



The London Coffee Shop Problem

Coursera capstone project presentation



The Problem

Where should I open a new coffee shop in London?

In order to find the perfect spot, we need to consider criteria that might be markers of a "good" location. The following criteria should help prospective coffee shop owners:

1. Few existing coffee shops (low competition)
2. Near students (caffeine demand)
3. Near retail, museums or other attractions (passing trade)



Data: London Locations

Using data from Wikipedia

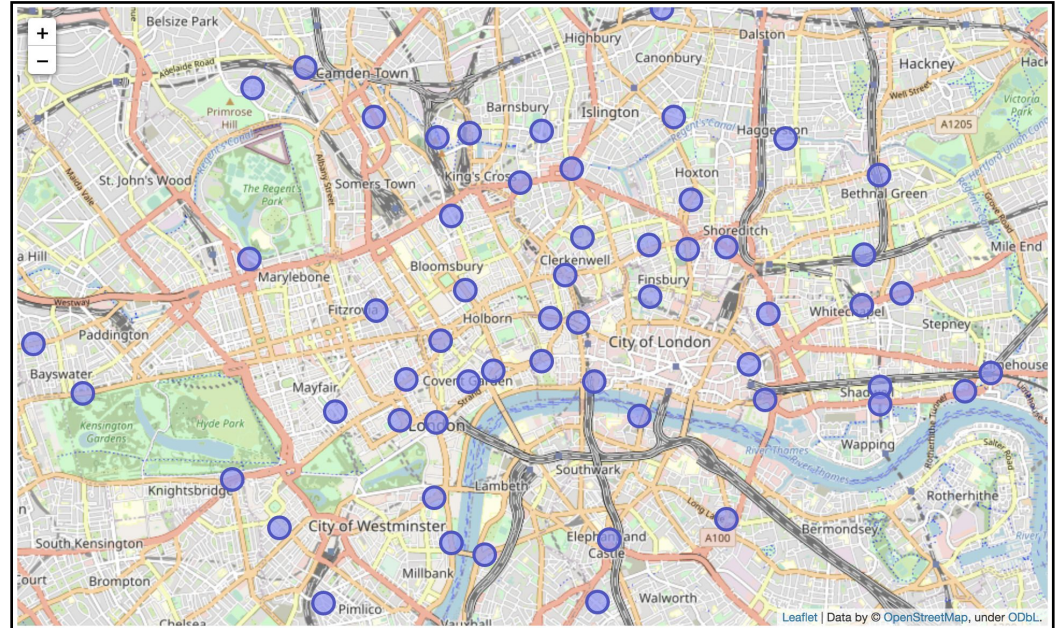
(https://en.wikipedia.org/wiki/List_of_areas_of_London), I searched for London locations within the central London postal area.

Next, I used the geocoder package to find their longitude and latitude (required for using the Foursquare API):

	Latitude	Longitude	Location
0	51.513312	-0.077765	Aldgate
1	51.512653	-0.118607	Aldwych
2	51.524730	-0.087540	Angel
3	51.508164	-0.095216	Bankside
4	51.520070	-0.093530	Barbican
5	51.536489	-0.110907	Barnsbury
6	51.510480	-0.184260	Bayswater
7	51.497050	-0.152750	Belgravia
8	51.497900	-0.081440	Bermondsey
9	51.524190	-0.059440	Bethnal Green

Data: London Locations

Here's a map of the 56 locations I am considering, visualised with folium:





Data: Foursquare API

The search terms I used to obtain venue information per location from Foursquare are: *["Coffee", "Colleges", "Shopping", "Fun"]*

For each search term, I counted the number of venues per location within a 500m radius for coffee shops and a 1km radius for the other queries.

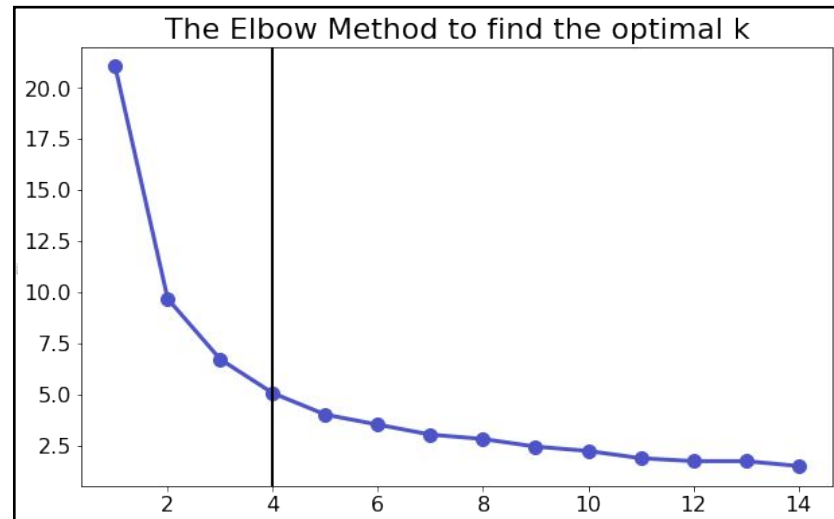
The resulting data was normalized between 0 and 1 per venue category to ensure that the scales were consistent:

Location	Latitude	Longitude	Coffee	Colleges	Fun	Shopping
Aldgate	51.513312	-0.077765	0.714286	0.000000	0.461538	0.000
Aldwych	51.512653	-0.118607	0.795918	1.000000	0.961538	0.625
Angel	51.524730	-0.087540	0.693878	0.166667	0.769231	0.125
Bankside	51.508164	-0.095216	0.530612	0.333333	0.346154	0.250
Barbican	51.520070	-0.093530	0.530612	0.166667	0.653846	0.375
Barnsbury	51.536489	-0.110907	0.102041	0.000000	0.307692	0.500
Bayswater	51.510480	-0.184260	0.061224	0.166667	0.115385	0.000
Belgravia	51.497050	-0.152750	0.142857	0.000000	0.269231	0.500
Bermondsey	51.497900	-0.081440	0.142857	0.166667	0.076923	0.000
Bethnal Green	51.524190	-0.059440	0.183673	0.000000	0.153846	0.125
Blackfriars	51.511600	-0.102408	0.387755	0.500000	0.500000	0.375
Bloomsbury	51.520740	-0.123100	0.714286	1.000000	1.000000	0.375
Cambridge Heath	51.532180	-0.056950	0.183673	0.000000	0.115385	0.000
Camden Town	51.537899	-0.137591	0.306122	0.833333	0.384615	0.250
Canonbury	51.548680	-0.091750	0.020408	0.000000	0.038462	0.125

Data: KMeans Clustering

I used KMeans clustering from sklearn to group the locations based on their relative frequency of coffee shops, attractions, shopping, and universities.

The Elbow Method was used to determine the number of appropriate clusters, which seemed to be 4:





Data: KMeans Clustering

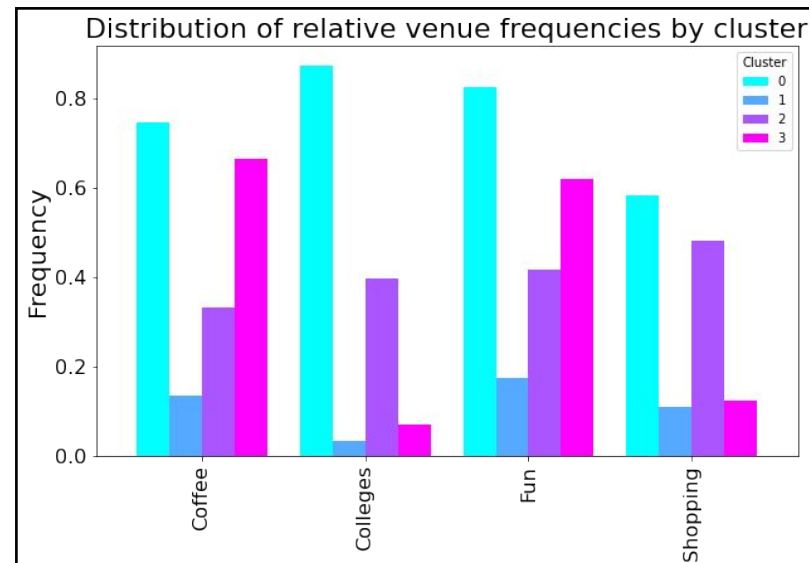
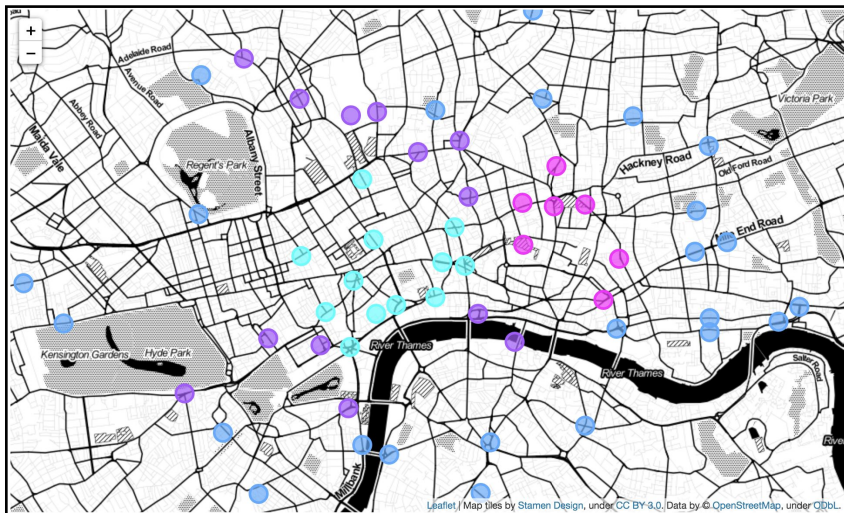
KMeans clustering with 4 clusters was run, grouping locations based on their relative frequency of coffee shops, colleges, attractions, and retail.

Each location was labelled according to its assigned cluster number:

Cluster	Location	Latitude	Longitude	Coffee	Colleges	Fun	Shopping
3	Aldgate	51.513312	-0.077765	0.714286	0.000000	0.461538	0.000
0	Aldwych	51.512653	-0.118607	0.795918	1.000000	0.961538	0.625
3	Angel	51.524730	-0.087540	0.693878	0.166667	0.769231	0.125
2	Bankside	51.508164	-0.095216	0.530612	0.333333	0.346154	0.250
3	Barbican	51.520070	-0.093530	0.530612	0.166667	0.653846	0.375
1	Barnsbury	51.536489	-0.110907	0.102041	0.000000	0.307692	0.500
1	Bayswater	51.510480	-0.184260	0.061224	0.166667	0.115385	0.000
1	Belgravia	51.497050	-0.152750	0.142857	0.000000	0.269231	0.500
1	Bermondsey	51.497900	-0.081440	0.142857	0.166667	0.076923	0.000
1	Bethnal Green	51.524190	-0.059440	0.183673	0.000000	0.153846	0.125

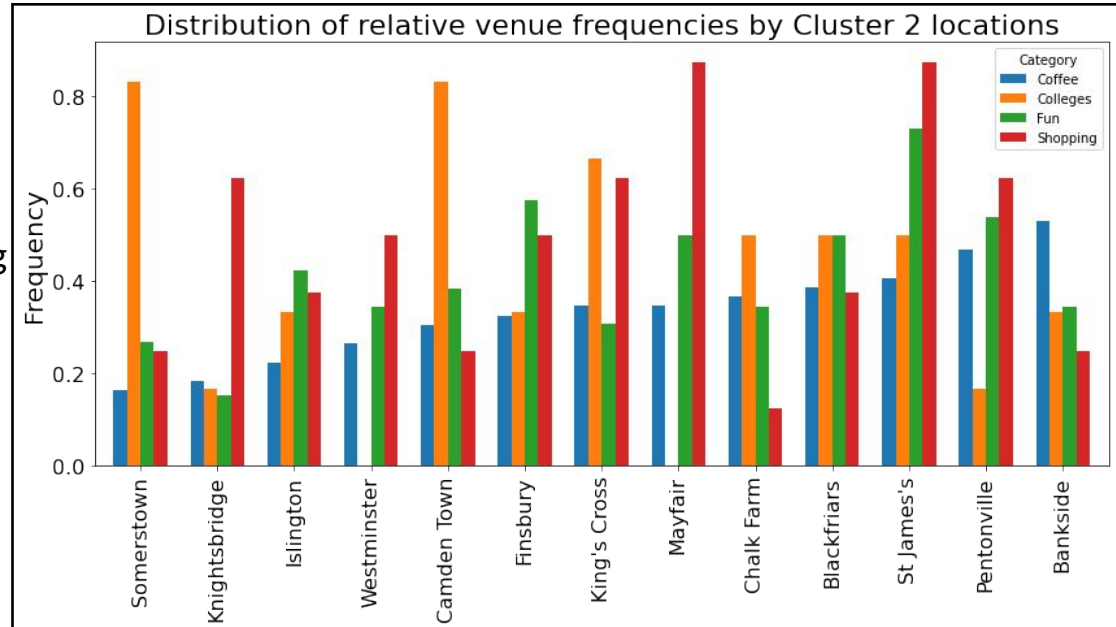
Results

Here are the London locations, grouped into their assigned clusters, as well as the mean frequency of each type of venue per cluster:



Results

The cluster analysis suggests that cluster 2 has some good candidate locations, with relatively few existing coffee shops but a reasonable frequency of colleges, attractions, and retail. Let's explore those locations further:





Conclusion

Using clustering with Foursquare venue data, I found 4 clusters of central London locations, with one particular cluster containing locations that, on average, had relatively few coffee shops already but reasonable proximity to students and entertainment. Within this cluster, 3 possible locations stand out to me as being the best:

- Camden Town, due to lack of existing competition and lots of students;
- King's Cross, with limited coffee shops and near to students and shopping;
- St James's, which has some students and lots of attractions close by.

Future work should consider other factors in order to fully specify the best location (e.g. unit availability, rent, and other overheads). Also, understanding the nearby residential population would be helpful.