Where should I book my AirBnB in London?

Introduction

I often get asked for recommendations by visiting family and friends as to what area of London they should stay in for an upcoming visit. The conversation always takes a predictable path whereby I ask what they like to do, what they want to see or eat, do they like quiet or noisy and then I make a recommendation. What I think would be super helpful is to categorize London by postcode into different types of neighborhood (e.g. quiet, trendy, lots of restaurants, etc.) and then have this list/map available for my friends, family, or just travelers in general to peruse themselves and they can then make their own informed decision.

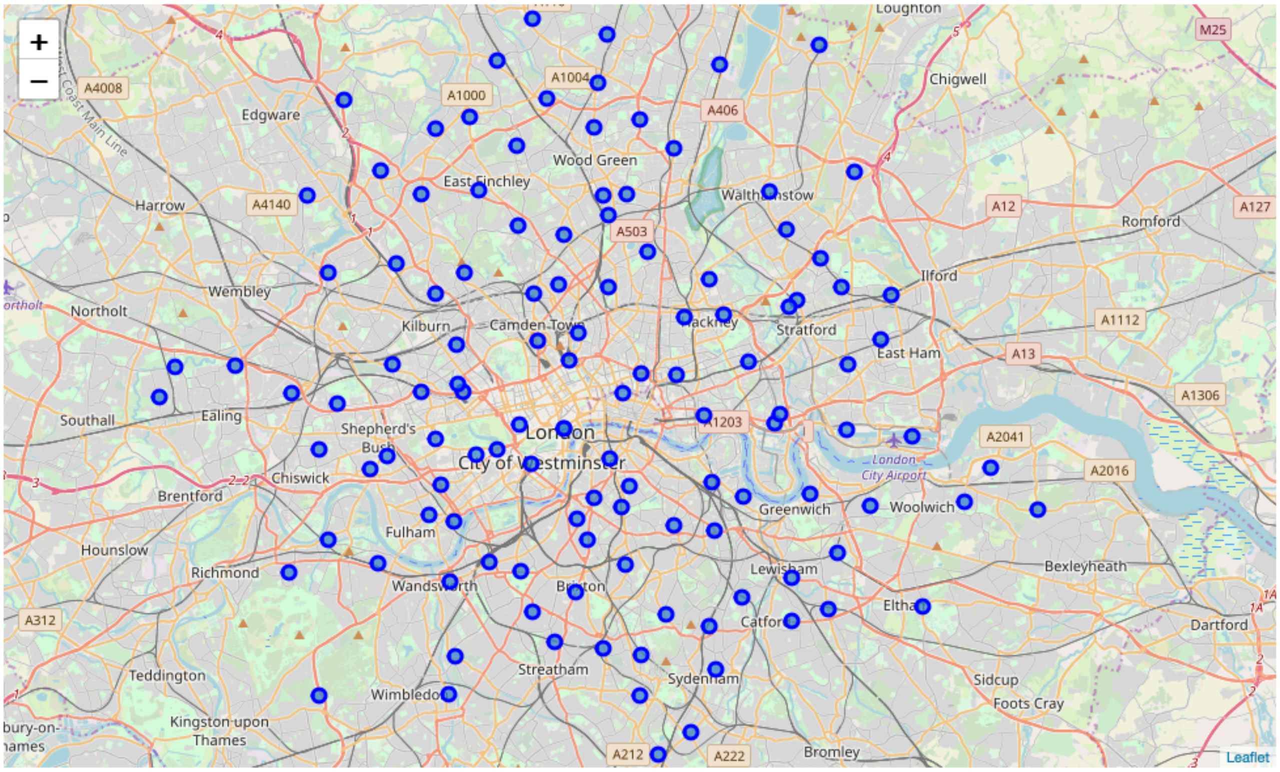
Data

I scraped a website[1] for the London postcode information, much like we did in the project analyzing Toronto. Once I had a list of the post codes, I found a website that would return an excel file of coordinate data given a list of UK postcodes. I exported a list of the London postcodes I had scraped, uploaded these to the geo information website[2], got the excel file with corresponding geo location data and then imported and merged it to my postcode data frame. Interestingly, there were 9 postcodes that didn't return any geo data from this website. Since there were only 9, I manually input this information into the excel file before I imported it. Had there been many more missing entries, I would have looked to find a new source for getting the geo data. After this, I connected to the Foursquare API and got the venue information for all of the London postcodes using their latitude and longitude data.

Methodology

I first started by superimposing each area onto a folium map of London. This gave a good visual idea of the scope of the project and what I was dealing with.

[1] https://www.milesfaster.co.uk/london-postcodes-list.htm
[2] https://gridreferencefinder.com/postcodeBatchConverter/



Next, I connected to the Foursquare API and called for all the venues (with a limit of 50) in a given radius of each areas' center point. This led to a data frame with 2,768 rows of venues coming from 121 area in London. The first five rows of the data frame shown below.

```
(2768, 7)
```

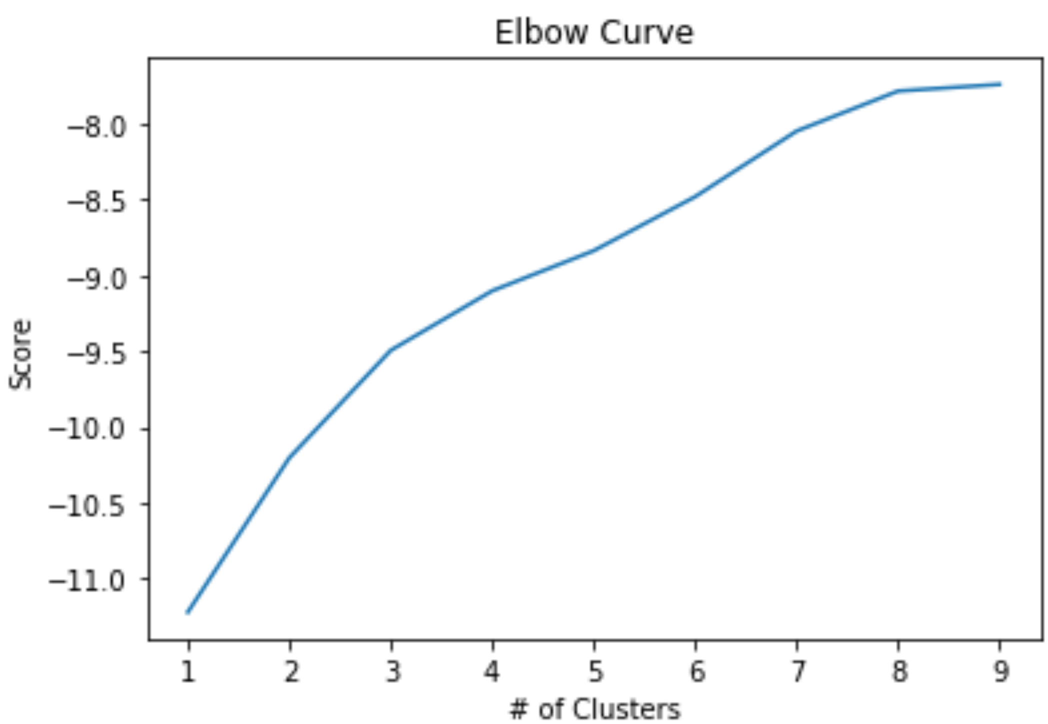
	Area	Area Latitude	Area Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Whitechapel, Stepney, Mile End	51.512497	-0.052098	George Tavern	51.514223	-0.053258	Pub
1	Whitechapel, Stepney, Mile End	51.512497	-0.052098	East London Food Centre	51.513591	-0.048786	Grocery Store
2	Whitechapel, Stepney, Mile End	51.512497	-0.052098	King Edward Memorial Park	51.508957	-0.048895	Park
3	Whitechapel, Stepney, Mile End	51.512497	-0.052098	St George's Leisure Centre	51.509930	-0.057905	Pool
4	Whitechapel, Stepney, Mile End	51.512497	-0.052098	Sainsbury's Local	51.514305	-0.054450	Grocery Store

Following this, and in preparation for using the unsupervised K-Mean clustering method, I used one hot encoding to give a numerical "dummy" value to the venues. Incidentally, we can also note that there were 297 unique types of venue in our London areas. Next, we group the data by area and average out how many of each type of venue there are in each respective area. This is the data we'll use to train our K-Means model.

Before training the model, a quick exploratory visualization of what the information we have looks like, giving a snapshot of the area of London and the top 10 most common venues in that area:

	Area	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	Abbey Wood	Campground	Yoga Studio	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Film Studio	Fish & Chips Shop	Fish Market	Flea Market	Flower Shop
1	Acton	Hotel	Coffee Shop	Gym	Grocery Store	Train Station	Fast Food Restaurant	Café	Park	Flea Market	Flower Shop
2	Archway, Tufnell Park	Martial Arts Dojo	Wine Shop	Pub	Pizza Place	Yoga Studio	Fish Market	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant
3	Balham	Italian Restaurant	Pub	Grocery Store	Pizza Place	Indian Restaurant	Gastropub	Wine Shop	Food Court	Food Stand	Food & Drink Shop
4	Barnes, Castelnau	Pub	Grocery Store	Coffee Shop	Farmers Market	Park	Gastropub	Train Station	Tea Room	Athletics & Sports	Lake

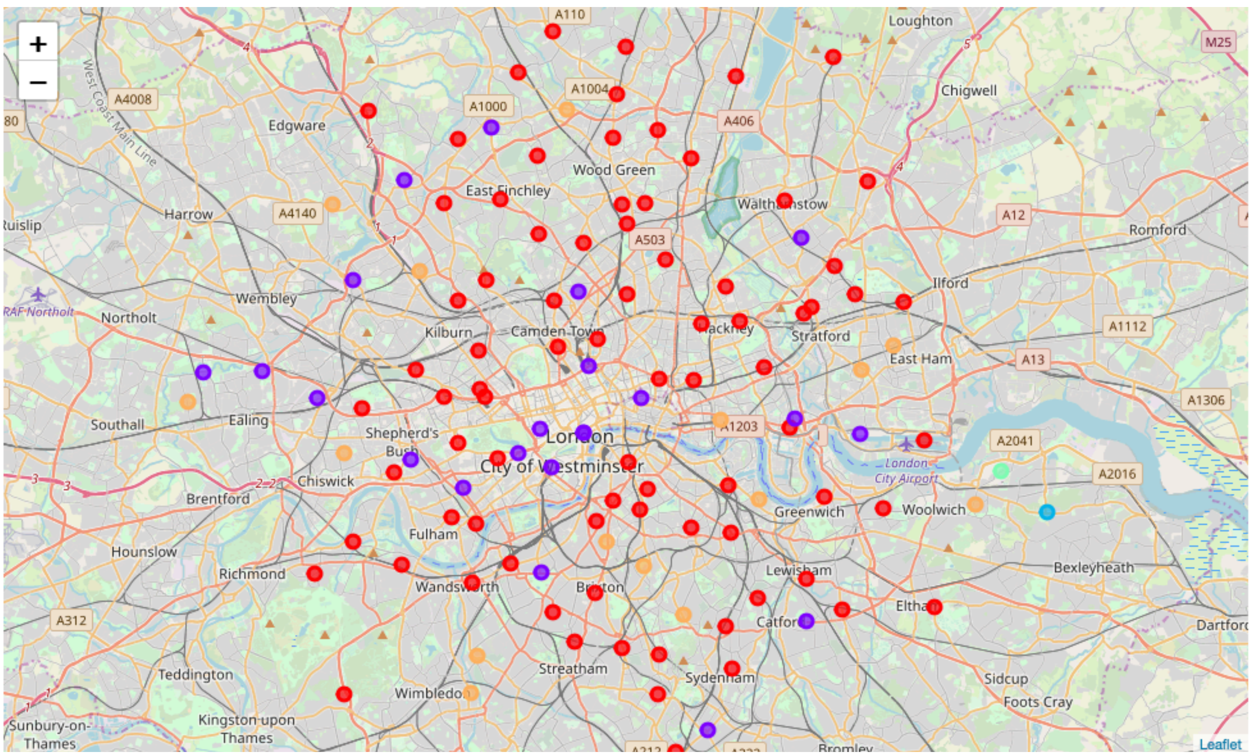
To train the model I first visualized the “elbow curve” in order to identify the best “k” to use. Typically there is something that appears as a line with a right angle and the right angle coincides with the best “k”. There wasn’t much of a kink in the curve so I chose k=5 because it looked as much like an inflection point as any without having too high of a “k.”



Next, I ran the K-Means model with 5 clusters and merged the resulting data frame, which categorized each area into a specific cluster group, back with the postcode and geospatial data frame resulting in a data frame that looked like this:

Postcode	Area	Latitude	Longitude	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
E1	Whitechapel, Stepney, Mile End	51.512497	-0.052098	4.0	Grocery Store	Park	Flower Shop	Sandwich Place	Harbor / Marina	Dive Bar	Coffee Shop	Hotel
E1W	Wapping	51.505800	0.057200	0.0	Café	Tea Room	Airport Service	Light Rail Station	Coffee Shop	Flea Market	Farmers Market	Fast Food Restaurant
E2	Bethnal Green, Shoreditch	51.526024	-0.066917	0.0	Café	Pub	Yoga Studio	Restaurant	Turkish Restaurant	Coffee Shop	Furniture / Home Store	Bagel Shop
E3	Bow, Bromley-by-Bow	51.530230	-0.028907	0.0	Pub	Café	Pizza Place	Grocery Store	Coffee Shop	Road	Bar	Bakery

With this information, we can color code each area onto our existing area map of London and get a nice visualization and sense of where the similarly grouped areas are in London.



Results

Cluster 0	Lots of café's, pubs, bars, restaurants (Denoted by **Red**)
Cluster 1	Lots of hotels, restaurants, bars (Denoted by **Purple**)
Cluster 2	Campgrounds, Markets, Fish related venues (Denoted by **Blue**)
Cluster 3	Grocery stores, Yoga, Outdoor markets (Denoted by **Green**)
Cluster 4	Lots of parks, café's and grocery stores (Denoted by **Orange**)

The results detailed in the table above give a general idea of the main types of venues in each cluster group. Interesting to note that cluster groups 2 and 3 only had one area, possibilities for which we'll discuss later. In general, it looks like cluster 0 can be categorized as "trendy, authentic London" as there are many, what appears to be, local shops, eateries and pubs.

Cluster 1 can be classified as “touristy” as there are lots of hotels (presumably because they are near to tourist attractions) and restaurants. Cluster 2 is on the outskirts of London, has lots of campgrounds and markets so I’d say can be classified as “outdoorsy”. Cluster 3 has lots of grocery stores, markets and yoga studios, so this can be classified as the “suburbs, commuter area.” Lastly, cluster 4 is categorized by lots of parks, café’s and grocery stores so I’d categorize it as “family-friendly,” especially as one can see that on the map these areas for the most part are outside of central London.

Discussion

One of the obvious discussion points is the fact that two of the clusters only had one area in them. This begs the question—should there have been fewer clusters? The answer is probably “yes” but as we saw from the “elbow curve” choosing which “k” to use wasn’t as obvious as it should have been.

It would also be interesting to introduce some pricing data, for example avg. hotel price, average income of residents, etc., to help train the model. It’s also important as travelers will be price sensitive and want to know where they can afford to stay.

Conclusion

In this exercise, I clustered similar groups of areas in London based on the type and number of venues. This resulted in data that gives a “feel” for what each area in London is like, where similar areas in London are, and what the most common venues in those areas are. People travelling to London can now get a better idea of where they should stay given their specific wishes—do they want a family-friendly area because they’re travelling with kids? Or perhaps a solo traveler hoping to take in lots of sights? Or party?

The information also gave rise to other useful applications. Perhaps someone needs to move within London and wants to find a similar area to where they currently live. Or maybe someone wants to explore London and is looking for something different to what they see every day. There are also investment implications—developers can use this information to make informed decisions on where they might want to build a hotel, or perhaps buy up individual properties for use as AirBnB rentals.