Housing Sales Prices and Venues Analysis for Liverpool

1. Introduction and Business problem

1.1 Description and Discussion of the Background

Liverpool is a city and metropolitan borough in Merseyside, England. The population in 2019 was approximately 498,042. Liverpool is the ninth-largest English district by population, and the largest in Merseyside and the surrounding region. It lies within the UK's sixth-most populous urban area, and its metropolitan area is the fifth-largest in the UK with a population of 2.24 million. Liverpool is the sixth-most-visited UK city. It is noted for its culture, architecture, and transport links. The city is closely associated with the arts, especially music; the popularity of the Beatles, widely regarded as the most influential music group in history, contributed to the city's status as a tourist destination. Since then, Liverpool has continued to produce many notable musical acts and record labels—musicians from Liverpool have produced 56 No. 1 hit singles, more than any other city in the world. Liverpool also has a long-standing reputation as the origin of various actors and actresses, artists, athletes, comedians, journalists, novelists, and poets. The city has the second-highest number of art galleries, national museums, listed buildings, and listed parks in the UK; only the capital, London, has more. The Liverpool Maritime Mercantile City includes the Pier Head, Albert Dock, and William Brown Street. In sports, the city is best known for being the home of Premier League football clubs Liverpool and Everton, with matches between the two being known as the Merseyside derby. The annual Grand National horse race takes place at Aintree Racecourse

Several areas of the city centre were granted World Heritage Site status by UNESCO in 2004, and its collection of parks and open spaces has been described as the "most important in the country" by the Register of Historic Parks and Gardens of Special Historic Interest. Its status as a port city has attracted a diverse population from a wide range of cultures and religions, primarily Ireland, Norway, and Wales. It is also home to the oldest black community in the UK and the oldest Chinese community in Europe. Natives and long-time residents of Liverpool are formally referred to as "Liverpudlians" but are more often called "Scousers", a reference to the form of stew made popular by sailors in the city and the most common name for the Liverpool accent and dialect. The city celebrated its 800th anniversary in 2007 and was named the 2008 European Capital of Culture, jointly with Stavanger, Norway. [1]

Liverpool is divided into approximately 50 postcode areas. This is a group of postcode districts in north-west England, which are subdivisions of four post towns. These cover most of Merseyside (including

Liverpool, Bootle and Prescot), part of west Lancashire (including Ormskirk) and a small part of north-west Cheshire

As mentioned before Liverpool is a city with a high population and population density hence it might be a challenge for investors to find a best place to open their business taking into account a lower real estate costs and depending on the nature of their business also less competitive area. On the other hand residents of the city might be more interested in a lower properties prices but areas that are more attractive due to majority of a social places. Unfortunately this information is not easy to find nowadays that will guide investors which district/areas to choose taking into account these indexes.

1.2 Data description

To find a solution of this problem I followed these steps to gather and analyse data:

- Obtained a list of postcodes and areas of City of Liverpool from doogal.co.uk [2]. The csv file consists of many valuable information and coordinates for each postcode areas. File required data cleaning but consist of coordinates of all areas.
- Obtained a data of average house prices (collected by postcodes) from website propertydata.co.uk [3], converted the data into a data frame and cleaned and removed unnecessary data.
- Combined two data frame together.
- Used **Foursquare API** [4] to get the most common venues of areas of Liverpool.
- Created a graphs, tables and maps to show the results.

2. Methodology

My data base has been downloaded as CSV file thanks to doogal.co.uk. This data consisted from all postcodes in UK. I cleaned the data by exploring only Liverpool postcodes using 'Region' column as a searching id and dropped the column that were not necessary for my analysis.¶

	Postcode	Latitude	Longitude	Town/Area	Region	Population	Households
0	L1	53.4026	-2.98031	Liverpool city centre	Liverpool	7096.0	3821.0
1	L2	53.4068	-2.99012	Liverpool city centre	Liverpool	935.0	622.0
2	L3	53.4089	-2.98554	Liverpool City Centre, Everton, Vauxhall	Liverpool	17396.0	7250.0
3	L4	53.4378	-2.96203	Anfield, Kirkdale, Walton	Liverpool	40091.0	18612.0
4	L5	53.4248	-2.97763	Anfield, Everton, Kirkdale, Vauxhall	Liverpool	12961.0	6476.0

I downloaded data of housing average price by postcode. Imported necessary libraries to download the data from website and read data as a dataframe. Cleaned and processed the data changing it's header, dropping unnecessary columns, changing the column name Area to "Postcode" be able to merge two dataframes together.

	Postcode	Avg yield	Avg price	£/sqft
0	L1	8.1%	£104,141	£234
1	L2	7.4%	£127,192	£285
2	L3	6.3%	£135,985	£205
3	L4	6.3%	£92,953	£73
4	L5	6.8%	£88,455	£96

I have used Folium library to visualize geographic details of Liverpool by postcodes and areas names. I have used latitude and longitude values to get the visual as below:

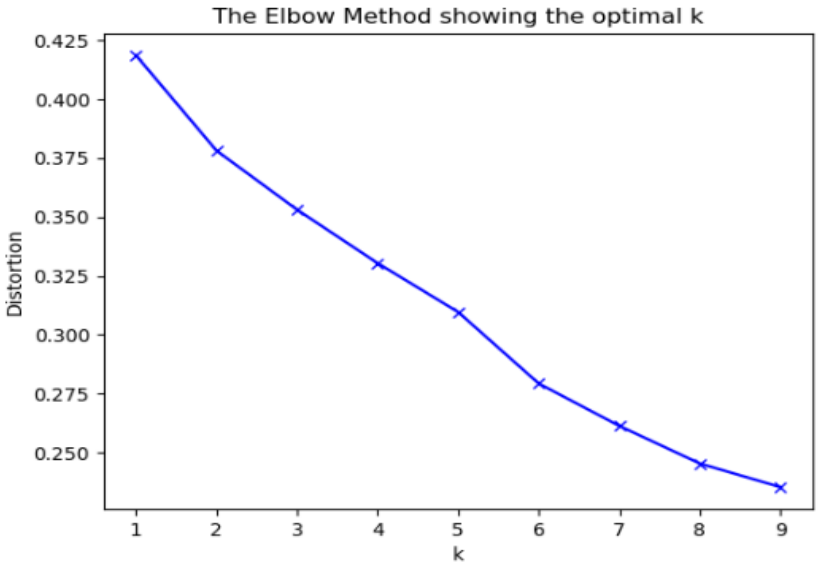
In the next step I have utilised Foursquare API to explore and segment the areas of Liverpool. I set up the limit to **100 venues** and radius to **750 m** for each area from latitude and longitude information obtained before. In total **100** venues and **113** unique categories have been returned from Foursquare. Here is the first 5 elements of the list of venues names, category latitude and longitude information:

	name	categories	lat	lng
0	Kazimier Garden	Beer Garden	53.402805	-2.981879
1	Leaf	Café	53.402869	-2.977708
2	Mowgli Street Food	Indian Restaurant	53.402828	-2.977600
3	FACT Cinema and Art Gallery	Indie Movie Theater	53.402314	-2.977378
4	Crust	Pizza Place	53.403690	-2.979528

And after merging areas and venues my data table look as below:

	Town/Area	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Liverpool city centre	53.4026	-2.98031	Kazimier Garden	53.402805	-2.981879	Beer Garden
1	Liverpool city centre	53.4026	-2.98031	Leaf	53.402869	-2.977708	Café
2	Liverpool city centre	53.4026	-2.98031	Mowgli Street Food	53.402828	-2.977600	Indian Restaurant
3	Liverpool city centre	53.4026	-2.98031	FACT Cinema and Art Gallery	53.402314	-2.977378	Indie Movie Theater
4	Liverpool city centre	53.4026	-2.98031	Crust	53.403690	-2.979528	Pizza Place

Due to common venue categories in areas of Liverpool I have applied the most common cluster method of unsupervised learning algorithm **K-Means** to cluster the areas**.** I found the optimum k for K-Means algorithm using Elbow method and applied **Euclidean** distance.

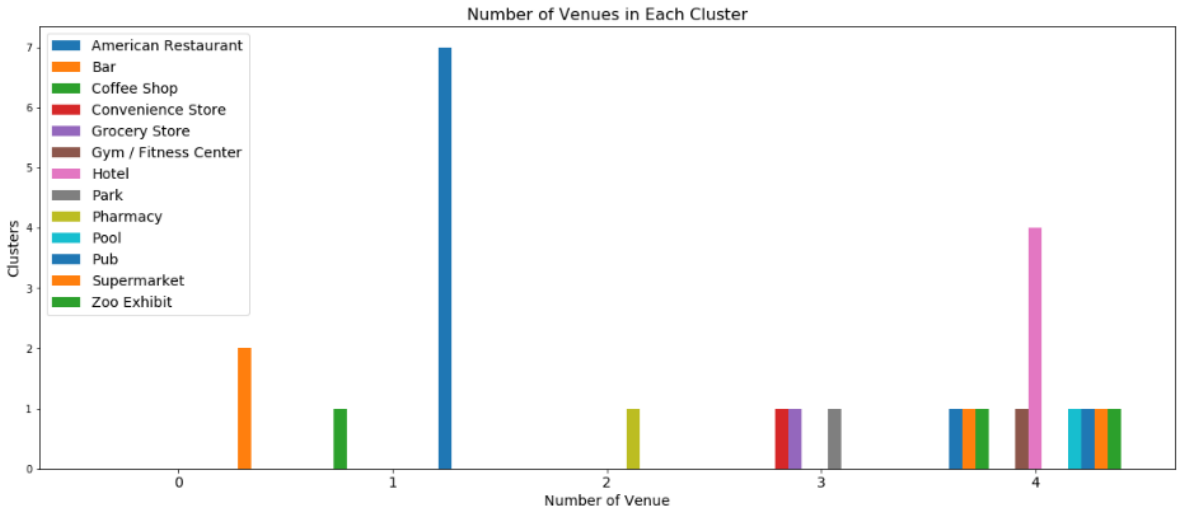


And my merged data frame with cluster labels looks like follow:

	Postcode	Latitude	Longitude	Town/Area	Region	Population	Households	Avg price	£/sqft	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	L1	53.4026	-2.98031	Liverpool city centre	Liverpool	7096.0	3821.0	£104,141	£234	4	Hotel	Coffee Shop	Pub	Café	Italian Restaurant	Cocktail Bar	Restaurant	Burger Joint	Shopping Mall	Sandwich Place
1	L2	53.4068	-2.99012	Liverpool city centre	Liverpool	935.0	622.0	£127,192	£285	4	Hotel	Coffee Shop	Pub	Café	Italian Restaurant	Cocktail Bar	Restaurant	Burger Joint	Shopping Mall	Sandwich Place
2	L3	53.4089	-2.98554	Liverpool City Centre, Everton, Vauxhall	Liverpool	17396.0	7250.0	£135,985	£205	4	Hotel	Pub	Italian Restaurant	Café	Coffee Shop	Sandwich Place	Bar	Restaurant	Fast Food Restaurant	Theater
3	L4	53.4378	-2.96203	Anfield, Kirkdale, Walton	Liverpool	40091.0	18612.0	£92,953	£73	1	Pub	Soccer Stadium	Sporting Goods Shop	Park	Concert Hall	Convenience Store	Cosmetics Shop	Cricket Ground	Department Store	Dessert Shop
4	L5	53.4248	-2.97763	Anfield, Everton, Kirkdale, Vauxhall	Liverpool	12961.0	6476.0	£88,455	£96	4	Gym / Fitness Center	Discount Store	Fast Food Restaurant	Supermarket	Zoo Exhibit	Flea Market	Cosmetics Shop	Cricket Ground	Department Store	Dessert Shop

I created a bar chart to be able to name the clusters.

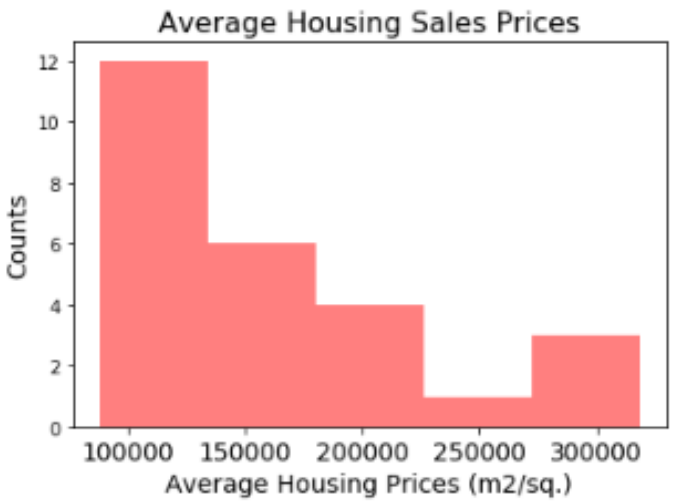
```
Text(0, 0.5, 'Clusters')
```



After analysing the above graph I have given a name to each cluster as follow:

	Clusters	Labels
0	0	Bars Venues
1	1	Pubs and Restaurants
2	2	Stores
3	3	Pharmacy Venues
4	4	Hotels Gyms and Accomodation Venues

In the next step I have created a data frame called “housing_data” that consist of 5 columns such as: Postcode, Area name, Avg price, £sqft and created cluster labels that enabled me to show on histogram the frequency of average housing sales price :



Next I defined the ranges as:

100000 AHP : "Low Level HSP"

100000-150000 AHP : "Mid-1 Level HSP"

150000-200000 AHP : "Mid-2 Level HSP"

200000-250000 AHP : "High-1 Level HSP"

>250000 AHP : "High-2 Level HSP"

In the next step I have added above ranges to my data frame as well as created a new data frame to show top 3 venues for each area as shown below:

	Town/Area	Top 3 Venues
0	Aigburth, Garston, Grassendale	2 Supermarket, 1 Hotel, 1 Pool
1	Aigburth, St Michael's Hamlet, Sefton Park	3 Pub, 2 Grocery Store, 2 Turkish Restaurant
2	Aintree, Fazakerley, Orrell Park, Walton	1 Café, 1 Grocery Store, 1 Park
3	Allerton, Mossley Hill	1 Coffee Shop, 1 Fast Food Restaurant, 1 Groce...
4	Anfield, Everton, Kirkdale, Vauxhall	1 Discount Store, 1 Fast Food Restaurant, 1 Gy...

3. Results

My final data frame after merging 3 top venues data frame and cluster label data frame looked as below:

	Postcode	Latitude	Longitude	Town/Area	Region	Population	Households	Avg price	£/sqft	Cluster Labels	...	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	Labels	Level_labels
0	L1	53.4026	-2.98031	Liverpool city centre	Liverpool	7096.0	3821.0	£104,141	£234	4	...	Café	Italian Restaurant	Cocktail Bar	Restaurant	Burger Joint	Shopping Mall	Sandwich Place	14 Hotel, 12 Coffee Shop, 10 Pub	Hotels Gyms and Accomodation Venues	Mid-1 Level HSP
1	L1	53.4026	-2.98031	Liverpool city centre	Liverpool	7096.0	3821.0	£104,141	£234	4	...	Café	Italian Restaurant	Cocktail Bar	Restaurant	Burger Joint	Shopping Mall	Sandwich Place	14 Hotel, 12 Coffee Shop, 10 Pub	Hotels Gyms and Accomodation Venues	Mid-1 Level HSP
2	L2	53.4068	-2.99012	Liverpool city centre	Liverpool	935.0	622.0	£127,192	£285	4	...	Café	Italian Restaurant	Cocktail Bar	Restaurant	Burger Joint	Shopping Mall	Sandwich Place	14 Hotel, 12 Coffee Shop, 10 Pub	Hotels Gyms and Accomodation Venues	Mid-1 Level HSP

The last step was to create a map with clustered area of Liverpool:

4. Discussion

As mentioned at the beginning of this project Liverpool is a big city with a high population density. Due to a big complexity many different approaches can be used to cluster and classify areas. It's worth to remember that not every approach will result in the same high quality.

I decided to use KMeans algorithm due to the fact that I have received an optimum k which gave me an insight of the best division of the areas. To this project I have used the areas from postcodes L1-L40 and this is due the fact that the data I obtained for average housing price was only for these postcodes. Nevertheless for more detailed and accurate guidance for investors or residents of the city who are looking for a new place to live that data set can be expanded to more areas and more detailed analysis can be obtained.

The next step of this project can be adding a choropleth map to show average housing sales price and additional features as areas, clusters names and top 3 venues. Unfortunately I was not able to obtained a correct map that would show me the wards of Liverpool in the areas as in my data set neither in JSON or SHP format.

The project finished on the stage of visualizing the data and clustering the information on the Liverpool map.

5. Conclusion

People are interested in Liverpool not only because the beauty of this city but also because if the history, culture, fantastic people, football souls, universities and lots of different attractions throughout the city. It is a good destination for investors to start a new business as well as for people to work here and live.

The information showed in this study can help people/investors/managers to understand the areas of Liverpool, their popular venues as well as the housing prices in Liverpool City.

6. References

[1] https://en.wikipedia.org/wiki/L_postcode_area

[2] https://propertydata.co.uk/cities/liverpool

[3] https://propertydata.co.uk/cities/liverpool

[4] https://foursquare.com/