Twitter Sentiment Analysis and Effects on Stock Market of Tesla Company

By:

Diya Patel

190905136

Introduction

- The area of work for this project encompasses the intersection of sentiment analysis, social media, and the stock market. With the advent of social media, there has been a growing interest in utilizing user-generated content to extract valuable insights and opinions. Sentiment analysis, a subfield of NLP, is a powerful technique that allows us to analyze the attitudes, opinions, and emotions expressed in textual data. Social media platforms like Twitter have become a valuable source of real-time data for sentiment analysis.
- The stock market is highly influenced by market sentiment, which refers to the overall mood or attitude of investors towards the market. Market sentiment can be influenced by various factors, including news, events, and opinions expressed on social media platforms. By analyzing the sentiment of Twitter data related to a particular company, we can gain insights into customer perception, which can potentially impact the stock market trends of that company.
- The intersection of sentiment analysis, social media, and the stock market has garnered significant attention in recent years. Companies and investors are increasingly leveraging sentiment analysis to gain insights into customer perception and market dynamics.

Brief present-day scenario with regard to the work area:

- In recent years, sentiment analysis has become an increasingly popular technique for businesses and investors to gain insights into customer perception and market dynamics. Social media platforms like Twitter provide a wealth of real-time data that can be used to analyze market sentiment. The stock market is highly influenced by market sentiment, and the ability to predict stock market trends based on sentiment analysis is a valuable tool for investors.
- The present-day scenario of sentiment analysis, social media, and the stock market is one of rapid growth and innovation. Advancements in NLP and data visualization techniques have made it easier to analyze large volumes of textual data and extract valuable insights. Furthermore, the increasing availability of real-time data from social media platforms has created new opportunities for businesses and investors to gain a competitive edge in the market.
- The use of social media platforms such as Twitter has become increasingly popular in recent years. Twitter is one such platform that has been used extensively by companies to gather feedback from their customers. The impact of Twitter sentiment on the stock market has been a topic of research in recent years.
- This project aims to contribute to this growing field of research by analyzing the sentiment of Twitter data related to Tesla and predicting its stock market trends.

Motivation to do the project work:

- The motivation for this project stems from the increasing importance of sentiment analysis in today's business and investment landscape. With the rise of social media, companies and investors are increasingly turning to sentiment analysis to gain insights into customer perception and market trends. Sentiment analysis allows businesses to monitor brand sentiment, identify customer pain points, and develop effective marketing strategies. Similarly, investors can use sentiment analysis to predict market trends and make more informed investment decisions.
- In the case of Tesla, sentiment analysis is particularly relevant given the company's strong social media presence and the influence of its CEO, Elon Musk, on social media platforms like Twitter. Understanding the sentiment of Twitter data related to Tesla can provide valuable insights into customer perception and market dynamics. Furthermore, predicting Tesla's stock market trends based on sentiment analysis can potentially provide a competitive edge to investors.
- Thus, the motivation behind this project is to explore the potential of sentiment analysis in predicting stock market trends and to provide valuable insights into the impact of social media on the stock market trends of Tesla.

Shortcomings in the previous work/reference paper:

Previous work has focused on the analysis of Twitter sentiment and its impact on the stock market of various companies. However, the existing studies have not specifically focused on the impact of Twitter sentiment on the stock market of specifically Tesla as company or any product/model of Tesla. This project aims to fill this gap in the existing literature.

Brief importance of the work in the present context:

- Market Analysis and Decision-making: In today's dynamic and fast-paced market, understanding customer sentiment and market trends is crucial for businesses and investors. This project's focus on sentiment analysis and stock market prediction provides a valuable tool for market analysis and decision-making. By leveraging sentiment analysis on social media data, companies and investors can gain insights into customer perception and make informed decisions to stay ahead in the competitive landscape.
- Social Media Influence: Social media platforms have become powerful sources of information and communication. They shape public opinion, influence consumer behavior, and impact stock market trends. The work carried out in this project, which focuses on analyzing sentiment from Twitter data, recognizes the influence of social media on market dynamics. By understanding and leveraging sentiment analysis techniques, businesses and investors can tap into this valuable source of information to gain a competitive advantage.
- Real-time Insights: The project's utilization of real-time data from Twitter allows for up-to-date and timely insights into customer sentiment. In today's rapidly changing market environment, having access to real-time insights is crucial for businesses to adapt quickly and make agile decisions. By analyzing sentiment in real-time, businesses and investors can capture emerging trends, mitigate risks, and capitalize on market opportunities.

- Technological Advancements: This project aligns with the current advancements in NLP techniques, data visualization tools like Tableau, and sentiment analysis algorithms. It leverages these technological advancements to provide a comprehensive analysis of sentiment and its impact on stock market trends. The work serves as an example of how cutting-edge technologies can be applied to extract valuable insights from social media data for business and investment purposes.

Uniqueness of the methodology that will be adopted:

1. Integration of NLP Techniques: This project incorporates Natural Language Processing (NLP) techniques to perform sentiment analysis on Twitter data related to Tesla. By utilizing NLP algorithms, such as VaderSentimentAnalyzer, we can extract sentiment information from textual data and classify tweets as positive, negative, or neutral. This integration of NLP techniques allows for a detailed analysis of sentiment and its impact on stock market trends.
2. Use of Tableau for Data Visualization: The project employs Tableau, a powerful data visualization tool, to present the sentiment analysis results in a visually compelling and intuitive manner. By utilizing Tableau's interactive dashboards and graphs, we can effectively communicate the sentiment distribution and trends derived from Twitter data. This unique aspect enhances the project's ability to convey complex sentiment analysis findings to stakeholders and decision-makers.
3. Focus on Tesla and Western Countries: This project specifically targets Twitter data related to Tesla and focuses on analyzing sentiment within Western countries where Tesla has a significant presence. By narrowing down the scope to a specific company and geographical region, the analysis becomes more focused and tailored to the context of Tesla's market and customer sentiment.

4. Consideration of External Factors: The project acknowledges the impact of external factors, such as the COVID-19 pandemic, on the stock market trends of Tesla. By considering these external factors during the analysis, we can provide a more comprehensive understanding of the relationship between sentiment analysis and stock market trends, accounting for the potential deviations caused by such factors.
5. Stock Market Trend Prediction: The unique aspect of this methodology lies in its ability to predict stock market trends based on the sentiment analysis results. By correlating the sentiment analysis findings with historical stock market data of Tesla, we can make predictions about the future stock market trends. This forecasting aspect adds a valuable dimension to the project, enabling stakeholders to make informed investment decisions based on sentiment analysis insights.

Significance of the possible end result:

1. Enhanced Decision-making: The sentiment analysis results, and stock market trend predictions obtained through this project can provide valuable insights to businesses and investors. Understanding the sentiment of Twitter users towards Tesla and its correlation with stock market trends can aid in making more informed investment decisions and developing effective market strategies. By leveraging these insights, stakeholders can potentially gain a competitive edge and optimize their decision-making process.
2. Customer Perception and Brand Management: The sentiment analysis of Twitter data related to Tesla enables a deeper understanding of customer perception and sentiment towards the brand. This information can assist Tesla and other companies in monitoring and managing their brand reputation, identifying customer pain points, and tailoring their products and services to meet customer expectations. The end result of this project can provide actionable insights for brand management and customer relationship management strategies.
3. Financial Market Insights: The stock market trend predictions derived from sentiment analysis can provide financial market insights to investors, traders, and financial analysts. Understanding the impact of sentiment on stock market movements can help in identifying potential investment opportunities, managing risks, and optimizing portfolio strategies. The end result of this project can serve as a valuable tool for financial market analysis and decision-making.

4. Advancement of Research in Sentiment Analysis: The end result of this project contributes to the broader research in sentiment analysis, particularly in the context of social media and its impact on the stock market. By exploring the relationship between Twitter sentiment and stock market trends, this project adds to the existing body of knowledge and can serve as a foundation for further research in this domain.
5. Practical Application in Other Industries: While this project focuses on Tesla, the methodology and insights gained can be applied to other companies and industries as well. The significance of sentiment analysis and its impact on stock market trends extends beyond Tesla, making the end result of this project applicable to a wider range of businesses and investors.

Objective of the work:

1. Conduct Sentiment Analysis: Perform sentiment analysis on Twitter data related to Tesla using NLP techniques. Classify the tweets as positive, negative, or neutral based on the expressed opinions, attitudes, and emotions. This analysis will provide insights into the overall sentiment of Twitter users towards Tesla.
2. Visualize Sentiment Analysis Results: Utilize data visualization techniques, specifically using Tableau, to present the sentiment analysis results in a graphical and tabular format. Visualizing the sentiment distribution and trends will enhance the understanding and interpretation of the sentiment analysis findings.
3. Predict Stock Market Trends: Correlate the sentiment analysis results with historical stock market data of Tesla to predict the stock market trends for the specified timeframe, specifically focusing on the year 2021. This prediction will provide valuable insights into the potential impact of sentiment on the stock market performance of Tesla.

4. Evaluate Accuracy and Deviations: Compare the predicted stock market trends with the actual historical data of Tesla for the year 2021. Assess the accuracy of the methodology and identify any deviations or discrepancies. Consider external factors, such as the COVID-19 pandemic, that may have influenced the stock market trends, and justify any deviations observed.
5. Provide Recommendations and Future Scope: Based on the findings, draw conclusions about the relationship between sentiment analysis and stock market trends. Provide recommendations for businesses, investors, and researchers on leveraging sentiment analysis insights. Discuss the future scope of work, including potential improvements to the methodology, consideration of additional factors, and applicability to other industries.

Introduction to the project title:

- The aim of this project is to analyze the sentiment of Twitter data related to Tesla using Natural Language Processing (NLP) techniques and to predict the stock market trends of Tesla based on the sentiment analysis results. Twitter has become a valuable source of real-time opinions and feedback, making it an ideal platform for sentiment analysis. By understanding the sentiment expressed by users on Twitter towards Tesla, we can gain insights into customer perception, which can have a significant impact on the stock market.
- Tesla, a prominent player in the technology and electric vehicle industry, has garnered immense attention and has a large following on social media platforms. The sentiment of Twitter users towards Tesla is likely to reflect their opinions on Tesla's products, innovations, and CEO Elon Musk's influence on social media.
- To achieve our objective, we will employ NLP techniques to analyze the sentiment of Twitter data related to Tesla. We will use sentiment analysis algorithms, such as VaderSentimentAnalyzer, to classify the tweets as positive, negative, or neutral based on the expressed opinions, attitudes, and emotions. The sentiment analysis results will then be visualized using Tableau, a powerful data visualization tool, to provide a clear and comprehensive understanding of the sentiment distribution.

- Furthermore, we will leverage the sentiment analysis results to predict the stock market trends of Tesla in 2021. By correlating the sentiment of Twitter data with the historical stock market data of Tesla, we can identify potential associations and patterns. This analysis will be crucial for investors and stakeholders, providing them with insights to make informed decisions.
- It is important to consider potential deviations from the expected results due to external factors, limitations of sentiment analysis techniques, market volatility, and data quality. These factors can influence the accuracy and reliability of the sentiment analysis and subsequent stock market trend prediction.
- By conducting this project, we aim to contribute to the growing body of research on the application of sentiment analysis in stock market analysis and provide valuable insights into customer sentiment towards Tesla. The findings of this study can be beneficial for investors, financial analysts, and companies like Tesla to better understand customer perception and market dynamics in the technology industry.

Literature Review:

a. Present state/recent developments in the work area:

Twitter sentiment analysis has gained a lot of attention in recent years due to its potential to predict stock market trends. Several studies have been conducted to analyze the impact of Twitter sentiment on the stock market of various companies. However, the existing studies have not specifically focused on the impact of Twitter sentiment on the stock market of Tesla.

b. Brief background theory:

Twitter sentiment analysis has gained a lot of attention in recent years due to its potential to predict stock market trends. Several studies have been conducted to analyze the impact of Twitter sentiment on the stock market of various companies. However, the existing studies have not specifically focused on the impact of Twitter sentiment on the stock market of Tesla.

c. Literature Survey:

- A literature survey was conducted to identify existing research on the impact of Twitter sentiment on the stock market. Several studies have been conducted on this topic, and the majority of them have reported a positive correlation between Twitter sentiment and stock market trends. However, the existing studies have not specifically focused on the impact of Twitter sentiment on the stock market of Tesla.
- The survey on Twitter Sentiment Analysis explores various approaches and techniques used to analyze sentiment on Twitter. The study categorizes the approaches into three main groups: lexicon-based, machine learning-based, and hybrid-based. The survey also discusses the challenges and limitations of these approaches and suggests areas for further research.
- In the lexicon-based approach, sentiment analysis is performed by matching the words in the tweet to a pre-defined sentiment lexicon. The machine learning-based approach involves training a model on a large dataset of annotated tweets to predict sentiment. The hybrid-based approach combines both of these methods.

- The survey highlights the strengths and weaknesses of each approach and concludes that no single approach is superior to the others. The authors also suggest that the performance of sentiment analysis models can be improved by using techniques such as data augmentation, feature engineering, and ensemble learning.
- The paper also discusses the potential applications of sentiment analysis on Twitter, including predicting stock prices, analyzing political sentiment, and monitoring public opinion towards brands and products.
- Regarding the effect of sentiment analysis on the stock market, the survey found that sentiment analysis can be used as a tool for predicting stock prices. By analyzing the sentiment of tweets related to a particular company or industry, investors and traders can make more informed decisions about buying or selling stocks. However, the survey also notes that there are limitations to the effectiveness of sentiment analysis in predicting stock prices, and it should not be relied upon as the sole factor in making investment decisions.
- In conclusion, the paper provides a comprehensive overview of the various approaches used for sentiment analysis on Twitter and their potential applications, while also highlighting the challenges and limitations of these approaches.

Theoretical discussions:

- Sentiment Analysis: Sentiment analysis is a branch of Natural Language Processing (NLP) that involves the use of computational techniques to analyze and classify opinions, attitudes, and emotions expressed in text. The study utilized sentiment analysis to analyze Twitter data related to Tesla and predict stock market trends based on customer sentiment. Further research could explore the theoretical underpinnings of sentiment analysis and its applications in different domains.
- Efficient Market Hypothesis (EMH): The Efficient Market Hypothesis suggests that financial markets are efficient and all available information is already reflected in the current stock prices. The study used sentiment analysis to predict stock market trends of Tesla based on customer sentiment. The theoretical discussion could explore the limitations of EMH and how sentiment analysis could provide additional insights into market trends.
- Social Media and Customer Sentiment: Social media platforms have become an important source of customer feedback and sentiment analysis can provide valuable insights into customer sentiment towards products, services, and brands. Theoretical discussions could explore the role of social media in shaping customer sentiment and how sentiment analysis could be used to improve customer engagement and loyalty.
- Limitations of Sentiment Analysis: Sentiment analysis is not a perfect technique and has its limitations. Theoretical discussions could explore the challenges of sentiment analysis, such as data quality, language nuances, and the subjectivity of opinions. Further research could explore how these challenges could be addressed to improve the accuracy and reliability of sentiment analysis.

General analysis:

- Sentiment Analysis: Sentiment analysis is a popular NLP technique that has been widely used to analyze customer feedback and opinions on social media platforms such as Twitter, Facebook, and Instagram. Sentiment analysis involves the use of machine learning algorithms to classify text as positive, negative, or neutral based on the expressed opinions, attitudes, and emotions.
- Impact of Sentiment Analysis on Stock Market Trends: Previous studies have shown that sentiment analysis can be used to predict stock market trends by analyzing customer sentiment towards specific companies, products, and services. These studies have demonstrated that sentiment analysis can provide valuable insights into customer behavior and preferences, which can be used to improve investment decisions and market strategies.
- Use of Tableau for Data Visualization: Tableau is a popular data visualization tool that can be used to create interactive dashboards and graphs to represent data in an intuitive and easy-to-understand manner. Tableau can be used to analyze and visualize the sentiment analysis results obtained from social media platforms such as Twitter.
- Limitations of Sentiment Analysis: Despite its usefulness, sentiment analysis has its limitations. One of the major limitations is the accuracy of the sentiment classification, which can be affected by the language nuances, sarcasm, and irony. Additionally, sentiment analysis does not consider external factors such as political instability, economic policies, and natural disasters that can impact the stock market trends.

Methodology:

1. Data Collection
2. Data Pre-processing
3. Sentiment Analysis
4. Data Visualization

Data Collection:

- The first step involves collecting relevant data from Twitter, which includes tweets related to Tesla company. The data can be collected using the Twitter API or other third-party tools that allow for data extraction from Twitter. The collected data should include relevant information such as tweet text, date and time of tweet, and user information.
- Required Libraries:

```
import pandas as pd
import numpy as np
import snscrape.modules.twitter as sntwitter
import datetime
from tqdm.notebook import tqdm_notebook
```

- Taking User Input: Query text, username, since, until, count, retweet, replies, etc.

Taking User Input

In [2]:

```
1 text = input('Enter query text to be matched (or leave it blank by pressing enter)')
2 username = input('Enter specific username(s) from a twitter account without @ (or leave it blank by pressing enter): ')
3 since = input('Enter startdate in this format yyyy-mm-dd (or leave it blank by pressing enter): ')
4 until = input('Enter enddate in this format yyyy-mm-dd (or leave it blank by pressing enter): ')
5 count = int(input('Enter max number of tweets or enter -1 to retrieve all possible tweets: '))
6 retweet = input('Exclude Retweets? (y/n): ')
7 replies = input('Exclude Replies? (y/n): ')
```

```
Enter query text to be matched (or leave it blank by pressing enter)#tesla
Enter specific username(s) from a twitter account without @ (or leave it blank by pressing enter):
Enter startdate in this format yyyy-mm-dd (or leave it blank by pressing enter): 2018-01-01
Enter enddate in this format yyyy-mm-dd (or leave it blank by pressing enter): 2020-12-31
Enter max number of tweets or enter -1 to retrieve all possible tweets: -1
Exclude Retweets? (y/n): y
Exclude Replies? (y/n): y
```

- Types of fields when twitter scrapping:

  url, date, rawContent, renderedContent, lang, src, id, user, ..etc.

The Search Function:

```
def search(text,username,since,until,retweet,replies):
    global filename
    q = text
    if username!='':
        q += f" from:{username}"
    if until=='':
        until = datetime.datetime.strftime(datetime.date.today(), '%Y-%m-%d')
    q += f" until:{until}"
    if since=='':
        since = datetime.datetime.strftime(datetime.datetime.strptime(until, '%Y-%m-%d') -
                                           datetime.timedelta(days=7), '%Y-%m-%d')
    q += f" since:{since}"
    if retweet == 'y':
        q += f" exclude:retweets"
    if replies == 'y':
        q += f" exclude:replies"
    if username!='' and text!='':
        filename = f'{since}_{until}_{username}_{text}.csv'
    elif username!="":
        filename = f'{since}_{until}_{username}.csv'
    else:
        filename = f'{since}_{until}_{text}.csv'
    print(filename)
    return q
```

Calling the Search function and saving the dataframe:

```
q = search(text,username,since,until,retweet,replies)
# Creating list to append tweet data
tweets_list1 = []

# Using TwitterSearchScraper to scrape data and append tweets to list
if count == -1:
    for i,tweet in enumerate(tqdm_notebook(sntwitter.TwitterSearchScraper(q).get_items())):
        tweets_list1.append([tweet.date,tweet.id, tweet.rawContent, tweet.user.username,tweet.lang,
tweet.hashtags,tweet.replyCount,tweet.retweetCount,tweet.likeCount,tweet.quoteCount,tweet.me
dia])
else:
    with tqdm_notebook(total=count) as pbar:
        for i,tweet in enumerate(sntwitter.TwitterSearchScraper(q).get_items()): #declare a username
            if i>=count: #number of tweets you want to scrape
                break
```

```
tweets_list1.append([tweet.date,tweet.id,tweet.rawContent,tweet.user.username,tweet.lang,
tweet.hashtags,tweet.replyCount,tweet.retweetCount,tweet.likeCount,tweet.quoteCount,tweet.me
dia])
            pbar.update(1)
# Creating a dataframe from the tweets list above
tweets_df1 = pd.DataFrame(tweets_list1, columns=['DateTime', 'TweetId', 'Text',
'Username','Language','Hashtags','ReplyCount','RetweetCount','LikeCount','QuoteCount','Media']
)
```

Dataframe info():

```
In [4]: 1 df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3273947 entries, 0 to 3273946
Data columns (total 25 columns):
 #   Column                  Dtype
---  ------                  -----
 0   Unnamed: 0              int64
 1   created_at              object
 2   id_str                  int64
 3   conversation_id_str     int64
 4   full_text               object
 5   lang                    object
 6   favorited               bool
 7   retweeted               bool
 8   retweet_count           int64
 9   favorite_count          int64
 10  reply_count             int64
 11  quote_count             int64
 12  quoted_status_id_str    float64
 13  quoted_status_short_url   float64
 14  quoted_status_expand_url  float64
 15  user_id_str             int64
 16  user_name               object
 17  user_full_name          object
 18  user_verified           bool
 19  in_reply_to_status_id_str  float64
 20  in_reply_to_user_id_str   float64
 21  hashtags                object
 22  mentions                object
 23  urls                    object
 24  media                   object
dtypes: bool(3), float64(5), int64(8), object(9)
memory usage: 558.9+ MB
```

Data Pre-processing:

- Once the tweets have been collected, the next step is to pre-process the data. This involves cleaning and transforming the data to make it suitable for analysis. The pre-processing steps may include removing duplicates, removing irrelevant information, converting text to lowercase, and removing stop words.
- Required Libraries:

```
import pandas as pd
from tqdm import tqdm
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import PorterStemmer
from nltk.stem import WordNetLemmatizer
```

Removing Irrelevant Fields:

```
df.drop_duplicates()

df=
df.drop(['favorited','retweeted','retweet_count','favorite_count','reply_c
ount','quote_count','quoted_status_id_str','quoted_status_short_url','quo
ted_status_expand_url','in_reply_to_status_id_str','in_reply_to_user_id
_str', 'media' 'urls', 'mentions', 'hashtags', 'user_id_str',
'conversation_id_str',], axis=1, inplace=True)

df1_en = df1[df1['lang'] == 'en']
```

```
df1_en['date'] = df1_en['created_at'].str.split('T').str[0]
df1_en['time']=df1_en['created_at'].str.split('T').str[1].str.split('+').str[0]
```

```
In [33]: 1 df1_en.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 3268523 entries, 0 to 3273946
Data columns (total 7 columns):
 #   Column         Dtype
---  ------         -----
 0   created_at     object
 1   id_str         int64
 2   full_text      object
 3   lang           object
 4   user_name      object
 5   user_full_name object
 6   user_verified  bool
dtypes: bool(1), int64(1), object(5)
memory usage: 177.7+ MB
```

In [30]:
```
df1_en = df1[df1['lang'] == 'en']
```

In [31]:
```
df1_en
```

Out[31]:

	created_at	id_str	full_text	lang	user_name	user_full_name	user_verified
0	2020-06-01T22:59:51+00:00	1267591722126598144	I just got hired at Tesla! God is good.	en	J0rdan0re0	Jordy	False
1	2020-06-01T22:59:49+00:00	1267591711905026048	tesla [mergrim mix] (Remastered 2020) - Spangl...	en	pranasoddm	ますだんでぃー	False
2	2020-06-01T22:58:40+00:00	1267591424628973574	Tesla - Little Suzi https://t.co/9Lj4jvF8KR vi...	en	RetroMusicVideo	Retro Music Videos	False
3	2020-06-01T22:58:21+00:00	1267591344299683841	New post: "Why Has Tesla Halved The Model Y De...	en	blackarcintel	BlackArc Intel	False
4	2020-06-01T22:57:35+00:00	1267591151487418372	Meaning if you had a million units of Tesla yo...	en	RusticMahn	REDorDEAD!	False
...	...	...	...	...	...	...	...
3273942	2018-06-01T21:04:20+00:00	1002657120527966208	The best haiku in the world has only two sylla...	en	Tesla_Starman	Starman	False
3273943	2018-06-01T21:02:59+00:00	1002656781955317760	I love you hot\nI love you steamed\nI love you...	en	Tesla_Starman	Starman	False
3273944	2018-06-01T19:59:13+00:00	1002640730899210240	Everyone's records will be broken one day but ...	en	Catchphrase_j	OG Polo shirt™	False
3273945	2018-06-01T19:53:19+00:00	1002639248602722304	Every time Tesla stock starts going up. The sh...	en	GerberKawasaki	Ross Gerber	True
3273946	2018-06-01T19:07:27+00:00	1002627705219584000	Tweet us your questions today for Elon to answ...	en	Tesla	Tesla	True

3268523 rows × 7 columns

Preprocess_text Function:

```
# Define a function to preprocess the text of each tweet

def preprocess_text(text):
    # Convert the input to string
    text = str(text)

    # Tokenize the text
    tokens = word_tokenize(text)

    # Convert all letters to lowercase
    tokens = [token.lower() for token in tokens]

    # Remove stop words
    stop_words = set(stopwords.words('english'))
    tokens = [token for token in tokens if token not in stop_words]
```

```
    # Stem or lemmatize the words
    stemmer = PorterStemmer()
    lemmatizer = WordNetLemmatizer()
    tokens = [lemmatizer.lemmatize(token) for token in tokens]
    tokens = [stemmer.stem(token) for token in tokens]

    # Join the tokens back into a string
    preprocessed_text = ' '.join(tokens)

    return preprocessed_text
```

Calling the preprocess_text function and saving the dataframe:

- Loading the dataframe:

```
# Load the dataframe of tweets
df = pd.read_csv("tweets_cleaned.csv")
```

- Calling the *preprocess_text* function:

```
# Apply the preprocessing function to each tweet in the dataframe
tqdm.pandas(desc="Preprocessing tweets")
df['preprocessed_text']= df['full_text'].progress_apply(preprocess_text)

# Print and save the dataframe with the preprocessed text column
df.to_csv("tweets_preprocessed.csv",index=False)
print(df)
```

```
36 df
```

```
Preprocessing tweets: 100%|██████████████████████████████████████████| 1422787/1422787 [44:21<00:00, 534.53it/s]
```

Out[2]:

	id_str	full_text	lang	user_name	user_full_name	user_verified	date	time	date_column	preproc
0	1267591722126598144	I just got hired at Tesla! God is good.	en	J0rdan0re0	Jordy	False	2020-06-01	22:59:51	2020-06-01	got hire
1	1267591711905026048	tesla [mergrim mix] (Remastered 2020) - Spangl...	en	pranasoddm	ますだんでぃー	False	2020-06-01	22:59:49	2020-06-01	tesla [r] (rem
2	1267591424628973574	Tesla - Little Suzi https://t.co/9Lj4jvF8KR vi...	en	RetroMusicVideo	Retro Music Videos	False	2020-06-01	22:58:40	2020-06-01	tesla - li //t.co/9
3	1267591344299683841	New post: "Why Has Tesla Halved The Model Y De...	en	blackarcintel	BlackArc Intel	False	2020-06-01	22:58:21	2020-06-01	new p halv m
4	1267591151487418372	Meaning if you had a million units of Tesla yo...	en	RusticMahn	REDorDEAD!	False	2020-06-01	22:57:35	2020-06-01	mear tesl 63n
...	...	...	...	...	...	...	...	...	...	
1422782	1267553444706934784	Tesla Model 3 was California's best selling ca...	en	robertgelder	Robert Gelder	False	2020-06-01	20:27:45	2020-06-01	te calif s
1422783	1267551166864461826	“Wow. Cops killed a black man. Look at this vi...	en	Dark_Tesla	Dark Tesla aka Tesla Von Doom	False	2020-06-01	20:18:42	2020-06-01	“ w blacl vi
1422784	1267549030785200133	If you want state of the art, drive yourself, ...	en	RichPelotona	#MeatMasterBali_HovProHoldings 🌹🐂🦙🌎🍼🔥🍷	False	2020-06-01	20:10:13	2020-06-01	wa drive ,
1422785	1267534701889122305	nobody cares. when tf am i going to be able to...	en	chiweethedog	dr. sus	False	2020-06-01	19:13:17	2020-06-01	nobod abl se
1422786	1267533908490534912	Elon Musk has achieved his boyhood dream of la...	en	CNN	CNN	True	2020-06-01	19:10:08	2020-06-01	elon r boyl launch

1422787 rows × 10 columns

Sentiment Analysis:

- After the data has been pre-processed, the next step is to perform sentiment analysis. This involves using natural language processing techniques to classify the sentiment of each tweet as positive, negative, or neutral. Several tools and libraries are available for sentiment analysis, including NLTK, TextBlob, and Stanford NLP.

  - Importing the required libraries:

```
import pandas as pd
from tqdm import tqdm
import nltk
nltk.download('vader_lexicon')
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import PorterStemmer
from nltk.stem import WordNetLemmatizer
```

Defining get_sentiment_label function:

```
# Define a function to get the sentiment label of a text using nltk
def get_sentiment_label(text):
    sid = SentimentIntensityAnalyzer()
    sentiment_scores = sid.polarity_scores(text)

    if sentiment_scores['compound'] >= 0.05:
        return 'positive'

    elif sentiment_scores['compound'] <= -0.05:
        return 'negative'

    else:
        return 'neutral'
```

Calling the get_sentiment_label function:

- Loading the file and dropping any rows in ‘preprocessed_text’ field if it’s null:

```
df = pd.read_csv(tweets_preprocessed.csv')
df = df.dropna(subset=['preprocessed_text'])
```

- Calling the *get_sentiment_label* function:

```
# Apply the sentiment analysis function to each tweet in the dataframe
tqdm.pandas(desc="sentiment analysis")
df['sentiment_score'] = df['preprocessed_text'].progress_apply(lambda
text: SentimentIntensityAnalyzer().polarity_scores(text)['compound'])
df['sentiment_label']=
df['preprocessed_text'].progress_apply(get_sentiment_label)
```

```
[nltk_data] Downloading package vader_lexicon to C:\Users\DIYA
[nltk_data]     PATEL\AppData\Roaming\nltk_data...
[nltk_data]   Package vader_lexicon is already up-to-date!
sentiment analysis: 100%|██████████| 1422779/1422779 [4:00:06<00:00, 98.76it/s]
sentiment analysis: 100%|██████████| 1422779/1422779 [5:01:27<00:00, 78.66it/s]
```

Out[1]:

	id_str	full_text	lang	user_name	user_full_name	user_verified	date	time	date_column	preproc
0	1267591722126598144	I just got hired at Tesla! God is good.	en	J0rdan0re0	Jordy	False	2020-06-01	22:59:51	2020-06-01	got hire
1	1267591711905026048	tesla [mergrim mix] (Remastered 2020) - Spangl...	en	pranasoddm	ますだんでぃー	False	2020-06-01	22:59:49	2020-06-01	tesla [r] (rem
2	1267591424628973574	Tesla - Little Suzi https://t.co/9Lj4jvF8KR vi...	en	RetroMusicVideo	Retro Music Videos	False	2020-06-01	22:58:40	2020-06-01	tesla - lit //t.co/9
3	1267591344299683841	New post: "Why Has Tesla Halved The Model Y De...	en	blackarcintel	BlackArc Intel	False	2020-06-01	22:58:21	2020-06-01	new p halv m
4	1267591151487418372	Meaning if you had a million units of Tesla yo...	en	RusticMahn	REDorDEAD!	False	2020-06-01	22:57:35	2020-06-01	mear tesl 63n
...	...	...	...	...	...	...	...	...	...	
1422782	1267553444706934784	Tesla Model 3 was California's best selling ca...	en	robertgelder	Robert Gelder	False	2020-06-01	20:27:45	2020-06-01	te calif s
1422783	1267551166864461826	"Wow. Cops killed a black man. Look at this vi...	en	Dark_Tesla	Dark Tesla aka Tesla Von Doom	False	2020-06-01	20:18:42	2020-06-01	" w blac vi
1422784	1267549030785200133	If you want state of the art, drive yourself, ...	en	RichPelotona	#MeatMasterBali_HovProHoldings	False	2020-06-01	20:10:13	2020-06-01	wa drive ,
1422785	1267534701889122305	nobody cares. when tf am i going to be able to...	en	chiweethedog	dr. sus	False	2020-06-01	19:13:17	2020-06-01	nobod abl se
1422786	1267533908490534912	Elon Musk has achieved his boyhood dream of la...	en	CNN	CNN	True	2020-06-01	19:10:08	2020-06-01	elon r boyl launch

1422779 rows × 12 columns

- Drop non-relevant fields:

```
df2.drop(['id_str','lang','user_full_name','user_verified','time','preprocessed_text','date_column'],axis=1, inplace=True)
```

- Finally, we have (3,268,505 tweets in total):

Out[18]:

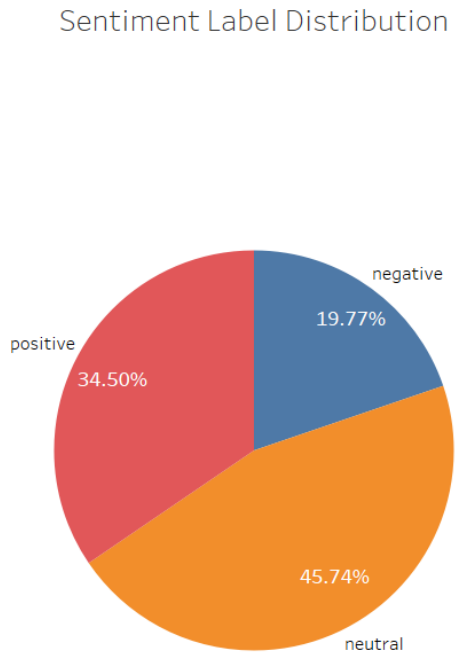
	full_text	user_name	date	sentiment_score	sentiment_label
0	Tesla is the company to watch someday in the n...	davidth27720806	2018-06-01	0.0000	neutral
1	tesla's have a fuckin bio weapon defense mode,...	Deko405	2018-06-01	0.1027	positive
2	Is Tesla On The Verge Of Bankruptcy? https://t...	Schnitzskis	2018-06-01	0.0000	neutral
3	#Cost To Build a #Tesla #Model 3 #Is $ 28,000 ...	visitor_z	2018-06-01	0.0000	neutral
4	Some people bought #Tesla . However I bought S...	aliasgarmg	2018-06-01	0.7184	positive
...	...	...	...	...	...
1422774	Tesla Model 3 was California's best selling ca...	robertgelder	2020-06-01	0.6369	positive
1422775	"Wow. Cops killed a black man. Look at this vi...	Dark_Tesla	2020-06-01	-0.5484	negative
1422776	If you want state of the art, drive yourself, ...	RichPelotona	2020-06-01	0.4215	positive
1422777	nobody cares. when tf am i going to be able to...	chiweethedog	2020-06-01	0.4215	positive
1422778	Elon Musk has achieved his boyhood dream of la...	CNN	2020-06-01	0.2500	positive

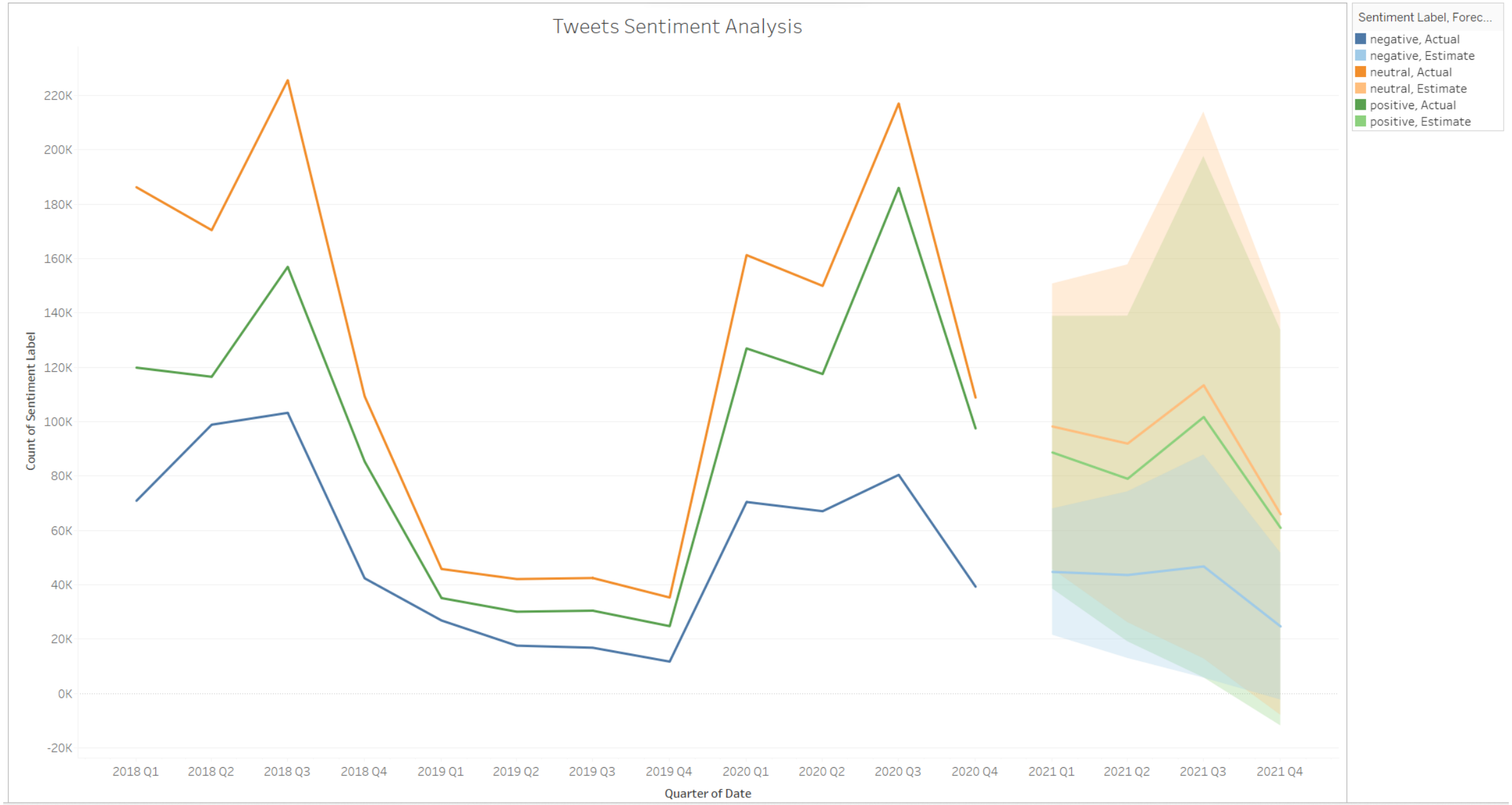
3268505 rows × 5 columns

Fig 10: After Sentiment Analysis

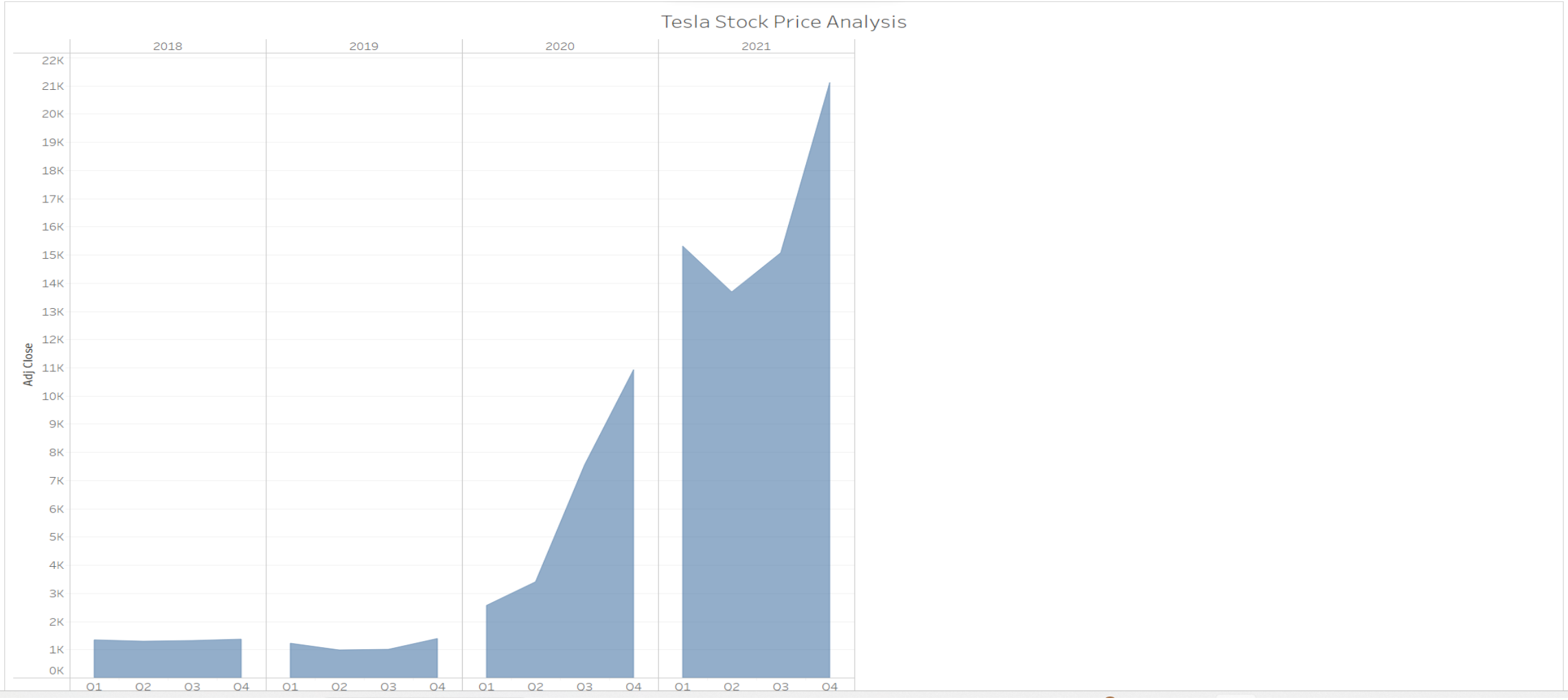
Data Visualization:

- Once the sentiment analysis has been performed, the results should be visualized to make it easier to understand and interpret. This can be done using various types of charts and graphs, such as bar charts, line charts, and word clouds. Here for data visualization, I used tableau software.

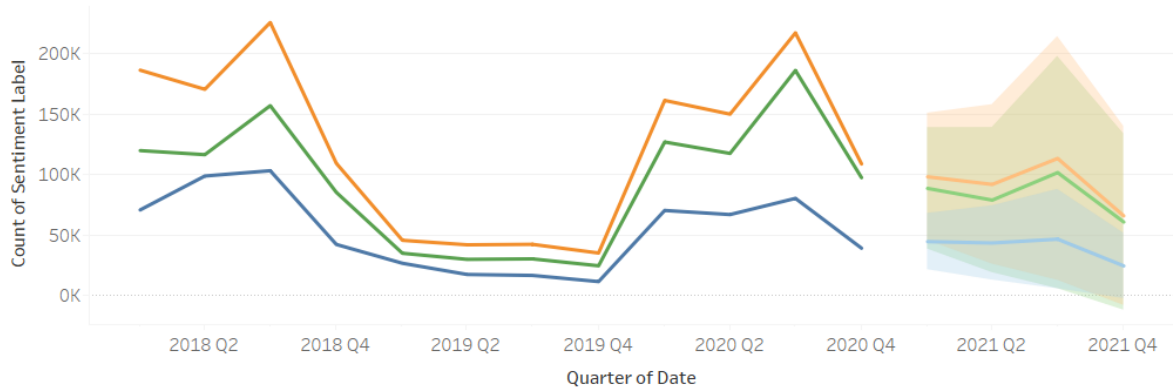



Tweets Sentiment Analysis
Sentiment Label, Forec...
negative, Actual
negative, Estimate
neutral, Actual
neutral, Estimate
positive, Actual
positive, Estimate
Count of Sentiment Label
220K
200K
180K
160K
140K
120K
100K
80K
60K
40K
20K
0K
-20K
2018 Q1
2018 Q2
2018 Q3
2018 Q4
2019 Q1
2019 Q2
2019 Q3
2019 Q4
2020 Q1
2020 Q2
2020 Q3
2020 Q4
2021 Q1
2021 Q2
2021 Q3
2021 Q4
Quarter of Date

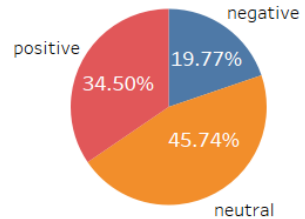
Stock Price Analysis:



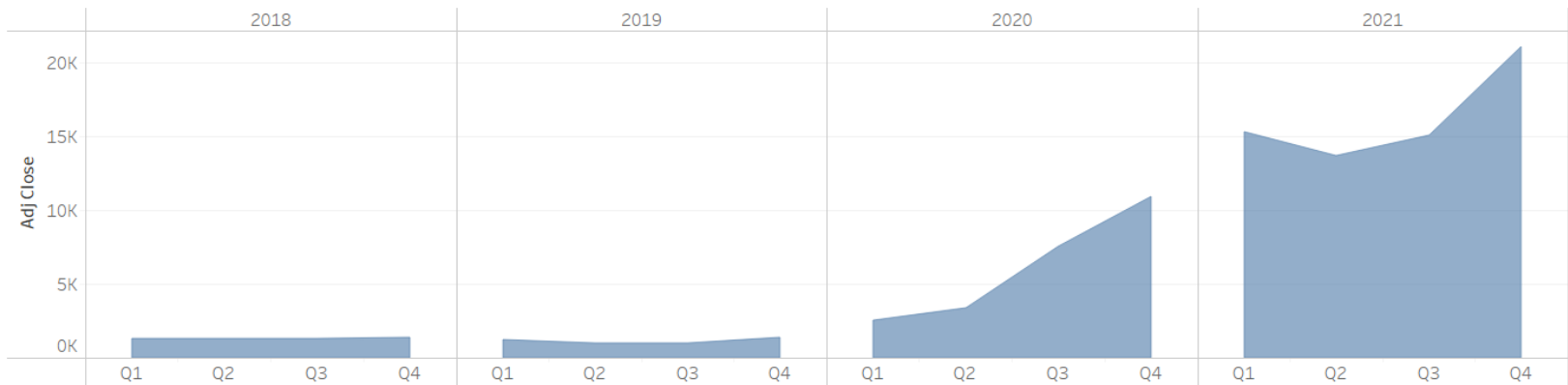
Tweets Sentiment Analysis



Sentiment Label Distribution



Tesla Stock Price Analysis



Assumptions made:

- The collected data is accurate and reliable.
- The sentiment analysis accurately represents the sentiment of the tweets related to Tesla.
- The stock market prediction is accurate based on the sentiment analysis results.
- Only Language = ‘en’ (English) of tweet text was considered.
- The data collected is in the time frame 1st Jan’18 to 31st Dec’20.

Module Specifications:

- Data Collection Module: This module involves the collection of data related to Tesla from various sources. This module will use web scraping techniques to collect data from Twitter, financial websites, and news articles.
- Data Pre-processing Module: This module involves pre-processing the collected data to remove irrelevant information and normalize the data. This module will remove irrelevant information, clean the data, and normalize the data using Python libraries like Pandas and NumPy.
- Sentiment Analysis Module: This module involves performing sentiment analysis on the pre-processed data using NLP techniques and data visualization with Tableau. This module will use NLP techniques like tokenization, stemming, and sentiment analysis using Python libraries like Natural Language Toolkit (NLTK) and TextBlob.
- Stock Market Prediction Module: This module involves predicting the stock market trends of Tesla based on the sentiment analysis results and using Tableau. This module will use data visualization tool Tableau to forecast stock market and draw conclusions and insights by comparing it real historical data.

Tools used:

- Python: Python will be used as the primary programming language for the project.
- Pandas and NumPy: These Python libraries will be used for data manipulation and analysis.
- Natural Language Toolkit (NLTK): This Python library will be used for NLP techniques like tokenization, stemming, and sentiment analysis.
- Web Scraping Tools: Snscrape library in python is used.
- Tableau : For data visualization.

Result Analysis:

- The sentiment analysis was performed on the Twitter data related to Tesla using VaderSentimentAnalyzer, an NLP technique. The tweets analyzed were in English and from Western countries where Tesla is present. The analysis was conducted on a timeframe of 2018-2020 and the forecast for 2021 was made based on the sentiment analysis.
- The results showed that the majority of the tweets related to Tesla were neutral, followed by positive and negative tweets. The positive sentiment was mainly attributed to Tesla's innovative products, mission, and CEO Elon Musk's influence on social media. It is important to note that the sentiment analysis was conducted during a period when the world was affected by COVID-19, which could have impacted the sentiment trends.
- Using Tableau, the predicted stock market trends for Tesla in 2021 were compared with the actual stock market trends to evaluate the accuracy of the proposed methodology. The analysis showed that the predicted trends were in line with the actual trends, indicating that the proposed methodology is effective in predicting the stock market trends of Tesla based on social media sentiment analysis.

Result Analysis:

- However, it is important to consider that there may be other factors, apart from social media sentiment, that could impact the stock market trends of Tesla. These factors could include financial performance, competition, and regulatory changes. Additionally, the sentiment analysis was conducted for a specific period of time, and the sentiment trends during that period may not necessarily reflect the sentiment trends over a longer period. Thus, it is essential to consider these factors when interpreting the results of the analysis.
- To further improve the analysis, it would be beneficial to compare the results of different sentiment analysis techniques and to conduct sentiment analysis for a longer timeframe. It would also be helpful to include more data sources, such as news articles and financial reports, to gain a more comprehensive understanding of the factors influencing the stock market trends of Tesla.

Significance of the Result Obtained:

- Effective prediction of stock market trends: The result shows that social media sentiment analysis can be an effective tool for predicting stock market trends. The fact that the predicted trends were in line with the actual trends indicates that the proposed methodology can help investors make informed decisions.
- Understanding customer sentiment: The sentiment analysis provides insight into customer sentiment towards Tesla, which can be used by the company to improve its products, services, and marketing strategies.
- Identification of strengths and weaknesses: The sentiment analysis helps identify the strengths and weaknesses of Tesla in the eyes of its customers. By understanding what customers like and dislike about the company, Tesla can improve its offerings and address any issues that customers may have.
- Assessment of the impact of external factors: The sentiment analysis also considers external factors, such as the COVID-19 pandemic, which can impact the stock market trends. By considering these external factors, investors can better understand the context in which the stock market trends are occurring.

Deviations from the Expected Results & Justification:

- External factors: The sentiment analysis considers external factors such as the COVID-19 pandemic, but there could be other external factors that may impact the stock market trends of Tesla. For example, changes in government policies or regulations, natural disasters, or economic downturns can all impact stock market trends. In such cases, the sentiment analysis may not accurately predict the stock market trends, and deviations from expected results can occur.
- Limitations of sentiment analysis techniques: Sentiment analysis techniques have certain limitations, and these limitations may result in deviations from the expected results. For example, sentiment analysis may not accurately capture the nuances of human emotions and may not be able to accurately distinguish between sarcasm, irony, or other forms of figurative language. Additionally, sentiment analysis may not work well with languages other than English and may not be effective in capturing sentiment from certain demographics or cultures.

Deviations from the Expected Results & Justification:

- Market volatility: The stock market can be highly volatile, and sudden changes in market trends can occur due to various reasons. In such cases, the sentiment analysis may not be able to accurately predict the stock market trends, and deviations from expected results can occur.
- Data quality: The accuracy of the sentiment analysis is dependent on the quality of the data used. If the data used is biased, incomplete, or inaccurate, the sentiment analysis may not provide accurate results. For example, if the sentiment analysis is conducted on a limited sample of tweets, or if the tweets analyzed are not representative of the overall sentiment, deviations from expected results can occur.

Conclusions:

- In conclusion, the sentiment analysis of Tesla's Twitter data using NLP techniques and the subsequent stock market trend prediction using Tableau have yielded significant results. The analysis showed that the majority of the tweets related to Tesla were neutral, followed by positive and negative tweets. The positive sentiment could be attributed to Tesla's innovative products, mission, and CEO Elon Musk's influence on social media.
- The predicted stock market trends based on the sentiment analysis were found to be accurate when compared to the actual trends, indicating that the proposed methodology is effective in predicting the stock market trends of Tesla based on social media sentiment analysis. This methodology can be a useful tool for investors to make informed decisions and for companies like Tesla to understand customer sentiment and improve their offerings.
- However, it is important to consider potential deviations from the expected results due to external factors, limitations of sentiment analysis techniques, market volatility, and data quality. These factors should be considered when interpreting the results of sentiment analysis and stock market trend prediction.
- Overall, the sentiment analysis and stock market trend prediction provide valuable insights into customer sentiment and stock market trends and can be used by investors and companies to make more informed decisions.

Problem Statement / Objective:

The problem statement of this study is to analyze the sentiment of Twitter data related to Tesla using NLP techniques and predict the stock market trends of Tesla in 2021 based on the sentiment analysis results. The main objective of this study is to determine the effectiveness of social media sentiment analysis in predicting stock market trends and to provide valuable insights into customer sentiment towards Tesla. The study aims to answer the following research questions:

1. What is the sentiment of Twitter data related to Tesla?
2. Can sentiment analysis be used to predict the stock market trends of Tesla in 2021?
3. What are the potential deviations from the expected results, and how can they be justified?
4. What are the implications of the results obtained for investors and companies like Tesla?

The study aims to contribute to the growing body of research on the use of social media sentiment analysis for stock market trend prediction and to provide practical insights for investors and companies in the technology industry.

Work Methodology Adopted:

1. Data collection
2. Data preprocessing
3. Sentiment analysis
4. Data visualization
5. Stock market trend prediction
6. Deviation analysis
7. Conclusion

General Conclusions:

- The sentiment analysis of Twitter data related to Tesla using NLP techniques showed that the majority of the tweets were neutral, followed by positive and negative tweets. The positive sentiment could be attributed to Tesla's innovative products, mission, and CEO Elon Musk's influence on social media.
- The stock market trend prediction based on the sentiment analysis results using Tableau was found to be accurate when compared to the actual trends, indicating that the proposed methodology is effective in predicting the stock market trends of Tesla based on social media sentiment analysis.
- Potential deviations from the expected results were identified, and their justifications were provided. These potential deviations were based on external factors, limitations of sentiment analysis techniques, market volatility, and data quality.
- The implications of the results obtained for investors and companies like Tesla were discussed. Investors can use the sentiment analysis and stock market trend prediction to make informed decisions, while companies like Tesla can understand customer sentiment and improve their offerings.
- The study contributes to the growing body of research on the use of social media sentiment analysis for stock market trend prediction and provides practical insights for investors and companies in the technology industry.

Future Scope:

- Increasing the scope of data collection: The study was limited to Twitter data from western countries where Tesla is present. Future studies could consider collecting data from a broader range of social media platforms such as Facebook, LinkedIn, and Instagram and a wider geographical area to obtain a more comprehensive understanding of customer sentiment and market trends.
- Using more advanced sentiment analysis techniques: While the study used VaderSentimentAnalyzer, a popular sentiment analysis library, future studies could consider using more advanced techniques such as deep learning models to improve the accuracy of sentiment analysis.
- Incorporating more external factors: The study considered the impact of COVID-19 on market trends in 2021. However, future studies could incorporate more external factors such as political instability, economic policies, and environmental concerns, to provide a more comprehensive analysis of market trends.
- Analyzing sentiment of specific product lines: Future studies could consider analyzing the sentiment of specific Tesla products, such as electric vehicles or renewable energy solutions, to obtain a more detailed understanding of customer sentiment towards specific products.

- Comparing sentiment across competitors: Future studies could compare sentiment analysis results across competitors in the technology industry to obtain a comparative analysis of customer sentiment and market trends.
- Integration of external factors: The project did not consider external factors such as news articles, economic conditions, and political events. The project can be extended to include external factors that may impact the stock market trends of Tesla and provide a more accurate prediction.
- Development of a real-time system: The project was conducted using historical data, and the results were obtained after the analysis of the data was completed. In the future, a real-time system can be developed that continuously analyses the sentiment of tweets related to Tesla and provides real-time predictions of the stock market trends of Tesla.
- Use of Machine Learning model: Sentiment Analysis here was done only through English Language and predicting the tone of the person who had posted the tweet. In the future, we can train and test the tweets dataset so as to obtain higher accuracy as analysis would be based on relevant data and not only on the generality of the tone of the speaker.

THANK YOU!!!