STAT 628 Module3: Analysis on Yelp Reviews Data

Huitong Kou, Zihang Wang, Peibin Rui
November 2020

1. Introduction

There are tens of thousands of reviews for different types of businesses. Our analysis mainly focuses on restaurants which serve steak in several US cities such as Madison and Pittsburgh. Among these restaurants, our specific goals are:

- Investigate the relationship between ratings and words in reviews from two different aspects: foods, non-food factors.
- Provide useful advice for improving the ratings of steak restaurants on Yelp based on our analysis.

2. Data Cleaning

Our dataset contains a subset of million reviews from restaurants in the U.S. such as Madison, Cleveland, Pittsburgh and Urbana-Champaign released by Yelp. The restaurants with at least 3 reviews older than 14 days are included and only reviews that were recommended at the time of the data collection are included

Limited by the computing power, we chose to focus on restaurants whose category contained the word “steak”. We followed the standard practice in Natural Language Processing (NLP) using the software package *nltk* in Python to clean the reviews. In order to match reviews with corresponding restaurants, we combined the tables created from *business.json* and *review.json*. To get some insights on how words in reviews are related with Yelp ratings, we created new variables by calculating the word frequencies in each review.

After filtering and combination, our primary cleaned data set *steak_cleaned.csv* has 33629 rows and 6085 columns.

3. Exploratory Data Analysis and Key Findings

3.1 Insights on different types of steak

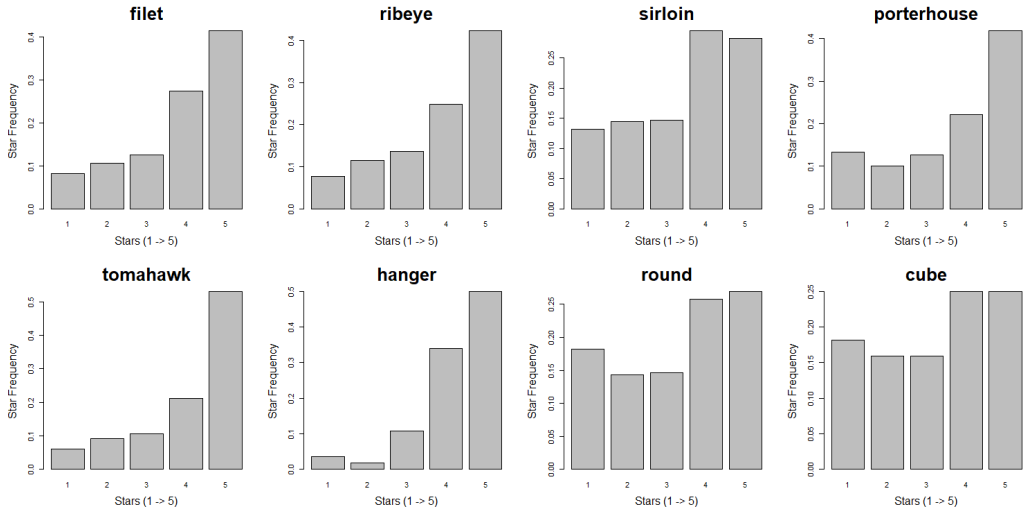

Figure 1

Firstly, we looked at the common types of steaks which appear many times in the reviews and investigated the distribution of their ratings. We found that reviews mentioned filet, ribeye, porterhouse, tomahawk and hanger steaks tend to have more 5-star ratings. The portion of their 5-star ratings is significantly over 45%.

However, although reviews mentioned sirloin, round, and cube steak also have a large proportion of high ratings, customers seem to be pickier since they also gave many low ratings on these steaks and the proportion of 5-star ratings is roughly less than 30%.

3.2 Insights on factors other than steak

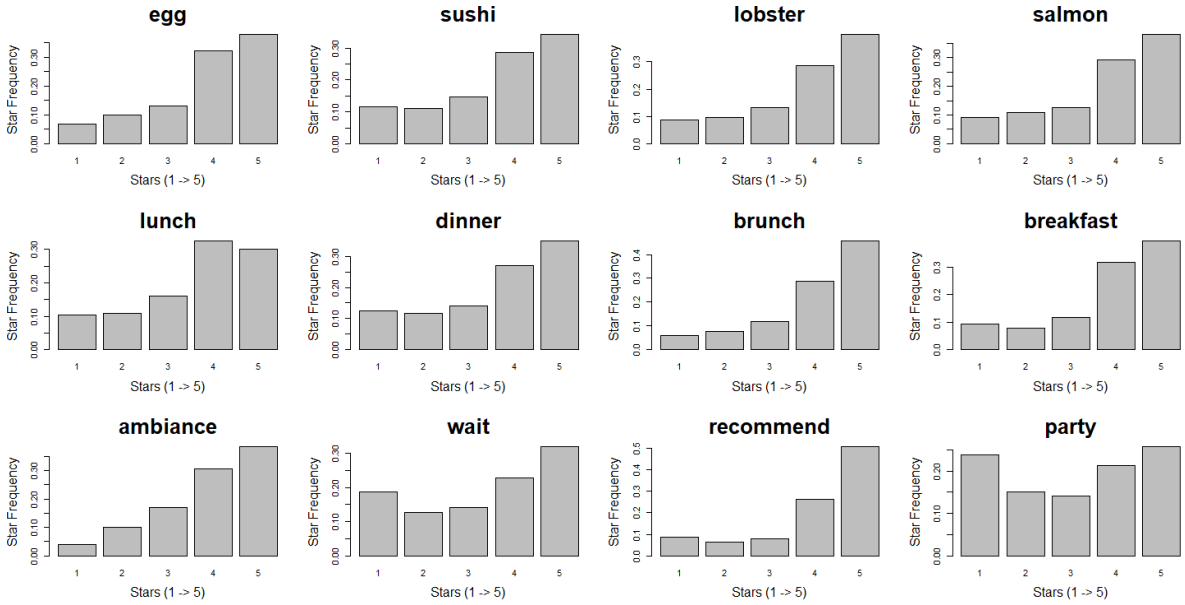

Figure 2

According to these plots of star distribution, we found that generally customers tend to give higher ratings if they mentioned some side orders other than steak. For examples, egg, sushi, lobster and salmon are more frequently mentioned as the star goes up, especially from 3-star to 4-star.

As for the meal time, all of them have a large proportion of high ratings. And brunch seems to be more popular than other meal times.

Non-food factors also play an important role in reviews' rating. Clearly, customers don't like waiting for seats or having meals when there is a party. Ambiance is one of the most important environmental factors that customers may pay more attention to or have higher expectation. A friendly and quiet ambiance can make people feel comfortable and be willing to give higher ratings. What's more, people also tend to give higher ratings if the steak restaurant is recommended by friends or relatives.

4. Statistical Analyses

4.1 T-tests on Business Attributes

Notice that the *business_city.json* contains many useful attributes of each restaurant. We decided to conduct t-tests on some of the attributes to see whether different levels of these attributes can make a statistically significant difference on a restaurant's rating.

We conducted the t-tests on different subsets since the attributes of each restaurant are differ from others. The first step is checking the variance equality of each 2 subsets. Then t-tests were

conducted accordingly using the *t.test* function in R.

Figure 3 shows the overall star distributions of all steak restaurants. Seven attributes were selected and p-values of each t-test are listed in Table 1. According to the results of t-tests, with significance level of 0.05, the Restaurants Reservations (True), Restaurants Attire (Dressy), Outdoor Seating (True) and Restaurants Delivery (False) can statistically lead to significant higher stars. The Noise Level (Quiet), WiFi (Free), Restaurants Good For Groups (True) don't statistically matter.

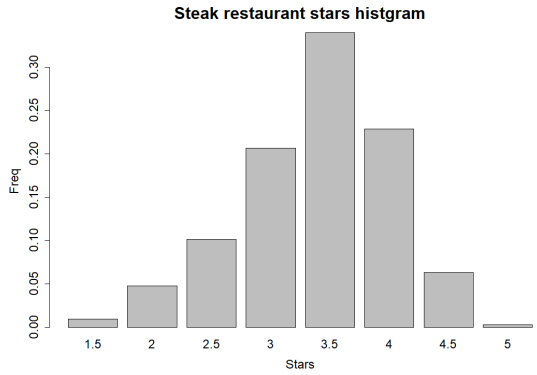

Figure 3

attribute	p.value
RestaurantsReservations	0.040536672387
NoiseLevel	0.115368255456
RestaurantsAttire	0.000007760033
WiFi	0.799610824593
OutdoorSeating	0.044275411757
RestaurantsDelivery	0.012348344280
RestaurantsGoodForGroups	0.454369986827

Table 1

4.2 Multiple Linear Regression Model

To quantitively analyze the influence of word frequency on review ratings, we fitted a multiple linear regression model. The dependent variables are star ratings and predictors are word frequencies in each review. Only words which occur more than 4000 times and we are interested in are included. So there are 93 predictors in total. The linear regression model has an R^2 of 0.3365, which implies that these predictors can explain 33.65% variation of review ratings. It is relatively low but reasonable since only a small part of whole words are considered in this model.

Using this regression model, we conduct t-test for each predictor to see whether having more of the word is significant in predicting ratings, while controlling for other words. Results of estimated slope and 95% CIs for some interesting words are shown as below:

Coefficients	Estimate	Lower Bound	Upper Bound	Standard Error	P-value
Intercept	3.3798	3.3567	3.4028	0.0118	$<2\times10^{-16}$
filet	0.0667	0.0312	0.1023	0.0182	2.374×10^{-4}
ribeye	0.0713	0.0231	0.1195	0.0246	3.732×10^{-3}
brunch	0.0626	0.0183	0.1069	0.0226	5.602×10^{-3}
ambiance	0.1319	0.0618	0.2020	0.0358	2.379×10^{-4}
recommend	0.2894	0.2488	0.3300	0.0207	3.313×10^{-44}
party	-0.0685	-0.1076	-0.0295	0.0199	5.850×10^{-4}
$R^2=0.3365$	$Adjusted\ R^2=0.3347$		$p-value: <2.2\times10^{-16}$		

Table 2

These results reinforce the key findings in EDA part. Mentioning words "filet", "ribeye",

"brunch", "ambiance" and "recommend" in reviews has a positive effect on ratings while mentioning word "party" may decrease the ratings. Take the result of word "recommend" as an example, mentioning "recommend" in reviews one more time will increase the ratings by 0.2894 on average.

As for model diagnostics, all VIFs are less than 2, which means there is little multicollinearity. Then we used QQ-plot to check the normality of residuals. Most of the points follow the diagonal, so the normality assumption holds. However, the residuals distributed not randomly. There exits some linear patterns on the residual plot. It is reasonable since the response of this model only has five outcomes. This is also the limitation of our multiple linear regression model. Ordinal logistic regression may be a better choice but it is more complicated and harder to interpret the results.

5. Data-Driven Recommendations and Actionable Plans

Based on our analytics we done above, we can provide some recommendations for steak restaurant owners or someone who would like to open a new restaurant serving steaks.

- For types of steaks, we suggest the business owners pay more attention on improving the quality of filet, ribeye, tomahawk, hanger and porterhouse steaks and emphasize them on their menus, since customers tend to be more satisfied with these dishes according to Figure 1 and results of regression model. At the same time, advertise their sirloin, round and cube steaks if one of them has brought their restaurant high-star comments. Since these steaks seem to be less welcomed, having highly rated them is praiseworthy and will make the restaurant more competitive.
- For food other than steaks, we suggest the steak restaurants consider providing or improving lobster and salmon as side orders since they are highly related with higher ratings. They should also consider hiring cooks to serve desserts, salad and cheese, especially sushi. The steak restaurants serving these foods generally get higher ratings on Yelp. What's more, providing brunches is also helpful for improving ratings.
- For other environmental factors, our suggestions are creating a comfortable inner ambiance and avoiding holding noisy activities such as parties. Besides, paying attention to the attire of their waiters/waitresses, improving reservation system and offering outdoor seating are also good ways to improve ratings. Last but not least, recommendation is one of the most important factors affecting review ratings. Steak restaurants can design some special offers to encourage customers to recommend them to friends and relatives.

Conclusion and Discussion

In summary, we cleaned reviews from Yelp and tried to get some insights through exploratory data analysis. Then we performed t-tests and built a multiple linear regression model to analyze the relationship between the ratings and reviews or attributes quantitively. Finally, based on our results, we proposed some recommendations and actionable plans for steak restaurant owners which could potentially help them improve their ratings on Yelp.

Due to time constrains, there are also some limitations in our analysis. For example, the data may not be completely cleaned up so that some observations may be misleading. And some conclusions may be reached too hastily and subjective considering the lack of our marketing knowledge.

Contributions

HK wrote/reviewed/edited the outline, data-cleaning, EDA, t-test, recommendations parts of summary; reviewed the corresponding slides; wrote/reviewed/edited the data-cleaning and t-test parts on Github page; reviewed/edited advice part in Shiny.
ZW wrote/reviewed/edited EDA, MLR, recommendations, conclusion parts of summary and did final editing; wrote/reviewed/edited the slides; wrote/reviewed/edited the EDA, MLR and image parts in Github; reviewed/edited plots and explanations in Shiny.
PR reviewed/edited introduction, MLR, and recommendations parts of summary; wrote/reviewed/edited the slides; wrote/reviewed/edited Shiny part in the Github; wrote/reviewed/edited codes and maintain the Shiny.