STAT 628 Yelp Dataset Project

SHUBO LIN, TIANYUE LUO, NILAY VARSHNEY, HENGRUI QU

December 2021

1 Introduction

1.1 Introduction

In this project, we analyze data from Yelp, focusing on cafés. Our goal is to explore factors extracted from attributes and reviews that could have an influence on business ratings. Furthermore, we provide useful and analytical suggestions to café owners in order to improve their Yelp ratings.
Our work can be mainly divided into two parts: attribute analysis and review analysis. For the first part, we count and choose the attributes that are provided by at least 80% of the businesses. Then, we fit an multi-way ANOVA model to find important attributes and use bar charts to compare their scaled ratings. For the second part, we do sentiment analysis to find informative nouns in reviews. Finally, we combine our findings from the two parts and give suggestions.

1.2 Data Cleaning

The original data consists of 5.36 million reviews on 155 thousand businesses. To get the information on cafés, we extract rows containing "cafe" in the "category" column in "business.json" and export them to a new file. Then, we merge this new file with "reviews.json" based on the "business_id" column. Our final data consists of 303,768 reviews on 3,023 businesses.

2 Attribute Analysis

2.1 Data Preprocessing

We count the attributes provided by the cafés using the "business.json" file. There are 35 attributes in total. However, not all businesses provide all 35 attributes. To avoid bias in our analysis, we choose the attributes that are provided by at least 80% of the businesses, including **RestaurantsTakeOut, BusinessParking, RestaurantsDelivery, WiFi, OutdoorSeating, RestaurantsPriceRange2, RestaurantsReservations, HasTV, Ambience.**

Then, we remove rows with no attributes. Since "BusinessParking" can be divided into five types, we expand the field to Garage/Street/Validated/Lot/Valet. Similarly, "Ambience" can be divided into nine types, so we expand the field to Romantic/Intimate/Touristy/Hipste r/Divey/Classy/Trendy/Upscale/Casual.

2.2 ANOVA

We apply an ANOVA model to fit the star ratings on the business attributes and find the attributes that significantly affect the rating of the businesses. The results from the model are shown below. We consider these attributes to be the key attributes.

Table 1: Significant Attributes from ANOVA

Attribute	DF	P-value
RestaurantsDelivery	2	0.03271 *
OutdoorSeating	2	8.60e-09 ***
RestaurantsPriceRange2	4	2e-16 ***
RestaurantsReservations	2	0.03405 *
HasTV	2	2e-16 ***
garage	2	9.22e-09 ***
street	2	2e-16 ***
lot	2	8.43e-05 ***
valet	2	0.00471 **
intimate	2	0.00928 **
touristy	2	0.04588 *
classy	2	0.00864 **
Residuals	2955	

Table 2: Comparison Within Attributes

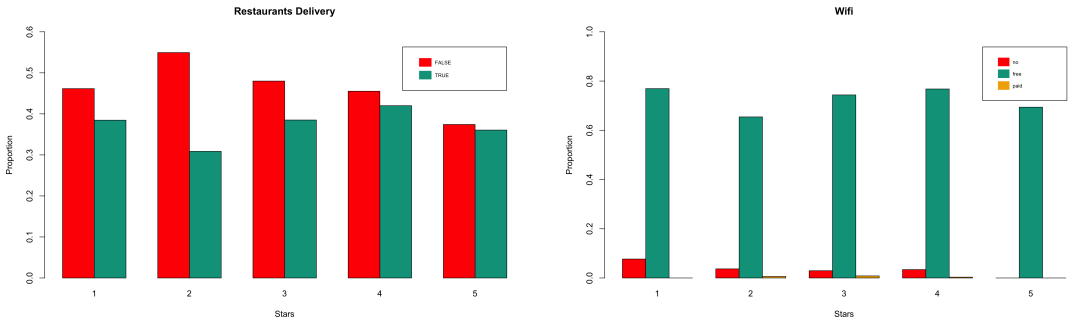
Attribute	Levels	P-value
RestaurantsDelivery	True-False	0.00646 **
OutdoorSeating	True-False	1.342e-07 ***
HasTV	True-False	5.314e-16 ***
ResstaurantReservations	True-False	0.3921
Garage	True-False	0.1597
Street	True-False	<2e-16 ***
Lot	True-False	0.06255 *
PriceRange	$-$$-$$$-$$$$	Only $-$$$$ >0.05
Valet	True-False	0.4359
Intimate	True-False	0.03744 *
Touristy	True-False	0.02825 *
Classy	True-False	1.05e-05 ***
RestaurantTakeout	True-False	0.7016
Wifi	No-Free-Paid	All p.value>0.05

2.3 Bar Charts to Compare Scaled Ratings

We plot each business attribute under different star ratings and visualize the difference between the different levels of each attribute. We combine these plots with pairwise t-tests to enhance our results. The bar charts can provide intuitive visual differences, while the t-tests can provide reliable statistical evidence. In this section, we show two examples with detailed explanations and corresponding plots.

First, we group star ratings into five levels by taking the floor of each rating. As a result, 1.5 star ratings are grouped with 1 star ratings, 2.5 star ratings are grouped with 2 star ratings, and so on. We use the proportion of businesses as the y-axis instead of the number of businesses since the number of businesses under each star is different, resulting in an imbalance in different levels under different attributes. The proportion is calculated as the number of businesses of one level (e.g. "FALSE") under a certain star rating divided by the number of all businesses under this star rating. To directly compare the proportions of each level of each attribute, we exclude the "NA" bar under each star rating.

Figure 1a shows the bar chart for ***Restaurant Delivery***. It is obvious that as the business rating increases from 2 stars to 5 stars, the proportion of restaurants without delivery services decreases while the proportion of businesses with delivery services increases. However, businesses with 1 star ratings go against this trend. The difference in average star rating between businesses with delivery services and businesses without delivery services is 0.072. Levene's Test shows the two samples have unequal variance ($p = 0.007478$). The results from Welch's t-test show that there is a statistically significant difference ($p = 0.00646$) between the average star rating between businesses with delivery services and businesses without delivery services. This suggests that if a business can provide delivery services, its star rating can greatly improve.



(a) Restaurant Delivery (b) Wi-Fi

Figure 1: Examples of Bar Charts Comparing Scaled Ratings

From the ANOVA results above, we know that there are a total of **12** significant attributes. The

other **11** attributes have test results similar to ***Restaurant Delivery***. **Figure 1b** shows the bar chart for ***Wifi***, which is not significant in the ANOVA model. The plot shows that the proportion of businesses that provide free Wi-Fi is much larger than the other two categories of businesses, and the differences among different star ratings are very small. The proportion of businesses with free Wi-Fi does not show any obvious trend, and as a result, the proportions of the other two categories of businesses do not show significant trends either. We use Tukey's HSD test to test the significance of the three paired differences, resulting in p-values of 0.471, 0.2101 and 0.497.

The results from Tukey's HSD test provide more accurate and reliable statistical analysis to support our point of view in the previous ANOVA part. Changes in attributes like *Delivery* and*Outdoor Seating* will have an important impact on the improvement of star ratings, while attribute like *Wifi* and *Take Out* will not. Therefore, we can provide some general suggestions for all businesses that can help them improve their star ratings.

3 Review Analysis

3.1 Word Selection

Using the tf-idf algorithm, we determine the three most frequent nouns mentioned in the reviews for each business. We pick nouns that are not trivial (i.e. stop words and words unrelated to cafés) and have high tf-idf scores for each individual business, where tf is the term frequency of the word in each individual review and idf is the inverse document frequency (i.e. the logarithm of the fraction of reviews containing the word) for each business.

3.2 Text Analysis

After determining the three most informative words for each café, we collect the star ratings of all reviews with these words and use a t-test to compare the star ratings with the star ratings of all reviews of the business. We set the significance level of the test to 0.1. This means that if the p-value of the t-test is greater than or equal to 0.1, then the business did better in this informative word and can maintain it. Meanwhile, if the p-value is less than 0.1, then the business needs to improve in this aspect. An example of the t-test is shown below.

Key word	Star Rating with Key Word	Star Rating of All Reviews	p-value
Breakfast	3.81	3.73	0.630
Service	3.55	3.73	0.297
Lunch	<2	3.73	<2e-16

In the example above, the business did well with regards to breakfast (p = 0.630) and service (p = 0.297), but they need to improve their lunch quality (p<2e-16).

If there are repeated words in the three words captured, and the star ratings of the corresponding reviews are completely opposite to each other, we conclude that the business's performance on this keyword is unstable and suggest that the business pay attention to the stability of quality in this aspect.

4 Suggestions and Conclusion

4.1 General Suggestions

Based on the results of the ANOVA table and bar chart comparison, we extracted several suggestions that can be changed in the attributes to yield better ratings. The general suggestions given in this section are based on the attributes of all businesses. In the Shiny app, we identify important attributes at certain levels from all cafés and give the businesses corresponding suggestions.
1. Providing delivery services can improve the average star rating by about 0.07 (p.value = 0.00646).
2. People prefer to sit outdoors outdoor when they drink coffee. Therefore, providing outdoor seating can improve the average star rating by about 0.15 (p.value = 1.342e-07).

3. People may prefer a quiet and cozy environment when having coffee. Therefore, removing TVs can improve the average star rating by about 0.21 (p.value = 5.314e-16).
4. Customers prefer street parking, so providing street parking can improve the average star rating by about 0.22 (p.value <2.2e-16). It may be hard to obtain street parking permits, but providing customers with convenient parking seems to be significant.
5. Wi-Fi, reservations, and take-out are not significant. Providing these services does not significantly affect ratings, so businesses do not need to pay attention to them.

4.2 Specific Suggestions

We provide three types of suggestions from different perspectives for each business. These suggestions are determined based on the selection of high-frequency words by the results from the tf-idf algorithm and t-test. The first type of suggestion describes what the business is doing well. We give this type of suggestion when the star ratings of reviews with a certain key word are higher than the ratings of all reviews of this businesses with a confidence of 0.1. The second type of suggestions describes what the business needs to improve. We give this type of suggestion when the the star ratings of reviews with a certain key word word is lower than the ratings of all reviews of this businesses. The third type of suggestion describes what the business needs to stabilize. We give this type of suggestion when we find that a certain key word is frequently used across all reviews and across negative reviews, indicating that some people are satisfied with this aspect with this, but others are critical of it. As a result, the business should maintain its stability with regards to that key word to ensure that every customer receives the same high level of service. We show an example of word-based suggestions below.

Suggestion 1: You are doing well in food, specifically coffee.Because comments related to coffee is not statistically lower than your average ratings.
Suggestion 2: You should offer more stable food quality, specifically chocolate.Because there are many positive and negative comments related to chocolate.

4.3 Conclusion

From the attribute analysis, we find that delivery services, outdoor seating, and street parking positively affect cafés' ratings, while TVs negatively affect ratings. From the review content analysis, we find that there are different important factors for different businesses, and that based on these factors, we can provide personalized suggestions for different types of businesses.

4.4 Shiny App

Our Shiny app can be found at this URL: `https://nvarshney2.shinyapps.io/recommendations/`

5 Strength and Weakness

5.1 Strengths

The results are easy to interpret and the recommendations are feasible. To verify the important attributes, we use statistical analysis and plots, providing credible and readable results. We also set our own reasonable word list, combined with tf-idf counting results, helping us provide customized suggestions for different business. Furthermore, the Shiny App can provide a great user experience.

5.2 Weaknesses

We do not study missing values in the attributes since we are unable to determine whether or not the corresponding businesses contain these attributes. Therefore, we are not sure about the specific relationship between these missing values and the star ratings. In addition, we use the star ratings in reviews to directly represent the attitude of the reviews containing the key words, but correlation tests should be used to determine whether the content of the reviews and their star ratings convey the same attitude.

A Contributions

Shubo Lin: Review Analysis
Tianyue Luo: Processing data and Attributes Analysis
Nilay Varshney: Shiny App
Hengrui Qu: Summary and Slides