Collection of Data from Twitter and Importing to R

Data Information

The Data was collected from twitter using the R ***twitteR*** library. The library provides an interface to the twitter web API.

The Data was collected for a week starting from 10th Feb 2016 and the tweets were stored in a *json* format on the local disk which could then be loaded back in the R environment for the analysis

Library methods

The methods from the twitteR library used were:

Setup_twitter_oauth: This method wraps the OAuth authentication functions for a TwitteR session, and must be provided with the user consumer key, secret key, access token and access secret provided for the application by the Twitter API.

Once the handshake was verified, *searchTwitter* method was used to specify the search parameters. In this implementation the search queries were related to rental and property related tweets in the New York area, and stored onto a local file in the disk. The tweets returned were in the format of a list of objects of twitteR class.

searchTwitter allows a user to issue a search of Twitter based on a supplied search string, in addition the search string user can also signify other parameters like *lang, since, geocode etc* to further filter the search scope. However, it was noticed that there is limit on the date of the tweets that are returned by the API which is independent of the *since* parameter specified. Hence the search space is limited to only a few days since the search is made.

twListToDF was used to convert the captured tweets into a single data frame, which could then be formatted as a *json* object and stored on the local disk.

From the *jsonlite* library , *toJSON* method was used to convert the objects into a single json format which was then written to the disk using the *write* method from the *base* library.

Importing *json* object into R

In order to load the object into R, *readlines* method was used. It reads all of the lines from the object and they were stored onto a data frame that could then be used to do the data analysis.