House Prices within the catchment of Outstanding Primary Schools in London, UK

Gaurav Kukreja

March 07, 2020

1. Introduction

1.1 Background

London is the capital and largest city of the United Kingdom, standing on the River Thames in the south-east of England. It's considered financial capital of the world and has been termed as one of the most innovative, investment friendly and popular for working professionals. London exerts a considerable impact upon the arts, commerce, education, entertainment, fashion, finance, healthcare, media, professional services, research & development, tourism and transportation. It is a house of highly ranked institutions for higher education including Imperial College London, London School of Economics, University College London and King's College etc. It attracts people from all around the globe to work and settle. As a result, there is constant influx of migration every year from all age groups, from inside and outside UK.

8.8m

live in London, 13% of the UK's population. London is growing at twice the rate of the UK as a whole.

The Government has a motto of providing free quality school education for every child, rich or poor equally and fairly. There are about 30k state schools all over the country. In order to maintain healthy competition and desirability among those, an assessment is conducted regularly, and Ofsted rating is assigned to each school i.e Outstanding, Good, Requires Improvement and Inadequate. They are rated on the areas of (a) leadership and management; (b) quality of teaching, learning and assessment; (c) personal development, behaviour and welfare and finally (d) outcomes for children and learners.

1.2 Problem

It is a well-known fact that the education is a key determinant of lifetime earnings, poverty risk and prosperity. The low pay and employment status are clearly associated with lower levels of education. There are independent (aka private) schools in operation as well including Eton, Harrow School etc., that are quite prestigious and boasts of delivering leaders to the country. But they come at a hefty cost which most can't afford.

In that case, it is a duty of a parent to ensure that their children get the best possible education at a state school, ideally in an Outstanding Rated School which offers higher probability of success in life ahead, either in further education, employment or entrepreneurship. It all starts with foundation which commences at a Primary School. UK got approx. 20k of those. Out of about 2k outstanding rated primary state schools, 342 of those reside in London where most professionals work and live. As per usual rules, a first child from a household is admitted to a specific school if residence is within the catchment of a school, which makes it critically important to reside within a certain radius of a school

The problem here is (1) It comes at a cost. The house prices within catchment of outstanding schools can cost somewhere between 10-20% premium; (2) Quite difficult to filter out which outstanding school to target out of 342 in London, especially for new immigrants; (3) Which area may be suitable for interest of parents (do they like parks, cafes, pubs, restaurants, theatre etc.) and most important (4) Affordability

1.3 Solution

We will use Data Science methodology to balance all the problem areas above to find a best possible solution for parents depending on their interests and affordability. The aim is to provide a simple and clear options to support well-informed decision making of buying a house in right area for them. After all, this is one of most important investment of life.

2. Data acquisition and cleaning

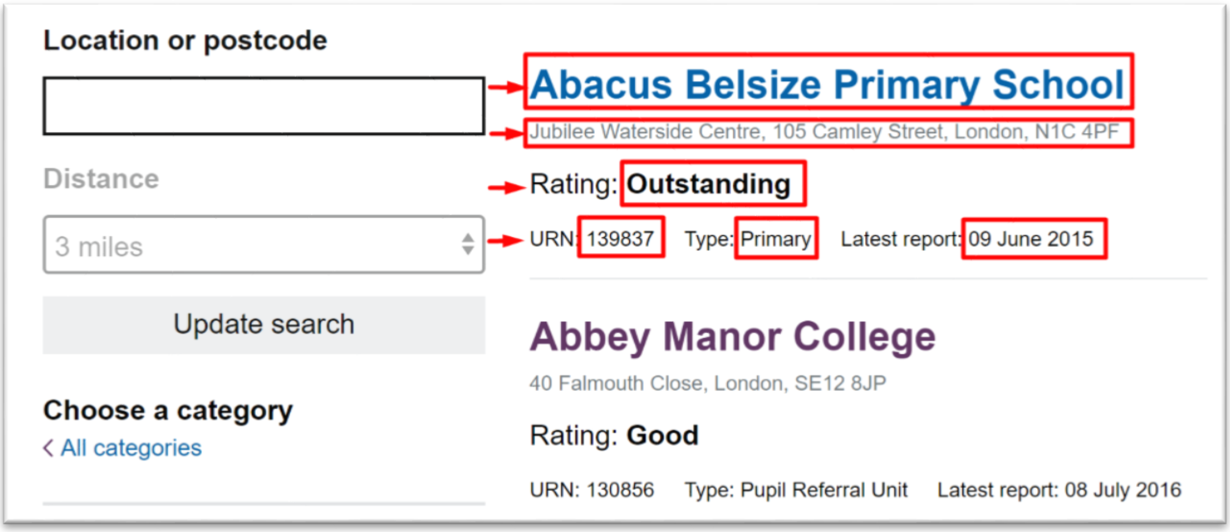
2.1 Data Sources

We will need below data sources available from different sources:

(a) Ofsted Report web portal

https://reports.ofsted.gov.uk/search?q=&location=&radius=&latest_report_date_start=&latest_report_date_end=&level_1_types=1

This shows details of all the schools across United Kingdom including State Schools, Independent Schools, Special Schools and Pupil Referral Units along with their relevant Ofsted ratings. The info is presented over web as html files on multiple pages. We will use BeautifulSoup Python library to pull it and convert into csv file.



(b) Wikipedia List of Areas in London

https://en.wikipedia.org/wiki/List_of_areas_of_London

This has a web table of Location, Borough, Town and PostCodes mainly. We will use pandas library to read it and convert it into dataframe.

Location	London borough	Post town	Postcode district	Dial code	OS grid ref
Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	020	TQ205805
Addington	Croydon[8]	CROYDON	CR0	020	TQ375645
Addiscombe	Croydon[8]	CROYDON	CR0	020	TQ345665
Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728

(c) Property Data

https://propertydata.co.uk/cities/london

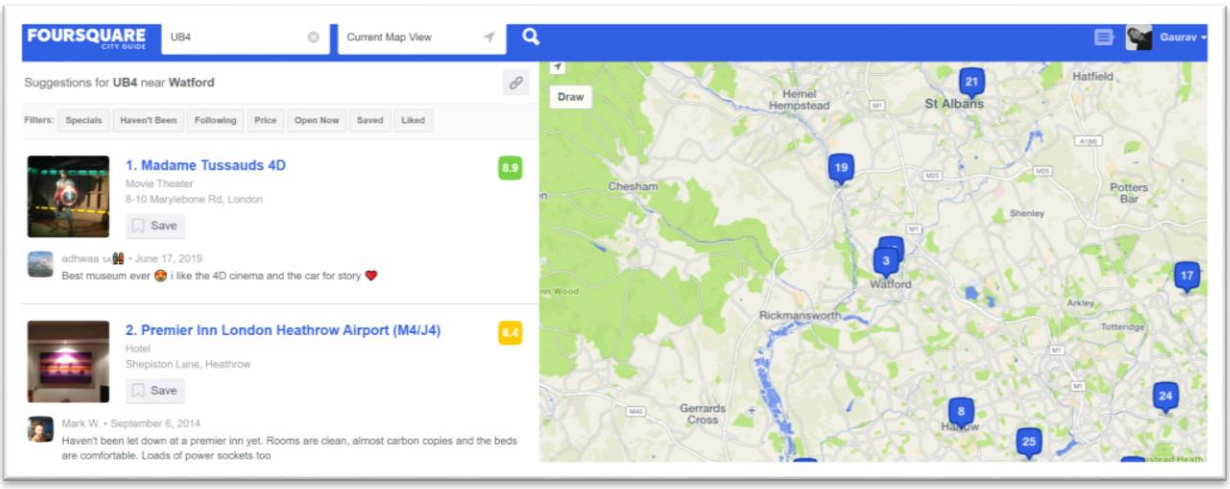
This produces a table with useful property price info at a postcode level. We will use pandas library to read and convert into dataframe.

Area	Avg yield	Avg price	£/sqft	5yr +/-	Explore data
BR1	3.7%	£447,159	£466	+25%	Explore data
BR2	3.1%	£513,115	£469	+27%	Explore data
BR3	3.7%	£456,224	£488	+26%	Explore data
BR4	-	£601,225	£460	+24%	Explore data

(d) Foursquare Venue Details

https://developer.foursquare.com/

This is an excellent developer portal to search or explore venues (restaurants, bars, hotels, parks etc.) near any location within certain radius of a location (by postcode, town, borough etc.). It also dispenses category of venues which is quite useful for our analysis. We will use developer API to extract the relevant info for our project.



2.2 Data Cleaning

Data downloaded or scraped from first 3 sources above were combined into 1 table after many adjustments to each set.

In Ofsted School Report data, there were several irrelevant information. For example, we are only aiming for State Primary schools thus we removed Independent & Special Schools as well as Pupils Referral Units. We also like to see Ofsted assessment for recent years (2018 or onwards) as assessment conducted many years ago may not sustain today. We also filtered down on Outstanding Schools only. Since other set of data has PostCode as common denominator, we combined multiple rows at that level, picking info from School addresses.

In property pricing data, we only need Postcode and Price, thus removed rest of the information. We had to remove commas and currency sign from numbers to make it work for further analysis.

In Wikipedia table for London areas, we had to clean up a lot to make it usable. We removed the reference #s given next to Boroughs in square brackets. Some postcodes had unnecessary spaces or no space, when needed, so adjusted that. Many postcodes were linked to same location, so we created multiple lines to join it with other set of data tables easily.

2.3 How will that be used

After combining 3 tables into 1, we would come down to required lines of data only for unique Postcodes/areas which we will search to get location coordinates using geocoder library and arcgis maps provider.

We will use those coordinates to get venues within a certain radius (1000m) using Foursquare API which will be used to cluster those locations into different labels based on category of venues nearby (Hotel, Bars, Gyms, Theatres etc.). We intend to use k-means algorithm which is most common cluster method of unsupervised learning. We will then display those clusters on a map using folium library.

We will also use pyplot chart to represent average house prices (adding 20% on top as premium for the school catchments) of each area within each cluster labels which will help parents to view different price points within their preferred cluster label (which is basically their interest types like Social Venues vs. Pub & Restaurants vs. Golf vs. Theatres or combination of those depending on what is produced by cluster labels using this algorithm). An example visualization of that is below.

3. Data Wrangling & Exploratory Data Analysis

3.1 No. of Schools, missing Data & extracting required Info

After converting html files into csv using BeautifulSoup library and then into dataframe using Pandas, the first step is to get an idea of number of schools across London and which ones are missing any critical value for analysis.

As we can see, we found NaN values under 2 important fields i.e Rating of School (Outstanding vs. Good etc.) and Type of School (Primary vs. Secondary etc.). Since we are specifically after Outstanding Primary Schools, we decided to drop these rows from the data taking more conservative approach to ensure we don't get any false positive.

```
Name
False    2375
Name: Name, dtype: int64

Address
False    2375
Name: Address, dtype: int64

Rating
False    2084
True      291
Name: Rating, dtype: int64

URN
False    2333
True       42
Name: URN, dtype: int64

Type
False    2333
True       42
Name: Type, dtype: int64

Latest_Report
False    2333
True       42
Name: Latest_Report, dtype: int64
```

We would also extract the first 3 letters of the postcodes from each address (see below). This is in order to combine other data tables to come up to House Prices and Location details.

	Name	Address	Rating	Type	Latest_Report	PostCode
0	Abacus Belsize Primary School	Jubilee Waterside Centre, 105 Camley Street, L...	Outstanding	Primary	09 June 2015	N1C
1	Abbey Manor College	40 Falmouth Close, London, SE12 8JP	Good	Pupil Referral Unit	08 July 2016	SE12
3	Abbotsbury Primary School	Abbotsbury Road, Morden, Surrey, SM4 5JS	Good	Primary	21 November 2017	SM4
4	Adamsrill Primary School	Adamsrill Road, Sydenham, London, SE26 4AQ	Good	Primary	07 February 2020	SE26
5	Addison Primary School	Addison Gardens, Blythe Road, London, W14 0DT	Good	Primary	12 September 2016	W14

We filtered out all the data in order to land onto Rating = 'Outstanding' and Type = 'Primary' which brought us to 75 rows (out of 2375 initial data). Since we had some outstanding schools within the same Postcode, we merged those rows with multiple school names separated by commas, reaching up to 62 rows. See final data below from first table.

```
No. of Schools that had assessment in 2018, 2019 or 2020:  1116
ONLY Primary Schools out of those:  771
ONLY Outstanding Schools out of those:  75
Grouping by PostCode:  62
```

	PostCode	Rating	Type	Name
0	BR1	Outstanding	Primary	Burnt Ash Primary School
1	BR3	Outstanding	Primary	Clare House Primary School,Harris Primary Acad...
2	BR5	Outstanding	Primary	Crofton Junior School
3	CR0	Outstanding	Primary	Chestnut Park Primary School,Harris Primary Ac...
4	CR5	Outstanding	Primary	Woodcote Primary School

3.2 Areas of London

The table extracted from Wikipedia on London areas had lot of anomalies including reference link #s within squared brackets, multiple postcodes within same rows & unrequired data columns. We used Pandas library to read html file in order to retrieve the table.

	Location	London borough	Post town	Postcode district	Dial code	OS grid ref
0	Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	020	TQ465785
1	Acton	Ealing, Hammersmith and Fulham [8]	LONDON	W3, W4	020	TQ205805
2	Addington	Croydon [8]	CROYDON	CR0	020	TQ375645
3	Addiscombe	Croydon [8]	CROYDON	CR0	020	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	020	TQ478728

We cleaned all it up to reach to below final table in order to be able to join it easily with School Data by Postcodes.

	PostCode	Location	Borough
0	BR1	Bromley,Plaistow,Sundridge,Widmore (also Widmo...	Bromley/Lewisham
1	BR2	Hayes,Keston,Leaves Green,Southborough	Bromley
2	BR3	Beckenham,Bickley,Bromley Common,Eden Park,Elm...	Bromley
3	BR4	Coney Hall,West Wickham	Bromley
4	BR5	Ruxley,Upper Ruxley,Derry Downs,Orpington,Pett...	Bexley/Bromley

3.3 Average House Prices

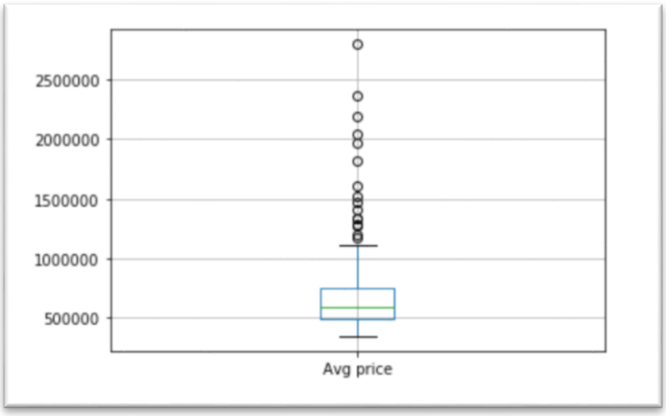
From a property price portal using Pandas to read html file, we managed to get average prices at a Postcode level (our common field in all tables to join) but lot of fields are redundant for us thus dropped and adjusted the current figures to better adjust later.

	Area	Avg yield	Avg price	£/sqft	5yr +/-	Explore data
0	BR1	3.7%	£447,159	£466	+25%	Explore data
1	BR2	3.1%	£513,115	£469	+27%	Explore data
2	BR3	3.7%	£456,224	£488	+26%	Explore data
3	BR4	-	£601,225	£460	+24%	Explore data
4	BR5	3.6%	£436,613	£426	+26%	Explore data

	PostCode	Avg price
0	BR1	447159
1	BR2	513115
2	BR3	456224
3	BR4	601225
4	BR5	436613

Since the property prices within catchment of outstanding schools are usually found at a premium price within the range of 10-20%, we will add 20% to these prices for our analysis and also then look at box plot just to get an idea. The first quartile ranges from £250 – 500k, 2nd one from £500-600k, 3rd one from £600-775k and 4th one is between £775-1.2M. There are many outliers as well.

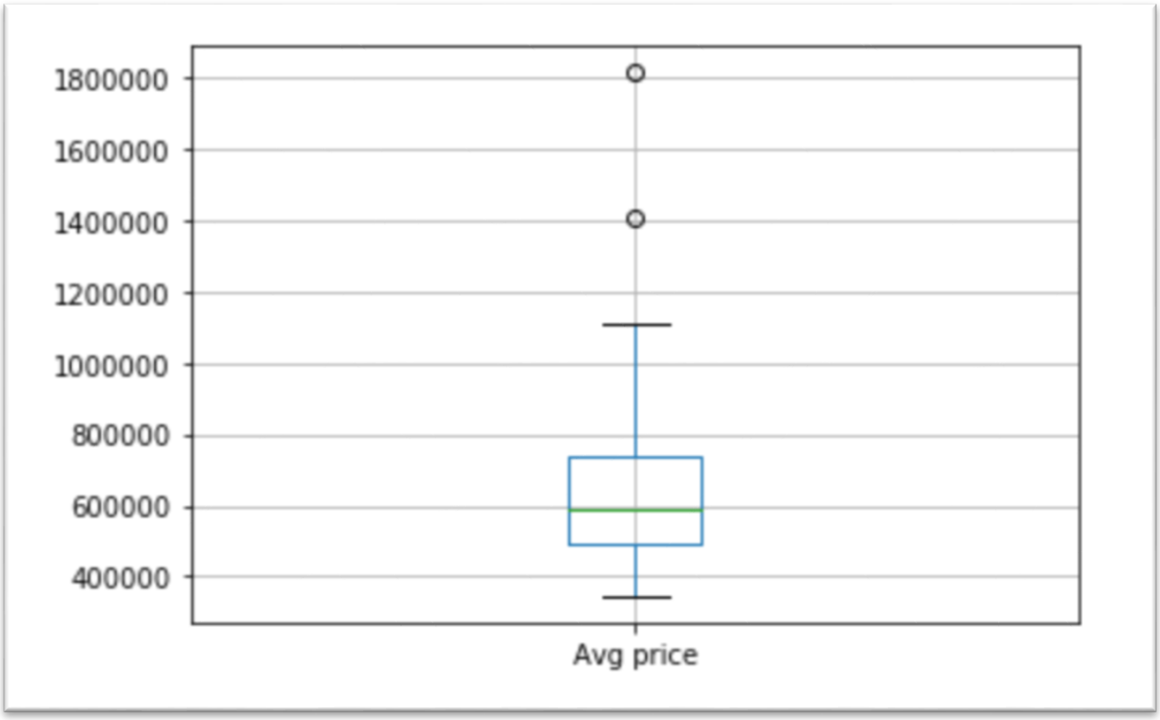
	PostCode	Avg price
0	BR1	537722.4
1	BR2	617064.0
2	BR3	541131.6
3	BR5	528127.2
4	BR6	669649.2



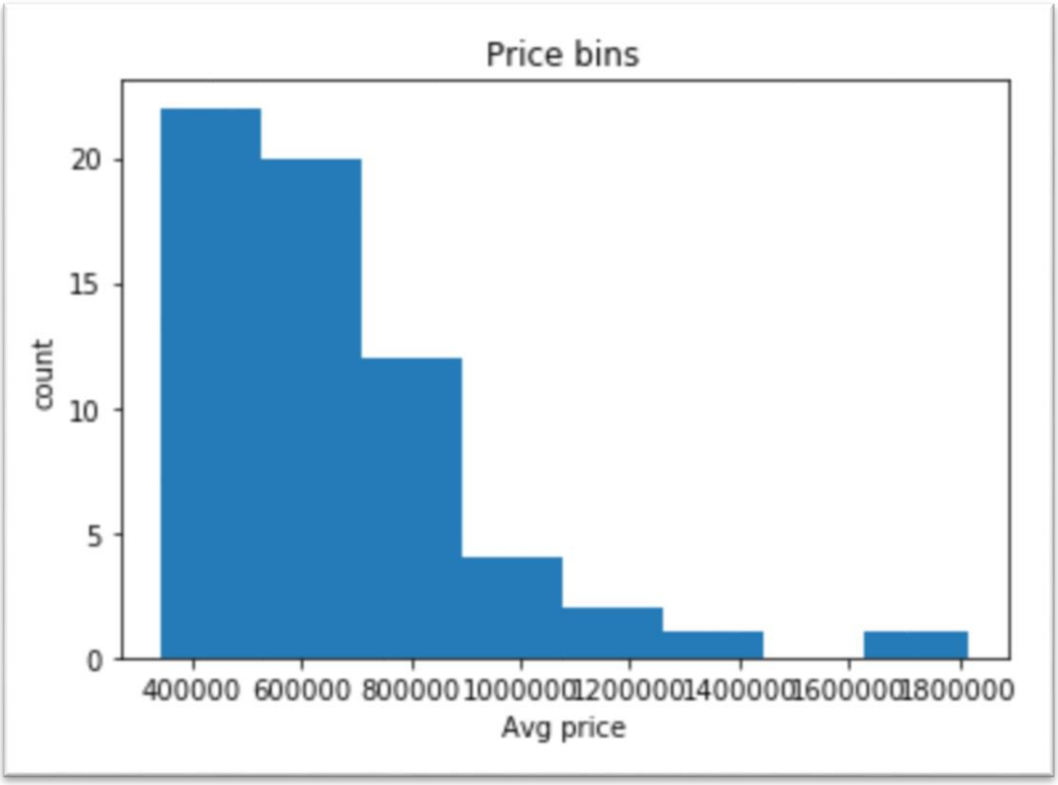
Now, we have 3 tables above, ready to join. We took Outstanding School one as a base and added information from other 2 tables based on PostCodes. See below.

	PostCode	Rating	Type	Name	Location	Borough	Avg price
0	BR1	Outstanding	Primary	Burnt Ash Primary School	Bromley,Plaistow,Sundridge,Widmore (also Widmo...	Bromley/Lewisham	447159
1	BR3	Outstanding	Primary	Clare House Primary School,Harris Primary Acad...	Beckenham,Bickley,Bromley Common,Eden Park,Elm...	Bromley	456224
2	BR5	Outstanding	Primary	Crofton Junior School	Ruxley,Upper Ruxley,Derry Downs,Orpington,Pett...	Bexley/Bromley	436613
3	CR0	Outstanding	Primary	Chestnut Park Primary School,Harris Primary Ac...	Addington,Addiscombe,Coombe,Croydon,Forestdale...	Croydon/Sutton	343852
4	CR5	Outstanding	Primary	Woodcote Primary School	Coulsdon,Old Coulsdon	Croydon	510381

Now since the house prices will only be relevant for selected postcodes which were available in previous tables (only where Outstanding primary schools reside), we notice different Boxplot.



We can also look at histogram bins for the potential Average house prices.



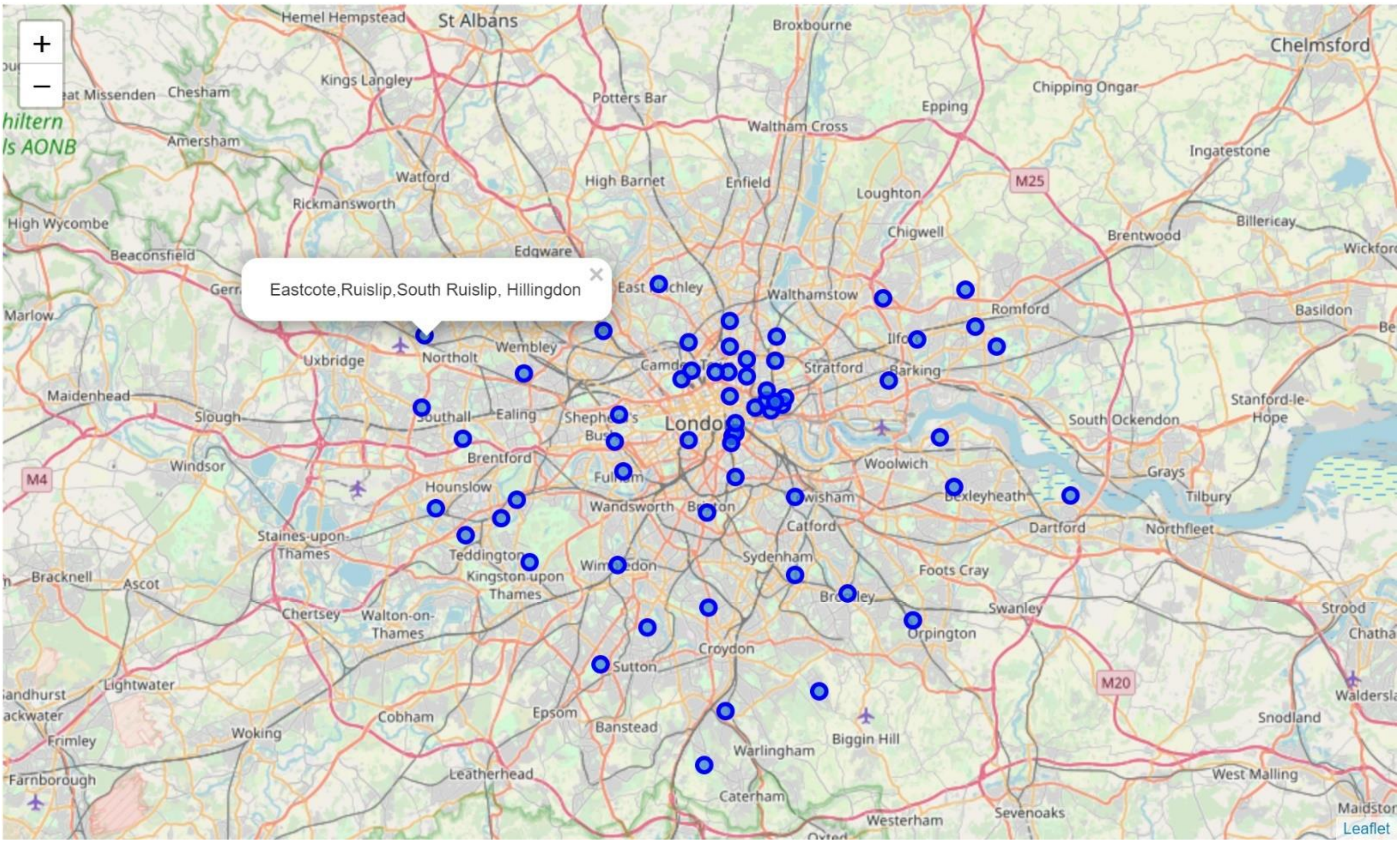
3.4 Location Coordinates & Visualization on a Map

Since we are dealing with location data and need to find out type of venues within the proximity of these areas, we will need latitude and longitude for precise results. That is a key element in order to use Foursquare API to explore various areas and segment them appropriately. I decided to use geocoder library and arcgis maps provider to find the coordinates.

	PostCode	Rating	Type	Name	Location	Borough	Avg price	Latitude	Longitude
0	BR1	Outstanding	Primary	Burnt Ash Primary School	Bromley,Plaistow,Sundridge,Widmore (also Widmo...	Bromley/Lewisham	447159	51.404240	0.015425
1	BR3	Outstanding	Primary	Clare House Primary School,Harris Primary Acad...	Beckenham,Bickley,Bromley Common,Eden Park,Elm...	Bromley	456224	51.415095	-0.035403
2	BR5	Outstanding	Primary	Crofton Junior School	Ruxley,Upper Ruxley,Derry Downs,Orpington,Pett...	Bexley/Bromley	436613	51.388315	0.078482

After careful selection of all necessary variables, it's time to visualise the locations on a map for which I decided to use Folium library which is quite popular for this purpose. Folium builds on the data wrangling strengths of the Python ecosystem and the mapping strengths of the leaflet.js library. Manipulate your data in Python, then visualize it in on a Leaflet map via folium. Using it, I created map of London, UK with locations superimposed on top. This step is quite important to get an idea whether the coordinates make sense against location/postcodes.

Alongside visual representation, it’s always wise to verify locations on a map by exploring popups randomly as well like above and proceed to next steps once we are comfortable.



3.5 Location Exploration using Foursquare API

In order to explore PostCodes with Outstanding primary Schools, Foursquare was used. I decided to limit the number of venues to 100 and radius of 1000m. While exploring the first PostCode, it was realised that required info about nearby venues is available under ‘Item’ key which needs to be used to obtain relevant info.

That was used to explore all areas and converted into Pandas dataframe. We only picked useful variables including Category of venues/places which is to be used in order to cluster and segment all the Postcodes.

That was combined with original dataset to further analyse to support our model evaluation. The shape of dataframe was found to be 3754 rows and 8 columns. See below.


```
{'meta': {'code': 200, 'requestId': '5e6234059388d7001b2aa05f'},
 'response': {'suggestedFilters': {'header': 'Tap to show:',
   'filters': [{'name': 'Open now', 'key': 'openNow'}]},
  'headerLocation': 'Bromley Town',
  'headerFullLocation': 'Bromley Town, London',
  'headerLocationGranularity': 'neighborhood',
  'totalResults': 56,
  'suggestedBounds': {'ne': {'lat': 51.413240412000086,
    'lng': 0.029825275198514866},
   'sw': {'lat': 51.39524039400006, 'lng': 0.0010247248015855422}},
  'groups': [{'type': 'Recommended Places',
    'name': 'recommended',
    'items': [{'reasons': {'count': 0,
       'items': [{'summary': 'This spot is popular',
         'type': 'general',
         'reasonName': 'globalInteractionReason'}]},
      'venue': {'id': '4ecd6c670aafd1868b867683',
       'name': 'Barrel & Horn',
       'location': {'address': '204-206 High Street',
        'lat': 51.40588697867192,
        'lng': 0.013701243249897743,
        'labeledLatLngs': [{'label': 'display',
          'lat': 51.40588697867192,
          'lng': 0.013701243249897743}],
        'distance': 218,
        'postalCode': 'BR1 1PW',
        'cc': 'GB',
        'city': 'Kent',
        'state': 'Kent',
        'country': 'United Kingdom',
        'formattedAddress': ['204-206 High Street',
         'Kent',
         'BR1 1PW',
         'United Kingdom']},
       'categories': [{'id': '4bf58dd8d48988d116941735',
         'name': 'Bar',
         'pluralName': 'Bars',
         'shortName': 'Bar',
         'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/nightlife/pub_',
          'suffix': '.png'},
```

(3754, 8)

	PostCode	School(s)	PostCode Latitude	PostCode Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	BR1	Burnt Ash Primary School	51.40424	0.015425	Barrel & Horn	51.405887	0.013701	Bar
1	BR1	Burnt Ash Primary School	51.40424	0.015425	unico	51.402189	0.015725	Ice Cream Shop
2	BR1	Burnt Ash Primary School	51.40424	0.015425	Apple Bromley	51.402940	0.016252	Electronics Store
3	BR1	Burnt Ash Primary School	51.40424	0.015425	Marks & Spencer	51.404017	0.015718	Department Store
4	BR1	Burnt Ash Primary School	51.40424	0.015425	Bromley Picturehouse	51.406740	0.012303	Indie Movie Theater

We also grouped the result by Postcodes to get an idea on number of venues found under parameter defined. This also supports to understand if we are getting too many or too less of results and adjust parameters accordingly, if needed.

PostCode	School(s)	PostCode Latitude	PostCode Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
BR1	58	58	58	58	58	58	58
BR3	43	43	43	43	43	43	43
BR5	15	15	15	15	15	15	15
CR0	8	8	8	8	8	8	8
CR5	4	4	4	4	4	4	4

As we know, clustering of dataset is based on different features of a categorical variable (Postcode in this case). Here those different features are type of venues (aka Venue Categories). I used onehot encoding to count number of venues within certain categories for a specific Postcode. It was further used to determine the mean of each category i.e. to figure out proportion or frequency of a specific venue category under a Postcode out of all. Then to analyse further, I decided to view top 5 venue categories of each Postcode (like below).

It's always useful to analyse this result with plain eye randomly to get an idea of (1) whether this make sense based on your experience; (2) how similar or dissimilar each group could be (what cluster they may fall under) and (3) to get some ideas on naming those clusters or group of Postcodes.

```
----BR1----
                venue  freq
0                 Pub  0.10
1         Coffee Shop  0.09
2      Clothing Store  0.09
3  Gym / Fitness Center  0.05
4    Indian Restaurant  0.03


----BR3----
           venue  freq
0  Grocery Store  0.09
1    Coffee Shop  0.07
2  Train Station  0.07
3           Park  0.05
4    Pizza Place  0.05


----BR5----
                     venue  freq
0              Supermarket  0.20
1                      Pub  0.13
2  Mediterranean Restaurant  0.13
3            Train Station  0.07
4       American Restaurant  0.07
```

In order to achieve next step of applying clustering algorithm, we need to switch this data into a dataframe so I converted this information as most common venues for each Postcode. I also limited it to most common 10 venues for each. See below.

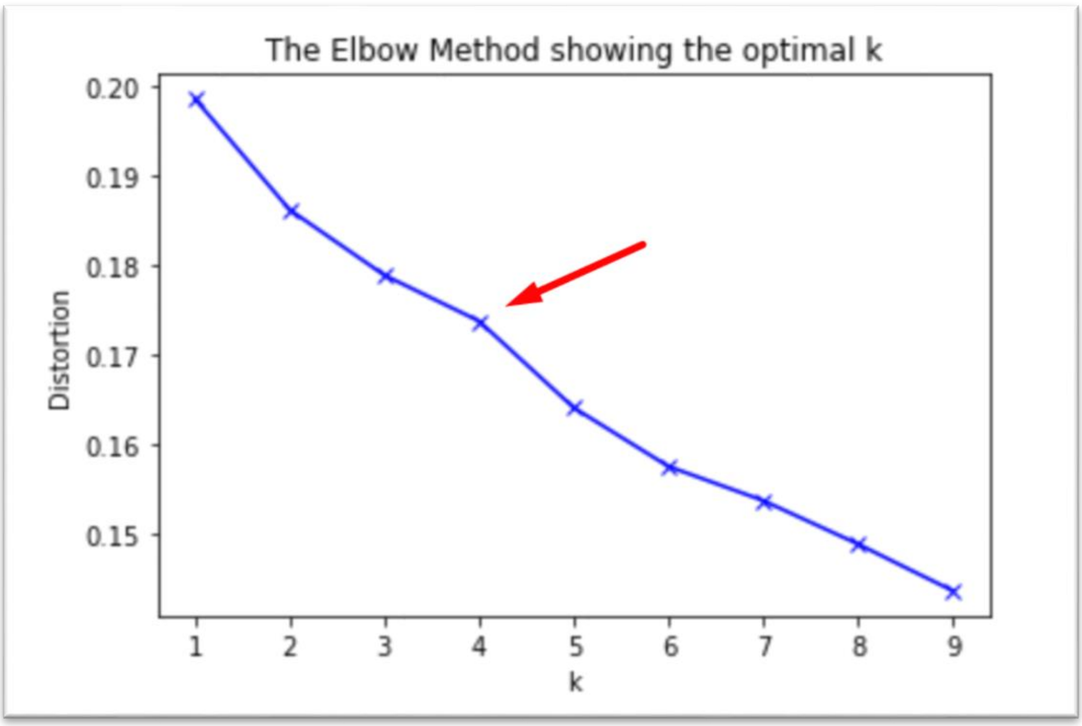
	PostCode	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	BR1	Pub	Clothing Store	Coffee Shop	Gym / Fitness Center	Café	Pizza Place	Indian Restaurant	Bar	Burger Joint	Supermarket
1	BR3	Grocery Store	Coffee Shop	Train Station	Pizza Place	Park	Bar	Tram Station	Pub	Tapas Restaurant	Bookstore
2	BR5	Supermarket	Pub	Mediterranean Restaurant	Coffee Shop	Indian Restaurant	American Restaurant	Fish & Chips Shop	Train Station	Thai Restaurant	Sushi Restaurant
3	CR0	Tram Station	Convenience Store	Grocery Store	Supermarket	Home Service	Furniture / Home Store	Gas Station	Farm	Entertainment Service	Ethiopian Restaurant
4	CR5	Fish & Chips Shop	Construction & Landscaping	Restaurant	Café	Filipino Restaurant	Ethiopian Restaurant	Event Space	Falafel Restaurant	Farm	Farmers Market

4. Clustering

There are various types of clustering algorithms such as partitioning, hierarchical and density-based clustering. Our dataset is unlabelled, and we need to divide them into different similar clusters in order to support understanding. For simplicity, performance and being inexpensive, I decided to apply K-means clustering. It divides the data into k non-overlapping subsets or clusters without any cluster internal structure or labels. This means, it's an unsupervised algorithm.

4.1 K-means Clustering

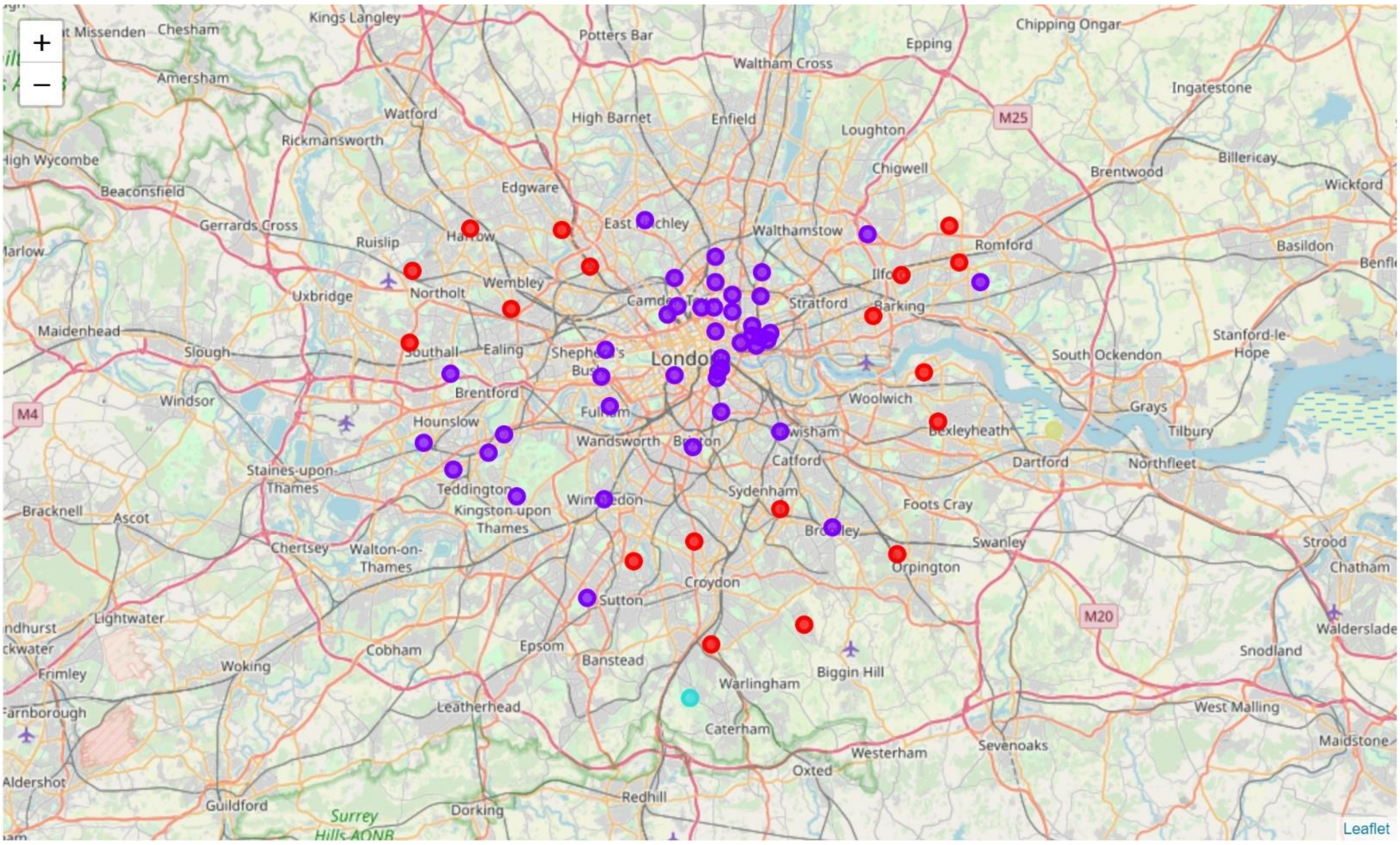
K-Means algorithm was applied to segment Postcodes into pre-determined 4 clusters. The optimality was also ensured using elbow method as we tried with various values of K.



The clusters are assigned to each Postcode as can be seen below. This needs to be examined further which we will do in the next steps.

	PostCode	Name	Location	Borough	Avg price	Latitude	Longitude	Cluster Labels	1st Most Common Venue
0	BR1	Burnt Ash Primary School	Bromley,Plaistow,Sundridge,Widmore (also Widmo...	Bromley/Lewisham	447159	51.404240	0.015425	2	Pub
1	BR3	Clare House Primary School,Harris Primary Acad...	Beckenham,Bickley,Bromley Common,Eden Park,Elm...	Bromley	456224	51.415095	-0.035403	1	Grocery Store
2	BR5	Crofton Junior School	Ruxley,Upper Ruxley,Derry Downs,Orpington,Pett...	Bexley/Bromley	436613	51.388315	0.078482	1	Supermarket
3	CR0	Chestnut Park Primary School,Harris Primary Ac...	Addington,Addiscombe,Coombe,Croydon,Forestdale...	Croydon/Sutton	343852	51.344980	-0.012265	1	Tram Station

Let’s visualize the clusters on a map (see next page).



4.2 Examine Cluster

We will now examine top results under each cluster (from 0 to 3) and name them based on type of venues in that segment. See below.

1st Cluster = Cluster Label 0 = 'Household Convenience'

	PostCode	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
1	BR3	0	Grocery Store	Coffee Shop	Train Station	Pizza Place	Park	Bar	Tram Station	Mediterranean Restaurant	Tapas Restaurant	Bookstore
2	BR5	0	Supermarket	Mediterranean Restaurant	Pub	Thai Restaurant	Sushi Restaurant	American Restaurant	Indian Restaurant	Coffee Shop	Turkish Restaurant	Fish & Chips Shop
3	CR0	0	Home Service	Tram Station	Convenience Store	Supermarket	Gas Station	Grocery Store	Cycle Studio	Entertainment Service	Food	Flower Shop
5	CR7	0	Grocery Store	Hotel	Bus Stop	Supermarket	Auto Garage	Optical Shop	Coffee Shop	Fast Food Restaurant	Ethiopian Restaurant	Event Space
6	CR8	0	Train Station	Platform	Cricket Ground	Breakfast Spot	Grocery Store	Pizza Place	Golf Course	Farmers Market	Entertainment Service	Ethiopian Restaurant

2nd Cluster = Cluster Label 1 = 'Pubs, Café, Restaurants & Hotel'

	PostCode	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	BR1	1	Pub	Clothing Store	Coffee Shop	Gym / Fitness Center	Park	Supermarket	Bar	Pizza Place	Burger Joint	Furniture / Home Store
9	E12	1	Coffee Shop	Hotel	Indian Restaurant	Pub	Grocery Store	Park	Pizza Place	Bar	Korean Restaurant	Bakery
10	E13	1	Pub	Grocery Store	Hotel	Coffee Shop	Gym / Fitness Center	Sandwich Place	Park	Bakery	Café	Indian Restaurant
11	E14	1	Pub	Coffee Shop	Park	Grocery Store	Brewery	Pizza Place	Café	Gym / Fitness Center	Sandwich Place	Beer Bar
12	E15	1	Coffee Shop	Café	Pub	Indian Restaurant	Flea Market	Restaurant	Middle Eastern Restaurant	Hotel	Brewery	Beer Bar

3rd Cluster = Cluster Label 2 = 'Green Area & variety Cuisines'

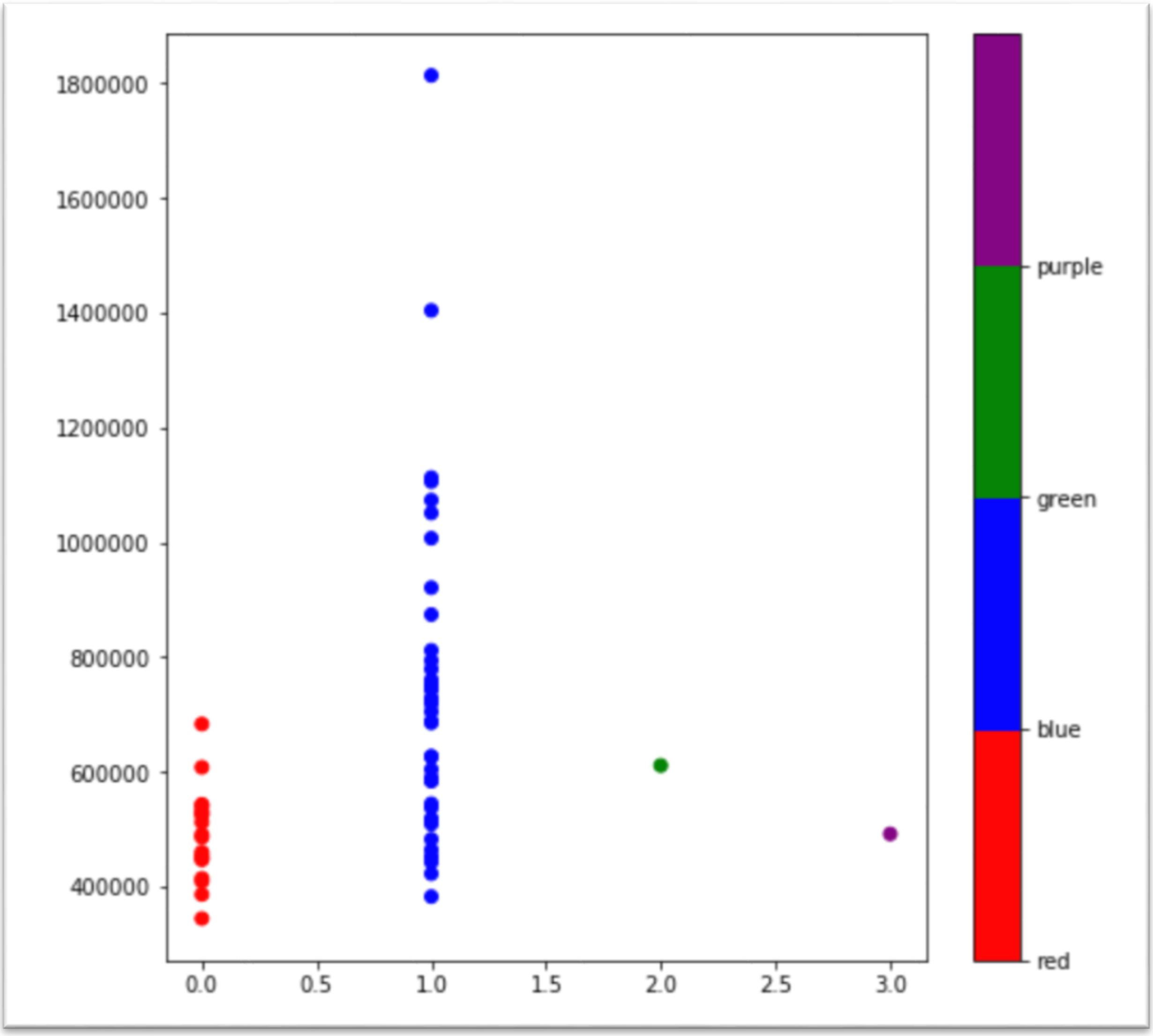
	PostCode	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	CR5	2	Fish & Chips Shop	Construction & Landscaping	Restaurant	Café	Filipino Restaurant	Ethiopian Restaurant	Event Space	Falafel Restaurant	Farm	Farmers Market

4th Cluster = Cluster Label 3 = 'Mixed Social'

	PostCode	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
7	DA15	3	Grocery Store	Hotel	Restaurant	Fast Food Restaurant	English Restaurant	Entertainment Service	Ethiopian Restaurant	Event Space	Falafel Restaurant	Farm

5. Result

Now based on the cluster labels (x axis below), we can visualize the house prices range (y axis below) against each of the cluster. As we can notice, 2nd cluster has most of the Postcodes or in other words, most postcodes fall within this cluster and subsequently most options of price range as well. We also notice that there is only 1 postcode each under cluster 3rd and 4th. Now based on interest of the parents i.e. which type of neighbourhood they like, an area can be targeted to buy a house also keeping in mind their affordability. In other words, this Data Science project supports to a great deal in taking a well-informed decision before making one of the most important investment in life.



6. Discussion

6.1 Observations

It can be noticed that the least expensive average price of a house can be obtained within Cluster 1 under £300k which represents Cluster 1 i.e. Household Convenience. That may not be very popular area but should meet the needs well if budget is a constraint and best primary education for the child is number 1 priority. The house prices range up to just under £700k within the same cluster if budget allows that interest lies for specific kind of venues within the same segment. What we can notice overall is that there are enough options to choose areas from.

The 2nd cluster offers most of the options for areas, but wider price range as well starting from about £300k. This cluster boasts of many restaurants, pubs, cafes, hotels, fitness centres and many other desirable venues. It may be most popular among parents aged between 25-34. The price range goes up to just under £1.8M.

The 3rd and 4th cluster only have 1 postcode/area each and price point for houses around £650k and 500k respectively.

6.2 Recommendations

There can't be a specific recommendation for parents. There seem to be plenty of options available for anyone's budget. It eventually comes down to the interest and priorities of everyone. By using this, they can at least assure their child(ren) best primary state education possible. Another point to consider may also be to imagine where they may want their child(ren) to study for next level i.e. secondary education and if they specifically like to buy within the catchment of Outstanding secondary schools as well which is out of scope for this Data Science project.

7. Conclusion

In this Data Science project, we tried to gauge if it's feasible to achieve a balance between affordability, outstanding primary education for children and preference over certain kind of area. We conclude that first those parents who seek to live in an area having a balance of pubs, bars, theatres, convenience etc., there are plenty of options stating from quite reasonable average property prices. However, desire to live in mixed social or green areas have very limited choices. Overall conclusion still favours Parents with multiple options within their budget thus London wins to have balance of life.