Predicting the best location to open a Restaurant in London

Kanishk Chauhan

February 23, 2020

1. Introduction

1.1 Background

Thousands of people dream of starting and owning a successful restaurant. Starting a restaurant business is no less than a herculean task. In a market where everybody wants to open a restaurant there are few things which restaurateurs have to keep in mind before starting up the business which includes cuisine, ambience, quality, services and location. In almost every case the path to entrepreneurial success starts with choosing the right location for their new restaurant.

Choosing a happening place could help restaurateurs keep an upper hand in the market. A happening place would grant you with happening consumers. Not everyone belongs to everywhere. Same, not every place is for restaurant. It is really difficult than some people think it might. A place could take you to the heights or could even turn into dud. Like, how would you reach to the consumers if it’s not in the right place? No consumers mean No business. It is as straight as that.

1.2 Problem

Clearly choosing the right location to start a restaurant is not easy, though, since every year thousands of restaurants fail. But some don't, even for first-time restaurant owners. This project aims to recommend locations in London with which offer good prospectus for starting and running a restaurant successfully. The key word, of course, is "successful".

1.3 Interest

Audience interested in the output of this project would include entrepreneurs looking to start a new business specifically restaurant in the city of London. This would also provide information to certain extent to the individuals looking for exploring the neighbourhoods of London.

2. Data

For this project we would primarily use Foursquare location data. Foursquare is a technology company that built a massive dataset of accurate location data. Foursquare powers location data for Apple Maps, Uber, Snapchat, Twitter and many other. We will use Foursquare

location data, for example, for getting popular venues like train station, gym, park, hotels, etc. for the neighbourhood we are exploring in London.

For leveraging Foursquare we first need to create a free developer account. Once the account is created we are provided with unique client-id and client secret id which can be used to make requests to extract location information from Foursquare. Though there are some limits on requests made per day, the current limit is sufficient for the purpose of this project.

We will also use location data present in 'geonames.org' website to extract coordinates of diverse neighbourhoods in London. This website is very useful and has coordinates information for almost all countries in the world.

3. Methodology

3.1 Data Pre-processing

First the basic data regarding the neighbourhoods of London was fetched from a Wikipedia page '*https://en.wikipedia.org/wiki/List_of_areas_of_London*'. The data on neighbourhoods was in the form of table on this page. To read the data into data frame 'read_html' method from Pandas library was used. Then dataframe was formatted to remove the unnecessary columns and keep only the columns listed below:

1) Postal Code
2) Borough
3) Neighborhood

Also Neighbourhood name concatenation was done to have unique values for Postal code.

Then in the second step, coordinates data for neighbourhoods of London was fetched from 'geonames.org' website. The same process was followed to read the data into a dataframe from a table on this website. But here the data was messy and had to follow many steps like separating the latitude and longitude values from coordinates and converting them to float values.

This coordinates dataframe was then joined to neighbourhood dataframe to create a final dataframe for neighbourhood along with latitude and longitude values. Also, the missing values in the form of NaN were handled by removing the rows that contained NaN.

Thus data pre-processing involved many steps to format the data, remove null values and join the data to be used for further analysis.

3.2 Exploratory Data Analysis

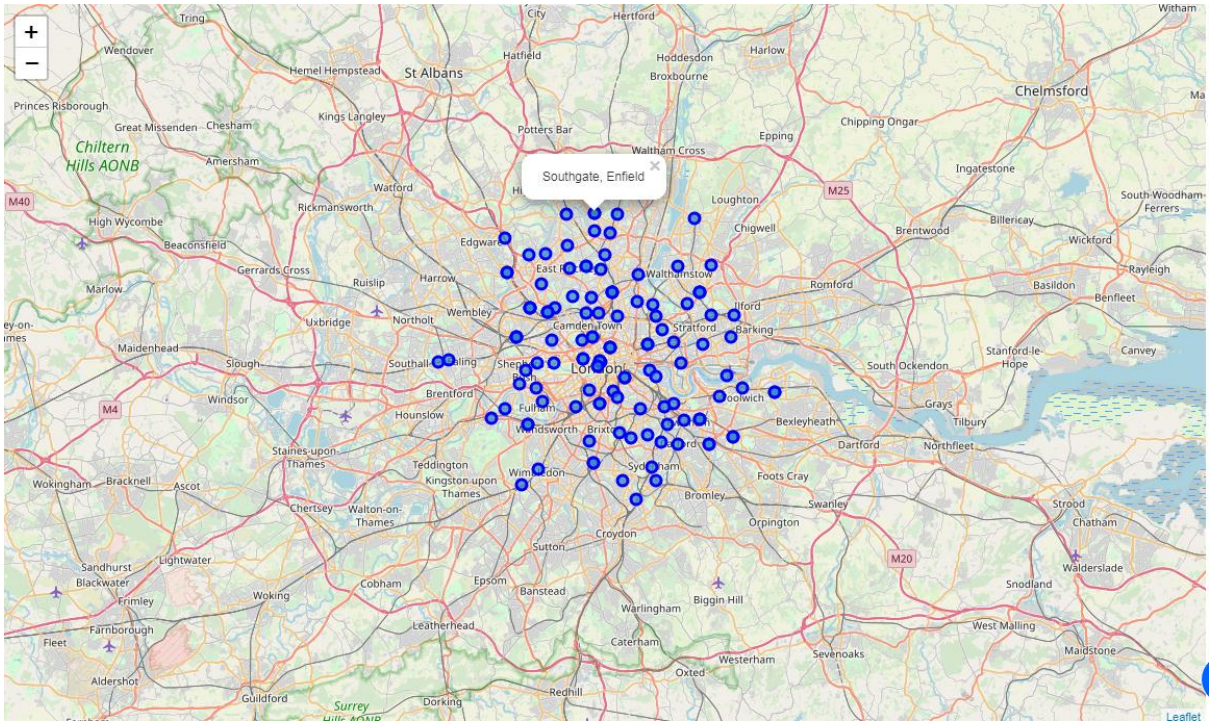
The next step after getting the neighbourhood data was to explore the neighbourhoods. Here we performed exploratory analysis leveraging Foursquare location platform data.

First we defined a function to establish a connection with Foursquare database using Client-id and Secret Client-id. Then using this function we passed latitude and longitude for each neighbourhood and received list of nearby venues in the form of a json file. The relevant information from this json file was saved in a dataframe.

The column of interest was venue.category which gave information on what type of venue was nearby the neighbourhood like Grocery store, Playground, Park, Theatre, Café, Train station, etc. Further exploration revealed that there were 265 unique categories in the entire dataset.

Then leveraging the folium library of python we were able to visualize the neighbourhood and explore it further.

Below is a visualization generated through folium.



Here we could also zoom in and out to explore a particular neighbourhood.

3.3 Machine Learning Model

Now that we have information on nearby venues for each neighbourhood, the next task was to cluster the neighbourhoods so that similar neighbourhoods are identified. For this we used K-means clustering algorithm.

K-means clustering algorithm basically calculates the distance between the entities based on the feature set. The method used for calculating the distance can be Euclidean distance method. K-mean is a iterative algorithm were initially the cluster centres are selected randomly, but then in each iteration these centres are updated based on the distance from

members belonging to that cluster. Finally the algorithm converges when centres location stoop updating, i.e. there is no change in the centre.

In our case the feature set was nearby venues to a neighbourhood. So first this data was converted into numerical data using one hot encoding technique. This is required to calculate distance in K-means algorithm. Then the mean for each category for each neighbourhood was calculated to bring data to row level for neighbourhood. Then finally the clusters were calculated.

4. Results

Now to open a restaurant it is very important to understand "complementary businesses".

By "complementary businesses", we're referring to the neighbouring businesses that could share their market with you, but aren't competition. The activities of complementary businesses actually drive diners to you. Simply being in close proximity to them can increase your foot traffic. Complementary businesses include but aren't limited to: stadiums, theatres, business districts, transportation hubs, and malls.

How complementary businesses benefit you:

Office buildings: Hungry professionals could come to your restaurant for lunch.

Train stations: Cafes around transport hubs can attract caffeine deprived commuters.

Theatres and stadiums: Restaurants can capture attendees pre-and-post performance or a big game

Looking at above essential characteristics we zeroed down a cluster which was a perfect match. It consisted of below 5 neighbourhoods.

Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Brunswick Park,Friern Barnet,New Southgate	Bus Stop	Park	Beer Bar	Metro Station	Pool
Canning Town,Custom House,North Woolwich,Silve...	Airport Service	Sandwich Place	Duty-free Shop	Italian Restaurant	Theater
Charlton	Bus Stop	Grocery Store	IT Services	Thai Restaurant	Pet Store
Forest Gate	Grocery Store	Train Station	Pub	Bakery	Café
Seven Sisters,Tottenham Green,West Green	Coffee Shop	Pub	Bus Stop	Café	Train Station

5. Discussion

We have achieved the goal of this project. However, it is to be noted that this recommendation was based solely on nearby venues and location information. This means the demographic information of the people in the area or price of rentals or any other such information was not considered. Nevertheless this recommendation could be a strong starting point to filter locations of interest and then do further analysis to come up with exact location of choice.

Also note that this model can be tweaked to recommend locations for business other than restaurants. For example, for recommending location to start a gym for instance may require tweaking the filters like proximity to office spaces so that one can hit the gym immediately before or after office.

6. Conclusion

In this project we explored the neighbourhoods of London with the help of location data from Foursquare. We found the popular venues in the vicinity of these neighbourhood. Using this information we then clustered the neighbourhoods and found unique characteristics for each cluster. This helped us in identifying similar neighbourhoods of London. We then found the cluster best suitable to start a Restaurant based on market intelligence. In short we selected neighbourhoods where complimentary businesses support a restaurant.