**Recommending Restaurants to investors by Predicting User Ratings and Overall trend of Business growth**

**Abstract:**I am proposing a system for investors who want to invest in the restaurants. Investor will be provided with the data suggesting top 10 restaurants in US to invest in. We will be considering different factors for this recommendation system.

Objective :

* Based on Yelp Data Set of User Reviews for Restaurants, we will recommend top 10 restaurants for investment. Business value for this project are:
  + To provide recommendation on top Business in restaurant
  + To increase chances of profit for investor.
* This project involved parsing json data and converting it to csv file.
* Created csv files are then processed
* Building regression modell and prediction data as well as predicting the trend of business growth.
* Predict top 10 restaurants to invest.

Presetup:  
1) Download all the source code from the zip folder provided.

2) Please download and extract Yelp Dataset Challenge data, yelp\_dataset\_challenge\_academic\_dataset from the following link: [Yelp Data Set](https://www.yelp.com/dataset_challenge)

a) Extract the zip file download.

b) Unziped file will be again the zip file. Change the extension of unzipped file to .tar

c) Unzip this file aswell. You will see 5 json files in the unzipped folder

3) Name the data folder as: yelp\_dataset\_challenge\_round9. This folder should directly have all your Jason files

4) Put this data folder in the same directory as that of the r code.

5) run file “InstallPackages.R”. This will install all the required packages:  
 install.packages("gdata")

install.packages("ggplot2")

install.packages("streamR")

install.packages("jasonlite")

install.packages("readr")

install.packages("dplyr")

install.packages("caret")

install.packages("qdap")

install.packages("quantmod")

6) Use “convertJsonToCsv.py” to convert all json files to csv files.

Run: python convertJsonToCsv.py filepath/JsonFileName.json  
for each json file in database.  
  
7) Now we have csv file ready to use in our folder yelp\_dataset\_challenge\_round9

8) Now run “preprocess.R”

This will do all preprocessing work on the files and data analysis of raw data. Initial graphs will be generated

9) Now run data integration : “DataIntegration.R”

This will carry out integration work

10) Now run “Analysis.R”

This will carry out all analysis wrk and output with top 10 restaurants will be displayed.

**Reference:**

* <https://en.wikipedia.org/wiki/Linear_regression>
* https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&cad=rja&uact=8&ved=0ahUKEwiynIC8x8rTAhXBOyYKHe7VAM0QFgg7MAM&url=http%3A%2F%2Fcs229.stanford.edu%2Fproj2014%2FYun%2520Xu%2C%2520Xinhui%2520Wu%2C%2520Qinxia%2520Wang%2C%2520Sentiment%2520Analysis%2520of%2520Yelp%27s%2520Ratings%2520Based%2520on%2520Text%2520Reviews.pdf&usg=AFQjCNHCJ83fnobr9CpBPBtziGuFRDAVSw
* <http://minimaxir.com/2014/06/reviewing-reviews/>
* <https://en.wikipedia.org/wiki/Generalized_linear_model>
* <https://www.slideshare.net/ohassta/critical-thinking>
* <http://leonidzhukov.net/hse/2011/seminar/papers/chi09-tie-gilbert.pdf>
* <http://www.utdallas.edu/~ryoung/phdseminar/DataMiningComparison-Melody.pdf>
* http://www.criticalthinking.org/pages/defining-critical-thinking/766