

# Battle of the Neighborhoods

## Finding the best neighborhood for opening an Arts & Crafts store in London using Data Science

This project aims to use Data Science concepts learned in the IBM Data Science Professional Course to solve a business problem.

In particular, we will use Foursquare location data and clustering of venue information to determine what might be the ‘best’ borough in London to open an Arts & Crafts store.

In this project we will proceed in a step by step manner, from problem definition to a conclusion that can be leveraged by the business stakeholders to make their decisions.

## Table of Contents

1. Introduction
2. Data Overview
3. Methodology
4. Results
5. Discussion
6. Conclusion

## 1. Introduction

The city of London is famous for many things – its magnificent ancient buildings, its nightlife, its multiculturalism, its role in international trade,

its growing foodie scene, its being the setting of many films and novels, and so on.

As one of the world's creative capitals it is not surprising that London:

- is home to more than 1.000 art galleries, many of which are rated amongst the best in the world;
- has also been the home of and inspiration for countless artists and creative movements;
- plays hosts to multiple world-renowned art schools, with students that travel across the globe to join their varied and inspiring courses.

The objective of this project is to determine which might be the 'best' borough in London for an entrepreneur to open an Arts & Crafts store for selling artist materials and where to organize creative workshops.

## 2. Data Overview

The data we will use to conduct our analysis comes from multiple sources:

- a list of boroughs in London (via Wikipedia. See link [https://en.wikipedia.org/wiki/List\\_of\\_London\\_boroughs](https://en.wikipedia.org/wiki/List_of_London_boroughs)). The Wikipedia page provides different information about each borough, including its name, its geographical location, its area, its population (2013 estimation) and so on. Since the data is not in a format that is suitable for direct analysis, scraping of the data was done from this site;
- a population projection for year 2020 for each London borough (via London Datastore, a free and open data-sharing portal where

anyone can access data relating to the capital. See link <https://data.london.gov.uk/dataset/london-borough-profiles>). The file was in Excel format, so we could load it directly;

- GIS boundaries for each borough in London (via London Datastore. See link <https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london>). The file was in ESRI format, so we could load it directly;
- venue data pertaining to Arts & Crafts store, but also Art Schools and Art Museums (via Foursquare). The venue data will help us find which borough is best suitable to open our store.

### 3. Methodology

After all the data was collected and put into data frames, cleansing and merging of the data was required to start the process of analysis.

The following image shows the first rows of the data frame created from the Wikipedia page:

	Borough	Inner	Status	Local authority	Political control	Headquarters	Area (sq mi)	Population (2013 est) [1]	Co-ordinates	Nr. in map	Population(2011 est)	Nr. inmap
0	Barking and Dagenham [note 1]	NaN	NaN	Barking and Dagenham London Borough Council	Labour	Town Hall, 1 Town Square	13.93	194352.0	51°33'39"N 0°09'21"E / 51.5607°N 0.1557°E	25.0	NaN	NaN
1	Barnet	NaN	NaN	Barnet London Borough Council	Conservative	Barnet House, 2 Bristol Avenue, Colindale	33.49	369088.0	51°37'31"N 0°09'06"W / 51.6252°N 0.1517°W	31.0	NaN	NaN
2	Bexley	NaN	NaN	Bexley London Borough Council	Conservative	Civic Offices, 2 Watling Street	23.38	236687.0	51°27'18"N 0°09'02"E / 51.4549°N 0.1505°E	23.0	NaN	NaN
3	Brent	NaN	NaN	Brent London Borough Council	Labour	Brent Civic Centre, Engineers Way	16.70	317264.0	51°33'32"N 0°16'54"W / 51.5588°N 0.2817°W	12.0	NaN	NaN
4	Bromley	NaN	NaN	Bromley London Borough Council	Conservative	Civic Centre, Stockwell Close	57.97	317899.0	51°24'14"N 0°01'11"E / 51.4039°N 0.0198°E	20.0	NaN	NaN

When getting the data from Wikipedia, the population of each borough was an estimation relative to year 2013, so we used the data gathered from London Datastore to have a population projection for the year 2020. The following image shows the first rows of the data frame after merging and cleansing:

	Borough	Latitude	Longitude	Population Projections (2020)
0	Barking and Dagenham	51.5607	0.1557	218800
1	Barnet	51.6252	-0.1517	403000
2	Bexley	51.4549	0.1505	247100
3	Brent	51.5588	-0.2817	342400
4	Bromley	51.4039	0.0198	332000

Next, we used the Foursquare API to get a list of all the venues in London belonging to specific categories. In addition to Arts & Crafts Store we included: Art Gallery, Art Museum, College Arts Building and Art Studio. In fact, in deciding where to open a new Arts & Crafts store, we think it would be convenient to consider the presence, in that particular borough, of other activities related to Art that can contribute to the success of our store.

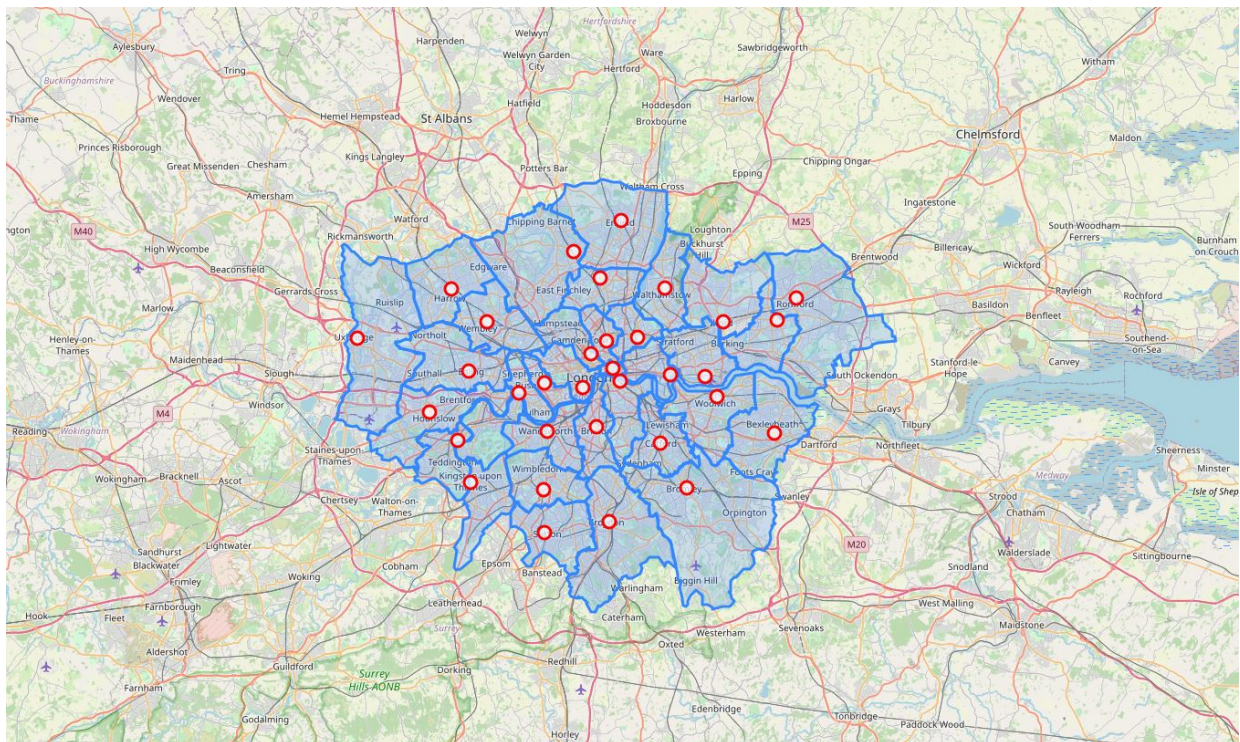
The following code snippet shows the venue categories we searched for using the Foursquare API:

```
# List of venue categories to search for (extracted from https://developer.foursquare.com/docs/build-with-foursquare/categories/)
categoryIDs = {
    'Art Gallery': '4bf58dd8d48988d1e2931735',
    'Art Museum': '4bf58dd8d48988d18f941735',
    'College Arts Building': '4bf58dd8d48988d199941735',
    'Art Studio': '58daa1558bbb0b01f18ec1d6',
    'Arts & Crafts Store': '4bf58dd8d48988d127951735'
}
```

We used the boundaries of each borough to assign each venue returned by Foursquare API to the borough it's located into. In fact, if you query Foursquare to search for venues nearby the coordinates of each borough you can't tell if all the venues returned are located into that specific borough; it can also happen that some venues are outside all London borough. If a venue is not assigned to any borough then it is ignored. The following image shows the first rows of the geodataframe containing all the GIS boundaries of each London borough:

	NAME	GSS_CODE	HECTARES	NONLD_AREA	ONS_INNER	SUB_2009	SUB_2006	geometry
0	Kingston upon Thames	E09000021	3726.117	0.000	F	None	None	POLYGON ((516401.600 160201.800, 516407.300 16...
1	Croydon	E09000008	8649.441	0.000	F	None	None	POLYGON ((535009.200 159504.700, 535005.500 15...
2	Bromley	E09000006	15013.487	0.000	F	None	None	POLYGON ((540373.600 157530.400, 540361.200 15...
3	Hounslow	E09000018	5658.541	60.755	F	None	None	POLYGON ((521975.800 178100.000, 521967.700 17...
4	Ealing	E09000009	5554.428	0.000	F	None	None	POLYGON ((510253.500 182881.600, 510249.900 18...

and this is the map of London with superimposed the boundaries of each borough:



And these are the first rows of the dataset containing the venues located in London boroughs returned by Foursquare API:



	Venue	Category	Venue_Latitude	Venue_Longitude	Borough
0	National Portrait Gallery	Art Gallery	51.509438	-0.128032	Westminster
1	Nicolaes Maes	Art Gallery	51.508853	-0.128864	Westminster
2	Andy Warhol	Art Gallery	51.507869	-0.099663	Southwark
3	Magus & the Fool - Art Café	Coffee Shop	51.531153	-0.039047	Tower Hamlets
4	Turning Earth	Art Gallery	51.570680	-0.037996	Waltham Forest
5	Bold Tendencies	Art Gallery	51.470771	-0.067811	Southwark
6	British Art 1540 To 1890	Art Gallery	51.490857	-0.128106	Westminster
7	British Art 1930-Now	Art Gallery	51.490664	-0.127205	Westminster
9	Pitzhanger Manor Gallery & House	Art Gallery	51.511170	-0.306750	Ealing
10	The Infinite Mix	Art Gallery	51.512493	-0.115048	Westminster
11	Arthaus Building	Art Gallery	51.543517	-0.057192	Hackney
12	East Wing Galleries	Art Gallery	51.510673	-0.117326	Westminster
13	Stockwell Graffiti Hall Of Fame	Art Gallery	51.468887	-0.116740	Lambeth
14	Matt's Gallery	Art Gallery	51.519127	-0.035948	Tower Hamlets
15	NOW Gallery	Art Gallery	51.500323	0.004654	Greenwich
16	Bussey Building	General Entertainment	51.469936	-0.067098	Southwark
17	Somerset House	Event Space	51.510786	-0.117899	Westminster
18	Martinspeed	Art Gallery	51.488914	-0.122176	Lambeth
19	Frieze Sculpture Park	Art Gallery	51.525525	-0.146465	Camden
20	Estorick Collection of Modern Italian Art	Art Museum	51.543971	-0.100671	Islington

Then, to analyze the venue's data returned by Foursquare we performed a technique in which categorical data is transformed into numerical data for Machine Learning algorithms: this technique is called One hot encoding. We obtained the following dataframe:

	Borough	Adult Education Center	Art Gallery	Art Museum	Art Studio	Arts & Crafts Store	Bar	Bike Shop	Coffee Shop	College Academic Building	College Arts Building	Coworking Space	Design Studio	Discount Store	Event Space	Fabric Shop	Frame Store	General College & University	General Entertainment	History Museum	Knitting Store	Museum	Performing Arts Venue	Print Shop	Pub	Stationery Store	Street Art	University
0	Barnet	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Brent	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	Bromley	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Camden	1	3	3	2	1	1	0	0	0	7	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	3	1
4	City of London	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0
5	Croydon	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	Ealing	0	1	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	Enfield	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	Greenwich	0	3	0	1	1	0	0	0	0	1	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	Hackney	0	3	2	3	4	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	Hammersmith and Fulham	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	Haringey	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	Havering	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	Hillingdon	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	Hounslow	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
15	Islington	0	0	2	0	5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
16	Kensington and Chelsea	0	1	6	1	0	0	0	0	0	3	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
17	Lambeth	0	3	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
18	Levensham	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	Merton	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	Newham	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
21	Redbridge	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
22	Richmond upon Thames	0	1	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
23	Southwark	0	4	4	4	0	0	0	0	0	3	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
24	Tower Hamlets	0	4	2	3	0	0	0	1	1	4	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
25	Waltham Forest	0	4	1	0	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
26	Wandsworth	0	0	1	1	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
27	Westminster	0	15	10	4	2	0	0	1	0	7	0	0	0	1	1	0	1	0	1	0	2	0	0	0	0	3	0

After this, we grouped rows of our data frame by borough and, by taking the mean of the frequency of occurrence of each venue's category, we created a new data frame listing the top 10 venues for each borough, making the data much simpler to analyze. The following is the resulting table:

	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	Barnet	Discount Store	Art Museum	College Arts Building	University	Art Gallery	Art Studio	Arts & Crafts Store	Bar	Bike Shop	Coffee Shop
1	Brent	Art Studio	University	Discount Store	Art Gallery	Art Museum	Arts & Crafts Store	Bar	Bike Shop	Coffee Shop	College Academic Building
2	Bromley	Arts & Crafts Store	University	Discount Store	Art Gallery	Art Museum	Art Studio	Bar	Bike Shop	Coffee Shop	College Academic Building
3	Camden	College Arts Building	Stationery Store	Art Gallery	Art Museum	Art Studio	History Museum	University	Street Art	Arts & Crafts Store	Bar
4	City of London	Stationery Store	Performing Arts Venue	Frame Store	College Arts Building	University	Design Studio	Art Gallery	Art Museum	Art Studio	Arts & Crafts Store

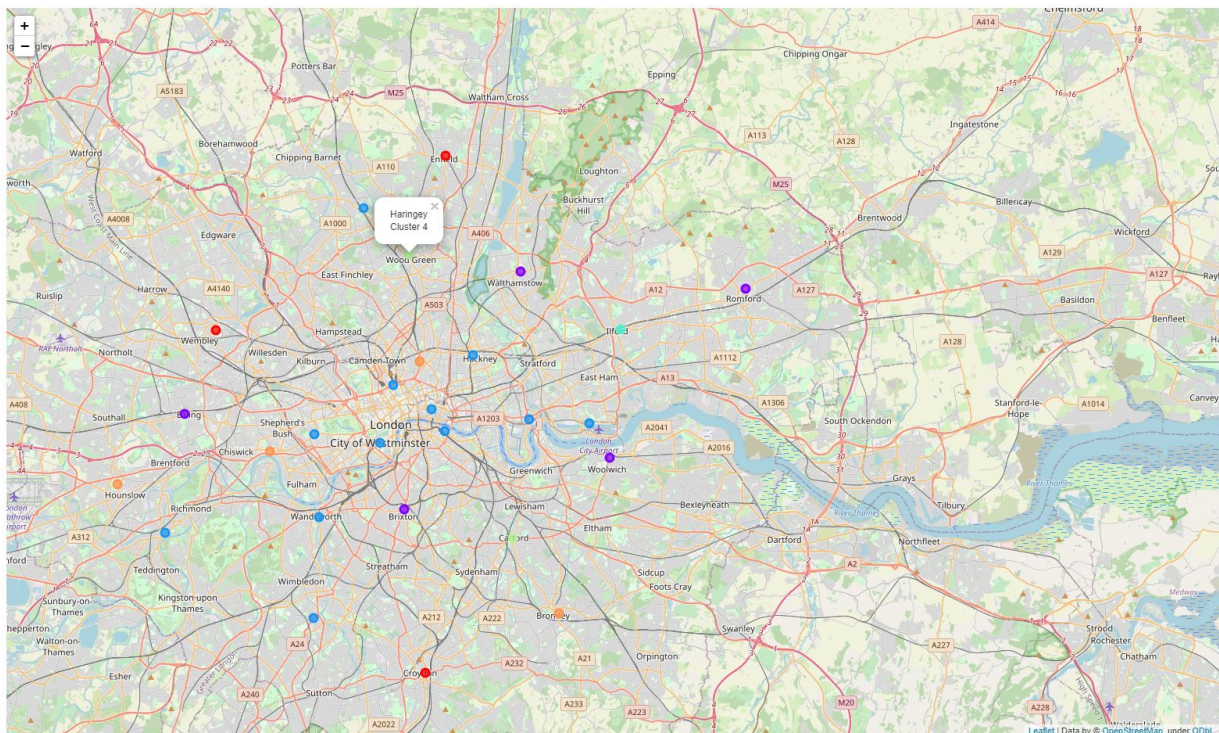
We could see that only 28 boroughs out of 33 have at least one venue falling in one of the categories we are interested into.

To make the analysis more interesting, we decided to cluster the boroughs based on the similarities of the frequency of occurrence of each venue's category in that borough. To do this we used *K-Means* clustering. To get our optimum *K* value we ran a test with different values of *K* and measured the accuracy of our model. In our case, we had the optimum at  $K = 6$ . That means we will have a total of 6 clusters.

After, we merged this data with the borough data frame to have a global vision on the distribution of venues of our interest among the various London boroughs. This new data frame represents (see the image below) the basis for analyzing new opportunities for opening a new Arts & Crafts store in London.

	Borough	Latitude	Longitude	Population Projections (2020)	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	Barnet	51.6252	-0.1517	403000	2	Discount Store	Art Museum	College Arts Building	University	Art Gallery	Art Studio	Arts & Crafts Store	Bar	Bike Shop	Coffee Shop
1	Brent	51.5588	-0.2817	342400	0	Art Studio	University	Discount Store	Art Gallery	Art Museum	Arts & Crafts Store	Bar	Bike Shop	Coffee Shop	College Academic Building
2	Bromley	51.4039	0.0198	332000	5	Arts & Crafts Store	University	Discount Store	Art Gallery	Art Museum	Art Studio	Bar	Bike Shop	Coffee Shop	College Academic Building
3	Camden	51.5290	-0.1255	248400	2	College Arts Building	Stationery Store	Art Gallery	Art Museum	Art Studio	History Museum	University	Street Art	Arts & Crafts Store	Bar
4	Croydon	51.3714	-0.0977	395600	0	Art Studio	University	Discount Store	Art Gallery	Art Museum	Arts & Crafts Store	Bar	Bike Shop	Coffee Shop	College Academic Building

Then we created a map using the Folium package in Python and each borough was colored based on the cluster label.



The above map shows the different clusters that had a similar mean frequency of the venue's categories we were interested into.



## 4. Results

From cluster analysis we get that only the following results:

### Cluster 1

Cluster 1

```
London_merged.loc[London_merged['Cluster_Labels'] == 0, London_merged.columns[list(range(3, London_merged.shape[1]))]]
```

	Population Projections (2020)	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	342400	0	Art Studio	University	Discount Store	Art Gallery	Art Museum	Arts & Crafts Store	Bar	Bike Shop	Coffee Shop	College Academic Building
4	395600	0	Art Studio	University	Discount Store	Art Gallery	Art Museum	Arts & Crafts Store	Bar	Bike Shop	Coffee Shop	College Academic Building
6	339200	0	Art Studio	University	Discount Store	Art Gallery	Art Museum	Arts & Crafts Store	Bar	Bike Shop	Coffee Shop	College Academic Building

Cluster 1 has three boroughs: Croydon, Brent and Enfield.

It we look at the most common venues one could suppose that this cluster has the highest average of Art Studios and Universities, but if you look at raw data you can find that there are only to Art Studios among all the venues we are interested to.

In the map, we can see that boroughs of this cluster are dispersed all throughout London making it one of the most sparsely populated cluster.

### Cluster 2

Cluster 2

```
London_merged.loc[London_merged['Cluster_Labels'] == 1, London_merged.columns[list(range(3, London_merged.shape[1]))]]
```

	Population Projections (2020)	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
5	360900	1	Art Gallery	Arts & Crafts Store	College Arts Building	University	Discount Store	Art Museum	Art Studio	Bar	Bike Shop	Coffee Shop
7	296000	1	Art Gallery	Discount Store	Art Studio	Arts & Crafts Store	College Arts Building	University	Art Museum	Bar	Bike Shop	Coffee Shop
11	262500	1	Arts & Crafts Store	College Arts Building	University	Discount Store	Art Gallery	Art Museum	Art Studio	Bar	Bike Shop	Coffee Shop
16	339200	1	Art Gallery	Arts & Crafts Store	College Arts Building	Museum	University	Discount Store	Art Museum	Art Studio	Bar	Bike Shop
24	282300	1	Art Gallery	Arts & Crafts Store	Art Museum	College Arts Building	University	Discount Store	Art Studio	Bar	Bike Shop	Coffee Shop

Cluster 2 has five boroughs: Lambeth, Greenwich, Ealing, Havering and Waltham Forest.

The most common venues are Art Galleries and Arts & Crafts stores: we have 11 Art Galleries and 8 Arts & Crafts stores.

## Cluster 3

Cluster 3

```
London_merged.loc[London_merged['Cluster Labels'] == 2, London_merged.columns[list(range(3, London_merged.shape[1]))]]
```

	Population Projections (2020)	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	403000	2	Discount Store	Art Museum	College Arts Building	University	Art Gallery	Art Studio	Arts & Crafts Store	Bar	Bike Shop	Coffee Shop
3	248400	2	College Arts Building	Stationery Store	Art Gallery	Art Museum	Art Studio	History Museum	University	Street Art	Arts & Crafts Store	Bar
8	284400	2	Arts & Crafts Store	Art Gallery	Art Studio	Art Museum	Bike Shop	Coworking Space	University	Discount Store	Bar	Coffee Shop
12	305600	2	Art Studio	College Arts Building	University	Discount Store	Art Gallery	Art Museum	Arts & Crafts Store	Bar	Bike Shop	Coffee Shop
15	160600	2	Art Museum	College Arts Building	Art Gallery	Art Studio	Museum	Design Studio	University	Discount Store	Arts & Crafts Store	Bar
18	211200	2	Art Studio	Arts & Crafts Store	College Arts Building	University	Discount Store	Art Gallery	Art Museum	Bar	Bike Shop	Coffee Shop
19	358100	2	Stationery Store	General Entertainment	University	Design Studio	Art Gallery	Art Museum	Art Studio	Arts & Crafts Store	Bar	Bike Shop
21	199400	2	Art Studio	Art Gallery	Knitting Store	College Arts Building	University	Discount Store	Art Museum	Arts & Crafts Store	Bar	Bike Shop
22	329000	2	Art Gallery	Art Museum	Art Studio	College Arts Building	General Entertainment	University	Discount Store	Arts & Crafts Store	Bar	Bike Shop
23	331000	2	Art Gallery	College Arts Building	Art Studio	Art Museum	Pub	Print Shop	Coffee Shop	College Academic Building	University	Design Studio
25	329000	2	Arts & Crafts Store	College Arts Building	Art Museum	Art Studio	University	Discount Store	Art Gallery	Bar	Bike Shop	Coffee Shop
26	248700	2	Art Gallery	Art Museum	College Arts Building	Art Studio	Stationery Store	Museum	Arts & Crafts Store	Event Space	History Museum	General College & University
27	9400	2	Stationery Store	Performing Arts Venue	Frame Store	College Arts Building	University	Design Studio	Art Gallery	Art Museum	Art Studio	Arts & Crafts Store

Cluster 3 has thirteen boroughs.

The most common venues are Art Galleries and College Arts Buildings.

In this cluster we have a total of 11 Arts & Crafts stores.

Boroughs of cluster 3 are mainly located in Inner London.

## Cluster 4

Cluster 4

```
London_merged.loc[London_merged['Cluster Labels'] == 3, London_merged.columns[list(range(3, London_merged.shape[1]))]]
```

	Population Projections (2020)	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
20	313800	3	Discount Store	University	Art Gallery	Art Museum	Art Studio	Arts & Crafts Store	Bar	Bike Shop	Coffee Shop	College Academic Building

Cluster 4 has only one borough: Redbridge.

In this cluster, among all the venues we are interested into, there is only a Discount store.

## Cluster 5

Cluster 5

```
London_merged.loc[London_merged['Cluster Labels'] == 4, London_merged.columns[list(range(3, London_merged.shape[1]))]]
```

	Population Projections (2020)	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
10	287400	4	College Arts Building	University	Discount Store	Art Gallery	Art Museum	Art Studio	Arts & Crafts Store	Bar	Bike Shop	Coffee Shop
17	314200	4	College Arts Building	University	Discount Store	Art Gallery	Art Museum	Art Studio	Arts & Crafts Store	Bar	Bike Shop	Coffee Shop

Cluster 5 has two boroughs: Lewisham and Haringey.

In this cluster there are two College Arts Buildings, one for each borough. No other venue of our interest.

## Cluster 6

Cluster 6

```
London_merged.loc[London_merged['Cluster_Labels'] == 5, London_merged.columns[list(range(3, London_merged.shape[1]))]]
```

	Population Projections (2020)	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
2	332000	5	Arts & Crafts Store	University	Discount Store	Art Gallery	Art Museum	Art Studio	Bar	Bike Shop	Coffee Shop	College Academic Building
9	190900	5	Arts & Crafts Store	Art Museum	University	Discount Store	Art Gallery	Art Studio	Bar	Bike Shop	Coffee Shop	College Academic Building
13	280700	5	Arts & Crafts Store	Stationery Store	Art Museum	University	Discount Store	Art Gallery	Art Studio	Bar	Bike Shop	Coffee Shop
14	239600	5	Arts & Crafts Store	Art Museum	Stationery Store	College Arts Building	University	Discount Store	Art Gallery	Art Studio	Bar	Bike Shop

Cluster 6 has four boroughs: Hounslow, Bromley, Islington and Hammersmith and Fulham.

In this cluster the most common venue are Arts & Crafts stores, with a total of 9 such stores, of which 5 in the Islington borough.

## 5. Discussion

All of the Arts & Crafts stores are located in cluster 2, 3 and 6, with a total of 28 stores, divided in the following way:

- 8 stores in cluster 2;
- 11 stores in cluster 3;
- 9 stores in cluster 6.

We have a total of 40 College Arts Buildings, of which 30 are located in cluster 3 and 6 in cluster 2.

We can note that in cluster 3, and in particular in the Westminster borough, we have the most part of Art Galleries and Art Museums.

In the Westminster borough there are also 7 College Arts Buildings, but here we can find only 2 Arts & Crafts stores. This gives us a good opportunity for opening in this borough a new Arts & Crafts store.

One of the main drawbacks of this analysis is that it is completely based on venue's data obtained from the Foursquare API, and I can tell for sure that many venues belonging to the category we were interested into are missing or misclassified. For example, I found that one of the biggest Arts & Crafts store in London, which is also renowned for its online

store, is classified as a miscellaneous store, so it has not be included in our analysis.

## 6. Conclusion

In conclusion, we had an opportunity to face a business problem and to tackle it using data science methods.

We used different Python libraries to fetch the information, control the content and visualize those datasets.

We learned how to gather data needed for our analysis from different sources like:

- Foursquare API, to investigate the venues in boroughs of London;
- Wikipedia, to get a list of all the London boroughs;
- London Datastore, to get some relevant information about London boroughs (GIS boundaries, population, etc.).

This project can be greatly improved with a better data source, rather than Foursquare, from which to retrieve venue's information filtered by a better classification into categories.

We can indeed use other information to guide our choice to where to open a new Arts & Crafts store, for example the proximity to public transport.

Anyway, this project is a starting point to investigate similar situations, for example opening a new restaurant, a new bookshop and so forth. This project acts as an initial direction to tackle more complex real-life problems using data science.