Executive Summary

1 Introduction

Yelp is a widely used public crowd-sourced website for customers to leave reviews about a business. In this project, we focus on what makes a bar successful by looking at attributes and what to improve by sentiment analysis and relevance of bigrams.

2 Data Pre-Processing

We chose to focus only on the business.json, reviews.json and the tip.json files provided by Yelp. Our first task was to filter through the business category dictionary list to select only the entities that were listed as cocktail, wine, whiskey, champagne and beer bars, which left us with 163 bars. We used this result to filter out unwanted businesses from the tip and review data and then kept only those located in California. For the review and tip datasets standard NLP practices were used to mine through the text using Python and R, respectively (i.e. tokenize words, remove numbers, punctuation, and stop words). To prep the attributes, we used Python to destruct the dictionary list and convert them to a dataframe. We then used R and Excel to make up the missing values randomly, since removing the rows will reduce the data too much.

3 Exploratory Data Analysis

3.1 Exploratory Data Analysis for Attributes

Our goal is to analyze the relationship between attributes and the rating to provide some insight as to what a successful bar looks like. Almost 130 bars out of 161 bars got ratings higher than 4, which is hard to find the difference of impact between the choices for each attribute. However, we still used boxplot to illustrate some obvious differences of some attributes. We notice from a box plot that for NoiseLevel when “loud”, results would probably lead to a lower distribution of ratings but when “very loud” there is a high rating and for WiFi boxplot, paid WiFi would be given a lower rating opposed to free or even no WIFI.

3.2 Exploratory Data Analysis for Review data

After data prep, we counted words for overall data and for each rating to initially investigate some interesting words. For sanity check, we also plot histograms of relevant words such as bad, worst, average and recommend against the rating to diagnose whether our NLP algorithm worked accurately. Lastly, we ran the word clouds both for overall and for each star. The primary key takeaways we obtained from word clouds are 1) wine and service have the highest mentions and 2) most negative reviews are related to table and ordering. These ideas are collected as assumptions for our further analysis.

3.3 Exploratory Data Analysis for Tips data

Our intuition was to run a sentiment analysis on the tip data as this is a common practice with text data, but the results of the preliminary sentiment analysis were uninteresting so we decided to analyze bigrams instead. Observed from a histogram of the top 20 most frequent bigrams, happy hour is the most mentioned phrase, but this result is not too surprising since our focal point is bars. For clarification, happy hour is a period of time where drinks from the bar are offered at a reduced price. This result may, however, hint that happy hour has high relevance to the rating a business receives. An example of a tip review containing happy hour is "New happy hour specials. Dropped the margarita for a seasonal sangria. Very nice change".

4 Results from Analysis

4.1 Attribute Analysis

We decide to use a linear regression model to explain the effect of the attributes. We began with fitting models with all attributes with no interaction terms and found the significant attributes. We decided to add interaction terms to improve the low R-squared value. Furthermore, we used ANOVA to determine whether new terms should be added into the model, then used a log transformation of the stars to normalize our data and improve the model.

$$rating = 1.48 + 0.08Validated - 0.08Lunch - 0.07Alcohol - 0.15hipster - 0.16(Valet * Lunch) - 0.18(Touristy * Karaoke) + 0.16(Touristy * Alcohol) + 0.14(Intimate * Hipster) - 0.17(Intimate * Street) + 0.1(Lunch * Hipster) + 0.09(Lunch * Alcohol)$$

Because of the cross-over interaction, some main effects are not significant but the interactions between them are significant. Take valet and lunch, in this case there is no overall effect of valet, but there are cross-over interactions between the two attributes. There are still some limitations in our linear model, like the R square is low at 33.95% and this model included too many cross-over interaction terms. However, this is probably a result from the original data which included only True and False for most attributes.

4.2 Review Results

Since the review data reveals how each customer feels when they visit a bar, it is practical to investigate what people think about the bar and to analyze their sentiments. The results for review analysis are divided into 4 parts:

4.2.1 Top keyword mentions in review

First and foremost, we would like to answer the question: what are customer reviews mentioning the most? To do this, we start obtaining nouns and sum up the total number of each word in the review. We also categorize them and group by each specific bar for more individual information. The interesting result is that most customers were mentioning a lot about food and service although these bars mainly offer alcohol. The top 10 most mentioned keywords include : *Food, Service, Wine, Staff, Drink, Price, Table, Beer, Seafood, Burger*

4.2.2 Using classification technique to classify positivity and negativity of adjectives

After we get the final review data set, we start obtaining adjectives using the NLTK package, and create a frequency matrix. There are approximately 18,000 adjectives from reviews. We continue classifying stars to positive and negative classes based on the response variable "stars" where we flag "positive" for 4 star reviews or more, otherwise, "negative" and then fit the model using the **Multinomial Naive Bayes Classification technique**. The example below is the prediction result which appears to be accurate and logical.

{reasonable': 'positive', 'friendly': 'positive', 'open': 'positive', 'grand': 'positive', 'cold': 'positive', 'not_disappointed': 'positive', 'perfect': 'positive', 'upset': 'negative'}

4.2.3 Obtaining useful nouns

We start by collecting all nouns in the review regardless of what bars they are referring to, then count the nouns and select the 100 most common words. It is also necessary to combine reviews for each individual business. Moreover, it is useful to select only words/nouns that are informative and can represent the nature of bar business. The following nouns are nouns which will be used for further analysis.

['food','drink','price','place','burger','seafood','salad','clam','chowder','crab','chicken','bread','chip','appetizer','taco','oyster','steak','pasta','lobster','wine','beer','cocktail','water','staff','table','reservation','waiter','bartender','waitress']

4.2.4 Sentiment analysis

We analyze the sentiment by creating a function which counts the number of positivity and negativity of each noun from the prediction of a sentiment based on the adjectives. The sentiment score is technically in the following expression:

$$\text{Sentiment score (\%)} = 100\% \times \frac{positive\ count}{positive\ count + negative\ count}$$

The following table for sentiment analysis is collected and grouped by business_id and the mentioned nouns (keyword in section 4.2.1).

	business_id	stars	positive_count	negative_count	sentiment_score_%	keyword	group
0	-BdYhP-12elmFV7oB1iv4A	5.0	7	0	100.000000	food	food
1	-e8RwknT5szoLk9uBZjzcw	4.0	1	0	100.000000	food	food
2	-ujBP1Dw0j1-Ffaz97-LXQ	4.0	33	1	97.058824	food	food

To solve this problem of NaN values in the data, we can calculate the average sentiment from the nouns under that category. For example, if the sentiment score for "drink" is empty for "abc" bar, we instead calculate the average of sentiment score from the words such as “beer”, “wine” and “water” etc. The following lists demonstrate how we group each word to the main categories:

Food: burger, seafood, salad, clam, chowder, taco, oysters, steak, pasta, lobster
Drink: wine, beer, cocktail, water
Service: staff, table, reservation, waiter, waitress, bartender

After fixing the issue, we can observe that most keywords extracted from reviews tend to be strongly positive with a score higher than 90, while, interestingly, some keywords related to service appear to have relatively lower sentiment scores. Therefore, we reach the conclusion that most negative reviews and sentiments are usually caused by service compared to other high level categories. Additionally, we diagnose the behavior of how keywords react to the rating by performing Pearson’s correlation test. It turns out that among 29 keywords, there are some keywords that indicate positive correlation, at a significance level of 0.05, to the rating which are **food, drink, service, burger, beer, staff, table, waiter**.

4.3 Tip Results

To measure relevance, we calculated the tf-idf of each bigram then ordered them from highest to lowest. Motivation behind using this metric was dual-purpose, one to quantify the relevance of the phrases but also to standardize word frequency. Upon doing this, we found that, much to our surprise, happy hour was not amongst the top relevant phrases. There are limitations to this process, such as, when creating the bigrams, this is an automated process by R so some phrases appear to be confusing and not aligned properly. For example, the review could read “I love their appetizers. It’s so cozy inside” but in the process of creating bigrams and removing stop words the bigram that was created was “appetizers cozy” instead of “love appetizers”, “cozy inside”. We chose to keep only bigrams that were given a 4, 4.5, or 5 rating since we want to highlight the positive things about the bar that already exist.

5 Recommendations and Conclusion

From the results of the model, we found that ambience, business parking, alcohol, and good for lunch are the most effective factors determining the rating scores. We suggest business owning or renting validated parking, valet, street parking, having a full bar alcohol, preparing wonderful menus for lunch, making intimate, touristy, and hipster ambience. We chose to analyze bigrams for the tips to inform the business what people already like about their bar. Without running an analysis, we know one of the best ways to bring in business is by marketing. The business owners will be able to evaluate the most relevant words for their business to understand what they should be marketing to the public, bringing in more business. Overall, people like a cozy atmosphere, live music and a wide range of food selection. Through our analysis of the review, tip, and business data from Yelp we were able to provide data driven suggestions to business owners and evaluate what makes a bar successful.

7 Contributions

AT: Merged data, analyzed tip data with bigrams, wrote all sections on tip analysis in executive summary and presentation, introduction and conclusion of executive summary, editing of summary, and shiny app

SS: Filtered the original datasets, conducted sentiment analysis of review text, wrote all sections of sentiment analysis in executive summary and presentation, transferred Google Slides to powerpoint software, and shiny app

JL: Deconstructed dictionary list to data frame, finished the data analysis of the new dataset, found an appropriate model to explain the data, wrote attributes section in summary, and shiny app.