

Identifying the best cities to set up a restaurant

Tejali Gangane

July 26, 2020

1. Introduction

1.1 Background

The idea of having setup your own restaurant may seem easy. But there are many factors to be considered. Owning a restaurant requires knowledge of the location, the type of food you will be serving, the hours required to be invested and more. To achieve all this there needs to be a plan according to which certain measures could be taken to set up a restaurant. Depending upon the type of food you will be serving you can decide which city will be optimal to set up your restaurant. There might be a possibility that people from certain cities are fond of certain cuisines and tend to follow them. Also, it might be more beneficial to setup restaurants near colleges or parks or companies as people might be able to visit during their breaks. There are so many factors to be taken to identify what city or place can the restaurant be setup in so as to generate more revenue. And ensuring that the food which will be served is worth it for the customers.

1.2 Problem

The data gathered will be consisting of venues with existing restaurants and how well they are doing which will be identified by the reviews given by the users. Also, identifying the types of cuisines these restaurants offer can help us understand what type of food people prefer the most.

1.3 Interest

This project will help the people who want to setup their own restaurant. It will help them in identifying which venue will be preferable to setup the restaurant. The factors which will might come in handy to identify the preferred location can be the type of food preferred by the crowd in certain location, what kind of food is served by the restaurants which are already present and how well they are doing. This can be identified by the ratings and reviews which will be given by customer. Owning a restaurant involves more work than we may think, but in the end it comes down to the location, the people, the food served and depending on these factors the success of the restaurant can be determined.

2. Data Acquisition and Cleaning

2.1 Data Sources

The attributes restaurant id, name, country, longitude, attitude, average cost, average rating and many more are found in the Kaggle dataset [here](#). This dataset lacks the information of all the restaurants in any area/city/country. Hence Foursquare API can be used.

Using the existing data, we can find all the restaurants that already exist in a particular place where the person wants to setup. And with the help of foursquare API we can find all the other list of restaurants which are not present in the dataset. Combining these two we can identify the cuisines served by these restaurants and compare with the type of cuisine that the person wants in their restaurant and identify which of the places will be suitable to setup the restaurant.

2.2 Data Cleaning

Data was downloaded from Kaggle. To further process the data, it was checked for missing values, none were found. It is important to find the missing values as while performing an analysis of the data we might not get proper results. Missing values can be dealt by different methods like imputation, i.e. replacing them by some other value or removing the row if it does not affect the data in any way. We can also drop the entire column if it is not an important feature.

The data types were also cross checked and were found to be perfect. This is done so as to ensure the integrity of the data. Whether the data present is accurate and consistent. Sometimes it might be possible that the date feature stores the date in a particular format and the one of the value consists of the date which is in some another format, this can lead to inconsistency of data and can provide inaccurate results or error when further processing.

No duplicate rows were present. Duplicate rows are checked so as to avoid data redundancy.

2.3 Feature Selection

After data cleaning there were 9,551 samples and 21 features in the data. After further examining each feature, it was observed that there were certain features which were unnecessary. Features like – ‘Address’, ‘Locality Verbose’, they are just detailed description of the location of the place at which the restaurant is present. And the kind of information which they had were a bit similar. Such features were discarded.

Other features like 'Has Table booking', 'Has Online delivery', 'Is delivering now', 'Switch to order menu', were not necessary in the context to what is to be achieved. The project aims to identify a suitable area for a person to setup their restaurant. The criteria required will be the location, the type of restaurants which are already present in that location; whether the idea of the person for the restaurant already exists in that place or is that place populated enough.

Now coming to the review of these restaurants, there are few features relating to it. They are - 'Aggregate rating', 'Rating color', 'Rating text', 'Votes'. The first 3 features mentioned are a bit similar to each other so we will just be keeping the feature 'Aggregate rating'. The feature 'Votes' indicates the number of ratings csted by people for a particular restaurant. We will keep this feature as it will help us to understand the number of people that liked/didn't liked the restaurant.