

Uncovering the Popular Locations in London

Chayapod Beekhaimook

January 10, 2021

Introduction

London has become a popular place among global and local travelers. It is home to the famous and gorgeous clock tower, Big Ben, and the giant iconic wheel near it, London Eye. Hyde Park, possible the largest and most famous park in London, is popular among travelers too. In addition to history, it is where the tale of Sherlock Holmes began at 221B Baker Street. Holmes is arguably the most famous detective in the fictional world and is feared by all criminals. Many visitors go to London mainly to visit his fictional dwelling place.



Figure 1: Popular Attractions In London (left) Big Ben and (right) Sherlock Holmes Museum At 221B Baker Street

Besides visiting all the famous places in London, visitors want to get the most out of their experience. Some will visit all the popular venues in various neighborhoods in London, while others will dwell at the most famous spot for hours. Regardless of the traveling style, it would be nice to have a list of popular venues in each area in London, so that visitors can plan their visit optimally. In this regards, data science can help us uncover these popular spots in London.

Business Problem

The goal of this project is to help visitors choose their destinations depending on their likings and the experiences that the neighborhoods have to offer. This also helps people make decisions if they are thinking about migrating to London or even if they want to relocate their neighborhoods within the city. Our findings will help stakeholders make informed decisions and address any concerns they have including the different kinds of cuisines, stores, leisure activities, and what the neighborhood has to offer.

Data Description

To determine the popular locations in London, we need the geographical location data of London, the popularity of each location, and the information of each location. Three sources of databases will be drawn upon as follows:

1. Wikipedia

Wikipedia provides us with information on all neighborhoods, although we will limit our analysis to London (https://en.wikipedia.org/wiki/List_of_areas_of_London). With the list of neighborhoods and postal codes, we can analyze the most popular venues in each category. Some of the data we will be using from this database are neighborhood, borough, and postal codes, such as Bexley Greenwich (borough) - London (town) - WE2 (post code).

borough: Name of Neighbourhood
town: Name of the borough
post_code: Postal codes for London.

2. ArcGIS API

ArcGIS API allows for the analysis regarding people and locations. With this API, we can use interactive maps to gain some insights. In this case, ArcGIS allows us to plot the various locations on the map using the latitude and longitude information, such as (for Bexley) 51.49245 deg latitude and 0.12127 deg longitude.

latitude: Latitude for Neighbourhood
longitude: Longitude for Neighbourhood

3. Foursquare API

After having the locations from the previous two databases, we can get more specific information using Foursquare API, including photos. Based on all the information collected, a clustering model will be utilized to uncover the patterns based on similar venue categories. Thus, a piece of important information from this database is the venue category of each location, such as (for Bexley) "supermarket", "historic site", and "coffee shop."

Neighbourhood : Name of the Neighbourhood
Neighbourhood Latitude : Latitude of the Neighbourhood
Neighbourhood Longitude : Longitude of the Neighbourhood
Venue : Name of the Venue
Venue Latitude : Latitude of Venue
Venue Longitude : Longitude of Venue
Venue Category : Category of Venue

Methodology

The approach to conduct the analysis is separated into 3 main steps as follows.

1. Gather the required data of London
2. Plot the map to show the neighborhoods being considered
3. Build our model by clustering all of the similar neighborhoods together and visualize the clusters

1. Gathering Data

1.1 London Neighborhoods

First, we begin by collecting the list of areas of London from the Wikipedia page as shown below.

	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
...
526	Woolwich	Greenwich	LONDON	SE18	020	TQ435795
527	Worcester Park	Sutton, Kingston upon Thames	WORCESTER PARK	KT4	020	TQ225655
528	Wormwood Scrubs	Hammersmith and Fulham	LONDON	W12	020	TQ225815
529	Yeadon	Hillingdon	HAYES	UB4	020	TQ115825
530	Yiewsley	Hillingdon	WEST DRAYTON	UB7	020	TQ063804

Figure 2: Collected Data About London From Wikipedia

Then, we process the data by removing spaces in the column titles and adding “_” instead. We only need some columns, so we will get rid of the rest. The required columns are London boroughs, Post town, and Postcode district. In addition, the information other than “London” are removed, since we are only interested in London. After ridding the unnecessary data columns, the following table is obtained.

	Borough	Town	Post_code
0	Bexley, Greenwich	LONDON	SE2
1	Ealing, Hammersmith and Fulham	LONDON	W3, W4
6	City	LONDON	EC3
7	Westminster	LONDON	WC2
9	Bromley	LONDON	SE20
...
521	Redbridge	LONDON	IG8, E18
522	Redbridge, Waltham Forest	LONDON, WOODFORD GREEN	IG8
525	Barnet	LONDON	N12
526	Greenwich	LONDON	SE18
528	Hammersmith and Fulham	LONDON	W12

Figure 3: Modified Data About London From Wikipedia. The “Postcode district”, “Dial code”, and “OS grid ref” are removed. Only the town “London” are selected.

1.2 Geopositions of London Neighborhoods

The latitudes and longitudes of London neighborhoods are required to plot in the geographical map. We will utilize the ArcGIS package in this regard. After obtaining the latitudes and longitudes of each borough, we append it to our data table from Wikipedia as shown below.

	Borough	Town	Post_code	Latitude	Longitude
0	Bexley, Greenwich	LONDON	SE2	51.49245	0.12127
1	Ealing, Hammersmith and Fulham	LONDON	W3, W4	51.51324	-0.26746
6	City	LONDON	EC3	51.51200	-0.08058
7	Westminster	LONDON	WC2	51.51651	-0.11968
9	Bromley	LONDON	SE20	51.41009	-0.05683
...
521	Redbridge	LONDON	IG8, E18	51.58977	0.03052
522	Redbridge, Waltham Forest	LONDON, WOODFORD GREEN	IG8	51.50642	-0.12721
525	Barnet	LONDON	N12	51.61592	-0.17674
526	Greenwich	LONDON	SE18	51.48207	0.07143
528	Hammersmith and Fulham	LONDON	W12	51.50645	-0.23691

Figure 4: The Data of Each Borough in London with Latitude and Longitude.

2. Visualizing London

We use the folium package to help visualize London, which is shown below.

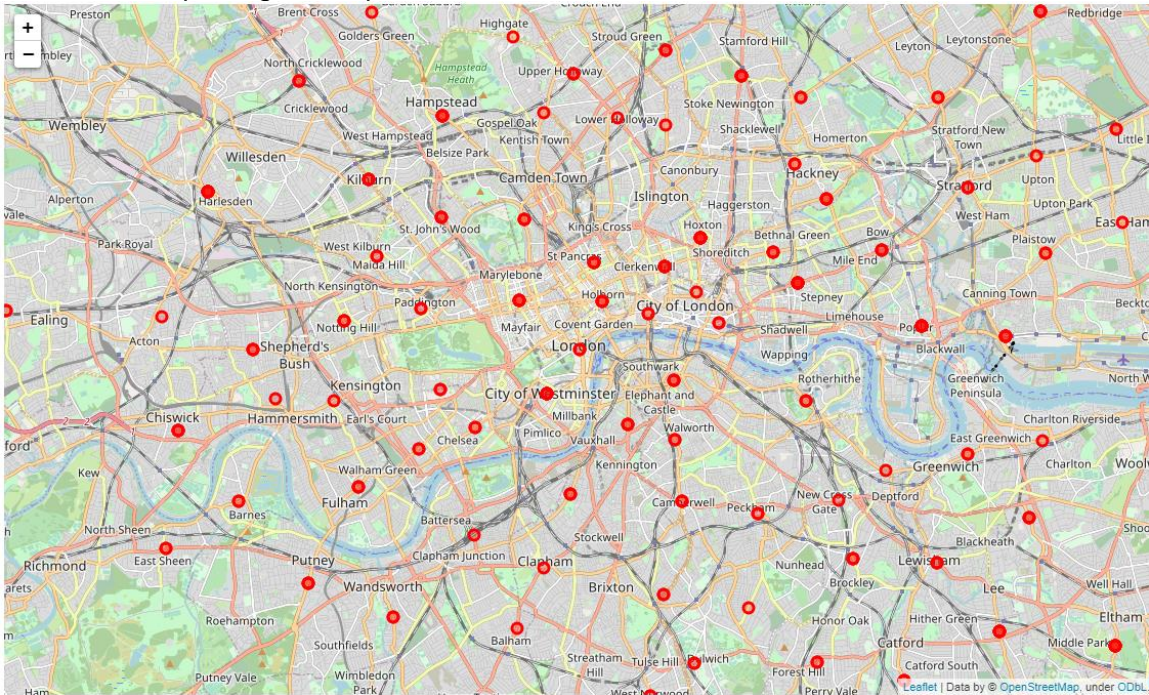


Figure 5: The Town of London And Its Boroughs

Once we have the visualization, it is time to integrate the Foursquare API. We will get the venues and their categories from the Foursquare API around the neighborhoods in London. All the venues and venue categories were acquired from Foursquare API as shown in the figure below.

	Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Category
0	Bexley, Greenwich	51.49245	0.12127	Lesnes Abbey	Historic Site
1	Bexley, Greenwich	51.49245	0.12127	Sainsbury's	Supermarket
2	Bexley, Greenwich	51.49245	0.12127	Lidl	Supermarket
3	Bexley, Greenwich	51.49245	0.12127	Abbey Wood Railway Station (ABW)	Train Station
4	Bexley, Greenwich	51.49245	0.12127	Bean @ Work	Coffee Shop
...
10405	Hammersmith and Fulham	51.50645	-0.23691	Mleczko Polish Deli	Deli / Bodega
10406	Hammersmith and Fulham	51.50645	-0.23691	Nut Case	Gourmet Shop
10407	Hammersmith and Fulham	51.50645	-0.23691	New Sweet'n'Sour Chinese Takeaway	Chinese Restaurant
10408	Hammersmith and Fulham	51.50645	-0.23691	The Vine Leaves Taverna	Greek Restaurant
10409	Hammersmith and Fulham	51.50645	-0.23691	QPR Superstore	Sporting Goods Shop

Figure 6: Venue and Venue Categories of All Neighborhoods in London from Foursquare API

3. Grouping by Venue Categories and K-Means Clustering

To get better results for clustering, we utilize the dummy variables. Then, we group the data by neighborhood and calculate the mean of each dummy variable for each neighborhood. Since there are many venues in each neighborhood, the top five most popular venues are selected for analysis. The examples of the most popular venues in each neighborhood are shown in the table below.

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Barnet	Coffee Shop	Café	Grocery Store	Pub	Italian Restaurant
1	Barnet, Brent, Camden	Clothing Store	Supermarket	Hardware Store	Gym / Fitness Center	Convenience Store
2	Bexley	Supermarket	Historic Site	Convenience Store	Coffee Shop	Train Station
3	Bexley, Greenwich	Daycare	Convenience Store	Food Service	Golf Course	Massage Studio

Figure 7: The Five Most Popular Venues In Each Neighborhood of London

K-Means Clustering Model

To classify the data even further, we will classify with 5 clusters with K-means technique based on the most popular venues in each neighborhood. After that, we visualize the data on the map. The results from implementing the K-Means model are appended to the original table along with the most popular venues. Then, each cluster is visualized on the map. Note that the clusters are not based on the distance, but rather the most common venues in each neighborhood.

Borough	Town	Post_code	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
Bexley, Greenwich	LONDON	SE2	51.49245	0.12127	4	Supermarket	Historic Site	Coffee Shop	Train Station	Platform
Ealing, Hammersmith and Fulham	LONDON	W3, W4	51.51324	-0.26746	2	Grocery Store	Indian Restaurant	Train Station	Breakfast Spot	Hotel
City	LONDON	EC3	51.51200	-0.08058	1	Hotel	Coffee Shop	Italian Restaurant	Gym / Fitness Center	Pub
Westminster	LONDON	WC2	51.51651	-0.11968	1	Hotel	Coffee Shop	Café	Pub	Sandwich Place
Bromley	LONDON	SE20	51.41009	-0.05683	1	Supermarket	Grocery Store	Hotel	Convenience Store	Fast Food Restaurant
...
Redbridge	LONDON	IG8, E18	51.58977	0.03052	1	Pub	Grocery Store	Café	Pizza Place	Bakery
Redbridge, Waltham Forest	LONDON, WOODFORD GREEN	IG8	51.50642	-0.12721	1	Theater	Hotel	Pub	Monument / Landmark	Garden
Barnet	LONDON	N12	51.61592	-0.17674	1	Coffee Shop	Café	Grocery Store	Pub	Italian Restaurant
Greenwich	LONDON	SE18	51.48207	0.07143	1	Pub	Grocery Store	Bus Stop	Indian Restaurant	Coffee Shop
Hammersmith and Fulham	LONDON	W12	51.50645	-0.23691	1	Pub	Coffee Shop	Café	Grocery Store	Gastropub

Figure 7: The Data Of Each Borough In London, Its Cluster, And The Most Popular Venues In Each Borough

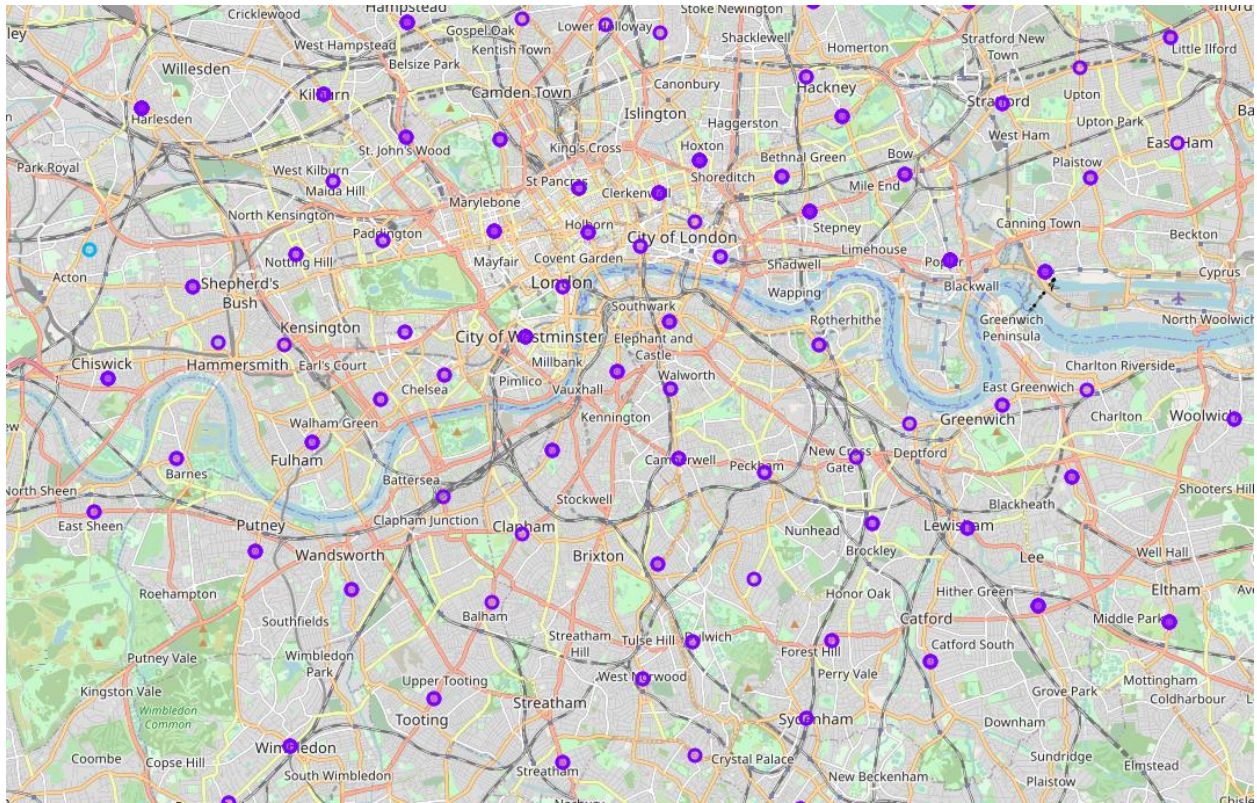


Figure 8: Visualization Of Each Cluster On The Map Of London

Results and Discussions

London has become more multicultural over the years. It has a variety of restaurants, bars, coffee shops, and breakfast places. In terms of shopping, London has many attractions, such as flower shops, fishing stores, and clothing stores. The most popular transportation methods are buses and trains. To relax, visitors and residents can go to parks, zoos, gyms, and historic sites. The city of London offers a multicultural and entertaining experience for all groups of people. In addition, since the clusters aren't based on the locations, but on the venue categories, it is nice to see that one can get a variety of experience (visualized with different clusters) within the nearby neighborhoods too.

Conclusions

The goal of this project is to explore London and look at how attractive it is to visitors and potential immigrants. It was explored based on the postal codes and common venues present in each of the neighborhoods are then determined. Finally, we used K-means to cluster similar neighborhoods together. London has a wide variety and uniqueness of experiences to offer. The cultural diversity allows ones to develop a sense of belonging in no time.

This approach can be used in the same manner with other cities that the reader is interested in. The author invites the reader to try the approach out and explore the city from the perspective of a data scientist.