Housing Sales Prices & Venues Data Analysis of Manchester

1. Introduction

Manchester is within the United Kingdom's second-most populous urban area, with a population of 545,500 as of 2017 and its population continues to grow rapidly, the Manchester City Council's 2018 State of the City Report shows that this growth is expected to continue rising up to 15% by 2025. The city attracts students, graduates and young professionals by employment opportunities, good quality accommodation, leisure and cultural offers. Manchester's economy has grown over the years, with its performance exceeding both Greater Manchester and the UK economy as a whole.

Many are flocking to Manchester because of the wide range of opportunities as well as economic and cultural diversity, many of those are young and have just started their career, in fact 65% of the graduates of universities in Manchester stay in the city after graduating. Thus, they need to find a place where the real estate values are lower but also surrounded by venues of their preference, which can be difficult to find.

This projects aims to create an analysis of features of Manchester neighbourhoods, including the median house sale prices in 2018, and density of venues in each area. This would help people that want to move to Manchester but do not know the area well, so they can make a more informed decision about which neighbourhood to live in. This study can also be beneficial towards investors who want to start a business in Manchester, for those it would be interesting to show the areas that present lower real estate and where the type of business they want is not very intense. Considering these points, I will create a segmented map and information charts including housing sale prices (HSP) and each area will be clustered according to its venue density.

2. Data acquisition and cleaning

2.1. Data Sources

For the longitude and latitude coordinates, the data can be found here. From this dataset, we will obtain the *postcode, name, latitude* and *longitude* of each neighbourhood.

For property sale prices, the data can be found here, the data includes the *names* and the *median house sale price* paid by each neighbourhood in England from 1995 to 2018. With the Python library *Pandas*, the data will be formatted, cleaned and tailored to what we actually need.

In order to get the most common venues in Manchester, we will use the Foursquare API, which provides diverse information about venues within a chosen radius. From this, we will be able to get information about the venues such as the venue's *neighbourhood, name, location and category*.

Finally, to build the *Choropleth* map, the neighbourhood boundaries will be extracted from the Manchester City Council website, and the file converted to the correct format.

2.2. Data Cleaning

Data downloaded from multiple sources were combined into a *pandas* data-frame. For the coordinate's dataset, there were postcodes that were not used anymore, so I selected only the ones that are still in use. It also included latitude and longitude for every postcode in Manchester, each area having many different postcodes, so I got the most central point to use as the neighbourhood coordinates. The house sale price data contained values for each neighbourhood in England from 1995 to 2018, selecting the information only from Manchester I decided to use the values from 2018 as it is the most recent year, dropping all the columns from previous years. The price values were saved as string objects, I converted them into integers so I could use them in the analysis. Luckily, the datasets in this project did not have any missing values. Both data-frames were merged according to the neighbourhood names, as it can be seen below:

	Ward	Median price by Ward	Latitude	Longitude	Postcode
0	Ancoats & Beswick	203250	53.481157	-2.213328	M4,M11,M12,M40
1	Ardwick	191995	53.465800	-2.217067	M1,M14,M12,M18,M13,M15
2	Baguley	150000	53.390994	-2.284337	M23
3	Brooklands	172000	53.406481	-2.296419	M33,M23
4	Burnage	192000	53.431485	-2.203522	M19,M20

Figure 1: Pandas data-frame containing the name of each neighbourhood in Manchester, median house sale price in 2018, latitude and longitude coordinates, and postcode.

Using Foursquare API, I limited the number of venues and the radius, however, after extracting the data I noticed that some neighbourhoods in Manchester had a very small number of venues, therefore, only the top ten areas with the most venues were considered in this analysis. After cleaning the data, there were 327 samples and 7 features, presenting 106 unique venue categories, as shown below:

	Ward	Ward Latitude	Ward Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Ardwick	53.4658	-2.217067	小樹林	53.464586	-2.223287	Chinese Restaurant
1	Ardwick	53.4658	-2.217067	Subway	53.465753	-2.215576	Sandwich Place
2	Ardwick	53.4658	-2.217067	McDonald's	53.467140	-2.217841	Fast Food Restaurant
3	Ardwick	53.4658	-2.217067	Greggs	53.465719	-2.216098	Bakery
4	Ardwick	53.4658	-2.217067	Booker	53.467352	-2.220819	Warehouse Store

Figure 2: Pandas data-frame containing the venues for each neighbourhood in Manchester, their coordinates, and categories.

3. Methodology

Using the python library *Geopy*, I was able to get the latitude and longitude coordinates of Manchester. With the *Folium library*, which allows us to create interactive maps using the coordinate dataset, I created a map of Manchester with its neighbourhoods superimposed on top:

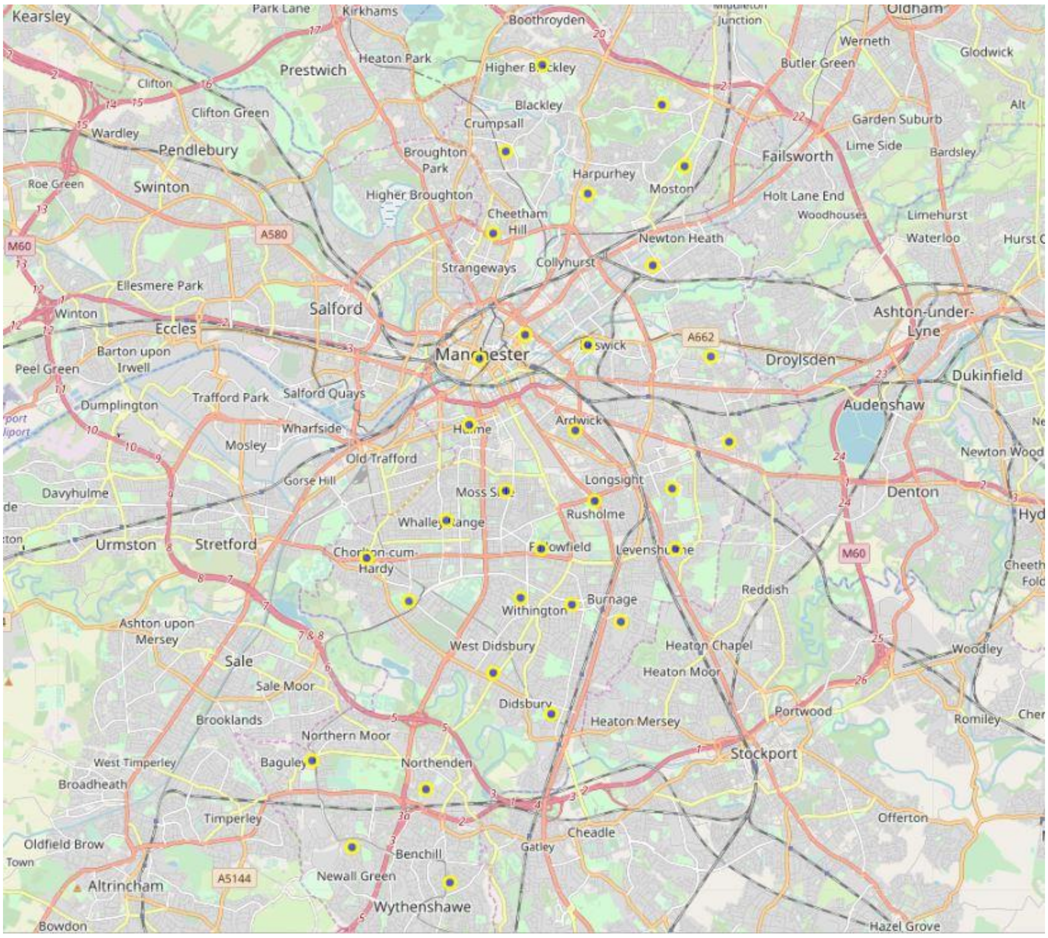

Figure 3: Map of Manchester with each neighbourhood superimposed on it, created using the python library *Folium.*

3.1. Venues analysis

With the help of Foursquare API, I limited the number of venues to 100 and the radius to 500m, and was able to obtain information about a total of 424 venues across Manchester, resulting in 131 unique venue categories. However, when we analyse the number of venues for each neighbourhood, we can see that some of them have a very small number of venues, as shown below:

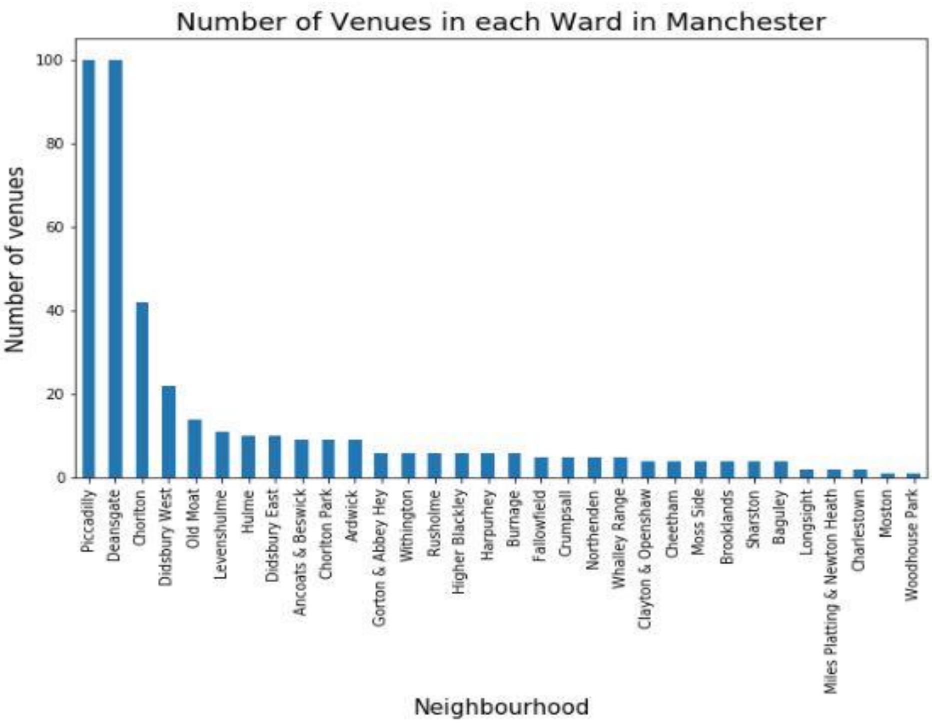

Figure 4: Bar chart showing the number of venues for each neighbourhood in Manchester, data acquired from Foursquare.

As some of the neighbourhoods have such a small number of venues, only the top 10 areas with the most venues will be considered in this analysis:

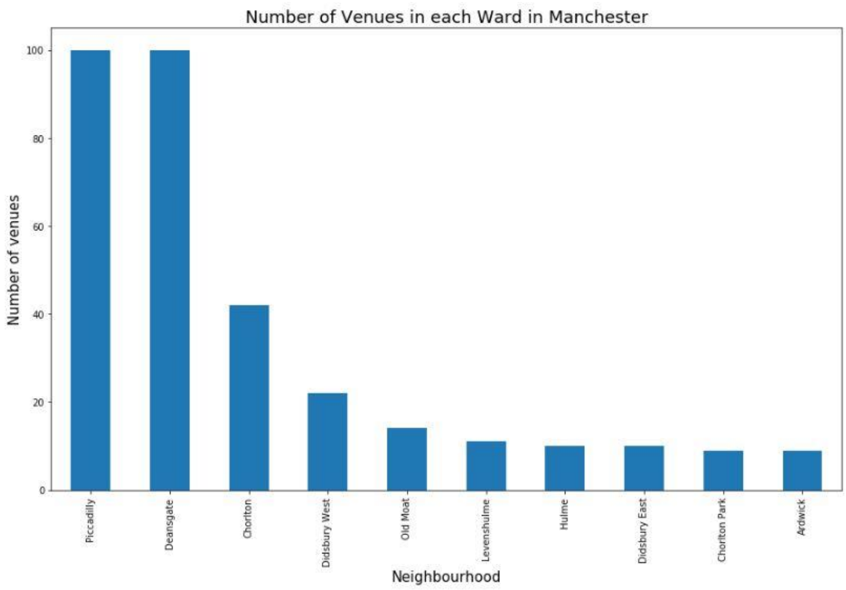

Figure 5: Bar chart showing the number of venues for the top 10 neighbourhoods with the most venues.

Considering these neighbourhoods, 106 unique venue categories were returned by Foursquare, and pubs & bars top the charts as we can see in the plot below:

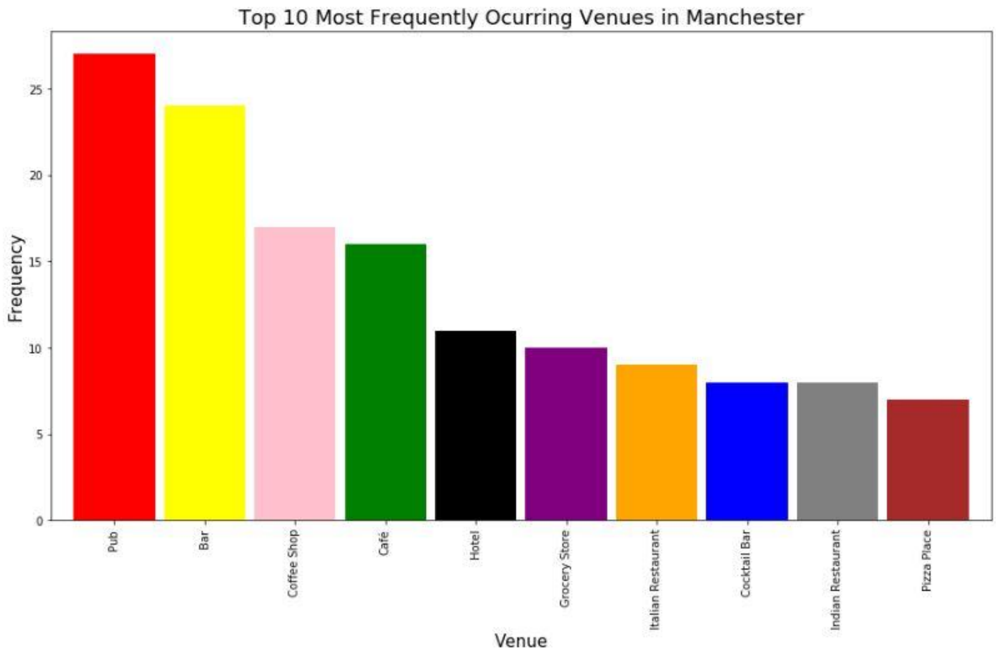

Figure 6: Bar chart showing the ten most frequently occurring venues in Manchester.

The new data-frame for the top 10 areas in Manchester contains a total of 327 samples and 7 features, as shown in the previous section, Figure *2*. Using one-hot encoding process I was able to convert the venues categories into binary vectors, which allowed us to get the frequency of occurrence of each category in each neighbourhood. The table below shows the top 10 most common venue for each neighbourhood in Manchester:

	Ward	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Ardwick	Pub	Warehouse Store	Chinese Restaurant	Pet Store	Bakery	Sandwich Place	Fast Food Restaurant	Bus Station	Park	Donut Shop
1	Chorlton	Pub	Bar	Café	Grocery Store	Pizza Place	Turkish Restaurant	Gastropub	Pharmacy	Mediterranean Restaurant	Deli / Bodega
2	Chorlton Park	Grocery Store	Convenience Store	Tram Station	Gas Station	Fish & Chips Shop	Park	Fabric Shop	Falafel Restaurant	Comic Shop	Concert Hall
3	Deansgate	Italian Restaurant	Hotel	Coffee Shop	Pub	Asian Restaurant	Bar	Plaza	Burger Joint	Cocktail Bar	Thai Restaurant
4	Didsbury East	Pub	Supermarket	Hotel	Fish & Chips Shop	Park	Coffee Shop	Chinese Restaurant	Gym / Fitness Center	Bakery	Fast Food Restaurant

Figure 7: Pandas data-frame containing the ten most common venues for each neighbourhood.

As shown above, there are some common venue categories in the areas. For this reason, we will use the unsupervised machine learning K-means algorithm to cluster the neighbourhoods.

3.2. Clustering the neighbourhoods

K-means is a relatively efficient method, which intends to partition the dataset into k clusters, minimizing the total intra-cluster variance (squared error function). So the aim of this process is to partition the neighbourhood set into groups that have similar characteristics, so the costumer can choose a specific group with the characteristics they are interested in. In order to do so, I used the following code:

```python
# set number of clusters
kclusters = 3

mcr_grouped_clustering = mcr_grouped.drop('Ward', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(mcr_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

```
array([0, 0, 2, 0, 0, 0, 1, 1, 1, 0], dtype=int32)
```

Figure 8: Python code for implementing the k-means algorithm, with number of clusters = 3.

Adding the cluster labels to the main data-frame:

	Ward	Median price by Ward	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Ardwick	191995	53.465800	-2.217067	0	Pub	Warehouse Store	Chinese Restaurant	Pet Store	Bakery	Sandwich Place	Fast Food Restaurant	Bus Station	Park	Donut Shop
1	Chorlton	334000	53.442839	-2.279974	0	Pub	Bar	Café	Grocery Store	Pizza Place	Turkish Restaurant	Gastropub	Pharmacy	Mediterranean Restaurant	Deli / Bodega
2	Chorlton Park	280650	53.435050	-2.267282	2	Grocery Store	Convenience Store	Tram Station	Gas Station	Fish & Chips Shop	Park	Fabric Shop	Falafel Restaurant	Comic Shop	Concert Hall
3	Deansgate	224975	53.478579	-2.246147	0	Italian Restaurant	Hotel	Coffee Shop	Pub	Asian Restaurant	Bar	Plaza	Burger Joint	Cocktail Bar	Thai Restaurant
4	Didsbury East	315000	53.414762	-2.224512	0	Pub	Supermarket	Hotel	Fish & Chips Shop	Park	Coffee Shop	Chinese Restaurant	Gym / Fitness Center	Bakery	Fast Food Restaurant

Figure 9: Pandas data-frame containing the names, the median house sale prices and the cluster labels for each neighbourhood in Manchester.

Using the *Folium* library, we can represent these 3 clusters on a map of Manchester where the red colour represents cluster 0, purple represents cluster 1 and green represent cluster 2:

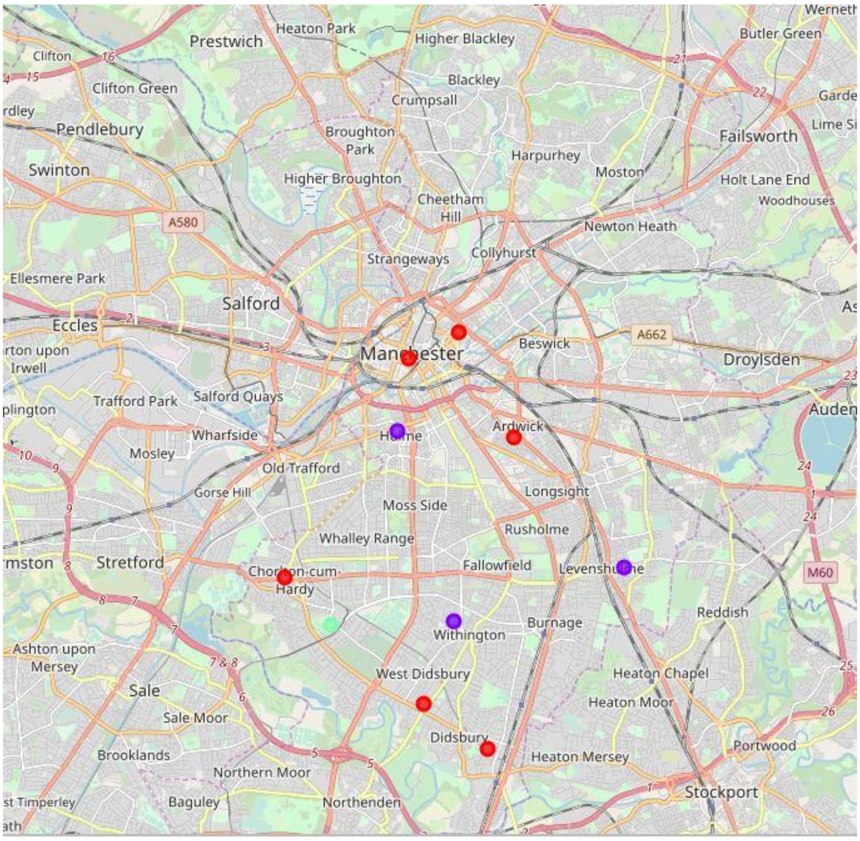

Figure 10: Map created with the Folium library, showing the segmentation of neighbourhood in Manchester into 3 different clusters.

Examining each cluster, we can plot the number of 1st Most Common Venue in each cluster, as follows:

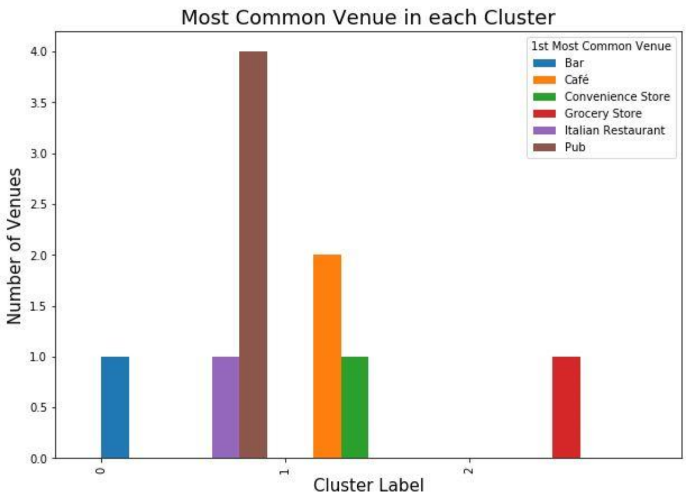

Figure 11: Bar chart representing the most frequently occurring venues in each cluster.

We can also analyse the following cluster data-frames that contain the top 3 venues for each neighbourhood:

Cluster 0

```
#get the ward and the most common venues columns
mcr_merged.loc[mcr_merged['Cluster Labels'] == 0, mcr_merged.columns[[0] + list(range(5, mcr_merged.shape[1]))]]
```

	Ward	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Top 3 Venues
0	Ardwick	Pub	Warehouse Store	Chinese Restaurant	Pet Store	Bakery	Sandwich Place	Fast Food Restaurant	Bus Station	Park	Donut Shop	Pub,Warehouse Store,Chinese Restaurant
1	Chorlton	Pub	Bar	Café	Grocery Store	Pizza Place	Turkish Restaurant	Gastropub	Pharmacy	Mediterranean Restaurant	Deli / Bodega	Pub,Bar,Café
3	Deansgate	Italian Restaurant	Hotel	Coffee Shop	Pub	Asian Restaurant	Bar	Plaza	Burger Joint	Cocktail Bar	Thai Restaurant	Italian Restaurant,Hotel,Coffee Shop
4	Didsbury East	Pub	Supermarket	Hotel	Fish & Chips Shop	Park	Coffee Shop	Chinese Restaurant	Gym / Fitness Center	Bakery	Fast Food Restaurant	Pub,Supermarket,Hotel
5	Didsbury West	Pub	Tennis Court	Bar	Bakery	Liquor Store	Bus Stop	Café	Deli / Bodega	Cupcake Shop	Italian Restaurant	Pub,Tennis Court,Bar
9	Piccadilly	Bar	Coffee Shop	Pub	Cocktail Bar	Café	Tea Room	Bookstore	Hotel	Indian Restaurant	Record Shop	Bar,Coffee Shop,Pub

Cluster 1

```
mcr_merged.loc[mcr_merged['Cluster Labels'] == 1, mcr_merged.columns[[0] + list(range(5, mcr_merged.shape[1]))]]
```

	Ward	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Top 3 Venues
6	Hulme	Café	Grocery Store	Vegetarian / Vegan Restaurant	Convenience Store	Deli / Bodega	Performing Arts Venue	Tanning Salon	Park	Fast Food Restaurant	Wine Bar	Café,Grocery Store,Vegetarian / Vegan Restaurant
7	Levenshulme	Café	Indian Restaurant	Pub	Fried Chicken Joint	Farmers Market	Grocery Store	Sandwich Place	Gym / Fitness Center	Antique Shop	Bakery	Café,Indian Restaurant,Pub
8	Old Moat	Convenience Store	Indian Restaurant	Grocery Store	Vegetarian / Vegan Restaurant	Pool	Asian Restaurant	Pub	Burger Joint	Sandwich Place	Park	Convenience Store,Indian Restaurant,Grocery Store

Cluster 2

```
mcr_merged.loc[mcr_merged['Cluster Labels'] == 2, mcr_merged.columns[[0] + list(range(5, mcr_merged.shape[1]))]]
```

	Ward	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Top 3 Venues
2	Chorlton Park	Grocery Store	Convenience Store	Tram Station	Gas Station	Fish & Chips Shop	Park	Fabric Shop	Falafel Restaurant	Comic Shop	Concert Hall	Grocery Store,Convenience Store,Tram Station

Figure 12: Analysis of the ten most common venues for each cluster.

With the information on Figure *11* and Figure *12*, we can label each cluster as:

- Cluster 0: “Drinking cluster”
- Cluster 1: “Eating cluster”
- Cluster 2: “Residential cluster”

3.3. House sale price

For the chosen neighbourhoods, the median house sale prices were plotted as follows:

Figure 13: Bar chart showing the median house sale price by neighbourhood in Manchester.

We can see that most of the neighbourhoods with a high number of venues also have high house sale price values. If we look into the frequency of median house sale price in Manchester, we get the following histogram:

Figure 14: Histogram showing the frequency of the median house sale prices in Manchester in 2018.

From Figure *14*, we can define price ranges as follows:

- Up to £140k: Low-1 Level House Price
- £140k-£190k: Low-2 Level House Price
- £190k-£240k: Mid-Level House Price
- £240k-£290k: High-1 Level House Price
- More than £290k: High-2 Level House Price

Now that we have all the labels we needed to classify each neighbourhood, we can merged it all together in the main data-frame, containing the cluster names as well as the price labels, as follows:

Ward	Median price by Ward	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Top 3 Venues	Price Labels	Cluster Names
Ardwick	191995	53.465800	-2.217067	0	Sandwich Place	Warehouse Store	Park	Bus Station	Pub	Chinese Restaurant	Pet Store	Fast Food Restaurant	Bakery	Wine Bar	Sandwich Place,Warehouse Store,Park	Mid-Level Price	Drinking cluster
Chorlton	334000	53.442839	-2.279974	0	Pub	Bar	Grocery Store	Café	Turkish Restaurant	Pizza Place	Gastropub	Cocktail Bar	Pharmacy	Deli / Bodega	Pub,Bar,Grocery Store	High-2 Level Price	Drinking cluster
Chorlton Park	280650	53.435050	-2.267282	2	Grocery Store	Convenience Store	Tram Station	Gas Station	Fish & Chips Shop	Fabric Shop	Park	Falafel Restaurant	Comic Shop	Concert Hall	Grocery Store,Convenience Store,Tram Station	High-1 Level Price	Residential Cluster
Deansgate	224975	53.478579	-2.246147	0	Hotel	Italian Restaurant	Coffee Shop	Pub	Bar	Plaza	Asian Restaurant	Thai Restaurant	Spanish Restaurant	Cocktail Bar	Hotel,Italian Restaurant,Coffee Shop	Mid-Level Price	Drinking cluster
Didsbury East	315000	53.414762	-2.224512	0	Pub	Supermarket	Fish & Chips Shop	Park	Coffee Shop	Gym / Fitness Center	Chinese Restaurant	Hotel	Bakery	Fast Food Restaurant	Pub,Supermarket,Fish & Chips Shop	High-2 Level Price	Drinking cluster

Figure 15: Pandas data-frame containing the cluster labels and names as well as the price range labels for each neighbourhood in Manchcester.

With the data-frame above, I was able to create a *Choropleth* map using the *Folium* library. This is a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map. It provides an easy way to visualize how a measurement varies across a geographic area or it shows the level of variability within a region. In this project, we will plot the different house sale price ranges for each neighbourhood in Manchester:

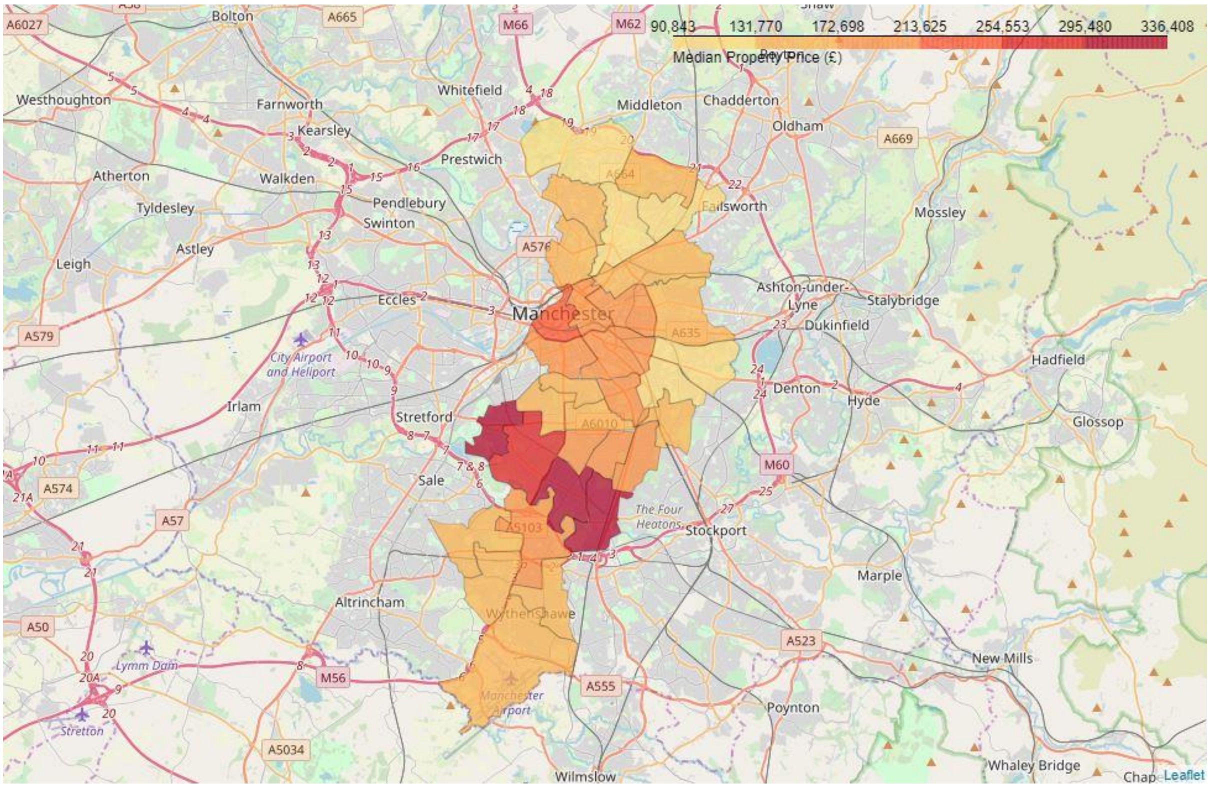

Figure 16: Choropleth map showing the different house sale prices across Manchester.

The darker colours on Figure *16* represent the neighbourhoods with the highest house sale price, Chorlton, Didsbury East and Didsbury West, which agrees with Figure *13*.

Finally, we can add the clusters to the *choropleth* map, having a pop-up label with information about the neighbourhood and cluster names and the house sale price range:

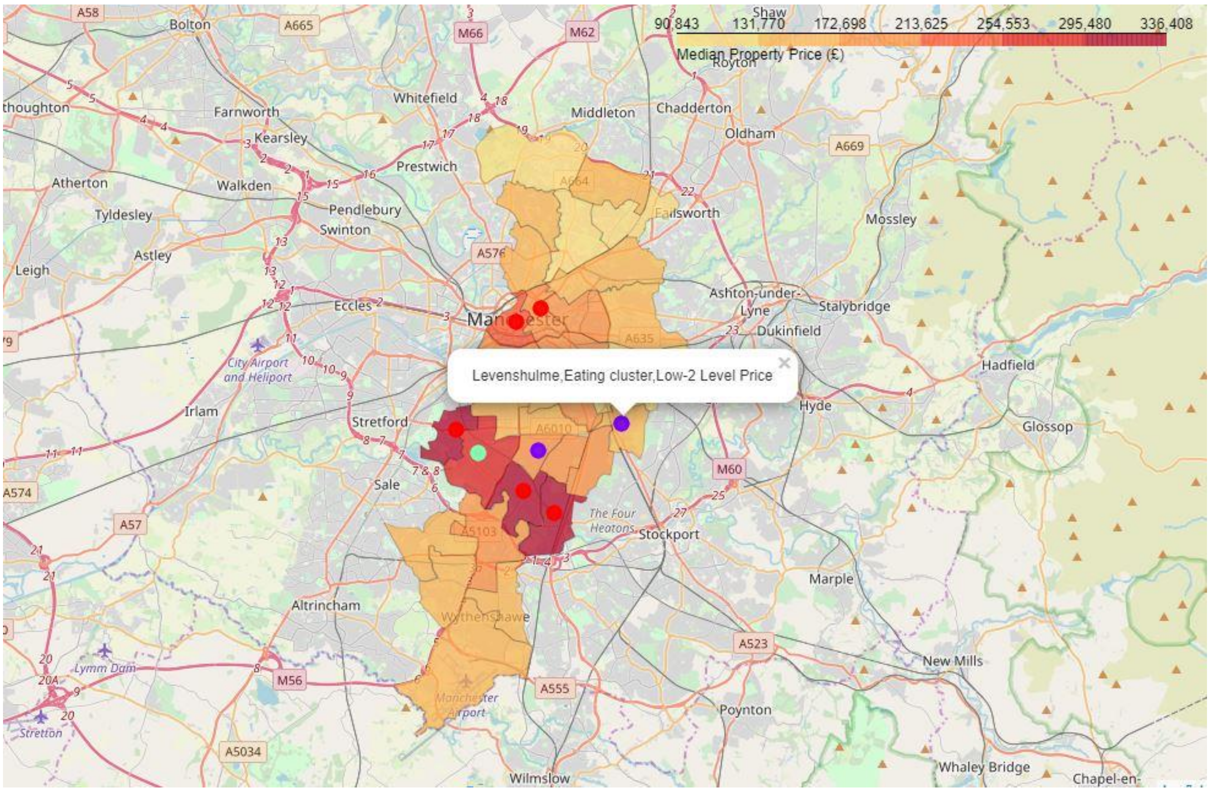

Figure 17: Choropleth map containing the different house sale prices across Manchester and the segmentation of its neighbourhoods according to their venues categories.

Analysing Figure *17*, for the people that are looking to buy a house where the real estate values are lower and there are more 'eating out' venues, Levenshulme would be a great option. And for the investors that are looking to open a pub or bar, stay out of the neighbourhood in the 'Drinking cluster' as this venue is very common in these areas.

4. Results & Discussion

This analysis showed the number and type of venues for 10 neighbourhoods in Manchester, as well as the median house sale price for each area, as we wanted to study the possibility of either opening a new venue or finding a good neighbourhood to live. With the use of data from web resources, python libraries and Foursquare API, I was able to set up a very realistic data-analysis scenario. With this project we found out that:

- The top venues in Manchester are pubs and bars.
- Chorlton, Didsbury East, Didsbury West and Chorlton Park have the highest house sale price and are also among the neighbourhood with the most venues in Manchester.
- Ardwick, Chorlton, Deansgate, Didsbury East, Didsbury West and Piccadilly are dominated by pubs and bars as the most common venue whereas Hulme, Levenshulme and Old Moat are dominated by cafes and restaurants as most common venues.

- Chorlton Park was separated from both of these clusters as grocery stores stand out as the most common venue, hence being considered a residential area.

According to this analysis, the neighbourhood which provides the least competition for a new social venue would be Chorlton Park, however, when you consider the prices, it might be better to consider what type of business it is so you can look into a lower price area.

As mentioned before, for the young people that are looking to buy a house where the real estate values are lower and there are more 'eating out' venues such as cafes and restaurants, Levenshulme is the perfect option.

A drawback of this analysis is that the clustering is completely based on the most common venues returned by Foursquare, and some neighbourhoods had a very small number of data, if we had more information we could have done an analysis of the whole city of Manchester, which would open up a lot more possible outcomes. A future direction for this project is to use the median house sale price data from the years 1995 to 2018 to predict the values for 2019 using regression.

5. Conclusion

In this study, I analysed the venues in 10 neighbourhoods in Manchester as well as the median house sale price for each area. I used python libraries to acquire, clean and analyse the data, Foursquare API to get information about the venues. Then, I analysed the results of segmentation of the neighbourhoods into clusters based on the most common venues using k-means algorithm, to finally build a *choropleth* map, which contains all the information acquired in this analysis: name of neighbourhood, name of cluster and house sale price range. Although these are not the only considerations when buying a property or opening a business, it certainly gives us some very important preliminary information about the neighbourhoods and would serve as guidance for the next steps.