Regression Analysis Abstract
10/9/2021

1. Team Members

- Colin Diggs
- Mingyuan Zhang
- Shivtej Shete
- Zachary Pete

2. Title

Can Yelp reviews help predict a restaurant's success? An analysis of predicting characteristics within reviewer data on Yelp

3. Brief description of the project's objectives

The goal of this project is to investigate the relationship of customer reviews on the website Yelp of restaurants across America. Specifically, we would like to predict the expected rating in stars of a restaurant with a certain set of characteristics such as location and food genre. Using a Yelp dataset download, we will clean and create a workable dataset of food reviews at different restaurants. Then using the tools and methods taught in class, we will investigate the relationship between food reviews and predicting characteristics through a regression analysis and potential advanced methods. Possible areas for exploration are weighting customers who review more, accounting for chain/non-chain restaurants, location choices, and the number of total reviews.

4. Data Source

Dataset (4.88 GB compressed) - https://www.yelp.com/dataset/download

Documentation - https://www.yelp.com/dataset/documentation/main

5. Team Tasks

- Colin Diggs - Outline report and create list of objectives and analysis of the dataset
- Mingyuan Zhang - Work on cleaning the data and beginnings of a regression analysis
- Shivtej Shete - Work on cleaning the data and initial data exploration and visualization
- Zachary Pate - Download and begin cleaning the dataset, feasibility of moving the dataset to AWS for us all to work on

All team members are expected to contribute to the work and analysis, access and use the data, and contribute to the final report.