Stat628: Module 3 - Group 7 Executive Summary

Introduction

Owning and operating a business requires knowledge of the customer in order to adequately adapt to a changing market. Yelp, the online businesses reviewing platform, gives business owners and operators information about how their business appeals to the customer market and provides an opportunity for response to feedback from the platform can increase customer satisfaction.

We desire to understand how certain business attributes affect a customer's rating of a breakfast or brunch restaurant. For these establishments our goal is to provide recommendations on what business attributes can increase a business Yelp rating.

Background Information

Data was provided from Yelp in four json files. We first analyzed the business data file to find which businesses were classified as having categories “breakfast”, “brunch”, “breakfast & brunch”, filtering the data set to 6,525 businesses. Secondly, we analyzed the review data file, filtering the business IDS of the customer review to only business IDS of the 6,525 restaurants. This resulted in 879,195 reviews for us to analyze. We filtered this data further by removing reviews that didn’t receive more than six votes in the “useful”, “funny”, “cool” columns. This was done to filter out reviews that didn’t receive enough attention from Yelp users. We believe that filtering out reviews that didn’t receive enough votes leaves us with the reviews that most affected a business and also provided feedback. This left us with 56,259 reviews to analyze.

We did not remove punctuation marks or non English words for our analysis as our NLP did not need these removed in order to function properly. To analyze the star rating of businesses we removed punctuation and keywords were searched for in the list of attributes. For example: the string “Casual True” was searched for and if found this business was added to a Casual Dataset, “Casual False” was searched for and added to a No Causal dataset.

Part 1: Analysis of Breakfast Restaurants on Yelp

Business data analysis showed the average star rating of the business to be a rating of 3.59 with a standard deviation of 0.85. The star rating of restaurants was standardized by Z-Score calculation of x - mean(x)/sd(x)/ resulting in Figure 2. The new standardized mean and standard deviation are -0.657and 0.845 respectively.

The data was then analyzed to show which attributes appeared most frequently in our list of breakfast restaurants attributes. A plot of the top 25 listed attributes is displayed below in Figure 1. From Figure 1 it can be seen that the attribute “casual” was listed most frequently. This is because there are two attributes that use the word “casual”, attire casual and casual true/false. From Figure 1, we decided to look at several other frequent attributes appearing in the attribute list such as Outdoor Seating, Restaurant Takeout, and Good for Groups. To analyze the effect of

Figure 1

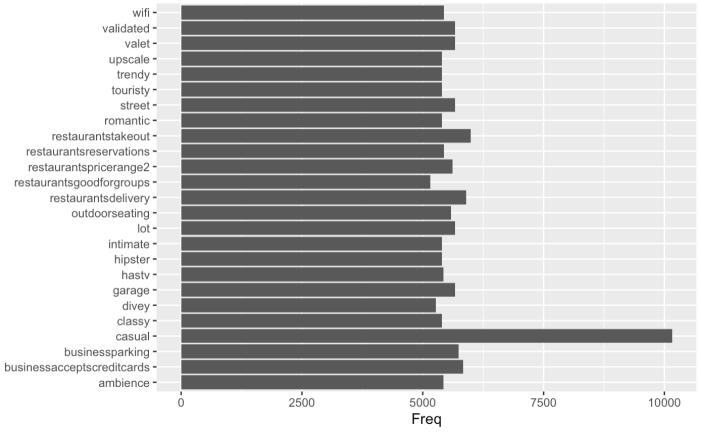
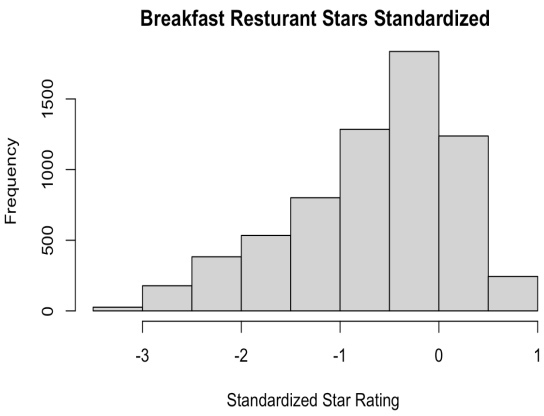


Figure 2



these attributes on the star rating of a restaurant, Welch two sample T-tests were conducted on the gour attributes. There was a large number of samples per group and the data appeared approximately normal when plotted in a histogram. The Two Sample T-test attributes were: Casual True, Outdoor Seating True, Restaurant Takeout True, Good for Group True and their respective False attributes listed in Table 1 as N(Attribute). Therefore our null hypothesis was that the True attribute and the False attribute have no effect on standardized rating of these restaurants. The alternative hypothesis was that the True attributes are more effective at increasing ratings than False attributes. We set a tolerance threshold at 0.05.

The analysis of the T-test for the Casual attribute found that this attribute did have a significant effect on the standardized star rating as the P value is well below our tolerance threshold . Similarly, the T-test analysis found that Outdoor Seating and Restaurant Takeout also had an effect on the average star rating. Good for Group attributes were not significant in affecting standardized rating as their P values were above our significance threshold. Table 1 lists the statistics produced from the two samples T-test with for the attributes listed.

Table 1

Attribute	Alternate Hypothesis	T - Statistic	DF	P Value	Mean X	Mean Y	95% Interval
Casual	C > NC	12.73	3750.5	2.2e-16	-0.5524513	-0.8504738	(0.25951 , Inf)
Outdoor Seating	O > NO	23.345	4291.1	2.2e-16	-0.4509378	-0.9580030	(0.47133 , Inf)
Takeout	T > NT	9.6593	293.78	2.2e-16	-0.6659300	-1.263581	(0.49556 , Inf)
Group	G > NG	-2.7618	871.9	.9971	-0.7434147	-0.6454435	(-0.15639 , Inf)

To analyze the review of the breakfast restaurants we conducted topic modeling. Our analysis with Topic Modeling found that our Yelp Reviews can be categorized into four topics: Food Quality, Serving Speed, Staff Attitude and Customer Loyalty. The Latent Dirichlet Allocation topic model provides us the frequency distribution of words in each topic. To identify the main

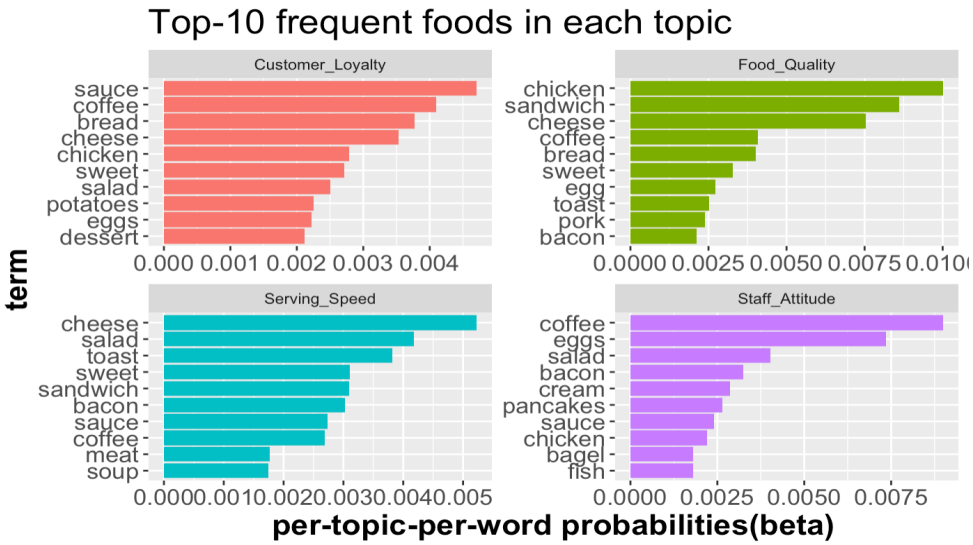
content in each topic properly, we first removed the common words like “breakfast”, then we filtered out the words which refer to specific foods like “sandwich”. After that we are able to name the topics by the high-frequency words, as the Table 2 shows:

Table 2

Topics	High-frequency words
Food Quality	meal, menu, delicious, hot, recommend
Serving Speed	wait, service, time/times, experience, stuff
Staff Attitude	nice, love, service, people, night
Customer Loyalty	time/times, service, delicious, experience, spot, day,

When the customers are talking about these topics, they mention different kinds of foods. As Figure 4 shows, for example, in the topic of food quality, the customers’ reviews suggest that people are more sensitive to the quality of meats, including chicken, sandwich, pork and bacon; For serving speed, people will have higher expectation on the simple foods like salad, toast and soup; The staff attitude are mainly talked when serving coffee and bakery, or we could say sweets connects emotionally with consumers more than other salty foods; And the loyal

Figure 3



customers are picky eaters but also will be charmed by details. They will mention more about the sauce, bread, dessert and even potatoes. By looking at the top frequent foods mentioned in each topic, we could present different views of target customer's behaviors and help restaurant managers to find bottlenecks in their menus.

One of the limitations of our topic model analysis is the identification of specialties. Since they may have uncommon or foreign names, we are not able to sum the frequency up for similar types of dishes. The other limitation is LDA commonly neglecting co-occurrence relations. For further analysis, finer tools that can process grammar structure may be adopted. In our attribute analysis

is that we made samples based on if the desired attribute was found true versus false. Some restaurants did not list these attributes so they were not included in the T-test analysis.

Part 2: Recommendation for Businesses

Based on our analysis of business attributes of breakfast and/or brunch restaurants we recommend the following for business owners: Businesses that have outdoor seating options and takeout options have an increase in star rating compared to businesses that don't have outdoor seating and don't have takeout options. Businesses that have a Casual setting listed in their attributes have a greater rating than those with Casual False listed. Therefore we recommend these businesses have an outdoor seating option, takeout option and a casual environment in order to increase the rating of these restaurants. This is due to these attributes having significant P values from the T Test, showing that they have an effect on standardized star rating. Table 1 shows that the significant attributes have higher samples mean, meaning True attributes have a higher rating. For example, Outdoor Seating True had a sample mean of -.45 and Outdoor Seating False had a sample mean of -.96.

We found that having a Good For Group attribute at these businesses was not impactful on business rating and therefore we are recommending resources be spent elsewhere than accommodating for this attribute. As shown in Table 1 this attribute was not significant because of the P value being above our threshold.

Based on our analysis of the reviews of breakfast and/or brunch restaurants we recommend the following for the respective business owners: If they feel like improving food quality, starting with entrees with chicken or bacon rather than sweets will help them spend less while achieving more as these words appear most frequently in the relevant topic. If their main business is coffee and bakery, investing more on staff training to make the atmosphere more friendly will be of higher priority than serving new items.To gain loyal customers, paying attention to details and quality control will be more effective than incline the limited resources to improve a certain aspect.

Some limitations of our recommendations are that recommendations are based on high-voted reviews and therefore do not stand for every customer's opinion. Another limitation is that some restaurants don't list every attribute so which limits our ability to accurately analyze impact for all breakfast restaurants.

Conclusion

Based on our analysis of breakfast/brunch restaurants on Yelp we found that the attributes of a casual setting, takeout ability and outdoor seating are impactful on a restaurant rating. We found that food Quality, Serving Speed, Staff Attitude and Customer Loyalty are the four topics discussed in reviews.

Contributions

Anna Hayes: Reviewed/Wrote/Edited Introduction, Background information, Part 1, Part 2, and Conclusion of Report. Reviewed/Wrote/Edited Slides 1:3 of Presentation 1. Reviewed/Wrote/Edited Slides 2:5 of Presentation 2. Wrote Data Cleaning Code and business attribute analysis code. Contributed to Github Repo.

David Gao: Reviewed/Edited Background of Report. Reviewed/Wrote/Edited Slides 11,12 of Presentation 1. Reviewed/Edited Slide 9,10 of Presentation 2. Produced Shiny App. Wrote Data Cleaning Code. Contributed to Github Repo.

Kaiyan Ma: Reviewed/Wrote/Edited Part 1, Part 2, and Conclusion of Report. Reviewed/Wrote/Edited Slides 4:10 of Presentation 1. Reviewed/Wrote/Edited Slides 6:8 of Presentation 2. Conducted NLP and Topic Analysis and wrote code for those analyses. Contributed to Github Repo.