**Project:** Customer review with sentimental analysis of Thai restaurants in Melbourne case from TripAdvisor

**Project Description**

As a metropolitan city, Melbourne has multi-national cuisine for customers to prefer. Thai food is also one of the favorite cuisines in the city. Not only in Melbourne, but Thai taste has also been favorable by Australians with its unique balance of sweet, spicy, and sour (*Redlemonthai Why do Australians love Thai food so much?*). In turn, there is a plethora of Thai restaurants in the CBD and the suburbs. Even though, Thai cuisine is famous and popular in Melbourne. However, there is minimal to none about customer review sentimental analysis in the Thai restaurant case. To the extent, there is no project that investigated specially in Melbourne context.

This project aims to perform the sentimental analysis on the customer review from all Melbourne-based Thai restaurants to understand the voice and trends of the people in Melbourne and expect to improve restaurants’ capability especially in customer service domain. This project has objectives as follows.

**Project Objectives**

* **Generalized Trends:** This project’s bottom-line is the need to deeply understand the majority trend of people in Melbourne regarding to Thai restaurants & cuisine and their feeling over the period to foresee the trend.
* **Customer voices & experiences:** This project expect to analyze the customer review and using sentimental analysis to extract feeling and experience towards Thai restaurants. To understand customer voice in which added value to the restaurant type business that relies on customer relationship.
* **Constructive Feedback:** This project wants to extract constructive feedback from positive and negative words associated that helpful not limited to specific restaurants but as the whole within Melbourne Metropolitan. This should pinpoint what are they currently have competitive advantage and what perceived as a weakness in which related to SWOT framework.

**Business Model**

This project aims to give insight into Thai restaurant owners. It will strengthen the understanding of the customers’ needs. Plus, addressing the improvement part on foods or services. Also, it can provide know-how on the decision-making part regarding the restaurant managerial approach. For instance, it may help the owners to desire which menu they should sell in bundles or upsell their target dish more.

Thus, this project aims to benefit the customers by knowing the key strengths of the restaurant and ensuring they have a good customer experience. In addition, it will give a key takeaway for those who are going to open Thai restaurants on how to continuously keep customer retention once they have opened the shop. **Project Challenge**

The challenging part of the project is the data source. The data that is available in TripAdvisor is expected to be massive as the project scope is the whole Melbourne Metropolitan area. Also, the customer review may be available in other languages like Thai. To overcome the complexity, the NLP (Neuro-linguistic programming) needs more optimization to capture and train to take a sentiment analysis from the raw data source.

A diagram of a diagram

Description automatically generated**Data Model**

**Characteristic and Analyzing Data**

**Data Source**

The data aimed to be obtained from TripAdvisor.au. This can be seen by filtering the data by setting it as “Restaurants in Melbourne, VIC”. Additionally, set the cuisine into Thai cuisine to achieve the target user. The required data metadata are the Restaurant’s business details and customer reviews.

**Data Type**

The data types in this project are expected to be two types: String and Int. Which, the Business details are a mixture of these two types. Whereas the customer reviews tend to be string type only that requires cleaning and extraction. The data is coming from CSV that has been collected the data from web scraping in which the schema like the data type has been already identified.

**Volume**

TripAdvisor has contained a massive amount of data of customer review. According to TripAdvisor press, the cumulative of TripAdvisor’s customer review up to 2022 has reached one billion records (*Travelers push TripAdvisor past 1 billion reviews & opinions!*). However, the scope of the project is relatively small and focused. The volume of the dataset is achievable and considered as a static dataset since the data has been collected to the present date and compiled within the project life cycle. The dataset of the customer review is massive but scalable as it resulted in 22859 records.

**Velocity**

The data in this project is considered as finite velocity as it has been collected up to date point rather than query for the real-time streaming. Likewise, the complexity of data obtaining has been reduced as it needs to deal with static data. However, the consideration regarding the velocity still mattered, as the model from the data analysis can be applied to real-time streaming if it has been implemented on the data streaming like Kafka or Cassandra.

**Variety**

The data has been considered as low variety since it was obtained only from one source: TripAdvisor. Thereby, the customer reviews are varied since it is a mixture of all Thai restaurants in Melbourne comment. In which, it is difficult to handle. As such, the data must undergo several processes prior to the data usage.

**Veracity**

The data from customer reviews has a high degree of veracity. For instance, the customer review might be biased or the review that has been posted might be affected by the negative experience that they have faced during their visit rather than good experience. Thus, some of the negative customer reviews might be considered as not essential as they lacked constructive feedback rather than intrinsic motivation. To identify and clean, it will filter out or minimize the degree of veracity of the available dataset.

**Visualization**

The data can be used from descriptive statistics to show simple trends like average score (mean), the number of restaurants, and the number of restaurants in each suburb. For the customer's reviews, it can be shown in the sentimental analysis like the ratio between positive, negative, and neutral experiences. Also, it might be compiled into a word cloud to see the most repetitive trend. To the extent, it might visualize with the machine learning like decision tree model to classify the hidden trend among the customer trends that represent the customer voice which is one of the business's added values.

**Data processing**

**Storage**

As a best practice, the data should be retained in the database. For the best option to store this kind of data, NoSQL like MongoDB can contain the data as granular and abide by ACID principles would the case. In Which, it can compress massive data into small sizes with JSON files. For the sake of simplicity, the data will be stored and used in the form of a CSV file.

**Software**

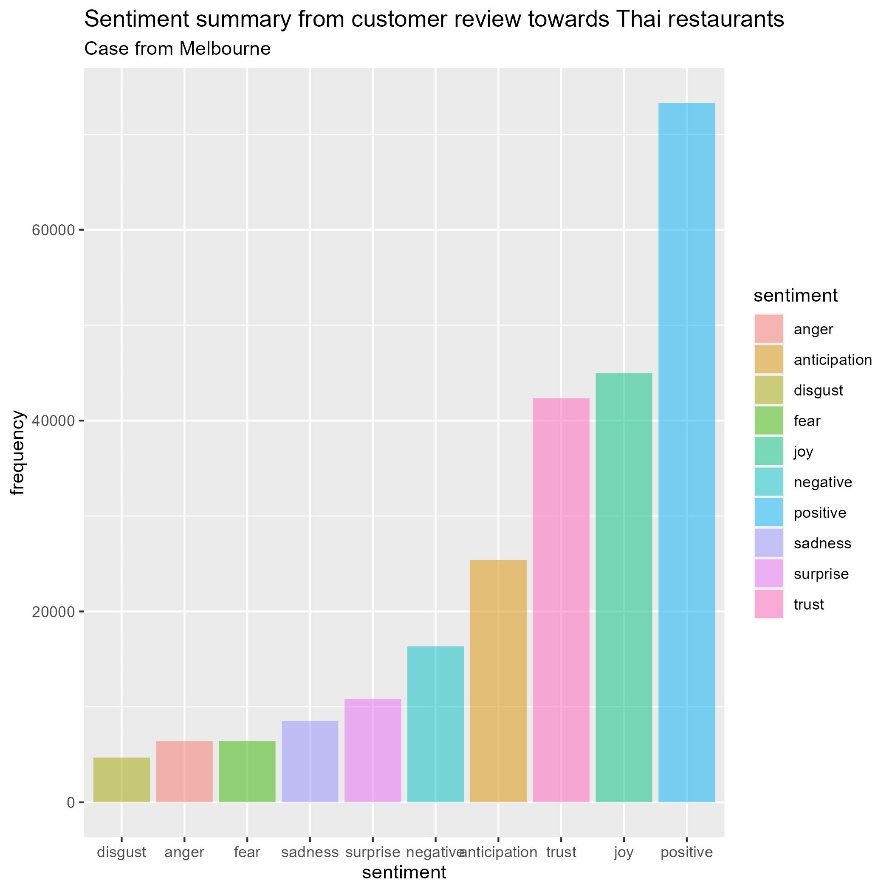
This project is considered a small-scale project. So, it did not require industrial-grade or enterprise software. As such, the software that is needed in a project can be divided into two types: Data Collection and Data Analysis. Data Collection can be obtained by using TripAdvisor API or using Python and R. In this project, R has been using it for web crawling. RSelenium and RCrawler packages have provided the tool to scrape the target website with a remote server. In which, the extensive understanding of HTML is a mandatory fundamental to working on it. For Data Analysis, R is preferable as it is built for complex analysis and yet a powerful visualization.

**Networking**

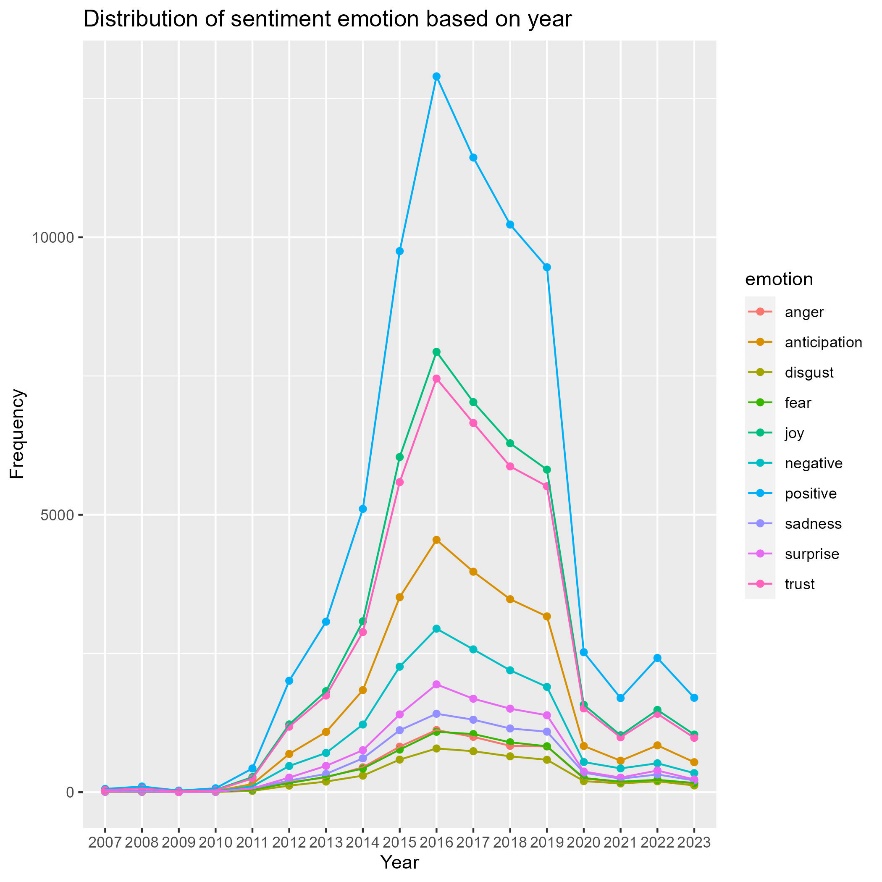
To perform Data Processing, a stable network is required. As the web crawling process, the bot (the script) is running in real-time to scrap the data in the target website via a remote server created by RSelenium which is based on Chromedriver.

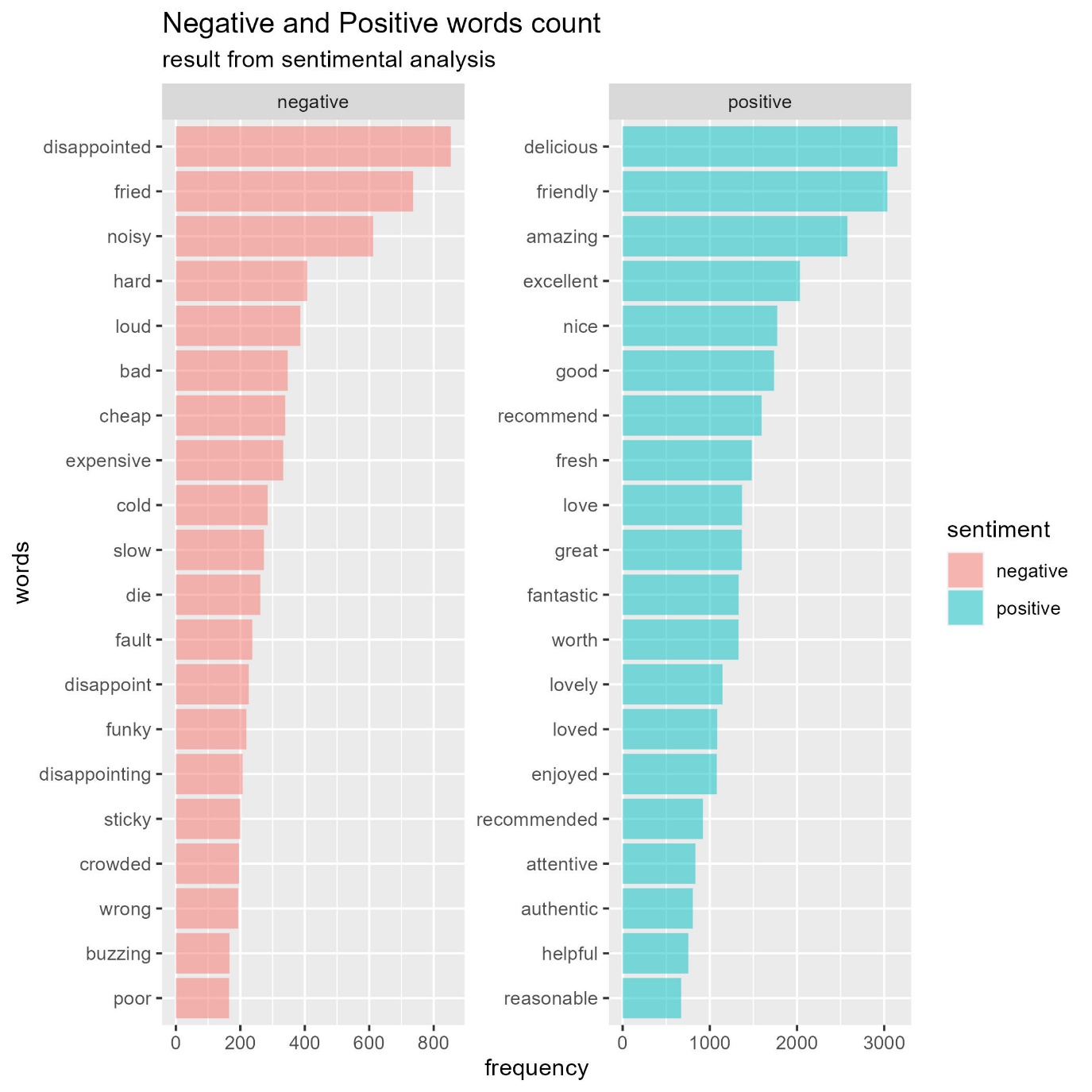
**Data Analysis**

To accomplish the project objective, the data (customer review) will be undergone through data wrangling to filter out stop words. After that, the node and the dictionary set will be defined to perform sentimental analysis. To fulfill objectives 2 and 3, the sentimental analysis package (syuzhet)will be in use where it will apply to cleaned data and query for the result to find the insight. For the first objective, it is possible to find by combining the prediction from the model and binding it with the review date and visualizing in a time-series model to foresee the trend of Thai restaurants in Melbourne. By compiling the data, the sentimental associated with the feeling can be summarized and plotted in a graph as follows.



The figure shows that the majority of feelings towards Thai restaurants are positive, joy, and trust respectively. It shows that Thai restaurants in Melbourne have built a positive image for the customers. Which, it is possible to dig down into this attribute to find what makes it positive. Likewise, the negative feelings should be investigated to see what the cause is and how to improve the current Thai restaurants.



By plotting a time-series line graph, it shows that customer reviews have decreased during the COVID-19 period. However, this graph has also generated the same insight as the first graph.

Next up, sentimental analysis will be implemented in which we aim to capture the negative and positive words and count the repetitiveness across the dataset. Then, the top 20 positive and negative association words will be captured and shown in the graph.

By performing an analysis, it appears that the most repetitive positive words are delicious, friendly, and amazing. It can be concluded that the strengths of Thai restaurants in Melbourne are delicious food, friendly customer service, and amazing experience. On the other side, the repetitive negative words are disappointing, noisy, and loud. It would be concluded that the customer might be disappointed with the taste or the slow waiting time and feel noisy as the restaurants are crowded.

As per analysis, this project finding illustrates that customers have overwhelmingly positive experiences towards Thai restaurants in Melbourne over a period. For the customer experience, they do love the Thai food tastes delicious yet excellent cuisine and friendly hospitality. For constructive feedback, the analysis shows that the restaurants should improve food processing time and importantly the restaurants should focus more on restaurant seating capacity as the owner should take into the decision that they would give up some seating capacity for a greater customer experience.

**Data Governance and Management**

The main consideration of this project is privacy and sensitivity. The collected data has involved high-sensitive data like customer profiles which are considered as digital footprint and are prone to vulnerability. To follow data compliance, the sensitive data will be confidential and not included in the data analysis section. For the governance and management aspect, accessibility to the data will be implemented. Only the author and the FIT teaching team will have access to the dataset. In addition, this project has defined a project lifecycle where the findings will be kept for educational purposes. Also, there are potential ethical concerns regarding data usage. As it already been stated, the data should be used for educational purposes only and it must not reveal the customer's identity, or be used in a harmful way, or be released for commercial use. Therefore, the findings of this project shall be used under fair use with non-profit interests or for research purposes.

**Appendix**

Melbourne Suburb list: [List of Melbourne Suburbs For Melbourne Web Designs | Melbourne Website Design, Hosting & Domain Names | zen10](https://zen10.com.au/melbourne-suburb-list/)

The data from TripAdvisor has been web crawling from [Restaurants in Melbourne - Tripadvisor](https://www.tripadvisor.com.au/FindRestaurants?geo=255100&cuisines=10660&broadened=true) and it contains 3 datasets: Business Details, Customer review part 1, and Customer review part 2.

The dataset available in this link. <https://drive.google.com/drive/folders/1TwRZBq_Tkn_HyJDN_otyQKV5ACkCYAJs?usp=sharing>

**Reference**

* Redlemonthai. “Why Do Australians Love Thai Food so Much?” Redlemonthai, redlemonthai, 2023, [www.redlemonthai.com.au/post/why-do-australians-love-thai-food-so-much#:~:text=Why%20Australians%20love%20these%20dishes,it%20appealing%20to%20many%20people](http://www.redlemonthai.com.au/post/why-do-australians-love-thai-food-so-much#:~:text=Why%20Australians%20love%20these%20dishes,it%20appealing%20to%20many%20people).
* Travelers push Tripadvisor past 1 billion reviews & opinions! (2022, February 1). <https://ir.tripadvisor.com/news-releases/news-release-details/travelers-push-tripadvisor-past-1-billion-reviews-opinions>