**Evaluation of the**

# Boroughs in London, UK

**in order to identify the**

# 'Best Borough to Live'

**Amila S Somarathna**

Final Assignment of the Applied Data Science Capstone by IBM/ Coursera

May 1, 2020

# Table of contents

# A: Introduction

### A.1: Background

London is considered to be one of the world's most important global cities and has been called the world's most powerful, most desirable, most influential, most visited, most expensive, innovative, sustainable, most investment-friendly, and most-popular-for-work city.[1]

Every year, thousands of people make the move to London both from within the UK and from overseas. They decide to move and settle down in London due to many reasons such as work commitment changes, looking for better living conditions, etc. However, there are certain things they have to consider before moving in. London housing and rental prices are among the highest in the world and can eat up to a significant portion of their income. Furthermore, other living costs such as public transport fares, owning a vehicle & driving are not cheap either. Considering these and many other facts, it is relatively tough matter to decide to where to settle down within London.

### A.2: Problem

London has 32 boroughs which vary from each other by many aspects: cost of living, housing prices, crime rates, etc. to name a few. Therefore, our problem here would be to find out the best London borough to live considering various facts & environments mentioned above.

### A.3: Interest

Any newcomer has to educate themselves beforehand about the things mentioned above to decide the best place for them to settle down. Furthermore, it is for any real estate agent's advantage that they are well updated on such matters whenever a client contacts them with such enquiry. Also, this knowledge will be welcomed by any property developer as it helps them on deciding best places to build their next housing scheme.

# B: Data

### B.1 Data Description

Based on the definition of our problem, there are certain data sets we have to consider:

- Various facts about the 32 London boroughs; ex: cost of living, house price, crime rate

- Other facilities that make them desirable neighbourhoods; ex: parks, shopping malls, attractions
- Since we are intend to elaborate these data using a choropleth map, we need to have boundaries of London boroughs

### B.2 Data Sources

Following data sources will be used to extract or generate the required information:

- Data and facts about London boroughs will be obtained as a .csv file from London Borough Profiles and Atlas, London Data Store website [2]
- Important venues and other desired locations will be obtained using Foursquare API [3]
- London borough boundaries will be obtained as a .json file from Statistical GIS Boundary Files for London, London Data Store website [4]

### B.3 Obtaining Data and Fine Tuning

The statistical data about London Boroughs was downloaded as a .csv file from London Data Store website. However, there are few problems with this data set.

There are too many unnecessary data fields (columns), thus they has to be dropped. We also have to consider dropping some rows as they are grouped data which is duplications.

Fortunately there is not any missing data for any field we decided to keep. However, the ranges of these data are too varied from each other. Therefore, all the numerical data will be normalized before any math was performing on them.

The cleaned and un-normalized data set looks as in figure 1:

| | Code | Borough | Inner_Outer | Businesses | Crime_rate | Fires_rate | House_Price | Council_tax | Greenspace | Public_Transport | GCSE | Life_satisfaction | Worthiness | Happiness |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | E09000002 | Barking and Dagenham | Outer London | 6560 | 83.4 | 3 | 243500 | 1354.03 | 33.6 | 3 | 58 | 7.1 | 7.6 | 7.1 |
| 1 | E09000003 | Barnet | Outer London | 26190 | 62.7 | 1.6 | 445000 | 1397.07 | 41.3 | 3 | 67.3 | 7.5 | 7.8 | 7.4 |
| 2 | E09000004 | Bexley | Outer London | 9075 | 51.8 | 2.3 | 275000 | 1472.43 | 31.7 | 2.6 | 60.3 | 7.4 | 7.7 | 7.2 |
| 3 | E09000005 | Brent | Outer London | 15745 | 78.8 | 1.8 | 407250 | 1377.24 | 21.9 | 3.7 | 60.1 | 7.3 | 7.4 | 7.2 |
| 4 | E09000006 | Bromley | Outer London | 15695 | 64.1 | 2.3 | 374975 | 1347.27 | 57.8 | 2.8 | 68 | 7.5 | 7.9 | 7.4 |

**Figure 1: Un-normalized data**

All long column names are shortened for the ease of use. The cleaned and normalized data set looks as in figure 2:

| | Code | Borough | Inner_Outer | Businesses | Crime_rate | Fires_rate | House_Price | Council_tax | Greenspace | Public_Transport | GCSE | Life_satisfaction | Worthiness | Happiness |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | E09000002 | Barking and Dagenham | Outer London | 0.118444 | 0.392655 | 0.750 | 0.202917 | 0.804419 | 0.566610 | 0.461538 | 0.822695 | 0.934211 | 0.962025 | 0.934211 |
| 1 | E09000003 | Barnet | Outer London | 0.472872 | 0.295198 | 0.400 | 0.370833 | 0.829989 | 0.696459 | 0.461538 | 0.954610 | 0.986842 | 0.987342 | 0.973684 |
| 2 | E09000004 | Bexley | Outer London | 0.163853 | 0.243879 | 0.575 | 0.229167 | 0.874759 | 0.534570 | 0.400000 | 0.855319 | 0.973684 | 0.974684 | 0.947368 |
| 3 | E09000005 | Brent | Outer London | 0.284283 | 0.370998 | 0.450 | 0.339375 | 0.818208 | 0.369309 | 0.569231 | 0.852482 | 0.960526 | 0.936709 | 0.947368 |
| 4 | E09000006 | Bromley | Outer London | 0.283380 | 0.301789 | 0.575 | 0.312479 | 0.800403 | 0.974705 | 0.430769 | 0.964539 | 0.986842 | 1.000000 | 0.973684 |

Figure 2: Normalized data

# C: Methodology

By looking at the above dataset, we can clearly distinguish 2 different categories of values:

- Values that a higher percentage makes a positive impact: Ex:- Greenspace, Public_Transport, GCSE, etc.
- Values that a lower percentage makes a positive impact: Ex:- Crime_rate, House_Price, Council_Tax, etc.

Therefore, we have to calculate 2 mean values: 1 Positive average, 1 Negative average, and find the Borough which satisfy both the criteria.

Let us append 2 more columns to the dataframe, named Pos_avg and Neg_avg, and calculate cumulative averages of particular columns. It looks as in figure 3.

| Pos_avg | Neg_avg |
|---|---|
| 0.685676 | 0.537498 |
| 0.790478 | 0.474005 |
| 0.692783 | 0.480701 |
| 0.702844 | 0.494645 |
| 0.801988 | 0.497418 |

Figure 3: Positive & Negative means

## C.1 Analysis with Charts

Now let us visualize these data using charts.

| Borough | Pos_avg |
|---|---|
| Westminster | 0.917172 |
| Bromley | 0.801988 |
| Camden | 0.793757 |
| Barnet | 0.790478 |
| Richmond upon Thames | 0.789202 |

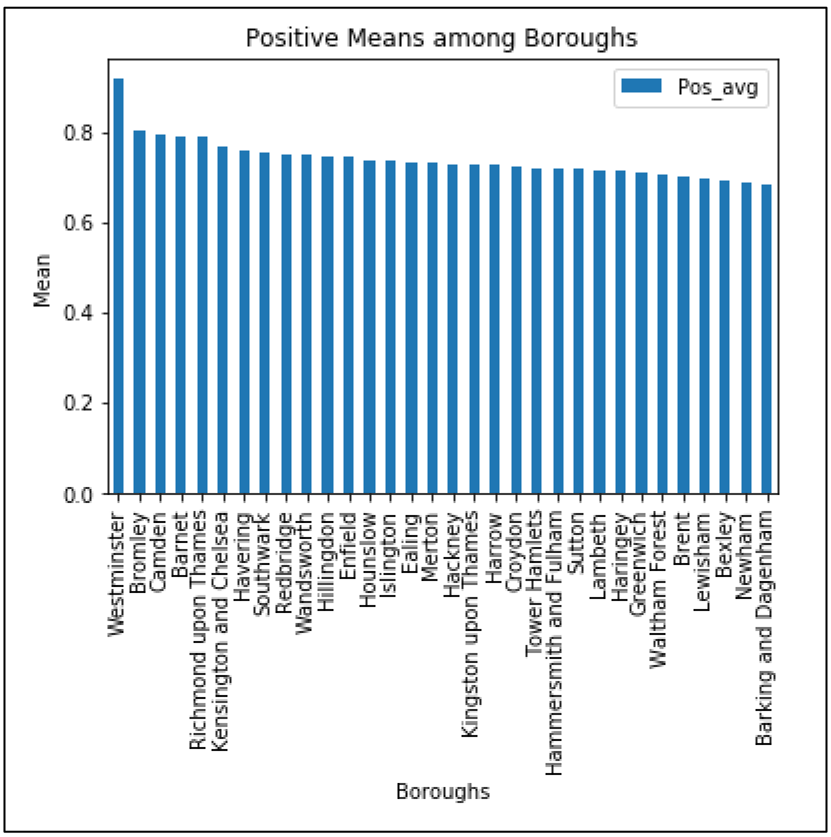


Figure 4: Top 5 Positve average Broughs

We can see Westminster winning this race with more than 90% positive average. The second and third places go to Bromley and Camden respectively.

| Borough | Neg_avg |
|---|---|
| Wandsworth | 0.414976 |
| Harrow | 0.454741 |
| Barnet | 0.474005 |
| Merton | 0.474552 |
| Sutton | 0.478172 |

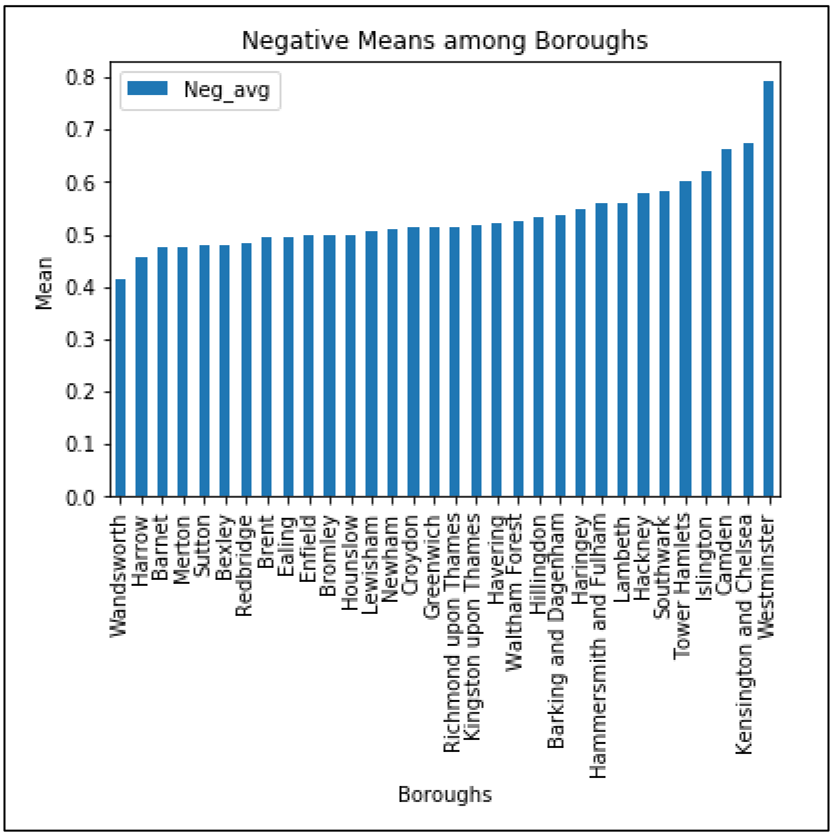


**Figure 5: Top 5 Negative average Boroughs**

Wandsworth will win this race followed by Harrow and Barnet.

Right, Let us append another column to ft2_df and include the difference between Pos_avg and Neg_avg. This way we can find which borough has the highest gap between Pos_avg and Neg_avg. Then we create the chart shown as in figure 6.

| Borough | Pos_neg_dif |
|---|---|
| Wandsworth | 0.335841 |
| Barnet | 0.316473 |
| Bromley | 0.304571 |
| Richmond upon Thames | 0.274372 |
| Harrow | 0.274239 |

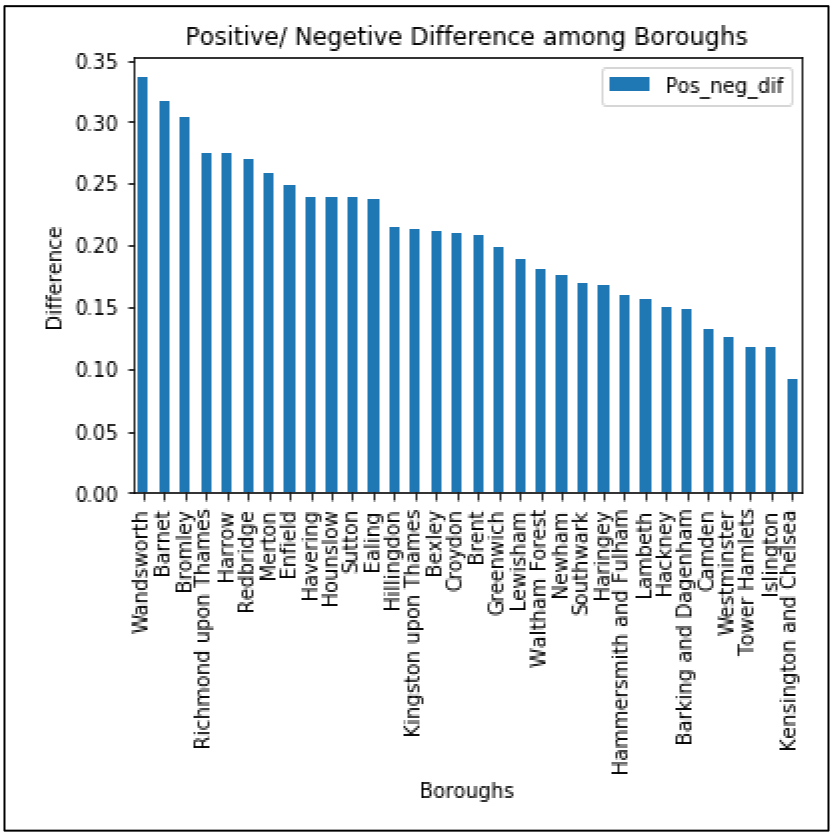


**Figure 6: Top 5 Boroughs with biggest Positive -Negative gap**

## C.2 Analysis with Choropleth Maps

To implements maps, we first have to get geo-locations for our Boroughs. For the simplicity, let us go for a center-point. Thus the following Wikipedia table [6] gives us the best way of achieving this.

Let us get rid of unnecessary data. We only need the Borough names and geo-locations. By looking at this dataset, we can clearly see that some of the Borough geo-locations are misleading. Namely: Bexley, Brent, Camden, Havering, Kensington and Chelsea, Sutton, Tower Hamlets, Waltham Forest and Westminster. Therefore, we have to get the actual location data for them and fix. We can use the Nominatim geocode() function to get actual geo-coding for each place.

So the corrected table will look as in figure 7.

| | Borough | Latitude | Longitude | Pos_neg_dif |
|---|---|---|---|---|
| **0** | Barking and Dagenham | 51.554117 | 0.150504 | 0.148179 |
| **1** | Barnet | 51.653090 | -0.200226 | 0.316473 |
| **2** | Bexley | 51.441679 | 0.150488 | 0.212081 |
| **3** | Brent | 51.563826 | -0.275760 | 0.208199 |
| **4** | Bromley | 51.402805 | 0.014814 | 0.304571 |

**Figure 7: Corrected geo-code data**

And the map looks as in figure 8.

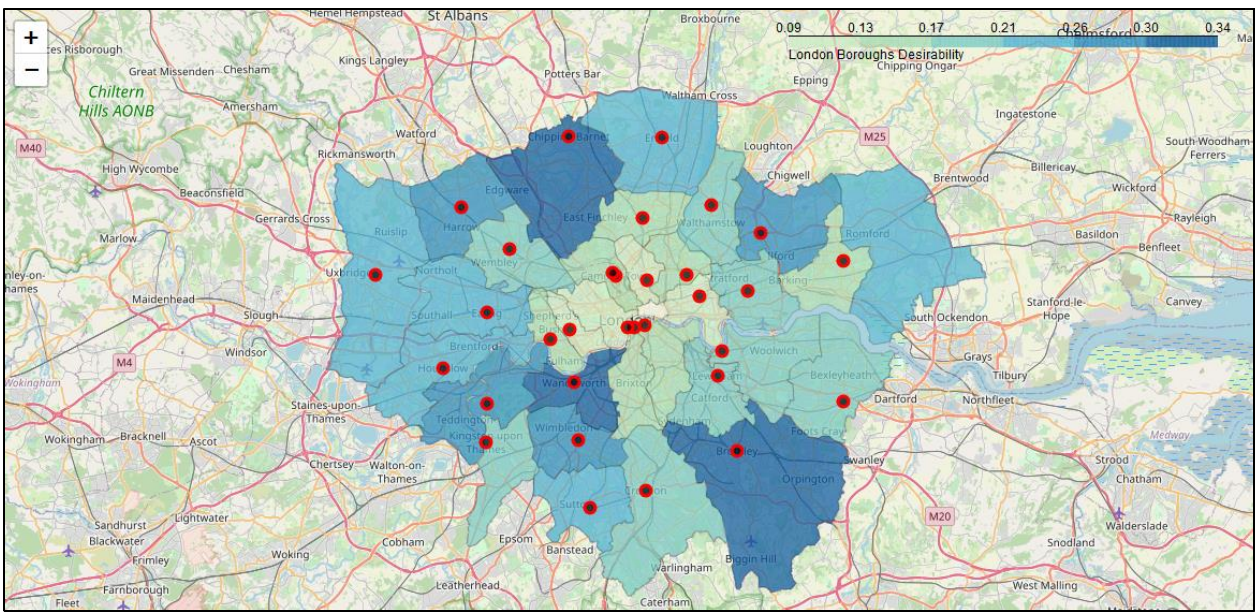


**Figure 8: London Boroughs as per their Positive -Negative difference**

By looking at the above map, we can see that:

- Outer Boroughs are scoring more against the Inner Boroughs. Probably because of the higher house prices and council taxes in the Inner Boroughs. Also because the Outer Boroughs are bigger, more spacious with lots of greenery and provides nature to enjoy.

- For some of the geo-locations we used to mark the Boroughs are actually not representing the center of that Borough.

With considering the size differences of Boroughs, (especially between Inner and Outer Boroughs) it is unwise to retrieve radius-wise venue data from Foursquare for each Borough and compare. Since a relatively smaller set radius will not cover enough area in bigger Boroughs, while a relatively bigger set radius might overlap venues for smaller Boroughs.

To solve this, and simplify the process, especially since now we have found out that Wandsworth is our go-to place, let us concentrate on Wandsworth alone and find out what are the attractive venues it provide to make it so popular.

Wandsworth is split into 20 areas or wards: (Balham, Bedford, Earlsfield, East Putney, Fairfield, Furzedown, Graveney, Latchmere, Nightingale, Northcote, Queenstown, Roehampton and Putney Heath, Shaftesbury, Southfields, St Mary's Park, Thamesfield, Tooting, Wandsworth Common, West Hill, West Putney)

So let us have an analysis among these 20 Wards.

NB: I tried to get geo-locations for each Ward using Nominatim geocode() function. However, after many tries and re-tries, it returned with wrong geo-locations for many Wards. Therefore, I have to get the geo-locations from Google Maps: Special thanks to the creator(s) of Wandsworth Borough Wards Google My Map [7] for providing geo-locations plus boundaries KML file to use in this project.

I also used the Average House Prices by Borough, Ward, MSOA & LSOA, London Data store [8] website to obtain the average house prices for these wards.

### C.3 Further Analysis with Foursquare Data

Building the Wandsworth data set; it looks as in figure 9.

| | Ward | Latitude | Longitude | Houseprice |
|---|---|---|---|---|
| **0** | Balham | 51.448691 | -0.150859 | 656.500 |
| **1** | Bedford | 51.434879 | -0.149078 | 592.875 |
| **2** | Earlsfield | 51.442106 | -0.186973 | 595.000 |
| **3** | East Putney | 51.455450 | -0.212334 | 575.000 |
| **4** | Fairfield | 51.458071 | -0.191678 | 610.000 |

**Figure 9: Wandsworth wards with geocodes and average house prices**

With this data, we acquire Foursquare™ venue data for each of the above ward. The data set looks as in figure 10.

| | Ward | Ward Latitude | Ward Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| 0 | Balham | 51.448691 | -0.150859 | Ciullosteria | 51.447144 | -0.148981 | Italian Restaurant |
| 1 | Balham | 51.448691 | -0.150859 | M1LK | 51.444450 | -0.150913 | Coffee Shop |
| 2 | Balham | 51.448691 | -0.150859 | We Brought Beer | 51.444324 | -0.150656 | Beer Store |
| 3 | Balham | 51.448691 | -0.150859 | Brickwood Coffee & Bread | 51.444509 | -0.151127 | Coffee Shop |
| 4 | Balham | 51.448691 | -0.150859 | Balham Bowls Club | 51.444984 | -0.152306 | Pub |

**Figure 10: Foursquare data for Wandsworth wards**

There are 106 unique categories. Let us group these data by the ward now and let see how many locations are there for each ward.

| | Ward Latitude | Ward Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|
| **Ward** | | | | | | |
| **Balham** | 38 | 38 | 38 | 38 | 38 | 38 |
| **Bedford** | 6 | 6 | 6 | 6 | 6 | 6 |
| **Earlsfield** | 29 | 29 | 29 | 29 | 29 | 29 |
| **East Putney** | 19 | 19 | 19 | 19 | 19 | 19 |
| **Fairfield** | 43 | 43 | 43 | 43 | 43 | 43 |

**Figure 11: Venue count for each Ward**

We can see that none of the Wards has returned with maximum number ie. 100. The maximum venue count for a Ward is for Thamesfield which is 62, followed by Northcote (51) and Fairfield (43). Also note that there are only 19 records. West Putney has not returned any venue records.

Now let us do one hot encoding on this data and group rows by Ward and by taking the mean of the frequency of occurrence of each category.

| | Ward | Animal Shelter | Antique Shop | Arts & Crafts Store | Asian Restaurant | Athletics & Sports | Bakery | Bar | Beer Bar | Beer Store | ... | Theater | Track | Trail | Train Station | Turkish Restaurant | Vietnamese Restaurant | Wine Bar |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Balham | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.078947 | 0.000000 | 0.000000 | 0.026316 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 |
| 1 | Bedford | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.166667 | 0.000000 | 0.166667 | 0.000000 | 0.000000 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 |
| 2 | Earlsfield | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 |
| 3 | East Putney | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.052632 | 0.000000 | 0.052632 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.052632 | 0.00 | 0.000000 |
| 4 | Fairfield | 0.00 | 0.000000 | 0.000000 | 0.046512 | 0.000000 | 0.023256 | 0.000000 | 0.000000 | 0.000000 | ... | 0.000000 | 0.00 | 0.000000 | 0.000000 | 0.000000 | 0.00 | 0.000000 |

**Figure 12: Venue data ne-hot encded and groupd**

Let us print each Ward along with the top 10 most common venues.

| | Ward | 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 | Balham | Grocery Store | Coffee Shop | Pub | Bakery | Indian Restaurant | Fast Food Restaurant | Breakfast Spot | Caucasian Restaurant | Fish & Chips Shop | Sandwich Place |
| 1 | Bedford | Hotel | Athletics & Sports | Bar | Tennis Court | Café | Bus Stop | Yoga Studio | Fried Chicken Joint | Deli / Bodega | Eastern European Restaurant |
| 2 | Earlsfield | Café | Coffee Shop | Thai Restaurant | Cocktail Bar | Platform | Pub | Grocery Store | Steakhouse | Burger Joint | Lounge |
| 3 | East Putney | Grocery Store | Coffee Shop | Thai Restaurant | Dance Studio | Park | Portuguese Restaurant | Record Shop | Sandwich Place | Café | Beer Store |
| 4 | Fairfield | Pub | Coffee Shop | Clothing Store | Hotel | Pizza Place | Supermarket | Sandwich Place | Breakfast Spot | Asian Restaurant | Gym / Fitness Center |

Figure 13: Top 10 most common venues for each Ward

Let us cluster the Wards; Run k-means to cluster the Wards into 5 clusters and analyze. Then Let us create a new data set that includes the clusters as well as the top 10 venues for each Ward. And finally, let us visualize the resulting clusters on a map.

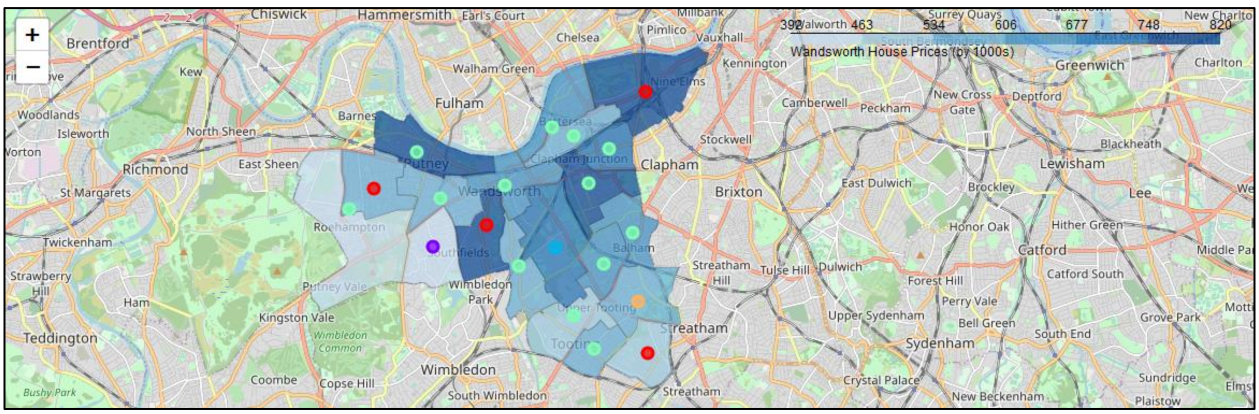


Figure 14: Wandsworth wards with venue clusters

# D: Result

The first part of this analysis was to find out the best London Borough to live, given on various facts. We obtained a London dataset with such stats and after cleaning, standardizing and normalizing these data, we analyzed these facts against each other and also among Boroughs. By doing so we found out that there are 2 types of data/ facts.

- The higher the number, increase the positive side of a Borough (Ex. Education standard, level of green space, etc.)
- The lower the number, increase the positive side of a Borough (Ex. Crime rate, House prices, Council Tax, etc.)

Therefore, we added 2 more columns: Positive mean and Negative mean by grouping these 2 different types of data and identified what are Boroughs with higher Positive means and what are the Boroughs with Lesser Negative means.

After that, we calculated the Positive - Negative difference for each Borough and found out what are the Boroughs with highest difference. Which gave us the answer for our question: What is the best London Borough to live?

According to that analysis, the Top 5 Boroughs were:

- Wandsworth
- Barnet
- Bromley
- Richmond upon Thames
- Harrow

Since we found out that Wandsworth is the best Borough to live, we shifted our focus to Wandsworth. We used Foursquare™ data to find out all the popular commercial venues located in each of the 20 Wandsworth Wards. Furthermore, we obtained average house prices for each of these Wards and compared that with the frequency of commercial venues.

Finally, as mentioned in the Data Description section, we visualized these data using Choropleth maps, both for London Boroughs and Wandsworth Wards.

# E: Discussion

As mentioned in the Introduction, London is one of the most important global cities plus one of the most desirable city to live as well; Having said that, London is also a huge city, with 32 Boroughs of different sizes spread over thousands of hectares.

Since the beginning, we clearly identified 2 different sets of Boroughs, which are:

- Inner London Boroughs
- Outer London Boroughs

Inner London Boroughs are relatively small in size, and being close to the epicenter of the city, are more expensive to live. Higher house prices, higher council taxes plus other high living costs associated with these Boroughs might make someone think twice of living in there regardless of many more positives such as good access to public transport and so on.

By comparison, Outer London Boroughs are bigger in size, less congested and offer more space to live in. Also they offer more green space which would delights the nature lovers. The living costs are relatively lower; however, they also have their own drawbacks too.

I believe that this analysis could be making better by breaking it into 2 parts. By analyzing Inner Boroughs and Outer Boroughs separately, we can have a better idea. Probably we will be able to find the 'Best Inner Borough' and the 'Best Outer Borough' to live in.

Having said all these, the winner of this contest was rather surprisingly, an Inner Borough, Wandsworth. Now, that is another thing to analyze.

By coincidence, while surfing through the Internet looking for data, I found out this newspaper article. Seems like they have found out an answer for that!

7 reasons why Wandsworth is the best place to live in London

# F: Conclusion

As mentioned in the Introduction these kind of analysis are very important to real estate agents and property developers. Also, anyone who wishes to move in to London will welcome such information.

People can take better decisions through accessing such information where they are provided freely and frequently.

# G: References


- [1] London Wikipedia Page, *https://en.wikipedia.org/wiki/London*
- [2] London Borough Profiles and Atlas, London Data Store, *https://data.london.gov.uk/dataset/london-borough-profiles*
- [3] Foursquare API, *https://developer.foursquare.com/*
- [4] Statistical GIS Boundary Files for London, London Data Store, *https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london*
- [5] City of London Wikipedia Page, *https://en.wikipedia.org/wiki/City_of_London*
- [6] List of London Boroughs, *https://en.wikipedia.org/wiki/List_of_London_boroughs*
- [7] Wandsworth Borough Wards, *https://www.google.com/maps/d/viewer?gl=us&ptab=2&ie=UTF8&oe=UTF8&msa=0&mid=1X*

BK2S1kOEmVeKPx3eFs325ewt4E&ll=51.45172462249272%2C-0.19283999999993284&z=12

- [8] Average House Prices by Borough, Ward, MSOA & LSOA, London Data Store, *https://data.london.gov.uk/dataset/average-house-prices*