

▼ Stock Sentiment Analysis

▼ STEP 1: IMPORTING LIBRARIES

```
!pip install wordcloud
!pip install gensim
!pip install nltk
!pip install numpy
!pip install pandas
!pip install seaborn
!pip install tensorflow
!pip install jupyterthemes
!pip install sklearn

Requirement already satisfied: Send2Trash>=1.8.0 in /usr/local/lib/python3.10/dist-packages (from notebook>=5.6.0->jupyterthemes) (1.8.0)
Requirement already satisfied: terminado>=0.8.3 in /usr/local/lib/python3.10/dist-packages (from notebook>=5.6.0->jupyterthemes) (0.1.2)
Requirement already satisfied: prometheus-client in /usr/local/lib/python3.10/dist-packages (from notebook>=5.6.0->jupyterthemes) (0.1.0)
Requirement already satisfied: nbclassic>=0.4.7 in /usr/local/lib/python3.10/dist-packages (from notebook>=5.6.0->jupyterthemes) (1.0.0)
Requirement already satisfied: platformdirs>=2.5 in /usr/local/lib/python3.10/dist-packages (from jupyter-core->jupyterthemes) (4.2.2)
Requirement already satisfied: parso<0.9.0,>=0.8.3 in /usr/local/lib/python3.10/dist-packages (from jedi>=0.16->ipython>=5.4.1->jupyter>=5.4.1)
Requirement already satisfied: jupyter-server>=1.8 in /usr/local/lib/python3.10/dist-packages (from nbclassic>=0.4.7->notebook>=5.6.0)
Requirement already satisfied: notebook-shim>=0.2.3 in /usr/local/lib/python3.10/dist-packages (from nbclassic>=0.4.7->notebook>=5.6.0)
Requirement already satisfied: lxml in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=5.6.0->jupyterthemes) (<4.9.0, >=4.8.0)
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=5.6.0->jupyter>=5.4.1)
Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=5.6.0->jupyterthemes)
Requirement already satisfied: defusedxml in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=5.6.0->jupyterthemes)
Requirement already satisfied: entrypoints>=0.2.2 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=5.6.0->jupyterthemes)
Requirement already satisfied: jupyterlab-pygments in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=5.6.0->jupyterlab>=1.0.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=5.6.0->jupyter>=5.4.1)
Requirement already satisfied: mistune<2,>=0.8.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=5.6.0->jupyter>=5.4.1)
Requirement already satisfied: nbclient>0.5.0 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=5.6.0->jupyter>=5.4.1)
Requirement already satisfied: pandocfilters>=1.4.1 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=5.6.0->jupyter>=5.4.1)
Requirement already satisfied: tinycc2 in /usr/local/lib/python3.10/dist-packages (from nbconvert>=5->notebook>=5.6.0->jupyterthemes)
Requirement already satisfied: fastjsonschema>=2.15 in /usr/local/lib/python3.10/dist-packages (from nbformat->notebook>=5.6.0->jupyter>=5.4.1)
Requirement already satisfied: jsonschema>=2.6 in /usr/local/lib/python3.10/dist-packages (from nbformat->notebook>=5.6.0->jupyterthemes)
Requirement already satisfied: ptyprocess>=0.5 in /usr/local/lib/python3.10/dist-packages (from pexpect>4.3->ipython>=5.4.1->jupyter>=5.4.1)
Requirement already satisfied: wctwidth in /usr/local/lib/python3.10/dist-packages (from prompt-toolkit!=3.0.0,!=>3.0.1,<3.1.0,>=2.0.0)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->matplotlib>=1.4.3->jupyter>=5.4.1)
Requirement already satisfied: argon2-cffi-bindings in /usr/local/lib/python3.10/dist-packages (from argon2-cffi->notebook>=5.6.0->jupyter>=5.4.1)
Requirement already satisfied: attrs>=22.2.0 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat->notebook>=5.6.0->jupyter>=5.4.1)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6.0->nbformat->notebook>=5.6.0->jupyter>=5.4.1)
Requirement already satisfied: referencing>=0.28.4 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat->notebook>=5.6.0->jupyter>=5.4.1)
Requirement already satisfied: rpdfs-py>=0.7.1 in /usr/local/lib/python3.10/dist-packages (from jsonschema>=2.6->nbformat->notebook>=5.6.0->jupyter>=5.4.1)
Requirement already satisfied: aiohttp<4,>=3.1.0 in /usr/local/lib/python3.10/dist-packages (from jupyter-server>=1.8->nbclassic>=0.4.1)
Requirement already satisfied: websocket-client in /usr/local/lib/python3.10/dist-packages (from jupyter-server>=1.8->nbclassic>=0.4.1)
Requirement already satisfied: cffi>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from argon2-cffi-bindings->argon2-cffi->notebook>=5.6.0->jupyter>=5.4.1)
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-packages (from beautifulsoup4->nbconvert>=5->notebook>=5.6.0->jupyter>=5.4.1)
Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-packages (from bleach->nbconvert>=5->notebook>=5.6.0->jupyter>=5.4.1)
Requirement already satisfied: idna>=2.8 in /usr/local/lib/python3.10/dist-packages (from aiohttp<4,>=3.1.0->jupyter-server>=1.8->nbclassic>=0.4.1)
Requirement already satisfied: snffio>=1.1 in /usr/local/lib/python3.10/dist-packages (from aiohttp<4,>=3.1.0->jupyter-server>=1.8->nbclassic>=0.4.1)
Requirement already satisfied: exceptiongroup in /usr/local/lib/python3.10/dist-packages (from aiohttp<4,>=3.1.0->jupyter-server>=1.8->nbclassic>=0.4.1)
Requirement already satisfied: pycparser in /usr/local/lib/python3.10/dist-packages (from cffi>=1.0.1->argon2-cffi-bindings->argon2-cffi-bindings->notebook>=5.6.0->jupyter>=5.4.1)
Installing collected packages: ply, lesscpy, jedi, jupyterthemes
Successfully installed jedi-0.19.1 jupyterthemes-0.20.0 lesscpy-0.15.1 ply-3.11
Collecting sklearn
  Downloading sklearn-0.0.post12.tar.gz (2.6 kB)
error: subprocess-exited-with-error

  × python setup.py egg_info did not run successfully.
    exit code: 1
    ↳ See above for output.

note: This error originates from a subprocess, and is likely not a problem with pip.
  Preparing metadata (setup.py) ... error
error: metadata-generation-failed

  × Encountered error while generating package metadata.
    ↳ See above for output.

note: This is an issue with the package mentioned above, not pip.
  hint: See above for details.
```

```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from wordcloud import WordCloud, STOPWORDS
import nltk
import re
from nltk.stem import PorterStemmer, WordNetLemmatizer
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize, sent_tokenize
import gensim
from gensim.utils import simple_preprocess
from gensim.parsing.preprocessing import STOPWORDS

# TENSORFLOW
import tensorflow as tf
from tensorflow.keras.preprocessing.text import one_hot, Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Flatten, Embedding, Input, LSTM, Conv1D, MaxPool1D, Bidirectional, Dropout
from tensorflow.keras.models import Model
from tensorflow.keras.utils import to_categorical

from jupyterthemes import jtplot
jtplot.style(theme = 'monokai', context = 'notebook', ticks = True, grid = False)

```

▼ STEP 2: LOADING THE STOCKS DATA

```
stock_df = pd.read_csv('/content/stock_sentiment.csv')
```

```
# VIEWING THE DATA
stock_df
```

1 to 25 of 5791 entries

index	Text	Sentiment
0	Kickers on my watchlist XIDE TIT SOQ PNK CPW BPZ AJ trade method 1 or method 2, see prev posts	1
1	user: AAP MOVIE. 55% return for the FEA/GEED indicator just 15 trades for the year. AWESOME.	1
2	user I'd be afraid to short AMZN - they are looking like a near-monopoly in eBooks and infrastructure-as-a-service	1
3	MNTA Over 12.00	1
4	OI Over 21.37	1
5	PGNX Over 3.04	1
6	AAP - user if so then the current downtrend will break. Otherwise just a short-term correction in med-term downtrend.	0
7	Monday's relative weakness. NYX WIN TIE TAP ICE INT BMC AON C CHK BIIB	0
8	GOOG - over trend line channel test & volume support.	1
9	AAP will watch tomorrow for ONG entry.	1
10	i'm assuming FCX opens tomorrow above the 34.25 trigger buy. still very much like this setup.	1
11	It really worries me how everyone expects the market to rally now, usually exact opposite happens every time we shall see soon bac spx jpm	1
12	AAP GAMCO's arry Haverty : Apple Is Extremely Cheap Great Video !!!	1
13	user Maykiljil posted that. I agree that MSFT is going higher & possibly north of 30	1
14	Momentum is coming back to ETFC Broke MA200 resistance on solid volume Friday. ong set-up	1
15	HA Hitting 35.65 means resume targeting 42 level ..	1
16	user gameplan shot for today but I liked on trend break from May or c+h break. OC weekly trend break back to July 2011	1
17	with FCX gapping well above ideal entry looking for a pull in to at least 35 on open for an entry	1
18	user great list again, particularly like FISV and SYK. All buy & hold types should check the free list out.	1
19	ATHX upper trend line	1
20	NG - nice PNF BY - breakout - need follow thru	1
21	Won't believe AAP uptrend is back until it crosses above MA(50)	0
22	X swing still on	1
23	SWY - 30% of float short and breaking out - ouch	1
24	BIOF wants 4.90's comin!!!	1

Show per page



Like what you see? Visit the [data table notebook](#) to learn more about interactive tables.

Next steps:

```
# DATAFRAME INFORMATION:
stock_df.info()

→ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 5791 entries, 0 to 5790
Data columns (total 2 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   Text        5791 non-null    object  
 1   Sentiment    5791 non-null    int64  
dtypes: int64(1), object(1)
memory usage: 90.6+ KB
```

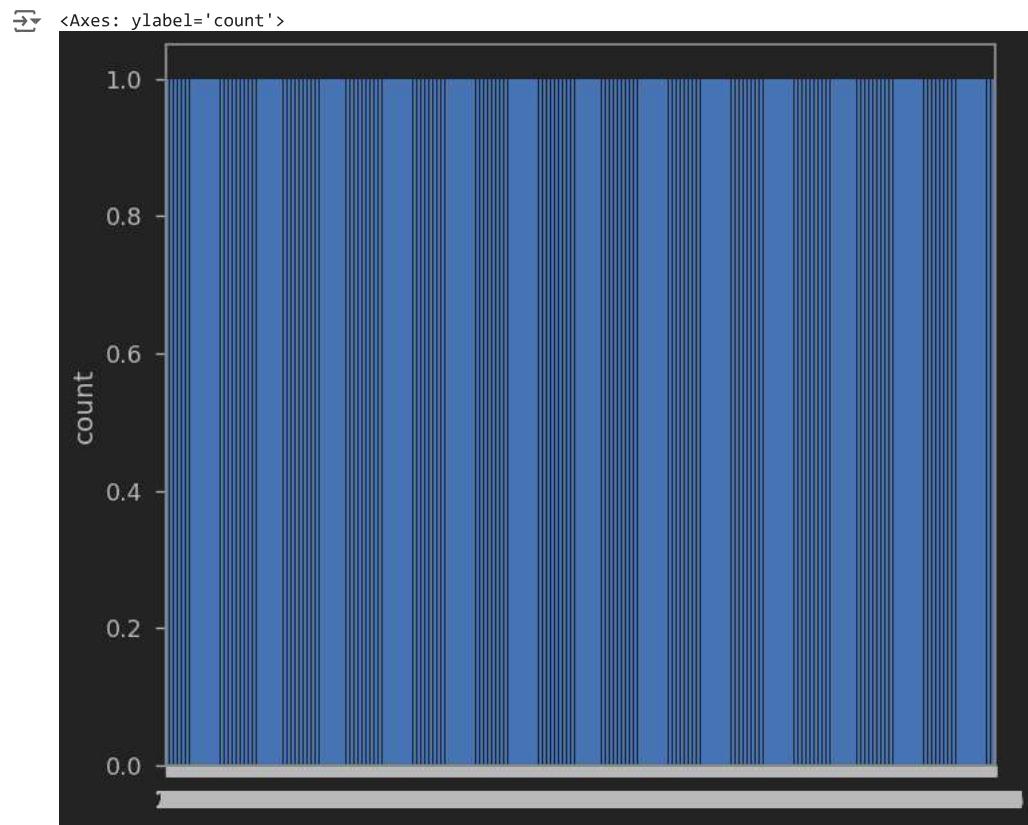
```
# CHECKING FOR NULL VALUES
stock_df.isnull().sum()
```

```
→ Text      0
Sentiment  0
dtype: int64
```

```
# CONFIRMING THE ELEMENTS PRESENT IN Sentiment Column
stock_df['Sentiment'].unique()
```

```
→ array([1, 0])
```

```
# VISUALIZING OUR DATASET
sns.countplot(stock_df['Sentiment'])
```



▼ STEP 3: PERFORMING DATA CLEANING

▼ A) REMOVING PUNCTUATIONS FROM TEXT

```
import string
string.punctuation

→ '!#$%&\'()*+,-./:;<=>?@[\\]^_`{|}~'
```

```
# DEFINING A FUNCTION TO REMOVE PUNCTUATIONS
def remove_func(message):
    Test_punc_removed = [char for char in message if char not in string.punctuation]
    Test_punc_removed_join = ''.join(Test_punc_removed)

    return Test_punc_removed_join

# REMOVING PUNCTUATIONS FROM OUR DATASET
stock_df['Text-Without-Punctuations'] = stock_df['Text'].apply(remove_func)
```

stock_df

1 to 25 of 5791 entries

index	Text	Sentiment	Text-Without-Punctuations
0	Kickers on my watchlist XIDE TIT SOQ PNK CPW BPZ AJ trade method 1 or method 2, see prev posts	1	Kickers on my watchlist XIDE TIT SOQ PNK CPW BPZ AJ trade method 1 or method 2 see prev posts
1	user: AAP MOVIE. 55% return for the FEA/GEED indicator just 15 trades for the year. AWESOME.	1	user AAP MOVIE 55 return for the FEAGEED indicator just 15 trades for the year AWESOME
2	user I'd be afraid to short AMZN - they are looking like a near-monopoly in eBooks and infrastructure-as-a-service	1	user Id be afraid to short AMZN they are looking like a nearmonopoly in eBooks and infrastructureasaservice
3	MNTA Over 12.00	1	MNTA Over 1200
4	OI Over 21.37	1	OI Over 2137
5	PGNX Over 3.04	1	PGNX Over 304
6	AAP - user if so then the current downtrend will break. Otherwise just a short-term correction in med-term downtrend.	0	AAP user if so then the current downtrend will break Otherwise just a shortterm correction in medterm downtrend
7	Monday's relative weakness. NYX WIN TIE TAP ICE INT BMC AON C CHK BIIB	0	Mondays relative weakness NYX WIN TIE TAP ICE INT BMC AON C CHK BIIB
8	GOOG - ower trend line channel test & volume support.	1	GOOG ower trend line channel test volume support
9	AAP will watch tomorrow for ONG entry.	1	AAP will watch tomorrow for ONG entry
10	i'm assuming FCX opens tomorrow above the 34.25 trigger buy. still very much like this setup.	1	im assuming FCX opens tomorrow above the 3425 trigger buy still very much like this setup
11	It really worries me how everyone expects the market to rally now,usually exact opposite happens every time we shall see soon bac spx jpm	1	It really worries me how everyone expects the market to rally nowusually exact opposite happens every time we shall see soon bac spx jpm
12	AAP GAMCO's arry Haverty : Apple Is Extremely Cheap Great Video !!!	1	AAP GAMCOs arry Haverty Apple Is Extremely Cheap Great Video
13	user Maykiljil posted that. I agree that MSFT is going higher & possibly north of 30	1	user Maykiljil posted that I agree that MSFT is going higher possibly north of 30
14	Momentum is coming back to ETFC Broke MA200 resistance on solid volume Friday. ong set-up	1	Momentum is coming back to ETFC Broke MA200 resistance on solid volume Friday ong setup
15	HA Hitting 35.65 means resume targeting 42 level ..	1	HA Hitting 3565 means resume targeting 42 level
16	user gameplan shot for today but I liked on trend break from May or c+h break. OC weekly trend break back to july 2011	1	user gameplan shot for today but I liked on trend break from May or ch break OC weekly trend break back to july 2011
17	with FCX gapping well above ideal entry looking for a pull in to at least 35 on open for an entry	1	with FCX gapping well above ideal entry looking for a pull in to at least 35 on open for an entry
18	user great list again, particularly like FISV and SYK. All buy & hold types should check the free list out.	1	user great list again particularly like FISV and SYK All buy hold types should check the free list out
19	ATHX upper trend line	1	ATHX upper trend line
20	NG - nice PNF BY - breakout - need follow thru	1	NG nice PNF BY breakout need follow thru
21	Won't believe AAP uptrend is back until it crosses above MA(50)	0	Wont believe AAP uptrend is back until it crosses above MA50
22	X swing still on	1	X swing still on
23	SWY - 30% of float short and breaking out - ouch	1	SWY 30 of float short and breaking out ouch
24	BIOF wants 4.90's comin!!!	1	BIOF wants 490s comin

Show 25 ▾ per page

 1 2 10 100 200 230 232

Next steps:

▼ B) REMOVING STOPWORDS

```
# DOWNLOADING STOPWORDS
nltk.download("stopwords")
stopwords.words('english')
```



```
will,  
'just',  
'don',  
"don't",  
'should',  
"should've",  
'now',  
'd',  
'll',  
'm',  
'o',  
're',  
've',  
'y',  
'ain',  
'aren',  
"aren't",  
'couldn',  
"couldn't",  
'didn',  
"didn't",  
'doesn',  
"doesn't",  
'hadn',  
"hadn't",  
'hasn',  
"hasn't",  
'haven',  
"haven't",  
'isn',  
"isn't",  
'ma',  
'mightn',  
"mightn't",  
'mustn',  
"mustn't",  
'needn',  
"needn't",  
'shan',  
"shan't",  
'shouldn',  
"shouldn't",  
'wasn',  
"wasn't",  
'weren',  
"weren't",  
'won',  
"won't",  
'wouldn',  
"wouldn't"]
```

```
from nltk.corpus import stopwords  
stop_words = stopwords.words('english')
```

```
# AFTER LOOKING AT THE DATA, IT WAS OBSERVED THAT SOME WORDS WERE FREQUENTLY REPEATED, THUS ADDING THEM IN THE STOPWORDS LIST (TO BE REMOVED)  
stop_words.extend(['from', 'subject', 'https', 're', 'edu', 'use', 'will', 'aap', 'co', 'day', 'user', 'stock', 'today', 'week', 'year'])
```

```
# DEFINING A FUNCTION TO REMOVE STOP WORDS AND SOME SHORT WORDS TOO (LESS THAN 3 CHARACTERS LONG), SINCE THEY DONT GIVE VALUE TO THE DATA
```

```
def preprocess(text):  
    result = []  
    for token in gensim.utils.simple_preprocess(text):  
        if len(token) >=3 and token not in stop_words:  
            result.append(token)  
    return result
```

```
# APPLYING THE PREPROCESSING FUNCTION:  
stock_df['Text_Pre-processed'] = stock_df['Text-Without-Punctuations'].apply(preprocess)  
stock_df
```

index	Text	Sentiment	Text-Without-Punctuations	Text_Pre-processed
0	Kickers on my watchlist XIDE TIT SOQ PNK CPW BPZ AJ trade method 1 or method 2, see prev posts	1	Kickers on my watchlist XIDE TIT SOQ PNK CPW BPZ AJ trade method 1 or method 2 see prev posts	kickers,watchlist,xide,tit,soq,pnk,cpw,bpz,trade,method,method,see,prev,posts
1	user: AAP MOVIE. 55% return for the FEA/GEED indicator just 15 trades for the year. AWESOME.	1	user AAP MOVIE 55 return for the FEAGEED indicator just 15 trades for the year AWESOME	movie,return,feageed,indicator,trades,awesome
2	user I'd be afraid to short AMZN - they are looking like a near-monopoly in eBooks and infrastructure-as-a-service	1	user Id be afraid to short AMZN they are looking like a nearmonopoly in eBooks and infrastructureasaservice	afraid,short,amzn,looking,like,nearmonopoly,ebooks
3	MNTA Over 12.00	1	MNTA Over 1200	mnta
4	OI Over 21.37	1	OI Over 2137	
5	PGNX Over 3.04	1	PGNX Over 304	pgnx
6	AAP - user if so then the current downtrend will break. Otherwise just a short-term correction in med-term downtrend.	0	AAP user if so then the current downtrend will break Otherwise just a shortterm correction in medterm downtrend	current,downtrend,break,otherwise,shortterm,correction,medterm,downtrend
7	Monday's relative weakness. NYX WIN TIE TAP ICE INT BMC AON C CHK BIIB	0	Mondays relative weakness NYX WIN TIE TAP ICE INT BMC AON C CHK BIIB	mondays,relative,weakness,nyx,win,tie,tap,ice,int,bmc,aon,chk,biib
8	GOOG - over trend line channel test & volume support.	1	GOOG over trend line channel test volume support	goog,over,trend,line,channel,test,volume,support
9	AAP will watch tomorrow for ONG entry.	1	AAP will watch tomorrow for ONG entry	watch,tomorrow,ong,entry
10	i'm assuming FCX opens tomorrow above the 34.25 trigger buy. still very much like this setup.	1	im assuming FCX opens tomorrow above the 3425 trigger buy still very much like this setup	assuming,fcx,opens,tomorrow,trigger,buy,still,much,like,setup
11	It really worries me how everyone expects the market to rally now,usually exact opposite happens every time we shall see soon bac spx jpm	1	It really worries me how everyone expects the market to rally nowusually exact opposite happens every time we shall see soon bac spx jpm	really,worries,everyone,expects,market,rally,nowusually,exact,opposite,happens,every,time,shall,see,soon,ba

	AAP GAMCO's arry Haverty : 12 Apple Is Extremely Cheap Great Video !!!	1	AAP GAMCOs arry Haverty Apple Is Extremely Cheap Great Video	gamcos,arry,haverty,apple,extremely,cheap,great,video
13	user Maykiljil posted that. I agree that MSFT is going higher & possibly north of 30	1	user Maykiljil posted that I agree that MSFT is going higher possibly north of 30	maykiljil,posted,agree,msft,going,higher,possibly,north
14	Momentum is coming back to ETFC Broke MA200 resistance on solid volume Friday, ong set-up	1	Momentum is coming back to ETFC Broke MA200 resistance on solid volume Friday ong setup	momentum,coming,back,etfc,broke,resistance,solid,volume,friday,ong,setup
15	HA Hitting 35.65 means resume targeting 42 level ..	1	HA Hitting 3565 means resume targeting 42 level	hitting,means,resume,targeting,level
16	user gameplan shot for today but I liked on trend break from May or c+h break. OC weekly trend break back to july 2011	1	user gameplan shot for today but I liked on trend break from May or ch break OC weekly trend break back to july 2011	gameplan,shot,liked,trend,break,may,break,weekly,trend,break,back,july
17	with FCX gapping well above ideal entry looking for a pull in to at least 35 on open for an entry	1	with FCX gapping well above ideal entry looking for a pull in to at least 35 on open for an entry	fcx,gapping,well,ideal,entry,looking,pull,least,open,entry
18	user great list again, particularly like FISV and SYK. All buy & hold types should check the free list out.	1	user great list again particularly like FISV and SYK All buy hold types should check the free list out	great,list,particularly,like,fisv,syk,buy,hold,types,check,free,list
19	ATHX upper trend line	1	ATHX upper trend line	athx,upper,trend,line
20	NG - nice PNF BY - breakout - need follow thru	1	NG nice PNF BY breakout need follow thru	nice,pnf,breakout,need,follow,thru
21	Won't believe AAP uptrend is back until it crosses above MA(50)	0	Wont believe AAP uptrend is back until it crosses above MA50	wont,believe,uptrend,back,crosses

Next steps:

[Generate code with stock_df](#)[View recommended plots](#)

▼ STEP 4: PLOTTING WORDCLOUD

```
# JOINING THE PREPROCESSED TEXT INTO A STRING
stock_df['Text_Pre-processed-joined'] = stock_df['Text_Pre-processed'].apply(lambda x: " ".join(x))
stock_df
```



1 to 25 of 5791 entries

Filter



index	Text	Sentiment	Text-Without-Punctuations
0	Kickers on my watchlist XIDE TIT SOQ PNK CPW BPZ AJ trade method 1 or method 2, see prev posts	1	Kickers on my watchlist XIDE TIT SOQ PNK CPW BPZ AJ trade method 1 or method 2 see prev posts
1	user: AAP MOVIE. 55% return for the FEA/GEED indicator just 15 trades for the year. AWESOME.	1	user AAP MOVIE 55 return for the FEAGEED indicator just 15 trades for the year AWESOME
2	user I'd be afraid to short AMZN - they are looking like a near-monopoly in eBooks and infrastructure-as-a-service	1	user Id be afraid to short AMZN they are looking like a nearmonopoly in eBooks and infrastructureasaservice
3	MNTA Over 12.00	1	MNTA Over 1200
4	OI Over 21.37	1	OI Over 2137
5	PGNX Over 3.04	1	PGNX Over 304
6	AAP - user if so then the current downtrend will break. Otherwise just a short-term correction in med-term downtrend.	0	AAP user if so then the current downtrend will break Otherwise just a shortterm correction in medterm downtrend
7	Monday's relative weakness. NYX WIN TIE TAP ICE INT BMC AON C CHK BIIB	0	Mondays relative weakness NYX WIN TIE TAP ICE INT BMC AON C CHK BIIB
8	GOOG - over trend line channel test & volume support.	1	GOOG over trend line channel test volume support
9	AAP will watch tomorrow for ONG entry.	1	AAP will watch tomorrow for ONG entry
10	i'm assuming FCX opens tomorrow above the 34.25 trigger buy. still very much like this setup.	1	im assuming FCX opens tomorrow above the 3425 trigger buy still very much like this setup
11	It really worries me how everyone expects the market to rally now, usually exact opposite happens every time we shall see soon bac spx jpm	1	It really worries me how everyone expects the market to rally nowusually exact opposite happens every time we shall see soon bac spx jpm

12	AAP GAMCO's arry Haverty : Apple Is Extremely Cheap Great Video !!!	1	AAP GAMCOs arry Haverty Apple Is Extremely Cheap Great Video	gamcos,arry,haverty,apple,extremely,cheap,gr
13	user Maykiljil posted that. I agree that MSFT is going higher & possibly north of 30	1	user Maykiljil posted that I agree that MSFT is going higher possibly north of 30	maykiljil,posted,agree,msft,going,higher,possit
14	Momentum is coming back to ETFC Broke MA200 resistance on solid volume Friday, ong set-up	1	Momentum is coming back to ETFC Broke MA200 resistance on solid volume Friday ong setup	momentum,coming,back,etfc,broke,resistance
15	HA Hitting 35.65 means resume targeting 42 level ..	1	HA Hitting 3565 means resume targeting 42 level	hitting,means,resume,targeting,level
16	user gameplan shot for today but I liked on trend break from May or c+h break. OC weekly trend break back to july 2011	1	user gameplan shot for today but I liked on trend break from May or ch break OC weekly trend break back to july 2011	gameplan,shot,liked,trend,break,may,break,we
17	with FCX gapping well above ideal entry looking for a pull in to at least 35 on open for an entry	1	with FCX gapping well above ideal entry looking for a pull in to at least 35 on open for an entry	fcx,gapping,well,ideal,entry,looking,pull,least,o
18	user great list again, particularly like FISV and SYK. All buy & hold types should check the free list out.	1	user great list again particularly like FISV and SYK All buy hold types should check the free list out	great,list,particularly,like,fisv,syk,buy,hold,type:
19	ATHX upper trend line	1	ATHX upper trend line	athx,upper,trend,line
20	NG - nice PNF BY - breakout - need follow thru	1	NG nice PNF BY breakout need follow thru	nice,pnf,breakout,need,follow,thru
21	Won't believe AAP uptrend is back until it crosses above MA(50)	0	Wont believe AAP uptrend is back until it crosses above MA50	wont,believe,uptrend,back,crosses
22	X swing still on	1	X swing still on	swing,still
23	SWY - 30% of float short and breaking out - ouch	1	SWY 30 of float short and breaking out ouch	swy,float,short,breaking,ouch
24	BIOF wants 4.90's comin!!!	1	BIOF wants 490s comin	biof,wants,comin

Show per page

1

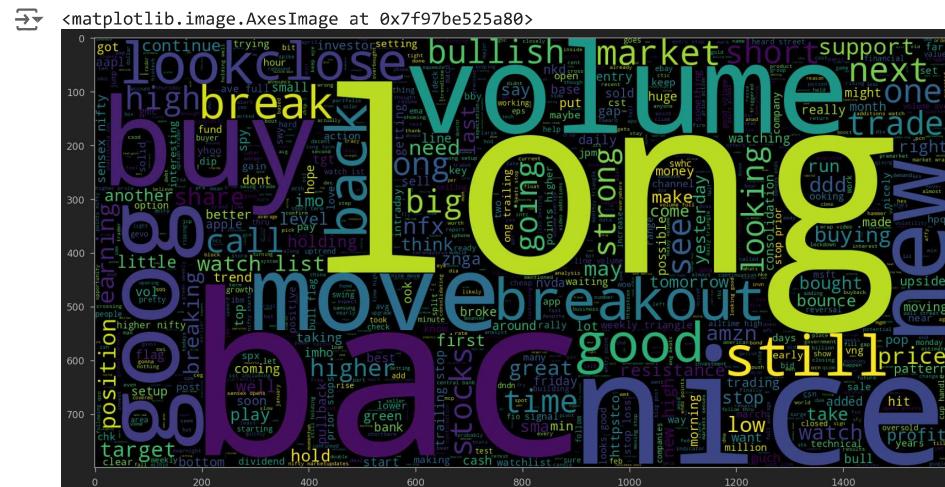
2 10 100 200 230 232

Like what you see? Visit the [data table notebook](#) to learn more about interactive tables.

Next steps: [Generate code with stock_dt](#)

View recommended plots

```
# PLOTTING A WORDCLOUD:  
plt.figure(figsize=(20,20))  
wc = WordCloud(max_words = 1000, width = 1600, height = 800).generate(" ".join(stock_df[stock_df['Sentiment']==1]['Text_Pre-processed-joined'])  
plt.imshow(wc)
```



▼ STEP 5: VISUALIZING CLEANED DATASET

```
nltk.download('punkt')
```

```
[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
True
```

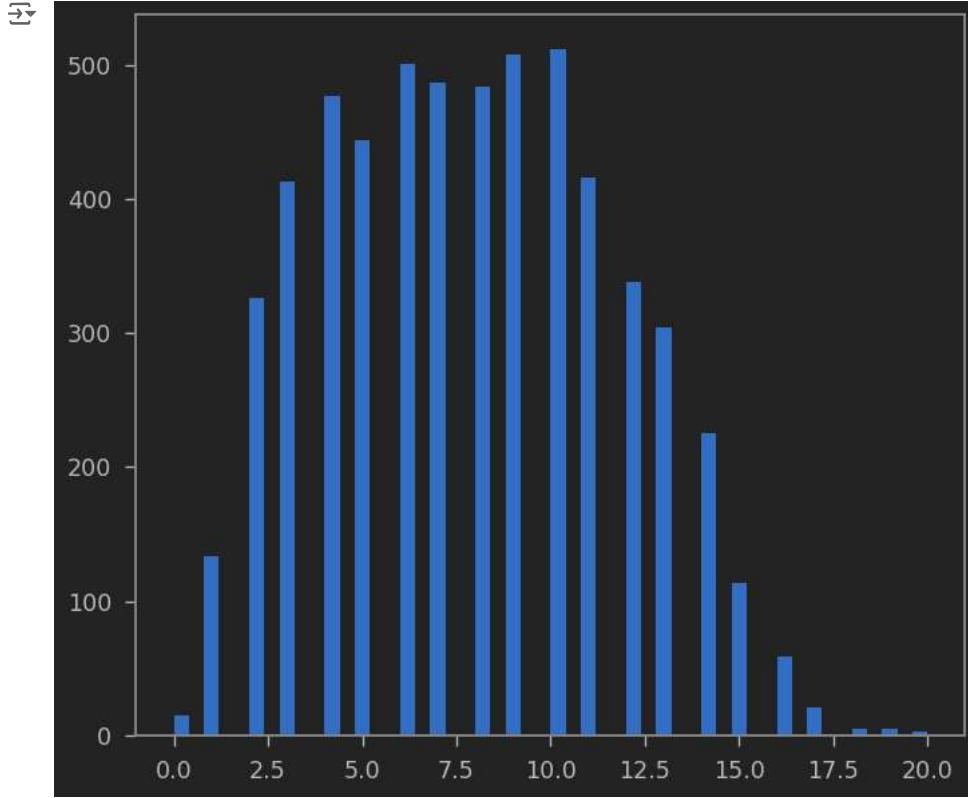
```
# TOKENIZING THE TEXT INTO WORDS
```

```
tweets_length = [ len(nltk.word_tokenize(x)) for x in stock_df['Text_Pre-processed-joined'] ]  
tweets_length
```

11

```
10,  
13,  
9,  
3,  
2,  
11,  
12,  
4,  
4,  
7,  
14,  
7,  
14,  
7,  
3,  
7,  
8,  
17,  
9,  
14,  
11,  
10,  
6,  
10,  
2,  
15,  
10,  
4,  
3,  
4,  
6,  
10,  
10,  
11,  
...]
```

```
# CREATING A HISTOGRAM TO VISUALIZE THE TOKENIZED DATA:  
plt.figure()  
plt.hist(tweets_length, bins =50)  
plt.show()
```



```
# THUS WE CAN OBSERVE THAT MOST OF OUR TWEETS EXIST BETWEEN 5 - 10 WORDS (THE HISTOGRAM RESEMBLES A BELL-CURVE)
```

▼ STEP 6: PREPARING THE DATA BY TOKENIZING AND PADDING

```
# A) TOKENIZING WORKS BY CONVERTING EACH TEXT INTO A SEQUENCE OF INTEGERS
```

```
# OBTAINING THE TOTAL NUMBER OF UNIQUE WORDS IN OUR DATASET:
```

```
list_of_words = []

for i in stock_df['Text_Pre-processed']:
    for j in i:
        list_of_words.append(j)

total_words = len(set(list_of_words))
total_words
```

9268

```
# SPLITTING DATA INTO TRAIN AND TEST
```

```
X = stock_df['Text_Pre-processed']
y = stock_df['Sentiment']

from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2)
```

```
X_train.shape
```

(4632,)

```
X_test.shape
```

(1159,)

```
X_train
```

```
4020 [bac, accumulated, may, calls, ooking, move, u...
5324 [srussolillo, company, wouldn't, admit, somethi...
3077 [innovation, dead, best, selling, computer, ip...
5640 [sensex, ends, points, higher, nifty, reclaims...
3855 [bac, gonna, boom]
...
5745 [rupee, hits, record, low, dollar, amid, coron...
2109 [james, white, jamba, juicy, turnaround, vimeo...
2164 [see, exit, csn]
2201 [ddd, soda, look, interesting, wait, earnings...
3494 [ndaq, triggered, pennies, yesterday, continue...
Name: Text_Pre-processed, Length: 4632, dtype: object
```

```
# CREATING A TOKENIZER TO TOKENIZE THE WORDS AND CREATE A SEQUENCE OF TOKENIZED WORDS (THAT CAN BE INTERPRETTED BY OUR MACHINE LEARNING ALG)
```

```
tokenizer = Tokenizer(num_words = total_words)

tokenizer.fit_on_texts(X_train)

train_sequences = tokenizer.texts_to_sequences(X_train)
test_sequences = tokenizer.texts_to_sequences(X_test)
```

```
train_sequences
```

...

```
[1586, 4313, 1, 55, 20, 4/4, 13, 185],
[4314, 83, 35],
[9, 4315, 13, 120],
[467, 45, 4, 4316, 1573],
[669, 1587, 598, 12, 1131, 12, 132],
[1319, 1983, 54, 191, 4317, 4318, 242, 1588],
[33, 2644, 36, 9, 75],
[115, 4319],
[4320, 976, 788, 458, 4321, 4322],
[191, 4323, 7, 4324, 2645],
[2646, 106, 794, 306, 4325, 381, 106],
[110, 4326, 1825, 929, 4327, 460, 1589, 1487],
[222, 2, 34, 539, 365, 594, 824, 113, 7, 55, 641],
[249, 40, 1117, 201, 81, 312, 749, 4328, 2, 568, 1286, 69, 55, 20],
[36, 1590],
[857, 1086, 22, 669, 1984, 19, 1591, 150, 113, 1435, 3, 96, 4329],
[576, 1592],
[519, 1, 378],
[2647,
292,
1593,
1985,
2648,
2649,
752,
4330,
508,
984,
4331,
1272,
4332,
4333],
[1986, 10, 429, 112, 1268, 1574, 11, 42],
[842, 316, 1101, 4334, 141, 566, 44],
[1752, 2650, 4335, 1215, 4336, 250],
[1987, 4337, 514],
...]
```

B) PADDING THE TRAINING AND TESTING DATA

```
# BASICALLY MAKING ALL OF THE DIGITS IN OUR TRAIN AND TEST DATA TO EQUAL LENGTH IN ORDER TO MAKE THE DEEP LEARNING MODEL MORE EFFICIENT (WE
padded_train = pad_sequences(train_sequences, maxlen=15, padding = 'post', truncating = 'post')
padded_test = pad_sequences(test_sequences, maxlen=15, padding = 'post', truncating = 'post')
```

CONVERTING DATA INTO CATEGORICAL 2D REPRESENTATIONS:

```
y_train_cat = to_categorical(y_train, 2)
y_test_cat = to_categorical(y_test, 2)
```

```
y_train_cat.shape
y_test_cat.shape
```

```
array((1159, 2))
```

```
y_test_cat
```

```
array([[0., 1.],
       [0., 1.],
       [1., 0.],
       ...,
       [0., 1.],
       [0., 1.],
       [1., 0.]], dtype=float32)
```

```
# THE IDEA OF CONVERTING Y_TRAIN AND y_test to categorical 2D REPRESENTATION IS TO MAKE SURE THAT THE OUTPUT WOULD HAVE TWO NEURONS
# THIS IS BECAUSE BASICALLY WE HAVE 2 CLASSES I.E. 0, AND 1
# SO, IF NEURON 1 WERE TO FIRE, WE'LL GET THE OUTPUT AS 1 AND 0; IF NEURON 2 WERE TO FIRE, WE'LL GET THE OUTPUT AS 0 AND 1
```

▼ STEP 7: BUILDING A CUSTOM-BASED DEEP NEURAL NETWORK TO PERFORM SENTIMENT ANALYSIS

```

model = Sequential()

model.add(Embedding(total_words, output_dim = 512)) # ADDING EMBEDDING LAYER

# BI-DIRECTIONAL RNN AND LSTM
model.add(LSTM(256))

model.add(Dense(128, activation = 'relu')) # ADDING A DENSE LAYER WITH 2 NEURONS AND ACTIVATION FUNCTION AS 'RELU'
model.add(Dropout(0.3)) # REGULARIZATION PARAMETER (TO PREVENT OVERFITTING)

model.add(Dense(2, activation = 'softmax')) # ADDING DENSE LAYER WITH 2 NEURONS AND ACTIVATION FUNCTION AS 'SOFTMAX'
model.compile(optimizer = 'adam', loss = 'binary_crossentropy', metrics = ['acc']) # COMPILING THE MODEL WITH OPTIMIZER - 'ADAM' AND WE WI
model.summary()

```

⤵ Model: "sequential"

Layer (type)	Output Shape	Param #
<hr/>		
embedding (Embedding)	(None, None, 512)	4745216
lstm (LSTM)	(None, 256)	787456
dense (Dense)	(None, 128)	32896
dropout (Dropout)	(None, 128)	0
dense_1 (Dense)	(None, 2)	258
<hr/>		
Total params: 5565826 (21.23 MB)		
Trainable params: 5565826 (21.23 MB)		
Non-trainable params: 0 (0.00 Byte)		

```
# TRAINING THE MODEL:
```

```
model.fit(padded_train, y_train_cat, batch_size =32, validation_split=0.2, epochs = 3)
```

⤵ Epoch 1/3
116/116 [=====] - 29s 228ms/step - loss: 0.5923 - acc: 0.6928 - val_loss: 0.4700 - val_acc: 0.7594
Epoch 2/3
116/116 [=====] - 24s 210ms/step - loss: 0.3129 - acc: 0.8742 - val_loss: 0.5359 - val_acc: 0.7540
Epoch 3/3
116/116 [=====] - 27s 228ms/step - loss: 0.1593 - acc: 0.9449 - val_loss: 0.5128 - val_acc: 0.7584
<keras.src.callbacks.History at 0x7f97585db070>

```
# THUS WE ACHIEVED AN ACCURACY OF 94.82% AND VALIDATION-ACCURACY AS 77.13%
```

▼ STEP 8: ASSESING THE TRAINING MODEL PERFORMANCE

```
pred = model.predict(padded_test)
```

```
prediction = []
for i in pred:
    prediction.append(np.argmax(i))

prediction
```

⤵

```
original = []
for i in y_test_cat:
    original.append(np.argmax(i))

original
```

```
^,  
1,  
0,  
1,  
1,  
1,  
0,  
0,  
1,  
1,  
0,  
0,  
1,  
0,  
1,  
1,  
0,  
1,
```

```
# CALCULATING ACCURACY SCORE FOR UNSEEN DATA (TEST DATA)  
from sklearn.metrics import accuracy_score, confusion_matrix, recall_score, precision_score, roc_auc_score, f1_score
```

```
accuracy = accuracy_score(original, prediction)  
accuracy
```

```
⤵ 0.7558239861949957
```

```
f1 = f1_score(original, prediction)  
precision = precision_score(original, prediction)  
recall = recall_score(original, prediction)  
auc = roc_auc_score(original, prediction)
```

```
print("f1 for the model is: {}".format(f1))  
print("Precision for the model is: {}".format(precision))  
print("Recall for the model is: {}".format(recall))  
print("AUC for the model is: {}".format(auc))
```

```
⤵ f1 for the model is: 0.8112074716477652  
Precision for the model is: 0.7845161290322581  
Recall for the model is: 0.8397790055248618  
AUC for the model is: 0.7279354797739253
```

```
cm = confusion_matrix(original, prediction)  
  
sns.heatmap(cm, annot = True)
```

```
⤵ <Axes: >
```

