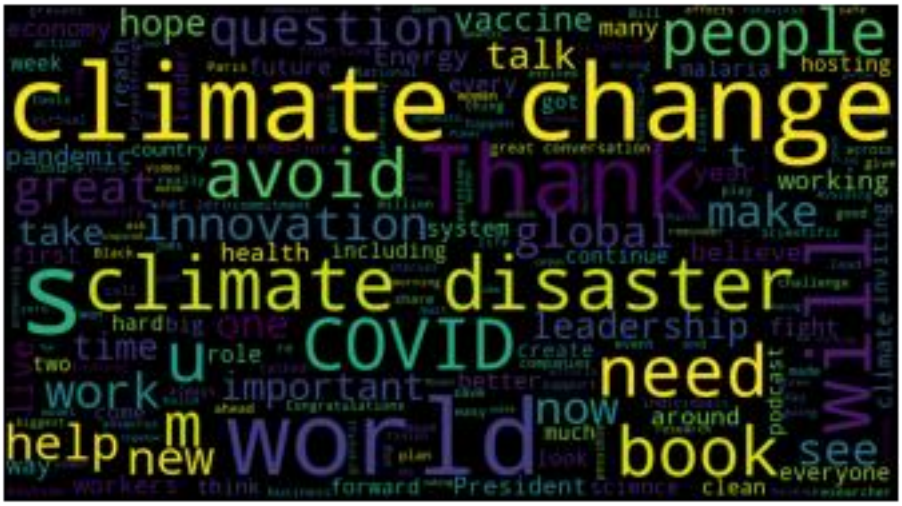

Illustration 1 - Word Cloud Analysis for Bill Gates

3.3.8 Pattern of Speech – Noun, Verb and Adjective Count Analysis

The overall amount of nouns that were used in all of Bill Gates tweets was 1068. This is a lot of nouns, but it also goes into conjunction with the numbers of verbs and adjectives. A lot of Bill Gates' tweets involve mentioning events, specific topics, and thus the result of the noun count is not surprising. The overall amount of verbs that were used in all of Bill Gates tweets was 474. Even though the number does not compare to the number of nouns within he tweets, there is an impressive amount of verbs. Bill Gates is certainly someone who puts an emphasis on more action, and it shows with the words that he uses within the verbs. The overall amount of adjectives that were used in all of Bill Gates tweets was 460.

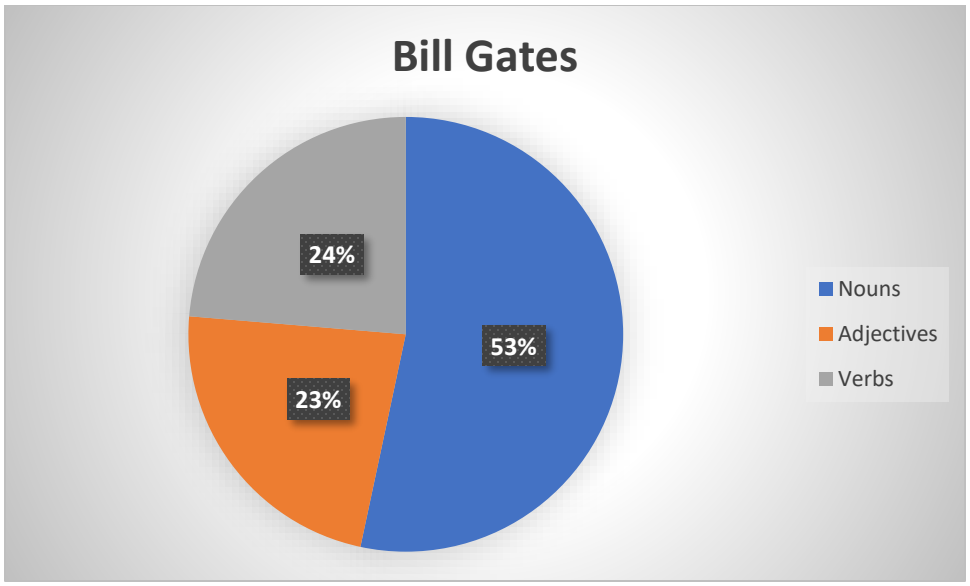

Figure 8 – Percentage of Text Consisting of Nouns, Adjectives and Verbs (Bill Gates)

3.3.9 User Mentions Frequency Analysis

In terms of user mentions, Bill Gates' tweets have overall mentioned other Twitters 7 times. However, this does not include user mentions that are in the body of the tweet itself. It also does not account for the times when Bill Gates responded to himself, adding onto the thread. If we were to count all of the user mentions that are not direct replies, there are about 20 overall user mentions.

Most of the user mentions on Bill Gates' Twitter comes from acknowledgement. This is usually in the form of thanks for interviews.

3.4 Case Study 2 – Elon Musk – CEO of Tesla Industries

This is the case study for Elon Musk, the CEO of Tesla Industries. Our feature engineering revealed the following characteristics about his tweet content based on the analysis of 100 scrapped tweets.

3.4.1 Sentiment Polarity Analysis

From a first-hand basis, it is clear that Elon Musk is an outlier. Looking at just the sentiment analysis, he has almost an even distribution of positive, negative and neutral. The overall sentiment

scores show that 44% of tweets have positive sentiments, 46% of tweets have neutral sentiments and 10% of the tweets have negative sentiments.

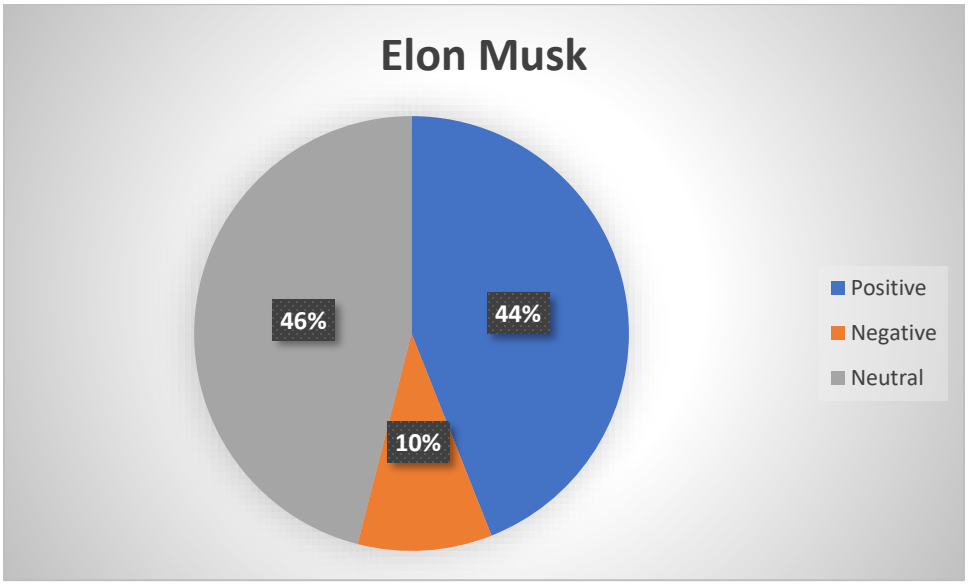

Figure 9 - Sentiment Analysis (Elon Musk)

While the skew towards neutral is not that big, it is the only time when the sentiment polarity leans more to the neutral than to the positive or the negative. However, the discrepancy between the two totals is not significant enough. In fact, one could predict that if that data gathered tweets from a few more days after what we had in our data, we could easily see more of a skew to the positive.

There is no sharp polarity or sharp change in the emotions when looking at Elon Musk's Twitter. They consistently go between positive and neutral. Although what should be noted is the extent of replies that play a big part in the stranger language he has a tendency of using, he does not use a lot of exact language, unless he is talking about something relating to his Tesla business or space travel.

3.4.2. Subjective Affectivity Analysis

The subjectivity of Elon Musk score swings wildly with plenty of "0" and "1"s. The subjectivity scores seem to be primarily lean towards the sentiment behind it. However, there are some scores that look strange. From taking a direct look at the affectivity, there are quite a few subjectivity scores that hit "1".

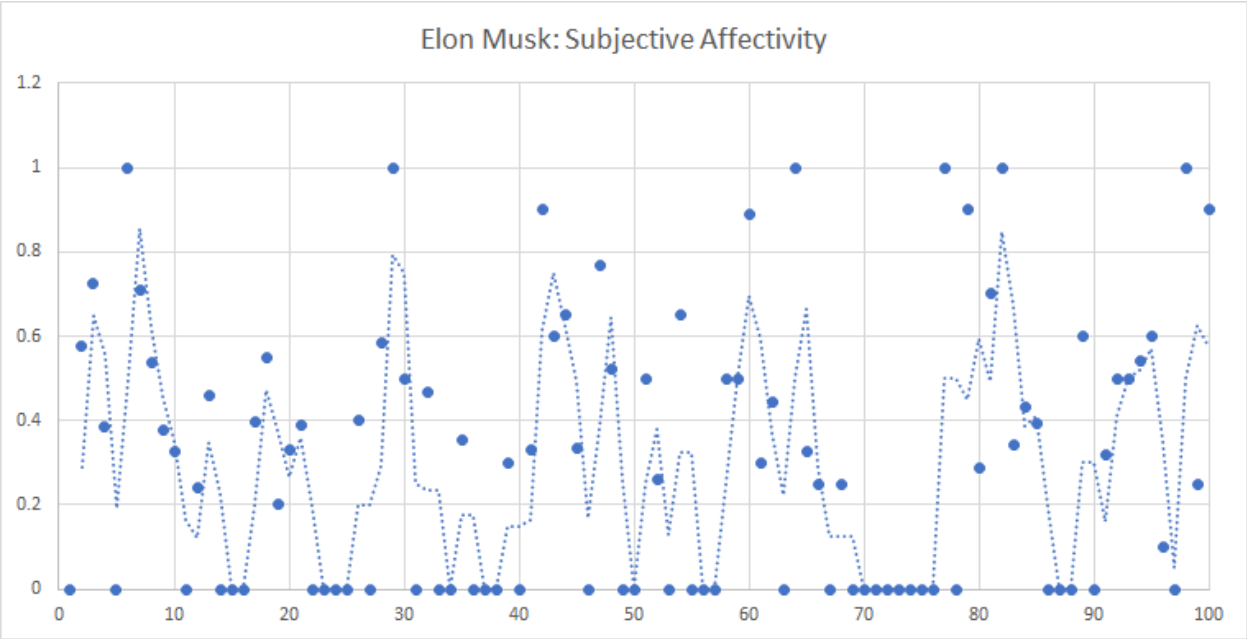

Figure 10 - Subjectivity Affectivity (Elon Musk)

It certainly shows that there is a lot of subjectivity and confusion when machine learning algorithm went through Elon Musk's tweets. Some of them are simply images without any text. Machine learning algorism was unable to determine if they are truly objective or not. However, it should be noted that his most popular tweet is also seen as his most subjective tweets.

3.4.3 Time Frequency Analysis

The posting periods that Elon Musk has is drastically different from everybody else in the sample size. The time span that the tweets are taken from, go from February 24 2021 to March 8 2021 for polarity. Elon Musk has the shortest time period (spanning for about a week) but also the highest frequency of tweets. Within these weeks, Elon Musk has posted an average of five tweets

every day. This likely explains why there are so many positive and neutral sentiments. There are quite a few tweets that are simply responses to other people. It would make sense that one would rather have a positive or neutral sentiment behind their responses. It would also explain why they are so quick and frequent, as well as why they are not as popular compared to his original tweets.

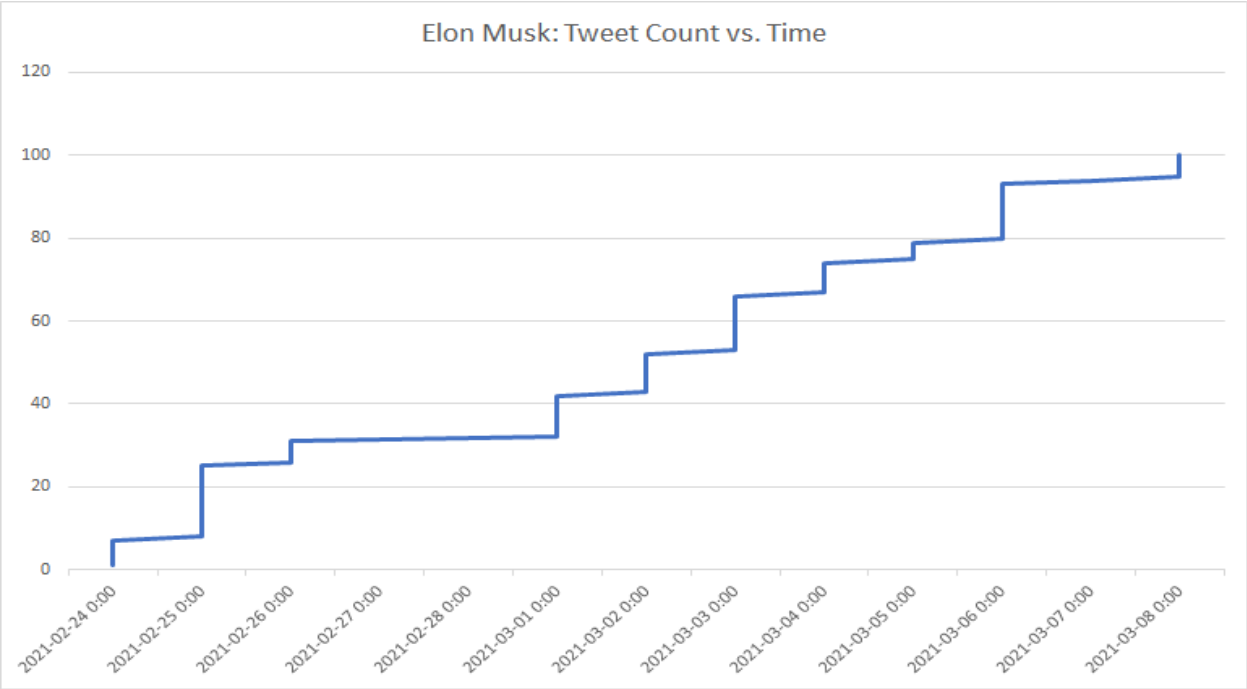

Figure 11 - Tweet Frequency by Time (Elon Musk)

Elon Musk is a good example to show the many instances where most replies and responses do not get as much interaction from users compared to original tweets. There are exceptions as sometimes a response to a tweet receives much more attention and user interaction than even the original. These cases are few and far between. Since the data does not check when other people quote retweeted Elon Musk, cannot be confirmed with the data we have for the time being.

3.4.4 Tweet Classification Analysis

Most of the tweets that Elon Musk created were original posts despite doubts that the images he shared are necessarily his own. There were not many posts that he retweeted from

elsewhere. The times when he retweets something will usually be business related, or with the cryptocurrency market.

Other than that though, Elon Musk tweet content contains a variety of features from image, meme to video links. When the code detects any tweet that shows an image but no description, the text comes up as blank. It will only share the link if the program can support the text in question.

3.4.5 Frequency of Tweets Analysis

Any indication about how often Elon Musk post is shown clear as day in the data for the tweet frequency. Since he does not have many tweets that are separated into months based on the time period, it focuses on intervals of three days age. Throughout each of them, he has a habit of posting a large amount. One could even argue that his posting frequency from March 7th to March 10th is a result from the data being cut off and not getting a chance to continue from where it was.

This is one part of the data that reads over 200 of his most frequent tweets at the time rather than 100. Although if anything, it is simply because it cannot make a graph otherwise. Elon Musk's data being a major outlier in that way makes it hard to talk about him in a sufficient way. However, as stated before, when calculating and analyzing the data itself, we only considered the first 100 most recent tweets.

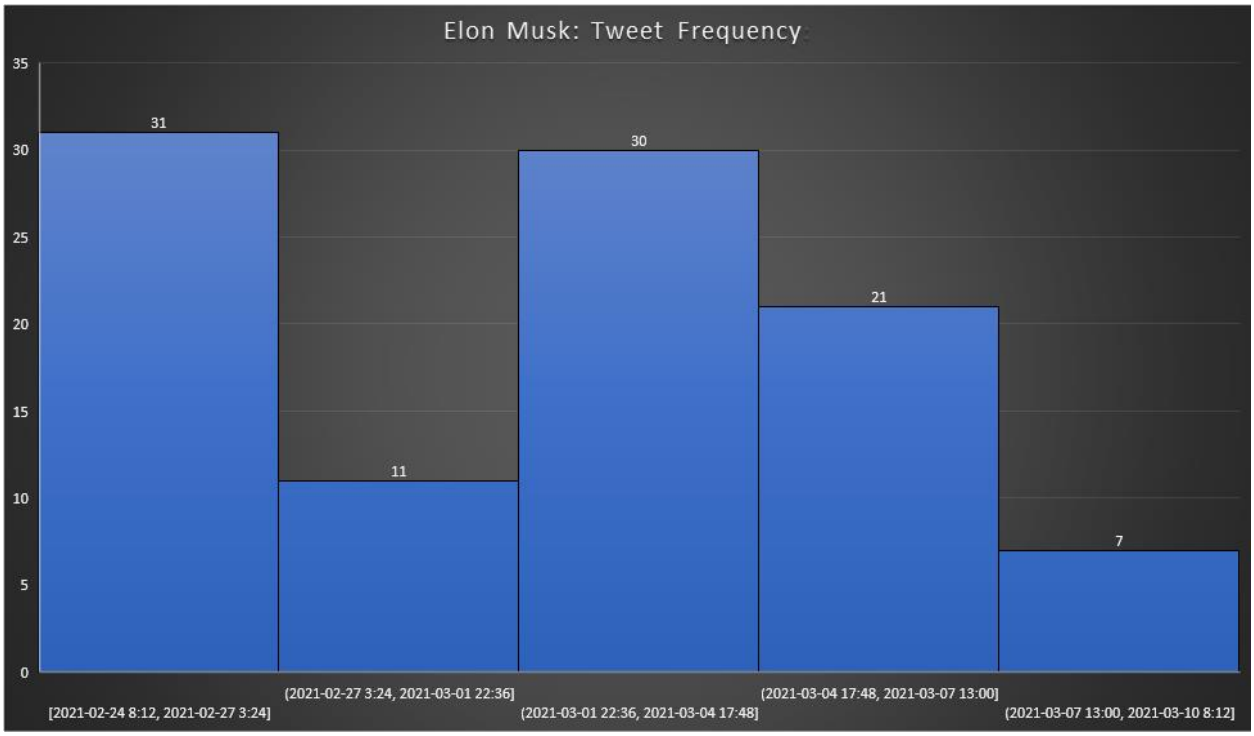

Figure 12 – Count of Tweets vs Time (Months) (Elon Musk)

3.4.6 Reaction Classification Analysis

Out of all of the interactions that Elon Musk's tweets have received, they have received 15,465,653 likes and 1,281,985 retweets. In total, all of them have gathered 16,747,638 interactions which is the highest out of all of the CEOs. His most popular tweet have 80,104 retweets and 845,223 likes.

The spikes in how many likes and retweets each tweet gets remains fairly consistent. Whenever there are instances of there being 0 likes, it tends to be towards tweets he himself retweeted. It has a strange way of working, but since those posts are not his own, those do not count his total amount of interactions.

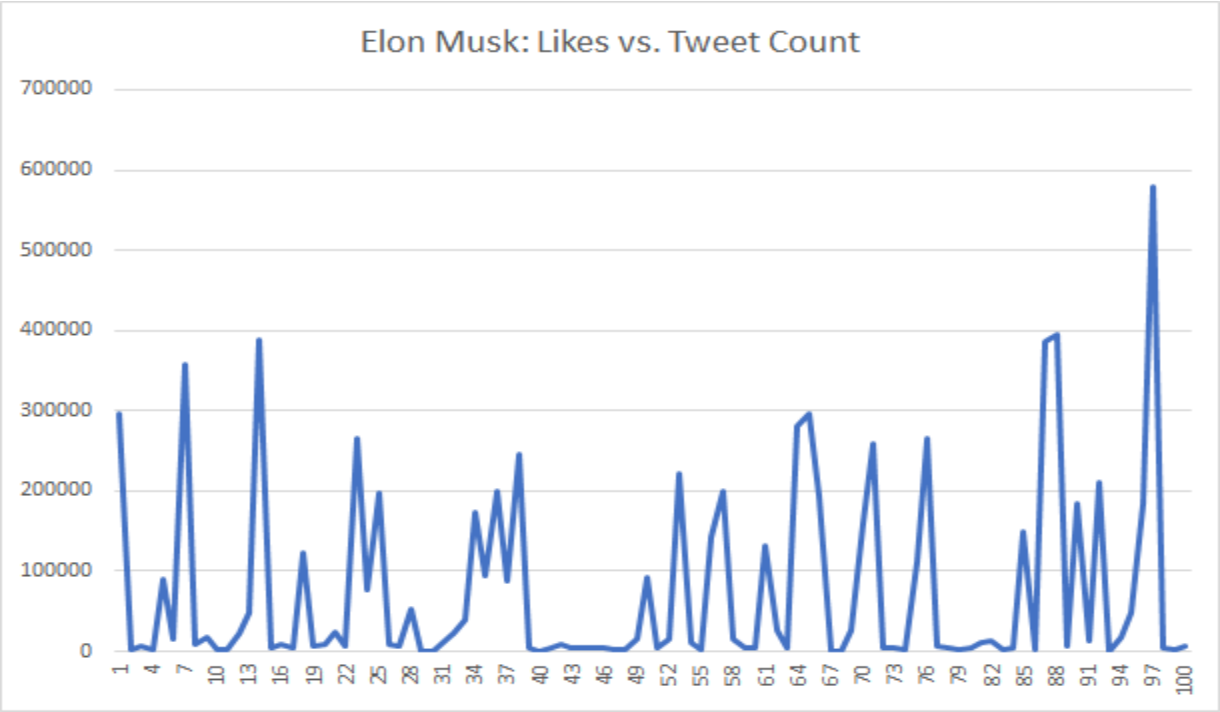

Figure 13 – Amount of Likes on Each Tweet (Elon Musk)

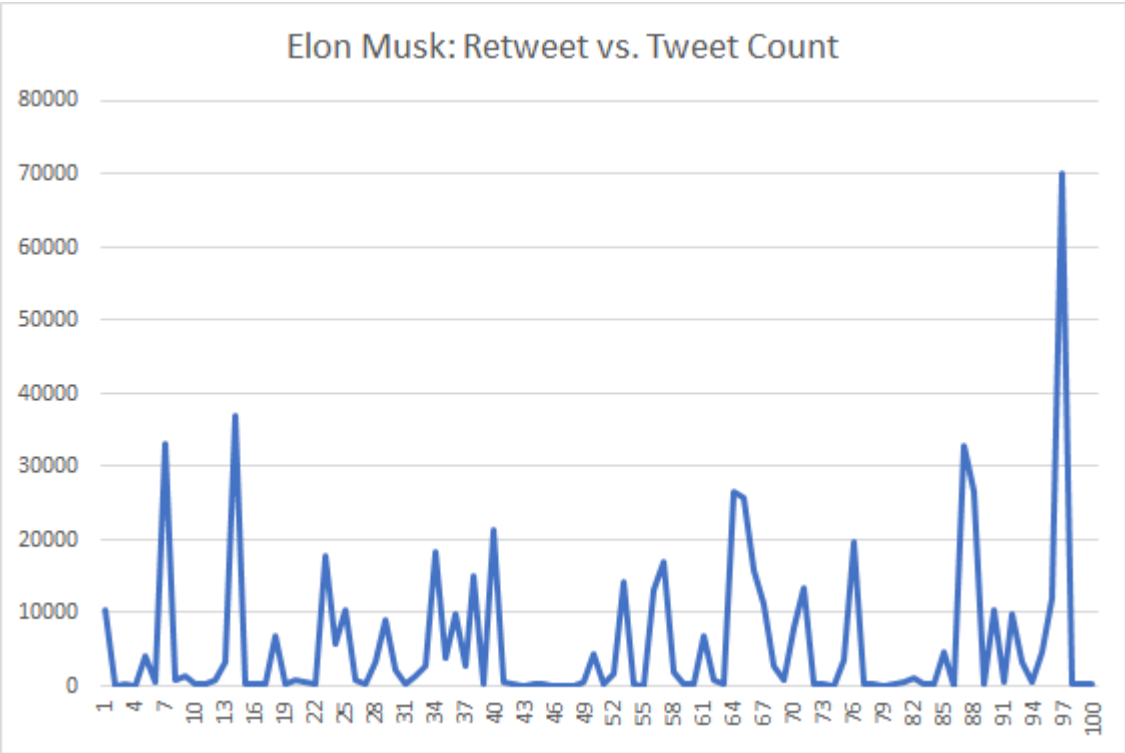

Figure 14 – Amount of Retweets on Each Tweet (Elon Musk)

From the most popular tweets that Elon Musk has, the one that has the most likes is simply an image. The image is a fairly simple “meme”. Although unlike a lot of the other memes that he posts, this one is very accessible to the mainstream. A lot of the different memes that Elon Musk posts are tangentially related to his own company or cryptocurrencies.

3.4.7 Word Frequency Analysis

From all of Elon Musk’s tweets, the most frequently appearing word is “will”. Despite this, the program itself states that “beta” is the word that appears the most. It is possible words such as “will” are common fill in the blank words that many people use. He talks to a lot of people, and also wants to inspire hope about the projects that he is working on or assisting such as space tourism and rockets. Another particularly interesting one is “Bitcoin” appearing on the Word Cloud, because he is notorious for being an advocate in investing on Bitcoin.

Illustration 2- Word Cloud Analysis for Elon Musk

The frequency of the words are not completely consistent. Although some of the words that appear on the Word Cloud show that Elon Musk is much more casual with his tweets. Since a good amount of them are replies to other people, it will include words that indicate he is making casual conversation with them.

3.4.8 Pattern of Speech: Nouns Verb and Adjective Count Analysis

The pattern of speech of Elon Musk's tweets is as follows:

- Total noun count = 89
- Total verb count = 40
- Total adjective count = 43

These numbers mean that Elon Musk's tweets are not usually very long. We noted his only long complete sentences are when he is explaining or advertising something of his own accord. As shown in the figure below, Elon Musk's tweets have a higher number of nouns. This shows that even for as strange as some of his tweets can be, there is some sort of grammatical structure. Although considering there are some tweets where there is no wording, and just an image, it could have an effect on the overall number.

From all of Elon Musk's tweets, there are only 40 verbs in the word count. It is very close to the amount of adjectives, which is at 43.

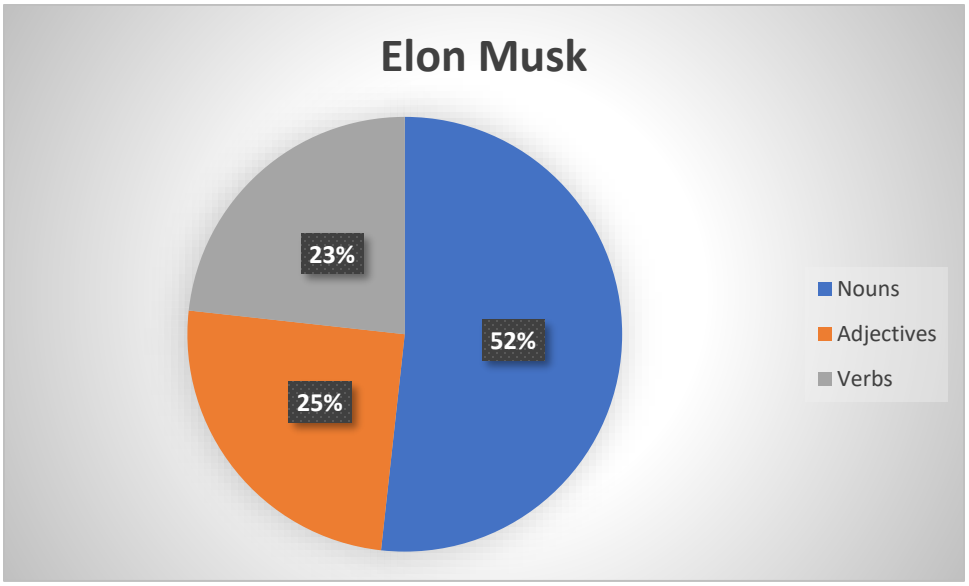

Figure 15 - Percentage of Text Consisting of Nouns, Adjectives and Verbs (Elon Musk)

3.4.9 User Mentions Frequency Analysis

Due to how many times Elon Musk has replied to people, there are a fair amount of user mentions frequency. Within all of his tweets, at least 20% of them involve mentioning someone. Whether this be in acknowledgement or in response varies, though most of them are simple replies to answer questions. Not all of these replies are serious, as they tend to relate to the cryptocurrencies that he focuses on as well.

3.5 Case Study 3 – Jeff Bezos – CEO of Amazon

This case study focuses on Jeff Bezos, the current CEO of Amazon. We selected 100 tweets from his Twitter. Our feature engineering indicated the following characteristics, sentiment and likes and retweets.

3.5.1 Sentiment Polarity Analysis

Based on the sentiment analysis alone, the tweets that Jeff Bezos has posted are fairly standard. The sentiment scores show that 75% of the tweets (74/99) have positive sentiments, 19% of the tweets (19/99) have neutral sentiments and 6% of the tweets (6/99) have negative sentiments.

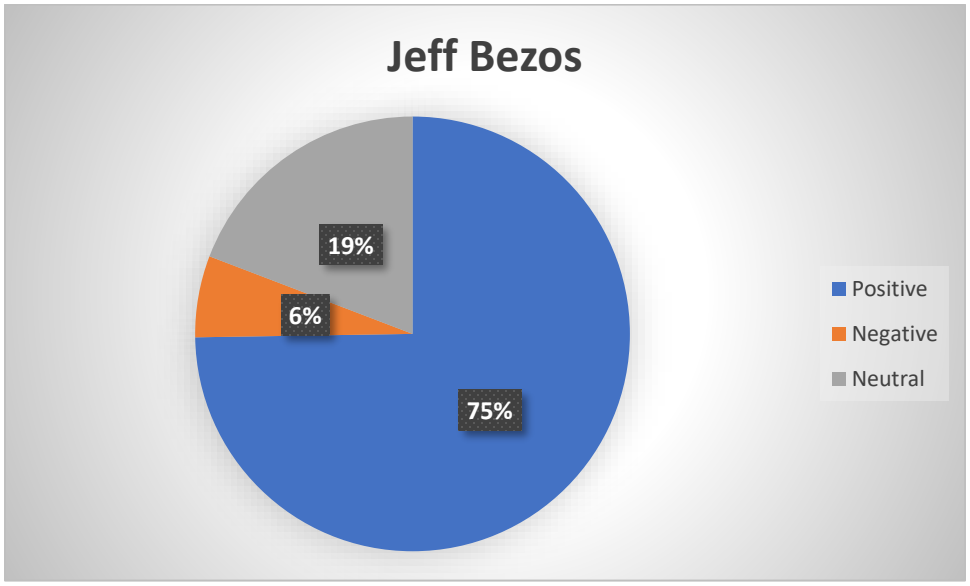

Figure 16 - Sentiment Analysis (Jeff Bezos)

3.5.2 Subjectivity Affectivity Analysis

When looking at the subjectivity scores for each of the individual Jeff Bezos tweets, a lot of his tweets are fairly subjective. They score fairly high in subjectivity, though most of his tweets reach a below average level of subjectivity to them. Only one post has a subjectivity score of "1", which means that a lot of what he says tends to have more objectiveness to them. The mean subjective affectivity score from all of the statements adds up to 0.46.

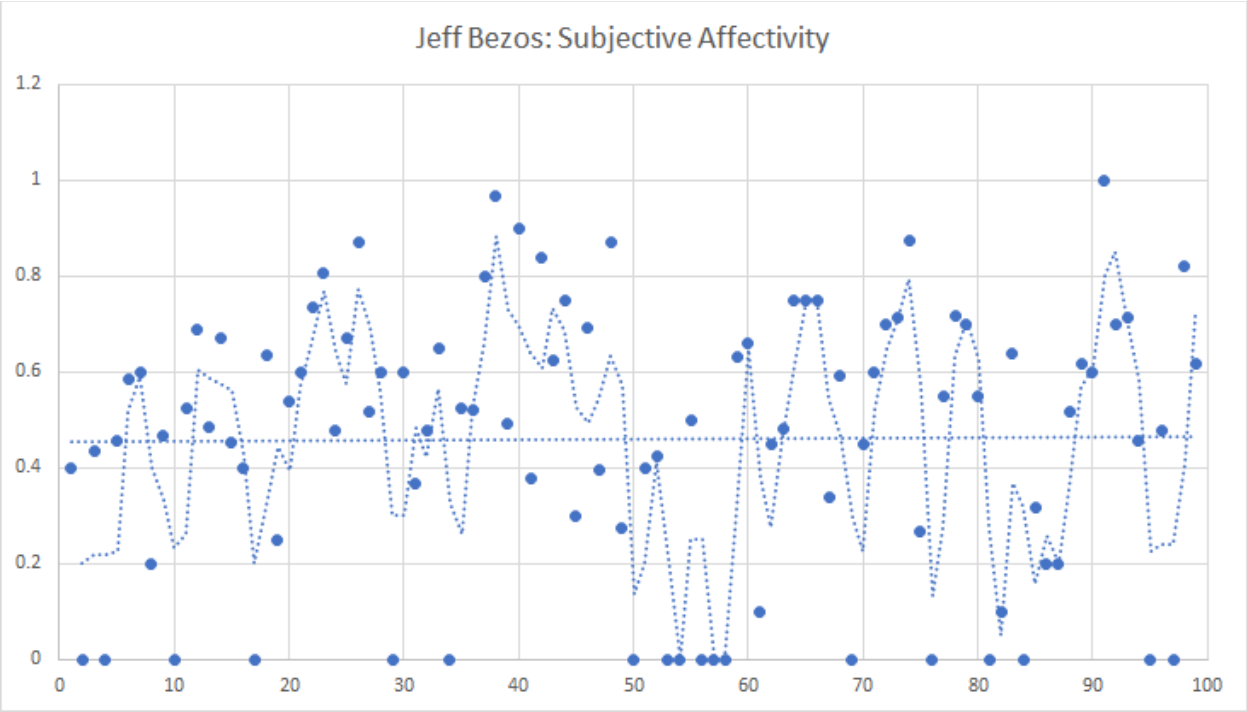

Figure 17 - Subjectivity Affectivity (Jeff Bezos)

From looking at the subjective affectivity scores, there is at least 1 post that has a subjective affectivity score of 1. The interesting thing about this tweet is that it contains a photo. However, Bezos himself is talking about how he is hanging out with a panda plush that his children have and talking about taking them somewhere later in the day. It is likely that this has a low subjectivity score because of the language used. It is hard to say that this is not an objective fact, as it does not state something that can be disputed as an "opinion".

There are several posts that have a subjectivity score of zero. However, he has more tweets that are simply photos. All of these are statements written by Bezos himself to explain specific situations, such as his and his ex-wife's divorce. The only reason that they cannot be made into tweets is because of the character limit.

3.5.3 Time Frequency Analysis

Jeff Bezos has the largest time gap making him an outlier in our selected CEOs. The time gap goes from February 20 2018 to February 28 2020. He is an outlier in terms of how infrequently he posts. He also does not have any recent tweets as opposed to the other CEOs.

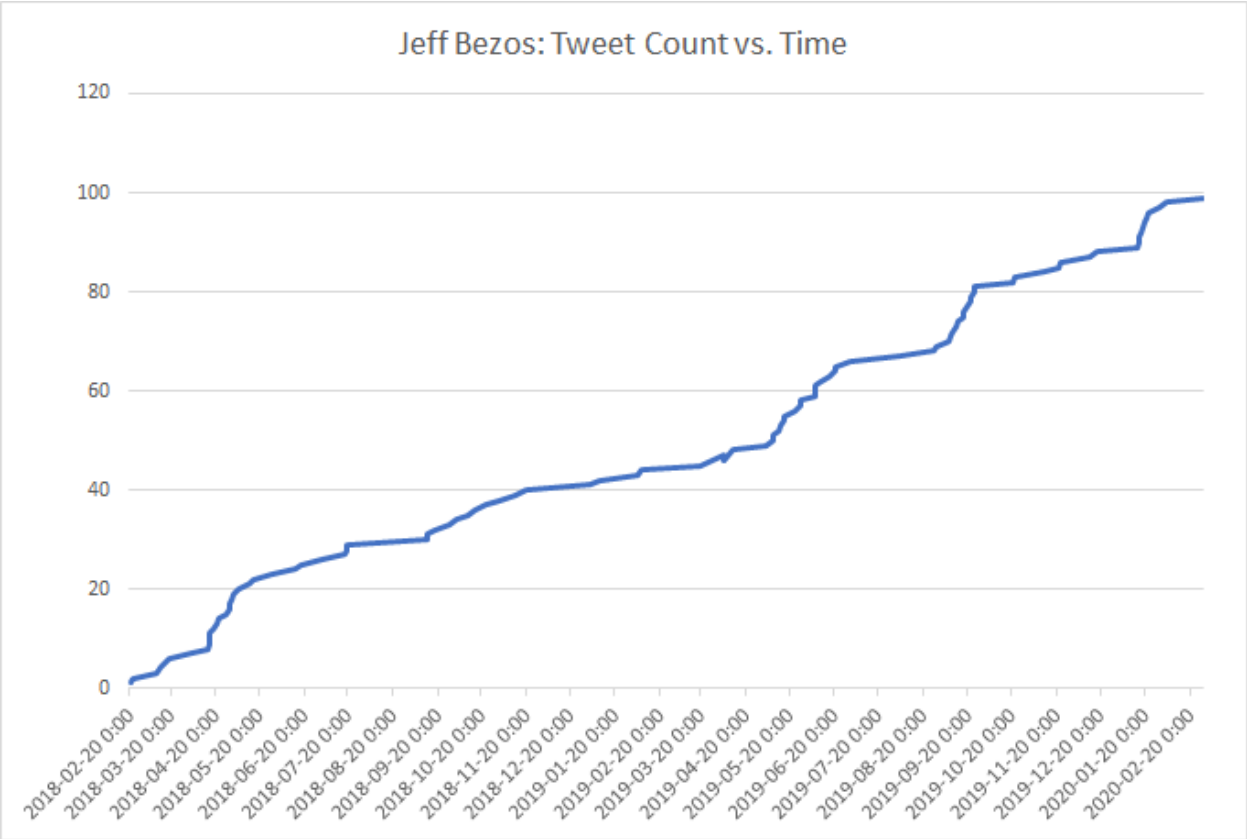

Figure 18 – Tweet Frequency by Time (Jeff Bezos)

3.5.4 Tweet Classification Analysis

Most of the tweets that Jeff Bezos posts are original tweets he made himself. There are a couple that contain videos or images. However, they have enough detail where they are not simply a single video or image.

3.5.5. Frequency of Tweets Analysis

When looking at the frequency of the tweets Jeff Bezos used to tweet fairly frequently back in 2018 and 2019. This started to wane around 2020, and by that time, he had not tweeted anything afterwards. This is no longer the case as of October 2021, but this study did not account for that time period.

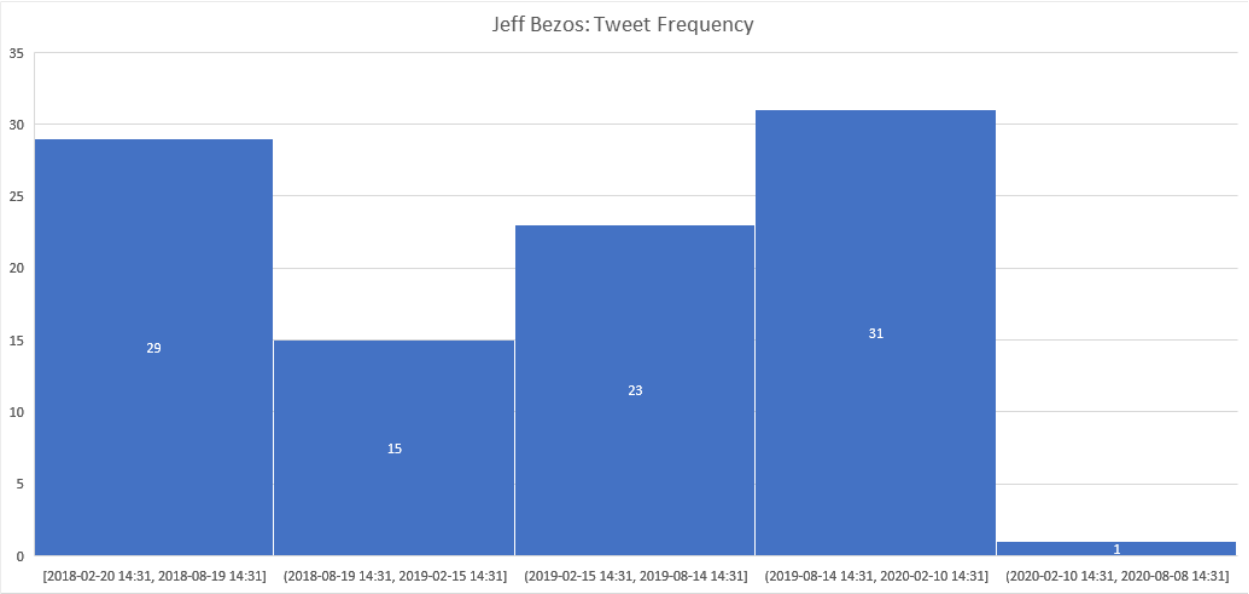

Figure 19 – Count of Tweets vs Time (Months) (Jeff Bezos)

3.5.6 Reaction Classification Analysis

In total, from Jeff Bezos' tweets have collected a grand total of 1,222,521 likes and 203,572 retweets. Out of all of his tweets, they have gathered 1,426,093 interactions in total. The overall count spikes at different times, garnering many likes for the first part of the tweeting period. It starts to wane off in the second half, most likely due to Jeff Bezos' decreasing activity on the account until he stopped using it in 2020.

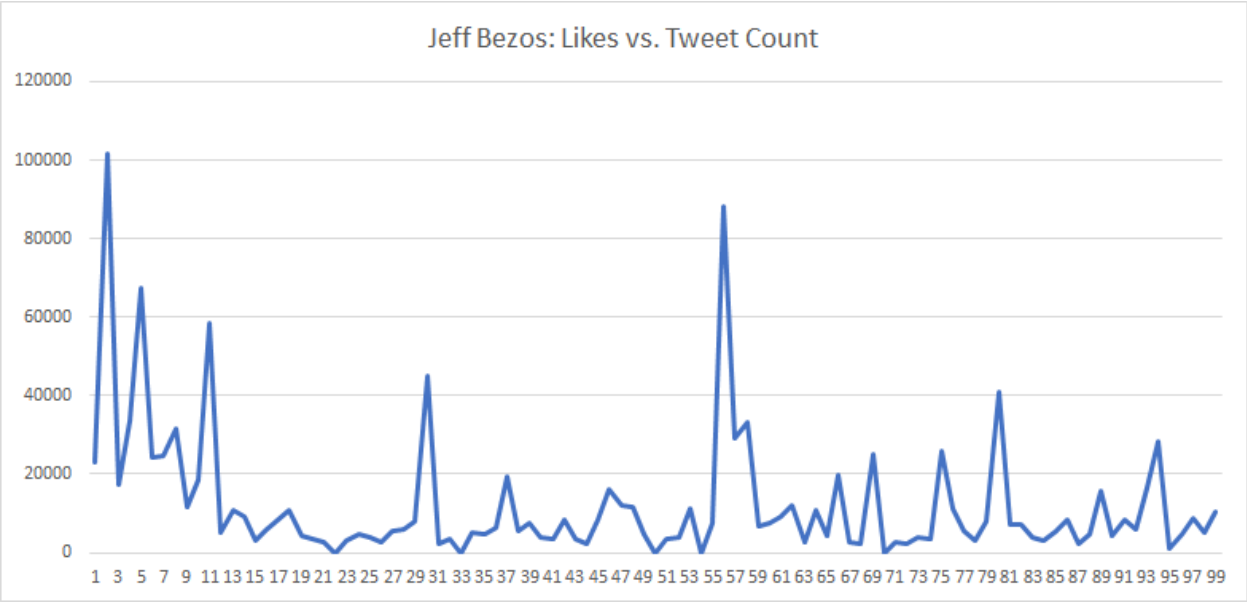

Figure 20 – Amount of Likes on Each Tweet (Jeff Bezos)

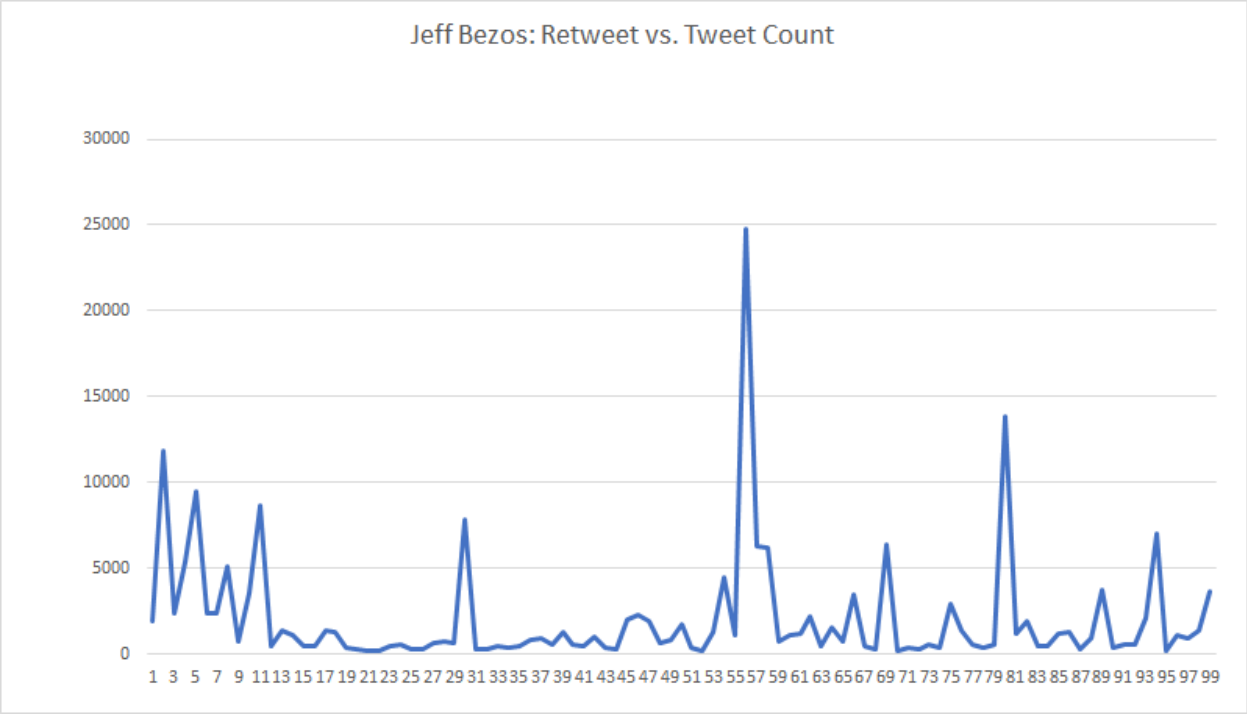

Figure 21 – Amount of Retweets on Each Tweet (Jeff Bezos)

3.5.7 Word Frequency Analysis

From the Word Cloud, it shows that the word that appears the most often on Jeff Beoz's Twitter is "Thank". Although other words that show up a lot include "GradatimFerociter" which

is a business term relating to Blue Origin, the motto of a company he owns. He also uses words like “kids” a lot, often relating his frequent references to his children in his tweets.

Illustration 3 - Word Cloud for Jeff Bezos’s Twitter

<u>3.5.8 Pattern of Speech: Noun Verb and Adjective Count Analysis</u>

Figure 24 shows the distribution of Bezos’s usage of nouns, verbs and adjective. There are a total of 659 nouns in all of Jeff Bezos’s tweets. This makes for 54% of the parts of speech in Jeff Bezos’s tweets consisting of nouns of some kind. He has decent count of 471 verbs and 439 adjectives. The pattern of speech is evenly distributed meaning he expressed his thoughts with – well-structured sentences and grammar structure.

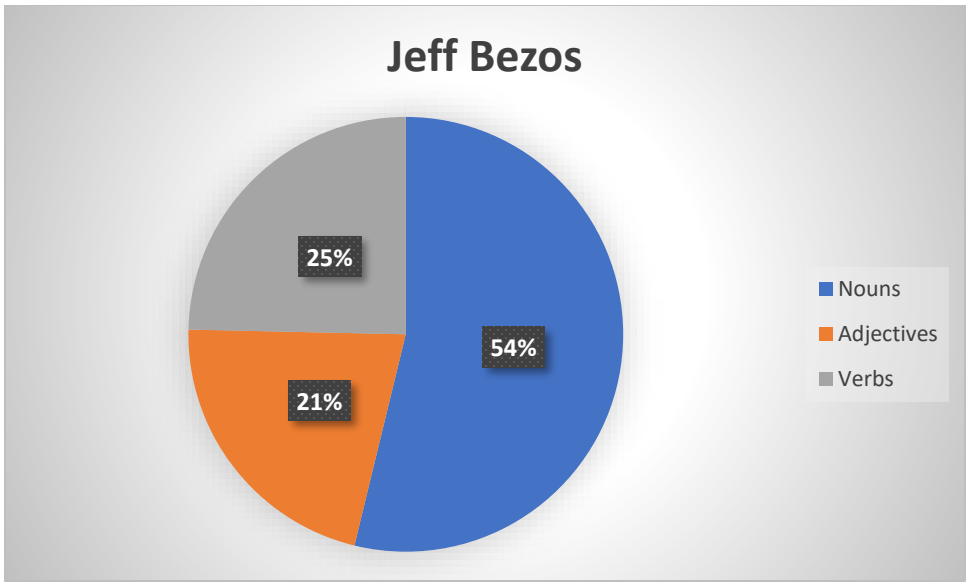

Figure 22 - Percentage of Text Consisting of Nouns, Adjectives and Verbs (Jeff Bezos)

3.5.11 User Mention Frequency Analysis

When looking at Jeff Bezos's tweets, just like any business CEO, he only mentions users who are relevant to the people he met. There are a lot of cases where the people he talks to are business partners. That, or he is congratulating someone and wanting to bring more attention to it to his followers.

3.6 Case Study 4 – Corie Barry, CEO of Best Buy

This case study focuses on Corie Barry, the CEO of Best Buy. We had collected 100 of her most recent tweets at the time. The following are the analyses results of her tweet content, sentiment and level of engagement.

3.6.1 Sentiment Polarity Analysis

The sentiment analysis scores for Corie Barry are fairly standard. From calculating the scores, we see that 74% of her tweets have positive sentiments behind them, 19% of them have neutral sentiments behind them and 7% have negative sentiments. Overall, it's a fairly average

scores. Nothing stands out in particularly on Corie Barry's Twitter. On the average, there was a polarity score of 0.388.

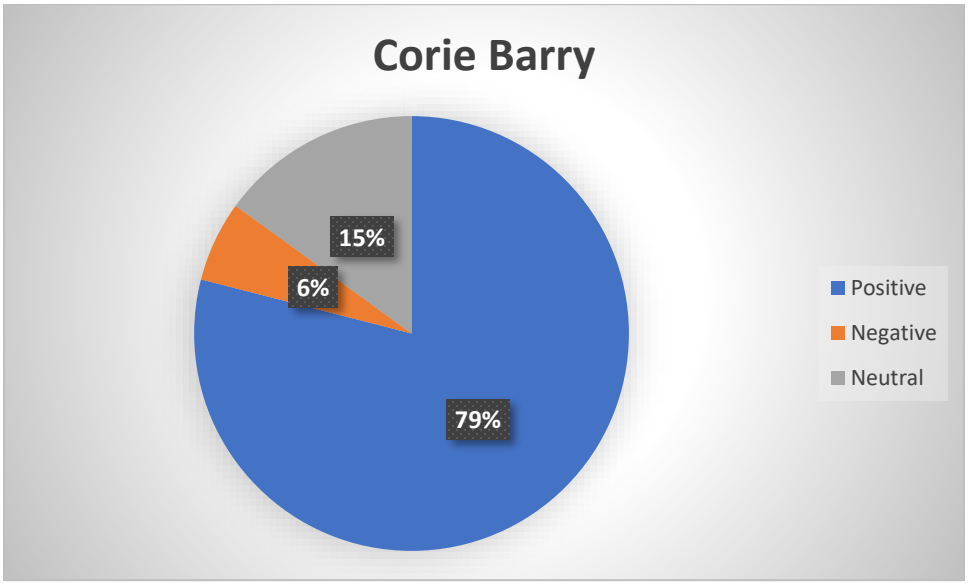

Figure 23 – Sentiment Analysis (Corie Barry)

When checking the time periods of when the positive, neutral and negative tweets are posted, there do not seem to be any major spikes in the polarity.

3.6.2 Subjective Affectivity Analysis

This subjective affectivity shows that the overall subjectivity goes everywhere. Most of what Corie Barry says is fairly objective, though not many of the scores stick to zero. The subjectivity is overall all over the place, but remains relatively in the middle and having an average score of 0.55.

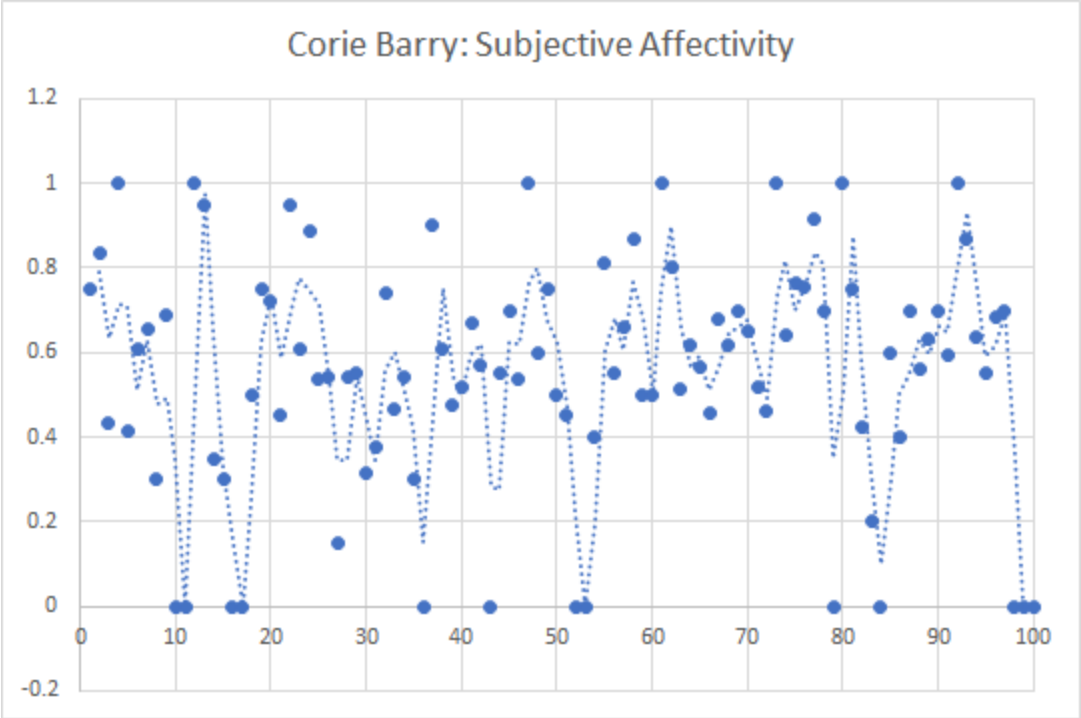

Figure 24 – Subjective Affectivity (Corie Barry)

Out of all of the tweets, 13% of them have a subjectivity score of 0. A lot of these tweets are sentiments of gratitude, such as replies saying "Thank you" or a variation like "Likewise". Others are simply stating objective information in order to spread awareness to other people who are watching two different parties interact with each other.

3.6.3 Time Frequency Analysis

The time period that the data goes from for Corie Barry is between April 23 2019 and December 23 2019. Like with Jeff Bezos, there is a very irregular posting pattern for Corie Barry. The only activity seen from Corie Barry comes from this period of time in 2019. And yet, it is still more consistent even throughout this one year. During this period, she only seemed to post about Best Buy, programs related to Best Buy and the team she typically works with when managing Best Buys. In fact, some examples of moments where she isn't talking about Best Buy come from replies when she thanks people or answers a question.

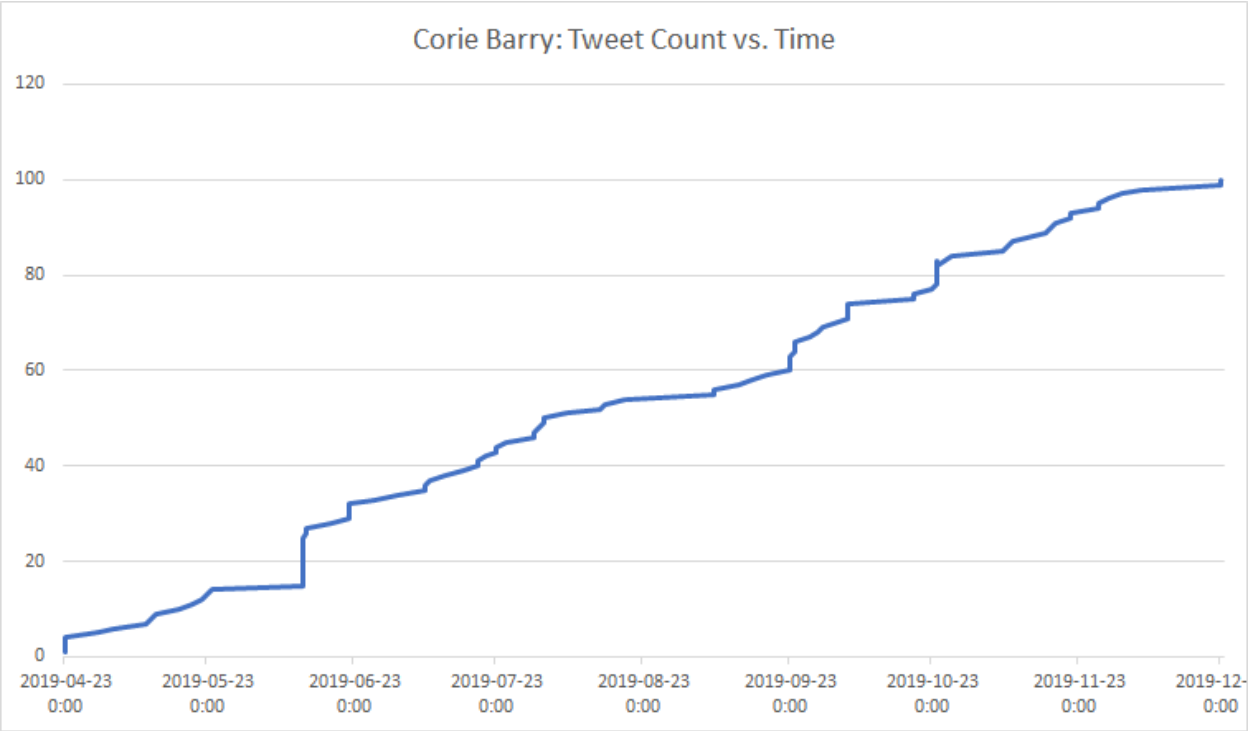

Figure 25 – Tweet Frequency by Time (Corie Barry)

There is a gradual rise, as to be expected from looking at the count of the tweets. Although her posting periods are fairly sporadic. Sudden jumps in the number of tweets appear fairly often, as stagnation amongst the line is also very common. After December 2019, Barry stopped posting.

3.6.4 Tweet Classification Analysis

A lot of the tweets that Corie Barry made are original tweets, not attached to any retweets or anything like that. A lot of these original tweets however, consist of replies. These ones are typically simple thanks to people who congratulate her, or congratulating others. Whenever there is a post that she has done on her own, she brings up Best Buy. There are a few videos that are posted as well, especially when showing off what Best Buy employees themselves are doing.

3.6.5 Tweet Frequency Analysis

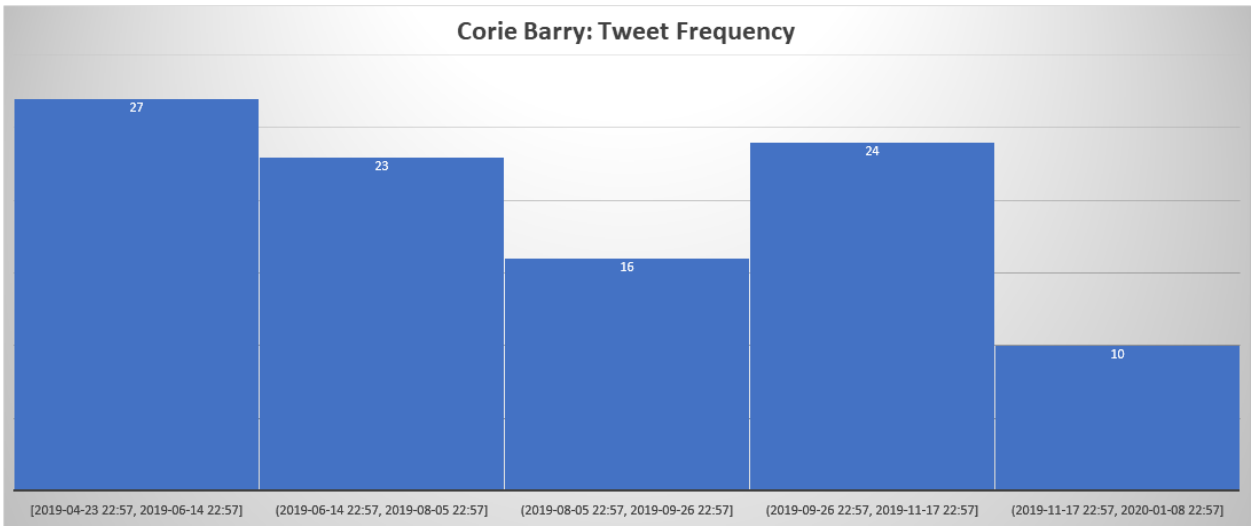

Figure 26 – Count of Tweets vs Time (Months) (Corie Bary)

During the posting period that we had for Corie Barry, she posted the most tweets between the end of April and the start of June.

3.6.6 Reaction Classification Analysis

Her most popular tweet has 126 likes and it is about a tweet stating she felt refreshed after spending a week in Las Vegas. She was glad to be back with her team and taking on new challenges, emphasized with a photo of her standing around with her employees.

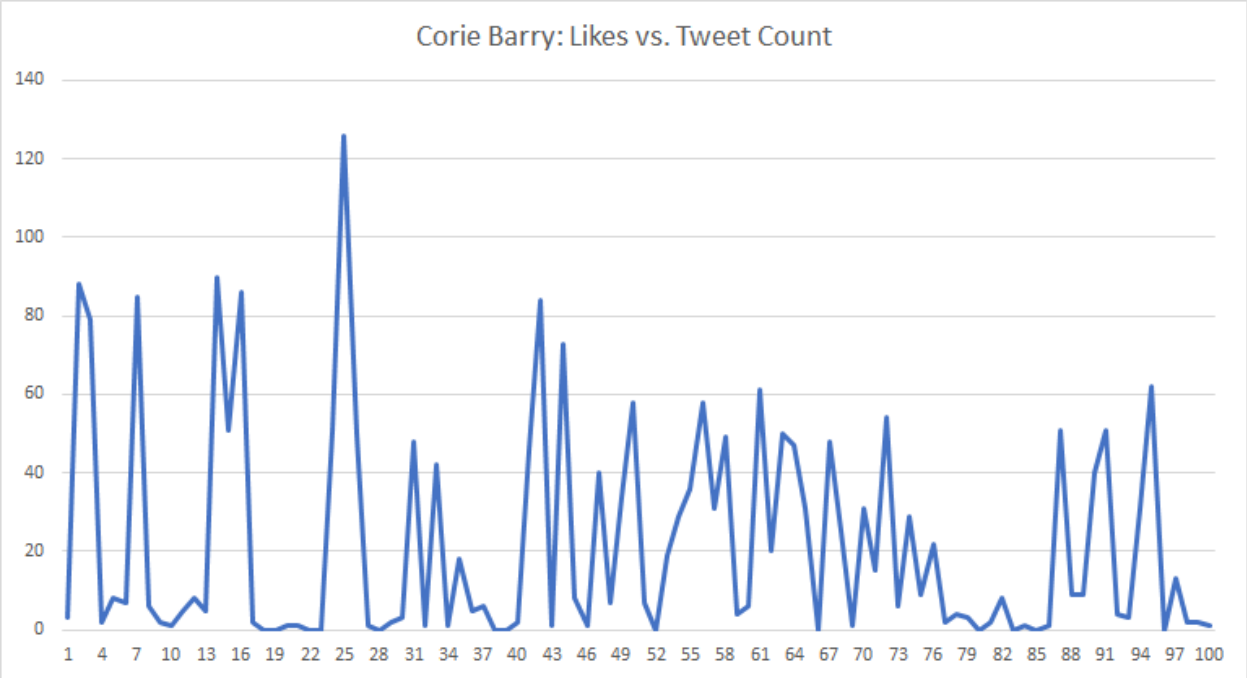

Figure 27 – Amount of Likes on Each Tweet (Corie Barry)

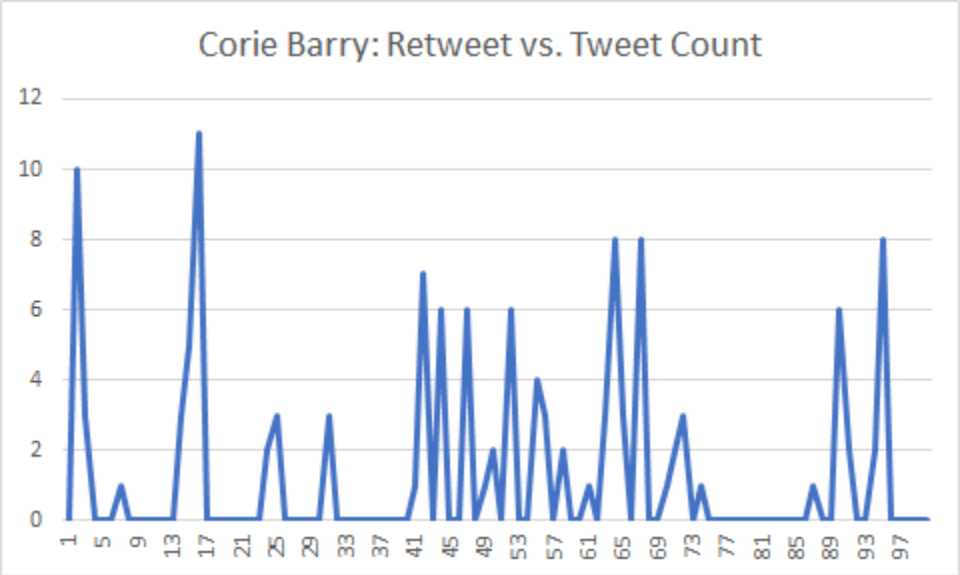

Figure 28 – Amount of Retweets on Each Tweet (Corie Barry)

3.6.7 Word Frequency Analysis

According to the Word Cloud analysis, the word that appears the most often in her Twitter is “Thank”. This is more likely than not, followed either as simply “thanks” or “you”. “Best Buy”

and “Family” are also words that appear the most. When added in conjunction with the fact “Thank” is there, it shows that lot of her tweets display gratitude to the staff.

Illustration 4 -Word Cloud for Corie Barry

<u>3.6.8 Parts of Speech: Noun, Verb and Adjective Count Analysis</u>

Corie Barry’s tweets show that there are 445 nouns in total. This makes up for 53% of the types of words that can be found in her tweets. There are a total of 201 verbs. This makes up 24% of the types of words that can be found in her tweets. Finally, adjectives makes up for 23% of the type of words with a total of 187 adjectives.

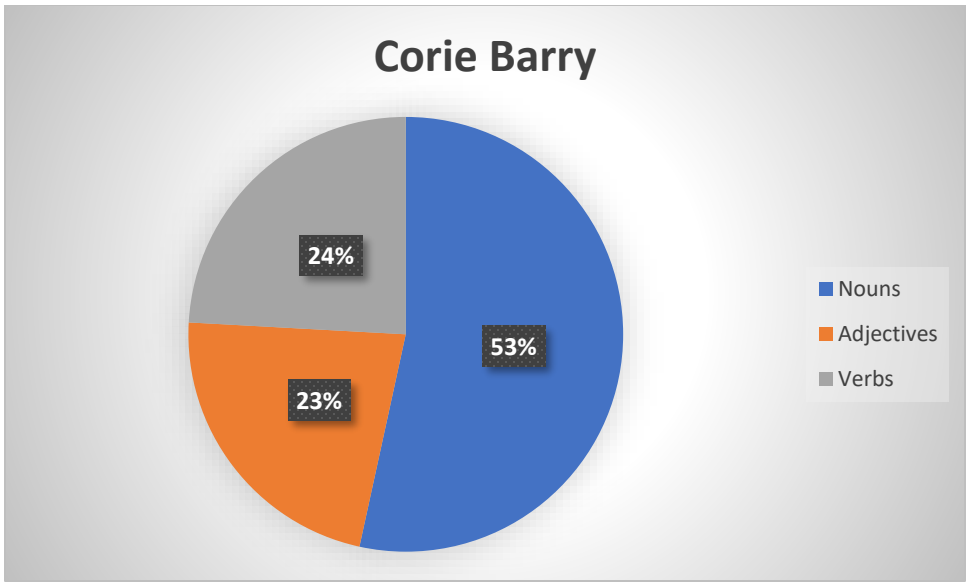

Figure 29 - Percentage of Text Consisting of Nouns, Adjectives and Verbs (Corie Barry)

3.6.9 User Mentions Frequency Analysis

57% of the tweets on Corie Barry's Twitter mentioned one or more users, most of them expressing congratulatory messages to her. Most of them are in reply to others, containing messages of gratitude of congratulating them on what they have done.

3.7 Case Study 5 – Ginni Rometty, CEO of IBM

Ginni Rometty retired CEO of IBM embedded IBM's A.I. platform including healthcare, finance and consulting, making IBM the most diverse company in 2018 [42]. Before her retirement Rometty was named the most powerful female CEO by many magazines. However Rometty is not an active user of Twitter. While we tried to gather 100 tweets from each of our cased studies, Ginni Rometty Twitter is the newest in terms of date of creation. As a result, she only has 39 tweets at the time of our data collection. Nevertheless, our feature engineering analysis reveal some interesting results about her content characteristics.

3.7.1 Sentiment Polarity Analysis

As Figure 31 shows, Rometty's tweets have a predominantly positive sentiment of 64%. In comparison, the amount of neutral and negative tweets make up the remaining percentage with 18% equally divided for them.

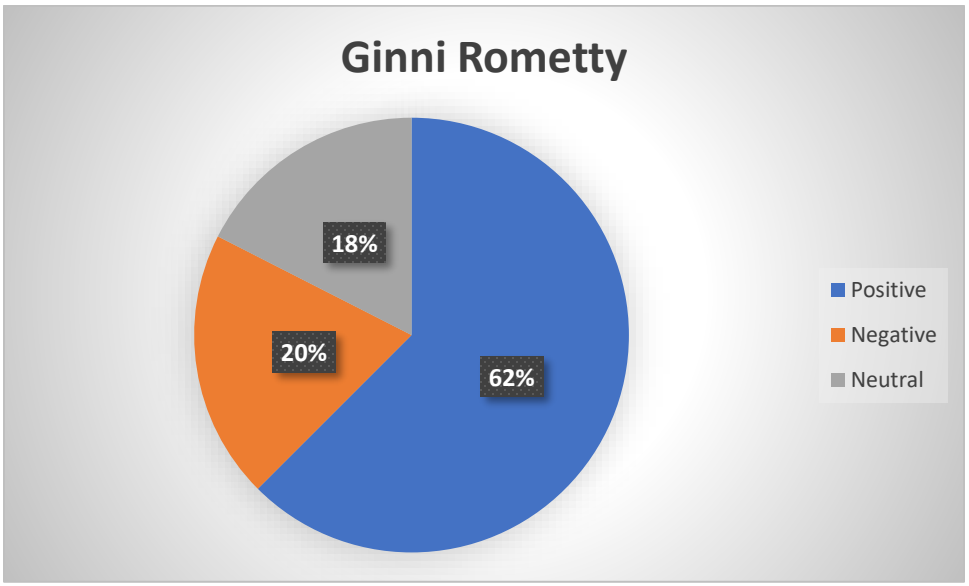

Figure 30 - Sentiment Analysis (Ginni Rometty)

Despite limited number of tweets from Rometty, a review of the 39 tweets indicate similar patterns like the other CEOs, most notably, the predominant positive sentiment and consistent polarity without significant swings.

3.7.2 Subjective Affectivity Analysis

Due to how few tweets there are, it is hard to tell what the exact subjectivity is. Regardless, we have a good amount of information that is depicted on the graph below. Based on what we can see, the subjective affectivity scores average out to about 0.39.

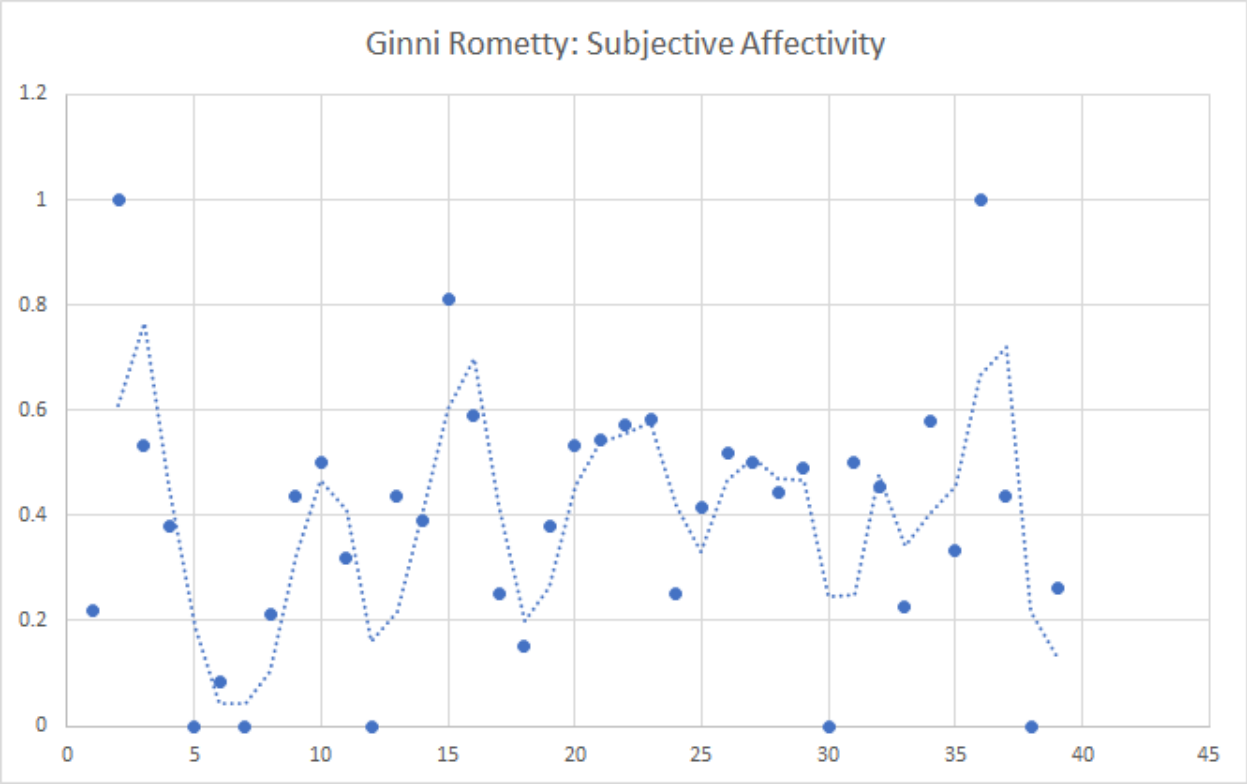

Figure 31 - Subjective Affectivity (Ginni Rometty)

3.7.3 Time Frequency Analysis

The time period that Ginni Rometty's tweets goes from December 28 2020 to March 8 2021. While it has the shortest time period, this can be accounted to the smaller amount of tweets that were collected. All of the tweets on her Twitter are original, with at least 8 of them being responses or retweets. However, at least three of the replies are replies to herself where she adds onto her previous points.

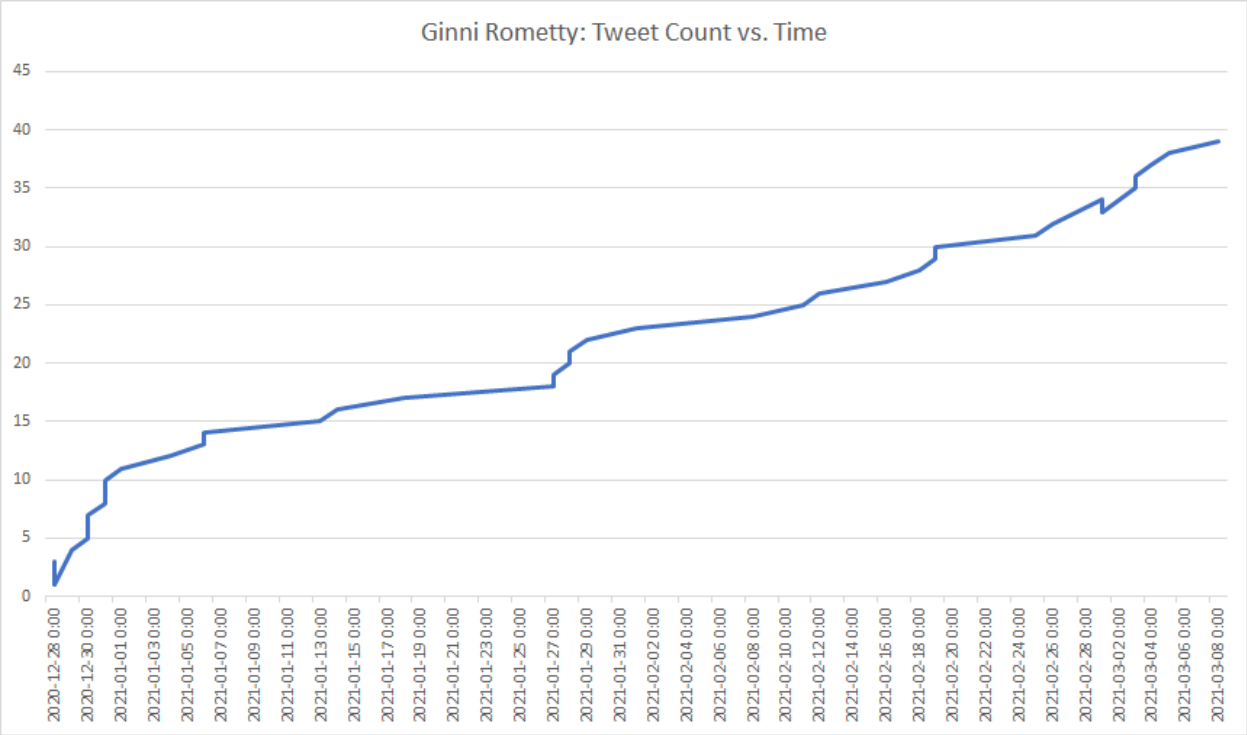

Figure 32 - Tweet Frequency by Time (Ginni Rometty)

3.7.4 Tweet Classification Analysis

When looking at the types of tweets that Ginni Rometty has posted on her Twitter, it is pretty fair between original tweets and retweets. From taking a look at the data that was collected, many of her tweets are original.

3.7.5 Tweet Frequency Analysis

Due to the smaller time window, it makes sense that there is not a lot of time for Ginni Rometty to show how many tweets she can post in certain periods. Figure 34 illustrates she tweeted pretty evenly in each of the 3 months where we collected data from.

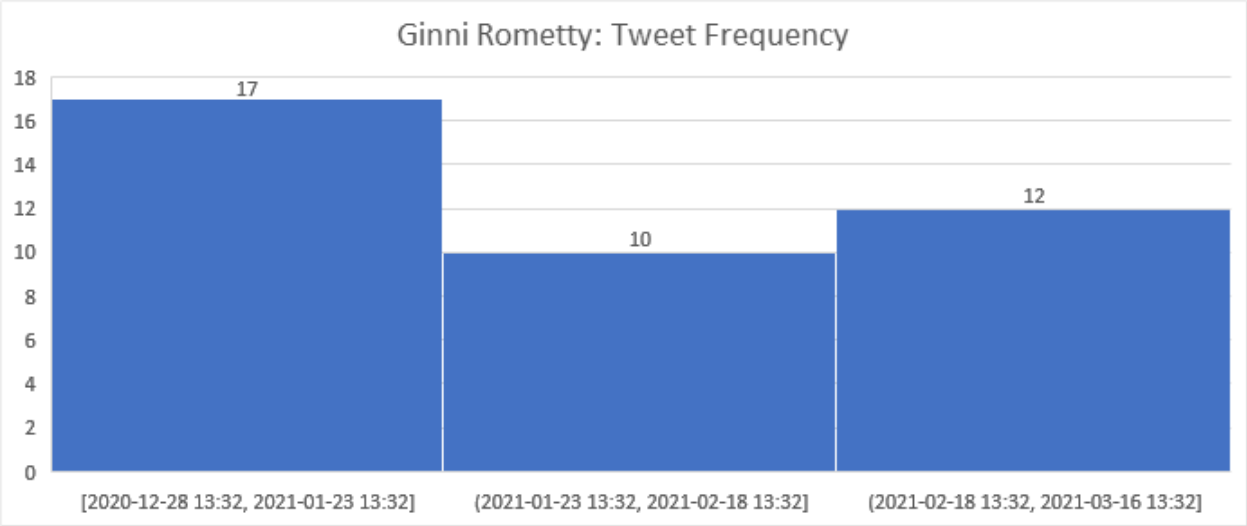

Figure 33 - Counts of Tweets vs Time (Months) (Ginni Rometty)

3.7.6 Reaction Classification Analysis

Overall, Ginni Rometty's tweets match a similar pattern to a lot of other bigger figures on Twitter. She has a level of 7,482 user interactions. All of the 39 tweets we looked at had attracted significant interactions with total of 6,601 likes and 881 retweets.

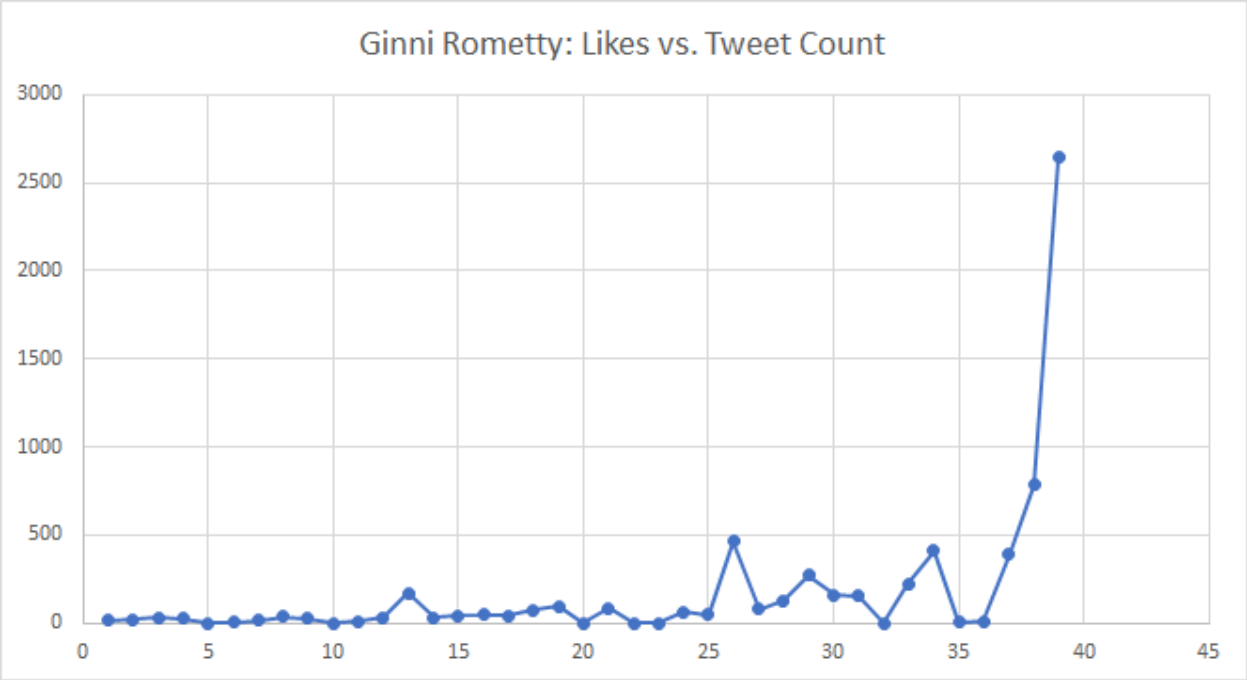

Figure 34 – Amount of Likes on Each Tweet (Ginni Rometty)

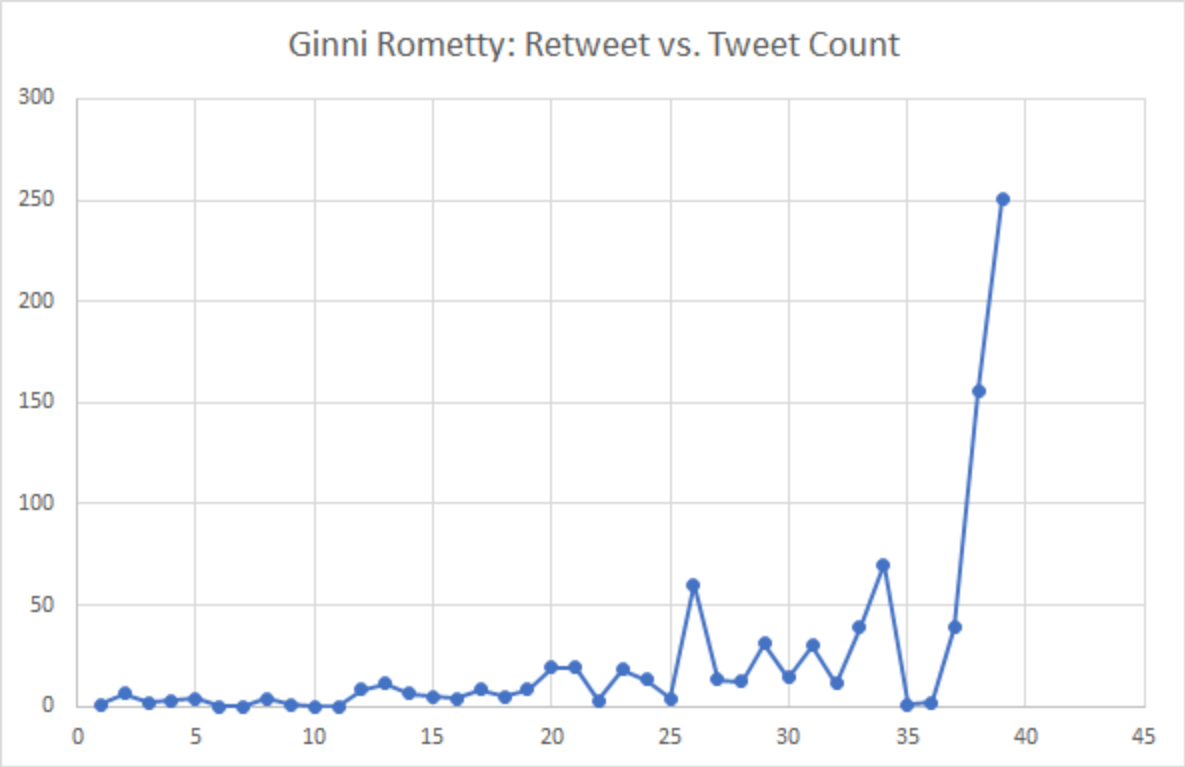

Figure 35 – Amount of Retweets on Each Tweet (Ginni Rometty)

The tweet that received the most likes and the most retweets is the exact same, being the very first tweet she posted on her account. This was a post that was wishing the best of luck to IBM after she had stepped down as the CEO. Afterwards though, there are not many other posts that had received the same amount of attention in either the Likes or the Retweets.

3.7.7 Word Frequency Analysis

The most common word that is depicted on Ginni Rometty's Twitter account is "Black", "first", and "will". These words are the results of Rometty's strong beliefs in social justice for black people and other minorities. From her 39 tweets, her passion for promoting social justice is obvious.

Illustration 5 - Word Cloud of Ginni Rometty's tweets

When looking at the topics that Ginni Rometty talks about, quite a few of them are words of encouragement. A lot of that encouragement goes towards students, potentially because that might be the demographic group that she develops.

3.7.8 Parts of Speech Analysis

As illustrated in Figure 38, the word count for Ginni Rometty's tweets shows a total of 184 nouns. This makes for 54% of the parts of the speech. This adds up with what we know when looking at the Word Cloud. There is more use of nouns compare to verbs and adjectives. In the word count for Ginni Rometty's tweets, there are a total of 79 verbs. This makes for 23% of the parts of the speech being verbs. In the word count for Ginni Rometty's tweets, there are a total of 79 adjectives. This makes for 23% of the parts of the speech being verbs. Coincidentally, it is also the exact same amount of verbs that were calculated.

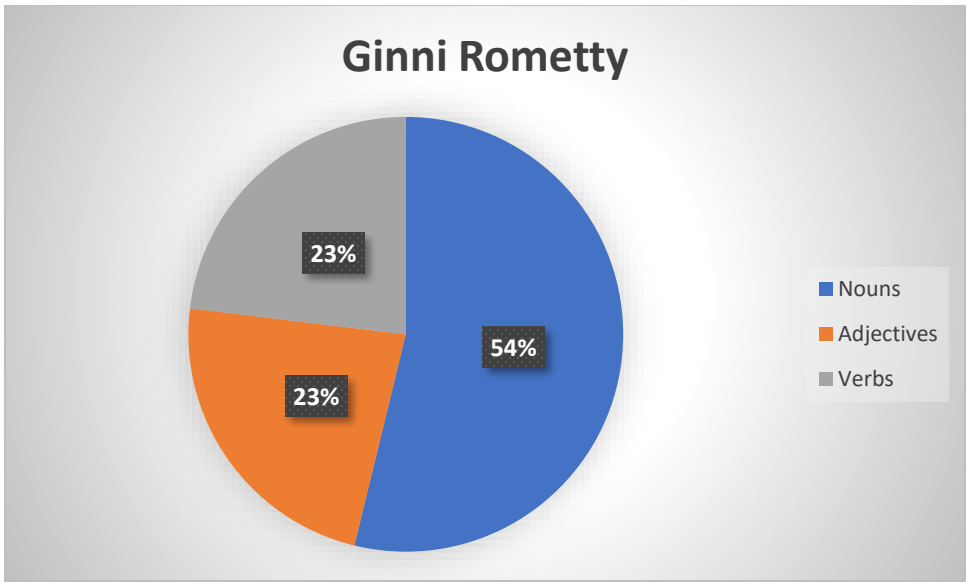

Figure 36 - Percentage of Text Consisting of Nouns, Adjectives and Verbs (Ginni Rometty)

3.7.9 User Mentions Frequency

Overall, Ginni Rometty mentioned about 5 Twitter users. All of them are unique, although one of them is not included due to it being a response to herself since it was a thread. Overall though, it should be noted that the low user mentions frequency is likely because of low number of tweets she has on her account.

3.8 Case Study 6 – Lisa Su, President of AMD

Dr. Lisa Su is the president of AMD, one of the most innovative tech companies. She is a computer engineer from MIT has been recognized by Fortune Magazine as one of the Most Powerful Women in Business. She is also the recipient of many awards and honors, including the International Peace Honors. This award recognized as the most outstanding global leaders and change-agents of our time and her philanthropic humanitarian work. Given the combination of her popularities, we expect Lisa Su's tweet analysis will show us some interesting features.

3.8.1 Sentiment Polarity Analysis

The sentiment analysis results in Figure 39 shows an astonishing level of positive sentiment with 97% of her tweets contain a positive sentiment, 3% of these tweets contain a neutral sentiment and none of them contain a negative sentiment.

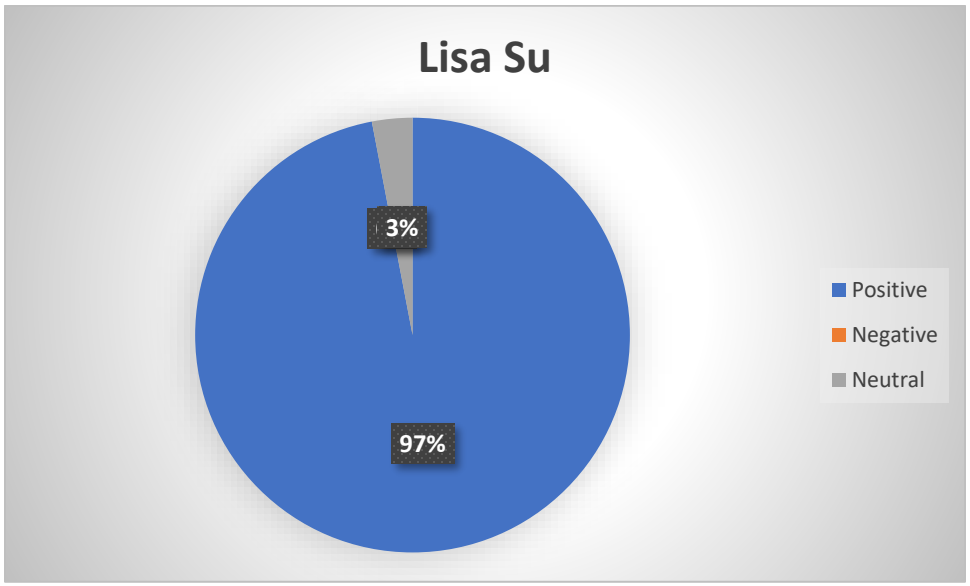

Figure 37 - Sentiment Analysis (Lisa Su)

This makes the polarity sentiment more difficult to tell. Without many or any tweets that have negative connotations, there are not a lot of comparisons that can be made between them.

3.8.2 Subjective Affectivity Analysis

The subjective affectivity scores for Lisa Su are fairly high. This is likely from the fact that a lot of her tweets are towards more positive sentiments. They have a tendency of leaning more towards being subjective. Although it is interesting to see that none of the positive tweets could have been interpreted as negative or neutral like many other tweets.

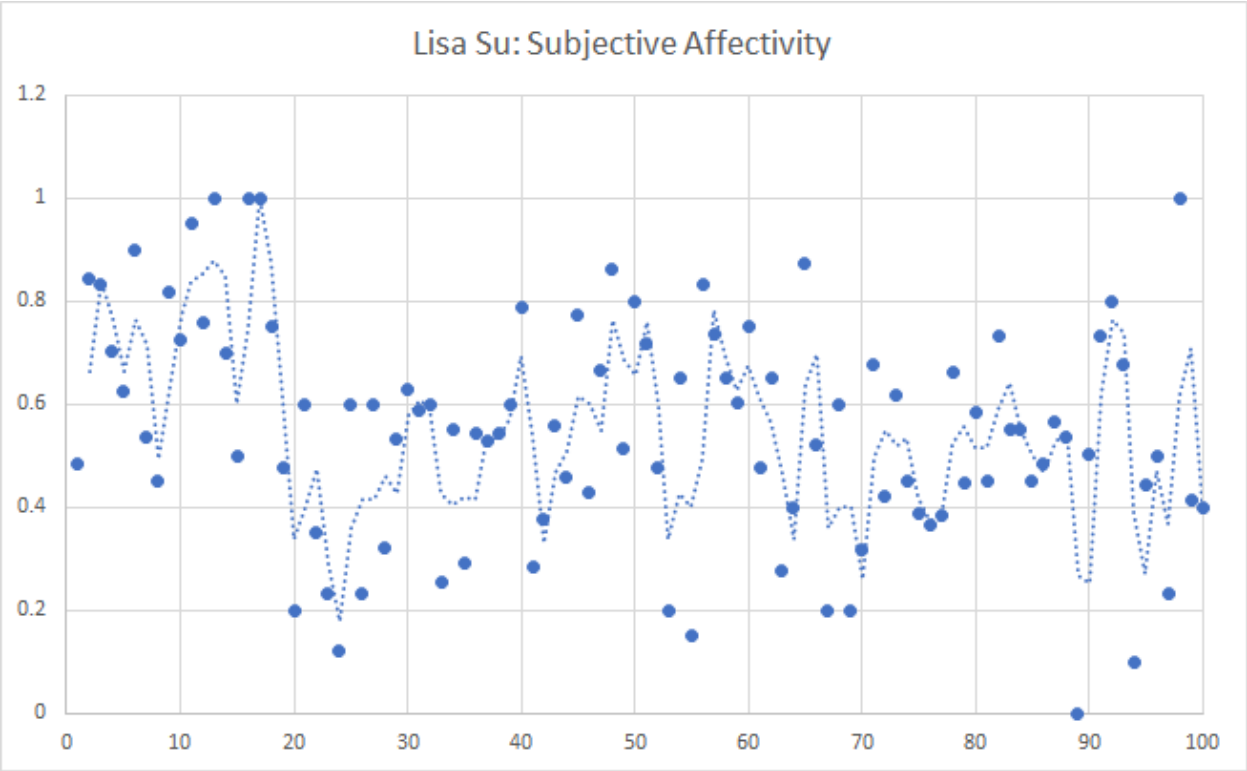

Figure 38 - Subjectivity Affectivity (Lisa Su)

The average mean number of Lisa Su's subjective affectivity score is 0.55. Only one of her tweets have ever hit a subjectivity score of "0". Only 4% of them had a subjective affectivity score of 1.

3.8.3 Time Frequency Analysis

The time period that Lisa Su's tweets take place goes from May 1 2019 to March 8 2021. When looking at how many tweets they have posted, it is clear that she is fairly consistent with how much she posts. A lot of the time, she tends to post in the afternoon.

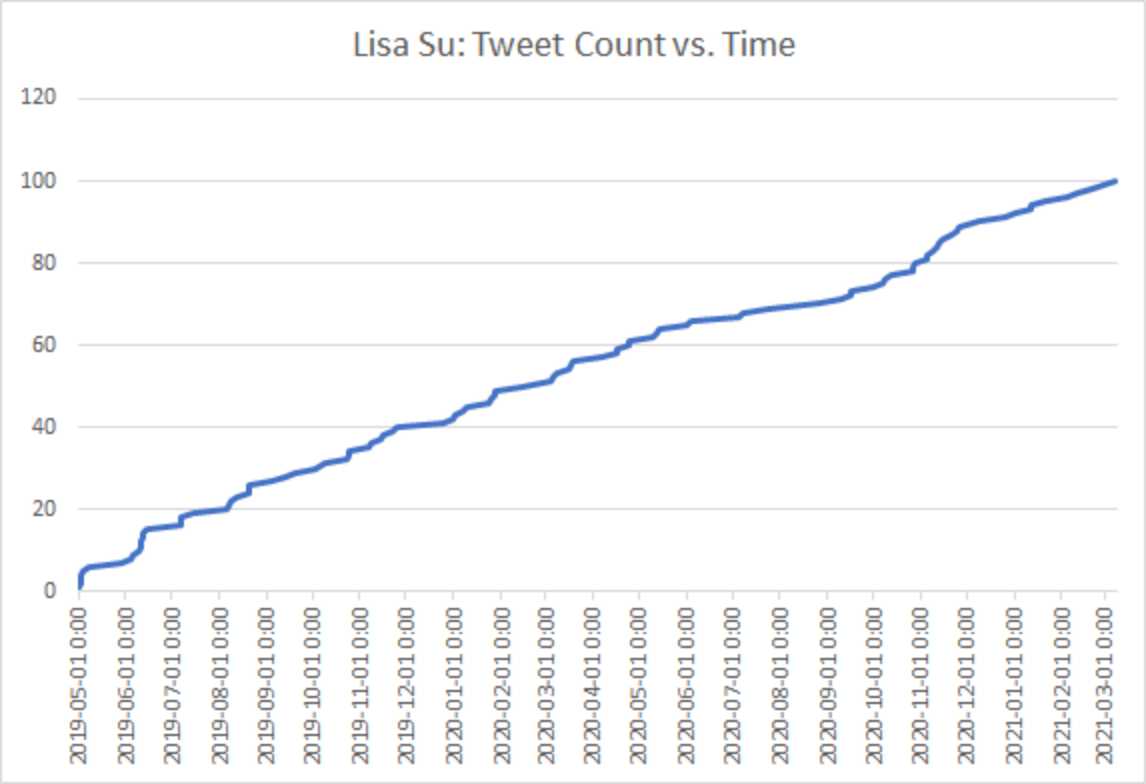

Figure 39 - Tweet Frequency by Time (Lisa Su)

Her tweet frequency is constant. There are not huge spikes, nor are there moments of drought when going through the tweets. There are a few moments where there are spikes when she tweeted about business events.

3.8.4 Tweet Classification Analysis

A good amount of the tweets that Lisa posts are some type of original tweet, meaning most of the time, Lisa Su is posting her own tweets. Despite that, a lot of these tweets can be a response of some kind, particularly if there is something important that is being brought up. There are a few times that the original tweets she posted are responses, usually expressing thanks to someone she had spoken to.

3.8.5 Tweet Frequency Analysis

Most of the tweets that Lisa Su had posted were in 2019, particularly during the time period of May 1 2019 to October 8 2019. This is evident when looking at the days and times when her tweets are posted for the most part.

A common thread in a lot of the tweets that she has is that she tends to tweet about important events or holidays. The more common these events are, then the more likely she is to tweet out about it.

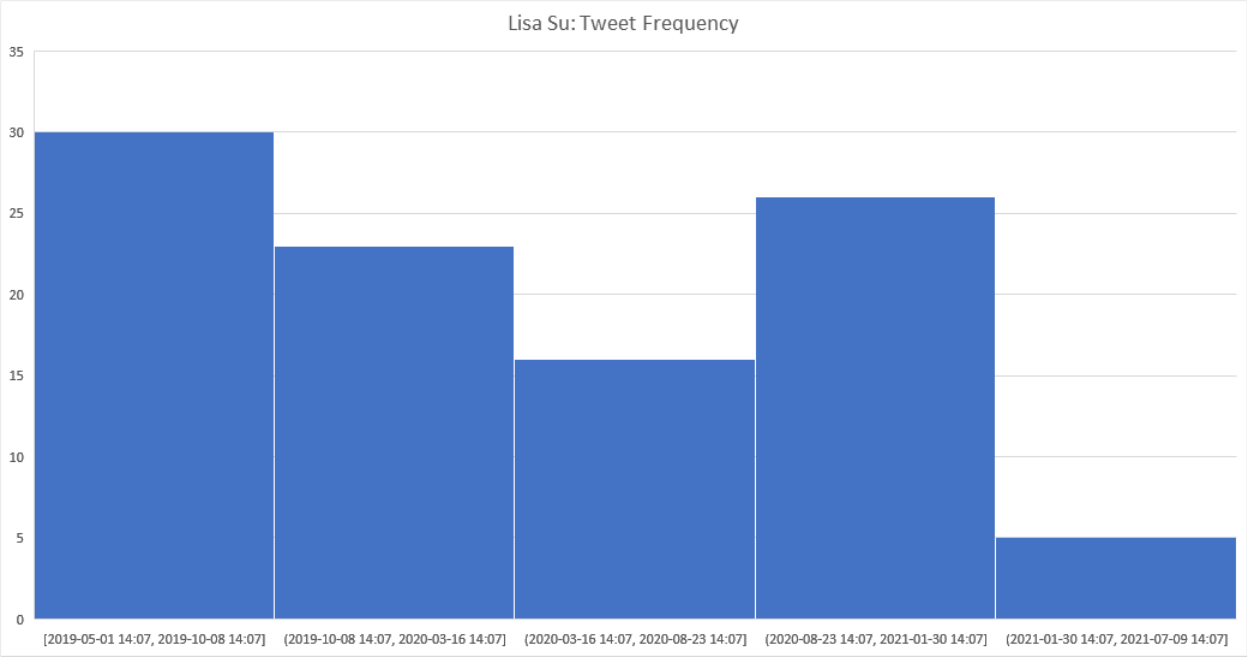

Figure 40 – Count of Tweets vs Time (Months) (Lisa Su)

3.8.6 Reaction Classification Traits

In total, Lisa Su's tweets have gathered a total of 209,290 interactions. From these, 22,956 of them are retweets and 186,334 of them are likes.

From looking at Lisa Su's Twitter account, there is a strange amount of spikes in the number of likes and retweets she receives on her tweets. The tweet that gathered the most retweets collected 2,389 retweets in total. This tweet is one advertising one of the kinds of technology that ADM is following, in this case, for video games. Video games, particularly the technology behind

it, is a very popular topic with people on the internet. Any news related to gaming and the hardware related to it does tend to receive quite a high amount of popularity.

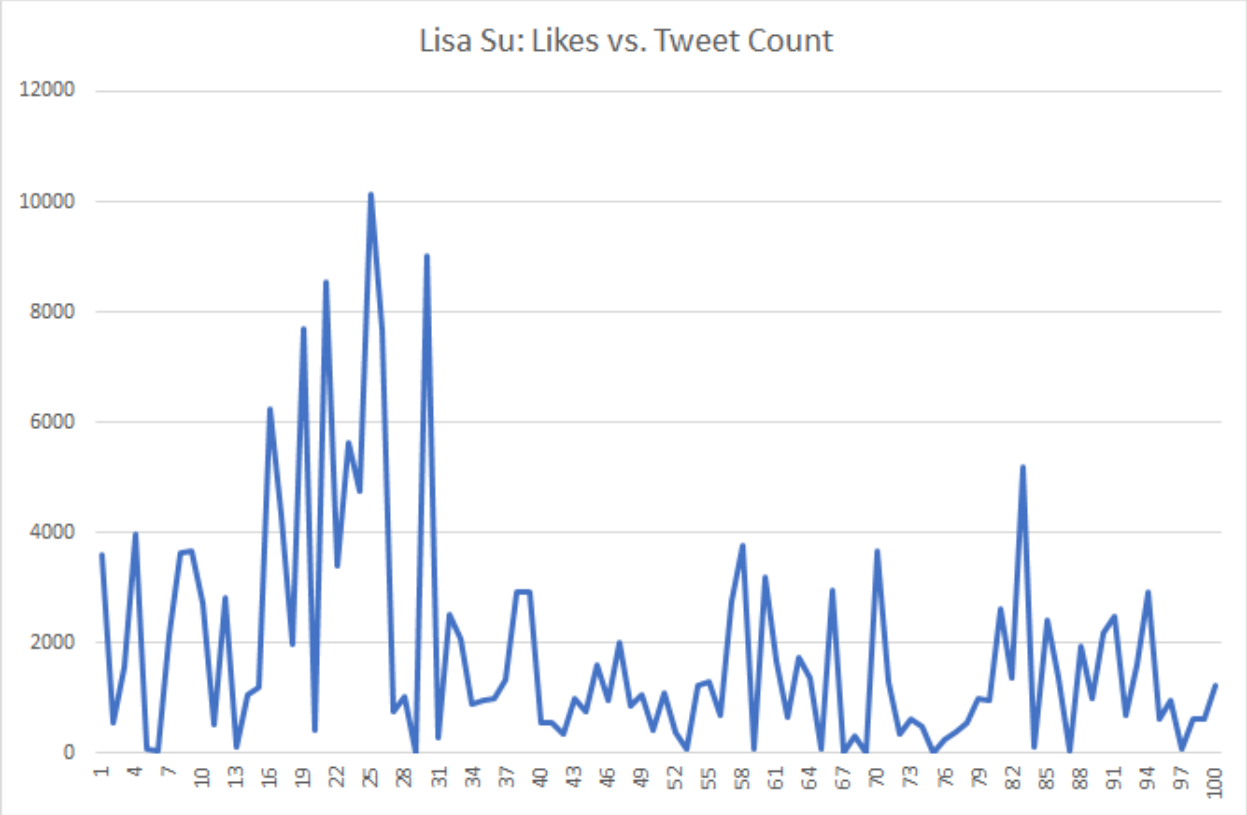

Figure 41 – Amount of Likes on Each Tweet (Lisa Su)

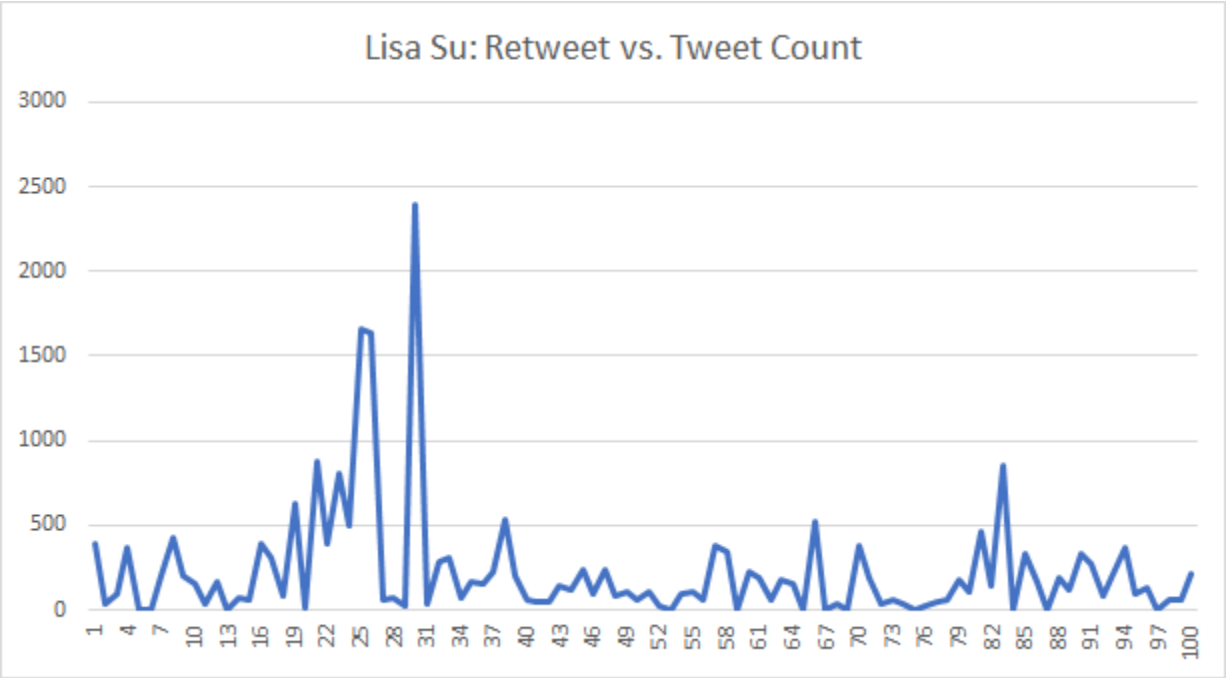

Figure 42 – Amount of Retweets on Each Tweet (Lisa Su)

The spikes between the retweets and likes are fairly consistent. This is likely because, when communicating between different business accounts, these business accounts like to retweet praise.

3.8.7 Word Frequency

From Lisa Su's Twitter account, the word that appeared the most often is "Thank" with quite a few instances including the word "you". However, the program does not take into the account for pronouns of any kind. This lines up with her sentiment analysis, as many of them have her saying her thanks to other people or companies.

Illustration 6-Word Cloud of Tweets from Lisa Su

There are other words of endearment like “friend” and “partner” regularly being used. These examples are more likely than not business related.

<u>3.8.8 Parts of Speech: Noun, Verb and Adjective Count Frequency</u>

When looking at all of Lisa Su’s tweets, there are 520 nouns in total. This makes up about 57% of the parts of the speech in the tweets. She also used 194 verbs in total. These make up for 22% of the parts of speech in the tweets. Finally, there are 194 adjectives in her tweets in total. These make up for 21% of the parts of the speech in the tweets.

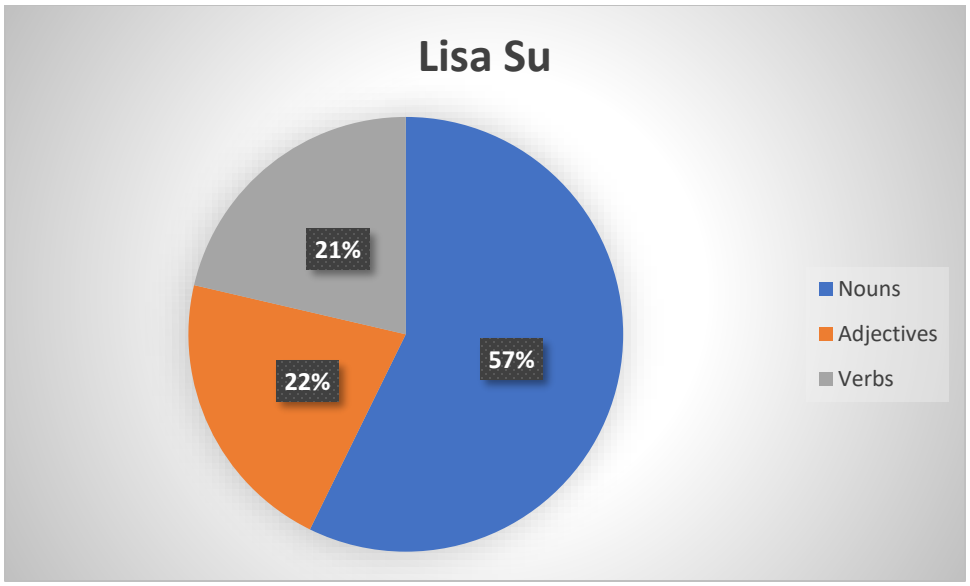

Figure 43 - Percentage of Text Consisting of Nouns, Adjectives and Verbs (Lisa Su)

3.8.11 User Mention Frequency Analysis

About 16% of Lisa Su's tweets directly mention someone in a response, meaning it's fairly frequent that someone gets mentioned. However, when we take a look at the mentions themselves, most of Lisa Su's tweets mention some other account.

3.9 Case Study 7 – Sundar Pichai, CEO of Alphabet Inc and Google

Sundar Pichai is the current CEO of Alphabet Inc. and Google. For this study, we collected 100 of his most recent tweets on his Twitter account. Sundar Pichai has been included in Time's annual list of the 100 most influential people in 2016 and 2020. Under his leadership, Google developed Chrome and Chrome OS and Google Drive.

3.9.1 Sentiment Polarity Analysis

On Sundar Pichai's Twitter, 78% of the tweets have a positive sentiment, 17% of them have neutral sentiments and 5% of them have negative sentiments. This is an average amount of tweets that shows both positive sentiments and sentiment sentiments.

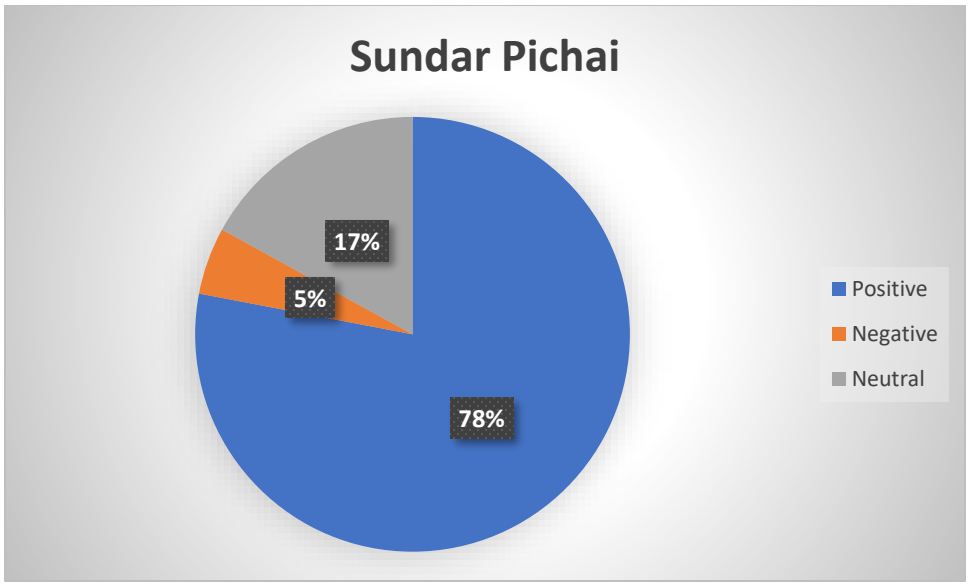

Figure 44 - Sentiment Analysis (Sundar Pichai)

There is quite a bit of talk about security, and what method can be used with increasing it. Likely referencing how browsers such as Google Chrome work hard in being able to increase the security to be the ideal browser for its users.

3.9.2 Subjective Affectivity Analysis

The following graph depicts the subjective affectivity scores from the 100 tweets we collected from Sundar Pichai's Twitter. On average, the subjective affectivity score for Sundar Pichai's tweets was about 0.5. Out of the subjectivity scores, 14% of them had a subjectivity score of 0. Meanwhile, 5% of them had a subjectivity score of 1. This averages out a lot of the numbers, making the subjectivity score fairly even.

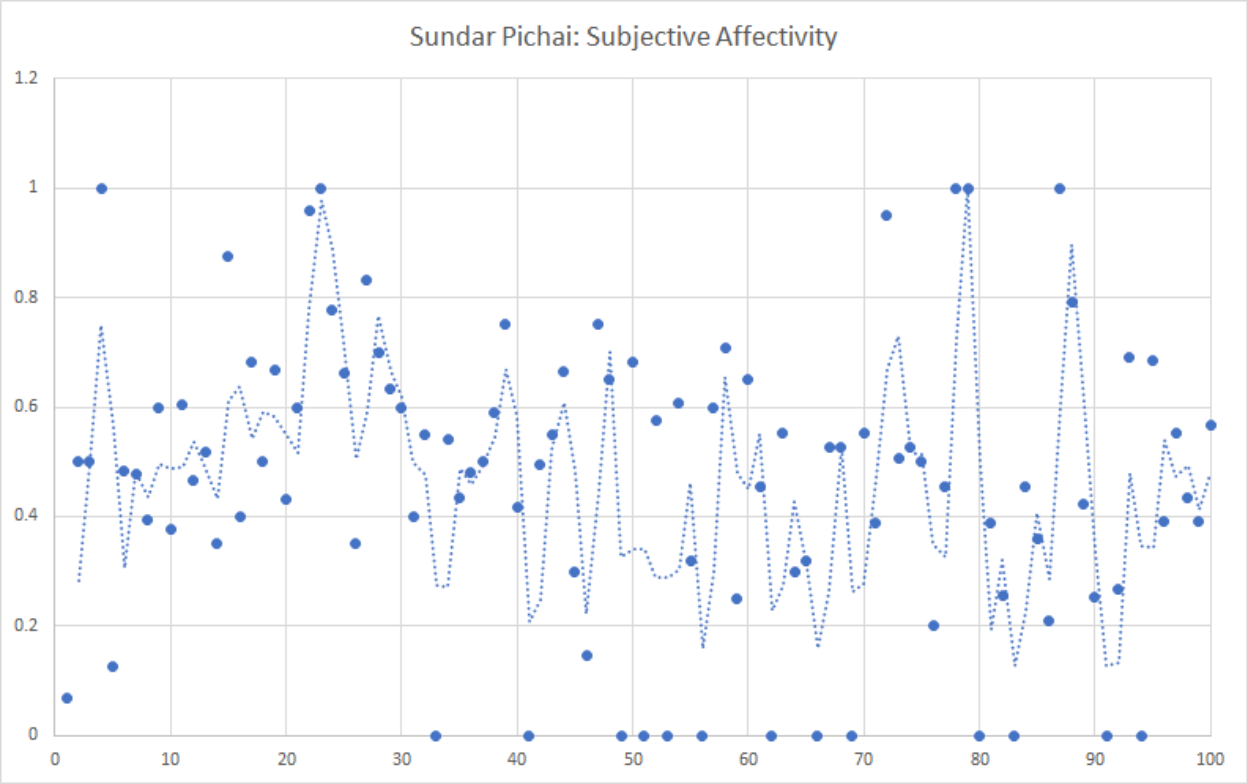

Figure 45 - Subjectivity Affectivity (Sundar Pichai)

When looking at the tweets individually, the types of posts that Sundar Pichai posted with a subjectivity score of 0 were speaking of information. They are relatively objective facts about things that had happened in history, or were meant to celebrate the accomplishments of others. Among the tweets that had a score of 1, there are more tweets using specific languages that specify the subjective affectivity as relatively stable.

3.9.3 Time Frequency Analysis

The time period for Sundar Pichai's tweets go from June 19 2020 to March 8 2021. He is consistently active on his Twitter, managing to post at least 5 times a month about various subjects. Sometimes they are posts about something Google contributed to, sometimes it's congratulating others based on their scientific breakthroughs.

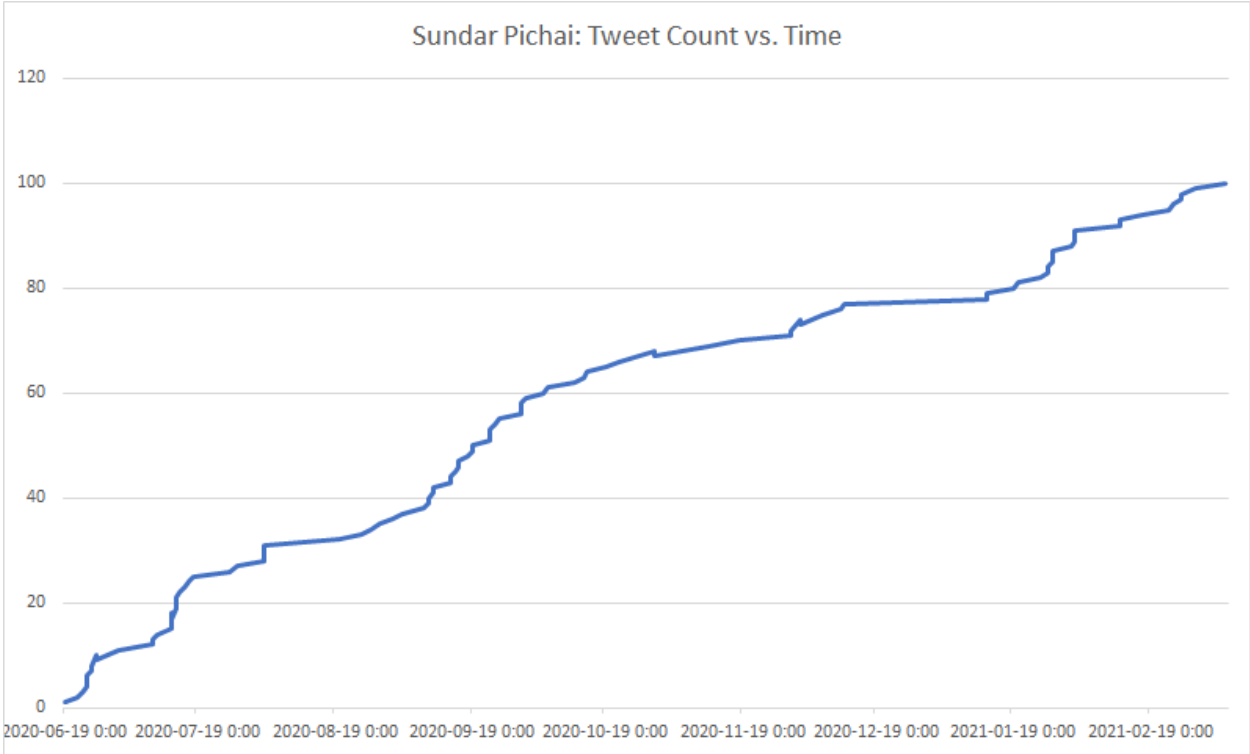

Figure 46 - Tweet Frequency by Time (Sundar Pichai)

The time they are more likely to post is between the afternoon and early evening. When checking the tweet count, it has a fairly consistent posting schedule. There are periods such as between July 2020 and August 2020 where nothing was posted. There are also a lot of bumpy roads during the time of November 2020 to December 2020. These may relate to the times when Sundar Pichai was called by the US Senate Committee on Commerce, Science and Transportation to testify.

3.9.4 Tweet Classification Analysis

Sundar Pichai posts many original tweets. Among these original tweets, a lot of them are inspirational in nature. Not many of them contain photos, although among the ones that do, they tend to be relevant to what he is talking about. Sometimes they are holiday related images, sometimes they are images that involve social justice.

Regardless, there are very few times when Sundar Pichai retweets. Out of the 100 tweets, only 27% of his tweets are retweets. 7% of his tweets are replies to someone, although since one of them is to himself, it is actually 6% based on the criteria we had set up. Some of the tweets are retweets form Susan Wojcicki, another high profile CEO on our sample list.

3.9.5 Tweet Frequency Analysis

The posting period of Sundar Pichai's tweets are in a time period from June 19 2020 to March 8 2021. Within this period, they were all divided into periods of two months each.

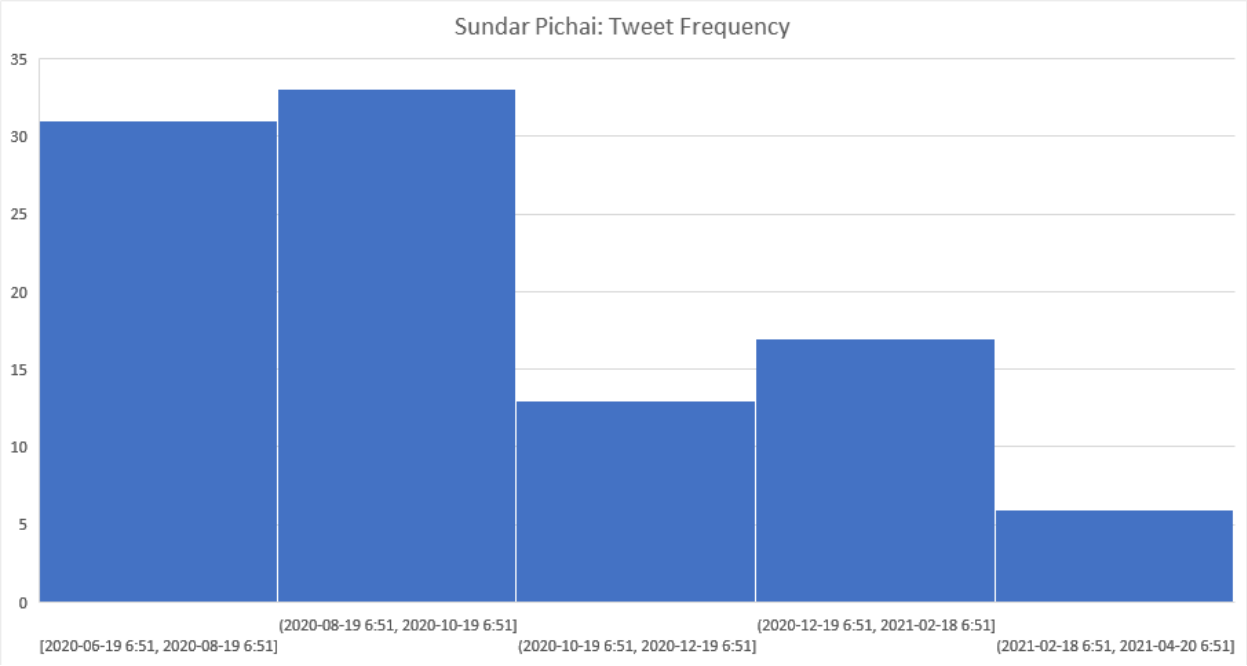

Figure 47 - Tweet Frequency for Sundar Pichai

The most tweets were posted in the time period of August 19 2020 to October 19 220. The least amount of tweets were in the most recent time period, February 18 2021 to April 20 2021.

3.9.6 Reaction Classification Analysis

Like with many of the tweets, there are random spikes of large amounts of likes and retweets that happen for Sundar Pichai's tweets. There are 970,299 likes and 201,382 retweets. In total, his tweets have collected 1,171,681 user interactions.

The tweet that received the most likes earned 316,512 likes. This tweet is one that congratulated India and Australian cricket game. This got a lot of attention, likely because it is a sports tweet. Whenever these big sporting events happen, a lot of users, especially if they have big name accounts, can attract a lot of attention.

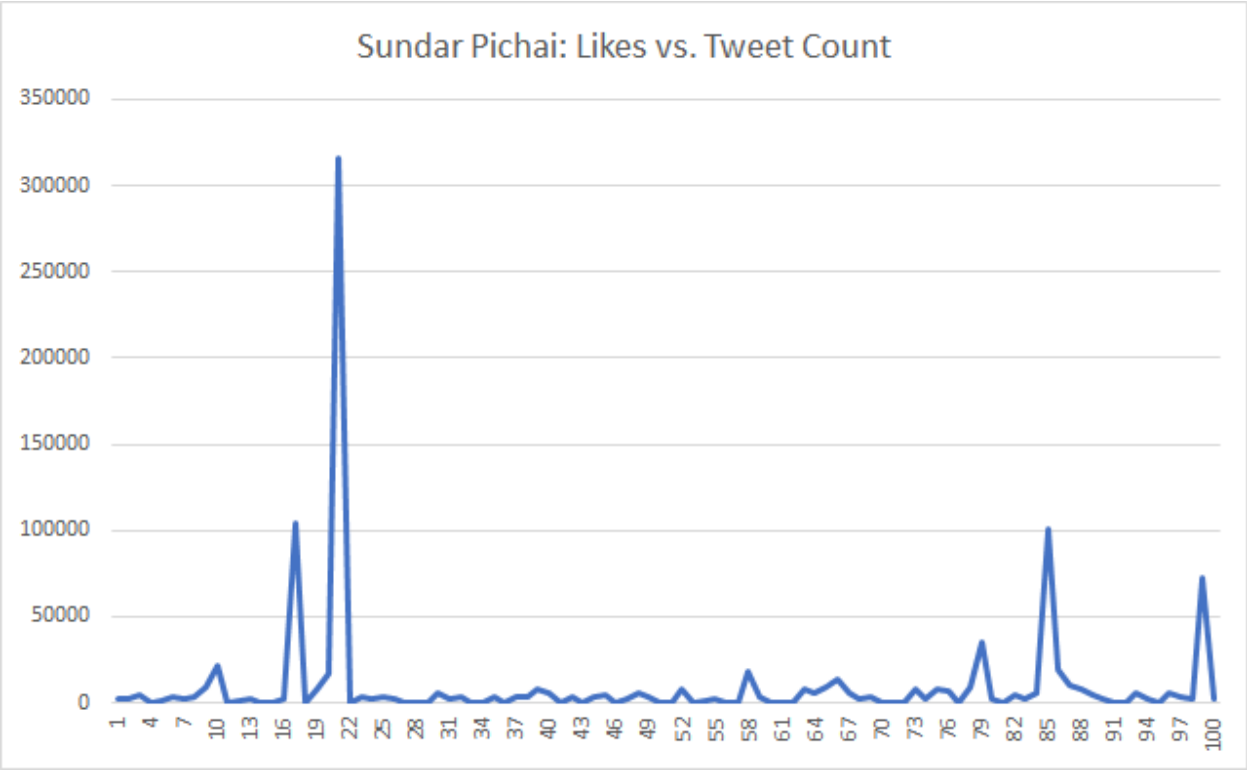

Figure 48 - Amount of Likes on Each Tweets (Sundar Pichai)

The tweet that received the most retweets collected 59,532 retweets in total. It is one that discusses Jewish traditions. However, same tweet has no Likes. That is because, his most retweeted tweet was retweeted from someone else. As a result, it will take note of the retweets because of how many people have retweeted the original. Although since the likes go to the original author of the tweet, it shows up with a total of zero.

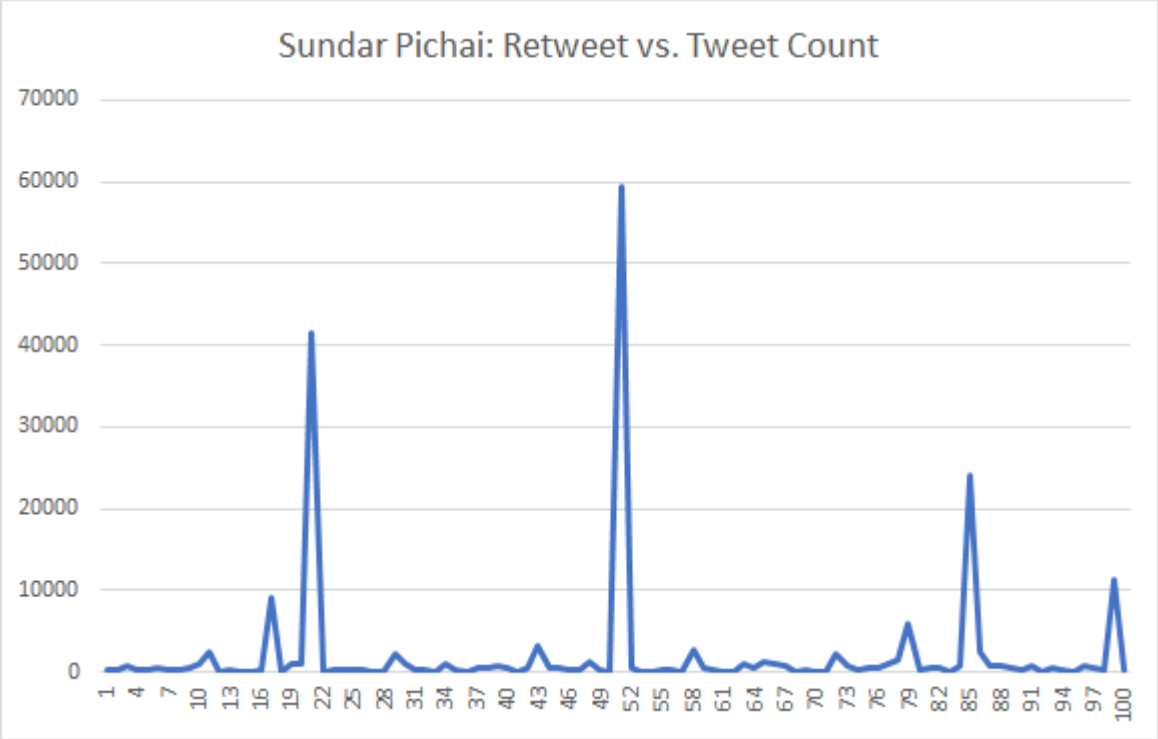

Figure 49 - Amount of Retweets on Each Tweet (Sundar Pichai)

3.9.7 Word Frequency and Lexical Richness Analysis

Based on the data, the word that appears the most commonly on Sundar Pichai's Twitter is Google. As the CEO of Google, the word expressed his preoccupation with Google.

Illustration 7 - Word Cloud Analysis of Sundar Pichai's Twitter

Other most frequently appeared words are related to wanting to help or be innovative. A focus is on the future with the word "will". Unsurprisingly, there is quite a bit of talk about COVID, especially in a time of uncertainty about the effects of COVID.

3.9.8 Parts of Speech: Noun, Verb and Adjective Count Analysis

Among all of Sundar Pichai's tweets, there are 1184 nouns in total. This makes up for 56% of the parts of speech in all of Sundar Pichai's tweets. There are 489 verbs in total. This makes up for 23% of the parts of speech in all of Sundar Pichai's tweets. There are 459 adjectives in total making up for 21% of the parts of speech.

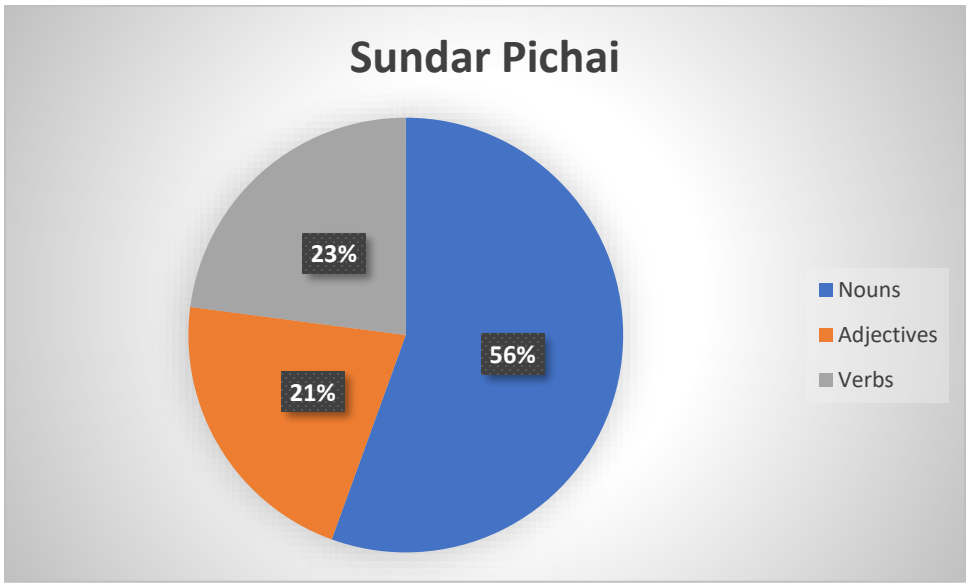

Figure 50 - Percentage of Text Consisting of Nouns, Adjectives and Verbs (Sundar Pichai)

3.9.9 User Mentions Frequency Analysis

User mentions from Sundar Pichai is low meaning he does not mention other Twitter users. 6% of his tweets are responses to someone else. This means that Sundar Pichai does not often reply to people, nor does he mention many other accounts. While it is not mentioned in the list of replies due to being a retweet, two of his tweets are retweets by Susan Wojcicki who is also mentioned in our case studies.

3.10 Case Study 8 – Susan Wojcicki, CEO of YouTube

Susan Wojcicki, the CEO of YouTube, is also recognized by Time and Forbes magazines. [42] She is also known for her promoting of women and girls in technology sphere. 100 of her most recent tweets were collected and then ran through our feature engineering. The following describes what was discovered.

3.10.1 Sentiment Polarity Analysis

Susan Wojcicki's Twitter's tweets shows the following when put through a sentiment analysis test. 70% of the tweets have a positive sentiment, 20% of them have neutral sentiments and 10% of them have negative sentiments. Like the other CEOs we see so far, Wojcicki has fairly high number of positive tweets as well.

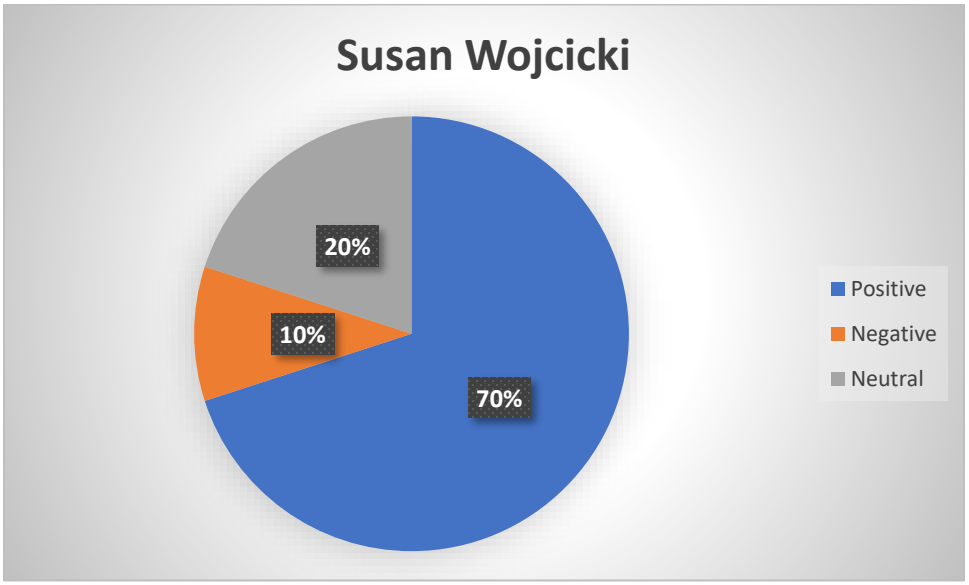

Figure 51 – Sentiment Analysis (Susan Wojcicki)

It is likely that a good amount of Susan's tweets are positive because she signs praises of popular YouTubers and musical artists as well as the YouTube platform itself. Like many others, she wants to express positive sentiments to the company and people she works with.

3.10.2 Subjective Affective Analysis

Susan Wojcicki's average subjective affectivity score came to a 0.48. Out of the subjectivity scores, 4% of all of the tweets had a subjectivity score of 1. Among the ones with a subjectivity score of 0, 18% of them were that low in their subjectivity scores.

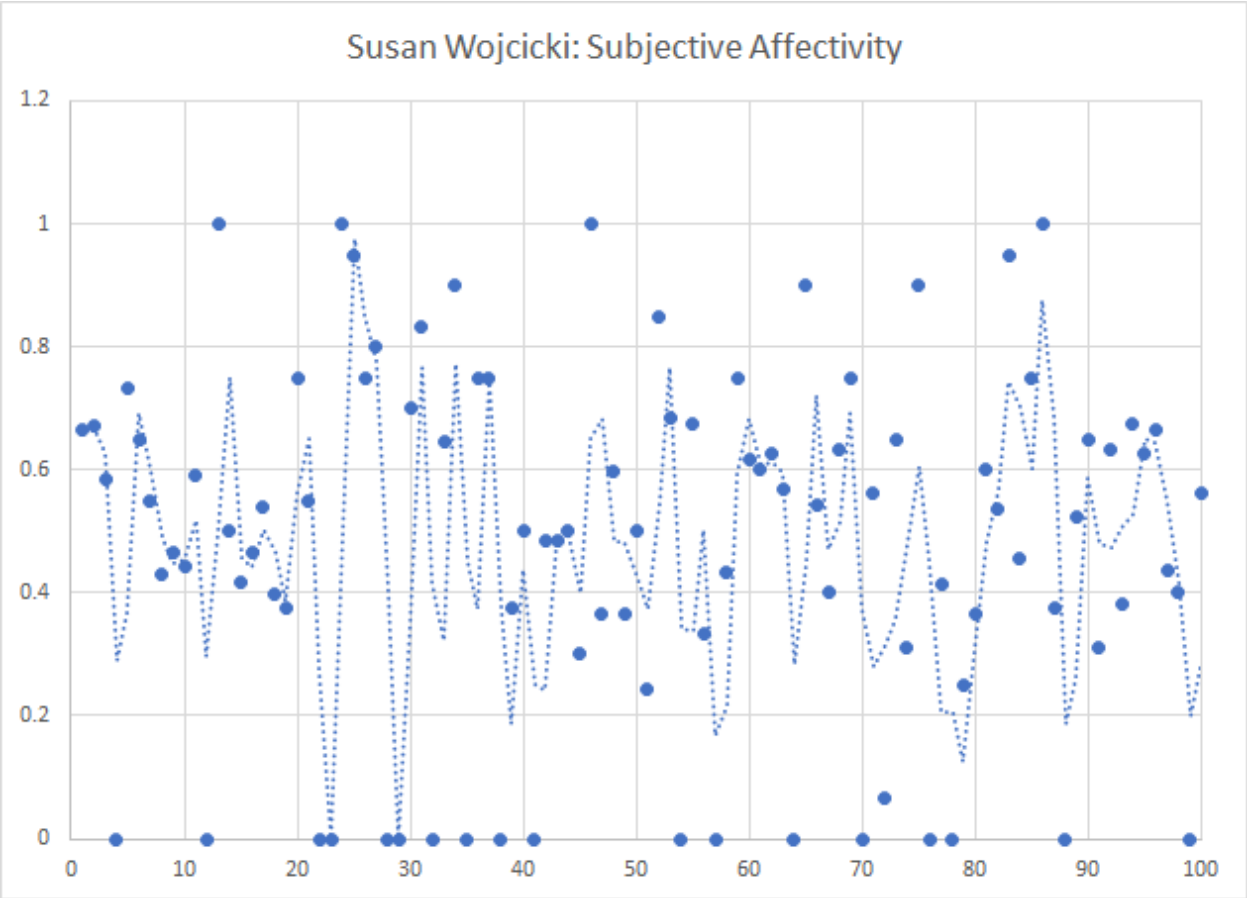

Figure 52 - Subjectivity Affectivity (Susan Wojcicki)

Susan's tweets are fairly subjective, but they earn high subjectivity scores for a particular reason. She shows a bias towards the most popular YouTube creators because of the revenue and traffic they gain for the platform. They are big names and since they have a lot of subscribers, they are the ones who have a lot of followers who will react positively to praise.

On the other hand, there is a lot of different points and graphs that spike at completely different times. There is not much suggesting that the time of day affects whether or not what she says is more subjective than others.

3.10.3 Time Frequency Analysis

The posting period that was collected with the data goes from October 9 2020 to March 8 2021. She primarily posts sometime in the early evening, with her latest tweet being posted at night.

Contrary to the popular assumptions, she did not receive higher interactions even that she manages the biggest video sharing site on the Internet. The only tweets that receive the most attention and interaction come from tweets relates to her acknowledgement or involve big YouTubers or musical acts. It is likely that the high engagement numbers that come from her tweets congratulating people on the platform comes from the people themselves.

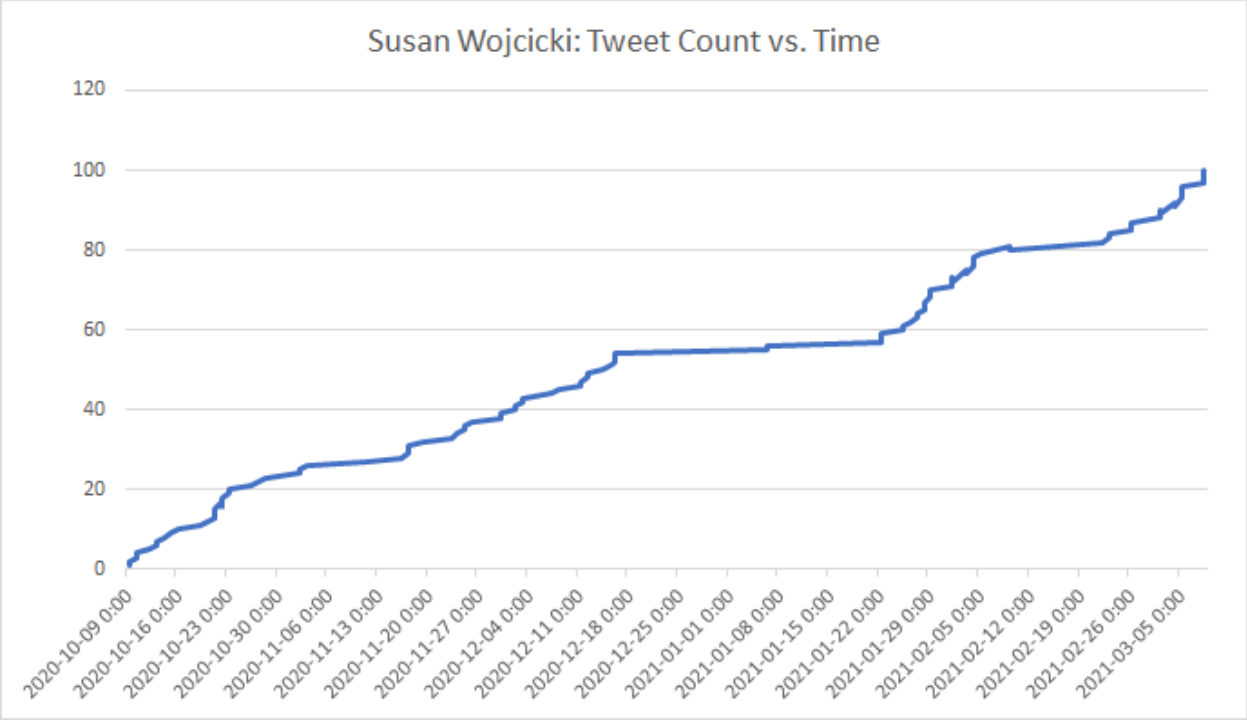

Figure 53 - Tweet Frequency by Time (Susan Wojcicki)

3.10.4 Tweet Classification Analysis

A lot of Susan Wojcicki's tweets are original posts she made herself. This is most common with the ones that prise YouTubers or advertise a watch event YouTube is hosting. There are not

that many retweets, although whenever she does it tends to be about YouTube and comes from the official account.

While many of her tweets are original, she also has a fair amount of tweets that are replies to someone. A lot of the people who she refers to tends to be towards people who are also big figures on YouTube. This can be not only YouTubers, but musicians, such as BLACKPINK, as well.

3.10.5 Frequency of Tweets Analysis

When analyzing Susan's tweet frequency, she has a tendency of posting quite a lot. The only exception was a period between December 24 2020 and January 31 2021. It is the only month that has a considerably less amount of tweets. Although considering it is usually the time for the holidays, it is inconclusive as to whether or not holiday is a factor.

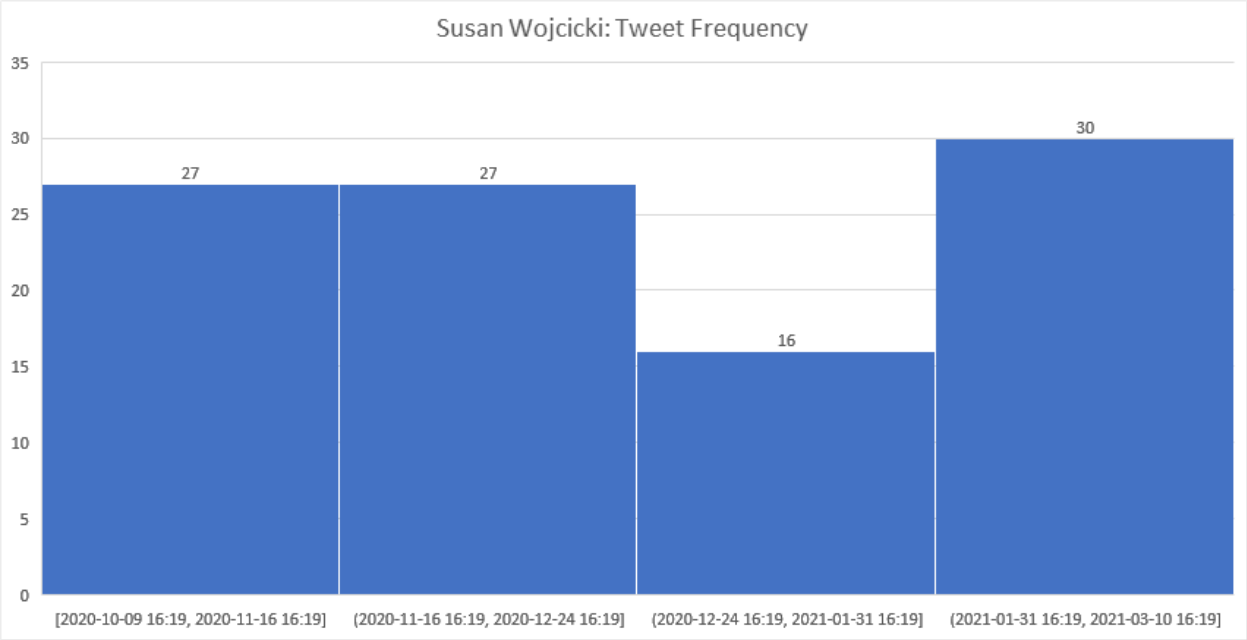

Figure 54 – Count of Tweets vs Time (Months) (Susan Wojcicki)

3.10.6 Reaction Classification Analysis

The reactions received on Susan Wojcicki's tweets fluctuate often. In total, from the 100 tweets that were collected from Susan's Twitter, there were 66,144 likes and 7800 retweets. As stated before, 0 is the minimum that one received. The most popular tweet received 12,731 likes. This tweet was wishing a happy birthday to three very popular YouTubers, who have a large following. It is likely that it was retweeted by one, two, or all three of them. Upon seeing this, many of the fans winded up liking the tweet because of the kind gesture from Susan. However, it was not the tweet that received the highest number of shares. Instead, that went to a tweet advertising an online concert that would be hosted on YouTube for the popular K-pop girl group, BLACKPINK.

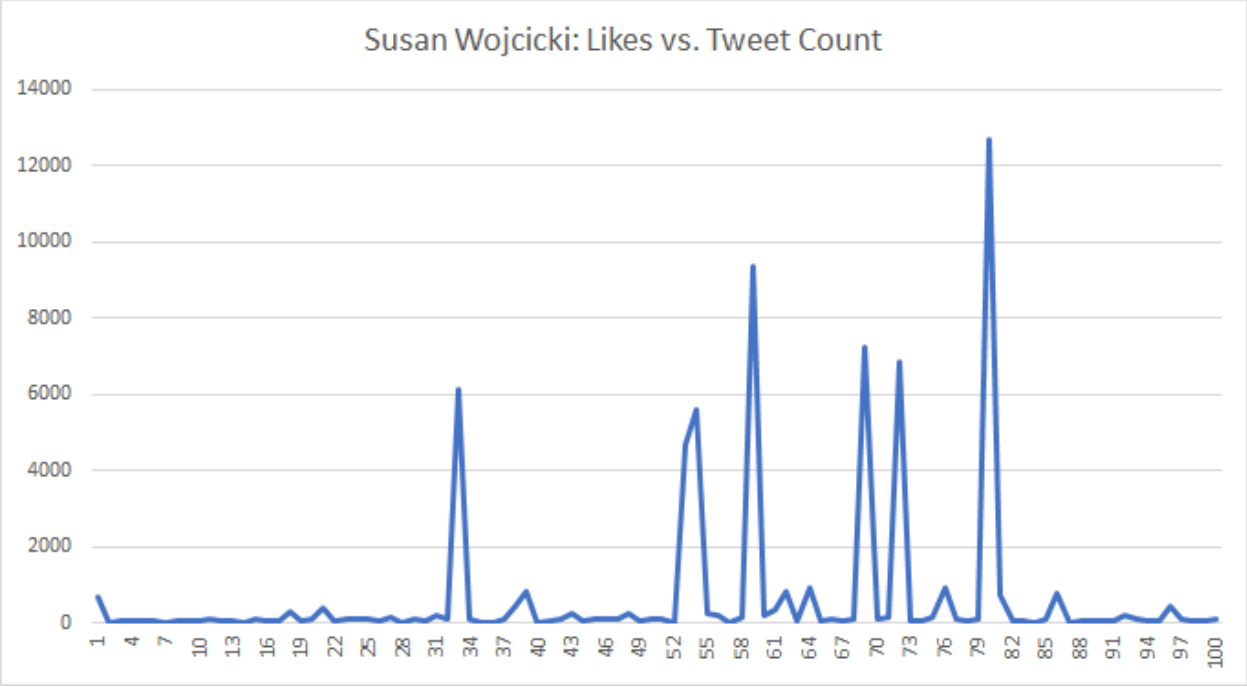

Figure 55 – Amount of Likes on Each Tweet (Susan Wojcicki)

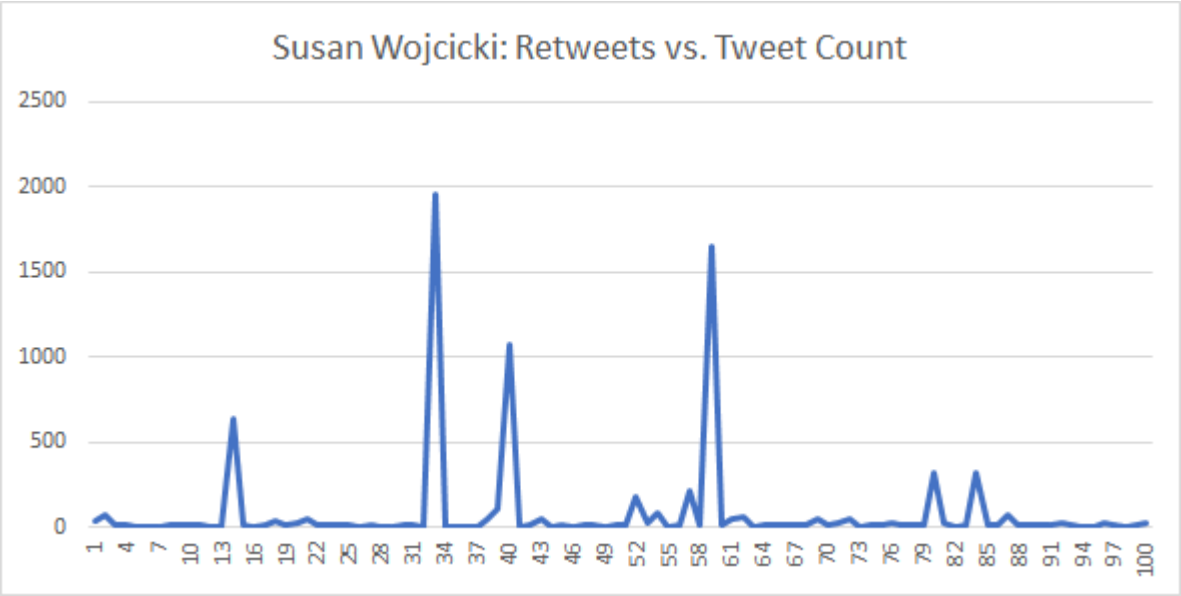

Figure 56 – Amount of Retweets on Each Tweet (Susan Wojcicki)

It is interesting to how much the numbers fluctuate for Susan Wojcicki. Perhaps this comes from the polarizing views that YouTube users have about her. It should also be noted that Susan Wojcicki is a contested figure among users on YouTube.

3.10.7 Word Frequency Analysis

The most frequent word on Susan's Twitter is "YouTube" as it appears 200 times. Whether she is speaking about creators (where that word was brought up 60 times) or about the website itself, it still makes sense that she could bring up the website itself.

Illustration 8 – Word Cloud for Susan Wojcicki's tweets

What is interesting to see is how many celebratory words are among her tweets, especially celebrating minorities or people of colour. Like other CEOs, words of gratitude such as "Thank" and "Congrats" are prominent themes in her text as well.

<u>3.10.8 Parts of Speech Analysis: Noun, Verbs and Adjective Count Analysis</u>

Between all of Susan Wojcicki's tweets, with all of the characters and words, there are 1070 nouns. This makes up for 54% of the parts of speech in Susan Wojcicki's tweets. The total number of verbs in Susan Wojcicki's tweets is 471. This makes up for 24% of the parts of speech in Susan Wojcicki's tweets. The total number of adjectives in Susan Wojcicki's tweets is 439. This makes up for 22% of the parts of speech in Susan Wojcicki's tweets. The number between the adjectives and the verbs are quite close. It is likely that there is not always the chance to be able to use descriptors.

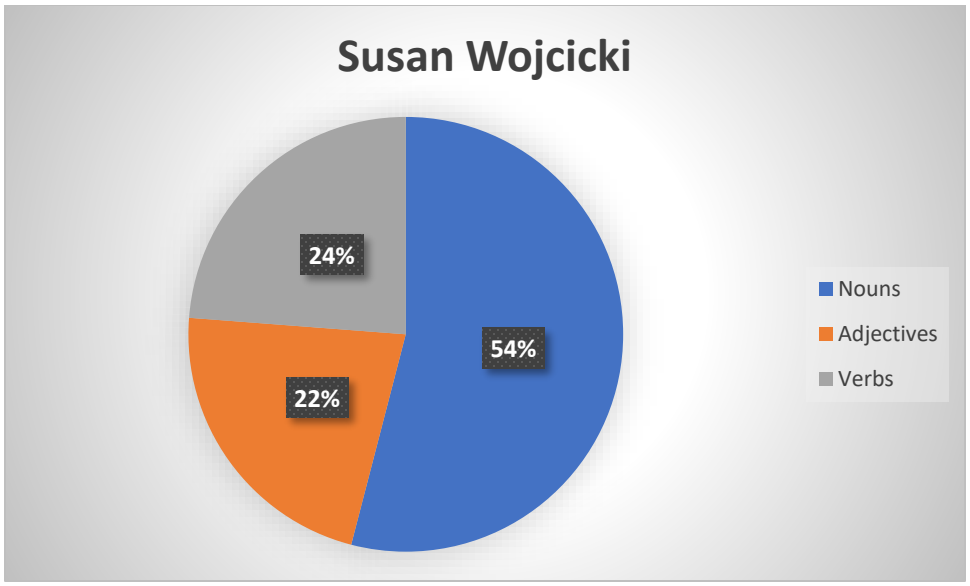

Figure 57 - Percentage of Text Consisting of Nouns, Adjectives and Verbs (Susan Wojcicki)

3.10.9 User Mentions Frequency Analysis

There are many user mentions that Susan Wojcicki engages in. A lot of them involve directly mentioning the Twitter user, rather than responding. Although these Twitter users also happen to be some of the bigger YouTubers at the time, or sharing the Twitter of the artist. Alternatively, one will also notice that another account Susan Wojcicki commonly mentions is the official YouTube Twitter itself. This is useful especially for people who think they can bring up their issues to Susan herself.

3.11 Case Study 9 – Tim Cook, CEO of Apple

This is the case study for Tim Cook, the CEO of Apple Inc,. In this section, we will be looking at 100 tweets collected from his Twitter account and see what the feature engineering picks up.

3.11.1 Sentiment Polarity Analysis

In the data for Tim Cook's tweets, 80% of the tweets have a positive sentiment, 13% of them have a neutral sentiment and 7% of them have a negative sentiment. There were no sudden

huge changes between Tim Cook's moods, and there does not seem to be a correlation between the polarity and the feeling.

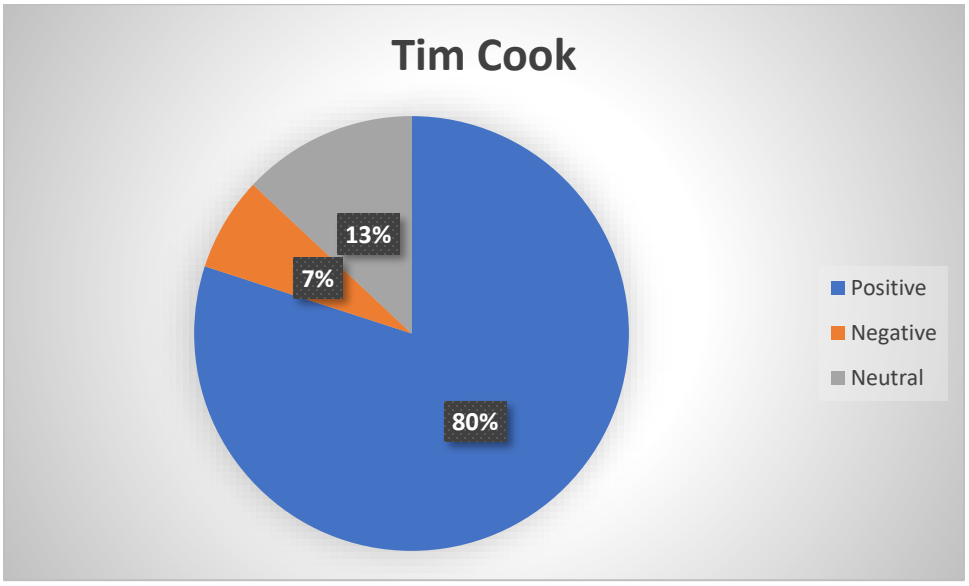

Figure 58 - Sentiment Analysis (Tim Cook)

Based on the time periods, the sentiment polarity did not show many major shifts between the hours.

3.11.2 Subjective Affectivity Analysis

The overall fluctuation of Tim Cook's subjective affectivity scores tends to be about average. When looking at the subjective affectivity, there are quite a few tweets that Tim Cook put out that have a subjective affectivity score of 1. This means they are either fully subjective, or they are simply images that he posted without a caption. As a result, that shows with the average mean score.

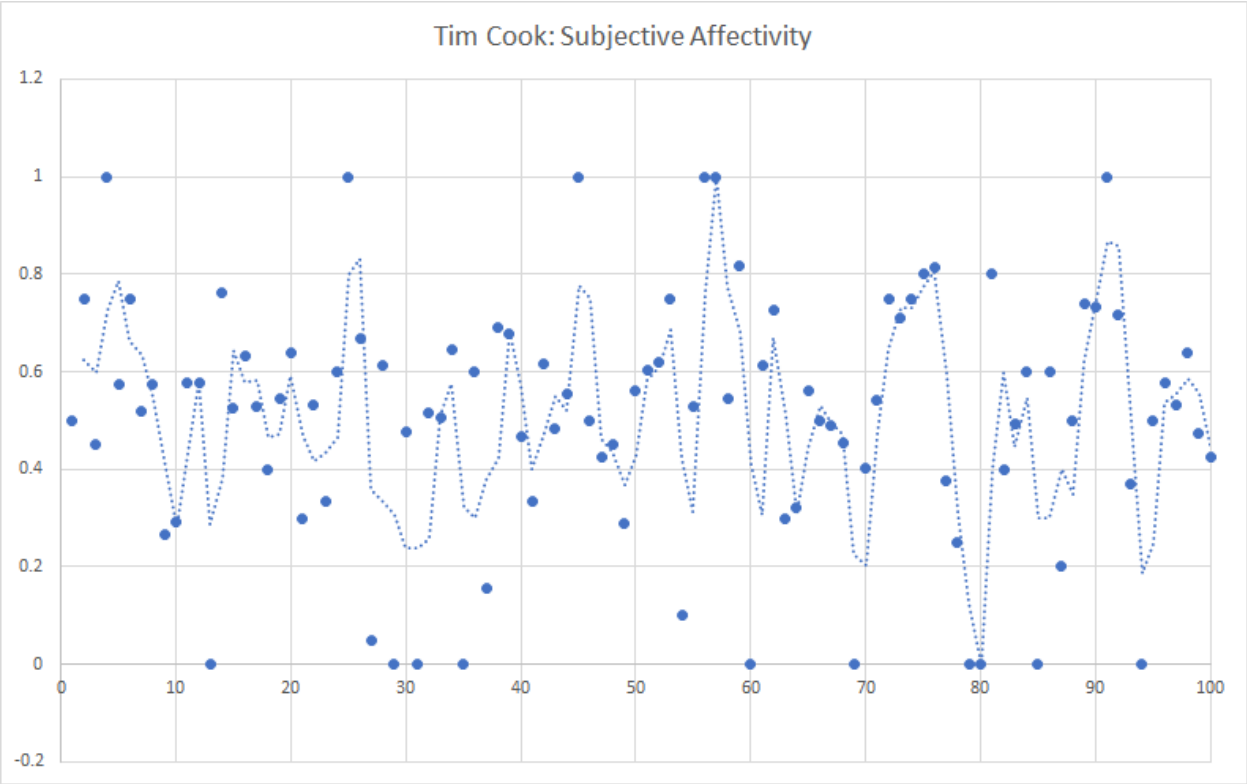

Figure 59 - Subjective Affectivity (Tim Cook)

When looking at the subjective affectivity score for Tim Cook, we see that there are a fair amount of posts that reach a subjectivity level of 0, making it hard to determine how subjective the tweet's content is.

3.11.3 Time Frequency Analysis

The time period of Tim Cook's tweets is between July 23 2020 to March 8 2021. It is similar to the time period for Sundar Pichai, although Tim Cook tweets much more frequently in comparison. The times that he posted the most frequently have been between the afternoon and the early evening.

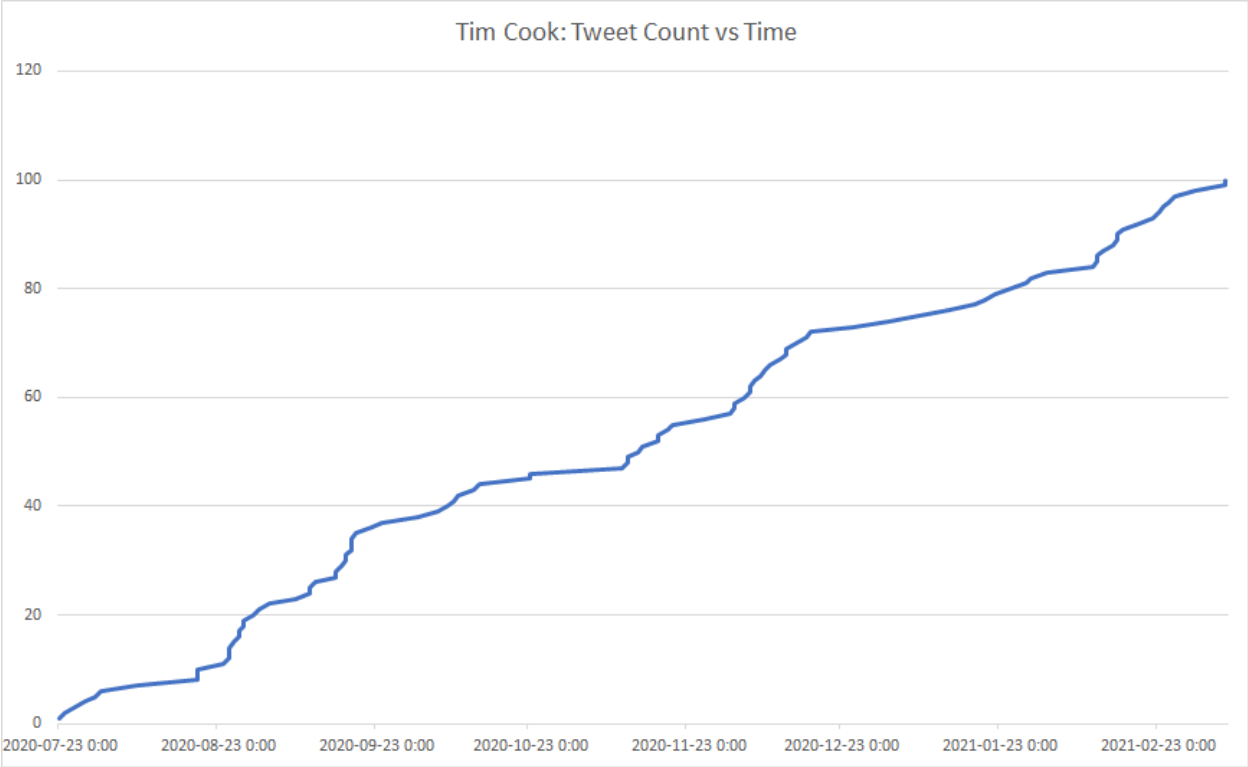

Figure 60 - Tweet Count by Time (Tim Cook)

3.11.4 Tweet Classification Analysis

Most of Tim Cook's tweets are all original tweets that he had written himself. Whether these be tweets that are replies, it is clear that he does not have much to retweet. Among all of his tweets, about 2% of them are retweets.

3.11.5 Frequency of Tweets Analysis

Not taking into account the posting period, it is apparent that Tim Cook posts fairly frequently in the time periods that are divided.

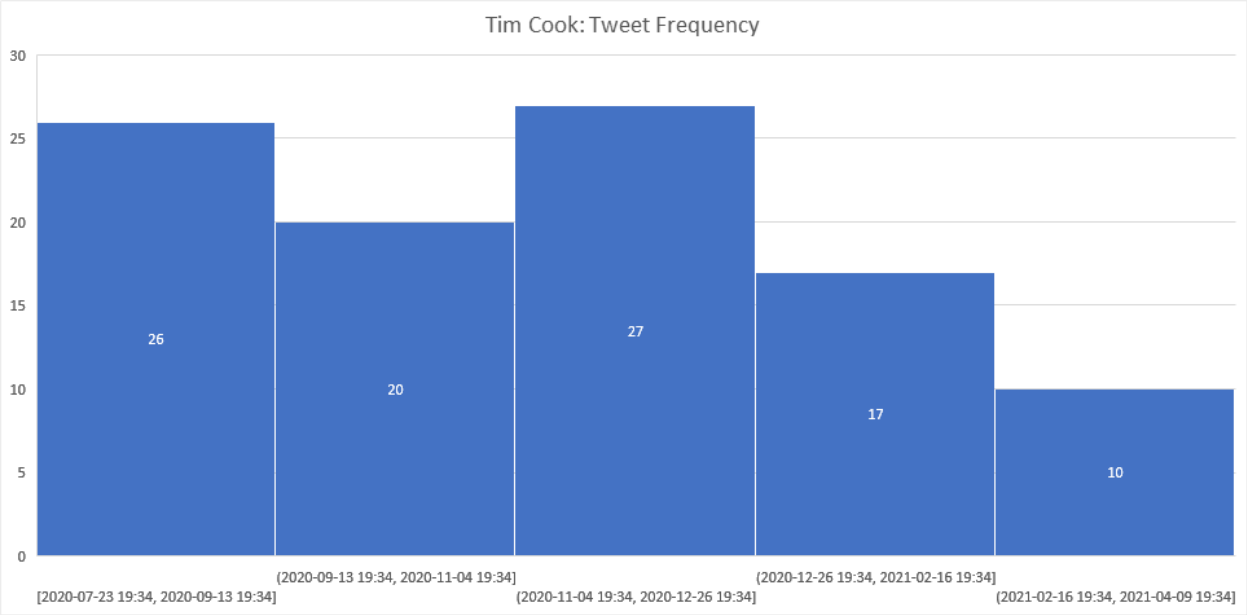

Figure 61 – Count of Tweets vs Time (Months) (Tim Cook)

3.11.6 Reaction Classification Analysis

Most of the interactions that appear for Tim Cook tend to gather into the thousands. There are very few exceptions, with some tweets not getting any likes or retweets. However, when they do get notes, many people seem to interact with them. In total, all 100 of the tweets have gathered 1,449,617 interactions. Out of these interactions, 1,303,354 are likes and 146,263 are retweets.

When taking a closer look at his tweets, the tweet that received the most likes and retweets is the same. The highest amount of likes and retweets is 113,888 likes and 17,060 retweets. The tweet in question is talking about how Tim Cook believes that users should be able to have a choice over their data. So the tweet was announcing how users will have a choice for app tracking rather than the feature being turned on by default. Since users can be quite protective of their data and how it is being used, it attracts a lot of attention especially from Apple consumers.

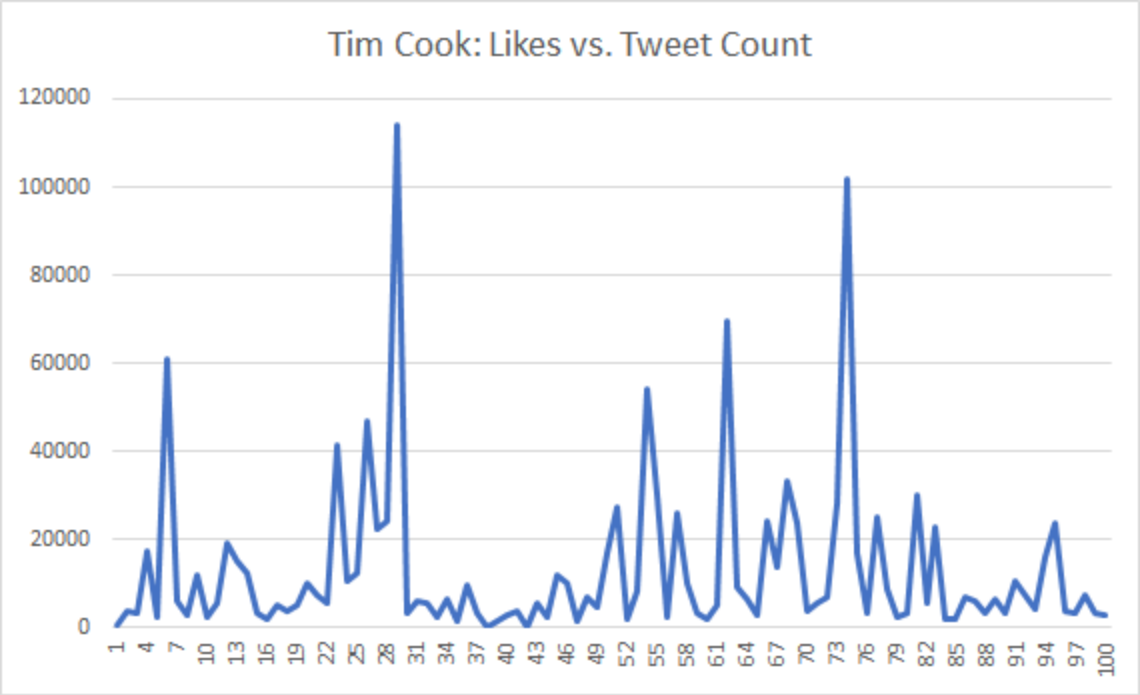

Figure 62 – Amount of Likes on Each Tweet (Tim Cook)

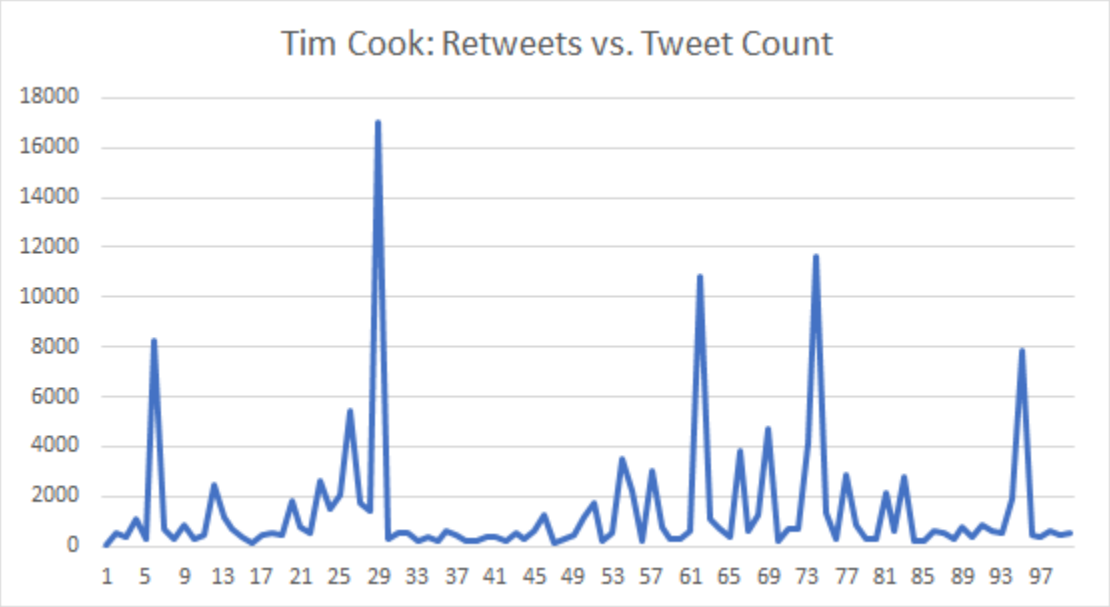

Figure 63 – Amount of Retweets on Each Tweet (Tim Cook)

We see a lot of random spikes from the amount of retweets and tweets that Tim Cook's Twitter receives. There is no consistency in the numbers, though we can easily establish that posts with lower amounts of interactions come from replies and the like.

3.11.7 Word Count Frequency Analysis

This word count frequency shows that the word that appears the most frequently for Tim Cook is “apple”. The case sensitivity shows that the word “Apple” primarily comes from the fact that it is the name of the company. Other words that have gathered a lot of instance of the words is ‘world’, ‘support, ‘health’ and ‘Thank’. For some reason, the singular letter ‘s’ has also garnered a lot of views on the Word Cloud.

Illustration 9 - Word Cloud Analysis of Tim Cook

3.11.8 Parts of Speech: Noun, Verb and Adjective Count Analysis

In terms of the numbers of nouns, there are a total of 1125 nouns. This makes up for 54% of the parts of speech in Tim Cook’s tweets. Among all of the words on Tim Cook’s tweets, there are a total of 479 verbs in total. This makes up for 23% of the parts of the speech in Tim Cook’s tweets. Finally, among all of the words in Tim Cook’s tweets, there are a total of 464 adjective in total. This makes up for 23% of the parts of the speech in Tim Cook’s tweets.

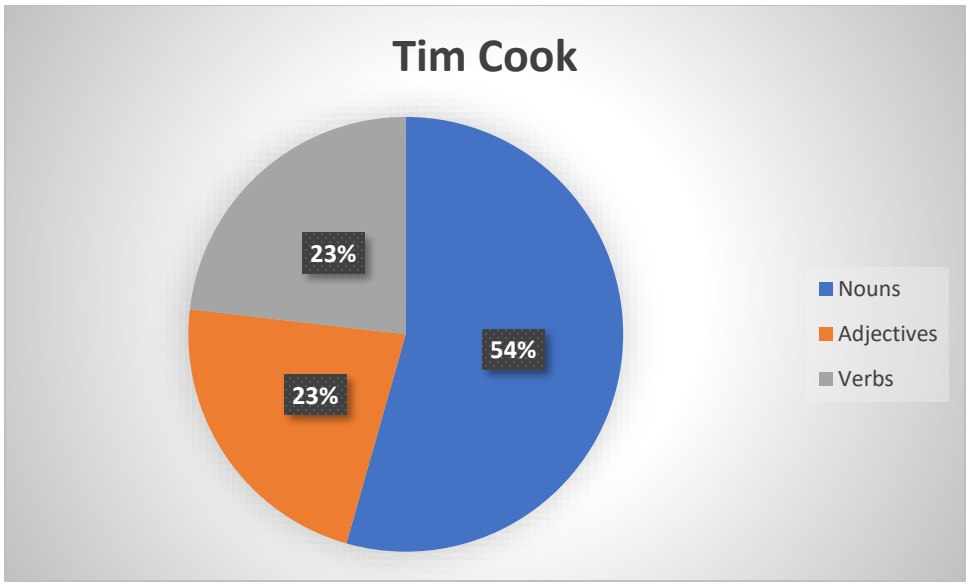

Figure 64 - Percentage of Text Consisting of Nouns, Adjectives and Verbs (Tim Cook)

3.11.9 User Mentions Frequency

Tim Cook's Twitter data has a fair amount of user mentions frequency. In total, he had done 39 total user mentions.

3.12 Case Study 10 – Whitney Wolfe Herd, CEO of Bumble

Whitney Wolfe Herd is the CEO of Bumble. In this section, we will be looking at 100 tweets collected from her Twitter account and look at the data when run through the feature engineering.

3.12.1 Sentiment Polarity Analysis

Based on the 100 tweets that were posted by Whitney Wolfe Herd, 67% of the tweets have a positive sentiment, 26% of the tweets have a neutral sentiment, and 7% of the tweets to have a negative sentiment. This is a fairly average score in terms of positive sentiments. It is interesting to see how many posts were defined as neutral on her Twitter in comparison to the others.

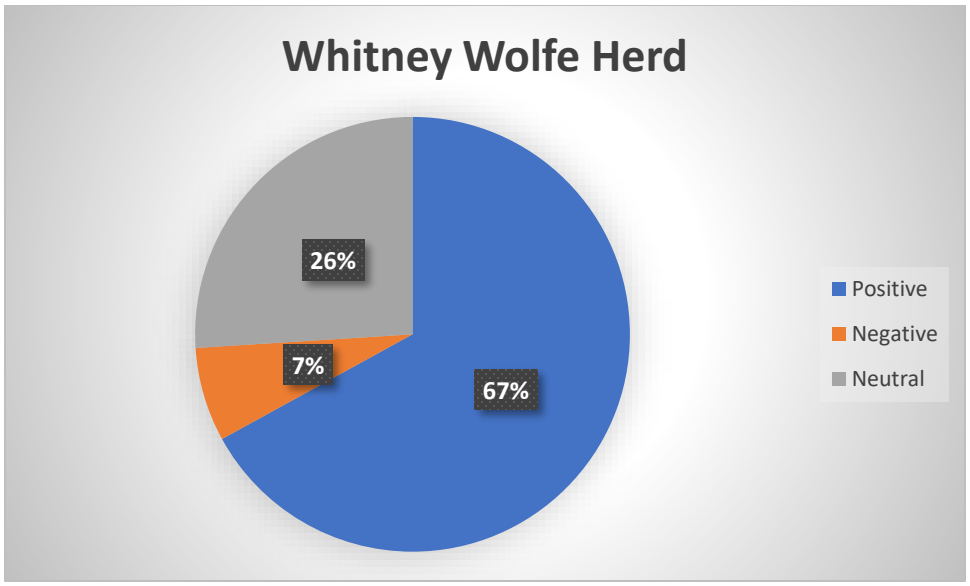

Figure 65 – Sentiment Analysis (Whitney Wolfe Herd)

This is one of the greater percentages of neutral tweets, especially in comparison to positive tweets. Some of these tweets are marked neutral for a good amount of her tweets talking about helping people. It should be noted, there are neutral tweets that were marked but should be marked more as positive. In this case, tweets that consist of a single hashtag #BelieveWomen, is in response to another case where abuse victims spoke out against their abusers. As a strong stand for supporting victims of intimate violence, machine learning tool was unable to build that connection. It is considered neutral in sentiment.

3.12.2 Subjective Affectivity Analysis

When analyzing Whitney Wolfe Herd's subjective affectivity, we can see that 17% of all of her tweets have a subjectivity score of 0. Meanwhile, 5% of them had a subjectivity score of 1. On average, the subjectivity of her tweets scored at 4.45.

The spread of subjectivity is relatively spread amongst her tweets, with the score overall being below average. Her tweets with a subjectivity score of 0, tend to be tweets relaying

information about important matters, such as sharing work information, politicians or the like. Rather than sharing her own viewpoint, it revolves more around how they are handling a situation.

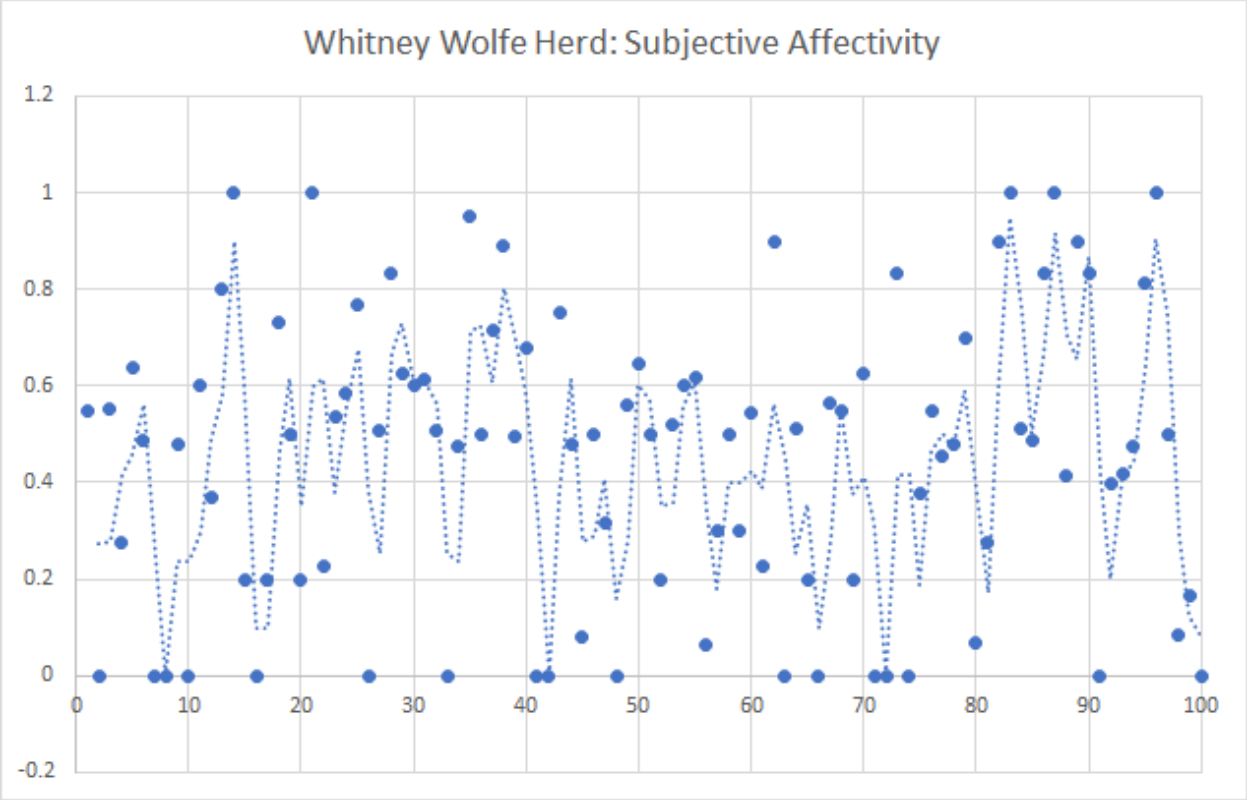

Figure 66 – Subjective Affectivity (Whitney Wolfe Herd)

3.12.3 Time Frequency Analysis

The frequency of tweets on Whitney Wolfe Herd's Twitter goes from August 24 2018 to March 1 2021. Just like Jeff Bezos, there is a lot of time that goes in between how often she posts on her Twitter account, although she has recent tweets in comparison. While there are a lot of tweets in the earlier months, it starts to wane as the time gets farther and farther.

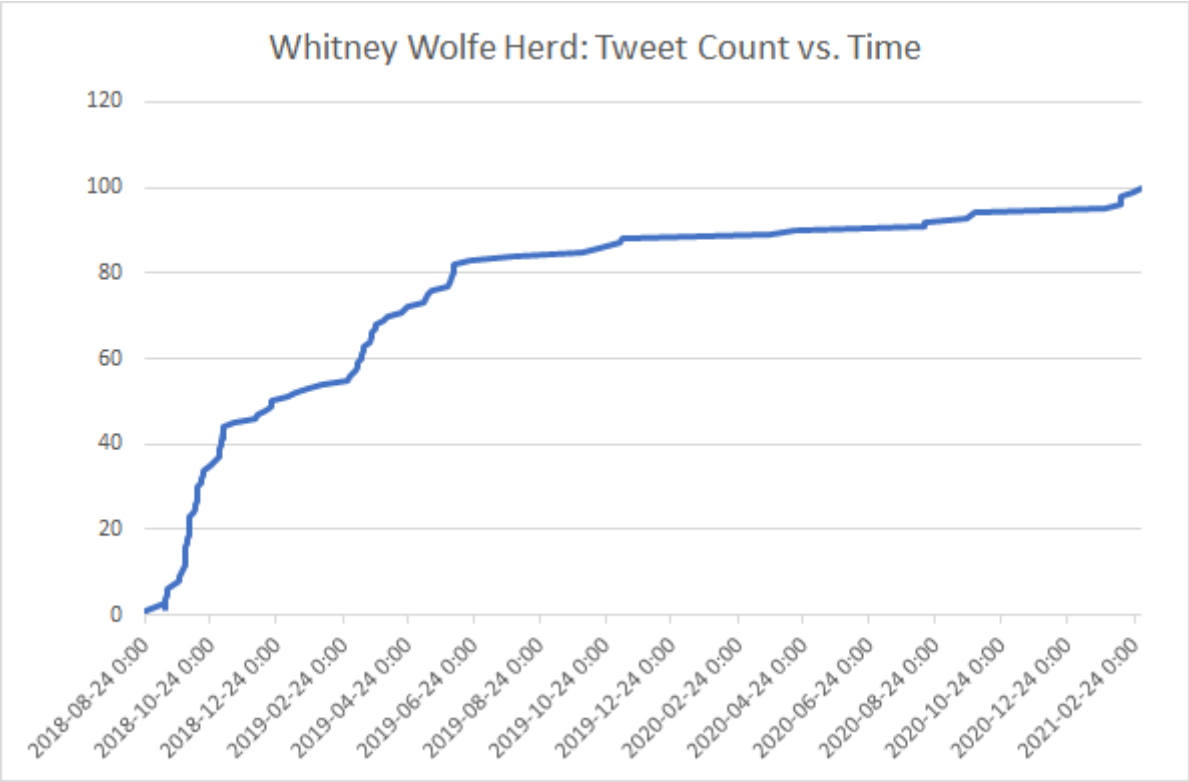

Figure 67 - Tweet Frequency by Time (Whitney Wolfe Herd)

According to the graph, her posting was consistent from August 2018 to October 2018. However, it starts to wane off with more of the months that pass. Starting from May 2019-June 2019, you can start to see how the posting patterns begin to stagnate.

3.12.4 Tweet Classification Analysis

From the 100 tweets, only a few of them are not retweets or responses of some extent. There are many user mentions overall which usually comes from replies and the like. Whitney Wolfe-Herd also retweeted a lot of articles, primarily ones that featured her as an interview subject. She would also retweet prominent female figures such as Serena Williams.

3.12.5 Frequency of Tweets Analysis

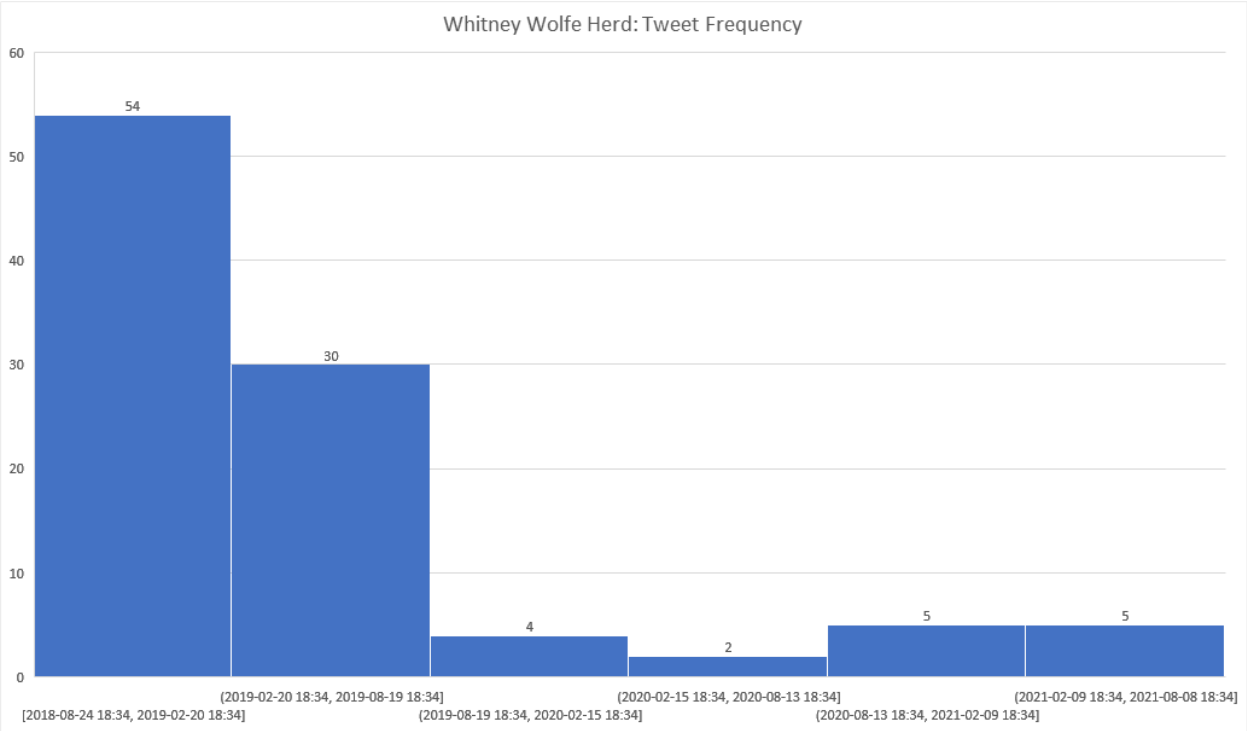

Figure 68 – Count of Tweets by Time (Months) (Whitney Wolfe Herd)

When dividing up the month periods for Whitney Wolfe Herd's tweets, they were split into periods of about six months in-between each. The tweet frequency was at its highest during the period of August 24 2018 to February 20 2019, where there were a total of 59 tweets posted overall. It takes a massive dip once we go into any of the other time periods.

Starting from August 19 2019 to February 15 2020, there are considerably less tweets being posted into these periods, showing a sudden drop.

3.12.6 Reaction Classification Analysis

Out of all of the tweets that are on Whitney Wolfe Herd's Twitter, the overall interactions reach up to 26,782 total interactions. Of these interactions, 4,549 of them are retweets and 22,233 of them are likes.

There are a lot of strange spikes in likes for Whitney Wolfe Herd's Twitter. The interactions that are involved are consistently low, only spiking near the beginning. The tweets that received

the most likes and retweets are both tweets that talk about the future of Bumble, the very company she deals with.

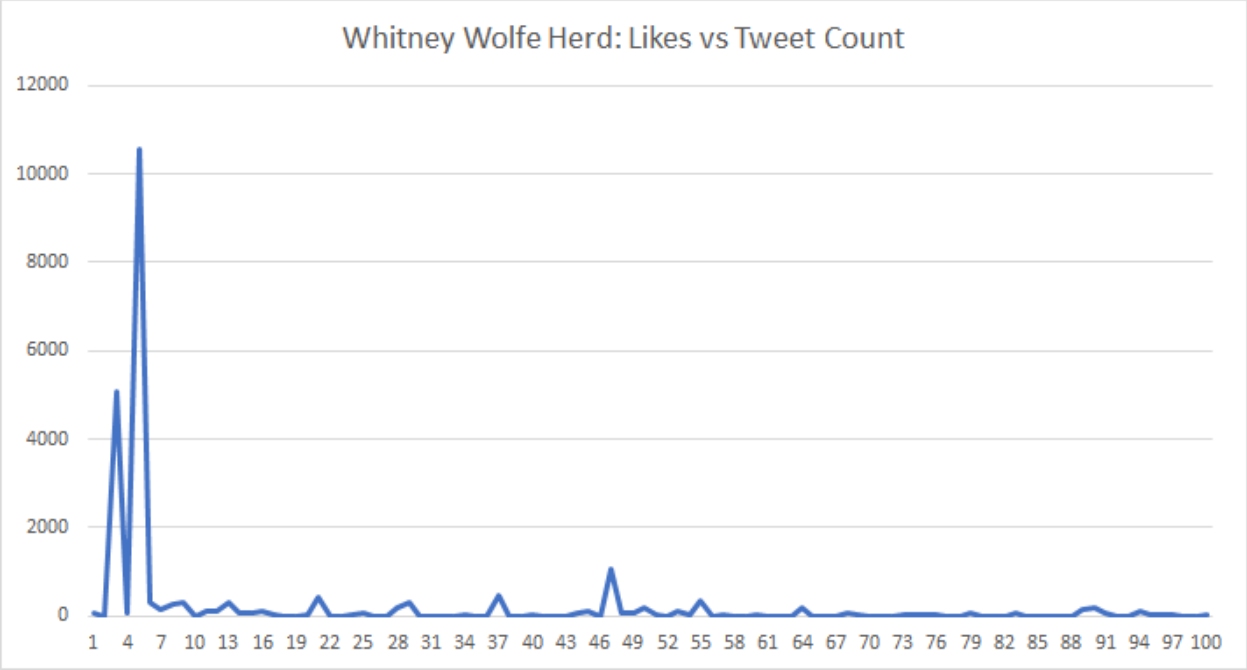

Figure 69 – Amount of Likes on Each Tweet (Whitney Wolfe Herd)

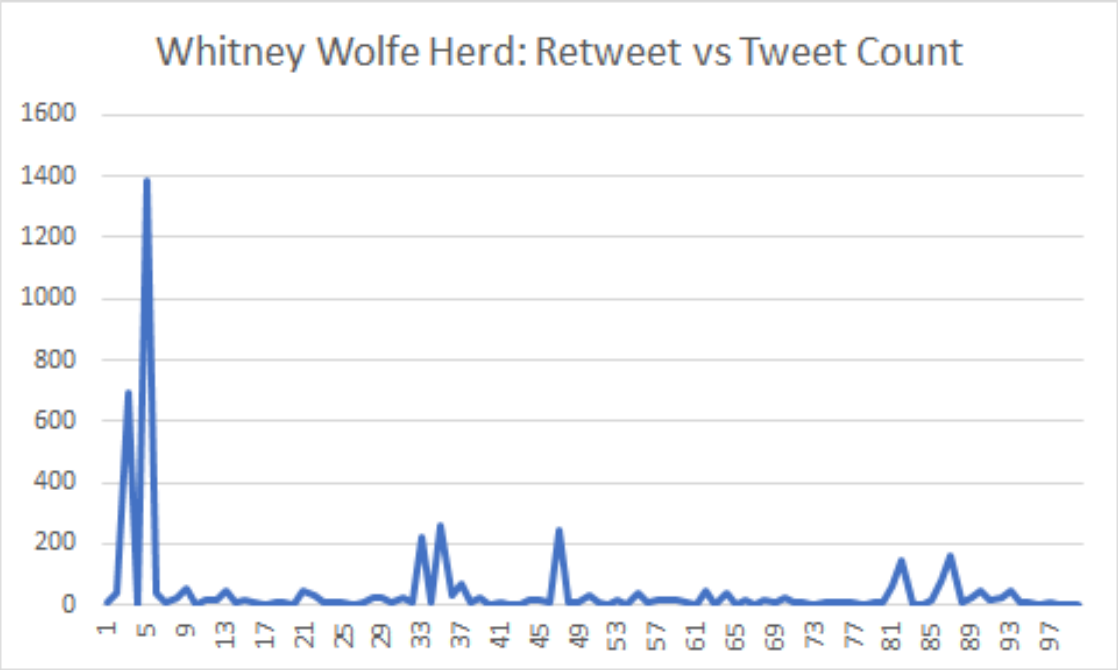

Figure 70 – Amount of Retweets on Each Tweet (Whitney Wolfe Herd)

There are a lot of retweets and quote retweets that appear on her Twitter. Most of them are tweets that talk about getting Bumble involved in other tech-based projects. As a result, she wants to share it on her own private Twitter account. This makes it hard to determine how many tweets can count as her own, as she shares many posts. It does explain why the consistency of the likes and retweets her posts receive are usually pretty low.

An interesting point is that, whenever Whitney Wolfe-Herd talks about social justice, the tweets receive slightly more replies. Couple of examples include her tweet about testifying in front of court for a bill to be passed that makes it illegal to send unsolicited nudes

3.12.7 Word Frequency Analysis

From all of the words that appear on Whitney Wolfe Herd's Twitter, the word that appears the most is "women". A look at the word cloud below shows the lexical richness of the text is widely varied. Another word that appears the most often is "Thank".

Illustration 10 - Word Cloud for Whitney Wolfe Herd

This Word Cloud properly reflects Whitney Wolfe Herd's tweets in a very good and succinct manner. It describes that she is very progressive and is all about encouraging women to be able to do their best. Business is an important topic as well.

3.12.8 Parts of Speech: Noun Verbs and Adjective Count Analysis

The total number of nouns among the tweets is 542. This makes up for 56% of the parts of speech in Whitney Wolfe Herd's tweets. A lot of it comes from the fact that there is a high word count among Whitney Wolfe Herd's Twitter account. The verbs count is 226. This makes up for 23% of the parts of the speech in her tweets. The total number of adjectives among Whitney Wolfe Herd's tweets is 204. This makes up for 21% of the parts of the speech in her tweets.

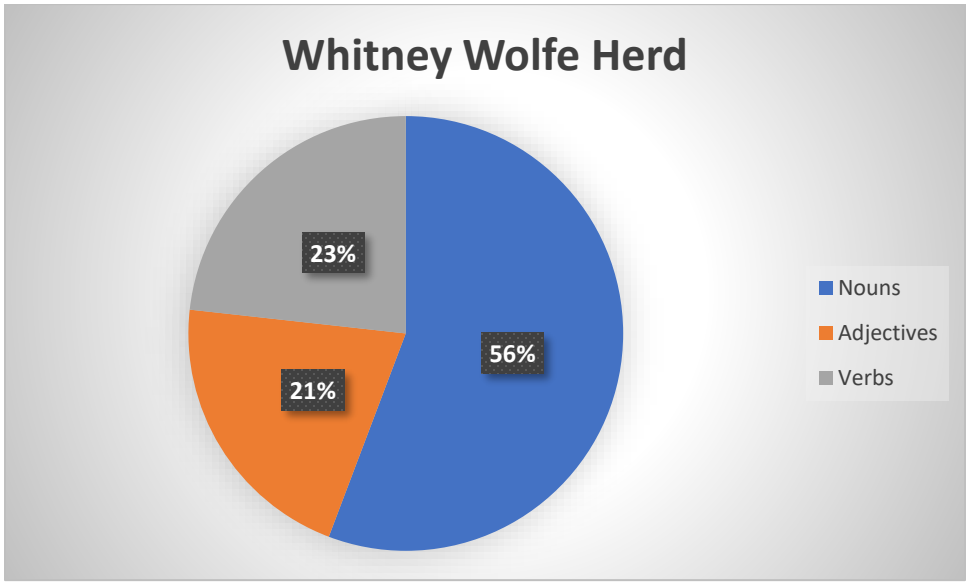

Figure 71 - Percentage of Text Consisting of Nouns, Adjectives and Verbs (Whitney Wolfe Herd)

3.12.9 User Mentions Frequency Analysis

About 7% of Whitney Wolfe Herd's tweets involve her mentioning someone, whether it be in a response or as a mention in the body of the tweet. Out of all of the people that she mentioned, the user that she references the most often is the official Bumble twitter account. This makes sense, as the CEO of Bumble, she is more concerned about focusing on mentioning the more relevant people. Whether this be the official account for Bumble, or people who she had spoken to and learned something interesting about.

Chapter 4 - Observations and Comparisons

Our case study was aiming to find out the content characteristics of the tweets and any preliminary evidence as to the effect of content characteristics on user engagement. We focus on the characteristics such as sentiment, word frequency, pattern of speech, and posting habits such as time of the day, frequency and types of tweets. Some individual differences in terms of their content characteristics are unveiled. For example all 10 CEOs tweets accounts contain information that relate to major events or business information that matters to the everyday life. Most of them have positive sentiment in their tweets. But some have a more even distribution of positive, negative and neutral sentiment. Most CEOs have high scores in terms of pattern of speech meaning their thoughts are expressed with complete sentences and good grammatical structure. However some do not. The following sections present the characteristics of the tweet contents and comparison of results from individual case study.

4.1 Comparisons for Sentiment Polarity

Among all 938 tweets consists of (691/938) of an average of 74% of them being positive, (178/938) 19% of them being neutral and 7% (69/938) of them being negative. This shows that most of the CEOs are likely to post more tweets that lean towards positive sentiments.

Table 5 – Comparison for Sentiment Polarity (%)

CEO	Positive	Negative	Neutral
Bill Gates	0.77	0.11	0.12
Elon Musk	0.44	0.10	0.46
Jeff Bezos	0.75	0.06	0.19
Corie Barry	0.79	0.06	0.15
Ginni Rometty	0.63	0.2	0.18
Lisa Su	0.97	0	0.03
Sundar Pichai	0.78	0.05	0.17
Susan Wojcicki	0.70	0.10	0.20
Tim Cook	0.80	0.07	0.13
Whitney Wolfe Herd	0.67	0.07	0.26

Lisa Su's tweets are extremely positive with a score of 97%. For Ellen Musk's only 44% of his tweets are positive while 46% are neutral sentiment. From our individual case study, however, both CEOs receive high number of likes and retweets.

4.2 Comparisons for Subjectivity Assessment

The content of CEO tweets have a fairly high subjectivity score. As a result, adding up all of the subjective assessment scores from all 10 of the CEOs, the range for the averages is between 0.3 and 0.6. The subjective affectivity averages for all the CEOs is shown in Table 6.

Table 6 – Subjectivity Assessment Averages for all CEOs

Subjective Affectivity	Averages
Bill Gates	0.4419
Elon Musk	0.31138
Jeff Bezos	0.45681
Corie Barry	0.54681
Ginni Rometty	0.39321
Lisa Su	0.54867
Sundar Pichai	0.46605
Susan Wojcicki	0.480
Tim Cook	0.49669
Whitney Wolfe-Herd	0.44818

Most of the time, when there was a subjective affectivity score of 1, it had a correlation to instances where very few or a lot of adjectives were used. One of the biggest problems is trying to figure out what words the code considers 'subjective' and how it is calculated.

For the subjectivity scores of 0, it could mean several reasons. The reason to why there was a subjectivity score of 0 all varied. For some, there are images not text in the content. The program cannot determine the subjectivity of an image. Hence the 0. However, it is worth noting that images do not always have to be visual. Some studies found posting image could bypass the Twitter character limit and avoid a long message to keep people's attention [4][11].

4.3 Comparisons for Time Frequency

Most of the 938 tweets were posted during the hours of early or late evening. The evening does tend to be a time when a lot of people are active, particularly users in younger demographics [9] [11] [19]. Afternoon appears to be the second popular time for CEOs to tweet. One exception is Elon Musk, who has a tendency of posting whenever he wants no matter the time of day.

4.4 Comparisons for Tweet Classification

Out of all of the tweets in our dataset, most of them were original tweets written by the content creators. Table 7 shows the overall distribution of the number of all original tweets, replies and retweets.

Table 7 – Overall Tweet Classification Numbers

Type of Tweet	Number of Tweet Types
Original Tweet	637
Retweet	119
Reply	183

By individual case study, Sundar Pichai is the only one who had more replies or retweets. It should also be noted that Elon Musk has retweets and the like among his tweets. Retweets affect our data, as they make it that the number of likes do not show up.

4.5 Comparisons for Frequency of Tweets

The period we include in our research is 2018-2021. During this time Jeff Bezos stopped tweeting in 2020 while all of Corie Barry's tweets only in 2019. Ginni Rometty's account was closed in January 2021. Elon Musk's posting period is entirely in 2021. When checking the tweets themselves, he posts quite frequently. There are certain days where he goes without posting a

single thing. However, when he does post, he posts an average of five posts a day. Whether this is related to cryptocurrencies, sharing funny images, Tesla Industries or talking about his rockets or family. He is the most frequently tweeting CEO in our selected sample of CEOs. Other CEOs do not post frequently like him. Rather, their posting frequency is sporadic.

4.6 Comparisons for Reaction Classification

A lot of the tweets tend to have a large amount of likes, usually in spikes. Not many of them tend to have tweets that consistently receive a large amount of likes or retweets. There are many different spikes of interactions, whether they be likes or dislikes, to specific tweets at certain periods of time. In total, there have been 12,024,209 likes and 1,261,132 retweets.

Table 8 – Total Number of Likes and Retweets

CEOs	Total Likes	Total RTs	Max Likes	Max RTs
Bill Gates	549,987	74,565	87,923	20,494
Elon Musk	7,694,511	599,036	579,050	70,187
Jeff Bezos	1,222,521	203,572	101,606	24,797
Corie Barry	2225	128	126	11
Ginni Rometty	6601	881	2645	251
Lisa Su	186,334	22,956	10,151	2389
Sundar Pichai	970,299	201,382	316,512	59,532
Susan Wojcicki	66,144	7800	12,713	1958
Tim Cook	1,303,354	146,263	113,888	17,060
Whitney Wolfe Herd	22,233	4549	10,561	1384
Sums	12,024,209	1,261,132		

Previous studies found users reacted more positively towards tweets that were more humorous and can be easily shared or they acknowledged groups with large followings. This primarily shows with Elon Musk and Susan Wojcicki, whose most popular tweets were either a

funny meme, or from acknowledging popular figure heads on YouTube. Most of the tweets on smaller accounts with a focus primarily on business matters and congratulating workers behind the scenes, do not nearly get as many interactions.

Based on the content characteristics identified in the case study, amusing meme/image significantly affect user engagement. It is no surprise that Elon Musk's most popular tweet consists of an amusing image. It goes into theories that images create a hedonic motivation to motivate users to interaction by likes or retweets [4] [9].

However, the traits of the image itself makes sense to why this, out of all of the memes that Elon Musk has posted, received the most attention. It is a fairly generic and innocuous image that does not require a lot of thinking. Even if it is posted mostly for his followers, when those followers share the image on their own Twitters by retweeting, people outside of the inner circle can understand it easily. As studies show [9] [11] [12] [16] accessibility is a key for high engagement as the attention span in today's information consumption is short. The amusement and joyful function of Elon Musk's tweets seems to relate to his high level of engagement.

4.7 Comparisons for Lexical Richness – Word Clouds

The Word Clouds for each CEO's account identified range of "words" that appear more frequently. Each of them tend to focus on mostly positive words or using words that represent camaraderie (words such as "friend" or "partner"). More times than not though, it is usually a word that involves their company (Tesla for Elon Musk, YouTube for Susan Wojcicki) or praising the workers.

In addition, "Thank" followed by the word "you" and other words of gratitude appeared frequently on CEOs word cloud. These are found, especially in replies or commented tweets to

show the tweet that involves working together, or gaining praise from others. It is notable to see that there are quite a few words that involve social justice or social change.

4.8 Comparisons for Parts of Speech – Noun, Verb and Adjective Counts

It is noticed that with the exception of Elon Musk, the CEOs used high number of nouns, verbs and adjectives as shown in Table 9.

Table 9 - Parts of Speech

CEOs	Nouns	Verbs	Adjectives
Bill Gates	1068	474	460
Elon Musk	89	40	43
Jeff Bezos	659	302	264
Corie Barry	445	201	187
Ginni Rometty	184	79	79
Lisa Su	520	194	194
Sundar Pichai	1184	489	459
Susan Wojcicki	1070	471	439
Tim Cook	1125	479	464
Whitney Wolfe Herd	542	226	204

With the exception of Elon Musk, the CEOs used high number of nouns, verbs and adjectives indicating their thoughts and opinions are expressed with sentences with good grammatical structures. Some research found the use of nouns, verbs and adjectives ensure faster sharing of tweets through retweets [36]. However, this does not explain Elon Musk whose most liked and retweeted content is a one word post – “Literally!”

4.9 Comparisons for User Mentions Frequency

The frequency of user mentions tends to vary. While we exclude instances where the user responds to themselves (usually because they are adding more about a thought on an existing tweet), there is not a single CEO account that does not mention a user in some way.

Table 10 – Comparison for User Mentions

Name of CEO	Number of User Mentions
Bill Gates	52
Elon Musk	99
Jeff Bezos	57
Corie Barry	114
Ginny Rometty	41
Lisa Su	222
Sundar Pichai	56
Susan Wojcicki	141
Tim Cook	39
Whitney Wolfe-Herd	83

Most of the time, they are defined as retweets from a post that they had done. For our study, user mentioned are expressed when they mention a specific individual. User mention is expressed as @ certain individual.

4.10 Results from Case Study

The results of the case study gave us the raw numbers for the content characteristics we had the code extract for the database. We collected and organized the numbers into categories, found the means for most of them, and then compared the raw numbers and averages of the CEOs to each other. All of this was to determine what the case study can possibly tell us about the effect content characteristics have to the user engagement.

We have summarized our findings as follows:

(A) **Topic of the tweet**: In most cases, the CEOs provide information about a major event, such as Covid-19, climate change or update about their companies; in smaller cases the CEO tweets provide the value of entertainment and humour. The words "Thank" and "you" are most likely to appear in the CEOs word clouds expressing "gratitude" and "appreciation".

(B) **Positive sentiment**: The CEO tweets were categorized as being predominantly positive all around the board – with Lisa Su having the most positive sentiment tweets. While Elon Musk seems like an exception, the percentage of his tweets classified as neutral and positive are very close, with neutral barely edging out ahead. Even still, there could be more or less tweets with positive sentiments, as the sentiment analysis part of our code did not seem to properly identify all the tweets correctly. Thanks to some words having negative connotations and the code does not consider overall context, it deduces that those

(C) **Pattern of speech**: The CEOs keep professional presence and expressed their thoughts/opinions with a balanced sentence structure.

(D) **Time and posting frequency**: The CEOs post their tweets mostly in the evening followed by the afternoon. The patterns and frequency of the tweets varied, with two CEOs, being Corie Barry and Jeff Bezos, having not posted for a long time when we collected the data. The most egregious example is Elon Musk who posts very frequently, and therefore, has the most recent tweets.

(E) **Tweet Classification**: Most of the tweets in our database were original tweets, however, there were quite a few retweets as well. They might have affected the numbers for user reactions, as likes pop up as zero and retweets remain as the number of the original tweet. However, there was not enough of them to find a particularly significant effect.

(F) **User reactions**: The tweets of the male CEOs received a large number of likes overall. In comparison to the female CEOs, the number of likes overshadow them significantly. The female CEO who had the highest amount of likes on a tweet, is only comparable to the male CEO who had the least amount of likes on a tweet (See Table 11). There is no clear explanation to why – as it can pertain to a confounding variable or the content characteristics of the overall tweets.

Although our case study aimed to identify preliminary evidence as to the effect of content characteristics on user engagement, the results are not conclusive. For example, some of our observations appear to contradict widely found relationship between positive sentiment and positive effect on user engagement [1][3][4][6][8][9][11]. It is noted that Elon Musk, who received the highest amount of likes and retweets, has tweets that display neutral sentiment and do not contain large number of nouns and adjectives. The following table, Table 11, provides insight and raw numbers about the relationship between content characteristics and level of engagement. The words that have been bolded are meant to signify the topic or hashtag in the tweet.

Table 11– Tweet of CEOs that received the most likes and retweets

CEO	Tweet (w/ highest amount of Likes and Retweets)	Likes	RTs
Bill Gates	It's great to see India in scientific innovation and **vaccine** manufacturing capability as the world works to end the **COVID-19** world pandemic	87,923	20,494
Elon Musk	Literally	579,050	70,187
Jeff Bezos	I just took a DNA test turns out I'm 100% **Lizzo'**s biggest fan	101,606	11,854
Corie Barry	Somehow, I go home more energized and fulfilled after a FULL week in Vegas getting ready for holiday. Thank you, **#BestBuyFamily**, for all you do. I love being on the team with you.	126	3

Ginni Rometty	Almost 40 years ago, I came to **@IBM** determined to make an impact on the world. As I step down from my role as executive chairman, I know that **IBM** will still be part of my next chapter - it will live on in the values and purpose it has instilled in me.	2645	251
Lisa Su	Welcome to the world, @AMDRyzen 5000 Series, our first processors with the amazing new "Zen 3" architecture. So proud of the **@AMD** global engineering team. We love **gamers**! **#GameOnAMD**	10,151	1,658
Sundar Pichai	One of the greatest test series wins ever. Congrats India and well played Australia, what a series **#INDvsAUS**	316,512	41,416
Susan Wojcicki	Happy birthday, **@teamtreesofficl**! **@MrBeastYT** and @MarkRober, thank you for showing us that big things can happen when the **YouTube** community comes together. 22 million trees and counting - I can't wait to see what you'll do next!	12,713	318
Tim Cook	We believe users should have the choice over the data that is being collected about them and how it's used. **Facebook** can continue to track users across apps and websites as before, App Tracking Transparency in **iOS 14** will just require that they ask for your permission first.	17,060	113,888
Whitney Wolfe-Herd	Today, **@Bumble** becomes a public company. This is only possible thanks to the more than 1.7 billion first moves made by brave women on our app - and the pioneering women on our app - and the pioneering women who paved the way for us in the business world. To everyone who made today possible: Thank you. **#BumbleIPO**	10,561	1,384

Overall, the case study answers our first research question. However, it also calls for further investigation on what motivates user engagement. The case study tells us that a multi-method, dimensional conceptualization of emotional responses should be made to create a more elaborate understanding of the relationship between the various characteristics. This is needed to answer the rest of the research questions posed in this thesis.

Chapter 5 – Design and Implementation of a Cognitive Inspired Machine Learning Solution

This chapter describes a novel solution to further understand whether factors embedded in the tweets that could affect the liking and retweeting behaviours. The solution is inspired by the principles of the Cognitive Behaviour Therapy (CBT) which emphasized the interconnection of thoughts, emotions and behaviour. The recognition guided this research to extend the feature engineering to create a more elaborate understanding of user to user engagement. To implement the proposed solution, a feature engineering architecture is employed. The implementation steps are explained under three layers: thoughts, emotions and behaviour.

5.1 Design of the Cognitive Inspired Solution

The research solution design for this study is inspired by the cognitive behaviour triangle which explores the interconnections of thoughts, feelings and behaviour. Developed by Beck in 1964, the cognitive triangle model hypothesises that "people's emotions and behaviours are influenced by their perceptions of events"[37]. How people perceive a situation leads to emotion/feelings and that leads to behaviour. In Cognitive Behaviour Therapy, the cognitive triangle framework is used to design technics to care for people with common mental health disorders by helping them to develop more adaptive cognitions and behaviours [27]. This theory has also inspired studies of human behaviours in economics where human decision making occurs under uncertainty [26]. Cognitive inspired models arise in AI applications. They can also be used to build artificial agents such as industrial chatbots that are better aligned with human cognitive mechanism such as preferences and association [7]. The figure below illustrates the cognitive triangle and elements in each bubble:

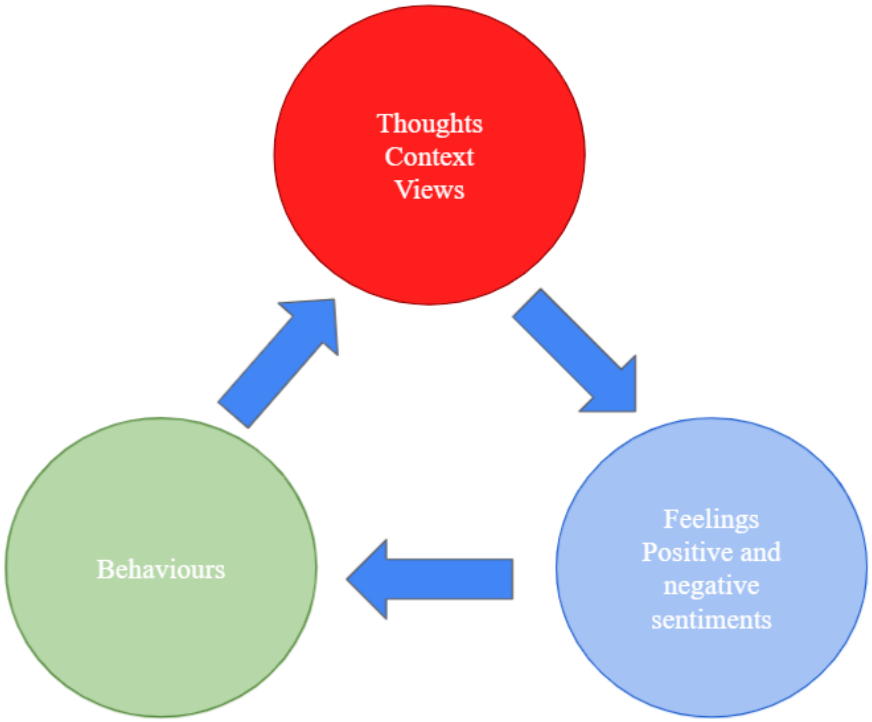

Illustration 11- Generic cognitive model based off Beck's negative cognition triad

In this illustration, it is suggested that when users scroll through social media – such as reading tweets - they give weight to the message with a cognitive perception and thoughts. The perception and thoughts generate feelings/emotions in the users. As a response to the emotions, users take actions or behave accordingly. Embedding this cognitive behaviour inspired model within the context of social media analysis could provide novel insights on the effect of emotional responses, social and cognitive relationships on user-to-user engagement, thus expanding the understanding of the factors influencing the behaviours of likes and sharing.

5.1.1 Why do we need a cognitive model?

Our case study evaluation identified several features that reveal the content characteristics of the selected tweets in our dataset. However, the results from the case study are not conclusive about how they (content characteristics) affect emotion of the users (RQ2) and whether content

characteristics and emotional responses from the users contribute to the behaviour of liking and sharing (retweets) (RQ3 and RQ4).

The emotional effects of social media content is not a well researched area according to a 2020 mega study [4]. Previous research in this area states that most stay within marketing and information research. Their finding confirmed how emotional responses from users could affect their decision making process in terms of liking and sharing. For example, a research found positive emotions drove increasing engagement behaviour on Instagram using a triangle framework which hypothesized that content characteristics (stimulus) generated perceptions and positive emotion in users (organism) which in turn influenced their affective commitment and interactions (responses) [8]. One of the cognitive models was proposed by Adikari et al to detect emotion awareness in industrial chatbots [7]. This model was successful in detecting perceived emotions from conversation data and predict transition of emotion from one state to another. The Chatobox model describes a generic cognitive process with "input" and "output" and flows through three layers: cognitive, emotions and responses. Other cognitive based research emphasize the importance of message content- how "thoughts" are expressed in the tweets and whether the content is for providing information or for inspiration or entertainment, and how the emotional response could affect engagement [9][12][25]. In marketing research, studies used self-reports to obtain data on categorical emotional responses to content stimulus. They found when users perceived joyful emotion on perceived "ease of use" of a product, users would respond with positive behavioral intention such as liking and sharing [4]. However, these researches almost all took place in the setting of marketing and branding, the motivation factors involved are different from user-to-user engagement setting where understanding is around individual users reacting to other individual users. The assumptions for user-to-user engagement are 1) there is a single person

managing the social media account, and 2) the person is not using the account to represent a company or a group when to interact with consumers on social media. This difference is important as the techniques used for increasing engagement with a brand or company are different and factors that affect engagement could differ greatly. This research needs a novel solution that could provide insights on individual user level.

In addition, the size of our data could limit the ability of the machine learning based model to predict engagement behaviour. Machine learning tools are capable of identifying stances and trends but requires large training samples to predict with precision. Cognitive approaches could improve the accuracy of model prediction because "humans can generalize a new concept from small and biased samples." For example, an infant can obtain the characteristics of an elephant just by seeing one elephant. A human does not need large samples to tell the difference between one object from another because of human cognitive mechanism such as "biases" and "association" [27][28]. In modelling of human decision making, researchers typically go around the data scarcity problem by applying cognitive evaluation, perception and inference to improve machine learning models to predict and explain results. Thus a cognitive inspired machine learning model is needed given the lack of understanding of emotional responses and that as access to private user tweet data for user-to-user interaction research is limited with privacy restrictions.

5.1.2 The Proposed Model: Cognitive-emotion-behaviour solution

We designed a generic cognitive- emotion – behaviour model to understand additional features of content, emotional responses and user engagement behaviour. The model has 3 layers, each one representing a part of the cognitive triangle model – thoughts, emotions, and behaviour. For this research model, tweets content represent thoughts/perception/opinions expressed by the CEOs. The characteristics include topics and themes, language features, and content format form

the “thought” layer. Emotions are regarded as mental states aroused by cognitive evaluations of “thoughts” [1][2][4]. Based on the contextual situation, the evaluations can result in emotion and requires actions in response to the emotional arousal [5][8]. In particular, previous studies found feelings such as inspired, delight, excitement, happiness, joy, pleasure and satisfaction can be regarded as positive emotions [4][6][8][12][14], and that positive emotions have been shown to be variable important for the development of better experiences in online game design in cognitive behaviour therapy [27], as well as engagement with the content. Therefore, the emotion layer is represented by sentiment and subjectivity affectivity. The behaviour layer is user engagement as an output of active processing and interpretation of information received, and emotional responses. In this research, the user engagement is measured by the count of likes and retweets. Due to access issues, we did not include replies. The generic model allows us to analyze content of the tweets as the input to identify motivation factors that can trigger thoughts within the cognitive layer, trigger emotion thus require responses, which in turn compel behaviour. The proposed research model is illustrated below:

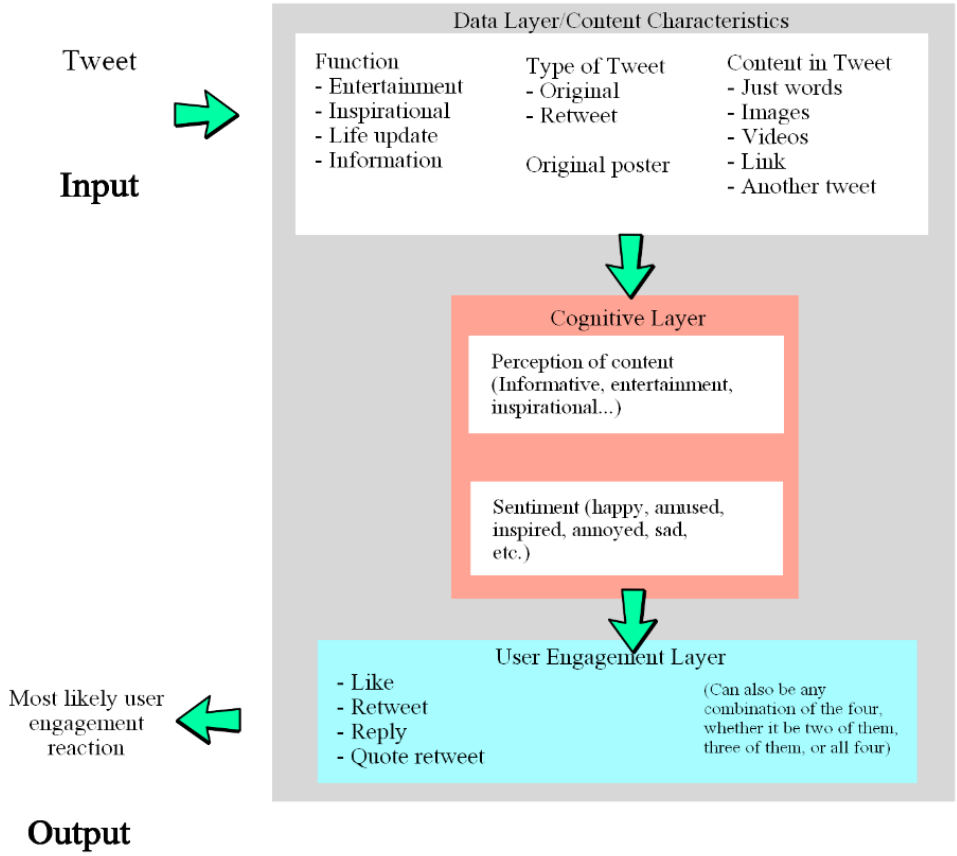

Illustration 12 – Novel Research Model Inspired by CBT

5.1.3 The novelty of the solution

The main novelty of the proposed solution is that it is a hybrid model aiming to combine a machine learning model with cognitive behavioural theory to tackle more complexities in human emotion when looking into user engagement. This framework also allows us to detect these complexities and also detect when they matter in user engagement. The model allows abstractions while allowing for detailed analysis for vertical layers individually based on the cognitive-emotion- behaviour triad. It is therefore extending the study of user to user engagement beyond

traditional probabilistic models, computation and system behaviour. This way of looking at user to user interaction on social media is novel and has the potential to be used to understand the fast growing AI environment where algorithm and artificial emotions exercise significant social influences [6][7].

Second, the proposed research model includes the cognitive aspects of perceived utility of thoughts/opinions expressed in tweets, as well as the aspects of emotion's role in influencing decisions and behaviours. This is different from most of the existing research on user engagement, where the impact of cognitive perception is rarely considered although the perceived utility of social media content is established, i.e. whether a tweet is informative or entertaining. Such cognitive perception or evaluation could generate positive feelings such as happiness, enjoyment, hence increasing the likelihood to like or share. Our proposed model goes further to consider subjectivity and sentiment to find more depth in opinions and thoughts expressed in tweet data. This proposed model considers the potential emotion and feelings aroused from perceived utility of the tweets and makes predictions based on those aspects. Since it was based on the cognitive behaviour model, engagement with companies, subjective affectivity and sentiment polarity, as well as time, and other posting habits are all considered. We have not found a study that included this wide range of features. Thus, it can be argued that the proposed model has potential to complement other techniques to provide a more complete understanding of user engagement behaviour.

The proposed cognitive model could be effective for experiments or exploratory studies of user engagement on social media. Since social media data is a massive digitalized human cognitions [6], our proposed cognitive model would be useful to investigate perception and information diffusion online. For example, our model could be used to investigate the relationship

between emotional content and social dynamics through looking at the mindsets of online influencers and the behaviour of their followers in the form of user engagement. Investigating how concepts are negatively structured by online users and how that has led to a widespread of misinformation on COVID would require consideration of cognitive dimension of a user's semantic frames. With the cognitive priors our proposed model established, we could understand how information is cognitively perceived, framed and shared [4][6][9][26].

5.1.4 Implementation

The following illustration is our roadmap for the proposed solution under a user level architecture. The goal is to detect the relevant factors that have positive affect on user engagement as well to predict engagement behaviours likes and retweets with reasonable accuracy and precision. The explanation of processes working at the background with the required testing and training and validation. The step involved at each stage are discussed in the following sections. We present our modeling framework also the roadmap for this effort in the following illustration.

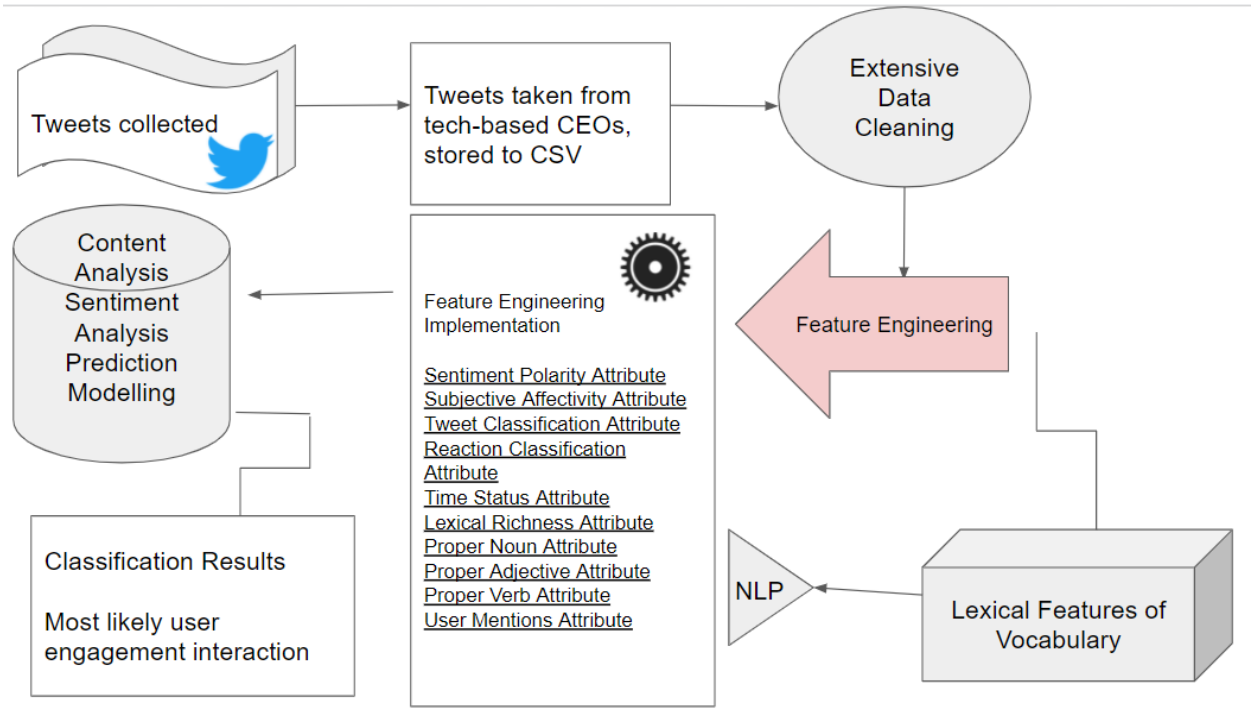

Illustration 13 – Roadmap for Implementing a Cognitive Inspired Machine Learning Model

5.2 Dataset Extraction

We utilized the same dataset we had for the case studies to analyze the cognitive model as a mean to see if the results are different. We focus on the CEOs as individual content creators, specifically because their user accounts are run by themselves and not by company accounts. The tweets are also always readily available for the purpose of observing user-to-user engagement since the accounts are available to the public. As seen in the case study, the selected tweets are within a time span of March 2018 to March 2021. The main reason that it wound up being this way is that Some CEO closed the account during this time or an account was not in use for a lengthy of time. These factors prohibited us from getting the max 100 tweets from each one of them.

Data was extracted and classified based on the features that were engineered into the program. Due to small data size, data cleaning was not substantial. The following data cleaning was completed:

A) Removal of URLs: The links related to post and pictures, videos or other articles are removed because they are not related to this research. Instead they may screw the count of nouns.

B) Removal of emojis: Emoticons used in the tweets because of specific characters when translating to CSV.

C) Removal of empty lines: there are a lot of empty lines that do not provide any insights to this study. Therefore we removed them.

D) Removal of characters that cannot be recognized like Japanese or Chinese characters used by some CEOs.

We also used a Python code to create a heatmap that shows the common topics from tweets with hashtags. For outliers like Elon Musk who did not use hashtags in any of his tweets, but deliberately used image and @ which may correlate with other factors to influence engagement, standard search function was used to identify topics that he tweeted. For measures of other features, machine learning models such as logistic regression and NLP were applied.

5.3. Feature Engineering Implementation

This section describes the feature engineering that we implemented to find useful predictors of user engagement. In our research, the feature engineering attributes focus on topic and themes, language, the type of tweets, the amount of engagement, lexical features of vocabulary, sentence structure, sentiment polarity analysis, and other features that can be analyzed and extracted from user data. The behaviour that we focus on in this research is retweets and likes. We are also examining the role each feature has in influencing our cognitive behavioural data layers.

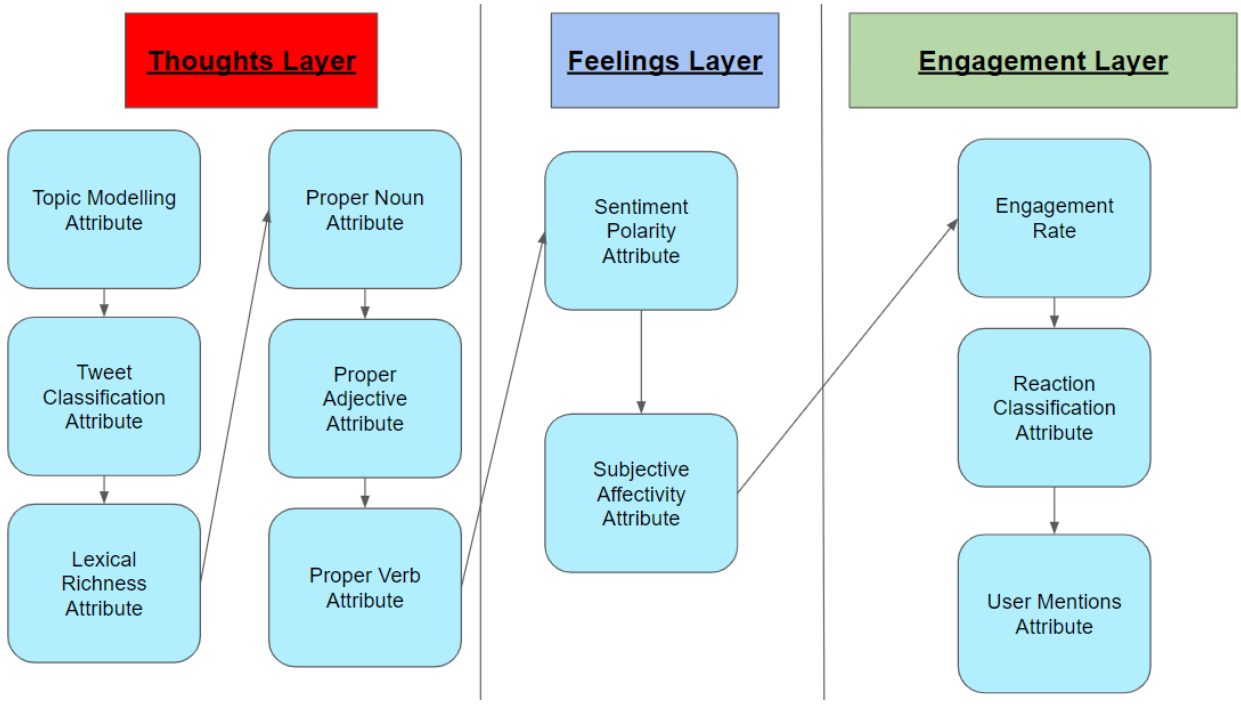

Illustration 14 – The Workflow of the Extensive Feature Engineering

5.3.1 The Thought Layer – Content Characteristics

In a communication process, content refers to a message that the sender (user A) directs towards potential receiver (user B) [4]. Expression of thoughts on social media such as Twitter has been studied extensively [4] [14]. One of the most important content feature is topic [20]. Our model is cognitive inspired, therefore it is important to understand whether content characteristics has affective power on user engagement.

Topic modeling: On Twitter, topics have been used to understand user preferences in order to generate higher volumes of engagement [20] [14]. Topics can be in the form of words, phrases, and hashtags. As such, hashtags is a quick way to detect topics that may have influence on engagement. Our case study evaluation revealed each CEO appeared to have their own individual preferences of topics to comment on. But the case study evaluation did not reveal whether some topics are more influential than others and why that is the case. Based on these, a topic modeling was adopted by this research aiming to see if certain topics are commonly tweeted by CEOs and if those topics can be used to predict engagement. However, in our sample, Elon Musk does not use hash tags while others do. We excluded his tweets for topic modeling. We also excluded Ginni Rometty who did not have enough tweets to make a correlation. Therefore our dataset for topic modelling is n=800.

Tweet Classification Attribute: The type of tweet posted by a user can influence the amount of engagement the posts receives. Therefore, it is important to classify the tweets we have collected. We assigned the following classifiers to distinguish the types of tweets.

- **Original post**: A tweet that was written and published by the original user (in this case, the respective CEO). Original posts tend to receive most of the user engagement.
- **Retweet/shared posts**: A post that was created from another user and shared on your own account. These types of tweets can get some engagement with the person who retweeted it, but compared to original posts, it is much lower. Twitter does not provide a feature that shows the number of likes a retweet receives. However, it does not mean that the retweet has no engagement. It is still an important social media type that has impact on level of engagement.
- **Replies**: A post created by the original poster in response to a different post. Usually these can be gratitude, congratulations, or informative. While they may have not a lot of engagement on the surface, replies are an important form of engagement between users.

1. **Proper Noun Attribute**: Proper nouns per sentence used in a tweet, in a percentage compared to the total number of words in an individual tweet. The purpose of this attribute is to determine how many instances of proper noun usage occurs in a tweet body.
   They will be explained as percentages and then calculated with the following formula:
   Number of proper nouns / Total number of words (for individual tweets) x 100
2. **Proper Adjective Attribute**: Proper adjectives used per sentence in a tweet, in percentage compared to the total number of words. The purpose of the attribute is to determine how many instances of proper adjective usage occurs in the tweet body.
   They will be depicted as percentages and then calculated with the following formula:
   Number of proper adjective / Total number of words (for individual tweets) x 100

3. **Proper Verb Attribute**: Proper verbs used per sentence in a tweet, in percentage compared to the total number of words. The purpose of this attribute is to determine how many instances of proper verb usage occurs in the tweet.

   They will be explained as percentages and then calculated with the following formula:

   Number of proper verb / Total number of words (for individual tweets) x 100

   1. **Lexical Richness Attribute**: Lexical richness is a broad concept that measures the quality of content in a tweet. This feature can help to identify how effective thoughts are expressed and how the quality of tweet content may influence level of engagement. For our study, we measure lexical richness of tweet content of the 10 CEOs by using Python. An empty list is created, and as the code runs through reading the words in a tweet, it will add the first iterance of a word into the list. In subsequent tweet readings, more unique words will be added until all the tweets of the particularly CEO are read. Afterwards, the output will show the number of unique words.

      Let UW be the number of unique words in a tweet.

$$Lexical\ Richness\ Attribute = \frac{Total\ number\ of\ words\ in\ UW}{Total\ number\ of\ words\ in\ tweet}$$

Although our study focuses on text other formats of expression will be considered as well, for example the use of mem, image, video link etc. These recent format used to expressed thoughts have shown to have positive affect on user engagement in various studies [4][7][8][12].

5.3.2 The Emotion Layer – Sentiment Analysis

In order to study the emotion reactions that come from reading the content characteristics above, sentiment analysis has to be done. Our features from feature engineering look at the sentiment polarity first and foremost.

The typical polarity score is determined by going back to the dictionary to mark which of the tweets have primarily positive, neutral or negative sentiments. Similarly to the case study analysis back in Chapter 4, we will be looking at what amount of tweets make up of positive sentiments, negative sentiments, and neutral sentiments. With these scores, the general sentiment will be evaluated.

4. **Sentiment Polarity Attribute**: Throughout someone's day, there will be times when someone's mood or thoughts change completely. This attribute will detect that shift in polarity between positive sentiments and negative sentiments based on the tweet. Usually, the fluctuation in this context will imply that some kind of event had taken place. Specifically, it will read through the tweets on someone's account and determine if they are positive or negative. The polarity points will be detected and determined from the time of day; both the positive polarity and negative polarity.

In particular, we want to dictate the overall sentiments of the tweet content; whether it be positive or negative. We also want to compare that to the time of day and whether or not there is a major shift between a tweet with a positive or negative sentiment.

For the user U_i, the set of n tweets for each Twitter user is TS = $\{t_1, t_2, t_3 \ldots\ldots\ldots t_n\}$

Let w be the set of z words for (TS), where W = $\{w_1, w_2, w_3 \ldots\ldots\ldots\ldots w_n\}$

Let the number of negative tweets in (TS) = neg_nc

Let the number of positive tweets in (TS) = pos_pc

Let the number of positive words in a tweet be (W) = pos_wc

Let the number of negative words in a tweet be (W) = neg_wc

Let β be the hyperparameter that depends on the number of posts.

$$Polarity\ Contrast\ (PC) = \frac{(\beta * pos_{pc} + pos_{wc}) - (\beta * pos_{nc} + neg_{wc})}{(\beta * pos_{pc} + pos_{wc}) + (\beta * pos_{nc} + neg_{wc})}$$

5. **Subjective Affectivity Attribute**: This attribute is to determine the subjectivity of a tweet. Emotions are complex, so the purpose of the attribute is to figure out the subjectivity score. With this score, it would be possible to determine how much subjectivity a tweet might have. The sentiment affectivity score is a sub-formula of the sentiment polarity. It is a float value within the range of [0.0 to 1.0] where 0.0 means it is very objective and 1.0 means it is very subjective. A value of the subjective affectivity will be done by finding the overall average of all the scores.

5.3.3 The Engagement Layer – Likes and Retweets

6. **<u>Engagement rate:</u>**

Threshold: One of the issues that we run into is being able to properly define the thresholds for the engagement. While engagement may seem easy to determine, the problem is what constitutes as high, medium and low engagement is highly contextual. If an account does not usually get a lot of engagement, then having 10 on its own might be considered as high on its own. Other important metrics include

- **Follower Count:** The assumption is larger number of followers, larger degree of attention that an account receives, and in turn will increase the likelihood of retweets and likes. On Twitter, any accounts that have more than 1000 followers are likely to attract much more attention, sharing and liking behaviour. Although even then, there are exception. The important distinction with follower count is how popular these people are.

- **Levels of Engagement:** The levels of engagement refers to the amount of engagement that a social media post receives. There are at least four levels that we are defining within our model. However, the engagement ranges might be different for all accounts, due to factors such as the number of followers an account has as of March 8 2021. As a result, we have created a formula to allow us to determine the proper range depending on the account.
- The formula that we will use to calculate the engagement rate for a tweet is as follows [71]:

$$engagement\ rate = \frac{(likes + retweets)}{number\ of\ followers}\ x\ 100$$

  - **No engagement**: Self-explanatory, but this is for times when the engagement is at zero. This does not really apply to users, as all of the accounts have at least 10 total engagements. However, this does apply to individual posts and might be more content-based.
  - **Low engagement**: It would entail that the amount of interactions on a post are lower than the amount of followers. While this doesn't mean it's only being seen by followers, even when it is being shared, it does mean that very few people are interacting with it when they see it on their respective timelines.
  - **Medium engagement:** The amount of interactions is close to, exact, or a little more than the number of followers on the account. On average, medium engagement is quantified by having an amount of engagements equal to, or slightly less or higher, than the number of followers on an account.
  - **High engagement:** It would usually entail that the amount of interactions are higher than the amount of followers. This means it is more likely that the post is being shared around and noticed by people who do not already follow.

7. **Reaction Classification Attribute**:

Likes and retweets are engagement behaviours in responses to the cognitive and emotional input described above. In our database, each tweet has an observable level of "likes" and "retweets" by month and year. Observing the increases or decrease of "likes' and "retweets" could reveal to what factors the changes are responding to. For Twitter in particular, a post with a lot of likes and retweets/shares tends to indicate that post has reached a high engagement level.

5.4. Predictive Model

When you know what factors are more present, it helps to further determine

The point of the model is to capture as many factors as possible to expand the understanding of user engagement and content characteristics. Although it is a calculated model, the results from the model could influence how engagement could be predicted. This is because the features with the highest correlation to user engagement will be used in the prediction model. We take inspiration from Li et., al [9] in the following formula:

$$Eng_i = \gamma_0 + \gamma_1 Cont\_direct_i + \gamma_2\, Cont\_link_i + \gamma_3 LexicalRichness_i + \gamma_4 At_i + \gamma_5 Like_i + \gamma_6 Retweet_i + \gamma_7 Time_i + \gamma_8 Polarity_i + \gamma_9 Subjectivity_i + \gamma_{10} Nouns_i + \gamma_{11} Verbs_i + \gamma_{12} Adjectives_i + \varepsilon_i$$

- γ_0 is the intercept
- Cont_direct is the dummy variable that has to do with the general content characteristics of the formula. While $Cont\_link_i$ is when a tweet in question contains a link that leads to either an image, another tweet, an article link or a video.
- $LexicalRichness_i$ refers to the uniqueness of words and text in the respective tweet. This, along with $Nouns_i$, $Verbs_i$ and $Adjectives_i$ all fall under the category of "parts of speech".
- $Time_i$ refers to the day, month, year, hour, minute and second that a tweet was posted on.

- At_i refers to the indication or appearance of the @ symbol within a tweet. This can be for people being mentioned in a tweet or when a reply is written.
- $Like_i$ and $Retweet_i$ refer to the number of likes and retweets that a post has received overall.
- $Polarity_i$ and $Subjectivity_i$ talk about the sentiment polarity of a tweet (whether or not it's positive, negative and neutral) as well as how subjective a tweet is.
- The last variable, ε_i, is to take in the possibility of any errors that may occur with the formula.

Chapter 6 –Discussion of Findings and Evaluation of the Solutions

Following the cognitive modeling framework, three sets of features were used to examine which components of thoughts and emotions could affect user to user engagement in the forms of likes and retweets. In this chapter, the results from our cognitive inspired solution are discussed. Through extensive feature engineering, our model consider the content characteristics and sentiment features on user engagement: (1) tweet topic and themes, linguistic features and other format used to express thoughts (2) tweet sentiment (3) engagement rate, retweets and likes. The correlation between the factors and engagement behaviours is established and presented as well.

Included in this chapter are detailed results from applying our cognitive model to 10 datasets established for this research. They belong to 10 former or present CEOs of tech based firms and industries. An attempt has been made by this study to explain the outlier datasets using mainly findings from previous research instead of re-testing. We then summarize similarities and differences between the results of the two models, showing what the cognitive model can tell us that the case study cannot.

In order to properly compare the cognitive model findings to the case study findings, we split the dataset of 938 tweets to 10 individual datasets. The cognitive model was implemented on them 10 CEOs account. This approach was taken due to the outlier accounts where their characteristics are distinct. This approach was taken by a few previous studies such as the study of use of memes and affective user responses [12], use of image and its impact on engagement [9], and predicting social engagement from word use [10]. Comparison was made to the results from cognitive model and the case study approach used in the first part of the research, aiming to find out:

1. Which form of engagement is more likely in these tweets

2. What is the characteristic content and emotional or other category that affects the result the most?

In order to properly validate our model, we used accuracy, precision and recall. We took a look at the true positive (when the model predicts the post will get a lot of likes and it does), true negative (when the model predicts a post will get a lot of retweets and it does), false positive (when the model predicts a post will get a lot of likes but it gets a lot of retweets) and false negatives (when the model predicts a post will get a lot of retweets but it got a lot of likes). With the results listed for each of the individuals, we still need to validate our cognitive model.

We want to make sure we properly count all the instances that the model predicts the positive result – likes in this case. Since the model has been tested, we can rely to the Excel sheets to see what the data from the case studies has presented with the raw numbers and do a side-by-side comparison, to see if the model's prediction was correct.

6.1 Cognitive model results and validation

The following sections are the results and findings from implementing the cognitive inspired machine learning model for each of the CEOs. Included in each of the discussion is the result of validation of our model's performance. To emphasize, the cognitive model was applied to 10 different datasets. We then compared the results to those found through case studies. Finally, the model was validated with precision, recall and accuracy.

6.1.1. Bill Gates – Cognitive Model Results

The overall result that we received from looking at the engagement rate, all the correlations, and the code shows that the topics that Bill Gates touches upon affects the user engagement the most. Perhaps it has to do with the type of topics that he discussed during the year. After all, 2020

was very active when it came to talking about COVID-19 and he was one of the people who were leading research into a vaccine.

6.1.1.1 Topic Modelling Themes

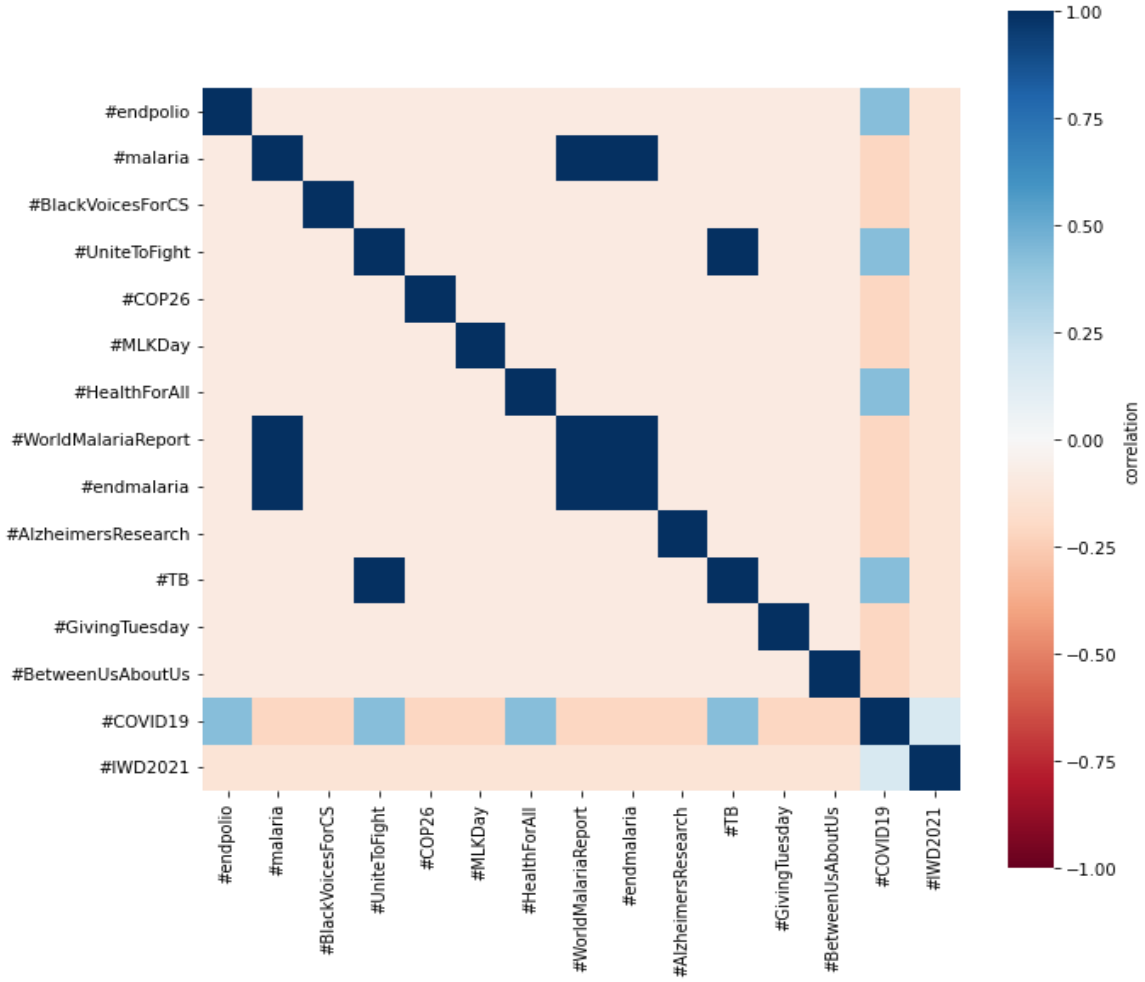

Figure 72 - Topic Modelling Heatmap (Bill Gates)

In the topic modelling heat map for Bill Gates, the hot topics are mainly related to advancement of medical research. Since a lot of these tweets were posted in the period of 2020 – 2021, a large focus is on discussing COVID-19, specifically the vaccine. We quantified the effects of topic on the number of likes and retweets using our model estimates [9] [14]. A couple of Tweet without hashtags and a tweet with hashtags was randomly selected to observe the effect. The tweets

were posted one after another. We estimate the effect of the hot topic by first of all obtaining the expected number of likes and retweets of a tweet without this topic to build a baseline. Then it is compared with the expected number of likes and retweets with the tweet with the hot topic. A huge percentage change is identified suggesting the hashtag topic received a lot less than the tweets that did not have a hashtag. However, this randomly selected set of examples provide evidence for a strong negative effect of the hot topic on likes and retweets.

Table 12 - Estimation Results for User Engagement – Topic and Themes (Bill Gates)

Tweet content- topic/theme	Likes	Retweets
Tweet w/o "hot topic" "Thank you @BarackObama for the kind words. Your leadership at the 2015 Paris climate conference helped prove that global cooperation on climate is possible and gives me hope that we can reach our goal of zero emissions. (Gratitude)	5,703	518
Tweet w "hot topic" Virologist Ruth Bishop pioneered the research that led to the discovery of rotavirus. Her legacy continues both as a role model for other researchers and in the millions of children's lives that have been saved because of her heroic work #IWD2021 (Informational)	1,306	201
Percentage change	-336.68%	-157.71%

6.1.1.2 Engagement Rate

The average engagement rate for Bill Gates' tweets reached an average of **0.015**, as determined from our engagement rate formula. This may not seem like a very high engagement rate, but it should be remembered that Bill Gates has a high amount of followers.

6.1.1.3 Predicted Engagement Rate

When looking at Bill Gates' predicted engagement, we see that a lot of his tweets are more likely to get likes. On average, his prediction engagement rate was at **0.748.**

The content characteristics show that the sentiment polarity and the subjectivity have the highest correlation to the prediction engagement rate, with a moderate correlation rate of 0.53. There is a higher significance suggesting that they affect his overall prediction engagement more than any of the other content characteristics.

6.1.1.4 Comparing to Case Study Results

Similar to case study results, the cognitive model results also show positive sentiment and subjectivity (content characteristics) are related to engagement levels. As well the cognitive model results also predicted Bill Gates is more likely to receive "likes" than "retweets". But the case study results did not tell us which content characteristics most influenced engagement behaviours. The cognitive model provides evidence that 'topic and theme' (thought characteristics) has the strongest affect compared to other factors for Bill Gates. In addition, the cognitive model results showed the role of cognitive perception of the tweet can be a behaviour driver. The perceived "utility" as informative and inspiring related to the topics of Covid vaccine is an example we noted in the cognitive model application. It is interesting to see the negative and neutral are almost equivalent although their numbers are small compare to the positive ones. The positive sentiment appears to relate to Bill Gates frequent posts of trending topics such as major health events during the time period under our exam. This is consistent to previous studies [20][35] which found trending topics leads to positive sentiment then leads to higher level of engagement.

In addition, applying the cognitive model allows a clear predicted idea about the overall rate and whether or not these factors have a large effect on the likelihood of likes or retweets. This is something that we could not determine from the case study alone, as case study results provide raw numbers and identified patterns and trends. It does not speak to relationships.

6.1.1.5 Validation Results

Bill Gates	Likes	Retweets
Predicted as Likes	70	9
Predicted as Retweets	11	10

Table 13 - Validation Results for Bill Gates

True Positive (TP): According to the model, 70 tweets were correctly predicted as more likely to get likes compared to retweets.

True Negative (TN): According to the model, 10 tweets were correctly predicted as more likely to get retweets compared to likes.

False Positive (FP): According to this model, 9 tweets were incorrectly predicted as more likely to get likes compared to retweets.

False Negative (FN): According to this model, 11 tweets were incorrectly predicted as more likely to receive retweets compared to likes.

With these scores in mind, the accuracy rate for the cognitive model to Bill Gates' tweets have a 70% accuracy rate of being predicted as more likely to receive likes. Among this rate, the model is precise 87.5% of the time and capable of getting the proper recall 88.61% of the time. When it comes to evaluating the numbers and tweets in our database, it has a pretty good accuracy, precision and recall rate.

6.1.2. Elon Musk – Cognitive Model Results

The following sections describes the cognitive model results we gathered on Elon Musk's tweets.

6.1.2.1 Topic Modelling Themes

Due to Elon Musk not adding many hashtags onto his tweets, a topic modelling table could not be made by our code. While this is not to say that he does not touch upon any kind of topics, it is worth noting. Following the methodology of Rieger and Klimnt [12] in their study of themes and affective user responses, we identified Elon Musk's content characteristics as lexically rich with use of image, meme and special letters.

For example, for all the tweets that Elon Musk posted, there is one tweet that received a sentiment and polarity score of 1, which was a meme about cryptocurrencies. It is unclear which word the data determines as the "positive" sentiment. There are quite a few tweets that share this trait. Research about the role of meme in engagement level found social media meme provides 'entertainment' and hedonic enjoyable element such as humor. Thus many attract attention for many users who access his tweet with the intention to consume entertainment offering [23].

One of the most common topics that we see in Elon Musk's tweet tend to be the SpaceX business and cryptocurrency. These topics garner a lot of interest, thus receive large amount of attention often being perceived as "exciting" and "inspiring". In addition, Elon Musk made lot of jokes or meme-like tweets that reference cryptocurrency such as Dogecoin. In fact, it is primarily something such as Dogecoin as he is one of the people who popularized it within the community.

Table 14 - Estimation Results for User Engagement – Topic and Themes (Elon Musk)

Tweet content- topic/theme	Likes	Retweets
Tweet w/o Creating the city of Starbase, Texas	220,864	14,188
Tweet w "meme" Doge meme shield (legendary item)" (entertainment)	280,340	26,667
Percentage change	21.22%	46.80%
Tweet w/o Launch abort on slightly conservative high thrust limit & recycling or propellant for another flight attempt today.	95,285	3,914

Tweet w "image" Starship SN10 landed in one piece! [image]	172,518	18,329
Percentage change	44.77%	78.65%

6.1.2.2 Engagement Rate

When looking at all of Elon Musk's tweets, we can determine that his average engagement rate is at **0.17056**. It seems like a fairly low engagement rate, but it makes sense when one considers how often his posts are retweeted. Similar to Bill Gates, there are views of their tweets from those who are not their followers. Musk's count increases often and quickly, suggesting robots accounts.

6.1.2.3 Predicted Engagement

The cognitive model has determined that the average prediction engagement that appears for Elon Musk, yields a score of **0.823**. This leans heavily in favor of Elon Musk more likely getting likes as the typical engagement on his tweets. In addition, it is one of the higher totals that we have seen.

The model had predicted a correlation score of 0.68 with Subjectivity, which is connected to the Thought Layer. It also shows that the content characteristics are the factor that strongly affect the engagement level for Musk the most. There was also a fairly significant correlation of 0.532 in relation to topic modelling and the engagement level.

6.1.2.4 Comparing to Case Study Results

Elon Musk shows us an interesting case when comparing his cognitive model results to his case study results. There are likely a lot of aspects that affect his tweets, but on a case study, it is pretty hard to tell. We do not have an answer to precisely why he has so many tweets with a high neutral sentiment. However, the case study also shows us that he is someone who is able to gather

a lot of engagement regardless of what he tweets. No matter what language he uses, how his sentences of structured, or what kind of words he uses, it does not affect how much engagement Elon Musk gets. Without a clear pattern, it comes off as very random and he remains an outlier within the data.

The cognitive model shows us that there is a high correlation with Elon Musk's tweets and subjective affectivity – as stated before. This makes sense when we look at the different topics that he touches upon. There are quite a few that make references to entertainment and popular memes on the Internet. This is especially clear from how he expresses himself.

6.1.2.5 Validation Results

Elon Musk	Likes	Retweets
Predicted as Likes	68	11
Predicted as Retweets	16	5

Table 15 – Validation Results for Elon Musk

True Positive (TP): According to the model, 68 tweets were correctly predicted as more likely to get likes compared to retweets.

True Negative (TN): According to the model, 5 tweets were correctly predicted as more likely to get retweets compared to likes.

False Positive (FP): According to this model, 11 tweets were incorrectly predicted as more likely to get likes compared to retweets.

False Negative (FN): According to this model, 16 tweets were incorrectly predicted as more likely to receive retweets compared to likes.

With these scores in mind, the accuracy rate for the cognitive model to Elon Musk's tweets have a 68% accuracy rate of being predicted as more likely to receive likes. Among this rate, the

model is precise 93.15% of the time and capable of getting the proper recall 86.08% of the time. When it comes to evaluating the numbers and tweets in our database, it has a pretty good accuracy, precision and recall rate.

6.1.3 Jeff Bezos – Cognitive Model Results

In this section, we describe the results for Jeff Bezos' cognitive model results.

6.1.3.1 Topic Modelling Themes

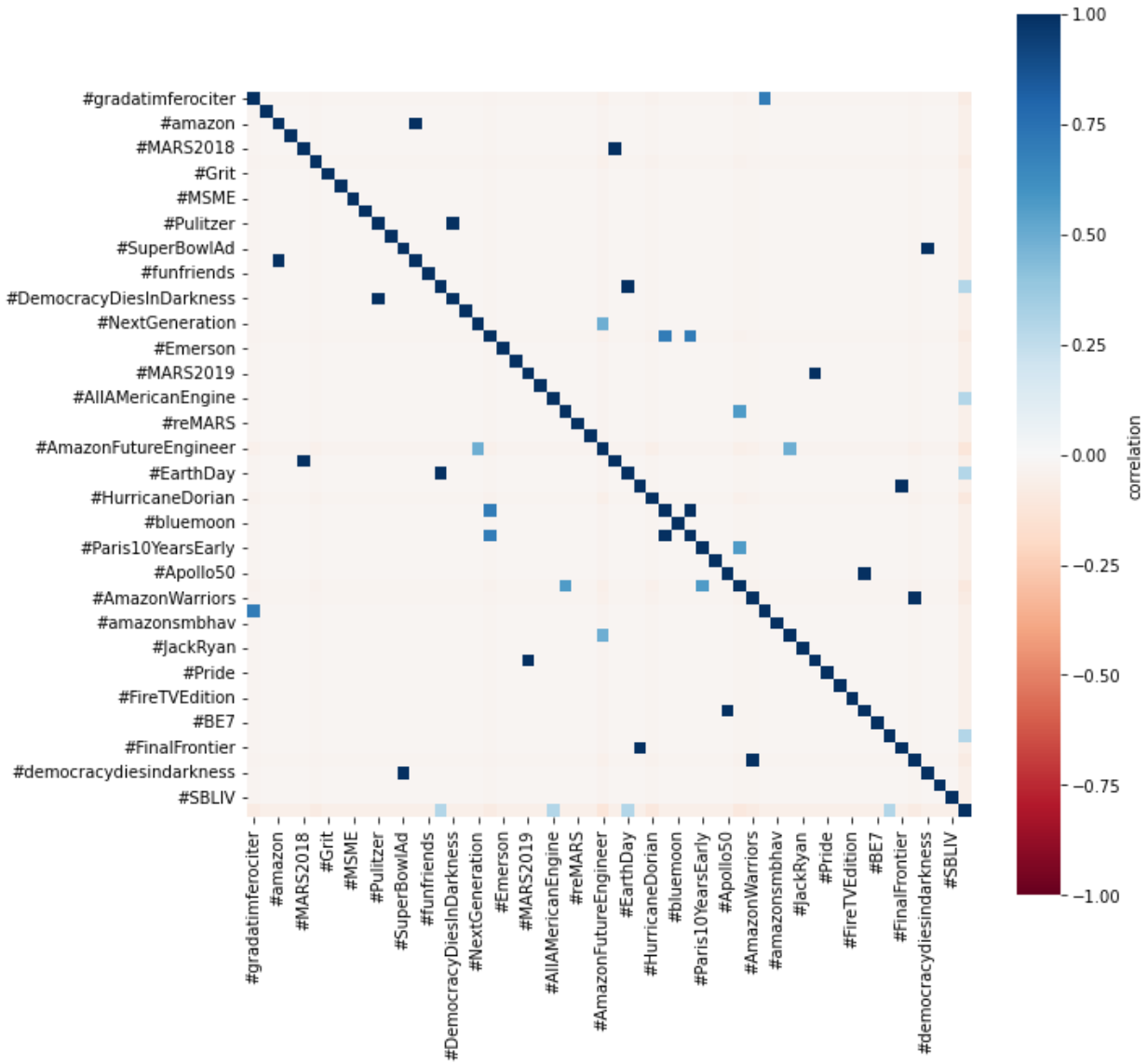

Figure 73 – Topic Modelling Heatmap (Jeff Bezos)

When looking at Jeff Bezos' topic modelling heat map, it is noted that Bezos' most "liked" tweets are perceived informative. The topics unveiled with topic modeling appears heavily on things relating to Amazons, from partnerships to the Amazon products themselves. There is at least one tweet related to entertainment as well, which featured him spending time and hanging out with famous R&B singer, Lizzo.

We quantified the effects of topic on the number of likes and retweets using our model estimates. The effect of the hot topic is first by obtaining the expected number of likes and retweets of a tweet without this topic to build a baseline. Then it is compared with the expected number of likes and retweets with the tweet with the hot topic. A percentage change is identified for the likes and retweets respectively. The results provide evidence for the positive effect of the hot topic on likes and retweets.

Table 16 - Estimation Results for User Engagement – Topic and Themes (Jeff Bezos)

Tweet content- topic/theme	**Likes**	**Retweets**
Tweet w/o "hot topic" "Thanks for the Super Bowl ad"	17,457	2399
Tweet w "hot topic" "I just took a DNA test, turns out I'm 100% @lizzo's biggest fan. #SBLIV	101,606	11,854
Percentage change	**82.82%**	**79.76%**
Tweet w/o "hot topic" Proud of the program we have in India to hire and train deaf associates at our delivery stations. Managers learn basic sign language to be able to communicate. Meeting this group was a moving experience for me.	24,461	2,335
Tweet w "hot topic" Hey, India. We're rolling out our new fleet of electric delivery rickshaws. Fully electric. Zero carbon. #ClimatePledge	67,390	24,797
Percentage change	**63.70%**	**90.58%**

6.1.3.2 Engagement Rate

When looking at the engagement rate numbers on Jeff Bezos's tweets, it seems that on average, he winds up with an average of **0.7005**.

6.1.3.3 Prediction Engagement

The average prediction engagement score that Jeff Bezos received was about 0.6926. His tweets are more likely to receive likes than retweets.

For the most part, it seems that the content characteristic that gained the most attention from Jeff Bezos' tweets was the sentiment analysis, with a correlation of **0.734**. This is a fairly high correlation in comparison. The characteristic that provided the least amount was subjectivity, at a correlation of about 0.32, suggesting that there is a lack of significance there.

6.1.3.4 Comparing to Case Study Results

The case study results of Jeff Bezos show that he has 1,426,093 interactions in the time frame. His tweets have positive sentiments (75%); he does not post frequently with a large gap of no posts. The case study also shows Jeff Bezos expressed his thoughts with well-structured sentences. However the case study results do not provide evidence for which factor most likely contributed to the level of interaction.

The cognitive model results provided such insights. First, through applying the model, we found sentiment has strongest possible positive correlation to the level of engagement (r=.73). Second, through looking at topic modeling, we estimated the changes of likes and retweets with tweet with and without. It is noted trending topics especially the one with perceived utility of entertainment affected the level of engagement with 82% jump in the number of likes compared to the previous post without the trending topic. Another example of using trending topic – zero

carbon rickshaws – dramatically increased the number of retweets which suggested a stronger perceived value of the tweet content as "informative" thus retweets 92% more than a similar content minus the topic of zero carbon.

6.1.3.5 Validating Results

Jeff Bezos	Likes	Retweets
Predicted as Likes	71	11
Predicted as Retweets	13	4

Table 17 – Validation Results for Jeff Bezos

True Positive (TP): According to the model, 71 tweets were correctly predicted as more likely to get likes compared to retweets.

True Negative (TN): According to the model, 4 tweets were correctly predicted as more likely to get retweets compared to likes.

False Positive (FP): According to this model, 13 tweets were incorrectly predicted as more likely to get likes compared to retweets.

False Negative (FN): According to this model, 11 tweets were incorrectly predicted as more likely to receive retweets compared to likes.

With these scores in mind, the accuracy rate for the cognitive model to Jeff Bezos' tweets have a 75% accuracy rate of being predicted as more likely to receive likes. Among this rate, the model is precise 94.67% of the time and capable of getting the proper recall 84.25 % of the time. When it comes to evaluating the numbers and tweets in our database, it has a pretty good accuracy, precision and recall rate.

6.1.4 Corie Barry – Cognitive Model Results

The cognitive model results for Corie Barry are the following:

6.1.4.1 Topic Modelling Themes

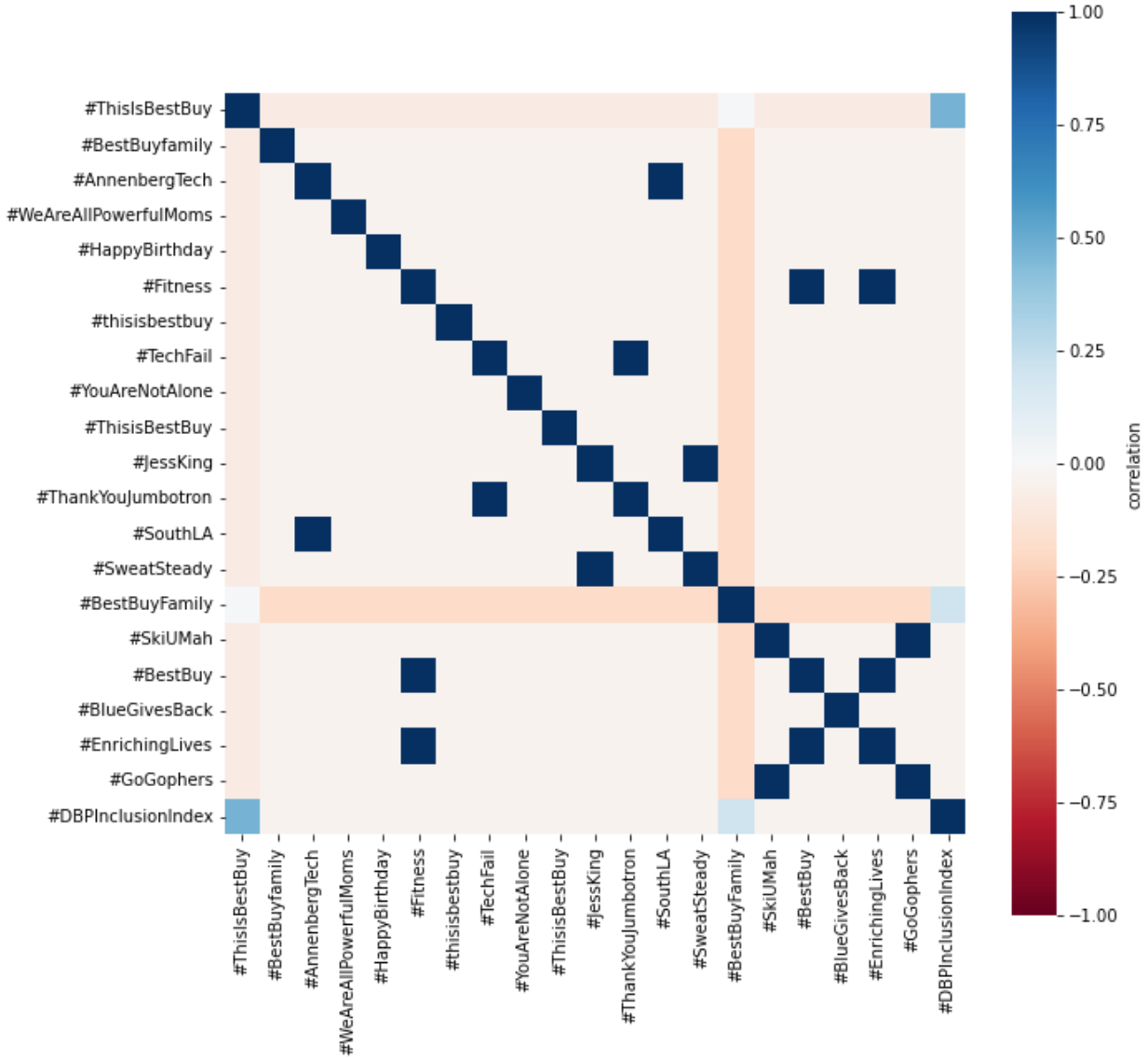

Figure 74 – Topic Modelling Heatmap (Corie Barry)

From heat map, it shows Corie Barry's hot topic is information related to Best Buy, and events related with the store. This matches up with all of the dark squares, because she has a tendency of including the same hashtags. It shows that the general topics she associates together are similar.

We quantified the effects of topic on the number of likes and retweets using our model estimates. We estimate the effect of the hot topic by obtaining the expected number of likes and retweets of a tweet without this topic to build a baseline. Then it is compared with the expected number of likes and retweets with the tweet with the hot topic. A percentage change is identified as indicated in the following table. The results provide evidence for the positive effect of the hot topic on likes and retweets.

Table 18 - Estimation Results for User Engagement – Topic and Themes (Corie Barry)

Tweet content- topic/theme	Likes	Retweets
Tweet w/o "hot topic" So proud of the amazing environment this team is creating. It is wonderful to see their hard work recognized and I am proud to be part of the #BestBuyFamily.	51	2
Tweet w "hot topic" Somehow, I go home more energized and fulfilled after a FULL week in Vegas getting ready for holiday. Thank you, #BestBuyFamily, for all you do. I love being on the team with you.	126	3
Percentage change	59.52%	33.33%

6.1.4.2 Engagement Rate

The engagement rate for Corie Barry is 0.71629. It is an interesting rate to see from Corie Barry, although it may not be completely accurate. Unlike the other CEOs, Corie Barry's follower count had to be taken from a more recent time. It is likely that the followers from previous years were not counted for.

6.1.4.3 Predicting Engagement

When run through the cognitive model, we see that the prediction engagement for Corie Barry is at 0.4349. It is one of the lower scores in comparison to the other CEOs, and seems to suggest that it leans much more towards retweets than it does likes compared to the others.

In addition to this, we have noticed that when it comes to the content characteristic that affect Corie Barry's tweets, we see that the content characteristic with the correlation with user engagement was sentiment analysis (r = 0.55).

6.1.4.4 Comparing to Case Study Results

The case study results show Cori Berry has extreme low engagement level with her most popular tweet with 126 likes. But like the other CEOs, the case study results show Cori Berry's tweets show a positive sentiment (74%), and she uses well-structured sentences to express her thoughts. Her tweeting habits are similar to Jeff Bezos very infrequent. Case study results include word frequency analysis which indicated her Twitter theme word is "thank you" and a lot of reference of her business Best Buy Family. The case study results are puzzling especially comparing with other CEOs who bear similar traits while receiving much higher level of interactions. However the case study itself did not show the association between any of the factors with the level of interactions.

With the inclusion of topic modelling, the cognitive model results provided insights on the content characteristics of the thoughts layer. We looked into topic modelling with hashtags as trending topics. With the inclusion of topic modelling, it showed us that Corie Barry's engagement rate leaning towards likes was heavily determined by the topic. It had the highest correlation to the engagement out of all the content characteristics, which is a surprising find. It is true that Corie Barry uses a lot of hashtags, particularly to promote the Best Buy business and brand. Though her most popular tweet is a photo of her and some colleagues, it relates to her work as she tags it as "BestBuyFamily".

6.1.4.5 Validating Results

Corie Barry	Likes	Retweets
Predicted as Likes	62	16
Predicted as Retweets	13	9

Table 19 – Validation Results for Corie Barry

True Positive (TP): According to the model, 62 tweets were correctly predicted as more likely to get likes compared to retweets.

True Negative (TN): According to the model, 9 tweets were correctly predicted as more likely to get retweets compared to likes.

False Positive (FP): According to this model, 16 tweets were incorrectly predicted as more likely to get likes compared to retweets.

False Negative (FN): According to this model, 13 tweets were incorrectly predicted as more likely to receive retweets compared to likes.

With these scores in mind, the accuracy rate for the cognitive model to Corie Barry's tweets have a 71% accuracy rate of being predicted as more likely to receive likes. Among this rate, the model is precise 87.32% of the time and capable of getting the proper recall 79.49% of the time. When it comes to evaluating the numbers and tweets in our database, it has a pretty good accuracy, precision and recall rate.

6.1.5 Ginni Rometty – Cognitive Model Results

The following discusses the cognitive model results for Ginni Rometty.

6.1.5.1 Topic Modelling Themes

Unfortunately, Ginni Rometty is one of the people who were not able to create a heat – map for out of the topic modeling function. This might be because she has about 39 tweets. As a result, there is not enough to be able to create a heat-map. Despite this, reviewing Ginny Rometty's 39 tweets clearly indicates most of her tweets are about inspiring others, whether it be for people who are looking for jobs in the future. Her tweets could also be addressed to other Americans in response to events such as the attempted siege on the Capitol Hill. We estimated effect of "topic" on the likes and retweets with two of her tweets. The topic/theme is her departure after 40 years working at the IBM. We labelled the first as being "inspirational" and the second as being "personal touch". Based on previous research [4] [19], it is inferred that a perceived utility of the same message is different and hence invoked different feelings which drove different level of engagement. A percentage change in the likes and retweets is identified as indicated in the following table. The results provide some evidence for the positive effect of the topic on likes and retweets.

Table 20 - Estimation Results for User Engagement – Topic and Themes (Ginni Rometty)

Tweet content- topic/theme	Likes	Retweets
Tweet w/o "hot topic" After nearly 40 years at @IBM, I'm sharing my lessons for IBMers with all of you. My first, never let anyone else define you. You write your own story.	408	70
Tweet w "hot topic" Almost 40 years ago, I came to @IBM determined to make an impact on the world. As I step down from my role as executive chairman, I know that IBM will still be part of my next chapter – it will live on in the values and purpose it has instilled in me.	2,645	251
Percentage change	84.57%	72.11%

6.1.5.2 Engagement Rate

First of all, Ginni Rometty's average engagement rate shows an average engagement score total of 0.544. It is a pretty mid-range engagement rate.

6.1.5.3 Prediction Engagement

When it comes to the prediction engagement, it seemed that she had an overall score of 0.6623 meaning that on average, her tweets attracted more engagement from likes than retweets.

Our model also predict positive subjectivity has positive affect on engagement level with 0.65 correlation score for Ginni Rometty.

6.1.5.4 Comparing to Case Study Results

The case study results provided some insights about Ginni Rometty's tweet content and tweeting habits. She only had 39 tweets in the period we selected to explore but the sentiment score is moderately high (62%). She expressed her thoughts with well-structured sentences and are professional. However, due to how many tweets were on her account, it was harder to make any possible conclusion from the case study based on observational data, to predict what affected her engagement the most.

In the cognitive model, sentiment analysis has the highest correlation to the engagement, and affects most of her tweets in getting more likes. Even with how few tweets Ginni Rometty had on her account at the time, many of them had a positive sentiment to them. This explains how positive sentiment could have contributed to most of her tweets on average attracting most of their likes thanks to the general positive perception of the tweet content. An exploration of topic impact followed the idea of perceived utility/value of a tweet which could generate emotion and responses to the emotion result in "likes" and "retweets".

6.1.5.5 Validating Results

Ginni Rometty	Likes	Retweets
Predicted as Likes	25	3
Predicted as Retweets	5	7

Table 21 – Validation Results for Ginni Rometty

True Positive (TP): According to the model, 25 tweets were correctly predicted as more likely to get likes compared to retweets.

True Negative (TN): According to the model, 5 tweets were correctly predicted as more likely to get retweets compared to likes.

False Positive (FP): According to this model, 3 tweets were incorrectly predicted as more likely to get likes compared to retweets.

False Negative (FN): According to this model, 7 tweets were incorrectly predicted as more likely to receive retweets compared to likes.

With these scores in mind, the accuracy rate for the cognitive model is 80%. Among this rate, the model is precise 87.32% of the time, and capable of getting a proper recall rate of 77.5% of the time. There is a chance this is reflective of the low amount of tweets that Ginni Rometty has.

6.1.6 Lisa Su – Cognitive Model Results

Listed below are the cognitive model results for Lisa Su.

6.1.6.1 Topic Modelling Themes

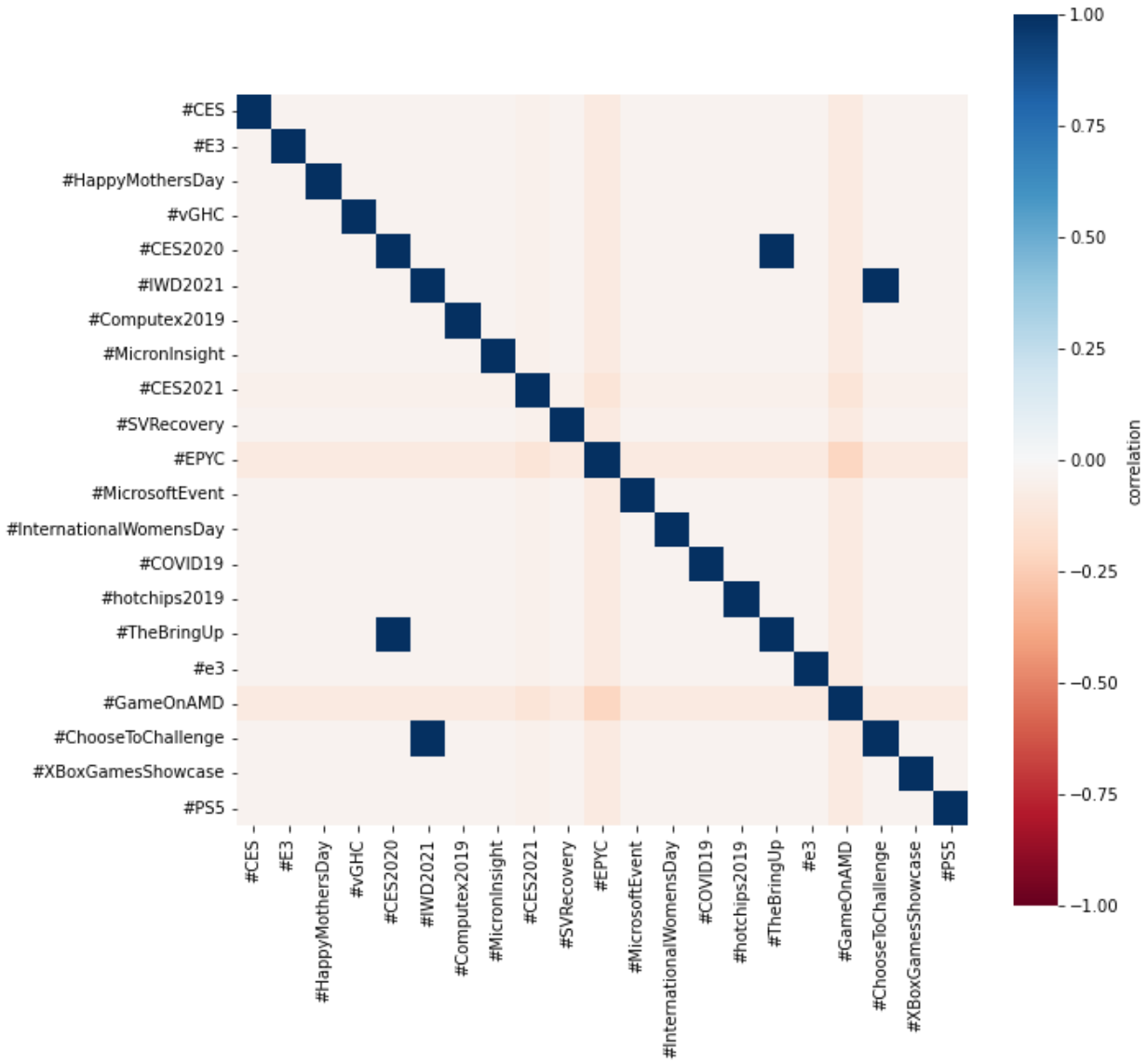

Figure 75 – Topic Modelling Heatmap (Lisa Su)

In the topic modelling heat map for Lisa Su, one can see a lot of informational topics in her tweets. We estimate the effect of the hot topic by obtaining the expected number of likes and retweets of a tweet without this topic to build a baseline. Then it is compared with the expected number of likes and retweets with the tweet with the hot topic. A percentage change is identified as indicated in the following table. The results provide evidence for the positive effect of the hot topic on likes and retweets.

Table 22 - Estimation Results for User Engagement – Topic and Themes (Lisa Su)

Tweet content- topic/theme	Likes	Retweets
Tweet w/o "hot topic" Thanks for all the excitement from our fans who joined us last week to launch @AMDRyzen 5000 series. I look forward to seeing all of you again on Oct 28th as we show off our "Big Navi", @Radeon RX 6000 series !!	4754	499
Tweet w "hot topic" "Welcome to the world, @AMDRyzen 5000 Series, our first processors with the amazing new "Zen 3" architecture. So proud of the @AMD global engineering team. We love gamers! #GameOnAMD" (Informational + Gratitude)	10,151	1658
Percentage change	53.17%	69.90%
Tweet w/o "hot topic" Honored and humbled to be in such great company. Thanks @barronsonline @firstadopter for the recognition!	1370	162
Tweet w "hot topic" What a super cool gift to receive at #e3. Thanks to our friends at Pictographs and Heromods who worked on this secret project and found a way to surprise me! Found a place in my home office for my new friend	1917	187
Percentage change	28.53%	13.37%

6.1.6.2 Engagement Rate

Out of all of the CEOs, Lisa Su had the highest average engagement scores with a total engagement at 2.247%. It creates an interesting result for her, as it means that on the daily basis, there are a decent amount of people who engage with her tweets. In addition, according to high number of followers also means higher attention which has potential to increase the number of likes and retweets. However this study did not include "attention" and "dwell time" as variables to study the effect. The role of user attention network was found influential to the engagement level in a 2020 study on user modeling and reading satisfaction for news recommendation [38].

6.1.6.3 Predicted Engagement

Our cognitive model predicted that her tweets will most likely to receive “likes” with an overall scores of 0.89.

There is statistical significance between the correlation of the sentiment analysis and the overall engagement that Lisa Su receives. From what we can see, she has a high correlation between her sentiment analysis and to her engagement at a correlation of **0.86**. This matches up with what we saw in the raw numbers for her likes and retweets.

6.1.6.4 Comparing to Case Study Results

In the case study, one could observe Lisa Su received high level of user engagement, 209,290 interactions in total. Case study results also showed her tweets have an astonishing high positive sentiment (97%). The high sentiment is related to the most frequent use of words shown in the case study as happy, gratitude, proud, partner, great, love. This is consistent with the high sentiment score. However case study results do not provide insights as to which of the factors contribute to the high engagement.

The cognitive model results confirmed the strong influence of positive sentiment on engagement with a correlation score of 0.58. Through the results of the topic modeling and estimates, we saw the topics that are highly exciting and motivating which are often about the gaming industry inventions and new products with recognition of partners strongly increases the number of likes and retweets. In addition, the cognitive model results show the positive sentiment came from perceived value that a reader might have going over Lisa Su’s tweets. AMD is a company that focuses on hardware for gaming consoles, which is important to a lot of people. Her use of hashtags during gaming convention events likely also garnered a lot of attention, and as stated before, this is shown with the topic modelling having a correlation score of **0.63** to the user engagement.

6.1.6.5 Validating Results

Lisa Su	Likes	Retweets
Predicted as Likes	82	7
Predicted as Retweets	5	5

Table 23 – Validation Results for Lisa Su

True Positive (TP): According to the model, 82 tweets were correctly predicted as more likely to get likes compared to retweets.

True Negative (TN): According to the model, 5 tweets were correctly predicted as more likely to get retweets compared to likes.

False Positive (FP): According to this model, 7 tweets were incorrectly predicted as more likely to get likes compared to retweets.

False Negative (FN): According to this model, 5 tweets were incorrectly predicted as more likely to receive retweets compared to likes.

With these scores in mind, the accuracy rate for the cognitive model is 87%. Among this rate, the model is precise 95.51% of the time, and capable of getting a proper recall rate of 90.43% of the time.

6.1.7 Sundar Pichai – Cognitive Model Results

Here is the cognitive model results information for Sundar Pichai.

6.1.7.1 Topic Modelling Themes

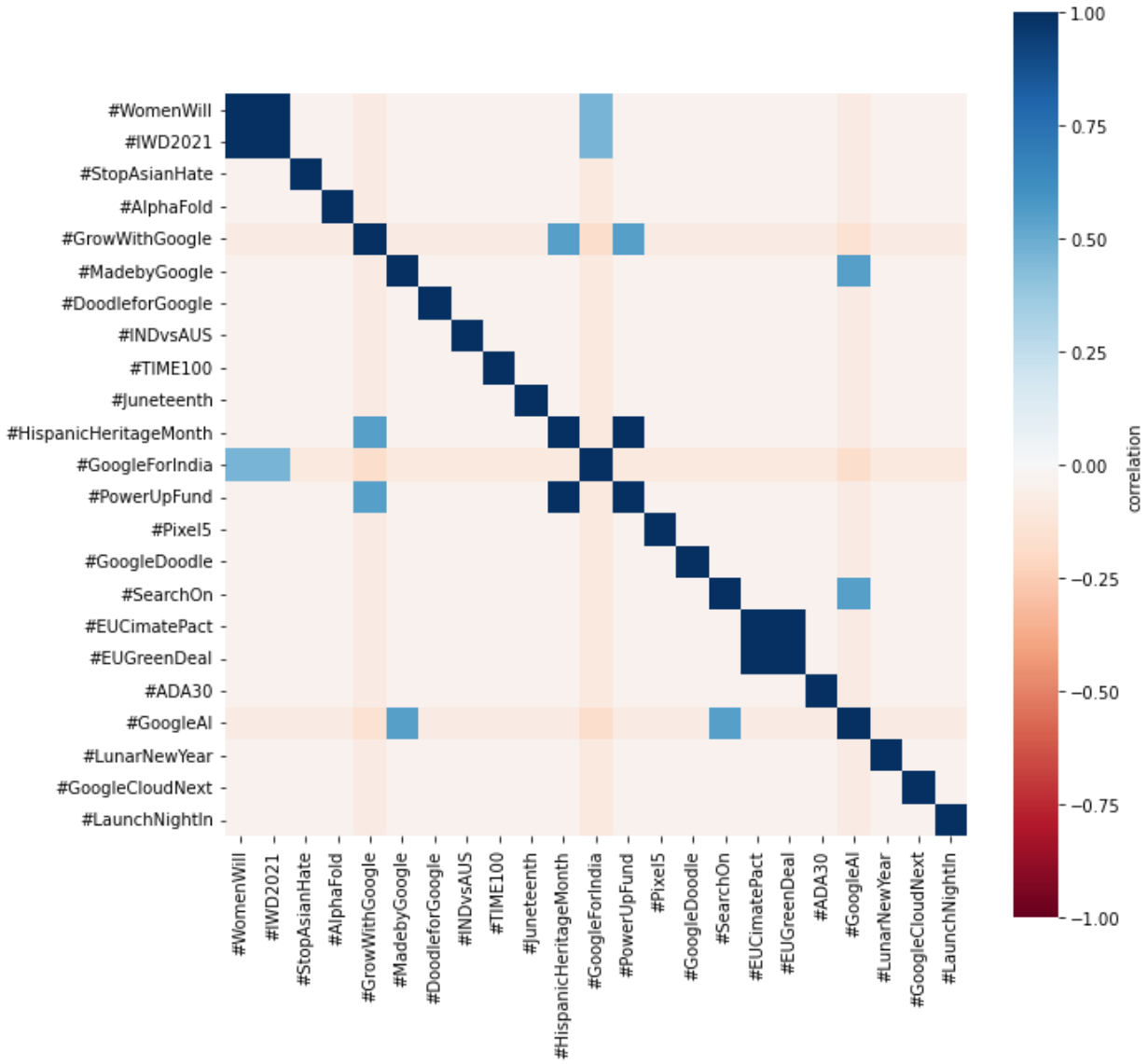

Figure 76 – Topic Modelling Heatmap (Sundar Pichai)

The topic modeling shows most of Sundar Pichai's hot topics are about "Google" We estimate the effect of the hot topic by obtaining the expected number of likes and retweets of a tweet without this topic to build a baseline. Then it is compared with the expected number of likes and retweets with the tweet with the hot topic. A percentage change is identified as indicated in the following table. The results provide evidence for the positive effect of the hot topic on likes and retweets.

Table 24 - Estimation Results for User Engagement – Topic and Themes (Sundar Pichai)

Tweet content- topic/theme	Likes	Retweets
Tweet w/o "hot topic" We applaud @POTUS's quick action on COVID relief, the Paris Climate Accord, and immigration reform. Google has supported action on these important issues & we look forward to working with the new administration to help the US recover from the pandemic + grow our economy.	16,883	948
Tweet w "hot topic" One of the greatest test series wins ever. Congrats India and well played Australia, what a series #INDvsAUS	316,512	41,416
Percentage change	94.67%	97.71%
Tweet w/o "hot topic" Yep, glad cricket is back!	10,598	684
Tweet w "hot topic" Today at #GoogleForIndia we announced a new $10B digitization fund to help accelerate India's digital economy. We're proud to support PM @narendramodi's vision for Digital India – many thanks to Minister @rsprasad & Minister @DrRPNishank for joining us.	101,250	24,080
Percentage change	89.53%	97.16%

6.1.7.2 Engagement Rate

Upon looking at his average engagement rate, we see that Sundar Pichai has an average engagement rate of about 0.3283.

When applying the cognitive model results to Sundar Pichai's tweets, we are introduced to some interesting aspects. He seems to match as the "most average" of all the male CEOs, in both the case studies and the cognitive model, topic seems to be a big factor into the engagement that Sundar Pichai gets.

6.1.7.3 Prediction Engagement

On average, when it came to the prediction engagement, Sundar Pichai received a score of 0.723, which leans more towards likes. Therefore, our cognitive model predicts Sundar Pichai is more likely to receive likes on his tweets.

Our model also found that sentiment has the strongest correlation on Sundar Pichai's engagement level (r = **0.6398**). This is a fairly moderate correlation, in comparison to the other content characteristics. Though based on the other correlations, there was as close of a correlation in the topic modelling (**r = 0.563**).

6.1.7.4 Comparing to Case Study Results

The case study results show that Sundar Pichai had 1,171,681 user interactions in the time frame. His tweets have positive sentiments (78%); his posting frequency is consistent and more evenly distributed throughout the observed time. The case study also shows he expressed his thoughts with well-structured sentences. However the case study results do not provide evidence for which factor most likely contributed to the high level of interaction.

The cognitive model results showed the strongest possible positive correlation to the level of engagement (r = 0.63) followed by use of trending topic (r = 0.62). Through looking at topic modeling, we estimated the changes of likes and retweets with tweet with and without. It is noted trending topics especially the one with perceived utility of entertainment affected the level of engagement with 94.6% jump in the number of likes compared to the previous post without the trending topic.

6.1.7.5 Validating Results

Sundar Pichai	Likes	Retweets
Predicted as Likes	68	11

Predicted as Retweets	16	5

Table 25 – Validation Results for Sundar Pichai

True Positive (TP): According to the model, 68 tweets were correctly predicted as more likely to get likes compared to retweets.

True Negative (TN): According to the model, 5 tweets were correctly predicted as more likely to get retweets compared to likes.

False Positive (FP): According to this model, 11 tweets were incorrectly predicted as more likely to get likes compared to retweets.

False Negative (FN): According to this model, 16 tweets were incorrectly predicted as more likely to receive retweets compared to likes.

With these scores in mind, the accuracy rate for the cognitive model to Sundar Pichai's tweets have a 84% accuracy rate of being predicted as more likely to receive likes. Among this rate, the model is precise 93.15% of the time and capable of getting the proper recall 86.08% of the time. When it comes to evaluating the numbers and tweets in our database, it has a pretty good accuracy, precision and recall rate.

6.1.8 Susan Wojcicki – Cognitive Model Results

These are the cognitive model results for Susan Wojcicki.

6.1.8.1 Topic Modelling Themes

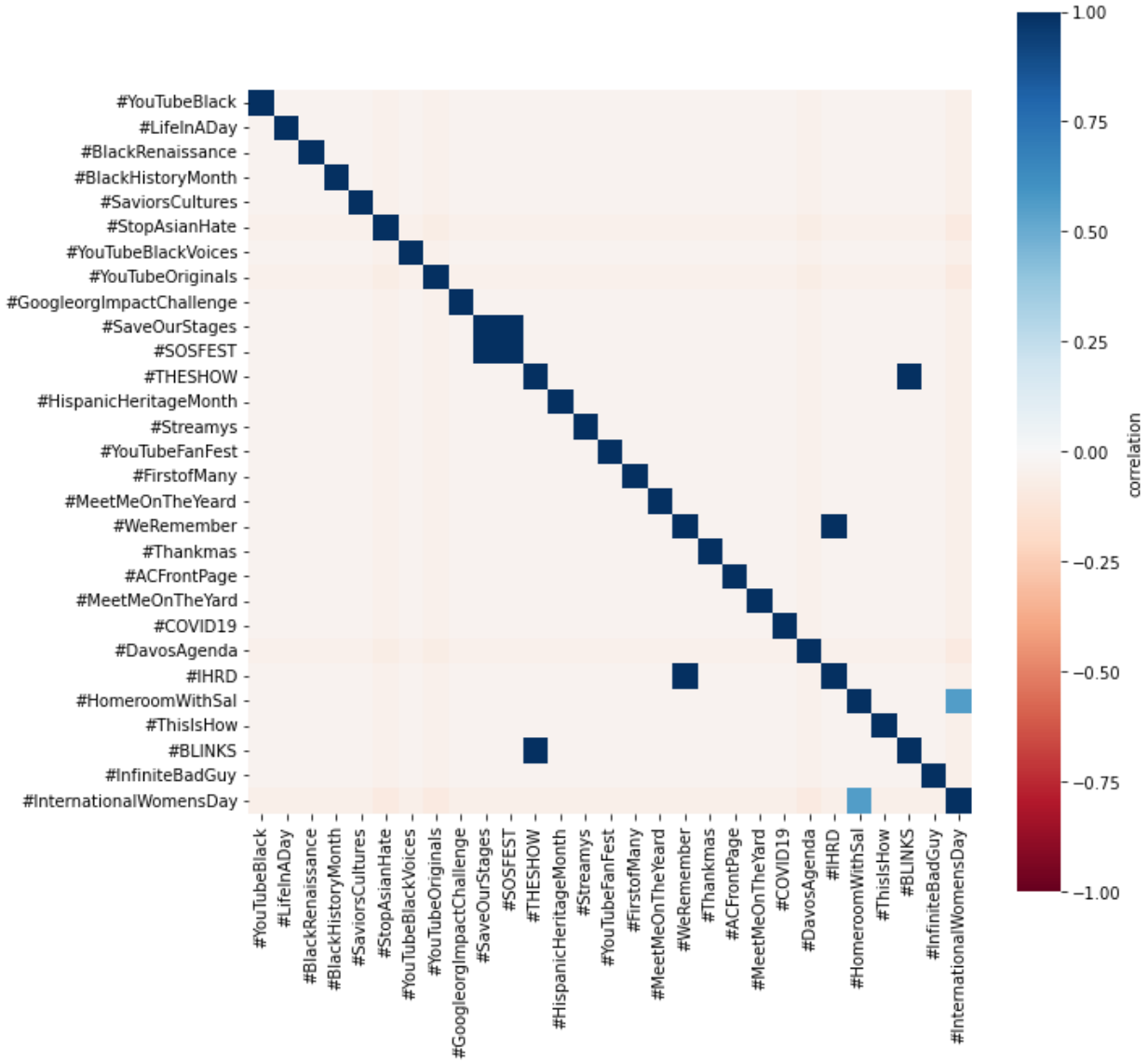

Figure 77 – Topic Modelling Heatmap (Susan Wojcicki)

When it comes to Susan Wojcicki, most of her tweets are overwhelmingly gratitude and entertainment. Being the CEO of YouTube, it is common that she will talk up a lot of people who regularly use the platform or celebrities who are hosting livestreams. In particular, she highlights the creators who have had quite a good time on the platform in terms of their involvement or numbers.

We estimate the effect of the hot topic by obtaining the expected number of likes and retweets of a tweet without this topic to build a baseline. Then it is compared with the expected number of likes and retweets with the tweet with the hot topic. A percentage change is identified as indicated in the following table. The results provide evidence for the positive effect of the hot topic on likes and retweets.

Table 26 - Estimation Results for User Engagement – Topic and Themes (Susan Wojcicki)

Tweet content- topic/theme	Likes	Retweets
Tweet w/o "hot topic" Check out @danielhowell's recap to see what world leaders – and @RobertDowneyJr! – and are saying at @wef about food shortages, populism & data security:	98	4
Tweet w "hot topic" So excited about #THESHOW tomorrow. Our teams have worked hard to make this livestream happen! #BLINKS, check out this exclusive peek into the behind-the-scenes prep in the work diary of YouTube's Sun Lee. @BLACKPINK	6145	1958
Percentage change	98.41%	99.80%
Tweet w/o "hot topic" Congrats @Valkyrae on Content Creator of the Year at The Game Awards!!! It is fantastic to see your hard work rewarded with this amazing honor.	4691	18
Tweet w "hot topic" Looking forward to #Thankmas, @Jack_Septic_Eye!	5617	77
Percentage change	16.49%	76.62%

6.1.8.2 Engagement Rate

Upon looking at Susan Wojcicki's numbers, her overall average engagement rate on all of her tweets is 0.2398. There are a lot of things that seem to affect the results that we get for Susan Wojcicki. The most important thing seems to be that content characteristic certainly has an effect.

6.1.8.3 Prediction Engagement

Overall, the prediction engagement for Susan Wojcicki's average prediction engagement is at 0.7902. It shows that she has a relatively moderate average prediction engagement that suggest her tweets are more likely to receive likes than they are retweets. This relates with the scores that we earned from Chapter 5 as well.

Out of all the content characteristics, we can see that the highest correlation from the content characteristics came from the sentiment analysis of her tweets. The correlation of the sentiment analysis scores to the engagement stands at 0.6644. This implies there is a moderate significance between them. Compared to the other content characteristics, those are the highest amounts.

6.1.8.4 Comparing to Case Study Results

The case study results show that Susan Wojcicki had 73,944 user interactions in the time frame. 70% of her tweets have positive sentiments; her posting frequency is frequent and consistent throughout the observed time. The case study also showed she expressed her thoughts with well-structured sentences. The most frequent word appeared in her tweets are brand related ones such as "YouTube" and "Creator". However the case study results do not provide evidence for which factor most likely contributed to the high level of interaction.

The cognitive model results provided some statistical results showing the modest positive correlation of emotion/sentiment to the level of engagement ($r = 0.66$) followed by use of trending topic ($r = 0.49$). Through looking at topic modeling, we estimated the changes of likes and retweets with tweet with and without. It is noted trending topics especially the one with perceived utility of "entertainment" affected the level of engagement with 98.1% jump in the number of likes compared to the previous post without the trending topic.

6.1.8.5 Validating Results

Susan Wojcicki	Likes	Retweets
Predicted as Likes	62	18
Predicted as Retweets	9	11

Table 27 – Validation Results for Susan Wojcicki

True Positive (TP): According to the model, 62 tweets were correctly predicted as more likely to get likes compared to retweets.

True Negative (TN): According to the model, 11 tweets were correctly predicted as more likely to get retweets compared to likes.

False Positive (FP): According to this model, 18 tweets were incorrectly predicted as more likely to get likes compared to retweets.

False Negative (FN): According to this model, 9 tweets were incorrectly predicted as more likely to receive retweets compared to likes.

With these scores in mind, the accuracy rate for the cognitive model is 71%. Among this rate, the model is precise 87.32% of the time, and capable of getting a proper recall rate of 77.5% of the time. This is a fairly moderate for the accuracy rate onto Susan Wojcicki's number.

6.1.9 Tim Cook – Cognitive Model Results

The following section has the cognitive model results that we had received for Tim Cook.

6.1.9.1 Topic Modelling Themes

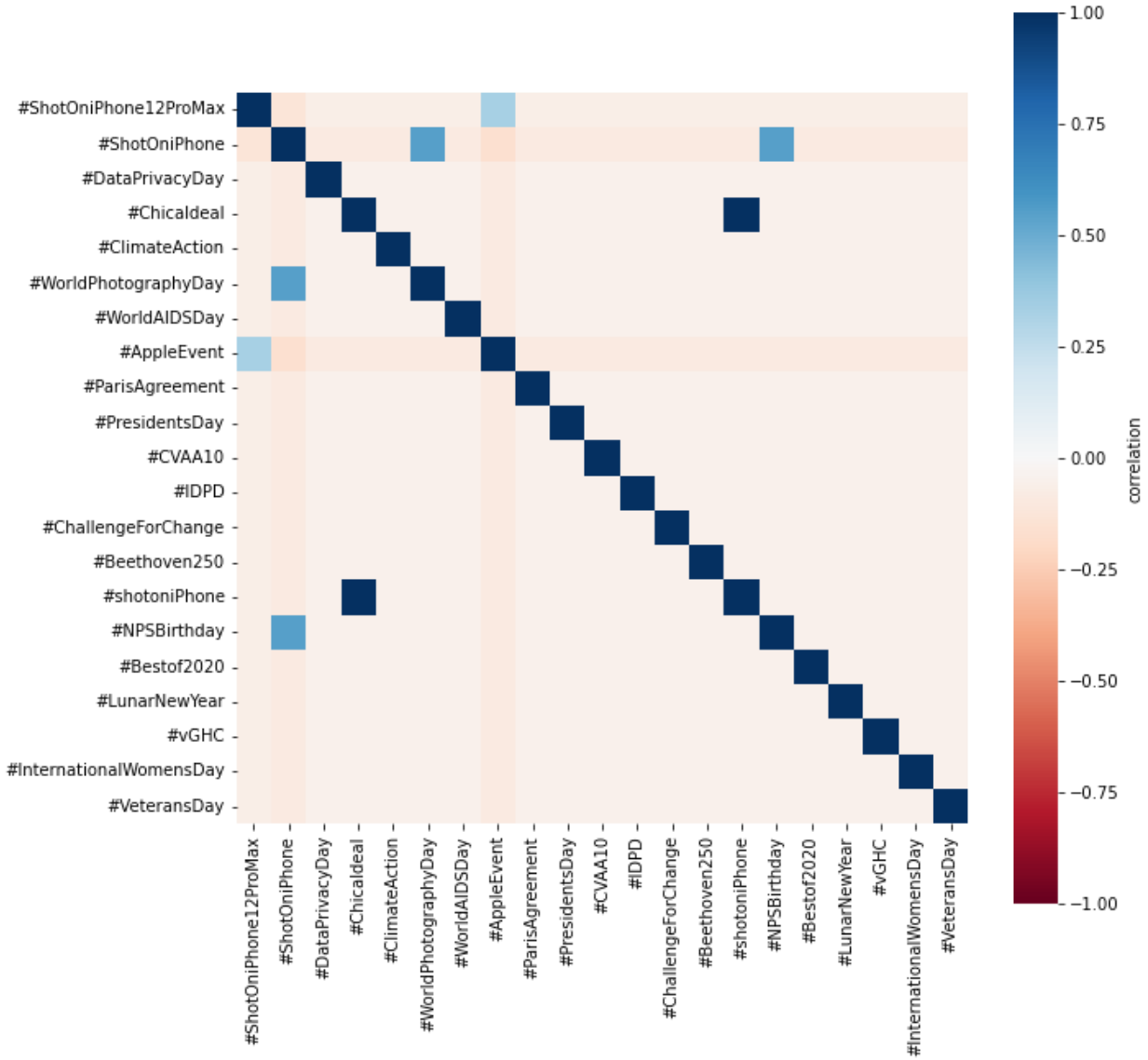

Figure 78 – Topic Modelling Heatmap (Tim Cook)

From looking at the modelling heat map for Tim Cook, a lot of the topics are informational in some way. Whether it be celebrating a specific holiday like Lunar New Year or Veterans' Day, or talking about the iPhone for business purposes, Cook cares immensely about the business aspect on his Twitter. That being said, there are some tweets based more in the entertainment aspect of things.

We estimate the effect of the hot topic by obtaining the expected number of likes and retweets of a tweet without this topic to build a baseline. Then it is compared with the expected number of likes and retweets with the tweet with the hot topic. A percentage change is identified as indicated in the following table. The results provide evidence for the positive effect of the hot topic on likes and retweets.

Table 28 - Estimation Results for User Engagement – Topic and Themes (Tim Cook)

Tweet content- topic/theme	Likes	Retweets
Tweet w/o "hot topic" "It's time to get started!" (Informational, leads to a link to the event mentioned in the tweet w/ "hot topic")	28,289	4,123
Tweet w "hot topic" Good morning! T-minus 6 hours! #AppleEvent	101,904	11,598
Percentage change	72.24%	64.45%
Tweet w/o "hot topic" Tiffany Williams is a veteran, a principal, and an inspiration. I've had the honor of visiting Tuskegee Public School, where she's using technology to preserve the legacy of her community and the heroic Tuskegee Airmen through storytelling on iPad." (Informational; to share the legacy of this person)	2,028	238
Tweet w "hot topic" One more sleep. #AppleEvent #ShotOniPhone12ProMax	54,127	3,482
Percentage change	96.25%	93.16%

6.1.9.2 Engagement Rate

When checking the average engagement rate for Tim Cook on his tweets, we see that his average engagement rate is 0.114. He manages to separate himself from the other CEOs, and does not have any potential overlap of data.

6.1.9.3 Prediction Engagement

After putting Tim Cook's tweets through the cognitive model, we see that his average engagement rate predictor has a scores of 0.62. This means that his tweets are more likely to receive likes as a form of engagement, compared to retweets.

The correlation numbers are a different manner, as aspects such as the subjectivity and polarity have the highest correlation out of all of his numbers (**r = 0.60**). This shows considering it is likely that a CEO would want to show enthusiasm for the product. In addition, since he has a lot of tweets with positive sentiments behind them, it attracts a lot more attention.

There is a possibility that this number would be much higher due to the fact he has a couple of tweets incorrectly marked as Negative when they would be Neutral or Positive.

6.1.9.4 Comparing to Case Study Results

The case study results show that Tim Cook had 1,449,617 interactions. 80% of his tweets have positive sentiments (80%); his posting frequency is frequent and consistent throughout the observed time. The case study also showed he expressed her thoughts with well-structured sentences. The most frequent word that appeared in his tweets are "apple" related and words like "support" "thank you". However the case study results do not provide evidence for which factor most likely contributed to the high level of interaction.

The cognitive model results told us the most positive correlation score came from subjectivity (r = 0.60) providing evidence that the brand "apple" and his CEO position on the brand most affect the likes and retweets received. Once again, the perceived utility of the tweets being informative about a highly sought after product coming from the CEO might have generated excitement which results in more positive engagement behaviours. Topic modeling as part of the cognitive model confirmed the same finding. It is noted trending topics about Apple Event affected

the level of engagement with 96% jump in the number of likes compared to the previous post without the trending topic.

6.1.9.5 Validating Results

Tim Cook	Likes	Retweets
Predicted as Likes	68	12
Predicted as Retweets	6	14

Table 29 – Validation Results for Tim Cook

True Positive (TP): According to the model, 68 tweets were correctly predicted as more likely to get likes compared to retweets.

True Negative (TN): According to the model, 14 tweets were correctly predicted as more likely to get retweets compared to likes.

False Positive (FP): According to this model, 12 tweets were incorrectly predicted as more likely to get likes compared to retweets.

False Negative (FN): According to this model, 6 tweets were incorrectly predicted as more likely to receive retweets compared to likes.

With these scores in mind, the accuracy rate for the cognitive model is 74%. Among this rate, the model is precise 82.93% of the time, and capable of getting a proper recall rate of 85% of the time.

6.1.10 Whitney Wolfe-Herd – Cognitive Model Results

These are the cognitive model results for Whitney Wolfe-Herd.

6.1.10.1 Topic Modelling Themes

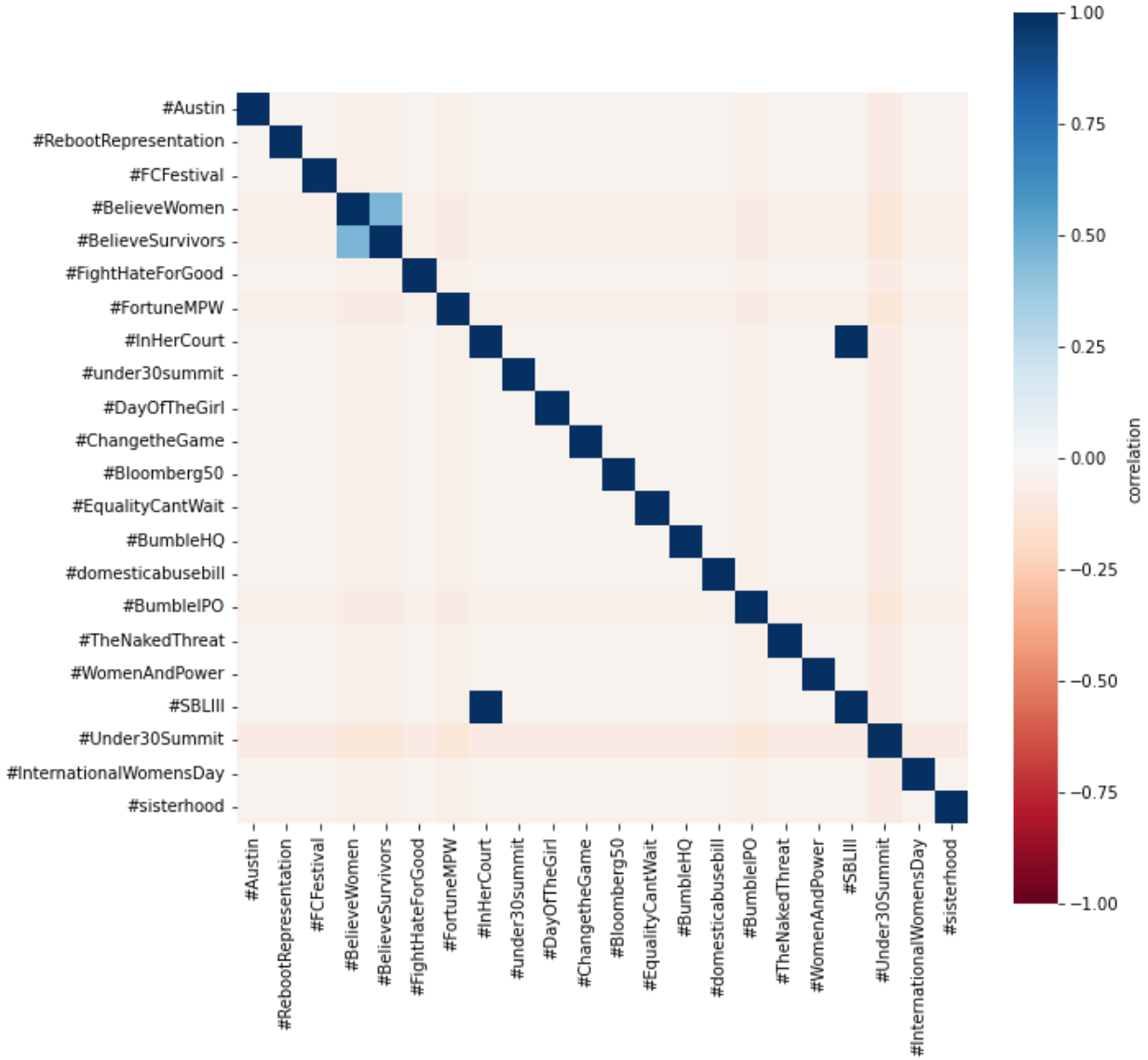

Figure 79 – Topic Modelling Heatmap (Whitney Wolfe-Herd)

Whitney Wolfe-Herd is very vocal in her support of social movements, particularly with feminist movement which puts a lot of her topics under inspirational. The topics reflect this a lot, many of them being supportive of women and advocating for women's rights. A common topic within her tweets is celebrating the Excellency of women, and celebrating women's success.

We estimate the effect of the topics by obtaining the expected number of likes and retweets of a tweet without this topic to build a baseline. Then it is compared with the expected

number of likes and retweets with the tweet with the hot topic. A percentage change is identified as indicated in the following table. The results provide evidence for the positive effect of the hot topic on likes and retweets.

Table 30 - Estimation Results for User Engagement – Topic and Themes (Whitney Wolfe Herd)

Tweet content- topic/theme	Likes	Retweets
Tweet w/o "hot topic" This week, @Bumble is making donations to @PPFA and the @ACLU in honor of Supreme Court Justice Ruth Bader Ginsburg's fight for gender equality. Thank you, RBG. The work continues.	267	25
Tweet w "hot topic" Today, @Bumble becomes a public company. This is only possible thanks to the more than 1.7 billion first moves made by brave women on our app — and the pioneering women who paved the way for us in the business world. To everyone who made today possible: Thank you. #BumbleIPO (Gratitude)	10,561	1384
Percentage change	97.47%	98.19%
Tweet w/o "hot topic" https://t.co/X09SsDjucc (Informational, leads to a link about Whitney Wolfe-Herd's appreciation of Serena Williams)	80	11
Tweet w "hot topic" Today, @Bumble's first US TV ad will air during #SBLIII. While the message is women-first, on a very male-dominated day in sports, we're truly striving for equality for all genders. #InHerCourt	1069	245
Percentage change	92.52%	95.51%

6.1.10.2 Engagement Rate

When checking the average user engagement rates for Whitney Wolfe-Herd, we see that it shows a rate of 0.9039%. It seems that she has a generally active rate, although this might be heavily affected by her follower count.

6.1.10.3 Prediction Engagement

The average engagement for the prediction engagement would be at **0.633**. This means that her tweets are more likely to receive likes.

The main correlation score that we saw for Whitney Wolfe-Herd came with her topics and sentiment. A significant relation between the engagement rate and sentiment analysis is quite strong, and the highest compared to the others. This is likely due to how passionate she is about the subjects that she talked about.

6.1.10.4 Comparing to Case Study Results

The case study results show that Whitney Wolfe-Herd's tweets scored moderate positive sentiment score of 67%. It is noted that one of the hashtag tweet was "miscoded" as "neutral" while it should have been "positive". She had 26,782 user interactions in the selected time frame. During the time, she also was more active at the beginning of the time and much less so later on. The case study also showed she expressed her thoughts with well-structured sentences. The most frequent word appeared in her tweets are Women and "Thank" "Bumble" (which is her business brand). However the case study results do not provide evidence for which factor most likely contributed to the level of interaction of 26,782.

The cognitive model results provided some statistical results showing the modest positive correlation of emotion/sentiment to the level of engagement ($r = 0.65$) followed by use of trending topic ($r = 0.57$). Through looking at topic modeling, we estimated the changes of likes and retweets with tweet with and without. In this case, the cognitive model results indicate that the use of hashtag positively influence the number of likes and retweets. It is noted using hashtag especially the one inspiring women had a jump of 97.5% jump in the number of likes compared to the previous post without the hashtag.

6.1.10.5 Validating Results

Whitney Wolfe-Herd	Likes	Retweets
Predicted as Likes	60	11
Predicted as Retweets	14	15

Table 31 - Validation Results for Whitney Wolfe-Herd

True Positive (TP): According to the model, 60 tweets were correctly predicted as more likely to get likes compared to retweets.

True Negative (TN): According to the model, 15 tweets were correctly predicted as more likely to get retweets compared to likes.

False Positive (FP): According to this model, 11 tweets were incorrectly predicted as more likely to get likes compared to retweets.

False Negative (FN): According to this model, 14 tweets were incorrectly predicted as more likely to receive retweets compared to likes.

With these scores in mind, the accuracy rate for the cognitive model is 74%. Among this rate, the model is precise 81.08% of the time, and capable of getting a proper recall rate of 84.51% of the time.

6.1.11 Cognitive Model Results Summary

In general, when looking at the results of the cognitive model, it is observed that the content characteristics with the most influence on the user engagement tends to be between the sentiment analysis score and topic modelling. The moderate to high correlation that those characteristics receive in comparison is noteworthy – and certainly shows what seems to be the most important thought for users [Table 23].

CEOs	Sentiment	Subjectivity	Topic	Class.	Time	L.R.	Nouns	Verbs	Adj.
Bill Gates	0.52	0.50	0.48	0.39	0.36	0.31	0.34	0.25	0.26
Elon Musk	0.50	0.68	0.53	0.44	0.41	0.38	0.43	0.21	0.21
Jeff Bezos	0.73	0.32	0.51	0.41	0.37	0.33	0.45	0.23	0.24
Corie Barry	0.55	0.44	0.50	0.49	0.43	0.34	0.41	0.22	0.21
Ginni Rometty	0.65	0.48	0.50	0.42	0.45	0.39	0.43	0.24	0.22
Lisa Su	0.58	0.54	0.63	0.46	0.42	0.35	0.42	0.24	0.23
Sundar Pichai	0.63	0.55	0.62	0.40	0.39	0.36	0.45	0.21	0.20
Susan Wojcicki	0.66	0.52	0.49	0.39	0.38	0.35	0.46	0.22	0.22
Tim Cook	0.58	0.60	0.64	0.40	0.37	0.33	0.43	0.23	0.21
Whitney Wolfe-Herd	0.65	0.53	0.57	0.43	0.36	0.35	0.41	0.22	0.20

Table 32 – Correlation of Content Characteristics to Engagement

There is also some evidence that the perceived utility of topic content as a behaviour driver. Following the topic modelling, we estimated the effect it has on user engagement. We noted that trending topics affects the number of likes and retweets, determining how much engagement there is. For example, writing a tweet with a trending topic makes it more like for users to notice and engage with it – especially when the message is perceived as informative or entertaining. We found a similar results for the biggest outlier in the data, Elon Musk, where we estimated the impact through looking at a tweet with an image and a tweet without. We found as consistent results as the others.

As stated before, this is something that the case study simply could not show us. The detail and idea that we get about the correlation to the engagement that we receive is very important. It is easy to say that the case study can help us determine that, but it simply does not go into the same detail as the cognitive model. We cannot completely determine from the case study what makes something popular. For example, we can look at the time of day all we want, but likely because of

confounding factors, they do not have to worry about what time they post about something unless it pertains to an event like Tim Cook and Lisa Su do.

6.2 Comparison Summary

With all of our cognitive results put down, now we summarize our comparisons and findings after all is said and done.

6.2.1 Predicted Engagement: More likely to receive likes

When looking at our average engagement rates, overall, we see that they are more likely to lean towards other users engaging with content by liking it rather than retweeting. Most of the scores lean more in favor of likes. In the case study, we observed similar trends and patterns – that a tweet is more likely to receive more likes than retweets. However, the results from the cognitive study clearly indicated that the content characteristics – especially topic and sentiment – more likely influenced the likelihood of a tweet receiving likes.

On the other hand, it is also likely that the actions of followers wind up conflating the number of likes and affect scores we see with other content characteristics. When a tweet is retweeted, the time that it was originally posted does not matter because a new iteration exists on an individual user's account.

CEOs	Predicted Engagement
Bill Gates	0.75
Elon Musk	0.82
Jeff Bezos	0.70
Corie Barry	0.44
Ginni Rometty	0.66
Lisa Su	0.90
Sundar Pichai	0.72
Susan Wojcicki	0.80
Tim Cook	0.62
Whitney Wolfe-Herd	0.63

Table 33 – The Predicted Engagement for all the 10 CEOs

6.2.2 Topic and Theme could generate positive emotions which could lead to likes

Based on what we gathered from the topic modelling process with the 938 tweets, we see that topics and themes can lead to a higher generation of likes. Taking all of the tweets into account, we created a heat map to indicate the unique topics among all of them, whenever they appeared two or more times. In total, the heat map indicated there are 36 unique topics but also indicated that there was no particular topic shared between the CEOs that they expressed their thoughts or opinions on.

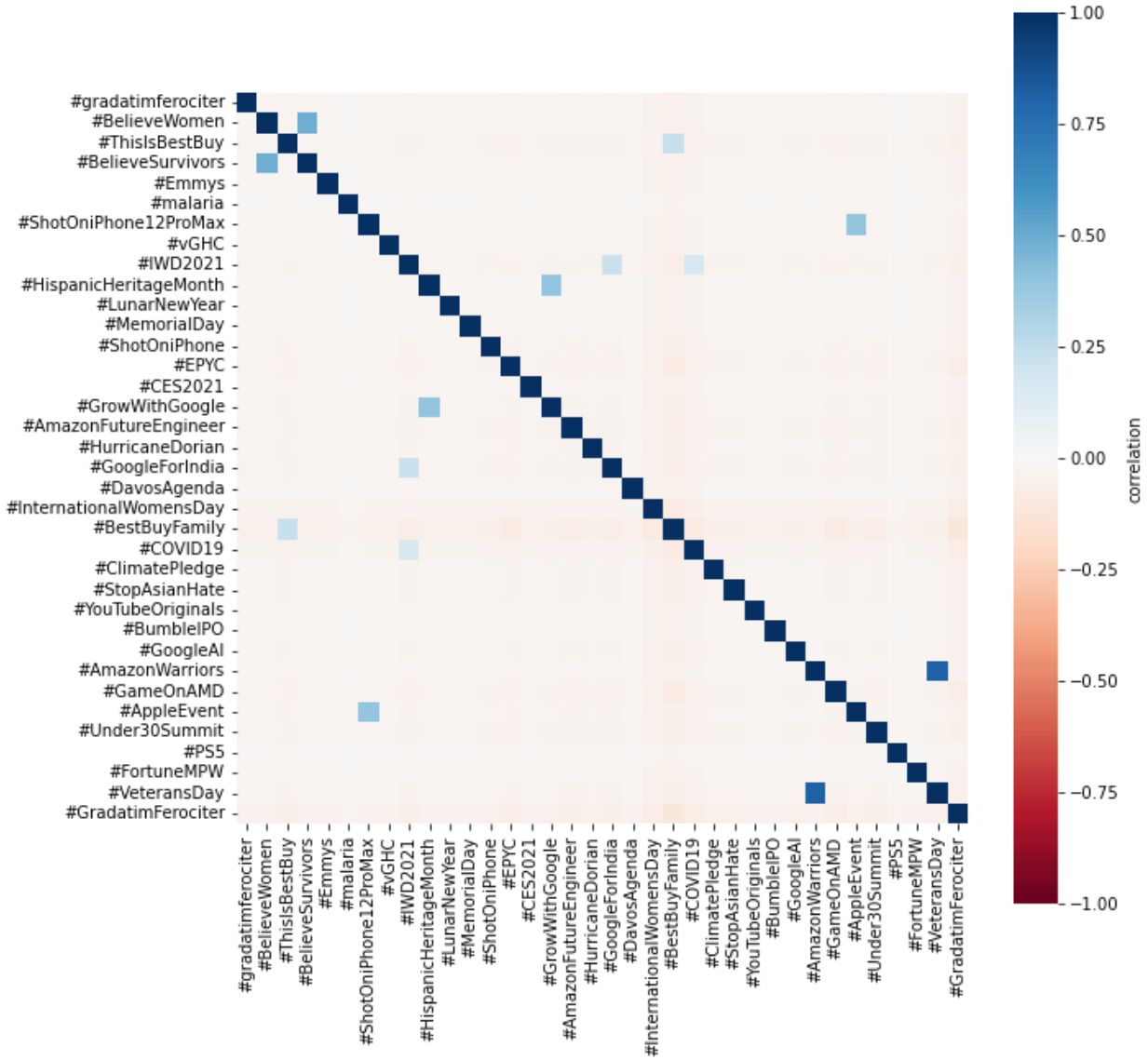

Figure 80 - Overall Topic Modelling Heatmap for all CEOs (lists hashtags that appear twice)

Thanks to determining the overall feelings of the forms of engagement, we have concluded that the topic and theme of a tweet has a high correlation with the engagement between users. Based on the analysis that we ran, we were able to map the tweets in three different categories: Informative tweets (Informative), tweets expressing gratitude (Gratitude) and entertaining tweets (Entertainment) [15] [25].

Due to the people in our database being CEOs, a lot of the tweets leaned much more on the informative. This can count as anything that is meant to convey information to the public, such as tweets that talk about diseases or advertising business conferences. In particular, considering our time period, there were a lot of tweets about medicine and trying to share information about COVID-19. As a result, a lot of the tweets that are informative have a higher effect on overall user engagement. Tweets that were entertaining could be about celebrities that the CEO had met, or related to something that is much more jovial. Just like the informative one, and considering the usual use of social media, tweets that are entertaining also have a higher effect on the user engagement on their own. Gratitude is self-explanatory, as it makes up a majority of the tweets that were replies, with people giving their "Thanks" or "Congratulations" for what they have done. Since a lot of them are replies, we see that the engagement is not as high in comparison.

6.2.3 Sentiment Analysis is Moderate to the Changes of Engagement level

There are two different levels when it comes to evaluating the sentiment analysis and its relation to engagement level. As a whole, sentiment analysis is a content characteristic that correlates very highly with the user engagement of each of the CEOs. The sheer effect of positive sentiment analysis helped to raise the average correlation of the positive sentiment scores to 0.73. When it came to the relation of neutral tweets to the engagement level, we also saw that there was a moderately high correlation for the neutral tweets and negative tweets and how they affect engagement. This proves that a negative sentiment tweet does not necessarily imply that the overall engagement rate will be low.

It is important to point out the limitations of current method to detect sentiment from Twitter tweets. Human emotion is complex and sentiment expressed in opinions can also be hard to detect through machine learning or Twitter API. For example, detecting sarcasm remains to be

a challenge for automatic detection data tools. Machine learning and other automatic tool also need improvement detecting sentiment from image, memes and videos. In the case of Elon Musk, the way he writes his tweets can make it hard to determine a specific sentiment from them. These challenges are illustrated in the following illustration.

Elon Musk	@Jason cgi irl @spacex
Elon Musk	@PPathole @SpaceX SpaceX team is doing great work! One day, the true measure of success will be that Starship flights are commonplace.
Elon Musk	@Erdayastronaut @SpaceX RIP SN10, honorable discharge
Elon Musk	Starship SN10 landed in one piece! https://t.co/lO4AF47MaN
Elon Musk	Launch abort on slightly conservative high thrust limit. Increasing thrust limit & recycling propellant for another flight attempt today.
Elon Musk	5 mins to Starship test flight attempt
Elon Musk	March 3rd
Elon Musk	Starship launch tomorrow. Window opens at 9am.
Elon Musk	@LuvMyModel3 @billhuang688 @F9Block5 @PPathole Yes, beta will be offered much more widely when version 9 is done, hopefully next month
Elon Musk	RT @yousuck2020: ã€ç™ºè¡¨ã€'2023å'´ã«ãºˆå®šã—ã¦ã„ã‚‹ãœœˆæ—…è¡Œã«ã®æ§˜ã‚'ã"æ‹å¼…ã—ã¾ã™ã€,æœ¬æ—¥ã,ˆã,Šå¿œå‹Ÿå—ã»˜é–‹å§‹ã—ã¾ã™ï¼ã¿ã“ã®ã§æ¥½ã—ã„æ—…è¡Œã«ã—ã¾ã—ã‡ã†ðŸš€ðŸŒðŸŒ• #dearMoon https://t.co/xBtOyJMIUh
Elon Musk	@billhuang688 @F9Block5 @PPathole FSD beta build V8.1 normally drives me around with no interventions. Next versionâ€¦ https://t.co/y9MwQQ1JxR
Elon Musk	@LabPadre @SpaceX Youâ€™re welcome

Figure 81 – Detecting sentiment – Is it positive, negative or neutral?

There was another interesting aspect about the effect of topic and theme to user engagement, and that was the number of times it also correlated well with the sentiment analysis. When it came to tweets with positive sentiments, there also happened to be a moderate to high effect from the topic that the tweet is about to the user engagement. This was not a subject we tested, though it is interesting to see just how similar the effect for sentiment analysis and topic and tweet theme were to the user engagement.

6.2.4 Summary of Other Content Characteristics and Relation to Predicted Engagement

While the other content characteristics are not as significant in relation to predicted engagement, there are some interesting observations to gather from them.

When calculating the relation of predicted engagement rate to the time of day, on average, it did not show a very high correlation. The correlation leaned towards weak or moderate, though on average, there was significantly more effect during the afternoon hours. While it seems

significant, it can also be argued that it is the most affected by confounding variables thanks to other content characteristics such as the existing likes and retweets count.

Lexical richness and the parts of speech had some of the weakest statistical significance out of all of the content characteristics. The effect and results were very inconsistent, with sudden spikes for some of the tweets such as Jeff Bezos, but then low scores in others like with Elon Musk. Lexical richness tended to affect them the most, by typically having correlations scores from 0.4 to 0.67. However, it is unclear if it just from lexical richness alone or if it is affected by aspects such as the noun count. Verb count and adjective count, however, ranked the lowest overall, typically having correlation scores between 0.2 to 0.49.

Chapter 7 – Conclusions, Limitation and Future Research

7.1 Conclusion

In this thesis, the concept of user to user engagement on Twitter is examined using 938 tweets posted by 10 CEOs from February 2018 to March 2021. Findings from this research provide evidence to specifically answer research questions: what content characteristics are found in user-to-user engagement, and in particular which of the factors affect the engagement levels. The review of recent literature on this topic showed despite a growing body of research on user engagement, such research is mostly done within the context of marketing research and brand recognition. While exploring existing methods to define and measure user engagement, it is noted that the topic of user to user engagement is fast expanding to include investigation of psychological, social and cognitive factors that potentially influence the direction of user engagement. It is noted that since 2020, there is a fast growing research exploring on-line social cognition and its effect on emotional responses and engagement behaviours. Findings from the proposed case study and a cognitive inspired model using feature engineering provide evidence that allow the following conclusions.

The results of case study unveiled useful statistics, trends, and patterns in the content of their tweets. Their tweets have predominantly positive sentiment, well-structured language usage, and uses of words of "gratitude" to express their thoughts. The CEOs posting habits are not consistent. Some post more often while others post more sporadically. However the level of engagement does not seem to be related to the time they tweet.

Results through our cognitive inspired machine learning model provided insights to the questions as to what factors motivate user behaviour of liking and retweeting. We draw several conclusions from these results. First, content characteristics, especially "topic and theme" and sentiment analysis, could influence the direction of engagement behaviours in the forms of likes

and retweets. Using hashtags (trending topics) could affect the behaviour of likes and retweets. Second, our cognitive inspired model provide evidence that positive sentiment has moderate to strong positive correlation to engagement level.

Perhaps the most important conclusion we can draw is the positive impact of the cognitive perception on likes. Although not statistically evaluated, cognitive evaluation, affect commitment (emotional responses) and liking and sharing behaviours could be used to explain that cognitive evaluation of topic, positive emotion generated could lead to behaviours in forms of likes. Results from the cognitive model showed initial indications for looking into the effects of emotional responses when perceiving certain utility of the tweets.

7.2 Benefits of This Research

This research is a pioneer study that tests out user-to-user engagement with a multi-method design and a cognitive lens. It highlights the role of cognitive perception and evaluation of users when they interact with the content. That perception may provoke positive feelings which may provokes emotional response as liking and sharing. It addresses the gaps that machine learning and quantitative approach cannot do accurately yet [4][6][7]. The research considers cognitive-emotion-behaviour triad and logic to design a model that look into content features, linguistic features to determine what emotions the tweets could generate, and the likelihood of them affecting engagement. The benefits of this research is we can use cognitive theory to explain outliers like Elon Musk, Corie Barry and Ginni Rometty.

Investigating perception of ideas/thoughts could then be recast into the problem of analyzing how semantic association was made around topics in social discourse. The lens of cognitive perception and semantic frame could be used to develop highly effective online discourse materials to promote public health and other public policy measures. Understanding the cognitive

mechanism behind how content is reconstructed could help unearth concepts used to promote misinformation, and therefore develop counter strategies.

7.3 Limitations of the Research

Several limitations are associated with this research. First our database consists of 10 tech elites who are current or former CEOs of giant US and Canadian companies. This brings in confounding factors that are beyond the control of this research. For example, their personal influence on users' action and behaviour is huge. The ties between the elite users such as the CEOs and general users might have dictated the level of engagement they receive regardless of the topic similarity, perception of theme function or the emotional responses or posting times. We have definitely seen the influence of their persona in the case of Elon Musk. No matter what he posts and at whatever frequency, Elon Musk gets a huge number of likes and retweets [57]. Secondly, the selection of female CEOs accounts for the purpose of balancing gender representation might have further skewed our data as they have much smaller followers count, as well as smaller number of likes and retweets. As well two of the accounts belong to the female CEOs are new or not frequently used. This research does not include gender difference as a variable, therefore the assessment about the use of words, language pattern may not be accurate without including the gender lens [58]. Lastly, the study is limited and suffered some computational issues in classifying discrete and categorical data such as labelling the themes from the topic. The study also suffers from some computational issues in characterizing the dimensionality of emotional responses. Due to these limitations, findings from our model may not be applicable to other set of user-to-user engagement relationship.

Nevertheless, this research could still contribute to the understanding of the user to user engagement by making aware the interconnection of thought, emotion and behaviours. The

proposed cognitive inspired machine learning model has potential to improve predictive power as it does not predict engagement in isolation.

7.4. Future Work

This research could spark future efforts, concepts, new database and more complex hybrid architecture including using AI and more advanced data analytic tools to unveil additional factors that contribute to user engagement.

First, future efforts could start with a better dataset, one that include the followers of different ages, gender and background. This would reduce the sampling biases. The small dataset and only the posting users not the follower users only tell one side story.

Second, while sentiment analysis is a key part of social media research, our study could have conducted the subjectivity affectivity with a word embedding model that identifies "semantically similar words [that] capture the context of words" and enabled the researchers to "extract domain specific words as well as synonyms" [7]. A word embedded model could improve the sentiment analysis a lot because the nuances of text can be hard to determine by a computer model. There are a few tweets in our data incorrectly marked by the program. Some examples of tweets use words in the context where it would actually be a positive sentiment. However, due to the language used, and how the vocabulary is typically used, the program flags it as a negative sentiment. As a result, word embedding and perhaps the natural language processing need to be improved to train the machine to pick up more nuance in emotion in text. We run into the same problem as many of our research articles ran into, where there is a lack of nuance in the emotion depicted in the text.

Finally, for future research, there should be some control when selecting sample of "content creator". For example, future research should carefully consider gender, age and other

characteristics of the content creators. In our sample, the discrepancy in number of followers, likes and retweets is huge, making direct comparisons numerically and visually, difficult.

Chapter 8 - Bibliography

1) L. Sigerson and C. Cheng. 2018. *Scales for Measuring user Engagement with Social Network Sites: A systematic review of psychometric properties.* Computers in Human Behaviour, 83, 10.1016/j.chb.2018.01.023
2) Diaz-Faes et al, 2019. *Towards a Second Generation of Social Media Metrics*. PLoS One, 14(5), doi: 10.1371/journal.pone.0216408
3) Borges-Tiago, Maria T. Flavio Tiago, Carla Cosme. (2019). *Exploring user's motivations to participate in viral communication social media*. *Journal of Business Research. 574–582*(2017): 43-49. https://doi.org/10.1016/j.jbusres.2018.11.011
4) Melanie Schreiner M., Thomas Fischer, Rene Riedl, 2021, T., and Ried, 2021. *Impact of content characteristics and emotion on behavioral engagement in social media*: literature review and research agenda, Electronic Commerce Research, 21, 329-345. https://link.springer.com/article/10.1007/s10660-019-09353-8
5) Sylvia M. Chan-Olmsted, Lisa-Charlotte Wolter, 2018, *In Perception and Practices of Social Engagement: A global perspective,* International Journal on Media Management, 20:1, 1-24 DOI: 10.1080/14241277.2017.1402183
6) Stella, Massimo 2021, *A Cognitive Network Science for Understanding Online Social Cognitions: A Brief Review*, Topics in Cognitive Science, DOI: 10.1111/tops.12551
7) Adikari, Achini, Daswin De Silva, Damminda Alahakoon, Singhuo Yu, 2019, *A Cognitive Model for Emotion Awareness in Industrial Chatbots*, IEEE Xplore, 10.1109/INDIN41052.2019.8972196

8) Luis V.Casaló, Carlos Flavián, Sergio Ibáñez-Sánchez, 2021. *Be creative, my friend! Engaging users on Instagram by promoting positive emotions*, Journal of Business Research, 130, 416-425, https://doi.org/10.1016/j.jbusres.2020.02.014

9) Li, Yiyi and Ying Xie, 2019, *Is a Picture worth a Thousand Words? An Empirical Study of Image Content and Social Media Engagement*, SAGE, DOI: 10.1177/0022243719881113; journals.sagepub.com/home/mrj

10) Mahmud Jalal, Jilin Chen, Jeffrey Nichols, 2014. *Why are you more engaged? Predicting Social Engagement from Word Use,* IBM Research San Jose

11) Wallace Chipidza and Jie Yan, 2021; *Does Flagging POTUS's Tweets Lead to Fewer or More Retweets? Preliminary Evidence from Machine Learning Models, Claremont Graduate University*. SocArXiv Papers, 10.31235/osf.io/69hkb

12) Diana Rieger and Christoph Klimnt, 2019, *The Daily Dose of Digital Inspiration 2: Themes and Affective user responses to meaningful memes in social media.* Journal of Media Psychology, 21 (10). DOI: 10.1177/1461444819842875

13) Haupt, Michael and Alex Jinich-Diamant, Li Jiawei, Mattew Nali, Tim K. Mackey, 2021. *Characterizing twitter user topics and communication network dynamics of the "Liberate" moment during Covid-19 using unsupervised machine learning and social network analysis*. Online Social Networks and Media, 21, https://doi.org/10.1016/j.osnem.2020.100114

14) Aldous, K. K., An, J., & Jansen, B. J. (2019). *View, Like, Comment, Post: Analyzing User Engagement by Topic at 4 Levels across 5 Social Media Platforms for 53 News Organization*s. Proceedings of the International AAAI Conference on Web and Social Media, 13(01), 47-57. https://ojs.aaai.org/index.php/ICWSM/article/view/3208

15) Sadri, A.M., & Hasan, S.*, 2019, Joint Inference of User Community and Interest Patterns in Social Media Networks, Social Network Analysis and Mining;,* Social Network Analysis and Mining, 9 (1), https://doi.org/10.1007/s13278-019-0551

16) Ellen Simpson, 2018 *Integrated & Alone: The use of Hashtags in Twitter Social Activism,* CSCW '18: Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing, 237–240, https://doi.org/10.1145/3272973.3274064

17) Seong Jae Min, 2019. *From algorithmic disengagement to algorithmic activism: charting social media users' response to news filtering algorithms*, Telematics and Informatics. 43, 101251, https://doi.org/10.1016/j.tele.2019.101251

18) Brena, G., Brambilla, M., Ceri, S., Di Giovanni, M., Pierri, F., & Ramponi, G. (2019). *News Sharing User Behaviour on Twitter: A Comprehensive Data Collection of News Articles and Social Interactions*. Proceedings of the International AAAI Conference on Web and Social Media, 13(01), 592-597. https://ojs.aaai.org/index.php/ICWSM/article/view/3256

19) Probst YC, Peng Q. 2019. *Social media in dietetics: Insights into use and user networks*. Nutr Diet. 76(4):414-420. doi: 10.1111/1747-0080.12488.

20) Tian Yang, Yilang Peng, 2020, *The Importance of Trending Topics in the Gatekeeping of Social Media News Engagement: A Natural Experiment on Weibo,* Communication Research, https://doi.org/10.1177/0093650220933729

21) Kumar, A., et al*,* 2019. *Stages of user engagement on Social Commerce Platforms: Analysis with the Navigational Clickstream Data,* International Journal of Electronic Commerce. 23 (2), 179-211, https://doi.org/10.1080/10864415.2018.1564550

22) K. Shu, X. Zhou, S. Wang, R. Zafarani and H. Liu, 2019. *The Role of User Profiles for Fake News Detection*, 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 436-439, doi: 10.1145/3341161.3342927

23) Apuke, Oberiri Desitny, Bahiyah Omar. 2021. *Fake news and COVID 19: modelling the predictors of fake news sharing among social media users.* Telematics and Informatics 56, https://doi.org/10.1016/j.tele.2020.101475

24) Francisca Adoma Acheampong, Chen Wenyu, Henry Nunoo-Mensah, 2020, *Text-based emotion detection: Advances, challenges, and opportunities,* 2(7). https://doi.org/10/1002/3ng2.12189

25) Hilde A. M. Voorveld, Guda van Noort, Daniël G. Muntinga & Fred Bronner 2018, *Engagement with Social Media and Social Media Advertising: The Differentiating Role of Platform Type,* Journal of Advertising*,* 47(1), 38-54, DOI: 10.1080/00913367.2017.1405754

26) David D. Bourgin, Joshua C. Peterson, Daniel Reichman, Stuart Russel, Thomas L. Griffiths, 2019, *Cognitive Model Priors for Predicting Human Decisions,* Machine Learning.

27) Assigana, E, Eric Chang, Seungsuk Cho, Vivek Kotecha, Bing Liu, Hannah Turner, Yan Zhang, Michael G. Christel, Scott M. Stevens, 2014, *TF-CBT Triangle of Life: A Game to Help with Cognitive Behavioral Therapy,* Entertainment Technology Center, Carnegie Mellon University, Pittsburgh, PA USA

28) Taniguchi H, Sato H, Shirakawa T. 2018. *A machine learning model with human cognitive biases capable of learning from small and biased datasets*. Sci Rep.; 8(1):7397. doi: 10.1038/s41598-018-25679-z

29) María del Rocío Bonilla , Eva Perea , José Luis del Olmo & August Corrons. (2020) *Insights into user engagement on social media. Case study of a higher education institution, Journal of Marketing for Higher Education*, 30:1, 145-160, DOI: 10.1080/08841241.2019.1693475

30) Sydral & Briggs Appel, Helmut, Alexander L. Gerlach, and Jan Crusius. 2018. *Engagement with Social Media: A Qualitative Exploration*. Journal of Marketing Theory and Practice. 26(1-2), 44-49. https://doi.org/10.1080/10696679.2017.1389243

31) Xiao, X, Antonio Attanasio, Silvia Chiusano and Tania Cerquitelli, 2017, *Twitter data laid almost bare: An insightful exploratory analyse,* Expert Systems With Applications, 30, 501-517, https://doi.org/10.1016/j.eswa.2017.08.017

32) Jonathon Hutchinson, 2019. *Micro-platformization for digital activism on social media*, Information, Communication & Society, 24 (1), 35-51, DOI: https://doi.org/10.1080/1369118X.2019.1629612

33) Rathje S, Van Bavel JJ, van der Linden S., 2021. *Out-group animosity drives engagement on social media*. Proc Natl Acad Sci USA,118(26), doi: 10.1073/pnas.2024292118.

34) Smith MA, Williamson LD, Bigman CA. 2020. *Can Social Media News Encourage Activism? The Impact of Discrimination News Frames on College Students' Activism Intentions*. Social Media + Society. doi:10.1177/2056305120921366

35) L. M. d. Oliveira and O. Goussevskaia, 2020, *Topic trends and user engagement on Instagram*, 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 488-495, doi: 10.1109/WIIAT50758.2020.00073.

36) Kushwha, Amit Kumar, Arpan Kumar Kar, P. V. Ilavarasan, 2020, *Predicting Information Diffusion on Twitter a Deep Learning Neural Network Model Using Custom Weighted Word Features*. Responsible Design, Implementation and Use of Information and Communication Technology, 456-468, DOI:10.1007/978-3-030-44999-5_38

37) Fenn K, Byrne M. 2013. *The key principles of cognitive behavioural therapy*. InnovAiT. 6(9), 579-585. doi:10.1177/1755738012471029

38) Wu, Chuhan, Fangzhao Wu, Tao Qi and Yongfen Huang, 2020. *User Modelling with Click Preference and Reading Satisfaction for News Recommendation*, Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence (IJCAI-20), DOI:10.24963/ijcai.2020/414

39) Srivatsan, Bharath and Kelly Zhou, Andrea LaPaugh, 2016, *An Algorithmic Approach to Trending Tweet Prediction and Recommendation*, COS 435 Project https://bharathsrivatsan.com/files/tweets.pdf

40) Jonathan Rose, 2017, *Brexit Trump, and Post-Truth Politics*, Public Integrity.19 (6), 555-558, DOI: 10.1080/10999922.2017.1285540

41) Durazzi, Francesco, Martin Muller, Marcel Salathe & Daniel Remondini, 2021, *Clusters of science and health related Twitter users become more isolated during the Covid-19 pandemic*. Scientific Reports, 11(1), DOI:10.1038/s41598-021-99301-0

42) Time Magazine, The 20 Most Influential People in Tech Right Now, 2018, https://time.com/4759416/most-influential-people-tech

43) Wikipedia, Lisa Su, https://en.wikipedia.org/wiki/Lisa_Su

44) Saner, E., 2019, YouTube's Susan Wojcicki: '*Where is the Line of free speech – are you removing voices that should be heard?*', The Guardian, https://www.theguardian.com/technology/2019/aug/10/youtube-susan-wojcicki-ceo-where-line-removing-voices-heard

45) Pew Research Center, 2019, *How Twitter Users Compare to the General Public*? https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/

46) Dumais, Susan, Robin Jeffries, Daniel Russell, Diane Tang, Jaime Teevan, 2013 *Understanding User Behavior Through Log Data and Analysis*, In Proceedings of CSCW 2013

47) Fenn, Kristina, Dr. Majella Byrne, 2013, *The key principles of cognitive behavioural therapy*, King's College, University of London DOI: 10.1177/1755738012471029

48) Donyavi, V. Mehdi Rabiei, Masound Nikfarjam, Amir Rahnejat, 2015," Develop and Validate a Metacognitive-Cognitive behaviour Model for Body Dysmorphic Disorder, Open Journal of Medical Psychology, 2015, 45-52

49) Paez, Thomas L. 2008. Int*roduction to Model Validation*, Sandia National Laboratories, SAND2008-7314

50) Mandle, Robert, *Validation Methods and Terms*, www.cbrlabs.com/assay-validation.html

51) Wallach, Kyrstin, 2014 “A Content Analyesis of Twitter Use: Factors that Might Increase Sales During an Award Show?”, the Elon Journal of Undergraduate Research in Communications Vol 5. No.1, Spring 2014

52) Silva, C., J. Vieira, Jose C., Rui C., & Ribeiro, A., 2021, Development and Validation of a Descriptive Cognitive Model for Predictability Usability Issue sin a Low-Code Development Platform. Human Factors, 63 (6), pg 1012-1032, DOI:10.1177/0018720820920429

53) Beck, J. S., & Fleming, S. (2021). A Brief History of Aaron T. Beck, MD, and Cognitive Behavior Therapy. Clinical Psychology in Europe, 3(2), 1-7. https://doi.org/10.32872/cpe.6701

54) Topic Modelling in Python Unsupervised Machine Learning to Find Tweet Topics. https://ourcodingclub.github.io/tutorials/topic-modelling-python/

55) Velotio Technologies, 2019. The Ultimate Beginner’s Guide to Jupyter Notebooks. https://medium.com/velotio-perspectives/the-ultimate-beginners-guide-to-jupyter-notebooks-6b00846ed2af

56) Salmon, Felix, 2018, Elon Musk and the Unnerving Influence of Twitter’s Power Users – Twitter has increasingly become a place where a small number of elite users commandeer huge followings – and they don’t always use their power for good, WIRED June 15, 2018, https://www.wired.com/story/elon-musk-and-the-unnerving-influence-of-twitters-power-users/

57) Maxwell, Yael, October 14, 2019, “Study Unearths Gender Disparities in Twitter Influence, TCTMD News article

https://www.tctmd.com/news/study-unearths-gender-disparities-twitter-influence

58) Nilizadeh, Shirin, Anne Groggel, Peter Lista, S.Das, Yong-Yeol Ahn, Apu Kapadia, Fabio Rojas, 2016, "Twitter's Glass Ceiling: the Effect of Perceived Gender on Online Visibility" Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016)

59) Dugar, Diva, May 16, 2019, Feature Engineering in Python, published in Towards Data Science https://towardsdatascience.com/feature-engineering-in-python-part-i-the-most-powerful-way-of-dealing-with-data-8e2447e7c69e

60) Zeya LT, November 9, 2021, Precision and Recall Made Simple, published in Towards Data Science https://towardsdatascience.com/precision-and-recall-made-simple-afb5e098970f

61) Socialblade, Bill Gates' Twitter account statistics, https://socialblade.com/twitter/user/billgates

62) Socialblade, Elon Musk's Twitter account statistics, https://socialblade.com/twitter/user/elonmusk

63) Socialblade, Jeff Bezos' Twitter account statistics, https://socialblade.com/twitter/user/jeffbezos

64) Socialblade, Corie Barry's Twitter account statistics, https://socialblade.com/twitter/user/corie_barry

65) Socialblade, Ginni Rometty's Twitter account statistics, https://socialblade.com/twitter/user/GinniRometty

66) Socialblade, Lisa Su's Twitter account statistics, https://socialblade.com/twitter/user/lisasu

67) Socialblade, Sundar Pichai's Twitter account statistics, https://socialblade.com/twitter/user/sundarpichai

68) Socialblade, Susan Wojcicki's Twitter account statistics, https://socialblade.com/twitter/user/SusanWojcicki

69) Socialblade, Tim Cook's Twitter account statistics, https://socialblade.com/twitter/user/tim_cook

70) Socialblade, Whitney Wolfe-Herd's Twitter account statistics, https://socialblade.com/twitter/user/WhitWolfeHerd

71) What is the Engagement Rate?, https://corporatefinanceinstitute.com/resources/knowledge/ecommerce-saas/engagement-rate/