

# STAT 628 Module 3 Project

Mengkun Chen<sup>1</sup>, Haoran Lu<sup>1</sup>, Rui Huang<sup>1</sup>

<sup>1</sup>Department of Statistics – University of Wisconsin-Madison

mchen373@wisc.edu, @wisc.edu, rhuang95@wisc.edu

## 1. Introduction

Affected by the covid-19 epidemic, the turnover of steakhouses that originally served dine-in food as their mainstay fell sharply compared with previous years. In this special period, how to change the steak house's business strategy and restaurant structure to adapt to the new needs of customers has become a top priority. In this project, Our goal is analysis the yelp data for steakhouses and give the business owners data-driven recommendations.

## 2. Data Cleaning

The yelp data of steakhouses business in OH, PA, WI, and IL were mainly analyzed, and we mainly used 'attributes' and 'tip' features. There are 353 steakhouse business records and a total of 35 attributes. Some steakhouse businesses miss the values for particular attributes. For example, 291 records miss the responds to whether dogs are allowed in the restaurants. Because of the large missing proportion, we drop those attributes whose missing rate exceed 10% to keep enough records for analysis.

We calculate the number of each category for the attribute. We find that almost all the steakhouse business are good for groups and support payment by credit cards. Thus, we drop these attributes. Ambience and BusinessParking have serveral sub-attributes. We further define these sub-attributes as True/False variables. "None" in Ambience is considered as missing value and "False" for BusinessParking. Alcohol has three categories-none, beer\_and\_wine and full\_bar. We map Alcohol to 1 if the business only provides beer and wine, to 2 if it is a full bar and 0 otherwise. For BusinessAttire, casual is mapped to 0 and dressy to 1. The remaining True/False attributes are re-coded as 1/0.

We tried to impute the incomplete records using the steakhouse business nearby, but the result is not satisfying. In order to make things simple, we directly drop all the incomplete cases. Thus, The pre-processed dataset has 305 records and 24 variables, showing the name, ID, geographical information and extracted attributes of each record.

## 3. Exploratory Data Analysis (EDA)

### 3.1. Tips Analysis

Our goal is to analyze word frequency from tips, so as to consider which factors and aspects customers pay most attention to. We combined all texts from *tip.json* and removed all punctuation marks and numbers. Then we removed all stop words and meaningless words, such as 'is', 'a' or 'I'. After counting the word frequency, we found that the remaining words still have a large number of unspecific adjectives and verbs, such as 'good', 'great' or 'love'. After above processing, we found that the left words are mostly some nouns that related to a steakhouse and they could be mainly separated to two groups, food- or service-related. Note that some words related to opening time, such as 'brunch', are also classified into the service-related group.



**Figure 1. World Cloud**

With this two groups, we computed their word frequencies and are able to check what customers cares. In order to clearly show the results, we draw word clouds for food- and service-related key words (Figure 1). We will further discuss the findings from tips in the next section.

### 3.2. Attributes Analysis

Star ratings distribution and frequency of each attribute under different categories are shown in the plots.

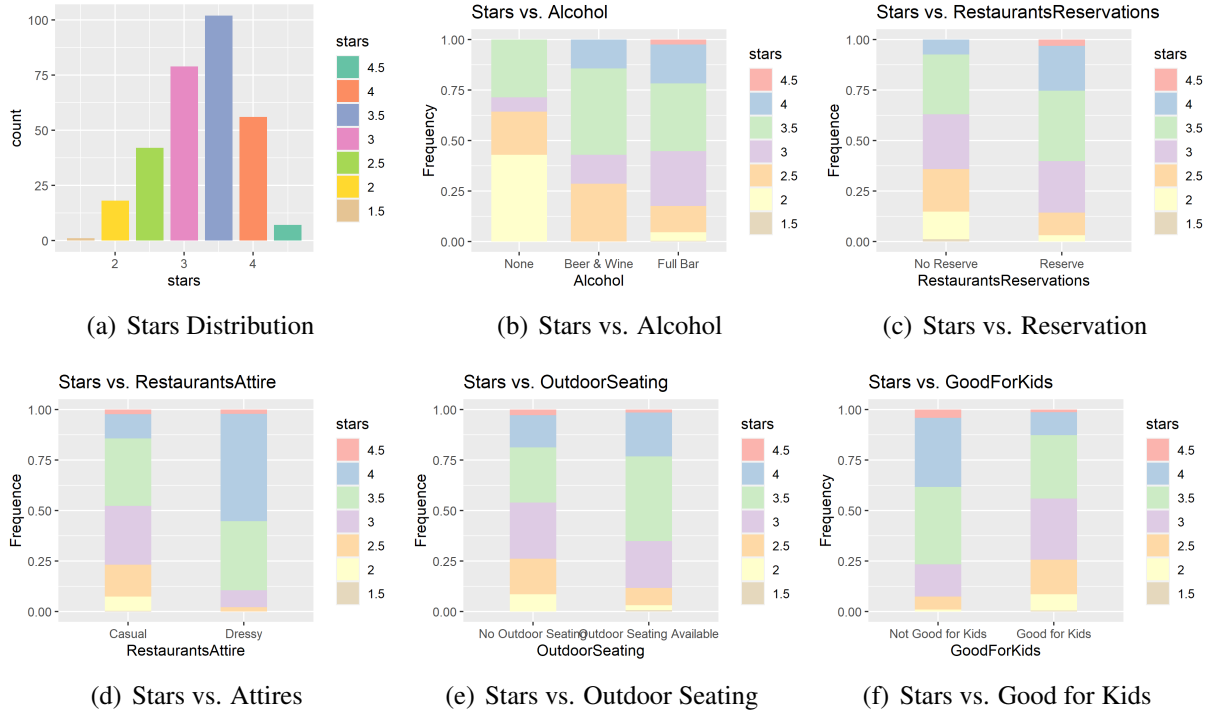
Figure 2(a) shows that most steakhouse businesses have star ratings between 2.5 and 4. Figure 2(b)-(f) indicate frequency of each star ratings are correlated with these attributes except for OutdoorSeating.

The plots provide us with an overview of the attributes' contributions to star ratings. After doing ANOVA, we find Alcohol, GoodForKids, RestaurantsAttire, RestaurantsDelivery, RestaurantsReservation and two types of Ambience-casual and classy-are the most influential attributes. We originally speculated that price, parking service were quite important in star ratings of restaurants. However, the result is opposite. Although stack bar plots indicate higher ratings in some categories of these attributes, like restaurants whose price ranges are 4 or which provide valet parking service, the difference in star ratings is masked by the unequal number of records under different categories.

## 4. Key Findings About Steakhouse Businesses

### 4.1. Tips Analysis

For food-related words, not surprisingly, the top one is 'steak' and other words concerning the main procedure of a dinner also showed up frequently, such as appetizer, side dish, and desserts. This implied that the quality of the steak is undoubtedly the most important factor. Secondly, meat other than beef also showed up many times, such as chicken, lamb, and especially sea foods. This implies customers also have diverse needs for meat other than steaks, and seafood is especially popular. Moreover, we surprisingly found 'sushi' is the third frequentest



**Figure 2. Star Ratings by Attributes**

word. Although sushi is not a traditional partner for steaks, it turns out that they are very popular among customers in steak houses. Since we have also seen Italian food, we think that offering overseas specialties is also a good way to attract customers. Finally, drinks are also a major focus, especially customers care about the bartenders and cocktails.

For service-related words, ‘atmosphere’ is found to be the most important non-food factor for a steak house. Comfortable music, warm lighting and beautiful tableware etc., will make a pleasant dining atmosphere. The second is ‘service’ and it was accompanied with ‘friendly’, ‘wait’, ‘reservation’ and ‘delivery’, so it was not hard to conclude fast and friendly service, convenient reservation and delivery are what the customers want. Also, prices are always important, based on the content of specific comments, people care more about whether the price is worth comparing rather than the absolute price. Lastly, we noticed people care about opening time such as ‘dinner’, ‘lunch’, and ‘brunch’. Since most restaurants opens at lunch and dinner time, it will be a good idea to serve brunch to expand business for those currently do not.

## 4.2. Attributes Analysis

We do linear regression of star ratings on all the 7 influential attributes. The adjusted  $R^2$  is 0.2481, which does not imply a good fit. Several reasons may explain for the poor fit. Star rating is discrete and ranges from 1.5 to 4.5. This violates normality assumption, which is convinced by residuals QQ-plot. Another reason is that star rating is not decided by these attributes completely. It is also affected by sentimental preference, geographical information or other factors that are difficult to quantify. We also checked the plot of residuals against fitted values, which indicates no specific patterns exist. Thus, the model is sufficient and equal variance assumption holds. Although the linear regression is not satisfactory, we still get some insights from it.

Some general suggestions based on attributes are as follows. With all other attributes

fixed, the average star ratings will increase 0.5789 if the steakhouse business provides only beer and wine and increase 0.4049 if the steakhouse business is a full bar; decrease 0.1579 if the steakhouse business is good for kids; increase 0.2031 if employees in the steakhouse business dress more formally; decrease 0.1437 if delivery is available; increase 0.1605 if reservation service is available; increase 0.1991 if the ambience is more casual and increase 0.3301 if the ambience is classier.

## **5. Recommendations for Steakhouse Businesses**

Based on our analysis of attributes and tips, we have general recommendations from two aspects for the steakhouse business.

The first is about food and drinks. It is undoubted that the quality of steaks is naturally the most important. However, we have some special suggestions in other ways. We find that desserts are especially popular, so we suggest that restaurants design innovative desserts to attract customers. Also, it's a good idea to have some unique cuisine, such as high-quality seafood, or delicious sushi. Finally, reasonable pricing is an important aspect. We recommend setting relatively high prices for high-quality ingredients and designing some affordable dishes. This allows customers to choose for themselves. For drinks, we found simply serving alcohol would make a great contribution to the star ratings. Also, people care about the quality of the liquor pretty much, as they comment about bartenders and cocktails very frequently.

The second is service-related, besides friendly and fast service, we highly recommend business owners to make the ambience classier and ask the waiters to dress more formally. This is because we found that people prefer steaks with a more casual atmosphere than a formal one. Secondly, providing reservation will greatly improve the star rating and delivery service is also very much in need.

After all, it's better for steakhouse business owners to provide more information on Yelp, such as restaurant ambience, alcohol provided and whether the restaurants are good for kids. This will help customers find steakhouse businesses that best fit their preference.

## **6. Discussions**

### **6.1. Strengths**

1. ANOVA analysis for Attributes is easy for interpretations, and the results are very helpful for making suggestions
2. The word clouds gave vivid visualization for customers' feedback and helps owners to gain an intuitive idea about important aspects.

### **6.2. Weakness**

1. Star ratings data violate normality assumption for linear regression models due to discrete values and limit range.
2. Reviews, locations or other information related should be considered in further discussion, and more complex word processing methods can be used

## 7. Contributions

MC conducted attribute analysis-extract\_attributes.ipynb, attributes\_analysis.R and attributes\_suttestions.ipynb, created README.md, wrote attribute analysis part in the summary, worked on and presented slides page 8 to 19.

HL wrote Exploratory Data Analysis, Key Findings and Recommendations, edited Introduction, Data Cleaning and Discussion in the summary, worked on and presented slides page 1 to 7, created and revised wordcloud\_tip.R in GitHub, wrote contents in 'General Food Advice', 'General Service Advice' pages of the shiny App.

RH conducted the data pre-cleaning work on the raw json files, edited global.R, sever.R and ui.R on the RShiny application and provided general feedback on the Shiny app, edited in the summary and presented slides page 20 to 28.

## 8. Source

[1] <https://docs.developer.yelp.com/referencebusiness-update>