Stat222 Revised Research Proposal

Hui-Fang Ko, Hsin-Wei Tsao

Feb 19, 2014

1 Introduction

As consumers, we really want to have good meal experiences which are highly correlated to the restaurants. So we try to make a more solid model to evaluate restaurants not only based on the meal and service they provide, but also taking hygiene issues into account.

2 Resources (Dataset)

2.1 Restaurant Scores in San Francisco

This is a dataset provided by the San Francisco of Department of Public Health at `https://data.sfgov.org/Public-Health/Restaurant-Scores/stya-26eb`. The Health Department has developed an inspection report and scoring system. After conducting an inspection of the facility, the Health Inspector calculates a score based on the violations observed. So this dataset contains information of 6073 restaurants in San Francisco area and their score in the inspection conducted by the department.

2.2 Yelp API

With Yelp API, we can get the consumers' reviews of those restaurants listed in the dataset above. Moreover, we can get some details about the restaurants including price, utilities(ex: Wi-Fi), etc..

2.3 Others

More information about the neighborhood (by zip code) including average housing price, criminal rate, etc..

3 Overall Research Questions

1. Is there any relationship(positive or negative) between the restaurant scores by Health Department and consumers' preferences on Yelp? Also, do they have the same distribution?

2. How does location influence on restaurants' score?

3. How do other criteria effect on restaurant evaluation?

4 Approaches

To answer our research questions, we are going to build a model to evaluate restaurants. There are few methods we wll use:

1. Linear Regression: We set the customers review score on Yelp as the response variable, and explanatory variables are restaurant score by Public Health Department, area crimal rate, housing price, numbers of reviews, texts in reviews, wifi, price rate, hours, take credit card or not.

2. Clustering: Use the same variables as above.

3. RFM Model: apply the RFM Model to analysis customers' reviews on yelp. The elements we will use are:

   - Recency - How recently did the customer post review?
   - Frequency - How often do they post?
   - Monetary Value - How many texts in reviews?

5 Anticipated Results

1. Those two scores might have positive relationship, but might not have same distribution.

2. We believe that locations have good influence on restaurants. So maybe we could build a food map recommending nice restaurants with both high ratings on Yelp and nice, clean environment.

6 Anticipated Figures & Tables

1. The scope of relationship of 2 different scores.

2. A table shows the weights that we used to evaluate a restaurants and the scores we give.

3. The map of recommended area which has higher chance to get better restaurants. (By our scores.)

4. A comparison with our scores and customers'.

7 Reference

1. 10 things to Know About Choosing a Restaurant Location http://restaurants.about.com/od/location/a/10-Things-To-Know-About-Choosing-A-Restaurant-Location.htm

2. Consumers Feast on Restaurant Ratings http://www.gsb.stanford.edu/news/research/stratman_consumerinfo.shtml

3. 8 Marketing Technologies that Affect Customer Restaurant Choices http://www.foodservicewarehouse.com/education/restaurant-marketing/8-marketing-technologies-that-affect-customer-restaurant-choices/c28057.aspx