Statistics 628 Module 3 Group 10 Executive Summary

Brian Tsai, Tinghui Xu, Lanxin Xiang

December 9, 2021

1. Introduction

In this project we used datasets from Yelp to try to extract business insights. We chose to focus on one of the largest Mexican restaurant chains in the US: Chipotle. The goal of the project is to analyze the reviews and ratings for the restaurants and try to find out the aspects that affect the ratings and what a business owner can do to improve their ratings in the future.

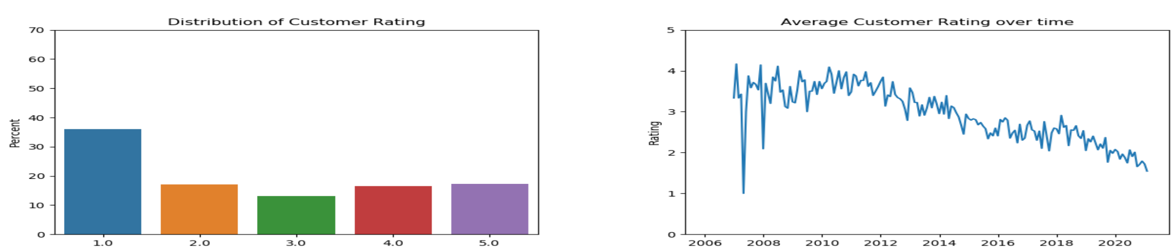
2. Data Cleaning/Filtering

For the data cleaning and filtering part, we first used the business.json file to filter out the business ids that belong to Chipotle restaurants.This was done by first filtering out Mexican restaurants that are open, then using the store name 'Chipotle', we filtered out the businesses that belong to this chain. We then narrowed it down by only looking at restaurants that have 10 or more reviews because we thought less than that, it wouldn't be worth analyzing. In the end, we extracted 173 Chipotle restaurants from the file. Next, we use the review.json file and extract the reviews for the restaurants that we extracted previously. This was done by filtering the reviews according to the business id of the restaurants that we extracted previously. Next, we discarded the reviews that aren't in English by using the langdetect module in Python. Finally, we obtained 12696 reviews for the 173 Chipotle restaurants.

3. Exploratory Data Analysis (EDA)

For EDA, we wanted to explore how the ratings Chipotle restaurants changed over time and also what the general distribution of the ratings are like, as well as look at word clouds to see what words get mentioned more in the reviews. We also conducted sentiment analysis of the reviews to see general trend of the sentiment from the reviews.

3.1 Ratings



(a) Chipotle Ratings Distribution

(b) Chipotle Avg. Rating over time

Figure 1: Chipotle Ratings EDA

We first looked at the distribution of ratings for Chipotle as a whole (Figure 1a). We see that most of the reviews are negative, with more than 50% of the reviews having only 1 or 2 stars, and only around 30% of the reviews having positive 4 or 5 stars rating. Next, we looked at the time series plot of the average ratings over time (Figure 1b). From the time series plot, we see that the average ratings used to be quite high at around 4, but starting from around 2012, there's been a consistent decreasing trend in the average ratings, and now the average rating is at around 2 stars, and could be expected to drop further in the future. This is a very worrying trend for Chipotle.

3.2 Word Clouds

(a) Chipotle Nouns Word Cloud (b) Chipotle Adj. Word Cloud

Figure 2: Chipotle Word Cloud

We made 2 word clouds, one for nouns and another for adjectives. This was done by merging all reviews text into one big chunk of text, converting to lower case, removing stop words, then filtering text according to noun and adjective. The word clouds give us a basic idea of what gets mentioned the most in reviews. For nouns, we see words such as food, time, order, location, and for adjectives, we see fresh, good, bad, etc... Based on the results, we can then look deeper into the reviews and see why these words are coming up so frequently.

3.3 Preliminary Sentiment Analysis

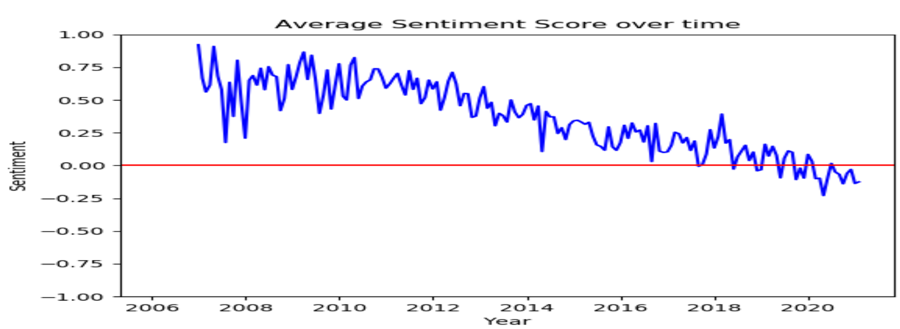


Figure 3: Chipotle Sentiment Score over time

Sentiment analysis of review text was done using VADER from nltk module. The basic principle for this model is each word has a 'valence' score ranging from -4 to 4, and words with greater intensity would have a higher score. For example, 'okay' is 0.9 whereas 'great' would be 3.1. To calculate the score for a sentence, we used the 'compound' score method, where the scores of each word is summed up then normalized to a value between -1 and 1, so positive reviews would have a positive score, and negative reviews would have a negative score in principle. Applying this on all the reviews, we get the average sentiment score from the reviews over time (Figure 3). We see that the general trend matches what we saw previously for the average ratings over time.

4. Aspect based Sentiment Analysis

The natural next step after the preliminary sentiment analysis is to single out specific aspects from the reviews and try calculating sentiment scores for each aspect for each individual restaurant, and the goal is to give recommendations based on the scores of these aspects. To identify the topics, we used Latent Dirichlet Allocation.

4.1 Latent Dirichlet Allocation(LDA)

To identify important aspects of the reviews, we used Latent Dirichlet Allocation(LDA) to find different topics and corresponding word frequencies under each topic. We then figure out the aspects by looking at the top 30 words under each topic in the output of LDA. For example, one unlabeled topic output from LDA may include words such as: guac, salad, cheese, burrito, etc... as the most frequent words. From the words, we can infer that the aspect is food. Going through the same process for the other topic outputs, we determined 4 important aspects of the reviews are: Food, Service, Ambience, and Price, hence these are the aspects we will be focusing our subsequent analysis on.

4.2 Relevant Words for Each Aspect

Our next step is to assign relevant words to each aspect, and we will be calculating aspect scores by using these words. To determine what words to allocate to each aspect, we referred to the results of LDA from earlier as well as our own judgement. For example, some words associated with the Food aspects are: guac, salad, cheese, burrito, fresh, ingredients... ;words for the Service aspect are: attitude, wait, management, pick-up...; words for the Ambience aspects are: environment, clean, seat, wifi...; and for Price: cheap, expensive, cost, refund, etc... So now for each aspect, we have a list of words associated with it. Next we calculate a score to each individual 'word' under an aspect according to the reviews.

4.3 Calculating Score for a Relevant Word

To calculate the score, we used the VADER sentiment score model from before applied to cleaned up text that has been converted to lowercase and with stopwords removed. Basically, each review is made up of many segments, and if a relevant 'word' is mentioned in a segment, we calculate the sentiment score for that segment. The score for that 'word' in that review would then be the average score of all segments that mentioned the 'word'. If the 'word' isn't mentioned at all in a review, then the score for that 'word' is assigned 0. Finally, the overall score for the 'word' would be the average score over all reviews with non-zero scores. What we end up with is a overall score for each relevant word associated with an aspect. An example of the output after this process for one business is shown in figure 4. The next step is then to actually calculate the overall score for an aspect.

business_id	black	guac	amount	...	door	table	...	service	time	...	price	free	...
__a6cve2cpTdOP6uNXhRMw	0	0.444	0.359	...	-0.024	-0.340	...	-0.089	0.137	...	0	0.462	...

Figure 4: Word scores

4.4 Calculate Aspect Score

To calculate the score for a particular aspect, we sum up the weighted scores of all the relevant words under that aspect. The scores for each word under that aspect is weighted according to how frequent they are mentioned in the reviews. Suppose for an aspect, we have relevant words a_i, each with an overall score x_i, with frequency f_i, where frequency is defined as the number of non-zero scores word a_i has over all reviews, then the overall aspect score is $S = \sum x_i \cdot \frac{f_i}{F}$, where $F = \sum f_i$. Using this weighting system, we can prevent the scenario where a word with very positive score but appears rarely in reviews would have too much influence on the overall word score or vice versa with a negative word.

business_id	Food score	Ambiance score	Service score	Price score
__a6cve2cpTdOP6uNXhRMw	0.174	-0.127	0.072	0.462

Figure 5: Aspect scores

5. Example of Recommendation based on Aspect Scores

To give suggestion for a specific restaurant, we compared their aspect scores against the 75% quantile and mean of the aspect score of all restaurants. If their aspect score was higher than the 75% quantile score, we would deem the restaurant to be doing really well in this area, and if their aspect score was lower than the mean, then we would say this aspect is lacking and give suggestions to improve this aspect. In the Shiny App, if you click on a specific aspect and a specific relevant word under it, you can see reviews that is associated with it, and business owners would be able to read those reviews and decide what actions to take based on it.

6. Strengths and Weaknesses

The strength in our approach is that we are able to explicitly calculate scores for 4 main aspects of reviews. We are able to then make comparisons between different Chipotle restaurants and give suggestions based on how one restaurant is performing compared to overall performance of all restaurants. We are also able to give specialized advice for each aspect for each restaurant. Apart from knowing how frequently a relevant word under an aspect appears in the reviews, we also obtained a specific score for that word, and from the score we can see if that particular word is positively or negatively contributing to the overall score for an aspect as well as the magnitude of contribution. The main weakness in our approach is that the aspect scores are calculated from list of words that we deemed to be relevant based on LDA output and our own judgement. It is possible that a a relevant word was not selected and may be missing from the list, and that would impact the calculation for the scores. Also, it is possible that a relevant word can be part of 2 aspects depending on the context. Another possible issue is we may not be able to accurately capture relationship between different segments. For example, one segment may have mentioned a food item, but the next segment may mentioned it was bad, then the score for that food item wouldn't take into account the adjective 'bad' which was part of the different segment.

7. Conclusion

In conclusion, we were able to extract 4 main aspects from the reviews, and give tailored advice to each restaurant based on the scores for the aspects, as well as the scores for relevant words under each aspect. A business owner would be able to see how each aspect of his/her restaurant is performing compared to other Chipotle restaurants and look at specific reviews that associated with a relevant word under an aspect. We expect that by using the Shiny App, a business owner would be able to make a business plan that can improve aspects that are lacking, and that in turn will help raise future Yelp review ratings for the restaurant.

8. Contributions

Brian Tsai: Data cleaning, EDA, slides, and report;
Tinghui Xu: LDA, slides, and Shiny app development;
Lanxin Xiang: Aspect based Sentiment analysis, slides

References

[1] Ben Chamblee. Building a sentiment analysis model using yelp reviews and ml ensemble methods. https://towardsdatascience.com/building-a-sentiment-analysis-model-using-yelp-reviews-and-ml-ensemble-methods-80e45 Oct 2021.

[2] Kaimenluo. kaimenluo/customerreviewsummarizesystem: 用户评价情感分析系统. https://github.com/kaimenluo/CustomerReviewSummarizeSystem.

[3] Elva Xiao. Text mining and sentiment analysis for yelp reviews of a burger chain. https://towardsdatascience.com/text-mining-and-sentiment-analysis-for-yelp-reviews-of-a-burger-chain-6d3bcfcab17b, Jan 2021.