IBM DATA SCIENCE CAPSTONE

BATTLE OF NEIGHBOURHOODS

TABLE OF CONTENTS :

- INTRODUCTION
- DATA SECTION
- METHODOLOGY
- RESULTS
- DISCUSSION
- ACKNOWLEDGEMENT
- CONCLUSION

INTRODUCTION :

- London is one of the most happening places in the UK. Famous for its lavish lifestyle and architecture ,it is one of the most preferred places to reside in.
-
- However , owing to the expensive residencies sided by the current economic and financial crisis , housing prices in London have been on an unprecedent rise causing real estate agencies to back out during the downfall.
- Owing to the current scenario , it has indeed become difficult for families wanting to live in independent homes to buy houses of their choice at affordable prices.
- Thus the question posed is – Is there affordable housing in London? if so where ?Is the locality good and how is the proximity with respect to facilities such as schools , hospitals and restaurants.

AIM OF THE PROJECT :

The project aims at rendering support to those interested in buying their own homes,to help the make wise and effective decisions under ongoing financial and economic constraints.

Target audience :

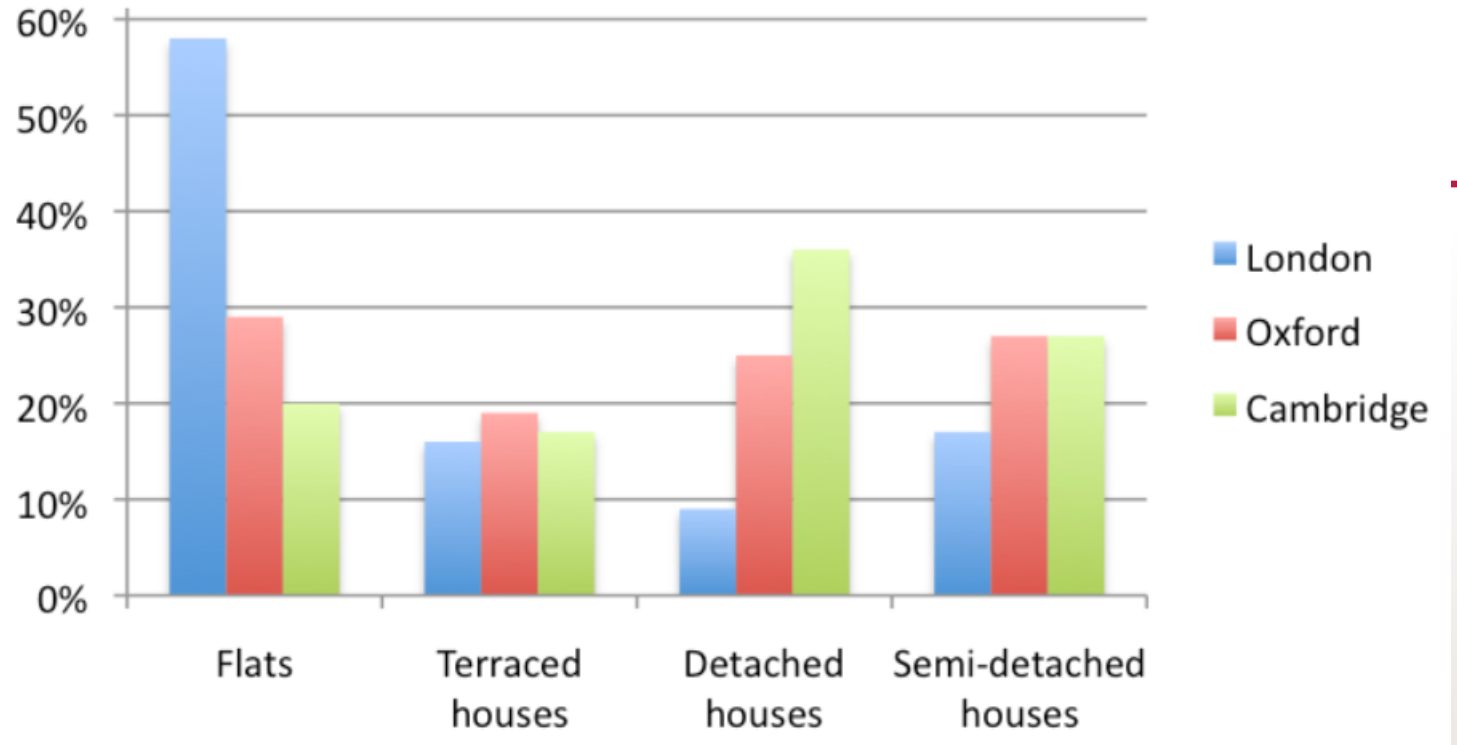
- Families or individuals wanting to purchase independent homes
- Real estate agencies on the lookout for house sales

Business problem :

The question at hand is 'Are there medium priced houses available for purchase in London? if so , is the construction sound or compromised in lieu with the minimal cost. Are there houses available in good localities with access to amenities such as hospitals , gyms , schools etc'. This project aims to address the above issue.

HOUSING AND PRICING PREFERNCES :

- The average purchase price for a first time buyer in Greater and Central London is around £509,047 while in South East London it costs around £339,510.However , considering the average pricing of some of the most exemplary localities in London a broad spectrum average of £2,202,125 was determine to set our price range in order to find residencies which fall in the same boundary.
- It was also worth taking note of the fact that most people in London , almost 60% prefer residing in flats or condominiums as it was less expensive causing many people to opt out of their preferred choices as shown by the graph on the next slide.



HOUSING PREFERNCES:

IT CAN BE OBSERVED THAT FLATS HAVE A HIGHER PREFERENCE OVER INDEPENDENT HOUSING.

DATA SECTION:
SOURCES

1. Data on house purchases in the UK were obtained from http://landregistry.data.gov.uk/
2. Other data sources used for reference were : https://www.testbig.com/ielts-writing-task-i-essays/chart-illustrates-housing-preferences-among-people-who-lived-british
3. https://www.homesandproperty.co.uk/property-news/buying/the-best-areas-to-start-your-search-for-a-family-house-with-good-links-to-the-city-a116886.html
4. https://www.zoopla.co.uk/discover/first-time-buyers/london/#4jJFWixLL2Hl0cuD.97

Data acquisition to solve the business problem:

- Data was acquired for the following purpose-
- House purchase real estate record data of UK from which a list of London streets and their corresponding average purchase prices can be obtained.
- Average house costing data in Greater London to get an average price estimate that we are looking for
- Geopy to obtain geographic coordinates for visualization
- Foursquare API to obtain venues near each of the localities
- Folium map to mark and observe the neighbourhoods

DATA PREPARATION :

Required packages such as pandas , numpy , geopy , folium , foursquare were imported.
Step 1 : Determining average house pricing in great London. Data was webscraped and converted to a data frame. Average price was calculated.

	PostCode	Area	Average House price	Average deposit required	Average household income required
0	Postcode	Area	Average house price	Average deposit required	Average household income required
1	DA18	Erith	£173,609	£43,177	£32,608
2	DA8	Erith	£242,668	£60,352	£45,579
3	DA17	Belvedere	£243,193	£60,482	£45,678
4	IG11	Barking	£245,019	£60,936	£46,021
5	RM6	Romford	£249,221	£61,981	£46,810
6	RM5	Romford	£261,202	£64,961	£49,060
7	UB1	Southall	£261,245	£64,972	£49,068
8	IG3	Ilford	£262,274	£65,228	£49,262
9	EN3	Enfield	£263,694	£65,581	£49,528

From the above table it can be observed that the average price of some of the well known localities in greater London , one of the most expensive of places, is £2,202,125. The amount was used as a base estimate to determine a suitable pricing range chosen to be between £2,000,000 and £3,000,000.

Data Pre-processing:

Step 2:

- The data used for the project was obtained from the UK real estate website. However , in order to analyse , redundant and insignificant data was removed. It involved :
- Importing data and renaming /formatting columns
- Dropping data / columns that would not be required for further analysis
- Rearranging data according to the date or year in order to work with recent information.
- Filtering out LONDON specific information by grouping data based on streets that contained the word LONDON.
-
- Determining average House pricing in each street and filtering out those in the range of £2,000,000 and £3,000,000.
- Obtaining the geographical coordinates of these localities using geopy.
- Obtaining a general visualization of the streets on the map of London using Folium.

	Locality	Avg_Price	CityCoords	combined
4	ABBEY GARDENS	2.767950e+06	(53.81877655, -1.60767235427371, 0.0)	(53.81877655, -1.60767235427371)
13	ABBEY TRADING POINT	3.000000e+06	(49.18179205, -2.07997384418976, 0.0)	(49.18179205, -2.07997384418976)
51	ABINGDON VILLAS	2.816000e+06	(51.4979279, -0.1948568, 0.0)	(51.4979279, -0.1948568)
118	ADMIRAL SQUARE	2.700000e+06	(51.4762442, -0.1801569, 0.0)	(51.4762442, -0.1801569)
138	AINGER ROAD	2.787500e+06	(51.5411992, -0.1588228, 0.0)	(51.5411992, -0.1588228)
146	AIREDALE AVENUE	2.022500e+06	(53.8289048, -1.8310423, 0.0)	(53.8289048, -1.8310423)
196	ALBION SQUARE	2.450000e+06	(-41.27375755, 173.289393239104, 0.0)	(-41.27375755, 173.289393239104)
197	ALBION STREET	2.096667e+06	(42.9863466, -81.261524, 0.0)	(42.9863466, -81.261524)
391	ANHALT ROAD	2.435000e+06	(51.4803265, -0.1667607, 0.0)	(51.4803265, -0.1667607)
406	ANSDELL TERRACE	2.250000e+06	(51.4998899, -0.1891027, 0.0)	(51.4998899, -0.1891027)
413	ANTHONY WAY	2.000000e+06	(42.7303787, -73.8993373, 0.0)	(42.7303787, -73.8993373)
422	APPLEGARTH ROAD	2.400000e+06	(53.7486539, -0.3266704, 0.0)	(53.7486539, -0.3266704)
484	ARKWRIGHT ROAD	2.964927e+06	(51.5509322, -0.1833781, 0.0)	(51.5509322, -0.1833781)
554	ASHCHURCH PARK VILLAS	2.150000e+06	(51.5000507, -0.2421733, 0.0)	(51.5000507, -0.2421733)
640	AUBREY WALK	2.950000e+06	(51.5068003, -0.1996792, 0.0)	(51.5068003, -0.1996792)
673	AVENUE ROAD	2.143471e+06	(51.4067969, -0.049519, 0.0)	(51.4067969, -0.049519)
715	BACK LANE	2.850000e+06	(51.9548168, -1.9710412, 0.0)	(51.9548168, -1.9710412)

Streets in LONDON with specified average house pricing within range and their respective coordinates.

METHODOLOGY:

After the required data to work was obtained with the use of tools like FOURSQUARE API the most common venues in the vicinity were determined.

The methodology section aimed at :

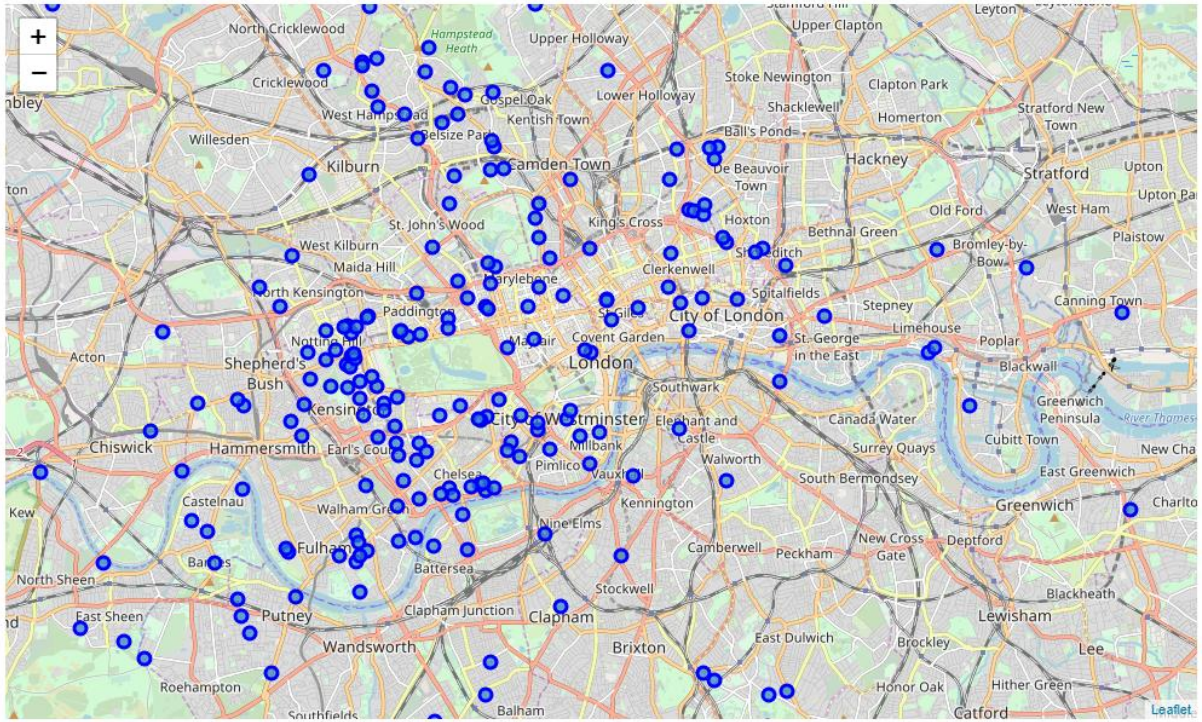
- Obtaining the most common venues in the vicinity of each street in order to find those with relevant amenities using FOURSQUARE.
- Explore the neighbourhoods to see if hospitals , schools and gyms and other facilities of importance are within a 2km radius in order to combine pricing with amenities.
- Exploratory analysis - Statistical comparison of amenities using box plots
- Clustering through a machine learning algorithm called K-means clustering to segment the streets into similar groups that can be prioritized for recommendation.
- Plotting the clusters to obtain a visualizations to draw inferences.

	Street	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	ABBEY GARDENS	Pub	Historic Site	Deli / Bodega	Outlet Store	Coffee Shop	History Museum	Discount Store	Supermarket	Pet Store	Optical Shop
1	ABBEY TRADING POINT	Supermarket	Hotel	Zoo Exhibit	Filipino Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Fabric Shop	Factory
2	ABINGDON VILLAS	Café	Italian Restaurant	Hotel	Bakery	Pub	Modern European Restaurant	Sushi Restaurant	Garden	Cupcake Shop	Burger Joint
3	ADMIRAL SQUARE	Pub	Hotel	Italian Restaurant	Restaurant	Spa	Café	Furniture / Home Store	Modern European Restaurant	Gastropub	Middle Eastern Restaurant
4	AINGER ROAD	Pub	Café	Coffee Shop	Italian Restaurant	Asian Restaurant	Furniture / Home Store	Scenic Lookout	Park	Market	Greek Restaurant

Obtaining frequency of occurrence of all venues for each street and finally consolidating them into the
10 most common venues for each street.

STREETS OF LONDON : FOLIUM

Out[30]:



K-MEANS CLUSTERING :

The areas were segmented into five clusters and visualized in order to draw conclusions on which cluster or localities would be preferred over the other.

	Locality	Avg_Price	latitude	longitude	Cluster Labels
4	ABBEY GARDENS	2767950.0	53.818777	-1.607672	1
13	ABBEY TRADING POINT	3000000.0	49.181792	-2.079974	3
51	ABINGDON VILLAS	2816000.0	51.497928	-0.194857	3
118	ADMIRAL SQUARE	2700000.0	51.476244	-0.180157	1
138	AINGER ROAD	2787500.0	51.541199	-0.158823	3

Adding cluster labels

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

1]:

	Avg_Price	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
146	2.022500e+06	Construction & Landscaping	Harbor / Marina	Other Repair Shop	Fish & Chips Shop	Zoo Exhibit	Fast Food Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor
197	2.096667e+06	Baseball Field	Vietnamese Restaurant	Breakfast Spot	Café	Skating Rink	Mexican Restaurant	Park	Baseball Stadium	Salon / Barbershop	Rock Club
413	2.000000e+06	Farm	Zoo Exhibit	Filipino Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Fabric Shop	Factory	Falafel Restaurant
554	2.150000e+06	Grocery Store	Pub	Indian Restaurant	Park	Bakery	Coffee Shop	Mediterranean Restaurant	Fish & Chips Shop	Bar	Middle Eastern Restaurant
673	2.143471e+06	Tapas Restaurant	Grocery Store	Tram Station	Park	Zoo Exhibit	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Fabric Shop

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

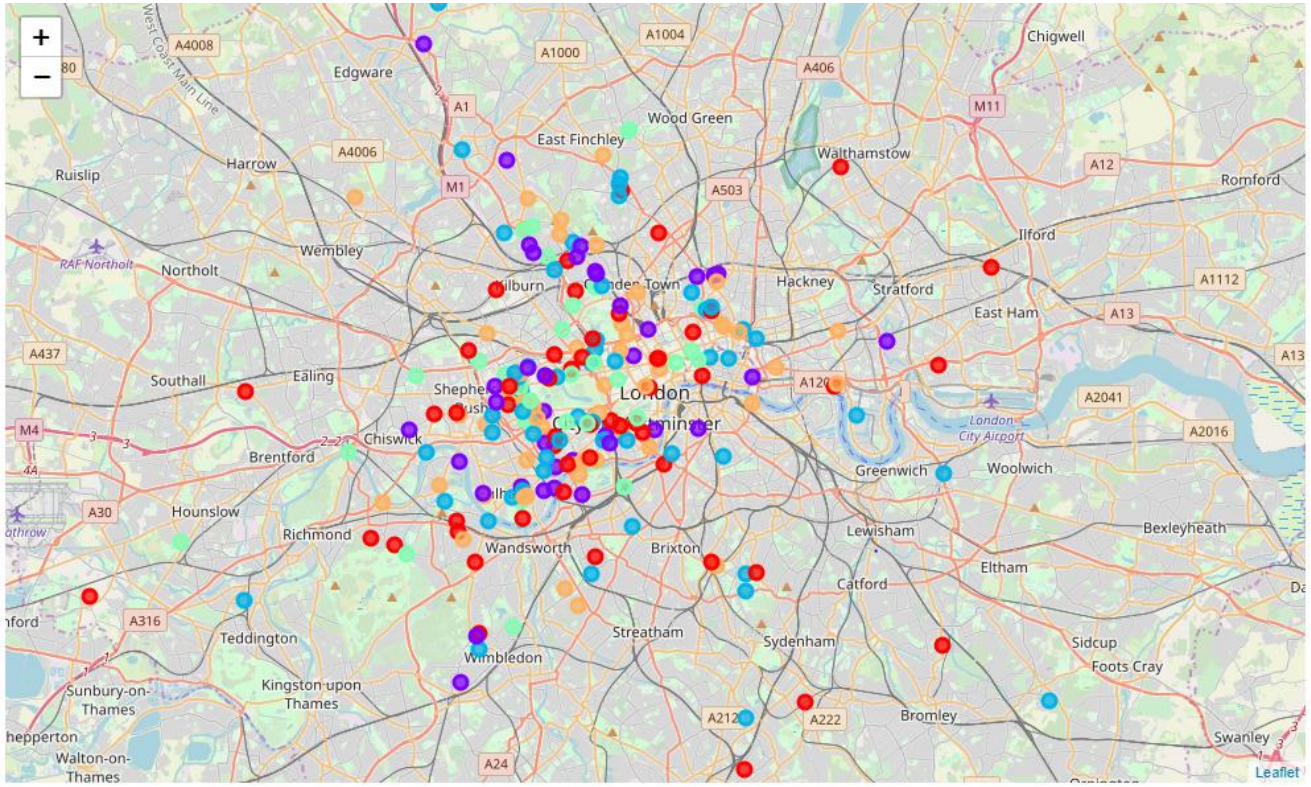
2]:

	Avg_Price	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
4	2767950.0	Pub	Historic Site	Deli / Bodega	Outlet Store	Coffee Shop	History Museum	Discount Store	Supermarket	Pet Store	Optical Shop
118	2700000.0	Pub	Hotel	Italian Restaurant	Restaurant	Spa	Café	Furniture / Home Store	Modern European Restaurant	Gastropub	Middle Eastern Restaurant
837	2745000.0	Hotel	Gym / Fitness Center	Diner	Entertainment Service	Ethiopian Restaurant	Event Space	Exhibit	Eye Doctor	Fabric Shop	Factory
1545	2650000.0	Italian Restaurant	Burger Joint	Pizza Place	Café	Bakery	Japanese Restaurant	Shopping Plaza	Restaurant	Peruvian Restaurant	Juice Bar
2106	2655000.0	Gym	Indian Restaurant	Train Station	Fast Food Restaurant	Gym / Fitness Center	Platform	Falafel Restaurant	Electronics Store	English Restaurant	Entertainment Service

Cluster 1 and 2

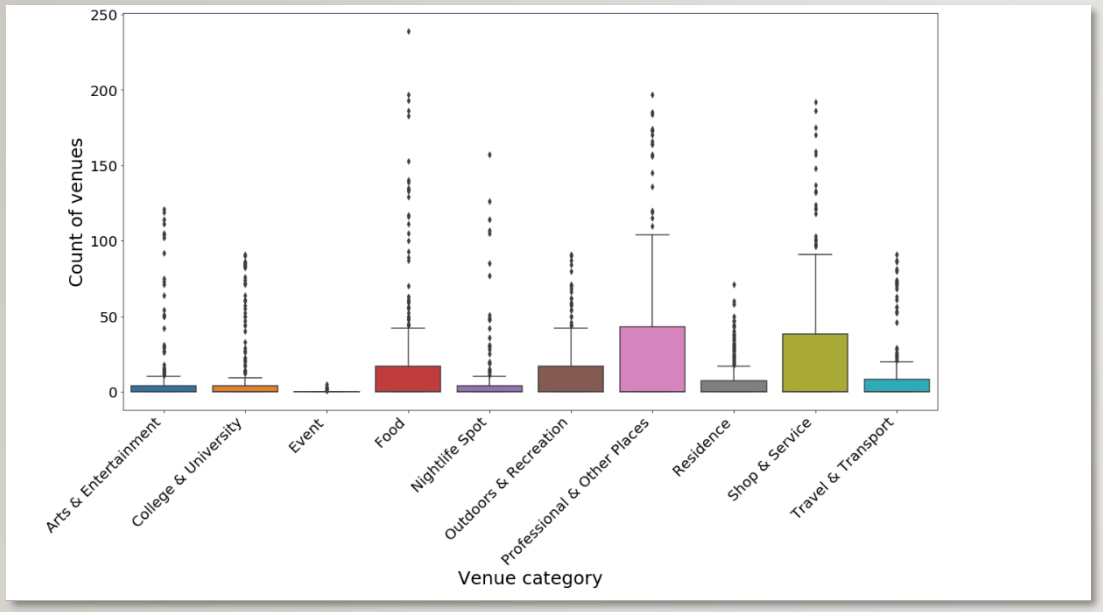
CLUSTER

VISUALIZATION :

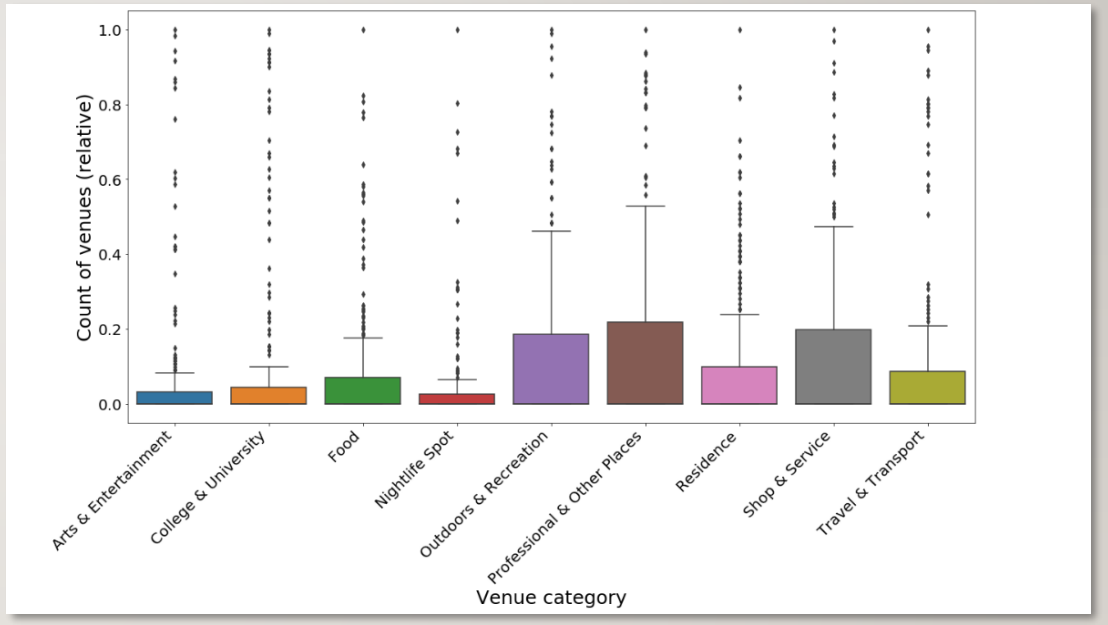


EXPLORATORY ANALYSIS:

In order to understand the number of occurrences of each category and to use it for comparison , exploratory analysis was conducted using box – plots both normal and transformed.

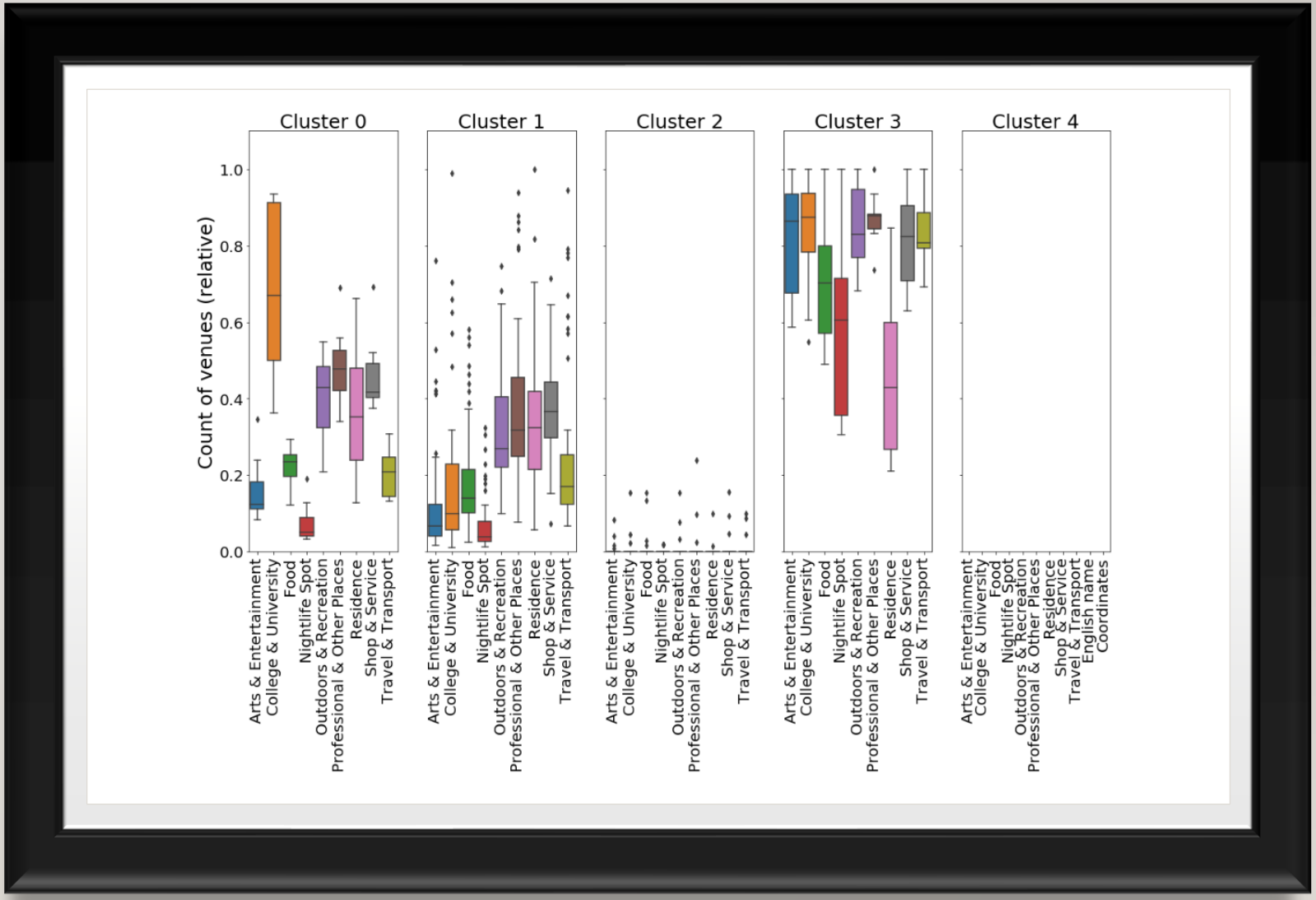


Venue categories based on count of venues



Transformation of venue categories

COMPARISON OF CLUSTERS BASED ON NUMBER OF OCCURENCES OF EACH CATEGORY



As it can be clearly observed clusters 0,1 and 3 have most of the amenities. Cluster 0 has a larger number of educational institutions in the vicinity , a residential place with moderately appropriate facilities. Cluster 1 has quite a large number of outliers and moderately placed amenities. Cluster 3 is observed to have higher counts of all categories though their mean and quartile ranges are smaller. Hence , the localities have been clustered , priced and observed and the following results have been drawn.

Results :

As per the scope of this project the following results were observed :

Average house considered for the project is between £2,000,000 - £3,000,000 which was determined based on mean price of £2,202,125

From the pricing statistics , it can be observed that 50% of the localities have a mean price of 2,400,000 pounds and a major 75% make up for a price value of 2,650,000 pounds which is a good price representative range.

The streets of London were segmented into five clusters.

- Cluster 0 – approximate price range (2,000,000-2,150,000)
- Cluster 1 – approximate price range (2,650,000-2,800,000)
- Cluster 2 – approximate price range (2,150,000-2,350,000)
- Cluster 3 – approximate price range (2,810,000-3,000,000)
- Cluster 4 – approximate price range (2,350,000-2,500,000)

All clusters had ample restaurants and recreational facilities in the vicinity.

The clusters were viewed on the map.

With respect to proximity of prominent amenities in each location , from a general perspective it can be observed that cluster 0 ,cluster 1 and cluster 3 have the most prominent number of venue categories.

cluster 0 has the least pricing , more educational institutions , preferable transportation and professional locations in the vicinity. cluster 1 has a large number of outliers with a comparatively small range of facilities. However , it is still a preferred location for residence and has a comparatively higher pricing range. cluster 3 has the highest pricing range in the average scale , however it is still affordable and has enumerable benefits and facilities with the highest occurrence of amenities - educational institutions, food , nightlife, transportation and highest number of residencies as well.

The clusters were compared in order to draw inferences on the best cluster using exploratory analysis - box plots

DISCUSSION:

Despite the economic crisis , from the data acquired , it can be observed that there is good affordable housing within a range of about £2,000,000 - £3,000,000 in London with almost 414 streets predominantly falling in the range. In order to increase the stake of preferences the venues in the vicinity of each street were viewed and it can be observed that there are plenty of restaurants and eateries in the vicinity of each street. When it comes to schools, local stores , work place, transportation etc , which are extremely important to family residents it can be observed that cluster 3 has the most number and occurrence of various amenities followed by cluster 0.In terms of pricing however cluster 0 is cheaper than cluster 3. This helps with deriving conclusions on the order of precedence of clusters as follows: In terms of pricing ,order of precedence would be :

cluster 0
cluster 1
cluster 3

In terms of pricing and amenities combined :

cluster 3
cluster 0
cluster 1

It depends on the preference of the customer , However places like Croydon, Colindale, Bromley, Hornchurch, Woolwich, Leytonstone, Brentford and Crystal Palace belonging to these above mentioned clusters can definitely be recommended.

CONCLUSION:

The scope of the project was to address the question of preferable housing suggestions to real estate agencies and real estate clientele interested in purchasing their own independent homes. In lieu with the expectation , localities with affordable residences were clustered , visualized and the ones with best amenities were deduced in order to render valuable suggestions to the target audience. Through the course of the project it was observed that comparatively cheaper and good quality homes are available in London and just that one needs to be on the lookout for it. The most dominant cluster was cluster 3 followed by cluster 0 while taking both amenities and pricing into consideration and some of the most predominant localities were found to be Croydon, Colindale, Bromley, Hornchurch, Woolwich, Leytonstone, Brentford and Crystal Palace.

Thus , I believe the project has addressed 60% of the housing issue. However there is a broader scope for further in depth analysis to suggest housing not just based on vicinity and pricing but also on the type of housing ,transportation, design and alternatives to amazing and yet affordable architectural constructs such as rooftop housing , bridged housing etc with more advanced exploratory analysis and machine learning techniques..

THANK YOU