COURSERA CAPSTONE PROJECT
THE BETTLE OF NEIGHBORHOODS

Table of Contents

The Battle of Neighbourhoods

Part 1: A description of the problem and a discussion of the background?

Part-1: Introduction: How to choose the Best Location to start an Asian Restaurant in London?

As a restaurant owner, to find the perfect location to help make your operation a true success. It's important to remember that a restaurant's location is as important to its success as great food and service. You should really put location at the top of your to to-do list and start looking right away.

If you can define your restaurant type and identify your target demographic and its most populated areas, you'll be well on your way to choosing a restaurant location that sets your business up for success. There's a lot of work, planning and preparation that goes into opening a restaurant. Here are a few tips on how to open a restaurant in London and some steps you'll need to take in order to achieve your goal.

1. Restaurant style:

The first thing you need to decide is what type of restaurant you want to open and the style it will hold. Having this set in stone will make the other steps a lot easier. To make things a little easier, here are three categories you need to consider:

- **Food**: What type of food will you be serving? Will you specialise in a certain cuisine?
- **Service**: What type of service will you be providing? Waiter service? Self-service?
- **Atmosphere**: What type of atmosphere do you want your establishment to hold? Do you want to open a casual, self-service cafe? Or a fine-dining, high-end restaurant?

2. Target Market:

In order to open an Asian restaurant in London, you need to be familiar with exactly who you are aiming to bring into your venue. Are you targeting audience from Indian, Pakistani, Bangladeshi and Chinese origin? You need to also be aware of the age group, the location, the amount of money they are willing to spend.

Suppose, if you want to open a classy restaurant, it will not make a lot of send to open it in an area frequented by young people. And if you put up an informal restaurant and entertain customers with live music until late at night, you would not open in a quiet residential area.

3. Who are your competitors?

As well as familiarising yourself with your audience, you must do so with your competitors. Look at similar businesses in your area and well-renowned restaurants of the same category. Consider the ways in which they market themselves and the service they provide. What works and what doesn't? You'll need to be able to compete with these establishments, so take on board what they're already doing and better it.

It's important to avoid price wars or other conflicts and trouble. Furthermore, being compared to another restaurant day after day is exhausting. The best way for you is to try to have your own personality and differentiate your restaurant from all the others. In that way, you not only have a unique business but show respect for the other enterprises in the area.

Be aware of the type of your establishment and the number of similar restaurants in the area you are looking to open it. If there is already a lot of them, chances are you're entering business in an over-saturated area and the potential for you to succeed quickly may be diminished. Consider opening in an area where there is a gap in the market and demand for your service. However, you must also consider the opposing. If there are no restaurants in the area similar to your concept, why? It could simply be that nobody has followed that path yet, or because it will not work. Do your research to ensure you have the right location to suit your restaurant?

4. Create your menu:

The menu is the core of any restaurant. It's important for you to get it right, as it's the deciding factor for customers deliberating a visit. Your service could be impeccable, decor and atmosphere exquisite, but without a menu to match, you shall have no such luck in success. So when you are targeting the Asian ethnicity audience, your menu should include cuisines from India, Pakistan, Bangladesh and China.

5. Location and Premises:

There are many factors that will contribute to where you decide your premises to be. It can be challenging to find a venue that will factor in all of your conditions, so it is likely you will have to compromise on a few things. However, here are the main factors you should consider when finding premises for your new restaurant:

- **Location**: How accessible is the location? Are there many competitors nearby?

  The perfect restaurant location is not necessarily the city centre or near a tourist attraction. One critical factor to incorporate in your search is that depending on your location, your restaurant should be easily accessible by car and/or has plenty of foot

traffic as well. Another consideration is that your restaurant location should be easy to explain on the telephone as well as in your advertising and posters.

- **Building type**: What type of building would you like your restaurant to be in?
- **Target market**: Make sure it is in an area where your target market is.
- **Cost**: Your ideal location may prove to be a little costly. Make sure that there is a realistic price point.

6. Parking Lot

You need to attract new customers, but if you don't have a parking lot nearby, your restaurant will suffer the effects. Few customers will consider the presence of a parking lot a determining factor in choosing one restaurant over another, but nevertheless, when a customer has a hard time finding a parking spot nearby, you're losing out. And it's important to remember the difficulties disabled people in wheelchairs may encounter trying to get to your restaurant.

As a final consideration on parking, remember: Customers want to feel safe and added to that, they want to park their car where they can have little worry that someone will dent or scratch their vehicle while they're dining. A scratched vehicle means you're unlikely to see that customer again.

7. Transportation

If you're seeking your fortune in the food and beverage field in a large city like London, you're fortunate in that just about any place has nearby transportation. But if you choose to locate in a small city, it's important to consider public transportation.

Part 2: A description of the data and how it will be used to solve the problem.

Part 2: Data Description

London is one of the most ethnically diverse cities in the world. At the 2011 census, London had a population of 8,173,941.The demography of London is analysed by the Office for National Statistic and data is produced for each of the Greater London wards, the City of London and the 32 London boroughs, the Inner London and Outer London statistical sub-regions, each of the Parliamentary constituencies in London, and for all of Greater London as a whole.

For our restaurant problem, we will focus on the Boroughs of London and work on getting the data from all the Boroughs. There are 32 London Boroughs with a population of around 150,000 to 300,000.

To solve our problem of finding a best location to start an Asian restaurant in London, we need the datasets based on various parameters such as:

1. Population of target audience in all the boroughs of London based on their:
   • Asian ethnicity
   • Age
   • Gender
   • Marital Status
   • Employment Status
   • Income
2. We also need the data about the required Business floor space and Rateable Value Statistics of each borough.
3. Considering the competitors factor, we also need the data of existing Licensed Restaurants in each borough.
4. And lastly we will also consider the borough level tourist and domestic annual spend estimates.

All the above required information is available at LONDON DATASTORE, which is a free and open data-sharing portal where anyone can access data relating to the city. The data is available in XLS and CSV format, which we can download and can use as-is for solving our problem.

The link for the LONDON DATASTORE, –
https://data.london.gov.uk/ or
https://data.london.gov.uk/census/data/

The link for List of Boroughs:
https://en.wikipedia.org/wiki/List_of_London_boroughs

The link for List of areas of London:
https://en.wikipedia.org/wiki/List_of_areas_of_London

The Licensed Data:
https://data.london.gov.uk/dataset/pubs-clubs-restaurants-takeaways-borough

The Ethnic Group by borough data:
https://data.london.gov.uk/dataset/ethnic-groups-borough

The Rated value data:
https://data.london.gov.uk/dataset/commercial-and-industrial-floorspace-borough

Earnings Data:
https://data.london.gov.uk/dataset/earnings-place-residence-borough

Along with the above datasets we will also use the Foursquare location data to solve our Problem.

Part 3: Methodology

To work on the solution, I have used Pandas library to read the data in XLS format and convert into panda's data frame. Extensive data exploration analysis is done, where lot of data is cleaned and presented in a suitable format.

Machine Learning Algorithm Simple Linear Regression is used to predict the data for Rated Value for the year 2018 for the selected borough. The dependant variable would be the rated value for year 2018 and the independent variables are the earnings of each borough and the existing restaurants in each borough.

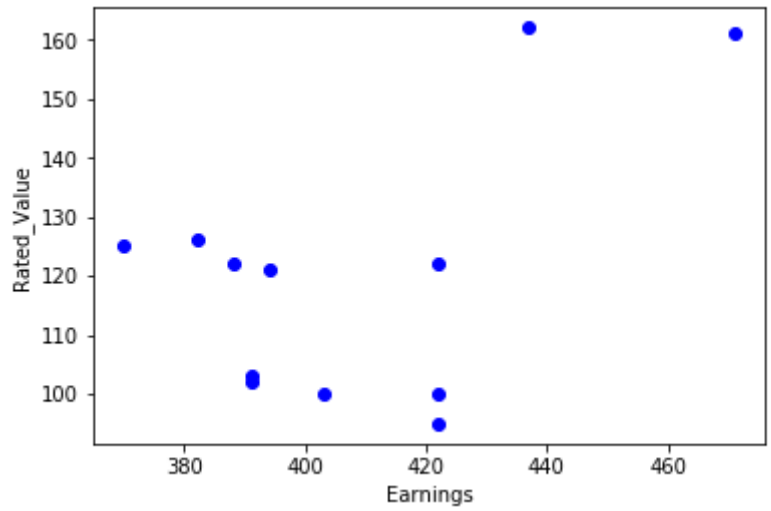

Figure 1 Earnings vs. Rated Value

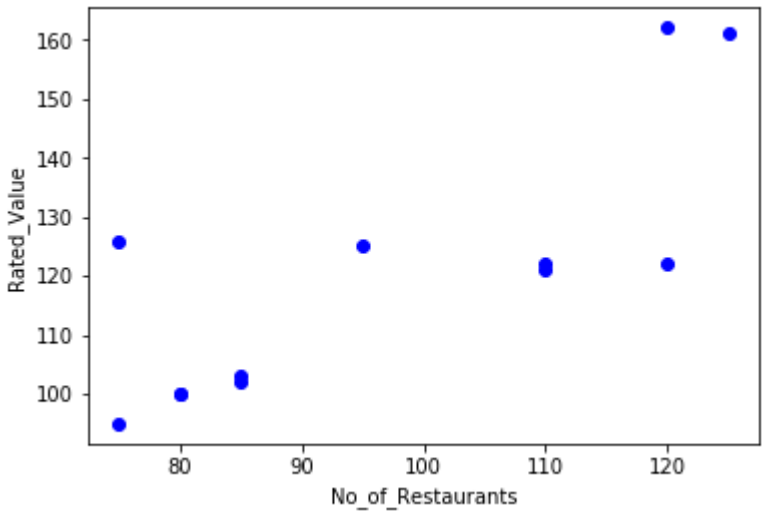

Figure 2 No of Restaurants vs. Rated Value

Two scatter plots are plotted between these variables and from these 2 plots, it is observed that the linear relationship exists between the Restaurants and the Rated Value. A scatter plot clearly shows the relation between variables where changes in one variable explain or possibly cause changes in the other variable. Also, it indicates that these variables are linearly related.

Simple Linear Regression fits a linear model with coefficients Ø = (Ø1, Ø2… Øn) to minimize the residual sum of squares between the independent X in the dataset and dependant Y by the linear approximation.

Coefficient and Intercept in the Simple Linear Regression are the parameters of the fit line. Given that it is simple linear regression with only 2 parameters and knowing that the parameters are the intercept and slope of the line, using the python library SciKit Learn, we can estimate them directly from our data. The available data is divided into Train and Test data. The train data is used to train the model and the test data is used to evaluate the model.

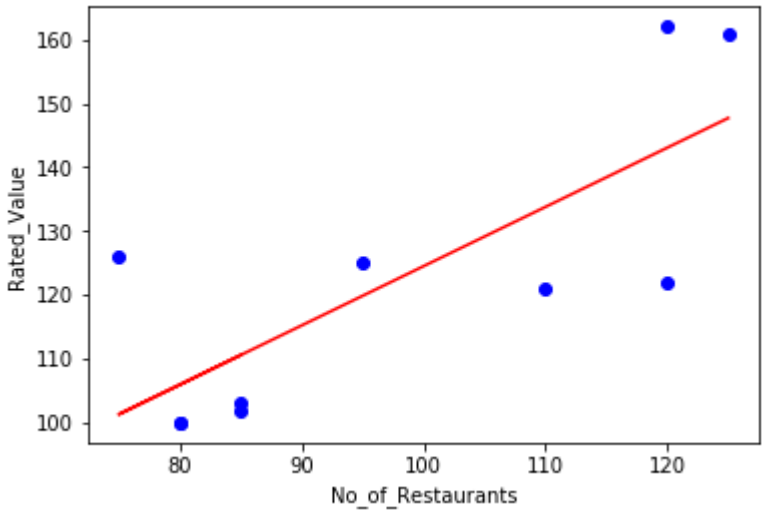

Figure 3 Linear Regression Model

Evaluation of the model is performed using the Evaluation Metrics such as Mean Absolute Error, Mean Squared Error and R-Squared. Due to very less available test data, the R-Squared for our model is not that great, but still we can consider our model for the prediction of the Rated Value for the year 2018.

Part 4: Segmenting and Clustering

After the prediction of rated value per sq. meter of a retail space is completed and when we are convinced that a particular borough will be the preferred location for the restaurant, we have to get the necessary data of that borough.

First we need to get the geo-coordinates of the borough and the geo-coordinates of the neighbourhoods of the borough from the web. I have used the Wikipedia pages to get this data.

The link for List of Boroughs:
https://en.wikipedia.org/wiki/List_of_London_boroughs

The link for List of areas of London:
https://en.wikipedia.org/wiki/List_of_areas_of_London

To read data from these URLs, I have used the requests, urllib and Beautiful Soup libraries of python.

After I have the geo-coordinates information of the borough and its neighbourhoods, I need the other data such as the venues or places of the neighbourhoods, the venue categories, working hours and so on. All this data is called Location data, and to get this data I need a reliable and

efficient location data providers and hence I am using Foursquare as the data provider. I have used the Foursquare API to explore the neighbourhoods in London city. I have also used the Explore function to get the most common venue categories in each neighbourhood and then use this feature to group the neighbourhoods into clusters. To cluster the neighbourhoods I am using K-means Clustering algorithm.

Geopy module and Nominatim library is used to convert a given address into the latitude and longitude values.

To visualize the neighbourhoods, the library Folium is used, to display the maps of London, with the boroughs super imposed on it and to display the map of borough with the neighbourhoods superimposed on it.

A python function getNearbyVenues() is created , to give the venue details like venue name, venue latitude, venue longitude, venue category along with neighbourhood name, latitude and longitude for each neighbourhood.

After the venue data for each neighbourhood of the Newham borough is generated , One-Hot encoding is applied on the venue category data, so that the analysis of the data will be easy in grouping the neighbourhoods based on the frequency of occurrence of each venue category.

Once the neighbourhoods are grouped based on the frequency of occurrence of the venue category, the top 10 venues of each neighbourhood are displayed as a dataframe.

After all the above data exploration and analysis and top 10 venues of each neighbourhood are identified, the K-means Clustering algorithm is applied to the resultant data frame to segment the data into 5 Clusters and all these 5 clusters are visualised in a map using the Folium library and finally the 5 clusters are examined to determine the discriminating venue categories that distinguish each cluster.

Result:

From the data sets of Asian population, we found that Newham borough is having highest Asian population and from the rated value data set, the prices in this borough are less compared to other boroughs. These 2 factors influenced more on my decision of choosing Newham borough as the preferred location for our restaurant. The Newham borough has 146 existing restaurants and taking this as an independent variable (X variable) I have predicted the rated value per sq. meter (dependant variable) between 160 to 165 using the Linear Regression model. I have also calculated the MAE and R-Squared with the test data, though we got less values for these metrics due to less available test data.

In the Segmenting and Clustering section, the neighbourhoods of Newham borough are explored, and the top 10 venues of each neighbourhood are listed. The neighbourhoods are clustered into 5 clusters using K-means algorithm and their most common neighbourhoods are identified. After applying the K-means algorithm the 5 neighbourhoods Beckton, Custom House, Maryland, Eastham and Manor Park are identified as best locations to open or start an Asian restaurant.

Discussion:

My observation after doing this analysis is the model we used could have given better results, if we had huge data to train and test our model. In spite of that this model gives us a better insight for our problem and also help us to gain better results. From the clustering results our problem finds a better solution of identifying the best location for the Asian restaurant. We could explore all the neighbourhoods of the borough and could list the most common venues based on their frequency of occurrence. From these results I can strongly recommend the Beckton, Custom house and few other neighbourhoods as a preferred location for our restaurant, as these areas have the restaurant venue as the most common venue.

Conclusion:

There is always room for improvement and hence the above solution I have provided can also be improved and the machine learning models can be trained and tested for best results depending upon the data we have.