



**COLLECTING TWEETS USING TWITTER STREAMING API’S**

**Sub: PRINCIPLES OF BIG DATA**

|  |  |
| --- | --- |
| Sai Sri Narne | (16272589) |
| Bhargavi Saipoojitha Chennupati | (16271057) |

Instructor

Dr. PRAVEEN RAO, Ph.D

# Phase2 of project deals with Analysis of created data and also its visualization based on various parameters. Here we come with the quires as mentioned in the interim report.

**ABSTRACT:**

The fundamental point of this task is to break down the huge information gathered from online networking (twitter). In this task, we have gathered twitter information (tweets) on "NFL(National Football League)" and we have broken down the gathered enormous information utilizing Apache Spark. We have actualized distinctive questions utilizing Spark Data outlines and an open API to break down the gathered information and drawn some fascinating yields from inquiry investigation.

**IMPLEMENTATION:**

* + Collected twitter data (tweets) related to “NFL(National Football League)” in JSON format.
  + Developed the environment IntelliJ for Scala and Spark development.
  + Queries has been written and displayed as per the analysis.
  + Explanation of the ten queries and their outputs (captured screenshots) are documented.

**SETTING UP OF ENVIRONMENT:**

In our undertaking, we utilized IntelliJ for Scala and Spark improvement. IntelliJ Scala blend is the best,

free setup for Scala and Spark improvement. To run IntelliJ we require Java JDK introduced in our

Framework. Also, by utilize Spark APIs make Scala question and import Spark shakes as library

conditions in IntelliJ lastly add some Spark API calls to the made protest. Presently IntelliJ for Scala and

Spark improvement condition is setup and we are prepared to actualize distinctive questions (Spark RDDs

and Data frames) on our gathered stream of tweets for examination.

In this increment of project, we have taken the JSON file from the first phase and stored it in the form of

Main table and queries are written in the Spark SQL language for the extraction of the outputs and hash table is assigned for the output designed.

Ten queries are written in the Spark SQL language and executed in the SCALA code which was written for execution.

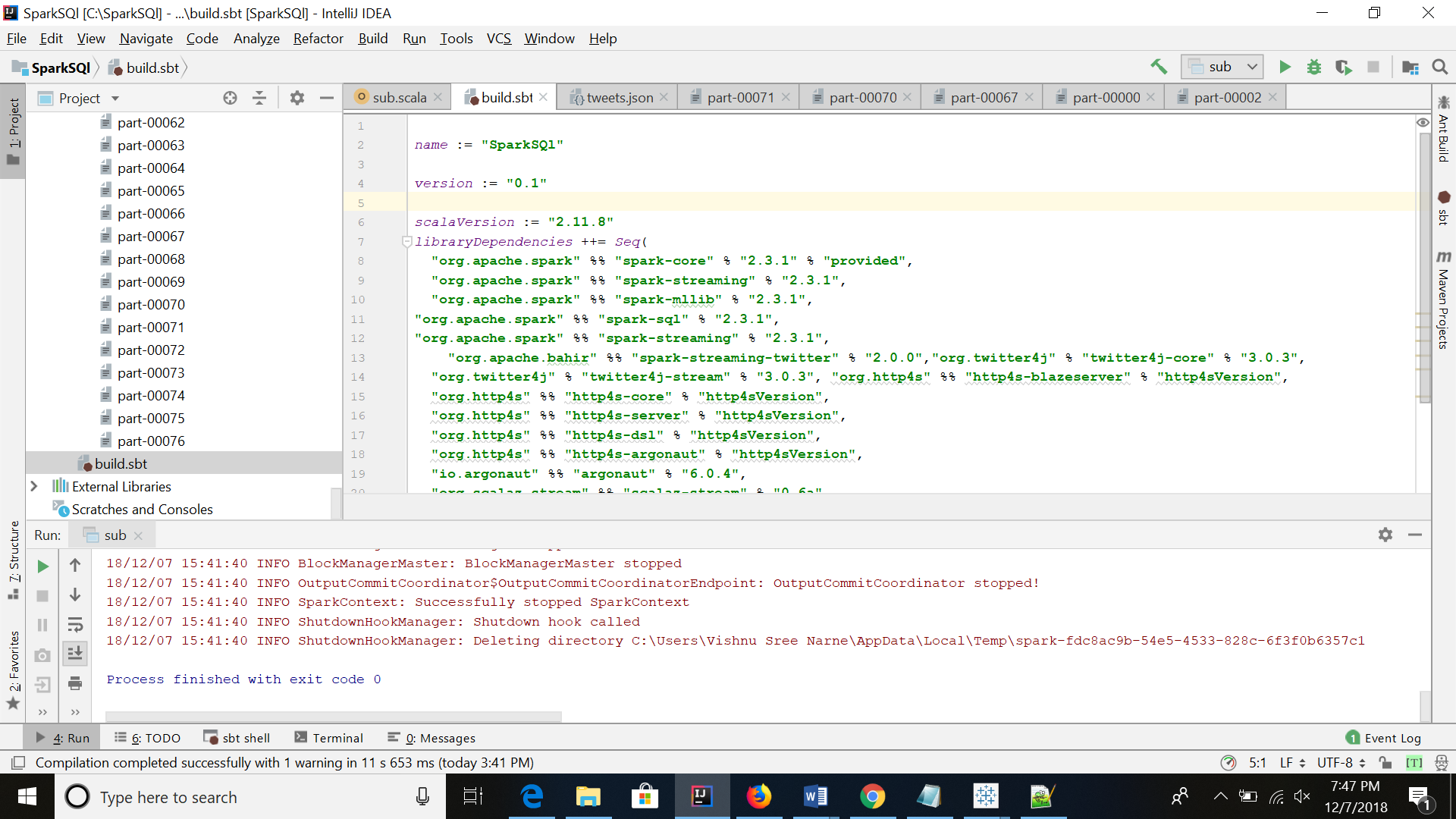
# TECHNOLOGIES USED:

# IntelliJ IDEA- Scala, Spark

# Tableau

# Queries and Analysis:

# Before the queries are analyzed in IntellijIDEA the .sbt file should be changed and modified in such a way that the support file is created between all versions.



# Scala Spark Code

# 

# Working of Spark Jobs

# 

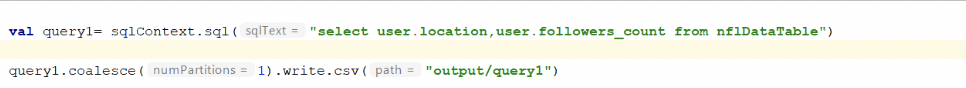
# 

# ANALYTICAL QUERIES:

# QUERY 1:

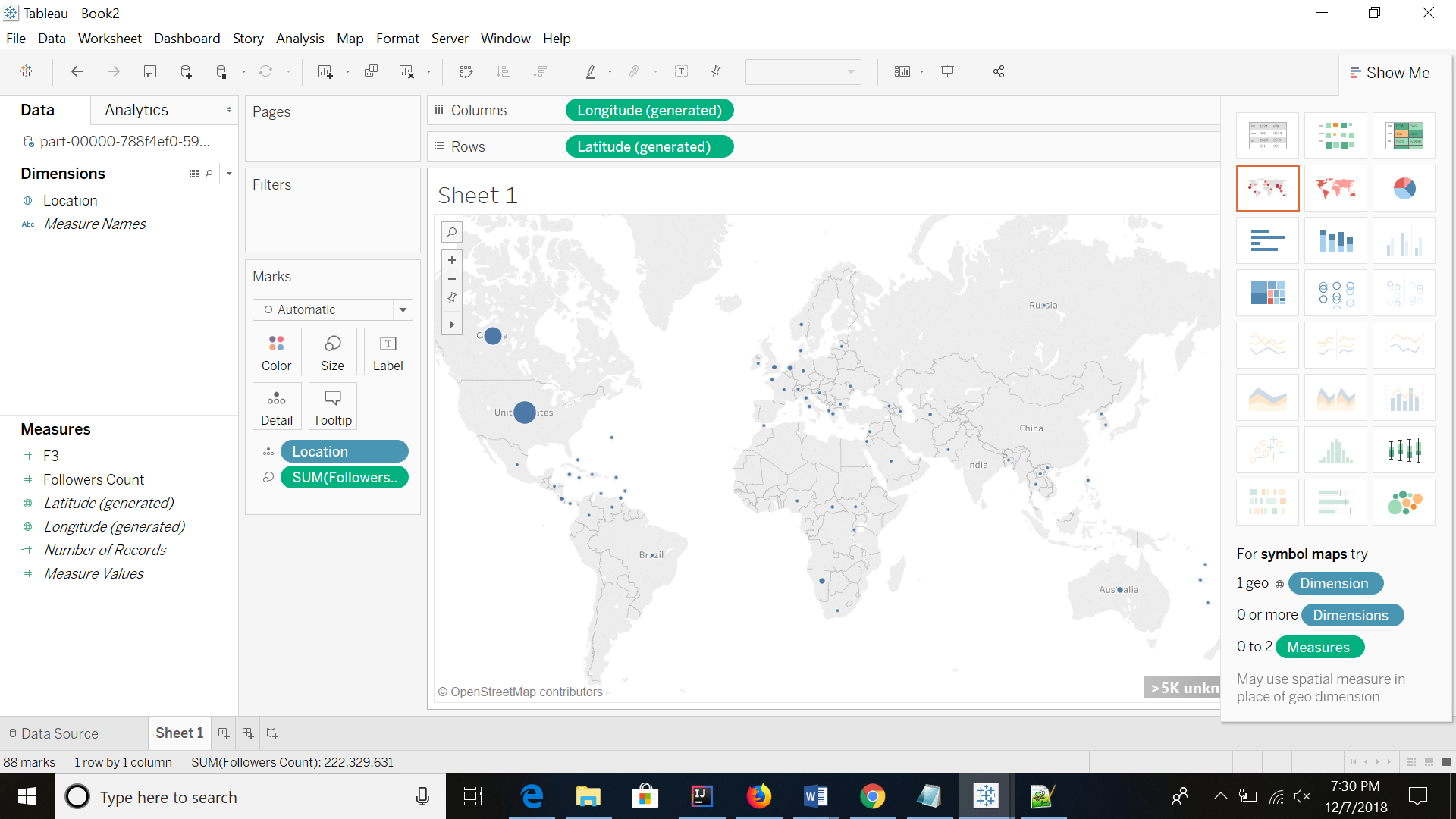
In this query it tells the location from which there are many tweets about NFL:

Code and analysis of data in IntellijIDEA:



Visualization:

Here, United States has the maximum followers count about Syria War.



# 

# Query 2:

In this Query we find the user with max number of followers about NFL

Code for analysis of data in IntellijIDEA:

# 

Visualization:

# 

# 

# 

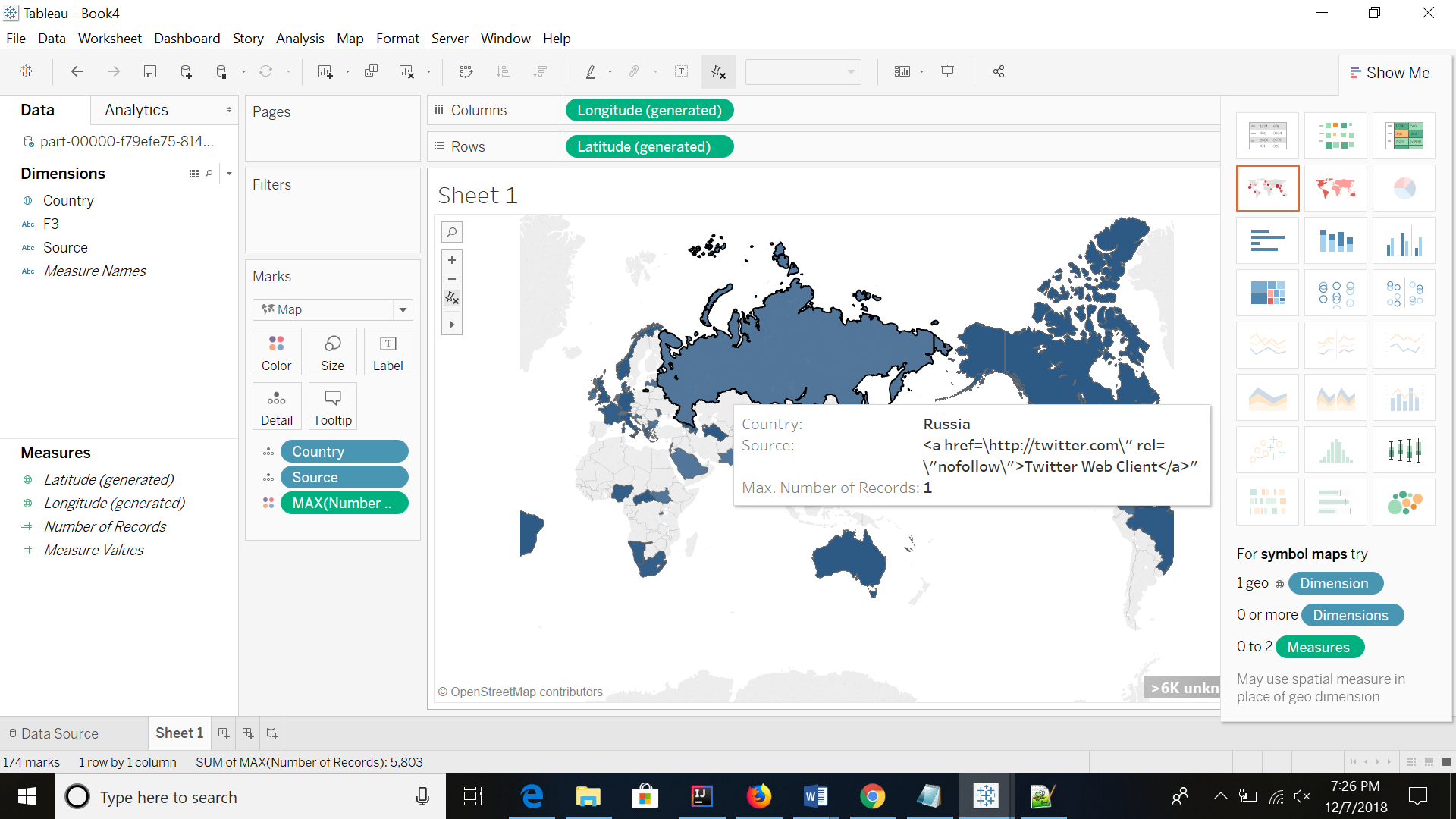
# Query 3:

In this Query we find the location from which there were more number of tweets.

Code for analysis of data in IntellijIDEA:



Visualization:

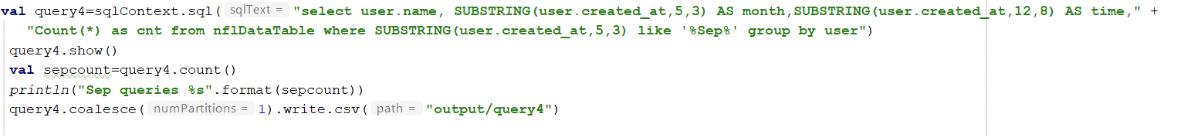


# 

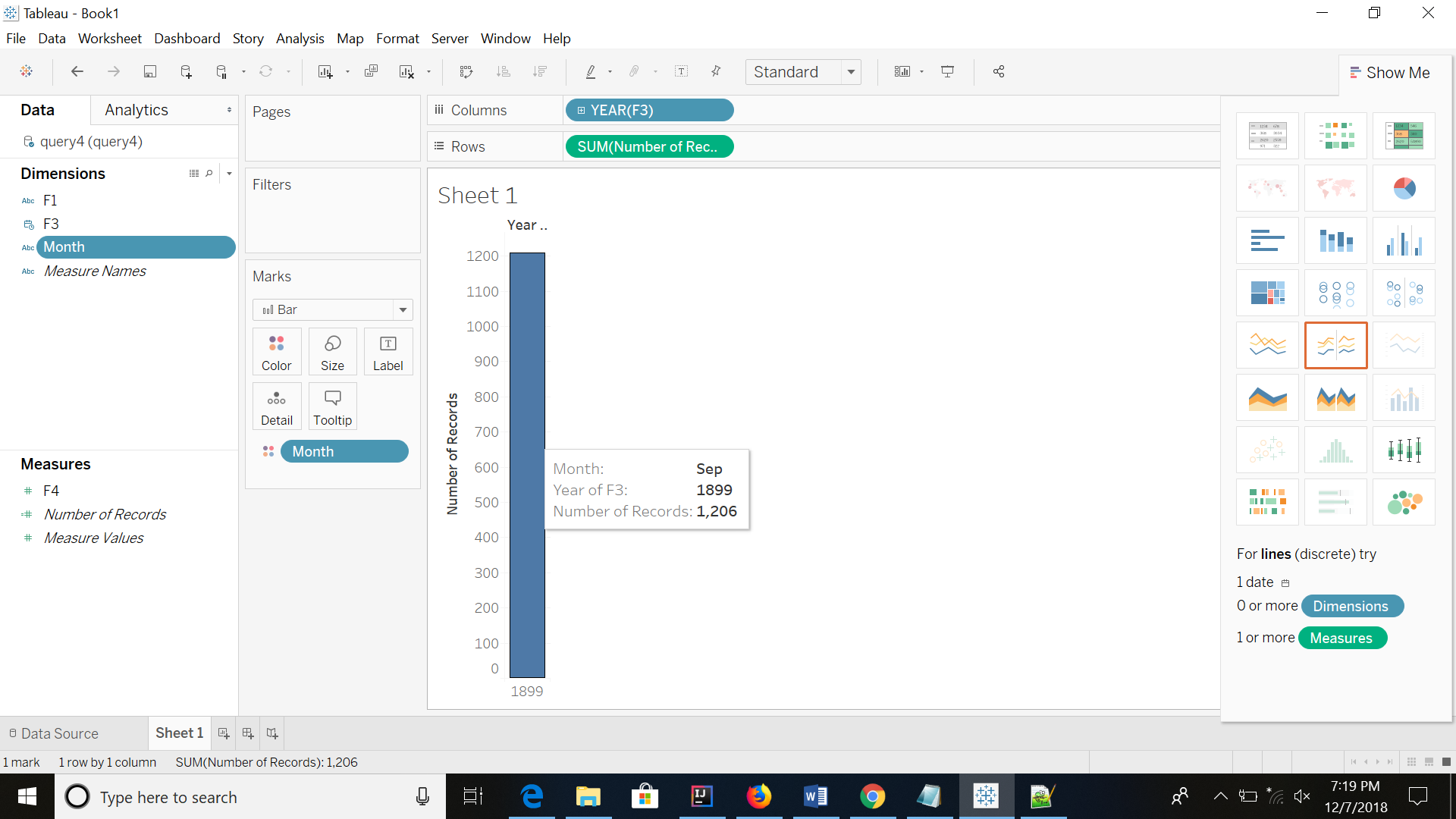
# Query 4:

In this Query we find the number of tweets generated in September.

Code for analysis of data in IntellijIDEA:



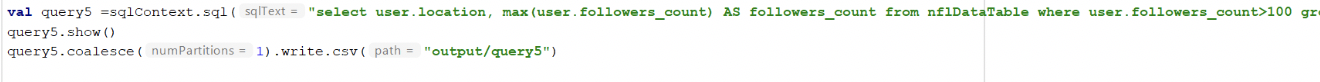
Visualization:



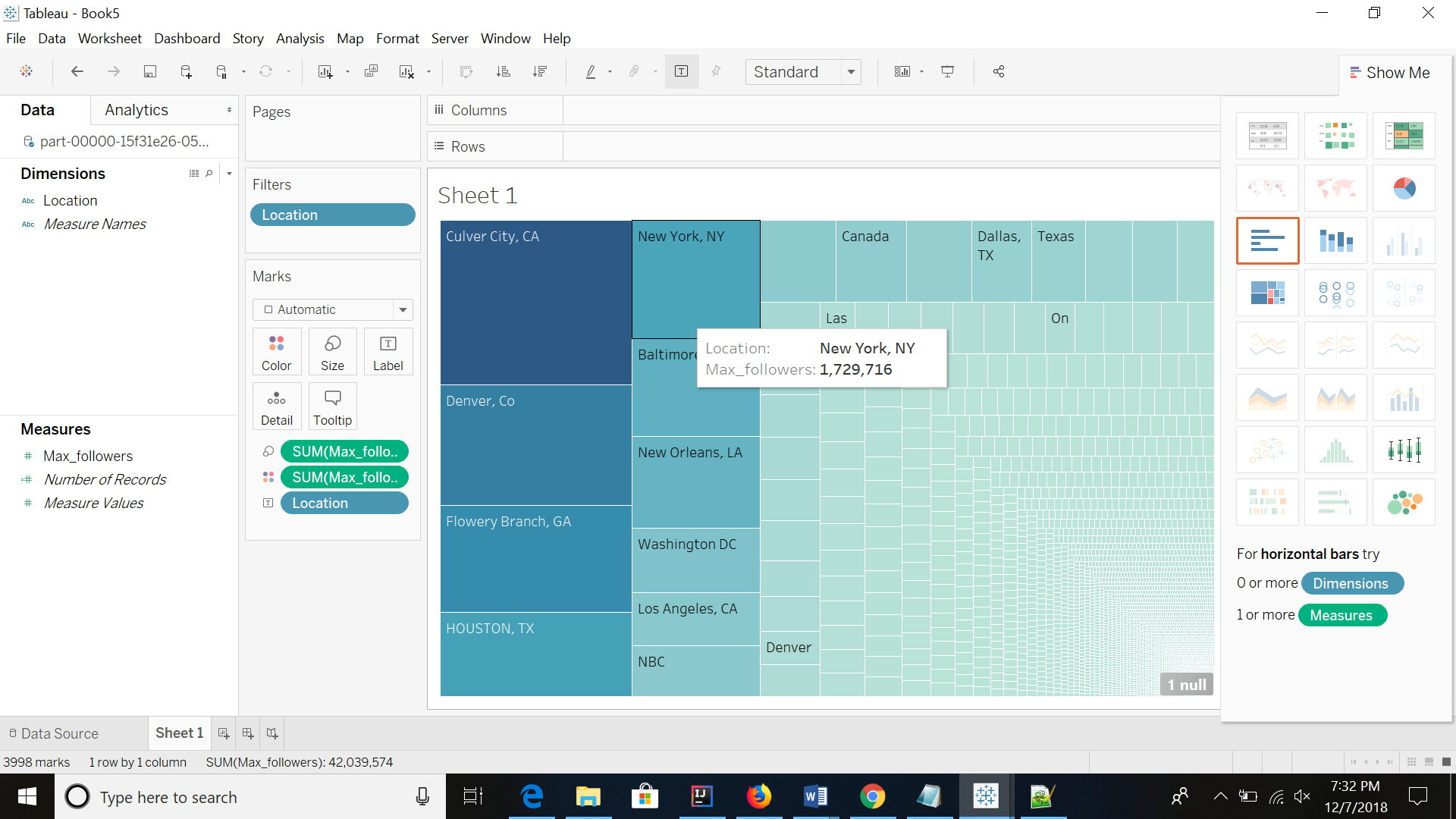
# Query 5:

This Query, it finds out the location in which the users have maximum followers about NFL

Code for analysis of data in IntellijIDEA:



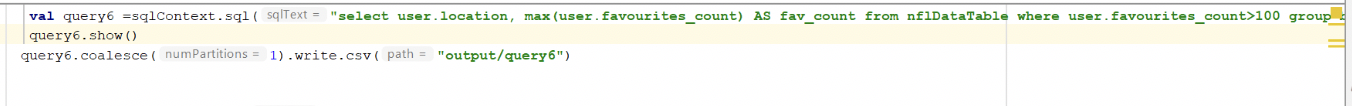
Visualization:



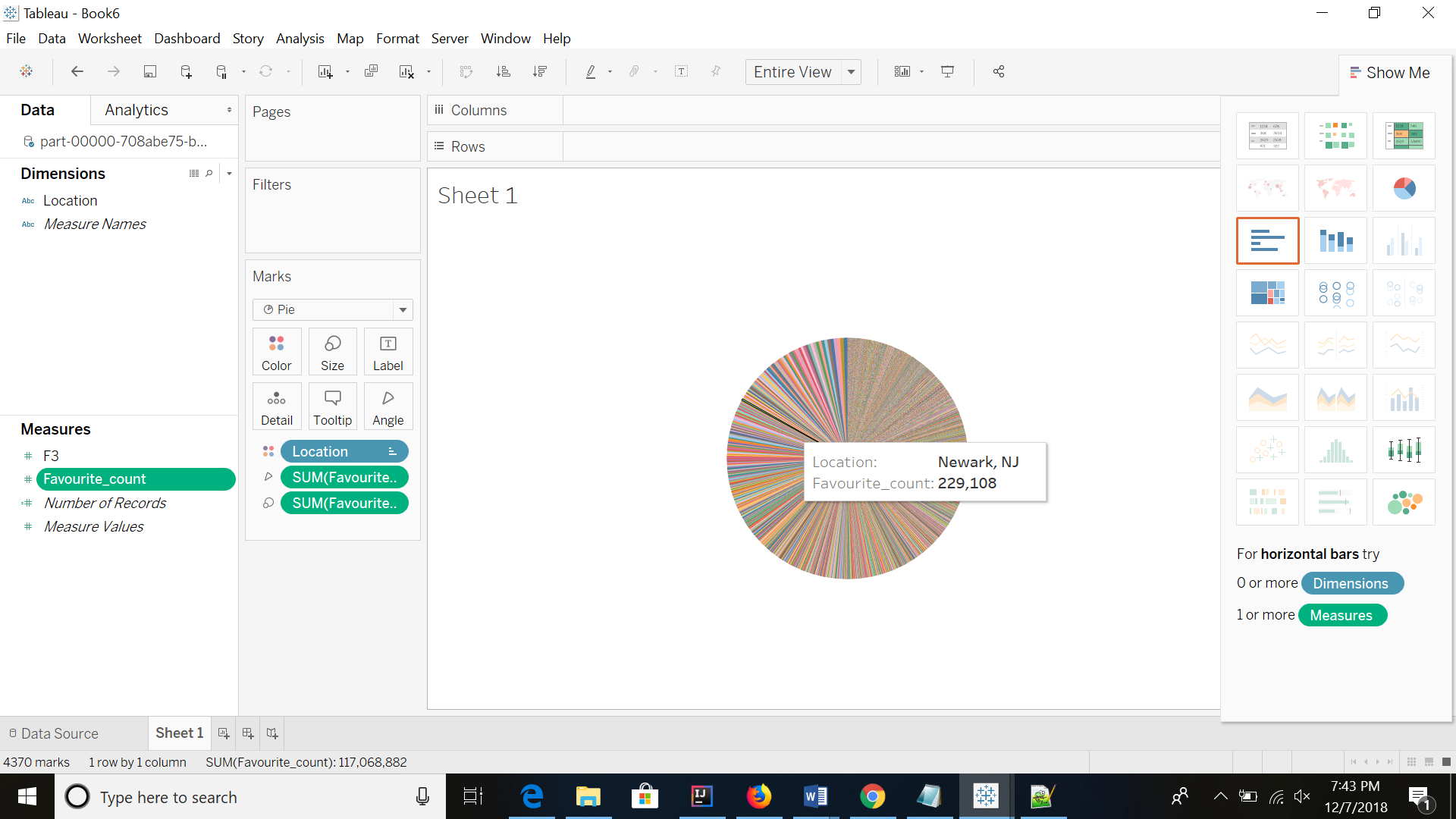
# Query 6:

In this Query which finds out the location in which the users have maximum favourites about NFL

Code for analysis of data in IntellijIDEA:



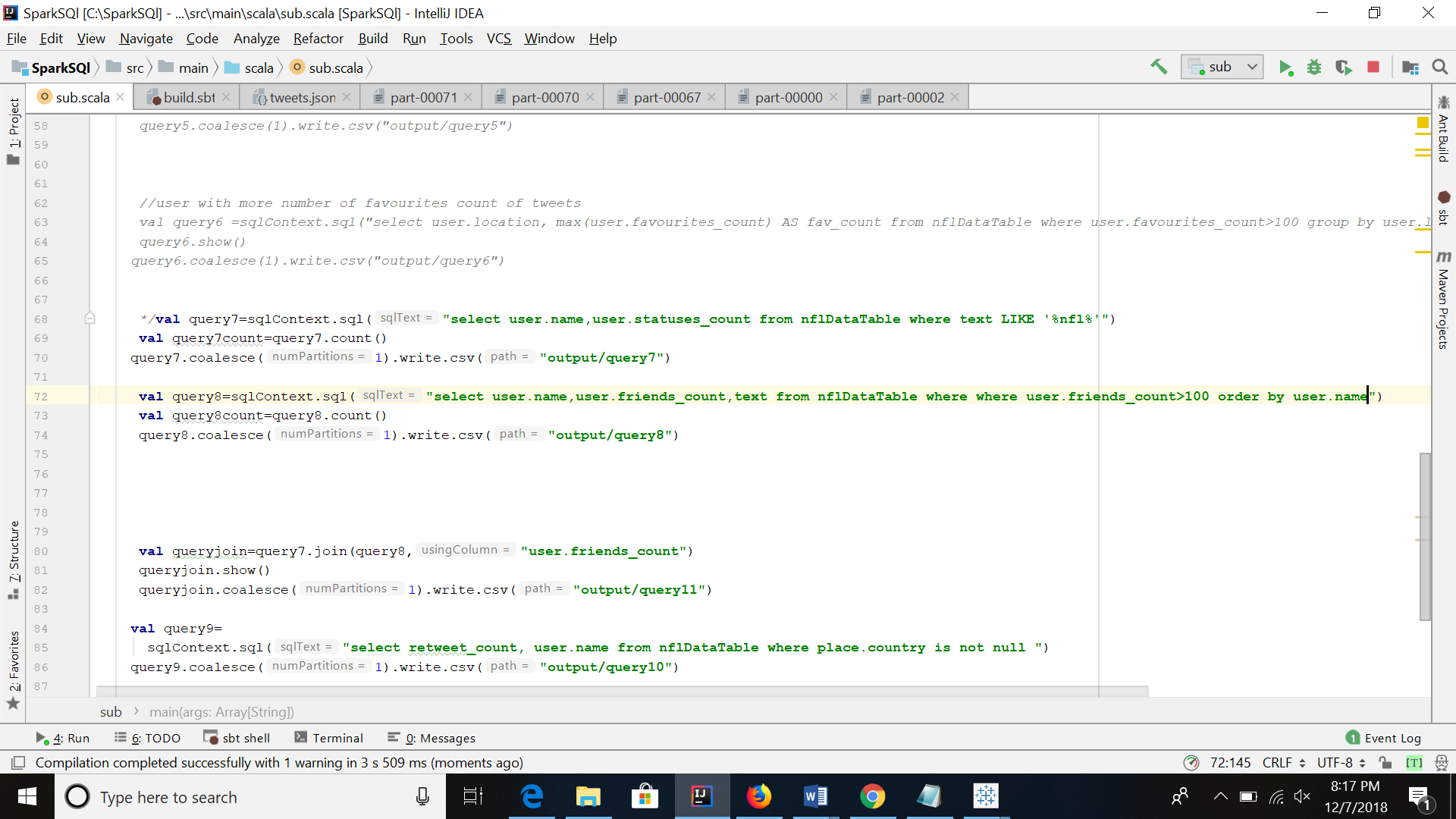
Visualization:



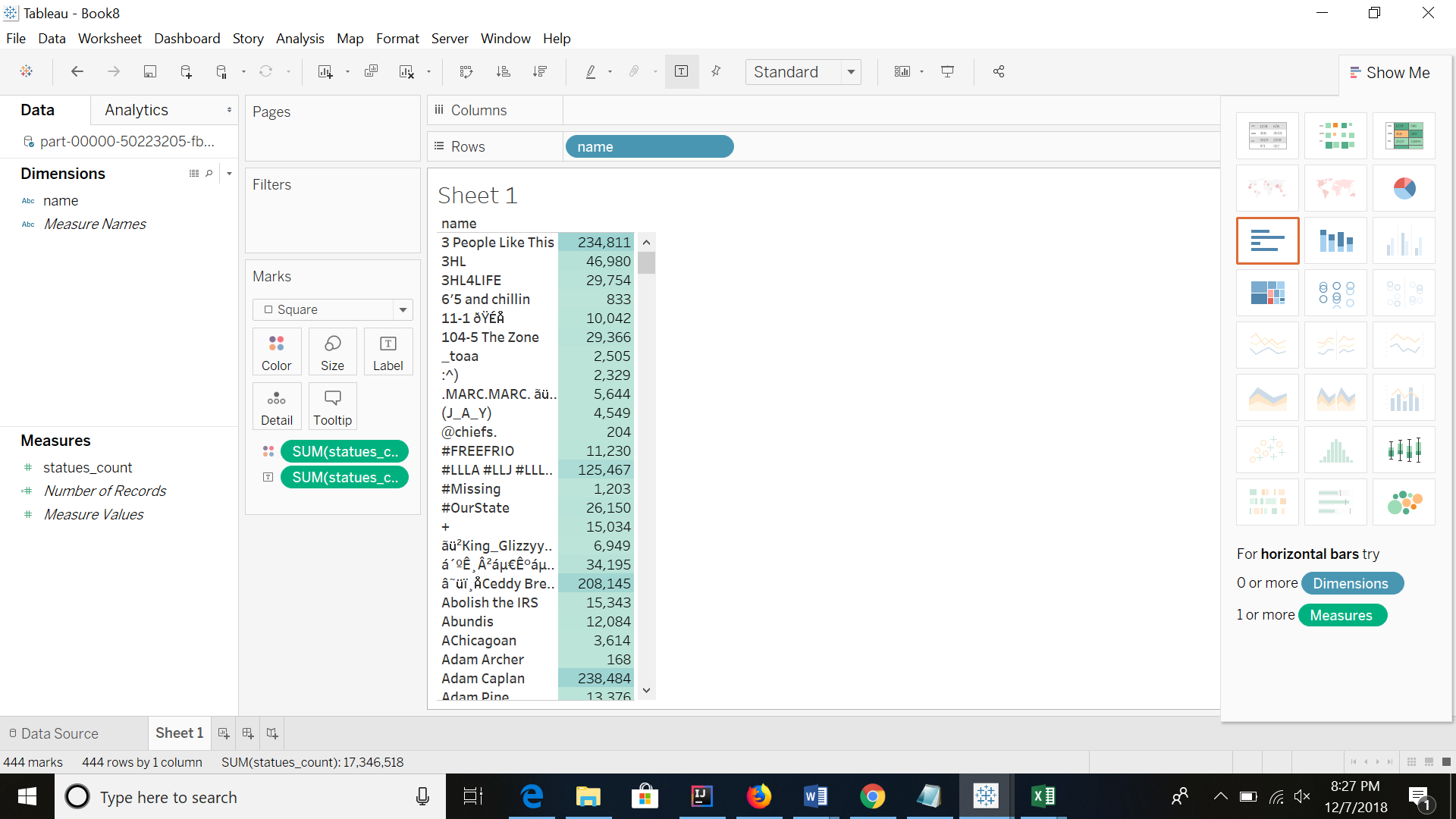
# Query 7:

In this Query which finds out the location in which the users have maximum Statuses about NFL

Code for analysis of data in IntellijIDEA:



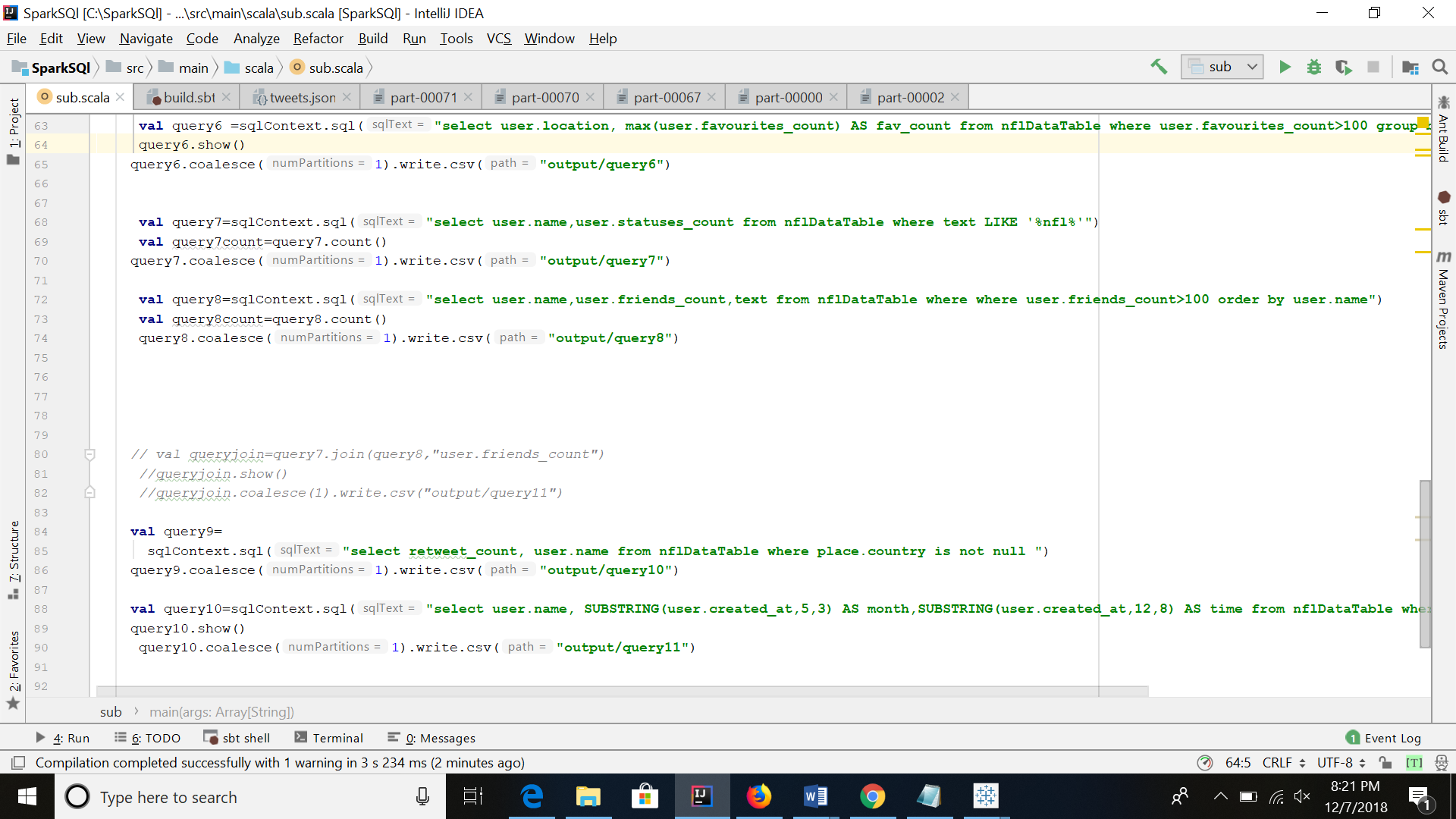
Visualization:

­­­­­­

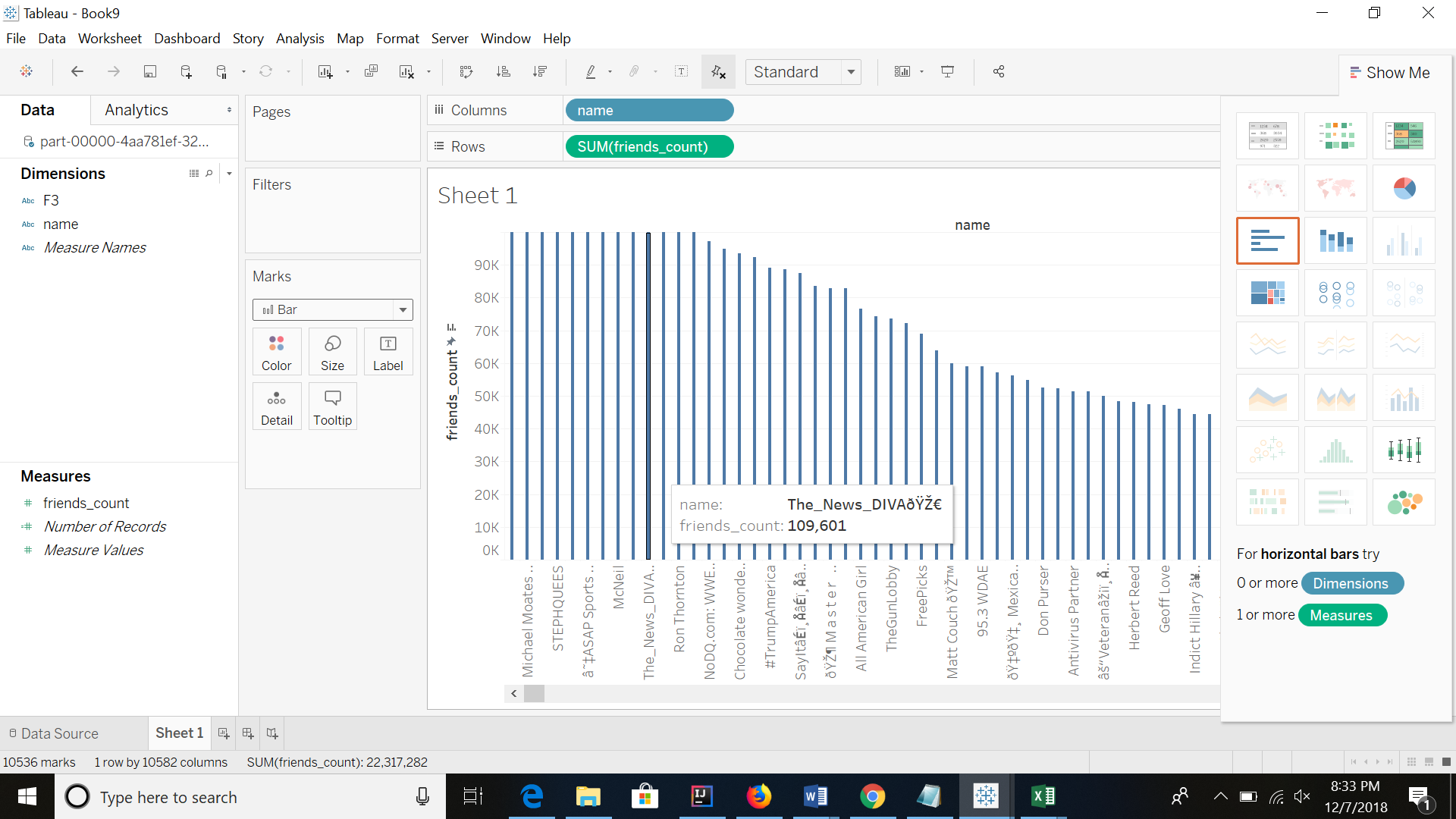
# Query 8:

In this Query to find out number of friends for twitter users that had friends greater than 100

Code for analysis of data in IntellijIDEA:



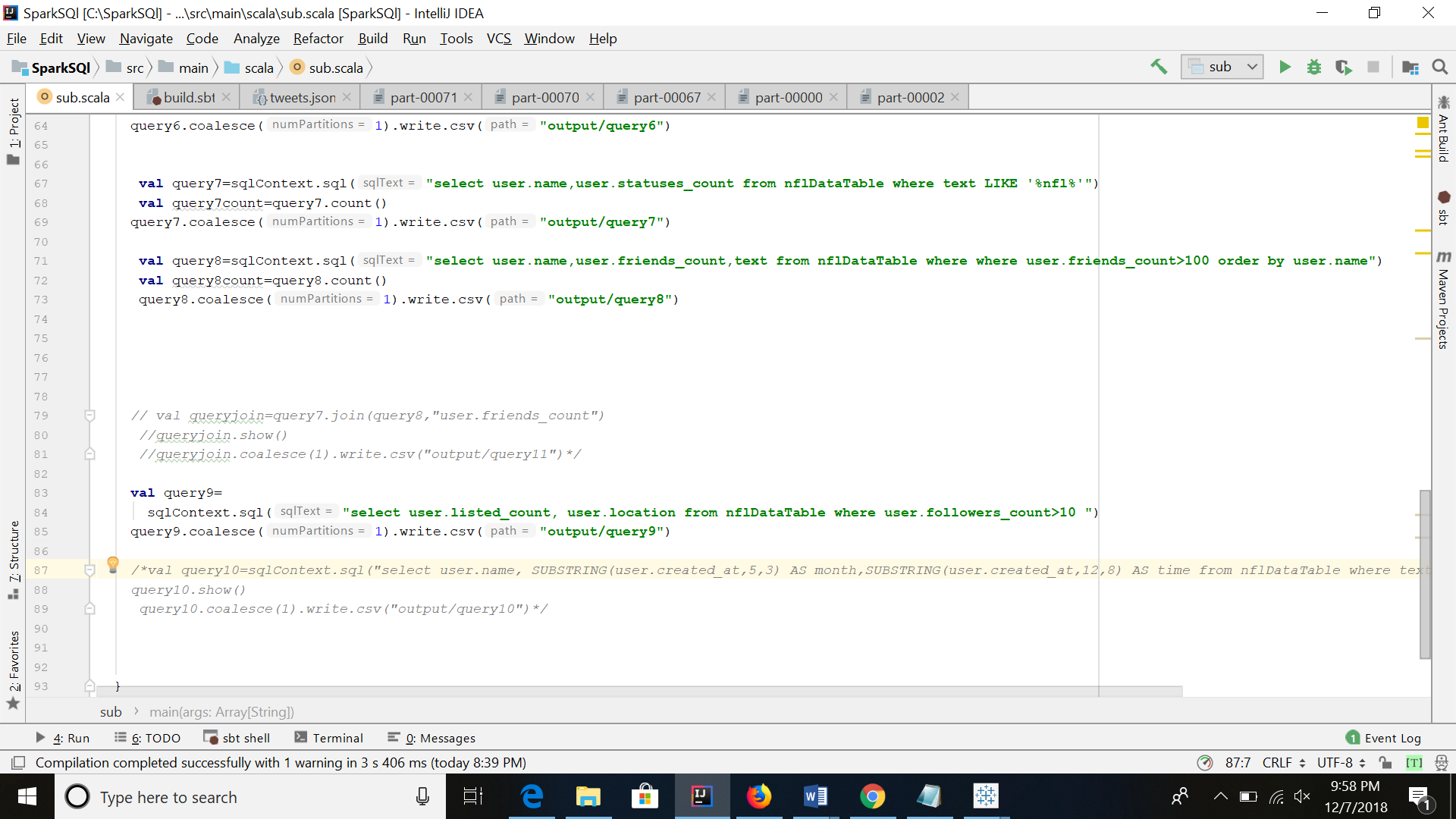
Visualization:



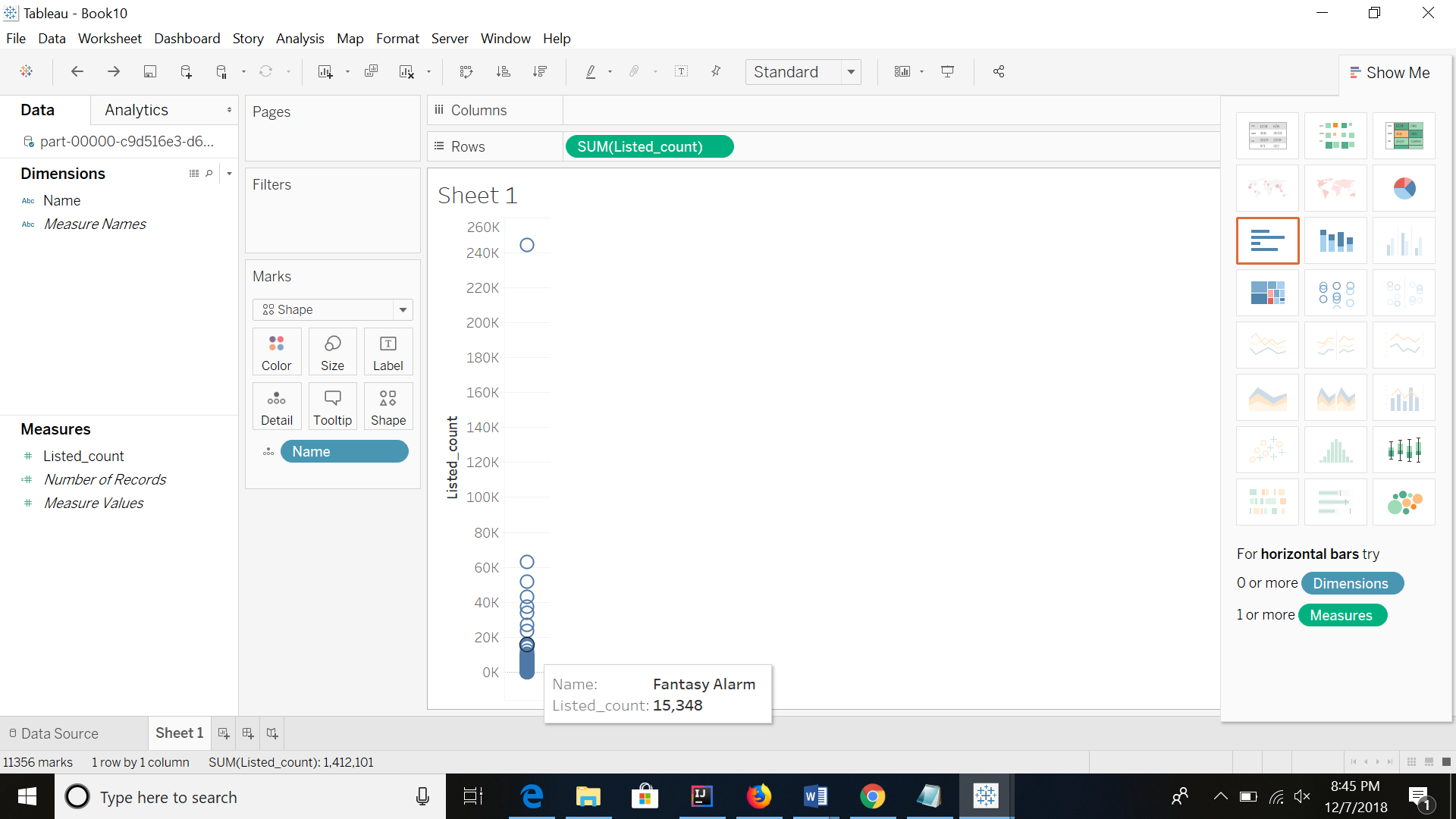
# Query 9:

In this Query we can see how the listed count increase with location

Code for analysis of data in IntellijIDEA:



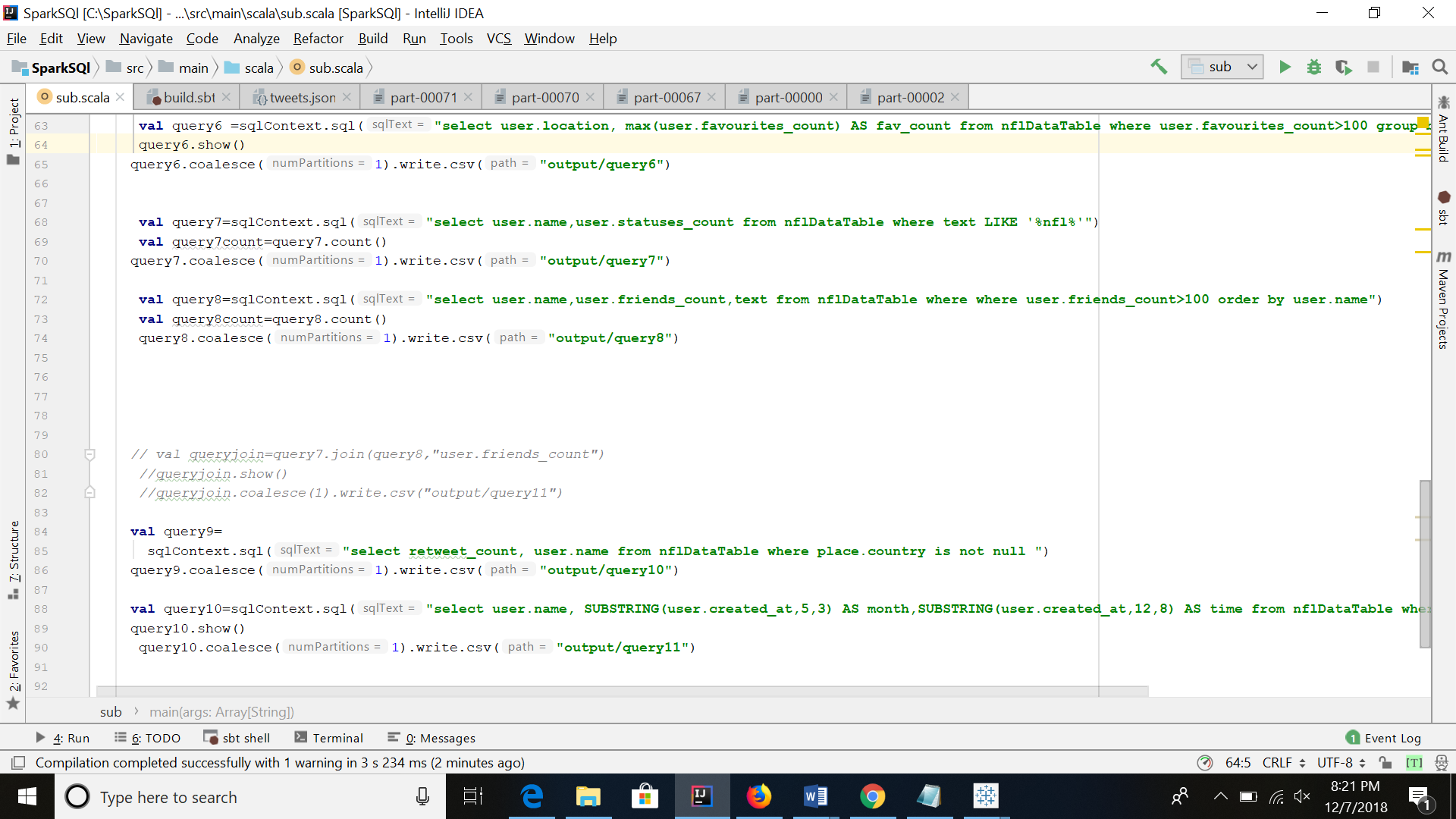
Visualization:



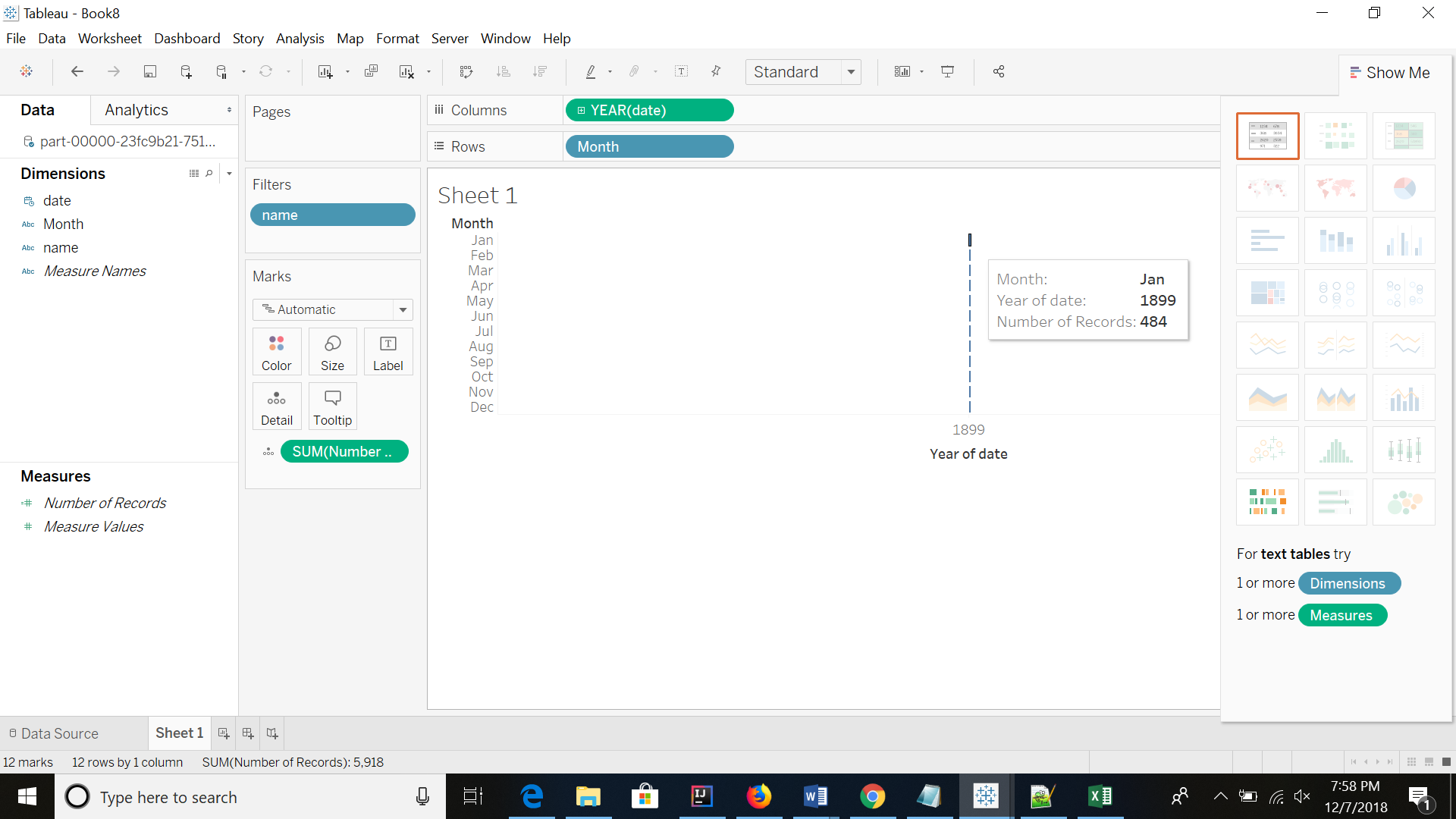
# Query 10:

In this query, we find the number of tweets analyzed on a monthly basis.

Code for analysis of data in IntellijIDEA:



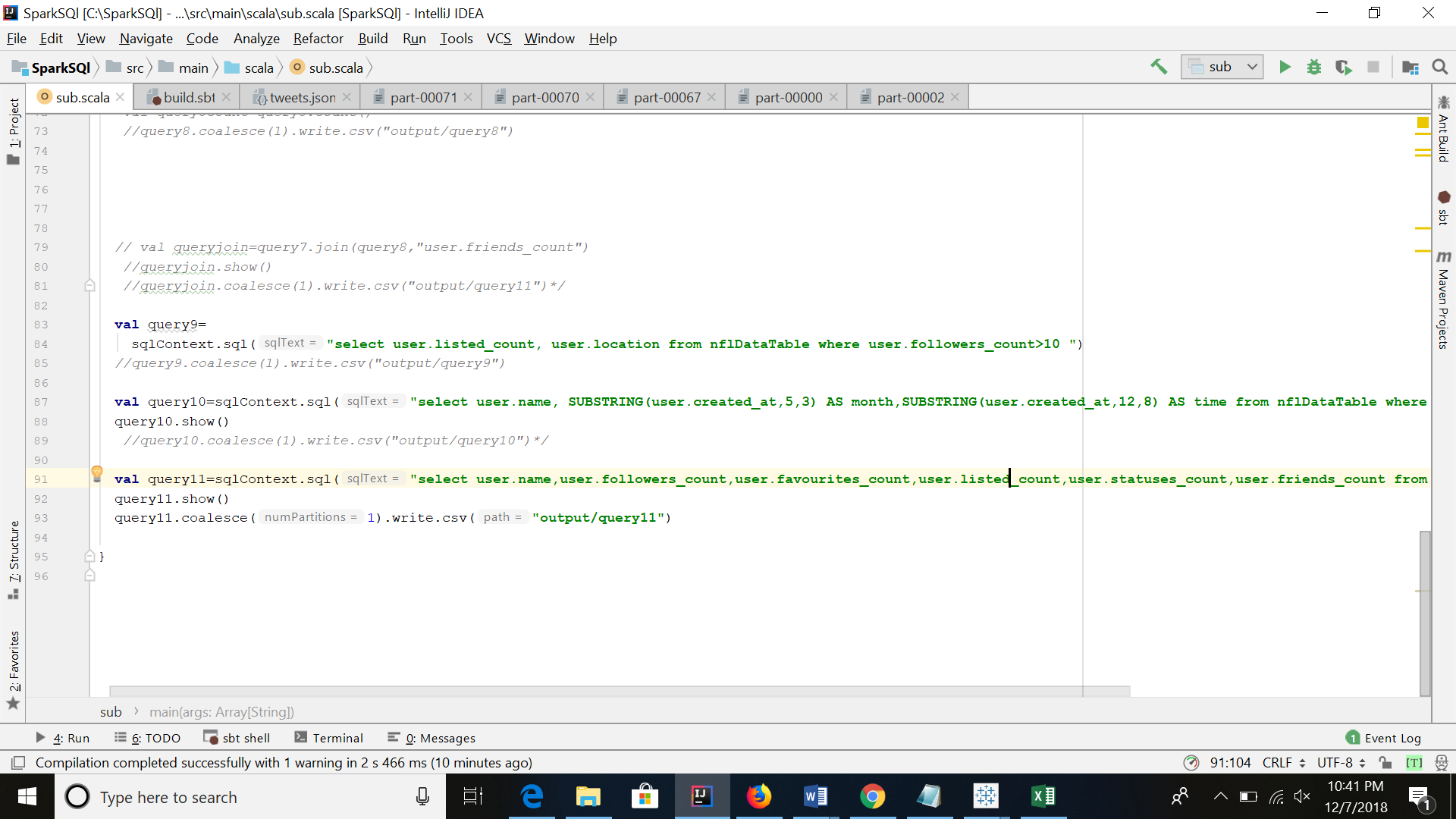
Visualization:



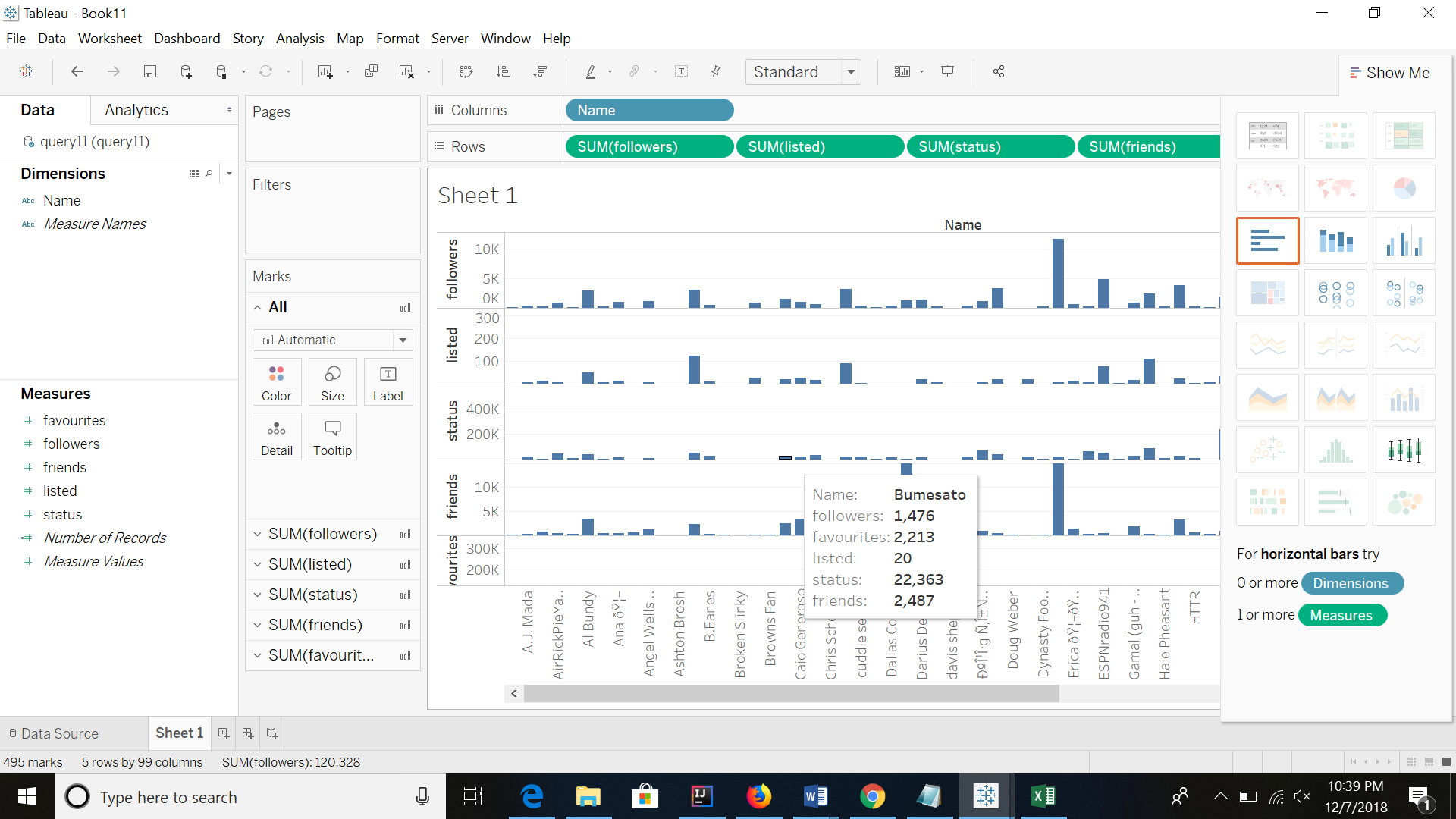
**Query 11**-

In this query, we find the number of friends, followers, favorites, listed, statuses count for all the user who are talking about NFL

Code for analysis of data in IntellijIDEA:



**Visualization-**



Link- <https://app.box.com/s/rvjfmheo8guuvnnhz7y7p7h0mm5s3g0z>