



**University of Puerto Rico, Mayagüez Campus
School of Engineering
Department of Electrical and Computer Engineering**

**Project #2 – Streaming for Twitter Analytics for
CIIC 8995: Big Data Analytics**

Presented by:
Cristian Garzón Alfonso

Presented for:
Manuel Rodríguez

Repo

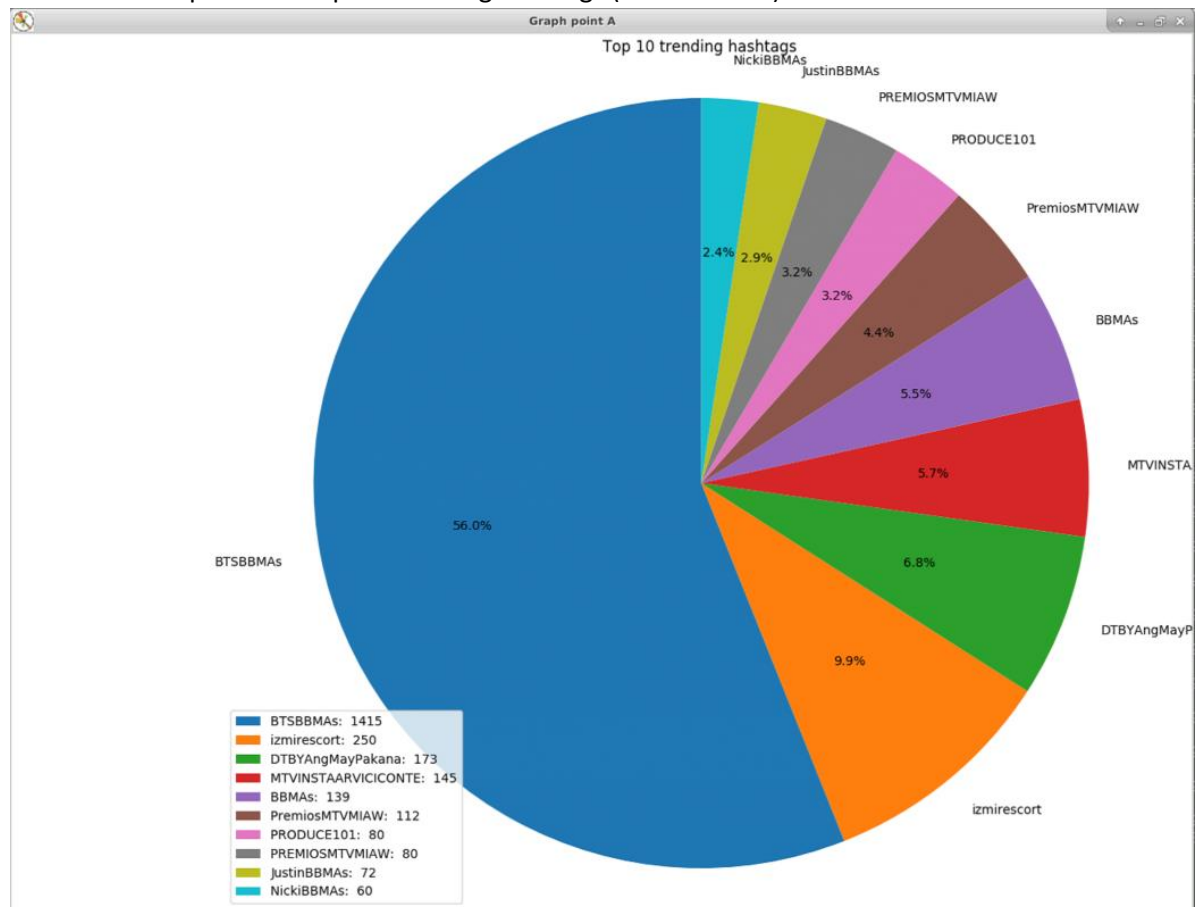
The link to get access to the repository in github is:

<https://github.com/CristianCGarzon/BigDataProject2.git>

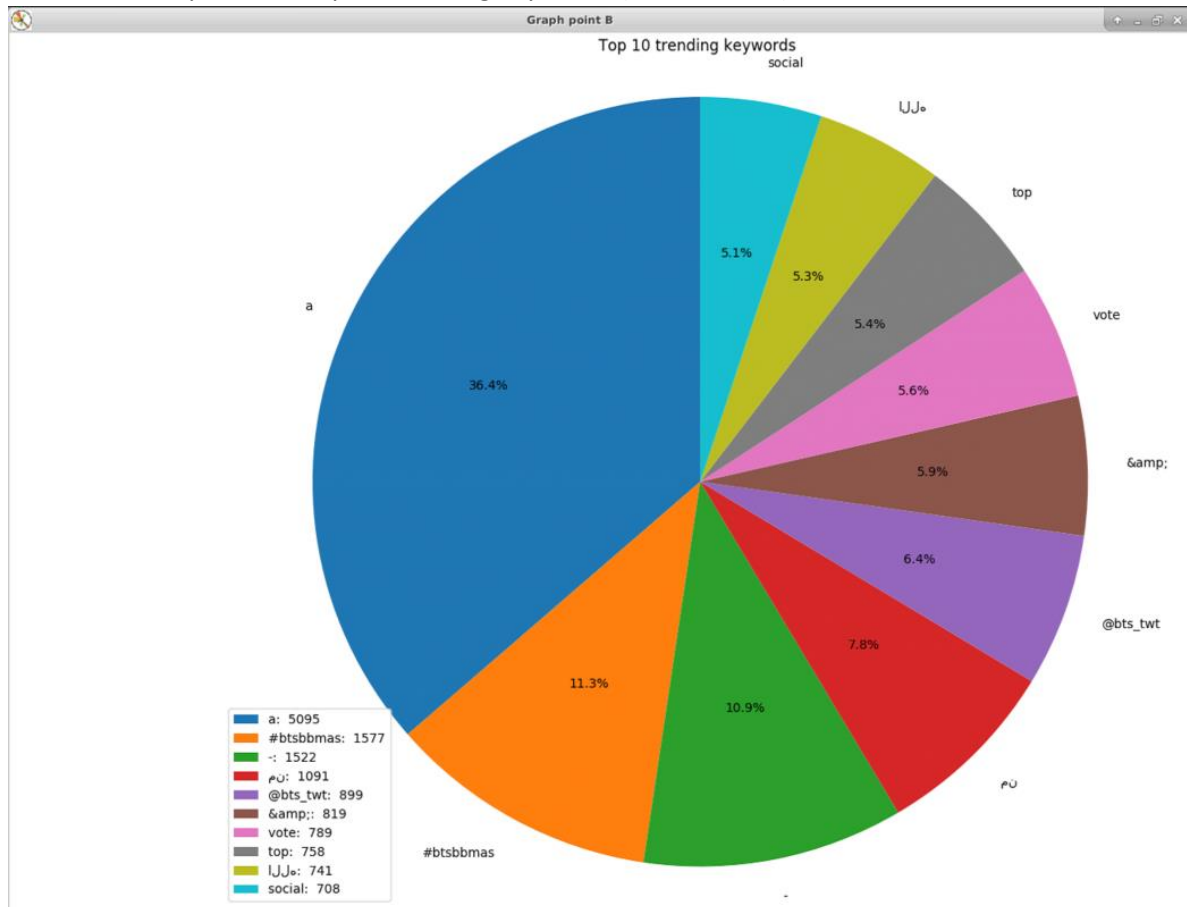
There is all the code to get tweets from Twitter API and put in kafka topic, once the tweets are in the topic we can start to consume and manipulate to save them in our tables in hive, then the information is ready to graph with matplotlib and pandas.

Graph's screenshots

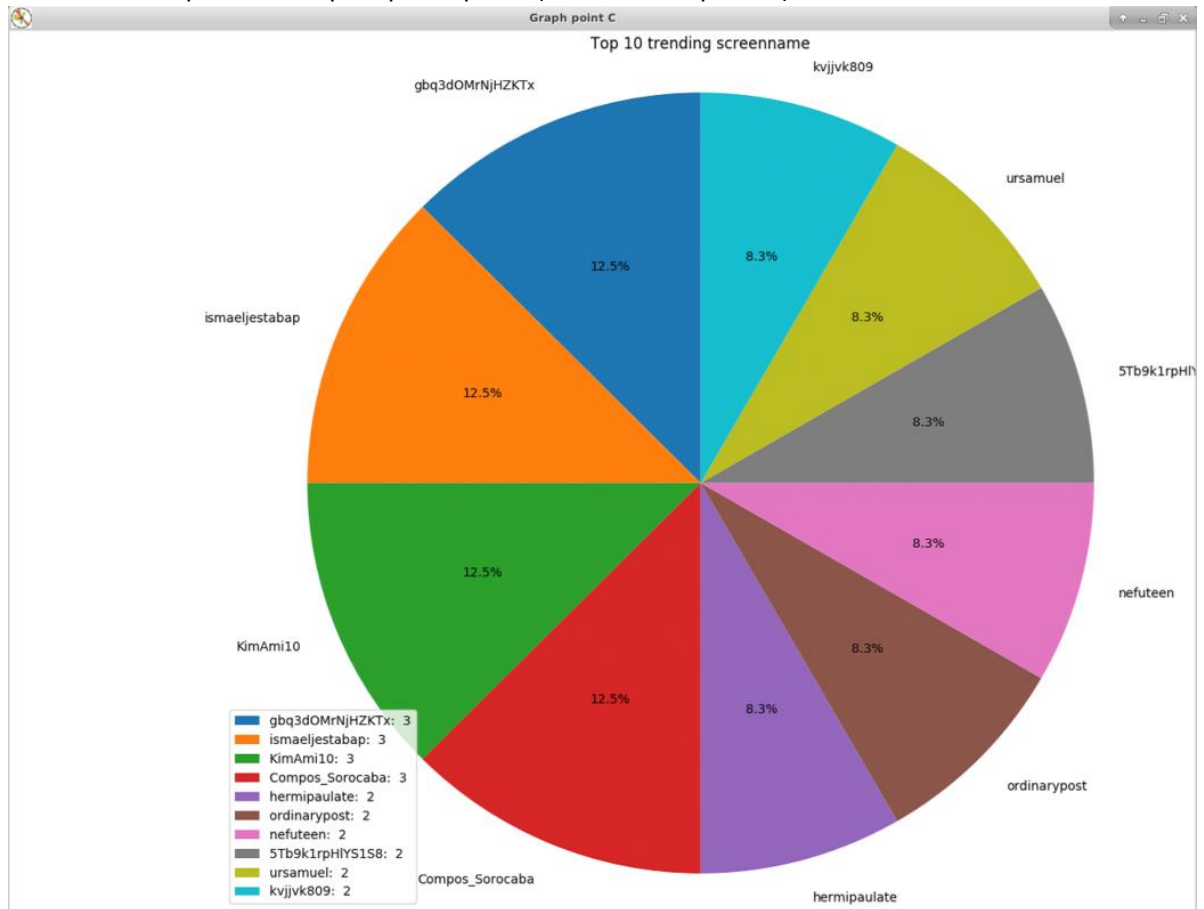
- Point 4 -> a. Capture the top 10 trending hashtags (most viewed) in the last 60 minutes.



- Point 4 -> b. Capture the top 10 trending keywords (most viewed) in the last 60 minutes



- Point 4 -> c. Capture the top 10 participants (most tweets posted) in the last 12 hours



- Point 5. Count the number of occurrences for these keyword, in intervals of 1 hours, on each day.

