Capstone Project - The Battle of Neighborhoods

Lina Grusliene

Introduction and data overview

New Real Estate Agency in Central London

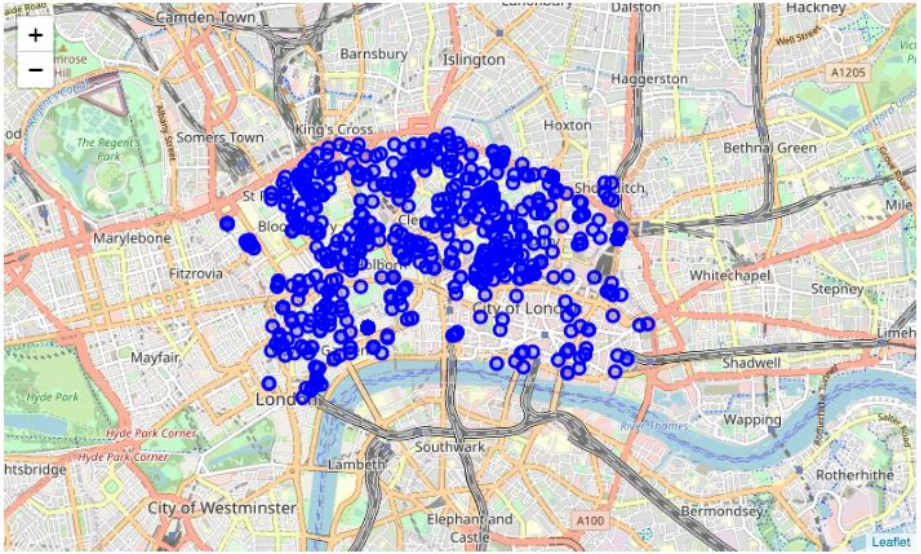
- London is a big and vibrant city, where lots of people dream to live in.
- Property in London could be very expensive, but prices vary as well as customers' wealth.
- The real estate market moves fast and is very competitive.
- The aim of this project is to find the best location in Central London to open a new Real Estate Office targeting luxury class property sellers and buyers.

Data acquisition and cleaning

1. HM Land Registry Price Paid Data containing property sales data in England and Wales in 2020, including price and postcode (neighbourhood)
2. A list of UK postcodes and their geographical coordinates latitude and longitude from FreeMapTools.
3. Real Estate Offices near sold properties from FourSquare.

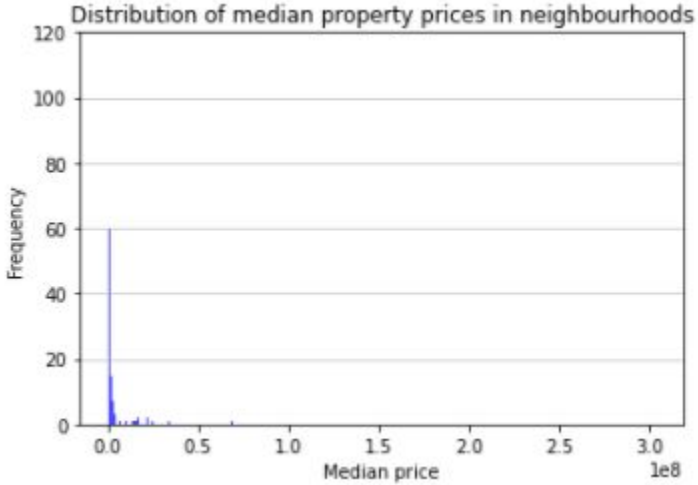
We have 1037 property sales in Central London with geographical coordinates and a number of real estate offices nearby.

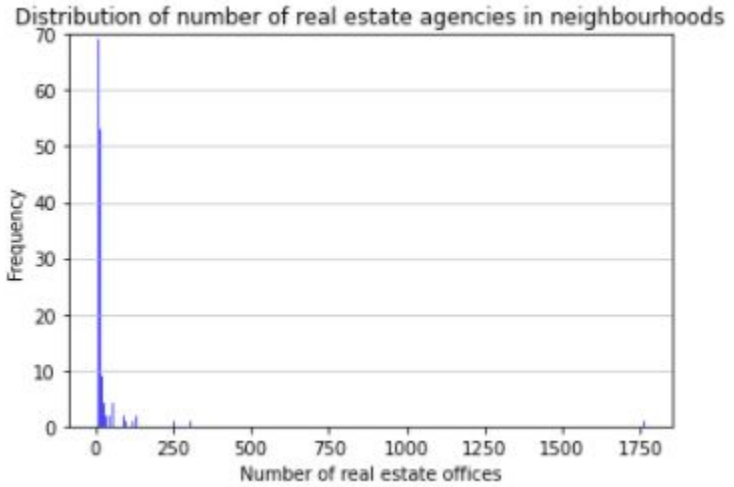
Property sales in Central London



Methodology

Distribution of real estate offices and property prices - skewed data



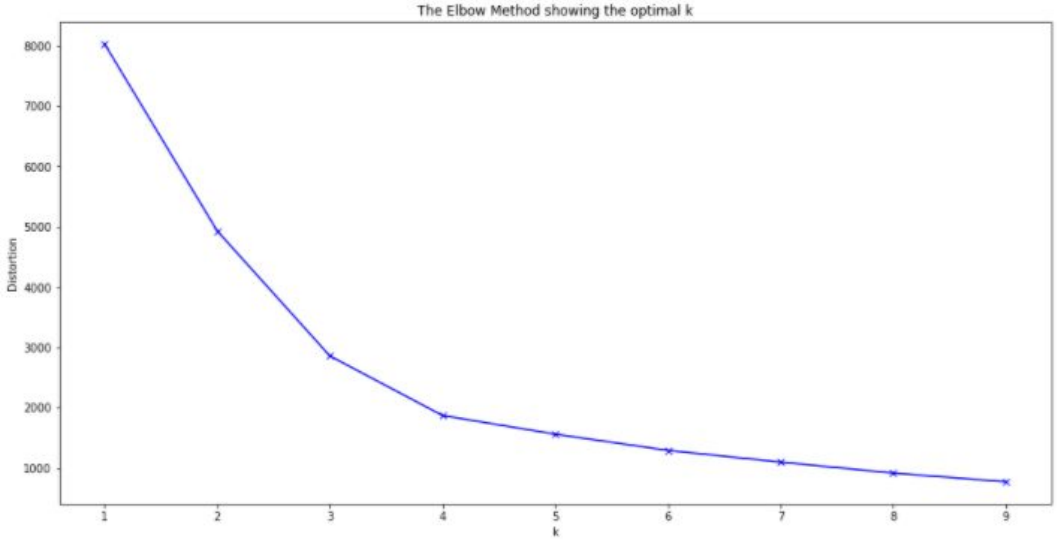


Using quantiles to deal with skewed data

	Neighbourhood	Borough	price	latitude	longitude	price quantiles range	price quantiles	Venue	agencies quantiles range	agencies quantiles
0	EC1A 2DJ	CITY OF LONDON	14900000.0	51.516995	-0.102727	(7650000.0, 303470000.0]	9	16	(15.0, 17.0]	6
1	EC1A 4HU	CITY OF LONDON	746000.0	51.518236	-0.097184	(701112.0, 830000.0]	4	18	(17.0, 20.0]	7
2	EC1A 7AB	CITY OF LONDON	649950.0	51.519512	-0.098774	(575000.0, 701112.0]	3	17	(15.0, 17.0]	6
3	EC1A 7BB	CITY OF LONDON	1465000.0	51.517906	-0.098956	(1320000.0, 1990000.0]	7	36	(20.0, 36.0]	8
4	EC1A 7BD	CITY OF LONDON	1500000.0	51.518864	-0.098559	(1320000.0, 1990000.0]	7	17	(15.0, 17.0]	6
...	...	...	...	...	...	...	...	...	...	...
476	WC2R 1HA	CITY OF WESTMINSTER	1569500.0	51.512084	-0.118208	(1320000.0, 1990000.0]	7	18	(17.0, 20.0]	7
477	WC2R 1JA	CITY OF WESTMINSTER	1100000.0	51.511471	-0.118233	(985000.0, 1320000.0]	6	9	(7.0, 9.0]	1
478	WC2R 3DX	CITY OF WESTMINSTER	1195000.0	51.511505	-0.113562	(985000.0, 1320000.0]	6	10	(9.0, 11.0]	2
479	WC2R 3JF	CITY OF WESTMINSTER	1500000.0	51.512940	-0.112840	(1320000.0, 1990000.0]	7	10	(9.0, 11.0]	2
480	WC2R 3JJ	CITY OF WESTMINSTER	1500000.0	51.513140	-0.112840	(1320000.0, 1990000.0]	7	10	(9.0, 11.0]	2

Both price and number of offices variables are split into quantiles and used for modelling to avoid outliers having too much impact.

Modelling - KMeans



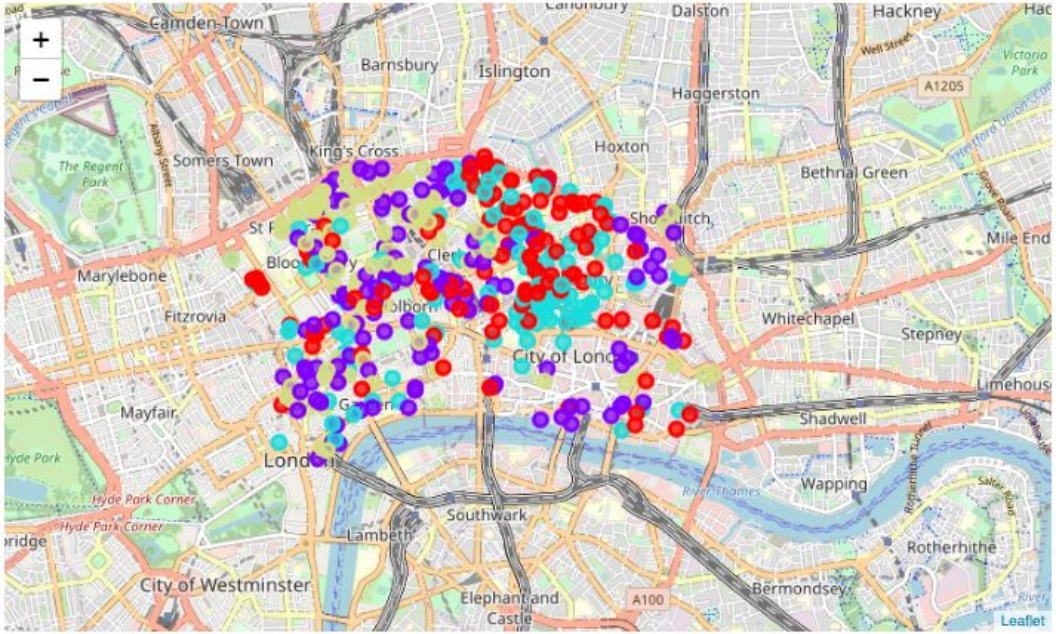
- KMeans Clustering Machine learning algorithm was used to cluster similar neighbourhoods together and then choose the best one for opening a new real estate office.
- Elbow and Silhouette score methods used to find that optimal number of clusters is 4.

KMeans - clustered data

	Cluster Labels	Neighbourhood	Borough	price	latitude	longitude	price quantiles range	price quantiles	Venue	agencies quantiles range	agencies quantiles
0	2	EC1A 2DJ	CITY OF LONDON	14900000.0	51.516995	-0.102727	(7650000.0, 303470000.0]	9	16	(15.0, 17.0]	6
1	2	EC1A 4HU	CITY OF LONDON	746000.0	51.518236	-0.097184	(701112.0, 830000.0]	4	18	(17.0, 20.0]	7
2	0	EC1A 7AB	CITY OF LONDON	649950.0	51.519512	-0.098774	(575000.0, 701112.0]	3	17	(15.0, 17.0]	6
3	2	EC1A 7BB	CITY OF LONDON	1465000.0	51.517906	-0.098956	(1320000.0, 1990000.0]	7	36	(20.0, 36.0]	8
4	2	EC1A 7BD	CITY OF LONDON	1500000.0	51.518864	-0.098559	(1320000.0, 1990000.0]	7	17	(15.0, 17.0]	6

Results, Discussion and Conclusions

Clustered neighbourhoods



Clustered neighbourhoods - descriptions

We have successfully clustered Central London neighbourhoods by property prices and density of real estate agencies:

- Cluster 1: Lower prices, high number of agencies
- **Cluster 2: High prices, low number of agencies**
- Cluster 3: High prices, high number of agencies
- Cluster 4: Low prices, low number of agencies

Discussion and Conclusions

Some neighbourhoods are more popular than others between real estate companies and property prices vary as well. Modelling could be used to find the best areas to invest.

All neighbourhoods were clustered based on property prices and number of real estate offices in close proximity using KMeans algorithm and the best areas to consider for the new office are in Cluster 2.

The area selection could be further narrowed down by considering other factors important to stakeholders such as:

- Rental prices
- Other types of luxury class venues nearby
- Traffic