**Boston Common Restaurant Recommendation**

Junhong Li

July, 22, 2021

# Introduction

## Background

The tourism and catering industry positively affects the promotion of a city's economy and visibility. However, due to the COVID-19 pandemic in 2020, it hit the global travel and catering industry hard. As more people are vaccinated, restrictions in countries and regions are gradually opening up. As a result, people began to have the desire to resume travel. As a city with rich culture and history on the east coast of the United States, Boston attracts many tourists every year. Four Season Hotel Boston, located in Boston Common, is the perfect place for many tourists. Four Season Hotel Boston wants to write a new Boston travel guide for its guests, including Boston attractions and dining around the Hotel. For the recommended introduction of scenic spots, you can follow the previous year's guide to do some appropriate content updates; However, due to the impact of the COVID-19 pandemic on the restaurant industry and recent changes in surrounding restaurants, this section needs to be rewritten.

## Problem

Due to the location of Four Seasons Hotel Boston, there are a large number of restaurants of various types nearby. The Hotel hopes to recommend high-quality restaurants of the overall type to increase tourists' overall satisfaction with the Hotel service while avoiding the presence of low-quality restaurants in the guide to affect the Hotel rating.

# Data acquisition and cleaning

## Data sources

The data comes from Foursquare, the location map data record of Four Seasons Hotel Boston up to July 22, 2021.

## Data cleaning

Convert the raw data of the network to a Dataframe, as shown in the figure. According to the shape function, there are 92 records of the original data. Where venue. The name represents the name of the merchant, venue. Venue. Location refers to the geographic coordinates of the location.Text, table

Description automatically generated with medium confidence

There are some problems with the raw data:

1. Every column's name contains unnecessary content that needs to be removed. The name of columns with the venue prefix makes it difficult to understand data.
2. It contains categories other than restaurants, which are not required by the project and need to be dropped.
3. The original data lacks effective quantitative indicators for restaurant evaluation.

The original data only includes the geographical location information of merchants, and quantitative indicators of customer evaluation need to be added.

After data processing, a new dataset that can be used for analysis is obtained. The new Colum name is name, ID, Categories, LAT (Latitude), LNG (Longitude), and Total Like. Total Likes are the number of comments that consumers have given to graph likes obtained from the Foursquare API by ID. After clearing, the dataset has 42 records in total.Graphical user interface, text, email

Description automatically generated

# Exploratory Data Analysis

## Data distribution of Total Likes

## Through calculation, I found that when people evaluated the 42 restaurants around The Four Seasons Hotel Boston, there found that there were 513 likes with the most value, and only one likes with the least value. The median value of likes was 62.5. Thus, the mean of people giving "like" to restaurants around the hotel is about 113.Text, table Description automatically generated

## As shown by histogram:Chart, histogram Description automatically generated

## Get Rating Level

In order to group total Likes more reasonably, it's nedd performed a qualitative analysis. The level is divided by calculating the first Quartile, median, and third Quartile of total Likes. 26,62.5 and 173.5, respectively.Graphical user interface, text, application

Description automatically generated

The restaurant with total Likes less than 26 is rated as poor; the restaurant with total likes between 26 and 62.5 is rated as below average; the restaurant with total likes between 62.5 and 173.5 is rated as average. Total likes of more than 173.5 are considered great restaurants. The figure below adds new columns total likes cat and creates a new data frame.Graphical user interface, text, application, email

Description automatically generated

## Set new restaurant category

## To avoid too detailed types of restaurants and facilitate the exploration of popular food types of restaurants around the hotel, the categories of visits are further summarized as follows: Bars, other, Asian food, Mexican food, American food, Italian food, French food. See the figure below.Graphical user interface, text, application, email Description automatically generated

From the histogram, we can see that another food type restaurant dominates nearby hotels, followed by Italian food. American food and Asian food are similar in quantity.Chart

Description automatically generated

# Clustering Modeling

# The commercial demand of the hotel is to know the most popular types of surrounding restaurants to recommend high-quality restaurants to customers to obtain better satisfaction more accurately. Therefore, I chose clustering modeling here to analyze the dataset and get the characteristics of restaurants with different user rating levels.

# One hot

# Transform the dataset into one Hot, where 0 represents negative, and 1 means positive, to facilitate the subsequent analysis of K Clusters. As below figure is shown.A picture containing background pattern Description automatically generated

* 1. **K-means Clustering**

Since I have previously graded our customer ratings as great, average, below average, and poor, we have a K value of 4. The run k means clustering results are as follows. We find that the added labels fall into 0,1,2,3.Graphical user interface, text, application, email

Description automatically generated

## Labeling

## First, the data with label 0 has been filtered, and the result is shown in the figure. We can find that all records with label= 0 are restaurants with a rating below average.Graphical user interface, application, table Description automatically generated

## As the histogram shown, we can find that the most significant number of Italian foods rating levels are below average, up to 5.Chart, bar chart Description automatically generated

## Then, the data label is 0 is extracted, and the result is shown in the figure. We can find that the record of label= 1 is the restaurant with a great rating level.

## Graphical user interface, text, application, table Description automatically generated

## From the histogram, it can be found that among great restaurants, except other, Asian food is served in the most significant three restaurants, followed by American food in the second place: 2 restaurants.Chart, bar chart Description automatically generated

## Then the data with label 2 is extracted, and the result is shown in the figure. Label 2 records with rating level poor.Graphical user interface, application, table Description automatically generated

## From the histogram, we can see that rating level poor restaurants, majority of them are in the other category.Chart, bar chart Description automatically generated

## Finally, the data with label three is extracted, and the result is shown in the figure. It can be found that all records with label= 3 are restaurants with an average rating level.Graphical user interface, text, application, email Description automatically generated

## According to the histogram, among the restaurants on the average rating, the other category takes up a maximum of 4 restaurants.Chart, bar chart Description automatically generated

## Mapping

## The analysis results are visualized by map. The red dot represents below Average, the yellow dot represents Average, the cyan dot represents poor, and the purple dot represents great.

Map

Description automatically generated

# Conclusions

# There are plenty of dining resources around the Four Seasons, with 42 restaurants in total. These include Chinese restaurants, Japanese restaurants, American restaurants, Italian restaurants, French restaurants, bars, sandwich places, and cafe.

# Of the 42 restaurants, the most popular overall and the highest-rated were those running Asian food. Although many Restaurants are operating Italian food around the hotel, the Overall Score of Italian restaurants is below average.

# Suggestions & insights

Four Seasons Hotel can mainly recommend restaurants that offer Asian food when recommending nearby restaurants to customers. First, the few asian restaurants receive the most consumer likes in competition with other kinds of restaurants. Secondly, although there are many Italian restaurants around the hotel, it is not recommended to recommend Italian restaurants because they have received "likes below average." Finally, if a hotel wants to increase the variety of food available in its restaurant, it can consider adding Asian Food Dishes. This insight will improve the customer satisfaction of the hotel restaurant, help the hotel get a better rating, and succeed in adding food Dishes.