Universiteit Leiden
The Netherlands

Computer Science & Economics

CEOs' use of visual cues in tweets and market reactions

D.M.D. Macquiné

First Supervisor: Dr. A.H. Zohrehvand
Second Supervisor: Dr. L. Cao

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

July 6, 2023

Abstract

Social media has provided a new way for CEOs to communicate with the public. Through social media CEOs can share previously undisclosed information with the public. Investors can use this information to determine the impact on the company and move the stock price. This shows how CEO communication on social media can have important consequences. In this research, I examined visual cues in tweets of CEOs of S&P 1500 companies and stock market reactions. I used event study to estimate the impact of CEOs' tweets on the stock, machine learning methods to identify features in pictures of CEOs, and a fixed effects model to link visual cues in CEOs' tweets to changes in the stock price. I found that CEOs' usage of images containing people has a significant negative effect on the stock price of their employing company. The finding is an important contribution to the literature on CEO communication and stock market reactions.

Acknowledgements

I would like to express my deepest gratitude to my supervisor Miros Zohrevand. His guidance during the thesis project has made the project a truly educational experience. He has provided thought-provoking research ideas and methodological approaches. Moreover, the feedback provided by Miros has helped me in refining my ideas and greatly contributed to the research quality. All in all, Miros' expertise on the research topic as well as the insights provided have proven to be of invaluable worth to a successful conclusion of the project for which I am truly grateful.

My sincere thanks also go out to my second supervisor, Dr. Lu Cao. Her expertise in computer vision methodologies as well as her aid in general have been of great value for this study.

Additionally, I owe a great debt of gratitude to my family without whom I could not have fulfilled it. Their continuous encouragement and support during the project as well as their love, patience, and belief in me have been of inestimable value to me.

Together the guidance of my supervisors and the support of my family have enabled me to complete the thesis successfully. I feel honored to have had such marvelous people surrounding me during the project and am ever grateful for their contribution to my academic and personal growth.

Contents

1 Introduction

The impact of CEOs on their companies' stock prices has drawn the attention of researchers for a considerable amount of time. Studies have shown CEOs are a major determinant of stock price movements (Fahlenbrach, 2009), (You, Srinivasan, Pauwels, & Joshi, 2020), (Harrison, Thurgood, Boivie, & Pfarrer, 2020) (Zhang & Wiersema, 2009),(Akansu, Cicon, Ferris, & Sun, 2017). Furthermore, research has found that CEO communication can significantly impact stock prices, For example, (Flam, Green, & Sharp, 2020) linked visual cues in CEOs' facial expressions during interviews to stock price movements.

The emergence of social media has provided a way for CEOs to directly engage with large audiences without constraints on frequency. For example, Jack Dorsey, Elon Musk and Tim Cook all have follower bases of millions of people on Twitter and Elon Musk has posted over 26,000 tweets at the time of writing. CEOs' tweets can also have a direct impact on the stock markets. In one instance, Musk tweeted he believed Tesla's stock price was overpriced, leading to a 9% decline in the stock price (Randewich, 2020). This individual case demonstrates how CEO communication on social media can influence stock markets.

The role of CEO communication on social media and its potential impact on stock markets has not gone unnoticed by researchers, who are studying structural links between CEOs' social media communication and market movements. For example, research found that CEOs' tweets overall have a positive impact on stock prices (Knipmeijer, 2020). Other research (Napora, 2019) differentiated between tweets about CSR, personal content, and operating performance and found the impact on the stock price depended on the textual content of the tweet. However, the media present in CEOs' tweets in regard to stock price performance has remained unexplored as far as we know. By not including media, an important part of the information present in tweets is omitted. There are good reasons to research images in CEOs' tweets since research finds tweets containing images can influence investment decisions differently than text-only tweets (Brown, Elliott, & Grant, 2019). In this research, we contribute to filling this gap left open by considering the media present in CEOs' tweets and relating these visual cues to deviations in the stock price. This leads to the research question:

"*To what extent do visual cues in CEOs' tweets impact the stock market?*"

To investigate the research question, we first identified the CEOs from the S&P 1500 that were active on Twitter. Subsequently, we gathered the CEOs' tweets, including media, using the Twitter API. Clustering was used to create clusters of images with similar characteristics. These clusters were then manually labeled based on an overarching topic. As a result, a cluster containing images of people was identified. Furthermore, we had to obtain the corresponding stock price data for the companies. To assess the impact of CEOs' tweets on stock prices, we conducted an event study analysis. This approach allowed us to determine the unexplained price movements of the stock, which are attributed to the CEOs' tweets. Finally, regression analysis was used to examine the relationship between the presence of images of people in a CEO's tweet and stock price movements.

The answer to the research question contributes to the literature on CEO communication on social media (Zohrehvand, 2022), (Knipmeijer, 2020), (Napora, 2019) by improving our understanding of CEO communication and the impact of their communication. In addition, the conclusions of this study are of interest to CEOs and their public relations teams, as it can help guide them in the decision-making process regarding including certain types of images the CEOs' tweets. Moreover, investors may find the answer interesting as well for making informed investment decisions.

2 Theoretical Background

2.1 The Importance of CEOs' Social Media Communication

One of the ways in which CEOs can impact the stock market is through their social media communication. CEOs have an information advantage as they receive information in a bottom-up manner, allowing them to have a comprehensive understanding of the company's state. In contrast, outsiders lack access to this same level of information. By using social media, CEOs can bridge the information asymmetry between them and outsiders by sharing previously undisclosed information.

CEOs sharing previously undisclosed information should impact the stock price of their company, since, according to the efficient market hypothesis, a company's stock price should reflect all relevant available information (Malkiel & Fama, 1970). Examples of information that may be relevant to firm valuation could be how the firm has performed in the last quarter or an outlook for the coming year. When this information is new to investors, it may contribute to their idea of the (future) state of the company and allows them to adjust their valuation. (Guindy & Riordan, 2017) found indications that companies using Twitter release information pertinent to firm valuation more quickly. (Bruinsma, 2022), (Malhotra & Malhotra, 2016) found that CEOs frequently discuss their business on Twitter, which opens up the possibility of them disclosing important information that could impact the share price. Additionally, (Malhotra & Malhotra, 2016) found that CEOs also use Twitter to share product information, external validation, and customer stories. They argue that these topics resonate with their followers, increasing the likelihood of retweets and potentially even reaching new customers. New customers may lead to improved business performance which may translate to an increased stock price.

Studies researching CEOs' tweets have found them to have a positive significant effect on the company stock price, suggesting that overall CEOs are able to use social media to add value (Napora, 2019),(Knipmeijer, 2020). Specifically, (Napora, 2019) found that tweets about CSR, personal views, and company operational information, on average, all result in positive abnormal returns. This shows that tweets sharing relevant information about a company's future performance are incorporated in the share price, in accordance with the efficient market theory.

2.2 The Cognitive Processing of Images and Text

Research shows that the cognitive processing of images and text is distinct. dual-coding theory (Paivio, 1971) hypothesizes that the brain processes text and images through different channels. According to the theory, concepts can be encoded into both imagery and verbal representations in the brain. When storing or retrieving information either or both channels may be utilized simultaneously, thus allowing for interaction between the imagery and textual systems. Several studies have found support for this theory. For example, studies found that some areas of the brain are active both when processing words and images while there are also text- and image-specific areas (Chee et al., 2000),(Ganis, Kutas, & Sereno, 1996). Other research found people to create mental visual representations to the same extent when thinking verbally and visually. However, a verbal representation is primarily formed when thinking verbally and to a lesser extent when thinking visually, indicating stronger visual than verbal abilities embedded in the brain (Amit, Hoeflin, Hamzah, & Fedorenko, 2017).

2.3 The Benefits of Adding Image Features to Models

Different studies demonstrate that using image features in models adds value compared to using only textual ones. For instance (Hu & Flaxman, 2018) found that combining textual and visual sentiment analysis more accurately predicts the emotional state of a Tumblr user than using just either one.
likewise, (Obaid & Pukthuanthong, 2022) and (Chiah, Hu, & Zhong, 2022) show how using images can be beneficial in predicting market returns. The studies examined the sentiment of photos in the news around the world and the effect on stock market returns. They utilized a measure called "Photo Pessimism" to capture market sentiment, which calculates the proportion of negative photos in the news on a given day. Notably (Obaid & Pukthuanthong, 2022) found that usually the sentiment of text and photos is highly correlated and so they can be substituted. However, during times of high anxiety, photos in the news serve as a better predictor for market returns than the text does. The authors link this to earlier finding (Chemtob et al., 1999) that images are more effective than text in transmitting traumatic events.
These examples illustrate that textual and visual features are complementary to each other because they are able to capture different information, so using both can result in more accurate models.

2.4 The Impact of Images on Investment Decisions

A number of studies show that images in tweets influence investors' decisions. For example, research finds that non-GAAP-compliant images [1] in company earnings tweets influence investment decisions. (Brown et al., 2019). On the other hand, the scholars find no such effect for tweets presenting earnings in text not complying with GAAP standards. The authors blame this on investors unintentionally processing and acting on information in non-GAAP-compliant images. (Nekrasov, Teoh, & Wu, 2020) examines the use of visuals in earning announcement tweets by companies. Among their findings is that tweets with visuals lead to higher user engagement among investors. Additionally, the researchers note that the inclusion of visuals in tweets initially triggers relatively strong market reactions but comparatively milder reactions later on compared to tweets that only contain text. The researchers link the findings to the limited attention theory (Hirshleifer & Teoh, 2003) which posits that investors' attention is limited making them focus on only a portion of the available information.

[1] non-GAAP-compliant images are images that do not adhere to the GAAP rules for presenting accounting information

2.5 Investors' Perception of CEOs' Usage of People Images

CEOs are among the key drivers of their companies' CSR and ESG performance (Wernicke, Sajko, & Boone, 2022), (Crace & Gehman, 2023). A systematic review found that firms' CSR efforts enhance their financial performance (Coelho, Jayantilal, & Ferreira, n.d.). However, studies focusing on the humane dimensions of CSR show mixed evidence. For example, (Tsai & Wu, 2022) found that firms' CSR activities related to human rights improve financial performance, while (Brammer, Brooks, & Pavelin, 2006) found that CSR activities related to the community negatively affect the stock price. In contrast, (Barnett & Salomon, 2006) found that activities related to the community increase financial performance, while labor-related activities decrease it. (Oddo, 2022) found that firms scoring better on the social dimension of ESG have a decreased valuation.

I argue that socially aware CEOs demonstrate their engagement with social topics by discussing them in their tweets, and expect, they will often use images of people when doing so because of the inherent connection between people and social topics. Investors may perceive a CEO's use of images of people in their tweets as a proxy for socially aware CEOs devoted to doing business in a socially responsible way. Consequently, investors may infer from the CEOs' use of images of people in their tweets that the CEO will focus on doing business in a socially responsible way. investors can then assess the expected impact on financial performance and adjust their investments accordingly.

This leads to the hypothesis that CEOs' use of images of people in their tweets significantly impacts stock prices. However, since the evidence on whether the impact of firms' social activities on firm performance is mixed, both positive and negative impacts are tested. Hence the following hypotheses:

HYPOTHESIS 1: *Tweets containing images of people have a positive effect on the stock price*

HYPOTHESIS 2: *Tweets containing images of people have a negative effect on the stock price*

2.6 Research Gap

This research adds to the existing work in several ways. While different studies examine the effect of CEOs' tweets on their firms' stock prices, they do not consider images in those tweets. Other research examines the effect of images in tweets on investor behavior, but not in the context of CEOs' tweets. Therefore, the novelty of this research comes from using the visual information in CEOs' tweets and studying the effect on the share price.

3 Method

3.1 Data Gathering

The data used in this research was obtained from different resources in several steps. First, an overview of all the CEOs of the (S&P 1500) companies was retrieved from Compustat Execucomp which includes all the CEOs of the S&P 1500 companies since Twitter was established. This overview consists of 5468 rows of CEOs and their companies. For every CEO it was necessary to check if they had a Twitter account. A manual search query consisting of the full name of the CEO followed by the name of the company and ending with "Twitter" was used to find out [2]. When a CEO's Twitter account was found the handle was noted down. In the end, this resulted in 494 unique CEOs having Twitter accounts. However, some of the CEO's never tweeted so the number useful of Twitter accounts is lower. Of the CEOs in the sample 25 led multiple companies [3]. This could pose a problem because a CEO could be tweeting on behalf of or about either of the companies, making it difficult to determine which stock price the tweet could affect. In this research this problem does not exist [4].

Some cleaning on the data of CEOs' terms was necessary before their tweets could be gathered. The data should include a starting date and a leaving date (in case they quit) for every CEO. For some CEOs, this data was missing or incorrect. The dates were checked and corrected to the extent possible. Dates certainly falling before Twitter's founding were not corrected since the time frame from which tweets should be gathered is not altered. In the left cases, the date could often be determined from press releases. In the cases where even this was impossible, the exact date could not be determined but the month could be. In these cases, for the starting date, the first day of the month in which the job was taken on is used. For the date of leaving the first day following the month of resignation is used. This method was chosen because it is most common to start a position on the first day of the month and end a position on the last day of the month. if the actual date of starting (leaving) falls after (before) the recorded date, tweets from outside the tenure will accidentally be included. It is expected that the effects of this are minor since the number of tweets from this period will usually be only a small proportion of the total number of tweets from when a CEO is in office. Checking whether a CEO has left has been done until the 9th of April 2022.

Subsequently, the Twitter IDs of the CEO's corresponding Twitter accounts were gathered because the Twitter API needs an account ID instead of a Twitter handle to request a user's tweets. Getting the ID of an account was done by sending HTTP requests to https://tweeterid.com/ with the Twitter handle of a CEO to identify the account. For three CEO's the Twitter ID could not be requested anymore since their accounts had been deleted in the meantime. Because of this, the tweets of these CEOs could not be gathered anymore and so the number of unique CEOs in the

[2] Searching for CEO's Twitter accounts was split up between this study and (Bruinsma, 2022)

[3] Of the 25 CEOs, 23 are actually two times in our used sample. One of the CEOs that has worked for two companies is Christopher J. Nassetta. Between gathering the tweets for one of the companies and the other he deleted his Twitter account. This does not pose a problem because his term at the company for which the tweets had not been gathered yet ended before Twitter was founded so no tweets for this company would have been included anyway. The other CEO with two terms at different companies of which only one is included is Anthony M. Jabbour. He has terms at two different companies but only used his Twitter account for one of them so the tweets are only used with respect to that company.

[4] There are a few cases of CEOs that have overlapping terms at different companies: Anthony M. Jabbour has overlapping terms at Black Knight Inc and Dun & Bradstreet Holdings Inc. but he only tweeted in the name of Black Knight and so Dun & Bradstreet is left out of the research. James A. Hyde has overlapping periods for companies he worked for but since there are no tweets for both of these periods this is not a problem. Richard Brian Handler has been CEO of both Jefferies Group LLC and Jefferies Financial Group Inc. Jefferies Group was acquired by Leucadia and since is not a publicly traded company on its own anymore. Leucadia was renamed to Jefferies Financial Group Inc in 2018. Richard Brian Handler his Twitter account exists since 2013 and so only Jefferies Financial Group Inc should be taken into account since by this time Jefferies Group was already acquired by Leucadia, now known as Jefferies Financial Group.

sample dropped to 491.

Next, we began gathering the tweets of CEOs using the Twitter API. The fields gathered from the tweets are; the text of the tweet, videos and[5] images in the tweets and the date and time of the tweet[6]. When gathering the media and text from a tweet, exactly what can be seen within the tweet in terms of text and images or videos on `twitter.com` is gathered. This means that the text and media included directly by the author are taken into account but also media referenced by the author from another tweet is taken into account if it was quoted or retweeted[7]. If the author of a tweet is referencing someone else's tweet by replying to that tweet, the media is not included since the media from that tweet can not be seen in the author's tweet. This also holds for an author that replies to their own tweet since in the new tweet the media can not be seen. The date and time of each tweet are stored to match a tweet with the right stock price. Furthermore, when gathering the tweets of CEOs, only the tweets of the period in which a CEO was employed for the company have been gathered[8]. Only tweets posted before the 9th of April 2022 are considered. Until this date, it has been checked if a CEO was still in office and so after this date, it might be the case the CEO left.

100 tweets of a CEO were gathered per request sent to the Twitter API because this is the maximum amount allowed per request. To take all the tweets of an CEO pagination is used. If there are more tweets from the CEO, the response of the API has a token that is used in a subsequent request to get the next "page" with tweets. The API has a limitation that allows only the 3200 most recent tweets of a user to be requested. This means that if a CEO has placed more than 3200 tweets, only the most recent 3200 can be retrieved. As a consequence, if the CEO of a company posted more than 3200 tweets after leaving, no tweets of the period in which the person was CEO can be retrieved anymore. For each of a CEO's tweets, we also needed any referenced tweets and media so they have been gathered as well. The API has a rate limit of 1500 requests per 15 minutes so after every request 2 seconds were waited to prevent reaching the limit.

The gathering of the tweets results in 185.439 tweets from 514 CEOs. 23 unique individuals were CEOs of two companies so the number of unique individuals in the sample adds up to 491. The number of gathered images (either video previews or just images) from the tweets adds up to

[5]For videos that are included in tweets, only a preview image of the video can be retrieved.

[6]Media is returned by the API as a hyperlink to the location of the media. For twelve media items, the hyperlink did not work anymore. For every tweet the number of missing items is stored

[7]If a tweet is retweeted, sometimes it is the case that the media from that tweet is already in the media from the author's tweet but sometimes this media is only or also present in the referenced tweet. No clear reason for when which one is the case has been found but if the media is already in the author's tweet, it is not taken also from the referenced tweet since this way the media would be included twice for these tweets.

[8]A limitation of the Twitter API is that it is not possible to specify a start or end time that lays before the 6th of November of 2010 for the tweets that should be included. If the end date is set to this date or later and the start date from when tweets should be included is left out, it is possible to get tweets from before this date too. In cases where a CEO became and left both before the 6th of November 2010, the date of leaving is set to the 6th of November 2010 to prevent tweets after this date to be included in the API response which would be unnecessary. In case a CEO became CEO before the 6th of November 2010 but left after, for the date until which tweets should be included the actual date of leaving is taken and the date from which tweets should be included is left out to also include tweets from before the 6th of November 2010. In this case, the needed tweets are filtered out. In the last case, where the date of becoming CEO and leaving are both after the 6th of November 2010, the API can just be used as intended with the date from until when the tweets should be retrieved.

71207. Of the CEOs, 350 placed at least one tweet, meaning that 164 of them have no tweets on their account or the tweets from their period in the office are superseded by more than 3200 tweets so that they can not be retrieved anymore.

3.2 Dependent variable

The dependent variable used in the analysis is an approximation of the deviation in the stock price from the expected evolution. this difference is then attributed to the CEO's tweets.

The effect of CEOs' tweets on the stock price is estimated using event study. In order to do so, the tweet(s) of a CEO need to be matched to the stock price of the company. To obtain the companies' stock prices Wharton Research Data Services (WRDS) is used. A tweet is not always posted at a time when the company stock can be traded. Therefore every tweet needs to be matched to a date on which the stock can be traded, referred to as the trading date of the tweet. Only closing prices of a stock on a trading date are available. As a consequence, it was necessary to match tweets to the first available closing price after the tweet.

Before matching CEOs' tweets to a trading date, the trading hours of the corresponding stock must be known. Stocks traded "over the counter" (OTC) or "Other" (OTH) are excluded due to unclear trading hours making it impossible to match them to a trading date. The other stocks are traded on the Nasdaq (NAS), New York Stock Exchange (NYSE), or the American Stock Exchange (ASE) which are all located in the Eastern time zone. For each company, at least one day of stock prices must be available. If this is not the case, the CEO's tweets are not being used since the stock price is needed for the event study. If the preceding conditions are met, the CEO's tweets can be matched to the right trading date.

To match CEOs' tweets to the appropriate trading date, the time of the tweet must be converted from UTC(+0) to Eastern Time. Depending on whether it is daylight saving time or standard time it is respectively 4 or 5 hours earlier in Eastern Time than UTC and so this amount of time is subtracted from the tweeting time in UTC to find the time of tweeting in Eastern Time. If the tweet is posted after market closure (4 p.m.) the tweet is moved to the next trading day because the first day the tweet could influence the stock prices would be the market opening of the next trading day. If the tweet is posted anywhere between 12 a.m. and 4 p.m. the date is left unchanged if the current day is a trading day since the tweet could already influence the stock price of that day. If the day is not a trading day the tweet is matched to the first next trading day. All the tweets that coincide on a trading date together form an event.

Next, it needs to be determined how the stock price of the company would have developed if the event never had taken place. This is called the expected return of the stock. The used model to calculate the expected returns is the Famma-French model as explained by the equation below.

$$r = R_f + \beta_1(R_m - R_f) + \beta_2(SMB) + \beta_3(HML) + \alpha$$

In this model, r is the expected return of the stock, R_f is the risk-free return rate, R_m is the return rate of the market portfolio, SMB is Small Minus Big which is the excess return that companies with a small market capitalization make over companies with a large one, HML stands for High

Minus Low and is calculated as the excess return of stocks with a high book-to-market ratio over stocks with a low book to-market-ratio (FAMA & FRENCH, 1992). The used R_f, R_m, SMB and HML come from a dataset that is part of a data library that French provides on his Dartmouth website. $\beta_{1,2,3}$ denotes how strongly the returns of the stock react to respectively; the market risk premium the excess returns of small-cap stocks and the excess returns of high book-to-market ratio stocks. Lastly, α is the excess return of the stock over the market portfolio. To find the $\beta_{1,2,3}$ and α coefficients an ordinary least squares model (OLS) is being trained on the r of the company, R_f, R_m, SMB, and HML for a period from 147 trading days before the event until 7 trading days before the event. This period is the so-called estimation window. For some of the events, the stock price of the company or the factors needed for the Famma-French model are missing. If this is the case the event is being dropped and not used in the event study. At this point, the model to calculate the expected return is obtained and can be used to calculate the expected return on the trading day of the event. Subtracting the expected returns from the actual share returns for the trading day yields abnormal returns that can be attributed to the event. In some cases, the actual returns of the stock price or the input data for the fama-french model on the trading day are missing in which case the expected returns can not be subtracted from it. If this is the case the event is dropped.

The data required to conduct event studies was present for 318 CEOs and a total of 71.738 events which in total contain 158.512 tweets and 61.332 media files that are used in the remainder of the research.

3.3 Independent Variables

The analysis uses two independent variables: the number of images in the event containing people, called "people". The other variable called "business" is equal to the number of tweets in an event in which the topic is business-related. The people variable is used to examine the effect of CEOs posting images of people on the stock price. To count the number of images in the event containing images of people it needs to be identified whether there is a person in an image. The business variable is used to find out whether the impact of images of people posted by a CEO is different when used in a business-specific context. Investors may perceive a CEO's use of images of people when he/she is discussing business-related matters as different from when discussing other topics. In case the CEO uses images of people when discussing business, investors may view this as a CEO's commitment towards the people related to the business, as when discussing other topics investors may not care about images of people since they don't show a commitment towards the people related to the business. Therefore the effect of images of people may depend on the topic of the tweet. this can be examined by examining the interaction between the "business" and "people" variables.

The approach to determining if an image contains people or if a tweet is about business is similar. Briefly explained the images and texts of tweets are turned into embeddings representing them. A dimensionally reduced representation of those is then clustered. A label is manually assigned to each cluster by examining the particular contents of a cluster and determining what the common topic of the cluster is. Ultimately, a cluster of images containing people and a combination of text clusters about business are obtained. If a tweet is assigned to any of the clusters which got the label business, the tweet gets business as a topic. The steps taken are discussed in more detail below.

3.3.1 Pre-Processing

3.3.1.1 Pre-Processing Tweet Text Contents

Before clustering, the text in the tweets had to be pre-processed. Tweets contain unique features not found in regular texts. As a result, the algorithm used to create the embeddings might not be able to cope with the tweet-specific features and represent them in embeddings.

The following preprocessing is applied to the text in the tweets:

- removing the "RT" tags from the text of the tweet since they are not used in regular texts.
- remove hyperlinks from the text. Hyperlinks are not widely used in normal texts.
- remove just the hashtag signs ("#") from the word they are in front of. The model might not know how to conceive the hashtag but the word accompanying the hashtag might carry important information.
- remove "@" and the name coming after it. The model may not understand the concept of mentions. Mentions often stand alone in texts and are not important for understanding the text in the tweet.

3.3.1.2 Turning Tweet Contents into Embeddings

To cluster tweets' images and texts, they are first turned into embeddings. The embeddings are numerical vector representations of the inputted data. A Hugging Face implementation of the CLIP (Contrastive Language-Image Pre-Training) model is used to create the embeddings. This model is a neural network trained on image-text pairs to predict what caption best fits an image (Radford et al., 2021). With this model, both texts and images can be converted into embeddings living in the same vector space. This means that image- and text embeddings close to each other in the vector space are also semantically close to each other. The model comes in different versions. In this research, the clip-ViT-L-14 model is used because of the best accuracy (on zero-shot ImageNet)[9].

3.3.1.3 Dimensionality-Reduction

After obtaining the embeddings, they could not directly be clustered due to the curse of dimensionality. The initial embeddings constitute 768 dimensions, with such a high number of dimensions the distance between the embeddings tends to be similar. For clustering the embeddings this is a problem since the distance between embeddings is what determines to which cluster an embedding is assigned.

To reduce the dimensionality of the embeddings UMAP (Uniform Manifold Approximation and Projection) UMAP was used. UMAP is a relatively novel technique for dimensionality reduction

[9]A downside of the model is that it is only capable of handling English texts. Since the tweets in this research are only from tweets of CEOs from American companies this does not pose a big problem. The tweets that could not be processed by the model are dropped. There is also a version of the model that is capable of handling different languages but this model has a lower accuracy and since the number of tweets that have non-English text is low, it was chosen to use the non-multilingual model.

that is faster than T-SNE and better preserves the global structure of the data that it is performed on (McInnes, Healy, & Melville, 2018).

3.3.2 Clustering

We are using HDBSCAN to cluster the embeddings into groups that share commonalities. HDBSCAN is a density-based clustering algorithm which means that embeddings in dense areas are assigned the same cluster. This aligns well with the UMAP method used to create dense representations of the embeddings. The UMAP algorithm has a setting called minimum distance which when set to a low value cramps similar embeddings together while increasing the distance to dissimilar embeddings. As a result, the representations outputted by UMAP have dense regions which help HDBSCAN to assign embeddings to the right cluster.

3.3.2.1 UMAP & HDBSCAN Joint Hyperparameter Tuning

The best settings for HDBSCAN and UMAP were determined by conjointly optimizing their hyperparameters using grid search [10] . The hyperparameter optimization is executed only on the image embedding. The text embeddings are clustered using the same parameter settings for HDBSCAN and UMAP as obtained by optimizing the parameters for the image embeddings. This decision was made because the information captured by the text and image embeddings is the same as they were both generated using the same CLIP model and live in the same vector space. Hence, the same parameters should also work well on the text embeddings.

Conjoint hyperparameter optimization was performed simultaneously for both UMAP and HDBSCAN. This approach was necessary because the output of UMAP serves as the input for HDBSCAN. If the UMAP parameters had been optimized first while keeping the HDBSCAN parameters constant, there is a possibility that the identified UMAP parameters would only work well with the specific combination of HDBSCAN parameters and less so with other settings for the HDBSCAN parameters. There may exist other HDBSCAN parameter settings that, when paired with different UMAP parameter settings outperform the combination found by sequentially optimizing the UMAP and HDBSCAN parameters. Therefore, by optimizing both UMAP and HDBSCAN parameters simultaneously, we aim to find the best combination of parameter settings.

Evaluating different combinations of parameter settings was done by comparing DBCV (Density Based Cluster Validity) scores of the clustering results. There are several reasons for using this evaluation method. Firstly, the DBCV score takes into account the embeddings that were considered noise by HDBSCAN (and therefore are not clustered). Secondly, DBCV was designed for density-based clustering methods like HDBSCAN. DBCV works by composing a score of the interconnectivity of points within a cluster and the connectivity between clusters. DBCV will return a good score if the points in a cluster are closely connected and clusters to each other are loosely connected (Moulavi, Jaskowiak, Campello, Zimek, & Sander, n.d.).

[10] Due to time constraints every combination of parameters is evaluated just once. UMAP is a stochastic algorithm and so the results will differ between runs resulting in a slightly different evaluation score. Because of this, it might be the case that there exists another combination in the grid that is better than for which the highest score was found.

The hyperparameters optimized for UMAP are the minimum distance, the number of neighbors, and the number of components. The values attempted for the parameters during the grid search are determined through a combination of experimentation with UMAP and logical reasoning.

The minimum distance determines how closely points may be crammed together by UMAP, in the grid search only 0 and 0.1 are attempted for this setting. Low values for the minimum distance setting result in points being packed together tightly. This way points that are near to each other are clumped to each other while points that are distant from each other become far away from each other. This behavior is desirable for the clustering because the embeddings that are like each other, are near while the ones not alike are further away. This results in alike embeddings clustered into one cluster without the dissimilar embeddings ending up in the same cluster.

The number of neighbors parameter in UMAP determines how many other data points are considered by UMAP in creating the dimensional reduced embedding. With high values, the resulting embedding will retain more of the global structure while with low settings the emphasis will be more on the local structure of the original data. The attempted values for the number of neighbors were 50, 100, and 200. The tried values for the number of neighbors all favor the retention of the global structure. In this research, an emphasis on global structure fits best since we are interested in finding the major topics in the texts and images of CEOs. Retaining the global structure of the embeddings accommodates HDBSCAN in finding the major topics when clustering.

The last optimized UMAP parameter is the number of components, which determines the number of dimensions in the output of UMAP. The attempted values for the number of components are; 10, 25, and 50. Selecting the appropriate values to try for this parameter involves a trade-off between keeping the number of components low to prevent the curse of dimensionality and keeping the number of components sufficiently high to sustain enough information for clustering.

The only parameter optimized for HDBSCAN was the minimum cluster size which determines the minimum number of images in each cluster. The attempted minimum cluster sizes are 500, 1000, 1500, and 2000. These options are determined by manually trying out different cluster sizes with HDBSCAN and investigating the results. Selecting an appropriate minimum cluster size involves striking a balance. Too low values for the minimum size results in too many clusters that do not generalize well. On the other hand, setting the minimum size too high results in a very low number of clusters, essentially lumping all kinds of images/texts together.
Another setting of HDBSCAN is the minimum number of samples, which is set to the same value as the minimum cluster size by default. We did not modify this parameter since the default value is suitable for creating the envisioned clusters. When the minimum number of samples is set to high values, clusters with a general topic will be created, and instances that do not fit the topic well are discarded as noise. This behavior is desired because it yields clusters with evident topics that apply to the majority of its instances. For this research, these clear-cut clusters are preferred since we are interested in relating images and texts with a well-defined topic to the impact they have on stock prices.

Table 2 presents an overview of all the values evaluated for the different settings of UMAP and HDBSCAN during the hyperparameter optimization.

Hyperparameter Grid		
Algorithm	**Setting**	**Attempted values**
UMAP	min_dist	0, 0.1
UMAP	n_neighbours	50, 100, 200
UMAP	n_components	10, 25, 50
HDBSCAN	min_cluster_size	500, 1000, 1500, 2000

Table 1: An overview of the algorithm parameters and the values tried.

The hyperparameter optimization yielded the following values for the parameters: 100 for the number of neighbors in UMAP, 0 for the minimum distance, 50 for the number of components, and 500 for the minimum cluster size in HDBSCAN. with these settings, 74.5% of the images could be clustered and the obtained DBCV score is 0.68, this value is the highest DBCV score encountered when optimizing the hyperparameters. The score is only a relative score to other runs of the algorithm with different parameter combinations and therefore can not be used as an indicator of the quality of the clusters.

3.3.2.2 Image Clustering Results

Clustering the images resulted in 7 different clusters and one group containing all the images which were discarded as noise and so not assigned to any cluster. The clusters were manually labeled by examining the images in the cluster and determining what they have in common. This resulted in the following topics for the 7 clusters; one cluster containing images of people, one containing images related to space, one containing images related to sport, one containing images related to telecom, one containing posters, one containing landscapes, and lastly one containing food[11]. Appendix A shows samples of images in each cluster. the images were manually selected to show the diversity of images that are part of a cluster and to show instances where images do not fit well with the label of the cluster. Appendix B contains visualizations of the clustering results.

Image clusters								
Cluster	People	Space	Sport	Telecom	Poster	Landscape	Food	*unclustered*
Number	25383	1570	1872	877	13379	2021	567	15574

Table 2: The image clusters and the number of images assigned to it

3.3.2.3 Text Clustering Results

Clustering the text resulted in 38 different clusters, as well as a group of texts that could not be clustered. For each of the 38 clusters, a subset of the tweets assigned to the cluster was manually

[11] It is important to note that certain images and texts possess properties that would make them suitable for multiple clusters. However, the HDBSCAN clustering algorithm assigns each image to exactly one cluster.

examined to find an overarching topic that describes the cluster. Some clusters are similar in content and were therefore given the same topic. Once each cluster was assigned a topic, clusters related to business matters were taken together to form a merged "business" cluster. If a tweet got assigned to any of the clusters in this merged cluster, business is assigned as topic to it. The clusters merged into the "business" cluster are the ones with the following topics:

- **own business:** CEOs discuss matters related to their own company such as their products, the people working for the company, and partners and customers of the company.
- **public speaking:** CEOs discussing on Twitter having publicly spoken at conferences, summits, events, etc.
- **leadership:** CEOs discussing how to lead a company.

Appendix C contains a table with for every cluster the top 10 words identifying that cluster based on their tf-idf scores to get an idea of the contents of the cluster. Furthermore, the table contains the assigned topic to each cluster and the clusters merged together in the business cluster.

3.4 Control variables

Two control variables were included in the analysis. The first control variable "experience", captures the CEOs' tweeting expertise and is measured as the logarithm of the number of tweets the CEO has posted up until the current tweet in the sample. By including this variable, we control for differences in CEOs' communication styles over time arising from CEOs becoming better at composing tweets in such a way to align with their strategic aim of the tweet. Since a CEO is not expected to get better at composing their tweets in a linear way to the number of tweets, the number of tweets is logged.[12]

The second control variable "nTweets", equals the number of tweets in an event. Controlling for this variable is done because CEOs' tweets can possess additional properties not captured by the other variables used in the analysis. These other tweet properties may have their own impact on the stock prices. The decision to use the number of tweets as control variable results from the fact that every single tweet can have its own impact on the stock price, and as demonstrated by (Knipmeijer, 2020) and (Napora, 2019), the impact of CEOs' tweets on average is positive.

3.5 Analysis

The relationship between abnormal returns and the features derived from the tweets is analyzed using a fixed effects model. Specifically, fixed entity effects are being used to capture abnormal returns outcomes not explained by other variables in the model and can be attributed to the entity,

[12] There are some limitations to measuring the CEO's experience in this way. One is that the sample of tweets for a CEO only consists of the tweets that he has posted while in the position of CEO for the investigated company. This means that tweets someone might have posted before becoming CEO are not incorporated while these contribute to someone's experience with tweeting. Ensuing, someone might already have been using Twitter as CEO or executive of another company, gaining experience not captured by the current measure. In addition, only the CEO's last 3200 tweets are captured, if, before these 3200 tweets, someone had already taken the position of CEO then the first tweet in the sample is used as the first tweet to calculate the experience of a CEO instead of the actual first tweet of a CEO.

in this case, the CEO. It is expected that differences exist in how CEOs express themselves on Twitter reflecting their unique personalities and communication styles. For example, one CEO might be more introverted while another one might be more extravert. Previous research found different CEOs to have distinct communication styles (Choudhury, Wang, Carlson, & Khanna, 2019), it makes sense to expect that these differences in CEOs' communication styles are also reflected in their tweets. An assumption of the model is that the effects specific to a CEO are stable over time.
The standard errors are clustered at the same CEO level in order to get better estimates of the standard errors. Observations regarding a particular CEO/company may be correlated to each other due to unobserved properties specific to the CEO. For example, it may be the case that the abnormal returns of a certain company might be larger on average than that of another company because the CEO is more prone than another one to share previously undisclosed information important for stock price valuation.

A fully specified model with all the variables is given by the equation below. To test the different hypotheses, various models are used that include a subset of the variables in the fully specified model. An overview of all the regression variables can be found in Appendix D

$$AR(0)_{i,j} = const + \beta_1 experience_{i,j} + \beta_2 nTweets_{i,j} + \beta_3 individual_j + \beta_5 people_{i,j} + \beta_6 business_{i,j} + \beta_7 (business * people)_{i,j} + \epsilon_{i,j}$$

Given the primary focus of this research on the impact on stock prices of visual cues in CEOs' tweets, only events containing at least one image are included in the analysis.

4 Results

4.1 Descriptive Statistics

The number of events included in the analysis is 32,694, meaning there are 32,694 events that contain at least one tweet with graphics in it. The number of tweets in the analysis is 91,209 bringing the average number of tweets per event to 2.79. The number of media items in the analysis is 61,332, so the number of media items on average present in an event comes down to 1.86. Furthermore of the 61,332 media items, 25,383 have a person on it. The number of CEOs in the analysis is 287, so for 287 CEOs all the necessary data was present, meaning that there are tweets available for their accounts, the abnormal returns of their tweets could be calculated and they have posted tweets containing media.

4.2 Correlation and VIF Analysis

Tables 5 and 6 in Appendix E show respectively the correlation matrix and variance inflation factors (VIFs) for the variables used in the regression analysis. There are no issues regarding multicollinearity since all the values of the VIFs are well below 5 (O'brien, 2007).

4.3 Regression Analysis Results

Table 3 reports the study's main results. Model 1 is a baseline model that includes only the control variables. Model 2 is used to find out what the effect of CEOs tweeting images of people has. Model 3 is used to find out whether CEOs discussing business matters in their tweets has an impact on the stock price. Model 4 includes both the business and people variables in order to create a baseline model to add the business∗people interaction. Model 5 builds on model 4 by adding an interaction variable created by multiplying the business and people variables. This model is used to find out whether there is interaction between CEOs tweeting images of people and CEOs discussing business matters.

	Dependent variable: AR0				
	(1)	(2)	(3)	(4)	(5)
people (imgs)		-1.780E-4**		-1.808E-4**	-1.308E-4
		(6.989E-5)		(7.107E-5)	(8.947E-5)
business (txts)			5.222E-5	6.739E-5	1.057E-4
			(1.859E-4)	(1.896E-4)	(1.928E-4)
business∗people					-1.777E-5
					(1.737E-5)
experience	-1.058E-4	-9.781E-5	-1.057E-4	-9.767E-5	-1.062E-4
	(3.449E-4)	(3.459E-4)	(3.447E-4)	(3.456E-4)	(3.445E-4)
nTweets	6.917E-5*	1.020E-4**	6.077E-5	9.165E-5**	9.331E-5**
	(3.691E-5)	(3.959E-5)	(3.838E-5)	(3.938E-5)	(3.990E-5)
const	3.729E-4	4.001E-4	3.616E-4	3.859E-4	3.564E-4
	(8.389E-4)	(8.425E-4)	(8.297E-4)	(8.342E-4)	(8.426E-4)
F Statistic	1.16 (df=289; 32405)	1.76 (df=290; 32404)	0.82 (df=290; 32404)	1.37 (df=291; 32403)	1.19 (df=292; 32402)
Observations	32694	32694	32694	32694	32694
R^2	7.180E-5	1.626E-4	7.560E-5	1.689E-4	1.839E-4
Residual Std. Error	0.03 (df=32405)	0.03 (df=32404)	0.03 (df=32404)	0.03 (df=32403)	0.03 (df=32402)

Note: *p<0.1; **p<0.05; ***p<0.01
Only events containing graphics are included in the analysis.

Table 3: Regression Analysis results

4.3.1 Baseline Variables Results

Model 1, the baseline model, includes only the nTweets and experience variables. In model 1 the nTweets variable shows to be significant against $p < 0.1$. The nTweets variable shows significant against a p-value of 0.05 in models 2, 4 and 5 and shows not significant in model 3. The positive sign of the nTweets variable means that CEOs' tweets on average result in positive abnormal returns This is in accordance with earlier research on the topic that found CEOs tweeting to have a positive effect on the stock price (Knipmeijer, 2020) and (Napora, 2019). Furthermore, the R^2 of the first model is 7.180E-5 which means the model explains 0.0072% of the variance in the abnormal returns.

4.3.2 Impact of CEOs Posting Images of People

In the theoretical background, two hypotheses were stated. hypothesis 1 stated that CEOs tweeting images of people will have a positive effect on the stock price while hypothesis 2 exactly contrary to this, stated that CEOs tweeting images of people will have a negative effect on the stock price. In models 2 and 4 in the results table the people variable shows significant against a value $p < 0.05$. The value of the coefficient in the models takes on a negative value with a value of -0.00018. This can be interpreted as images of people in CEOs' tweets having a negative impact on the stock price of the company with a negative impact of 0.018 percentage points per image posted containing a person. Therefore the results of models 2 and 4 support hypothesis 2. Furthermore, The R^2 of model 2 is 1.626E-4 or 0.016%, so 0.016% of the variance in the abnormal returns is explained by the model.

4.3.3 Effect of CEOs Posting Images of People in Business Context

Models 3-5 include the business variable counting the number of tweets in the event in which the CEO is discussing business topics. The variable shows not significant in any of the models so there is not enough evidence to conclude that CEOs discussing business content in their tweets has an impact on the stock price. Furthermore, the business-people interaction variable does not show significant either so there is no evidence found that there do exist interaction effects between CEOs posting images of people and discussing business in their tweets.

5 Discussion

This study asked the question of to what extent visual cues in CEOs' tweets impact the stock price of their companies. In the theoretical background, I expressed the expectation that in particular pictures of people would have an impact, with both positive and negative impacts as a possibility. To research the question; the Twitter API was used to gather the tweets and media present in the tweets of CEOs of S&P 1500 companies. Event study was used to calculate abnormal returns attributed to CEOs' tweets and a Fama–French three-factor model was used to calculate the expected return which could be subtracted from the actual returns to calculate the abnormal returns. Next, the images and texts present in CEOs tweets were turned into embeddings with the sentence transformers model, these embeddings were then reduced in dimensionality with UMAP and clustered with HDBSCAN. This resulted in several image and text clusters. Several text clusters were merged together, jointly forming the business cluster. One image cluster was identified containing the sought-after images of people. subsequently, a fixed effects model, which accounts for differences in abnormal returns explained by differences in CEOs was used to relate pictures of people in CEOs' tweets to abnormal returns.

5.1 Main findings and Contribution to Existing Literature

The main result found is that CEOs' usage of images of people is significantly linked to negative stock returns of the CEO's company. A possible explanation for this may be that investors see the usage of images of people by CEOs as their commitment to the people related to the business and doing business in a socially responsible way and they determine the financial impact on the company, as I argued in the theoretical background. However, it is not examined whether this is the case and future research would be necessary to determine the exact reason. In any case, the findings of this research contribute to knowledge about how CEO communication on social media can impact stock markets. Previous research on CEO social media communication has already shown that CEOs' tweets can impact the stock market (Knipmeijer, 2020) as well as showing the stock market to react differently depending on the topic of CEOs' tweets (Napora, 2019). This research adds to those existing works by showing visual cues in CEOs' tweets also to have an impact.

5.2 Variation in significance of nTweets variable

The nTweets variable showing significant against a p-value of 0.05 in models 2,4 and 5 instead of the 0.1 in model 1 may be due to the nTweets variable in model 1 taking up some of the variance which is explained by the people variable in the other models. The nTweets variable showing not signifcant in model 3 may be due to the correlation between the business variable and the nTweets variable. table 5 shows that the correlation between these two variables is 0.5, which may lead to both variables explaining some of the same variance.

5.3 Explanatory Power of People Variable

Another finding of the research is that model 2 in which the people variable was added explains 0.016% of the variance in the abnormal returns. The baseline model however only explains 0.00718% of the variation, so adding the people variable does help in explaining more of the variance. That the explanatory power of the models and the people variable is limited is expected. There could be a lot of different events happening concurrently other than a CEO tweeting, which all could create abnormal returns and are not captured.

5.4 Lack of Evidence for relation of business and people∗business Variables to Abnormal Returns

No evidence was found that CEOs discussing business specifically has an impact on the stock price. It may be the case that when CEOs are discussing business matters in their tweets, they don't disclose new information to the public deemed important by investors for firm valuation. Another possibility is that the business variable is noisy and contains also tweets that are not about business. The tweets in different clusters were not always clearly about business topics or not. Therefore, there may be clusters wrongfully included in the merged business cluster.

Another finding of this research is the lack of evidence for interaction between the people and business variables. The interpretation of this finding is that there is no evidence that the effect on the stock price of CEOs posting images of people depends on whether the CEO discusses business in their tweets or not. This could be due again to a noisy business variable. Another reason might be that investors do not care about whether an image of a person is posted in a business context or not. (Cronqvist, Makhija, & Yonker, 2012) finds CEOs act likewise on personal matters as business matters. If investors view CEOs' usage of images of people in their tweets as a sign of their commitment to social topics, it may not matter to them whether the images are used in business-related or non-business-related tweets. The usage in both cases could signal to investors that the CEO is engaged with social topics and will also reflect this in their corporate decisions. Investors could then infer the financial implications of a CEO's dedication to social issues in a business setting.

5.5 Practical Significance of CEOs Posting Images of People

From an investor's standpoint, the practical significance of the results is limited. A causal relationship has not been established so it is not certain that images of people cause changes in the stock price. For example, another preceding type of event might lead to both CEOs posting images of people and stock prices to changes in which case the change in the stock price might already have occurred by the time a CEO posts images of people. Even if a causal relationship would be established the practical significance to investors would be limited. The effect of the people coefficient is very small so a lot of money would have to be invested at once to make any money. An example can illustrate this: if one would sell $10,000 worth of stocks of a company when the CEO posts an image containing a person the gained value on average would be around 0.00018*10,000 = 1.8 dollars. From this should then still be distracted transaction costs. In addition, it would be necessary to monitor all CEO's Twitter accounts for tweets and scan them for containing images of people which would bring costs with it as well. Furthermore, according to the efficient market hypothesis (Malkiel & Fama, 1970) all new information publicly available should be directly incorporated into the stock price which would make it impossible to make a profit as soon as it is clear to the public that pictures of people in CEOs tweets have a negative effect on the stock price.

For CEOs and their PR teams, the practical significance of the results is bigger. Even though it is not certain that there is a causal relationship between images in CEOs' tweets and share price declines, CEOs may want to adhere to the adage, better safe than sorry. As a precaution to prevent any share price drops that may be caused by a CEO including images of people in his tweets, CEOs may want to refrain from using images of people in their tweets. Otherwise, the CEOs should at least be aware of how images of people in their tweets may be received by the stock market and be more cautious when composing their tweets and the message it sends.

6 Limitations

There are several limitations to this research that should be discussed. One of the limitations is that tweets are analyzed on a per-instance basis. By doing so the larger whole that a tweet might be part of is not considered as such. For example, a CEO might use multiple tweets in a thread to circumvent the character limit of one tweet to create a bigger story. In this case, the whole story as a whole might have an effect on the stock price of the company that can not be properly inferred from the individual tweets. Another case is CEOs respond to another user's tweet, their response is analyzed in isolation, without considering the context of the original tweet. Once again, the collective tweets might hold important information for stock price valuations that can not be extracted from just the CEOs' response alone.

Another limitation of this research is that the abnormal returns can be calculated only once for each trading date because only the closing price of stocks is available. Because of this, a substantial time may pass between a tweet and the point in time the abnormal returns attributed to the tweet are calculated for. This is especially the case if a tweet was not posted outside trading hours or on a non-trading day, so it has to be matched with the following first trading day. During the intermediate time, numerous other events may take place which also have an effect on the stock price, making the relationship between the tweet and its' effect on the stock price less clear.

The lack of establishing a causal relationship between CEOs' usage of images of people and stock price declines is another limitation of this research, which makes it impossible to state that the former causes the latter. For instance, It might be the case that a particular type of preceding event prompts CEOs to post images of people and cause stock price declines. An example might be a company whose working conditions have been badly reported in the news. This might cause CEOs to post good things about the working conditions in the company and post pictures of people to counter the negative publicity. At the same time, the negative publicity might cause the company's stock to fall. Future research could attempt to establish a causal relationship, for example by testing for Granger causality.

In addition to the limitations already mentioned, the data analyzed in this study may be biased and not representative of all types of companies. The companies included in the analysis are all listed in the United States, it is possible that the results are not applicable to companies listed in other countries. Furthermore, not all the CEOs of the companies in the analyzed S&P 1500 were on Twitter. There exists bias regarding which CEOs are on Twitter, (Napora, 2019), (Oh & Bunkanwanicha, 2016) find that CEOs active on Twitter are predominantly CEOs of companies active in the technology/IT and consumer goods sectors. Other research (Bruinsma, 2022) finds that CEOs active on Twitter are mostly male between 45 and 65. Therefore the inferences made in this research might not be valid in general.

The correction of CEOs' leaving/starting date constitutes a limitation to the research, this was, however, not identified in time to still be addressed. For some CEOs, the date of leaving/starting could not be exactly determined. In these cases, the starting date is designated as the first day of the month in which the role as CEO was initiated, while the leaving date is determined as the day immediately following the month of resignation. This results in the possibility of tweets from outside a CEO's term also ending up in the tweets sample. In future research, it would be preferred to keep the inclusion of tweets from outside a CEO's term to a minimum. One way to achieve this would be to set the date of starting (leaving) to the first date of the next (preceding) month. The rationale behind minimizing the number of tweets outside a CEO's tenure is that those tweets will likely not have an impact on the stock price of the company anymore and the scope of the research is only the period that a CEO is in office. This approach would mean that fewer tweets from inside the CEO's term end up in the tweet sample, but this is regarded as less severe than taking into account tweets from when a CEO does not hold the position.

7 Future work

Future research could build on this study in several ways. One of the improvements could be to examine the specific circumstances under which images of people in CEOs' tweets negatively impact stock prices. It is conceivable that effects depend on the context in which the images are used. This possibility has already been touched upon in this research by examining whether the impact of images of people depends on the tweet discussing business matters.

Future work could also aim at adapting the models used to extract features from CEOs' tweets for the specific use of tweets. The majority of the models used to extract features from the tweets are not trained specifically on tweets. Tweets may contain elements that are specific to them such as hashtags, abbreviations and have a character limit that makes them different from other sources. Adapting the models or creating new ones that are specifically trained on Twitter data may improve the features used in the models and, consequently, the results.

Furthermore, future work could explore the relationship between stock price changes and other types of images than those having people on them. To investigate this, the cluster methods of this research could be used. Another option would be to use image classification models which could be used to classify the contents of an image. This way it is also possible to get multiple topics for an image, capturing the different subtleties in it which all may differently impact stock prices. This makes it possible to relate more specific aspects of images to the returns and thereby more precisely determine the effect of an image on the stock market.

8 Conclusion

This study investigated the impact of visual cues used by CEOs in their Twitter communication on stock prices. The results show that CEOs' usage of images featuring people, negatively affects the stock price of their company. This research is an important contribution to the locus of research on CEOs' social media communication and stock market reactions since it examined the previously unexplored relation between visual cues in CEO communication and stock market reactions. The approach taken in this research to study visual cues can be applied to investigate visual cues in CEO communication more broadly as well as to study other visual cues in CEO communication that may impact the stock market.

References

Akansu, A., Cicon, J., Ferris, S. P., & Sun, Y. (2017). Firm performance in the face of fear: How ceo moods affect firm performance. *Journal of Behavioral Finance*, *18*(4), 373-389. Retrieved from https://doi.org/10.1080/15427560.2017.1338704 doi: 10.1080/15427560.2017.1338704

Amit, E., Hoeflin, C., Hamzah, N., & Fedorenko, E. (2017). An asymmetrical relationship between verbal and visual thinking: Converging evidence from behavior and fmri. *NeuroImage*, *152*, 619-627.

Barnett, M. L., & Salomon, R. M. (2006, November). Beyond dichotomy: the curvilinear relationship between social responsibility and financial performance. *Strategic Manage. J.*, *27*(11), 1101–1122.

Brammer, S., Brooks, C., & Pavelin, S. (2006). Corporate social performance and stock returns: Uk evidence from disaggregate measures. *Financial Management*, *35*(3), 97–116. Retrieved 2023-06-10, from http://www.jstor.org/stable/30137803

Brown, N. C., Elliott, W. B., & Grant, S. M. (2019). Gaming regulation with image-based tweets. *Behavioral & Experimental Accounting eJournal*.

Bruinsma, M. (2022). *Exploring themes of ceo social media communication* [Thesis]. https://theses.liacs.nl/pdf/2022-2023-BruinsmaM.pdf.

Chee, M. W., Weekes, B., Lee, K. M., Soon, C. S., Schreiber, A., Hoon, J. J., & Chee, M. (2000). Overlap and dissociation of semantic processing of chinese characters, english words, and pictures: Evidence from fmri. *NeuroImage*, *12*(4), 392-403. Retrieved from https://www.sciencedirect.com/science/article/pii/S1053811900906311 doi: https://doi.org/10.1006/nimg.2000.0631

Chemtob, C. M., Roitblat, H. L., Hamada, R. S., Muraoka, M. Y., Carlson, J. G., & Bauer, G. B. (1999). Compelled attention: The effects of viewing trauma-related stimuli on concurrent task performance in posttraumatic stress disorder. *Journal of Traumatic Stress*, *12*(2), 309-326. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1023/A%3A1024728626505 doi: https://doi.org/10.1023/A:1024728626505

Chiah, M., Hu, X., & Zhong, A. (2022). Photo sentiment and stock returns around the world. *Finance Research Letters*, *46*, 102417. Retrieved from https://www.sciencedirect.com/science/article/pii/S1544612321004074 doi: https://doi.org/10.1016/j.frl.2021.102417

Choudhury, P., Wang, D., Carlson, N. A., & Khanna, T. (2019). Machine learning approaches to facial and text analysis: Discovering ceo oral communication styles. *Strategic Management Journal*, *40*(11), 1705-1732. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/smj.3067 doi: https://doi.org/10.1002/smj.3067

Coelho, R., Jayantilal, S., & Ferreira, J. J. (n.d.). The impact of social responsibility on corporate financial performance: A systematic literature review. *Corporate Social Responsibility and Environmental Management*, *n/a*(n/a). Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/csr.2446 doi: https://doi.org/10.1002/csr.2446

Crace, L., & Gehman, J. (2023). What really explains esg performance? disentangling the asymmetrical drivers of the triple bottom line. *Organization & Environment*, *36*(1), 150-178. Retrieved from https://doi.org/10.1177/10860266221079408 doi: 10.1177/10860266221079408

Cronqvist, H., Makhija, A. K., & Yonker, S. E. (2012). Behavioral consistency in corporate finance: Ceo personal and corporate leverage. *Journal of Financial Economics*, *103*(1), 20-40. Retrieved from https://www.sciencedirect.com/science/article/pii/S0304405X11001851 doi: https://doi.org/10.1016/j.jfineco.2011.08.005

Fahlenbrach, R. (2009). Founder-ceos, investment decisions, and stock market performance. *Journal of Financial and Quantitative Analysis*, *44*(2), 439–466. doi: 10.1017/S0022109009090139

FAMA, E. F., & FRENCH, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, *47*(2), 427-465. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1992.tb04398.x doi: https://doi.org/10.1111/j.1540-6261.1992.tb04398.x

Flam, R., Green, J., & Sharp, N. Y. (2020). Do investors respond to CEO facial expressions of anger during television interviews? *SSRN Electronic Journal*. Retrieved from https://doi.org/10.2139/ssrn.3740755 doi: 10.2139/ssrn.3740755

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for common sense: An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, *8*, 89-106.

Guindy, M. A., & Riordan, R. (2017). Tweeting the good news: Returns and price informativeness. *Social Science Research Network*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2999443.

Harrison, J. S., Thurgood, G. R., Boivie, S., & Pfarrer, M. D. (2020). Perception is reality: How ceos' observed personality influences market perceptions of firm risk and shareholder returns. *Academy of Management Journal*, *63*(4), 1166-1195. Retrieved from https://doi.org/10.5465/amj.2018.0626 doi: 10.5465/amj.2018.0626

Hirshleifer, D., & Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, *36*(1), 337-386. Retrieved from https://www.sciencedirect.com/science/article/pii/S0165410103000648 (Conference Issue on) doi: https://doi.org/10.1016/j.jacceco.2003.10.002

Hu, A., & Flaxman, S. (2018, jul). Multimodal sentiment analysis to explore the structure of emotions. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery &amp data mining.* ACM. Retrieved from https://doi.org/10.1145%2F3219819.3219853 doi: 10.1145/3219819.3219853

Knipmeijer, K. (2020, Apr). *The effect of ceo tweets on the stock market activity* [thesis]. https://thesis.eur.nl/pub/53916/Master-Thesis-416147.pdf.

Malhotra, C. K., & Malhotra, A. (2016). How ceos can leverage twitter. *MIT Sloan Management Review*, *57*, 73-79.

Malkiel, B. G., & Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work*. *The Journal of Finance*, *25*(2), 383-417. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1970.tb00518.x doi: https://doi.org/10.1111/j.1540-6261.1970.tb00518.x

McInnes, L., Healy, J., & Melville, J. (2018). *Umap: Uniform manifold approximation and projection for dimension reduction.* arXiv. Retrieved from https://arxiv.org/abs/1802.03426 doi: 10.48550/ARXIV.1802.03426

Moulavi, D., Jaskowiak, P. A., Campello, R. J. G. B., Zimek, A., & Sander, J. (n.d.). Density-based clustering validation. In *Proceedings of the 2014 siam international conference on data mining (sdm)* (p. 839-847). Retrieved from https://epubs.siam.org/doi/abs/10.1137/1.9781611973440.96 doi: 10.1137/1.9781611973440.96

Napora, P. (2019, Aug). *The changing dynamics of corporate messaging – how ceos' twitter posts affect their companies' stock prices?* [thesis]. https://arno.uvt.nl/show.cgi?fid=150422.

Nekrasov, A., Teoh, S. H., & Wu, S. (2020). Visuals and attention to earnings news on twitter. *Review of Accounting Studies*, *27*, 1233 - 1275.

Obaid, K., & Pukthuanthong, K. (2022). A picture is worth a thousand words: Measuring investor sentiment by combining machine learning and photos from news. *Journal of Financial Economics*, *144*(1), 273-297. Retrieved from https://www.sciencedirect.com/science/article/pii/S0304405X21002683 doi: https://doi.org/10.1016/j.jfineco.2021.06.002

O'brien, R. M. (2007, Oct 01). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*(5), 673-690. Retrieved from https://doi.org/10.1007/s11135-006-9018-6 doi: 10.1007/s11135-006-9018-6

Oddo, R. (2022, August). *To what extent do esg scores affect firm valuation? : Evidence based on large dutch enterprises* [thesis]. Retrieved from http://essay.utwente.nl/92863/

Oh, A. J., & Bunkanwanicha, P. (2016, Apr). *Ceos on twitter.* Aachen : R. Piskac c/o Redaktion Sun SITE, Informatik V, RWTH Aachen. Retrieved from https://ceur-ws.org/Vol-1691/paper_04.pdf

Paivio, A. (1971). *Imagery and verbal processes.* New York: Holt, Rinehart & Winston.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., ... Sutskever, I. (2021). Learning transferable visual models from natural language supervision. *CoRR*, *abs/2103.00020*. Retrieved from https://arxiv.org/abs/2103.00020

Randewich, N. (2020, may). *Tesla tumbles after musk tweets stock too high.* Thomson Reuters. Retrieved 2023-06-03, from https://www.reuters.com/article/us-tesla-stocks-musk-idUSKBN22D5V7

Tsai, H.-J., & Wu, Y. (2022, Jul 01). Changes in corporate social responsibility and stock performance. *Journal of Business Ethics*, *178*(3), 735-755. Retrieved from https://doi.org/10.1007/s10551-021-04772-w doi: 10.1007/s10551-021-04772-w

Wernicke, G., Sajko, M., & Boone, C. (2022). How much influence do ceos have on company actions and outcomes? the example of corporate social responsibility. *Academy of Management Discoveries*, *8*(1), 36-55. Retrieved from https://doi.org/10.5465/amd.2019.0074 doi: 10.5465/amd.2019.0074

You, Y., Srinivasan, S., Pauwels, K., & Joshi, A. (2020, Nov 01). How ceo/cmo characteristics affect innovation and stock returns: findings and future directions. *Journal of the Academy of Marketing Science*, *48*(6), 1229-1253. Retrieved from https://doi.org/10.1007/s11747-020-00732-4 doi: 10.1007/s11747-020-00732-4

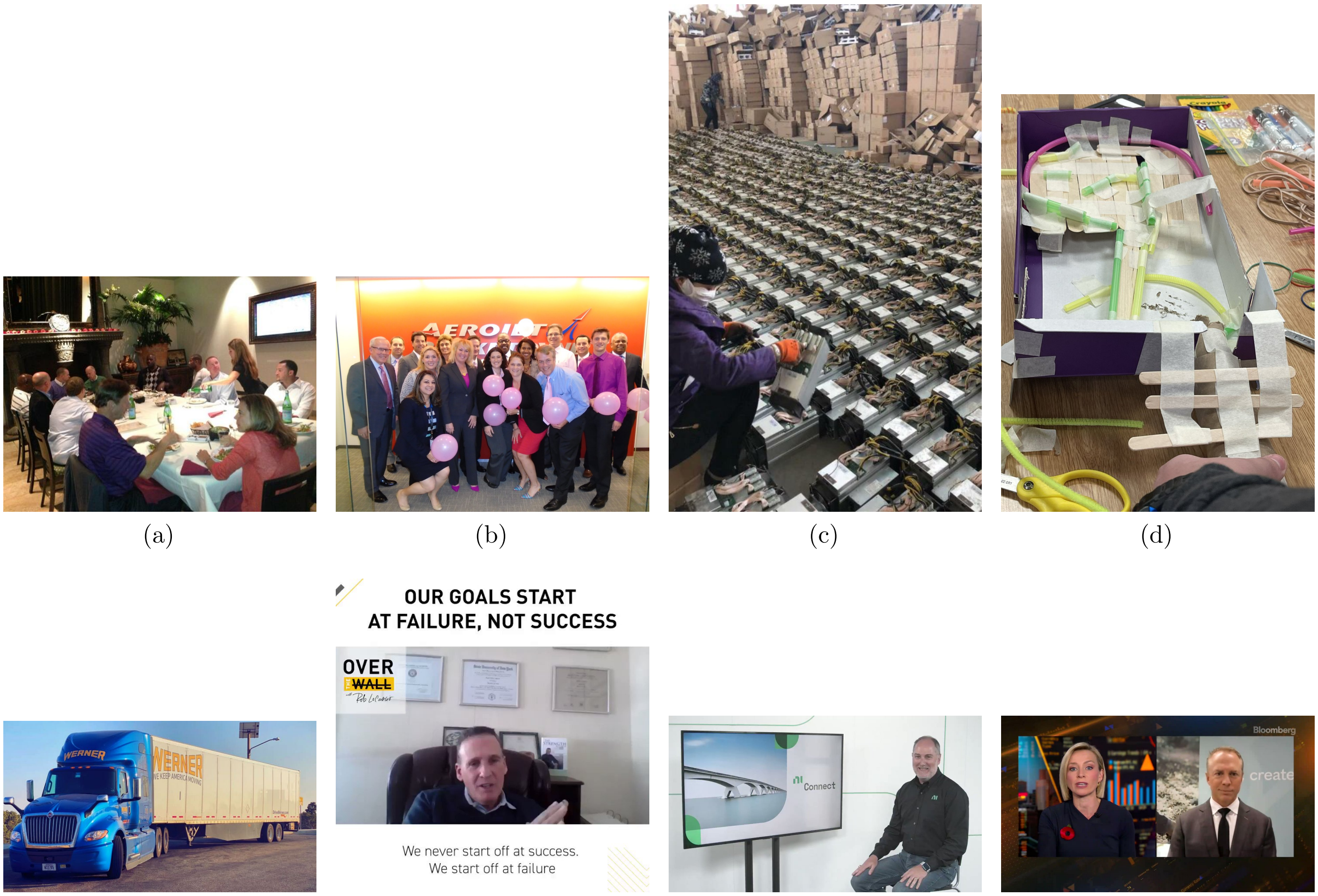
Zhang, Y., & Wiersema, M. F. (2009). Stock market reaction to ceo certification: the signaling role of ceo background. *Strategic Management Journal*, *30*(7), 693-710. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1002/smj.772 doi: https://doi.org/10.1002/smj.772

Zohrehvand, A. (2022, June). *Fifty Million Followers Can't Be Wrong, or Can They? Effects of Social Media Feedback on CEO Communication.* Retrieved from https://ssrn.com/abstract=4141444 doi: http://dx.doi.org/10.2139/ssrn.4141444

A Image Cluster Samples

This section contains samples of images of every cluster. The images are manually selected to give an overview of the variety of pictures that are assigned to a cluster. For example, the Sports cluster (Figure 3) does include images of people sporting but also of logos, clubs, sports items, statistics, and tournament overviews. In addition, pictures were selected that do not fit well with the label given to the cluster. For example, Figure 1e, part of the people cluster contains a lorry and no human being so it does not fit with the "People" label of the cluster. Additionally Figure 8, a sample of the unclustered images shows that there are also images that are not clustered that would have fitted under the label of some of the clusters. For example figure 8f would have fitted with the label "Space", Figure 8g with "Sport" and Figure 8h with "Food".

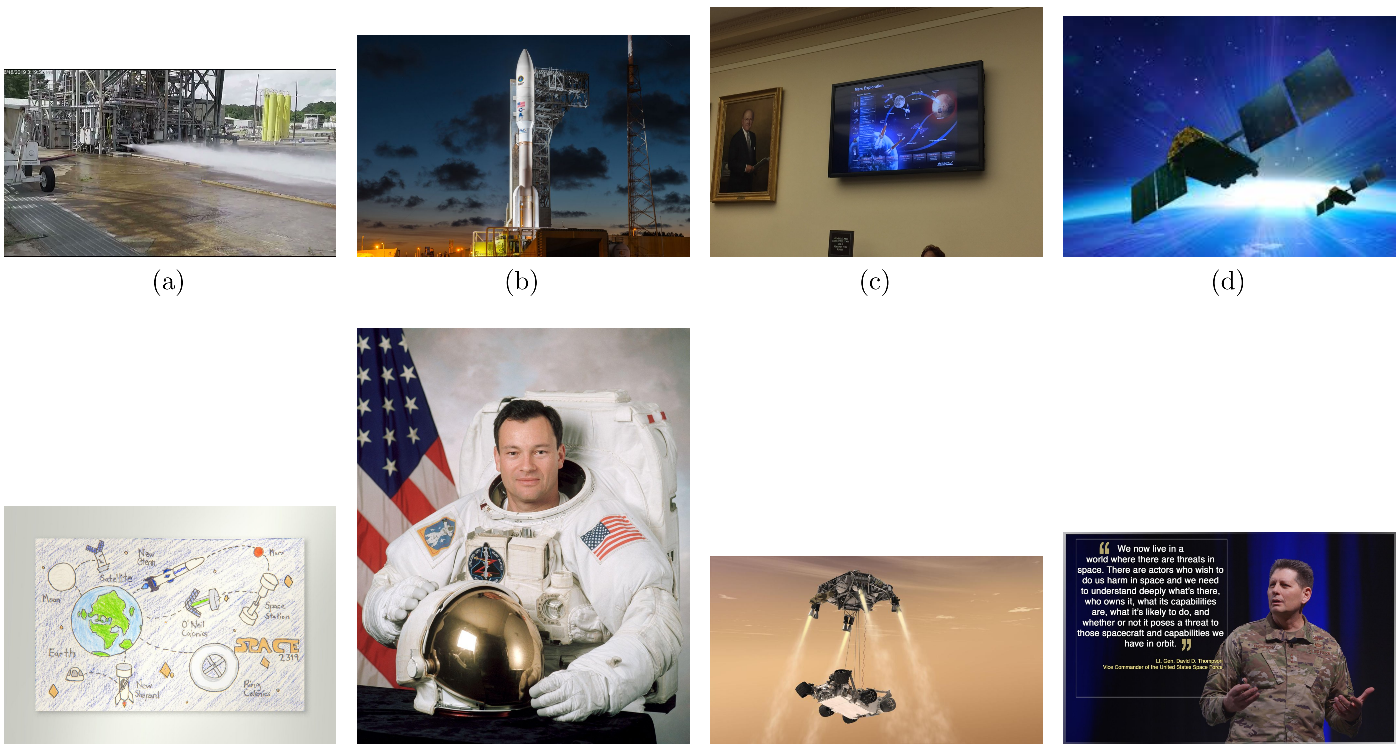
Because the images in this sample were selected manually on variety and non-fittingness, the samples are not representative of the cluster as a whole. They do however show that the labels given to clusters are not matching all the images in that cluster and that clusters can have differing kinds of images that relate to the same topic. To get a representative overview of the images in a cluster one may look at the following dataset that consists a larger sample of 100 randomly selected images of each cluster.



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 1: Image sample of the "People" cluster



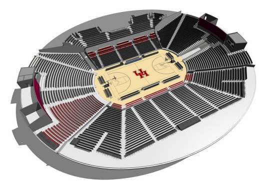
(a) (b) (c) (d)

(e) (f) (g) (h)

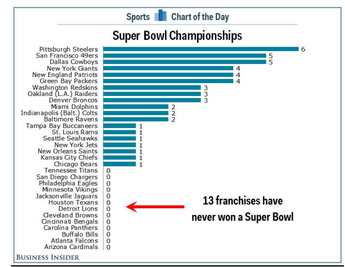
Figure 2: Image sample of the "Space" cluster

(a)

(b)



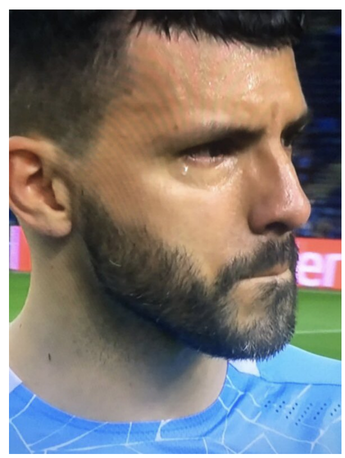
(c)



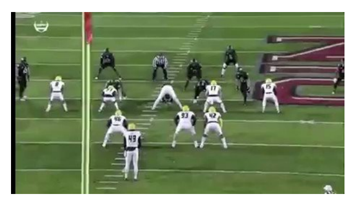
(d)

(e)

(f)

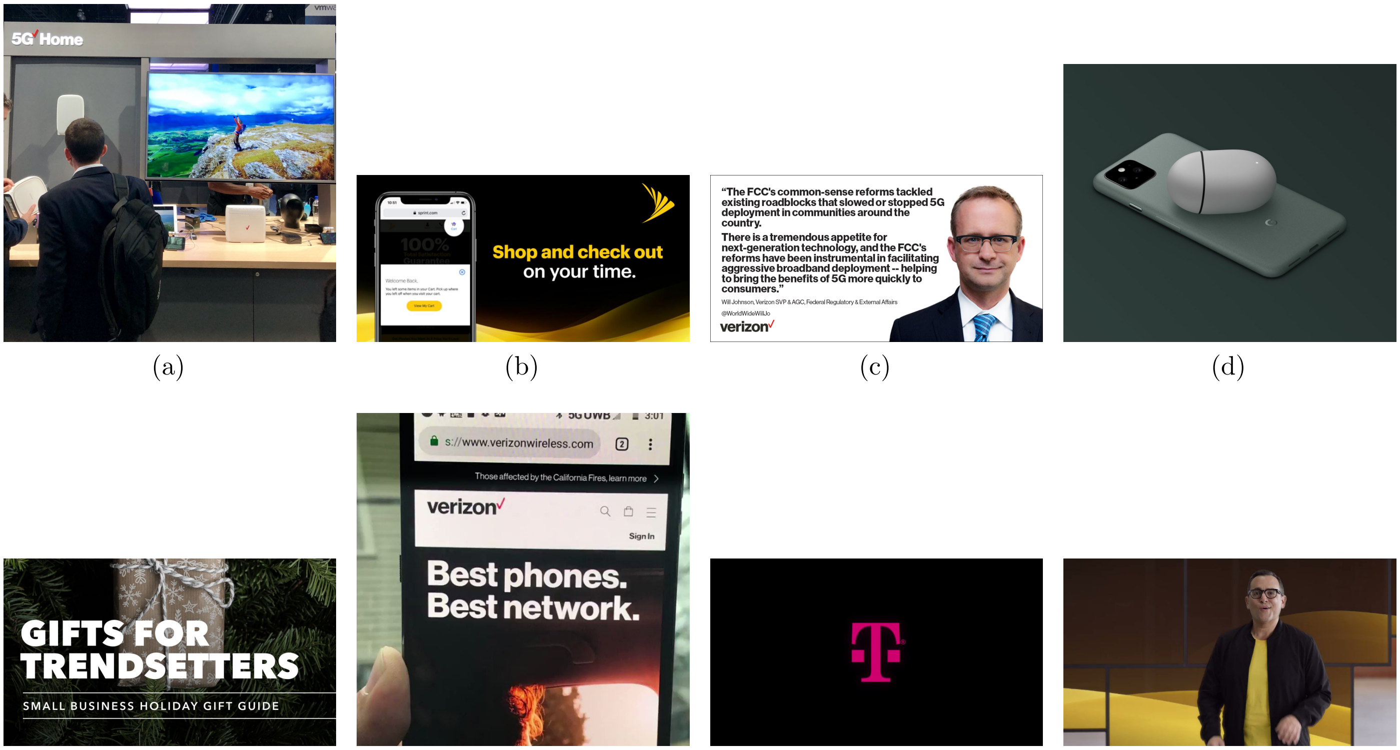


(g)



(h)

Figure 3: Image sample of the "Sport" cluster



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 4: Image sample of the "Telecom" cluster

(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 5: Image sample of the "Poster" cluster

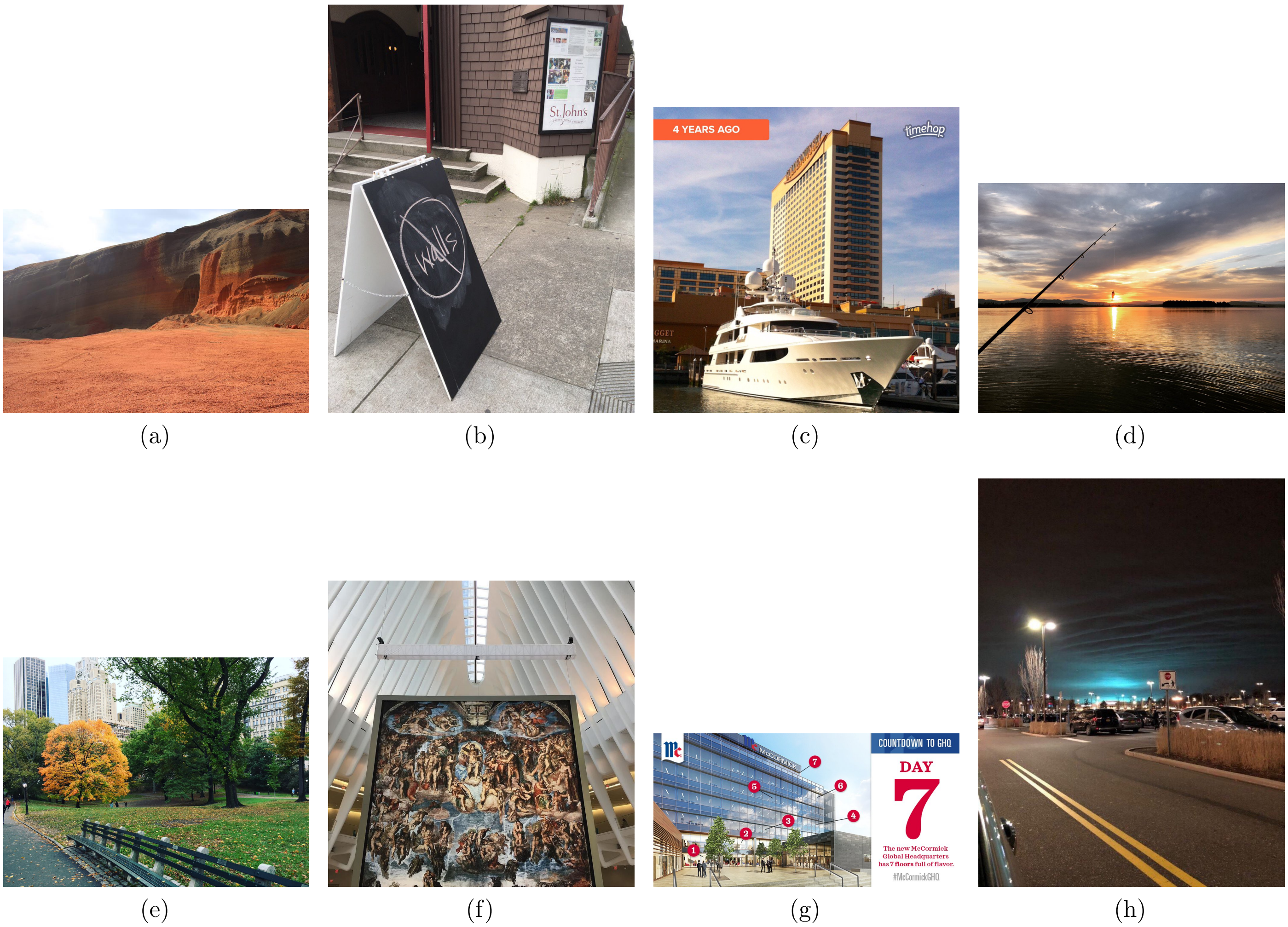


Figure 6: Image sample of the "Landscape" cluster

(a)

(b)

(c)

(d)

(e)

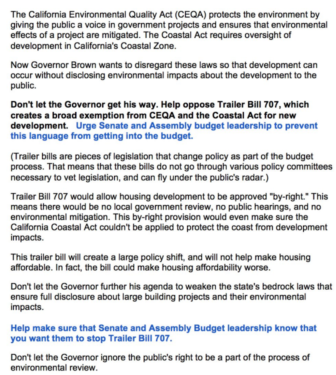
(f)

(g)

(h)

Figure 7: Image sample of the "Food" cluster

(a)



The California Environmental Quality Act (CEQA) protects the environment by giving the public a voice in government projects and ensures that environmental effects of a project are mitigated. The Coastal Act requires oversight of development in California's Coastal Zone.

Now Governor Brown wants to disregard these laws so that development can occur without disclosing environmental impacts about the development to the public.

Don't let the Governor get his way. Help oppose Trailer Bill 707, which creates a broad exemption from CEQA and the Coastal Act for new development. **Urge Senate and Assembly budget leadership to prevent this language from getting into the budget.**

(Trailer bills are pieces of legislation that change policy as part of the budget process. That means that these bills do not go through various policy committees necessary to vet legislation, and can fly under the public's radar.)

Trailer Bill 707 would allow housing development to be approved "by-right." This means there would be no local government review, no public hearings, and no environmental mitigation. This by-right provision would even make sure the California Coastal Act couldn't be applied to protect the coast from development impacts.

This trailer bill will create a large policy shift, and will not help make housing affordable. In fact, the bill could make housing affordability worse.

Don't let the Governor further his agenda to weaken the state's bedrock laws that ensure full disclosure about large building projects and their environmental impacts.

Help make sure that Senate and Assembly Budget leadership know that you want them to stop Trailer Bill 707.

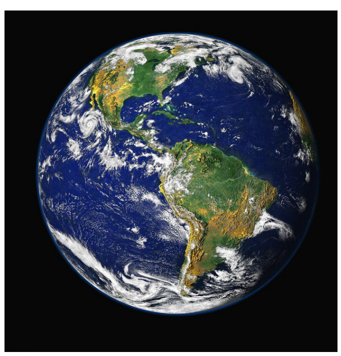
Don't let the Governor ignore the public's right to be a part of the process of environmental review.

(b)

(c)

(d)

(e)



(f)

(g)

(h)

Figure 8: Image sample of the "Unclustered" images

B Image Cluster Visualizations

This appendix contains visualizations of the image clusters. To create the visualizations the embeddings had to be turned to two-dimensional. As to cluster the images, UMAP has been used to reduce the number of dimensions. The settings used for UMAP are the same as for clustering except for the number of components which has been set to two so as to end up with just two dimensions. Reducing the dimensions to two instead of the 50 used for clustering results inevitably in the loss of data that was present when clustering and in different distances between points from the 50-dimensional representations that clustering was executed on. Because of this, the visualizations are not a perfect reflection of the clustered data, and also because UMAP is not deterministic. As a consequence of this it might look like some points are assigned to a different cluster than would be expected from the visualizations. Still, the visualizations are a useful tool for analyzing the clusters.

figure 9 is a visualization of all the clusters except for the images that could not be clustered. From this figure one might tell that there are points that lie on the border between different clusters, these are expected to be points that share properties of different clusters. It can also be seen that there are clusters that have a little tangential relationship with other clusters. For example, the Space cluster seems to be distant from all the other clusters. Figure 10 is a visualization of all the clusters, also including the images that could not be clustered. Judging from this visualization it looks like a part of the unclustered images are in dense regions that could have been clusters of their own. Because of the settings used with HDBSCAN, these regions will contain not enough images to become clusters and so they are discarded. There are also unclustered data points that seem to fall into one of the clusters. For these, it might be the case that they are too far of from data points in the cluster to get assigned to it because there are not enough pictures in the cluster that are close enough to it.

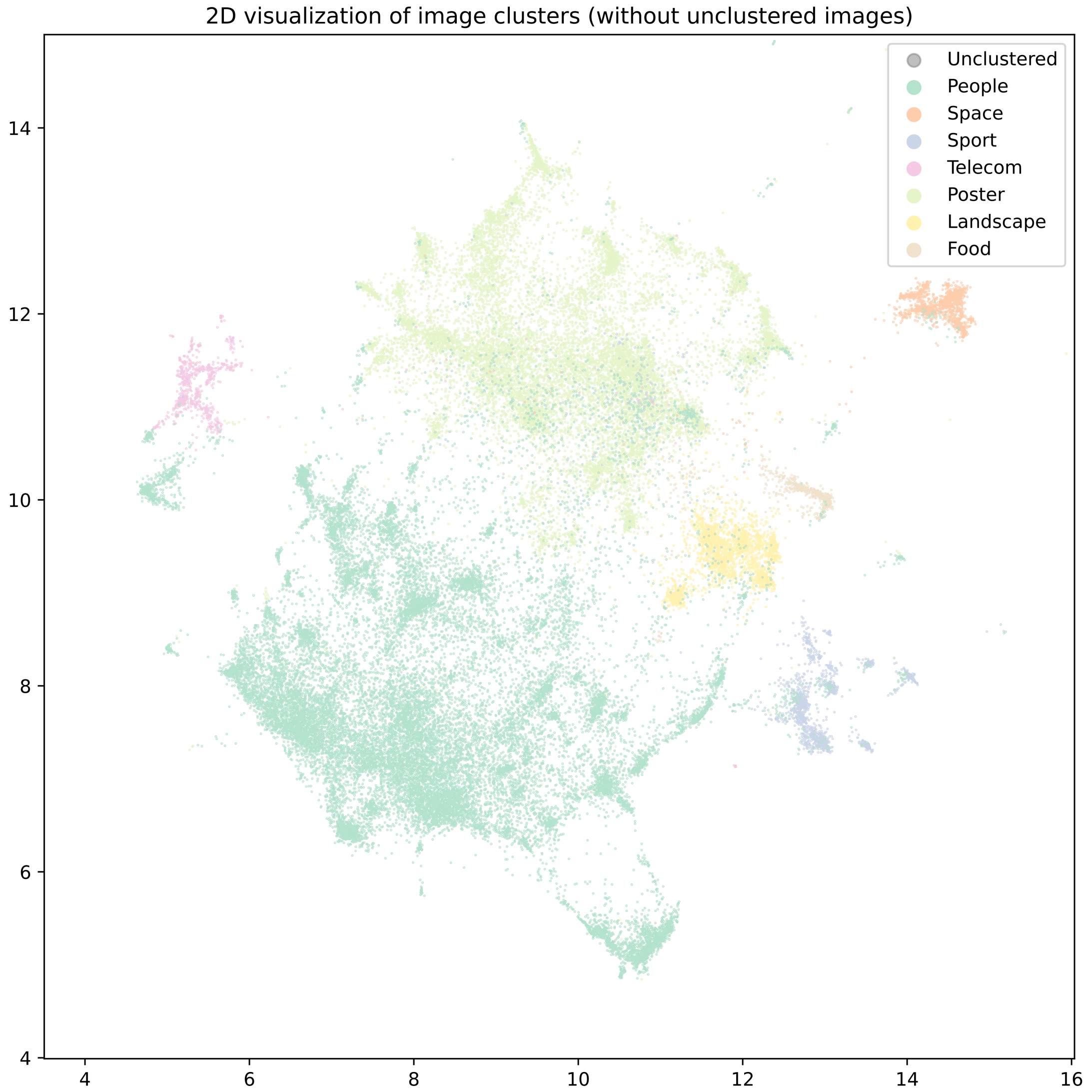


Figure 9: Visualization of the image clusters excluding the unclustered images

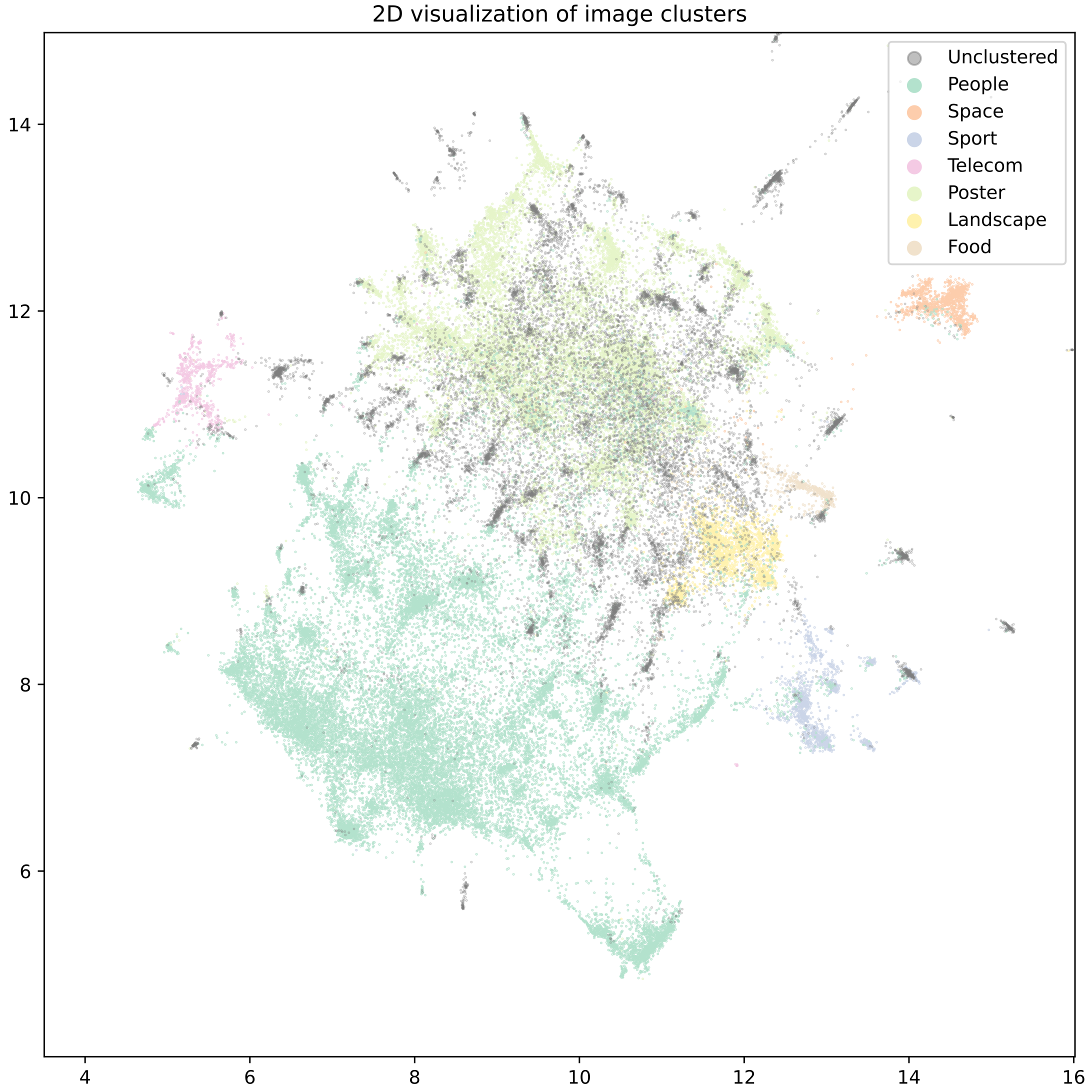


Figure 10: Visualization of all image clusters as well as the unclustered images

C Text Clusters

This appendix contains a table with all the text clusters. the "cluster" column contains a numerical identifier for each cluster. The "tfidf 10 most important words in cluster" column contains the ten most important words in the cluster are displayed to get an idea of the contents of the cluster. The 10 most important words in the cluster are determined by calculating the tf-idf score for every word in the cluster and ranking them. The "Assigned cluster topic" column contains a topic that was assigned to the topic by manually looking at tweets in the text cluster and searching for a common topic. The "overarching topic" column contains a topic for multiple clusters together that those clusters together in common as a higher-order subject. The "comment" column as last contains some notes about the cluster to get a better understanding of the cluster contents.

cluster	tfidf 10 most important words in cluster	Assigned cluster topic	Overarching topic	comment
0	billiondollarbuyer, see, deal, thanks, think, new, make, let, business, good	other		the tweets are all about billion dollar buyer, a reality tv show from Tillman Feritta, CEO of Landry's. He tries new horecaproducts in the show
1	launch, space, iridium, engine, rocket, satellite, mission, today, next, mars	own business	business	tweets about their own (space) product, their own business
2	bitcoin, digital, energy, money, gold, world, asset, property, time, ha	other		mostly talking about bitcoin/equity investing. Not per se about their own company
3	davos, wef, world, wef14, wef17, great, wef18, wef19, global, wef15	other		tweets about the world economic forum
4	great, people, one, ceo, business, new, make, best, say, via	other		tweets are all about time, future, present, past etc
5	yes, wow, de, lol, yep, ha, yup, thx, la, en	social interaction	social	tweets are mostly very short reactions on others, abbreviations of one or a few words and texts that are not English
6	ceridian, great, ceridian-insights, dayforce, team, proud, amazing, makesworklifebetter, hcm, customer	own business	business	All tweets are about ceridian a HRM company, their own business mostly
7	new, health, covid19, covid, patient, drug, vaccine, healthcare, 19, help	multi topic		All tweets are about health care, some are from CEOs of companies in healthcare, others are more CEO activism of CEOs thanking the nurses etc
8	5g, sprintfam, customer, new, get, great, network, sprint5g, today, team	own business	business	All tweets are about telecom (own business)

9	wa, today, king, life, family, great, dr, rip, legacy, one	showing compassion	social	Tweets about those that are diseased, mostly positive things about them
10	game, great, win, go, team, congrats, season, good, tonight, let	public relations	social	Tweets about sports
11	kodak, film, movie, thank, shot, new, great, see, theatre, via	own business	business	Tweets about their own (movie related)business. For example CEOS of Kodak, Disney, Movie theaters etc
12	thought, affected, help, impacted, family, prayer, safe, support, hurricane, relief	showing compassion	social	showing empathy for everyone affected by diseasters
13	thank, veteran, service, veterans, day, country, sacrifice, today, served, honor	showing compassion	social	Showing empathy for veterans
14	driver, werner, professional, thank, team, truck, million, free, accident, today	own business	business	mostly tweets about own (logistics) business
15	happy, year, thanksgiving, christmas, holiday, new, family, wishing, merry, everyone	public relations	social	Tweets about holidays, both about the own company as wishing other people a good one
16	live, nation, via, music, concert, tour, new, festival, congrats, show	own business	business	mostly tweets from CEOs about their own business (media, entertainment, events)
17	great, kw2013, thanks, cf2014, day, looking, team, today, forward, ceo	public speaking	business	tweets about conferences/summits/keynotes
18	dell, delltechnologies, great, new, thanks, customer, dellworld, partner, team, nxtwork	own business	business	mostly tweets about the own business (Dell)

19	jobs, hiring, employment, job, employer, outlook, unemployment, meos, us, survey	multi topic		mostly tweets about the labor market, a lot from the same CEO of Manpowergroup that is active in the business, also from other CEOs discussing the topic
20	redhat, red, hat, rhsummit, openstack, open, great, cloud, rackspace, source	own business	business	Mostly tweets about their own business (Cloud, software, hardware)
21	water, letssolvewater, world, proud, great, help, day, challenge, colleague, solve	own business	business	Mostly tweets about their own business (water solutions)
22	car, via, tesla, vehicle, new, auto, electric, cars, driving, best	own business	business	Mostly tweets about their own business (cars)
23	infrastructure, energy, solar, future, need, via, city, power, new, investment	multi topic		Both tweets from CEOs that are in the business of renewable energy (consulting firms, solar panel firms) and CEO activism tweets
24	team, great, woman, proud, today, work, thank, women, year, world	own business	business	CEOs tweeting about the internals of the company as well as partnerships with others
25	ceo, today, great, team, new, thank, year, join, proud, excited	own business	business	CEOs tweeting about the internals of the company as well as partnerships with others
26	people, business, leader, company, work, leadership, culture, employee, help, make	leadership	business	Tweets about leadership and entrepeneurship in general, not about the own business. Also linking to articles
27	trump, wa, like, would, one, people, ha, time, get, know	CEO activism	social	Mostly tweets about politics
28	ai, via, human, new, machine, learning, intelligence, data, robot, technology	multi topic		Tweets about how AI impacts CEOs own business but also how it impacts society

29	google, search, antitrust, ha, ad, via, new, tech, apple, android	CEO activism	social	Mostly CEO activism about BigTech (a few tweets from the ceos of these companies themselves)
30	cloud, security, data, new, cybersecurity, customer, via, business, solution, great	multi topic		Both tweets of CEOs from cybersec firms as well as CEOs talking about the impact of security on their own business and society as a whole
31	day, great, morning, beautiful, good, new, today, run, nyc, back	public relations	social	Tweets of CEOs about landscapes, buildings, views
32	retail, store, via, new, shopping, holiday, retailer, cre, experience, online	multi topic		Mainly tweets of one CEO that is in the business of mall investments, also tweets from CEOs of firms in the business of analyzing the retail sector
33	leadership, via, leader, people, great, leaders, culture, work, team, make	leadership	business	Tweets about leadership and entrepeneurship in general, not about the own business. Also quoting famous people
34	via, marketing, data, customer, social, business, digital, contact, email, constant	multi topic		Both tweets from CEOs of marketing companies as from CEOs of analytical companies giving advice on marketing
35	love, agree, one, good, yes, great, get, know, right, awesome	social interaction	social	Short tweets containing positive words (agree, great, right etc.)
36	great, thanks, today, thank, show, team, meeting, forward, morning, time	public speaking	business	CEOs tweeting about talking at shows/meetings/radio etc
37	thanks, thank, congrats, congratulations, team, great, proud, well, us, much	social interaction	social	thanking people and congratualting other people

Table 4: Text clusters label assignments

D Regression Variables Overview

- **i** and **j** are indices where j is an index for the CEO and i is an index for the event.
- $\mathbf{AR(0)}_{i,j}$ Are the abnormal returns for the stock price of the company on the trading day of the event.
- $\boldsymbol{\beta_k}$ are the coefficients of the variables in the model.
- **const** is the intercept of the regression model.
- $\boldsymbol{\epsilon}_{i,j}$ is a term for the error between the value predicted by the model and the observed value.
- $\mathbf{individual}_j$ is the fixed effects term that is added to take into account CEO-specific traits that are stable over time.
- $\mathbf{experience}_{i,j}$ is constructed by taking the log of the average number of tweets from the first tweet in the CEO sample to the beginning of the event till the end of the event. So the calculation is: $log(\frac{tweet_t+tweet_{t+n}}{2})$ where t is the number of tweets tweeted at the beginning of the event and n is the number of tweets in the event itself.
- $\mathbf{nTweets}_{i,j}$ is the number of tweets that the event comprises of.
- $\mathbf{people}_{i,j}$ This variable is equal to the total number of images of people that a CEO has posted on the trading day.
- $\mathbf{business}_{i,j}$ This variable is equal to the number of tweets in the event that are part of one of the clusters together forming the merged "business" cluster.
- $(\mathbf{business{*}people})_{i,j}$ this variable is an interaction variable constructed by taking the product of $\mathbf{people}_{i,j}$ and $\mathbf{business}_{i,j}$

E Variables Correlation Matrix and VIFs

E.1 Correlation Matrix

	Correlation Matrix			
	people	**business**	**business∗people**	**experience**
business	0.25			
business∗people	0.63	0.46		
experience	0.09	0.18	0.07	
nTweets	0.36	0.5	0.39	0.23

Table 5: Correlations between variables used in analysis

E.2 VIFs

Variance Inflation Factors	
VIF Factor	**feature**
1.75	people
1.54	business
2.02	business∗people
1.07	experience
1.48	nTweets

Table 6: Variance Inflation Factors of the variables used in the analysis