Coursera Capstone Project

Hrushik Tadvai

June 2020

1 Introduction

Restaurants and Cafes are a great place to relax, eat and socialise with friends and family. The restaurant industry in the UK is worth in excess of £38 billion and more than 86,000 restaurants participate in the sector with most Brits spending a weekly average of around £18 on eating out. They not only become a place for meetup but if successful they become a tourist attraction befitting the local neighbourhood. Restaurants are capable to accelerate tourism growth in the local area and provide an enhanced value-added experience for visitors and local people alike.

A venue such as this generates great income for property developers. My client wishes to enter this booming sector in his local borough Harrow invest into a venue.

Harrow is a well ethically diverse part of London and there is already a tense competition in all types of cuisine. It has a great population of 250,000 with a great infrastructure and multiple successful shopping outlets so there is always a high demand for good quality restaurants.

1.1 Business Problem

My client wishes to invest into a restaurant sector which is proven to be successful, hence risk averse and open where there is less competition from other vendors nearby in the local borough.

My objectives are a follows:

- Identify the best locations In Harrow to establish a new venue.
- Identify which type of cuisine is most popular among locals.

To do this I shall employ use of Visualisation and Exploratory Data Analysis and machine learning techniques such as clustering to identify this and meet my clients objectives.

1.2 Target Audience

This Project is of particular use for property developers, Investors and restaurant owner hopefuls to analyse successful restaurants and Identify the locations where investment and projects should be made. This data and report will only be relevant to the borough of Harrow as each Borough have different demographics and hence different preferences for venues.

2 Data

The following page: https://en.wikipedia.org/wiki/List_of_areas_of_London contains detail of every borough and its neighbourhoods in London. To extract this data I shall use the BeautifulSoup4 and request packages to employ web-scraping techniques to achieve this.

We can then get the latitude and longitude coordinates for each of these neighbourhoods by using the geocoder package in python. This is to help plot the data in a map.

For the venue details I will be using the Foursquare API to gather this data. This will provide me with venues in Harrow and their categories such as location and the type of cuisines which I will then be using for my selection.

3 Methodology

I first extracted the data from the Wikipedia page using request and Beautiful soup packages. The data came out as a long array. I then split this data into multiple lists and then combined them into a pandas data frame. Afterwards I cleaned the data to prevent any repeats and sort the missing cells.

We cannot use this data with the Foursquare API information without the latitude and longitude coordinates. To do so I used the geocoder package which identifies the neighbourhood in my data frame and then subsequently pinpoint its corresponding latitude and longitude coordinates. We can then visualise this using the folium package to create a map and check whether the neighbourhoods are accurately plotted in the map of London.

I then selected Harrow as the centre and then using the FourSquare API I Identified every venue within 2000m radius. Here I conducted some exploratory data analysis to identify the number of venues at each location and identify the most common ones and identify the number of unique categories curated from all the venues. Foursquare will do this. We can use this data to identify the most popular restaurants.

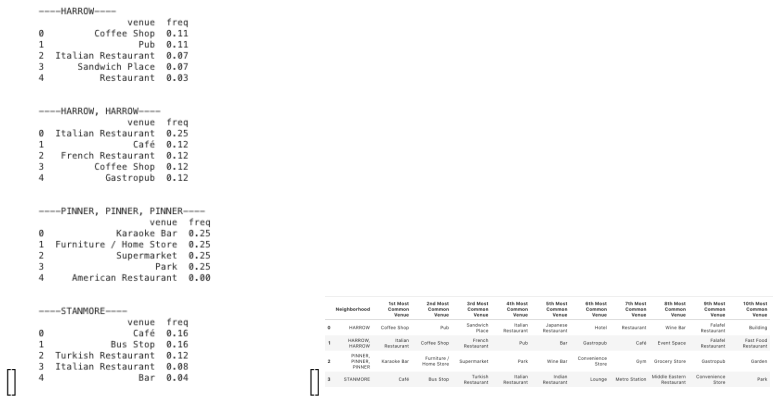
After this we can use the K-clustering machine learning algorithm to identify which neighbourhoods have the highest concentration of venues; this will help answer questions regarding my objectives.

4 Results

In the exploratory data analysis we have identified that the neighbour hood of Pinner has the least amount of venues as shown in the table below.

Neighborhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
HARROW	100	100	100	100	100	100
HARROW, HARROW	24	24	24	24	24	24
PINNER, PINNER, PINNER	4	4	4	4	4	4
STANMORE	25	25	25	25	25	25

Furthermore, through the exploratory data analysis we have identified that the most common types of venue is an Italian restaurant or a cafe.



```
----HARROW----
                venue  freq
0         Coffee Shop  0.11
1                 Pub  0.11
2  Italian Restaurant  0.07
3      Sandwich Place  0.07
4          Restaurant  0.03


----HARROW, HARROW----
                venue  freq
0  Italian Restaurant  0.25
1               Café   0.12
2   French Restaurant  0.12
3         Coffee Shop  0.12
4           Gastropub  0.12


----PINNER, PINNER, PINNER----
                venue  freq
0        Karaoke Bar  0.25
1  Furniture / Home Store  0.25
2         Supermarket  0.25
3                Park  0.25
4    American Restaurant  0.00


----STANMORE----
                venue  freq
0               Café  0.16
1           Bus Stop  0.16
2  Turkish Restaurant  0.12
3  Italian Restaurant  0.08
4                 Bar  0.04
```

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	HARROW	Coffee Shop	Pub	Sandwich Place	Italian Restaurant	Japanese Restaurant	Hotel	Restaurant	Wine Bar	Falafel Restaurant	Building
1	HARROW, HARROW	Italian Restaurant	Coffee Shop	French Restaurant	Pub	Bar	Gastropub	Café	Event Space	Falafel Restaurant	Fast Food Restaurant
2	PINNER, PINNER, PINNER	Karaoke Bar	Furniture / Home Store	Supermarket	Park	Wine Bar	Convenience Store	Gym	Grocery Store	Gastropub	Garden
3	STANMORE	Café	Bus Stop	Turkish Restaurant	Italian Restaurant	Indian Restaurant	Lounge	Metro Station	Middle Eastern Restaurant	Convenience Store	Park

Figure 1:

5 Discussion

The observation clearly indicate that most of the malls are near the harrow neighbourhood therefore to achieve the client's objectives, we select the neighbourhood with least venues which is Pinner. Furthermore K cluster algorithm proved to be ineffective as this is over a small area and there are only 4 neighbourhoods involved so the data from it is irrelevant. From the frequency table in figure 1 it is clear that the client should invest into a cafe or an Italian or Indian restaurants as they proved to be frequent and popular.

Property developers are advised to not invest in harrow neighbourhood where there are multiple venues which indicate that the prices will be high and competition will also be high so unless you are confident in your business, this will prove to be a risky investment

6 Conclusion

In this project I have used many analytical techniques to determine factors which best meet my objectives. I have identified the location and type of cuisine for my objectives and I used the most accurate data available to plot maps and determine the frequency of the categories of each type of venue. The findings on this project will help relevant project managers, chefs and investors capitalize on opportunities in high potential locations where we are avoiding areas of high competition while identifying which type of cuisine is popular in this area.

This project is only relevant to people looking to invest in Harrow as it has a different demographic from other areas of London. In this project we have only considered few factors such as frequency but there are other factors which affect this such as the population of each neighbourhood, average income in each neighbourhood and the demographics of the neighbourhood. Furthermore I had to rely on a free software of foursquare to get my data on venues and it is clear that many were missed so in future projects I would consider whether to invest into a paid service to gather more accurate data.