

# The Battle of the Neighbourhoods – London

By James Barnard



## Introduction

London has 33 boroughs (including the City of London) and is home to over 8 million people and an additional 0.5m people commute to jobs there from outside London.

For this project, we want to look at the boroughs of London and classify them. Some neighbourhoods are mostly residential, some have more business or commercial spaces surrounding them. The venues closest to the centre of a neighbourhood determine why and how people use it.

By analysing the venues data, we can classify boroughs by their primary usage. This data could be useful for city planners and residents alike: it could help plan further city development, or help individuals where to live or set up a business.

## Data

We'll need data on the location of the boroughs and on the venues closest to them.

1. A list of boroughs is available from Wikipedia and their central location coordinates can be obtained using Nominatim.
2. We can use the Foursquare API to explore venue types in each borough. Foursquare outlines these high-level venue categories with more sub-categories:
  - Arts & Entertainment (4d4b7104d754a06370d81259)
  - College & University (4d4b7105d754a06372d81259)
  - Event (4d4b7105d754a06373d81259)
  - Food (4d4b7105d754a06374d81259)
  - Nightlife Spot (4d4b7105d754a06376d81259)
  - Outdoors & Recreation (4d4b7105d754a06377d81259)

- Professional & Other Places (4d4b7105d754a06375d81259)
- Residence (4e67e38e036454776db1fb3a)
- Shop & Service (4d4b7105d754a06378d81259)
- Travel & Transport (4d4b7105d754a06379d81259)

We'll be querying the number of venues in each category in a 1000m radius around the centre of each borough. This radius was chosen because 1000m is a reasonable walking distance.

## Methodology

Here I use Nominatim to geocode all London boroughs before making a dataframe:

	Borough	Latitude	Longitude
0	Barking and Dagenham	51.554117	0.150504
1	Barnet	51.648784	-0.172913
2	Bexley	51.441679	0.150488
3	Brent	51.584778	-0.299180
4	Bromley	51.402805	0.014814
5	Camden	51.542305	-0.139560

Create a map of London with boroughs superimposed:



# Venues and Categories

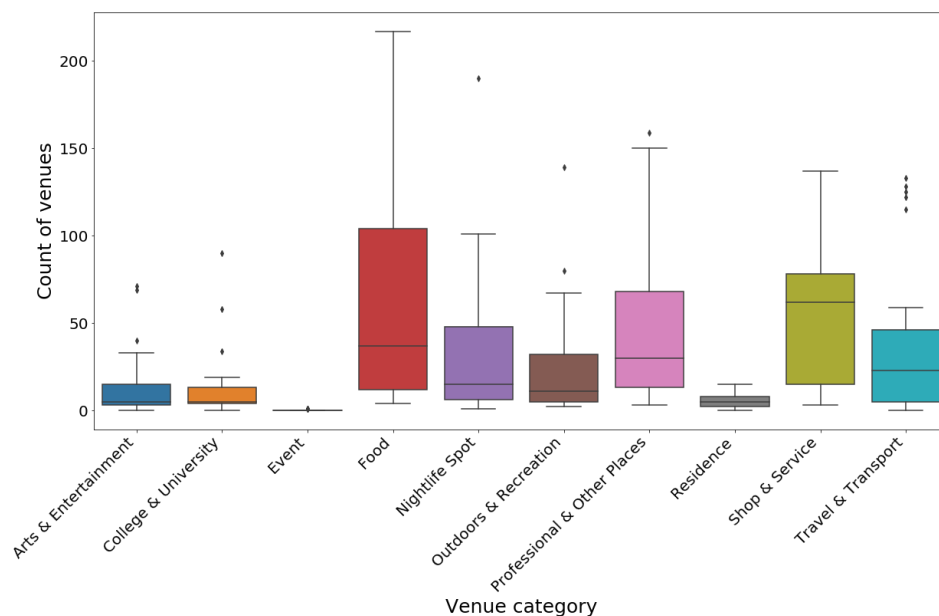
We will use Foursquare API to explore venue categories in each borough. Foursquare outlines these high-level venue categories with more sub-categories:

- Arts & Entertainment (4d4b7104d754a06370d81259)
- College & University (4d4b7105d754a06372d81259)
- Event (4d4b7105d754a06373d81259)
- Food (4d4b7105d754a06374d81259)
- Nightlife Spot (4d4b7105d754a06376d81259)
- Outdoors & Recreation (4d4b7105d754a06377d81259)
- Professional & Other Places (4d4b7105d754a06375d81259)
- Residence (4e67e38e036454776db1fb3a)
- Shop & Service (4d4b7105d754a06378d81259)
- Travel & Transport (4d4b7105d754a06379d81259)

We can use the foursquare explore API with categoryId to query the number of venues of each category in a specific radius.

	Borough	Latitude	Longitude	Arts & Entertainment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport
0	Barking and Dagenham	51.554117	0.150504	1	5	0	5	4	8	4	1	10	3
1	Barnet	51.648784	-0.172913	2	4	0	8	4	5	13	3	9	4
2	Bexley	51.441679	0.150488	1	1	0	8	6	2	5	0	15	3
3	Brent	51.584778	-0.299180	1	4	0	4	4	5	8	2	5	3
4	Bromley	51.402805	0.014814	7	1	1	44	12	9	28	5	75	12
5	Camden	51.542305	-0.139560	33	12	0	126	75	33	68	7	115	56

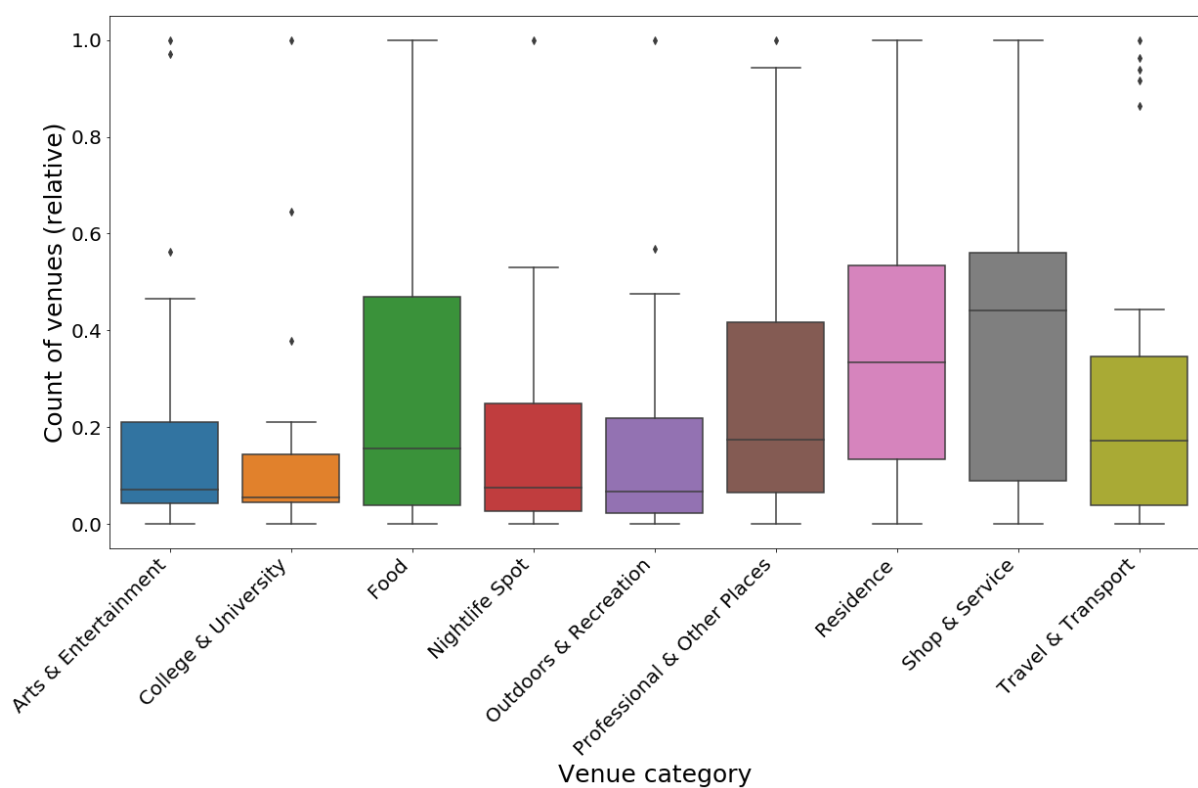
Plotting these venue categories as boxplots:



The most frequent venue categories are professional, shop&service and food. Event has very little data, let's discard it from both the dataframe and the list of categories.

## Data Preparation

Let's normalize the data using MinMaxScaler (scale from 0 to 1). This scales the data and provides an easy to interpret score at the same time. Visualizing the scaled data:

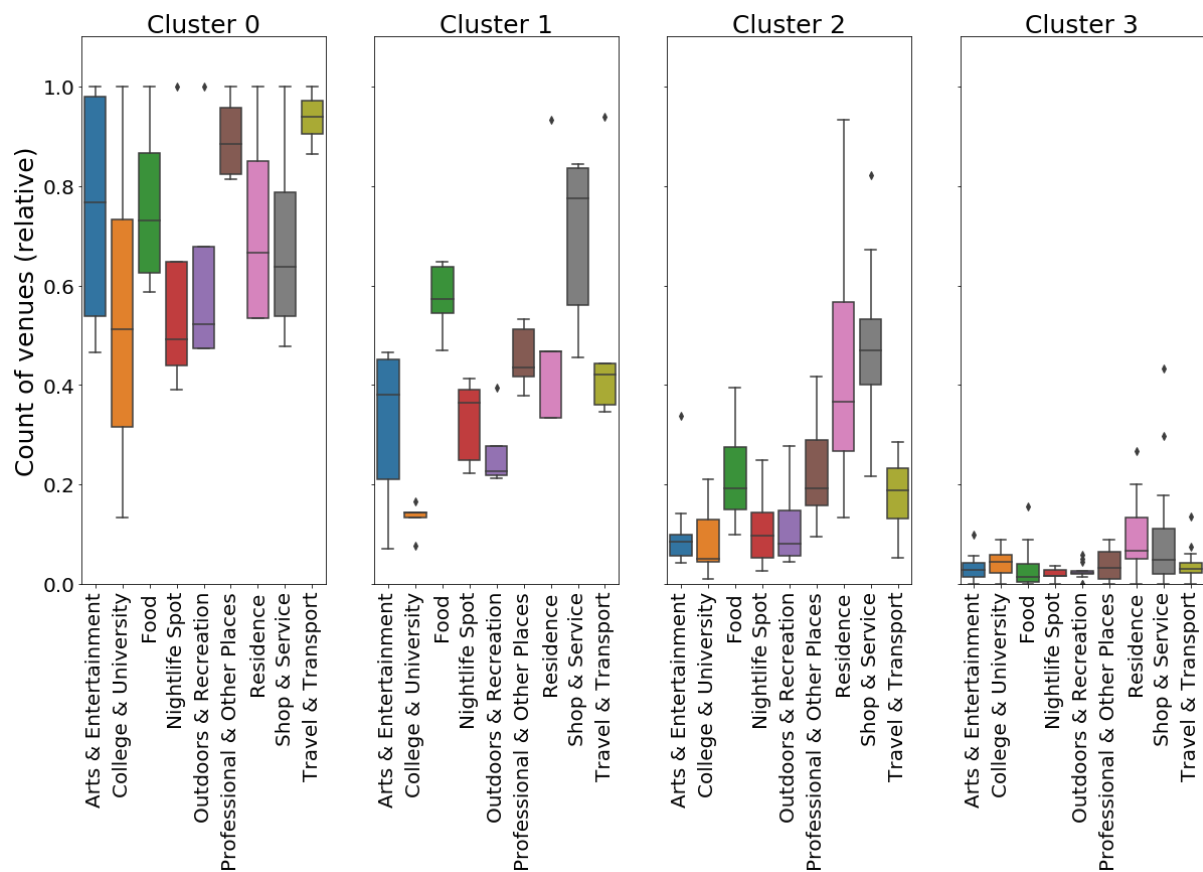


# Clustering

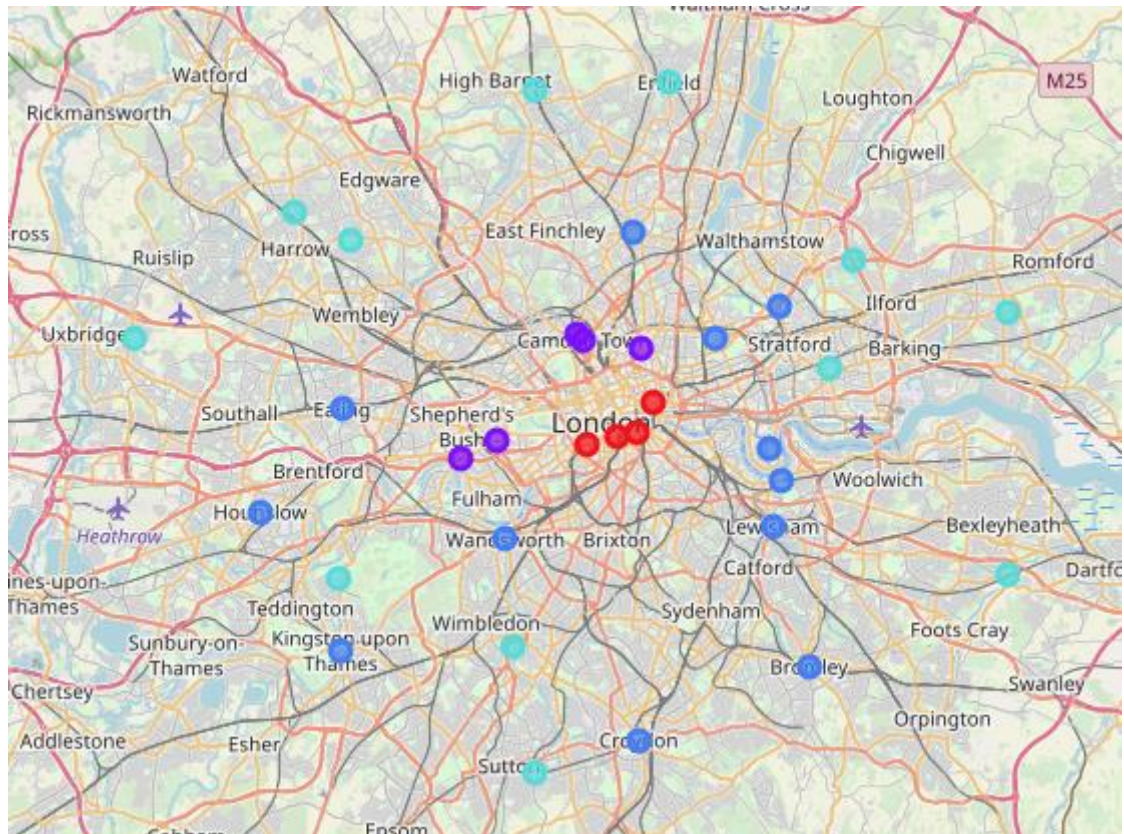
We'll be using k-means clustering in order to group similar boroughs together. This will enable analysis of the boroughs to be done more simply.

Preliminary analysis was done in order to determine the number of clusters, in the end opting for 4 clusters which gave a good spread of characteristics across the clusters and also making sense geographically.

Visualizing the clusters again with boxplots:



We can also map our data again, this time updated with the cluster data:



# Results

Here is how we can characterize the clusters by looking at venue scores:

- Cluster 0 (Red) has consistently high scores for all venue categories. This is the most diversely developed part of the city
- Cluster 1 (Purple) has highest marks for Shops & Services as well as food.
- Cluster 2 (Blue) has high marks in Residence and Shop&Service, indicating these are primarily residential areas.
- Cluster 3 (Light blue) has low marks across the board, but also have relatively high marks in Residence and Shop&Service, again indicating these are primarily residential areas, but with fewer (or more disparate) venues.

Plotting the clusters on a map shows us that the clusters can be seen as sort of concentric circles:

- Cluster 0 is the oldest central part of the city
- Cluster 1 is a little bit further out and to the north
- Cluster 2 is a little bit further out again
- Cluster 3 boroughs are the furthest from the centre

The hypothesis is that opportunities for business and residential will likely be minimal in Cluster 0 in the city centre (and/or expensive), but as you progress outwards into Clusters 1, 2 and 3 respectively, there will be increased opportunities for business and residential development.

## Discussion

To be fair, Foursquare data isn't all-encompassing. The highest number of venues are in the Food and Shop & Service categories. The data doesn't take into account a venue's size (e.g. a university building attracts a lot more people than a hot dog stand – each of them is still one Foursquare "venue").

## Conclusion

Foursquare data is limited but can provide insights into a city's development. This data could be combined with other sources (e.g. city data on number of residents) to provide more accurate results.