IDENTIFYING NEIGHBOURHOOD TYPES IN MANCHESTER (UK) FOR COMPANY EMPLOYEE RELOCATIONS

IBM Data Science Professional Capstone Project

Abstract

This project is for the completion of the IBM Data Science Professional Certification via Coursera. The project demonstrates a method for the segmentation of the districts within Manchester (UK) by type of venues within.

Dave Webb

dave_webb86@live.co.uk

Aug - 2020

Introduction

Relocating for a job in a new city can be a stressful experience for anyone and a move that goes badly can quickly lead to an employee leaving the company. [1] The role of a good HR department is to ease this process for any new hires to help them integrate into their new location quicker. A question that is often asked by relocating employees is “where are the best neighbourhoods?” to live in near their new location of employment. This is not a simple question to answer as different people have different needs, some may crave the hustle and bustle of a city centre with lots of night time activity while others may want a nice quiet neighbourhood to relax in after work.

This project aims to demonstrate a method for analysing the locations of a new city to allow informed recommendations of neighbourhoods to set up home based on the preferences of a moving employee using a worked example of the UK city of Manchester. The goal is to demonstrate that companies and HR departments can easily develop tools to aid relocating personnel with the aim to increase staff retention following such moves. The tool will categorise prospective neighbourhoods according to their key features to allow the quick and easy and easy generation of recommendation lists based on employee preferences. In addition, the average house price for each of the neighbourhoods will be sourced to allow a user to assess the affordability of each of the recommended neighbourhoods whilst making comparisons.

Data

The following data sets have been identified as required to build the tool:

- Name and location for all the neighbourhoods in Manchester
- Average house prices for each of the neighbourhoods
- List of the venue types and their frequency in each of the neighbourhoods

Open accessible sources identified to find this data along with the first 5 rows of each data set as obtained are shown below:

- Name of each of the neighbourhoods to be web scraped from Wikipedia, which contains a comprehensive list of the areas of greater Manchester separated by postcode outcodes [2] (these are the first section of standard UK postcodes)

	Postal district	Post town	Coverage	Local authority
0	M1	MANCHESTER	Manchester central	Manchester
1	M2	MANCHESTER	Manchester central	Manchester
2	M3 (Sectors 1, 2, 3, 4 and 9)	MANCHESTER	Manchester central	Manchester
3	M3 (Sectors 5, 6 and 7)	SALFORD	Salford Quays	Salford
4	M4	MANCHESTER	Manchester central	Manchester

- Location of each borough will be sourced from freemaptools.com [3] which contains downloadable CSV files of UK locations defined by latitude and longitude, separated by postcode

	A	B	C	D
1	id	postcode	latitude	longitude
2	2	AB10	57.13514	-2.11731
3	3	AB11	57.13875	-2.09089
4	4	AB12	57.101	-2.1106
5	5	AB13	57.10801	-2.23776

- Average UK house pricing is to be obtained from the UK gov.uk website [4] containing an up to date list of properties sold in the UK in the current year with postcode information

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	{A2479555-3559-74C7-E053-6B04A8C0887D}	299950	31/01/2020 00:00	EX17 3FL	E	EX	S	Y	F	7		YEO CRESCENT		CREDITON	MID DEVON	DEVON	A	A
2	{A2479555-5A44-74C7-E053-6B04A8C0887D}	280000	28/02/2020 00:00	B60 1DU	B	B6	T	N	F	21		COTTAGE LANE	MARLBROOK	BROMSGROVE	BROMSGROVE	WORCESTERSHIRE	A	A
3	{A2479555-5A46-74C7-E053-6B04A8C0887D}	267000	03/03/2020 00:00	WR5 2PB	W	WR	S	N	F	16		REGIMENT CLOSE	BROOMHALL	WORCESTER	WYCHAVON	WORCESTERSHIRE	A	A
4	{A2479555-5A47-74C7-E053-6B04A8C0887D}	164000	28/02/2020 00:00	HR2 7DD	H	HR	T	N	F	14		WATERFIELD ROAD		HEREFORD	HEREFORDSHIRE	HEREFORDSHIRE	A	A
5	{A2479555-5A48-74C7-E053-6B04A8C0887D}	72500	21/02/2020 00:00	B61 8AB	B	B6	F	N	L	ST. JAMES COURT, 30	FLAT 11	THE STRAND		BROMSGROVE	BROMSGROVE	WORCESTERSHIRE	A	A

- Neighbourhood venue types will be sourced using the Foursquare API [5] which allows the search for venues in a location given the geographic co-ordinates, these venues can then be arranged by type and frequency using simple manipulation in python

Methodology – Initial Data Wrangling & Exploration

The Manchester neighbourhood data was sourced from Wikipedia using the python pandas library read_html function before converting to a csv file. The location and house pricing data were downloaded as CSV files from the sources already identified.

Data wrangling was quickly put together using MS Excel. Initially to ease data processing the two entries for postcode M3 in the neighbourhood dataset were merged and the non-residential postcodes removed (for example Manchester airport). The geographic co-ordinates were simply joined onto the end of the neighbourhood table via Excel vlookup functions. House prices were first reduced to Manchester only data via the Excel sort and filter function before averaging the property price for each of the Manchester outcodes. Finally, the averaged data was amended to the neighbourhood database again using vlookup functions. The first 5 rows of the final cleaned data are shown below.

	Postal district	Neighbourhood	Local authority	Latitude	Longitude	Average Sold House Price, 000s
0	M1	Manchester central	Manchester	53.4773	-2.2351	346
1	M2	Manchester central	Manchester	53.4800	-2.2426	4008
2	M3	Manchester central, Salford Quays	Manchester	53.4836	-2.2502	263
3	M4	Manchester central	Manchester	53.4846	-2.2291	328
4	M5	Seedley, Weaste	Salford	53.4790	-2.2848	217

Visualisation of the geographical locations of the Manchester neighbourhoods was performed using the folium library on python, with each of the individual postal districts represented by a blue dot (see Figure 1). It is noted that districts outside of the Manchester ring road to the east and the south of the city do not appear to be represented. This is due to the postal codes being classified as different cities, Oldham to the east and Stockport to the south. This is a quirk of England as there are often many cities in close proximity with densely populated areas. A potential improvement for the model may be to gather the postal district codes for the entire region and then filter based on the distance form the city centre, although time constraints have limited this improvement for this report.

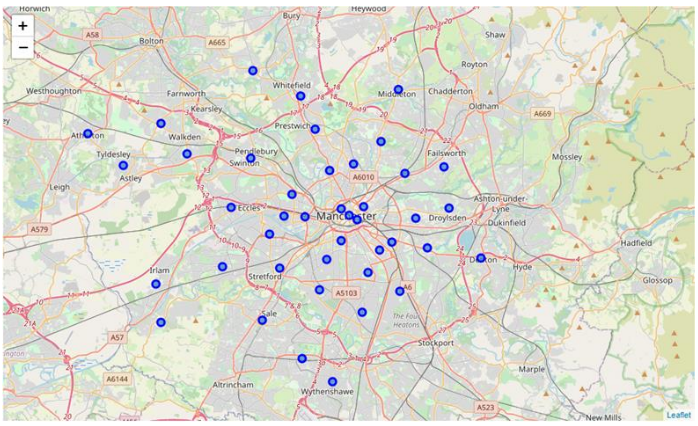

Figure 1: Map detailing the central location of each of the postal districts in Manchester, UK

Initial data exploration was also undertaken on the house prices in each of the postal districts and it was noted that M2 and M17 appear to be obvious outliers from the rest of the explored areas (see Figure 2). Further analysis revealed that these districts contained the least number of house sales for the region in 2020 leaving them open for heavy distortion. For example M2 only had 3 property sales, although the data contained no further insights on the reason of the distortion this district is identified as the centre of the city so reasonable assumption would be that these are commercial property sales which would distort the pricing . As the house price data is only designed to be an informative feature of the final data rather than being used to segment the data it was deemed acceptable to use the data as discovered with this highlighted caveat identifying the outliers.

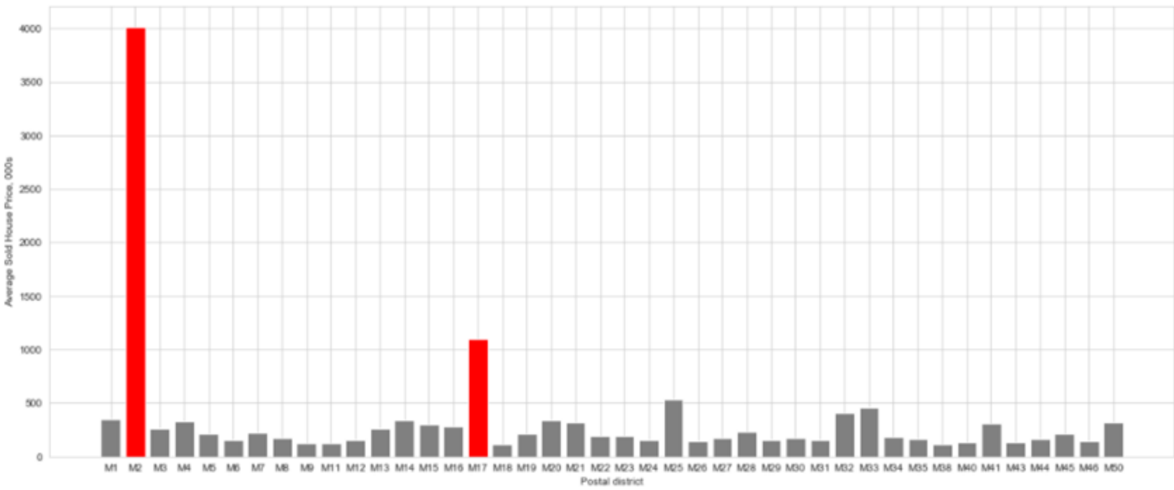

Figure 2: Average sold property prices for 2020 in each of the Manchester postal districts. M2 and M17 have been identified as outliers from the rest of the data.

Methodology – Sourcing Neighbourhood Feature Types and Clustering

Venue data was generated via the Foursquare API [5] in Python using the central point of each Manchester postcode area as the location to search for active Foursquare venues. The search was limited to an area of radius 1000m around the central point and limited to 100 venues. This may have led to some overlap of venues in the heavily clustered city centre but would be well suited to the spaced out suburb areas. Once complete the data was assessed for the number of venues in each location (see Figure 3) and it can be seen that outside of a few areas in the city centre the number of identified venues is fairly limited at <20. This suggests that Foursquare may not be a comprehensive guide to Manchester outside of the city centre. Utilisation of the Foursquare API is mandatory for completion of this IBM Data Science Professional Certification module report however so the rest of the investigation will be completed using this tool.

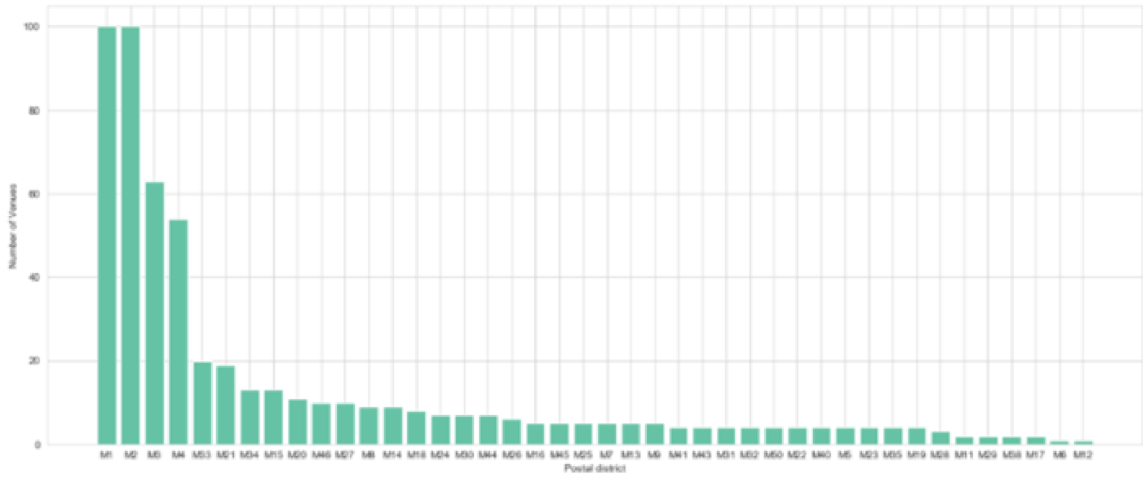

Figure 3: Number of venues for each of the Manchester postcode areas as identified by the Foursquare API.

After sourcing the venues the metadata for each location was combed to source the designated venue type for each data point (for example "Indian restaurant" or "Pub"). This information was then used to group similar venues and create a frequency of each location for all the Manchester postcodes. 135 different venue types were identified within Manchester which is a significant amount of variety in the 553 different venues identified for the region. It is out of the scope for this report but planned improvement for the data is to group similar venues into categories, for example "Bar", "Pub" and "Gay bar" would all be collated into a "Nightlife" group while all the various restaurant types could be collated into an "Eatery" group.

Postal district	Adult Boutique	American Restaurant	Art Gallery	Arts & Crafts Store	Asian Restaurant	Australian Restaurant	Auto Garage	Bagel Shop	Bakery
M1	0.01	0.0	0.01	0.0	0.0	0.0	0.0	0.0	0.02
M11	0.00	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.50
M12	0.00	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.00
M13	0.00	0.0	0.00	0.0	0.2	0.0	0.0	0.0	0.00
M14	0.00	0.0	0.00	0.0	0.0	0.0	0.0	0.0	0.00

Figure 4: Slice of the Manchester venue data where the venues type have been collated to relative frequency for each of the postal districts.

Clustering of the different neighbourhoods was done using the K-means clustering tool from the SciKit Learn library available in python. [6] Optimization of the tool was done using the standard elbow method for clustering [7] where the inertia (or distortion) of the final system is plotted against the number of clusters chosen. Typically this plot produces a distinct elbow where the rate of reduction in inertia rapidly decreases for further increases in the number of clusters selected. This was no the case however for the Manchester neighbourhood data (see Figure 5), with no clear elbow visible in the data. The rate of inertia change does appear to shallow slightly at 4 clusters however so this value was chosen to implement the final algorithm on the data and cluster the Manchester neighbourhoods.

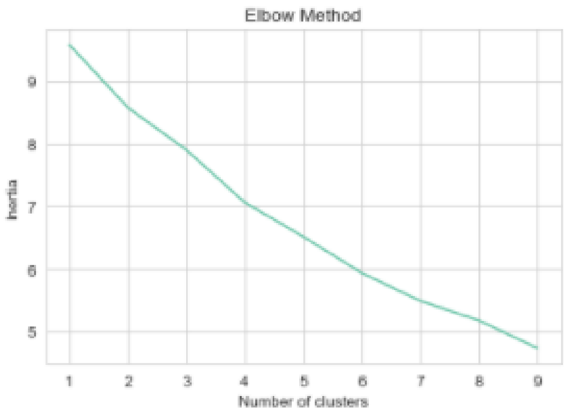

Figure 5: Results of applying the elbow optimisation method to the K-means clustering algorithm applied to the Manchester neighbourhood venue data.

Results

Despite the optimisation method for the clustering tool not working ideally the results from the algorithm came out well. A visual representation of the clusters within Manchester was generating using the folium tool as before, shown in Figure 6. As can be seen most the neighbourhoods in Manchester have been clustered into the same group with a few smaller numbers of grouped neighbourhoods around the city. A full list of each of the clusters along with the average house price and the top 5 most common venues in each neighbourhood is shown in the appendix. Below shows a brief description of each of the clusters along with comments.

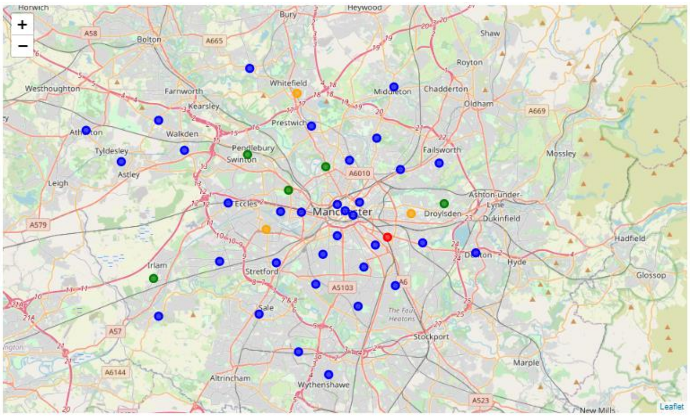

Figure 6: Map detailing the central location of each of the postal districts in Manchester (UK) after sorting into coloured clusters based on the common venue types in each location.

Cluster 1 – Blue

This is the most dominant cluster seen in Manchester. The defining features of the neighbourhoods in this cluster is the presence of a pub or grocery store (this is the UK equivalent to a convenience store for US readers) within the top 5 venues. This demographic of venues is common for built up urban areas in the UK that have an active nightlife alongside residential areas. Interestingly the centre of the city is categorised with a lot of the suburb areas in the city which indicated that despite the clustering technique used the spread of neighbourhood types in this group is still large and would require refinement for someone moving to the city to make an informed decision.

Cluster 2 – Red

This cluster is clearly the anomaly from the rest of the categorised neighbourhoods. The only identified feature here is a "Gaming café" which doesn't appear in any other neighbourhoods which is why it has been categorised by itself.

Cluster 3 – Yellow

This cluster looks to be defined by the presence of "Indian restaurant" and "Bakery" venues in the top 5 locations for each neighbourhood. Indian restaurants and bakeries are common in both commercial and residential areas in the UK, but the lack of pubs for most of the locations shows that these areas may be less reliant on nightlife than other Manchester locations so could potentially be quieter for people looking to move there.

Cluster 4 – Green

This cluster is defined by the presence of "Supermarket" venues within the top 5 locations for each neighbourhood. For the UK areas with lots of supermarkets are typically in heavily residentially based areas as it is common for UK residents to only travel short distances for weekly grocery shopping. This would indicate that this cluster would not be expected to have an active nightlife and may be an ideal location for anyone moving to the city with a young family to settle down.

Discussion

A model has been demonstrated for clustering the neighbourhoods in and around Manchester (UK) but the results leave a lot to be desired to be used as a useful tool for a HR operative aiding staff moving to the area. Major issues identified with potential future work solutions are as follows:

- The Foursquare API does not provide a good overview of the venues in the neighbourhoods away from Manchester city centre
    - Use of another tool such as the Google Maps Location API [8] may provide a better set of venue data
- Segmentation of the neighbourhoods in Manchester is not optimal with a failure in the elbow method to define the ideal number of clusters and the creation of a large group of neighbourhoods that is hard to define as having an overall theme

    - It has already been identified in the project that a lot of the identified venues could easily be collated into categories to further define the venue types in each district, this additional categorisation may help with the final clustering algorithm
- The data still requires a manual search through the segmented groups to make a decision on which neighbourhood would suit a new employee
    - Once the other improvements have been made to the system a recommender system would provide a very useful addition to the tool, allowing a user to put in the preferences of features in a neighbourhood and then provide a list of places they might enjoy in the city

Time constraints and coding skill level have limited my ability to enact these changes to improve the tool in this report but the changes will be completed and an update issued on my data science blog. [9]

Conclusion

This report set out to demonstrate the possibility of building a tool to aid HR workers help relocating members of staff choose the correct neighbourhood in their new resident city using the city of Manchester in the UK as an example. In the most basic sense of the task the project has been a success but as identified it is clear that several refinements need to be made to the tool before it can be considered truly useful addition to a HR operatives.

Appendix

Cluster 1

	Postal district	Neighbourhood	Local authority	Latitude	Longitude	Average Sold House Price, 000s	Cluster_Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	M1	Manchester central	Manchester	53.4773	-2.2351	346	0	Gay Bar	Hotel	Bar	Pub	Coffee Shop
1	M2	Manchester central	Manchester	53.4800	-2.2426	4008	0	Gay Bar	Coffee Shop	Hotel	Bar	Pub
2	M3	Manchester central, Salford Quays	Manchester	53.4836	-2.2502	263	0	Bar	Coffee Shop	Pub	Indian Restaurant	Restaurant
3	M4	Manchester central	Manchester	53.4846	-2.2291	328	0	Coffee Shop	Bar	Pub	Tea Room	Beer Bar
4	M5	Seedley, Weaste	Salford	53.4790	-2.2848	217	0	Indian Restaurant	Pizza Place	Food Truck	Brewery	None
7	M8	Crumpsall, Cheetham Hill	Manchester	53.5088	-2.2383	174	0	Light Rail Station	Asian Restaurant	Burger Joint	Park	Tram Station
8	M9	Harpurhey, Blackley	Manchester	53.5214	-2.2127	122	0	Pub	Hotel	Furniture / Home Store	Park	Bar
11	M13	Longsight	Manchester	53.4603	-2.2139	264	0	College Cafeteria	Asian Restaurant	Hookah Bar	Indian Restaurant	Sports Club
12	M14	Fallowfield, Moss Side, Ladybarn, Rusholme	Manchester	53.4477	-2.2244	341	0	Gym / Fitness Center	Bus Station	Coffee Shop	Fried Chicken Joint	Bed & Breakfast
13	M15	Hulme, Manchester Science Park	Manchester	53.4656	-2.2501	297	0	Grocery Store	Café	Discount Store	Park	Garden Center
14	M16	Firswood, Moss Side, Old Trafford, Whalley Range	Trafford, Manchester	53.4548	-2.2636	281	0	Grocery Store	Sandwich Place	Bus Stop	Shopping Mall	Yoga Studio
16	M18	Abbey Hey, Gorton	Manchester	53.4613	-2.1687	117	0	Gym / Fitness Center	Hotel	Bakery	Supermarket	Sandwich Place
17	M19	Levenshulme, Burnage	Manchester	53.4370	-2.1942	216	0	Gym / Fitness Center	Park	Optical Shop	Butcher	None
18	M20	Didsbury, Withington	Manchester	53.4252	-2.2303	336	0	Pizza Place	Café	Deli / Bodega	Bus Station	Bus Stop
19	M21	Chorlton Cum Hardy, Firswood	Manchester	53.4377	-2.2710	316	0	Grocery Store	Spanish Restaurant	Indian Restaurant	Japanese Restaurant	Falafel Restaurant
20	M22	Wythenshawe, Northenden, Sharston Industrial Area	Manchester	53.3856	-2.2586	192	0	Tram Station	Coffee Shop	Supermarket	Discount Store	None
21	M23	Baguley, Roundthorn Industrial Estate	Manchester	53.3990	-2.2873	190	0	Clothing Store	Fast Food Restaurant	Garden	Park	None
22	M24	Middleton, Alkrington	Rochdale	53.5511	-2.1962	153	0	Grocery Store	Warehouse Store	Stadium	Gym	Pub
23	M25	Prestwich	Bury	53.5285	-2.2746	531	0	Pub	Frozen Yogurt Shop	Pizza Place	Breakfast Spot	Grocery Store
24	M26	Radcliffe, Stoneclough	Bury, Bolton	53.5616	-2.3337	147	0	Pub	Pool	Fast Food Restaurant	Furniture / Home Store	Grocery Store
26	M28	Worsley, Walkden, Mosley Common, Wardley Indus...	Manchester, Salford	53.5147	-2.3967	234	0	Sports Club	Park	Convenience Store	None	None
27	M29	Tyldesley, Astley	Wigan	53.5082	-2.4569	156	0	Construction & Landscaping	Grocery Store	None	None	None
28	M30	Eccles	Salford	53.4842	-2.3542	174	0	Grocery Store	Gym	Pizza Place	Fast Food Restaurant	Train Station
29	M31	Carrington, Partington	Trafford	53.4193	-2.4214	164	0	Grocery Store	Convenience Store	IT Services	Yoga Studio	None
30	M32	Stretford	Trafford	53.4502	-2.3087	403	0	Park	Warehouse Store	Gym / Fitness Center	Bus Stop	None
31	M33	Sale, Brooklands	Trafford	53.4205	-2.3251	460	0	Pub	Grocery Store	Pizza Place	Chinese Restaurant	Sandwich Place
32	M34	Denton, Audenshaw	Tameside	53.4559	-2.1176	181	0	Supermarket	Clothing Store	Gym Pool	Café	Shopping Plaza
33	M35	Failsworth, Castle Industrial Estate	Oldham	53.5071	-2.1523	165	0	Pub	Bar	Gym / Fitness Center	Supermarket	None
34	M38	Little Hulton	Salford	53.5317	-2.4210	115	0	Pub	Asian Restaurant	None	None	None
35	M40	Collyhurst, Miles Platting, Moston, New Moston...	Manchester	53.5035	-2.1900	130	0	Pub	Café	Tram Station	Bus Stop	None
36	M41	Urmston	Trafford	53.4510	-2.3629	308	0	Pub	Playground	Convenience Store	Indian Restaurant	None
40	M46	Atherton	Wigan	53.5260	-2.4903	145	0	Pub	Supermarket	Bar	Sandwich Place	Pharmacy
41	M50	Salford Quays	Salford	53.4794	-2.3043	323	0	Construction & Landscaping	Tram Station	Paper / Office Supplies Store	Brewery	None

Cluster 2

	Postal district	Neighbourhood	Local authority	Latitude	Longitude	Average Sold House Price, 000s	Cluster_Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
10	M12	Ardwick, Longsight	Manchester	53.4648	-2.2019	151	1	Gaming Cafe	None	None	None	None

Cluster 3

	Postal district	Neighbourhood	Local authority	Latitude	Longitude	Average Sold House Price, 000s	Cluster_Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
9	M11	Clayton, Openshaw	Manchester	53.4783	-2.1793	122	2	Bakery	Indian Restaurant	None	None	None
15	M17	Trafford Park, The Trafford Centre	Trafford	53.4591	-2.3179	1100	2	Bakery	Sandwich Place	None	None	None
39	M45	Whitefield	Bury	53.5473	-2.2883	217	2	Indian Restaurant	Pub	Bridal Shop	Bakery	Deli / Bodega

Cluster 4

	Postal district	Neighbourhood	Local authority	Latitude	Longitude	Average Sold House Price, 000s	Cluster_Label	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
5	M6	Claremont, Irlams O' Th' Height	Salford	53.4916	-2.2970	158	3	Supermarket	None	None	None	None
6	M7	Higher Broughton, Cheetwood	Salford, Manchester	53.5062	-2.2609	218	3	Bakery	Pizza Place	Pub	Supermarket	River
25	M27	Swinton, Clifton, Pendlebury, Wardley	Salford	53.5122	-2.3363	173	3	Supermarket	Pizza Place	Park	Discount Store	Clothing Store
37	M43	Droylsden	Tameside	53.4838	-2.1477	137	3	Discount Store	Supermarket	Pharmacy	Soccer Stadium	None
38	M44	Irlam, Cadishead	Trafford	53.4409	-2.4261	159	3	Supermarket	Gym	Warehouse Store	Gas Station	Optical Shop

References

[1] ‘Managing Employee Relocation’, *SHRM*, Aug. 16, 2019. https://www.shrm.org/resourcesandtools/tools-and-samples/toolkits/pages/managingemployeerelocation.aspx (accessed Aug. 07, 2020).

[2] ‘M postcode area’, *Wikipedia*. Aug. 07, 2020, Accessed: Aug. 16, 2020. [Online]. Available: https://en.wikipedia.org/w/index.php?title=M_postcode_area&oldid=971595438.

[3] ‘Download UK Postcodes with Latitude and Longitude’. https://www.freemaptools.com/download-uk-postcode-lat-lng.htm (accessed Aug. 16, 2020).

[4] ‘Price Paid Data’, *GOV.UK*. https://www.gov.uk/government/statistical-data-sets/price-paid-data-downloads (accessed Aug. 16, 2020).

[5] ‘Foursquare Developer’. https://developer.foursquare.com/ (accessed Aug. 16, 2020).

[6] ‘2.3 Clustering’, *SciKit Learn*. https://scikit-learn.org/stable/modules/clustering.html#k-means (accessed Aug. 31, 2020).

[7] ‘Elbow method (clustering)’, *Wikipedia*. Apr. 21, 2020, Accessed: Aug. 31, 2020. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Elbow_method_(clustering)&oldid=952286034.

[8] ‘Overview | Places SDK for Android’, *Google Developers*. https://developers.google.com/places/android-sdk/overview (accessed Aug. 31, 2020).

[9] ‘DataPig86 - A Data Science Blog’. https://datapig86.blogspot.com/ (accessed Aug. 31, 2020).