PRINCIPLES OF BIG DATA MANAGEMENT

PHASE 2
Twitter Analysis on Football Data Report

Team Member:
Niharika Reddy Voodem(16319416)
V H N V R S Sai Pavan (16296461)
Sai Krishna Reddy Katta (16297003)

Link

Tweets data:
https://drive.google.com/file/d/1PNxu_YzMjVvnzpCAodXf65msFHZklCIi/view?usp=sharing

Github Code Link:

https://github.com/Niharika-0706/Principles-of-Big-Data

OBJECTIVE

- Stream tweets from Twitter (we collected more than 500,000 tweets).
- Write at least 11 analytic queries to explore and understand the collected tweets.
- Create Visualizations on the results.
- Create a web page to view the visualizations.

SOFTWARES PLANNING TO USE

- Apache Spark
    - Spark Streaming to collect tweets.
    - Spark SQL to store and execute queries.
- Matplot, Seaborn and word cloud for visualization
- HTML, CSS, Bootstrap and JavaScript for web application

IMPLEMENTATION

- Initially, we wrote a python program to stream the tweets and save them into JSON format which will have details of the tweets.
- The extracted JSON tweets are persisted into the Apache SparkSQL in the form of Views.
- We used PyCharm IDE to write queries and visualize the outputs using tables and charts.
- The web application is developed using HTML, CSS and Java script.

Libraries used:

- Matplot for visualizations
- Text Blob to get sentimental score
- Flask for web application

Code for Collecting the Tweets:

```
# # # # TWITTER STREAMER # # # #
class TwitterStreamer():
    """
    Class for streaming and processing live tweets.
    """

    def __init__(self):
        pass

    def stream_tweets(self, fetched_tweets_filename, hash_tag_list):
        # This handles Twitter authetification and the connection to Twitter Streaming API
        listener = StdOutListener(fetched_tweets_filename)
        auth = OAuthHandler(CONSUMER_KEY, CONSUMER_SECRET)
        auth.set_access_token(ACCESS_TOKEN, ACCESS_TOKEN_SECRET)
        stream = Stream(auth, listener)

        # This line filter Twitter Streams to capture data by the keywords:
        stream.filter(track=hash_tag_list)


# # # # TWITTER STREAM LISTENER # # # #
class StdOutListener(StreamListener):
    """
    This is a basic listener that just prints received tweets to stdout.
    """

    def __init__(self, fetched_tweets_filename):
        self.fetched_tweets_filename = fetched_tweets_filename
```

```
    def on_data(self, data):
        try:
            print data
            with open(self.fetched_tweets_filename, 'a+') as tf:
                tf.write(data)
            return True
        except BaseException as e:
            print("Error on_data %s" % str(e))
        return True

    def on_error(self, status):
        print(status)


if __name__ == '__main__':
    # Authenticate using config.py and connect to Twitter Streaming API.
    hash_tag_list = [ "la liga", "santander","football","messi", "leo","ronaldo","cristiano",
                      "cr7","goat","iefa","best player","golden shoes","awards","fifa","fcb",
                      "soccer","calcio","nfl","footballseason","footballgame","real madrid",
                      "fc barcelona","psg","hala madrid"]
    fetched_tweets_filename = "football_data.json"
    twitter_streamer = TwitterStreamer()
```

HTML and JavaScript code for UI:

```
<!DOCTYPE html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <title>PB Phase2 Project</title>
    <link rel="stylesheet" href="https://stackpath.bootstrapcdn.com/bootstrap/4.3.1/css/bootstrap.min.css">
    <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.4.1/jquery.min.js"></script>
    <script src="https://maxcdn.bootstrapcdn.com/bootstrap/4.4.1/js/bootstrap.min.js"></script>
    <style>img {max-width:100%}
    .model-dialog {max-width: 90% !important}
    </style>
    <script type="text/javascript">
      $(document).ready(function () {
            $(".modal-dialog").css({"width":"90%","max-width":"90%"});
      });
      </script>
</head>
<body>
<center><h2>Twitter Analysis on Football Data</h2></center>
<table class="table">
  <thead class="thead-dark">
    <tr>
      <th scope="col">Sno</th>
      <th scope="col">Query Question</th>
        <th scope="col">Visualization</th>
    </tr>
  </thead>
  <tbody>
    <tr>
      <th scope="row">1</th>
      <td> Tweet count from different countries</td>
        <td><a href="#" data-toggle="modal" data-target="#myModal"><img src="graphs.PNG"></a>
        </td>

    </tr>
    <tr>
      <th scope="row">2</th>
      <td>Top tweeted text and its retweet count</td>
      <td><a href="#" data-toggle="modal" data-target="#myModal1"><img src="graphs.PNG"></a></td>

    </tr>
   <tr>
      <th scope="row">3</th>
      <td> Tweet count from different countries</td>
      <td><a href="#" data-toggle="modal" data-target="#myModal2"><img src="graphs.PNG"></a></td>

    </tr>
  <tr>
```

```
      <th scope="row">4</th>
      <td> Users with more no of retweets for his tweet</td>
      <td><a  href="#" data-toggle="modal" data-target="#myModal3"><img src="gr..!!P-hs.PNG"></a></td>
    </tr>
  <tr>
      <th scope="row">5</th>
      <td> Tweets based on the different league matches</td>
      <td><a  href="#" data-toggle="modal" data-target="#myModal4"><img src="gr..!!P-hs.PNG"></a></td>
    </tr>
  <tr>
      <th scope="row">6</th>
      <td> User with more no of followers</td>
      <td><a  href="#" data-toggle="modal" data-target="#myModal5"><img src="gr..!!P-hs.PNG"></a></td>
    </tr>
  <tr>
      <th scope="row">7</th>
      <td> Tweets from different Sources</td>
      <td><a  href="#" data-toggle="modal" data-target="#myModal6"><img src="gr..!!P-hs.PNG"></a></td>
    </tr>
  <tr>
      <th scope="row">8</th>
      <td> On 1<hich Day More Tweets are posted</td>
      <td><a  href="#" data-toggle="modal" data-target="#myModal7"><img src="gr..!!P-hs.PNG"></a></td>
    </tr>
  <tr>
      <th scope="row">9</th>
      <td> Tweet count based on language</td>
      <td><a  href="#" data-toggle="modal" data-target="#myModal8"><img src="gr..!!P-hs.PNG"></a></td>
    </tr>
  <tr>
      <th scope="row">10</th>
      <td> Tweet count on hourly bases</td>
      <td><a  href="#" data-toggle="modal" data-target="#myModal9"><img src="gr..!!P-hs.PNG"></a></td>
    </tr>
  <tr>
      <th scope="row">ll</th>
      <td>20 Most Occured \,ords</td>
      <td><a  href="#" data-toggle="modal" data-target="#myModal10"><img src="gr..!!Jlhs.PNG"></a></td>
    </tr>
  </tbody>
</table>
 <div class="modal" id="myModal">
    <div class="modal-dialog">
```

```
    <div  class="modal-content">

      <!-- Modal Header-->
      <div  class="modal-header">
        <h4 class="modal-title">Tweet count from different countries</h4>
        <button  type="button" class="close" data-dismiss="modal">&times;</button>
      </div>

      <!-- Modal body-->
      <div  class="modal-body">
          <center><img src="http://localhost:5000/query.Ll"/></center>
      </div>

      <!-- Modal footer-->
      <div  class="modal-footer">
        <button  type="button" class="btn  btn-danger" data-dismiss="modal">Close</button>
      </div>

    </div>
  </div>
 </div>


<div class="modal" id="myModall">
  <div  class="modal-dialog">
    <div  class="modal-content">

      <!-- Modal Header-->
      <div  class="modal-header">
        <h4 class="modal-title">Top tweeted text and its retweet count</h4>
        <button  type="button" class="close" data-dismiss="modal">&times;</button>
      </div>

      <!-- Modal body-->
      <div  class="modal-body">
          <center><img src="http://localhost:5000/queryfl"/></center>
      </div>

      <!-- Modal footer-->
      <div  class="modal-footer">
        <button  type="button" class="btn  btn-danger" data-dismiss="modal">Close</button>
      </div>

    </div>
  </div>
 </div>
```

```
<div class="modal" id="myModal2">
    <div class="modal-dialog">
      <div  class="modal-content">

        <!-- Modal Header-->
        <div class="modal-header">
          <h4 class="modal-title">Tweet count from different countries</h4>
          <button  type="button"  class="close'' data-dismiss="modal">&times;</button>
        </div>

        <!-- Modal body-->
        <div class="modal-body">
            <center><img src="htt12://localhost:5000/query Q "/></center>
        </div>

        <!-- Modal footer-->
        <div class="modal-footer">
          <button  type="button"  class="btn  btn-danger"  data-dismiss="modal">Close</button>
        </div>

      </div>
    </div>
  </div>

<div class="modal" id="myModal3">
    <div class="modal-dialog">
      <div  class="modal-content">

        <!-- Modal Header-->
        <div class="modal-header">
          <h4 class="modal-title">Users with more no of retweets for his tweet</h4>
          <button type="button" class="close" data-dismiss="modal">&times;</button>
        </div>

        <!-- Modal body-->
        <div class="modal-body">
            <center><img src="httP-://localhost:5000/query..L.!"/></center>
        </div>

        <!-- Modal footer-->
        <div class;''modal-footer''>
          <button type="button" class="btn btn-danger" data-dismiss="modal">Close</button>
        </div>

      </div>
    </div>
  </div>
```

```
<div class="modal" id="myModal4">
    <div class="modal-dialog">
      <div  class=''modal-content'>

        <!-- Modal Header-->
        <div class="modal-header">
          <h4 class="modal-title">Tweets based on the different league matches</h4>
          <button type="button" class="close" data-dismiss="modal">&times;</button>
        </div>

        <!-- Modal body-->
        <div class="modal-body">
            <center><img src="htt12://localhost:5000/queryJ2"/></center>
        </div>

        <!-- Modal footer-->
        <div class="modal-footer">
          <button type="button" class="btn btn-danger" data-dismiss="modal">Close</button>
        </div>

      </div>
    </div>
  </div>


<div class="modal" id="myModalS">
    <div class="modal-dialog">
      <div class="modal-content">

        <!-- Modal Header-->
        <div class="modal-header">
          <h4 class="modal-title">User with more no of follm,ers</h4>
          <button type="button" class="close" data-dismiss="modal">&times;</button>
        </div>

        <!-- Modal body-->
        <div class="modal-body">
            <center><img src="htt12://localhost:5000/queryl..2"/></center>
        </div>

        <!-- Modal footer-->
        <div  class="modal-footer">
          <button type="button" class="btn btn-danger" data-dismiss="modal">Close</button>
        </div>

      </div>
    </div>
  </div>
```

```
<div class="modal" id="myModal6">
    <div class="modal-dialog">
      <div class="modal-content">

        <!-- Modal Header-->
        <div class="modal-header">
          <h4 class="modal-title">Tweets from different Sources</h4>
          <button type='button' class='close" data-dismiss="modal''>&times;</button>
        </div>

        <!-- Modal body-->
        <div class="modal-body">
            <center><img src="httR://localhost:5000/queryfl"/></center>
        </div>

        <!-- Modal footer-->
        <div class=''modal-footer''>
          <button type="button" class="btn btn-danger" data-dismiss="modal">Close</button>
        </div>

      </div>
    </div>
  </div>

<div class="modal" id="myModal7">
    <div class="modal-dialog">
      <div class="modal-content">

        <!-- Modal Header-->
        <div class="modal-header">
          <h4 class="modal-title">On which Day More Tweets are posted</h4>
          <button type="button" class="close" data-dismiss="modal">&times;</button>
        </div>

        <!-- Modal body-->
        <div class="modal-body">
            <center><img src="httR://localhost:5000/query_LJ!"/></center>
        </div>

        <!-- Modal footer-->
        <div class="modal-footer">
          <button type="button" class="btn btn-danger" data-dismiss="modal">Close</button>
        </div>

      </div>
    </div>
  </div>
```

```
<div class="modal" id="myModal8">
    <div  class="modal-dialog">
      <div class="modal-content">

        <!-- Modal Header-->
        <div class="modal-header''>
          <h4 class="modal-title">Tweet count based on language</h4>
          < button  type="button"  class="close "  data-dismiss="modal ">&times; </button>
        </div>

        <!-- Modal body-->
        <div class="modal-body">
            <center><img src="htt12://localhost:5000/queryfl"/></center>
        </div>

        <!-- Modal footer-->
        <div class="modal-footer''>
          <button type="button" class="btn btn-danger" data-dismiss="modal''>Close</button>
        </div>

      </div>
    </div>
  </div>

<div class="modal" id="myModal9">
    <div class::::"modal-dialog">
      <div class="modal-content">

        <!-- Modal Header-->
        <div class="modal-header''>
          <h4 class="modal-title">Tweet count on hourly bases</h4>
          < button  type="button"  class="close "  data-dismiss="modal ">&times;</button>
        </div>

        <!-- Modal body-->
        <div class="modal-body">
            <center><img src="htt12://localhost:5000/queryi.1.!l"/></center>
        </div>

        <!-- Modal footer-->
        <div class="modal-footer''>
          <button type="button" class="btn btn-danger" data-dismiss="modal''>Close</button>
        </div>

      </div>
    </div>
  </div>

<div  class;::"modal"  id="myModal10">
    <div class="modal-dialog">
      <div class="modal-content">

        <!-- Modal Header-->
        <div class="modal-header''>
          <h4 class="modal-title">20 Most Occured Words</h4>
          <button type="button" class="close" data-dismiss="modal">&times;</button>
        </div>

        <!-- Modal body-->
        <div class="modal-body">
            <center><img src="h.tip://localhost:5000/qu_y_L.ll"/></center>
        </div>

        <!-- Modal footer-->
        <div class="modal-footer">
          <button type="button" class="btn btn-danger" data-dismiss="modal''>Close</button>
        </div>

      </div>
    </div>
  </div>
</body>
</html>
```

Code For 10 Queries:

```
from flask import Flask, send_file
import seaborn as sns
from pyspark.sql import SparkSession
import matplotlib.pyplot as plt
import pandas
from io import BytesIO
from wordcloud import WordCloud

app = Flask(__name__)


@app.route("/query/<id>", methods=['GET'])
def hello(id):
    if id == "1":
        query1 = spark.sql(
            "select place.country,count(*) As Count from tweet_table where place.country is not null GROUP BY "
            "place.country ORDER BY count "
            "DESC limit 10")
        pd_query1 = query1.toPandas()
        pd_query1.plot.pie(y='Count', labels=pd_query1.country.values.tolist(), figsize=(10, 5), autopct='%.2f')
        img = BytesIO()
        plt.savefig(img)
        img.seek(0)
        return send_file(img, mimetype='image/png')
    if id == "2":
        query2 = spark.sql(
            "select user.name,retweeted_status.text as Retweet_Text,retweeted_status.retweet_count as Retweet_Count "
            "from tweet_table where retweeted_status.retweet_count is not null order by "
            "retweeted_status.retweet_count desc limit 10")
        pd_query2 = query2.toPandas()
        pd_query2.plot()
        img = BytesIO()
```

```
        plt.savefig(img)
        img.seek(0)
        return send_file(img, mimetype='image/png')
    if id == "3":
        query3 = spark.sql("select user.location,count(text) as tweet_count from tweet_table where "
                           "place.country='United States'and "
                           "user.location is not null Group By user.location ORDER BY tweet_count DESC LIMIT 15")
        pd_query3 = query3.toPandas()
        pd_query3.plot.area(x="location", y="tweet_count", figsize=(11, 5))
        img = BytesIO()
        plt.savefig(img)
        img.seek(0)
        return send_file(img, mimetype='image/png')
    if id == "4":
        query4 = spark.sql("select user.screen_name,text,retweeted_status.retweet_count from tweet_table order by "
                           "retweeted_status.retweet_count DESC limit 10")
        pd_query4 = query4.toPandas()
        pd_query4.plot.scatter(y="retweet_count", x='screen_name', color='DarkBlue', figsize=(14,5))
        # plt.title("Users with more no of retweets for his tweet")
        img = BytesIO()
        plt.savefig(img)
        img.seek(0)
        return send_file(img, mimetype='image/png')
    if id == "5":
        query5 = spark.sql(
            "select count(*) as count,q.text from (select case when text like '%fcb%' then 'fc barcelona' when "
            " text like '%real madrid%' then 'real madrid' when text like '%fifa%' then 'fifa' when text"
            " like '%la liga%' then 'la liga' when text like '%nfl%' then 'nfl' when text like '%arsenal%' "
            " then 'arsenal' when text like '%chelsea%' then 'chelsea fc' when text like '%manchester%' "
            "then 'manchester' when text like '%psg%' then 'psg' when text like '%premier league%' then 'premier "
            "league' "
            " end as text from "
```

```
            " tweet_table)q group by q.text")
        pd_query5 = query5.toPandas()
        sns.catplot(x="text", y="count", kind="point", data=pd_query5.dropna(),aspect=3)
        # plt.title("Tweets based on the different league matches")
        img = BytesIO()
        plt.savefig(img)
        img.seek(0)
        return send_file(img, mimetype='image/png')
    if id == "6":
        query6 = spark.sql(
            "select user.screen_name, max(user.followers_count)as followers_count from tweet_table where text "
            " like '%football%' group by user.screen_name, user.lang order by followers_count desc limit 10")
        pd_query6 = query6.toPandas()
        sns.catplot(x="screen_name", y="followers_count", kind="violin", data=pd_query6, height=5, aspect=3)
        # plt.title("User with more no of followers")
        img = BytesIO()
        plt.savefig(img)
        img.seek(0)
        return send_file(img, mimetype='image/png')
    if id == "7":
        query7 = spark.sql(
            "select count(*) as NumberOfTweets, 'Android' as Source from tweet_table where source like '%Twitter for "
            "Android%' UNION select count() as NumberOfTweets, 'IPhone' as Source from tweet_table where source like "
            "'%Twitter for iPhone%' UNION select count() as NumberOfTweets, 'IPad' as Source from tweet_table where "
            "source like '%Twitter for iPad%' UNION select count() as NumberOfTweets, 'Web' as Source from "
            "tweet_table where source like '%Twitter Web App%'")
        pd_query7 = query7.toPandas()
        pd_query7.plot.line(y="NumberOfTweets", x="Source")
        img = BytesIO()
        plt.savefig(img)
        img.seek(0)
        return send_file(img, mimetype='image/png')
```

```
if id == "8":
    day_data = spark.sql("SELECT substring(user.created_at,1,3) as day from tweet_table where text is not null")
    day_data.createOrReplaceTempView("day_data")
    days_final = spark.sql(
        """ SELECT Case
          when day LIKE '%Mon%' then 'WEEKDAY'
          when day LIKE '%Tue%' then 'WEEKDAY'
          when day LIKE '%Wed%' then 'WEEKDAY'
          when day LIKE '%Thu%' then 'WEEKDAY'
          when day LIKE '%Fri%' then 'WEEKDAY'
          when day LIKE '%Sat%' then 'WEEKEND'
          when day LIKE '%Sun%' then 'WEEKEND'
           else
           null
           end as day1 from day_data where day is not null""")
    days_final.createOrReplaceTempView("days_final")
    query8 = spark.sql("SELECT day1 as Day,Count(*) as Day_Count from days_final where day1 is not null group by "
                       "day1 order by count(*) desc")
    pd_query8 = query8.toPandas()
    pd_query8.plot.pie(y="Day_Count", labels=pd_query8.Day.tolist(), autopct='%.2f')
    img = BytesIO()
    plt.savefig(img)
    img.seek(0)
    return send_file(img, mimetype='image/png')
if id == "9":
    query9 = spark.sql("select count(*) as count,lang from tweet_table where text like '%la liga%' group by lang "
                       "order by count desc")
    pd_query9 = query9.toPandas()
    sns.catplot(x="lang", y="count", data=pd_query9)
    # plt.title("Tweet count based on language")
    img = BytesIO()
    plt.savefig(img)
```

```
    img.seek(0)
    return send_file(img, mimetype='image/png')
if id == "10":
    timehour = spark.sql("SELECT SUBSTRING(created_at,12,2) as hour from tweet_table where text is not null")

    timehour.createOrReplaceTempView("timehour")
    timeAnalysis = spark.sql(
        "SELECT Case when hour>=0 and hour <4 then 'midnight' when hour>=4 and hour <7 then 'earlymorning' "
        " when hour>=7 and hour <12 then 'Day-time' when hour>=12 and hour <15 then 'afternoon' when hour>=15 and "
        " hour <18 then 'evening' when hour>=18 and hour <=23 then 'Night-time' end as time from timehour")

    timeAnalysis.createOrReplaceTempView("timeAnalysis")
    query10 = spark.sql("SELECT time as hour,Count(*) as tweets_count from timeAnalysis where time is not null "
                        "group by time order by count(*) desc")
    pd_query10 = query10.toPandas()
    sns.catplot(x="hour", y="tweets_count", kind="violin", style="smoker", split=True, data=pd_query10)
    # plt.title("Tweet count on hourly bases")
    img = BytesIO()
    plt.savefig(img)
    img.seek(0)
    return send_file(img,mimetype='image/png')
if id == "11":
    query11 = spark.sql(
        "select count(*) as count,text from tweet_table where text like '%football%' group by text order by count "
        "desc limit 20")
    pdquery11 = query11.toPandas()
    train_text = " ".join(pdquery11.text)
    wordcloud = WordCloud().generate(train_text)
    plt.figure()
    plt.subplots(figsize=(7,7))
    wordcloud = WordCloud(
        background_color="Black",
        max_words=len(train_text),
        max_font_size=30,
        relative_scaling=.5).generate(train_text)
    plt.imshow(wordcloud)
    plt.axis("off")
    img = BytesIO()
    plt.savefig(img)
    img.seek(0)
    return send_file(img, mimetype='image/png')


if __name__ == "__main__":
    spark = SparkSession.builder.appName("Phase 2 querying and plotting").getOrCreate()
    sc = spark.sparkContext
    df = spark.read.json(r"C:\Users\jagad\Downloads\pavan\football_data.json")
    df.createOrReplaceTempView("tweet_table")
    app.run(debug=True)
```

Query Execution and Results:

Home Page

Twitter Analysis on Football Data

Sno	Query Question	Visualization
1	Tweet count from different countries	
2	Top tweeted text and its retweet count	
3	Tweet count from different countries	
4	Users with more no of retweets for his tweet	
5	Tweets based on the different league matches	
6	User with more no of followers	
7	Tweets from different Sources	
8	On which Day More Tweets are posted	
9	Tweet count based on language	
10	Tweet count on hourly bases	
11	Generating a word cloud on top 20 text that includes name Football	

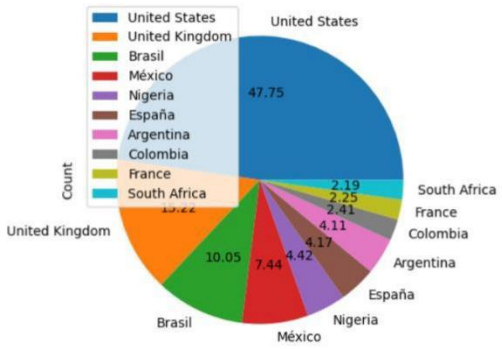
We have executed and visualized totally 10 queries.

1). Tweet count from different countries using Pie Chart.

Query result in table form :

	country	Count
0	United States	3056
1	United Kingdom	974
2	Brasil	643
3	México	476
4	Nigeria	283
5	España	267
6	Argentina	263
7	Colombia	154
8	France	144
9	South Africa	140

Tweet count from different countries

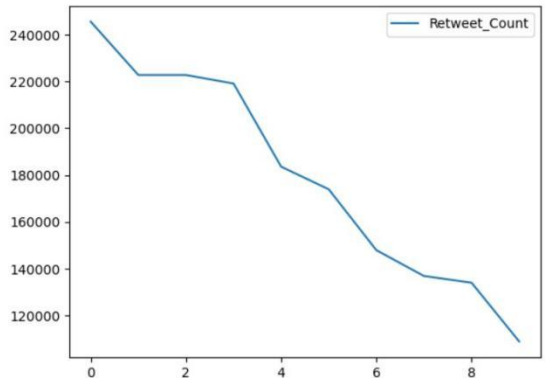


2) Query to display the Top tweeted text and its retweet count using “line plot”.

Result:

	name	Retweet_Text	Retweet_Count
0	ash	Our story. Our awards. Our songs. Our albums. ...	245540
1	Nat Nat🖤	So a month ago I dropped a ring & a clip d...	222756
2	Summer	So a month ago I dropped a ring & a clip d...	222754
3	Lucas Ferkol	If golf and soccer switched announcers... http...	219128
4	💜REBECA becky💜	[#오늘의방탄] 아미와 방탄.. 방탄과 아미... 우리 존재 자체가 사랑..💜 20...	183615
5	May 22nd	This son bought his dad a truck after his dad'...	173951
6	Adelaide	Brit Awards. Thank you to everyone who voted. ...	147956
7	Andrea loves Z	This kid is so cool. May Allah bless him. Dinn...	136926
8	Ruben	Si Mbappé ficha por el Real Madrid este verano...	134065
9	💜BTSpopie⁷💜	thank you 💞 https://t.co/ESmo67cUek	108975

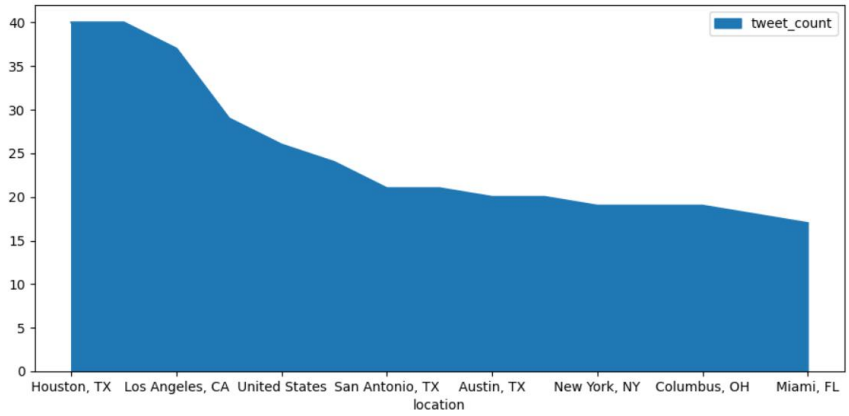
Top tweeted text and its retweet count



3). Tweet count from different countries using “Area Chart”
Result:

	location	tweet_count
0	Houston, TX	40
1	Chicago, IL	40
2	Los Angeles, CA	37
3	Atlanta, GA	29
4	United States	26
5	Dallas, TX	24
6	San Antonio, TX	21
7	Philadelphia, PA	21
8	Florida, USA	20
9	Austin, TX	20
10	New York, NY	19
11	Columbus, OH	19
12	Nashville, TN	19
13	California, USA	18
14	Miami, FL	17

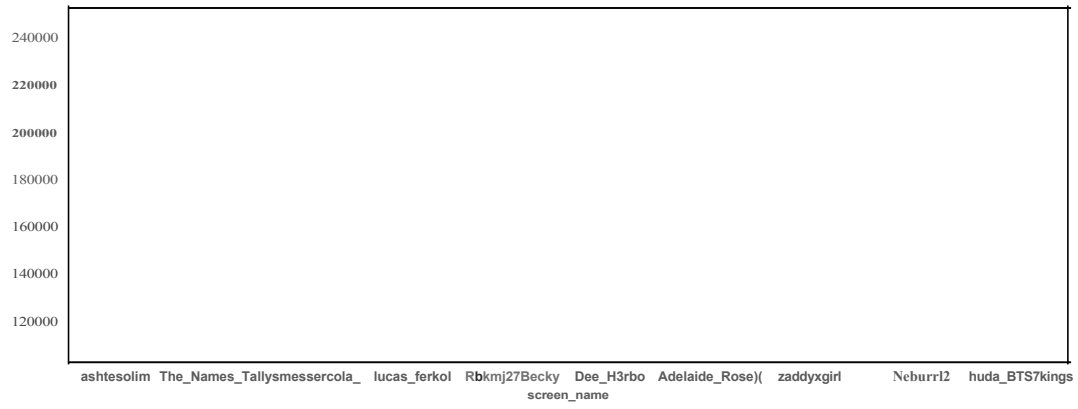
Tweet count from different countries



4). Users with more no of retweets for his tweet using “Scatter plot”
Result:

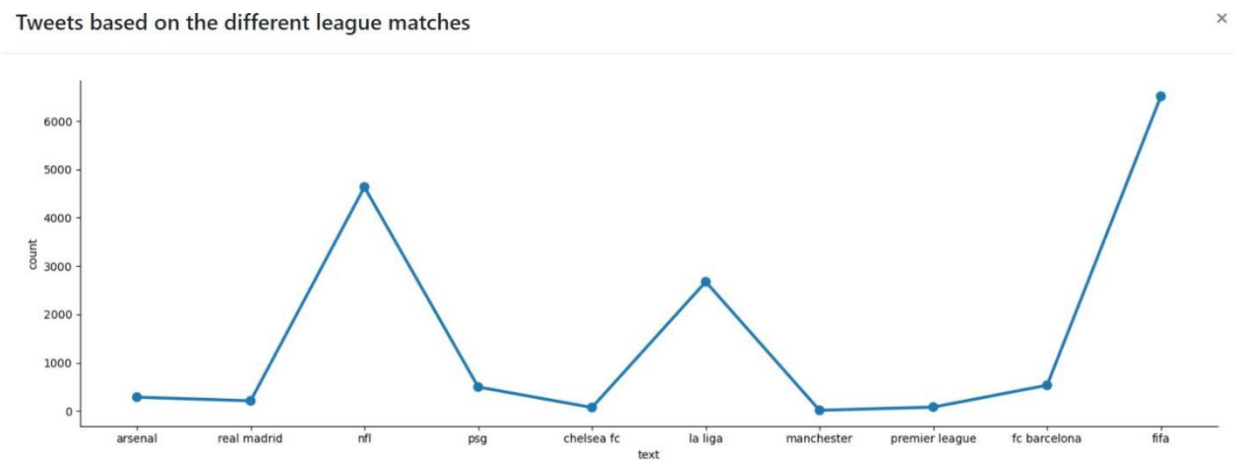
	screen_name	text	retweet_count
0	ashtesolim	RT @justinbieber: Our story. Our awards. Our s...	245540
1	The_Names_Tally	RT @gi11aawilsonn: So a month ago I dropped a **r...**	222756
2	srnessercola_	RT @ginaawilsonn: So a month ago I dropped a r...	222754
3	lucas fer'kol	RT@johnbcrist: If golf and soccer switched an...	**219128**
4	**Rbkmj27Becky**	RT@bts_bighit: [#.2. 9-I]0 □12-r .. t!-ilr 0r□l..2–1...	183615
5	Dee_ H3rbo	RT@Charles43631438: This son bought his dad a...	173951
6	Adelaide_ RoseX	RT@Harry_Styles: Brit Awards. Thank you to ev...	147956
7	zaddyxgirl	RT@iamsrk: Tlhis kid is so cool. May Allah ble...	136926
8	Neburr12	RT@DjMaRiiQ_90: Si Mbappe ficha por el Real M...	134065
9	iluda_BTS7kings	RT@BTS_twt: thank you l'ittps:J/t.co/ESmo67cUek	108975

Users with more no of retweets for his tweet



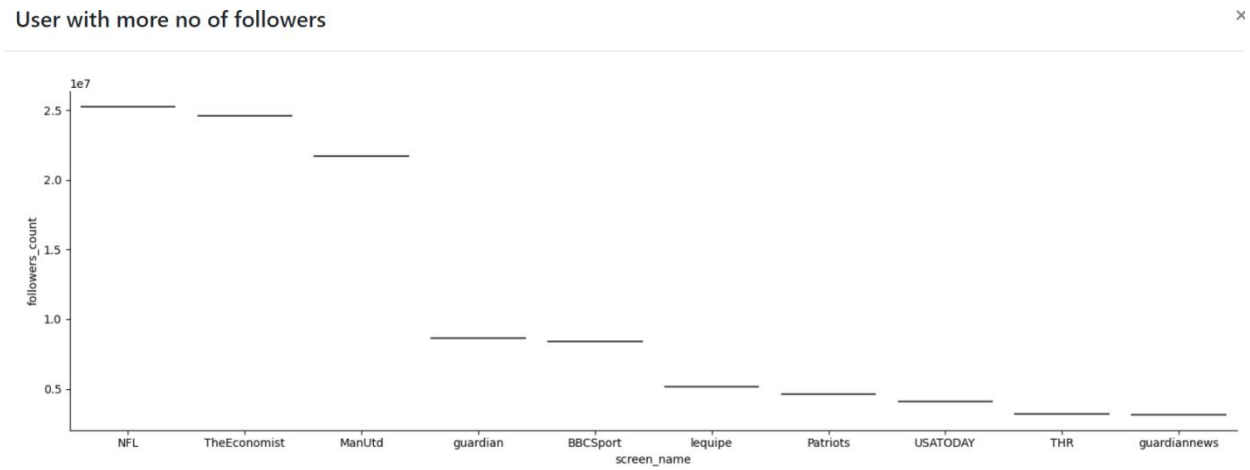
5). Tweets based on the different league matches using “Seaborn Point Code”
Result:

	count	text
0	533598	None
1	289	arsenal
2	213	real madrid
3	4636	nfl
4	499	psg
5	75	chelsea fc
6	2676	la liga
7	17	manchester
8	83	premier league
9	536	fc barcelona
10	6513	fifa



6). User with more no of followers using “Seaborn violin Graph”
Result:

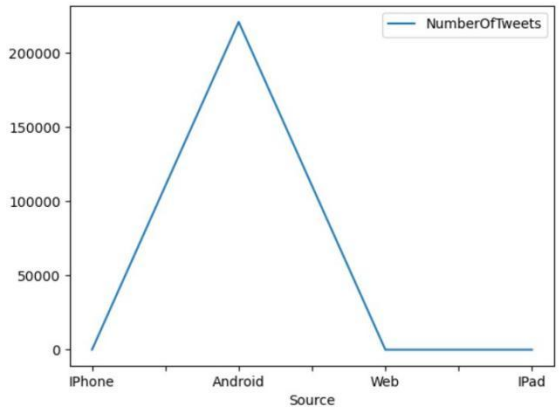
	screen_name	followers_count
0	NFL	25264898
1	TheEconomist	24595917
2	ManUtd	21682621
3	guardian	8639812
4	BBCSport	8375940
5	lequipe	5151027
6	Patriots	4603729
7	USATODAY	4061642
8	THR	3223404
9	guardiannews	3138150
10	BBCMOTD	2917306
11	RealPaigeWWE	2556725
12	SkySport	2459605
13	chrislhayes	2103570
14	notthefakeSVP	2037971



7). Tweets from different Sources using (Line Graph)
Result:

	NumberOfTweets	Source
0	0	IPhone
1	220842	Android
2	0	Web
3	0	IPad

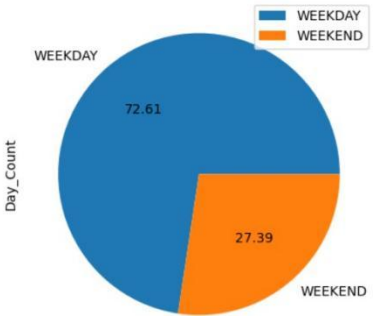
Tweets from different Sources



8). On which Day More Tweets are posted using “Pie Graph”
Result:

	Day	Day_Count
0	WEEKDAY	398728
1	WEEKEND	150403

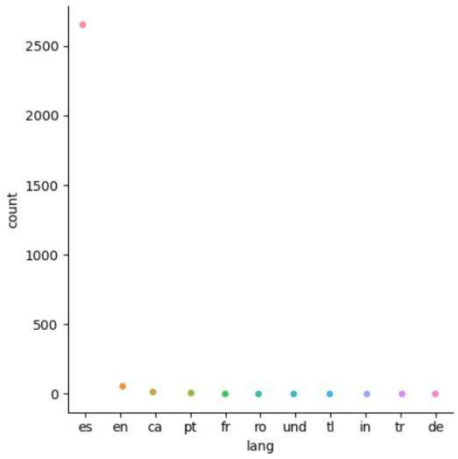
On which Day More Tweets are posted



9). Tweet count based on language using “Seaborn Category Plot)
Result:

	count	lang
0	2653	es
1	58	en
2	19	ca
3	13	pt
4	6	fr
5	2	ro
6	1	und
7	1	tl
8	1	in
9	1	tr
10	1	de

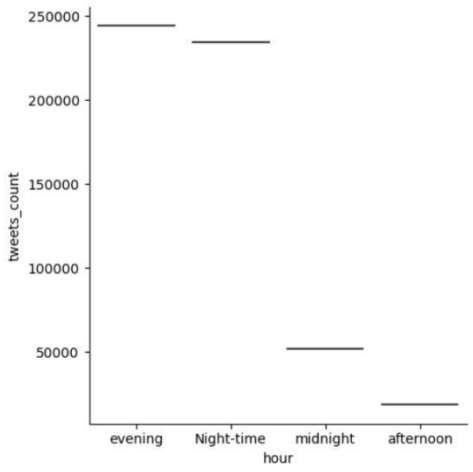
Tweet count based on language



10). Tweet count on hourly bases using “Seaborn Category plot”
Result:

	hour	tweets_count
0	evening	244177
1	Night-time	234567
2	midnight	51924
3	afternoon	18463

Tweet count on hourly bases



11). 20 Most Occurred Words using “Seaborn Category plot”
Result:

	co1111rl:	text
0	6551	RT @llJtlerfyTC: My wire threw my new foo1ball o...
11	337	Rli @Alhle teT veelils: Bra:don Miller had one of ...
2	790	RT @orfootball: Just anolher weekeml missingL
3	522	RT @90sfootballt: Ronaldo in his prime was simp...
4	508	RT@b:rfoolball: Foo1ball icons' football icons...
5	472	RT @ManU'ld: (3 •1miss il because that's my job...
6	463	RT @StigEzeh: Even if it's intemational foO'lb...
7	453	RT @FoolyHumo11 r:lih rov,bacik to when Die,go Forla ...
8	423	RT @FIFAWC2018: I eiould \!a ch Neyrnar playooL
9	358	RT @FulbolBrble: If you started your foofuall ...
10	338	RT@SEConCBS: Saturdays are better with coUeg...
111	337	RT@RapSheet: Akey parl oflhis: •In a contra ...
12	301	RT @Arsenal: Saturdays withou1 foo1b all C:nn \n
13	239	RT @brl'ootball: 13 year s ago lod ay, Cristia no ...
14	237	Rli @uitdlBr elt: Thread of the best (football) tw...
15	235	RT@gri_lay: In llhe 11ovell Political Science wo...
16	234!	RT@mrvi11cecarter15: Thank you lo the@Buocane...
17	2•66	RTbrfoo1ba'll: It's been 13 years si11ce Ro11aL..

20 Most Occured Words