



DATA MINING

Assignment 2

Kanika Kushwaha

Student No: 40366629

Word Count : 1648

Index

1.0 Introduction.....	3
2.0 Methodology.....	3
3.0 Results And Discussions	4
3.1 Sentiment Analysis	9
3.2 Model Interpretation.....	10
4.0 Conclusion.....	11
5.0 Bibliography	12
6.0 Appendix – R Code	13

1.0 Introduction

The study involves analysing Tweets by a renowned person known as Elon Musk who is a pioneer in the field of Artificial Intelligence, Engineering and so many other fields. In this study, the Twitter channel was chosen to analyse new and previous tweets made by Elon Musk in 2022 which can help identify the latest trends and upcoming technologies and overall happenings in the world.

This report aims to identify the right set of factors and relationships among attributes involving Twitter data and analyse the sentiments of the tweets by Elon Musk. It concludes with finding the best approach and model to predict the number of likes obtained by a tweet using regression models like multiple linear regression, decision tree regression and random forest regression. It closely examines the methodology and later focuses on the results and findings and concludes with the right strategies corporations can follow to enhance their Twitter marketing strategies.

2.0 Methods

The Elon Musk dataset contains 3060 observations and 4 variables related to the tweets made by Elon Musk in the year 2022 and the information about the engagement around the tweets. The data set was checked for missing data and outliers using summary statistics. The duplicate values were checked and removed. The Date column was split into Dates and Time that contained date and time respectively. The length of each tweet was calculated and stored in a new variable called length. Further, several steps were taken to clean the tweets. The tweets were converted to lowercase, and all URLs and other HTML contents including images and emoticons were removed. Stopwords which are frequently used words in English were also removed to avoid hindrance during analysis. Further, any tweet of a length of fewer than 5 words was also removed as they aren't utilized as much useful in analysis. Tokenization was also performed on the tweets which is a process of splitting into a list of tokens. Further, Stemming was also performed that removed frequently occurring words.

Using matplotlib and seaborn libraries, visualisations were made to analyse the relationships between the dataset's variables. Further, functions were created to analyse the most frequently occurring words in the tweets. A word cloud was also created that showed the most frequently used words in the tweets. To further interpret the data, sentiment analysis on the most liked tweets was also performed.

To predict the number of likes with the selected feature set, multiple linear regression, random forest regression and support vector regression were used. Features like the number of URLs and the presence of images were extracted from the tweets for regression models to better fit the data. Further, the cleaned tweet was encoded using CounterVectorization. The categorical variables were converted to numeric data using dummies. The length, number of words and sentiment scores were later added to the feature set.

According to (Uyanık & Güler, 2013), regression analysis is a statistical technique for analysing associations between variables with a cause-and-effect relationship. Regression models with one dependent variable and numerous independent variables are referred to as multiple regressions. (Uyanık and Güler, 2013).

The multiple regression model can be formulated as follows

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + \dots + \alpha_p X_p$$

A form of tree-based structure called decision tree regression is used to forecast the dependent variable's numerical results (Rathore and Kumar, 2016).

The method describing the ensemble's fundamental predictors is tree-structured trees, and each of these trees is built using an infusion of randomness in Random Forest (Segal, 2004).

The Elon musk dataset was randomly divided into training and testing data. The training dataset consists of 1758 records with 3307 variables. In the test dataset, there were around 1172 records with 3307 variables. A correlation matrix was calculated specifying the strength of relationships between variables. Firstly, feature extraction using counter vectorization of the tweets was performed. Further, the number of URLs and the presence of images were also calculated. The feature set included the length of the tweets, number of words, cleaned tweets, tokenized tweets, number of URLs, contains images, sentiment and time. To predict the number of likes, 3 types of regression algorithms were used - Multiple linear regression, decision tree regression and random forest regression. The comparison of the regression models to find the best model was done using mean squared error and accuracy. The following section discusses the results and findings of the study being carried out.

3.0 Results

The summary statistics were performed on the dataset (Figure 1). Figure 2 (See Appendix) shows that the majority of tweets had a length between 10 – 50.

	Tweets	Retweets	Likes	Date
count	3060	3060.000000	3.060000e+03	3060
unique	2994	NaN	NaN	3029
top	@BillyM2k 🌈	NaN	NaN	23/08/2022 15:52
freq	13	NaN	NaN	2
mean	NaN	5847.451634	7.074538e+04	NaN
std	NaN	21662.545853	1.965653e+05	NaN
min	NaN	41.000000	9.330000e+02	NaN
25%	NaN	300.750000	5.914750e+03	NaN
50%	NaN	786.000000	1.521050e+04	NaN
75%	NaN	3366.250000	5.649675e+04	NaN
max	NaN	681707.000000	4.780787e+06	NaN

Figure 1: Summary Statistics of Elon musk tweets dataset

Figures 3 and 4 show that the highest number of likes and retweets were in the month of May 2022.

As the retweets, likes and Twitter length is left-skewed, it can be interpreted that just a small percentage of the tweets will do well in terms of reader interaction while the majority of them will produce little interaction.

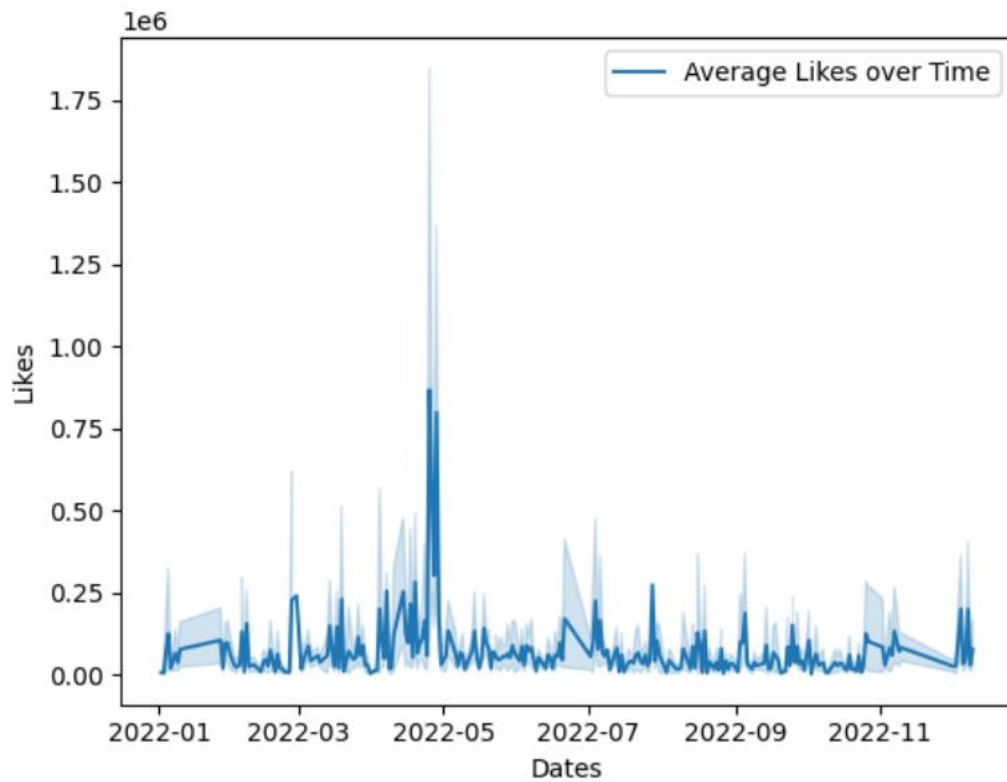


Figure 3: Number of likes obtained by tweets by date

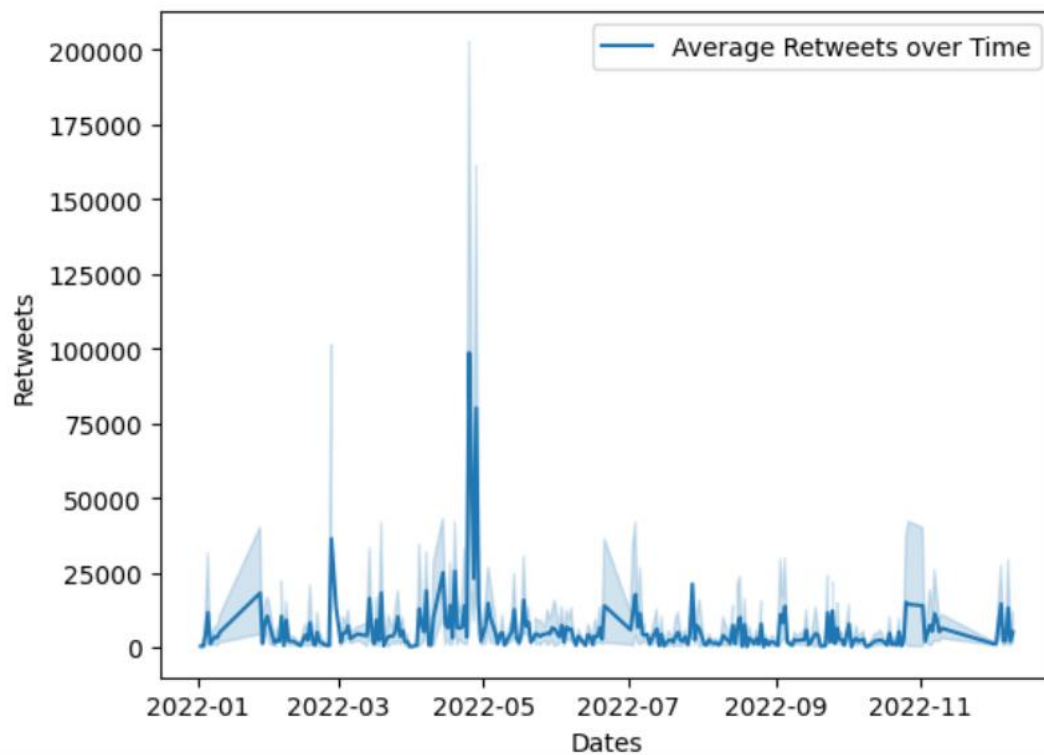


Figure 4: Number of retweets obtained by tweets by date

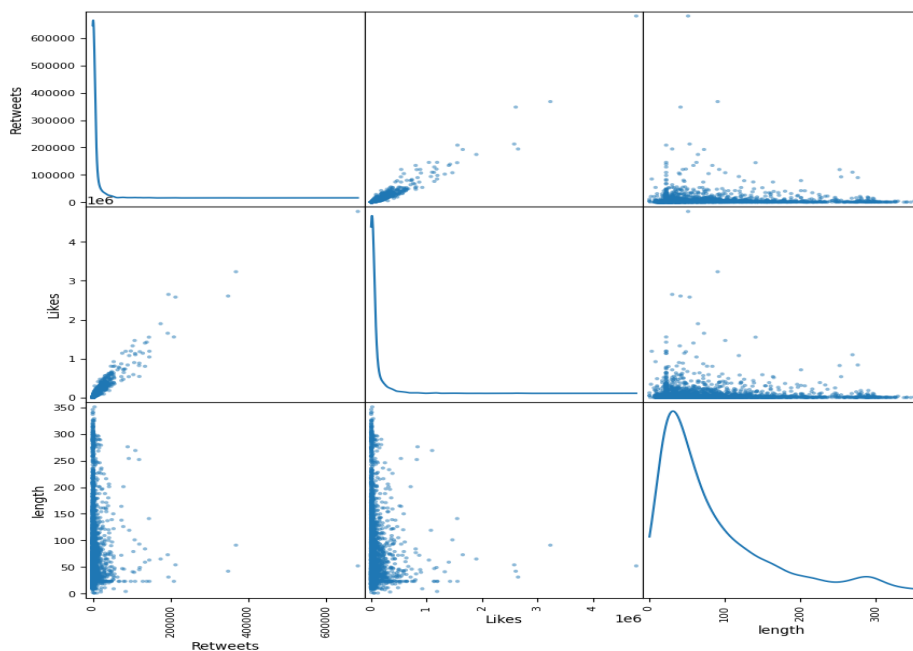


Figure 5: Scatter plot of various attributes of tweets

Figure 5 indicates a strong relationship between likes and retweets. Figure 6 shows that as the number of words increases, engagement (likes and retweets) decreases. Therefore, a smaller number of words are preferred and have higher engagement.

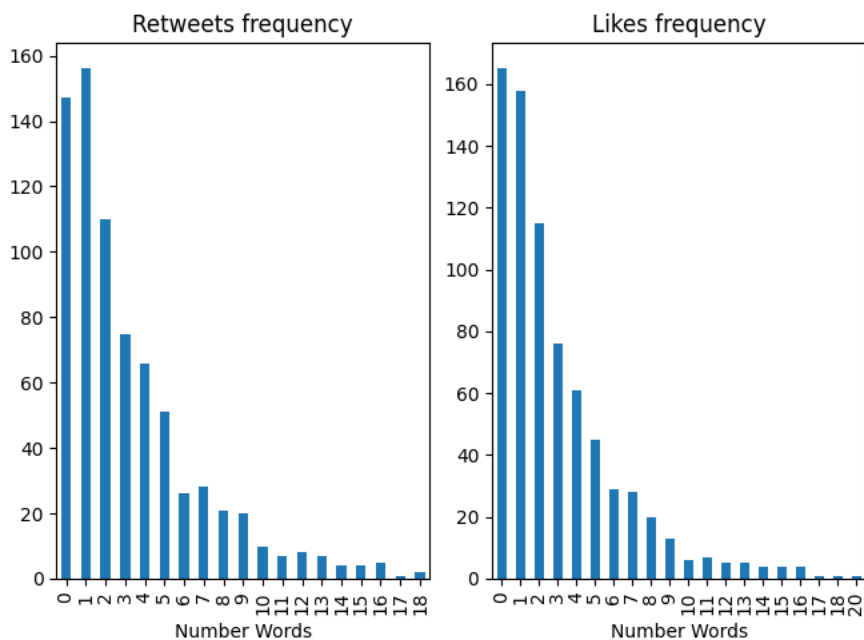


Figure 6: Bar graph of retweets and likes w.r.t number of words

Figure 7 shows the most used words in the tweets with the top 25% of likes and retweets. The top 4 words for both likes and retweets were people, Twitter, Starlink and SpaceX. Further, the frequency distributions of tweets were calculated.

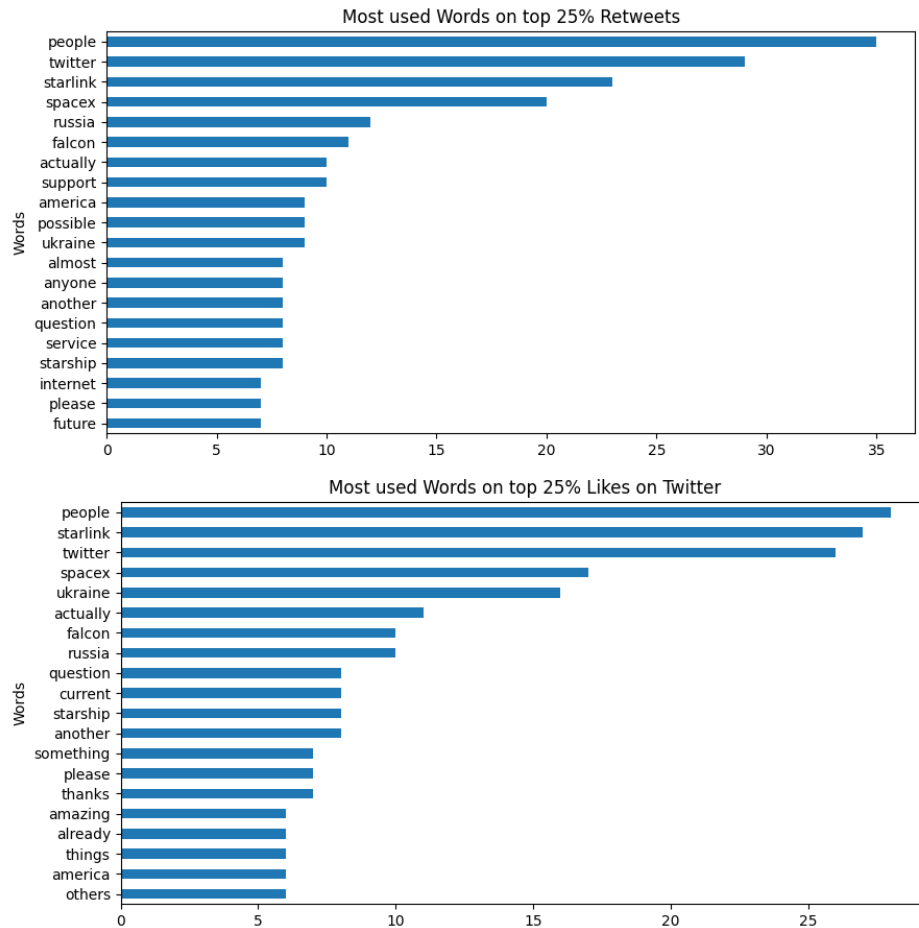


Figure 7: Most used words in the top 25% of liked and retweeted tweets

Figure 8 shows a bar graph of the most common words found in tweets. Figure 9 shows the word cloud which shows a similar trend of words as concluded by the frequency distribution and common words in tweets.

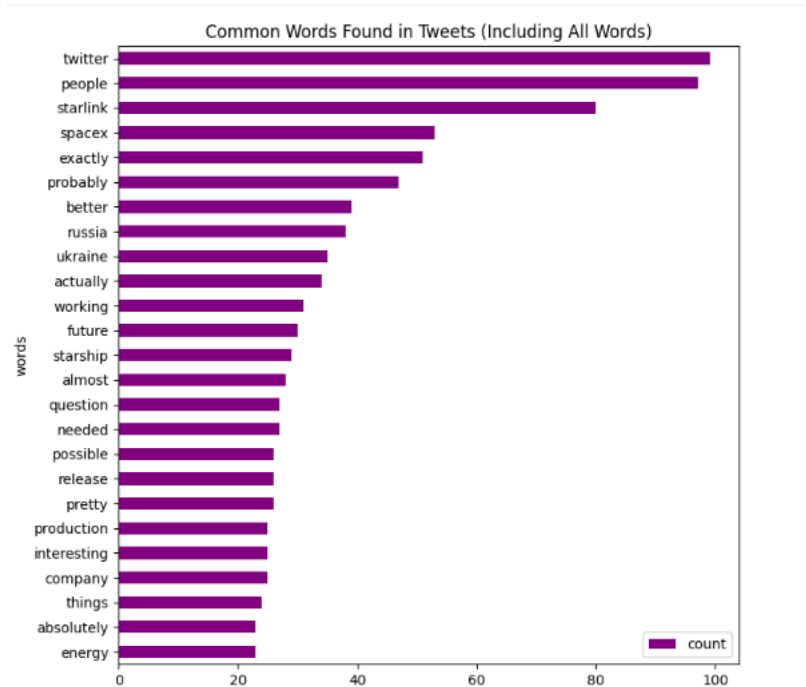


Figure 8: Bar graph representing Common Words found in Tweets

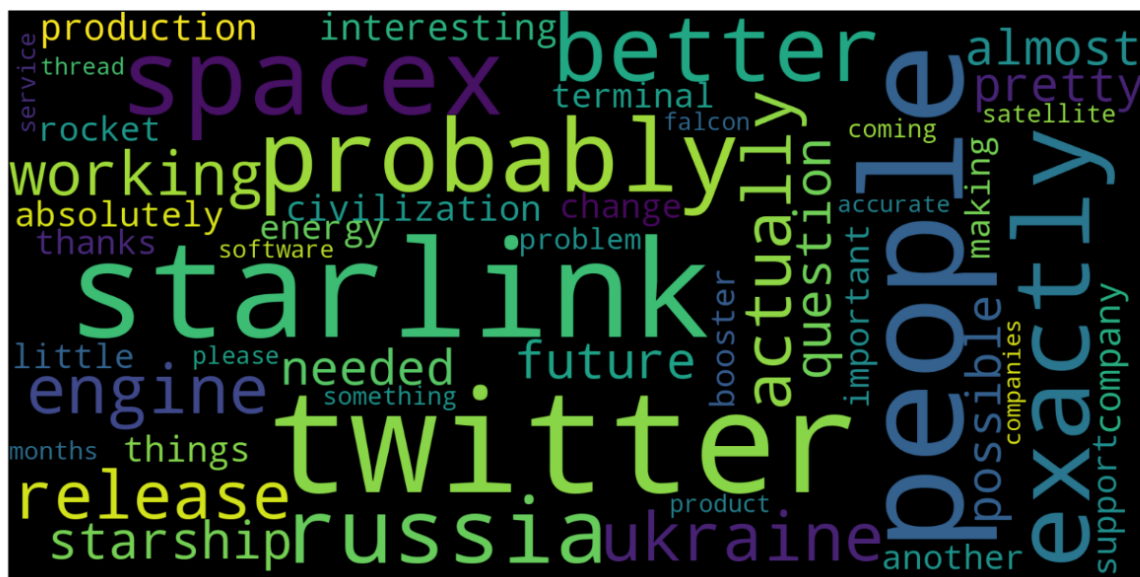


Figure 9: Word Cloud for tweets

3. 1 Sentiment Analysis of Tweets

While performing the sentiment analysis on the top 5 liked tweets, it revealed that the top 5 liked tweets have overall a neutral sentiment with only one tweet with a negative sentiment. When the sentiment was calculated for the whole dataset, it was

revealed that 73.25% of tweets were of neutral sentiment while 18.90% and 7.85 % were positive and negative respectively (Figure 10). Therefore, the overall sentiment of the tweets is neutral in nature.

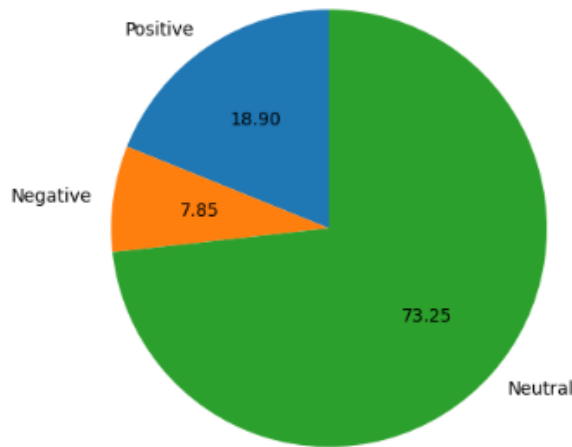


Figure 10: Percentage distribution of sentiments of tweets

3.2 Model Interpretation

Figure 12 shows the heatmap representing the various strengths of relationships between variables. It shows a higher and positive correlation between retweets, no_of_urls and likes and a negative correlation between length, number of words, sentiment and likes. The retweets have a positive correlation with likes, number words and no_of_urls while a negative correlation with length and sentiment. The length has a high positive correlation with the number of words and sentiment.

Regression Models	R ²	RMSE
Multiple Linear Regression	0.639	130084
Decision Tree Regression	0.829	7695568511
Random Forest Regression	0.835	49932042893

Table 1 : Regression Models Summary

Table 1 provides a summary of accuracy across the classification models used.

From Table 1, it can be seen that all the three-regression models have negative r square values. This means that these regression models have produced regression lines that are worse than using the mean value and the models have fitted the data very badly. Out of the three, the best model for predicting the number of likes is the multiple linear regression with the smallest R- Square value of 0.63 which, in turn, determines the average error between the actual and predicted values. The root means square error which is the standard deviation of residuals is 130084 which is the lowest among the three models. The decision tree regression has performed better with R square 0.829 and root mean square error 7695568511 than the random forest regression with R square 0.835 and root mean square error 49932042893 which is a very small difference.

	Coefficient
contains_image	-0.001143
Dates_2022-01-04	-2563.816818
Dates_2022-01-05	-23679.543045
Dates_2022-01-06	19514.118596
Dates_2022-01-08	35788.434989
...	...
Retweets	10.620495
no_of_urls	17111.278039
Number words	-1864.678130
length	54.084301
sentiment	1246.610277

Figure 13: Coefficient values obtained through multiple linear regression

Figure 13 shows the coefficients of the various attributes obtained from multiple linear regression. Retweets have a positive relationship with likes. Thus, as the number of retweets increases the number of likes will also increase. The no_of_urls has a very strong positive relationship with likes. Therefore, as the number of URLs in tweets increases, the number of likes also increases. The number of words has a large negative relationship with the number of likes. Therefore, as the number of words decreases, the number of likes will increase. The length and sentiment also have a positive relationship with likes. Therefore, as the sentiment score increases, the number of likes also increases. Similar is the case with the length attribute. Lastly, common words like Twitter, people have a very strong positive relationship with likes and therefore if these words are used more often in tweets by Elon Musk, it is very likely that the number of likes will increase.

4.0 Conclusion

In conclusion, analysing Twitter data is useful in terms of finding what the audience sentiments are for a particular topic or a person and analysing various attributes to consider while formulating marketing strategies on Twitter. In this study, the overall sentiments were neutral for Elon Musk.

Attributes of the tweets like retweets, number of URLs, length and sentiment all have a positive impact on the number of likes. Thus, corporations can take advantage of these factors to enhance their marketing strategies by increasing engagement around the tweets. The number of URLs has a really strong positive impact on the number of likes. Therefore, companies can make sure to include a lot of URLs in the tweets. As the sentiment score will be more, the tweets will be more positive in nature and thus will garner more likes. The number of words used in the tweets should be less due to its negative relationship with the number of likes.

5.0 Bibliography

Rathore, S.S. and Kumar, S. (2016) 'A Decision Tree Regression Based Approach for the Number of Software Faults Prediction', *SIGSOFT Softw. Eng. Notes*, 41(1), pp. 1–6. Available at: <https://doi.org/10.1145/2853073.2853083>.

Segal, M.R. (2004) 'Machine learning benchmarks and random forest regression'.

Uyanık, G.K. and Güler, N. (2013) 'A Study on Multiple Linear Regression Analysis', *Procedia - Social and Behavioral Sciences*, 106, pp. 234–240. Available at: <https://doi.org/10.1016/j.sbspro.2013.12.027>.

6.0 Appendix

data-mining-assignment-2

April 27, 2023

[210]: `pip install nltk`

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Requirement already satisfied: nltk in /usr/local/lib/python3.9/dist-packages
(3.8.1)
Requirement already satisfied: tqdm in /usr/local/lib/python3.9/dist-packages
(from nltk) (4.65.0)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.9/dist-
packages (from nltk) (2022.10.31)
Requirement already satisfied: joblib in /usr/local/lib/python3.9/dist-packages
(from nltk) (1.2.0)
Requirement already satisfied: click in /usr/local/lib/python3.9/dist-packages
(from nltk) (8.1.3)
```

```
[211]: ## importing packages

import matplotlib.pyplot as plt
%matplotlib inline
import numpy as np
import pandas as pd
import seaborn as sns
import nltk
import html
import re
import random
import string
from wordcloud import WordCloud, STOPWORDS
from PIL import Image
import itertools
import collections
import tweepy
from textblob import TextBlob
nltk.download('words')
words = set(nltk.corpus.words.words())

nltk.download([
    "names",
```

```

        "stopwords",
        "words",
        "punkt"
    ])

from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem import PorterStemmer

!pip install vaderSentiment
from vaderSentiment.vaderSentiment import SentimentIntensityAnalyzer

analyser = SentimentIntensityAnalyzer()

!pip install datashader
import datashader as ds
import datashader.transfer_functions as tf

```

```

[nltk_data] Downloading package words to /root/nltk_data...
[nltk_data] Package words is already up-to-date!
[nltk_data] Downloading package names to /root/nltk_data...
[nltk_data] Package names is already up-to-date!
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Package stopwords is already up-to-date!
[nltk_data] Downloading package words to /root/nltk_data...
[nltk_data] Package words is already up-to-date!
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Package punkt is already up-to-date!

```

```

Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Requirement already satisfied: vaderSentiment in /usr/local/lib/python3.9/dist-
packages (3.3.2)
Requirement already satisfied: requests in /usr/local/lib/python3.9/dist-
packages (from vaderSentiment) (2.27.1)
Requirement already satisfied: charset-normalizer~=2.0.0 in
/usr/local/lib/python3.9/dist-packages (from requests->vaderSentiment) (2.0.12)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in
/usr/local/lib/python3.9/dist-packages (from requests->vaderSentiment) (1.26.15)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.9/dist-
packages (from requests->vaderSentiment) (3.4)
Requirement already satisfied: certifi>=2017.4.17 in
/usr/local/lib/python3.9/dist-packages (from requests->vaderSentiment)
(2022.12.7)
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-
wheels/public/simple/
Requirement already satisfied: datashader in /usr/local/lib/python3.9/dist-
packages (0.14.4)

```

Requirement already satisfied: numba>=0.51 in /usr/local/lib/python3.9/dist-packages (from datashader) (0.56.4)

Requirement already satisfied: pillow in /usr/local/lib/python3.9/dist-packages (from datashader) (8.4.0)

Requirement already satisfied: pandas in /usr/local/lib/python3.9/dist-packages (from datashader) (1.5.3)

Requirement already satisfied: requests in /usr/local/lib/python3.9/dist-packages (from datashader) (2.27.1)

Requirement already satisfied: datashape in /usr/local/lib/python3.9/dist-packages (from datashader) (0.5.2)

Requirement already satisfied: scipy in /usr/local/lib/python3.9/dist-packages (from datashader) (1.10.1)

Requirement already satisfied: param in /usr/local/lib/python3.9/dist-packages (from datashader) (1.13.0)

Requirement already satisfied: xarray in /usr/local/lib/python3.9/dist-packages (from datashader) (2022.12.0)

Requirement already satisfied: toolz in /usr/local/lib/python3.9/dist-packages (from datashader) (0.12.0)

Requirement already satisfied: numpy in /usr/local/lib/python3.9/dist-packages (from datashader) (1.22.4)

Requirement already satisfied: dask in /usr/local/lib/python3.9/dist-packages (from datashader) (2022.12.1)

Requirement already satisfied: pyct in /usr/local/lib/python3.9/dist-packages (from datashader) (0.5.0)

Requirement already satisfied: colorcet in /usr/local/lib/python3.9/dist-packages (from datashader) (3.0.1)

Requirement already satisfied: llvmlite<0.40,>=0.39.0dev0 in /usr/local/lib/python3.9/dist-packages (from numba>=0.51->datashader) (0.39.1)

Requirement already satisfied: setuptools in /usr/local/lib/python3.9/dist-packages (from numba>=0.51->datashader) (67.7.2)

Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.9/dist-packages (from dask->datashader) (23.1)

Requirement already satisfied: pyyaml>=5.3.1 in /usr/local/lib/python3.9/dist-packages (from dask->datashader) (6.0)

Requirement already satisfied: cloudpickle>=1.1.1 in /usr/local/lib/python3.9/dist-packages (from dask->datashader) (2.2.1)

Requirement already satisfied: click>=7.0 in /usr/local/lib/python3.9/dist-packages (from dask->datashader) (8.1.3)

Requirement already satisfied: partd>=0.3.10 in /usr/local/lib/python3.9/dist-packages (from dask->datashader) (1.4.0)

Requirement already satisfied: fsspec>=0.6.0 in /usr/local/lib/python3.9/dist-packages (from dask->datashader) (2023.4.0)

Requirement already satisfied: python-dateutil in /usr/local/lib/python3.9/dist-packages (from datashape->datashader) (2.8.2)

Requirement already satisfied: multipledispatch>=0.4.7 in /usr/local/lib/python3.9/dist-packages (from datashape->datashader) (0.6.0)

Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.9/dist-packages (from pandas->datashader) (2022.7.1)

Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.9/dist-packages (from requests->datashader) (3.4)
 Requirement already satisfied: charset-normalizer~=2.0.0 in /usr/local/lib/python3.9/dist-packages (from requests->datashader) (2.0.12)
 Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.9/dist-packages (from requests->datashader) (2022.12.7)
 Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.9/dist-packages (from requests->datashader) (1.26.15)
 Requirement already satisfied: six in /usr/local/lib/python3.9/dist-packages (from multipledispatch>=0.4.7->datashape->datashader) (1.16.0)
 Requirement already satisfied: locket in /usr/local/lib/python3.9/dist-packages (from partd>=0.3.10->dask->datashader) (1.0.0)

```
[212]: elon_data = pd.read_csv("Elon musk.csv") ## read in data
```

```
[213]: elon_data
```

```
[213]:
```

		Tweets	Retweets	Likes	\
0	@PeterSchiff thanks		209	7021	
1	@ZubyMusic Absolutely		755	26737	
2	Dear Twitter Advertisers https://t.co/GMwHmInPAS		55927	356623	
3	@BillyM2k		802	19353	
4	Meeting a lot of cool people at Twitter today!		9366	195546	
...	
3055	@LimitingThe @baglino Just that manganese is a...		171	3173	
3056	@incentives101 @ICRicardoLara Exactly		145	4234	
3057	@ICRicardoLara Your policies are directly resp...		421	6144	
3058	@ICRicardoLara You should be voted out of office		484	7029	
3059	CB radios are free from govt/media control		11302	113429	

	Date
0	27/10/2022 16:17
1	27/10/2022 13:19
2	27/10/2022 13:08
3	27/10/2022 02:32
4	26/10/2022 21:39
...	...
3055	27/01/2022 22:01
3056	27/01/2022 21:23
3057	27/01/2022 21:13
3058	27/01/2022 21:12
3059	27/01/2022 21:00

[3060 rows x 4 columns]

```
[214]: elon_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```

RangeIndex: 3060 entries, 0 to 3059
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0    Tweets      3060 non-null   object
1    Retweets     3060 non-null   int64
2    Likes        3060 non-null   int64
3    Date         3060 non-null   object
dtypes: int64(2), object(2)
memory usage: 95.8+ KB

```

```

[215]: ##Checking for null values

np.sum(elon_data.isnull().any(axis=1))

```

[215]: 0

```

[216]: elon_data.describe(include = 'all')

```

```

[216]:
           Tweets      Retweets      Likes      Date
count          3060    3060.000000  3.060000e+03    3060
unique           2994             NaN             NaN    3029
top    @BillyM2k             NaN             NaN  23/08/2022 15:52
freq           13             NaN             NaN         2
mean           NaN    5847.451634  7.074538e+04         NaN
std           NaN    21662.545853  1.965653e+05         NaN
min           NaN     41.000000  9.330000e+02         NaN
25%           NaN     300.750000  5.914750e+03         NaN
50%           NaN     786.000000  1.521050e+04         NaN
75%           NaN    3366.250000  5.649675e+04         NaN
max           NaN   681707.000000  4.780787e+06         NaN

```

```

[217]: elon_data.shape

```

[217]: (3060, 4)

```

[218]: ##cleaning tweets

###removing duplicates

elon_tweets = elon_data.drop_duplicates("Tweets",keep="first") #delete the
↳ duplicates by dropping them and store the result value to a new variable
elon_tweets

```

```

[218]:
           Tweets  Retweets  Likes \
0    @PeterSchiff  thanks      209  7021
1    @ZubyMusic  Absolutely    755  26737

```

```

2      Dear Twitter Advertisers https://t.co/GMwHmInPAS      55927  356623
3                                     @BillyM2k              802    19353
4      Meeting a lot of cool people at Twitter today!      9366  195546
...
3055 @LimitingThe @baglino Just that manganese is a...      171    3173
3056                @incentives101 @ICRicardoLara Exactly      145    4234
3057 @ICRicardoLara Your policies are directly resp...      421    6144
3058 @ICRicardoLara You should be voted out of office      484    7029
3059                CB radios are free from govt/media control 11302  113429

```

```

                Date
0      27/10/2022 16:17
1      27/10/2022 13:19
2      27/10/2022 13:08
3      27/10/2022 02:32
4      26/10/2022 21:39
...
3055  27/01/2022 22:01
3056  27/01/2022 21:23
3057  27/01/2022 21:13
3058  27/01/2022 21:12
3059  27/01/2022 21:00

```

[2994 rows x 4 columns]

```

[219]: https://stackoverflow.com/questions/35595710/splitting-timestamp-column-into-separate-date-and-time-columns

elon_tweets['Dates'] = pd.to_datetime(elon_tweets['Date']).dt.date
elon_tweets['Time'] = pd.to_datetime(elon_tweets['Date']).dt.time

```

<ipython-input-219-645aecb2b8d6>:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
elon_tweets['Dates'] = pd.to_datetime(elon_tweets['Date']).dt.date

<ipython-input-219-645aecb2b8d6>:4: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
elon_tweets['Time'] = pd.to_datetime(elon_tweets['Date']).dt.time

```

[220]: elon_tweets

```

[220]:

		Tweets	Retweets	Likes	\
0	@PeterSchiff	thanks	209	7021	
1	@ZubyMusic	Absolutely	755	26737	
2	Dear Twitter Advertisers	https://t.co/GMwHmInPAS	55927	356623	
3	@BillyM2k		802	19353	
4	Meeting a lot of cool people at Twitter today!		9366	195546	
...		
3055	@LimitingThe @baglino	Just that manganese is a...	171	3173	
3056	@incentives101 @ICRicardoLara	Exactly	145	4234	
3057	@ICRicardoLara	Your policies are directly resp...	421	6144	
3058	@ICRicardoLara	You should be voted out of office	484	7029	
3059	CB radios are free from govt/media control		11302	113429	

	Date	Dates	Time
0	27/10/2022 16:17	2022-10-27	16:17:00
1	27/10/2022 13:19	2022-10-27	13:19:00
2	27/10/2022 13:08	2022-10-27	13:08:00
3	27/10/2022 02:32	2022-10-27	02:32:00
4	26/10/2022 21:39	2022-10-26	21:39:00
...
3055	27/01/2022 22:01	2022-01-27	22:01:00
3056	27/01/2022 21:23	2022-01-27	21:23:00
3057	27/01/2022 21:13	2022-01-27	21:13:00
3058	27/01/2022 21:12	2022-01-27	21:12:00
3059	27/01/2022 21:00	2022-01-27	21:00:00

[2994 rows x 6 columns]

```
[221]: # We wont be needing the date column now
elon_tweets = elon_tweets.drop(columns='Date',axis=1)
elon_tweets
```

[221]:

		Tweets	Retweets	Likes	\
0	@PeterSchiff	thanks	209	7021	
1	@ZubyMusic	Absolutely	755	26737	
2	Dear Twitter Advertisers	https://t.co/GMwHmInPAS	55927	356623	
3	@BillyM2k		802	19353	
4	Meeting a lot of cool people at Twitter today!		9366	195546	
...		
3055	@LimitingThe @baglino	Just that manganese is a...	171	3173	
3056	@incentives101 @ICRicardoLara	Exactly	145	4234	
3057	@ICRicardoLara	Your policies are directly resp...	421	6144	
3058	@ICRicardoLara	You should be voted out of office	484	7029	
3059	CB radios are free from govt/media control		11302	113429	

	Dates	Time
0	2022-10-27	16:17:00

```

1      2022-10-27  13:19:00
2      2022-10-27  13:08:00
3      2022-10-27  02:32:00
4      2022-10-26  21:39:00
...
3055   2022-01-27  22:01:00
3056   2022-01-27  21:23:00
3057   2022-01-27  21:13:00
3058   2022-01-27  21:12:00
3059   2022-01-27  21:00:00

```

[2994 rows x 5 columns]

```
[222]: elon_tweets.describe(include = 'all')
```

```
[222]:
```

	Tweets	Retweets	Likes	Dates \
count	2994	2994.000000	2.994000e+03	2994
unique	2994	NaN	NaN	251
top	@PeterSchiff thanks	NaN	NaN	2022-05-20
freq	1	NaN	NaN	40
mean	NaN	5964.803273	7.195106e+04	NaN
std	NaN	21885.354243	1.985386e+05	NaN
min	NaN	41.000000	9.330000e+02	NaN
25%	NaN	306.250000	5.901000e+03	NaN
50%	NaN	813.000000	1.544400e+04	NaN
75%	NaN	3512.250000	5.840075e+04	NaN
max	NaN	681707.000000	4.780787e+06	NaN

```

Time
count      2994
unique     1185
top    17:36:00
freq        8
mean       NaN
std        NaN
min        NaN
25%        NaN
50%        NaN
75%        NaN
max        NaN

```

```
[223]: # Let's get the length of the messages
elon_tweets['length'] = elon_tweets['Tweets'].apply(len) ##https://medium.com/
@nikhil\_48887/
sentiment-analysis-on-twitter-dataset-positive-negative-neutral-clustering-85ee7ba75bcf
elon_tweets
```

```
[223]:
```

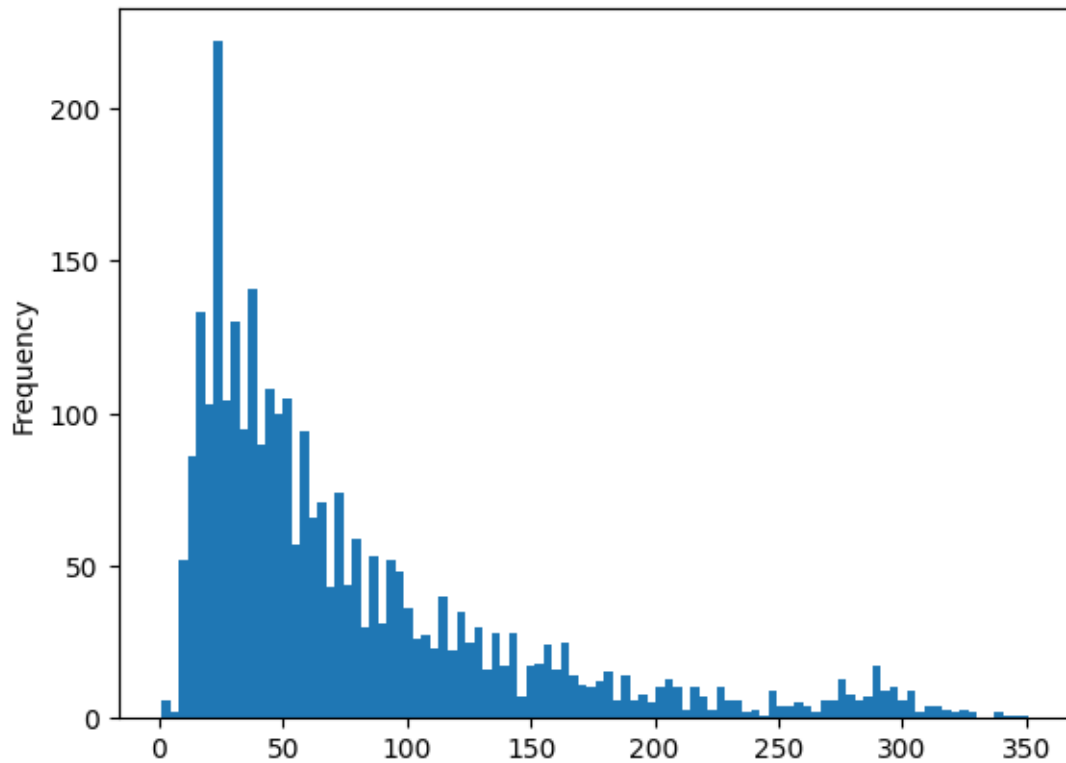
		Tweets	Retweets	Likes	\
0	@PeterSchiff	thanks	209	7021	
1	@ZubyMusic	Absolutely	755	26737	
2	Dear Twitter Advertisers	https://t.co/GMwHmInPAS	55927	356623	
3	@BillyM2k		802	19353	
4	Meeting a lot of cool people at Twitter today!		9366	195546	
...			
3055	@LimitingThe @baglino	Just that manganese is a...	171	3173	
3056	@incentives101 @ICRicardoLara	Exactly	145	4234	
3057	@ICRicardoLara	Your policies are directly resp...	421	6144	
3058	@ICRicardoLara	You should be voted out of office	484	7029	
3059	CB radios are free from govt/media control		11302	113429	

	Dates	Time	length
0	2022-10-27	16:17:00	21
1	2022-10-27	13:19:00	21
2	2022-10-27	13:08:00	48
3	2022-10-27	02:32:00	11
4	2022-10-26	21:39:00	46
...
3055	2022-01-27	22:01:00	135
3056	2022-01-27	21:23:00	37
3057	2022-01-27	21:13:00	119
3058	2022-01-27	21:12:00	48
3059	2022-01-27	21:00:00	42

[2994 rows x 6 columns]

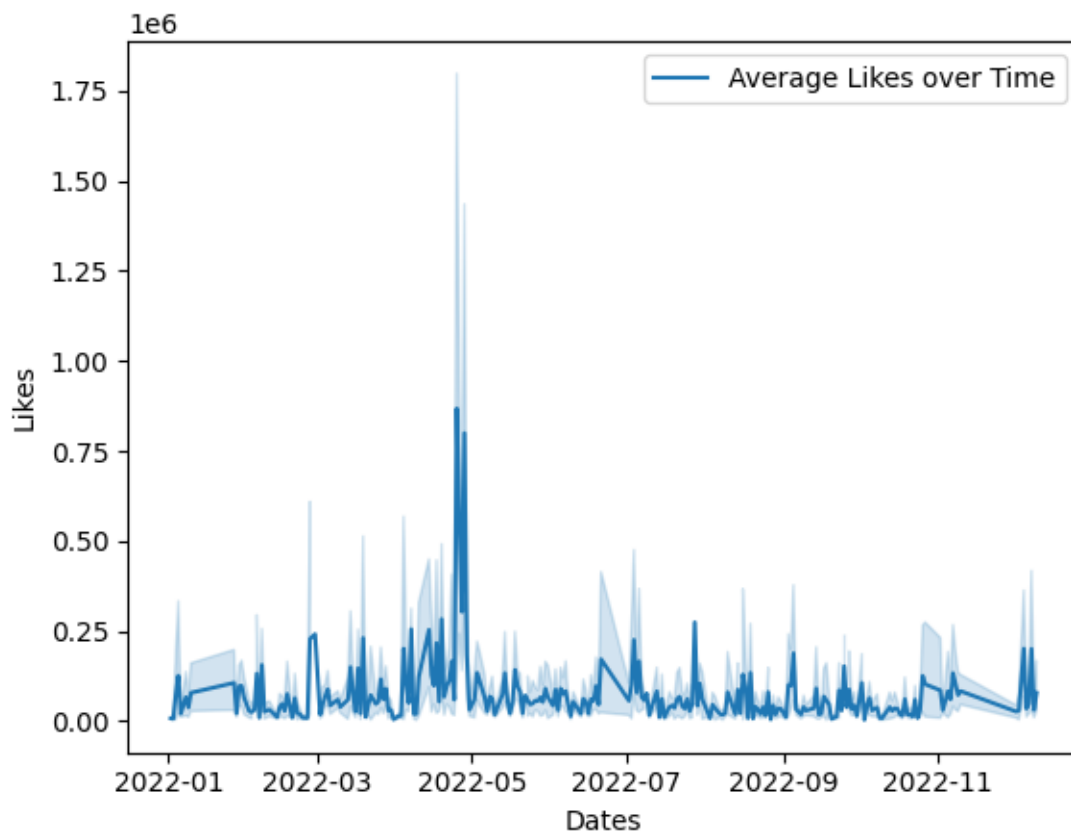
```
[224]: elon_tweets['length'].plot(kind='hist',bins=100)
```

```
[224]: <Axes: ylabel='Frequency'>
```



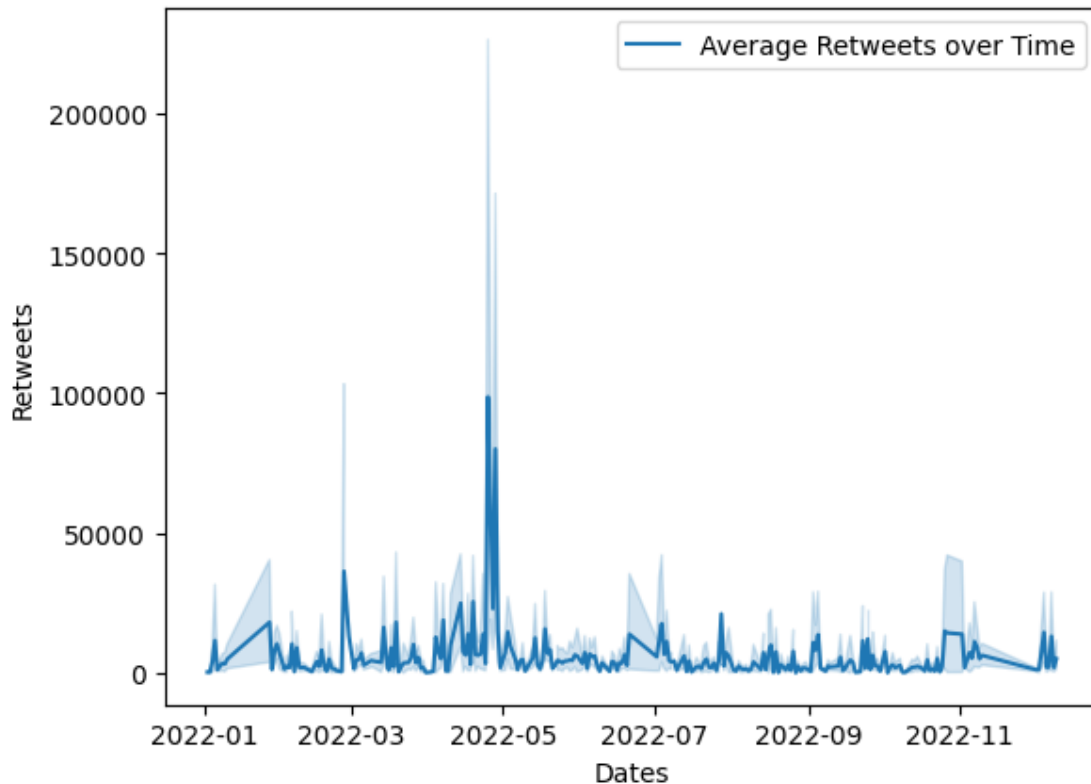
```
[225]: ##https://medium.com/@nikhil_48887/
      ↪sentiment-analysis-on-twitter-dataset-positive-negative-neutral-clustering-85ee7ba75bcf
sns.lineplot( x = 'Dates',
              y = 'Likes',
              data = elon_tweets,
              label = 'Average Likes over Time')
```

```
[225]: <Axes: xlabel='Dates', ylabel='Likes'>
```



```
[226]: https://medium.com/@nikhil\_48887/sentiment-analysis-on-twitter-dataset-positive-negative-neutral-clustering-85ee7ba75bcf
sns.lineplot( x = 'Dates',
              y = 'Retweets',
              data = elon_tweets,
              label = 'Average Retweets over Time')
```

```
[226]: <Axes: xlabel='Dates', ylabel='Retweets'>
```

```
[227]: elon_tweets.describe()
```

```
[227]:
```

	Retweets	Likes	length
count	2994.000000	2.994000e+03	2994.000000
mean	5964.803273	7.195106e+04	79.085170
std	21885.354243	1.985386e+05	69.219397
min	41.000000	9.330000e+02	1.000000
25%	306.250000	5.901000e+03	30.000000
50%	813.000000	1.544400e+04	54.000000
75%	3512.250000	5.840075e+04	103.750000
max	681707.000000	4.780787e+06	351.000000

```
[228]: ##https://github.com/flaviohenriquecbc/machine-learning-capstone-project/blob/
↳master/title-success-prediction.ipynb
title_length = 'length'
# set the columns that will show statistic and graph
columns = [{'column_name': 'Retweets', 'column_text': 'Retweets'},
           {'column_name': 'Likes', 'column_text': 'Likes on Twitter'},
           {'column_name': 'length', 'column_text': 'Tweets length on Twitter'}]

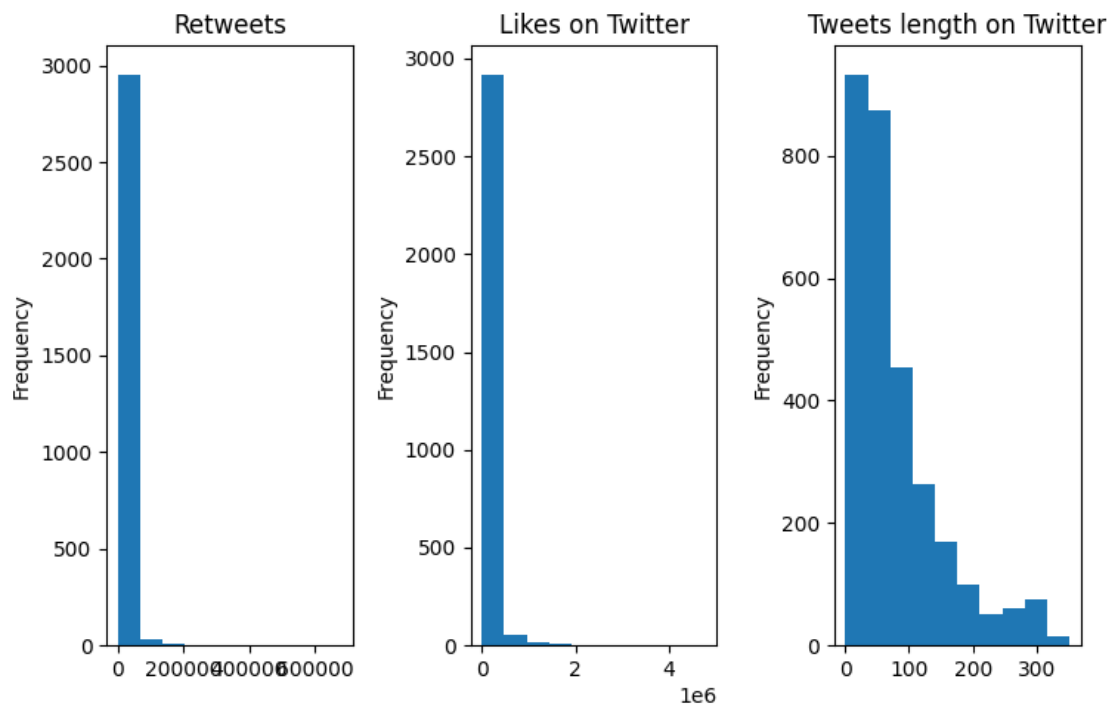
# plot histogram of columns
```

```

plt.figure(figsize=(10, 5))
for el in columns:
    class_name = el['column_name']
    column = elon_tweets[column_name]
    plt.subplot(1, 4, columns.index(el) + 1)
    plt.tight_layout()
    column.plot.hist(orientation='vertical', cumulative=False,
    title=el['column_text'])
    if class_name != title_length:
        print('Total {}: {} times'.format(class_name, column.sum()))

```

Total Retweets: 17858621 times
 Total Likes: 215421483 times



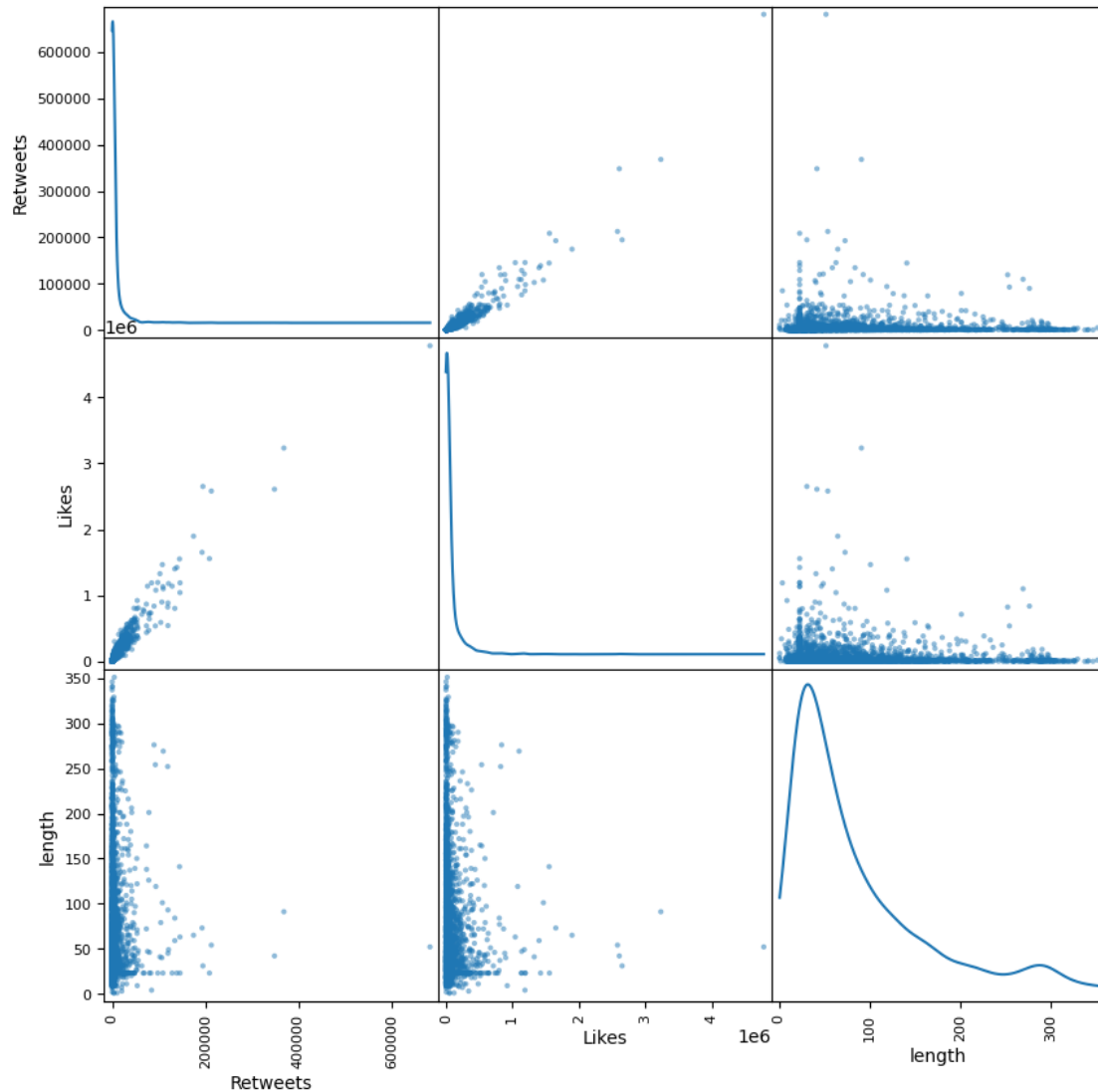
[229]: *### From the above graphs, Likes and retweets are positive-skewed as they are concentrated on the left part of the graph. Therefore, a small part of the articles will over-perform about readers' interaction and the biggest part of them will generate less interaction.*

[230]: *##https://github.com/flaviohenriquecbc/machine-learning-capstone-project/blob/master/title-success-prediction.ipynb*
Trying to gather information about multiple attributes of the twitter data

```

temp = pd.plotting.scatter_matrix(elon_tweets, diagonal="kde", figsize=(10, 10))

```



[231]: *### From the above a direct relationship between likes and retweets can be found.*

[232]: *###https://catriscode.com/2021/05/01/tweets-cleaning-with-python/
Data Cleaning*

```

tweet_list = elon_tweets.Tweets.to_list() ## changing tweets column to list

## Function to clean tweets

def clean_tweet(tweet):
    if type(tweet) == np.float:
        return ""

```

```

r = tweet.lower()
r = re.sub("'", "", r) # This is to avoid removing contractions in english
r = re.sub("@[A-Za-z0-9_]+", "", r)
r = re.sub("#[A-Za-z0-9_]+", "", r)
r = re.sub(r'http\S+', '', r)
r = re.sub('([()!?!])', ' ', r)
r = re.sub('\[.*?\]', ' ', r)
r = re.sub("[^a-z0-9]", " ", r)
r = r.split()
sw = stopwords.words("english") #you can adjust the language as you desire
sw.remove("not")
r = [w for w in r if not w in sw]
stopwords1 = ["for", "on", "an", "a", "of", "and", "in", "the", "to",
↳ "from"]
r = [w for w in r if not w in stopwords1]
r = [x for x in r if len(x) > 5] ## removing tweets with less than 5 words
r = " ".join(word for word in r)
return r

```

```

[233]: cleaned = [clean_tweet(tw) for tw in tweet_list]
token_cleaned = cleaned
elon_tweets["cleaned"] = cleaned

```

<ipython-input-232-053aba553372>:8: DeprecationWarning: `np.float` is a deprecated alias for the builtin `float`. To silence this warning, use `float` by itself. Doing this will not modify any behavior and is safe. If you specifically wanted the numpy scalar type, use `np.float64` here. Deprecated in NumPy 1.20; for more details and guidance: <https://numpy.org/devdocs/release/1.20.0-notes.html#deprecations>

```

if type(tweet) == np.float:

```

```

[234]: for i in range(len(cleaned)):
token_cleaned[i] = word_tokenize(cleaned[i])

```

```

[235]: ##https://www.analyticsvidhya.com/blog/2021/06/
↳ twitter-sentiment-analysis-a-nlp-use-case-for-beginners/
↳ #h-step-5-data-preprocessing
## Performing Stemming

import nltk
st = nltk.PorterStemmer()
def stemming_on_text(data):
    text = [st.stem(word) for word in data]
    return data
elon_tweets['cleaned'] = elon_tweets['cleaned'].apply(lambda x:
↳ stemming_on_text(x))
elon_tweets['cleaned'].head()

```

```
[235]: 0          thanks
      1          absolutely
      2    twitter advertisers
      3
      4    meeting people twitter
      Name: cleaned, dtype: object
```

```
[236]: elon_tweets["tokenized_tweets"] = token_cleaned
      elon_tweets
```

```
[236]:
```

		Tweets	Retweets	Likes	\
0	@PeterSchiff	thanks	209	7021	
1	@ZubyMusic	Absolutely	755	26737	
2	Dear Twitter Advertisers	https://t.co/GMwHmInPAS	55927	356623	
3	@BillyM2k		802	19353	
4	Meeting a lot of cool people at Twitter today!		9366	195546	
...	
3055	@LimitingThe @baglino	Just that manganese is a...	171	3173	
3056	@incentives101 @ICRicardoLara	Exactly	145	4234	
3057	@ICRicardoLara	Your policies are directly resp...	421	6144	
3058	@ICRicardoLara	You should be voted out of office	484	7029	
3059	CB radios are free from govt/media control		11302	113429	

	Dates	Time	length	\
0	2022-10-27	16:17:00	21	
1	2022-10-27	13:19:00	21	
2	2022-10-27	13:08:00	48	
3	2022-10-27	02:32:00	11	
4	2022-10-26	21:39:00	46	
...	
3055	2022-01-27	22:01:00	135	
3056	2022-01-27	21:23:00	37	
3057	2022-01-27	21:13:00	119	
3058	2022-01-27	21:12:00	48	
3059	2022-01-27	21:00:00	42	

		cleaned	\
0		thanks	
1		absolutely	
2		twitter advertisers	
3			
4		meeting people twitter	
...		...	
3055	manganese alternative phosphorus scaling catho...		
3056		exactly	
3057	policies directly responsible outrageously ins...		
3058		office	

```

3059                                radios control

                                tokenized_tweets
0                                [thanks]
1                                [absolutely]
2                                [twitter, advertisers]
3                                []
4                                [meeting, people, twitter]
...
3055 [manganese, alternative, phosphorus, scaling, ...
3056                                [exactly]
3057 [policies, directly, responsible, outrageously...
3058                                [office]
3059                                [radios, control]

[2994 rows x 8 columns]

```

```

[237]: https://github.com/flaviohenriquecbc/machine-learning-capstone-project/blob/master/title-success-prediction.ipynb

# function to plot the top performers
PERCENTAGE = 25

def remove_outliers(df, column):
    Q1 = df[column].quantile(0.25)
    Q3 = df[column].quantile(0.75)
    IQR = Q3 - Q1
    non_outliers = (df[column] >= Q1 - 1.5 * IQR) & (df[column] <= Q3 + 1.5 * IQR)
    # non_outliers = df[column] >= 0
    return df.loc[non_outliers]
# plot relation between the number of words and the retweets and likes
def plot_top_length_performers(col_name, is_grouped):
    plt.figure(figsize=(10, 5))
    for el in columns:
        class_name = el['column_name']
        column = elon_tweets[class_name]
        if class_name != title_length:
            plt.subplot(1, 3, columns.index(el) + 1)
            plt.tight_layout()
            # remove outliers
            title_stat_top = remove_outliers(elon_tweets, class_name)
            # get top PERCENTAGE%
            title_stat_top = title_stat_top.sort_values(by=[class_name],
            ascending=False).head(int(len(elon_tweets)*(PERCENTAGE/100.00)))
            if (is_grouped):

```

```

        temp = title_stat_top.groupby(pd.cut(title_stat_top[col_name],
↪bins=np.arange(0, 140+10, step=10))).count()[class_name]
        else:
            temp = title_stat_top.groupby(title_stat_top[col_name]).
↪count()[class_name]
            temp.plot.bar()
            #sns.despine()
            plt.title('{ } frequency'.format(el['column_name']));

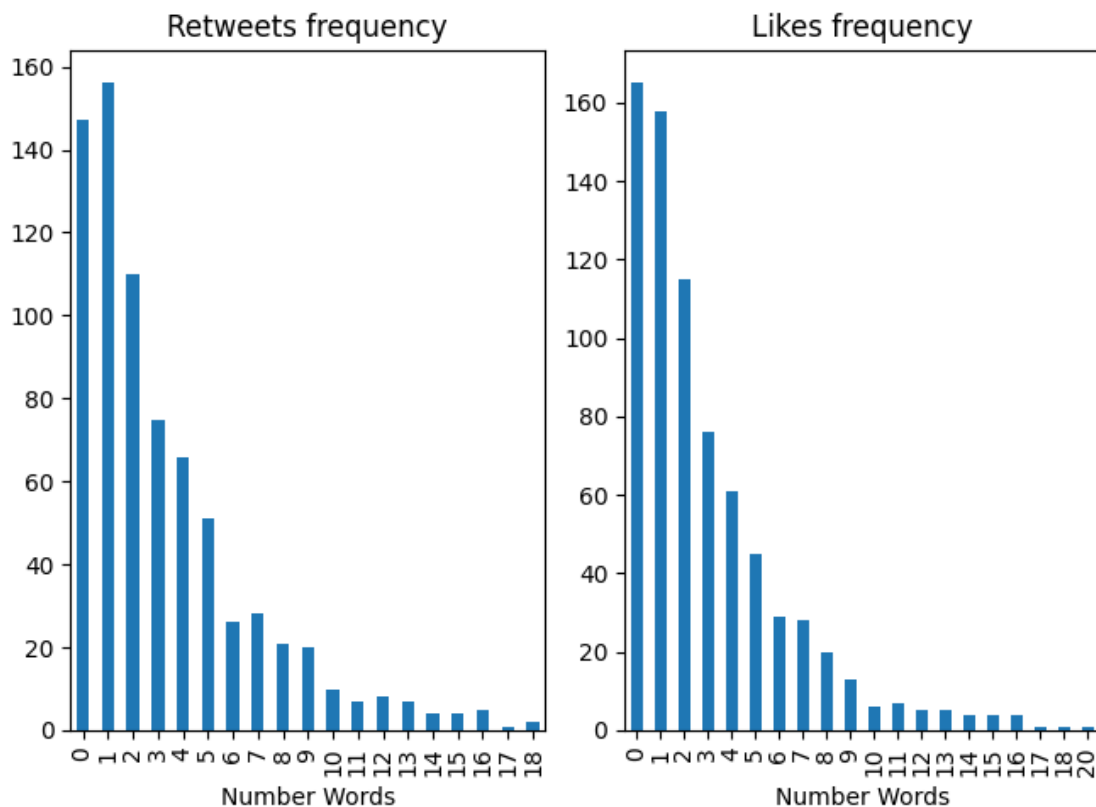
```

[238]: <https://github.com/flaviohenriquecbc/machine-learning-capstone-project/blob/master/title-success-prediction.ipynb>

```

number_words = 'Number Words'
elon_tweets[number_words] = elon_tweets['cleaned'].str.split().apply(len)
plot_top_length_performers(number_words, False)

```



[239]: <https://github.com/flaviohenriquecbc/machine-learning-capstone-project/blob/master/title-success-prediction.ipynb>

```

## plot relation between the words used in tweets and the number of retweet and
↪likes for the top PERCENTAGE% performers
def plot_top_performers(lst_col):

```

```

for el in columns:
    column = elon_tweets[el['column_name']]
    class_name = el['column_name']
    if class_name != title_length:
        plt.figure(figsize=(10, 5))
        # remove outliers
        elon_tweets_top = remove_outliers(elon_tweets, class_name)
        elon_tweets_top = elon_tweets_top.sort_values(by=[class_name],
↪ascending=False).head(int(len(elon_tweets_top)*(PERCENTAGE/100.00)))
        x = pd.DataFrame({
            col:np.repeat(elon_tweets_top[col].values,
↪elon_tweets_top[lst_col].str.len())
            for col in elon_tweets_top.columns.difference([lst_col])
        }).assign(**{lst_col:np.concatenate(elon_tweets_top[lst_col].
↪values)})[elon_tweets_top.columns.tolist()]
        temp = x.groupby(lst_col).count()[class_name].
↪sort_values(ascending=False).head(20).sort_values(ascending=True)
        temp.plot.barh()
        #sns.despine()
        plt.title('Most used {} on top {}% {}'.format(lst_col, PERCENTAGE,
↪el['column_text']));

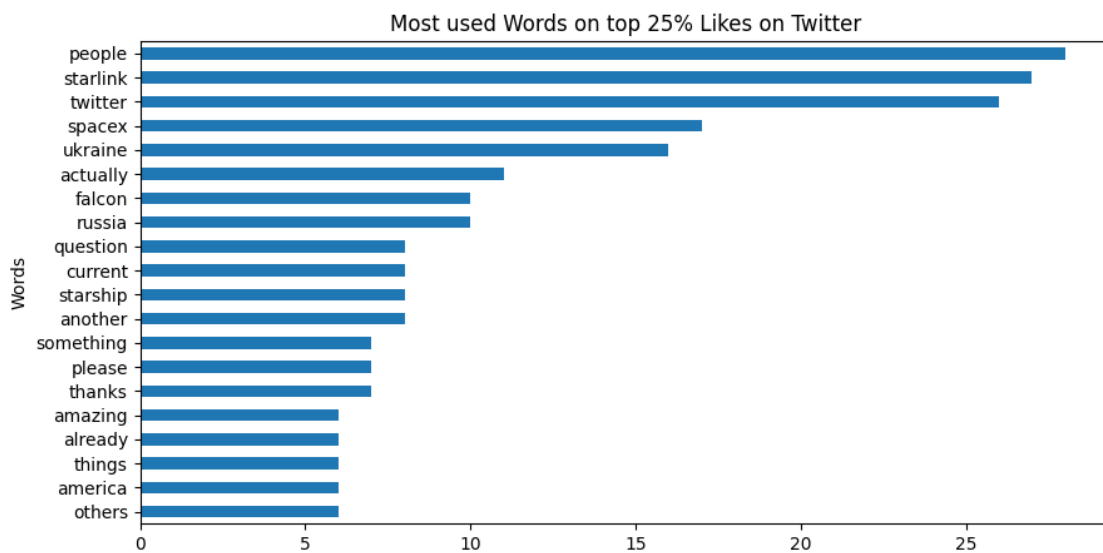
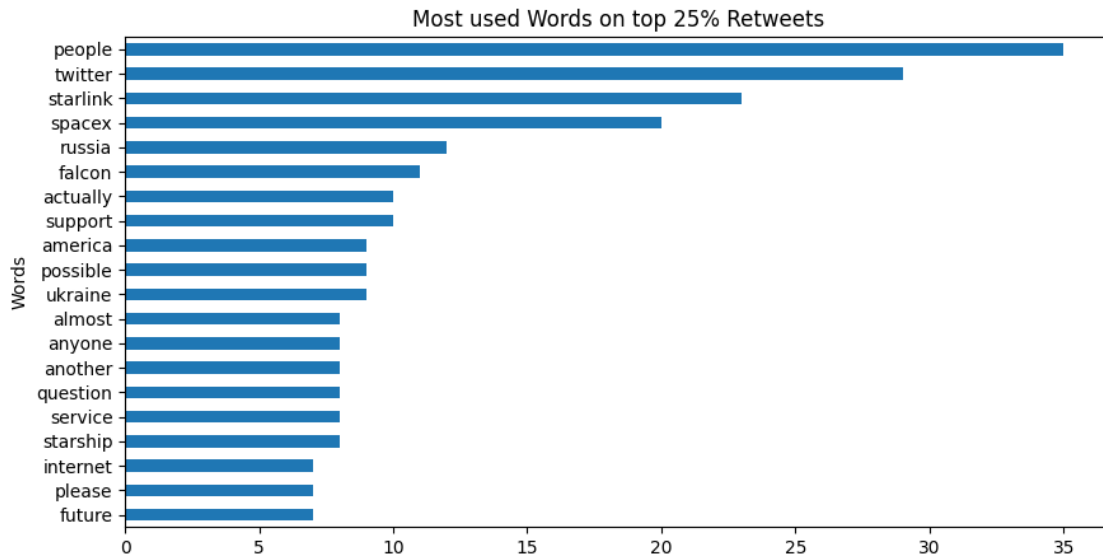
```

```

[240]: ###https://github.com/flaviohenriquecbc/machine-learning-capstone-project/blob/
↪master/title-success-prediction.ipynb

###create a column containing the words of the title
# remove special characters
temp = elon_tweets['cleaned'].str.lower().str.translate(str.maketrans('', '', '()!
↪@#$: "?,./+')).str.split()
# temp = title_stat['Title'].str.lower().str.split()
# remove common words (it, a, the, ...)
stop=set(stopwords.words('english'))
elon_tweets['Words'] = temp.apply(lambda x: [item for item in x if item not in
↪stop and not '\\\\' in item])
# title_stat['Words'] = temp
# print(stop) #uncomment to see removed words
# plot relation between the words used on medium title and the number of
↪retweet/favorite/claps. Just for the top PERCENTAGE% performers
#sns.despine()
plot_top_performers('Words')

```

```
[241]: ###https://www.earthdatascience.org/courses/use-data-open-source-python/
      ↪intro-to-apis/calculate-tweet-word-frequencies-in-python/

      # List of all words across tweets
      all_words_no_urls = list(itertools.chain(*token_cleaned))
      all_words_no_urls
```

```
[241]: ['thanks',
      'absolutely',
```

'twitter',
'advertisers',
'meeting',
'people',
'twitter',
'entering',
'twitter',
'definitely',
'closer',
'citizen',
'journalism',
'appreciated',
'prominence',
'twitter',
'nobody',
'beautiful',
'twitter',
'empowers',
'citizen',
'journalism',
'people',
'disseminate',
'without',
'establishment',
'citizen',
'journalism',
'according',
'unnamed',
'sources',
'matter',
'remain',
'anonymous',
'emerged',
'chaotic',
'global',
'politics',
'interventions',
'combustible',
'conflicts',
'sometimes',
'messaging',
'caused',
'problems',
'concerning',
'moving',
'object',
'fasting',

'double',
'vitalik',
'another',
'respect',
'authority',
'inhibits',
'innovation',
'lesson',
'longest',
'unlearn',
'twitter',
'broadly',
'inclusive',
'possible',
'serving',
'lively',
'occasionally',
'rancorous',
'debate',
'widely',
'divergent',
'beliefs',
'absolutely',
'necessary',
'important',
'exactly',
'thread',
'unless',
'damaged',
'testing',
'really',
'matters',
'booster',
'production',
'spooling',
'booster',
'incremental',
'design',
'improvements',
'bedroom',
'surfing',
'friend',
'houses',
'silicon',
'valley',
'decade',
'frankly',

'friends',
'simple',
'useful',
'technology',
'random',
'position',
'players',
'pieces',
'polytopia',
'addresses',
'limitations',
'mansion',
'douche',
'kasparov',
'almost',
'playing',
'iphone',
'otherwise',
'exciting',
'largest',
'animal',
'whales',
'bigger',
'biggest',
'dinosaur',
'welcome',
'answer',
'spacex',
'starlink',
'refused',
'provide',
'funding',
'absolutely',
'interesting',
'social',
'algorithm',
'people',
'decided',
'current',
'accepting',
'decides',
'current',
'question',
'exactly',
'hypocrites',
'neuralink',
'nothing',

'forever',
'change',
'manufacturing',
'awesome',
'happen',
'perpetuating',
'propaganda',
'concocted',
'industry',
'distract',
'trillion',
'dollars',
'subsidies',
'companies',
'receive',
'switzerland',
'belgium',
'chocolate',
'eccentric',
'british',
'artisanal',
'thread',
'actually',
'release',
'internal',
'testing',
'needed',
'public',
'although',
'improvement',
'decades',
'headline',
'misleading',
'starlink',
'obviously',
'robust',
'positioning',
'satellites',
'stronger',
'signal',
'problem',
'favorite',
'conspiracy',
'theories',
'actually',
'ailable',
'eventually',

'berlin',
'colors',
'specially',
'layers',
'giving',
'complexity',
'otherwise',
'possible',
'varies',
'spacex',
'positions',
'companies',
'recessions',
'silver',
'lining',
'companies',
'existing',
'removed',
'laboring',
'misapprehension',
'whoever',
'replaces',
'amenable',
'western',
'philosophy',
'unlikely',
'kremlin',
'olympics',
'russia',
'overrunning',
'ukraine',
'destabilize',
'entire',
'region',
'perhaps',
'russia',
'calamitous',
'defeat',
'conventional',
'warfare',
'something',
'strategically',
'critical',
'crimea',
'probability',
'nuclear',
'weapons',

'unlikely',
'closest',
'guessing',
'probably',
'spring',
'without',
'horrible',
'global',
'brutal',
'common',
'phrase',
'months',
'certainly',
'warrants',
'closer',
'scrutiny',
'bakhmut',
'pretty',
'humour',
'british',
'compromise',
'overwhelmingly',
'likely',
'outcome',
'question',
'destruction',
'happens',
'nature',
'debate',
'probable',
'outcome',
'horrible',
'attrition',
'destroys',
'ukraine',
'severely',
'damages',
'russia',
'massive',
'outcome',
'exactly',
'congrats',
'launch',
'falcon',
'record',
'launches',
'single',

'vehicle',
'exactly',
'spacex',
'active',
'satellites',
'combined',
'tracking',
'double',
'reports',
'fanfare',
'predictions',
'rarely',
'accelerating',
'sustainable',
'energy',
'primary',
'mission',
'require',
'exceptional',
'execution',
'possible',
'matter',
'awesome',
'starlink',
'airplane',
'starlink',
'another',
'airplane',
'worked',
'perfectly',
'populi',
'several',
'architectural',
'upgrades',
'release',
'sticking',
'sounds',
'definitely',
'giving',
'naming',
'internet',
'airplanes',
'accessing',
'internet',
'bottles',
'exquisite',
'perfume',

'already',
'unique',
'limited',
'edition',
'collector',
'working',
'release',
'something',
'several',
'countries',
'individuals',
'paying',
'terminals',
'thanks',
'spacex',
'losing',
'unpaid',
'service',
'related',
'enhanced',
'security',
'measures',
'cyberwar',
'defense',
'definitely',
'exactly',
'weapons',
'ideally',
'prediction',
'actually',
'summon',
'almost',
'thanks',
'optimus',
'precise',
'terminals',
'ukraine',
'present',
'paying',
'service',
'spacex',
'already',
'withdrawn',
'request',
'funding',
'analogy',
'america',

'humiliating',
'withdrawal',
'afghanistan',
'america',
'country',
'america',
'crimea',
'crucial',
'russia',
'russia',
'hawaii',
'crucial',
'america',
'probably',
'practice',
'perfect',
'russia',
'punched',
'forced',
'retreat',
'ukraine',
'russia',
'choice',
'humiliating',
'defeat',
'probability',
'obviously',
'disagree',
'probability',
'utterly',
'dangerous',
'bright',
'russia',
'crimea',
'scenario',
'comforted',
'thought',
'watching',
'mushroom',
'clouds',
'russia',
'destruction',
'defeat',
'respond',
'civilization',
'whether',
'crimea',

'absolutely',
'russia',
'russia',
'crimea',
'critical',
'national',
'security',
'importance',
'russia',
'southern',
'standpoint',
'losing',
'crimea',
'losing',
'hawaii',
'harbor',
'exactly',
'russia',
'choice',
'losing',
'crimea',
'battlefield',
'choose',
'latter',
'already',
'sanctioned',
'cutoff',
'russia',
'possible',
'russia',
'exceptionally',
'standard',
'performance',
'suspension',
'production',
'improved',
'comfort',
'without',
'affecting',
'handling',
'version',
'priority',
'standard',
'version',
'speeds',
'happening',
'satellites',

'upgrade',
'satellite',
'technology',
'speeds',
'density',
'locations',
'significant',
'increases',
'donate',
'option',
'starlink',
'normally',
'higher',
'others',
'donate',
'starlinks',
'places',
'happened',
'quieter',
'blower',
'possible',
'harder',
'sounds',
'machina',
'proceeding',
'carefully',
'starship',
'progress',
'months',
'object',
'lambos',
'rocketry',
'millipede',
'millennia',
'outfits',
'thread',
'tennis',
'company',
'clowns',
'lonely',
'dyslexic',
'secret',
'diagram',
'intersection',
'audiences',
'pretty',
'popular',

'pitching',
'crossover',
'hijinks',
'scratch',
'mostly',
'accurate',
'comments',
'thread',
'conspiracy',
'theorist',
'definitely',
'strangest',
'meetings',
'experienced',
'trying',
'hardest',
'escalate',
'situation',
'obviously',
'failing',
'damned',
'damned',
'damned',
'starship',
'starlink',
'satellites',
'working',
'hilarious',
'though',
'starlink',
'losing',
'companies',
'getting',
'billions',
'taxpayer',
'funding',
'ukraine',
'condition',
'anonymity',
'spineless',
'coward',
'ignorant',
'starlink',
'directly',
'requested',
'ukraine',
'government',

'worked',
'required',
'resources',
'lockdown',
'development',
'progress',
'resist',
'attacks',
'sophisticated',
'adversary',
'actively',
'provided',
'falcon',
'development',
'appreciated',
'certainly',
'context',
'course',
'signal',
'jamming',
'systems',
'getting',
'better',
'amount',
'spacex',
'requesting',
'battlefield',
'advantage',
'satellite',
'ironically',
'battlefields',
'signal',
'starlink',
'someday',
'humanity',
'stopped',
'pointing',
'destroy',
'almost',
'chance',
'launched',
'highest',
'curses',
'foiled',
'accurate',
'russia',
'ability',

'destroy',
'europe',
'utterly',
'nuclear',
'missiles',
'minutes',
'surprising',
'number',
'people',
'course',
'situation',
'potential',
'internet',
'towers',
'destroyed',
'starlink',
'clarifying',
'starlink',
'primary',
'communications',
'system',
'ukrainian',
'anyone',
'please',
'difference',
'warfront',
'starlink',
'system',
'working',
'warfront',
'others',
'russia',
'actively',
'trying',
'starlink',
'safeguard',
'spacex',
'diverted',
'massive',
'resources',
'towards',
'defense',
'starlink',
'concerning',
'citizens',
'nations',
'helping',

'ukraine',
'surely',
'reasonable',
'person',
'launch',
'nuclear',
'problem',
'dealing',
'reasonable',
'people',
'people',
'turbulence',
'commercial',
'airliner',
'handle',
'insane',
'amounts',
'bending',
'question',
'following',
'recommendation',
'falcon',
'probably',
'starlink',
'breakeven',
'starship',
'needed',
'addition',
'terminals',
'create',
'launch',
'maintain',
'replenish',
'satellites',
'ground',
'stations',
'telcos',
'access',
'internet',
'gateways',
'defend',
'cyberattacks',
'jamming',
'getting',
'harder',
'approaching',
'starlink',

'losing',
'insanely',
'difficult',
'communications',
'constellation',
'bankruptcy',
'company',
'starlink',
'conference',
'bankrupt',
'strange',
'nothing',
'leaked',
'competitors',
'launch',
'communications',
'lockheed',
'boeing',
'surprised',
'particular',
'individual',
'working',
'retires',
'corruption',
'finest',
'spacex',
'asking',
'recoup',
'expenses',
'can',
'not',
'existing',
'system',
'indefinitely',
'several',
'thousand',
'terminals',
'greater',
'typical',
'households',
'unreasonable',
'future',
'consciousness',
'driven',
'madness',
'starlink',
'requires',

'nothing',
'except',
'source',
'system',
'battery',
'generator',
'connectivity',
'anywhere',
'active',
'terminals',
'course',
'happens',
'almost',
'successfully',
'destroyed',
'viasat',
'ukraine',
'several',
'months',
'starlink',
'relentless',
'jamming',
'attacks',
'spacex',
'relentless',
'cyberwar',
'success',
'limited',
'starlink',
'growth',
'ukraine',
'opinion',
'saying',
'different',
'basically',
'starlink',
'roaming',
'capability',
'terminal',
'anywhere',
'counter',
'intuitive',
'actually',
'product',
'please',
'perfume',
'twitter',

'welcome',
'support',
'ukraine',
'change',
'stories',
'tomorrow',
'bottles',
'bottles',
'omnigender',
'product',
'getting',
'fragrance',
'business',
'inevitable',
'finest',
'fragrance',
'nobody',
'bremmer',
'finally',
'achieve',
'question',
'typically',
'ayahuasca',
'subtle',
'meditating',
'regularly',
'scheduled',
'programming',
'pointless',
'currently',
'overlap',
'russian',
'ukrainian',
'demands',
'spoken',
'months',
'subject',
'matter',
'talked',
'expressed',
'concerns',
'recent',
'request',
'hunter',
'strange',
'desire',
'become',

'involved',
'flying',
'strong',
'candidate',
'counter',
'intuitive',
'product',
'important',
'reading',
'anything',
'indeed',
'kinetic',
'energy',
'scales',
'square',
'velocity',
'appreciated',
'sealed',
'escape',
'shield',
'probably',
'terminals',
'ukraine',
'terminal',
'provide',
'internet',
'uplink',
'potentially',
'several',
'thousand',
'people',
'served',
'single',
'terminal',
'coming',
'trying',
'survive',
'invited',
'career',
'elsewhere',
'literal',
'outside',
'autopilot',
'office',
'inverse',
'poetry',
'making',

'metaphors',
'literal',
'interesting',
'appeared',
'identical',
'emails',
'actually',
'either',
'spaces',
'sentences',
'forming',
'binary',
'signature',
'identified',
'leaker',
'recursively',
'nuclear',
'probability',
'rising',
'rapidly',
'leaves',
'everyone',
'perhaps',
'starlink',
'provided',
'decisive',
'battlefield',
'advantage',
'trying',
'possible',
'escalate',
'pleasant',
'memory',
'double',
'header',
'agreed',
'beautiful',
'welcome',
'twitter',
'friend',
'actual',
'performance',
'difference',
'truncate',
'useless',
'zeroes',
'dragons',

'action',
'better',
'coming',
'trying',
'always',
'regular',
'contact',
'fedorov',
'reporting',
'article',
'falsely',
'claims',
'starlink',
'terminals',
'service',
'percentage',
'operation',
'spacex',
'exceed',
'happening',
'battlefield',
'classified',
'starlink',
'provide',
'majority',
'bandwidth',
'electric',
'propeller',
'mounted',
'faster',
'creative',
'design',
'generate',
'meaningful',
'thrust',
'norway',
'missed',
'satellite',
'constellation',
'capability',
'tonnage',
'capability',
'spacex',
'dragon',
'docking',
'extended',
'intense',

```
'indeed',
'personally',
'latter',
'content',
'software',
'engineering',
'server',
'operations',
'design',
'product',
'evolution',
...]
```

```
[242]: ##https://www.earthdatascience.org/courses/use-data-open-source-python/
      ↪ intro-to-apis/calculate-tweet-word-frequencies-in-python/
      # Create counter
      counts_no_urls = collections.Counter(all_words_no_urls)

      most_common = counts_no_urls.most_common(25)
```

```
[243]: clean_tweets_no_urls = pd.DataFrame(counts_no_urls.most_common(25),
      columns=['words', 'count'])

      clean_tweets_no_urls.head()
```

```
[243]:
```

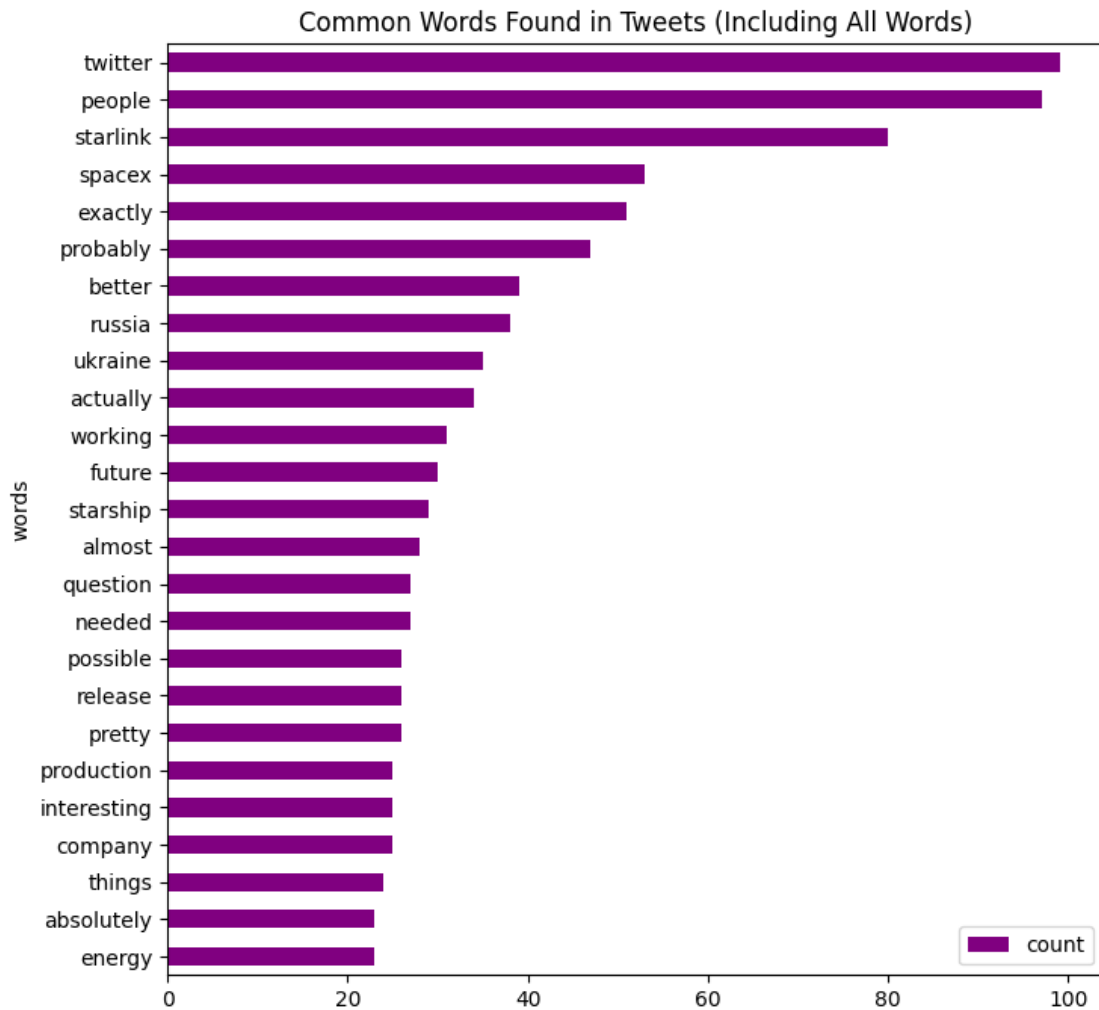
	words	count
0	twitter	99
1	people	97
2	starlink	80
3	spacex	53
4	exactly	51

```
[244]: ##https://www.earthdatascience.org/courses/use-data-open-source-python/
      ↪ intro-to-apis/calculate-tweet-word-frequencies-in-python/
      fig, ax = plt.subplots(figsize=(8, 8))

      # Plot horizontal bar graph
      clean_tweets_no_urls.sort_values(by='count').plot.barh(x='words',
      y='count',
      ax=ax,
      color="purple")

      ax.set_title("Common Words Found in Tweets (Including All Words)")

      plt.show()
```



```
[245]: collection_words = ['elon musk', 'elonmusk']
```

```
[246]: ##https://www.earthdatascience.org/courses/use-data-open-source-python/
      ↪ intro-to-apis/calculate-tweet-word-frequencies-in-python/
tweets_nsw_nc = [[w for w in word if not w in collection_words]
                  for word in cleaned]
```

```
[247]: ##https://www.earthdatascience.org/courses/use-data-open-source-python/
      ↪ intro-to-apis/calculate-tweet-word-frequencies-in-python/
# Flatten list of words in clean tweets
all_words_nsw_nc = list(itertools.chain(*tweets_nsw_nc))

# Create counter of words in clean tweets
counts_nsw_nc = collections.Counter(all_words_nsw_nc)
```



```
counts_nsw_nc.most_common(15)
```

```
[247]: [('twitter', 99),
        ('people', 97),
        ('starlink', 80),
        ('spacex', 53),
        ('exactly', 51),
        ('probably', 47),
        ('better', 39),
        ('russia', 38),
        ('ukraine', 35),
        ('actually', 34),
        ('working', 31),
        ('future', 30),
        ('starship', 29),
        ('almost', 28),
        ('question', 27)]
```

```
[248]: len(counts_nsw_nc)
```

```
[248]: 3479
```

```
[249]: clean_tweets_ncw = pd.DataFrame(counts_nsw_nc.most_common(15),
                                       columns=['words', 'count'])
clean_tweets_ncw.head()
```

```
[249]:
```

	words	count
0	twitter	99
1	people	97
2	starlink	80
3	spacex	53
4	exactly	51

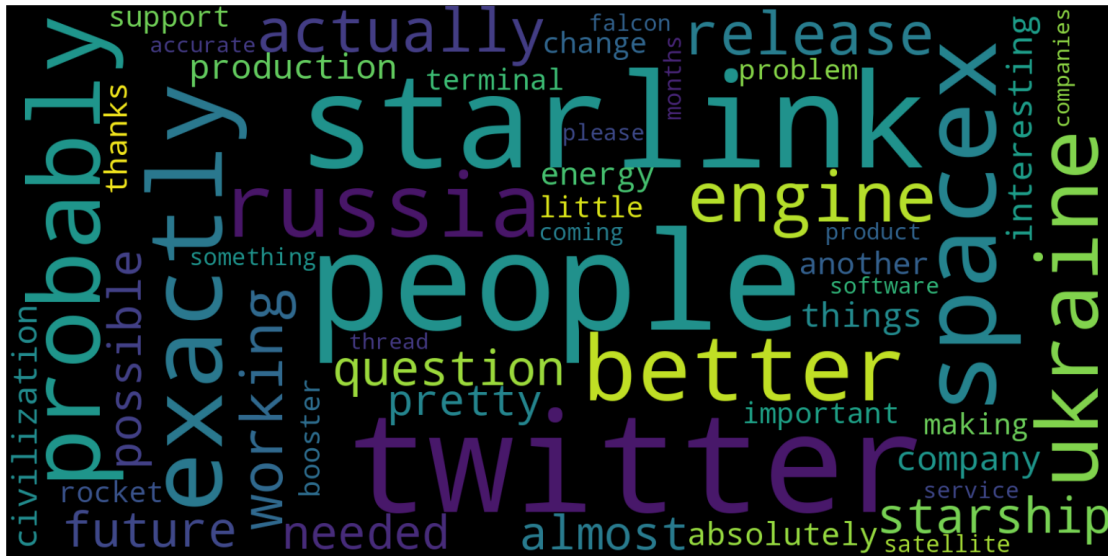
```
[250]: ### Word Cloud
```

```
[251]: string = pd.Series(all_words_no_urls).str.cat(sep=' ')
```

```
[252]: from wordcloud import WordCloud, STOPWORDS
import matplotlib.pyplot as plt
stopwords = set(STOPWORDS)
stopwords.update(["elon musk", "elonmusk"]) #adding our own stopwords
```

```
[253]: wordcloud = WordCloud(width=1600,
                             ↪stopwords=stopwords,height=800,max_font_size=200,max_words=50,collocations=False,
                             ↪background_color='black').generate(string)
plt.figure(figsize=(20,10))
plt.imshow(wordcloud, interpolation="bilinear")
```

```
plt.axis("off")
plt.show()
```



```
[254]: ### sentiment analysis
```

```
[255]: ##https://www.natasshaselvaraj.com/twitter-sentiment-analysis-with-python/
import nltk
nltk.download('vader_lexicon')
from nltk.sentiment.vader import SentimentIntensityAnalyzer

sid = SentimentIntensityAnalyzer()

list1 = []
for i in elon_tweets['cleaned']:
    list1.append((sid.polarity_scores(str(i)))['compound'])
```

```
[nltk_data] Downloading package vader_lexicon to /root/nltk_data...
```

```
[nltk_data] Package vader_lexicon is already up-to-date!
```

```
[256]: ##https://www.natasshaselvaraj.com/twitter-sentiment-analysis-with-python/  
elon_tweets['sentiment'] = pd.Series(list1)  
  
def sentiment_category(sentiment):  
    label = ''  
    if(sentiment>0):  
        label = 'positive'  
    elif(sentiment == 0):  
        label = 'neutral'
```

```

else:
    label = 'negative'
return(label)

elon_tweets['sentiment_category'] = elon_tweets['sentiment'].
    ↪ apply(sentiment_category)
elon_tweets

```

[256]:

		Tweets	Retweets	Likes	\
0	@PeterSchiff thanks		209	7021	
1	@ZubyMusic Absolutely		755	26737	
2	Dear Twitter Advertisers https://t.co/GMwHmInPAS		55927	356623	
3	@BillyM2k		802	19353	
4	Meeting a lot of cool people at Twitter today!		9366	195546	
...	
3055	@LimitingThe @baglino Just that manganese is a...		171	3173	
3056	@incentives101 @ICRicardoLara Exactly		145	4234	
3057	@ICRicardoLara Your policies are directly resp...		421	6144	
3058	@ICRicardoLara You should be voted out of office		484	7029	
3059	CB radios are free from govt/media control		11302	113429	

	Dates	Time	length	\
0	2022-10-27	16:17:00	21	
1	2022-10-27	13:19:00	21	
2	2022-10-27	13:08:00	48	
3	2022-10-27	02:32:00	11	
4	2022-10-26	21:39:00	46	
...	
3055	2022-01-27	22:01:00	135	
3056	2022-01-27	21:23:00	37	
3057	2022-01-27	21:13:00	119	
3058	2022-01-27	21:12:00	48	
3059	2022-01-27	21:00:00	42	

		cleaned	\
0		thanks	
1		absolutely	
2		twitter advertisers	
3			
4		meeting people twitter	
...		...	
3055	manganese alternative phosphorus scaling catho...		
3056		exactly	
3057	policies directly responsible outrageously ins...		
3058		office	
3059		radios control	

	tokenized_tweets	Number Words	\
0	[thanks]	1	
1	[absolutely]	1	
2	[twitter, advertisers]	2	
3	[]	0	
4	[meeting, people, twitter]	3	
...	
3055	[manganese, alternative, phosphorus, scaling, ...	7	
3056	[exactly]	1	
3057	[policies, directly, responsible, outrageously...	7	
3058	[office]	1	
3059	[radios, control]	2	

	Words	sentiment	\
0	[thanks]	0.4404	
1	[absolutely]	0.0000	
2	[twitter, advertisers]	0.0000	
3	[]	0.0000	
4	[meeting, people, twitter]	0.0000	
...	
3055	[manganese, alternative, phosphorus, scaling, ...	NaN	
3056	[exactly]	NaN	
3057	[policies, directly, responsible, outrageously...	NaN	
3058	[office]	NaN	
3059	[radios, control]	NaN	

	sentiment_category
0	positive
1	neutral
2	neutral
3	neutral
4	neutral
...	...
3055	negative
3056	negative
3057	negative
3058	negative
3059	negative

[2994 rows x 12 columns]

```
[257]: tweet_list = elon_tweets.Tweets.to_list()
```

```
[258]: #top = elon_tweets.groupby('Likes').head(50).reset_index(drop=True)

#top_tweets = elon_tweets[top]
#top_tweets
```

```
[259]: top_liked_tweets = elon_tweets.sort_values(by="Likes", ascending=False)

top_liked_tweets = top_liked_tweets[0:5]
top_liked_tweets
```

```
[259]:
```

		Tweets	Retweets	Likes \
2219	Next I'm buying Coca-Cola to put the cocaine b...	681707	4780787	
2244	I hope that even my worst critics remain on Tw...	368279	3232772	
2216	Let's make Twitter maximum fun!	194742	2650644	
2243	Yesss!!! https://t.co/OT9HzUHuh6	348158	2608578	
2215	Listen, I can't do miracles ok https://t.co/z7...	212854	2581112	

	Dates	Time	length	cleaned \
2219	2022-04-28	00:56:00	52	buying cocaine
2244	2022-04-25	16:12:00	91	critics remain twitter speech
2216	2022-04-28	01:53:00	31	twitter maximum
2243	2022-04-25	19:43:00	42	
2215	2022-04-28	01:57:00	54	listen miracles

	tokenized_tweets	Number Words \
2219	[buying, cocaine]	2
2244	[critics, remain, twitter, speech]	4
2216	[twitter, maximum]	2
2243	[]	0
2215	[listen, miracles]	2

	Words	sentiment	sentiment_category
2219	[buying, cocaine]	0.0000	neutral
2244	[critics, remain, twitter, speech]	0.0000	neutral
2216	[twitter, maximum]	-0.3612	negative
2243	[]	0.0000	neutral
2215	[listen, miracles]	0.0000	neutral

```
[260]: ###https://medium.com/@nikitasilaparasetty/
↳twitter-sentiment-analysis-for-data-science-using-python-in-2022-6d5e43f6fa6e

sentiment_objects = [TextBlob(tweet) for tweet in elon_tweets.cleaned]

sentiment_objects[0].polarity, sentiment_objects[0]
```

```
[260]: (0.2, TextBlob("thanks"))
```

```
[261]: ###https://medium.com/@nikitasilaparasetty/
↳twitter-sentiment-analysis-for-data-science-using-python-in-2022-6d5e43f6fa6e

sentiment_values = [[tweet.sentiment.polarity, str(tweet)] for tweet in
↳sentiment_objects]
```

```
# Print the value of the 0th row.
```

```
sentiment_values[0]
```

```
# Print all the sentiment values
```

```
sentiment_values[0:99]
```

```
[261]: [[0.2, 'thanks'],
        [0.2, 'absolutely'],
        [0.0, 'twitter advertisers'],
        [0.0, ''],
        [0.0, 'meeting people twitter'],
        [0.0, 'entering twitter'],
        [0.0, ''],
        [0.0, ''],
        [0.1, 'definitely closer citizen journalism appreciated prominence twitter'],
        [0.0, 'nobody'],
        [0.85,
         'beautiful twitter empowers citizen journalism people disseminate without
         establishment'],
        [0.0, 'citizen journalism'],
        [0.0, 'according unnamed sources matter remain anonymous'],
        [0.0,
         'emerged chaotic global politics interventions combustible conflicts sometimes
         messaging caused problems'],
        [0.0, ''],
        [0.0, ''],
        [0.0, ''],
        [0.0, 'concerning'],
        [0.0, 'moving object'],
        [0.0, 'fasting'],
        [0.0, 'double vitalik'],
        [0.0, ''],
        [0.0, 'another'],
        [0.0, ''],
        [0.0, ''],
        [0.0, 'respect authority inhibits innovation'],
        [0.0, ''],
        [0.0, ''],
        [0.0, 'lesson longest unlearn'],
        [0.0, ''],
        [-0.209375,
         'twitter broadly inclusive possible serving lively occasionally rancorous
         debate widely divergent beliefs'],
        [0.0, 'absolutely necessary'],
        [0.0, ''],
```

```

[0.0, ''],
[0.4, 'important'],
[0.0, ''],
[0.25, 'exactly'],
[0.0, ''],
[0.0, ''],
[0.0, 'thread'],
[0.0, ''],
[0.2,
  'unless damaged testing really matters booster production spooling booster
incremental design improvements'],
[0.0, 'bedroom surfing friend houses silicon valley decade frankly friends'],
[-0.06666666666666667,
  'simple useful technology random position players pieces polytopia addresses
limitations'],
[0.0, 'mansion douche'],
[0.0, 'kasparov almost playing iphone otherwise'],
[0.0, ''],
[0.0, ''],
[0.15, 'exciting largest animal whales bigger biggest dinosaur'],
[0.8, 'welcome'],
[0.0, 'answer spacex starlink refused provide funding'],
[0.2, 'absolutely'],
[0.5, 'interesting'],
[0.0, ''],
[0.03333333333333333, 'social algorithm people'],
[0.0, 'decided current accepting'],
[0.0, 'decides current question'],
[0.25, 'exactly hypocrites'],
[0.0, 'neuralink'],
[0.0, ''],
[0.0, ''],
[0.0, 'nothing forever'],
[0.0, ''],
[0.0, ''],
[1.0, 'change manufacturing awesome'],
[0.0, 'happen'],
[0.0, ''],
[-0.1,
  'perpetuating propaganda concocted industry distract trillion dollars
subsidies companies receive'],
[0.0, ''],
[0.0, 'switzerland belgium chocolate eccentric british artisanal'],
[0.0, ''],
[0.0, 'thread'],
[0.0, 'actually release internal testing needed public'],
[0.0, ''],

```

```

[0.0, ''],
[0.0, 'although improvement decades'],
[0.0, ''],
[0.0,
'headline misleading starlink obviously robust positioning satellites stronger
signal problem'],
[0.25, 'favorite conspiracy theories actually'],
[0.0, 'ailable'],
[0.0, ''],
[0.0, ''],
[0.0, 'eventually'],
[0.17857142857142858,
'berlin colors specially layers giving complexity otherwise possible'],
[0.0,
'varies spacex positions companies recessions silver lining companies
existing'],
[-0.09999999999999999,
'removed laboring misapprehension whoever replaces amenable western philosophy
unlikely kremlin olympics'],
[0.0, 'russia overrunning ukraine destabilize entire region perhaps'],
[-0.07142857142857142,
'russia calamitous defeat conventional warfare something strategically
critical crimea probability nuclear weapons'],
[0.0, ''],
[-0.5, 'unlikely closest'],
[0.0, 'guessing probably spring'],
[-0.5, 'without horrible global'],
[-0.5875, 'brutal common phrase months'],
[0.0, ''],
[0.0, ''],
[0.0, ''],
[0.21428571428571427, 'certainly warrants closer scrutiny'],
[0.0, ''],
[0.0, 'bakhmut']]

```

```

[262]: ##https://medium.com/@nikitasilaparasetty/
↪twitter-sentiment-analysis-for-data-science-using-python-in-2022-6d5e43f6fa6e
## Create a dataframe of each tweet against its polarity

sentiment_df = pd.DataFrame(sentiment_values, columns=["polarity", "tweet"])

sentiment_df

```

```

[262]:
   polarity tweet
0      0.20  thanks
1      0.20  absolutely
2      0.00  twitter advertisers

```


3	0.00	
4	0.00	meeting people twitter
...
2989	0.00	manganese alternative phosphorus scaling catho...
2990	0.25	exactly
2991	-0.40	policies directly responsible outrageously ins...
2992	0.00	office
2993	0.00	radios control

[2994 rows x 2 columns]

```
[263]: #https://medium.com/@nikitasilaparasetty/
↳twitter-sentiment-analysis-for-data-science-using-python-in-2022-6d5e43f6fa6e
# Save the polarity column as 'n'.

n=sentiment_df["polarity"]

# Convert this column into a series, 'm'.

m=pd.Series(n)

m
```

```
[263]: 0      0.20
1      0.20
2      0.00
3      0.00
4      0.00

...
2989   0.00
2990   0.25
2991  -0.40
2992   0.00
2993   0.00
Name: polarity, Length: 2994, dtype: float64
```

```
[264]: ##https://medium.com/@nikitasilaparasetty/
↳twitter-sentiment-analysis-for-data-science-using-python-in-2022-6d5e43f6fa6e
# Initialize variables, 'pos', 'neg', 'neu'.

pos=0
neg=0
neu=0

# Create a loop to classify the tweets as Positive, Negative, or Neutral.
# Count the number of each.
```

```

for items in m:
    if items>0:
        print("Positive")
        pos=pos+1
    elif items<0:
        print("Negative")
        neg=neg+1
    else:
        print("Neutral")
        neu=neu+1

print(pos,neg,neu)

```

Positive
 Positive
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Positive
 Neutral
 Positive
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Neutral
 Negative
 Neutral
 Neutral
 Neutral

Positive
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Positive
Positive
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral

Neutral
Positive
Neutral
Negative
Neutral
Negative
Neutral
Negative
Neutral
Negative
Negative
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Positive
Negative
Neutral
Neutral
Neutral
Positive
Negative
Positive
Neutral
Positive
Positive
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Neutral

Neutral
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Negative
Negative
Positive
Neutral
Positive
Positive
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative

Negative
Neutral
Positive
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Positive
Positive
Neutral
Negative
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Positive
Negative
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Negative
Negative
Neutral

[illegible]

Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Positive
Positive
Positive
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Negative
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Negative
Neutral
Positive
Neutral
Neutral
Neutral
Neutral

Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Positive

Neutral
Neutral
Negative
Negative
Negative
Neutral
Negative
Neutral
Neutral
Neutral
Negative
Positive
Neutral
Positive
Neutral
Positive
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Negative
Neutral
Neutral
Positive
Neutral
Positive
Negative
Neutral
Neutral
Negative
Neutral
Neutral
Positive
Positive

Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral

Negative
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Positive
Neutral
Neutral
Neutral
Negative
Neutral
Positive
Negative
Neutral
Neutral
Positive
Neutral
Positive
Neutral

Neutral
Negative
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Negative
Negative
Neutral
Positive
Neutral
Neutral
Negative
Neutral
Positive
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Negative

Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Negative
Positive
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Negative
Positive
Neutral
Neutral
Neutral
Negative
Neutral
Positive
Neutral
Neutral
Neutral
Neutral

Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Negative
Negative
Neutral
Neutral
Negative
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral

Neutral
Negative
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Positive
Positive
Positive
Negative
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral

Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Positive
Neutral
Negative
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Negative
Neutral
Neutral
Negative
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Positive
Positive
Positive
Positive
Neutral
Neutral

[illegible]

Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Positive
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral

Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral

Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Negative
Neutral
Neutral
Positive
Neutral

Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Negative
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Negative
Negative
Neutral
Neutral
Neutral
Positive
Negative
Negative
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Negative
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral

Neutral
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Negative
Negative
Neutral
Negative
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral

Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Negative
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Positive
Positive
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral

Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Negative

Negative
Neutral
Negative
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Negative
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Positive
Positive
Negative
Positive
Neutral
Neutral

Neutral
Positive
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Positive
Neutral
Negative
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Positive
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative

Neutral
Neutral
Neutral
Negative
Neutral
Positive
Negative
Neutral
Neutral
Positive
Neutral
Negative
Positive
Neutral
Neutral
Neutral
Positive
Positive
Positive
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral

Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Negative
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral

Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Negative

Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral

Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Negative
Negative
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Negative
Neutral
Positive
Negative
Neutral
Neutral
Neutral
Negative
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral

Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Negative
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Positive
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Negative
Negative

Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Positive
Positive
Neutral
Neutral
Positive
Neutral
Positive
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Negative
Neutral
Neutral
Neutral

Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Negative
Negative
Negative
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Negative
Neutral
Negative
Positive
Positive
Negative
Neutral
Neutral
Neutral
Neutral
Positive
Negative
Negative
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Negative
Negative
Neutral
Positive
Neutral
Neutral
Neutral

Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Negative

Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Positive
Neutral
Negative
Neutral
Negative
Neutral
Positive
Positive
Negative

Neutral
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Negative
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Positive
Neutral
Neutral

Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Positive
Neutral
Neutral
Positive
Neutral
Positive
Positive
Negative
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral

Positive
Neutral
Neutral
Neutral
Positive
Negative
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Negative
Positive
Neutral
Positive
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Neutral

Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Negative
Positive
Negative
Neutral
Positive
Neutral

Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Positive
Positive
Positive
Positive
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Negative
Neutral
Positive

Positive
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Negative
Neutral
Positive
Neutral
Negative
Neutral
Positive
Neutral
Positive
Neutral
Negative
Neutral
Neutral
Positive
Positive
Negative
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral

Neutral
Neutral
Negative
Negative
Neutral
Neutral
Neutral
Positive
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Positive
Neutral
Negative
Neutral
Neutral
Positive
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral

Negative
Negative
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Positive
Positive
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Negative
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Positive

Neutral
Negative
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Negative
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral

Neutral
Positive
Negative
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Negative
Neutral
Positive
Positive
Neutral
Positive
Neutral
Neutral
Negative
Neutral
Neutral
Positive

Positive
Positive
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Negative
Negative
Neutral
Neutral
Negative
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Negative
Positive
Neutral

Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Negative
Neutral
Negative
Neutral
Neutral
Neutral
Positive
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Negative
Positive
Neutral
Neutral
Positive
Neutral
Negative
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral

Neutral
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Negative
Negative
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Positive
Neutral

Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Negative
Positive
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Negative
Positive
Neutral
Neutral
Neutral
Neutral
Negative
Positive
Neutral
Positive
Neutral
Neutral
Neutral

[illegible]

Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Negative
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Negative
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral

Neutral
Neutral
Negative
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral

Positive
Positive
Neutral
Positive
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Positive
Neutral
Positive
Negative
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral

Negative
Positive
Neutral
Negative
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Negative
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Negative
Positive
Neutral
Positive
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Positive
Neutral
Neutral
Positive
Neutral

Neutral
Neutral
Negative
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Positive
Negative
Neutral
Negative
Neutral
Positive
Neutral
Positive
Negative
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Positive
Neutral

Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Negative
Positive
Positive
Neutral
Neutral
Negative
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Negative
Positive
Negative
Positive
Neutral
Positive
Positive
Negative
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Negative
Neutral
Neutral
Positive
Neutral
Negative
Neutral
Neutral
Neutral

Neutral
Positive
Neutral
Neutral
Positive
Positive
Neutral
Neutral
Positive
Negative
Neutral
Negative
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Positive
Negative
Negative
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral

Neutral
Neutral
Neutral
Positive
Neutral
Positive
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral

[illegible]

Negative
Negative
Positive
Positive
Neutral
Neutral
Neutral
Positive
Neutral
Negative
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Negative
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral

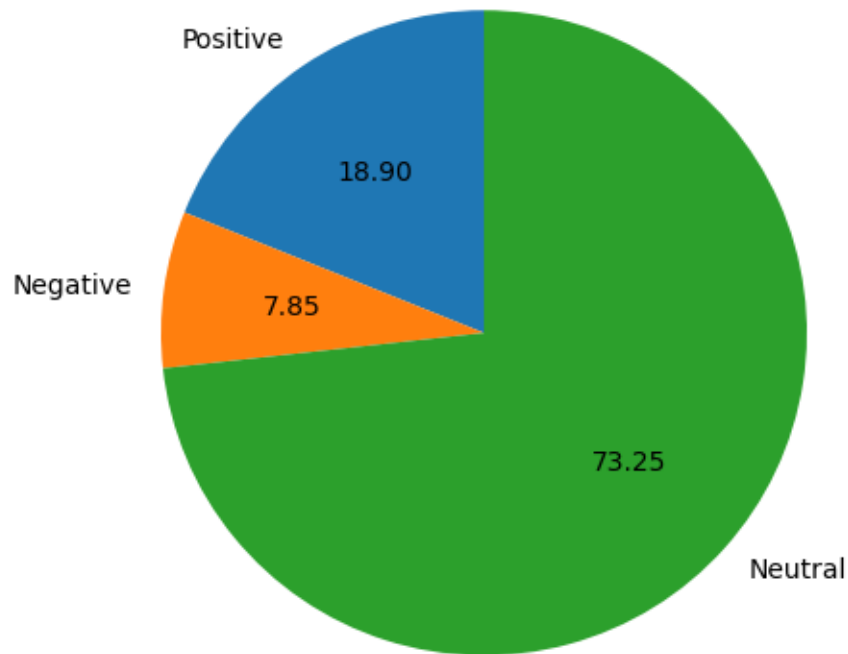
Positive
Neutral
Negative
Neutral
Neutral
Positive
Neutral
Neutral
Negative
Negative
Neutral
Neutral
Positive
Neutral
Neutral
Positive
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Neutral
Neutral
Negative
Positive
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral
Neutral

Positive
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Neutral
Neutral
Neutral
Negative
Neutral
Neutral
Positive
Neutral
Neutral
Neutral
Neutral
Positive
Neutral
Positive
Negative
Neutral
Neutral
566 235 2193

```
[265]: ###https://medium.com/@nikitasilaparasetty/  
        ↪twitter-sentiment-analysis-for-data-science-using-python-in-2022-6d5e43f6fa6e  
  
pieLabels=["Positive","Negative","Neutral"]  
  
populationShare=[pos,neg,neu]  
  
figureObject, axesObject = plt.subplots()  
  
axesObject.pie(populationShare,labels=pieLabels,autopct='%1.2f',startangle=90)  
  
axesObject.axis('equal')
```



```
plt.show()
```



```
[266]: ##### Machine Learning  
      ## Predicting number of likes
```

```
[267]: from urllib.parse import urlparse
```

```
[268]: def count_urls(row) :  
      text = row['Tweets']  
      parsed = urlparse(text)  
      if parsed.scheme and parsed.netloc:  
          return 1  
      else:  
          return 0  
  
      elon_tweets['no_of_urls'] = elon_tweets.apply(count_urls, axis = 1)  
  
      def contains_image(row):  
          text = row['Tweets']  
          if 'pic.twitter.com' in text:  
              return True  
          else:  
              return False
```

```
elon_tweets['contains_image'] = elon_tweets.apply(contains_image, axis = 1)

print(elon_tweets)
```

		Tweets	Retweets	Likes \
0	@PeterSchiff	thanks	209	7021
1	@ZubyMusic	Absolutely	755	26737
2	Dear Twitter Advertisers	https://t.co/GMwHmInPAS	55927	356623
3		@BillyM2k	802	19353
4	Meeting a lot of cool people at Twitter today!		9366	195546
...
3055	@LimitingThe @baglino	Just that manganese is a...	171	3173
3056	@incentives101 @ICRicardoLara	Exactly	145	4234
3057	@ICRicardoLara	Your policies are directly resp...	421	6144
3058	@ICRicardoLara	You should be voted out of office	484	7029
3059	CB radios are free from govt/media control		11302	113429

	Dates	Time	length \
0	2022-10-27	16:17:00	21
1	2022-10-27	13:19:00	21
2	2022-10-27	13:08:00	48
3	2022-10-27	02:32:00	11
4	2022-10-26	21:39:00	46
...
3055	2022-01-27	22:01:00	135
3056	2022-01-27	21:23:00	37
3057	2022-01-27	21:13:00	119
3058	2022-01-27	21:12:00	48
3059	2022-01-27	21:00:00	42

		cleaned \
0		thanks
1		absolutely
2		twitter advertisers
3		
4		meeting people twitter
...		...
3055	manganese alternative phosphorus scaling catho...	
3056		exactly
3057	policies directly responsible outrageously ins...	
3058		office
3059		radios control

	tokenized_tweets	Number Words \
0	[thanks]	1
1	[absolutely]	1

2	[twitter, advertisers]	2
3	[]	0
4	[meeting, people, twitter]	3
...
3055	[manganese, alternative, phosphorus, scaling, ...	7
3056	[exactly]	1
3057	[policies, directly, responsible, outrageously...	7
3058	[office]	1
3059	[radios, control]	2

	Words	sentiment	\
0	[thanks]	0.4404	
1	[absolutely]	0.0000	
2	[twitter, advertisers]	0.0000	
3	[]	0.0000	
4	[meeting, people, twitter]	0.0000	
...	
3055	[manganese, alternative, phosphorus, scaling, ...	NaN	
3056	[exactly]	NaN	
3057	[policies, directly, responsible, outrageously...	NaN	
3058	[office]	NaN	
3059	[radios, control]	NaN	

	sentiment_category	no_of_urls	contains_image
0	positive	0	False
1	neutral	0	False
2	neutral	0	False
3	neutral	0	False
4	neutral	0	False
...
3055	negative	0	False
3056	negative	0	False
3057	negative	0	False
3058	negative	0	False
3059	negative	0	False

[2994 rows x 14 columns]

```
[269]: ###https://stackoverflow.com/questions/45306988/
      ↪column-of-lists-convert-list-to-string-as-a-new-column
elon_tweets['cleaned_string'] = ','.join(map(str, l)) for l in_
      ↪elon_tweets['tokenized_tweets']]
elon_tweets
```

		Tweets	Retweets	Likes	\
0	@PeterSchiff	thanks	209	7021	
1	@ZubyMusic	Absolutely	755	26737	

2	Dear Twitter Advertisers https://t.co/GMwHmInPAS	55927	356623
3	@BillyM2k	802	19353
4	Meeting a lot of cool people at Twitter today!	9366	195546
...
3055	@LimitingThe @baglino Just that manganese is a...	171	3173
3056	@incentives101 @ICRicardoLara Exactly	145	4234
3057	@ICRicardoLara Your policies are directly resp...	421	6144
3058	@ICRicardoLara You should be voted out of office	484	7029
3059	CB radios are free from govt/media control	11302	113429

	Dates	Time	length	\
0	2022-10-27	16:17:00	21	
1	2022-10-27	13:19:00	21	
2	2022-10-27	13:08:00	48	
3	2022-10-27	02:32:00	11	
4	2022-10-26	21:39:00	46	
...	
3055	2022-01-27	22:01:00	135	
3056	2022-01-27	21:23:00	37	
3057	2022-01-27	21:13:00	119	
3058	2022-01-27	21:12:00	48	
3059	2022-01-27	21:00:00	42	

	cleaned	\
0	thanks	
1	absolutely	
2	twitter advertisers	
3		
4	meeting people twitter	
...	...	
3055	manganese alternative phosphorus scaling catho...	
3056	exactly	
3057	policies directly responsible outrageously ins...	
3058	office	
3059	radios control	

	tokenized_tweets	Number Words	\
0	[thanks]	1	
1	[absolutely]	1	
2	[twitter, advertisers]	2	
3	[]	0	
4	[meeting, people, twitter]	3	
...	
3055	[manganese, alternative, phosphorus, scaling, ...	7	
3056	[exactly]	1	
3057	[policies, directly, responsible, outrageously...	7	
3058	[office]	1	

3059	[radios, control]	2
------	-------------------	---

	Words	sentiment \
0	[thanks]	0.4404
1	[absolutely]	0.0000
2	[twitter, advertisers]	0.0000
3	[]	0.0000
4	[meeting, people, twitter]	0.0000
...
3055	[manganese, alternative, phosphorus, scaling, ...	NaN
3056	[exactly]	NaN
3057	[policies, directly, responsible, outrageously...	NaN
3058	[office]	NaN
3059	[radios, control]	NaN

	sentiment_category	no_of_urls	contains_image \
0	positive	0	False
1	neutral	0	False
2	neutral	0	False
3	neutral	0	False
4	neutral	0	False
...
3055	negative	0	False
3056	negative	0	False
3057	negative	0	False
3058	negative	0	False
3059	negative	0	False

	cleaned_string
0	thanks
1	absolutely
2	twitter,advertisers
3	
4	meeting,people,twitter
...	...
3055	manganese,alternative,phosphorus,scaling,catho...
3056	exactly
3057	policies,directly,responsible,outrageously,ins...
3058	office
3059	radios,control

[2994 rows x 15 columns]

[270]: `elon_tweets.info()`

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2994 entries, 0 to 3059
```

Data columns (total 15 columns):

#	Column	Non-Null Count	Dtype
0	Tweets	2994 non-null	object
1	Retweets	2994 non-null	int64
2	Likes	2994 non-null	int64
3	Dates	2994 non-null	object
4	Time	2994 non-null	object
5	length	2994 non-null	int64
6	cleaned	2994 non-null	object
7	tokenized_tweets	2994 non-null	object
8	Number Words	2994 non-null	int64
9	Words	2994 non-null	object
10	sentiment	2930 non-null	float64
11	sentiment_category	2994 non-null	object
12	no_of_urls	2994 non-null	int64
13	contains_image	2994 non-null	bool
14	cleaned_string	2994 non-null	object

dtypes: bool(1), float64(1), int64(5), object(8)
memory usage: 353.8+ KB

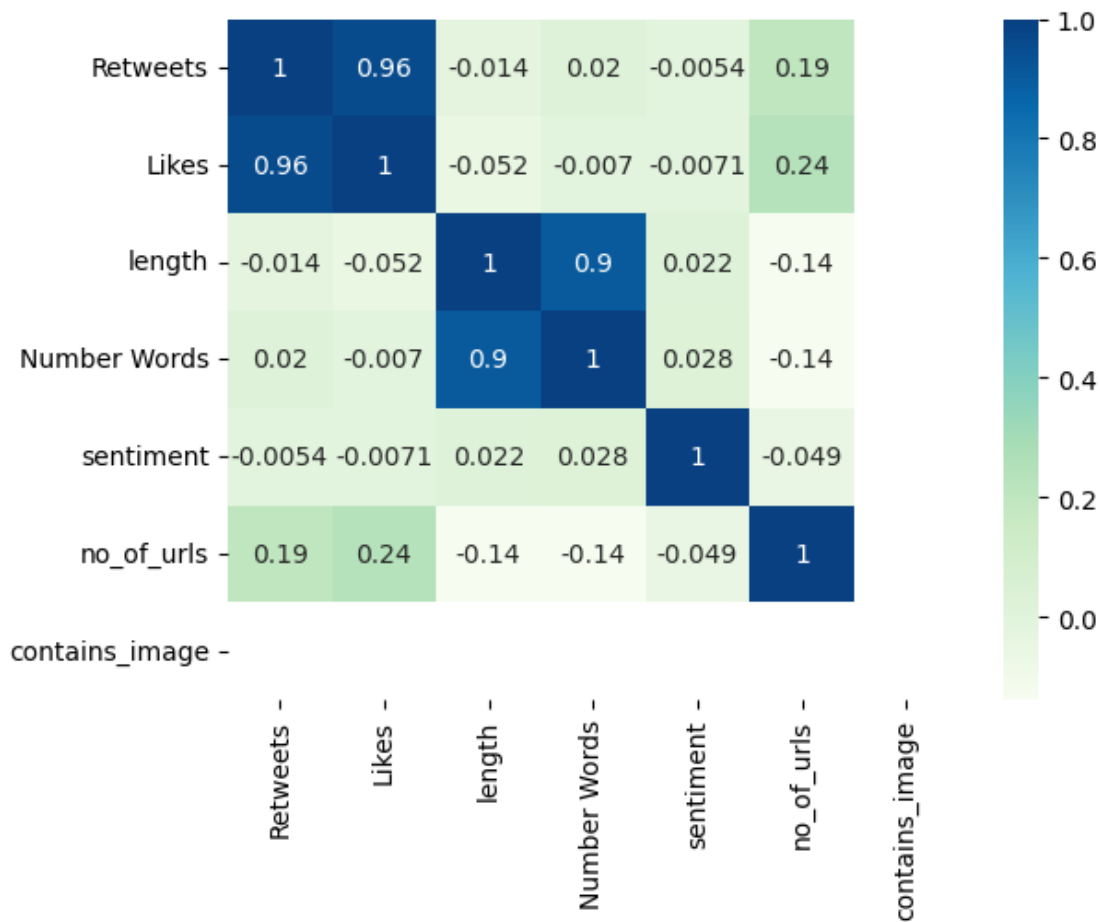
```
[271]: elon_tweets = elon_tweets.dropna()
```

```
[272]: #####https://www.visual-design.net/post/
      ↪ semi-automated-exploratory-data-analysis-process-in-python
      ## correlation between all numerical data
      correlation_num = elon_tweets.corr()
      sns.heatmap(correlation_num, cmap = "GnBu", annot = True)
```

<ipython-input-272-1d7c9e83deb3>:3: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.

```
correlation_num = elon_tweets.corr()
```

```
[272]: <Axes: >
```



```
[273]: ##https://github.com/niks92/Twitter-sentiment-analysis/blob/master/
      ↪Sentiment_vaderSentiment.ipynb
from sklearn.feature_extraction.text import CountVectorizer
# Define the cleaning pipeline we defined earlier
vectorizer = CountVectorizer()
tweets_countvectorizer = vectorizer.fit_transform(elon_tweets['cleaned'])
```

```
[274]: print(vectorizer.get_feature_names_out())
```

```
['ability' 'absolute' 'absolutely' ... 'zeroes' 'zillion' 'zukunft']
```

```
[275]: print(tweets_countvectorizer.toarray())
```

```
[[0 0 0 ... 0 0 0]
 [0 0 1 ... 0 0 0]
 [0 0 0 ... 0 0 0]
 ...
 [0 0 0 ... 0 0 0]]
```

```
[0 0 0 ... 0 0 0]
[0 0 0 ... 0 0 0]]
```

```
[276]: tweets_countvectorizer.shape
```

```
[276]: (2930, 3427)
```

```
[277]: x = pd.DataFrame(tweets_countvectorizer.toarray())
x
```

```
[277]:
```

	0	1	2	3	4	5	6	7	8	9	...	3417	\
0	0	0	0	0	0	0	0	0	0	0	0	...	0
1	0	0	1	0	0	0	0	0	0	0	0	...	0
2	0	0	0	0	0	0	0	0	0	0	0	...	0
3	0	0	0	0	0	0	0	0	0	0	0	...	0
4	0	0	0	0	0	0	0	0	0	0	0	...	0
...
2925	0	0	0	0	0	0	0	0	0	0	0	...	0
2926	0	0	0	0	0	0	0	0	0	0	0	...	0
2927	0	0	0	0	0	0	0	0	0	0	0	...	0
2928	0	0	0	0	0	0	0	0	0	0	0	...	0
2929	0	0	0	0	0	0	0	0	0	0	0	...	0
...
3418	0	0	0	0	0	0	0	0	0	0	0	...	0
3419	0	0	0	0	0	0	0	0	0	0	0	...	0
3420	0	0	0	0	0	0	0	0	0	0	0	...	0
3421	0	0	0	0	0	0	0	0	0	0	0	...	0
3422	0	0	0	0	0	0	0	0	0	0	0	...	0
3423	0	0	0	0	0	0	0	0	0	0	0	...	0
3424	0	0	0	0	0	0	0	0	0	0	0	...	0
3425	0	0	0	0	0	0	0	0	0	0	0	...	0
3426	0	0	0	0	0	0	0	0	0	0	0	...	0
...
2925	0	0	0	0	0	0	0	0	0	0	0	...	0
2926	0	0	0	0	0	0	0	0	0	0	0	...	0
2927	0	0	0	0	0	0	0	0	0	0	0	...	0
2928	0	0	0	0	0	0	0	0	0	0	0	...	0
2929	0	0	0	0	0	0	0	0	0	0	0	...	0

```
[2930 rows x 3427 columns]
```

```
[278]: y = elon_tweets['Likes']
y
```

```
[278]: 0      7021
1     26737
2    356623
3     19353
4    195546
...
```



```

2989      5617
2990    155063
2991      5432
2992    23318
2993      6709
Name: Likes, Length: 2930, dtype: int64

```

```

[280]: X = elon_tweets[['Dates', 'Time', 'contains_image', 'cleaned']]
H = elon_tweets[['Retweets', 'Dates', 'Time',
↳ 'contains_image', 'cleaned', 'no_of_urls', 'Number Words', 'length', 'sentiment'
↳ ]]

```

```

[284]: def to_num(contains_image):
        label = ''
        if(contains_image == 'False'):
            label = 0
        else:
            label = 1
        return(label)

X['contains_image'] = X['contains_image'].apply(to_num)

```

<ipython-input-284-4ef7f3725633>:9: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
X['contains_image'] = X['contains_image'].apply(to_num)

```

[329]: ###https://www.projectpro.io/recipes/
        ↳ convert-categorical-variables-into-numerical-variables-in-python
Z = pd.get_dummies(data=X, drop_first=True)
Z['Retweets'] = elon_tweets['Retweets']
Z['no_of_urls'] = elon_tweets['no_of_urls']
Z['Number words'] = elon_tweets['Number Words']
Z['length'] = elon_tweets['length']
Z['sentiment'] = elon_tweets['sentiment']
Z

```

```

[329]:
contains_image  Dates_2022-01-04  Dates_2022-01-05  Dates_2022-01-06  \
0              1                  0                  0                  0
1              1                  0                  0                  0
2              1                  0                  0                  0
3              1                  0                  0                  0
4              1                  0                  0                  0
...           ...                  ...                  ...                  ...

```

2989	1	0	0	0
2990	1	0	0	0
2991	1	0	0	0
2992	1	0	0	0
2993	1	0	0	0

	Dates_2022-01-08	Dates_2022-01-09	Dates_2022-01-10	Dates_2022-02-02 \
0	0	0	0	0
1	0	0	0	0
2	0	0	0	0
3	0	0	0	0
4	0	0	0	0
...
2989	0	0	0	1
2990	0	0	0	1
2991	0	0	0	1
2992	0	0	0	1
2993	0	0	0	1

	Dates_2022-02-03	Dates_2022-02-04	... cleaned_younger \
0	0	0	...
1	0	0	...
2	0	0	...
3	0	0	...
4	0	0	...
...
2989	0	0	...
2990	0	0	...
2991	0	0	...
2992	0	0	...
2993	0	0	...

	cleaned_youtube explain former members	cleaned_youtube nonstop \
0	0	0
1	0	0
2	0	0
3	0	0
4	0	0
...
2989	0	0
2990	0	0
2991	0	0
2992	0	0
2993	0	0

	cleaned_youtube playing annoying algorithm convinced audible \
0	0

1	0
2	0
3	0
4	0
...	...
2989	0
2990	0
2991	0
2992	0
2993	0

	cleaned_youtube	subscriptions	Retweets	no_of_urls	Number	words	\
0		0	209	0		1	
1		0	755	0		1	
2		0	55927	0		2	
3		0	802	0		0	
4		0	9366	0		3	
...			
2989		0	246	0		1	
2990		0	7909	0		3	
2991		0	235	0		1	
2992		0	1193	0		3	
2993		0	220	0		2	

	length	sentiment
0	21	0.4404
1	21	0.0000
2	48	0.0000
3	11	0.0000
4	46	0.0000
...
2989	53	0.0000
2990	37	0.0000
2991	67	0.0258
2992	75	0.0000
2993	63	0.0000

[2930 rows x 3308 columns]

```
[170]: #input = input[0:2930]
```

```
[73]: #input.info
```

```
[330]: Y = elon_tweets['Likes']
Y
```

```
[330]: 0          7021
      1          26737
      2         356623
      3          19353
      4         195546
      ...
      2989         5617
      2990        155063
      2991         5432
      2992        23318
      2993         6709
      Name: Likes, Length: 2930, dtype: int64
```

```
[338]: random.seed(14)
      ##https://medium.com/analytics-vidhya/
      ↪implementing-linear-regression-using-sklearn-76264a3c073c
      from sklearn.model_selection import train_test_split
      X_train, X_test, y_train, y_test = train_test_split(Z, Y, test_size=0.4,
      ↪random_state=101)
      print(X_train.shape)
      print(X_test.shape)
      print(y_train.shape)
      print(y_test.shape)
```

```
(1758, 3308)
(1172, 3308)
(1758,)
(1172,)
```

```
[339]: from sklearn.linear_model import LinearRegression
      model = LinearRegression()
      model.fit(X_train,y_train)
```

```
[339]: LinearRegression()
```

```
[340]: # print the intercept
      print(model.intercept_)
```

```
10907.3279860791
```

```
[349]: coeff_parameter = pd.DataFrame(model.coef_,Z.columns,columns=['Coefficient'])
      coeff_parameter
```

```
[349]:          Coefficient
contains_image    -0.001143
Dates_2022-01-04 -2563.816818
Dates_2022-01-05 -23679.543045
```

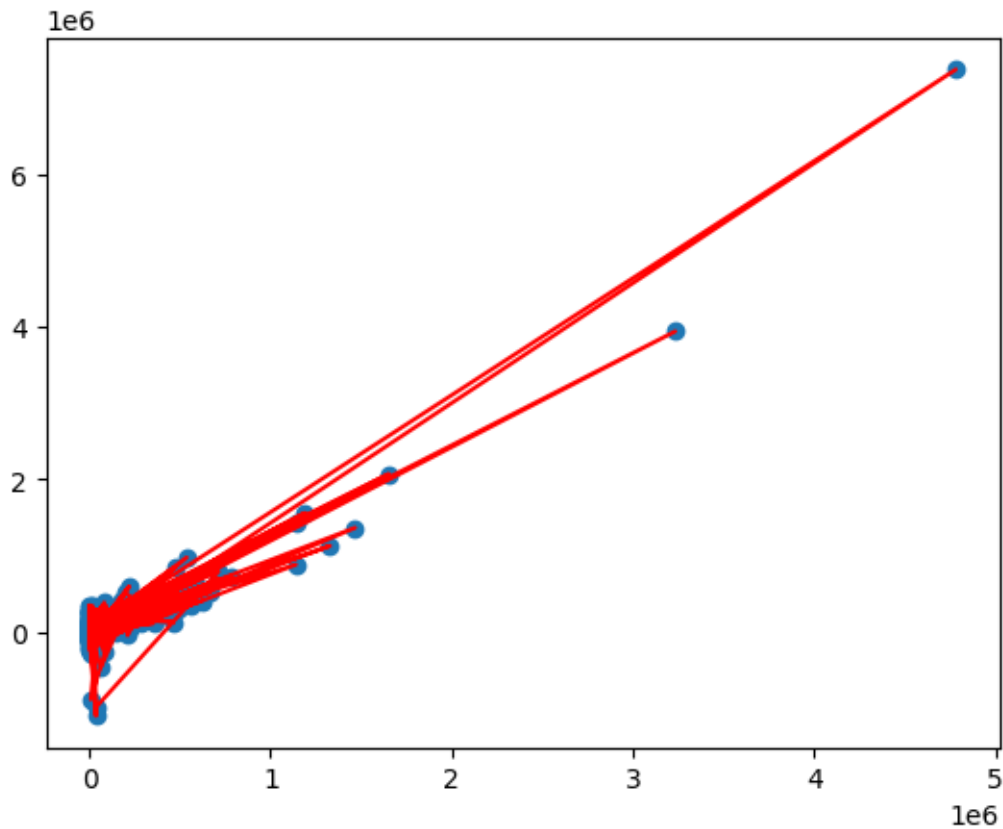
```
Dates_2022-01-06 19514.118596
Dates_2022-01-08 35788.434989
...
Retweets 10.620495
no_of_urls 17111.278039
Number words -1864.678130
length 54.084301
sentiment 1246.610277
```

```
[3308 rows x 1 columns]
```

```
[343]: predictions = model.predict(X_test)
       predictions
```

```
[343]: array([ 93739.4901209 ,  3891.26653225, 11058.85370781, ...,
        3743.20361381, 177990.97044688, -81643.82824398])
```

```
[344]: ##https://towardsdatascience.com/
       ↳linear-regression-in-6-lines-of-python-5e1d0cd05b8d
       plt.scatter(y_test, predictions)
       plt.plot(y_test, predictions, color='red')
       plt.show()
```



```
[345]: ###https://www.analyticsvidhya.com/blog/2021/05/
        ↪multiple-linear-regression-using-python-and-scikit-learn/
        ↪#How\_to\_Train\_a\_Model\_for\_Multiple\_Linear\_Regression?

from sklearn.metrics import r2_score

from sklearn.metrics import mean_squared_error

score=r2_score(y_test, predictions)

print("r2 socre is ",score)

print("mean_sqrd_error is==",mean_squared_error(y_test, predictions))

print("root_mean_squared error of is==",np.sqrt(mean_squared_error(y_test,
        ↪predictions)))
```

```
r2 socre is  0.6391935102722137
mean_sqrd_error is== 16921946521.435667
root_mean_squared error of is== 130084.38231177356
```

```
[346]: from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
from sklearn import metrics
from pandas import DataFrame, Series
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
import matplotlib
import matplotlib.pyplot as plt
from sklearn import linear_model
from sklearn.model_selection import train_test_split, cross_val_score,
        ↪cross_val_predict
import missingno as msno # plotting missing data
import seaborn as sns # plotting library
from sklearn import svm
```

```
[347]: ###https://dibyendudeb.com/
        ↪comparing-machine-learning-regression-models-using-python/
        ↪#Comparing\_regression\_models

dt_regressor = DecisionTreeRegressor(random_state = 0)
dt_regressor.fit(X_train,y_train)
#Predicting using test set
y_pred = dt_regressor.predict(X_test)
mae=metrics.mean_absolute_error(y_test, y_pred)
```

```

mse=metrics.mean_squared_error(y_test, y_pred)
# Printing the metrics
print('Decision Tree Regression Accuracy: ', dt_regressor.score(X_test,y_test))
print('R2 square:',metrics.r2_score(y_test, y_pred))
print('MAE: ', mae)
print('MSE: ', mse)

```

Decision Tree Regression Accuracy: 0.8299375385381481
R2 square: 0.8299375385381481
MAE: 21388.60324232082
MSE: 7975987018.227816

```

[348]: ##https://dibyendudeb.com/
      ↪comparing-machine-learning-regression-models-using-python/
      ↪#Comparing_regression_models

rf_regressor = RandomForestRegressor(n_estimators = 300 , random_state = 0)
rf_regressor.fit(X_train,y_train)
#Predicting the SalePrices using test set
y_pred = rf_regressor.predict(X_test)
mae=metrics.mean_absolute_error(y_test, y_pred)
mse=metrics.mean_squared_error(y_test, y_pred)
# Printing the metrics
print('Random Forest Regression Accuracy: ', rf_regressor.score(X_test,y_test))
print('R2 square:',metrics.r2_score(y_test, y_pred))
print('MAE: ', mae)
print('MSE: ', mse)

```

Random Forest Regression Accuracy: 0.8359165680175112
R2 square: 0.8359165680175112
MAE: 18951.041373720138
MSE: 7695568511.409369

```

[364]: import os
      os.chdir("/content/drive/My Drive/Colab Notebooks")
      !ls

```

'\$gdrive_home'	DataMiningAsssign1.pdf
colab_pdf.py	Sentiment_Analysis_Twitter.ipynb
'Copy of Data Mining Assignment 2'	title-success-prediction.ipynb
'Copy of Data Mining Assignment 2 (1)'	Untitled0.ipynb
'Copy of Data Mining Assignment 2 (2)'	Untitled1.ipynb
'Data Mining Assignment 2.ipynb'	Untitled2.ipynb
'DataMiningAsssign1 (1).ipynb'	Untitled3.ipynb
DataMiningAsssign1.ipynb	

```
[365]: !wget -nc https://raw.githubusercontent.com/brpy/colab-pdf/master/colab_pdf.py
from colab_pdf import colab_pdf
colab_pdf('Data Mining Assignment 2')
```

File 'colab_pdf.py' already there; not retrieving.

```
-----
ValueError                                Traceback (most recent call last)
<ipython-input-365-fbd6557fd239> in <cell line: 3>()
      1 get_ipython().system('wget -nc https://raw.githubusercontent.com/brpy/
    ↪ colab-pdf/master/colab_pdf.py')
      2 from colab_pdf import colab_pdf
----> 3 colab_pdf('Data Mining Assignment 2')

/content/colab_pdf.py in colab_pdf(file_name, notebookpath)
     20     # Check if the notebook exists in the Drive.
     21     if not os.path.isfile(os.path.join(notebookpath, file_name)):
--> 22         raise ValueError(f"file '{file_name}' not found in path
    ↪ '{notebookpath}'.")
     23
     24     # Installing all the recommended packages.

ValueError: file 'Data Mining Assignment 2' not found in path '/content/drive/
    ↪ MyDrive/Colab Notebooks/'.
```

```
[372]: !jupyter nbconvert --to html /content/Data Mining Assignment 2.ipynb"
```

```
/bin/bash: -c: line 0: unexpected EOF while looking for matching `"'
/bin/bash: -c: line 1: syntax error: unexpected end of file
```

```
[373]: !apt-get install texlive texlive-xetex texlive-latex-extra pandoc
!pip install py pandoc
```

```
Reading package lists... Done
Building dependency tree
Reading state information... Done
pandoc is already the newest version (2.5-3build2).
pandoc set to manually installed.
The following additional packages will be installed:
  dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono
  fonts-texgyre fonts-urw-base35 javascript-common libapache-pom-java
  libcommons-logging-java libcommons-parent-java libfontbox-java libfontenc1
  libgs9 libgs9-common libharfbuzz-icu0 libidn11 libijs-0.35 libjbig2dec0
  libjs-jquery libkpathsea6 libpdfbox-java libptexenc1 libruby2.7 libsynchronex2
  libteckit0 libtexlua53 libtexluajit2 libwoff1 libzzip-0-13 lmodern
  poppler-data preview-latex-style rake ruby ruby-minitest ruby-net-telnet
```


ruby-power-assert ruby-test-unit ruby-xmlrpc ruby2.7 rubygems-integration
tlutils teckit tex-common tex-gyre texlive-base texlive-binaries
texlive-fonts-recommended texlive-latex-base texlive-latex-recommended
texlive-pictures texlive-plain-generic tipa xfonts-encodings xfonts-utils

Suggested packages:

fonts-noto fonts-freefont-otf | fonts-freefont-ttf apache2 | lighttpd
| httpd libavalon-framework-java libcommons-logging-java-doc
libexcalibur-logkit-java liblog4j1.2-java poppler-utils ghostscript
fonts-japanese-mincho | fonts-ipafont-mincho fonts-japanese-gothic
| fonts-ipafont-gothic fonts-arphic-ukai fonts-arphic-uming fonts-nanum ri
ruby-dev bundler debhelper gv | postscript-viewer perl-tk xpdf | pdf-viewer
xzdec texlive-fonts-recommended-doc texlive-latex-base-doc python3-pygments
icc-profiles libfile-which-perl libspreadsheet-parseexcel-perl
texlive-latex-extra-doc texlive-latex-recommended-doc texlive-luatex
texlive-pstricks dot2tex prerex ruby-tcltk | libtcltk-ruby
texlive-pictures-doc vprerex default-jre-headless

The following NEW packages will be installed:

dvisvgm fonts-droid-fallback fonts-lato fonts-lmodern fonts-noto-mono
fonts-texgyre fonts-urw-base35 javascript-common libapache-pom-java
libcommons-logging-java libcommons-parent-java libfontbox-java libfontenc1
libgs9 libgs9-common libharfbuzz-icu0 libidn11 libijs-0.35 libjbig2dec0
libjs-jquery libkpathsea6 libpdfbox-java libptexenc1 libruby2.7 libsyntax2
libteckit0 libtexlua53 libtexluajit2 libwoff1 libzip-0-13 lmodern
poppler-data preview-latex-style rake ruby ruby-minitest ruby-net-telnet
ruby-power-assert ruby-test-unit ruby-xmlrpc ruby2.7 rubygems-integration
tlutils teckit tex-common tex-gyre texlive texlive-base texlive-binaries
texlive-fonts-recommended texlive-latex-base texlive-latex-extra
texlive-latex-recommended texlive-pictures texlive-plain-generic
texlive-xetex tipa xfonts-encodings xfonts-utils

0 upgraded, 59 newly installed, 0 to remove and 26 not upgraded.

Need to get 169 MB of archives.

After this operation, 537 MB of additional disk space will be used.

Get:1 <http://archive.ubuntu.com/ubuntu focal/main amd64 fonts-droid-fallback all 1:6.0.1r16-1.1> [1,805 kB]

Get:2 <http://archive.ubuntu.com/ubuntu focal/main amd64 fonts-lato all 2.0-2> [2,698 kB]

Get:3 <http://archive.ubuntu.com/ubuntu focal/main amd64 poppler-data all 0.4.9-2> [1,475 kB]

Get:4 <http://archive.ubuntu.com/ubuntu focal/universe amd64 tex-common all 6.13> [32.7 kB]

Get:5 <http://archive.ubuntu.com/ubuntu focal/main amd64 fonts-urw-base35 all 20170801.1-3> [6,333 kB]

Get:6 <http://archive.ubuntu.com/ubuntu focal-updates/main amd64 libgs9-common all 9.50~dfsg-5ubuntu4.7> [681 kB]

Get:7 <http://archive.ubuntu.com/ubuntu focal/main amd64 libidn11 amd64 1.33-2.2ubuntu2> [46.2 kB]

Get:8 <http://archive.ubuntu.com/ubuntu focal/main amd64 libijs-0.35 amd64 0.35-15> [15.7 kB]

Get:9 <http://archive.ubuntu.com/ubuntu> focal/main amd64 libjbig2dec0 amd64
0.18-1ubuntu1 [60.0 kB]
Get:10 <http://archive.ubuntu.com/ubuntu> focal-updates/main amd64 libgs9 amd64
9.50~dfsg-5ubuntu4.7 [2,173 kB]
Get:11 <http://archive.ubuntu.com/ubuntu> focal/main amd64 libkpathsea6 amd64
2019.20190605.51237-3build2 [57.0 kB]
Get:12 <http://archive.ubuntu.com/ubuntu> focal/main amd64 libwoff1 amd64
1.0.2-1build2 [42.0 kB]
Get:13 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 dvisvgm amd64
2.8.1-1build1 [1,048 kB]
Get:14 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 fonts-lmodern all
2.004.5-6 [4,532 kB]
Get:15 <http://archive.ubuntu.com/ubuntu> focal-updates/main amd64 fonts-noto-mono
all 20200323-1build1~ubuntu20.04.1 [80.6 kB]
Get:16 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 fonts-texgyre all
20180621-3 [10.2 MB]
Get:17 <http://archive.ubuntu.com/ubuntu> focal/main amd64 javascript-common all
11 [6,066 B]
Get:18 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 libapache-pom-java
all 18-1 [4,720 B]
Get:19 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 libcommons-parent-
java all 43-1 [10.8 kB]
Get:20 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 libcommons-logging-
java all 1.2-2 [60.3 kB]
Get:21 <http://archive.ubuntu.com/ubuntu> focal/main amd64 libfontenc1 amd64
1:1.1.4-0ubuntu1 [14.0 kB]
Get:22 <http://archive.ubuntu.com/ubuntu> focal-updates/main amd64 libharfbuzz-
icu0 amd64 2.6.4-1ubuntu4.2 [5,580 B]
Get:23 <http://archive.ubuntu.com/ubuntu> focal/main amd64 libjs-jquery all
3.3.1~dfsg-3 [329 kB]
Get:24 <http://archive.ubuntu.com/ubuntu> focal/main amd64 libptexenc1 amd64
2019.20190605.51237-3build2 [35.5 kB]
Get:25 <http://archive.ubuntu.com/ubuntu> focal/main amd64 rubygems-integration
all 1.16 [5,092 B]
Get:26 <http://archive.ubuntu.com/ubuntu> focal-updates/main amd64 ruby2.7 amd64
2.7.0-5ubuntu1.8 [95.6 kB]
Get:27 <http://archive.ubuntu.com/ubuntu> focal/main amd64 ruby amd64 1:2.7+1
[5,412 B]
Get:28 <http://archive.ubuntu.com/ubuntu> focal/main amd64 rake all 13.0.1-4 [61.6
kB]
Get:29 <http://archive.ubuntu.com/ubuntu> focal/main amd64 ruby-minitest all
5.13.0-1 [40.9 kB]
Get:30 <http://archive.ubuntu.com/ubuntu> focal/main amd64 ruby-net-telnet all
0.1.1-2 [12.6 kB]
Get:31 <http://archive.ubuntu.com/ubuntu> focal/main amd64 ruby-power-assert all
1.1.7-1 [11.4 kB]
Get:32 <http://archive.ubuntu.com/ubuntu> focal/main amd64 ruby-test-unit all
3.3.5-1 [73.2 kB]

Get:33 <http://archive.ubuntu.com/ubuntu> focal/main amd64 ruby-xlrrpc all 0.3.0-2 [23.8 kB]

Get:34 <http://archive.ubuntu.com/ubuntu> focal-updates/main amd64 libruby2.7 amd64 2.7.0-5ubuntu1.8 [3,532 kB]

Get:35 <http://archive.ubuntu.com/ubuntu> focal/main amd64 libsyntax2 amd64 2019.20190605.51237-3build2 [55.0 kB]

Get:36 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 libteckit0 amd64 2.5.8+ds2-5ubuntu2 [320 kB]

Get:37 <http://archive.ubuntu.com/ubuntu> focal/main amd64 libtexlua53 amd64 2019.20190605.51237-3build2 [105 kB]

Get:38 <http://archive.ubuntu.com/ubuntu> focal/main amd64 libtexluajit2 amd64 2019.20190605.51237-3build2 [235 kB]

Get:39 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 libzip-0-13 amd64 0.13.62-3.2ubuntu1 [26.2 kB]

Get:40 <http://archive.ubuntu.com/ubuntu> focal/main amd64 xfonts-encodings all 1:1.0.5-0ubuntu1 [573 kB]

Get:41 <http://archive.ubuntu.com/ubuntu> focal/main amd64 xfonts-utils amd64 1:7.7+6 [91.5 kB]

Get:42 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 lmodern all 2.004.5-6 [9,474 kB]

Get:43 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 preview-latex-style all 11.91-2ubuntu2 [184 kB]

Get:44 <http://archive.ubuntu.com/ubuntu> focal/main amd64 t1utils amd64 1.41-3 [56.1 kB]

Get:45 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 teckit amd64 2.5.8+ds2-5ubuntu2 [687 kB]

Get:46 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 tex-gyre all 20180621-3 [6,209 kB]

Get:47 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 texlive-binaries amd64 2019.20190605.51237-3build2 [8,041 kB]

Get:48 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 texlive-base all 2019.20200218-1 [20.8 MB]

Get:49 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 texlive-fonts-recommended all 2019.20200218-1 [4,972 kB]

Get:50 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 texlive-latex-base all 2019.20200218-1 [990 kB]

Get:51 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 texlive-latex-recommended all 2019.20200218-1 [15.7 MB]

Get:52 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 texlive all 2019.20200218-1 [14.4 kB]

Get:53 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 libfontbox-java all 1:1.8.16-2 [207 kB]

Get:54 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 libpdfbox-java all 1:1.8.16-2 [5,199 kB]

Get:55 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 texlive-pictures all 2019.20200218-1 [4,492 kB]

Get:56 <http://archive.ubuntu.com/ubuntu> focal/universe amd64 texlive-latex-extra all 2019.20200218-1 [12.5 MB]

```

Get:57 http://archive.ubuntu.com/ubuntu focal/universe amd64 texlive-plain-
generic all 2019.20200218-1 [24.6 MB]
Get:58 http://archive.ubuntu.com/ubuntu focal/universe amd64 tipa all 2:1.3-20
[2,978 kB]
Get:59 http://archive.ubuntu.com/ubuntu focal/universe amd64 texlive-xetex all
2019.20200218-1 [14.6 MB]
Fetched 169 MB in 8s (21.2 MB/s)
Extracting templates from packages: 100%
Preconfiguring packages ...
Selecting previously unselected package fonts-droid-fallback.
(Reading database ... 122400 files and directories currently installed.)
Preparing to unpack .../00-fonts-droid-fallback_1%3a6.0.1r16-1.1_all.deb ...
Unpacking fonts-droid-fallback (1:6.0.1r16-1.1) ...
Selecting previously unselected package fonts-lato.
Preparing to unpack .../01-fonts-lato_2.0-2_all.deb ...
Unpacking fonts-lato (2.0-2) ...
Selecting previously unselected package poppler-data.
Preparing to unpack .../02-poppler-data_0.4.9-2_all.deb ...
Unpacking poppler-data (0.4.9-2) ...
Selecting previously unselected package tex-common.
Preparing to unpack .../03-tex-common_6.13_all.deb ...
Unpacking tex-common (6.13) ...
Selecting previously unselected package fonts-urw-base35.
Preparing to unpack .../04-fonts-urw-base35_20170801.1-3_all.deb ...
Unpacking fonts-urw-base35 (20170801.1-3) ...
Selecting previously unselected package libgs9-common.
Preparing to unpack .../05-libgs9-common_9.50~dfsg-5ubuntu4.7_all.deb ...
Unpacking libgs9-common (9.50~dfsg-5ubuntu4.7) ...
Selecting previously unselected package libidn11:amd64.
Preparing to unpack .../06-libidn11_1.33-2.2ubuntu2_amd64.deb ...
Unpacking libidn11:amd64 (1.33-2.2ubuntu2) ...
Selecting previously unselected package libijs-0.35:amd64.
Preparing to unpack .../07-libijs-0.35_0.35-15_amd64.deb ...
Unpacking libijs-0.35:amd64 (0.35-15) ...
Selecting previously unselected package libjbig2dec0:amd64.
Preparing to unpack .../08-libjbig2dec0_0.18-1ubuntu1_amd64.deb ...
Unpacking libjbig2dec0:amd64 (0.18-1ubuntu1) ...
Selecting previously unselected package libgs9:amd64.
Preparing to unpack .../09-libgs9_9.50~dfsg-5ubuntu4.7_amd64.deb ...
Unpacking libgs9:amd64 (9.50~dfsg-5ubuntu4.7) ...
Selecting previously unselected package libkpathsea6:amd64.
Preparing to unpack .../10-libkpathsea6_2019.20190605.51237-3build2_amd64.deb
...
Unpacking libkpathsea6:amd64 (2019.20190605.51237-3build2) ...
Selecting previously unselected package libwoff1:amd64.
Preparing to unpack .../11-libwoff1_1.0.2-1build2_amd64.deb ...
Unpacking libwoff1:amd64 (1.0.2-1build2) ...
Selecting previously unselected package dvisvgm.

```

```

Preparing to unpack .../12-dvisvgm_2.8.1-1build1_amd64.deb ...
Unpacking dvisvgm (2.8.1-1build1) ...
Selecting previously unselected package fonts-lmodern.
Preparing to unpack .../13-fonts-lmodern_2.004.5-6_all.deb ...
Unpacking fonts-lmodern (2.004.5-6) ...
Selecting previously unselected package fonts-noto-mono.
Preparing to unpack .../14-fonts-noto-mono_20200323-1build1~ubuntu20.04.1_all.deb ...
Unpacking fonts-noto-mono (20200323-1build1~ubuntu20.04.1) ...
Selecting previously unselected package fonts-texgyre.
Preparing to unpack .../15-fonts-texgyre_20180621-3_all.deb ...
Unpacking fonts-texgyre (20180621-3) ...
Selecting previously unselected package javascript-common.
Preparing to unpack .../16-javascript-common_11_all.deb ...
Unpacking javascript-common (11) ...
Selecting previously unselected package libapache-pom-java.
Preparing to unpack .../17-libapache-pom-java_18-1_all.deb ...
Unpacking libapache-pom-java (18-1) ...
Selecting previously unselected package libcommons-parent-java.
Preparing to unpack .../18-libcommons-parent-java_43-1_all.deb ...
Unpacking libcommons-parent-java (43-1) ...
Selecting previously unselected package libcommons-logging-java.
Preparing to unpack .../19-libcommons-logging-java_1.2-2_all.deb ...
Unpacking libcommons-logging-java (1.2-2) ...
Selecting previously unselected package libfontenc1:amd64.
Preparing to unpack .../20-libfontenc1_1%3a1.1.4-0ubuntu1_amd64.deb ...
Unpacking libfontenc1:amd64 (1:1.1.4-0ubuntu1) ...
Selecting previously unselected package libharfbuzz-icu0:amd64.
Preparing to unpack .../21-libharfbuzz-icu0_2.6.4-1ubuntu4.2_amd64.deb ...
Unpacking libharfbuzz-icu0:amd64 (2.6.4-1ubuntu4.2) ...
Selecting previously unselected package libjs-jquery.
Preparing to unpack .../22-libjs-jquery_3.3.1~dfsg-3_all.deb ...
Unpacking libjs-jquery (3.3.1~dfsg-3) ...
Selecting previously unselected package libptexenc1:amd64.
Preparing to unpack .../23-libptexenc1_2019.20190605.51237-3build2_amd64.deb ...
Unpacking libptexenc1:amd64 (2019.20190605.51237-3build2) ...
Selecting previously unselected package rubygems-integration.
Preparing to unpack .../24-rubygems-integration_1.16_all.deb ...
Unpacking rubygems-integration (1.16) ...
Selecting previously unselected package ruby2.7.
Preparing to unpack .../25-ruby2.7_2.7.0-5ubuntu1.8_amd64.deb ...
Unpacking ruby2.7 (2.7.0-5ubuntu1.8) ...
Selecting previously unselected package ruby.
Preparing to unpack .../26-ruby_1%3a2.7+1_amd64.deb ...
Unpacking ruby (1:2.7+1) ...
Selecting previously unselected package rake.
Preparing to unpack .../27-rake_13.0.1-4_all.deb ...
Unpacking rake (13.0.1-4) ...

```

```

Selecting previously unselected package ruby-minitest.
Preparing to unpack .../28-ruby-minitest_5.13.0-1_all.deb ...
Unpacking ruby-minitest (5.13.0-1) ...
Selecting previously unselected package ruby-net-telnet.
Preparing to unpack .../29-ruby-net-telnet_0.1.1-2_all.deb ...
Unpacking ruby-net-telnet (0.1.1-2) ...
Selecting previously unselected package ruby-power-assert.
Preparing to unpack .../30-ruby-power-assert_1.1.7-1_all.deb ...
Unpacking ruby-power-assert (1.1.7-1) ...
Selecting previously unselected package ruby-test-unit.
Preparing to unpack .../31-ruby-test-unit_3.3.5-1_all.deb ...
Unpacking ruby-test-unit (3.3.5-1) ...
Selecting previously unselected package ruby-xmlrpc.
Preparing to unpack .../32-ruby-xmlrpc_0.3.0-2_all.deb ...
Unpacking ruby-xmlrpc (0.3.0-2) ...
Selecting previously unselected package libruby2.7:amd64.
Preparing to unpack .../33-libruby2.7_2.7.0-5ubuntu1.8_amd64.deb ...
Unpacking libruby2.7:amd64 (2.7.0-5ubuntu1.8) ...
Selecting previously unselected package libsyntax2:amd64.
Preparing to unpack .../34-libsyntax2_2019.20190605.51237-3build2_amd64.deb ...
Unpacking libsyntax2:amd64 (2019.20190605.51237-3build2) ...
Selecting previously unselected package libteckit0:amd64.
Preparing to unpack .../35-libteckit0_2.5.8+ds2-5ubuntu2_amd64.deb ...
Unpacking libteckit0:amd64 (2.5.8+ds2-5ubuntu2) ...
Selecting previously unselected package libtexlua53:amd64.
Preparing to unpack .../36-libtexlua53_2019.20190605.51237-3build2_amd64.deb ...
Unpacking libtexlua53:amd64 (2019.20190605.51237-3build2) ...
Selecting previously unselected package libtexluaajit2:amd64.
Preparing to unpack .../37-libtexluaajit2_2019.20190605.51237-3build2_amd64.deb
...
Unpacking libtexluaajit2:amd64 (2019.20190605.51237-3build2) ...
Selecting previously unselected package libzip-0-13:amd64.
Preparing to unpack .../38-libzip-0-13_0.13.62-3.2ubuntu1_amd64.deb ...
Unpacking libzip-0-13:amd64 (0.13.62-3.2ubuntu1) ...
Selecting previously unselected package xfonts-encodings.
Preparing to unpack .../39-xfonts-encodings_1%3a1.0.5-0ubuntu1_all.deb ...
Unpacking xfonts-encodings (1:1.0.5-0ubuntu1) ...
Selecting previously unselected package xfonts-utils.
Preparing to unpack .../40-xfonts-utils_1%3a7.7+6_amd64.deb ...
Unpacking xfonts-utils (1:7.7+6) ...
Selecting previously unselected package lmodern.
Preparing to unpack .../41-lmodern_2.004.5-6_all.deb ...
Unpacking lmodern (2.004.5-6) ...
Selecting previously unselected package preview-latex-style.
Preparing to unpack .../42-preview-latex-style_11.91-2ubuntu2_all.deb ...
Unpacking preview-latex-style (11.91-2ubuntu2) ...
Selecting previously unselected package t1utils.
Preparing to unpack .../43-t1utils_1.41-3_amd64.deb ...

```

```

Unpacking t1utils (1.41-3) ...
Selecting previously unselected package teckit.
Preparing to unpack .../44-teckit_2.5.8+ds2-5ubuntu2_amd64.deb ...
Unpacking teckit (2.5.8+ds2-5ubuntu2) ...
Selecting previously unselected package tex-gyre.
Preparing to unpack .../45-tex-gyre_20180621-3_all.deb ...
Unpacking tex-gyre (20180621-3) ...
Selecting previously unselected package texlive-binaries.
Preparing to unpack .../46-texlive-
binaries_2019.20190605.51237-3build2_amd64.deb ...
Unpacking texlive-binaries (2019.20190605.51237-3build2) ...
Selecting previously unselected package texlive-base.
Preparing to unpack .../47-texlive-base_2019.20200218-1_all.deb ...
Unpacking texlive-base (2019.20200218-1) ...
Selecting previously unselected package texlive-fonts-recommended.
Preparing to unpack .../48-texlive-fonts-recommended_2019.20200218-1_all.deb ...
Unpacking texlive-fonts-recommended (2019.20200218-1) ...
Selecting previously unselected package texlive-latex-base.
Preparing to unpack .../49-texlive-latex-base_2019.20200218-1_all.deb ...
Unpacking texlive-latex-base (2019.20200218-1) ...
Selecting previously unselected package texlive-latex-recommended.
Preparing to unpack .../50-texlive-latex-recommended_2019.20200218-1_all.deb ...
Unpacking texlive-latex-recommended (2019.20200218-1) ...
Selecting previously unselected package texlive.
Preparing to unpack .../51-texlive_2019.20200218-1_all.deb ...
Unpacking texlive (2019.20200218-1) ...
Selecting previously unselected package libfontbox-java.
Preparing to unpack .../52-libfontbox-java_1%3a1.8.16-2_all.deb ...
Unpacking libfontbox-java (1:1.8.16-2) ...
Selecting previously unselected package libpdfbox-java.
Preparing to unpack .../53-libpdfbox-java_1%3a1.8.16-2_all.deb ...
Unpacking libpdfbox-java (1:1.8.16-2) ...
Selecting previously unselected package texlive-pictures.
Preparing to unpack .../54-texlive-pictures_2019.20200218-1_all.deb ...
Unpacking texlive-pictures (2019.20200218-1) ...
Selecting previously unselected package texlive-latex-extra.
Preparing to unpack .../55-texlive-latex-extra_2019.202000218-1_all.deb ...
Unpacking texlive-latex-extra (2019.202000218-1) ...
Selecting previously unselected package texlive-plain-generic.
Preparing to unpack .../56-texlive-plain-generic_2019.202000218-1_all.deb ...
Unpacking texlive-plain-generic (2019.202000218-1) ...
Selecting previously unselected package tipa.
Preparing to unpack .../57-tipa_2%3a1.3-20_all.deb ...
Unpacking tipa (2:1.3-20) ...
Selecting previously unselected package texlive-xetex.
Preparing to unpack .../58-texlive-xetex_2019.20200218-1_all.deb ...
Unpacking texlive-xetex (2019.20200218-1) ...
Setting up javascript-common (11) ...

```

```

Setting up libharfbuzz-icu0:amd64 (2.6.4-1ubuntu4.2) ...
Setting up fonts-lato (2.0-2) ...
Setting up fonts-noto-mono (20200323-1build1~ubuntu20.04.1) ...
Setting up libwoff1:amd64 (1.0.2-1build2) ...
Setting up ruby-power-assert (1.1.7-1) ...
Setting up libtexlua53:amd64 (2019.20190605.51237-3build2) ...
Setting up libijs-0.35:amd64 (0.35-15) ...
Setting up libtexluajit2:amd64 (2019.20190605.51237-3build2) ...
Setting up libfontbox-java (1:1.8.16-2) ...
Setting up rubygems-integration (1.16) ...
Setting up libzip-0-13:amd64 (0.13.62-3.2ubuntu1) ...
Setting up fonts-urw-base35 (20170801.1-3) ...
Setting up poppler-data (0.4.9-2) ...
Setting up ruby-minitest (5.13.0-1) ...
Setting up tex-common (6.13) ...
update-language: texlive-base not installed and configured, doing nothing!
Setting up libfontenc1:amd64 (1:1.1.4-0ubuntu1) ...
Setting up ruby-test-unit (3.3.5-1) ...
Setting up libjbig2dec0:amd64 (0.18-1ubuntu1) ...
Setting up libidn11:amd64 (1.33-2.2ubuntu2) ...
Setting up libteckit0:amd64 (2.5.8+ds2-5ubuntu2) ...
Setting up libapache-pom-java (18-1) ...
Setting up ruby-net-telnet (0.1.1-2) ...
Setting up xfonts-encodings (1:1.0.5-0ubuntu1) ...
Setting up t1utils (1.41-3) ...
Setting up fonts-texgyre (20180621-3) ...
Setting up libkpathsea6:amd64 (2019.20190605.51237-3build2) ...
Setting up fonts-lmodern (2.004.5-6) ...
Setting up fonts-droid-fallback (1:6.0.1r16-1.1) ...
Setting up libjs-jquery (3.3.1~dfsg-3) ...
Setting up ruby-xmlrpc (0.3.0-2) ...
Setting up libsynchronet2:amd64 (2019.20190605.51237-3build2) ...
Setting up libgs9-common (9.50~dfsg-5ubuntu4.7) ...
Setting up teckit (2.5.8+ds2-5ubuntu2) ...
Setting up libpdfbox-java (1:1.8.16-2) ...
Setting up libgs9:amd64 (9.50~dfsg-5ubuntu4.7) ...
Setting up preview-latex-style (11.91-2ubuntu2) ...
Setting up libcommons-parent-java (43-1) ...
Setting up dvisvgm (2.8.1-1build1) ...
Setting up libcommons-logging-java (1.2-2) ...
Setting up xfonts-utils (1:7.7+6) ...
Setting up libptexenc1:amd64 (2019.20190605.51237-3build2) ...
Setting up texlive-binaries (2019.20190605.51237-3build2) ...
update-alternatives: using /usr/bin/xdvi-xaw to provide /usr/bin/xdvi.bin
(xdvi.bin) in auto mode
update-alternatives: using /usr/bin/bibtex.original to provide /usr/bin/bibtex
(bibtex) in auto mode
Setting up lmodern (2.004.5-6) ...

```



```

Setting up texlive-base (2019.20200218-1) ...
mktexlsr: Updating /var/lib/texmf/ls-R-TEXLIVEDIST...
mktexlsr: Updating /var/lib/texmf/ls-R-TEXMFMAIN...
mktexlsr: Updating /var/lib/texmf/ls-R...
mktexlsr: Done.
tl-paper: setting paper size for dvips to a4:
/var/lib/texmf/dvips/config/config-paper.ps
tl-paper: setting paper size for dvipdfmx to a4:
/var/lib/texmf/dvipdfmx/dvipdfmx-paper.cfg
tl-paper: setting paper size for xdvi to a4: /var/lib/texmf/xdvi/XDvi-paper
tl-paper: setting paper size for pdftex to a4:
/var/lib/texmf/tex/generic/config/pdftexconfig.tex
Setting up tex-gyre (20180621-3) ...
Setting up texlive-plain-generic (2019.202000218-1) ...
Setting up texlive-latex-base (2019.20200218-1) ...
Setting up texlive-latex-recommended (2019.20200218-1) ...
Setting up texlive-pictures (2019.20200218-1) ...
Setting up texlive-fonts-recommended (2019.20200218-1) ...
Setting up tipa (2:1.3-20) ...
Regenerating '/var/lib/texmf/fmtutil.cnf-DEBIAN'... done.
Regenerating '/var/lib/texmf/fmtutil.cnf-TEXLIVEDIST'... done.
update-fmtutil has updated the following file(s):
    /var/lib/texmf/fmtutil.cnf-DEBIAN
    /var/lib/texmf/fmtutil.cnf-TEXLIVEDIST
If you want to activate the changes in the above file(s),
you should run fmtutil-sys or fmtutil.
Setting up texlive (2019.20200218-1) ...
Setting up texlive-latex-extra (2019.202000218-1) ...
Setting up texlive-xetex (2019.20200218-1) ...
Setting up rake (13.0.1-4) ...
Setting up libruby2.7:amd64 (2.7.0-5ubuntu1.8) ...
Setting up ruby2.7 (2.7.0-5ubuntu1.8) ...
Setting up ruby (1:2.7+1) ...
Processing triggers for fontconfig (2.13.1-2ubuntu3) ...
Processing triggers for mime-support (3.64ubuntu1) ...
Processing triggers for libc-bin (2.31-0ubuntu9.9) ...
Processing triggers for man-db (2.9.1-1) ...
Processing triggers for tex-common (6.13) ...
Running updmap-sys. This may take some time... done.
Running mktexlsr /var/lib/texmf ... done.
Building format(s) --all.
    This may take some time... done.
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting py pandoc
  Downloading py pandoc-1.11-py3-none-any.whl (20 kB)
Installing collected packages: py pandoc
Successfully installed py pandoc-1.11

```

```
[376]: from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

```
[379]: !cp /content/drive/My Drive/Colab Notebooks/Data Mining Assignment 2.ipynb ./
```

```
cp: cannot stat '/content/drive/My': No such file or directory
cp: cannot stat 'Drive/Colab': No such file or directory
cp: cannot stat 'Notebooks/Data': No such file or directory
cp: cannot stat 'Mining': No such file or directory
cp: cannot stat 'Assignment': No such file or directory
cp: cannot stat '2.ipynb': No such file or directory
```

```
[383]: !wget -nc https://raw.githubusercontent.com/brpy/colab-pdf/master/colab_pdf.py
from colab_pdf import colab_pdf
colab_pdf('Data Mining Assignment 2.ipynb')
```

File 'colab_pdf.py' already there; not retrieving.

WARNING: apt does not have a stable CLI interface. Use with caution in scripts.

WARNING: apt does not have a stable CLI interface. Use with caution in scripts.

E: Unable to locate package texlive-generic-recommended
[NbConvertApp] WARNING | pattern '\$notebookpath\$file_name' matched no files
This application is used to convert notebook files (*.ipynb)
to various other formats.

WARNING: THE COMMANDLINE INTERFACE MAY CHANGE IN FUTURE RELEASES.

Options

=====

The options below are convenience aliases to configurable class-options,
as listed in the "Equivalent to" description-line of the aliases.

To see all configurable class-options for some <cmd>, use:

<cmd> --help-all

--debug

set log level to logging.DEBUG (maximize logging output)

Equivalent to: [--Application.log_level=10]

--show-config

Show the application's configuration (human-readable format)

Equivalent to: [--Application.show_config=True]

--show-config-json

Show the application's configuration (json format)
 Equivalent to: [--Application.show_config_json=True]

--generate-config
 generate default config file
 Equivalent to: [--JupyterApp.generate_config=True]

-y
 Answer yes to any questions instead of prompting.
 Equivalent to: [--JupyterApp.answer_yes=True]

--execute
 Execute the notebook prior to export.
 Equivalent to: [--ExecutePreprocessor.enabled=True]

--allow-errors
 Continue notebook execution even if one of the cells throws an error and include the error message in the cell output (the default behaviour is to abort conversion). This flag is only relevant if '--execute' was specified, too.
 Equivalent to: [--ExecutePreprocessor.allow_errors=True]

--stdin
 read a single notebook file from stdin. Write the resulting notebook with default basename 'notebook.*'
 Equivalent to: [--NbConvertApp.from_stdin=True]

--stdout
 Write notebook output to stdout instead of files.
 Equivalent to: [--NbConvertApp.writer_class=StdoutWriter]

--inplace
 Run nbconvert in place, overwriting the existing notebook (only relevant when converting to notebook format)
 Equivalent to: [--NbConvertApp.use_output_suffix=False]

--NbConvertApp.export_format=notebook --FilesWriter.build_directory=]

--clear-output
 Clear output of current file and save in place, overwriting the existing notebook.
 Equivalent to: [--NbConvertApp.use_output_suffix=False]

--NbConvertApp.export_format=notebook --FilesWriter.build_directory=

--ClearOutputPreprocessor.enabled=True]

--no-prompt
 Exclude input and output prompts from converted document.
 Equivalent to: [--TemplateExporter.exclude_input_prompt=True]

--TemplateExporter.exclude_output_prompt=True]

--no-input
 Exclude input cells and output prompts from converted document.
 This mode is ideal for generating code-free reports.
 Equivalent to: [--TemplateExporter.exclude_output_prompt=True]

--TemplateExporter.exclude_input=True

--TemplateExporter.exclude_input_prompt=True]

--allow-chromium-download
 Whether to allow downloading chromium if no suitable version is found on the system.
 Equivalent to: [--WebPDFExporter.allow_chromium_download=True]

```

--disable-chromium-sandbox
    Disable chromium security sandbox when converting to PDF..
    Equivalent to: [--WebPDFExporter.disable_sandbox=True]
--show-input
    Shows code input. This flag is only useful for dejavu users.
    Equivalent to: [--TemplateExporter.exclude_input=False]
--embed-images
    Embed the images as base64 dataurls in the output. This flag is only useful
for the HTML/WebPDF/Slides exports.
    Equivalent to: [--HTMLExporter.embed_images=True]
--sanitize-html
    Whether the HTML in Markdown cells and cell outputs should be sanitized..
    Equivalent to: [--HTMLExporter.sanitize_html=True]
--log-level=<Enum>
    Set the log level by value or name.
    Choices: any of [0, 10, 20, 30, 40, 50, 'DEBUG', 'INFO', 'WARN', 'ERROR',
'CRITICAL']
    Default: 30
    Equivalent to: [--Application.log_level]
--config=<Unicode>
    Full path of a config file.
    Default: ''
    Equivalent to: [--JupyterApp.config_file]
--to=<Unicode>
    The export format to be used, either one of the built-in formats
    ['asciidoc', 'custom', 'html', 'latex', 'markdown', 'notebook',
'pdf', 'python', 'rst', 'script', 'slides', 'webpdf']
    or a dotted object name that represents the import path for an
    ``Exporter`` class
    Default: ''
    Equivalent to: [--NbConvertApp.export_format]
--template=<Unicode>
    Name of the template to use
    Default: ''
    Equivalent to: [--TemplateExporter.template_name]
--template-file=<Unicode>
    Name of the template file to use
    Default: None
    Equivalent to: [--TemplateExporter.template_file]
--theme=<Unicode>
    Template specific theme(e.g. the name of a JupyterLab CSS theme distributed
as prebuilt extension for the lab template)
    Default: 'light'
    Equivalent to: [--HTMLExporter.theme]
--sanitize_html=<Bool>
    Whether the HTML in Markdown cells and cell outputs should be sanitized.This
should be set to True by nbviewer or similar tools.
    Default: False

```

Equivalent to: [--HTMLExporter.sanitize_html]

--writer=<DottedObjectName>
 Writer class used to write the results of the conversion

Default: 'FilesWriter'

Equivalent to: [--NbConvertApp.writer_class]

--post=<DottedOrNone>
 PostProcessor class used to write the results of the conversion

Default: ''

Equivalent to: [--NbConvertApp.postprocessor_class]

--output=<Unicode>
 overwrite base name use for output files.
 can only be used when converting one notebook at a time.

Default: ''

Equivalent to: [--NbConvertApp.output_base]

--output-dir=<Unicode>
 Directory to write output(s) to. Defaults to output to the directory of each notebook.

To recover previous default behaviour (outputting to the current working directory) use . as the flag value.

Default: ''

Equivalent to: [--FilesWriter.build_directory]

--reveal-prefix=<Unicode>
 The URL prefix for reveal.js (version 3.x).
 This defaults to the reveal CDN, but can be any url pointing to a copy of reveal.js.
 For speaker notes to work, this must be a relative path to a local copy of reveal.js: e.g., "reveal.js".
 If a relative path is given, it must be a subdirectory of the current directory (from which the server is run).
 See the usage documentation
 (<https://nbconvert.readthedocs.io/en/latest/usage.html#reveal-js-html-slideshow>)
 for more details.

Default: ''

Equivalent to: [--SlidesExporter.reveal_url_prefix]

--nbformat=<Enum>
 The nbformat version to write.
 Use this to downgrade notebooks.
 Choices: any of [1, 2, 3, 4]
 Default: 4
 Equivalent to: [--NotebookExporter.nbformat_version]

Examples

The simplest way to use nbconvert is

```
> jupyter nbconvert mynotebook.ipynb --to html
```

Options include ['asciidoc', 'custom', 'html', 'latex', 'markdown', 'notebook', 'pdf', 'python', 'rst', 'script', 'slides', 'webpdf'].

```
> jupyter nbconvert --to latex mynotebook.ipynb
```

Both HTML and LaTeX support multiple output templates. LaTeX includes

'base', 'article' and 'report'. HTML includes 'basic', 'lab' and 'classic'. You can specify the flavor of the format used.

```
> jupyter nbconvert --to html --template lab mynotebook.ipynb
```

You can also pipe the output to stdout, rather than a file

```
> jupyter nbconvert mynotebook.ipynb --stdout
```

PDF is generated via latex

```
> jupyter nbconvert mynotebook.ipynb --to pdf
```

You can get (and serve) a Reveal.js-powered slideshow

```
> jupyter nbconvert myslides.ipynb --to slides --post serve
```

Multiple notebooks can be given at the command line in a couple of different ways:

```
> jupyter nbconvert notebook*.ipynb
> jupyter nbconvert notebook1.ipynb notebook2.ipynb
```

or you can specify the notebooks list in a config file, containing::

```
c.NbConvertApp.notebooks = ["my_notebook.ipynb"]
```

```
> jupyter nbconvert --config mycfg.py
```

To see all available configurables, use `--help-all`.

[383]: 'File Download Unsuccessful. Saved in Google Drive'