In [229]:
```
#Intalling textblob library
!pip install textblob
```

```
WARNING: Ignoring invalid distribution -umpy (c:\users\lenovo\anacond
a3\lib\site-packages)
WARNING: Ignoring invalid distribution -umpy (c:\users\lenovo\anacond
a3\lib\site-packages)
WARNING: Ignoring invalid distribution -umpy (c:\users\lenovo\anacond
a3\lib\site-packages)

Requirement already satisfied: textblob in c:\users\lenovo\anaconda3\
lib\site-packages (0.15.3)
Requirement already satisfied: nltk>=3.1 in c:\users\lenovo\anaconda
3\lib\site-packages (from textblob) (3.5)
Requirement already satisfied: joblib in c:\users\lenovo\anaconda3\li
b\site-packages (from nltk>=3.1->textblob) (0.16.0)
Requirement already satisfied: click in c:\users\lenovo\anaconda3\li
b\site-packages (from nltk>=3.1->textblob) (7.1.2)
Requirement already satisfied: regex in c:\users\lenovo\anaconda3\li
b\site-packages (from nltk>=3.1->textblob) (2020.6.8)
Requirement already satisfied: tqdm in c:\users\lenovo\anaconda3\lib\
site-packages (from nltk>=3.1->textblob) (4.47.0)

WARNING: Ignoring invalid distribution -umpy (c:\users\lenovo\anacond
a3\lib\site-packages)
WARNING: Ignoring invalid distribution -umpy (c:\users\lenovo\anacond
a3\lib\site-packages)
WARNING: Ignoring invalid distribution -umpy (c:\users\lenovo\anacond
a3\lib\site-packages)
WARNING: You are using pip version 21.1.2; however, version 21.2.4 is
available.
You should consider upgrading via the 'c:\users\lenovo\anaconda3\pyth
on.exe -m pip install --upgrade pip' command.
```

In [9]:
```
#importing necessary libraries

import pandas as pd
from wordcloud import WordCloud, STOPWORDS
from PIL import Image
from nltk.sentiment.vader import SentimentIntensityAnalyzer
#from langdetect import detect
from nltk.stem import SnowballStemmer
from nltk.sentiment.vader import SentimentIntensityAnalyzer
from sklearn.feature_extraction.text import CountVectorizer
from textblob import TextBlob
```

In [13]:
```
dataframe = pd.read_csv("AppleTwitterData20_21.csv")
dataframe.head()
dataframe.columns
```

Out[13]:
```
Index(['Datetime', 'Text'], dtype='object')
```

In [14]:
```
dataframe.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 33368 entries, 0 to 33367
Data columns (total 2 columns):
 #   Column    Non-Null Count  Dtype 
---  ------    --------------  ----- 
 0   Datetime  33368 non-null  object
 1   Text      33367 non-null  object
dtypes: object(2)
memory usage: 521.5+ KB
```

In [35]:
```
df = dataframe.sort_values(by='Datetime',ignore_index=True)
df
```

Out[35]:

	Datetime	Text
0		NaN
1	2020-01-01 00:24:13+00:00	4 & 1 hour charts for all 78 instruments ...
2	2020-01-01 00:56:57+00:00	Top 10 trades of the decade: Number 4: Going l...
3	2020-01-01 01:24:41+00:00	4 & 1 hour charts for the Group 3 instrume...
4	2020-01-01 01:26:34+00:00	4 & 1 hour charts for all 78 instruments a...
...	...	...
33363	2021-09-20 22:10:27+00:00	AAPL 20210917 Weekly Price Pattern Coordinates...
33364	2021-09-20 22:15:53+00:00	Apple Inc price at close, 2021-09-20, is 142.9...
33365	2021-09-20 23:15:03+00:00	@MacRumors @julipuli Now, if only any of us ou...
33366	Datetime	Text
33367	Datetime	Text

33368 rows × 2 columns

In [36]:
```
df.drop_duplicates(keep='first')
df.head()
```

Out[36]:

	Datetime	Text
0		NaN
1	2020-01-01 00:24:13+00:00	4 & 1 hour charts for all 78 instruments ...
2	2020-01-01 00:56:57+00:00	Top 10 trades of the decade: Number 4: Going l...
3	2020-01-01 01:24:41+00:00	4 & 1 hour charts for the Group 3 instrume...
4	2020-01-01 01:26:34+00:00	4 & 1 hour charts for all 78 instruments a...

In [37]:
```
df.shape
```

Out[37]: (33368, 2)

In [38]:
```
df.isnull().sum()
```

Out[38]:
```
Datetime    0
Text        1
dtype: int64
```

In [39]:
```
df.dropna()
```

Out[39]:

	Datetime	Text
1	2020-01-01 00:24:13+00:00	4 & 1 hour charts for all 78 instruments ...
2	2020-01-01 00:56:57+00:00	Top 10 trades of the decade: Number 4: Going l...
3	2020-01-01 01:24:41+00:00	4 & 1 hour charts for the Group 3 instrume...
4	2020-01-01 01:26:34+00:00	4 & 1 hour charts for all 78 instruments a...
5	2020-01-01 01:30:11+00:00	Made one big mistake Day Trading Apple (AAPL) ...
...	...	...
33363	2021-09-20 22:10:27+00:00	AAPL 20210917 Weekly Price Pattern Coordinates...
33364	2021-09-20 22:15:53+00:00	Apple Inc price at close, 2021-09-20, is 142.9...
33365	2021-09-20 23:15:03+00:00	@MacRumors @julipuli Now, if only any of us ou...
33366	Datetime	Text
33367	Datetime	Text

33367 rows × 2 columns

In [40]:
```
def preprocessing_text(df):
    #put everythin in lowercase
    df['Text'] = df['Text'].str.lower()
    #Replace rt indicating that was a retweet
    df['Text'] = df['Text'].str.replace('rt', '')
    #Replace occurences of mentioning @UserNames
    df['Text'] = df['Text'].replace(r'@\w+', '', regex=True)
    #Replace links contained in the tweet
    df['Text']= df['Text'].replace(r'http\S+', '', regex=True)
    df['Text'] = df['Text'].replace(r'www.[^ ]+', '', regex=True)
    #remove numbers
    df['Text'] = df['Text'].replace(r'[0-9]+', '', regex=True)
    #replace special characters and puntuation marks
    df['Text'] = df['Text'].replace(r'[!"#$%&()*+,-./:;<=>?@[\]^_`{|}
~]\n', '', regex=True)
    return df
```

In [41]:
```
preprocessing_text(df)
```

Out[41]:

	Datetime	Text
0		NaN
1	2020-01-01 00:24:13+00:00	& hour chas for all instruments are av...
2	2020-01-01 00:56:57+00:00	top trades of the decade: number : going long...
3	2020-01-01 01:24:41+00:00	& hour chas for the group instruments a...
4	2020-01-01 01:26:34+00:00	& hour chas for all instruments are ava...
...	...	...
33363	2021-09-20 22:10:27+00:00	aapl weekly price pattern coordinates\nannota...
33364	2021-09-20 22:15:53+00:00	apple inc price at close, --, is .. #apple #aapl
33365	2021-09-20 23:15:03+00:00	now, if only any of us outside the us could ...
33366	Datetime	text
33367	Datetime	text

33368 rows × 2 columns

In [42]:
```
df.dropna()
```

Out[42]:

	Datetime	Text
1	2020-01-01 00:24:13+00:00	& hour chas for all instruments are av...
2	2020-01-01 00:56:57+00:00	top trades of the decade: number : going long...
3	2020-01-01 01:24:41+00:00	& hour chas for the group instruments a...
4	2020-01-01 01:26:34+00:00	& hour chas for all instruments are ava...
5	2020-01-01 01:30:11+00:00	made one big mistake day trading apple (aapl) ...
...	...	...
33363	2021-09-20 22:10:27+00:00	aapl weekly price pattern coordinates\nannota...
33364	2021-09-20 22:15:53+00:00	apple inc price at close, --, is .. #apple #aapl
33365	2021-09-20 23:15:03+00:00	now, if only any of us outside the us could ...
33366	Datetime	text
33367	Datetime	text

33367 rows × 2 columns

In [43]:
```
df.head(20)
```

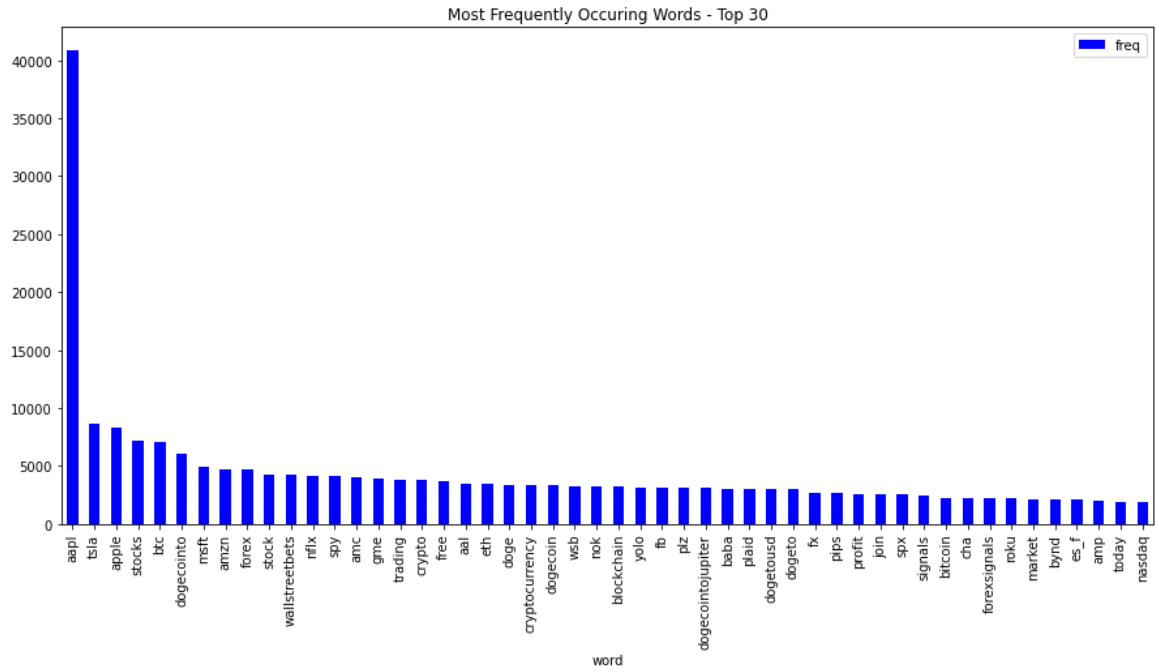
Out[43]:

	Datetime	Text
0		NaN
1	2020-01-01 00:24:13+00:00	& hour chas for all instruments are av...
2	2020-01-01 00:56:57+00:00	top trades of the decade: number : going long...
3	2020-01-01 01:24:41+00:00	& hour chas for the group instruments a...
4	2020-01-01 01:26:34+00:00	& hour chas for all instruments are ava...
5	2020-01-01 01:30:11+00:00	made one big mistake day trading apple (aapl) ...
6	2020-01-01 07:32:26+00:00	was a great year for me.. thanks to the unive...
7	2020-01-01 09:06:20+00:00	total returns..bitcoin: +nasdaq : +% ...
8	2020-01-01 16:40:06+00:00	made one big mistake day trading apple (aapl) ...
9	2020-01-01 16:45:09+00:00	$spy\n\n staed - what's next\n- i staed lookin...
10	2020-01-01 16:52:00+00:00	time for a wrap on my strava stats. 🚴+🏃= ,mi...
11	2020-01-01 17:40:00+00:00	#aapl - aapl +% en el - tradingview -
12	2020-01-01 17:46:33+00:00	happy productivity in with #pdfzone #app #mac...
13	2020-01-01 19:38:19+00:00	$es_f thu jobless claims/pmi\ngap up & go ...
14	2020-01-01 19:46:17+00:00	$es_f\n\nif there is a gap up open today (... ...
15	2020-01-01 19:52:39+00:00	$es_f\n\nhourly cha\n\nlook like upside is not...
16	2020-01-01 19:53:50+00:00	es_f\n\nanother view.\n#es_fspx #trading #f...
17	2020-01-01 19:59:02+00:00	$es_f\n\nso, for a bear case..\ngap was paiall...
18	2020-01-01 20:36:24+00:00	$es_f\n\nit look like a toss up\n\n\nor\n\n\n-...
19	2020-01-01 20:47:14+00:00	$aapl #aapl new years fireworks coming? buyers...

In [44]:
```
df = df.dropna()
```

In [45]:
```
from sklearn.feature_extraction.text import CountVectorizer
import matplotlib.pyplot as plt


cv = CountVectorizer(stop_words = 'english')
words = cv.fit_transform(df.Text)

sum_words = words.sum(axis=0)

words_freq = [(word, sum_words[0, i]) for word, i in cv.vocabulary_.ite
ms()]
words_freq = sorted(words_freq, key = lambda x: x[1], reverse = True)

frequency = pd.DataFrame(words_freq, columns=['word', 'freq'])

frequency.head(50).plot(x='word', y='freq', kind='bar', figsize=(15,
7), color = 'blue')
plt.title("Most Frequently Occuring Words - Top 30")
```

Out[45]: Text(0.5, 1.0, 'Most Frequently Occuring Words - Top 30')



In [46]:
```
from wordcloud import WordCloud

wordcloud = WordCloud(background_color = 'white', width = 1000, height
= 1000).generate_from_frequencies(dict(words_freq))

plt.figure(figsize=(10,8))
plt.imshow(wordcloud)
plt.title("WordCloud - Vocabulary from Reviews", fontsize = 22)
```

Out[46]:
```
Text(0.5, 1.0, 'WordCloud - Vocabulary from Reviews')
```

In [47]:
```
def getSubjectivity(df):
    return TextBlob(df).sentiment.subjectivity

def getPolarity(df):
    return TextBlob(df).sentiment.polarity

df['Subjectivity'] = df['Text'].apply(getSubjectivity)
df['Polarity'] = df['Text'].apply(getPolarity)

df
```

```
<ipython-input-47-72966b48b876>:7: 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/panda
s-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df['Subjectivity'] = df['Text'].apply(getSubjectivity)
<ipython-input-47-72966b48b876>:8: 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/panda
s-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df['Polarity'] = df['Text'].apply(getPolarity)
```

Out[47]:

	Datetime	Text	Subjectivity	Polarity
1	2020-01-01 00:24:13+00:00	& hour chas for all instruments are av...	0.400000	0.400000
2	2020-01-01 00:56:57+00:00	top trades of the decade: number : going long...	0.450000	0.218750
3	2020-01-01 01:24:41+00:00	& hour chas for the group instruments a...	0.400000	0.400000
4	2020-01-01 01:26:34+00:00	& hour chas for all instruments are ava...	0.400000	0.400000
5	2020-01-01 01:30:11+00:00	made one big mistake day trading apple (aapl) ...	0.100000	0.000000
...	...	...	...	...
33363	2021-09-20 22:10:27+00:00	aapl weekly price pattern coordinates\nannota...	0.000000	0.000000
33364	2021-09-20 22:15:53+00:00	apple inc price at close, --, is .. #apple #aapl	0.000000	0.000000
33365	2021-09-20 23:15:03+00:00	now, if only any of us outside the us could ...	0.395139	-0.010417
33366	Datetime	text	0.000000	0.000000
33367	Datetime	text	0.000000	0.000000

33367 rows × 4 columns

In [49]:
```
def getAnalysis(score):
    if score < 0:
        return 'Negative'
    elif score == 0:
        return 'Neutral'
    else:
        return 'Positive'

df['Analysis'] = df['Polarity'].apply(getAnalysis)
df
```

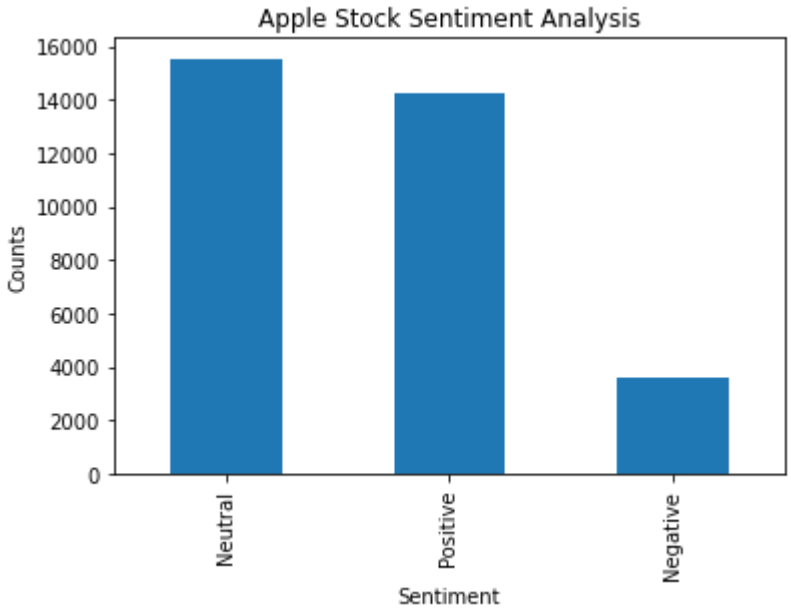
```
<ipython-input-49-657b92d30cda>:9: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/panda
s-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  df['Analysis'] = df['Polarity'].apply(getAnalysis)
```

Out[49]:

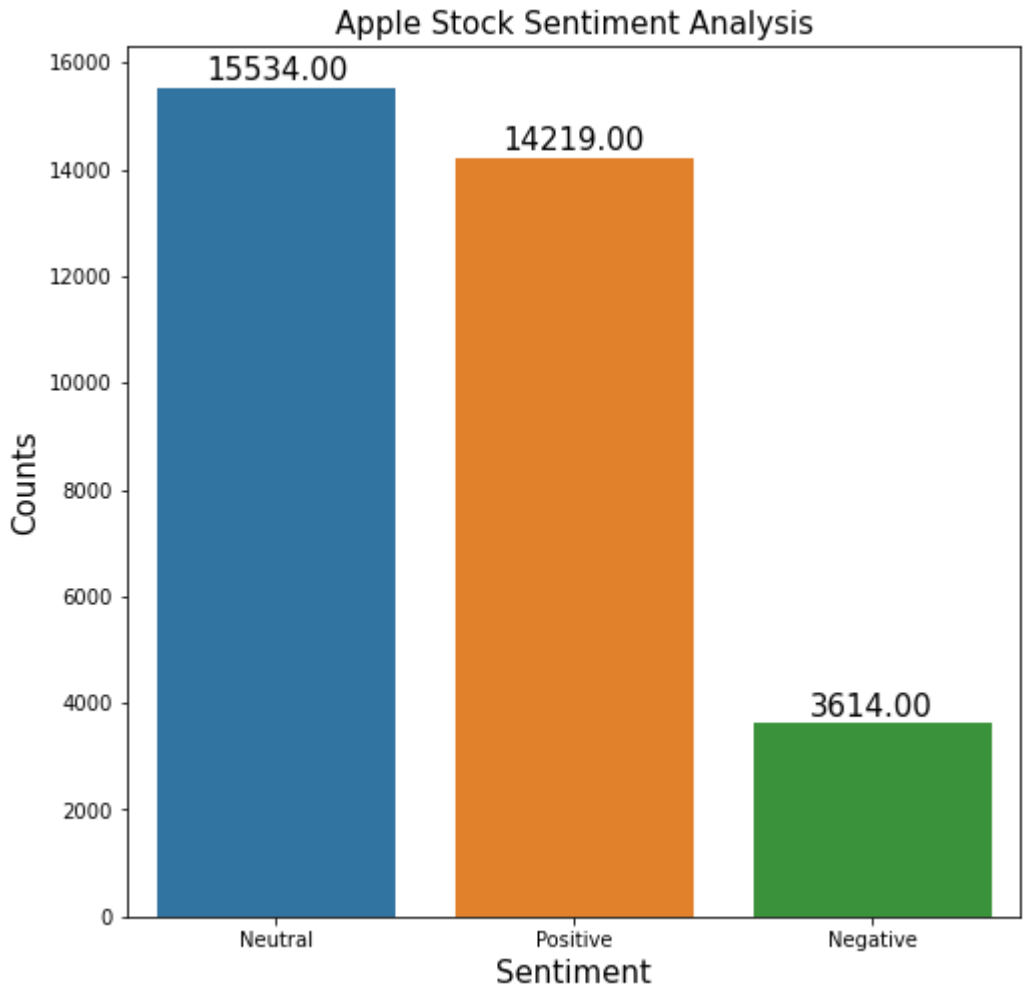
	Datetime	Text	Subjectivity	Polarity	Analysis
1	2020-01-01 00:24:13+00:00	& hour chas for all instruments are av...	0.400000	0.400000	Positive
2	2020-01-01 00:56:57+00:00	top trades of the decade: number : going long...	0.450000	0.218750	Positive
3	2020-01-01 01:24:41+00:00	& hour chas for the group instruments a...	0.400000	0.400000	Positive
4	2020-01-01 01:26:34+00:00	& hour chas for all instruments are ava...	0.400000	0.400000	Positive
5	2020-01-01 01:30:11+00:00	made one big mistake day trading apple (aapl) ...	0.100000	0.000000	Neutral
...	...	...	...	...	...
33363	2021-09-20 22:10:27+00:00	aapl weekly price pattern coordinates\nannota...	0.000000	0.000000	Neutral
33364	2021-09-20 22:15:53+00:00	apple inc price at close, --, is .. #apple #aapl	0.000000	0.000000	Neutral
33365	2021-09-20 23:15:03+00:00	now, if only any of us outside the us could ...	0.395139	-0.010417	Negative
33366	Datetime	text	0.000000	0.000000	Neutral
33367	Datetime	text	0.000000	0.000000	Neutral

33367 rows × 5 columns

In [50]:
```
df['Analysis'].value_counts()

plt.title('Apple Stock Sentiment Analysis')
plt.xlabel('Sentiment')
plt.ylabel('Counts')
df['Analysis'].value_counts().plot(kind = 'bar')
plt.show()
```



In [51]:
```
df_status = df.groupby('Analysis').size().reset_index(name="Counts").so
rt_values(by="Counts",ascending=False)

plt.figure(figsize=(8, 8))
plots = sns.barplot(x="Analysis", y="Counts", data=df_status)
for bar in plots.patches:

  # passing the coordinates where the annotation shall be done. x-coord
inate: bar.get_x() + bar.get_width() / 2, y-coordinate: bar.get_height
()
  # free space to be left to make graph pleasing: (0, 8)  # ha and va s
tand for the horizontal and vertical alignment
    plots.annotate(format(bar.get_height(), '.2f'),
                   (bar.get_x() + bar.get_width() / 2,
                    bar.get_height()), ha='center', va='center',
                   size=15, xytext=(0, 8),
                   textcoords='offset points')
plt.xlabel("Sentiment", size=15)
plt.ylabel("Counts",size=15)
plt.title("Apple Stock Sentiment Analysis", size = 15)
plt.show()
```



In [53]:
```
df1 = pd.get_dummies(df['Analysis'])
```

In [54]:
```
df1
```

Out[54]:

	Negative	Neutral	Positive
1	0	0	1
2	0	0	1
3	0	0	1
4	0	0	1
5	0	1	0
...	...	...	...
33363	0	1	0
33364	0	1	0
33365	1	0	0
33366	0	1	0
33367	0	1	0

33367 rows × 3 columns

In [55]:
```
df
```

Out[55]:

	Datetime	Text	Subjectivity	Polarity	Analysis
1	2020-01-01 00:24:13+00:00	& hour chas for all instruments are av...	0.400000	0.400000	Positive
2	2020-01-01 00:56:57+00:00	top trades of the decade: number : going long...	0.450000	0.218750	Positive
3	2020-01-01 01:24:41+00:00	& hour chas for the group instruments a...	0.400000	0.400000	Positive
4	2020-01-01 01:26:34+00:00	& hour chas for all instruments are ava...	0.400000	0.400000	Positive
5	2020-01-01 01:30:11+00:00	made one big mistake day trading apple (aapl) ...	0.100000	0.000000	Neutral
...	...	...	...	...	...
33363	2021-09-20 22:10:27+00:00	aapl weekly price pattern coordinates\nannota...	0.000000	0.000000	Neutral
33364	2021-09-20 22:15:53+00:00	apple inc price at close, --, is .. #apple #aapl	0.000000	0.000000	Neutral
33365	2021-09-20 23:15:03+00:00	now, if only any of us outside the us could ...	0.395139	-0.010417	Negative
33366	Datetime	text	0.000000	0.000000	Neutral
33367	Datetime	text	0.000000	0.000000	Neutral

33367 rows × 5 columns

In [56]:
```
df2 = pd.concat([df, df1], axis = 1)
```

In [57]:
```
df2
```

Out[57]:

	Datetime	Text	Subjectivity	Polarity	Analysis	Negative	Neutral
1	2020-01-01 00:24:13+00:00	& hour chas for all instruments are av...	0.400000	0.400000	Positive	0	0
2	2020-01-01 00:56:57+00:00	top trades of the decade: number : going long...	0.450000	0.218750	Positive	0	0
3	2020-01-01 01:24:41+00:00	& hour chas for the group instruments a...	0.400000	0.400000	Positive	0	0
4	2020-01-01 01:26:34+00:00	& hour chas for all instruments are ava...	0.400000	0.400000	Positive	0	0
5	2020-01-01 01:30:11+00:00	made one big mistake day trading apple (aapl) ...	0.100000	0.000000	Neutral	0	1
...	...	...	...	...	...	...	...
33363	2021-09-20 22:10:27+00:00	aapl weekly price pattern coordinates\nannota...	0.000000	0.000000	Neutral	0	1
33364	2021-09-20 22:15:53+00:00	apple inc price at close, --, is .. #apple #aapl	0.000000	0.000000	Neutral	0	1
33365	2021-09-20 23:15:03+00:00	now, if only any of us outside the us could ...	0.395139	-0.010417	Negative	1	0
33366	Datetime	text	0.000000	0.000000	Neutral	0	1
33367	Datetime	text	0.000000	0.000000	Neutral	0	1

33367 rows × 8 columns

In [58]:
```
df3 = df2[0:33365]
df3
```

Out[58]:

	Datetime	Text	Subjectivity	Polarity	Analysis	Negative	Neutral
1	2020-01-01 00:24:13+00:00	& hour chas for all instruments are av...	0.400000	0.400000	Positive	0	0
2	2020-01-01 00:56:57+00:00	top trades of the decade: number : going long...	0.450000	0.218750	Positive	0	0
3	2020-01-01 01:24:41+00:00	& hour chas for the group instruments a...	0.400000	0.400000	Positive	0	0
4	2020-01-01 01:26:34+00:00	& hour chas for all instruments are ava...	0.400000	0.400000	Positive	0	0
5	2020-01-01 01:30:11+00:00	made one big mistake day trading apple (aapl) ...	0.100000	0.000000	Neutral	0	1
...	...	...	...	...	...	...	...
33361	2021-09-20 20:32:57+00:00	is apple's stock..\n$aapl #aapl\n 📎👉 |	0.000000	0.000000	Neutral	0	1
33362	2021-09-20 20:57:29+00:00	#sdc leads today #reddit #wallstreetbets. stoc...	0.000000	0.000000	Neutral	0	1
33363	2021-09-20 22:10:27+00:00	aapl weekly price pattern coordinates\nannota...	0.000000	0.000000	Neutral	0	1
33364	2021-09-20 22:15:53+00:00	apple inc price at close, --, is .. #apple #aapl	0.000000	0.000000	Neutral	0	1
33365	2021-09-20 23:15:03+00:00	now, if only any of us outside the us could ...	0.395139	-0.010417	Negative	1	0

33365 rows × 8 columns

In [59]:
```
df3['Datetime']=pd.to_datetime(df3.Datetime, format='%Y/%m/%d %H:%M:%S.%f')
```

```
<ipython-input-59-5357662dc02d>:1: 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
  df3['Datetime']=pd.to_datetime(df3.Datetime, format='%Y/%m/%d %H:%M:%S.%f')
```

In [60]:
```
df4 = df3.groupby([df3['Datetime'].dt.date]).sum()
```

In [61]:
```
df4
```

Out[61]:

Datetime	Subjectivity	Polarity	Negative	Neutral	Positive
2020-01-01	8.872078	3.595666	2.0	8.0	10.0
2020-01-02	29.965398	7.373483	11.0	31.0	42.0
2020-01-03	16.466594	3.734829	11.0	12.0	21.0
2020-01-04	2.215873	1.043651	1.0	8.0	3.0
2020-01-05	4.615278	0.696528	1.0	9.0	4.0
...	...	...	...	...	...
2021-09-16	13.678165	5.780536	6.0	14.0	19.0
2021-09-17	10.251888	0.065447	8.0	14.0	9.0
2021-09-18	6.362626	2.328093	3.0	5.0	13.0
2021-09-19	7.614745	3.094349	3.0	9.0	12.0
2021-09-20	14.341612	3.641494	13.0	22.0	21.0

627 rows × 5 columns

In [62]:
```
df4.columns
```

Out[62]:
```
Index(['Subjectivity', 'Polarity', 'Negative', 'Neutral', 'Positive
'], dtype='object')
```

In [63]:
```
df5 = df4.drop(['Subjectivity','Polarity'],axis=1)
```

In [64]:
```
df5
```

Out[64]:

Datetime	Negative	Neutral	Positive
2020-01-01	2.0	8.0	10.0
2020-01-02	11.0	31.0	42.0
2020-01-03	11.0	12.0	21.0
2020-01-04	1.0	8.0	3.0
2020-01-05	1.0	9.0	4.0
...	...	...	...
2021-09-16	6.0	14.0	19.0
2021-09-17	8.0	14.0	9.0
2021-09-18	3.0	5.0	13.0
2021-09-19	3.0	9.0	12.0
2021-09-20	13.0	22.0	21.0

627 rows × 3 columns

In [65]:
```
df5.reset_index(inplace=True)
```

In [66]:
```
df5
```

Out[66]:

	Datetime	Negative	Neutral	Positive
0	2020-01-01	2.0	8.0	10.0
1	2020-01-02	11.0	31.0	42.0
2	2020-01-03	11.0	12.0	21.0
3	2020-01-04	1.0	8.0	3.0
4	2020-01-05	1.0	9.0	4.0
...	...	...	...	...
622	2021-09-16	6.0	14.0	19.0
623	2021-09-17	8.0	14.0	9.0
624	2021-09-18	3.0	5.0	13.0
625	2021-09-19	3.0	9.0	12.0
626	2021-09-20	13.0	22.0	21.0

627 rows × 4 columns

In [72]:
```
col_list = list(df5)
col_list
```

Out[72]:
```
['Datetime', 'Negative', 'Neutral', 'Positive']
```

In [73]:
```
col_list.remove('Datetime')
```

In [75]:
```
df5['Total_Tweets'] = df5[col_list].sum(axis=1)
```

In [76]:
```
df5
```

Out[76]:

	Datetime	Negative	Neutral	Positive	Total_Tweets
0	2020-01-01	2.0	8.0	10.0	20.0
1	2020-01-02	11.0	31.0	42.0	84.0
2	2020-01-03	11.0	12.0	21.0	44.0
3	2020-01-04	1.0	8.0	3.0	12.0
4	2020-01-05	1.0	9.0	4.0	14.0
...	...	...	...	...	...
622	2021-09-16	6.0	14.0	19.0	39.0
623	2021-09-17	8.0	14.0	9.0	31.0
624	2021-09-18	3.0	5.0	13.0	21.0
625	2021-09-19	3.0	9.0	12.0	24.0
626	2021-09-20	13.0	22.0	21.0	56.0

627 rows × 5 columns

In [77]:
```
df5.columns
```

Out[77]:
```
Index(['Datetime', 'Negative', 'Neutral', 'Positive', 'Total_Tweets
'], dtype='object')
```

In [78]:
```
df5
```

Out[78]:

	Datetime	Negative	Neutral	Positive	Total_Tweets
0	2020-01-01	2.0	8.0	10.0	20.0
1	2020-01-02	11.0	31.0	42.0	84.0
2	2020-01-03	11.0	12.0	21.0	44.0
3	2020-01-04	1.0	8.0	3.0	12.0
4	2020-01-05	1.0	9.0	4.0	14.0
...	...	...	...	...	...
622	2021-09-16	6.0	14.0	19.0	39.0
623	2021-09-17	8.0	14.0	9.0	31.0
624	2021-09-18	3.0	5.0	13.0	21.0
625	2021-09-19	3.0	9.0	12.0	24.0
626	2021-09-20	13.0	22.0	21.0	56.0

627 rows × 5 columns

In [79]:
```
df5.insert(5,'Stock','Apple')
```

In [80]:
```
df5
```

Out[80]:

	Datetime	Negative	Neutral	Positive	Total_Tweets	Stock
0	2020-01-01	2.0	8.0	10.0	20.0	Apple
1	2020-01-02	11.0	31.0	42.0	84.0	Apple
2	2020-01-03	11.0	12.0	21.0	44.0	Apple
3	2020-01-04	1.0	8.0	3.0	12.0	Apple
4	2020-01-05	1.0	9.0	4.0	14.0	Apple
...	...	...	...	...	...	...
622	2021-09-16	6.0	14.0	19.0	39.0	Apple
623	2021-09-17	8.0	14.0	9.0	31.0	Apple
624	2021-09-18	3.0	5.0	13.0	21.0	Apple
625	2021-09-19	3.0	9.0	12.0	24.0	Apple
626	2021-09-20	13.0	22.0	21.0	56.0	Apple

627 rows × 6 columns

In [81]:
```
df5['Datetime']=pd.to_datetime(df5.Datetime, format='%Y/%m/%d %H:%M:%S.%f')
```

In [82]:
```
df5['Year'] = df5['Datetime'].dt.year
df5['Month'] = df5['Datetime'].dt.month
df5['Day'] = df5['Datetime'].dt.day
```

In [83]:
```
df5
```

Out[83]:

	Datetime	Negative	Neutral	Positive	Total_Tweets	Stock	Year	Month	Day
0	2020-01-01	2.0	8.0	10.0	20.0	Apple	2020	1	1
1	2020-01-02	11.0	31.0	42.0	84.0	Apple	2020	1	2
2	2020-01-03	11.0	12.0	21.0	44.0	Apple	2020	1	3
3	2020-01-04	1.0	8.0	3.0	12.0	Apple	2020	1	4
4	2020-01-05	1.0	9.0	4.0	14.0	Apple	2020	1	5
...	...	...	...	...	...	...	...	...	...
622	2021-09-16	6.0	14.0	19.0	39.0	Apple	2021	9	16
623	2021-09-17	8.0	14.0	9.0	31.0	Apple	2021	9	17
624	2021-09-18	3.0	5.0	13.0	21.0	Apple	2021	9	18
625	2021-09-19	3.0	9.0	12.0	24.0	Apple	2021	9	19
626	2021-09-20	13.0	22.0	21.0	56.0	Apple	2021	9	20

627 rows × 9 columns

In [84]:
```
df5.drop(['Datetime'],axis=1)
```

Out[84]:

	Negative	Neutral	Positive	Total_Tweets	Stock	Year	Month	Day
0	2.0	8.0	10.0	20.0	Apple	2020	1	1
1	11.0	31.0	42.0	84.0	Apple	2020	1	2
2	11.0	12.0	21.0	44.0	Apple	2020	1	3
3	1.0	8.0	3.0	12.0	Apple	2020	1	4
4	1.0	9.0	4.0	14.0	Apple	2020	1	5
...	...	...	...	...	...	...	...	...
622	6.0	14.0	19.0	39.0	Apple	2021	9	16
623	8.0	14.0	9.0	31.0	Apple	2021	9	17
624	3.0	5.0	13.0	21.0	Apple	2021	9	18
625	3.0	9.0	12.0	24.0	Apple	2021	9	19
626	13.0	22.0	21.0	56.0	Apple	2021	9	20

627 rows × 8 columns

In [85]:
```
colum_names = ['Year','Month','Day','Stock','Positive','Negative','Neutral','Total_Tweets']

df5 = df5.reindex(columns=colum_names)
```

In [86]:
```
df5
```

Out[86]:

	Year	Month	Day	Stock	Positive	Negative	Neutral	Total_Tweets
0	2020	1	1	Apple	10.0	2.0	8.0	20.0
1	2020	1	2	Apple	42.0	11.0	31.0	84.0
2	2020	1	3	Apple	21.0	11.0	12.0	44.0
3	2020	1	4	Apple	3.0	1.0	8.0	12.0
4	2020	1	5	Apple	4.0	1.0	9.0	14.0
...	...	...	...	...	...	...	...	...
622	2021	9	16	Apple	19.0	6.0	14.0	39.0
623	2021	9	17	Apple	9.0	8.0	14.0	31.0
624	2021	9	18	Apple	13.0	3.0	5.0	21.0
625	2021	9	19	Apple	12.0	3.0	9.0	24.0
626	2021	9	20	Apple	21.0	13.0	22.0	56.0

627 rows × 8 columns

In [88]:
```
df5.to_csv("apple.csv",index=False)
```

In [ ]:
```

```