Aditya Shashank Chavan – 3416409224 – aschavan@usc.edu

INF 553: YelpProject_UPDATE 1

Accomplishments

- Setup environment, downloaded datasets, performed basic statistical analysis on given data
- the yelp_dataset consists of 6 json files, out of which we decided to use or start working with business.jsona and checkin.json
- there are 188593 business out of which for our project we needed only restaurants.
- Since we are predicting the most profitable area for a restaurant, so focused on the location and category attributes of the data.
- The main Idea was to split the data into training and testing sets. As not all restaurants are evenly divided based on location, we had to investigate the 'neighborhood' field to get a vague idea of the splitting criteria.

Plans

- Study the relationship between location and business categories
- Develop those train/test splits and identify which machine learning algorithms can be used.
- Filter data to get required results and remove redundant data.

Challenges

- Figure out which attributes of a business contribute to its success in a specific location.