Coursera IBM Applied Data Science Capstone Project

(The Battle of Neighbourhoods)

Airbnb Recommendation System for Manchester Neighbourhoods Using K-Means Clustering and Foursquare API in Python

By

Olumide Olumayegun

May 2020

Contents

1. Introduction and Business Problem

1.1 Introduction

Manchester is a city and metropolitan borough in Greater Manchester, England, with a population of 547,627 as of 2018 (making it the fifth most populous English district) [1]. Manchester is the third most visited city in the UK, after London and Edinburgh. In 2019, it surpassed Edinburgh to become the second most visited city in the UK after London [2]. It is notable for its architecture, culture, musical exports, media links, scientific and engineering output, social impact, sports clubs and transport connections. There are three universities in the City of Manchester: The University of Manchester, Manchester Metropolitan University and Royal Northern College of Music. The University of Manchester is the largest full-time non-collegiate university in the United Kingdom. Manchester is well known as a city of sport with two decorated Premier League football clubs bear the city name – Manchester United and Manchester City. Therefore, Manchester is very popular with visitors seeking short-term accommodations. These include sport fans, business visits, school open day's visits, conferences, e.t.c.

Our client, who is planning a visit to Manchester, would like us to recommend the best Airbnb listings to book for a short stay in Manchester. However, this is not a straight forward choice as the clients have preferences for Airbnb as well as the kind of neighbourhood the Airbnb is located.

1.2 Business Problem

To build segregation and clustering models that will recommend to an individual suitable Airbnb accommodations in Manchester City area based on available data of Airbnb listings and venues categories in Manchester neighbourhoods.

2. Data Description

The development and analysis of this system will require a variety of data sources. Below is a brief description of each dataset as well as links to the original sources where the data was downloaded from.

- **Airbnb listing for Greater Manchester, England:** This dataset contains information for Airbnb properties for the area of Greater Manchester. It is provided by Inside AirBnb. It was downloaded from the website at http://insideairbnb.com/get-the-data.html. It is available as listings.csv file under Greater Manchester, England, United Kingdom on

the website. This file was downloaded and saved in the working directory of this project. The dataset contains information and metrics for Airbnb listings in Greater Manchester. Some useful observations in the data include: Listing id, neighbourhood group, neighbourhood, latitude and latitude of the listing, room type, price, number of reviews for the listing, and availability. K-Means algorithm will be applied on this dataset to cluster the Airbnb listings into different groups.

- **Geographical coordinates of Manchester neighbourhoods:** This is the latitude and longitude of the 32 neighbourhoods in Manchester area. This was obtained from ArcGIS using geocoder function in Python. A table of Manchester neighbourhoods and their latitude and longitude was created by pulling the neighbourhood geographical coordinates with Python calls. The neighbourhood geographical coordinates will be used to generate maps of Manchester with Folium. It will also form part of API request for venues in the neighbourhoods.

- **List and categories of venues in Manchester neighbourhoods:** This dataset contains important information about venues in each Manchester neighbourhood. This was obtained through API requests to Foursquare location data. K-means algorithm will be applied on this dataset to cluster the Manchester neighbourhoods based on the categories of venues in the neighbourhood.

3. Methodology

3.1 Download and import required libraries

The first step is to download the libraries required for completing the data science project. The libraries used in this project include: Numpy, Pandas, json library to handle JSON files, request library to handle requests, geocoder library, matplotlib, scikitlearn K-Means library and folium for map rendering.

3.2 Load and clean the data

The Airbnb listings dataset for Greater Manchester will be loaded into pandas dataframe. The dataset is cleaned by removing irrelevant/unwanted observations and convert the data types to the appropriate one. A sample of the cleaned dataset is shown in Figure 1 below.

4

Table 1 Airbnb listings in Greater Manchester

:

	id	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	availability_365
0	68951	Bolton	Bolton District	53.56214	-2.53305	Entire home/apt	65.0	2	82	196
1	85109	Manchester	City Centre	53.48074	-2.23186	Private room	50.0	2	434	349
2	159189	Rochdale	Rochdale District	53.61537	-2.19885	Shared room	55.0	1	0	349
3	283495	Rochdale	Rochdale District	53.56259	-2.21945	Private room	60.0	3	10	300
4	299194	Stockport	Stockport District	53.37600	-2.04462	Entire home/apt	50.0	2	255	349

3.3 Descriptive statistical analysis

A quick statistical analysis of the dataset is shown in Table 2 below. The data showed that there are 4924 Airbnb listings in Greater Manchester as at 16th of May 2020. There are 41 neighbourhoods distributed across 10 neighbourhood groups (or Borough). Salford District neighbourhood has the highest number of Airbnb listings of 787. More than half of the listings are in the Machester neighbourhood group. The rooms are grouped into 4 types with the most popular room type being private room. Analysis also showed that there are four types of room available: Private room, Entire home/apt, Hotel room, and Shared room. The most common room type is private room.

The spread of the Airbnb on the map of Greater Manchester is shown in Figure 1. There is dense concentration of listings around the Manchester City area. Manchester city is a popular city with visitors such as footbool fans, business travels, tourists, e.t.c. So let assume that we are only interested in Airbnb listed in Manchester City neighbourhoods only. We will extract listings in Manchester neighbourhoods from the whole Greater Manchester listings and apply our analysis to this.

Table 2 Statistical analysis of Airbnb listings dataset

|	id	neighbourhood_group	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	availabili
count	4.924000e+03	4924	4924	4924.000000	4924.000000	4924	4924.000000	4924.000000	4924.000000	4924.0
unique	NaN	10	41	NaN	NaN	4	NaN	NaN	NaN	
top	NaN	Manchester	Salford District	NaN	NaN	Private room	NaN	NaN	NaN	
freq	NaN	2823	787	NaN	NaN	2412	NaN	NaN	NaN	
mean	2.855575e+07	NaN	NaN	53.472945	-2.249493	NaN	88.091998	3.518278	27.871649	153.7
std	1.143300e+07	NaN	NaN	0.045497	0.081715	NaN	173.412694	15.870997	51.629740	138.8
min	6.895100e+04	NaN	NaN	53.345830	-2.708880	NaN	0.000000	1.000000	0.000000	0.0
25%	1.996664e+07	NaN	NaN	53.448817	-2.275375	NaN	30.000000	1.000000	1.000000	0.7
50%	3.044096e+07	NaN	NaN	53.474685	-2.239015	NaN	52.000000	1.000000	7.000000	124.5
75%	3.888886e+07	NaN	NaN	53.488280	-2.218380	NaN	90.000000	2.000000	30.000000	323.0
max	4.345605e+07	NaN	NaN	53.677360	-1.976980	NaN	5539.000000	400.000000	545.000000	363.0
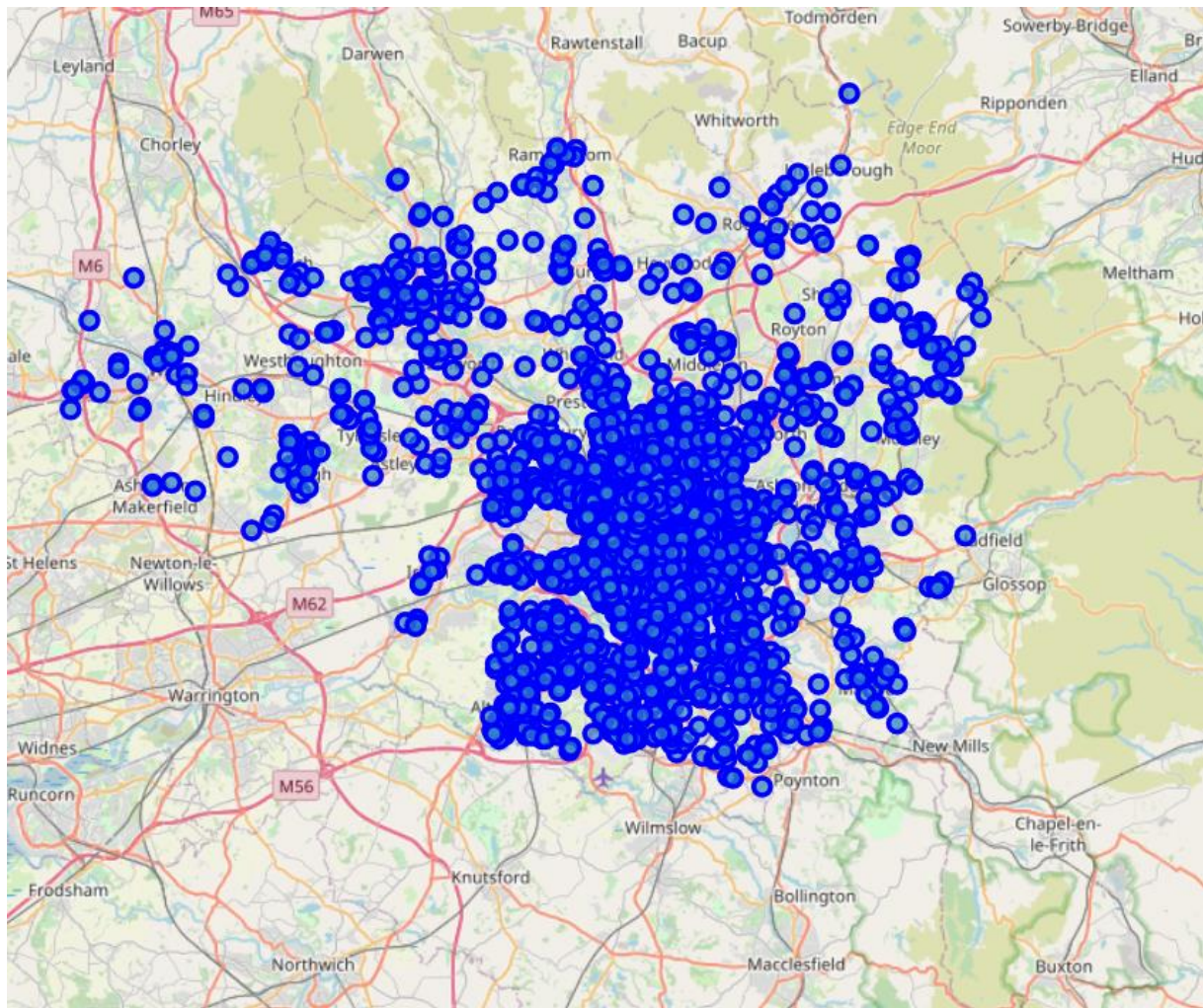
Figure 1 Spread of Airbnb listings in Greater Manchester

3.4 K-Means clustering of Airbnb in Manchester

We will segregate the listings in Manchester into ten clusters based on the room type, price, minimum nights, number of reviews and availability using K-Means clustering algorithm. The groupings should not be influenced by the geographical location of the listings. Hence, we won't be using the geographical coordinates as features for the clustering algorithm. The neighbourhood location of the listings will only be taken into consideration through the available amenities/venues in the neighbourhood. This will be investigated later through neighbourhood segregation and clustering.

A map of the ten clusters predicted is shown in Figure 2. Therefore, based on a client's preferences for Airbnb, we can suggest a group of Airbnb to the customer by using the trained K-Means model to predict appropriate cluster. Suppose our client is also interested in the facilities/venues in the neighbourhood of the listings in order to narrow down the choices

since there are still many options to choose from within each clusters. We will approach this by also segregating and clustering the Manchester neighbourhoods based on the venues.
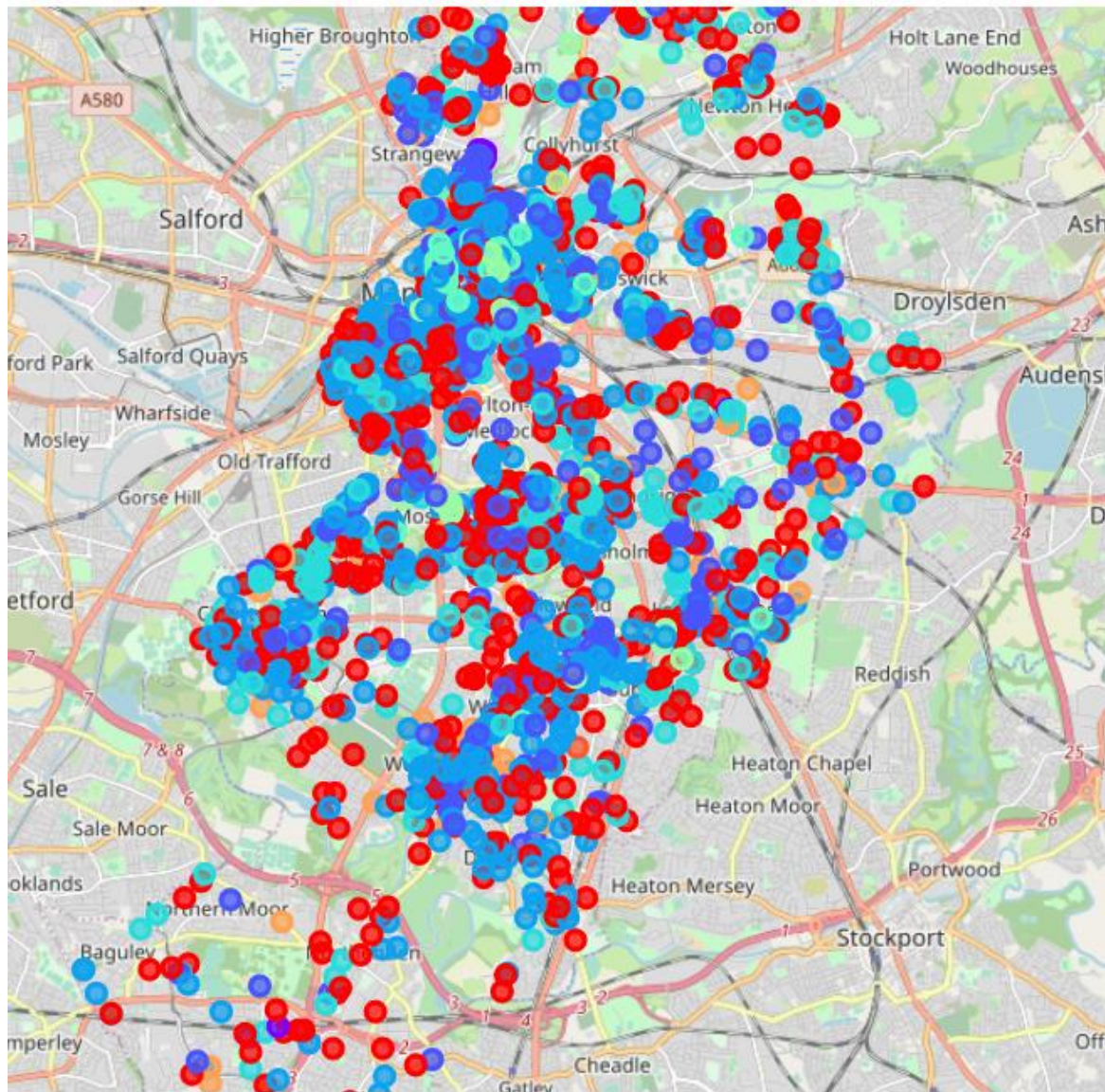
Figure 2 Airbnb clusters in Manchester

3.4 K-Means clustering of neighbourhoods in Manchester

We will obtain the latitude and longitude of the 32 neighbourhoods in Manchester using geopgraphical coordinates data call to ArcGIS with geocoder library. The results is presented in the form of in Table 3 below and their location on the map is shown in Figure 3. The information is passed to Foursquare API to obtain location data of the neighbourhoods, a sample of which is presented in Table 4.

Table 3 Geographical coordinates of Manchester neighbourhoods

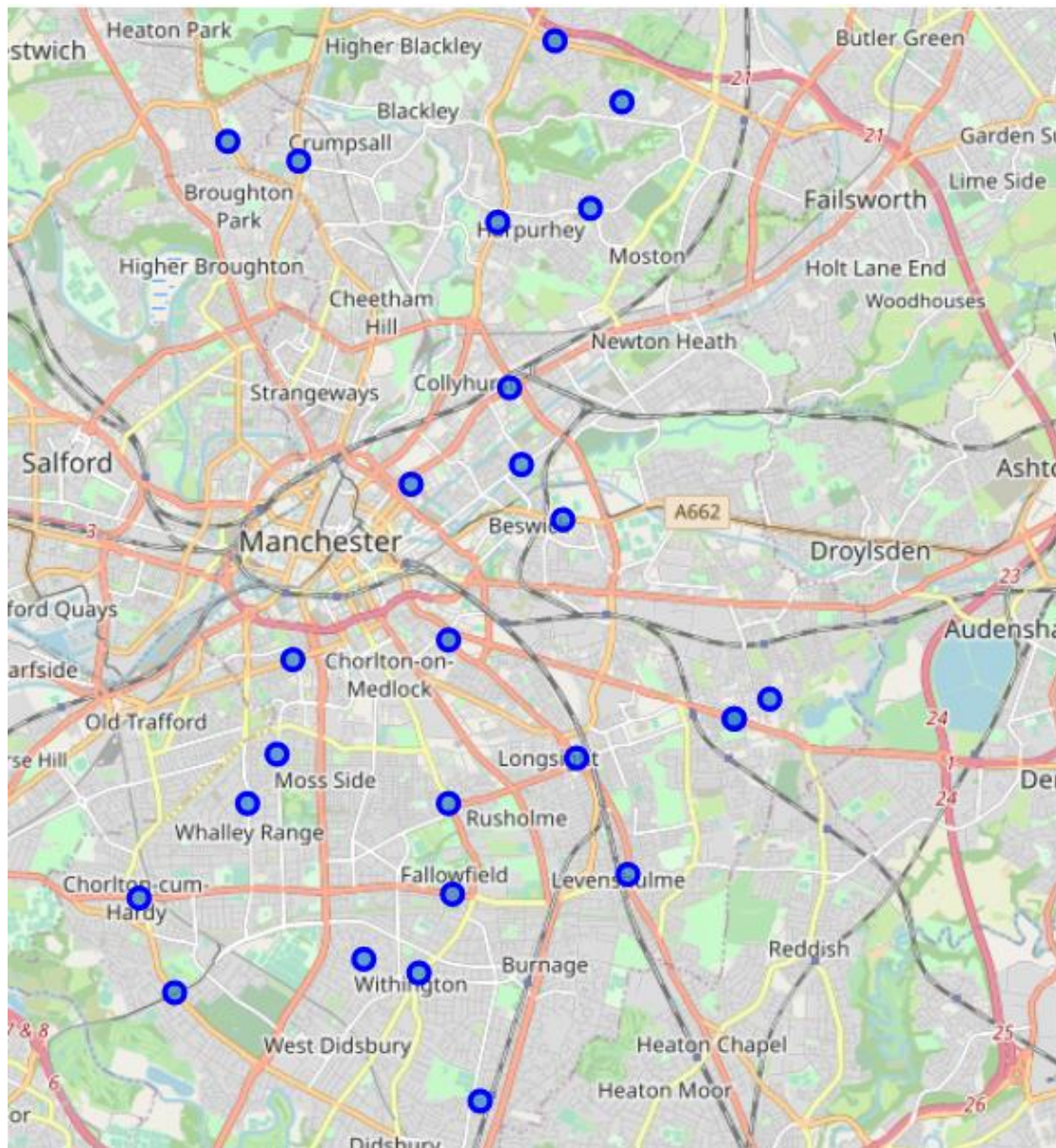|	Neighbourhood	Latitude	Longitude
0	City Centre	53.481740	-2.202640
1	Ancoats and Clayton	53.485350	-2.229200
2	Moss Side	53.457460	-2.252370
3	Bradford	53.487381	-2.209943
4	Hulme	53.467270	-2.249730
5	Chorlton	53.442510	-2.276560
6	Didsbury West	53.411967	-2.227542
7	Old Moat	53.436027	-2.237556
8	Burnage	53.421360	-2.217280
9	Levenshulme	53.444850	-2.191590
10	Didsbury East	53.405200	-2.221760
11	Withington	53.434610	-2.227860
12	Chorlton Park	53.432620	-2.270450
13	Longsight	53.456980	-2.200510
14	Whalley Range	53.452210	-2.257530
15	Moston	53.513940	-2.198020
16	Cheetham	53.552934	-2.190033
17	Higher Blackley	53.531350	-2.204250
18	Rusholme	53.452260	-2.222530
19	Ardwick	53.469320	-2.222620
20	Brooklands	53.520900	-2.261000
21	Gorton South	53.461110	-2.173060
22	Gorton North	53.461110	-2.173060
23	Fallowfield	53.442890	-2.221940
24	Miles Platting and Newton Heath	53.495350	-2.211990
25	Baguley	53.399090	-2.285610
26	Northenden	53.407150	-2.259350
27	Harpurhey	53.512610	-2.213980
28	Crumpsall	53.518890	-2.248740
29	Woodhouse Park	53.463107	-2.166657
30	Sharston	53.399663	-2.255452
31	Charlestown	53.525080	-2.192530

Figure 3 Map of Manchester neighbourhoods

Table 4 Samples of venues and venue categories in Manchester neighbourhood

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	City Centre	53.48174	-2.20264	Etihad Stadium	53.483373	-2.200592	Soccer Stadium
1	City Centre	53.48174	-2.20264	City Square	53.484490	-2.201418	Event Space
2	City Centre	53.48174	-2.20264	City Of Manchester Stadium Star Bar	53.483095	-2.200507	Bar
3	City Centre	53.48174	-2.20264	CityStore at the Stadium	53.484474	-2.200725	Sporting Goods Shop
4	City Centre	53.48174	-2.20264	Mary D's	53.480953	-2.197384	Pub
5	City Centre	53.48174	-2.20264	National Squash Centre & Regional Arena	53.484667	-2.202540	Athletics & Sports
6	City Centre	53.48174	-2.20264	Manchester Regional Arena	53.483845	-2.203875	Track Stadium
7	City Centre	53.48174	-2.20264	Manchester Tennis and Football Centre	53.486217	-2.202394	Tennis Court
8	City Centre	53.48174	-2.20264	Magic Muffin	53.481708	-2.200761	Café
9	City Centre	53.48174	-2.20264	The Mancunian	53.483082	-2.200424	Restaurant
10	City Centre	53.48174	-2.20264	1894 Club	53.483194	-2.200378	Lounge
11	City Centre	53.48174	-2.20264	Tony's Stadium Chippy	53.481104	-2.197516	Fish & Chips Shop
12	City Centre	53.48174	-2.20264	Sportcity	53.484658	-2.198017	Athletics & Sports
13	City Centre	53.48174	-2.20264	Ashbury Meadow Childrens Centre	53.477892	-2.205378	Child Care Service
14	Ancoats and Clayton	53.48535	-2.22920	Rudy's	53.484167	-2.229419	Pizza Place
15	Ancoats and Clayton	53.48535	-2.22920	The Jane Eyre	53.484121	-2.229092	Cocktail Bar
16	Ancoats and Clayton	53.48535	-2.22920	Bem Brasil	53.484092	-2.231634	Brazilian Restaurant
17	Ancoats and Clayton	53.48535	-2.22920	Chaology	53.483473	-2.230088	Tea Room
18	Ancoats and Clayton	53.48535	-2.22920	Ancoats Coffee Co	53.483536	-2.227407	Coffee Shop
19	Ancoats and Clayton	53.48535	-2.22920	Ancoats General Store	53.483647	-2.230384	Convenience Store

We will run K-Means clustering algorithm to group the Manchester neighbourhood into 5 clusters. The clusters are shown on the map of Manchester in Figure 4. An examination of the clusters showed that:

- Cluster 1: Neighbourhoods popular for Chinese restaurants, pubs and convenience stores
- Cluster 2: Neighbourhoods for different varieties of restaurants, grocery stores and bars/pubs
- Cluster 3: Only one neighbourhood in this cluster with Men's store, Warehouse store and concert hall
- Cluster 4: Also one neighbourhood with Food truck, Construction and landscaping
- Cluster 5: Neighbourhood with Chinese and fast food restaurants and supermarkets/markets
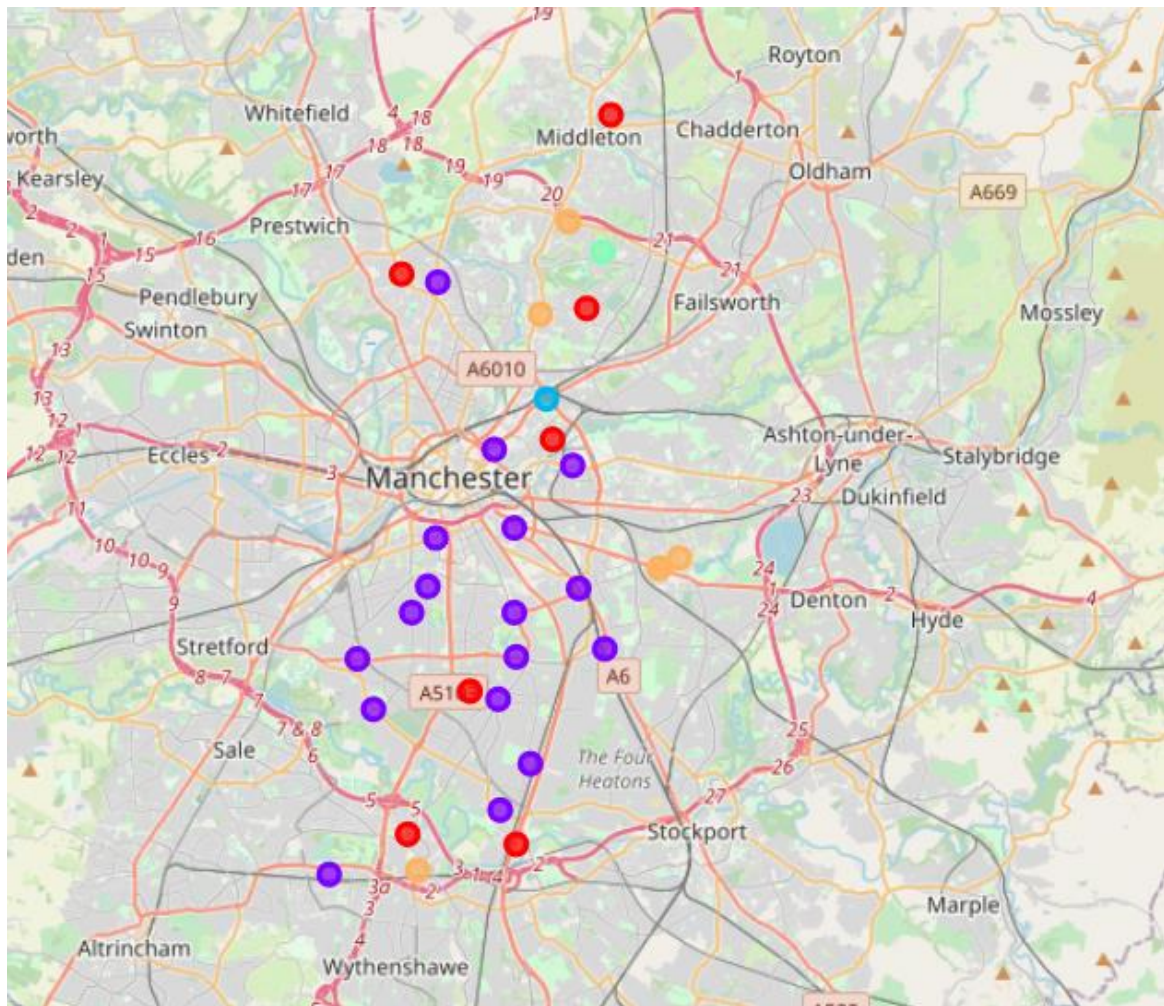
Figure 4 Clustering of Manchester neighbourhoods

4. Results and Discussions

4.1 Use case scenario

The result and application of this system will be demonstrated with a use case scenario.

Suppose a client is looking to book Airbnb in Manchester area. The customer preferences for Airbnb accommodation is as follows:

- Room type: private room
- Price: £55
- Minimum night: 2
- Number of reviews: 100
- Availability: 340 days

In addition, the customer prefers accommodations in a neighbourhood with a lot of Chinese and fast food restaurants and supermakets/markets.

We will use the clustering models developed to make recommendations of Airbnb listings to this customer.

Application of our K-Means model for Airbnb listings predicted that based on the client's criteria for Airbnb the best Airbnb cluster for our client is Cluster labelled 4. However, since the customer also want listing in neighbourhoods with Chinese and fast food restaurants and supermarkets/markets, we will restrict our recommendations to listings in neighbourhood Cluster 5 according to the K-Means model of neighbourhood clusters . This neighbourhoods are: Higher Blackley, Gorton South, Gorton North, Harpurhey and Sharston. After applying these conditions to the Airbnb listings, we were able to extract the best Airbnb listings to recommend to our client.

Some of the recommended listings is shown in Table 5 below . A visual representation of the recommended Airbnb on the map of Manchester is shown in Figure 5.

Table 5 Some of the recommended Airbnb for the client

|	id	neighbourhood	latitude	longitude	room_type	price	minimum_nights	number_of_reviews	availability_365	bnb_cluster
82	4054992	Gorton South	53.45932	-2.17265	Private room	27.0	1	34	337	4
83	4070469	Gorton North	53.46072	-2.17283	Private room	25.0	2	58	329	4
85	4161025	Gorton North	53.46593	-2.19135	Private room	24.0	2	83	255	4
87	4276085	Gorton South	53.44839	-2.18699	Private room	29.0	1	30	214	4
155	7393366	Gorton South	53.45963	-2.17161	Private room	50.0	2	20	204	4
201	9555264	Gorton North	53.46485	-2.18963	Private room	24.0	3	113	349	4
234	10666909	Harpurhey	53.51657	-2.20673	Private room	31.0	1	111	348	4
407	14660317	Harpurhey	53.50042	-2.21434	Private room	23.0	2	42	288	4
444	15312708	Sharston	53.37631	-2.24357	Private room	90.0	7	0	349	4
592	18070300	Gorton North	53.46167	-2.18030	Private room	40.0	1	3	349	4
774	20975369	Gorton South	53.44505	-2.19071	Private room	15.0	4	19	350	4
778	21010933	Gorton South	53.45627	-2.16015	Private room	23.0	1	63	349	4
798	21333796	Gorton North	53.46024	-2.17398	Private room	20.0	1	8	349	4
978	23970587	Sharston	53.38516	-2.25902	Private room	40.0	1	0	349	4
1116	26012263	Gorton North	53.46796	-2.15763	Private room	100.0	1	0	349	4
1165	26831335	Gorton South	53.44877	-2.18150	Private room	50.0	1	14	349	4
1169	26874816	Gorton South	53.45012	-2.18341	Private room	20.0	6	27	349	4
1184	27161917	Gorton South	53.45197	-2.17303	Private room	80.0	1	0	349	4
1188	27222554	Gorton North	53.46874	-2.15930	Private room	60.0	1	0	349	4
1268	28503615	Harpurhey	53.51590	-2.20544	Private room	210.0	1	1	350	4
1296	28872426	Gorton North	53.46851	-2.15807	Private room	130.0	3	1	349	4
1301	29027662	Gorton South	53.44944	-2.18292	Private room	25.0	6	15	349	4
1361	29737801	Higher Blackley	53.52641	-2.21559	Private room	20.0	1	1	350	4
1679	34041192	Gorton North	53.45776	-2.18935	Private room	30.0	2	9	350	4
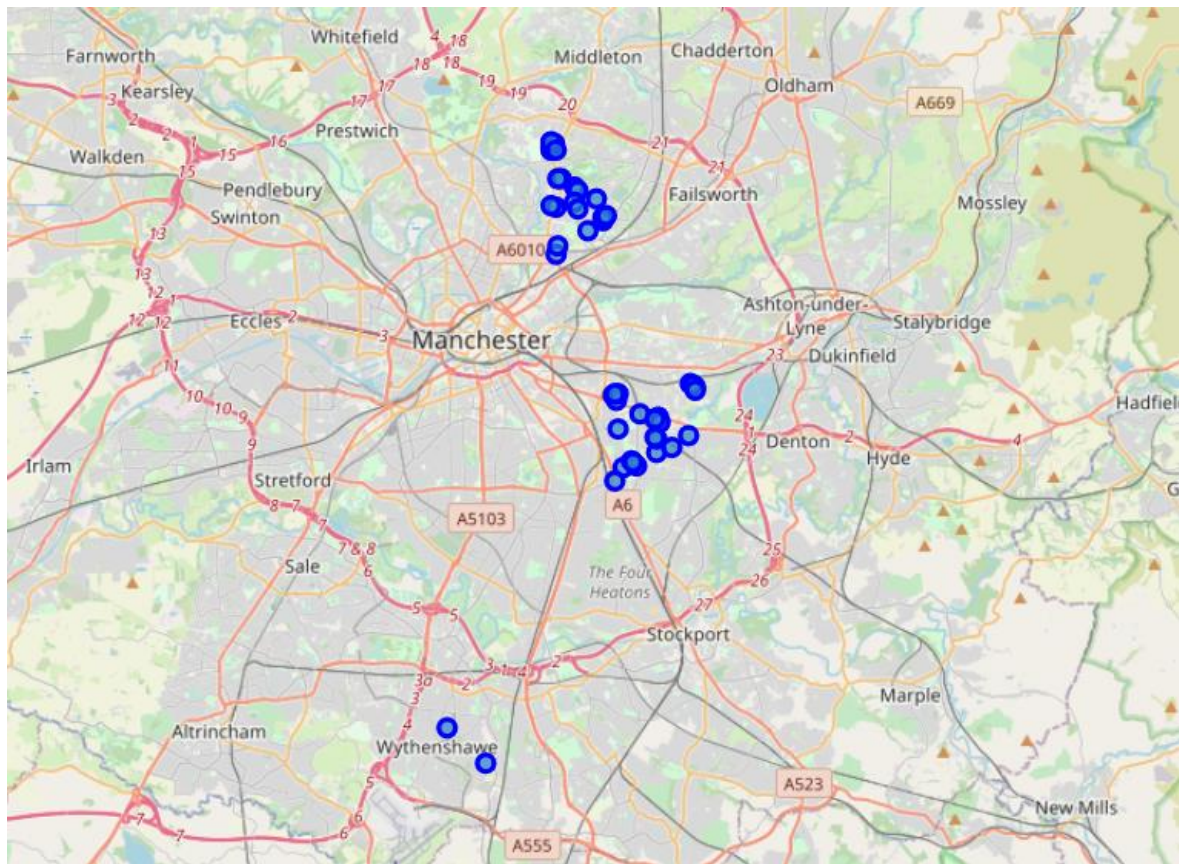
Figure 5 Map of recommended Airbnb in Manchester neighnourhood

5. Conclusions

In this project, we have been able to deploy data science powers of dataset wrangling, machine learning models (K-Means clustering) and geolocation APIs to develop a system to recommend Airbnb listings to our clients looking to book accommodation in Manchester for short stays. We were able to satisfy the problem statement set out at the start of the project, which is to build a segregation and clustering models able to recommend suitable Airbnb in Manchester based on data of Airbnb listings and neighbourhood venues. The capability of the models were demonstrated through a use case scenario. Out of the over 3000 Airbnb listings, our system was able to pick about 50 listings suitable for our client. It should be emphasized that this 50 listings can be further narrowed down for the client by a stricter application of the criteria.

References

[1] "Population estimates for the UK, England and Wales, Scotland and Northern Ireland mid-2018". www.ons.gov.uk. Office for National Statistics. 26 June 2019. Retrieved 27 May 2020.

[2] "London visited by 50% of UK's tourists". BBC News. Archived from the original on 8 June 2013. Retrieved 21 May 2020.