What makes a Good Business?

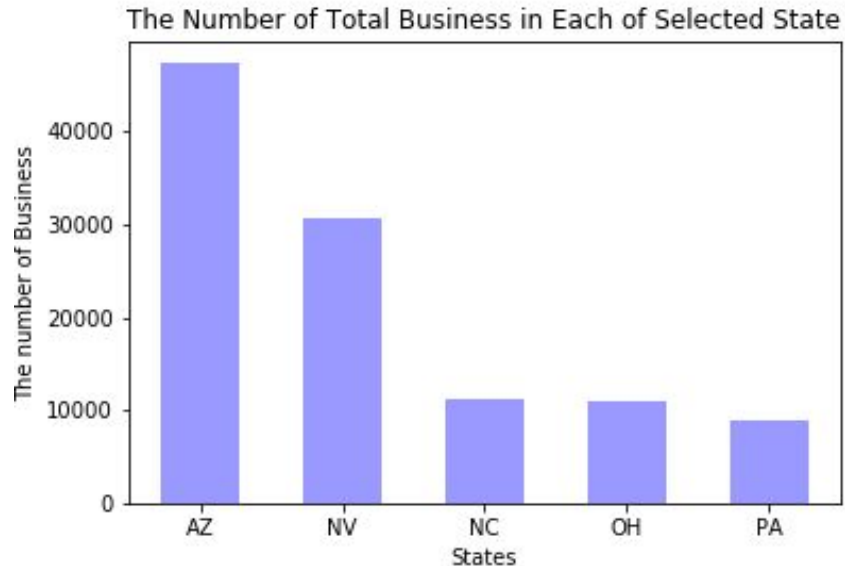CS 287 Final Project
Chen Ling
12/13/2017

Intro

 The dataset I am using is from Yelp.com, which is a subset of business reviews, business information and user's information.

In this project, I intended to find commonplaces between high-rated business, including the type of business, location and the quality of reviews. Besides, I intended to analysis the star rating system.
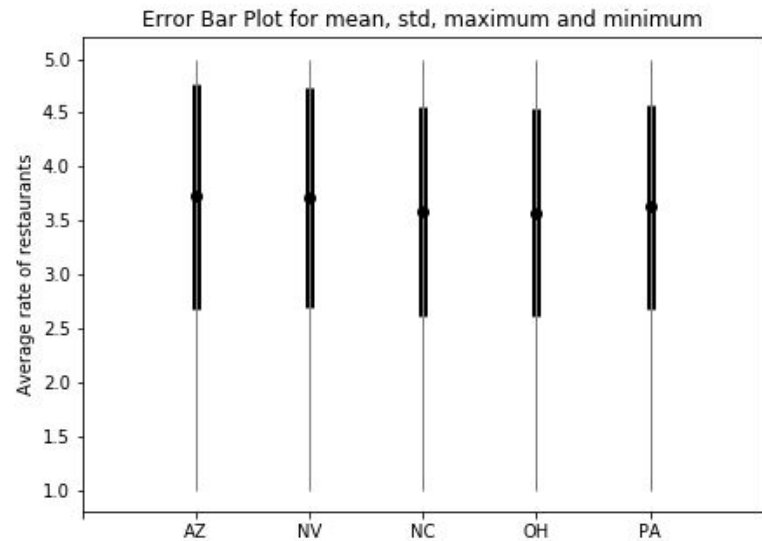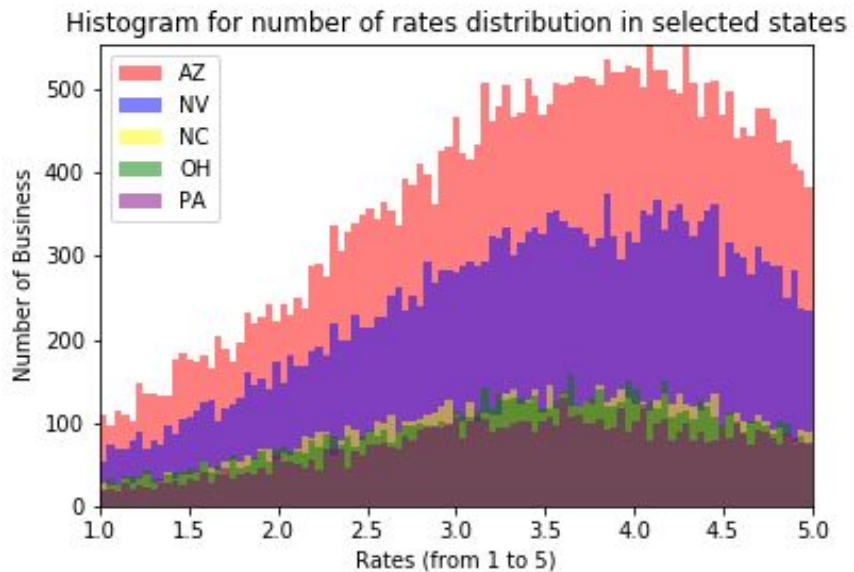
Last but not least, based on the Text Sentiment API, I want to analysis the subjectivity and polarity of reviews.

I chose top 5 states which has the most number of registered business ID, and they are AZ, NV, NC, OH, and PA. The following graph is the number of business from Yelp in each of five states.
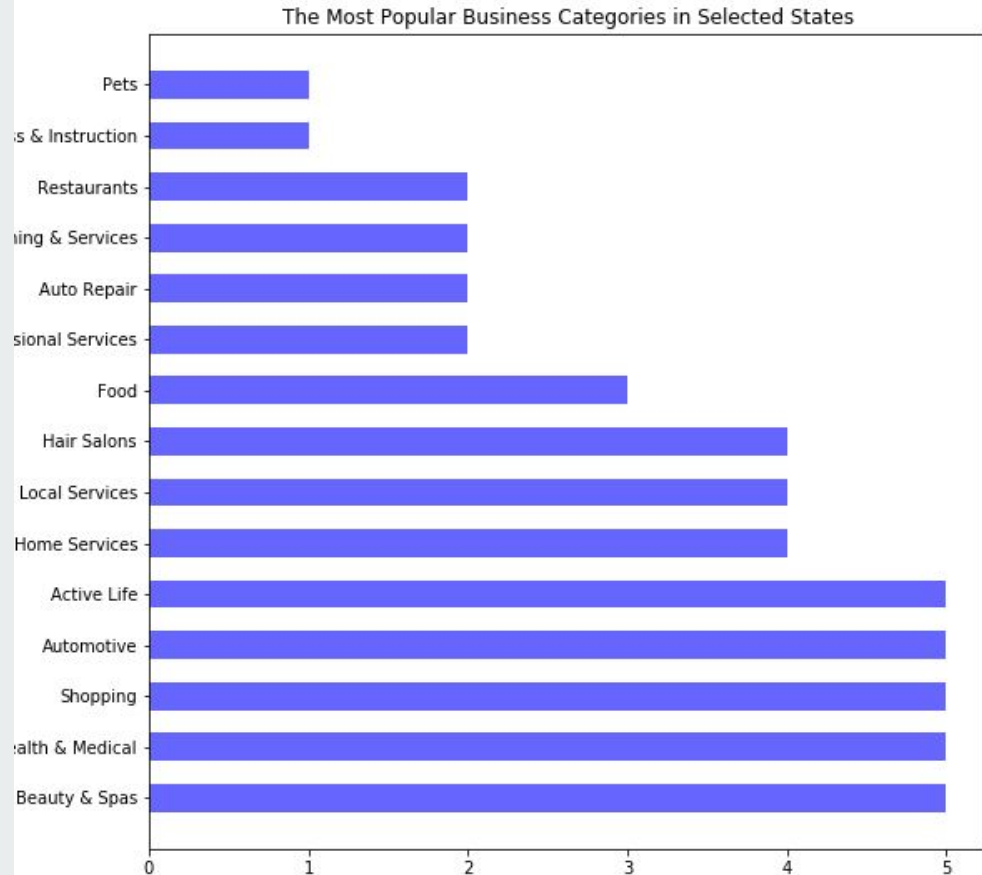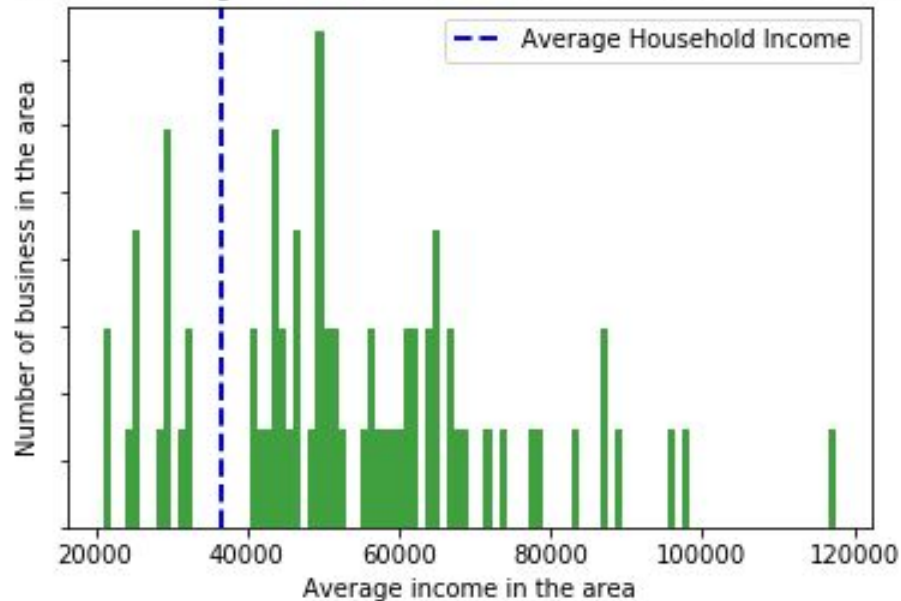

The Number of Total Business in Each of Selected State

Graphs

Histogram for number of rates distribution in selected states

Error Bar Plot for mean, std, maximum and minimum

The most popular Business Categories in the selected states

Active Life, Automotive, Shopping, Health & Medical, and Beauty & Spa

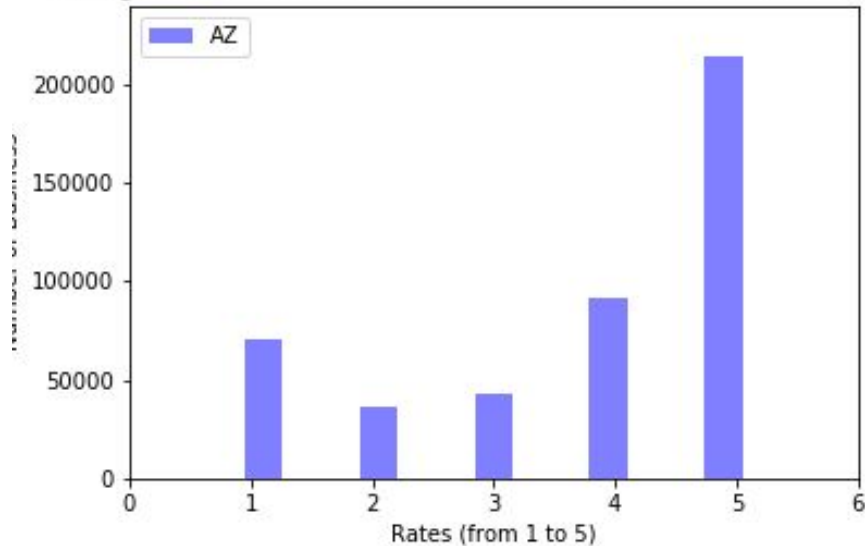The Most Popular Business Categories in Selected States

Location of Business and Area Average Income

Location of High-rated Business VS. Household Annual Income
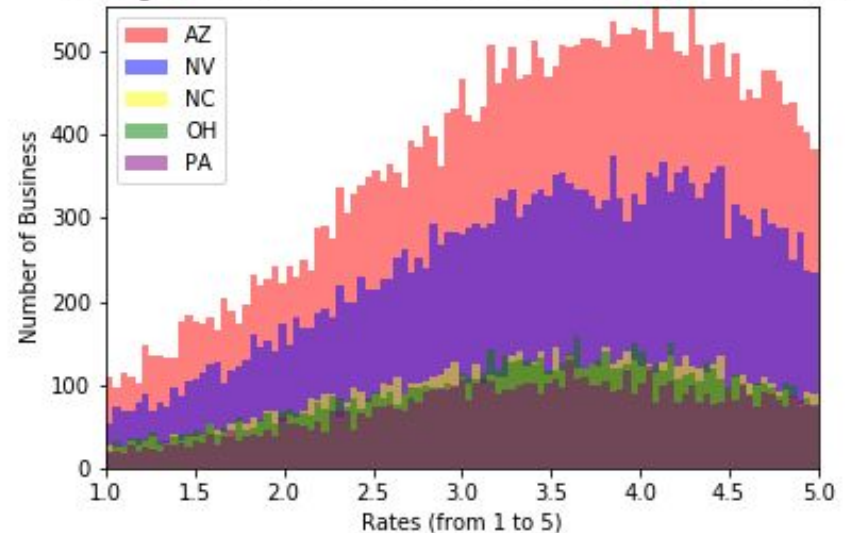
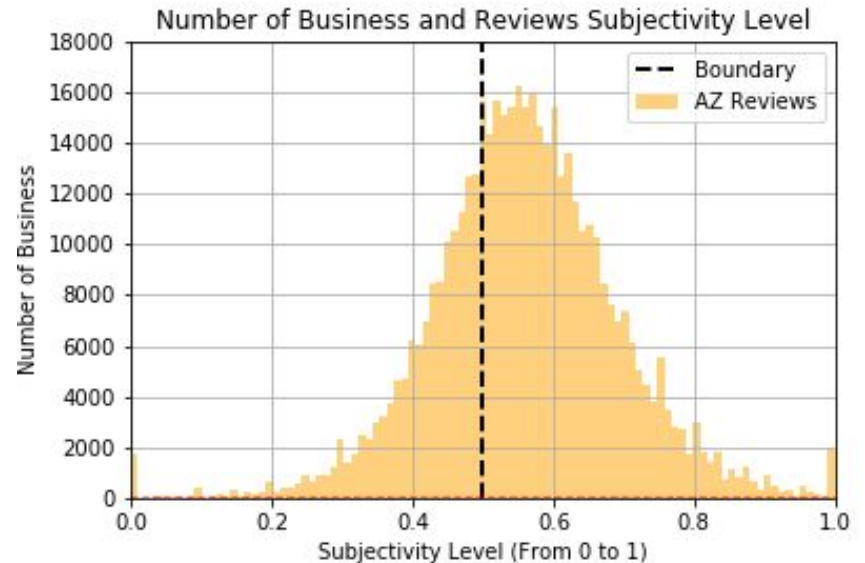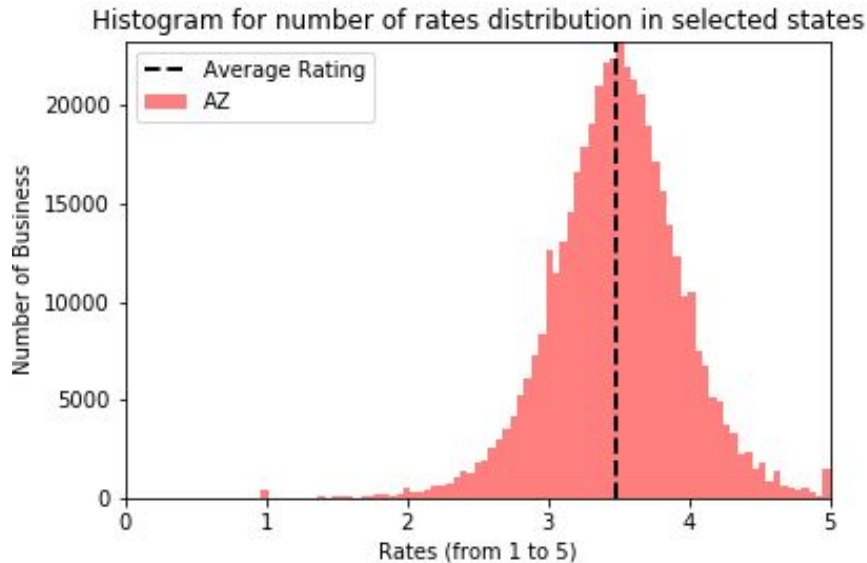# Comparison of Aggregated rating and actual rating

Histogram for number of actual stars distribution in selected states

Histogram for number of rates distribution in selected states

Text Sentiment Analysis Result

Result and Conclusion

1. high-rated businesses are more likely to exist in the relatively wealthy area; and high-rated businesses category reflects that people nowadays care more about personal care and living quality.

2. Star rating system has been considered harmful, which is not accurate and may contain personal subjectivity. A simple thumb up and thumb down would solve lots of problems