Submitted by:

Tiancheng Xie

Zeng Yue

Josh Oguntimehin

**Principles of Big Data Management**

**Phrase-2**

Spring 2019

**Analyzing and Visualizing Twitter data on Mobiles**

**Objective:**

**Applications/Software’s Used:** Scala, Apache Spark SQL, Vegas, Twitter API, Python.

**Collecting tweets from Twitter:**

* Firstly, we have created a developer account in Twitter using below link. https://apps.twitter.com/
* We have written python program that is used to fetch tweets in JSON format. (TweetsExtract.py)
* The tweet data is collected on the concept based on to analyze and visualize the data regarding various mobile phones.

Sample tweets are collected for the key words by using python program

**#iphone'**,**'#Samsung'**,**'#Moto'**,**'#Redmi'**,**'#Xiaomi'**,**'#Nokia'**,**'#lenovo'**,**'#oppo'**,**'#OnePlus'**,**'#BlackBerry'**,**'#HTC**

1. Using spark to stream the tweet,the data can be -ingested – from different sources like twitter and can perform high level complex algorithms like queries and the processed data can be pushed out to the file systems.To stream the twitter data twitter.utils contains all the built in functionality -to- stream data from twitter.

* **SqlContext which is a new- session, with separated SQL configurations, registered functions temporary tables (data frames to store the all relational functionality in Spark SQL).**

The main theme of this project is to do big data analytics on the Mobile phones. Based on the twitter tweets, we predicted few interesting query analysis and visualization on mobile phone twitter data. First, we collected the tweets regarding the mobile phone data from the twitter API. By using the collected data, we build 10 interesting queries which results a data analysis on the twitter data. Visualization is realized by using Vegas which is an awesome resource for matplotlib.

**Architecture diagram:**

**A close up of a map

Description generated with very high confidence**

**Sample JSON Object Structure:**

**The tweets are saved and stored with the object format as JSON.**

{

"created\_at":"Fri Nov 10 16:11:34 +0000 2017",

"id":929018704041349120,

"id\_str":"929018704041349120",

"text":"You know #SteveJobs hustled his way up to @Apple (and they still got the best hustle! You got that #iphoneX don\u2019t y\u2026 https:\/\/t.co\/GqUSbaqhFF",

"display\_text\_range":[

0,

140

],

"source":"\u003ca href=\"https:\/\/mobile.twitter.com\" rel=\"nofollow\"\u003eTwitter Lite\u003c\/a\u003e",

"truncated":true,

"in\_reply\_to\_status\_id":null,

"in\_reply\_to\_status\_id\_str":null,

"in\_reply\_to\_user\_id":null,

"in\_reply\_to\_user\_id\_str":null,

"in\_reply\_to\_screen\_name":null,

"user":{

"id":923408384358940674,

"id\_str":"923408384358940674",

"name":"Retired Ratchet",

"screen\_name":"retiredratchet",

"location":"Las Vegas, NV",

"url":null,

"description":"#Blitter for the culture:\nRatchet is a slang term in hip hop that in the strictest sense refers to an uncouth female & is a Louisianan regiolect of \"wretched\"",

"translator\_type":"none",

"protected":false,

"verified":false,

"followers\_count":156,

"friends\_count":224,

listed\_count":0,

"favourites\_count":993,

"statuses\_count":618,

"created\_at":"Thu Oct 26 04:38:09 +0000 2017",

"utc\_offset":null,

"time\_zone":null,

"geo\_enabled":false,

"lang":"en",

"contributors\_enabled":false,

"is\_translator":false,

"profile\_background\_color":"F5F8FA",

"profile\_background\_image\_url":"",

"profile\_background\_image\_url\_https":"",

"profile\_background\_tile":false,

"profile\_link\_color":"1DA1F2",

"profile\_sidebar\_border\_color":"C0DEED",

"profile\_sidebar\_fill\_color":"DDEEF6",

"profile\_text\_color":"333333",

"profile\_use\_background\_image":true,

"profile\_image\_url":"http:\/\/pbs.twimg.com\/profile\_images\/923427538038173696\/tBE9KGIV\_normal.jpg",

"profile\_image\_url\_https":"https:\/\/pbs.twimg.com\/profile\_images\/923427538038173696\/tBE9KGIV\_normal.jpg",

"profile\_banner\_url":"https:\/\/pbs.twimg.com\/profile\_banners\/923408384358940674\/1510329291",

"default\_profile":true,

"default\_profile\_image":false,

"following":null,

"follow\_request\_sent":null,

"notifications":null

},

"geo":null,

"coordinates":null,

"place":null,

"contributors":null,

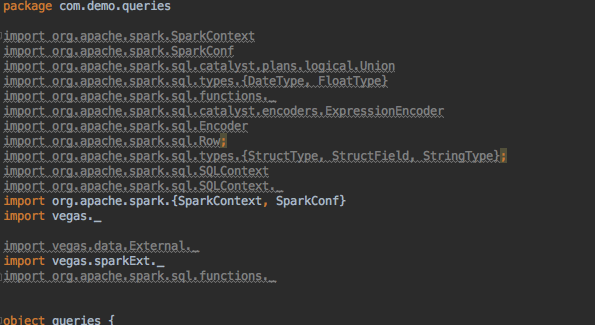
"quoted\_status\_id":927672133739835392,..

}

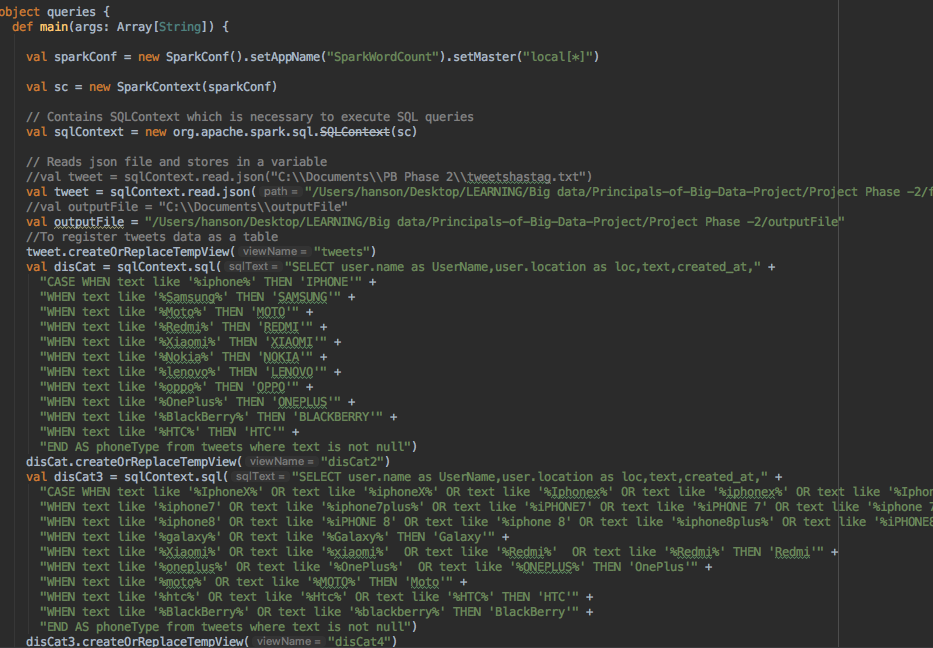
**2)Queries and outputs:**

**1.Illustraing the library we are using in this lab for reference**

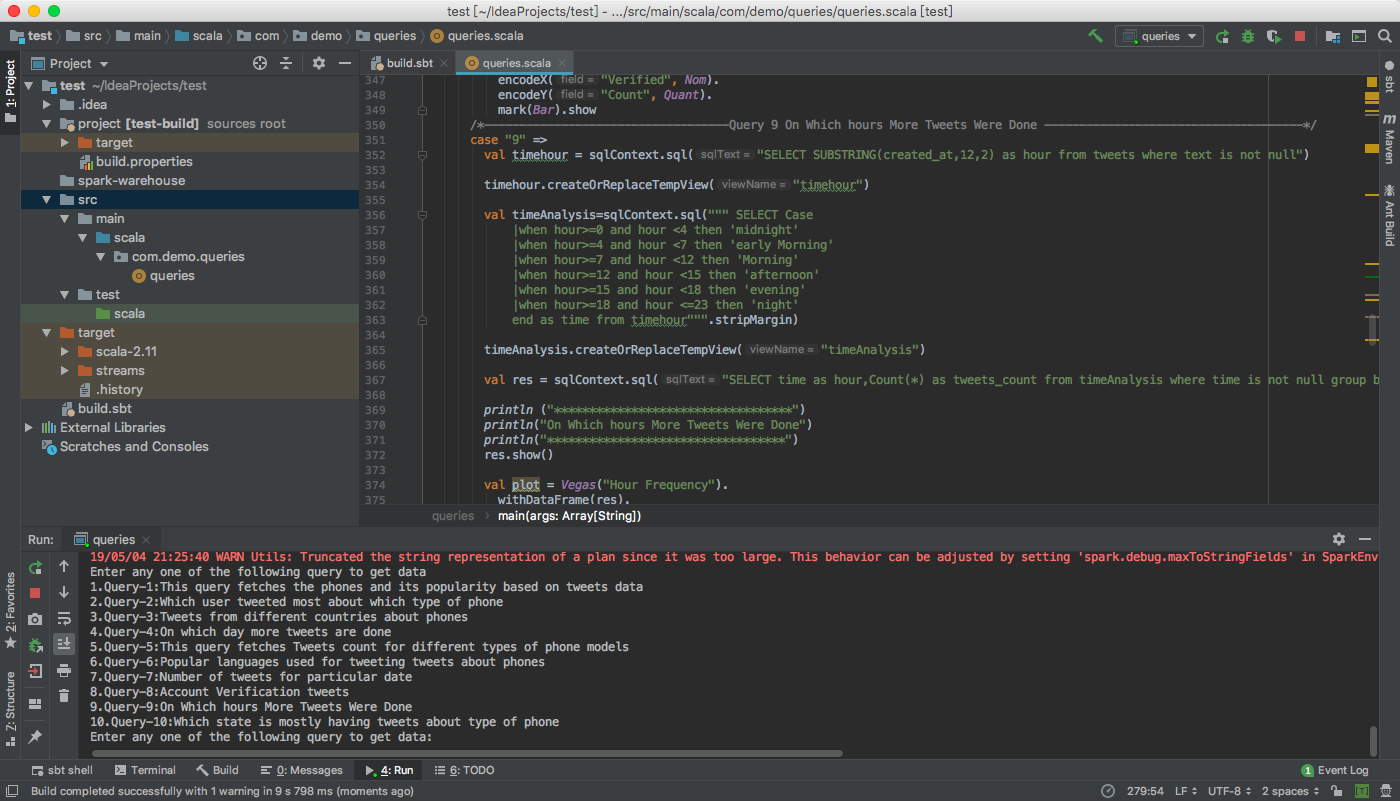
**Code:**

****

**2.The Menu of the program, for reading the source file, selecting the query we want to process.**

****

**Output:**

****

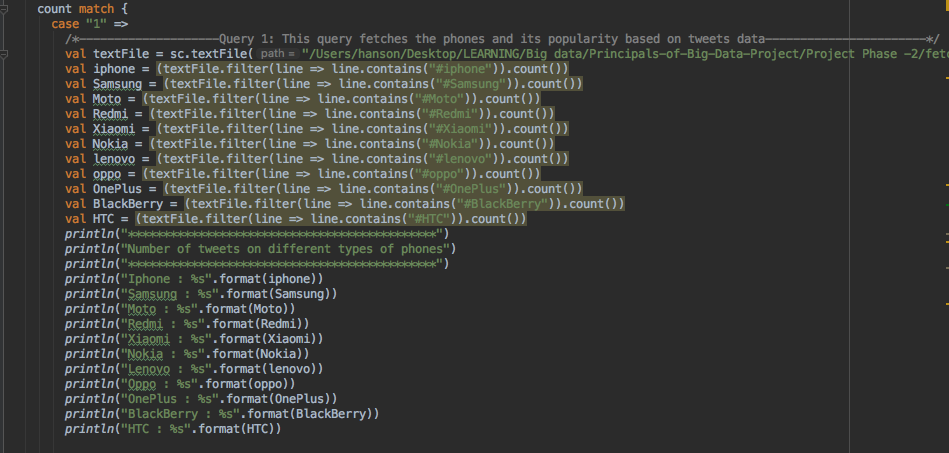
**Query 1: Query for fetching the tweets corresponding to phones and its popularity count on the tweets depending up the on tweets data collected.**

**Answer:** This query is built to analyze the tweets – what the count of each model which are tweeted by the users in the collected tweets which directly reflects the popularity of each phone model.

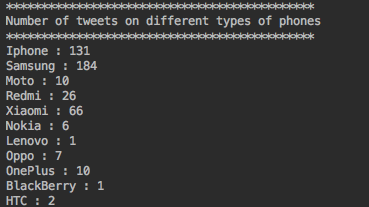
As we collected the data related to mobile phones **#iphone'**,**'#Samsung'**,**'#Moto'**,**'#Redmi'**,**'#Xiaomi'**,**'#Nokia'**,**'#lenovo'**,**'#oppo'**,**'#OnePlus'**,**'#BlackBerry'**,**'#HTC**

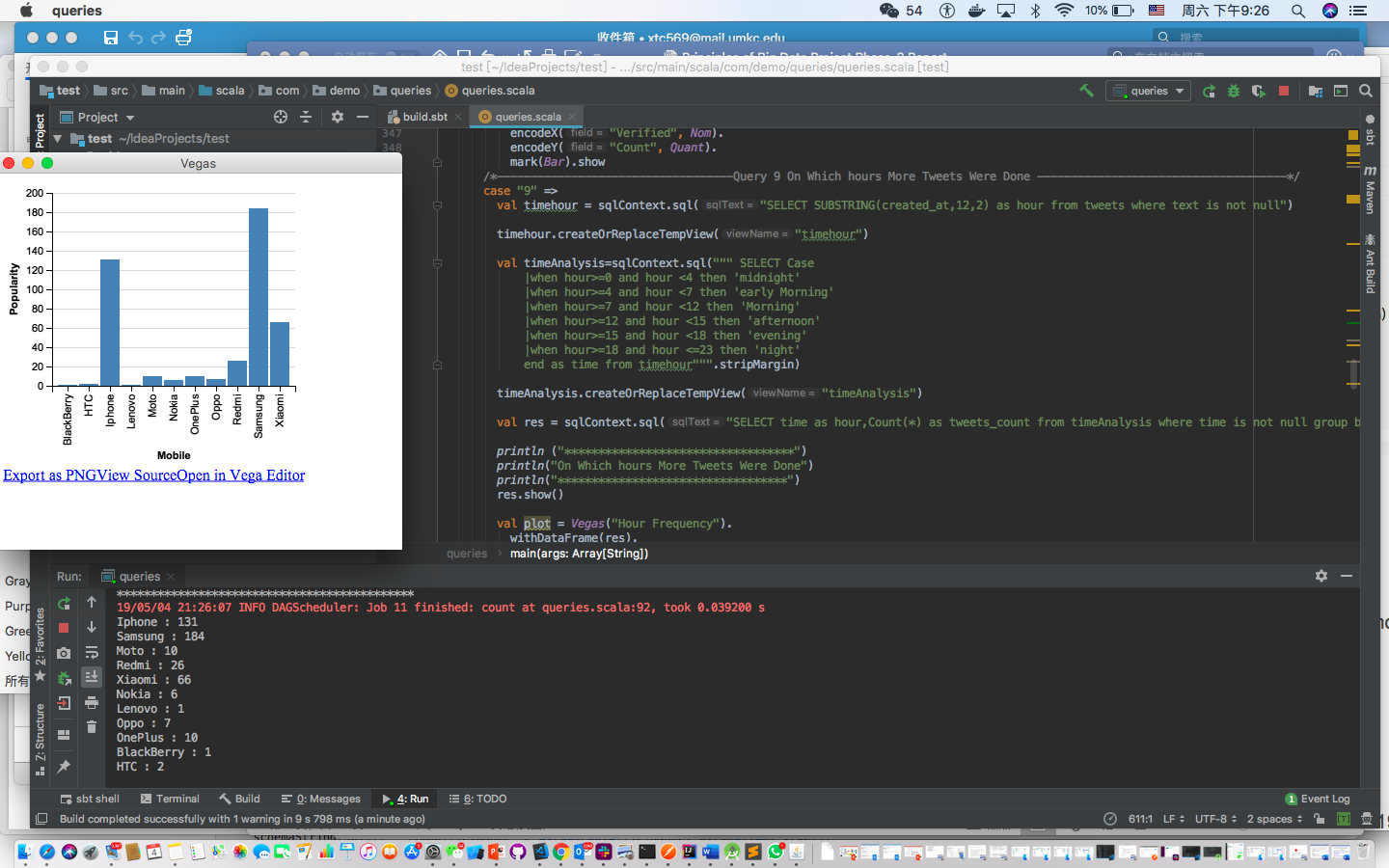
Such that built a query to perform the count operation on each model. So, it results the count -how many times a phone model appears in the tweets.

**Code:**

****

**Output (Include visualization):**

****

****

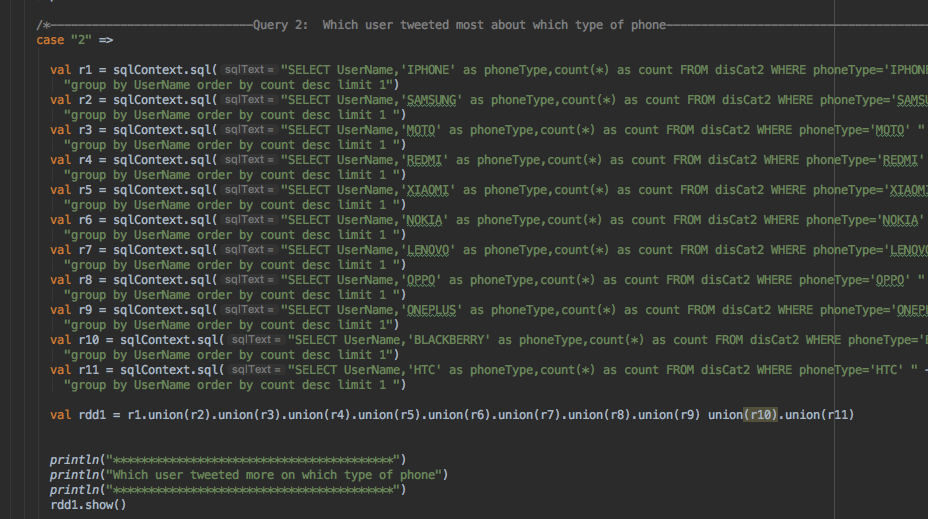
**Query 2: Query for fetching which user tweeted most about which type of phone.**

**Answer:** This query is built to analyze the tweets – based on which user tweeted most about which type of phone.

As we collected the data related to mobile phones **#iphone'**,**'#Samsung'**,**'#Moto'**,**'#Redmi'**,**'#Xiaomi'**,**'#Nokia'**,**'#lenovo'**,**'#oppo'**,**'#OnePlus'**,**'#BlackBerry'**,**'#HTC**

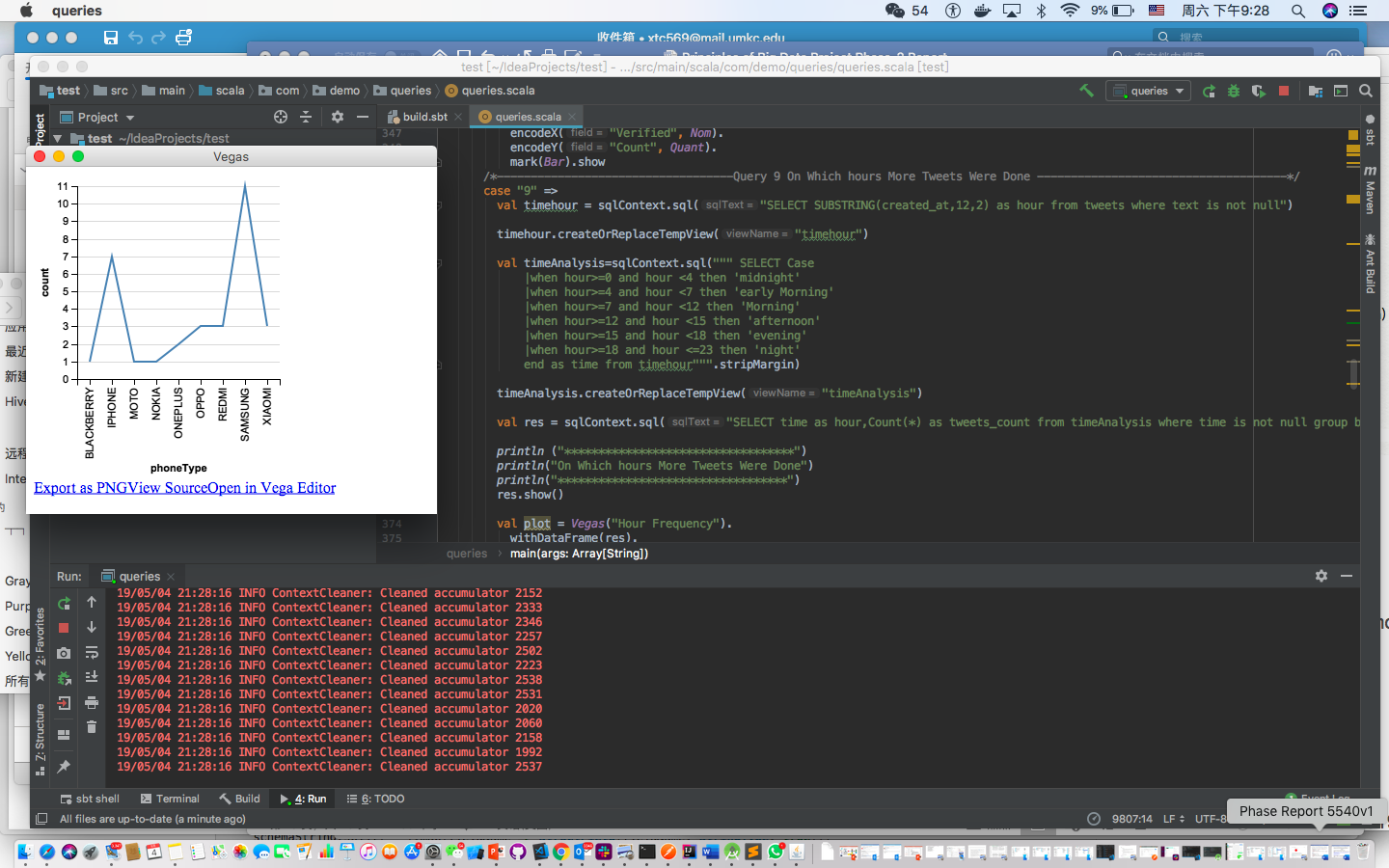
Such that built a query to perform the user tweeted most about which type of phone. So, it results the count -how many times a user tweeted at most for each mobile phone.

**Code:**

****

**Output(include visualization):**

****

****

**Query 3: Query for fetching Tweets from different countries about phones**

**Answer:** This query is built to analyze the tweets – based on which country tweeted more about on phones.

As we collected the data related to mobile phones **#iphone'**,**'#Samsung'**,**'#Moto'**,**'#Redmi'**,**'#Xiaomi'**,**'#Nokia'**,**'#lenovo'**,**'#oppo'**,**'#OnePlus'**,**'#BlackBerry'**,**'#HTC**

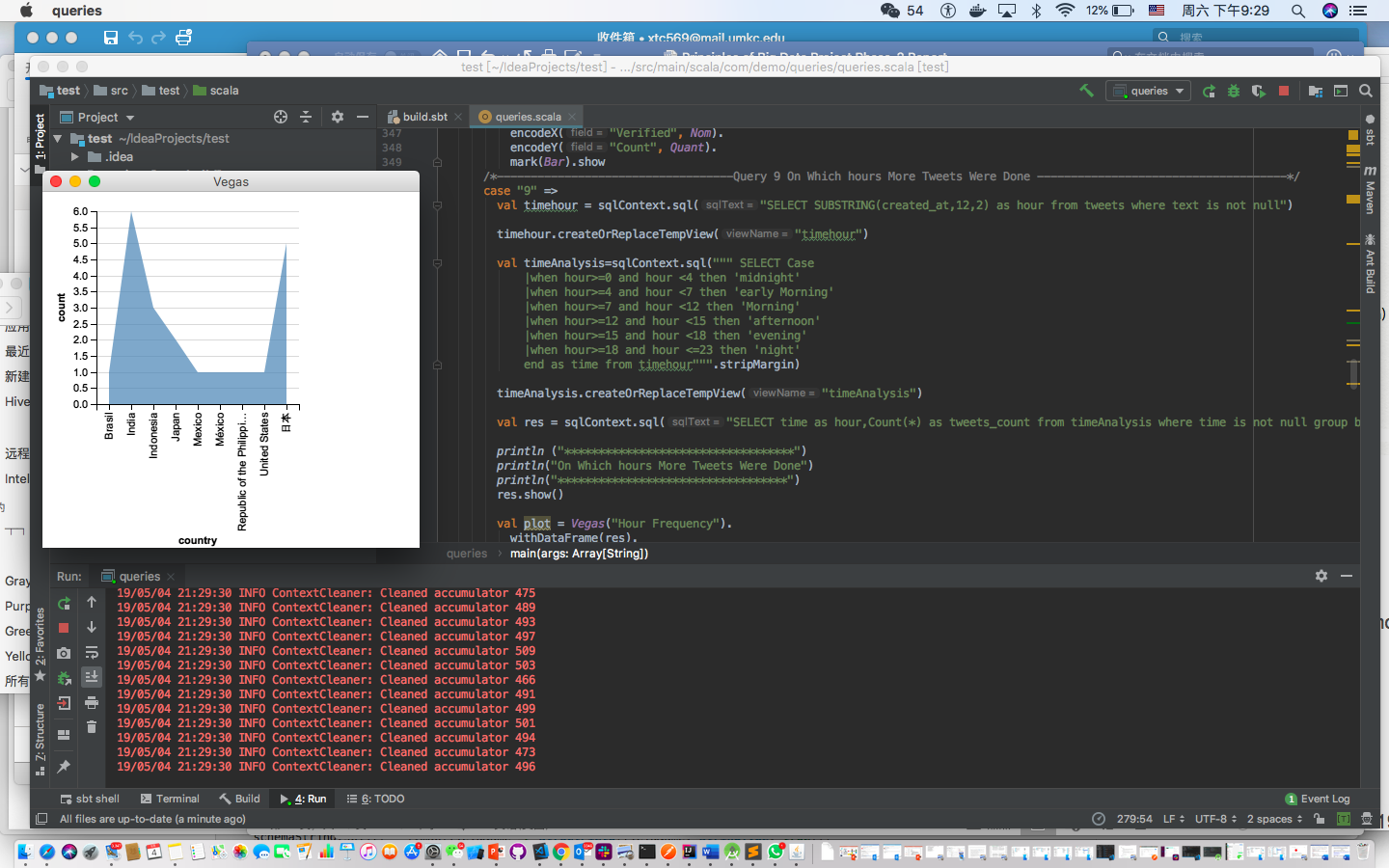
Such that built a query for finding out tweets from different countries about phones. So, it results the count -how many tweets from different countries posted about phones.

**Code:**

****

**Output:**

****

****

**Query 4: Query for on which day more tweets are done.**

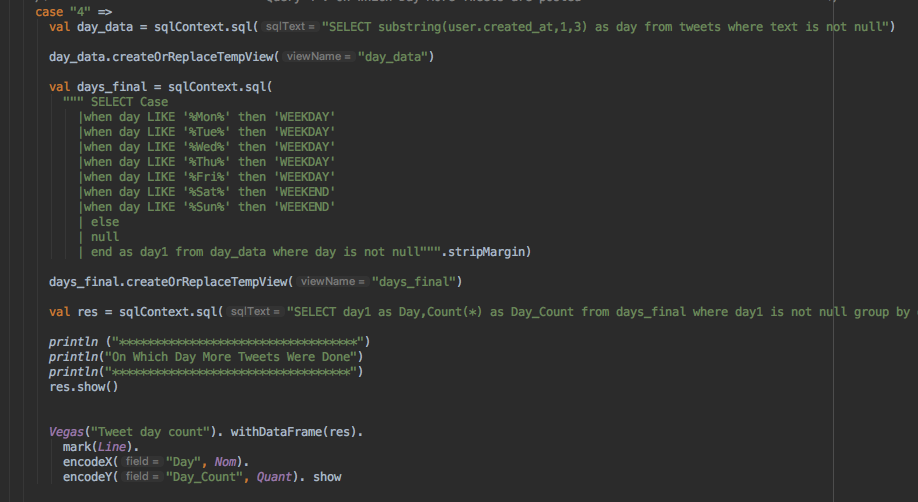
**Answer:** This query is built to analyze the tweets – based on which day more tweets are done.

MONDAY, TUESDAY, WEDNESDAY, THURSDAY,FRIDAY as Weekday

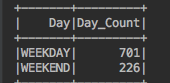
SATURDAY, SUNDAY as Weekend

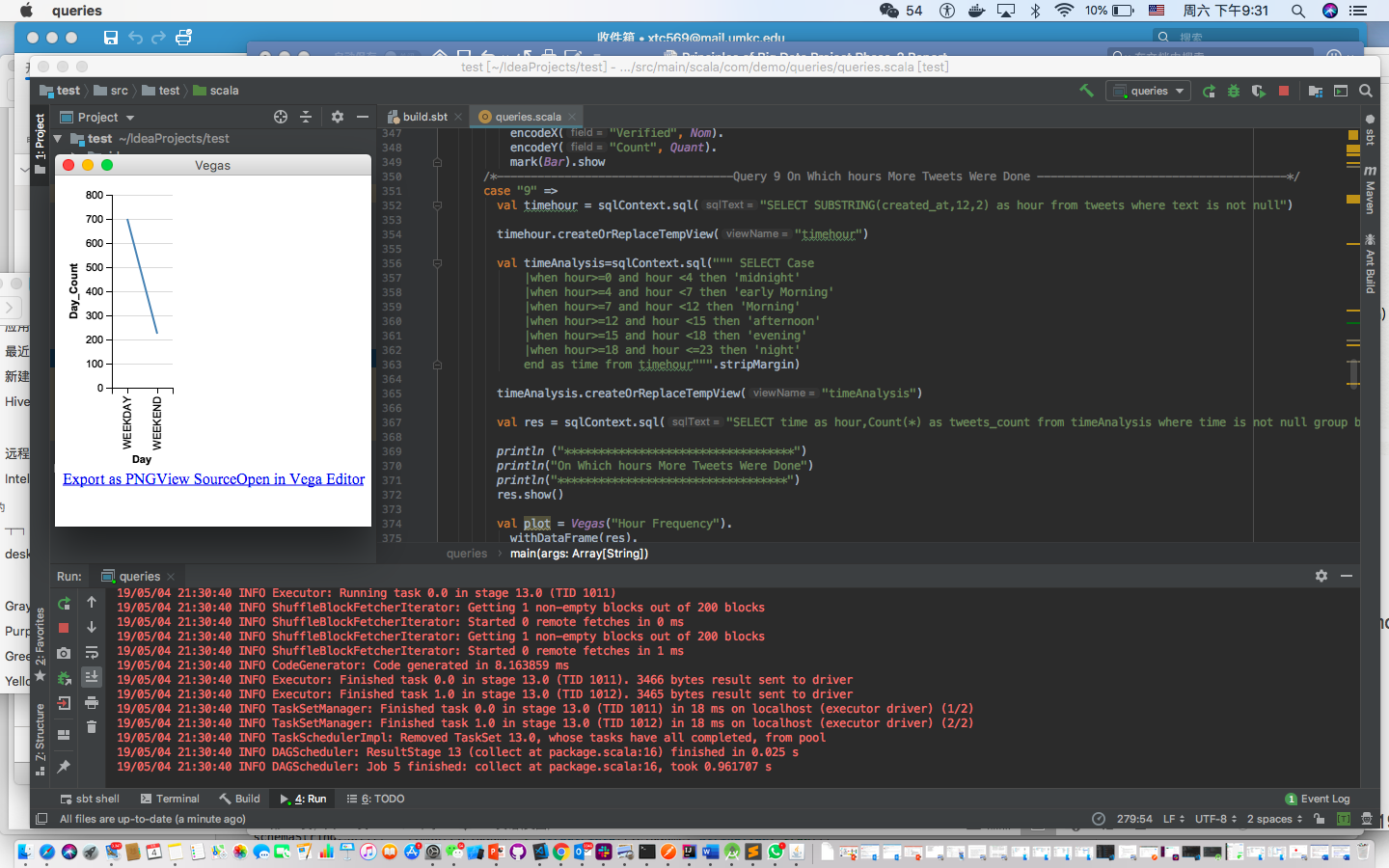
As we collected the data related to mobile phones **#iphone'**,**'#Samsung'**,**'#Moto'**,**'#Redmi'**,**'#Xiaomi'**,**'#Nokia'**,**'#lenovo'**,**'#oppo'**,**'#OnePlus'**,**'#BlackBerry'**,**'#HTC**

Such that built a query to perform for analyzing on which day the more tweets are posted. So, it results the count – about giving a figure of on what day & how many tweets are done.

**Code:**

**Output(include visualizaton):**

****

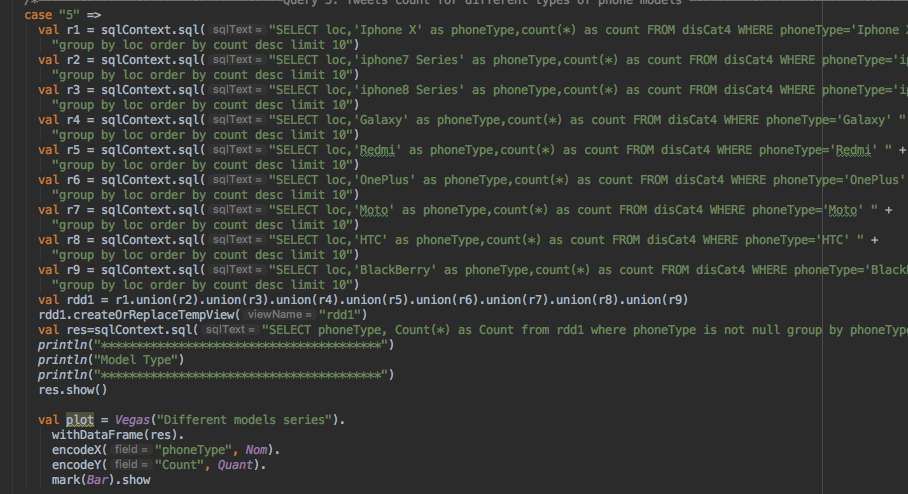
****

**Query 5: This query fetches Tweets count for different series of phone models.**

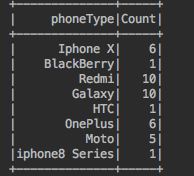
**Answer:** This query is built to analyze the tweets – for the individual series of each mobile phone.

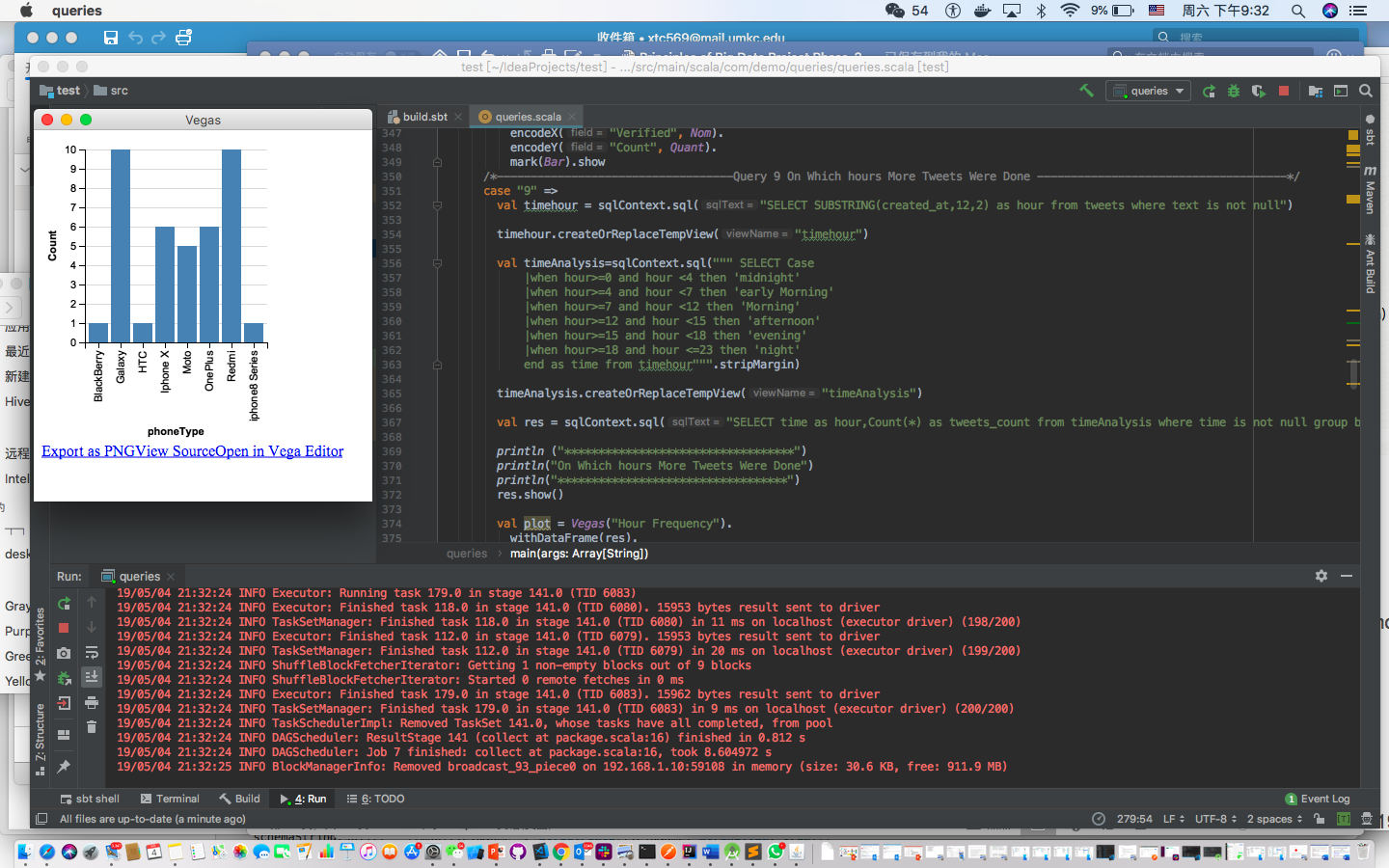
As we collected the data related to mobile phones **#iphone'**,**'#Samsung'**,**'#Moto'**,**'#Redmi'**,**'#Xiaomi'**,**'#Nokia'**,**'#lenovo'**,**'#oppo'**,**'#OnePlus'**,**'#BlackBerry'**,**'#HTC**

Such that built a query to perform the user tweeted most which type of series for mobile phone. So, it results the tweets count -for different series of phone models.

**Code:**

**Output(Visualization):**

****

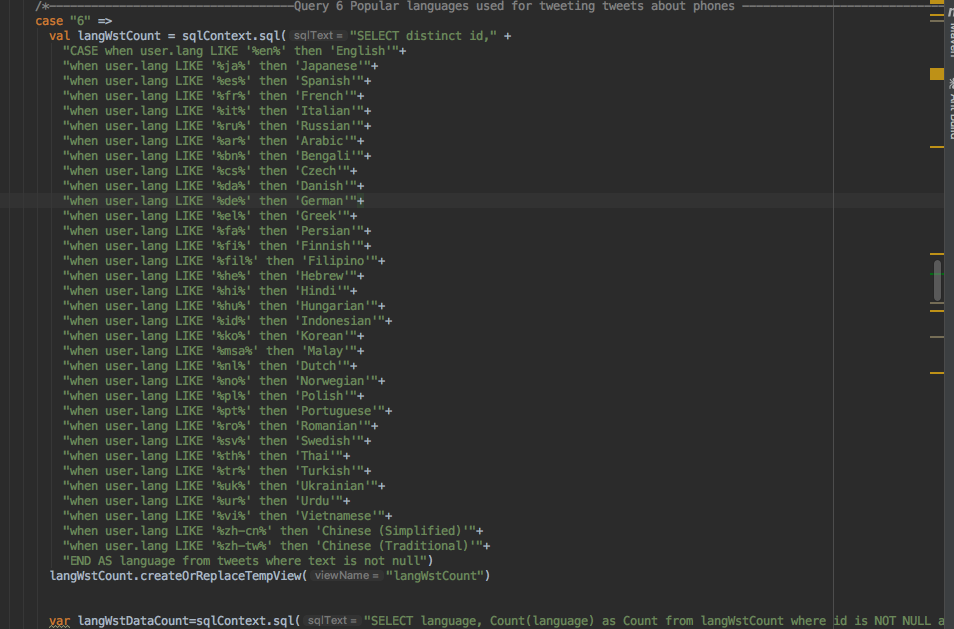
****

**Query 6: Query for fetching Popular languages used for tweeting tweets about phones.**

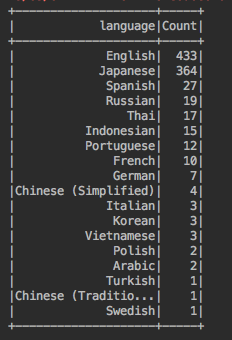
**Answer:** This query is built to analyze the tweets – popular languages used for tweeting tweets about phones.

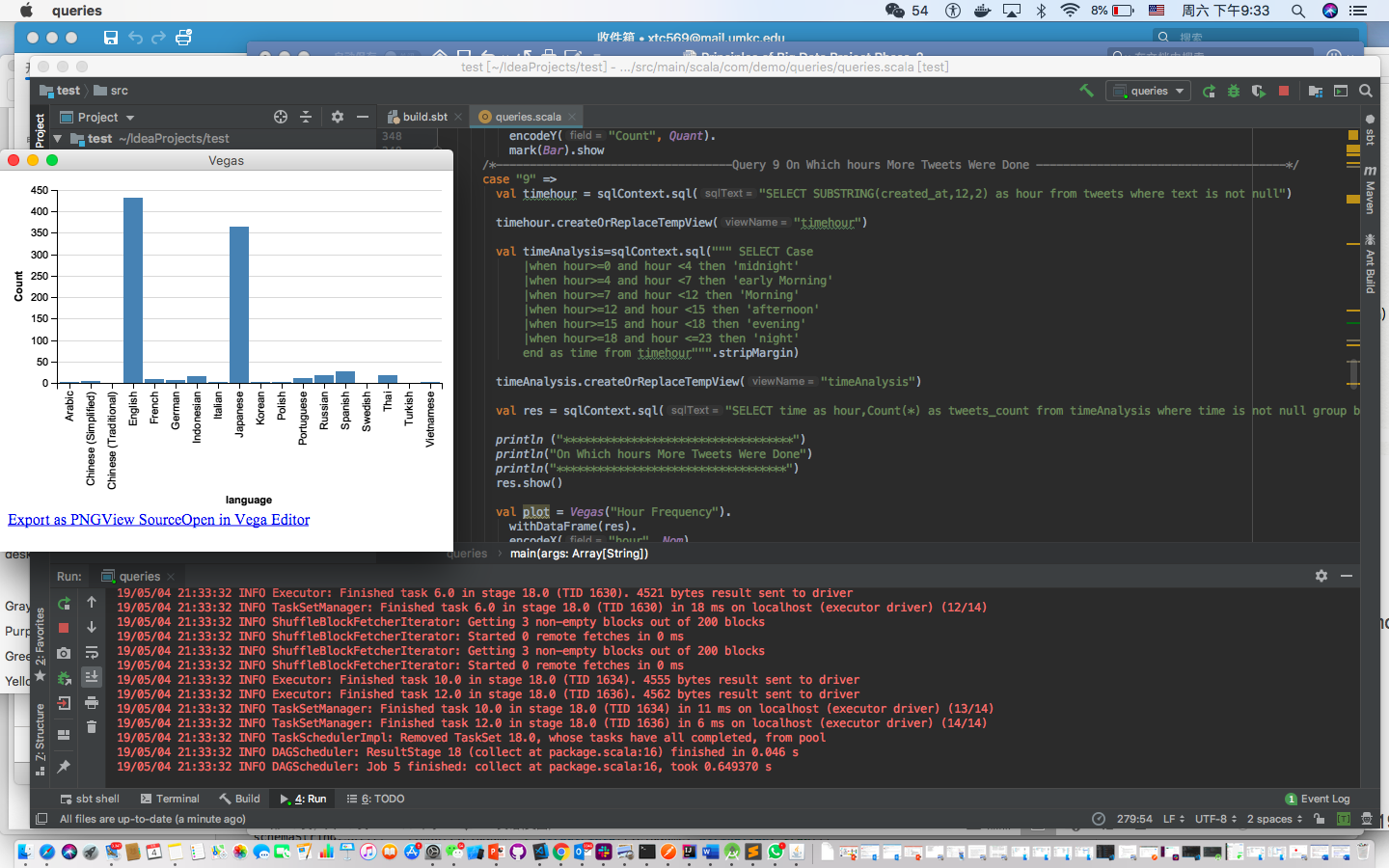
As we collected the data related to mobile phones **#iphone'**,**'#Samsung'**,**'#Moto'**,**'#Redmi'**,**'#Xiaomi'**,**'#Nokia'**,**'#lenovo'**,**'#oppo'**,**'#OnePlus'**,**'#BlackBerry'**,**'#HTC**

Such that built a query to analyze the popular languages used. So, it results the count – of most used language by the users.

**Code:**

**Output(Visualization):**

****

****

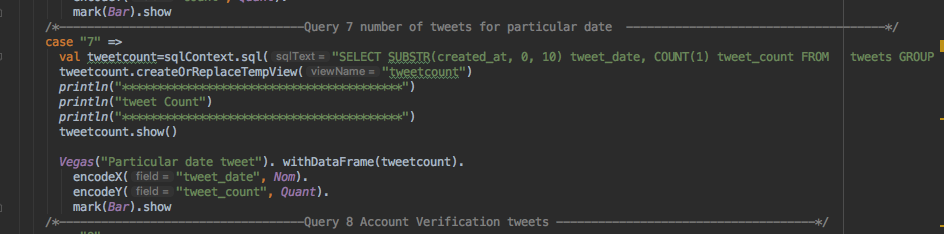
**Query 7: Query for fetching Number of tweets for date.**

**Answer:** This query is built to analyze the tweets – based on each individual date how many tweets are posted.

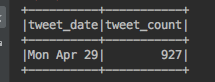
As we collected the data related to mobile phones **#iphone'**,**'#Samsung'**,**'#Moto'**,**'#Redmi'**,**'#Xiaomi'**,**'#Nokia'**,**'#lenovo'**,**'#oppo'**,**'#OnePlus'**,**'#BlackBerry'**,**'#HTC**

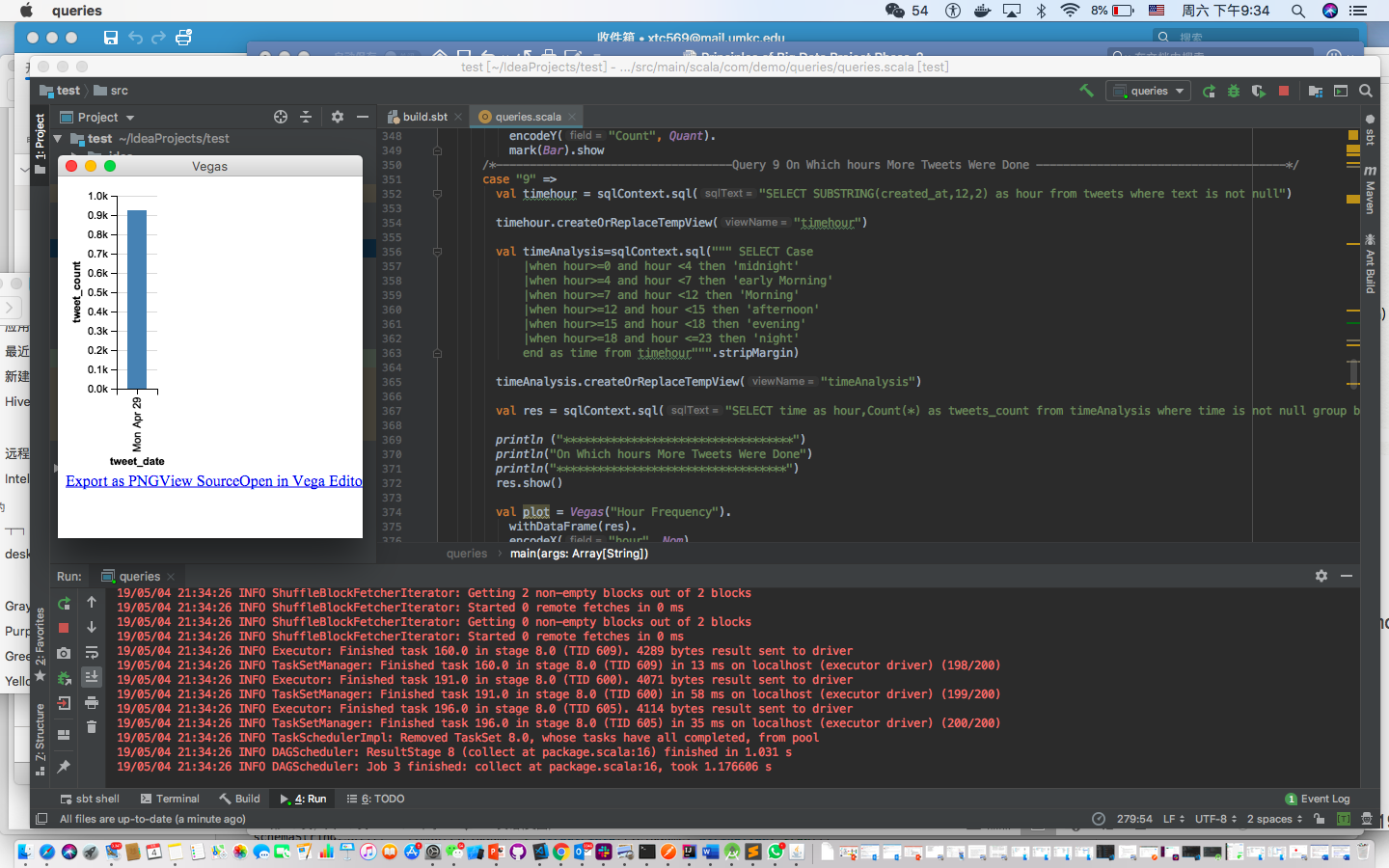
Such that built a query to analyze, depending up on the collected data for each date how many tweets are posted .

**Code:**

****

**Output(Visualization):**

****

****

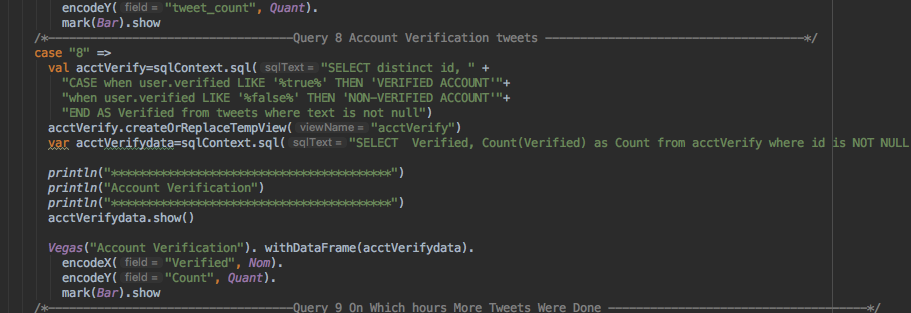
**Query 8: Query for fetching Account Verified tweets.**

**Answer:** This query is built to analyze the tweets – based on account verified tweets

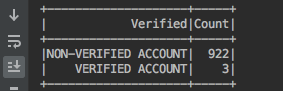
As we collected the data related to mobile phones **#iphone'**,**'#Samsung'**,**'#Moto'**,**'#Redmi'**,**'#Xiaomi'**,**'#Nokia'**,**'#lenovo'**,**'#oppo'**,**'#OnePlus'**,**'#BlackBerry'**,**'#HTC**

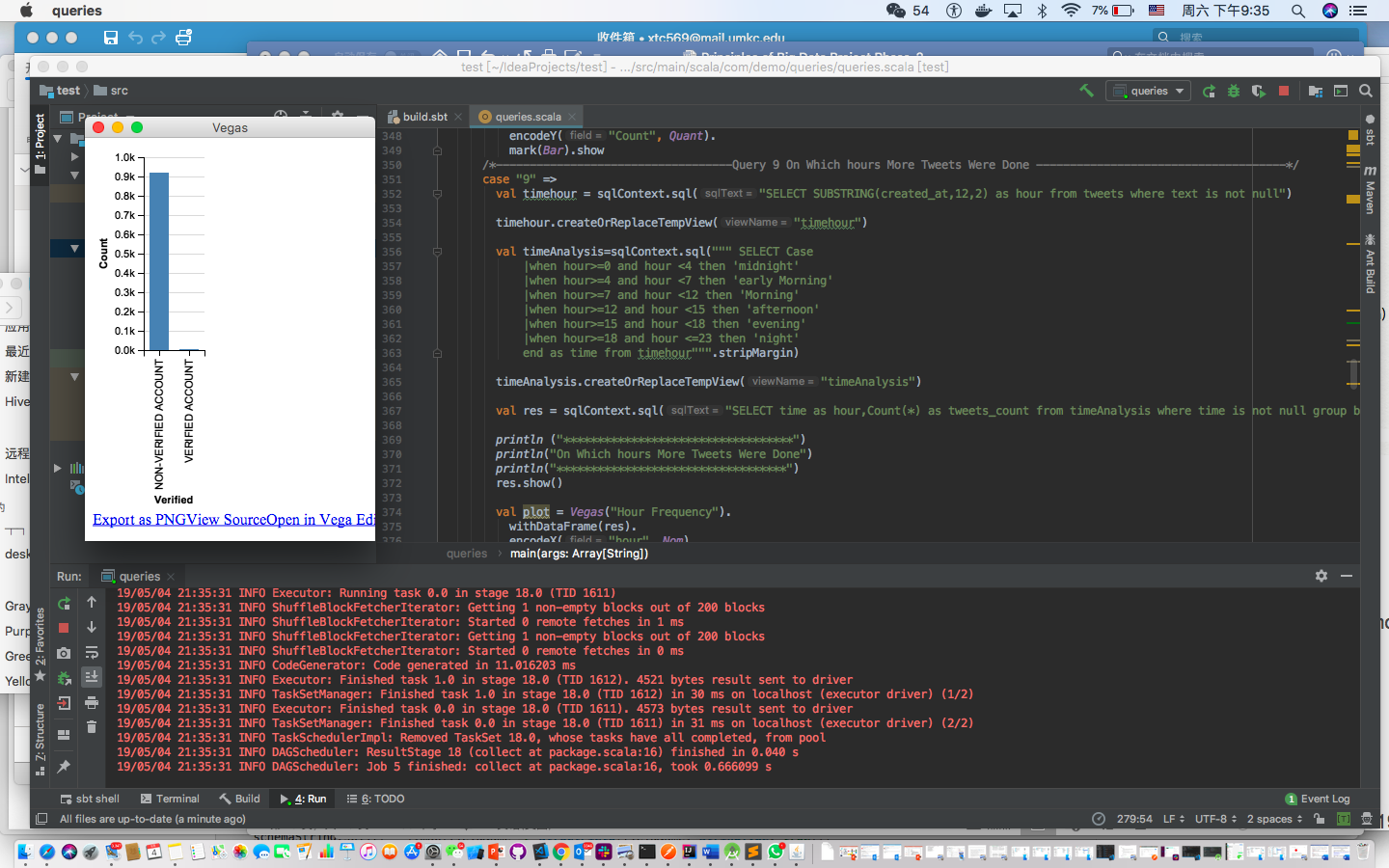
Such that built a query to analyze how many users are verified users. The accounts mainly differentiated by official check. Thus, counting the tweets which are posted from official accounts.

**Code:**

****

**Output(Visualization):**

****

****

**Query 9: Query for fetching On Which Hours More Tweets Were Done.**

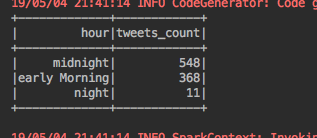
**Answer:** This query is built to analyze the tweets – based on hours of the tweets.

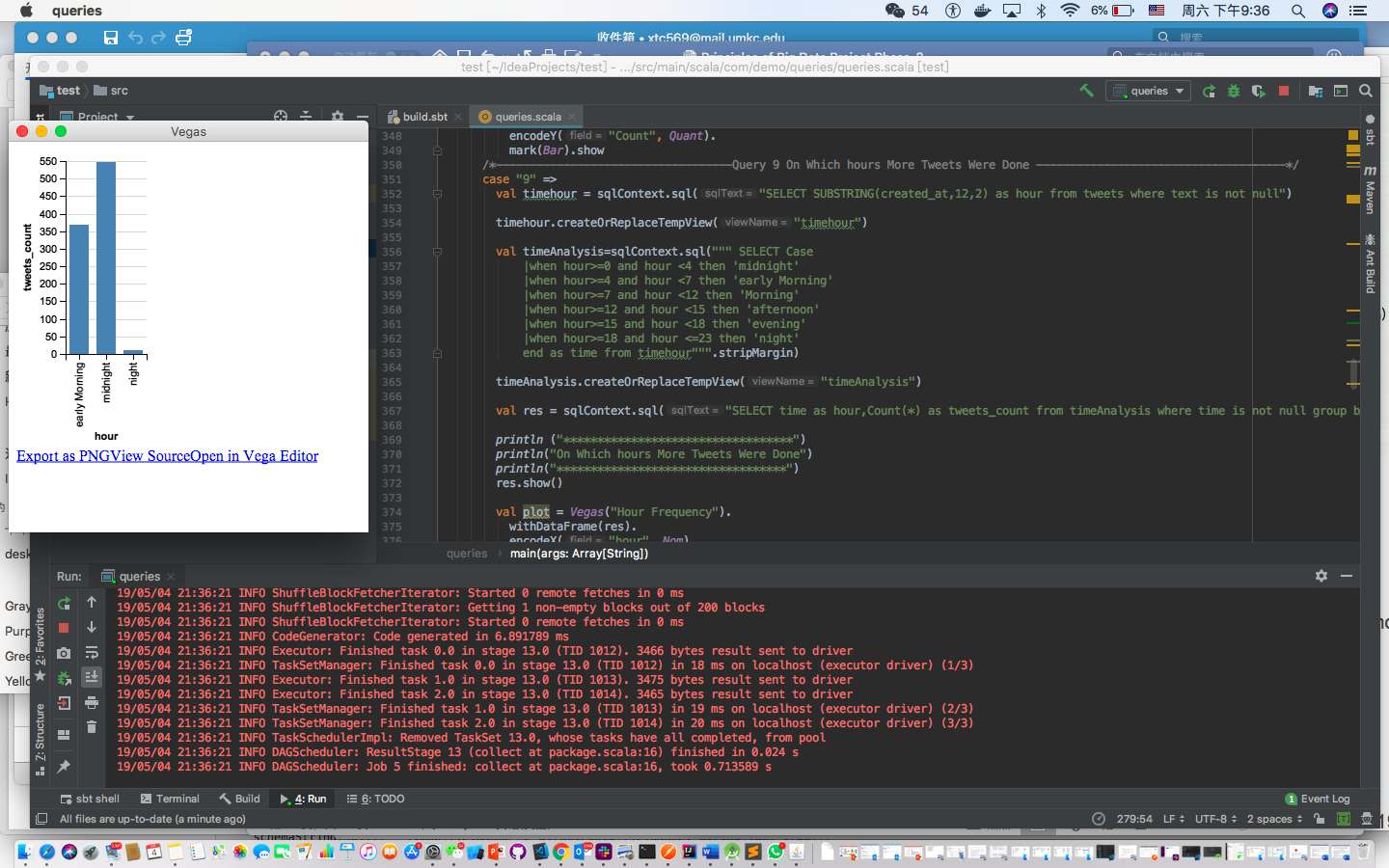
As we collected the data related to mobile phones **#iphone'**,**'#Samsung'**,**'#Moto'**,**'#Redmi'**,**'#Xiaomi'**,**'#Nokia'**,**'#lenovo'**,**'#oppo'**,**'#OnePlus'**,**'#BlackBerry'**,**'#HTC**

Such that built a query to analyze when the tweets are tweeted like-Morning, afternoon, night. So ,it results the time analysis of tweets

**Code:**

**Output(Visualization):**

****

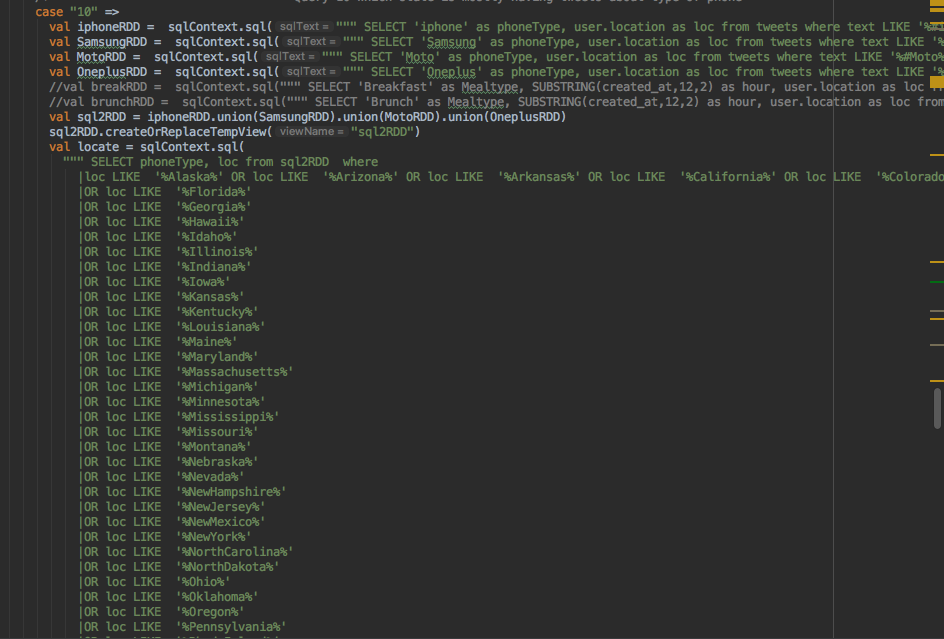
****

**Query 10: Query for fetching Which state is mostly having tweets about type of phone.**

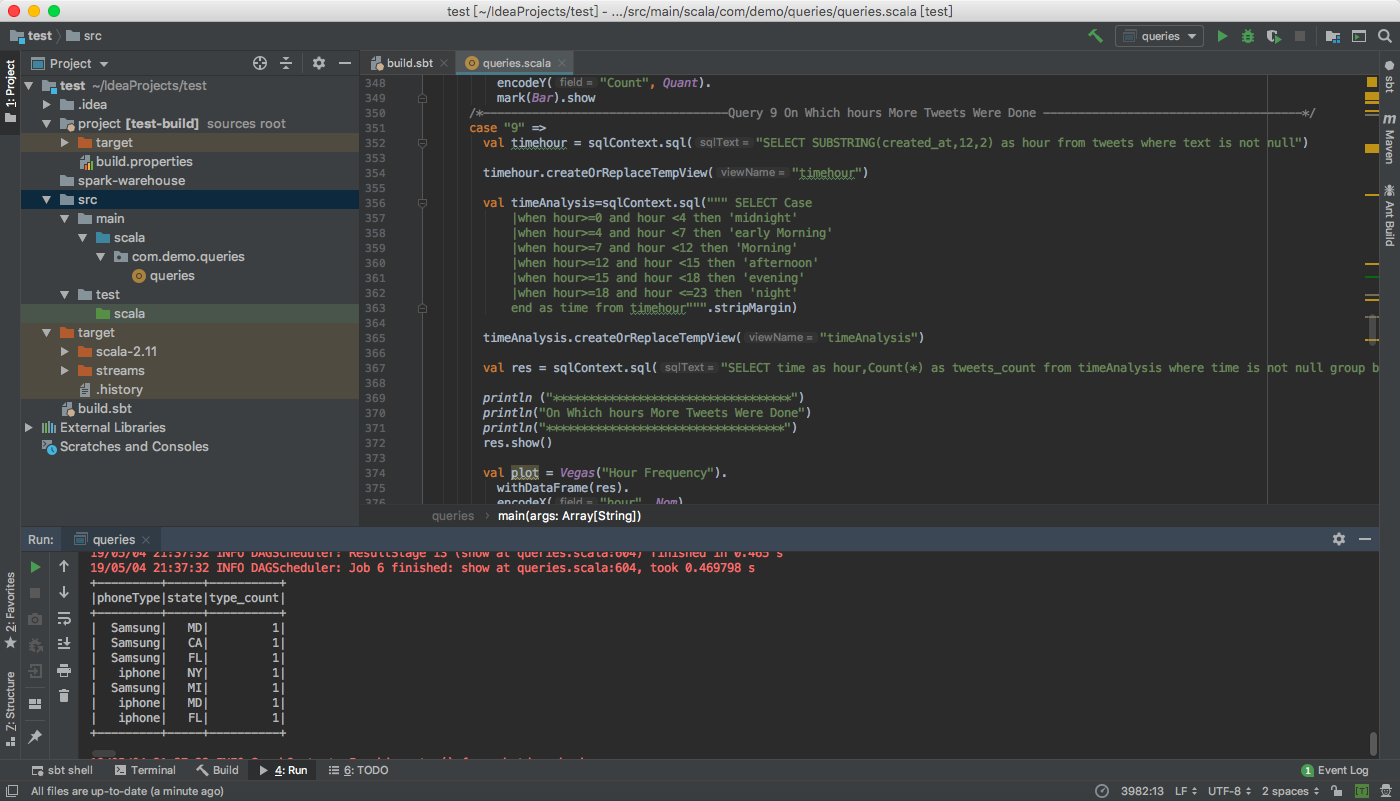
**Answer:** This query is built to analyze the tweets – **Which state is mostly having tweets about type of phone**

As we collected the data related to mobile phones **#iphone'**,**'#Samsung'**,**'#Moto'**,**'#Redmi'**,**'#Xiaomi'**,**'#Nokia'**,**'#lenovo'**,**'#oppo'**,**'#OnePlus'**,**'#BlackBerry'**,**'#HTC**

Such that built a query to perform

**Code:**

**Output:**

****

**Visualization:**

**A screenshot of a cell phone

Description generated with high confidence**

**Reference:**

<https://spark.apache.org/>

<https://hadoop.apache.org/>

<http://link.galegroup.com.contentproxy.phoenix.edu/apps/doc/A487904581/GIC?u=uphoenix_uopx&sid=GIC&xid=ead46665>

<https://www.vegas-viz.org/>

https://github.com/a190884810/CS5540