Coursera Capstone

IBM Applied Data Science Capstone

Opening a New Indian Restaurant in London, United Kingdom

By: Niveditha TR

June 2020

1.   Introduction

London is the biggest city in the UK, by density and by population. It is approximately 600 square miles (1500 square kilometres). As of 2020, London's population stands at approximately 9.3 million. London City in the United Kingdom with a long history of international immigration. They came from many parts of the world according to Wiki data. London approximately estimates a population of around 542,857 or 6.6% of the population. The majority are concentrated in West London, home to London's Hindu community, though populations can be found throughout London.

Indian curry was already very famous in England since 19th century. It's commonly assumed that London's first Indian restaurants were set up by Bangladeshi immigrants in the 1960s and '70s, but actually, the first one was established over 200 hundred years ago, in 1810.

2.   Business Problem

The objective of this capstone project is to analyze and select the best locations in the London city to open a new restaurant.

1. What is the best location for an Indian restaurant in London City?

2. In what Neighborhood should I open an Indian restaurant to have the best chance of being successful?

3.   Target Audience

➤ Business personnel who wants to invest or open a restaurant.
   Freelancer who loves to have their own restaurant as a side business.

➤ Finding the best location for opening a restaurant.

➤ Budding Data Scientists, who want to implement some of the most used Exploratory Data Analysis techniques to obtain necessary data, analyse it and, finally be able to tell a story out of it

# 4. Data Section

For this project we need the following data,

- London data that contains Borough, Neighbourhoods along with their latitudes and longitudes
  1) Data Source : https://en.wikipedia.org/wiki/List_of_areas_of_London

  2) London Boroughs csv data file from internet dataset

- These files contain the neighbourhoods and latitude and longitude data

- Indian restaurants in neighbourhood of London city.
  1) Data Source: Foursquare API
  2) Description: By using this API we will get all the venues in London neighbourhood. We can filter these venues to get only Indian restaurants.

5. Approach

Collect the London city data from

https://en.wikipedia.org/wiki/List_of_areas_of_London

* Using Foursquare API we will get all venues for each neighbourhood.

* Filter out all venues which are Indian Restaurants.

* Analysing using Clustering (Specially K-Means):

1. Find the best value of K
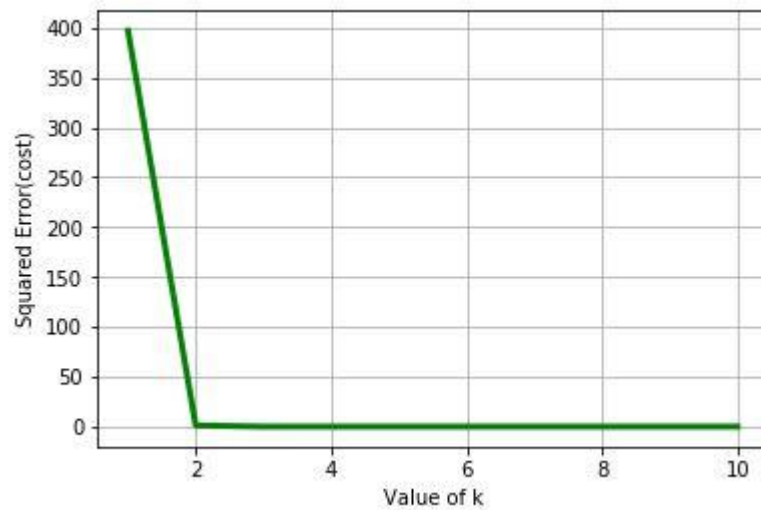
2. Visualize the neighbourhood with number of Indian Restaurants.

Cluster 0 Neighbourhoods

```
In [44]:  1  london_merged.loc[london_merged['Cluster Labels'] == 0]
```

t[44]:

	Neighborhood	Indian Restaurant	Cluster Labels	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 Com V
0	Addington	0.00000	0	Coffee Shop	Pizza Place	Pub	Italian Restaurant	Supermarket	Chinese Restaurant	English Restaurant	Dessert Shop	Diner	Dis
1	Addiscombe	0.00000	0	Coffee Shop	Pizza Place	Pub	Italian Restaurant	Supermarket	Chinese Restaurant	English Restaurant	Dessert Shop	Diner	Dis
2	Albany Park	0.00000	0	Coffee Shop	Clothing Store	Pub	Fast Food Restaurant	Pharmacy	Supermarket	Italian Restaurant	Warehouse Store	Plaza	Ch Resta
3	Aldborough Hatch	0.02381	0	Fast Food Restaurant	Hotel	Supermarket	Clothing Store	Grocery Store	Turkish Restaurant	Sandwich Place	Coffee Shop	Bakery	Depar
4	Aldwych	0.00000	0	Electronics Store	Breakfast Spot	Warehouse Store	Fish & Chips Shop	Dessert Shop	Diner	Discount Store	Donut Shop	Eastern European Restaurant	E Resta
5	Alperton	0.00000	0	Pub	Golf Course	Warehouse Store	English Restaurant	Department Store	Dessert Shop	Diner	Discount Store	Donut Shop	Ea Euro

Cluster 1 Neighbourhoods

Cluster 1

```
5]:  1  london_merged.loc[london_merged['Cluster Labels'] == 1]
```

5]:

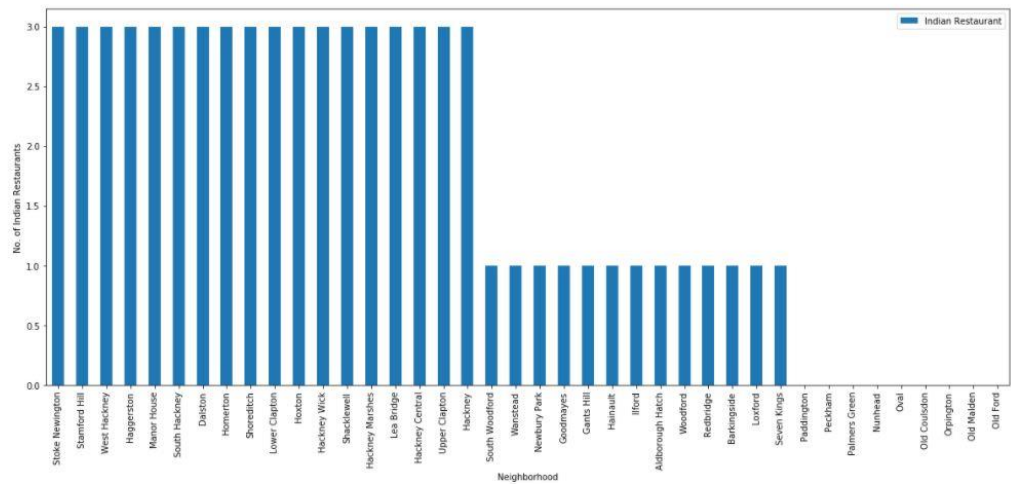|	Neighborhood	Indian Restaurant	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Londo Boroug
85	Cowley	0.0	1	Stables	Warehouse Store	Fast Food Restaurant	Department Store	Dessert Shop	Diner	Discount Store	Donut Shop	Eastern European Restaurant	Electronics Store	Hillingdo
115	Eastcote	0.0	1	Stables	Warehouse Store	Fast Food Restaurant	Department Store	Dessert Shop	Diner	Discount Store	Donut Shop	Eastern European Restaurant	Electronics Store	Hillingdo
162	Harefield	0.0	1	Stables	Warehouse Store	Fast Food Restaurant	Department Store	Dessert Shop	Diner	Discount Store	Donut Shop	Eastern European Restaurant	Electronics Store	Hillingdo
164	Harlington	0.0	1	Stables	Warehouse Store	Fast Food Restaurant	Department Store	Dessert Shop	Diner	Discount Store	Donut Shop	Eastern European Restaurant	Electronics Store	Hillingdo
165	Harmondsworth	0.0	1	Stables	Warehouse Store	Fast Food Restaurant	Department Store	Dessert Shop	Diner	Discount Store	Donut Shop	Eastern European Restaurant	Electronics Store	Hillingdo
176	Hillingdon	0.0	1	Stables	Warehouse Store	Fast Food Restaurant	Department Store	Dessert Shop	Diner	Discount Store	Donut Shop	Eastern European Restaurant	Electronics Store	Hillingdo
186	Ickenham	0.0	1	Stables	Warehouse Store	Fast Food Restaurant	Department Store	Dessert Shop	Diner	Discount Store	Donut Shop	Eastern European Restaurant	Electronics Store	Hillingdo
213	Longford	0.0	1	Stables	Warehouse Store	Fast Food Restaurant	Department Store	Dessert Shop	Diner	Discount Store	Donut Shop	Eastern European Restaurant	Electronics Store	Hillingdo
254	Northwood	0.0	1	Stables	Warehouse Store	Fast Food Restaurant	Department Store	Dessert Shop	Diner	Discount Store	Donut Shop	Eastern European Restaurant	Electronics Store	Hillingdo
289	Ruislip	0.0	1	Stables	Warehouse Store	Fast Food Restaurant	Department Store	Dessert Shop	Diner	Discount Store	Donut Shop	Eastern European Restaurant	Electronics Store	Hillingdo
302	Sipson	0.0	1	Stables	Warehouse Store	Fast Food Restaurant	Department Store	Dessert Shop	Diner	Discount Store	Donut Shop	Eastern European Restaurant	Electronics Store	Hillingdo
310	South Ruislip	0.0	1	Stables	Warehouse Store	Fast Food Restaurant	Department Store	Dessert Shop	Diner	Discount Store	Donut Shop	Eastern European Restaurant	Electronics Store	Hillingdo
												Eastern		

* Inference From these Results and related Conclusions

Lets visualize the maximum number of Indian Restaurants

```
In [46]:  1  london_graph=pd.DataFrame(london_onehot.groupby('Neighborhood')['Indian Restaurant'].sum())
          2  london_graph=london_graph.sort_values(by='Indian Restaurant', ascending=False)
          3  london_graph.iloc[:40].plot(kind='bar', figsize=(20,8))
          4  plt.xlabel('Neighborhood')
          5  plt.ylabel('No. of Indian Restaurants')
          6  plt.show
```

Out[46]: <function matplotlib.pyplot.show(*args, **kw)>

6. Results

The results of the exploratory data analysis and clustering is summarized below:

1. Cluster 0 neighbourhoods has the highest number of Indian restaurants.

2. Cluster 1 neighbourhoods has a high density of Indian restaurants.

3. I will open my restaurant in Hillington neighbourhoods because it is near to Heathrow International Airport. There are many chances all immigrants will come to the nearest restaurant. So, the profit will be more. The results of the clustering are visualized in the map below with cluster 0 in red colour and cluster 1 in purple colour
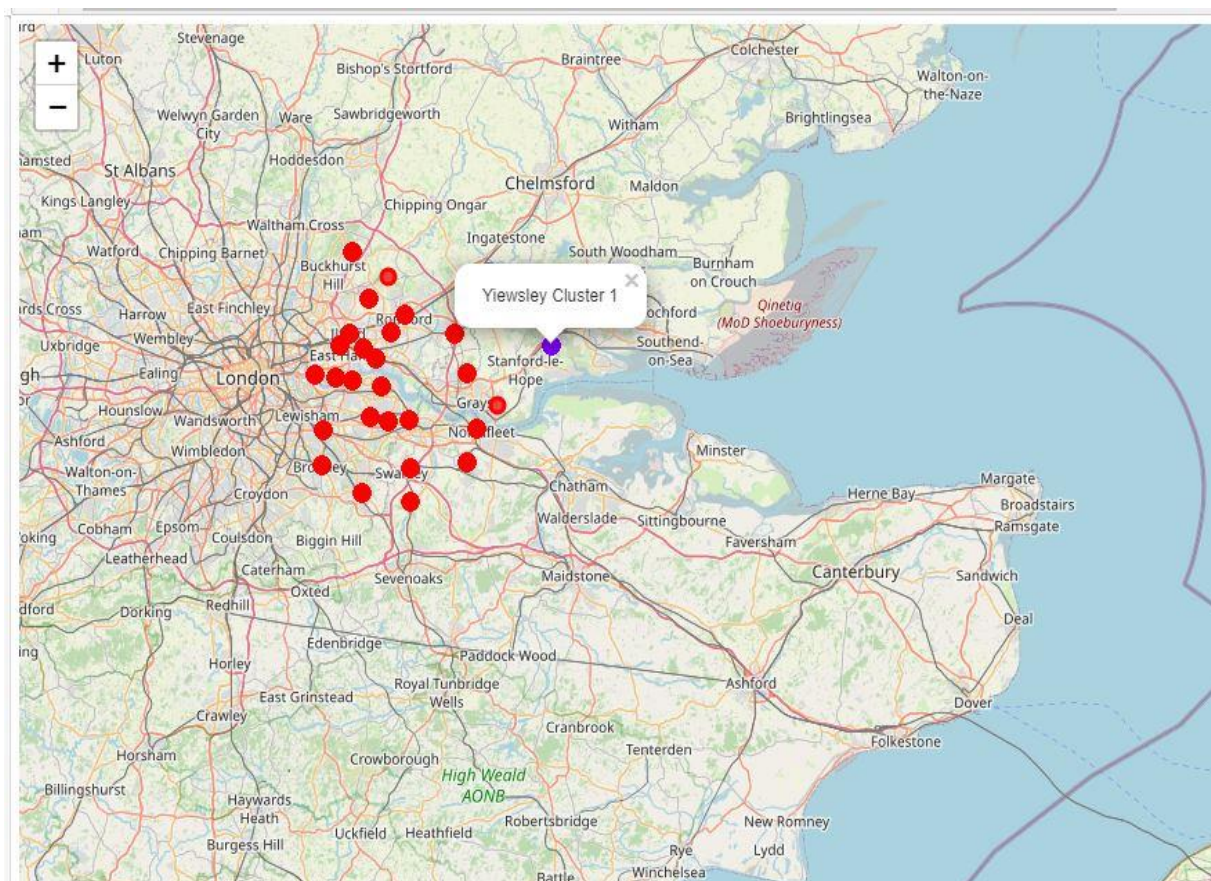
7.Discussion

According to the analysis, South west zone of London will provide the least competition for an upcoming Indian restaurant as the International Airport is close to this neighbourhood. So, all this is the best place for Indian immigrants for having lunch/dinner and the frequency of Indian restaurants is very low compared to other neighbourhoods.

Cluster 0 has the highest number of Indian restaurant and cluster 1 highly dense restaurants.

Some drawbacks of analysis are: the clustering is completely based on the data provided by Foursquare API. Since land price, the distance of venues from the closest station, the number of potential customers, could all play a major role and thus, this analysis is definitely far from being conclusory. However, it definitely gives us some very important preliminary information on the possibilities of opening restaurants in the London boroughs.

Also, another pitfall of this analysis could be consideration of only one major borough of London City, taking into account all the areas under the 5 major boroughs would give us an even more realistic picture. Furthermore, these results also could potentially vary if we use some other clustering techniques like DBSCAN.

8.Limitations and Future Suggestions

In this project, we only consider one factor i.e. frequency of occurrence of Indian restaurants, there are other factors such as population and income of residents that could influence the location decision of a new restaurant

Future research could devise a methodology to estimate such data to be used in the clustering algorithm to determine the preferred locations to open a new

restaurant. In addition, this project made use of the free Sandbox Tier Account of Foursquare API that came with limitations on the number of API calls. Future research could make use of paid account to bypass these limitations and obtain more results.

9.Conclusion

Finally, to conclude this project, we have got a small glimpse of how real-life Data science project looks like. I have used some frequently used python libraries to handle JSON file, plotting graphs, and other exploratory data analysis. Use Foursquare API to major boroughs of London City and their neighbourhoods. Potential for this kind of analysis in a real-life business problem is discussed in great detail. Also, some of the drawbacks and chances for improvements to represent even more realistic pictures are mentioned. As a final note, all of the above analyses are depended on the adequacy and accuracy of Four-Square data. A more comprehensive analysis and future work would need to incorporate data from other external databases.