**Project Title**: Rating Trends of Various Business Types

**Team Members**: Tracy Oguni

**Objective and overview of the project**

The objective of the project is to survey the different kinds of businesses in the Yelp Data Set and see which are more likely to get positive or negative reviews. For example, do people generally rate Dentists lower than restaurants? If so, why?

**Significance**

Any business that is concerned with its image (i.e. all of them) should be concerned with its reviews. It is therefore important to know how seriously to take a review. Part of that requires a reader to know just how important a negative (or positive) review is and comparing a business’ average review to that of others in its class is one way to do this.

**Data Mining Tasks**

The following data mining tasks will be performed:

1. Extracting the information about Businesses form the dataset.
2. Grouping the businesses based on business types.
3. Sorting and evaluating the different rating types for each business type.
4. Producing numerical measures of how each business type is being rated.
5. Predicting the type of a business from its ratings.

**Evaluating the Efficacy of the Solution**

A small portion of the Yelp Dataset will be left out of the initial calculations and later used to judge the accuracy of the predictions. This dataset (i.e. the test set) will consist of businesses whose types will be predicted based on the type, amount, and distribution of their reviews.

**Project Deliverable**

Deliverables at the end of the semester will include but not be limited to:

1. A chart showing the average rating for each business type.
2. A chart showing the rating frequency distribution for each business type
3. A figure showing how accurately prediction of a business’ type based on its rating was.

Those deliverables will be placed on the website. They will also be included in the Demo/Presentation.

**Tasks and Co-ordination**

The sole team member will perform all the tasks.

**Challenges**

**Challenges in this project**

The main challenge to be expected in this project is the question of just how to accurately predict the type of a business based on its ratings. The solution to the problem will lie in how much the different kind of businesses differ in their reviews. It is possible that predicting a business’ type from its rating will be nearly impossible.

**Addressing the Challenges**

One way to address the challenge is the collect as much different kinds of information about the ratings as possible. This will make it more likely that a rating pattern will be detected. Information to be collected will include average rating, average numbers of ratings, and frequency of rating types. Information on how many users leave ratings might also be useful.