Applied Data Science Capstone Report

The Battle of Neighborhoods, Week5 - A Data Journey to London

Ryo Kajiura

16.07.2020

INTRODUCTION

Background

Moving to a different country is always a difficult event in life, especially when you have not been there before. Nowadays it is a lot easier than before thanks to the internet and google, however, collecting proper information of the neighborhood is still not as easy as finding an apartment sored by the range of renting costs in the local website. It is worth checking the official database from the government and extract useful information but this is very time consuming.

The other day I received an email from a friend of mine in Japan and he is planning to move to London as he got an attractive offer. He has never been there until now and he has a family with a small child. So here is already a case of above…

Problem

My friend wants to find "good" areas in London to start searching for an apartment, however, he has no clue where and how to start it. He knows that the neighborhood needs to be safe and they like to have good access to Asian food. So I decided to help him with my recently learned skill of Data Analysis.

Interest

This research and analysis is to find the area(s) inside or around the center of London where the crime rate is low and the access to Asian cuisines are good.

Data Acquisition

Data Acquisition

The main data required for this research and analysis are taken from the websites shown below and .csv/.xls files are locally stored for the easiness of the access.

1. **The official crime record from the UK government**
   https://data.london.gov.uk/dataset/recorded_crime_summary
2. **The geographical coordinate of Boroughs and Ward in London**
   https://data.london.gov.uk/dataset/excel-mapping-template-for-london-boroughs-and-wards
3. **The venue data in the areas of interest**
   These are acquired through Foursquare API by using the sorted geometrical coordinates from 1 and 2 above.

Methodology

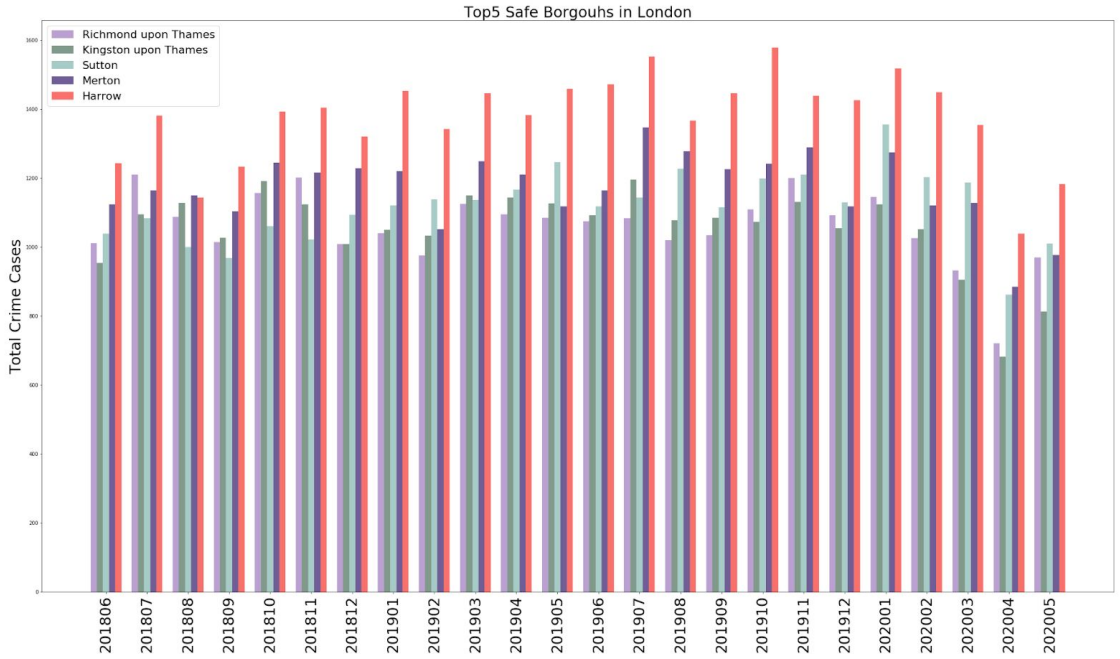
Data Cleaning and Analysis

The crime record from the government consists of the name of Boroughs, major categories of crimes and detailed categories. The case numbers are shown from June 2018 until May 2020. Firstly, the categorizations of crime are removed and the case numbers are sorted with Boroughs as we would like to see roughly how the crimes are distributed around London. After that, the deeper investigation of the category dependencies will be considered if needed.

There are 33 Boroughs inside and around the center of London. After applying the cleaning of data above, the table below is acquired. The data between October 2018 and December 2019 are not shown only for the visualization purpose.

Borough	201806	201807	201808	201809	202001	202002	202003	202004	202005	Ave.
London Heathrow and London City Airports	286	312	280	277	286	292	267	54	38	269.958333
Kingston upon Thames	953	1095	1127	1027	1124	1052	905	682	813	1054.625000
Richmond upon Thames	1011	1210	1087	1014	1145	1026	932	720	969	1058.750000
Sutton	1038	1083	999	968	1355	1203	1187	862	1010	1117.750000
Merton	1123	1163	1149	1103	1274	1120	1127	884	977	1171.458333
Harrow	1243	1381	1144	1232	1517	1448	1353	1039	1182	1375.666667
Bexley	1334	1238	1243	1263	1619	1461	1284	904	1209	1409.666667

From this table, we can see that “London Heathrow and London City Airports” is the safest borough in the area of our interests, however, it is obviously not a common place for accommodation. So this borough is excluded for further analysis and from the second to the six boroughs are used. In order to make sure that taking the average of monthly cases are reasonable, the total case numbers per month in each Borough are visualized as below. The crime cases in April and May 2020 are notably smaller than usual and this is highly likely due to the lock-down of the city of London because of Covid-19 countermeasure.

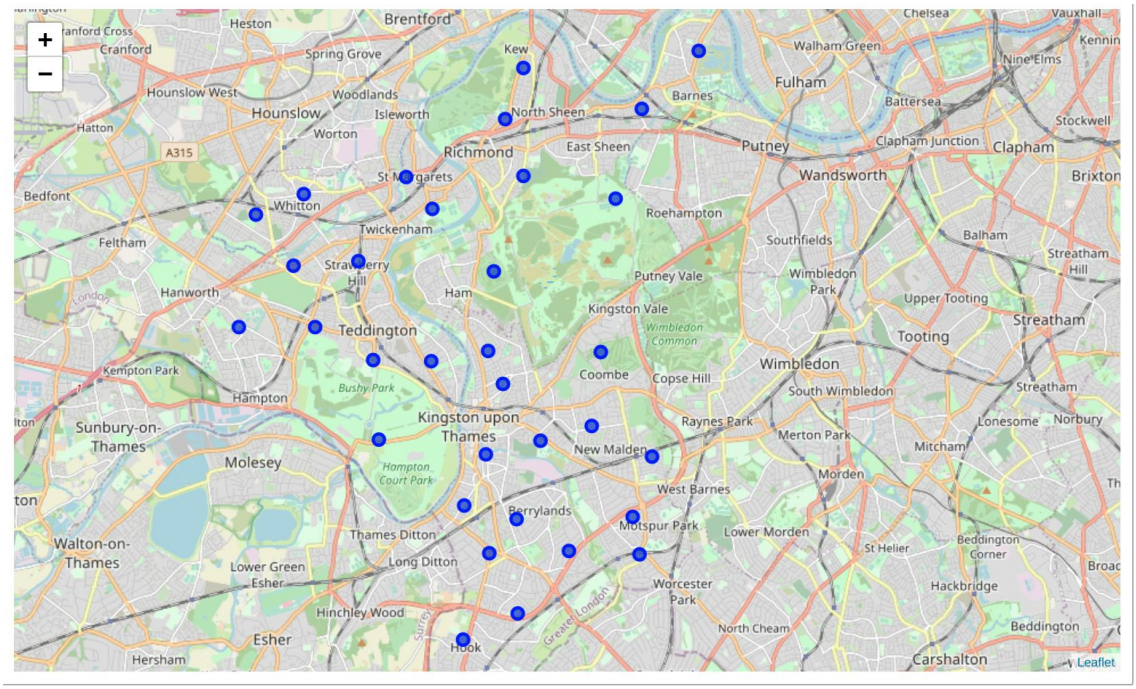


With all the discussion above, it should be safe to say that choosing the top two safe boroughs, **Richmond upon Thames and Kingston upon Thames**, for further analysis is reasonable.

Exploring around Richmond upon Thames and Kingston upon Thames

Mapping

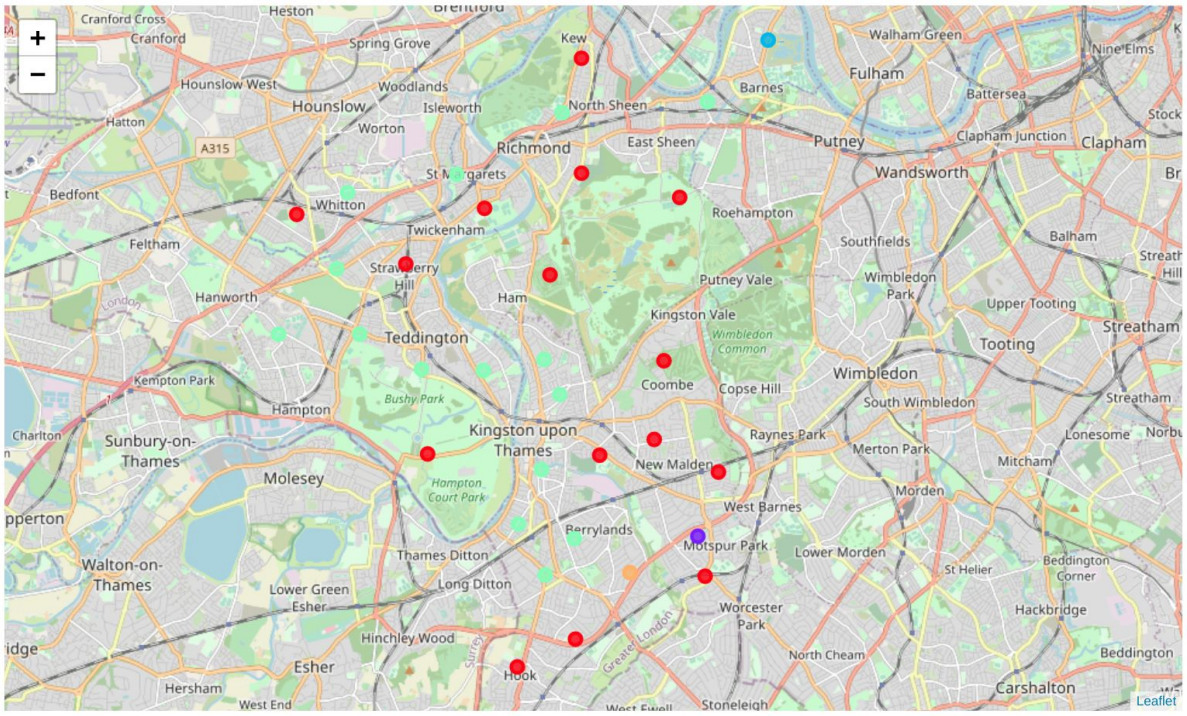
Now we need more detailed geographical information to feed to the Foursquare API. Unfortunately there was no one clean data containing boroughs, wards, longitude and latitude. After merging and cleaning the data from two xls files, the geographical coordinates of 34 Wards are acquired. The map below shows the locations of Wards of interests.



Venue Data Acquisition through Foursquare API

The data of venues around Wards are acquired by using Foursquare API. The date I chose is 10th of January 2020 as the information after February might be already biased because of the lockdown of the city due to the Covid-19.

From 34 location data of Ward, I could get 322 venue data with 95 unique categories by applying 500m radius from the geographical centers of each Ward. Clustering of Wards was done with the number of k=5.



Result

After clustering and looking at the venue categories, I could find two Wards, **"Beverley" and "North Richmond",** to recommend to my friend to start a new career in London.

```
df_c0_asian
```

	Ward	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	Beverley	Korean Restaurant	Supermarket	Gym Pool	Café	Grocery Store	Gym / Fitness Center	Indian Restaurant	Japanese Restaurant	Karaoke Bar	Fast Food Restaurant
1	Chessington North And Hook	Park	Bakery	Supermarket	Indian Restaurant	Fish & Chips Shop	Wine Shop	Food	Diner	Exhibit	Farmers Market
5	Norbiton	Soccer Stadium	Gym / Fitness Center	Fried Chicken Joint	Hotel	Indian Restaurant	Food Truck	Breakfast Spot	Rental Car Location	Food	Fish & Chips Shop
6	Old Malden	Japanese Restaurant	Train Station	Grocery Store	Park	Bakery	Steakhouse	Flower Shop	Diner	Exhibit	Farmers Market
11	Heathfield	Gym	Soccer Field	Pizza Place	Playground	Indian Restaurant	Convenience Store	Department Store	Diner	Exhibit	Farmers Market
15	Twickenham Riverside	Café	Gastropub	French Restaurant	Art Gallery	Park	Playground	Pub	Chinese Restaurant	Bus Stop	Deli / Bodega

```
df_c3_asian
```

	Ward	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
3	St. Mark's	Coffee Shop	Pub	Grocery Store	Pharmacy	Breakfast Spot	Gym / Fitness Center	French Restaurant	Indian Restaurant	Fish & Chips Shop	Farmers Market
6	Fulwell and Hampton Hill	Grocery Store	Butcher	Italian Restaurant	Fast Food Restaurant	Coffee Shop	Pub	Chinese Restaurant	Wine Shop	Bistro	Boat or Ferry
10	North Richmond	Garden	Pub	Restaurant	Hotel	Rugby Stadium	Korean Restaurant	Italian Restaurant	Chinese Restaurant	Japanese Restaurant	Food Truck
11	St. Margaret's and North Twickenham	Pub	Comedy Club	Seafood Restaurant	Park	Deli / Bodega	Coffee Shop	Chinese Restaurant	Café	Grocery Store	Boat or Ferry

Discussion

The key conditions of this research was to identify the areas with low crime rate and good access to Aisan cuisine. It could have been better if other living factors were taken into account, such as renting cost and public transportation situations.

I could have included other Boroughs so that more Wards could have been taken into account and the clustering might have shown more interesting results.

Conclusion

I can now propose to my friend two Wards for possible accommodation inside London with the low crime rate and good access to Asian cuisines.