Data Science Neighbourhoods Analysis for Restaurant - Manchester

DCJ

Applied Data Science Capstone IBM

Introduction

LOVE FOOTBALL, WOULD LIKE TO LIVE THERE AND SEE POTENTIAL FOR RESTAURANT

BIGGEST CITIES OF UK AND BIG POPULATION

PEOPLE FROM ALL AROUND THE WORLD

Business Problem & Data

- Would like to see the oportunity for opening a restaurant and what neighbourhood would be the most likely to do it.
- Have the venue data for the Foursquare API
- Have also the geographical coordinates of the neighborhoods, using GeoPy Library
- Will retrieve the information about the neighbourhoods of Manchester through the Wikipedia page through data scraping

Methodology

Feature Extraction: One hot encoding

In [21]:
```
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()
```

Out[21]:

Methodology

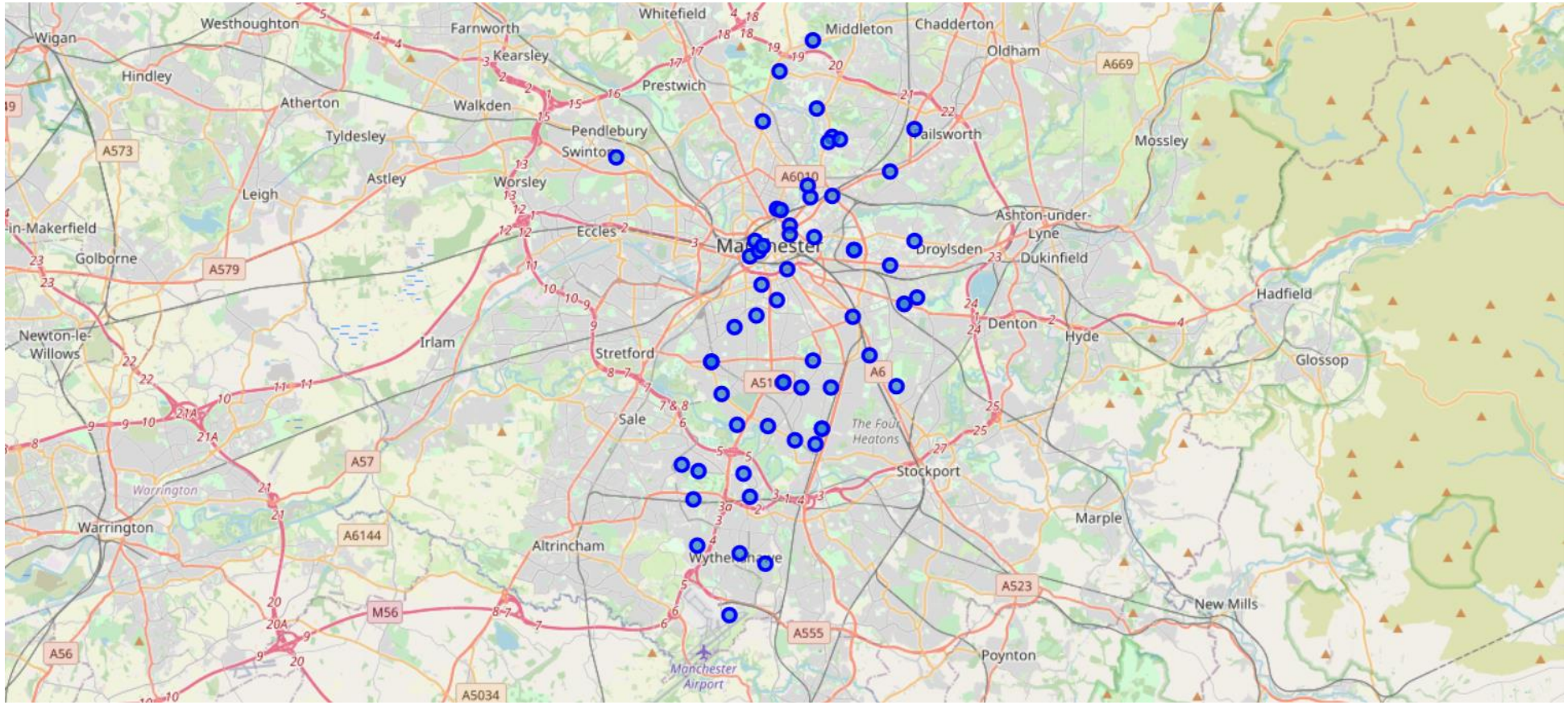
Unsupervised Learning: K- Means Klustering

In [27]:
```
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)
```

In [28]:
```
plot(max_range, scores, "No. of clusters", "Silhouette Score")
```

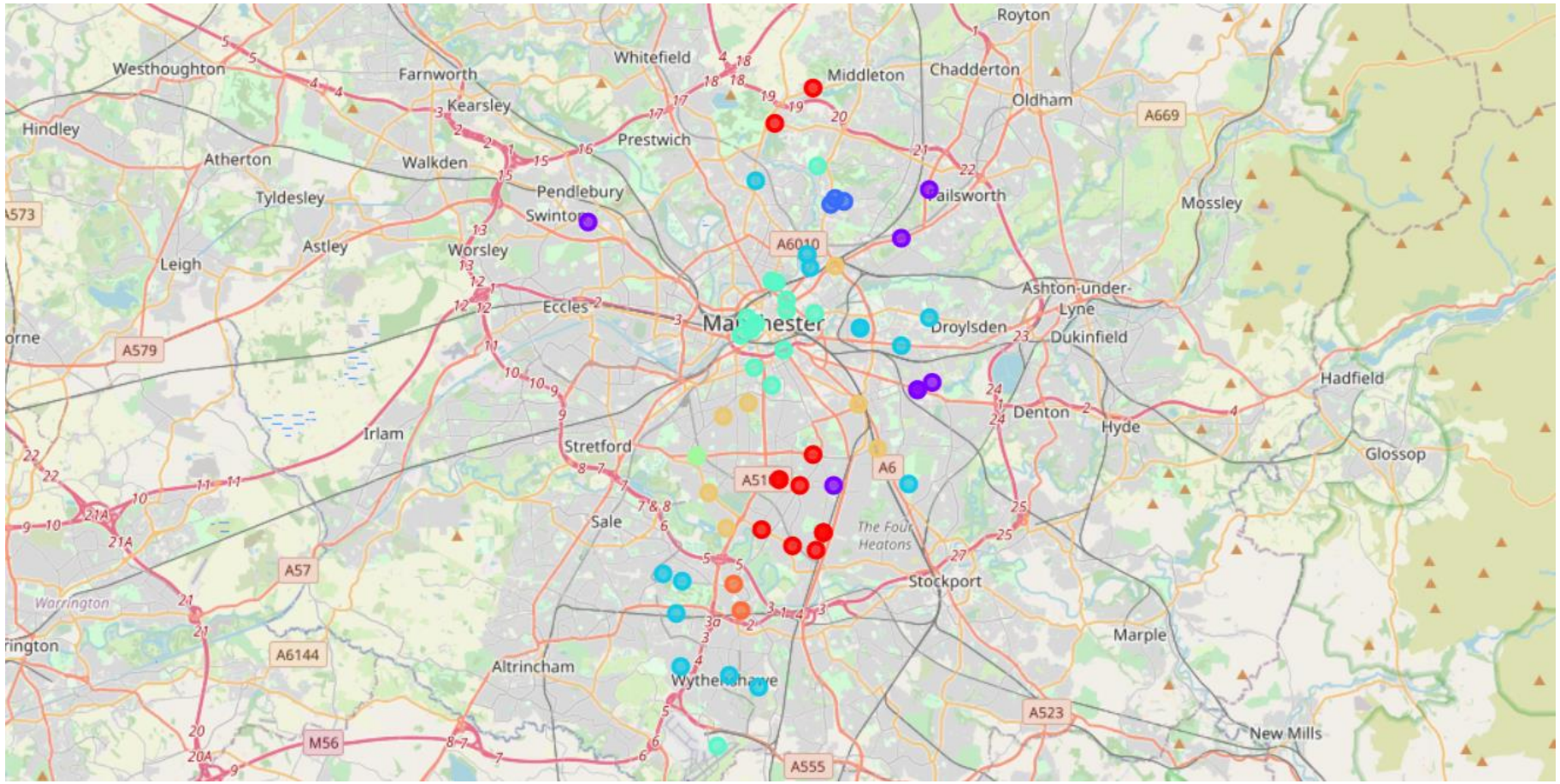
Methodology

Plotting: Folium



Results

Visualization of the Clusters



Business Discussion & Conclusions

- The client that would like to carefully decide and have into account all other factors to take a decision. Data analysis and machine learning were the principles to assist on decision making.
- After studying all of the four clusters it is recommendable to have an interest in setting up a restaurant and the best possibilities would be Barlow Moore or Brooklands, that fall into cluster four and if somebody is interested in opening a restaurant this information would be useful. For this conclusions, Python's inbuilt libraries such as BeautifulSoup, Folium and GeoPy were key to be able to determine.
- As mentioned, cluster 4 would be very attractive for opening a restaurant.
- The K-Means model worked really well and was successful in the clustering exercise.