

```
# Install JDK
from google.colab import drive
drive.mount('/content/drive')
```

Mounted at /content/drive

```
!apt-get update
```

```
Get:1 https://cloud.r-project.org/bin/linux/ubuntu bionic-cran40/ InRelease [3,626 B]
Ign:2 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86\_64 InRelease
Get:3 https://cloud.r-project.org/bin/linux/ubuntu bionic-cran40/ Packages [53.9 kB]
Get:4 http://ppa.launchpad.net/c2d4u.team/c2d4u4.0+/ubuntu bionic InRelease [15.9 kB]
Ign:5 https://developer.download.nvidia.com/compute/machine-learning/repos/ubuntu1804/x86\_64 InRel
Get:6 http://security.ubuntu.com/ubuntu bionic-security InRelease [88.7 kB]
Get:7 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86\_64 Release [697 B]
Hit:8 https://developer.download.nvidia.com/compute/machine-learning/repos/ubuntu1804/x86\_64 Relea
Get:9 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86\_64 Release.gpg [836
Hit:10 http://archive.ubuntu.com/ubuntu bionic InRelease
Get:12 http://archive.ubuntu.com/ubuntu bionic-updates InRelease [88.7 kB]
Ign:13 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86\_64 Packages
Get:13 https://developer.download.nvidia.com/compute/cuda/repos/ubuntu1804/x86\_64 Packages [770 kB]
Hit:14 http://ppa.launchpad.net/cran/libgit2/ubuntu bionic InRelease
Get:15 http://ppa.launchpad.net/deadsnakes/ppa/ubuntu bionic InRelease [15.9 kB]
Get:16 http://archive.ubuntu.com/ubuntu bionic-backports InRelease [74.6 kB]
Get:17 http://security.ubuntu.com/ubuntu bionic-security/universe amd64 Packages [1,410 kB]
Get:18 http://archive.ubuntu.com/ubuntu bionic-updates/main amd64 Packages [2,550 kB]
Get:19 http://ppa.launchpad.net/graphics-drivers/ppa/ubuntu bionic InRelease [21.3 kB]
Get:20 http://ppa.launchpad.net/c2d4u.team/c2d4u4.0+/ubuntu bionic/main Sources [1,759 kB]
Get:21 http://security.ubuntu.com/ubuntu bionic-security/restricted amd64 Packages [399 kB]
Get:22 http://security.ubuntu.com/ubuntu bionic-security/multiverse amd64 Packages [24.7 kB]
Get:23 http://security.ubuntu.com/ubuntu bionic-security/main amd64 Packages [2,119 kB]
Get:24 http://archive.ubuntu.com/ubuntu bionic-updates/universe amd64 Packages [2,182 kB]
Get:25 http://archive.ubuntu.com/ubuntu bionic-updates/restricted amd64 Packages [429 kB]
Get:26 http://archive.ubuntu.com/ubuntu bionic-updates/multiverse amd64 Packages [31.6 kB]
Get:27 http://ppa.launchpad.net/c2d4u.team/c2d4u4.0+/ubuntu bionic/main amd64 Packages [900 kB]
Get:28 http://ppa.launchpad.net/deadsnakes/ppa/ubuntu bionic/main amd64 Packages [40.8 kB]
Get:29 http://ppa.launchpad.net/graphics-drivers/ppa/ubuntu bionic/main amd64 Packages [47.2 kB]
Fetched 13.0 MB in 8s (1,595 kB/s)
Reading package lists... Done
```

```
!apt-get install openjdk-8-jdk-headless -qq > /dev/null
```

```
# Get Spark installer (check the path on spark.apache.org)
```

```
!wget -q http://apache.mirrors.pair.com/spark/spark-2.4.7/spark-2.4.7-bin-hadoop2.7.tgz
```

```
# Check path
```

```
!ls
```

drive sample_data spark-2.4.7-bin-hadoop2.7.tgz

```
#Untar the Spark installer
```

```
!tar -xvf spark-2.4.7-bin-hadoop2.7.tgz
```

```
spark-2.4.7-bin-hadoop2.7/sbin/stop-master.sh
spark-2.4.7-bin-hadoop2.7/sbin/start-history-server.sh
spark-2.4.7-bin-hadoop2.7/sbin/stop-mesos-shuffle-service.sh
spark-2.4.7-bin-hadoop2.7/sbin/stop-mesos-dispatcher.sh
spark-2.4.7-bin-hadoop2.7/sbin/stop-all.sh
spark-2.4.7-bin-hadoop2.7/sbin/start-mesos-shuffle-service.sh
spark-2.4.7-bin-hadoop2.7/sbin/start-slaves.sh
spark-2.4.7-bin-hadoop2.7/sbin/spark-config.sh
spark-2.4.7-bin-hadoop2.7/sbin/start-shuffle-service.sh
spark-2.4.7-bin-hadoop2.7/sbin/start-slave.sh
spark-2.4.7-bin-hadoop2.7/sbin/spark-daemons.sh
spark-2.4.7-bin-hadoop2.7/sbin/start-all.sh
spark-2.4.7-bin-hadoop2.7/sbin/spark-daemon.sh
spark-2.4.7-bin-hadoop2.7/sbin/slaves.sh
spark-2.4.7-bin-hadoop2.7/sbin/stop-slaves.sh
spark-2.4.7-bin-hadoop2.7/LICENSE
spark-2.4.7-bin-hadoop2.7/pvthon/
```

```

spark-2.4.7-bin-hadoop2.7/python/setup.py
spark-2.4.7-bin-hadoop2.7/python/pylintrc
spark-2.4.7-bin-hadoop2.7/python/README.md
spark-2.4.7-bin-hadoop2.7/python/run-tests.py
spark-2.4.7-bin-hadoop2.7/python/.coveragerc
spark-2.4.7-bin-hadoop2.7/python/run-tests-with-coverage
spark-2.4.7-bin-hadoop2.7/python/.gitignore
spark-2.4.7-bin-hadoop2.7/python/test_support/
spark-2.4.7-bin-hadoop2.7/python/test_support/userlibrary.py
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/_SUCCESS
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2014/
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2014/month=9/
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2014/month=9/day=1/
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2014/month=9/day=1/.part
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2014/month=9/day=1/part-
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2015/
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2015/month=10/
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2015/month=10/day=26/
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2015/month=10/day=26/par
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2015/month=10/day=26/.pa
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2015/month=10/day=25/
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2015/month=10/day=25/par
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2015/month=10/day=25/.pa
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2015/month=10/day=25/.pa
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2015/month=10/day=25/par
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2015/month=9/
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2015/month=9/day=1/
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2015/month=9/day=1/.part
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/year=2015/month=9/day=1/part-
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/_metadata
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/parquet_partitioned/_common_metadata
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/people_array.json
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/ages.csv
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/ages_newlines.csv
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/people_array_utf16le.json
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/people1.json
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/people.json
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/orc_partitioned/
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/orc_partitioned/_SUCCESS
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/orc_partitioned/b=0/
spark-2.4.7-bin-hadoop2.7/python/test_support/sql/orc_partitioned/b=0/c=0/

```

```

# Install findspark - a python library to find Spark
!pip install -q findspark

```

```

# Set environment variables
# Set Java and Spark home based on the location where they are stored
import os
os.environ["JAVA_HOME"] = "/usr/lib/jvm/java-8-openjdk-amd64"
os.environ["SPARK_HOME"] = "/content/spark-2.4.7-bin-hadoop2.7"

```

```

# Create a local Spark session
import findspark
findspark.init()
from pyspark.sql import SparkSession
spark = SparkSession.builder.master("local[*]").getOrCreate()

```

```

# Import the package
import pyspark
from pyspark import SparkContext
from pyspark.sql import SQLContext
import pandas as pd
import numpy as np
import matplotlib as plt
import string
import re
import nltk

```

```

# Create SparkContextz

```

```

sc = SparkContext.getOrCreate()
print("SparkContext : ", sc)

# Define spark
from pyspark.sql import SparkSession
spark = SparkSession.builder.master("local[*]").getOrCreate()

    SparkContext : <SparkContext master=local[*] appName=pyspark-shell>

# Mount with gdrive
from google.colab import drive
drive.mount('/content/gdrive')

    Mounted at /content/gdrive

# Set the directory
import os
os.chdir("/content/gdrive/My Drive/IDS561/IDS561Project/CodeData")
!ls

    Data1.csv  Data.xlsx  MasterData1.xlsx  MasterData2.xlsx

data = pd.read_excel("/content/gdrive/My Drive/IDS561/IDS561Project/CodeData/MasterData2.xlsx")
data['Date'] = pd.to_datetime(data['Date'])
data.head()

#data.count()

```

	id	Text	Country	Date
0	1386791055765469952	#influencers #celebrities #contentcreators #wo...	US	2021-04-20
1	1386789896115810048	One of the biggest reasons NOT to be a Cyberba...	US	2021-04-20
2	1386789197969769984	Now over a year into working remotely, we've g...	US	2021-04-20
3	1386785660846149888	8 U.S. Cities and Towns That Will Pay You To M...	US	2021-04-20
4	1386785645192940032	#Employment is one of the things that you can'...	US	2021-04-20

```
#creating spark dataframe
```

```

from pyspark.sql.types import StructType, StructField, DoubleType, StringType, DateType, TimestampType

schema = StructType([StructField("id", StringType(), True), StructField("Text", StringType(), True), StructField("Country", StringType(), True), StructField("Date", DateType(), True)])
sdf = spark.createDataFrame(data, schema=schema)
sdf.show(10)

```

```

+-----+-----+-----+-----+
|          id|          Text|Country|          Date|
+-----+-----+-----+-----+
|1386791055765469952|#influencers #cel...|US|2021-04-20 00:00:00|
|1386789896115810048|One of the bigges...|US|2021-04-20 00:00:00|
|1386789197969769984|Now over a year i...|US|2021-04-20 00:00:00|
|1386785660846149888|8 U.S. Cities and...|US|2021-04-20 00:00:00|
|1386785645192940032|#Employment is on...|US|2021-04-20 00:00:00|
|1386785400518330112|Hubstaff organize...|US|2021-04-20 00:00:00|
|1386784381344300032|Check this out FR...|US|2021-04-20 00:00:00|
|1386783345078579968|Studies show that...|US|2021-04-20 00:00:00|
|1386783031663500032|We're hiring! We ...|US|2021-04-20 00:00:00|
|1386782931885159936|We're hiring! We ...|US|2021-04-20 00:00:00|
+-----+-----+-----+-----+
only showing top 10 rows

```

```
#Cleaning the tweets
```

```
from pyspark.sql.functions import col, lower, regexp_replace, split
```

```
def clean_text(c):
    c = lower(c)
    c = regexp_replace(c, "#workfromhome", "")
    c = regexp_replace(c, "#work", "")
    c = regexp_replace(c, "#remote", "")
    c = regexp_replace(c, "#wfh", "")
    c = regexp_replace(c, "#working", "")
    c = regexp_replace(c, "#remotework", "")
    c = regexp_replace(c, "^rt ", "")
    c = regexp_replace(c, "(https:\/\/)\S+", "")
    c = regexp_replace(c, "@([A-Za-z0-9_]+)", "") # Remove usernames
    c = regexp_replace(c, "#([A-Za-z0-9_]+)", "") # Remove Hashtags
    c = regexp_replace(c, "[^a-zA-Z0-9\\s]", "")
    c = regexp_replace(c, "[\n]", " ")
# c = regexp_replace(c, "workfromhome ", "")
# c = split(c, "\\s+") #tokenization...
    return c
```

```
#viewing the data
clean_text_df = sdf.select( "id", clean_text(col("Text")).alias("tweet"), "Country", "Date")
clean_text_df.show(20, truncate=False)
```

```
clean_text_df.count()
```

```
+-----+-----+
|id      |tweet
+-----+-----+
|1386791055765469952|check my website for my top recommended sites
|1386789896115810048|one of the biggest reasons not to be a cyberbacker cyberbacker will make you
|1386789197969769984|now over a year into working remotely weve got some new tips and strategies th
|1386785660846149888|8 us cities and towns that will pay you to move there and work remote in 2021
|1386785645192940032|is one of the things that you cant trust who knows when you will be terminate
|1386785400518330112|hubstaff organized a virtual company retreat this year learn about the challen
|1386784381344300032|check this out free get paid 900 in 60 seconds athome
|1386783345078579968|studies show that has increased productivity by 5 and 1 in 5 executives expec
|1386783031663500032|were hiring we are on the hunt for individuals who are interested in bettering
|1386782931885159936|were hiring we are on the hunt for individuals who are interested in bettering
|1386781945510739968|happy and stay extra in and
|1386780874549580032|and has changed the landscape of aside from finding tips via the internet
|1386778176500740096|15 top ways to make money with digital marketing click here to see the post
|1386775879423040000|grow your expertise with a membership from the international society of six si
|1386775848645270016|tips amp tricks for working efficiently from home
|1386775592125829888|enhancing your work from home setup the way we work is changing rapidly man
|1386774587774569984|enjoying this monday with relaxing music thank you
|1386772350092420096|your remote work experience can be a positive and healthy one check out th
|1386772298531800064|learn how to determine which rooms are vacant in your facilities in spaceview
|1386772161294220032|awesome system just got updated and it now comes with 14 day email swipes trai
+-----+-----+
```

```
only showing top 20 rows
```

```
40645
```

```
from pyspark.sql.functions import substring
clean_text_df = clean_text_df.withColumn('Date', substring('Date',1,10))
clean_text_df.show(20)
```

```
+-----+-----+-----+-----+
|id      |tweet|Country|Date|
+-----+-----+-----+-----+
|1386791055765469952|check my we...|US|2021-04-20|
|1386789896115810048|one of the bigges...|US|2021-04-20|
|1386789197969769984|now over a year i...|US|2021-04-20|
|1386785660846149888|8 us cities and t...|US|2021-04-20|
|1386785645192940032|is one of the th...|US|2021-04-20|
|1386785400518330112|hubstaff organize...|US|2021-04-20|
|1386784381344300032|check this out fr...|US|2021-04-20|
|1386783345078579968|studies show that...|US|2021-04-20|
```

1386783031663500032	were hiring we ar...	US	2021-04-20
1386782931885159936	were hiring we ar...	US	2021-04-20
1386781945510739968	happy and stay ...	US	2021-04-20
1386780874549580032	and has change...	US	2021-04-20
1386778176500740096	15 top ways to ma...	US	2021-04-20
1386775879423040000	grow your experti...	US	2021-04-20
1386775848645270016	tips amp tricks f...	US	2021-04-20
1386775592125829888	enhancing your wo...	US	2021-04-20
1386774587774569984	enjoying this mon...	US	2021-04-20
1386772350092420096	your remote work...	US	2021-04-20
1386772298531800064	learn how to dete...	US	2021-04-20
1386772161294220032	awesome system ju...	US	2021-04-20

only showing top 20 rows

#Removing Duplicates

```
clean_text_df1 = clean_text_df.drop_duplicates(subset = ['id','tweet','Date','Country'])
clean_text_df1.count()
```

25476

sample how textblob works

```
testimonial = TextBlob("not great")
testimonial.sentiment
```

Sentiment(polarity=-0.4, subjectivity=0.75)

```
from textblob import TextBlob
```

#Textblob Algorithm

```
def get_sentiment(text):
    testimonial = TextBlob(text)
    return testimonial.sentiment
```

Get sentiment from tweet

```
from pyspark.sql.functions import udf
```

```
sentimentvalue = udf(lambda x: TextBlob(x).sentiment[0])
subjectivity = udf(lambda x: TextBlob(x).sentiment[1])
sendf = clean_text_df1.withColumn("polarity", sentimentvalue("tweet").cast("double"))\
    .withColumn("subjectivity", subjectivity("tweet").cast("double"))
sendf.show(20)
```

id	tweet	Country	Date	polarity	subjectivity
1386668393298789888	more than 50 of e...	US	2021-04-20	0.135	0.
1386528439117649920	with second wave...	US	2021-04-22	0.24583333333333335	0.4875000000000000
1386764264778739968	book now for summ...	US	2021-04-23	0.3	0
1386647617115909888	check out heliax ...	US	2021-04-23	-0.1	0
1386780000000000000	i love you	US	2021-04-23	0.5	0
1387018016660066310	6 days until its ...	US	2021-04-23	0.0	0
1387435800539897860	employees working...	US	2021-04-23	0.20000000000000004	0.4333333333333333
1387431987732942854	time for the morn...	US	2021-04-23	0.0	0
1385871633076785155	at acism we serve...	US	2021-04-23	0.0	0
1385765560496148486	does not mean w...	US	2021-04-23	0.428125	0.643
1385655102170701827	leadership strate...	US	2021-04-23	0.0	0
1385238233085214721	user experience i...	US	2021-04-23	0.19318181818181818	0.4772727272727272
1384950278470918146	work on an empty ...	US	2021-04-23	0.11666666666666665	0
1384824249664380929	working from home...	US	2021-04-23	0.0	0
1384644134934257664	1 in 4 workers ...	US	2021-04-23	0.0	0
1386993128805122049	remote work chall...	US	2021-04-24	-0.1	0
1386772229858463751	virtual celebrati...	US	2021-04-24	0.12878787878787878	0.3666666666666666
1385573696488845314	zoom call essenti...	US	2021-04-24	0.0	0
1385499765937344512	excited to drop a...	US	2021-04-24	-0.05347222222222223	0.43
1385340426354311171	i got a bike desk...	US	2021-04-24	0.375	0.

only showing top 20 rows

```
sendf.dtypes
```

```
[('id', 'string'),
 ('tweet', 'string'),
 ('Country', 'string'),
 ('Date', 'string'),
 ('polarity', 'double'),
 ('subjectivity', 'double')]
```

```
#defining sentiment logic
```

```
def categorysentiment(x):
    if x < 0:
        sentiment = "negative"
    elif x == 0:
        sentiment = "neutral"
    else:
        sentiment = "positive"
    return sentiment
```

```
sentiment_udf = udf(categorysentiment, StringType())
```

```
sendf1 = sendf.withColumn("sentiment", sentiment_udf("polarity"))
```

```
sendf1.show(10)
```

id	tweet	Country	Date	polarity	subjectivity
1386668393298789888	more than 50 of e...	US	2021-04-20	0.135	0.3
1386528439117649920	with second wave...	US	2021-04-22	0.24583333333333335	0.48750000000000000
1386764264778739968	book now for summ...	US	2021-04-23	0.3	0.
1386647617115909888	check out heliax ...	US	2021-04-23	-0.1	0.
1386780000000000000	i love you	US	2021-04-23	0.5	0.
1387018016660066310	6 days until its ...	US	2021-04-23	0.0	0.
1387435800539897860	employees working...	US	2021-04-23	0.20000000000000004	0.4333333333333333
1387431987732942854	time for the morn...	US	2021-04-23	0.0	0.
1385871633076785155	at acism we serve...	US	2021-04-23	0.0	0.
138576560496148486	does not mean w...	US	2021-04-23	0.428125	0.6437

only showing top 10 rows

```
neg10 = sendf1.filter(sendf1.polarity == -1)\
    .select('id','tweet')\
    .take(10)
```

```
neg10
```

```
[Row(id='1385597877905330177', tweet='something tells me commute to office is getting closer no its
Row(id='1388957304989655057', tweet=' i support the hybrid approach but until the flsa is rewritte
Row(id='1388252417871909888', tweet='in this day and age what a horrible thing to say '),
Row(id='1388263028097910016', tweet='he did sl so nasty in hs '),
Row(id='1387537623783247877', tweet='the truth about working from home in 24 shocking charts via
Row(id='1388229694881890048', tweet='this is horrific and i believe her'),
Row(id='1388265523159300096', tweet='i swear people who act like postal workers are the worst peop
Row(id='1385522043399983112', tweet='who says work from home is supposed to be boring get up dress
Row(id='1389143691567276033', tweet=' no network since morning we are already paying for recharg
Row(id='1388918813966553088', tweet='reminder you treated me dreadfully during covid i quit micro
```

```
# Enable Arrow-based columnar data transfers
```

```
spark.conf.set("spark.sql.execution.arrow.pyspark.enabled", "true")
```

```
# Convert the Spark DataFrame back to a Pandas DataFrame using Arrow
```

```
sendf2 = sendf1.select("*").toPandas()
```

```
#Loading results to csv
```

```
#sendf2.to_csv('/content/gdrive/My Drive/IDS561/IDS561Project/Results/Textblobfinal.csv', index = False,

#Creating three dataframes containing positive, negative, neutral tweets from final results dataframe
# Will use these dataframes to generate words and how many times they are used in positive, negative, neu

tweets_df_pos = pd.DataFrame(sendf2['tweet'][sendf2['sentiment']=='positive'])
tweets_df_neg = pd.DataFrame(sendf2['tweet'][sendf2['sentiment']=='negative'])
tweets_df_neu = pd.DataFrame(sendf2['tweet'][sendf2['sentiment']=='neutral'])

#Importing packages for reports

import matplotlib.pyplot as plt
from wordcloud import WordCloud, STOPWORDS
import string
%matplotlib inline

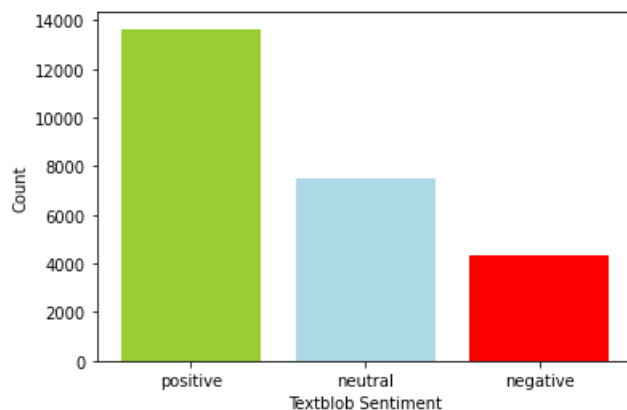
# Distribution of sentiment of the tweets

df_plot = sendf1.groupby("sentiment").count()
df_plot.show(5)

x = df_plot.toPandas()["sentiment"].values.tolist()
y = df_plot.toPandas()["count"].values.tolist()

plt.bar(x,y, color =['yellowgreen', 'lightblue', 'red'])
plt.xlabel("Textblob Sentiment")
plt.ylabel("Count")
plt.show()
```

```
+-----+-----+
|sentiment|count|
+-----+-----+
| positive|13649|
| neutral | 7492|
| negative| 4335|
+-----+-----+
```



```
#To create wordcloud with high frequency words in the tweets
```

```
def wordcloudreport(dfx, title):
    wordcloud = WordCloud(width=1600, height=800, random_state=1, max_words=100, background_color='white',
    wordcloud.generate(str(set(dfx['tweet'])))
    # declare our figure
    plt.figure(figsize=(20,10))
    plt.title(title, fontsize=20,color='Red')
    plt.imshow(wordcloud, interpolation='bilinear')
    plt.axis('off')
    plt.tight_layout(pad=10)
    plt.show()
```

[illegible][illegible]

```
wordcloudreport(tweets_df_neu, "Neutral Sentiment")
```


	tweet	sentiment	vadersenti_pos	vadersenti_neg	vadersenti_neu	vadersenti_compound
0	more than 50 of executives expect to start off...	positive	0.138	0.052	0.810	0.5574
1	with second wave

```
sendf2[['tweet', 'vadersenti_pos', 'vadersenti_neg', 'vadersenti_neu', 'vadersenti_compound', 'vader_senti
```

	tweet	vadersenti_pos	vadersenti_neg	vadersenti_neu	vadersenti_compound	vader_senti
0	more than 50 of executives expect to start off...	0.138	0.052	0.810	0.5574	positive
1	with second wave amp becoming quite the norm...	0.096	0.000	0.904	0.5106	positive
2	book now for summer 2 hours from la lake vie...	0.136	0.000	0.864	0.5106	positive
3	check out heliax theyre looking for a remote ...	0.086	0.000	0.914	0.1779	positive
4	i love you	0.677	0.000	0.323	0.6369	positive
5	6 days until its	0.000	0.000	0.700	0.4500	...

```
#defining sentiment logic

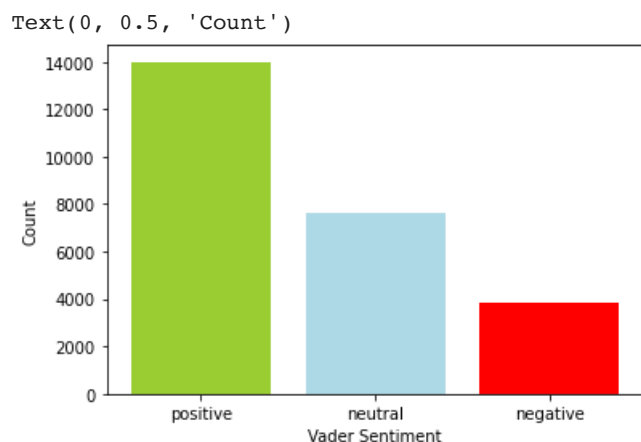
sendf2.loc[sendf2['vadersenti_compound'] > 0.05 , 'vader_senti'] = 'positive'
sendf2.loc[sendf2['vadersenti_compound'] < -0.05 , 'vader_senti'] = 'negative'
sendf2.loc[(sendf2['vadersenti_compound'] <=0.05) & (sendf2['vadersenti_compound'] >=-0.05) , 'vader_senti'] = 'neutral'

# Distribution of sentiment of the tweets

counts = sendf2.vader_senti.value_counts(normalize=False)

x=counts.index
y=counts

plt.bar(x, y, color=['yellowgreen', 'lightblue', 'red'])
plt.xlabel("Vader Sentiment")
plt.ylabel("Count")
```



```
#Count of Positive, negative, neutral tweets
print(sendf2.vader_senti.value_counts())

positive    14001
```

neutral	7637
negative	3838

```
#Creating three dataframes containing positive, negative, neutral tweets from final results dataframe
# Will use these dataframes to generate words and how many times they are used in positive, negative, nei
```

```
vader_df_pos = pd.DataFrame(sendf2['tweet'][sendf2['vader_senti']=='positive'])
vader_df_neg = pd.DataFrame(sendf2['tweet'][sendf2['vader_senti']=='negative'])
vader_df_neu = pd.DataFrame(sendf2['tweet'][sendf2['vader_senti']=='neutral'])
```

```
wordcloudreport(vader_df_pos, "Positive Sentiment")
```



```
wordcloudreport(vader df neg, "Negative Sentiment")
```