|  |
| --- |
| Yelp Inc. |
| Final project plan |
| Capstone DATA2206 |

|  |
| --- |
| Team members: Arunkumar Senthilkumar Student number: 100888951 Jayarani Rajesh Student number: 100818850 Rahul Sapparapu Student number: 100890409 |

**Table of Contents**

[**Project Significance:** 2](#_Toc126774802)

[**Anticipated Outcomes:** 2](#_Toc126774803)

[**Limitations:** 3](#_Toc126774804)

[**Directions of analysis:** 3](#_Toc126774805)

[**How this project fits the program of study?** 4](#_Toc126774806)

[**Project timeline** 4](#_Toc126774807)

# **Project Significance:**

The importance of this project can't be overstated as it will delve into what factors make customers satisfied or dissatisfied with restaurants serving seven different cuisine types. By examining customer reviews, making comparisons, and performing sentiment analysis, we will gain a better understanding of what customers want and what restaurants can do better. This information will be incredibly valuable to restaurants as it will provide them with practical ideas for improving the customer experience and growing their business.

In today's competitive environment, it's critical for restaurants to be in tune with their customers' desires and needs. The results of this project have the potential to shape industry best practices and drive growth and profit for restaurants of all sizes and types.

Additionally, this project will add to the existing body of knowledge in the field of customer satisfaction and restaurant analysis. Its findings have the potential to direct future research and move the industry forward.

To sum up, this project is significant for both its immediate practical applications and its contributions to the academic and business communities. The insights and recommendations it generate have the potential to have a lasting impact on the food industry and improve the dining experience for customers everywhere.

# **Anticipated Outcomes:**

Our project aims to deliver several key outcomes that will be beneficial for both the food industry and customers. These include:

1. Understanding Key Customer Satisfaction Factors: By analyzing customer reviews and conducting sentiment analysis, we aim to discover the key drivers of customer satisfaction for each of the seven cuisine types. This information will provide restaurants with a roadmap for improving the customer experience.
2. Comparative Analysis Across Cuisine Types: By comparing the results from our analysis of each cuisine type, we aim to gain a comprehensive understanding of customer preferences and expectations. This information will be useful for restaurants as they can tailor their offerings to better meet the needs of their target customers.
3. Best Practices for Customer Satisfaction: Our analysis will provide valuable insights into best practices for customer satisfaction that will help drive growth and profit for restaurants.
4. Areas for Improvement: Our project will also identify areas where restaurants can improve, such as menu offerings, service quality, or atmosphere.
5. Future Research Directions: Finally, our findings will inform future research in the field of customer satisfaction and restaurant analysis.

Overall, we aim to deliver actionable insights and recommendations that will positively impact the food industry and improve the dining experience for customers. Our comprehensive analysis across different cuisine types will provide valuable insights and resources for both academia and the business community.

# **Limitations:**

When it comes to this project, there are a few potential limitations that we should be aware of:

1. Quality of Reviews: Customer reviews can be subjective, and not always a true reflection of the dining experience. Additionally, some reviews may be fake or biased.
2. Methodological Challenges: The methods we use for sentiment analysis and statistical analysis may not completely capture the intricacies of customer preferences and experiences.
3. Local Relevance: Our findings may only apply to the specific geographic area and time period represented in the Yelp data.
4. Incomplete Information: It may be hard to consider all the variables that impact customer preferences and experiences, such as culture, socioeconomic status, and regional differences.
5. Personal Bias: Interpreting and making conclusions from customer reviews can be subjective, and personal biases may impact the analysis.

Recognizing these limitations and working to minimize their effects will help ensure the validity and usefulness of our results, both for the food industry and future research.

# **Directions of analysis:**

1. **Exploratory Data Analysis**

* We'll look at the different types of cuisine represented in the Yelp dataset, counting the number of each type of restaurant.
* We'll also keep track of the number of reviews for each cuisine type, to get an idea of what customers have been saying.
* To better understand how customers feel about each type of cuisine, we'll create visualizations that show the distribution of restaurant ratings by cuisine type.

1. **Review Analysis**

* We'll start by organizing the data in the business table, categorizing each restaurant according to its cuisine type.
* Next, we'll link the business table with the review table, pulling in all the reviews for each type of cuisine.
* We'll then perform sentiment analysis on these reviews to get a sense of the most common words used in positive and negative reviews.
* Using a support vector machine (SVM) model, we'll identify the most positive and negative words for each cuisine type.
* Our goal is to identify the top 10 positive and negative words for each cuisine type, to understand what customers love and what they dislike about each type of restaurant.
* By comparing the strengths and weaknesses of each cuisine type, we'll make recommendations for how each type of restaurant can improve in the future.
* These recommendations might touch on topics like service, food, decoration, and more, and will be based on our analysis of what customers are saying in their reviews.

# **How this project fits the program of study?**

As a data analytics student, this project aligns well with the curriculum and skills I've acquired throughout my program. Our aim is to utilize knowledge of data analysis techniques and tools to extract meaningful insights from the Yelp customer reviews

* Data Cleaning and Preparation: We will use various techniques to clean and prepare the Yelp dataset for analysis, including imputing missing values, identifying outliers, and scaling the features.
* Exploratory Data Analysis (EDA): We will conduct EDA on the customer reviews and ratings, using data visualization and summary statistics, to gain a deeper understanding of the data.
* Statistical Analysis: We will apply statistical methods such as hypothesis testing and regression analysis to compare the different cuisine types and determine the factors that influence customer experiences and preferences.
* Machine Learning: We will implement machine learning algorithms, such as decision trees or random forests, to predict customer satisfaction based on various variables and improve the accuracy of the results.

By utilizing these skills and tools, We aim to carry out a comprehensive analysis of the customer reviews and provide valuable recommendations for the improvement and growth of restaurants in different cuisine types.

# **Project timeline**

|  |  |
| --- | --- |
| **Activity** | **Completion Date** |
| Kick off meeting with teammates | 11/Jan/2023 |
| First meeting to discuss on draft proposal | 12/Jan/2023 |
| Preparation of draft proposal document | 27/Jan/2023 |
| Preparation of final proposal document | 03/Feb/2023 |
| Project plan document preparation | 08/Feb/2023 |
| Second meeting with team to discuss on further analysis | 10/Feb/2023 |
| Initial Analysis and exploration about Yepl company and the dataset | 15/Feb/2023 |
| Research which part of database would be a good fit to store data and its implementation with the models. | 17/Feb/2023 |
| Presenting the analysis done to client | 22/Feb/2023 |
| Exploring different ways and improving the performance of the SVM model | 01/Mar/2023 |
| Creating visualizations for the report | 15/Mar/2023 |
| Collating the work done and preparing a report covering all the aspects of the project | 29/Mar/2023 |
| Final meeting with the client and the delivery due of the project | 05/Apr/2023 |