DS-GA 1007 Final Project: Exploratory Data Analysis of Restaurants on Yelp:

Important features on How to Open a Good Restaurant

Group 2: Jie Gu, Wenjie Shao, Yuhan Liu, Xi Yang, Tiantian Zhang

Center for Data Science

New York University

{jg6617, ws2237, yl7576, xy2122, tz921}@nyu.edu

1. Introduction

1.1 Business problem and dataset:

According to the National Restaurant Association, the projected sales of the U.S. restaurant industry will reach $899 billion in 2020, and the total number of restaurants in the U.S. is over 600 thousand. Undoubtedly, the restaurant industry provides tons of opportunity for people, but at the same time, it becomes extremely competitive as customers have higher standards when determining a good restaurant. Therefore, in this project, we want to look for features that contribute to a successful restaurant from the aspect of data science and give business owners some advice on how to open and operate a restaurant that satisfies most customers.

The two datasets we are going to analyze to achieve our goal are from Yelp, which provide information of businesses registered on Yelp and their customers' reviews. In the first dataset, each row represents a business and contains features such as locations, review stars, price range, opening hours, parking information, categories of restaurants, and etc. Another dataset is about the review information and each row represents a review written by a yelp user, and the columns contain the user id and business id. From these two datasets, we can understand the success of some restaurants by applying exploratory analysis and generate suggestions for business owners from multiple dimensions.

1.2 Data preparation:

As the dataset includes all kinds of businesses on Yelp, we first filtered by business type and selected all restaurants to do data analysis. Then, in the origin dataset, most of the columns are dummy variables, so we aggregate restaurant type dummy variables into one categorical variable to make the visualization easier. In this way, we can analyze the relationship between various restaurant features and review stars which measure the success of a business.

2. Distribution of Restaurants

There are totally 20,042 restaurants in the dataset and they can be divided into 28 categories such as Mexican, Chinese, Pizza, Barbeque, and etc. To understand the distribution of restaurants in the dataset, the number of restaurants in each state and each city are shown in **Appendix 2**. Most of the restaurants are located in Arizona and Nevada, and Las Vegas as well as Phoenix are two cities that have more restaurants than other cities. To be more specific, among the 28 different types of restaurants, Mexican, pizza, and Chinese food are the most, whereas vegetarian, nightlife, Pakistani are the least. From these distributions we can see that the catering industry is particularly competitive in some regions and types, which will also give lots of opportunities.

3. Important Feature: Review Stars

3.1 distribution of review stars

The distributions of 'review stars' are shown in **Appendix 3.1 Figure 1**. It is obvious that review stars of most restaurants are in the range [3.0,4.0]. In order to better understand the underlying patterns of the distributions of 'review stars', we divided our restaurants into two categories, one with lower 'review stars' ranging from 0 to 3.5, and another one with higher 'review stars' ranging from 4 to 5, which will be considered as good restaurants in our study.

When deciding to open a new restaurant, the location and type of the restaurant are two of the most important factors that need to be considered. The average of review stars in each state and city can be found in **Appendix 3.1 Figure 2, 3.** Among the 260 cities in the dataset, the top 20 cities with most restaurants are selected to analyze the review stars. It is clear that restaurants in Edinburgh have higher review stars than restaurants in other cities. As for the types of restaurants, the distribution of 'review stars' of different types of restaurants can be seen in **Appendix 3.1 Figure 4**. Fast food restaurants have lower review stars than average, and middle eastern restaurants as well as vegetarian restaurants have higher review stars than average. Therefore, from the above analysis we can conclude that the average review star varies in different regions and different types. It may be more difficult for a business owner to open a good restaurant in cities that have large percentage of restaurants with high evaluation, because the market is more competitive and requires higher standards. However, we should also consider that the reason why some certain types of restaurants have high average review is that the total number of them is too low. So, it can be valuable opportunities but also challenges for business owners to enter a niche market.

3.2 Other Features Contribute to Review Stars

In further steps, the relationships between other features and 'review stars' are explored. Most of those features are binary features, which takes value 0 or 1 for whether restaurants provide this kind of service or not, respectively. **Appendix 3.2, Figure 1 to 4**, show that most of those features cannot make a big difference in review stars, no matter whether restaurants provide those services or not. However, whether the restaurant accepts credit card or not and whether the restaurant provides free Wi-Fi are two significant factors as there shows an obvious drop in review stars for restaurants without these services. So, these are two factors make great effect on the success of a restaurant.

Besides, there are some features that showed interesting patterns in contributing to the high review stars. In **Appendix 3.2**, **Figure 5**, we see the average of 'review stars' for restaurants with different noise levels. From this graph we can see that when a restaurant is recognized as very loud, it is less likely for customers to offer a high 'review star'. On the contrary, if a restaurant offers a quiet dining atmosphere, it is more likely that the customer would leave a higher review. We can also see from the graph that most restaurants lie in the average category with medium level of noise. Another interesting example is shown in **Appendix 3.2**, **Figure 6**. The graph shows the average of 'review stars' based on whether a restaurant offers drive-through service or not. It seems that restaurants that offer the drive-through service are less likely to be ranked with high reviews. One possible explanation is that most restaurants with this service are fast-food restaurants like McDonalds, KFC or Burger King. People are less likely to give high reviews to these types of restaurants as what they seek is not the high-quality dining experience but convenience and reasonable price.

3.3 Review stars over years

To investigate the trend of the review stars over time, we plotted the mean stars ratings against years. Since the number of reviews was relatively low before 2013 (**Appendix 3.3, Figure 1**), we only take star ratings from 2013 to 2017 into consideration. As shown in the plot (**Appendix 3.3, Figure 2**), there is a clear increasing trend of the mean star review ratings from 2013 to 2017, indicating an overall increase in the customer satisfaction among the Yelp's users with restaurants. The increase might be due to an overall improvement in the restaurants' foods and services or emerging good restaurants. Also, restaurant owners might become more aware of the star ratings and reviews at Yelp, which motivates them to improve their services. Therefore, business owners who would like to open new restaurants should keep their minds up-to-date and always improve their services and food qualities to maintain the success of their businesses.

4. Yelp Review Text Analysis

4.1 Most common words in 1-star and 5-star reviews

To explore what users' shared preferences are, we explored users' reviews corresponding to stars 1 (lowest) and 5 (highest). Specifically, we found the most frequent words in users' 5-star and 1-star reviews.

The plot (**Appendix 4.1, Figure 1)** shows the frequencies of 30 most common words in users' reviews with star 5 except the stop words that we have determined. From the words like 'great' and 'best', we can know that customers hold a positive attitude for the restaurants. Foods in these well-performed restaurants are often 'delicious' and 'fresh' and are popular items like 'Chicken' and 'pizza'. It is also reasonable to guess that customers feel pleased with the service from the words like 'service' and 'friendly'. Besides, due to the high frequency of 'service', we are able to say that service is a significant measurement for restaurants. In addition, since 'time' may refer to 'waiting time' or 'having a good time', it is hard to make a conclusion on it.

Another plot (**Appendix 4.1, Figure 2)** presents the most frequent words in star-1 reviews. Similar to the previous plot, the 'service' appears frequently, however, it might indicate that customers are disappointed with the service here. The word 'manager' also further proves this because it may refer to 'complain to the manager'. According to 'minutes' and 'wait', customers are likely to lose their patience after waiting for a long time. Moreover, we can see that customers are not satisfied with the 'location'. Thus, this would be insightful for people who are trying to open a good restaurant by avoiding these issues.

4.2 'Best' Restaurant analysis

For those who would like to open a good restaurant, it is worth time analyzing the restaurants that often get 5-star reviews and learning why it is highly appraised so that merchants can learn from it. We selected the top 6 restaurants with the most reviews with star 5 and chose the one having the highest average star among to be the 'best' restaurant. The figure (**Appendix 4.2, Figure 1)** is a word cloud of most frequent words in the reviews. The more frequently a word appears, the larger the word is. From this plot, we know that this is a Korean BBQ in Las Vegas. Except for the advantages we

mentioned in all reviews with 5 stars, the word ‘price’ may mean that the price in this restaurant is affordable. Moreover, it is a good idea for merchants to hold ‘happy hour’ since many customers seem to be satisfied with that for this ‘best’ restaurant.

We are also interested in some 1-star reviews from this ‘best’ restaurant. The table (**Appendix 4.2, Table1)** contains the first 3 1-star reviews for this restaurant. These reviews complain about the long waiting time and the servers’ rude attitude. The first one also points out that they offer a free dessert if customers review them, leading to a high score in yelp. In fact, it is reasonable for merchants to take this action in order to open a good restaurant. However, merchants should commit that no matter how customers review them, they will still show a good attitude and offer a free dessert. By doing this, restaurants can gain good reputation and learn their shortages at the same time.

4.3 Reviews for certain types of restaurant

Merchants who want to open a certain type of restaurant can gain insights from the reviews by successful restaurants within a specific type. Since our dataset has most Mexican food and Piazza reviews, we will take 5-star reviews from those restaurants as examples. Figure (**Appendix 4.3, Figure 1)** indicates that ‘taco’, ‘carne asada’, ‘burrito’ and ‘salsa’ are most popular among Mexican restaurants. From figure (**Appendix 4.3, Figure 2)**, we know that when people eat pizza, they may like to have ‘salad’, ‘wing’, ‘sauce’ and ‘drink’ as sides.

Generally speaking, merchants can learn customers’ preference and requirements for restaurants. They can adjust their restaurants’ food, service, price and location to make their restaurants ‘good’.

5. Conclusion and Future Work

In conclusion, it is worth analyzing the existing restaurant reviews and star ratings to get insights about customers’ preferences from both good and bad reviews. Through a series of exploratory data analysis, we can find that customers have different expectations on different types of restaurants in different regions. In terms of services, some of them highly affect customers’ dining experiences such as acceptance of credit card, Wi-Fi, and noise level. Also, by extracting key words in written reviews on Yelp, we also noticed some certain aspects of restaurants that customers care about, including servers’ friendly attitude, fresh food, short waiting time, and affordable prices. Therefore, business owners should carefully choose their business types, locations, apply successful factors of existing good restaurants and avoid issues leading to customers’ dissatisfaction.

While this project can give some suggestions for business owners, it is not absolutely reliable because we only used data from Yelp. As claimed in part 4.2, some restaurants use tricks to have higher review stars and better reviews online, which make the data unreliable. For future works, we can collect more data from multiple websites and have a more comprehensive overview of restaurants in the market. Most importantly, business owners should take their own research while using our result as references before opening new restaurants because of the uniqueness and specialty of every business.

Appendix

Part 2

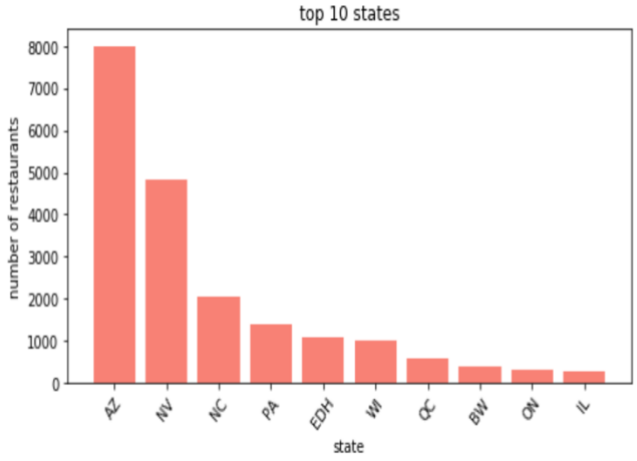

Figure 1

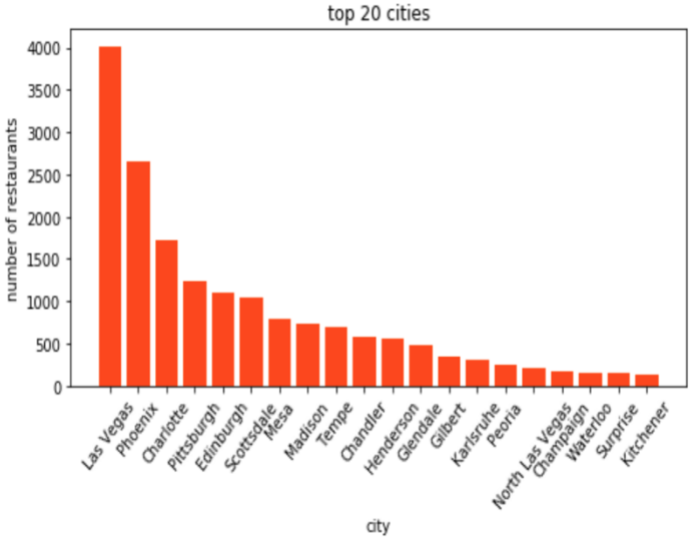

Figure 2

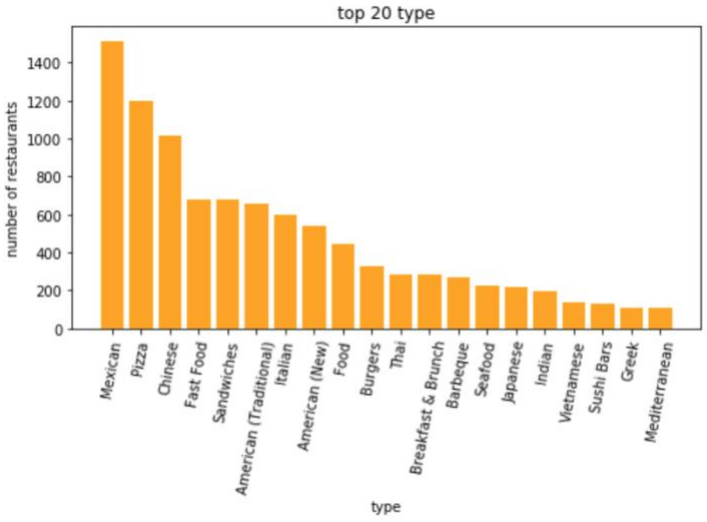

Figure 3

Part 3.1

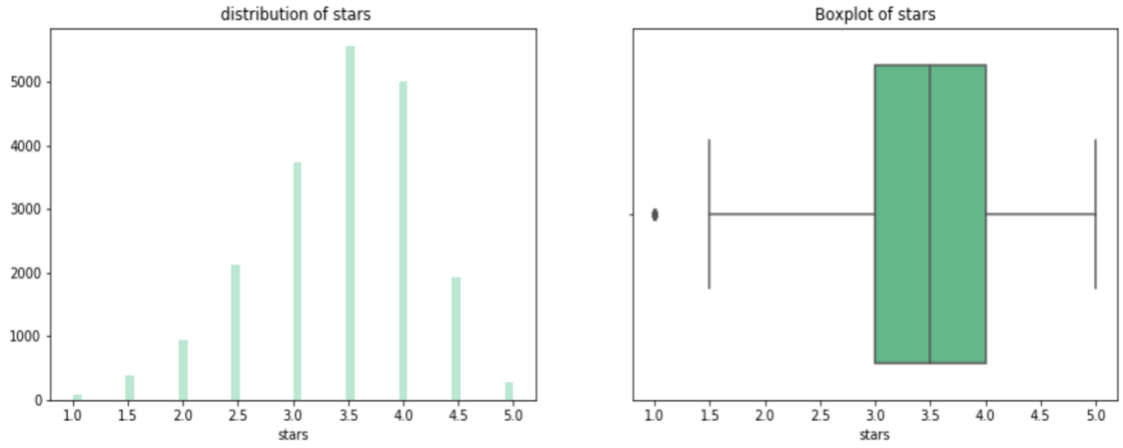

Figure 1

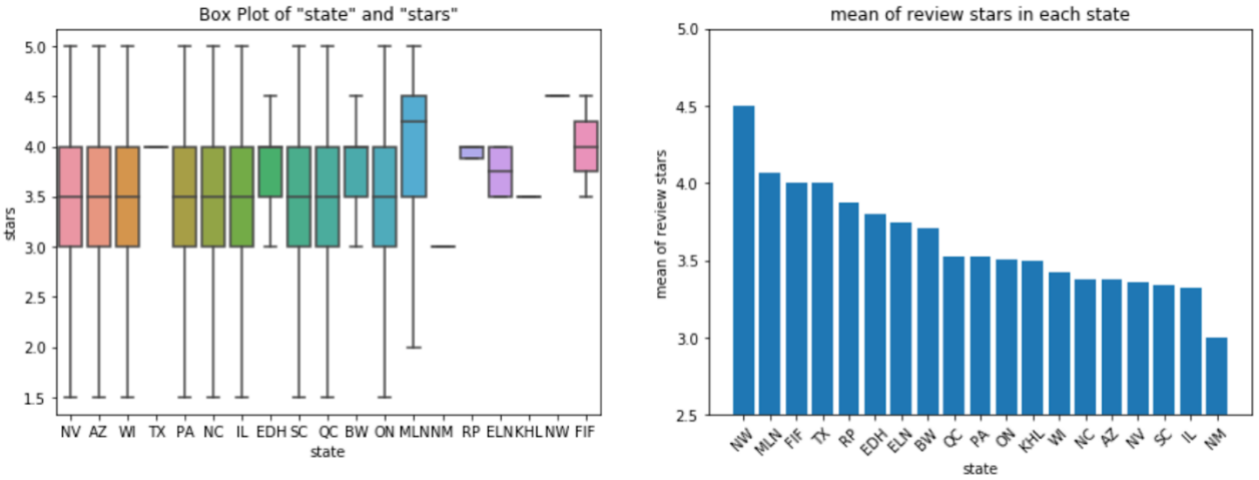

Figure 2

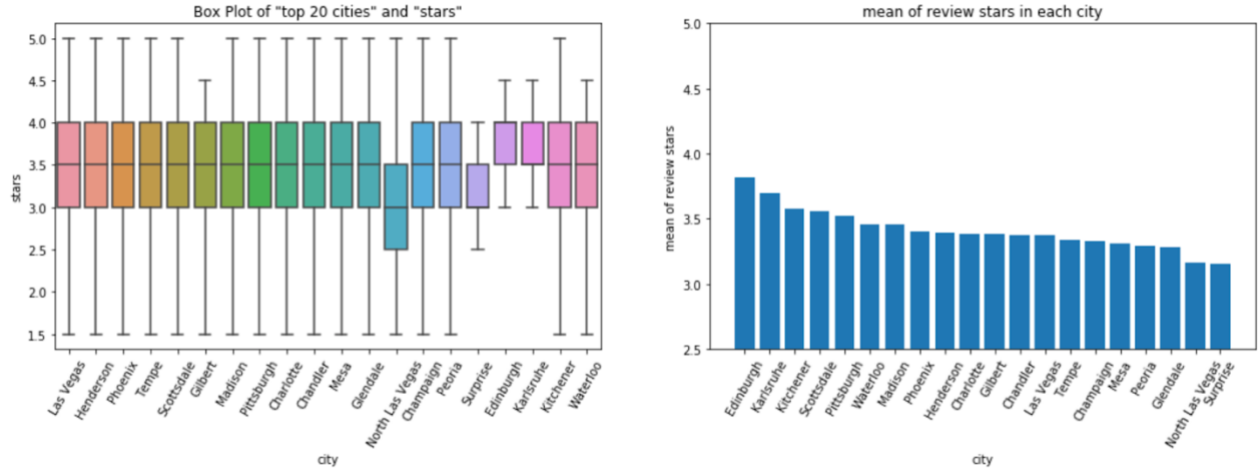

Figure 3

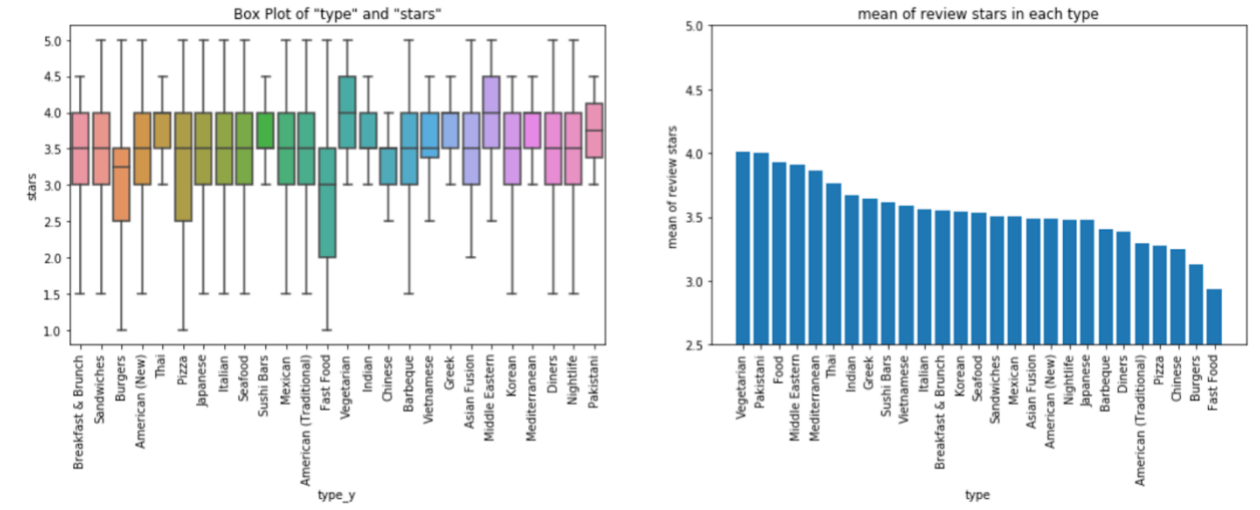

Figure 4

Part 3.2

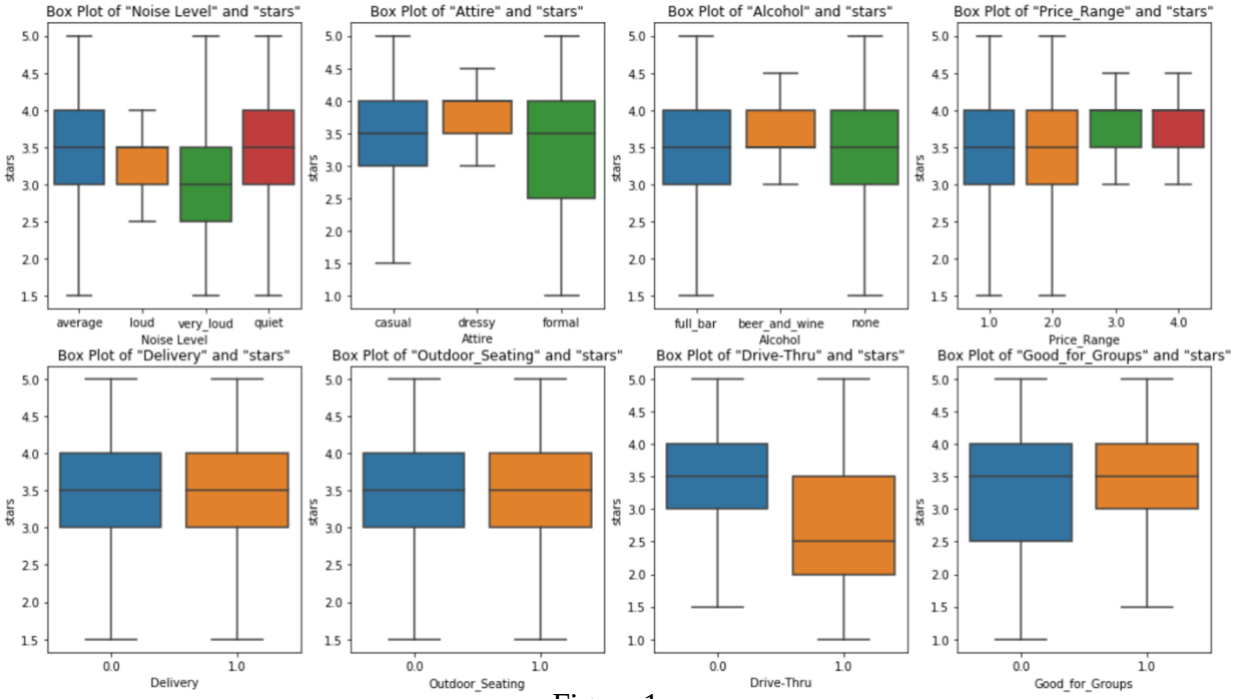

Figure 1

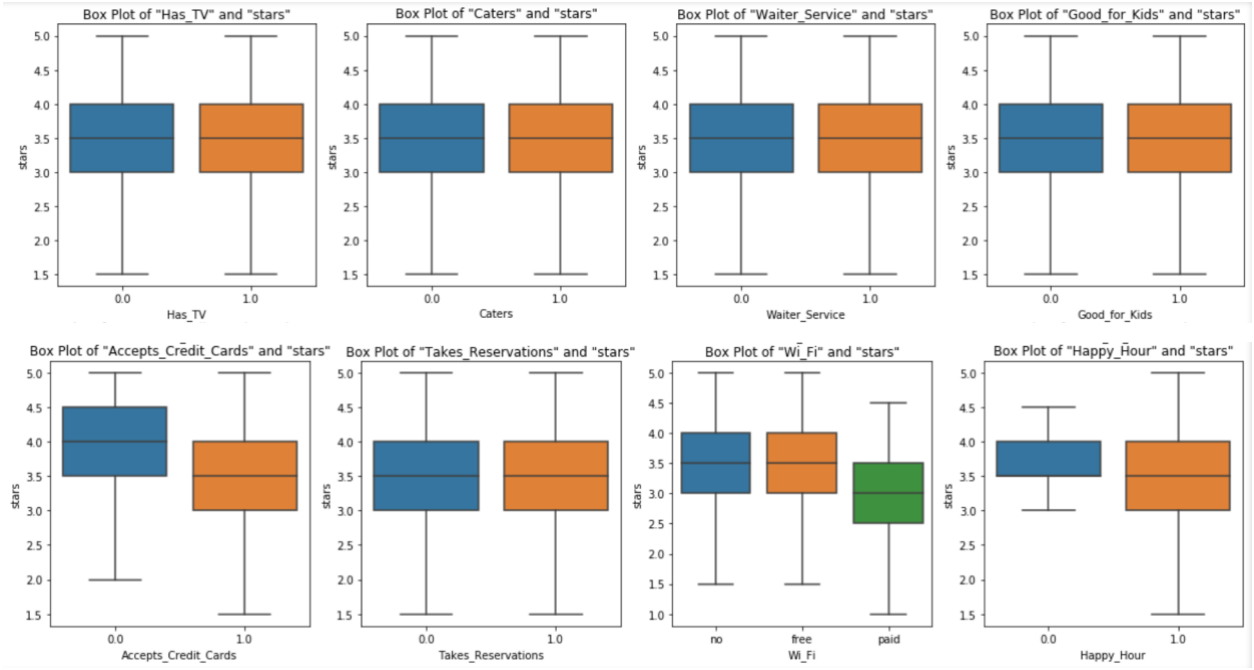

Figure 2

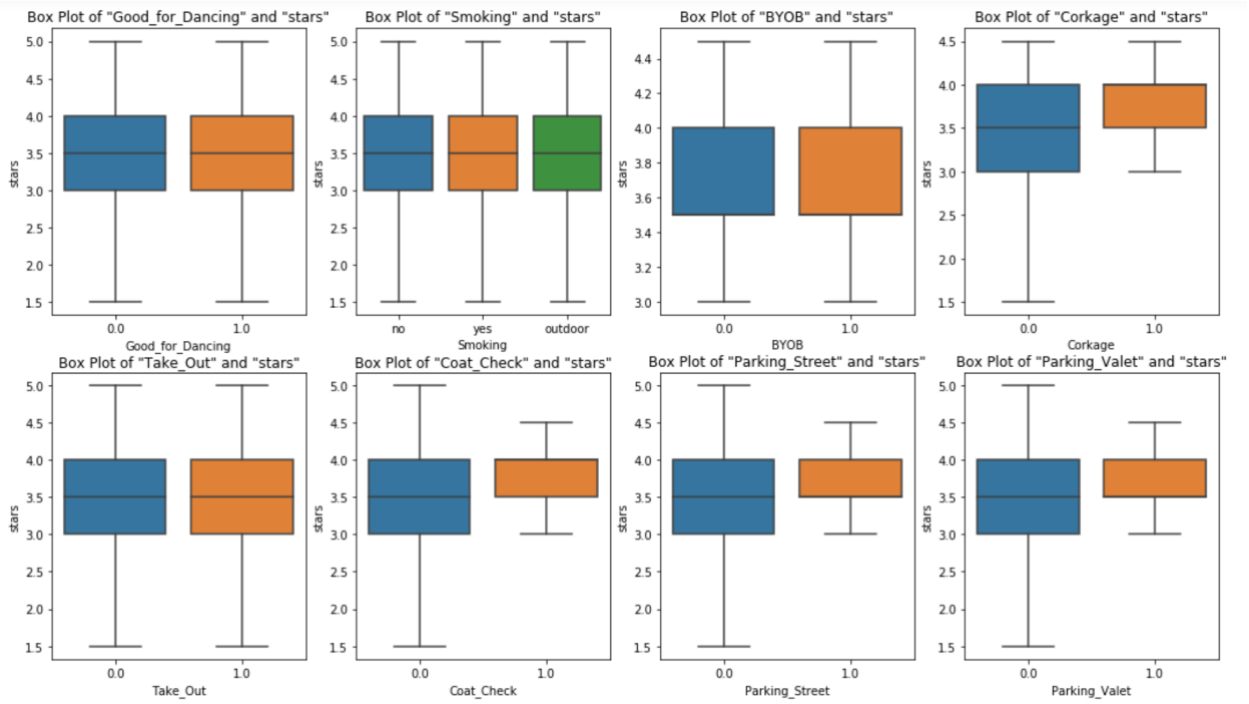

Figure 3

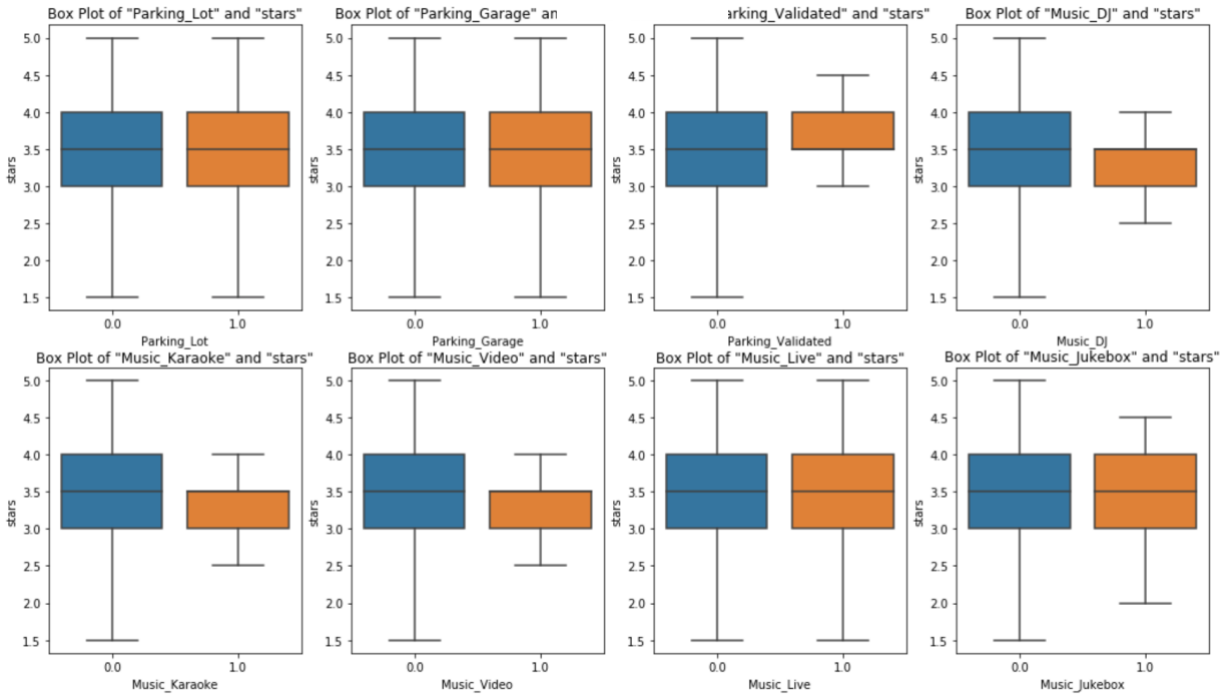

Figure 4

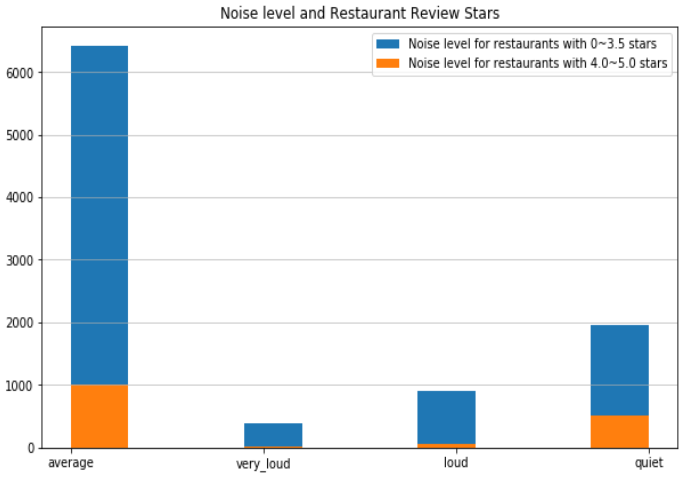

Figure 5

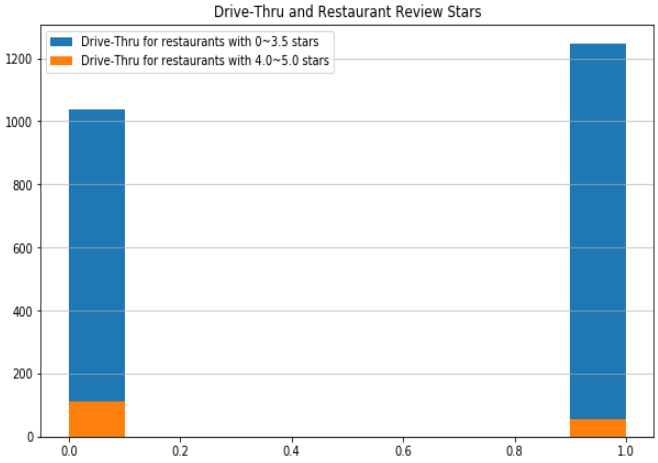

Figure 6

Part 3.3

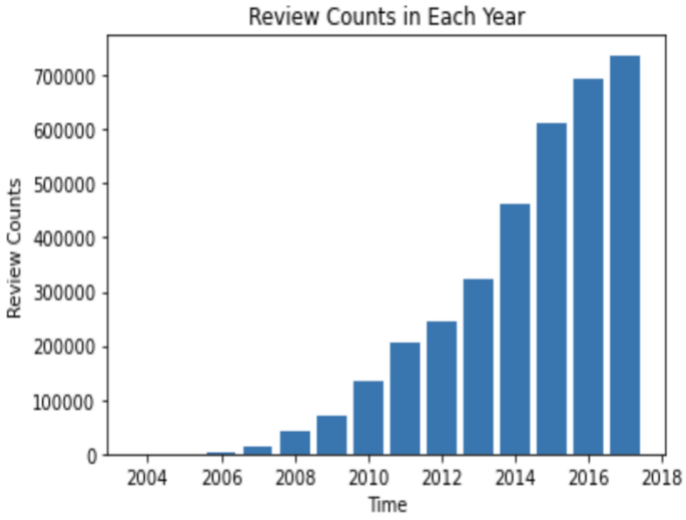

Figure 1

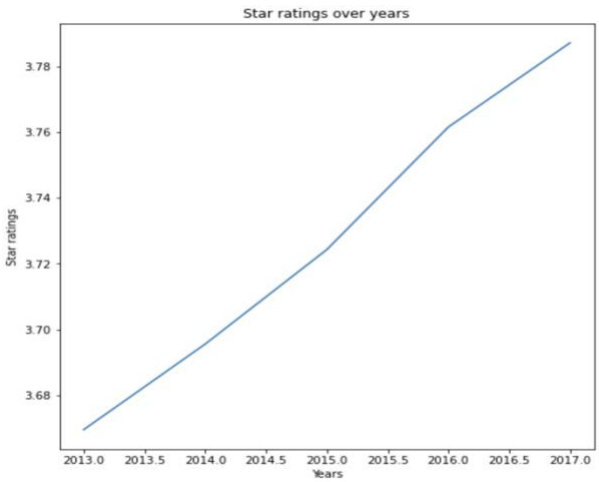

Figure 2

Part 4.1

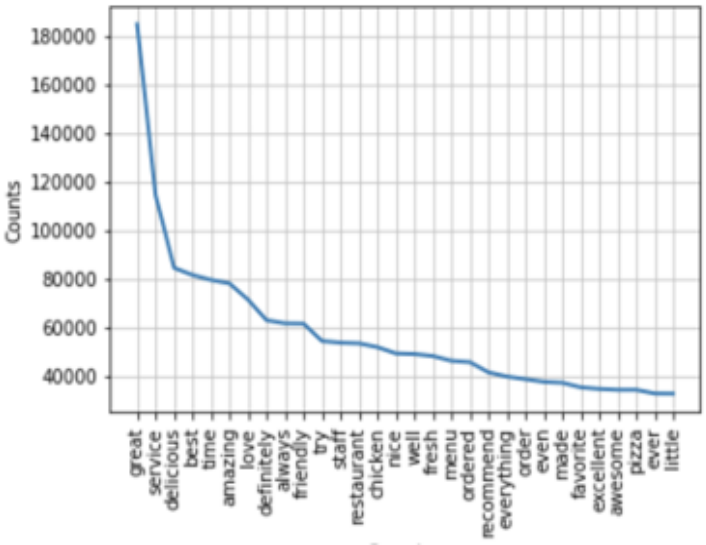

Figure 1

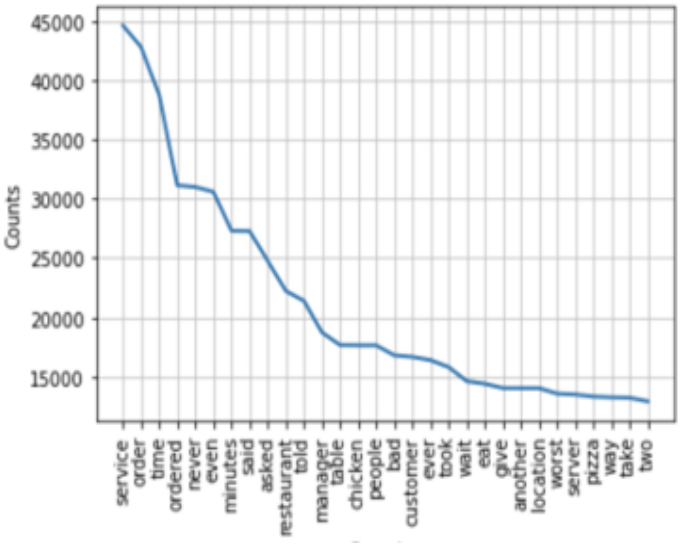

Figure 2

Part 4.2

Figure 1

)ut[236]:

	review_id	user_id	business_id	stars	date	text	useful	funny	cool	city	sta
519678	Oxhy2uXkLinvDVe6Mux-Sw	MOxfE9V_yHaN5ZgfSk1t4g	hihud--QRriCYZw1zZvW4g	1	2015-11-15	So came here for Lunch on a Saturday. Food was okay. However, the service was absolutely horrendous. They weren't very busy at all. It took 45 minutes to bring out our beef toro and we asked about it twice, and they said the kitchen is still slicing it. We ordered combo fried rice and specifically said only chicken and beef. They brought it out with shrimp too, and when we told them, our sever said he'll just take out the shrimp for us. Umm.. No.. The rice tastes like shrimp now. Overall service was horrible, has to request more water 3 times, request for the grill to be changed, request for my check after waiting 15 minutes.\n\nThe reviews on YELP are totally FAKE because they offer a free dessert if you review them. No body is going to show them a 1 star review to get a dessert.\n\nDefinitely do not recommend.	0	0	0	Las Vegas	N
519704	BHo81rRKLfn4X1GaV9QPwg	Bw9Xlh7M84gY1kBfPZFUxQ	hihud--QRriCYZw1zZvW4g	1	2016-04-13	Server told us to hurry up just as we were finish eating because there was a reservation. That's not our problem. That's very rude whether you are a server or not. Never coming here again and telling everyone about this bad service	2	0	0	Las Vegas	N
519725	GNaRj3G-1l0v5CcjwbHkzg	BKrJUkm8weTBm069zKUZgw	hihud--QRriCYZw1zZvW4g	1	2016-07-22	I haven't been able to try the food due to the overwhelming long line of 90 plus minute wait time. Just not worth the wait in Vegas! 1 star for the long wait. I would give it no star but that's not an option.	0	0	0	Las Vegas	N

Table 1

Part 4.3

Figure 1


will
ordered
back
people
slice
well
one
made
order
great
everything
food
even
friend
never
take
meal
say
topping
Came
pizza
got
eat
staff
friendly
tried
amazing
know
come
wait
love
place
service
time
awesome
good
menu
go
drink
crust
dinner
table
server
make
fresh
best
salad
nice
try
deliciou
restaurant
went
best pizza
sauce
excellent
want
fantastic

Figure 2