

Analysing Neighbourhoods of Manchester for starting a new restaurant.



Hend Emad
IBM Data Science Professional Certificate

Introduction



- ★ Manchester is one of the biggest cities in UK.
- ★ It is United Kingdom's second most populated urban area, with a population of 3 million that comprises people of various ethnicities from all over the world.

Business Problem

- Start a restaurant.
- Neighbourhood that is most likely to give a good business.

Data

- ❑ Neighbourhoods of Manchester
 - Neighbourhoods of Manchester wikipedia page through data scraping.
- ❑ Geographical coordinates of the neighbourhoods.
 - Using Geopy library.
- ❑ Venue data from FourSquare.
 - Using FourSquare API.

Methodology

1. Feature Extraction

(One Hot Encoding)

```
import pandas as pd
man_hot = pd.get_dummies(explore_man[['Venue Category']], prefix = "", prefix_sep = "")

# Add neighbourhood column back to dataframe
man_hot['Neighbourhood'] = explore_man['Neighbourhood']

#Move neighbourhood column to the first column
fixed_columns = [man_hot.columns[-1]] + man_hot.columns[:-1].value.tolist()
man_hot = man_hot[fixed_columns]

man_hot.head()
```

Unsupervised Learning

K-Means Clustering

```
max_range = 15 #max range 15 (number of clusters)
from sklearn.metrics import silhouette_samples, silhouette_score

indices = []
scores = []

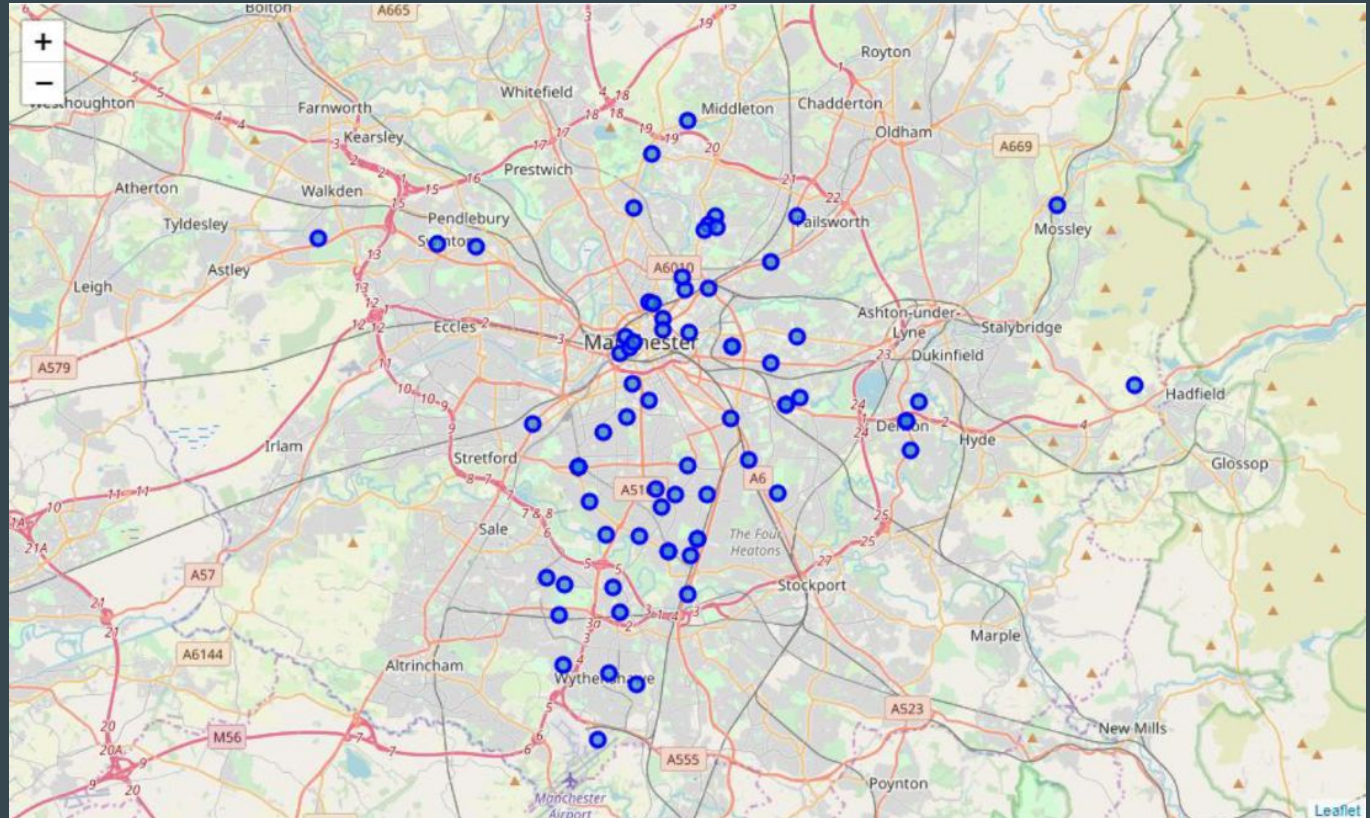
for man_clusters in range(2, max_range):
    # Run K-Means clustering
    man_gc = man_grouped_clustering
    kmeans = KMeans(n_clusters = man_clusters, init = 'K-means++', random_state = 0).fit_predict(man_gc)

    # Gets the score for the clustering operation performed
    score = silhouette_score(man_gc, kmeans)

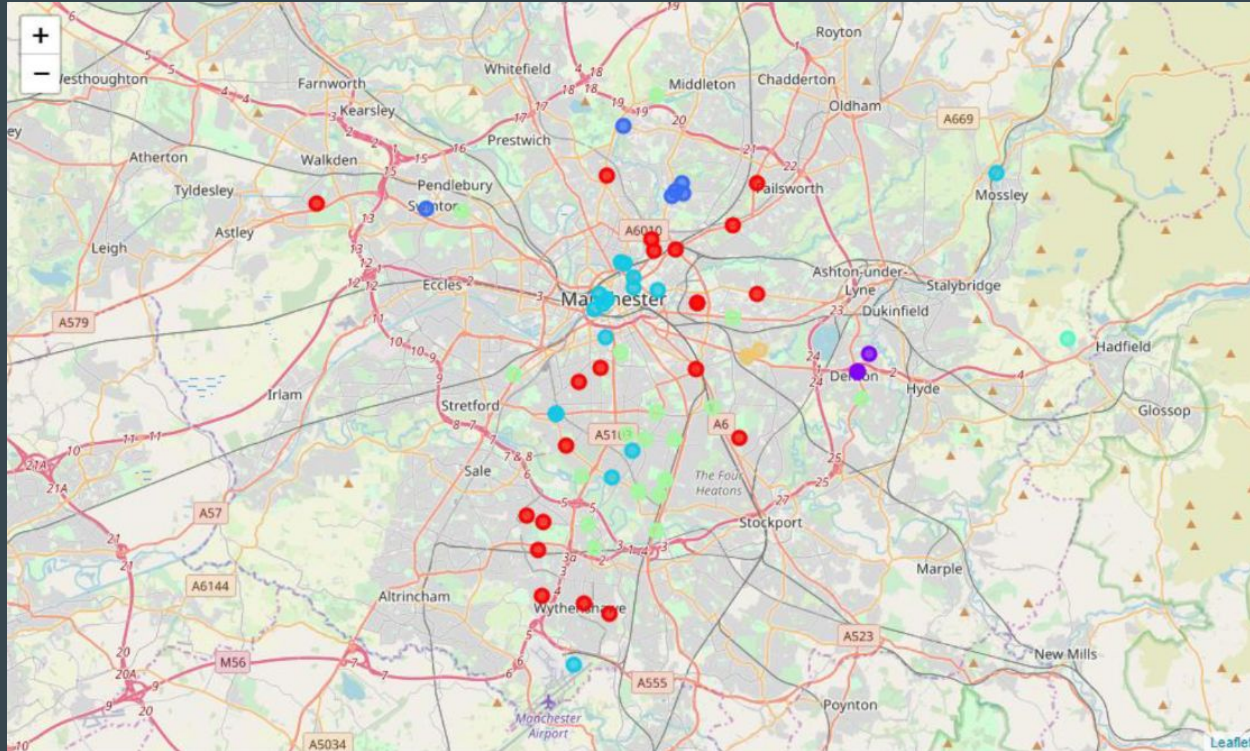
    # Appending the index and score to the respective lists
    indices.append(man_clusters)
    scores.append(score)
```

Plotting

- Folium



Results - - Visualization of clusters



- After visualizing the clusters, the individual clusters were studied and some important conclusions were derived. The neighbourhood that had the most number of restaurants was cluster number 4.

Discussion

- Most suitable neighbourhoods for starting the restaurant business are present in the cluster no. 4.
- Our K-Means model worked perfectly and successfully clustered similar neighbourhoods together.
- After studying all four clusters, it is recommended to the client that neighbourhoods such as Barlow Moor, bookelands and Hyde Newton (ward) that are in cluster 4 look like good locations for starting their restaurant business.
- The client can go ahead and make a decision depending on other factors like availability and legal requirements that are out of scope of this project.

Conclusion

- Data Analysis and machine learning techniques used in this project can be very helpful in determining solutions of certain business problems.
- Python inbuilt libraries such as Geopy, Folium and BeautifulSoup make it very easy and effective to analyse a geographical location.
- In this project we studied the neighbourhoods of Manchester city and came up with a recommendation of neighbourhoods where our client can start their restaurant business.