INSY 669 Text Analytics

Final Group Project

Google Review Analysis for Market Success: Understanding the Café Industry in Toronto

Dr. Lavisha Jindal

March 1st, 2023

Ke(Kristen) Chen

Shanshan Lao

Xinran(Iris) Liu

Yang(Olivia) Ma

Duoyi(Flora) Zhang

McGill University

1. Motivation

Thorough market research is crucial for starting a successful café in Toronto. The café industry is highly competitive, so identifying competitors' strengths and weaknesses can help determine the café's unique value proposition. Understanding the target market's demographics and preferences is important for tailoring offerings. Gathering customer feedback through surveys, focus groups, or online reviews can help design a café that appeals to the target audience. Overall, market research can help differentiate the café and meet customer needs for long-term success in Toronto.

2. Problem Statement

In this case, our team has been hired as data consultants to help a café identify a location that best suits their products. The two available location options are Queen St. E. and Bay St. This café not only provides coffee but also delicious food for breakfast and lunch. The owner is confident about their range of food options and believes it to be their core competitive advantage. However, due to funding limitations, the owner needs to decide between the interior decoration of the café or the number of employees they can hire at the early stage to ensure each customer gets a better experience. After solving these two questions, they also want any other suggestions we can provide to help them succeed.

3. Data Acquisition

In order to gather information on the café industry in Toronto, we used a web scraping technique to extract reviews from Google Maps for 6 cafes each located in two different areas: Queen St E. (near the beaches), and Bay St. (in the Financial District). We created a function called "getReviews" which takes in the name and street of the café, as well as its Google Maps URL. This function utilizes the Selenium package in Python to simulate a web browser and interact with the Google Maps interface to scroll through reviews and extract relevant data such as the user's name, date of review, star rating, and review text. We then stored this data in a Pandas DataFrame for analysis. By using this method, we were able to collect 3262 reviews on 12 cafes in Toronto, which can provide valuable insights for market research and analysis.

4. Review Analysis

The WordCloud Analysis aims to showcase the most commonly repeated words in certain cases, helping restaurateurs comprehend which attributes were most significant to customers based on their Google Map comments. Figure 1 and 2 are the most frequently utilized words that describe the features customers appreciate in Good* and Bad* restaurants, which have a compound rate over and below 0.

To help choosing between locations, Figure 3 and 4 are the most frequently utilized words that describe the features customers appreciate in Good* and Bad* restaurants in Queen St E. area; Figure 5 and 6 are the most frequently utilized words that describe the features customers appreciate in Good* and Bad* restaurants in the Bay St. area.

Furthermore, to further help our client to understand customers' feedback and needs we constructed a logistic regression model and a Naive Bayes model for classification. The results show that the most informative words for positive reviews are: amazing, downtown, delicious, awesome, friendly. And the most informative words for negative reviews are: server, overprice, use, salad and bad.

5. Business Insight

The results of WordClouds show that, generally, customers in Queen St. E pay more attention to food. The words 'food' and other words related to food are mentioned very frequently in both the positive and negative reviews, as shown in Figures 3 and 4. In comparison, they are hardly mentioned in the Bay St. area, as shown in Figures 5 and 6. This makes sense, as most customers in Bay St. are employees working in the surrounding area, thus having a shorter time available for coffee and even less time for food. Whereas those near the beach usually have more time to sit down and enjoy not only coffee but also a good meal. Therefore, since our client defines 'the range of food options' as their main competitive advantage, we recommend they open their cafe in Queen St. E. In addition, the word 'food' is mentioned relatively more frequently in the negative reviews than in the positive reviews, which means that customers in that area are not greatly satisfied with the food currently available. Thus, our client can take this as an advantage to outperform the other cafes in the nearby area.

In terms of deciding between better decoration or better service in the early stage, we suggest the owner allocate more funds to hiring more employees to provide better customer experience. Words related to interior decoration are not mentioned frequently in either the Queen St. E or Bay St. area, whereas keywords like 'service' and 'staff' appear constantly in the reviews. Especially in Queen St. E, 'service' is one of the most frequent words in the negative comments. Thus, if our client can ensure good service, they are likely to out-compete their close competitors.

According to our analytical results, we have two other suggestions. Firstly, breakfast is mentioned more than lunch; therefore, our client should focus on improving their breakfast cuisines to attract more customers. Some recipe suggestions include cheese, eggs, and salad, as these are mentioned frequently in the word cloud as well as in the classification models. Secondly, the word 'time' is mentioned a lot in the negative comments in the Queen St. E area; this could likely mean that people in the nearby area have lots of complaints about the waiting time. Our client should focus on coming up with a strategy to ensure a shorter waiting time.

6. Future Work

Expanding the scrapped cafe list to include multiple regions can provide valuable insights into the best locations to open a cafe and what kind of services or products are most in demand. By analyzing data from a variety of regions, cafe owners can identify patterns in customer preferences, such as popular menu items, preferred ambience, and preferred price points.

Additionally, we would like to enhance our model's predictive capabilities and provide deeper insights by integrating time-series data. We aim to leverage real-time streaming data to analyze social media sentiment related to the restaurant industry.These information can be used to develop a cafe concept that is tailored to the target market's preferences and increase the chances of success.

7. Summary

This report details the market research conducted by our team of data consultants to help a café identify a suitable location in Toronto. To conduct analysis, we utilized two locations as demonstration. We first web-scraped to collect reviews from Google Maps and conducted sentiment analysis and classification models. Then,

we visualise the results using WordCloud. Our findings revealed that people in Queen St. East paid more attention to food and were not highly satisfied with the food they were getting. It was recommended that the café open in this location and focus on improving the breakfast cuisines and the waiting time. For future endeavors, our team suggested that the café integrate time-series data and social media sentiment to further enhance their predictive capabilities.

8. Appendices

Figure 1. WordCloud of most frequently mentioned in the positive reviews

Figure 2. WordCloud of most frequently mentioned in the negative reviews

Figure 3. WordCloud of most frequently mentioned in the positive reviews in Queen St. E

Figure 4. WordCloud of most frequently mentioned in the negative reviews in Queen St. E

Figure 5. WordCloud of most frequently mentioned in the positive reviews in Bay St.

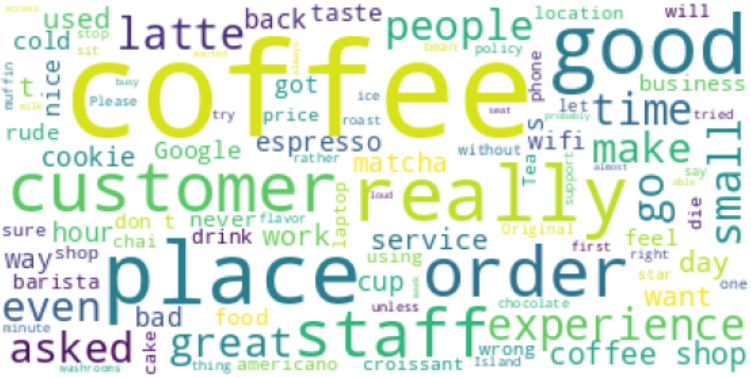

Figure 6. WordCloud of most frequently mentioned in the negative reviews in Bay St.