Project -1

Problem 1 – Collecting Tweets using R

CSE – 587

Instructor – Bina Ramammurthy

Submitted By:-

NALIN KUMAR

Person – 50170479

Collecting Tweets Using R and Importing from a file on disk

The problem specified to collect tweets on a specific subject in the first part followed by saving those tweets on the disk. In the next step, tweets collected in the first step were imported from disk and printed in JSON format. The data source used is Twitter API and the tweets were collected on the subject related to prices associated with renting and buying apartments in New York, specifically in the most expensive areas such as Manhattan and Brooklyn in valid JSON format. The data has been collected for 7 days beginning from 25th February 2016. While converting the tweets in valid JSON format, this involved using different packages such as jsonlite, RJSONIO, rjson and their respective methods such as toJSON() and fromJSON(). These methods were used primarily to convert to and from JSON data format. Initially the packages were installed using install.packages("package-name") command and once installed, subsequently, the installed packages were loaded into memory in RStudio using the library() command. Examples of other packages which were used in Problem1 include ROAuth, twitteR and RCurl. ROAuth package was used to validate the access to twitter API for the purpose of data collection. RCurl package was used to generate HTTP requests while twitteR package provides an interface to the twitter web API and make search requests into twitter based on certain parameters such as keywords, language, number of tweets etc.

The most challenging part in this problem was to convert data in list format into data in JSON format. The data returned by searchTwitter() method comes in list format which cannot be directly converted to JSON format in R. For this, initially we converted the list data into data frame R object using lapply and rbind utility functions. Subsequently, this data frame object was converted into JSON object using the toJSON() function of jsonlite library which takes an argument in data.frame format. Finally, the JSON object was saved into a file on the disk and in the next part we imported this file from the disk and printed the JSON object.

The following image displays sample tweets collected and printed in JSON format in RStudio