

Introduction

Bars are an essential part in the restaurant industry and a good place for people to either kill some time and have a drink, or get some food with friends. In this project, we will focus on the reviews of bars in Madison, WI. By analyzing Yelp review information, our goal is to produce business insights to these bars, convert them into data-driven action plans and help them achieve a higher rating scores on Yelp.

Data Pre-Processing

In our analysis, the original given data includes two parts:

- *business_city.json*: basic information of restaurants such as name, city, state and type
- *review_city.json*: review information, such as stars customers gave and some comments

We started by merging *business_city.json* and *review_city.json* to create a csv file with only parameters of interest: star rating, text of comments, name of the restaurants, address, city, state, postal code, attributes and categories of restaurants.

Because our concentration is on **Bars** in **Madison, WI**, we refined the scope of our data and kept only bars in Madison. During this step, an extra filter was added by searching for “night” in the categories. This was because when we were investigating our data, many sushi bars were also included, which differed from the traditional bars for night life and alcohols.

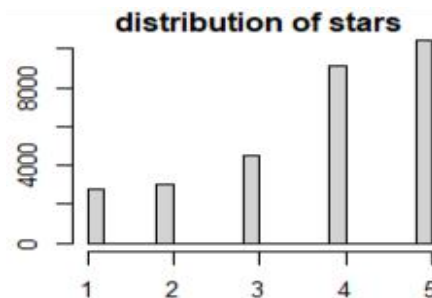
After we cleaned and combined data, we imported *nlTK* in python to deal with the review text. We lowered the text, imported *string* to get rid of punctuation, tokenized words and got rid of the *stop words*. Then we lemmatized the words (e.g. ‘kisses’ to ‘kiss’) to get the original form of every word. Lastly, we built a big arrow to store the frequencies of each word appeared and combined it with our merged data frame.

Exploratory Data Analysis (EDA)

- Analysis of data in general

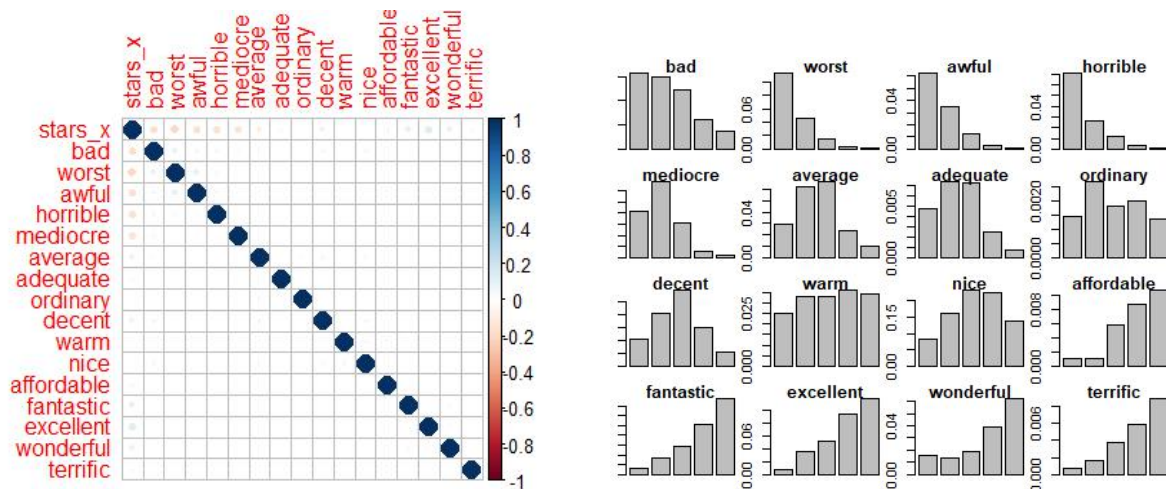
Basic information of our cleaned data and some explanation of the variables we used.

There are in total 6060 columns & 29825 rows. Each row refers to a specific review and the related information (*business_id*, *rating*, *attributes*, etc.). The first column, *stars_x*, is the rating of the corresponding review. From the 12th column to the 6060th column, they are the word arrow we built and combined during cleaning data.



Via calculation, we obtained the mean of ratings of bars in Madison is 3.717, which is slightly lower than the average ratings of service in Madison on Yelp, 3.722. And down below shows the distribution of our ratings. We can tell that people tend to give positive feedback more often than negative ones towards bars.

The plot left below shows the correlation between *stars_x* and some positive/negative words occurred in the text. And we can tell that positive words are positive correlated with higher ratings. Vice versa.

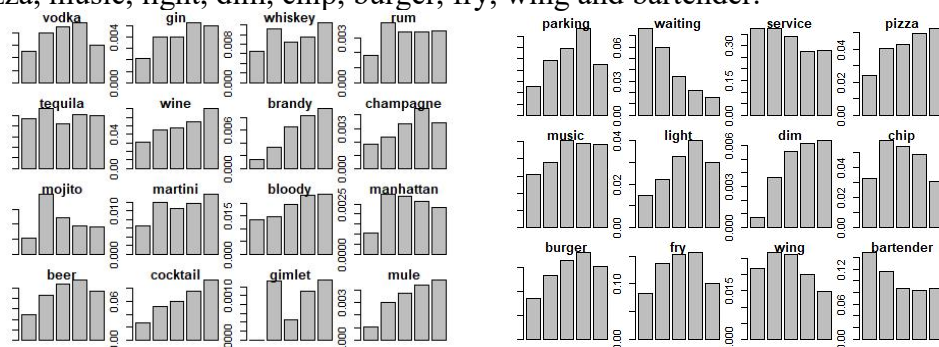


The plot right above is a sanity check. The table proves the correlation check we did in section a. Also, it shows that our data is quite reliable. However, due to the limitation of our ability, we didn't look up into cases like 'not good', which has resulted in some poor correlation with ratings. For example, 'clean' only has a correlation of 0.01355 with ratings, far less than words 'terrific', 'excellent'.

- Analysis of words related to bars

Now that our data is reliable, we can look up into if there's a relationship between the ratings of the reviews having some specific words. So, we drew a few histograms (below) to illustrate the distribution of the ratings of reviews having these words.

Choice of words included: vodka, gin, whiskey, rum, tequila, wine, brandy, champagne, mojito, martini, bloody (for 'bloody Marry'), manhattan, beer, cocktail, gimlet, mule, parking, waiting, service, pizza, music, light, dim, chip, burger, fry, wing and bartender.



Words like "waiting" and "bartender" shows an obvious pattern that more negative reviews are given using this word, whereas words like "wine", "brandy" and "cocktail" have a positive association with the reviews and stars. Besides, "tequila", "vodka" have shown no obvious relationship with ratings.

Analysis of Bars in Madison on Yelp

As shown in the previous step, we have discovered some words whose frequencies in reviews were positively, negatively or hardly associated with the ratings of bars. To check if our intuition from previous step provides significant statistical meaning, we performed t-test and compared if there exists difference between the mean of ratings with and without a word the reviews.

After we performed t-test on all the words mentioned above, some showed statistically significant difference (p-value lower than 0.05 and Confidence Interval do not include 0), others showed no big difference (p-value quite large). We have converted the test results into tables below for more straightforward understanding.

- Words with positive association: reviews mentioning these words have higher stars

word	average star without word	average star with word	difference (with-without)	p-value
brandy	3.71	4.17	0.45	5e-11
gimlet	3.72	4.13	0.42	0.036
cocktail	3.69	4.08	0.39	2.2e-16
dim	3.71	4.08	0.37	5e-06
wine	3.7	3.99	0.29	2.2e-16
pizza	3.71	3.93	0.23	9.26e-13
beer	3.67	3.85	0.18	2.2e-16
whiskey	3.71	3.9	0.18	0.01
burger	3.69	3.86	0.17	2.2e-16
music	3.71	3.85	0.14	1.9e-05

- Words with negative association: reviews mentioning these words have lower stars

word	star without this word	star with this word	difference (with-without)	p-value
waiting	3.74	2.91	-0.83	2.2e-16
discount	3.72	3.08	-0.64	6.273e-5
bartender	3.74	3.5	-0.24	2.2e-16
wing	3.72	3.48	-0.24	8.14e-06
service	3.78	3.56	-0.22	2.2e-16
chip	3.72	3.57	-0.15	1.7e-05

- Words with little association

word	star without this word	star with this word	difference (with-without)	p-value
tequila	3.72	3.73	0.01	0.92
vodka	3.72	3.70	-0.02	0.85
mojito	3.72	3.60	-0.12	0.45

With the p-values given, we can infer that many reviewers have positive opinions about brandy (4.17 star), cocktail (4.08) and gimlet (4.13) as we assumed, and negative opinions about long waiting time (2.91), discount (3.08) and bartender service (3.5). Also, we have noticed that, test regarding vodka & tequila have p-values 0.85 & 0.92 respectively, indicating they have little influence on the rating. In addition, cocktail positively influenced ratings, while mojito, as a category of cocktail, slightly negatively influenced the ratings.

In the food section, pizza & burgers skew positively, with stars 3.93 & 3.86 respectively, whereas wings & chips skew negatively, with stars 3.48 & 3.57 respectively. Besides, reviews with “dim” tend to have an average increase of 0.37 in ratings and have a p-value of 5e-06.

Recommendations for Businesses

We have discovered the positive/negative association between the rating of a bar and words used by Yelp reviewers. Therefore, we made some recommendations to improve ratings.

- Alcohol/Beverage

Importing more brandy, cocktail and gimlet than other alcohol, but import less mojito; reviews with brandy and gimlet have 0.4 more stars (p-value<<0.05), mojito has 0.12 less.

- Service

Train your bartenders; reviews regarding bartenders have 0.24 less stars than without it (p-value<<0.05).

- Food

Serve pizza & burgers more, chips & wings less; burgers & pizza increase rating by 0.2 star, while chips & wings decrease by 0.2 approximately.

- Ambiance

Dim atmosphere makes the rating higher; reviews mentioning dim have an average of 0.2 more rating than without it.

- Others

Set more tables and give out one free drink per customer for those waiting more than 30 mins; reviews mentioning waiting has 0.83 LESS star than without it (p-value<<0.05).

Give out better discount to regular customers; Surprisingly, reviews with discount have 0.64 less star than without it (p-value<<0.05). Either no discounts or little ones.

Conclusion/Discussion:

In our study, we discussed the relationship between the word frequencies of customers' reviews, and the ratings on Yelp. We found that some words, like brandy and cocktail, show up more frequently in positive reviews with high ratings, whereas words like waiting, service and discount, are mentioned more in the negative reviews with low ratings. Still, there are words like tequila and vodka which haven't shown much difference in the ratings.

We then examined the difference of average ratings between reviews using a certain word and those without this word. Finally, by performing t-test, we came out with several words that can make statistically significant difference in ratings, and little or no difference in ratings. Lastly, we proposed suggestions for improvement in different aspects, such as service and food quality.

Contribution

Summary: JX started and built the frame, worked on Intro, Data Cleaning, EDA, Analysis, Recommendation and Conclusion; HT worked on Data Pre-processing & cleaning part, helped with Recommendation & Table-building part.

Presentation: JX created the PowerPoint and performed page 9-12; HT performed page 5-8; SL performed page 1-4.

Github: JX organized the repository, updated code and readme files; HT wrote step1 & step3 codes, and helped with step4 code. SL drew some plots.

Shiny: JX helped with one function, and finalized the app; SL helped build app; HT helped check the app.