

Analysing Neighbourhoods of Manchester For Starting A New Restaurant

Asif Khan Shakir

IBM Applied Data Science

12.11.2020

Introduction



- Manchester: One of the biggest cities in UK
- United Kingdom's second-most populous urban area, with a population of 2.9 million
- The population comprises of 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

- Feature Extraction

- One Hot Encoding

```
man_1hot = pd.get_dummies(explore_man[['Venue Category']], prefix="", prefix_sep="")

# Add neighbourhood column back to dataframe
man_1hot['Neighbourhood'] = explore_man['Neighbourhood']

# Move neighbourhood column to the first column
fixed_columns = [man_1hot.columns[-1]] + man_1hot.columns[:-1].values.tolist()
man_1hot = man_1hot[fixed_columns]

man_1hot.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 = []

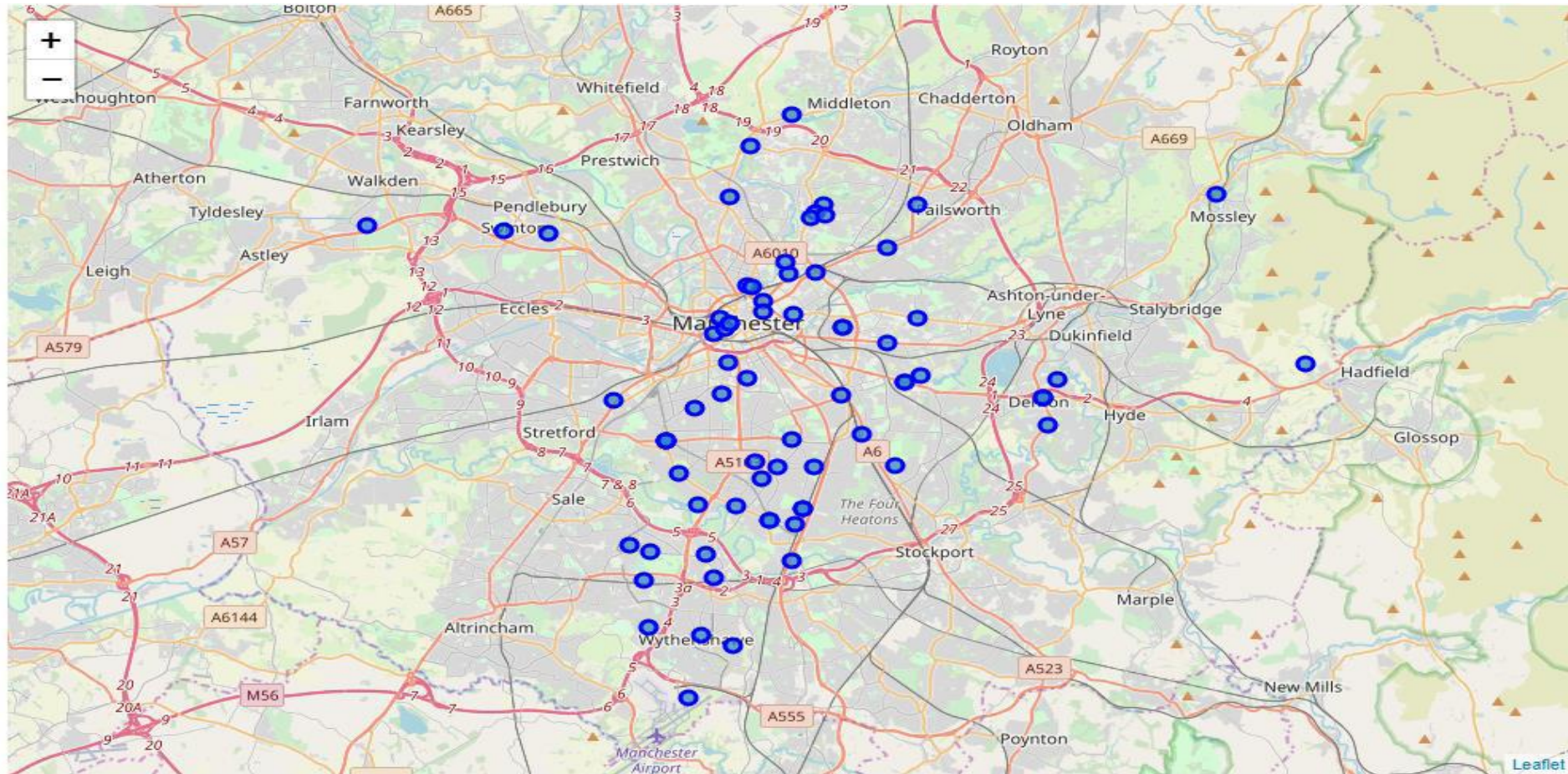
for man_clusters in range(2, max_range) :

    # 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)

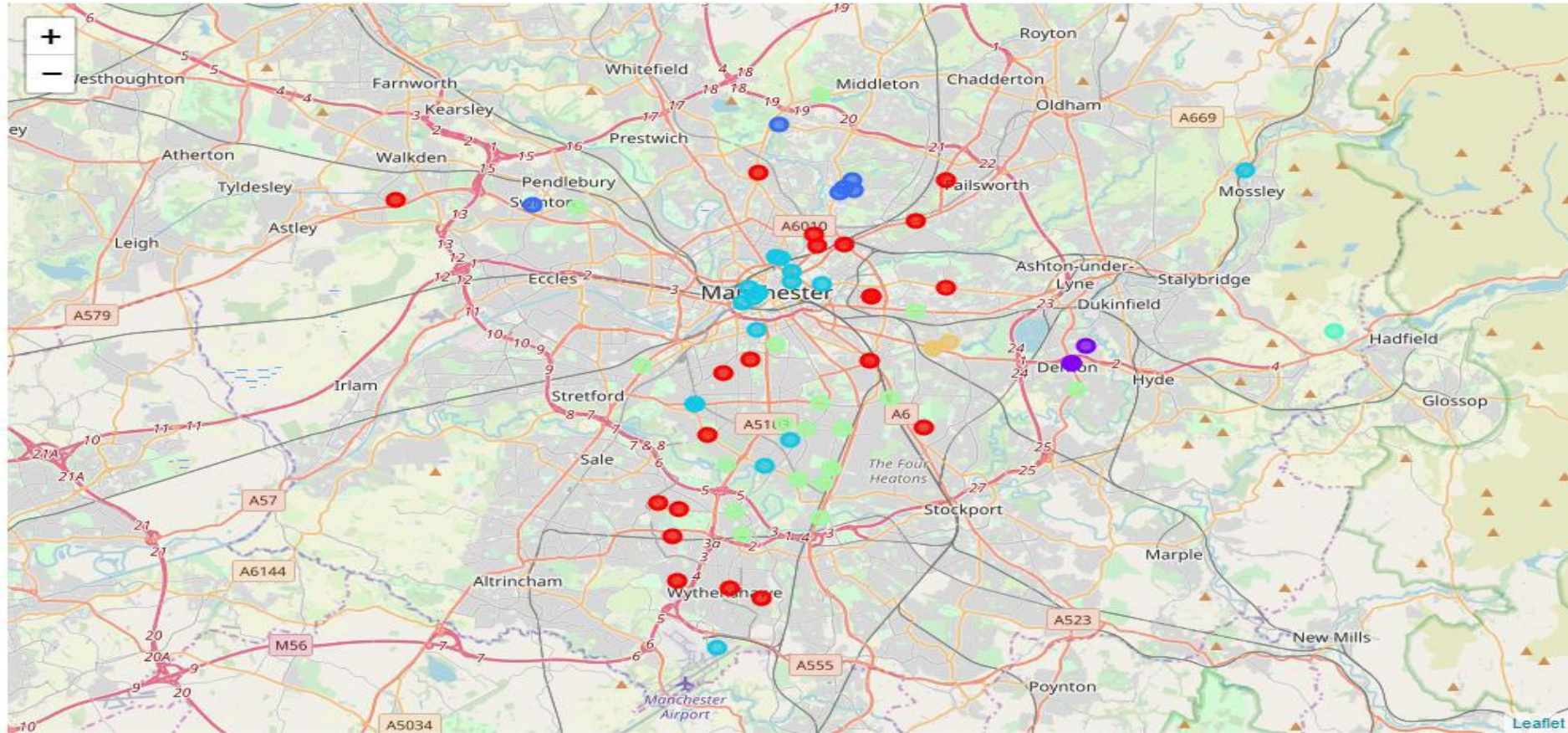
    # Appending the index and score to the respective lists
    indices.append(man_clusters)
    scores.append(score)
```


- Plotting
–Folium



Results

- Visualization of clusters



After visualising 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 number 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, Brooklands and Hyde Newton (Ward) that fall 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's 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.