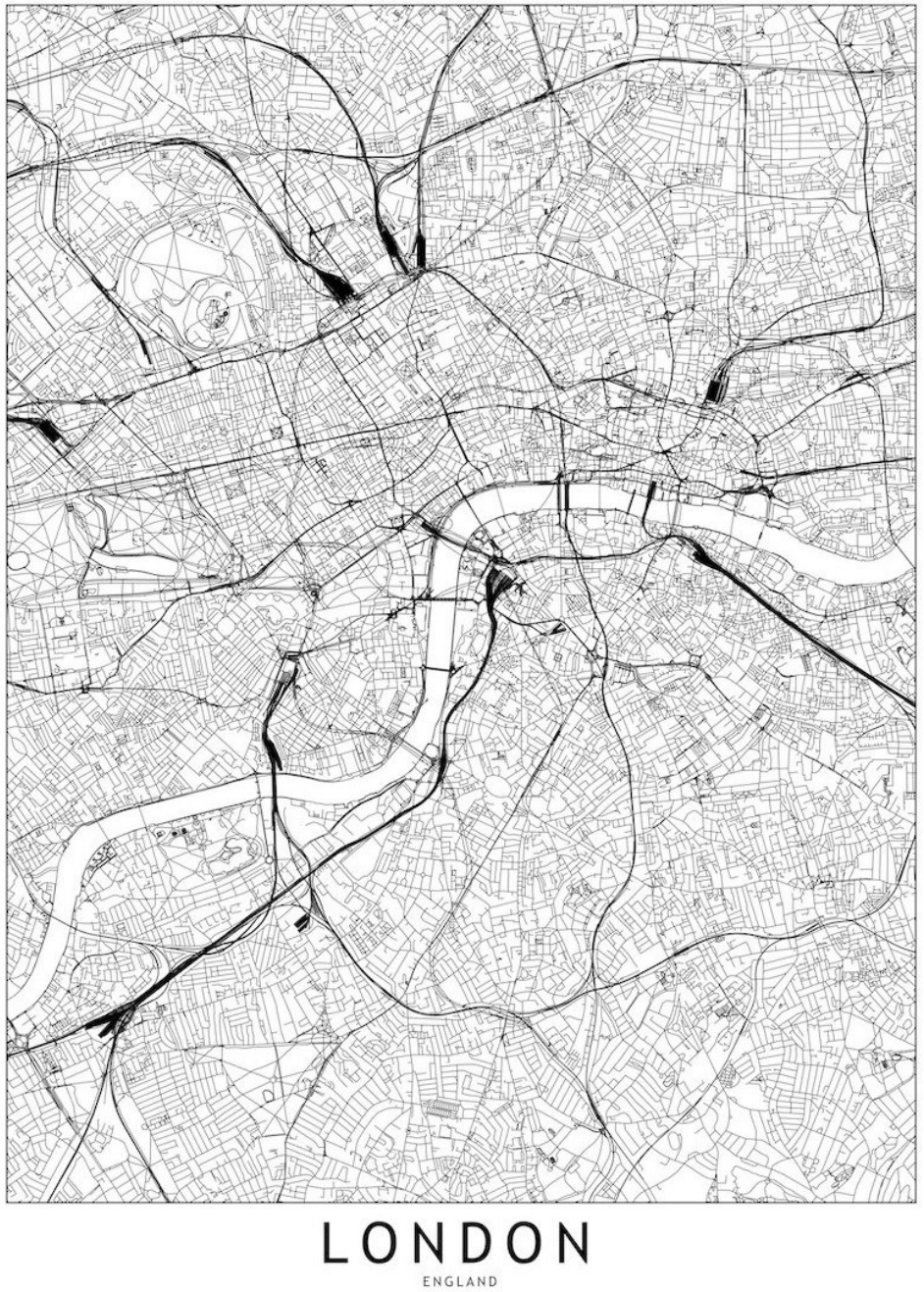


Applied Data Science Capstone

Finding the best place to open a new restaurant in London

---

by
Pegah Bangiantabrizi

Introduction

The restaurant and leisure industry in London is growing exponentially, The demand in the culinary industry has become very high, and as a result, has the extent of competition, in order to open a restaurant in a trendy area of this city.

Business Problem

- The primary factor that you should consider is to find an appropriate location for this new restaurant and, as well as to be aware of the possible nearby and local competitors.

- In this work, we will implement the fundamental analysis, in order to find the optimal London Borough, in which the restaurant can be opened. This is conducted according to the criteria

Data
The Data Description

To find a solution to the problem, and to be able to build a recommender model, first we consider the following:

1) Its geographical coordinates (i.e., latitude and longitude) to find out where exactly the venues are located.

2) In order to access the location of a restaurant, its latitude and longitude should be known, so that we can point to its coordinates, and further create a map displaying all the restaurants with its labels, respectively.

Data
The Data Cleaning

Based on the criteria listed above, the following data is utilized in this analysis:

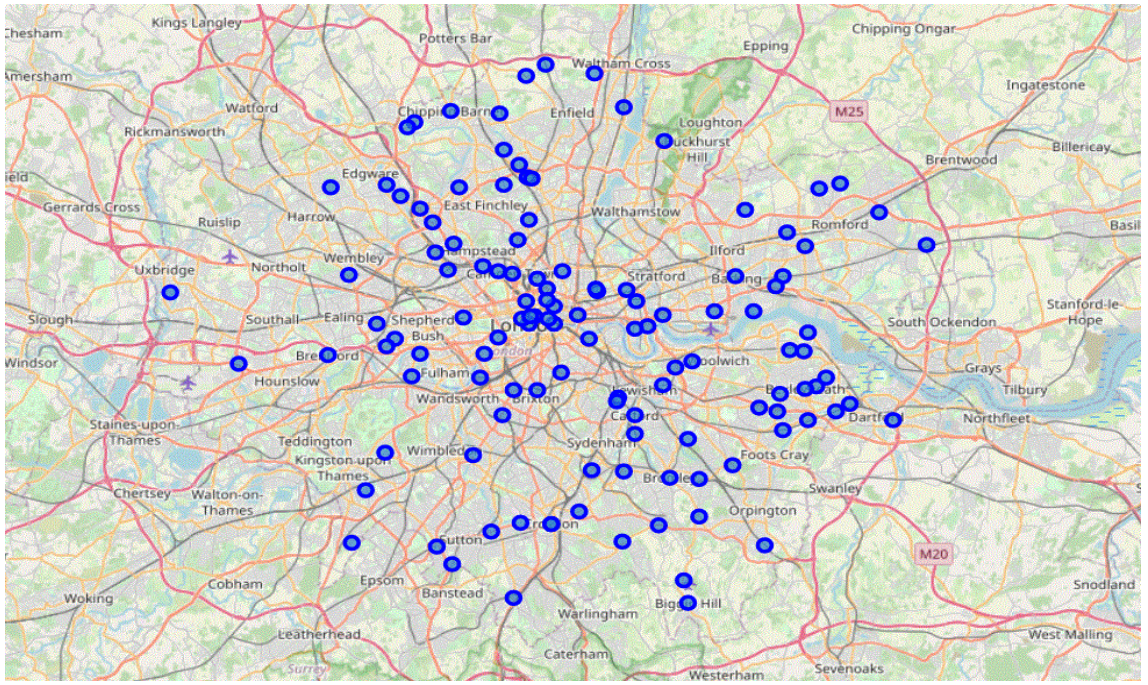
The deployment of BeautifulSoup, in order to extract the data from Wikipedia, and to further provide the relevant information on the London boroughs, i.e., also known as the local authority districts. Besides, the local areas or neighbourhoods are considered for each borough for the detailed analysis.

The Foursquare API https://foursquare.com/, to extract the relevant information on the available restaurants, for a given neighbourhood and borough in London. This API also provided information about the restaurant styles based on cuisine.

The utilized data provided by the UK Government available at data.london.gov.uk to get detailed insights on the London boroughs.

Methodology

Geopy and Folium libraries coordinates of all the locations is achieved and mapped geospatial data on the London map.



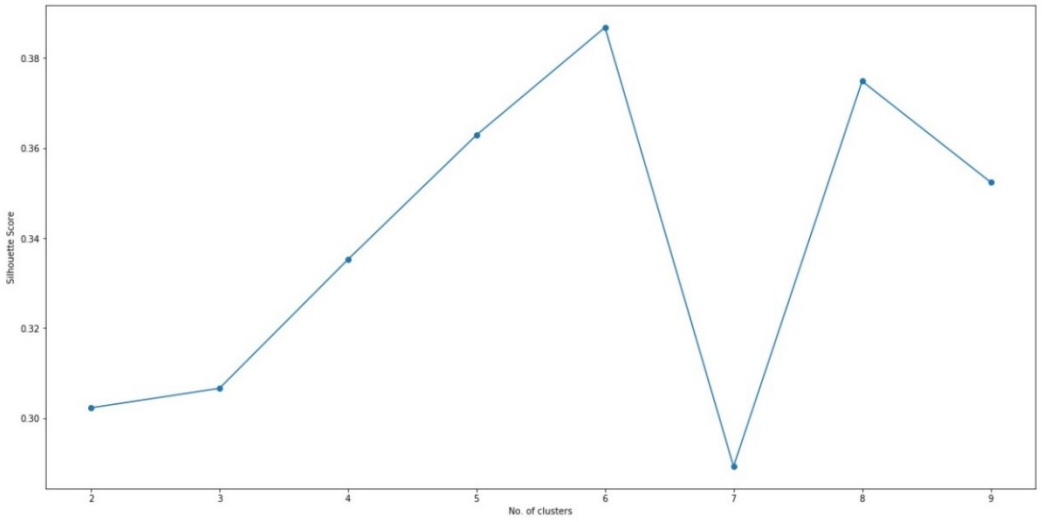
Feature Extraction

```
# one hot encoding
London_onehot = pd.get_dummies(London_restaurants[['Venue Category']], prefix="", prefix_sep="")

# add neighborhood column back to dataframe
London_onehot['Neighborhood'] = London_restaurants['Neighborhood']

# move neighborhood column to the first column
fixed_columns = [London_onehot.columns[-1]] + list(London_onehot.columns[:-1])
London_onehot = London_onehot[fixed_columns]
```

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Austrian Restaurant	Brazilian Restaurant	Cantonese Restaurant	...	Sushi Restaurant
9	Acton	0	0	0	0	0	0	0	0	0	...	0
16	Acton	0	0	0	0	0	0	0	0	0	...	0
24	Addington	0	0	0	0	0	0	0	0	0	...	0
32	Addiscombe	0	0	0	0	0	0	0	0	0	...	0
34	Addiscombe	0	0	0	0	0	0	0	0	0	...	0

5 rows × 68 columns

Unsupervised Learning: K-means




```
opt = np.argmax(scores) + 2 # Finds the optimal value
opt
```

```
6
```

```
# set number of clusters
kclusters = opt

London_grouped_clustering = London_grouped.drop('Neighborhood', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(London_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

```
array([0, 3, 1, 2, 4, 4, 2, 1, 4, 4], dtype=int32)
```

Results

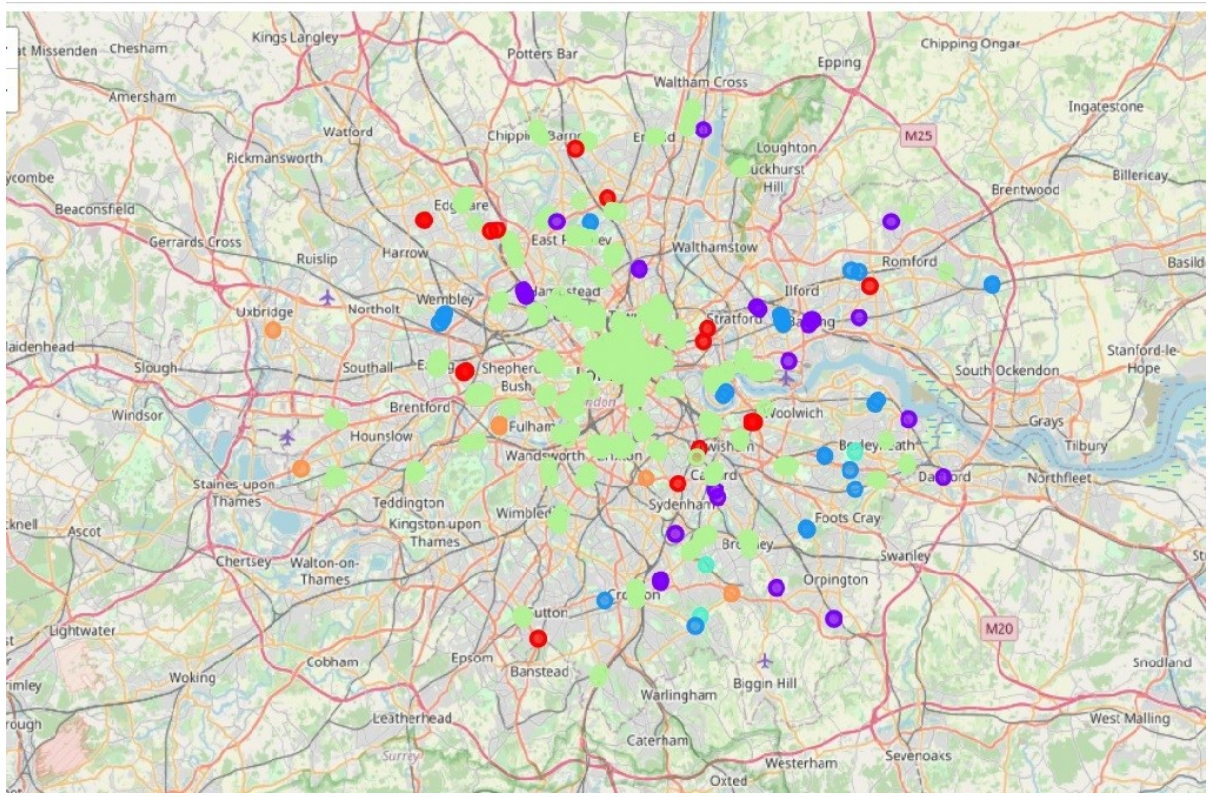
- *The results show that there are 971 restaurants in London with 67 different style of cuisines.*
- *The detail shows the number of neighborhoods assigned to each cluster.*
- *Cluster 4 indicates neighborhoods with the highest concentration of restaurants with the amount of 889*
- *Cluster 3 indicates neighborhoods with the least number of restaurants with the amount of 3.*

Number of venues belonging to each cluster:

```
London_merged['Cluster Labels'].value_counts()
```

```
4    889
2     29
1     27
0     17
5      6
3      3
Name: Cluster Labels, dtype: int64
```

Map of Clusters



Discussion

- The analysis of the results shows that London overall has a high frequency of restaurants within its neighbourhoods, In terms of the clustered data.
- By further looking at the frequency of restaurants based on locations, the most common restaurants in each cluster have been located and as a result, The best boroughs to open a restaurant with a specific cuisine (i.e., most favourite cuisine of the borough) can be chosen.
- The analysis recommends neighbourhoods in a specific cluster can be chosen to open a restaurant which is the least common in that specific region, which would result in profitability and less competition.

Conclusion

- The presented work has been conducted based on a small sample cluster, for the sole purpose of depicting the program's feasibility potential. It can further be extended into larger data sets, and as a result, the program would result in more output results and findings.
- The outcomes would also provide the relevant information on the number and intensity of the restaurants in certain boroughs, which can be interpreted as which areas would be of higher demands in terms of the popular regions for the intended cuisine and leisure industry.
- Moreover, a number of other factors regarding these places can be included in order to distinguish or narrow down the results, such as the restaurants' type (e.g., Italian, Asian, etc.) or hygiene rate.

Thank You!