Data Science Capstone Project

Exploring Pizza restaurants in London using Foursquare API & Machine Learning

Name: Priya Kate

Introduction : Business problem

There are many Pizza restaurants in London who have rating and user tips. As a part of data science consulting project, I have been assigned analytics work to find optimal location from Central London having good ratings and maximum user tips. Since, there are lots of pizza restaurants in London, my project reports will help Pizza lovers who are new in London. We would also prefer locations as close to city center as possible.

Data Discussion:

The data has been collected from Foursquare API. After importing necessary libraries, to fetch the data from Foursquare API, we used the client ID and client secret. We added address as London and found the longitude and latitude.

Then with a suitable radius we added "Pizza" as a search query and the via URL we got the data in Jason file which in turn we converted into a dataframe

We found 30 pizza restaurant near London city center with their location details & address. Following is a cut snapshot of the data we found:

	id	name	categories	referralId	hasPerk	location.address	location.lat	location.lng	location.labeledLatLn
0	4b5061def964a520cb2127e3	Pizza Express	[{'id': '4bf58dd8d48988d1ca941735', 'name': 'P...	v-1624102622	False	450 Strand	51.508464	-0.126157	[{'label': 'display', '! 51.5084640038466
1	4bc6005242419521ca66031d	Pizza Hut	[{'id': '4bf58dd8d48988d1ca941735', 'name': 'P...	v-1624102622	False	56/59 Strand	51.509410	-0.123401	[{'label': 'display', '! 51.50941028020455
2	4b5e09fef964a520747a29e3	Pizza Express	[{'id': '4bf58dd8d48988d1ca941735', 'name': 'P...	v-1624102622	False	85 Victoria Street	51.497458	-0.135103	[{'label': 'display', '! 51.4974582, 'lng
3	4c41b1e6520fa5935af3c9ac	Pizza Hut	[{'id': '4bf58dd8d48988d1ca941735', 'name': 'P...	v-1624102622	False	Piccadilly Circus, 29-31 Regent Street	51.509348	-0.134752	[{'label': 'display', '! 51.5093476677347
4	4bcd8f10fb84c9b6f022223e	Pizza Hut	[{'id': '4bf58dd8d48988d1ca941735',	v-1624102622	False	19 Leicester Square	51.510849	-0.129475	[{'label': 'display', '! 51.510849, 'lng

Methodology:

In Python, we used seaborn and matplotlib libraries for exploratory data analysis (EDA). The following horizontal bar graph indicates the Pizza restaurants with column length as distances from city center.
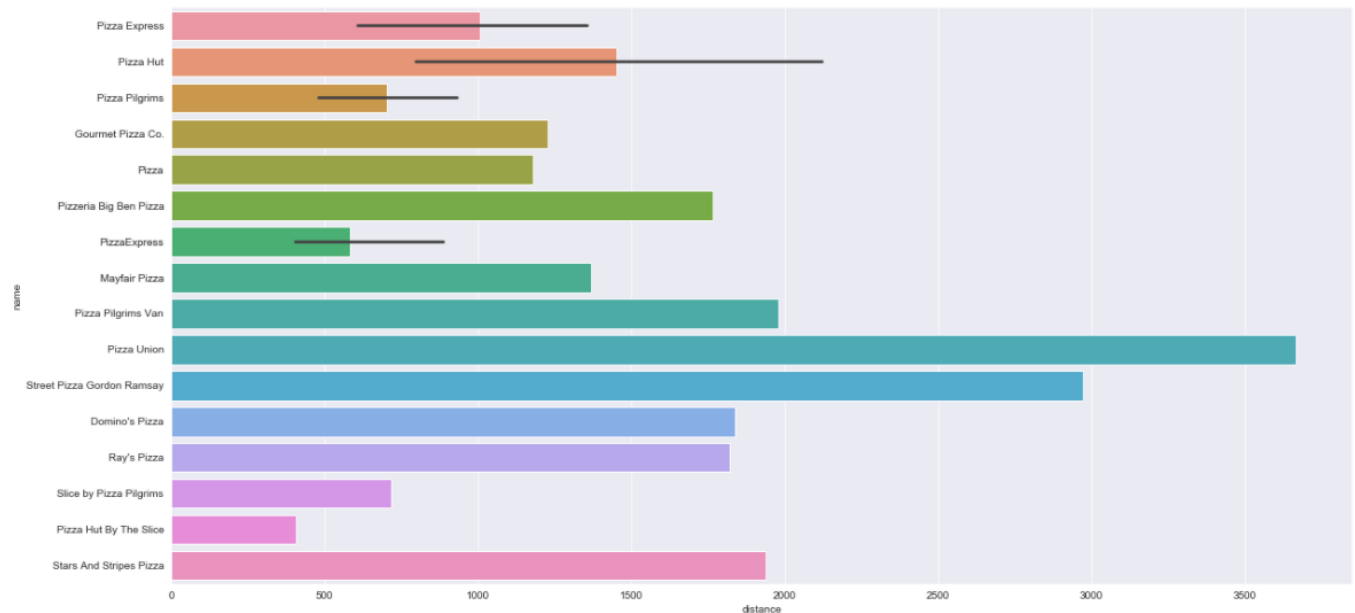
Then we used Folium library to visualize the geographical details of these restaurants on city map of London. This is mentioned in below map:
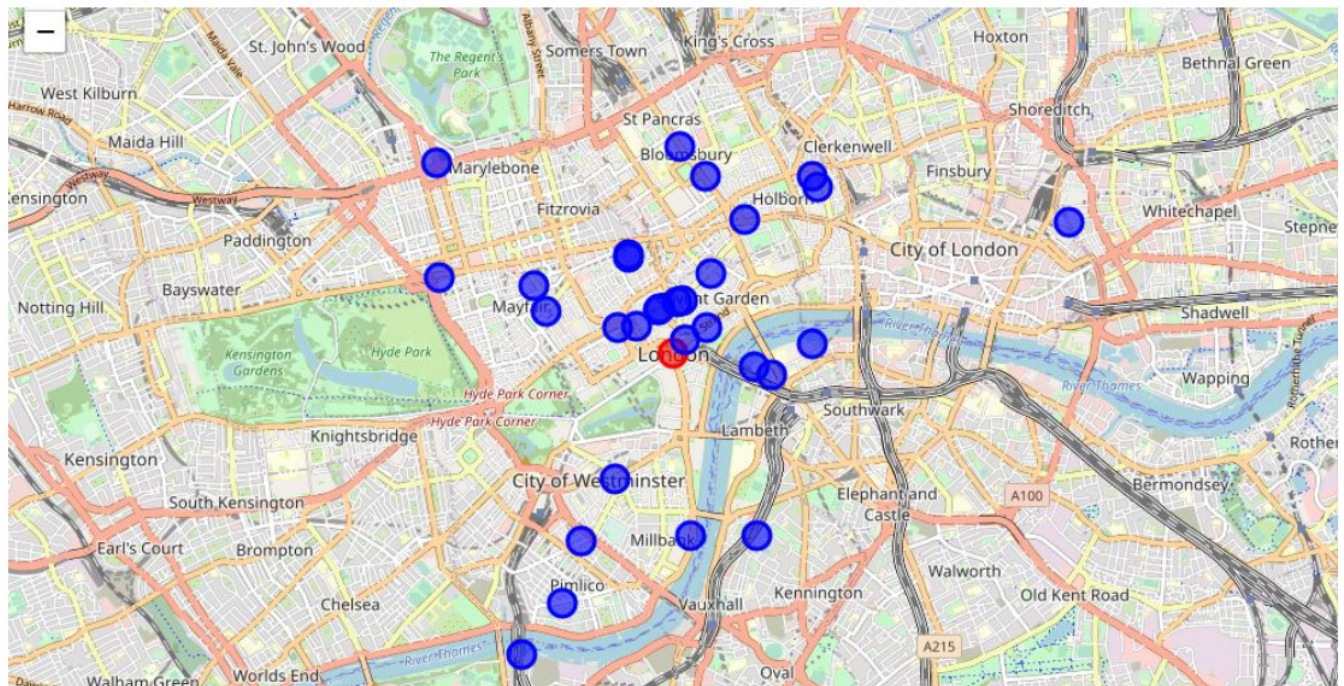
Then we explored the closest Pizza Restaurants using the venue ID which resulted into all the details of the particular restaurant.

Then we checked the rating for each restaurant. We also found the number of tips each restaurant has received from customers.

Then we explored tips and list of associated features followed by formatting the column width and display all tips as below:

	text	agreeCount	disagreeCount	id	user.firstName	user.lastName	user.id
0	The best Neapolitan-style pizza you can have in London. The interior gets quite warm during the summer – have a seat outside or choose take away and go to Soho Square. Good service, great value.	4	0	53e25059498e1550f3853f85	Lele	Buonerba	6164189

Then we defined the URL, send "GET request" and displayed features associated with user by user ID with most agree counts and complete profile as below:

	id	prefix	suffix	width	height
0	5864cf4ee386e329a41e9078	https://fastly.4sqi.net/img/general/	/6164189_eHrCO8bvc-DQQgUE3He8x8jnuiePgiHZWcdnhgrg6Tc.jpg	1440	1920

After the user profiling, we explored all the tips provided by that user for exploring addition places and get expert tips on dishes for the chose restaurants.

	text	agreeCount	disagreeCount	id
0	Pizza with tomato sauce and a whole form of burrata cheese is spectacular. Highly recommend this place! Get a reservation.	0	0	58612bf92321c6675b596457
1	Free wifi is only up to 30min. Terminal 8 AA bag drop off has long lines but they move quickly. Security was fine. Below average food options at T8 – McDonalds or Aunt Annie's are your best choices.	1	0	57de8c2c498ea040bf1eba40
2	Good: you can get here with a train from Centrale, Garibaldi or Cadorna. Bad: if you're flying to the US (I was on AA) the Terminal 1 check-in area's placement makes no sense at all. Okay overall.	0	0	57de8b83498ea711b76a4822
3	My favorite discotheque in Milan. Beautiful atmosphere, I especially love the smaller room (pictured). Don't miss out on NUL – Friday nights, €15 entry with a drink and coat check, great fun and DJs.	1	0	57de8aef498e40323c6a22a5
4	Absolutely loved this place! After three hours of dancing at Output's rooftop, delicious Mediterranean food was just what I needed. It's not on the list but you can order arak! Very peculiar drink.	0	0	57de89c5498e40323c67ec68
5	Loved the cortado! Great atmosphere but I didn't have the patience to wait to be seated for Sunday brunch, although it looks like it's well worth it!	0	0	57de88f8cd104ad3b630c178
6	The best record store I've been to in New York. Fantastic selections of secondhand records in very good condition. I bought a whole lot of disco 12" for extremely fair prices (most for just $3).	1	0	57de88ae498ed3f735ef44f0

It also includes the total number of agreements by other users on each tip.

It also helps in exploring nearby places in London.

Then we got the venue for the tip with the greatest number of agree counts. And with the tip ID we found the most recommended restaurant. In our case most of the agreements on 2 different tips indicated the following restaurant:

```
Bottega Baretti
{'address': "Via Sant'Anselmo 28/F", 'crossStreet': 'Via Baretti', 'lat': 45.057895312490075, 'lng': 7.680486155038875, 'labele
dLatLngs': [{'label': 'display', 'lat': 45.057895312490075, 'lng': 7.680486155038875}], 'postalCode': '10125', 'cc': 'IT', 'cit
y': 'Torino', 'state': 'Piemonte', 'country': 'Italia', 'formattedAddress': ["Via Sant'Anselmo 28/F (Via Baretti)", '10125 Tori
no Piemonte', 'Italia']}
```

Following is the datatypes and null values present in our dataframe:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 18 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   ID               30 non-null     int64
 1   Name             30 non-null     object
 2   Distance         30 non-null     int64
 3   Average_rating   30 non-null     float64
 4   Number_of_tips   30 non-null     int64
 5   address          28 non-null     object
 6   lat              30 non-null     float64
 7   lng              30 non-null     float64
 8   postalCode       29 non-null     object
 9   labeledLatLngs   30 non-null     object
 10  cc               30 non-null     object
 11  city             30 non-null     object
 12  state            30 non-null     object
 13  country          30 non-null     object
 14  formattedAddress 30 non-null     object
 15  neighborhood     2 non-null      object
 16  crossStreet      8 non-null      object
 17  id               30 non-null     object
dtypes: float64(3), int64(3), object(12)
memory usage: 4.3+ KB
```

Following is a high-level statistical summary of all the variables:

	ID	Distance	Average_rating	Number_of_tips	lat	lng
count	30.000000	30.000000	30.000000	30.000000	30.000000	30.000000
mean	14.500000	1341.033333	4.560000	23.433333	51.509119	-0.127926
std	8.803408	863.541924	2.908442	36.081023	0.010481	0.015841
min	0.000000	163.000000	0.000000	0.000000	51.483481	-0.157777
25%	7.250000	585.250000	1.250000	1.250000	51.506779	-0.135016
50%	14.500000	1220.500000	5.800000	17.000000	51.510796	-0.127090
75%	21.750000	1832.500000	6.250000	25.000000	51.515069	-0.119754
max	29.000000	3666.000000	8.500000	177.000000	51.523658	-0.077416

There is a strong correlation between "Average user rating" and "Number of tips".

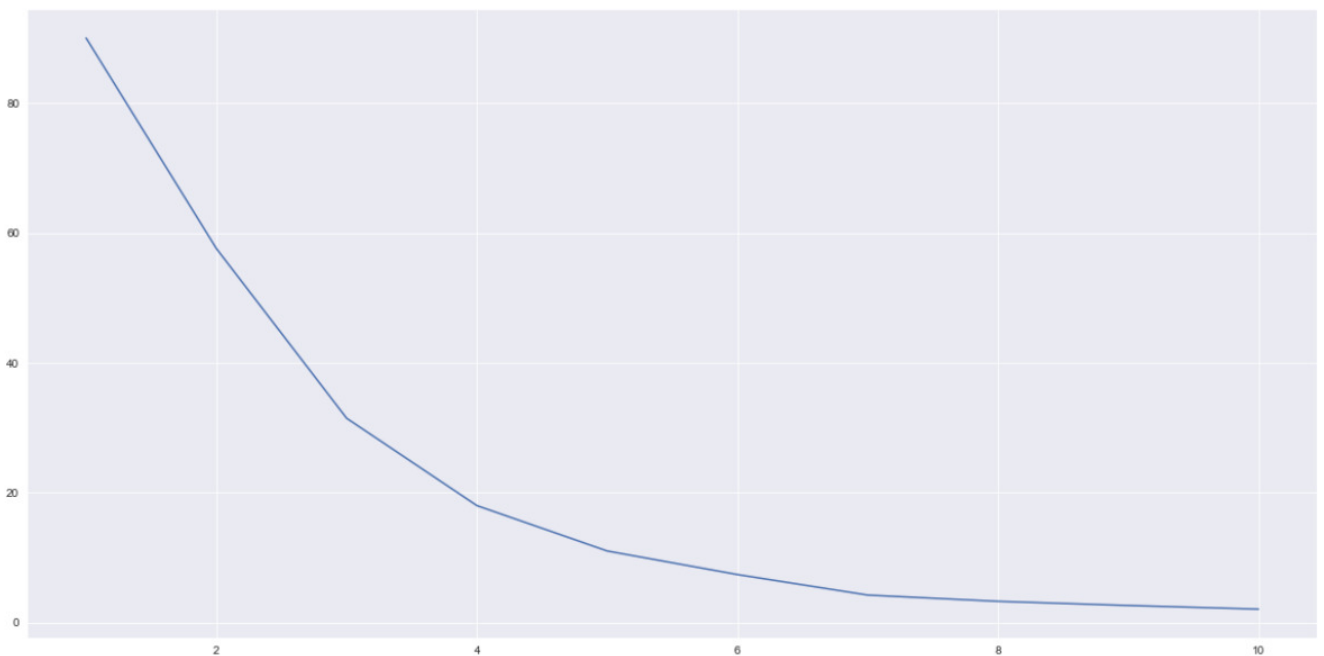Distance has a vey less relationship with ratings or tips.

This is evident form the following the Heatmap:
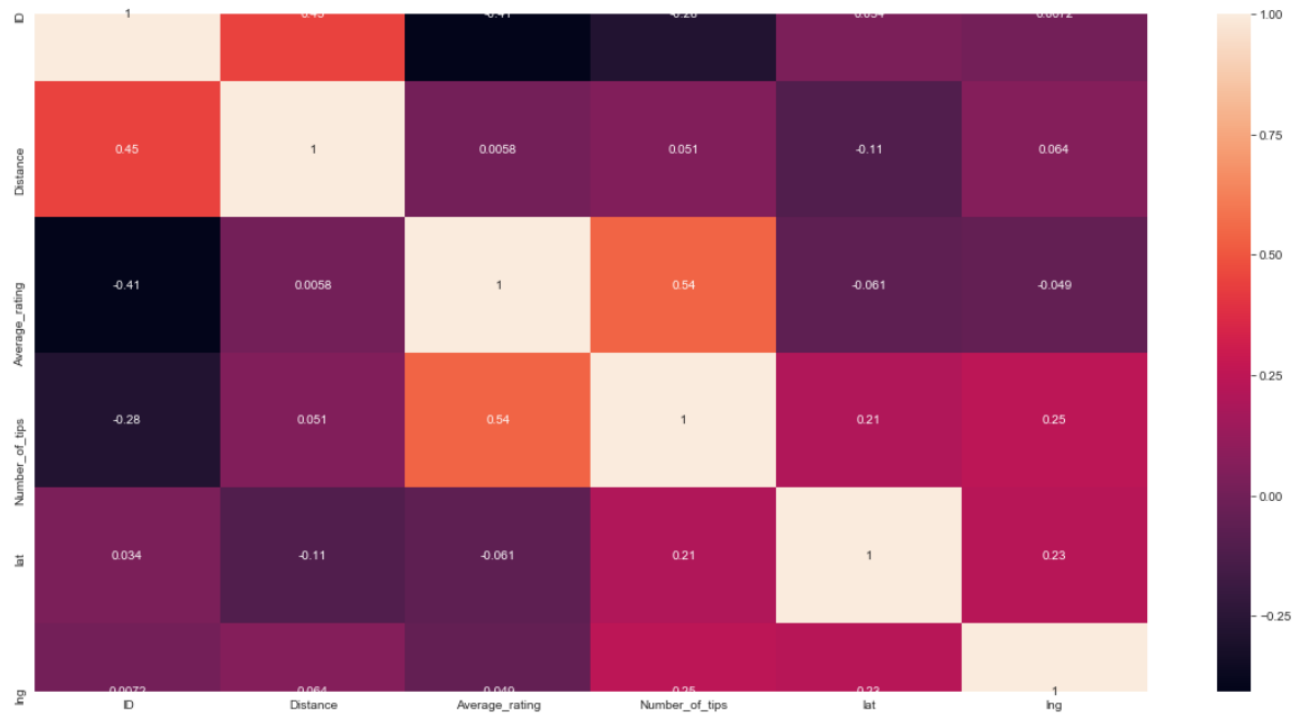
Clustering:

We used K mean clustering for grouping the pizza restaurants based on distance, ratings and number of tips.

Following is a WSS plot based on inertia values for each number of clusters.

We can observe that, based on the elbow of the graph, there should be 4 number of clusters.

We verified this by calculating a silhouette_score.

The silhouette_score for:

For 3 number of Clusters is = `0.51643`
For 4 number of Cluster is = `0.51767`

For 4 number of clusters the silhouette_score is higher so, we chose 4 clusters.

Results & Discussion:

Following is the dataframe along with appended cluster column:

	Name	Distance	Average_rating	Number_of_tips	Clus_kmeans3
0	Domino's Pizza	1837	0.0	0	1
1	Gourmet Pizza Co.	1228	6.3	45	0
2	Mayfair Pizza	1370	6.6	28	0
3	Pizza	1178	0.0	0	1
4	Pizza Express1	163	5.3	24	0
5	Pizza Express2	1213	5.8	30	0
6	Pizza Express3	772	6.3	12	0
7	Pizza Express4	953	5.9	36	0
8	Pizza Express5	1330	5.1	18	0
9	Pizza Express6	1606	5.9	16	2
10	Pizza Hut By The Slice	405	0.0	0	1

Following is the summary of Cluster Profiles:

Clus_kmeans3	Distance	Average_rating	Number_of_tips	freq
0	755.857143	5.978571	24.785714	14
1	1455.000000	0.000000	0.750000	8
2	2235.500000	6.066667	10.500000	6
3	2298.000000	8.350000	143.500000	2

- The First Cluster has highest number of restaurants and have shortest distance and average ratings and tips.
- The Second cluster has least ratings and tips
- The third cluster has average distance, ratings and tips
- The fourth cluster has highest rating and tips but is farthest from the city center of London.

Conclusion:

- The customer should prefer the 4th cluster restaurants if they have time to spare because these two restaurants have great user ratings and highest number of tips.
- The customer in hurry should choose the first cluster of restaurants because they are nearest and have good rating
- The second cluster restaurants should be avoided in all circumstances
- The third cluster restaurants can be visited based on the cost and other factors. Its more of a choice of the customer.

The study conducted in this project is very helpful for people who are new in London and want to explore good quality restaurants. This study recommends the restaurants based on various parameters such as distance form city center, Average user rating and total number of tips from reviewers. It also indicates the location of restaurants on the city map of London for easy accessibility of customers.

In this way, this project provide good insights on Pizza restaurants in London for tourists.