

STATS 140SL Final Report

Yelp Data Analysis

December 11 2017
Fall 2017

Xiwen Jiang, Yu Ying, Zhe Zhang, Junhyuk Jang, Patrick Ran

STATS 140SL Final Report

Yelp Data Analysis

Introduction

In the project, we assume that a client who like to invest a large amount of money into a business and ask us to find the best choice on it. Potential, profitable, and sustainable are three main factors to consider to the investment. Hence, we connect with the Yelp data and some external source data. We start to consult this case in three directions:

1. The amount of customers;
2. The consumption leve;
3. The environment safety

Visualization & Analysis

Count of Business

Based on the graph of data output, the most popular types of business are restaurants, shopping, food, beauty & spas, health & medical, home service, and nightlife. However, we need to choose one as the final type of bushiness, so we do more detail analysis based on the performance included count and stars.

Comparing with the performances of six types of business, we would like to choose the shopping as our target business types. First reason is that the shopping has the most average stars and large amount of counts, which means that there are a lot of potential customers and high level of consumption. Secondly, although the count of restaurant and food is higher than the count of shopping, the average star is much lower than the average star of shopping, so we decide to the choose the business type which has the less competitive environment. Less environment is very important factor for a new company to consider.

ANOVA Output & Analysis:

Ho: There is no relationship between the average star and the count of the business.

Ha: There is a significant relationship between the average star and the count of the business.

The anova table shows that the p value is 0.303, which is greater than 0.05. Thus, we accept the null hypothesis and conclude that there is no significant relationship between the average star and the count of business in 95% confident level.

Analysis the AVG Stars per State

- **★Average star for each type of business:**

Graph of AVG Stars per states demonstrates that Alabama (AL) and New Hampshire (NH) have the highest and perfect star rate, 5.0, but the review counts are not enough. Inversely, the Nevada and Arizona have the most largest amount of review counts, but their star rate are kind of low or very low, only around 3.7. We concludes the reasons that less review counts may cause higher star rates, but the most review counts may cause lower star rates.

Determination on the location of the business

- **★GDP with external data:**

The largest GDP states include California, Texas, Florida, and New York. These states are near by the coast, we would like to think that the residents lived in these states could be become the potential customers, since they have more economics environment. Therefore, the residents live in the state near by coasts have more interesting and willing to eat outside or at restaurants, so if the investors want to open the restaurants, we suggest opening in the coastal states.

- **★Population Density**

- **★ The density of the population in each state:**

The graph of population density illustrates the same information with the previous slide, GDP for each state. The highest population density states still include California, Texas, Florida, and New York. We may need to reconsider the previous conclusion because the larger populations usually produce higher GDP. Hence, the higher GDP maybe neither be equal to the real

economics environment and nor represent the income level. On the other hand, it is not completely bad news. The good signal of larger population proves the more potential capacity of marketing, which is also beneficial for a new restaurant.

- ★ **Criminal Rate**
- ★ Criminal Rate with External Data:

The states, which have the highest criminal rate, include New Mexico and Washington, around 4% criminal rate. Criminal is the one of important factors to be consider because its affect the restaurants' reputation fame. Also, the customers really care about the safety condition. If the locations of restaurants have better to choose some states whose criminal rates are less than 3%. It guarantees the future development of new restaurants.

Rank for each States:

Count of the number of Business:

Shopping (Most popular, High Potential Les Competitive)

Average Stars:

AL, NH, CA (Better Commercial Atmosphere)

GDP:

CA, TX, FL, NY (High Consumption Level)

Population Density:

CA, TX, FL, NY (More Potential Customers)

Criminal Rate:

NO, NM, WA (Safer Environment)

Conclusion

- **Type of the Business:**

Compared to other four industries, we believe **Shopping**, with relatively more reviews, has the higher potential on profitable ability. Even though it has less review than restaurant, it should be less competitive.

- **Location of the Business:**

Based on average stars, GDP, population density and criminal rate, we believe **California** and **New York** are our best choice, since they both have a well--conditioned business environment and more potential customers with high consumption level.