



When we approach or start with any type of analysis, first job should be getting the basic counts and knowing your data. One can be a great data scientist if they know how to understand their data well.

From the above dashboard, it is clearly understood that we processed 79,623 tweets live. It also shows number of topics via hashtags (11,750) discussed in 2 hours of Friday night. It shows number of active users in USA (43,135). These tweets were coming in from 6,190 places. Here, places are the obtained from place field of these users. Dip in tweet timeline is because we stop our stream to reduce the threshold of incoming batch of data so as to use Spark Streaming efficiently. Our dataset mainly comprises of english tweets.

Now, there are two different unique counts for location: place and geo points. Twitter provides 3 different ways in which one can geolocate themselves. First is by using point location. Here we get exact latitude and longitude of the user. Second is bounding box. Here user give approximate area within which user is. It is a bounding box four co-ordinates. Finally there is place. Here, User gives his plac. This can be either area level, city level, state level, etc. We use both geo points and place for visualizations.