SUDHARCHITH SONTY

UBID: 50169912

MAIL: sudharch@buffalo.edu

CSE-487/587

DATA INTENSIVE COMPUTING

PROJECT 1 – PART 1 – REPORT

COMPLETENESS OF INFORMATION

- The data is collected in the JSON format and consists of the following fields

  - text
  - favorited
  - favoriteCount
  - created
  - truncated
  - id
  - statusSource
  - screenName
  - retweetCount
  - isRetweet
  - retweeted

PROCESS OF RETRIEVING TWEETS

Data Structures Used: Lists, arrays and Data Frames

Data Source: Twitter

Period of Time Data has been collected: 2016-02-27 to 2016-03-05

PACKAGES USED:

I have used the following packages to collect Data as a part of this project

twitteR: This package provides an interface to the Twitter API

Jsonlite: The jsonlite package provides fast and efficient methods to retrieve tweets in the JSON format by using the API. I have used the toJSON function (to convert the incoming data into JSON format) and the fromJSON function (to read the data that has been previously saved to disk in the JSON format)

Stringr: The stringr package is used to provide a set of wrappers for common string operations and with respect to this application I used it to create the filenames and append them with the current timestamp to store them onto the local disk

ROAuth: This package is used to establish and provide the Authentication for the application that is being developed.