



THE DATA- DRIVEN GUIDE TO LIVING IN LONDON

**IBM - APPLIED DATA SCIENCE
CAPSTONE PROJECT**

Worawan Thanasanchiraphong
October 2019



Introduction

London is considered to be one of the world's most important global cities and has been termed the world's most powerful, most desirable, most influential, most visited, most expensive, innovative, sustainable, most investment friendly, and most popular for work city in the world. London has a diverse range of people and cultures, and more than 300 languages are spoken in the region. Its estimated mid-2018 municipal population (corresponding to Greater London) was 8,908,081. [1] London's population has grown 7.5% in just five years. [2]

Moving to a new city can be difficult, so the purpose of this project is to guide those people who are moving to London to find a nice, affordable area to live. We will analyze **32 boroughs, plus City of London** by creating a choropleth map with average rent (one-bedroom apartment) in each borough, plotting London Underground stations and finding most common venues in each borough.



Data Description

To analyze each borough in London, I gather all data below:

- Collecting London borough names from **London Councils** [3] and using **Google Maps** [4] to manually get coordinates in each borough then put it in CSV file (Sometimes Geocoder doesn't provide accurate coordinates)
- Downloading data of average rent in London by borough in CSV format from **London Datastore** [5] (In this project, we will use only average rent of 1-bedroom apartment)
- Using BeautifulSoup to scrape London Underground stations and coordinates (both Public Domain Tube Stations & Public Domain DLR Stations List) from **OpenStreetMap** [6]
- Using **Foursquare API** [7] to obtain surrounding venues in every borough in London
- Using geojson file from **carto.com** [8] to draw polygons for London boroughs (Geojson file will be used for creating choropleth map)

Methodology

First, import the list of average rent in London, by borough and import the list of borough names and borough coordinates. Then create dataframes from these two lists and merge these two dataframes together. Use **BeautifulSoup** and **Requests** to scrape a list of London Underground stations and locations from OpenStreetMap Wikipedia.

To create a map, use **Geocoder** to fetch a location of London city. Create a choropleth map with **Folium** to show average rent in each borough from the dataframe and use geojson file to draw polygons of London boroughs. Add a layer of clusters showing density of London Underground stations to the map.

Use **Foursquare API** to get the top 100 venues in a radius of 1000 meters of particular boroughs. Prepare the data for **K-means** clustering by grouping the dataframe by boroughs. K-means clustering is performed on this data set to return 3 clusters, or characteristics of boroughs in terms of most common type of venues.

Lastly, plot center location of each borough including City of London with different color depending on most common venue cluster the borough has. The popups are added for showing borough name, most common venues, and average rent.

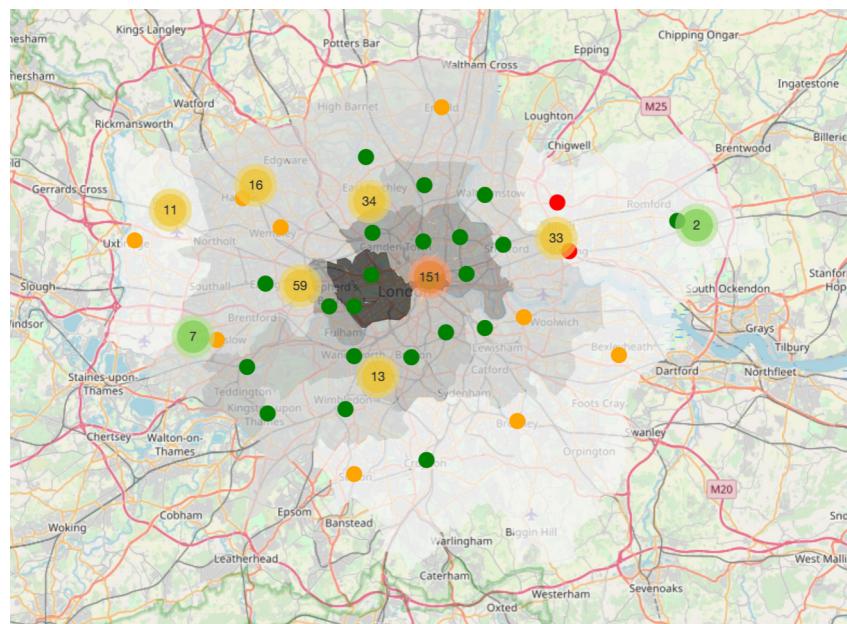
Result

The output from the clustering algorithm shows the division of all the boroughs into 3 groups in terms of most common venues present in that area.

- Cluster 0: Boroughs with 'Pizza and Hotel Venues'
- Cluster 1: Boroughs with 'Intensive Pub & Coffee Shop Venues'
- Cluster 2: Boroughs with 'Pub & Coffee Shop Venues'

The choropleth map was created to show average rent by borough with clusters showing London Underground station density. Borough center locations were added to the map with popups contain information of borough name, most common venues, and average rent.

The results are visualised as follows:



- Cluster 0 is in red
- Cluster 1 is in green
- Cluster 2 is in orange

Discussion

Those boroughs which are closer to the center of London city tend to have more expensive average rent, higher density of Underground stations and a significant number of pub and coffee shop venues.

And of course, those boroughs which are far away from the city center have cheaper average rent, lower density of Underground stations and less number of pub and coffee shop venues.

In London borough of Sutton, Croydon, Bromley, and Bexley does not have any Underground stations.

For improvement in the future, this project could be done in neighborhood level with unlimited venue data from Foursquare API. It would give more accurate result.

Conclusion

Every year, many people move into London to work or study. To choose an area for living, it totally depends on their budget, choices of transportation, type of surrounding venues, borough characteristics, etc. For this reason, people can make a decision based on data provided and their personal preferences.

References:

- [1] London - Wikipedia
<https://en.wikipedia.org/wiki/London>
- [2] Trust for London
<https://www.trustforlondon.org.uk/data/people-moving-and-london/>
- [3] London Councils
<https://www.londoncouncils.gov.uk/node/1938>
- [4] Google Maps
<https://www.google.co.th/maps/>
- [5] London Datastore
<https://data.london.gov.uk/dataset/average-private-rents-borough>
- [6] List of London Underground stations - OpenStreetMap
https://wiki.openstreetmap.org/wiki/List_of_London_Underground_stations
- [7] Foursquare API
<https://developer.foursquare.com/>
- [8] London Boroughs Geojson
https://cooleyb.carto.com/tables/london_boroughs_boroughs_kml/public