**Sprint 3- User Story #2- Geographical Distribution of Tweets**

**Introduction:**

This document explains our analysis of twitter APIs for location related information and the methodology followed by us to collect the information about the location from where tweet originated to understand the geographical distribution of tweets.

**Analysis of Twitter Stream and Search APIs for location related information:**

Twitter search and stream APIs return a series JSON objects for each tweet when called for a specific keyword or hashtag. The JSON objects contain multiple parameters out of which 3 parameters are useful for retrieving location information about the tweet. These parameters are-

1. **Place**- This parameter is associated with the tweet itself and provides info about the location from where the tweet came.
2. **Location**- This parameter is associated with the user who created the tweet and provides info about the location entered by the user while creating the twitter account.
3. **Time Zone**- This parameter is also associated with the user who created the tweet and provides the info about the geographical time zone to which the user belongs.

Of the 3 parameters listed above, Place and Location are null in most of the JSON responses while the parameter Time Zone is filled in a considerable number of responses. Therefore, we have chosen time zone parameter to retrieve info about the location of tweet. Also the time zone of the user can be assumed as the location from where the tweet originated for following reasons.

1. We are looking at geographical distribution of tweets across various regions on the earth and not across specific cities or countries i.e. we are looking at the distribution at much higher level. The time zones easily map to geographical regions.
2. Time zones cover a fairly large areas. Therefore, we can assume that in most of the cases the user of the tweet is located in the same time zone as the one with which he had registered on twitter i.e the tweet originated from the same time zone.
3. While there can be users who have migrated to different countries or regions and hence are no longer in the same time zone as the one with which they registered, the no of such users is negligible when we are looking at the tweets on global scale.

**Compiling a set of time zones:**

We have compiled a set of 168 time zones. These time zones are the ones that have appeared in tweet responses from search and stream APIs and the time zones available on android smartphones. For the time zones that have appeared on twitter so far, we used data from following source-

<https://github.com/StevenMaude/twitter-time-zones>.

We also computed the latitudes and longitudes for the locations represented by these time zones. For latitudes and longitudes , we used following website-

http://www.latlong.net/

The time zones along with their latitudes and longitudes are saved in a file LocationsNew.csv.

(Note: Latitude and longitude information is required to plot the locations on Open Street Map.)