**Introduction**

There are many ways to get information about businesses now a day. Yelp is a mobile app for finding businesses, making reservation, or giving out reviews for businesses. People can rate the businesses based on their products or services. The maximum of the star rating is 5 and the rating review can be rated by other users as well. Reviews are written down after the customers got their services done. By using the reviews and the ratings, the sentiment of the customers can be determined.

**Analysis**

The dataset is from the Yelp, and it was download from Kaggle website. The datasets used are the Yelp business dataset and the Yelp review dataset. There are 50000 data were pull from the dataset and then use merge function to merge two datasets with business ID. After merging the dataset, the missing values were found and dropped. In the below picture show the Exploratory Data Analysis (EDA) were performed on the merged dataset. Also, the outlier was removed for better view of the graphs for the number variables.

Chart

Description automatically generated with medium confidence

Graphical user interface, application

Description automatically generated with medium confidenceGraphical user interface, application, Excel

Description automatically generated with medium confidence

Graphical user interface, application, table, Excel

Description automatically generated with medium confidenceChart, histogram

Description automatically generatedChart, bar chart, histogram

Description automatically generated

The graph showed the percentage of the state inside the dataset. In the picture above, it showed that most of the state inside the dataset were from Massachusetts.

To count the words inside reviews and categories. The reviews and categories data were tokenized, then stop words and punctuations were removed as well. Afterward, the Lambda function was used to count the tokenized word inside review texts and categories.

**Result**

This is the word cloud for the review text, the larger the word inside the word cloud meaning the higher the frequency of the reviews appeared inside the review dataset. In this case, food, good, and place are the top three among the review text data. Text

Description automatically generated

For sentiment analysis, the ratings that were above and equal to 3 are set to 1, and the ratings that were below that are set to -1. Then it was separated to positive and negative dataset to make the word cloud. The first picture below shows the positive word cloud, and the second picture shows the negative word cloud. It was astonishing that two of the word clouds looked extremely similar, but the word count datasets show the difference that are inside the plots.

Text

Description automatically generated

Text

Description automatically generated

Table

Description automatically generated with medium confidenceTable

Description automatically generated with medium confidenceTable

Description automatically generated

The picture above shows the review text words count, positive word count and the negative word count. Most of the words used in the high rating reviews are oblivious to interpret, since the words used are all positive. For the low rating reviews, there are some mixes of the words that are not negative were used in the dataset. But, looking at the bottom half of the negative word count list, there are lots of words that contain negative sentiment which means people are not upset about the business.

Chart, bar chart

Description automatically generated

The picture above shows the bar plot of the count dataset for Yelp review.

Text, chat or text message

Description automatically generated

This is the word cloud for the categories inside the yelp dataset.

Chart, bar chart, histogram

Description automatically generated

The above picture is the graph for the categories count. Looks like most of the Yelp user rate the businesses that provide foods for them. Thus, the restaurant being in the top of the categories among the datasets.

Research Questions:

1. What are the most used words for a good rating? (rating > = 3)

Good

1. How often the positive word being use have a high rating?

21905

1. What are the most used words for a bad rating? (rating < 3)

Food

1. How often the negative word being use have a low rating?

6376

1. Which type of categories are the most among the datasets?

Restaurant

**Conclusion**

Overall, most of the Yelp user used Yelp to find restaurants near them. After visiting and having the service or foods, customers gave out reviews and rating based on the service that they received. In which, in the reviews, the customers told us the sentiment they have after receiving the service or foods in the businesses. These gave us good amounts of information to tell whether the businesses are worth to visit or not by the review text, not by the rating of the businesses.