Introduction

London has one of the most sought-after real-estate markets. It's cosmopolitan and vibrant lifestyle makes it also when the of the world's top destination for tourists.

The aim of this project was to segment London using the most common venues in each area and compare it with the average home prices in the same area to understand if there is a relationship between the type and frequency of venues and home prices.

The analysis would potentially help real-estate investors understand if the relationship between some venues will lead to an increase in the prices of properties and therefore return on their investments.

Data Used

Three datasets are used:
1. Venues and places data from Foursquare to identify the top common places in each area
2. London's average home prices per area from London DataStore, which will be combined with the common places to identify an interesting insight.
3. London's areas coordinate from London DataStore, which will be used to get the common places in from Foursquare

Methodology

1. Explored the areas covered in the datasets to ensure consistency in both datasets, specifically with regards to the names of areas.
2. Applied a simple histogram plot to explore the distribution of the prices of the data. As shown in the plot below, home prices are skewed towards 250,000 to 750,000 range, with fewer areas in the 1,000,000 + range
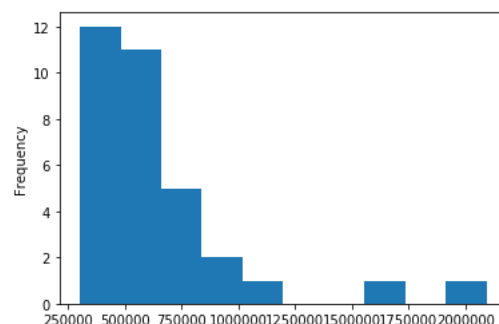
3. Applied some descriptive analysis to show the distribution of the data as shown below:

```
count    3.300000e+01
mean     6.654116e+05
std      3.725934e+05
min      3.015180e+05
25%      4.599220e+05
50%      5.739380e+05
75%      6.839870e+05
max      2.092485e+06
```

4. Based on the previous two steps, the areas were categorised according to the average prices in each area. The purpose of this step was to see if insights can be identified on a category level (See below table).
   a. Price Cat 1: Areas above the 95% percentile
   b. Price Cat 2: Areas between 75% and 95% percentile
   c. Price Cat 3: Areas between 50% and 75% percentile
   d. Price Cat 4: Areas between 25% and 50% percentile
   e. Price Cat 5: Areas between 0% and 25% percentile

Area	Latitude	Longitude	Value	Price Cat
Kensington and Chelsea	51.498517	-0.187363	2092485	1
Westminster	51.513203	-0.148677	1718124	1
Camden	51.534370	-0.143292	1099876	2
Hammersmith and Fulham	51.492079	-0.216469	972231	2
City of London	51.514287	-0.091435	950760	2
Richmond upon Thames	51.449029	-0.316954	819044	2
Wandsworth	51.455802	-0.176793	818443	2
Islington	51.538980	-0.108733	778290	2
Haringey	51.589738	-0.105810	683987	3
Barnet	51.605532	-0.207715	667593	3
Southwark	51.483583	-0.078566	641210	3

5. Data was fetched from Foursquare for the top venues in each area and plotted on a bar chart to better understand the distribution and which areas had the most venues.

	Neighborhood	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	Barking and Dagenham	Park	Construction & Landscaping	Home Service	Yoga Studio	Falafel Restaurant	Fried Chicken Joint	French Restaurant	Food Court	Fish & Chips Shop	Field
1	Barnet	Gym / Fitness Center	Supermarket	Gas Station	Metro Station	Bakery	Grocery Store	Farm	Fried Chicken Joint	French Restaurant	Food Court
2	Bexley	Pub	Recreation Center	Bar	Pizza Place	Chinese Restaurant	Fast Food Restaurant	Supermarket	Bowling Alley	Thrift / Vintage Store	Mexican Restaurant
3	Brent	Scandinavian Restaurant	Furniture / Home Store	Supermarket	Portuguese Restaurant	Sandwich Place	Fast Food Restaurant	Event Space	French Restaurant	Food Court	Fish & Chips Shop
4	Bromley	Gym / Fitness Center	Park	Soccer Stadium	Fruit & Vegetable Store	French Restaurant	Food Court	Fish & Chips Shop	Field	Fast Food Restaurant	Farmers Market

a. As seen in the chart below, Westminster and City of London are the only two areas that have delivered the maximum, 100 venues.
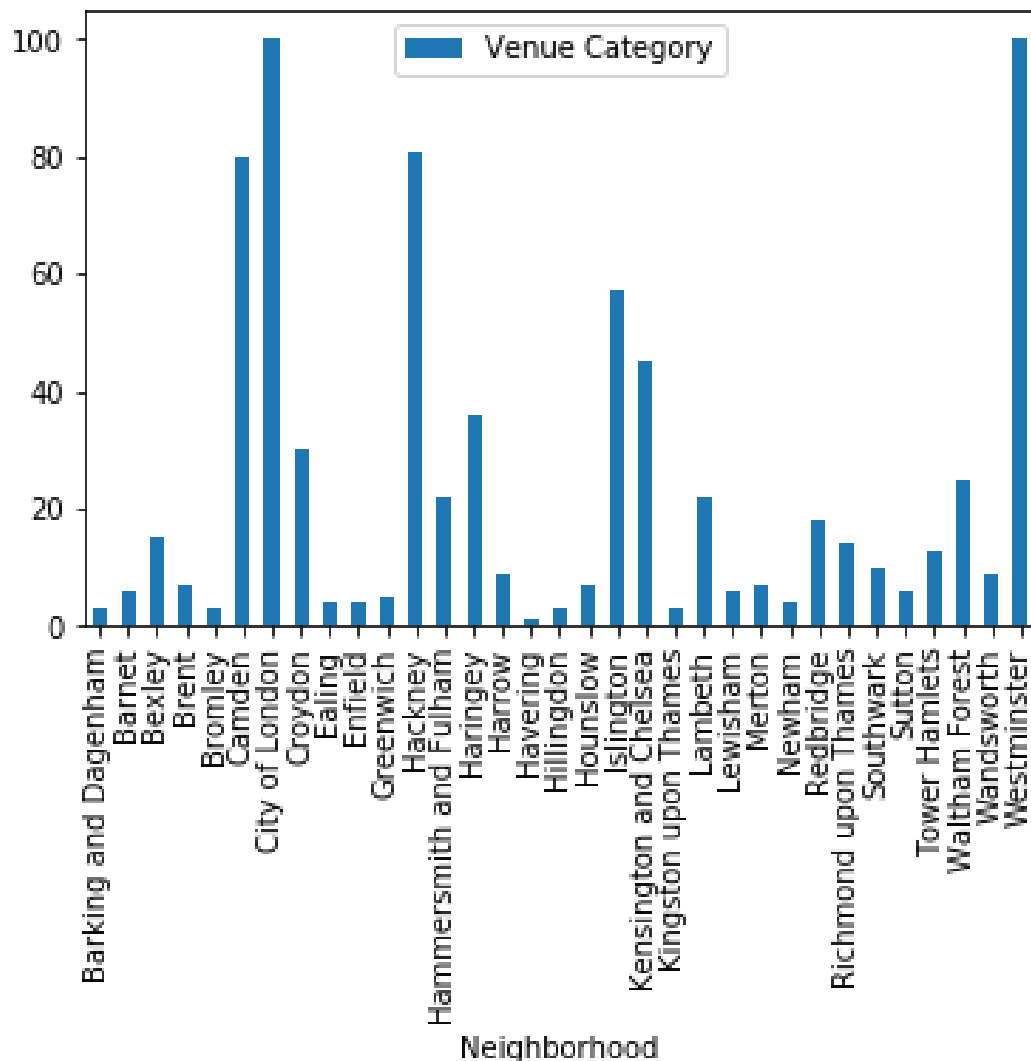

6. K-means was applied to data after one-hot encoding int. Multiple values for K were tested with little impact to the results, until settling on 6.

Results

K-means model delivered the following labels to the 33 areas based on their top 10 venues

```
array([0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 2, 1, 1, 1, 1, 4, 3, 1, 1, 1, 1, 1
,
       2, 5, 2, 1, 2, 2, 2, 0, 1, 2, 1], dtype=int32)
```

The clusters were then visualised in on the map below for further inspection.
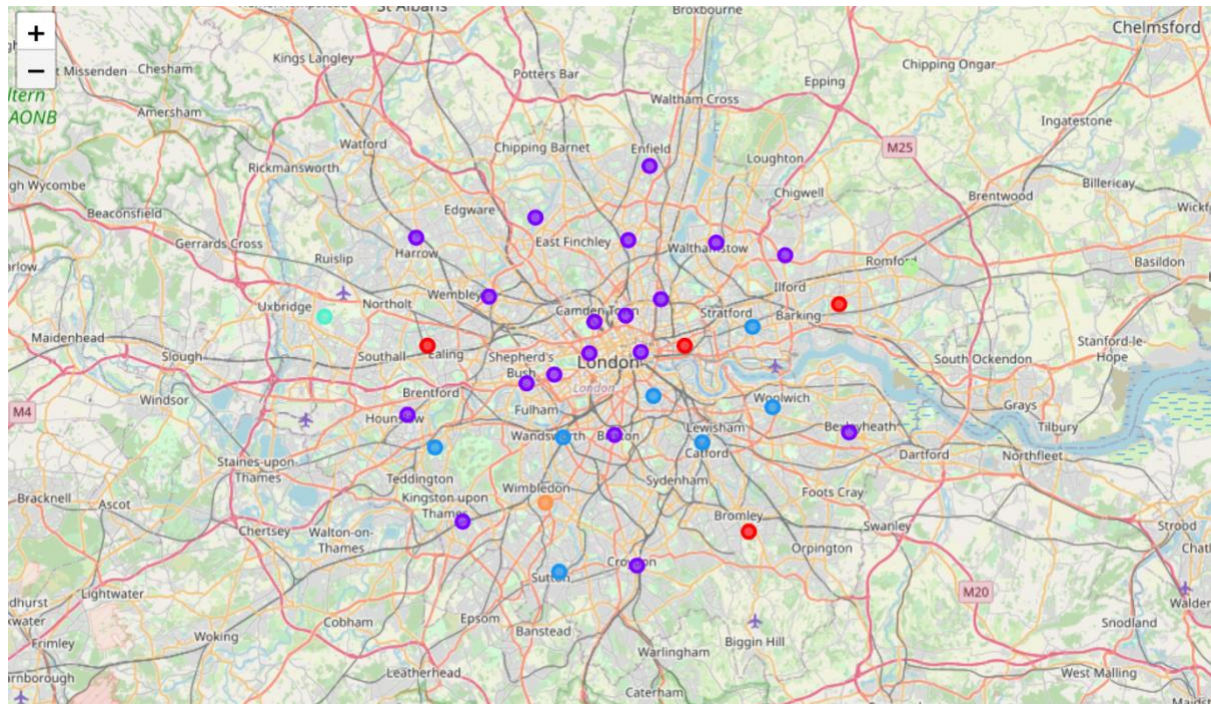
Discussion

Clusters
Figure below shows Cluster 0, which I can only assume that it was clustered on the basis of having Park as the number 1 or 2 common place.

	Area	Latitude	Longitude	Value	Price Cat	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 Cor
0	Barking and Dagenham	51.546468	0.124998	301518	5	0	Park	Construction & Landscaping	Home Service	Yoga Studio	Falafel Restaurant	Fried Chicken Joint	French Restaurant	Food Court	
4	Bromley	51.391800	0.026393	502623	4	0	Gym / Fitness Center	Park	Soccer Stadium	Fruit & Vegetable Store	French Restaurant	Food Court	Fish & Chips Shop	Field	Fas Rest
8	Ealing	51.518032	-0.324967	578110	3	0	Train Station	Park	Tennis Court	Shop & Service	Event Space	French Restaurant	Food Court	Fish & Chips Shop	
29	Tower Hamlets	51.517782	-0.043682	554009	4	0	Park	Fried Chicken Joint	Turkish Restaurant	Sandwich Place	Grocery Store	Farm	Bakery	Thai Restaurant	Rest

The model clustered most of the areas as Cluster 1 as shown below. It's unclear from the results the basis on which the model did this segmentation. Perhaps it was the number of coffee shops, or just poor application of the model/ an unapparent insight.

	Area	Latitude	Longitude	Value	Price Cat	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
1	Barnet	51.605532	-0.207715	667593	3	1	Gym / Fitness Center	Supermarket	Gas Station	Metro Station	Bakery	Grocery Store	Farm
2	Bexley	51.459171	0.136321	357779	5	1	Pub	Recreation Center	Bar	Pizza Place	Chinese Restaurant	Fast Food Restaurant	Supermarket
3	Brent	51.551800	-0.257501	578705	3	1	Scandinavian Restaurant	Furniture / Home Store	Supermarket	Portuguese Restaurant	Sandwich Place	Fast Food Restaurant	Event Space
5	Camden	51.534370	-0.143292	1099876	2	1	Coffee Shop	Greek Restaurant	Japanese Restaurant	Pizza Place	Park	Italian Restaurant	Pub
6	City of London	51.514287	-0.091435	950760	2	1	Coffee Shop	Gym / Fitness Center	Italian Restaurant	Wine Bar	Clothing Store	Seafood Restaurant	Restaurant
7	Croydon	51.368197	-0.096495	399645	5	1	Pub	Mediterranean Restaurant	Italian Restaurant	Indian Restaurant	Bar	Seafood Restaurant	Coffee Shop
9	Enfield	51.639907	-0.082701	463806	4	1	Indian Restaurant	Tennis Court	Garden	Pizza Place	Discount Store	Farmers Market	Dim Sum Restaurant
11	Hackney	51.549614	-0.069847	614955	3	1	Turkish Restaurant	Cocktail Bar	Pub	Café	Coffee Shop	Restaurant	Bar
12	Hammersmith and Fulham	51.492079	-0.216469	972231	2	1	Café	Hotel	Pub	Cocktail Bar	Coffee Shop	Gym	Grocery Store
13	Haringey	51.589738	-0.105810	683987	3	1	Café	Women's Store	Clothing Store	Fast Food Restaurant	Coffee Shop	Pharmacy	Sandwich Place
14	Harrow	51.591820	-0.338085	527206	4	1	Indian Restaurant	Afghan Restaurant	Sandwich Place	Fast Food Restaurant	Platform	Bakery	Coffee Shop
17	Hounslow	51.471441	-0.347016	507876	4	1	Bus Stop	Indian Restaurant	Portuguese Restaurant	Coffee Shop	Pizza Place	Pub	Dive Bar
18	Islington	51.538980	-0.108733	778290	2	1	Pub	French Restaurant	Mediterranean Restaurant	Gastropub	Bakery	Furniture / Home Store	Park

The figure below shows the results for cluster 2. The majority of the areas in this cluster have Café as their most common venue.

	Area	Latitude	Longitude	Value	Price Cat	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	
10	Greenwich	51.476668	0.051810	462820	4	2	Bus Stop	Stationery Store	Soccer Field	Fruit & Vegetable Store	Fried Chicken Joint	French Restaurant	Food Court	Fish & Chips Shop	
22	Lewisham	51.452592	-0.024830	475142	4	2	Indian Restaurant	Pub	Park	Café	Athletics & Sports	Tree	Farm	Fried Chicken Joint	R
24	Newham	51.531290	0.030812	409477	5	2	Café	Market	Pub	Yoga Studio	Fried Chicken Joint	French Restaurant	Food Court	Fish & Chips Shop	
26	Richmond upon Thames	51.449029	-0.316954	819044	2	2	Café	Park	Gastropub	Boat or Ferry	Garden	Playground	Historic Site	Beer Garden	
27	Southwark	51.483583	-0.078566	641210	3	2	Café	Pub	Building	Garden	Park	Convenience Store	Lake	Bus Stop	Pho
28	Sutton	51.364229	-0.181871	402121	5	2	Pub	Bar	Park	Café	Grocery Store	Farm	Fried Chicken Joint	French Restaurant	Fc
31	Wandsworth	51.455802	-0.176793	818443	2	2	Café	Bus Stop	Grocery Store	Park	Pub	Garden Center	Gastropub	Pizza Place	Fc

Clusters 3, 4, and 5 shows below do not show any distinctive information to why the model had clustered them in this manner.

	Area	Latitude	Longitude	Value	Price Cat	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 Mc Comm Ven
16	Hillingdon	51.537793	-0.438383	452272	5	3	Sporting Goods Shop	Construction & Landscaping	Playground	Yoga Studio	Falafel Restaurant	Fried Chicken Joint	French Restaurant	Food Court	Fish Ch Sh

```
ldn_clus[ldn_clus['Cluster Labels'] ==4]
```

	Area	Latitude	Longitude	Value	Price Cat	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
15	Havering	51.570939	0.203535	387535	5	4	Park	Yoga Studio	Falafel Restaurant	Fried Chicken Joint	French Restaurant	Food Court	Fish & Chips Shop	Field	Fast Food Restaurant

```
ldn_clus[ldn_clus['Cluster Labels'] ==5]
```

	Area	Latitude	Longitude	Value	Price Cat	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 Cor
23	Merton	51.411093	-0.196325	638519	3	5	Tram Station	Pizza Place	Pub	Park	Event Space	French Restaurant	Food Court	Fish & Chips Shop	Field	Fas Rest

Relationship with Average Prices

The model did not produce any beneficial results for our analysis with average home prices. It's quite likely the modelling approach was inaccurate or unsuitable for the problem at hand.

Conclusion

The approach to modelling the problem was not suitable. A different approach is recommended as next steps. Perhaps one that looks into correlating the number of venues within each category with the increase in prices. Perhaps a combination of time series analysis that reflects price increase with changes in frequency of types of venues.