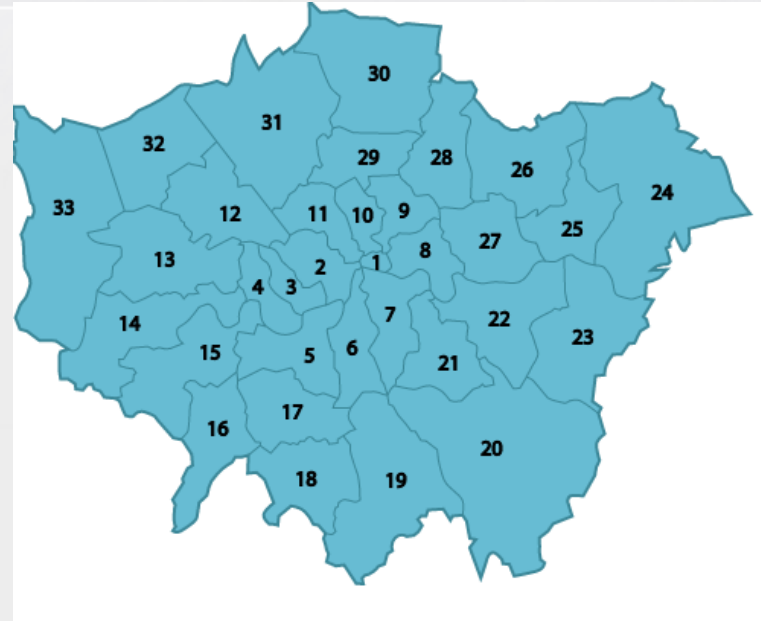


Analysis of the London districts

common venues, average house prices

Introduction

- **Object**
 - London districts (boroughs)
- **Problem**
 - The boroughs differ from each other in a variety of parameters
 - We need identity differences using information about common venues and average property prices.
- **Interest:**
 - If you want to open an office in London you will be interested in information (in a convenient view) about the districts infrastructure (venues) and the average house prices.



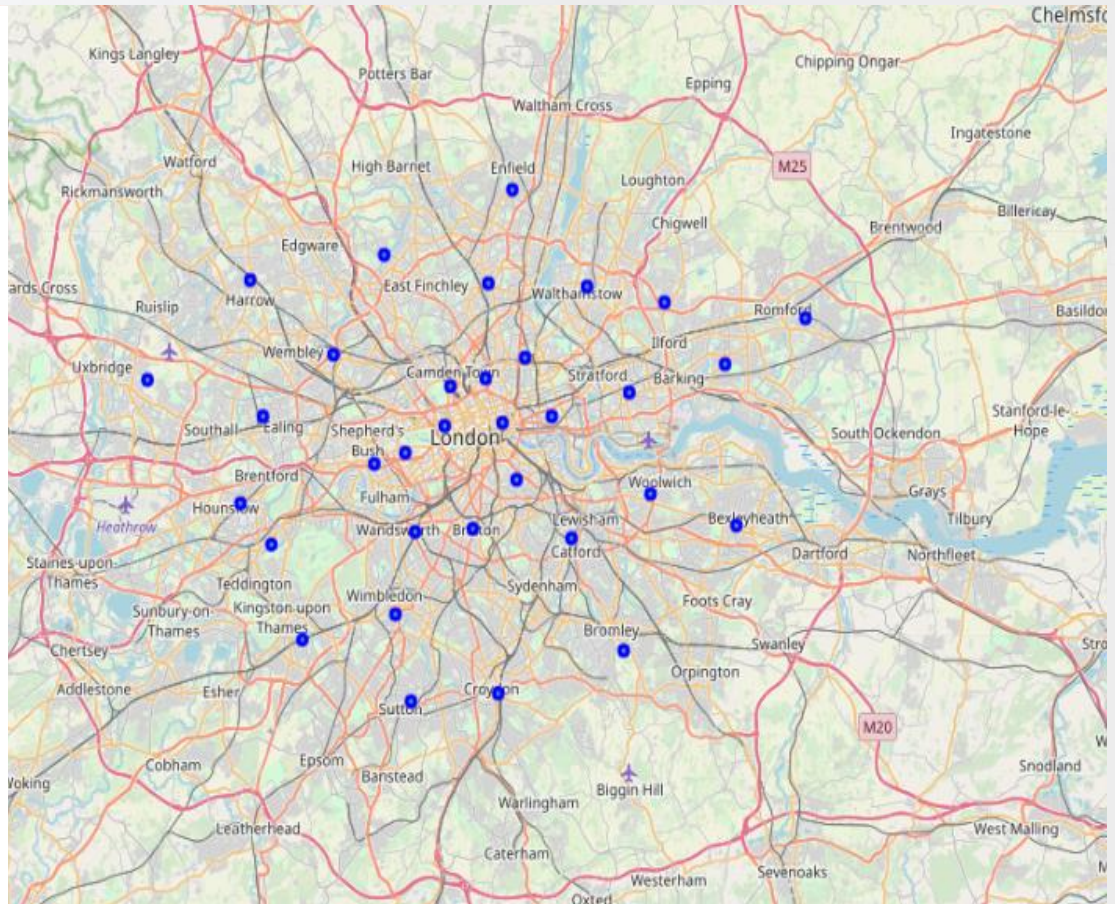
Data

- Postcodes by administrative area
 - coordinates of the all UK areas
 - [AdministrativeAreas.php](#)
- Average house prices
 - information about average house prices of the all UK areas from period 1968-2019.
 - [Average-prices-2019-04.csv](#)
- Foursquare API
 - necessary to get the most common venues
 - <https://api.foursquare.com>
- Country geojson file
 - necessary to create choropleth map with level of average prices
 - [london_boroughs.geojson](#)

Methodology (1 of 4) – master data

- Master data

	Borough	Latitude	Longitude	Average Price
0	Barking and Dagenham	51.546501	0.125051	2.990287e+05
1	Barnet	51.605598	-0.207649	5.167502e+05
2	Bexley	51.459202	0.136265	3.333173e+05
3	Brent	51.551800	-0.257490	4.745404e+05
4	Bromley	51.391800	0.026386	4.302424e+05
5	Camden	51.534302	-0.143282	8.143753e+05
6	City of London	51.514301	-0.091442	6.997945e+05
7	Croydon	51.368198	-0.096448	3.699863e+05
8	Ealing	51.518002	-0.325004	4.597960e+05
9	Enfield	51.639900	-0.082636	3.932527e+05
10	Greenwich	51.476601	0.051824	3.965288e+05
11	Hackney	51.549599	-0.069849	5.418023e+05
12	Hammersmith and Fulham	51.492100	-0.216492	7.187671e+05
13	Haringey	51.589699	-0.105783	5.115430e+05
14	Harrow	51.591801	-0.338094	4.456364e+05
15	Havering	51.570999	0.203626	3.577700e+05
16	Hillingdon	51.537800	-0.438400	4.103858e+05
17	Hounslow	51.471401	-0.347104	3.981704e+05
18	Islington	51.539001	-0.108697	6.137179e+05
19	Kensington and Chelsea	51.498501	-0.187347	1.288685e+06
20	Kingston upon Thames	51.397900	-0.287269	4.678941e+05
21	Lambeth	51.457699	-0.120317	4.886103e+05
22	Lewisham	51.452599	-0.024825	4.066142e+05
23	Merton	51.411098	-0.196242	5.021075e+05
24	Newham	51.531300	0.030927	3.642919e+05
25	Redbridge	51.579300	0.066536	4.039113e+05
26	Richmond upon Thames	51.449001	-0.317008	6.508974e+05
27	Southwark	51.483601	-0.078522	4.750521e+05
28	Sutton	51.364201	-0.181904	3.702000e+05
29	Tower Hamlets	51.517799	-0.043724	4.478946e+05
30	Waltham Forest	51.588001	-0.009436	4.331580e+05
31	Wandsworth	51.455799	-0.176848	5.698327e+05
32	Westminster	51.513199	-0.148674	9.989116e+05



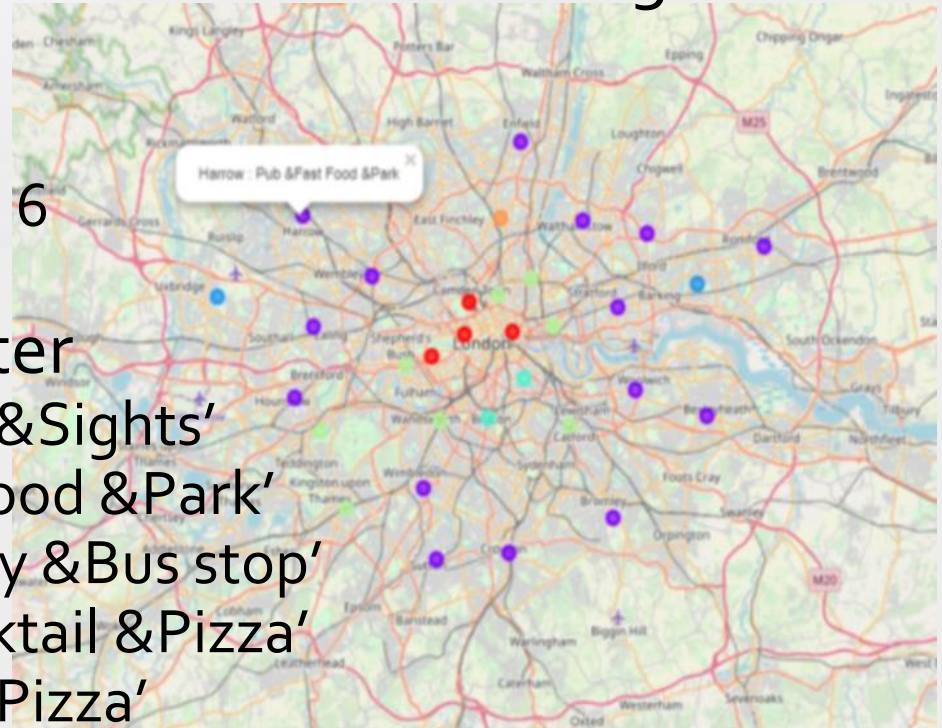
Methodology (2 of 4) – segment boroughs

- Use Foursquare API
- Set for each borough
 - number of venues = 100
 - radius = 2000 meters
- Explore the boroughs and segment them.

	Borough	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	Grocery Store	Pub	Bus Stop	Soccer Field	Park	Supermarket	Recreation Center	Go Kart Track	Café	Movie Theater
1	Barnet	Café	Indian Restaurant	Park	Restaurant	Coffee Shop	Supermarket	Turkish Restaurant	Golf Course	Pub	Convenience Store
2	Bexley	Pub	Grocery Store	Coffee Shop	Supermarket	Fast Food Restaurant	Clothing Store	Hotel	Convenience Store	Café	Pizza Place
3	Brent	Coffee Shop	Grocery Store	Hotel	Sandwich Place	Fast Food Restaurant	Sporting Goods Shop	Warehouse Store	Clothing Store	Supermarket	Café
4	Bromley	Pub	Coffee Shop	Pizza Place	Gym / Fitness Center	Park	Clothing Store	Grocery Store	Indian Restaurant	Supermarket	Electronics Store

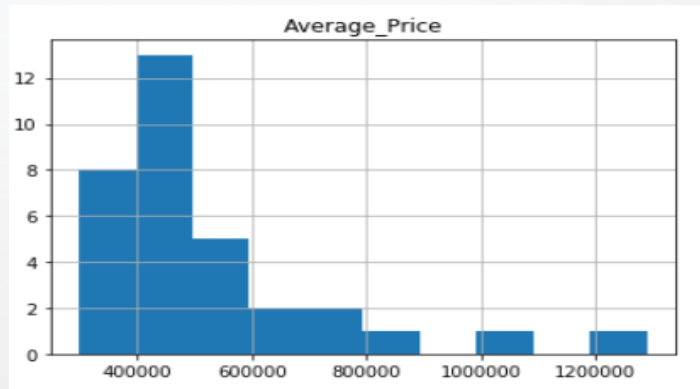
Methodology (3 of 4) - cluster the boroughs

- Use K-means algorithm to cluster the boroughs
- Set parameters
 - number of clusters equal 6
- Set names to each cluster
 - Cluster 0 – 'Hotel &Cafe &Sights'
 - Cluster 1 – 'Pub & Fast Food &Park'
 - Cluster 2 – 'Pub &Grocery &Bus stop'
 - Cluster 3 – 'Coffee &Cocktail &Pizza'
 - Cluster 4 – 'Pub &Café &Pizza'
 - Cluster 5 – 'Cafe &Turkish&Indian Food &Park'

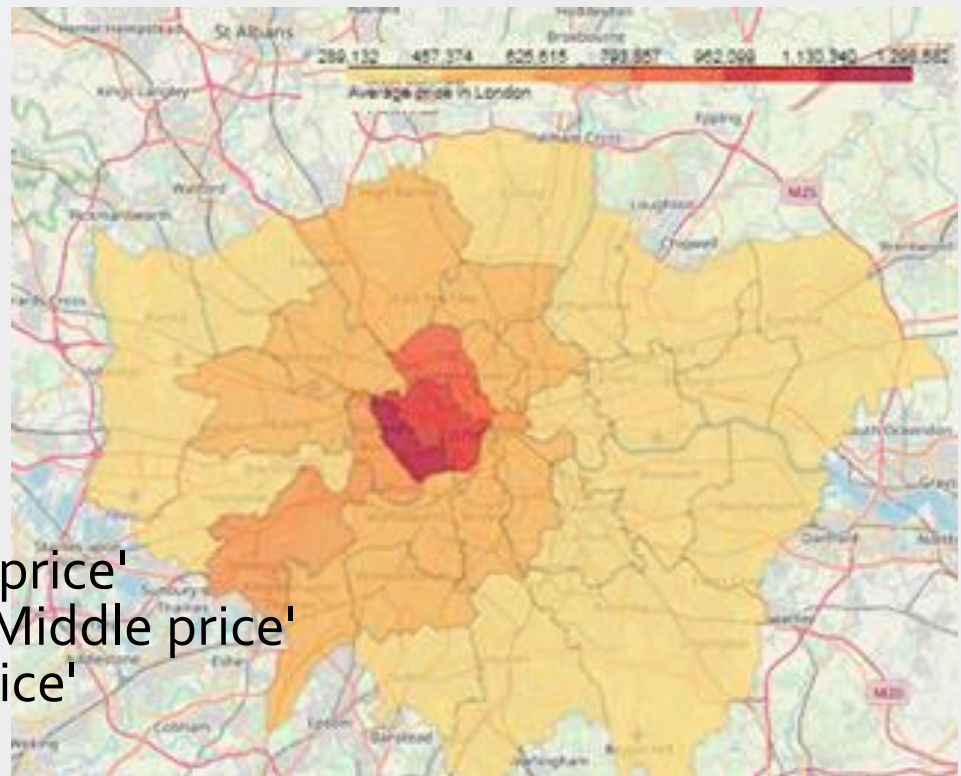


Methodology (4 of 4) –ranges of average prices

- Create histogram with level of average prices:



- Define the ranges
 - 400 000: 'Low price'
 - 400 000–500 000: 'Middle price'
 - 500 000–600 000: 'Above Middle price'
 - 600 000–800 000: 'High price'
 - 800 000: 'Very high price'

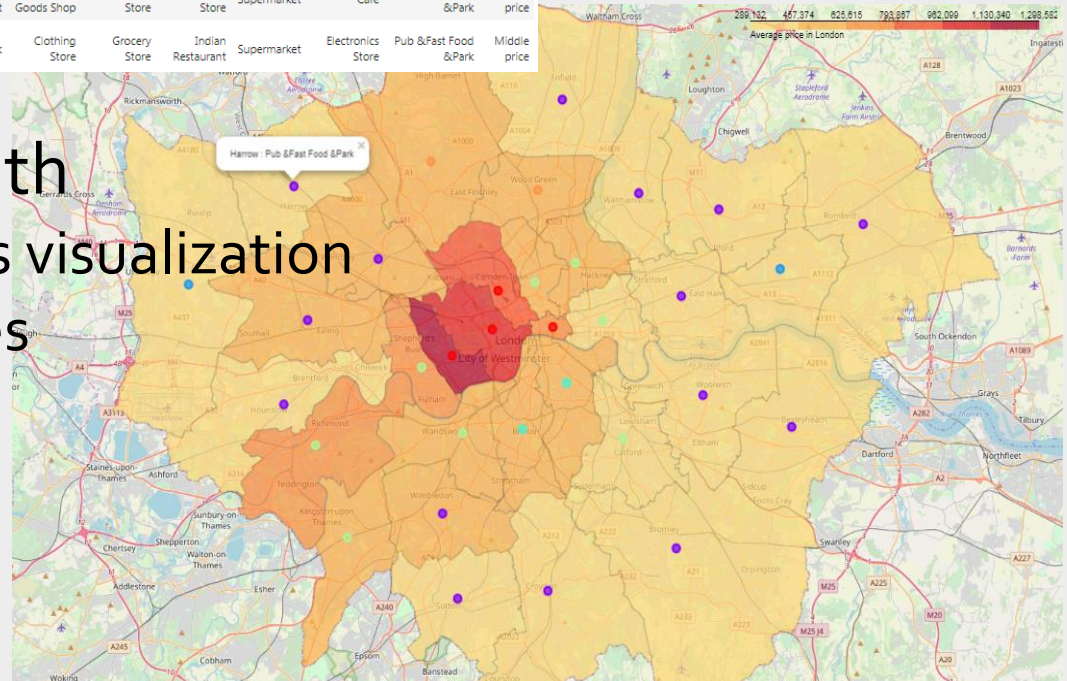


Results

- Master data with clusters and price categories

Borough	Latitude	Longitude	Average_Price	Cluster Labels	251 House Common Venue	2112 House Common Venue	3112 House Common Venue	4111 House Common Venue	5111 House Common Venue	6111 House Common Venue	7111 House Common Venue	8111 House Common Venue	9111 House Common Venue	10111 House Common Venue	Cluster Name	Level prices
0 Barking and Dagenham	51.546501	0.125051	299028.7074	2	Grocery Store	Pub	Bus Stop	Soccer Field	Park	Supermarket	Recreation Center	Go Kart Track	Café	Movie Theater	Pub & Grocery & Bus stop	Low price
1 Barnet	51.605598	-0.207649	516750.1635	5	Café	Indian Restaurant	Park	Restaurant	Coffee Shop	Supermarket	Turkish Restaurant	Golf Course	Pub	Convenience Store	Cafe & Turkish & Indian Food & Park	Above Middle price
2 Bexley	51.459202	0.136265	333317.2861	1	Pub	Grocery Store	Coffee Shop	Supermarket	Fast Food Restaurant	Clothing Store	Hotel	Convenience Store	Café	Pizza Place	Pub & Fast Food & Park	Low price
3 Brent	51.551800	-0.257490	474540.4375	1	Coffee Shop	Grocery Store	Hotel	Sandwich Place	Fast Food Restaurant	Sporting Goods Shop	Warehouse Store	Clothing Store	Supermarket	Café	Pub & Fast Food & Park	Middle price
4 Bromley	51.391800	0.026386	430242.4197	1	Pub	Coffee Shop	Pizza Place	Gym / Fitness Center	Park	Clothing Store	Grocery Store	Indian Restaurant	Supermarket	Electronics Store	Pub & Fast Food & Park	Middle price

- Map with markers with
 - average house prices visualization
 - markers about venues



Discussion and Conclusion

- **Discussion**

- We used K-means algorithm with number of clusters equal 6
 - change number of clusters
 - change number of top venues.
- We used only one pair of coordinates for each of 33 boroughs
 - dataset can be extended
 - drill down the dataset to the streets.
- We used average house price data on a certain period
 - use actual data about prices.

- **Conclusion section**

- Useful information for companies and entrepreneurs for choosing the office location
 - Convenient visualization (map with colors and markers)
 - Consolidation information on realty prices and the necessary infrastructure availability in different London boroughs