

# Capstone Project – Battle of the Neighborhoods

## Introduction

The HR department of a large multinational company is helping employees who are unfamiliar with London to find an area to live in. These employees might be transferring from another office outside of London/UK, or who lived in one area of London for a few months and now looking for a more long-term place.

The aim of this project is to provide insight into different districts in London based on similarities and dissimilarities, which an employee can use to focus their accommodation search on. This project will analyze the districts in the Greater London area, UK, and take in the key considerations the employees might have when looking for areas to live in.

## Data

The data used for this project will be acquired from publicly available sources, such as Wikipedia and/or government sites, as well as venue information through Foursquare API's search feature. The dataset for location points consists of postcodes in the Greater London area, postcode districts and area for each postcode and coordinate information for each postcode.

The focus will be on unique postcode district codes, they consist of groups of postcodes representing population centers and neighborhoods in each wider district within the Greater London area. Additional information for each postcode district, such as population, average housing price in the 2019, will also be used alongside local venue data to provide insight into the qualities and affordability of a neighborhood based on postcode district codes.

Data sources, cleaning and feature selection:

*Greater London postcodes, coordinates and population:*

<https://www.doogal.co.uk/UKPostcodesCSV.ashx?area=London>

The link above downloads the CSV file that contains a list of all postcodes in the Greater London area along with additional information for each. The relevant columns from the original CSV are:

Postcode – individual postcodes

In Use? – Whether the postcode is still in use or not. This column is used to filter out postcodes that are no longer in use.

County – Some of the postcodes are under more than one county, so this column will be used to filter out duplicates by taking the wider Greater London county only

Ward – The name of the ward the postcode is in

Postcode area – The area code for the postcode, typically the first few letters in a UK postcode before the first number

Population – residential population recorded for each postcode. As the focus of this project is in potential residential areas, postcodes without population data (i.e. NaN) or less than population of 10 will be assumed to have no or little residential property available and filtered out.

Latitude and Longitude – coordinates for each postcode

	Postcode	In Use?	County	Postcode area	Postcode district	Population	Latitude	Longitude
0	BR1 1AA	Yes	Greater London	BR	BR1	NaN	51.401546	0.015415
1	BR1 1AB	Yes	Greater London	BR	BR1	NaN	51.406333	0.015208
2	BR1 1AD	No	Greater London	BR	BR1	NaN	51.400057	0.016715
3	BR1 1AE	Yes	Greater London	BR	BR1	34.0	51.404543	0.014195
4	BR1 1AF	Yes	Greater London	BR	BR1	NaN	51.401392	0.014948

*Fig 1: Example of postcode data extracted from raw CSV*

Individual postcodes are then grouped by the 'Postcode district' column, which typically is the first section of an individual postcode. The new 'Postcode' column will contain the count of postcodes in a postcode district, the new 'Population' column will be the sum of all population in each postcodes and the average latitude and longitude were taken as the new coordinates for the postcode district under the new relevant columns. Note that as postcodes with less than 10 population are excluded before grouping, the average coordinates will gravitate less towards areas with little or no residential properties.

	Postcode district	Postcode area	County	Postcode	Population	Latitude	Longitude
0	BR1	BR	Greater London	920	55392.0	51.413635	0.021312
1	BR2	BR	Greater London	818	44401.0	51.387739	0.021657
2	BR3	BR	Greater London	879	46548.0	51.404795	-0.029514
3	BR4	BR	Greater London	316	19184.0	51.375202	-0.007653
4	BR5	BR	Greater London	813	45773.0	51.391719	0.102623

*Fig 2: Example grouping of postcodes by postcode districts*

Average sold house prices by postcode districts in 2019:

<https://landregistry.data.gov.uk/app/standard-reports/download-report?utf8=%E2%9C%93&report=avgPrice&areaType=county&area=GREATER+LONDON&aggregate=pcDistrict&period%5B%5D=2019&age=any>

Another feature that will be included is the relative cost of moving into a property in one neighborhood versus other neighborhoods. This project will use land registry data on sold house prices as a metric to compare difference in cost to move into a new area. The above link leads to the page where the data on the average sold prices for all properties in each postcode districts in 2019 can be downloaded.

The raw file includes a breakdown of type of property sold, number of sales of that type and the average prices. The property type will be used to determine if properties available in an area is high in flats/apartments, which tend to be smaller, or houses, which will be more suitable for families.

	Unnamed: 0	Detached	Sales	Semi-det	Sales.1	Terraced	Sales.2	Flat/mais	Sales.3	Overall average	Total sales
0	BR1	938364	81	564206	119	401489	269	329418	194	475197	663
1	BR2	872935	122	557675	211	444082	153	331216	263	506304	749
2	BR3	980797	59	679700	116	536987	206	361486	297	523146	678
3	BR4	753619	35	593491	132	503479	48	213450	36	544098	251
4	BR5	677339	70	446640	278	334842	158	246505	86	415007	592

Fig 3: Example of raw data from the average sold house prices by postcode districts

UK postcode area names:

[https://en.wikipedia.org/wiki/List\\_of\\_postcode\\_areas\\_in\\_the\\_United\\_Kingdom](https://en.wikipedia.org/wiki/List_of_postcode_areas_in_the_United_Kingdom)

Lastly, to provide some additional information on the wider area a postcode district falls under, the postcode area names will be provided. Here, the area name will be taken from the first letters of the postcode district (i.e. BR from BR1, BR2...), and matched with the list of names in the Wikipedia page link above.

	Postalcode area	Postcode area name	Code formation
0	AB	Aberdeen	
1	AL	St Albans	
2	B	Birmingham	
3	BA	Bath	
4	BB	Blackburn	

Fig 4: Example of list of Postcode area names from the Wikipedia page