Goldman Sachs

CORPORATE DISCLOSURES ANALYSIS UNSTRUCTURED DATA ANALYSIS of 10K REPORTS

FOR THE YEARS (2009 -2019)

Contents

[**Introduction**](#_heading=h.aomrt3r4zm3h) **2**

[**Corporate Disclosures Analysis**](#_heading=h.30j0zll) **2**

[Analysis / Comparison of Financial Statement Metrics with Sentiment](#_heading=h.1fob9te) 2

[Analysis / Comparison of Financial Statement Metrics with Complexity](#_heading=h.aywe0rra8lmq) 3

[Management’s Discussion and Analysis (MD&A) Sentence Sentiment and Gunning Fog Complexity Comparison](#_heading=h.3znysh7) 4

[WordCloud](#_heading=h.5kw79iha75qi) 5

[Forecast](#_heading=h.vxkf3dkbir2t) 5

[**Twitter Data Analysis**](#_heading=h.8rvr9k7wck6t) **5**

[Introduction](#_heading=h.4xxpqbewwuo2) 6

Sentiment Analysis 7

[Goldman Sachs Twitter Word Cloud for Tweets After 2019-01-01](#_heading=h.tyjcwt) 8

[**Combined Observations From Section1 and Section2**](#_heading=h.3n9fbo6u4nqr) **9**

[Positive Sentiment](#_heading=h.exuqtikbh4rd) 10

[Research and Technology](#_heading=h.8tvl1j31bai2) 10

[Structured Data Analysis: Pros and Cons](#_heading=h.co80hwqja854) 10

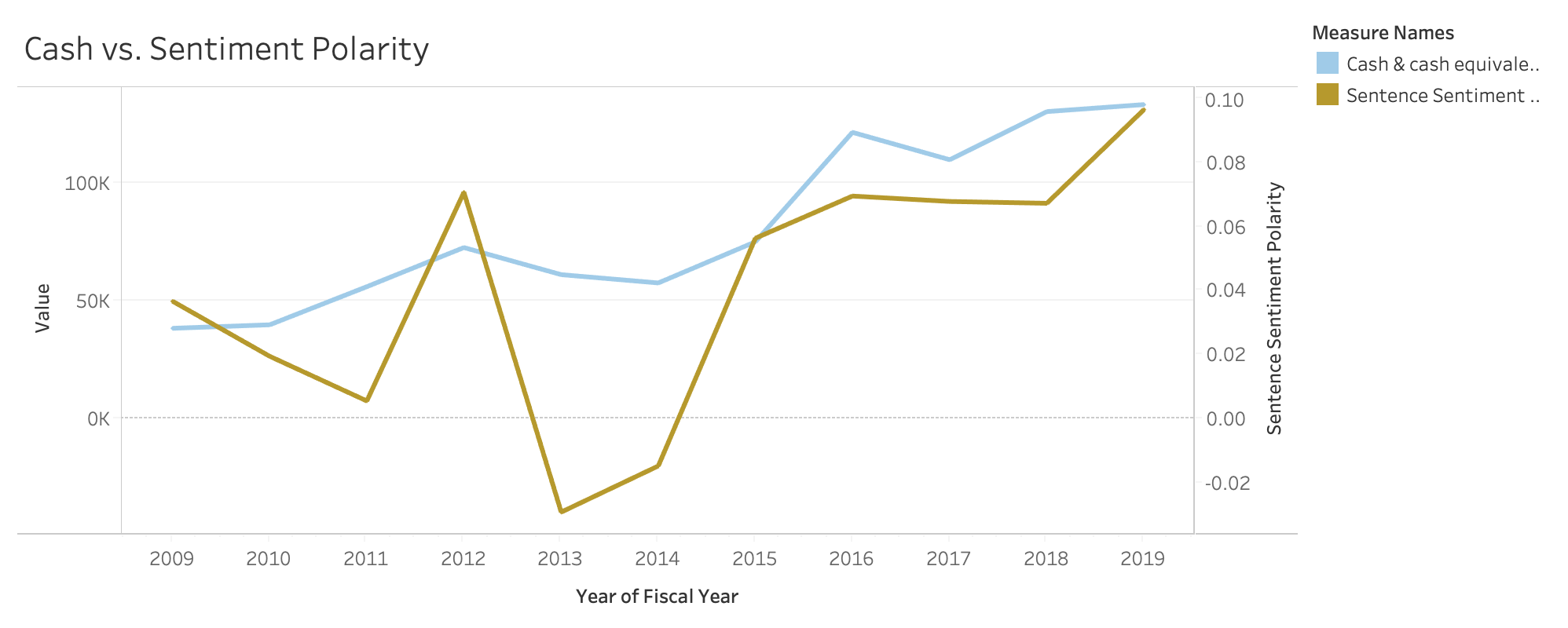
# Introduction

We decided to analyze Goldman Sachs Group. First, we analyzed the text from the Management’s Discussion and Analysis sections of their form 10-Ks from the year 2009 through 2019. Next, we examined the company’s financial performance during this same period. We looked for similarities and differences between these different types of data. Then we analyzed the recent Twitter data that the company has posted since the beginning of 2019.

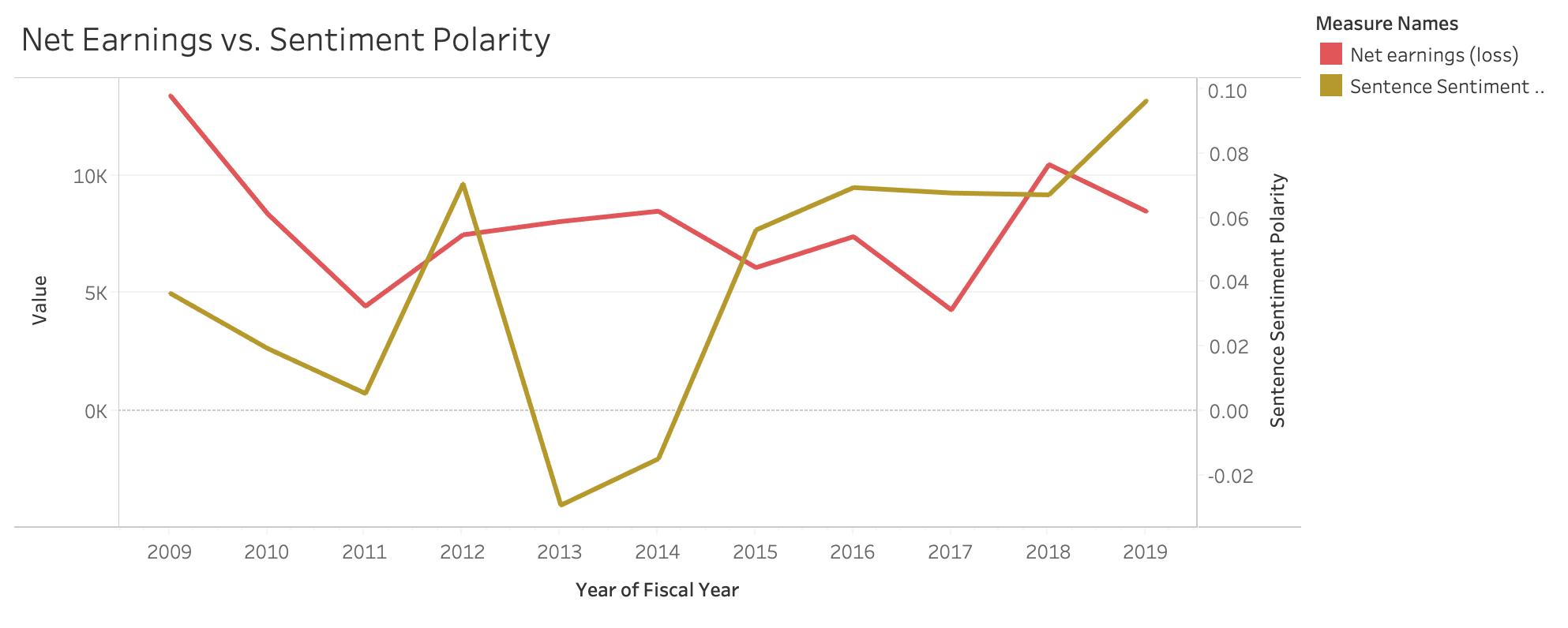
# **Corporate Disclosures Analysis**

## Analysis / Comparison of Financial Statement Metrics with Sentiment

The first graph here shows the trends from 2009–2019 in cash and cash equivalents and in sentiment polarity of the text in the MD&A section of the 10-Ks. In most years, the direction of the change in one factor is the same as the direction of the change in the other factor. They start off differently, with cash and cash equivalents increasing the first two years while sentiment polarity becomes more less positive. However, from 2011 onward, the trend lines tend to move in the same direction as each other. It is noteworthy that in 2013 and 2014 the downward trend is much more significant in the sentiment polarity.



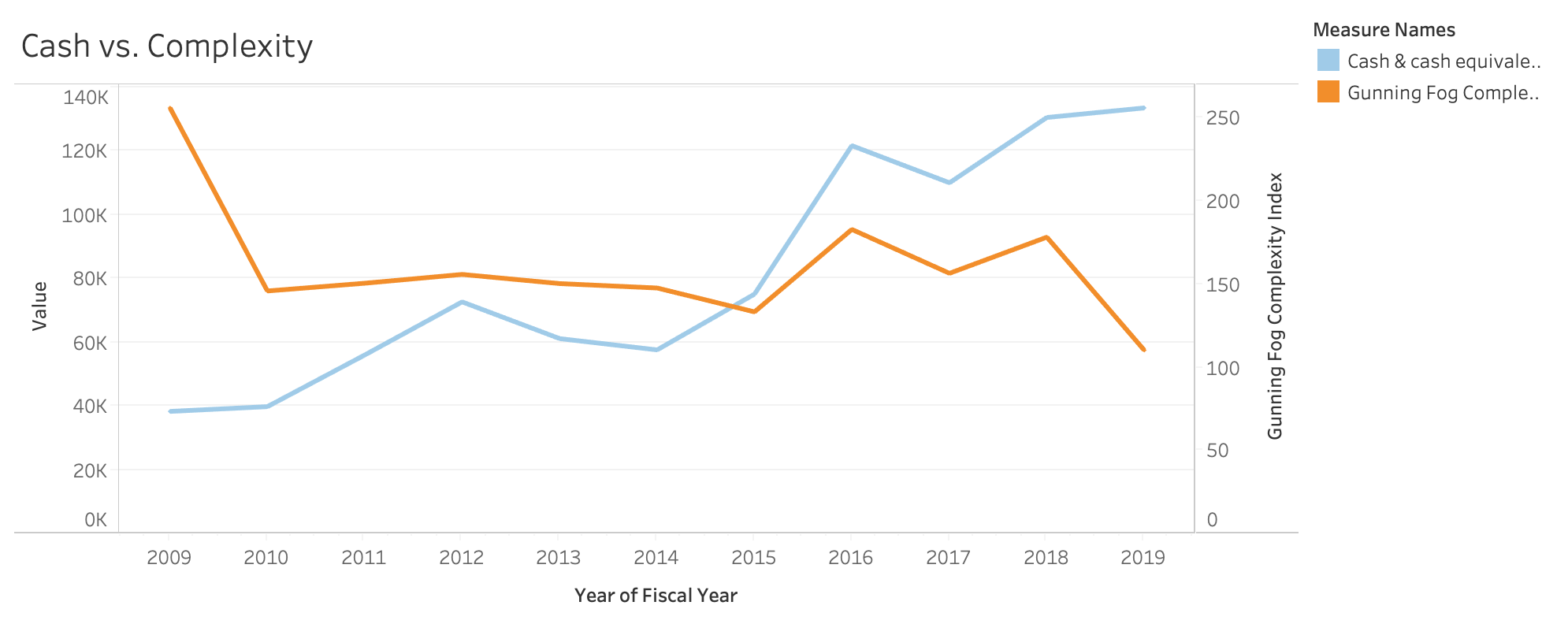
The next graph shows the comparison of the company’s net earnings versus the sentiment polarity over the same years. Goldman Sachs Group always had a positive net earning during these years. The trend in net earnings frequently changed from one year to the next. For example, 2011 shows a change from decreasing to increasing, 2014 shows a change from increasing to decreasing, and 2015, 2016, 2017, and 2018 all show a change in direction of the line graph.



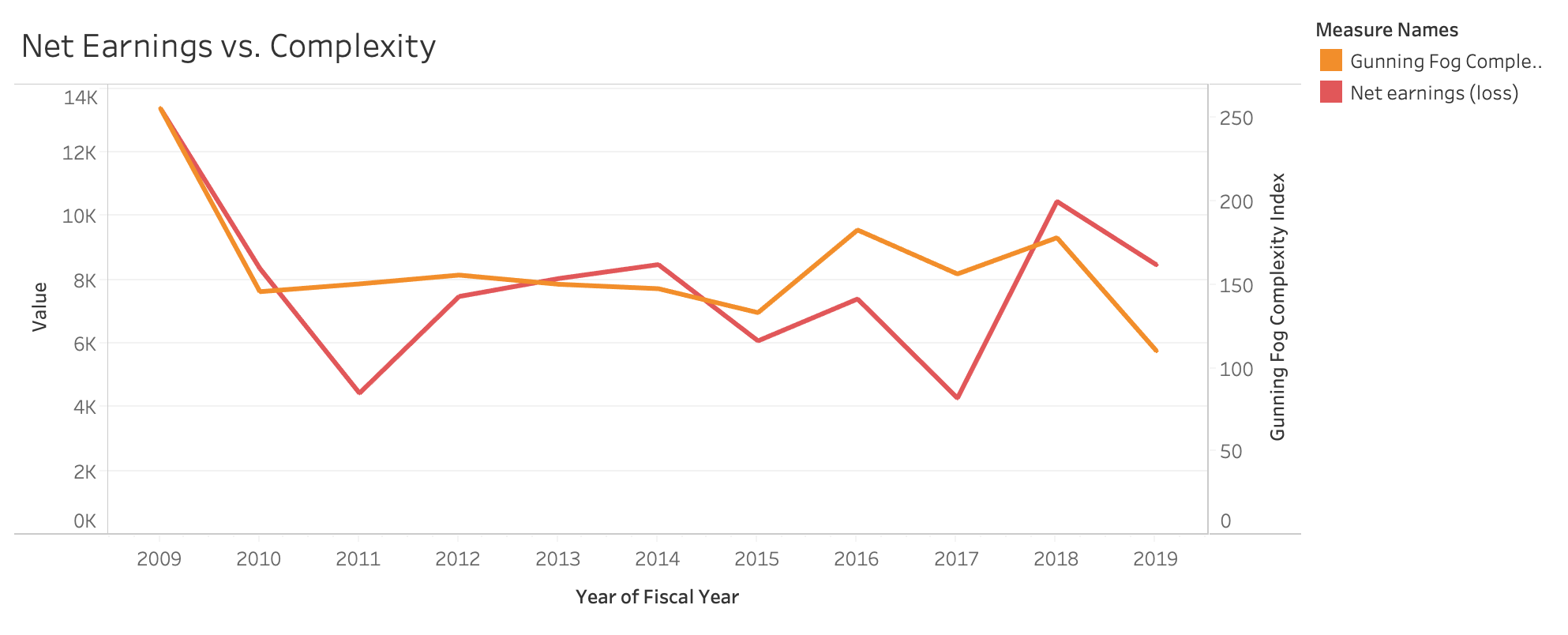
In both cases, the sentiment polarity showed greater volatility than the financial metrics. The cash and cash equivalents metric seems to be slightly more correlated with the changes in sentiment polarity.

## Analysis / Comparison of Financial Statement Metrics with Complexity

The graph below shows a negative correlation between cash and cash equivalents and text complexity. This indicates that Goldman Sachs Group may have tried to distract or confuse readers of its 10-Ks more in years when their cash position was less favorable.



When comparing net earnings and text complexity, we see a more positive correlation. Perhaps this is because the company benefited more when it pursued more complex operations, and then the 10-K required more complex language to explain what they had done during the year.



The volatility of text complexity more closely resembles that of cash and net earnings than sentiment polarity did. Between the two graphs involving complexity, the cash graph is more consistent with our expectations about the relationships between these factors.

## Management’s Discussion and Analysis (MD&A) Sentence Sentiment and Gunning Fog Complexity Comparison

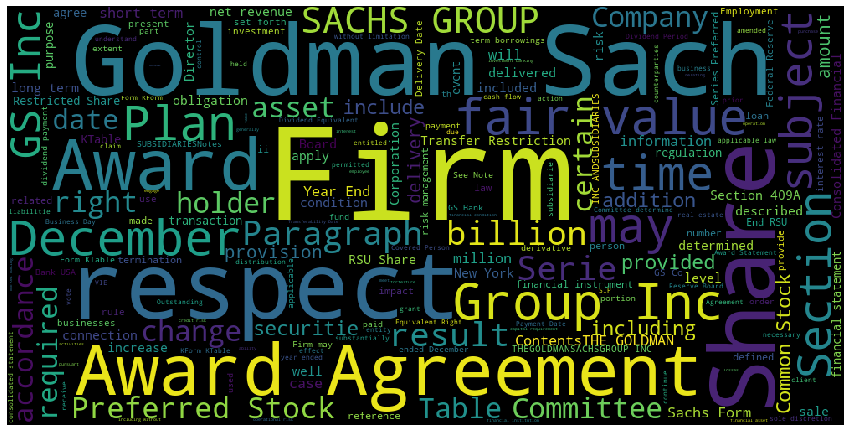
|  |  |  |
| --- | --- | --- |
| Year | Sentiment Score | Gunning \_Fog  Complexity |
| 2009 | 0.036 | 205.94 |
| 2010 | 0.019 | 145.83 |
| 2011 | 0.005 | 150.49 |
| 2012 | 0.071 | 155.79 |
| 2013 | -0.029 | 150.27 |
| 2014 | -0.014 | 147.62 |
| 2015 | 0.056 | 133.26 |
| 2016 | 0.069 | 182.86 |
| 2017 | 0.067 | 156.46 |
| 2018 | 0.067 | 178.21 |
| 2019 | 0.096 | 110.22 |

The “***Balance Sheet Analysis****”* of the MD&A section was mainly used for the text analysis as it contains the performance of the company specific to the year .

Since selecting a paragraph of 100 words for which the gunning fog calculation is more accurate would not reflect the discussion content ,a sample of around 500 words was selected for the calculation.

*It can be observed that Discussions with higher positive sentiment are less Complex, as in 2019 .*

## WordCloud



## Forecast

We can use the data to predict the company’s financial performance in the near future. The cash position has been on a gradual but relatively steady increasing trend over the 11 years we examined. This is a good sign for the company in terms of liquidity and its ability to cover its short-term obligations. The net earnings graph has changed direction frequently. The highest value was at the beginning of the span we examined, in 2009. Optimistically, we can say that Goldman Sachs Group can expect to earn between $5 billion and $10 billion per year with consistency. However, we are concerned that they have not found a way to achieve stable growth in net earnings.

# **Twitter Data Analysis**

## Introduction

Twitter data is the typical unstructured data. Twitter is used by companies as an important platform to communicate with the public. We are asked to use twitter data to observe technology awareness by detecting technology keywords.

Sentiment analysis is the automated process of analyzing text data and sorting it into sentiments positive, negative, or neutral. Using sentiment analysis tools to analyze opinions in Twitter data can help companies understand how people are talking about their brand.

Why sentiment analysis?

* Business: In marketing field companies use it to develop their strategies, to understand customers’ feelings towards products or brand, how people respond to their campaigns or product launches and why consumers don’t buy some
* products.
* Politics: In the political field, it is used to keep track of political views, to detect consistency and inconsistency between statements and actions at the government level. It can be used to predict election results as well!
* Public Actions: Sentiment analysis also is used to monitor and analyse social phenomena, for the spotting of potentially dangerous situations and determining the general mood of the blogosphere.

DATA ANALYSIS METHODS AND RESULTS

For the given task, we have used Python programming language for the data analysis.

Sentiment Analysis is the process of ‘computationally’ determining whether a piece of writing is positive, negative or neutral. It’s also known as opinion mining, deriving the opinion or attitude of a speaker.

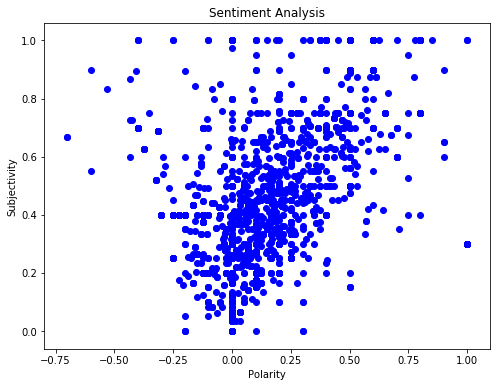
Based on the background knowledge, following are the steps below we used to conduct the analysis:

1. Downloaded the twitter data for the Goldman Sachs company and filtered the data after 2019-01-01.

We first created an API account and collected the required keys and tokens for collecting the data from Twitter. We grabbed all possible tweets from the account and transformed them into a csv and filtered as per the required dates.

We then cleaned the data through a function removing all punciations, number, hyperlinks, spaces, brackets etc. and refined the whole data frame to be used for analysing.

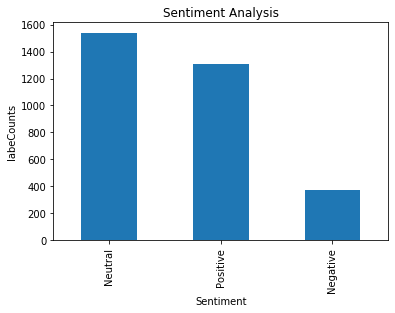
2. Sentiment Analysis

We created functions to read all positive tweets, negative tweets, understand the subjectivity and polarity as follows:

Polarity is a float which lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement. Subjective sentences generally refer to personal opinion, emotion or judgment whereas objective refers to factual information. Subjectivity is also a float which lies in the range of [0,1]. We calculate the sentiment using TextBlob or Vader. Based on the polarity and subjectivity, you determine whether it is a positive text or negative or neutral. For TextBlog, if the polarity is >0, it is considered positive, <0 -is considered negative and ==0 is considered neutral

We then calculated and visualised the counts of tweets for each sentiment: Neutral, Positive and Negative.

In the below figure, we see that most of the tweets are of neutral sentiment and positive.



3. Generated word clouds based on the twitter text data for that company and explained our finding reasonably:

Word clouds or tag clouds are graphical representations of word frequency that give greater prominence to words that appear more frequently in a source text. ... Most word cloud generators have features that allow users to change colors, font, and exclude common or similar words

## Goldman Sachs Twitter Word Cloud for Tweets After 2019-01-01



The most common words used in the tweets for GS are as follows: Disucsse, Research, GS, XBB, XAB, XAE,CEO, David, TalksAtGS, small businesses, explain, WATCH, PODCAST, David Soloman, pandemic, people, etc.

Each color in the word cloud is part of a legend of count of occurrences in the tweets data.The top three in decreasing order are as follows:

* PURPLE: Research, Watch, future, CEO, XBC( Xebec Adsorption Inc) , XAB
* SHADES OF YELLOW: Discusse, TalksAtGS, market, Listen,
* SHADES OF GREEN: GS, Watch, people, CEO, David, Small businesses,explain etc.

Overall, from the word cloud we can say about what the public and the company are talking about.

Goldman Sachs Research features original insights on the economy, markets and industries, drawn from research teams around the world. As on Business Insider, Nov 19 2020 article, Goldman Sachs has been going through some massive changes under CEO David Solomon.

It's taken big steps involving transparency and inclusion to change up its culture. After its first-ever investor day in early 2020, the firm is looking to execute on targets including multi-year cost-cutting plans and it's making big pushes into small businesses like wealth management and consumer banking.

David Solomon said Goldman Sachs “absolutely” wants to get on the blockchain bandwagon.

The bank’s CEO told French newspaper Les Echos in an article published Friday that Goldman is conducting “extensive research on tokenization” and he believes the future of payments systems is bound to live on blockchain. Solomon said it was too early to tell which digital coin or platform will ultimately be the winner. The company believes that big financial players will seek to digitally tokenize their operations to reduce costs and serve clients better, according to a person with knowledge of Goldman’s position.

4. Detected the keywords , generated the keyword frequency (Aggregated and Individual Occurrence)

TECHNOLOGY KEYWORD FREQUENCY FOR GS TWEETS

|  |  |
| --- | --- |
| KEYWORD | FREQUENCY |
| Big Data | 2 |
| Information system | 101 |
| Blockchain | 110 |
| Artificial Intelligence | 15 |
| Machine Learning | 1 |
| OpenSource | 0 |
| Digital Banking | 28 |
| Digital Signature | 0 |
| Neural Network | 361341 |
| Analytics | 18 |

# Combined Observations From Section1 and Section2

#### Positive Sentiment

It can be observed that the ***Sentiment*** of the **MD&A** section of 10K reports and the tweets are mostly ***Positive*** .

#### Research and Technology

From the ***WordCloud***  it can be inferred that the company is focussed on ***Research*** ,***Future*** which also reflects in the Technology Terms Usage like *Blockchain ,Neural Networks*

#### Structured Data Analysis: Pros and Cons

|  |  |
| --- | --- |
| **Pros** | **Cons** |
| Easy to analyze (when compared to unstructured data) as less processing is required | Life doesn't always fit into neat, uniform boxes – the types of things you can record in a structured way are limited |
| Computers can easily read and handle structured data, since this practice has been around for a long time | There is more unstructured data than structured data out there. Ignoring unstructured means missing out on a big part of the picture |