Analysis of the London districts (common venues, average house prices)

Alexander Malkov

June, 2018

1. Introduction

1.1. Background

London is considered to be one of the world's most important global cities and has been termed the world's most powerful, most desirable, most influential, most visited, most expensive, innovative, sustainable, most investment friendly, most popular for work, and the most vegetarian-friendly city in the world. London exerts a considerable impact upon the arts, commerce, education, entertainment, fashion, finance, healthcare, media, professional services, research and development, tourism and transportation. London ranks 26 out of 300 major cities for economic performance. It is one of the largest financial centres and has either the fifth or sixth largest metropolitan area GDP. [1]

London divided into 32 local authority districts that make up Greater London (London boroughs) and City of London. The London boroughs were all created at the same time as Greater London on 1 April 1965 by the London Government Act 1963 and are a type of local government district. Twelve were designated as Inner London boroughs and twenty as Outer Londonboroughs.

The London boroughs have populations of around 150,000 to 300,000. Inner London boroughs tend to be smaller, in both population and area, and more densely populated than Outer London boroughs. The London boroughs were created by combining groups of former local government units. [2]

1.2. Problem

The London boroughs differ from each other in a variety of parameters. For different purposes it is necessary to select parameters to compare London boroughs.

This project aims to identity differences of the London boroughs using information about common venues and average property prices.

1.3. Interest

If businessman is planning to open an office in London he would be interested in information about the districts infrastructure (venues) and the average house prices.

---

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

[2] https://en.wikipedia.org/wiki/London_boroughs

2. Data

Data were used to solve the problem:

- Postcodes by administrative area[3].

  The csv-file contains information about coordinates of the all UK areas. It is necessary to filter the data for areas of London.

- Average house prices[4]

  The csv-file contains information about average house prices of the all UK areas from period 1968-2019. It is necessary to filter the data for areas of London and last period (2019-04-01). After that, for further analysis it is necessary to merge the data (coordinates and average prices).

- Forsquare API[5]

  It is necessary to get the most common venues of given Borough of London.

- Country geojson file[6]

  It is necessary to create choropleth map with level of average prices.

---

[3] https://www.doogal.co.uk/AdministrativeAreas.php

[4] http://publicdata.landregistry.gov.uk/market-trend-data/house-price-index-data/Average-prices-2019-04.csv

[5] https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}

[6] https://github.com/radoi90/housequest-data/raw/master/london_boroughs.geojson

3. Methodology section

The master data includes the required for analysis components (borough name, borough latitude and longitude, average house price in the borough):

	Borough	Latitude	Longitude	Average_Price
0	Barking and Dagenham	51.546501	0.125051	299028.7074
1	Barnet	51.605598	-0.207649	516750.1635
2	Bexley	51.459202	0.136265	333317.2861
3	Brent	51.551800	-0.257490	474540.4375
4	Bromley	51.391800	0.026386	430242.4197

Using the library Folium we can created a map of London with markers (boroughs):
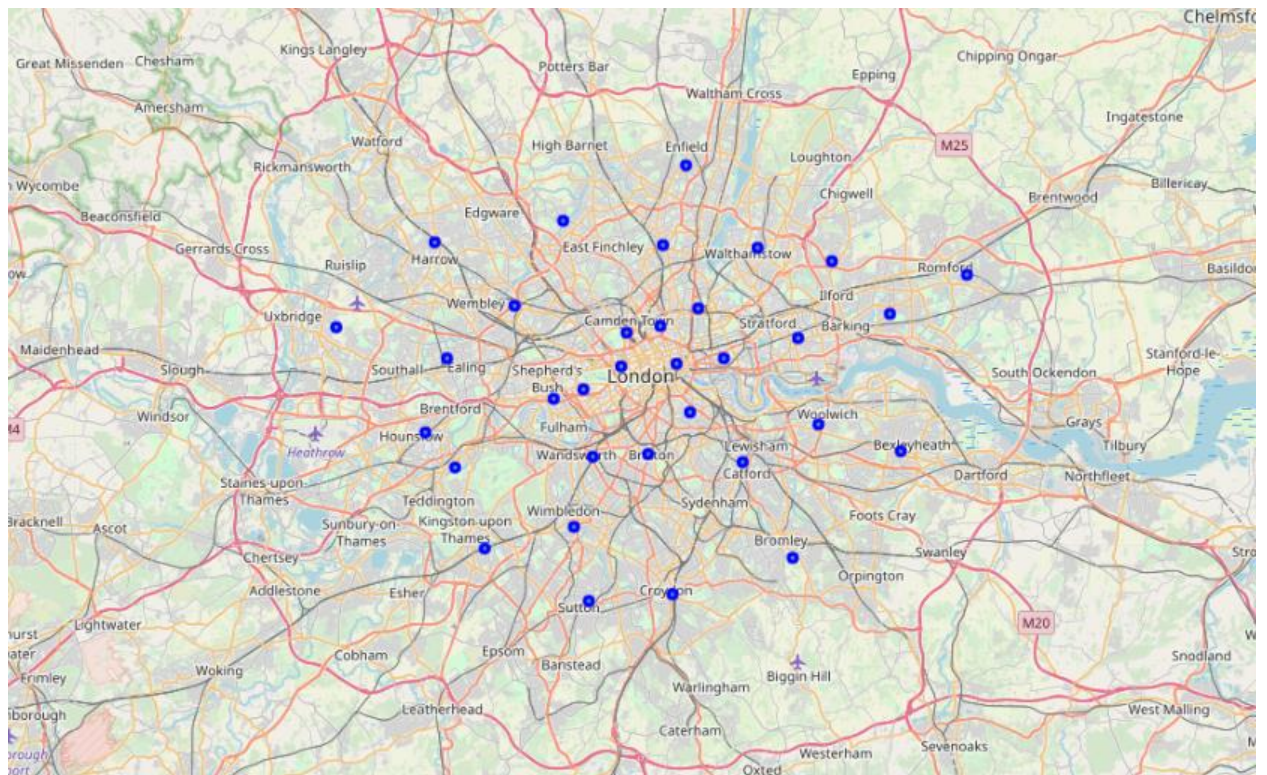
Using the Foursquare API we can explore the boroughs and segment them.

If determine the limit of number of venues equal 100 and radius equal 2000 meters for each borough for its latitude and longitude values we can get a list of venues:

	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Barking and Dagenham	51.546501	0.125051	Capital Karts	51.531792	0.118739	Go Kart Track
1	Barking and Dagenham	51.546501	0.125051	Mayesbrook Park	51.549842	0.108544	Park
2	Barking and Dagenham	51.546501	0.125051	Co-op Food	51.540093	0.127522	Grocery Store
3	Barking and Dagenham	51.546501	0.125051	Vue	51.532149	0.135000	Movie Theater
4	Barking and Dagenham	51.546501	0.125051	Lidl	51.551934	0.112965	Supermarket

By sorting to quantity of venues we observe that some boroughs have less than 100 venues:

Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Hillingdon	25	25	25	25	25	25
Barking and Dagenham	39	39	39	39	39	39
Redbridge	43	43	43	43	43	43
Barnet	59	59	59	59	59	59
Greenwich	64	64	64	64	64	64
Bexley	67	67	67	67	67	67
Bromley	71	71	71	71	71	71
Sutton	75	75	75	75	75	75
Harrow	77	77	77	77	77	77
Newham	78	78	78	78	78	78
Enfield	83	83	83	83	83	83
Waltham Forest	96	96	96	96	96	96
Hounslow	98	98	98	98	98	98
Haringey	100	100	100	100	100	100
Brent	100	100	100	100	100	100

This is because we used the restriction of only one pair of coordinates and on the radius for each borough.

Select 10 top venues for each borough:

	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

Since the K-means algorithm is one of the most common cluster unsupervised learning methods, we can use this algorithm to cluster the boroughs.

We set number of clusters equal 6 and run k-means clustering. The merged table with cluster labels for each borough:

	Borough	Latitude	Longitude	Average_Price	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	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
1	Barnet	51.605598	-0.207649	516750.1635	5	Café	Indian Restaurant	Park	Restaurant	Coffee Shop	Supermarket	Turkish Restaurant	Golf Course	Pub	Convenience Store
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
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
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

We can examine each cluster and determine the discriminating venue categories that distinguish each cluster. Based on the defining categories, we can assign a name 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'

We can also examine the frequency of average housing prices in different ranges. We create histogram with level of average prices:
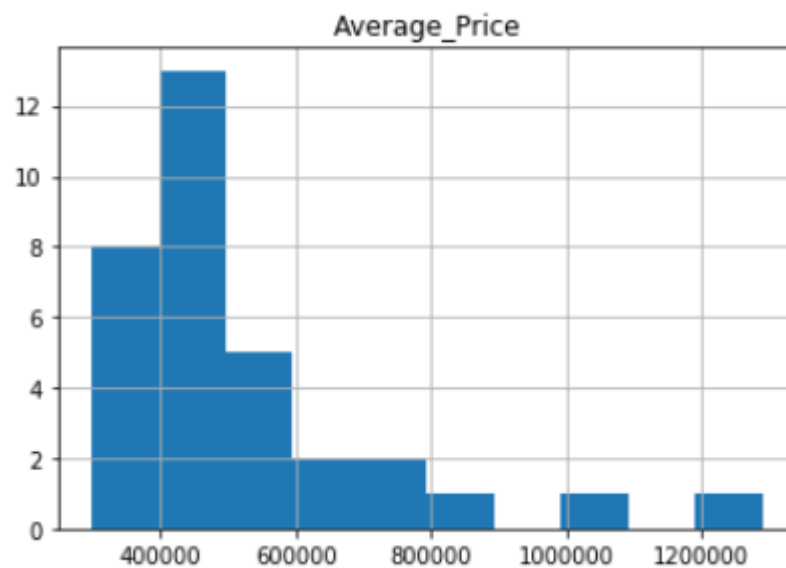
Now we can define the ranges as below:

- 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'.

4. Results section

Let's merge those new variables with related cluster information in our main master table.

	Borough	Latitude	Longitude	Average_Price	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	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

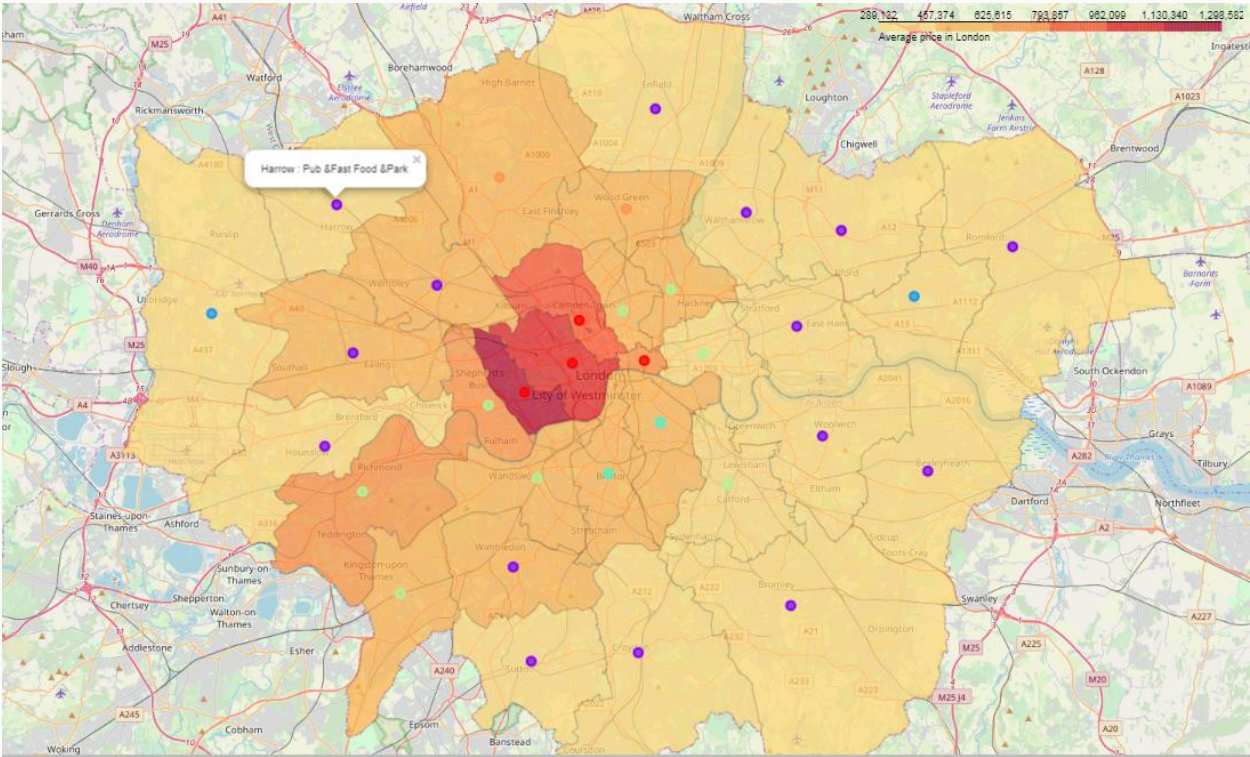We can see a cluster map of London boroughs below:
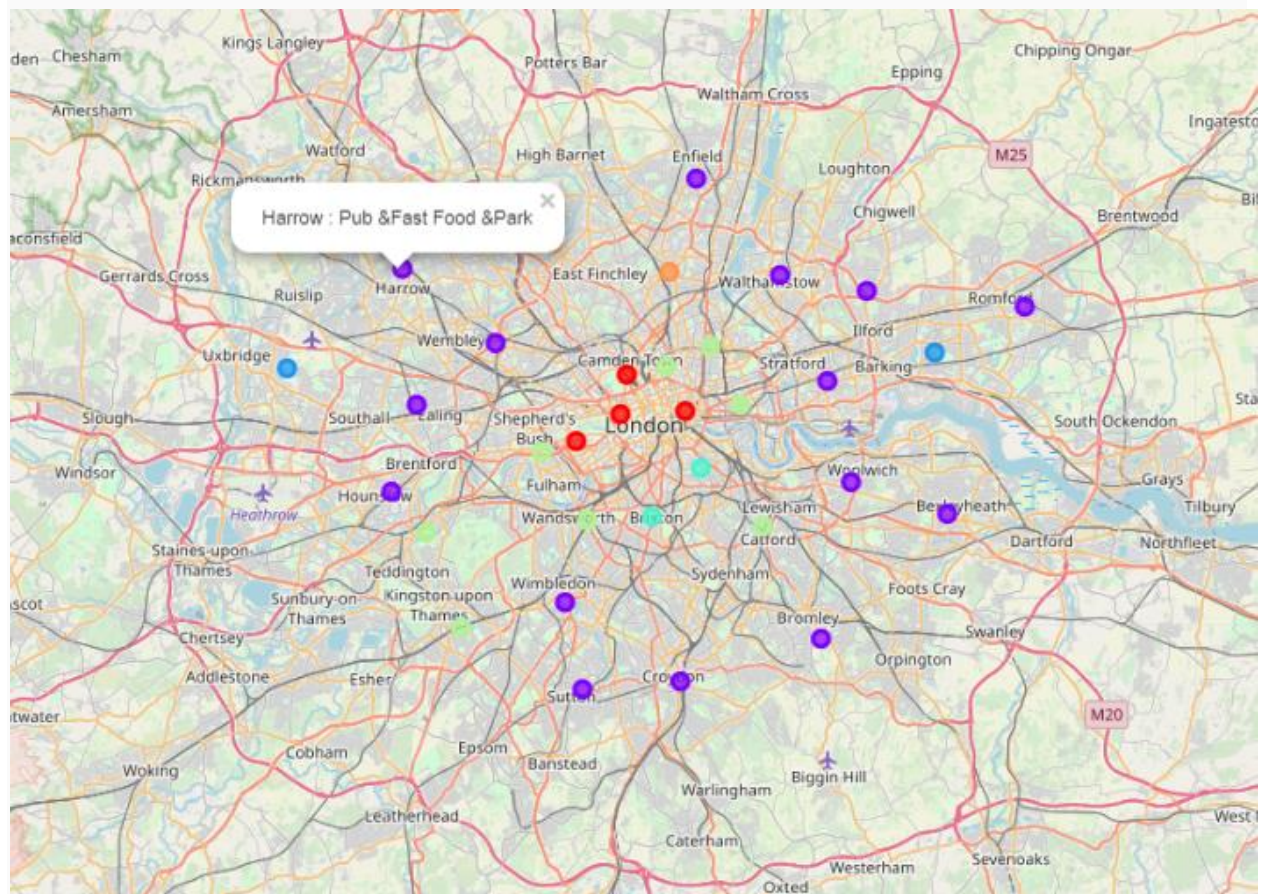
We can also visualize the average house prices using choropleth map:
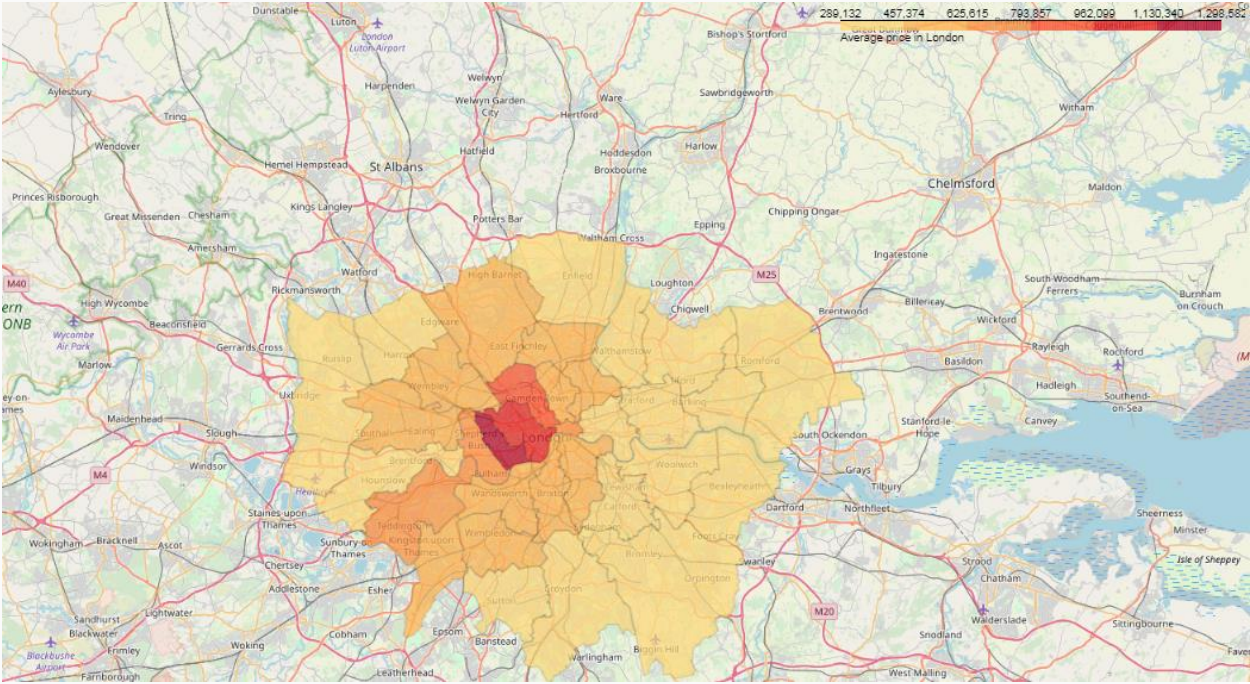
Finally we can add markers about venues to the choropleth map with level of average prices

5. Discussion section

In our research of London boroughs we used K-means algorithm with number of clusters equal 6.

We used only one pair of coordinates for each of 33 boroughs. However, the boroughs are very different in area: some boroughs are too large for radius (2000 meters). For more detailed and accurate studies, the dataset can be extended, for example, we drill down the dataset to the streets. We can also try using a different value for the number of clusters and the number of top venues.

We also used average house price data on a certain period. For more accurate model we have to use actual data about prices.

6. Conclusion section

Often companies and entrepreneurs open offices in financial centers. London is one of the largest financial center.

Consolidated visualized information about the realty prices and the necessary infrastructure availability in different London boroughs will be useful for choosing the office location.