Battle of the Neighbourhoods

Martin Macke

January 2021

1. Introduction

London is city that is diverse in culture and famous for its numerous eateries and restaurants. It is made up of 32 Boroughs and the City of London. As an investor who is interested in opening up a business in one of the London Boroughs, it would be good to know what venues are popular in each borough and if restaurants are popular, what sort of cuisine is popular in that borough.

2. Data Description

London has a total of 32 Boroughs. The names of the boroughs were collected from the Wikipedia page: https://en.wikipedia.org/wiki/London_boroughs .
The python geopy library was used to gain individual latitude and longitude co-ordinates for each borough. A new dataset was created by merging the London borough names and their respective co-ordinates. Using the co-ordinates, and the FourSquare API, the top venue categories for each borough were obtained and the corresponding data was transformed into the number of unique categories curated from all the returned venues.

3. Methodology

3.1 Data Cleaning

Several steps were taken to extract the data.
The beautifulsoup library was used to extract the names of 32 london boroughs from the Wikipedia page.
Using the details of the extracted boroughs, the geopy library was used to obtain individual co-ordinates for each borough.
The visualisation library Folium was used to mark each borough on the map of London
Using the Foursquare API, the neighbourhoods were explored in order to obtain the top venues in each borough.
The resulting venues information was transformed into a pandas dataframe.

	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Barking and Dagenham	51.554117	0.150504	Central Park	51.559560	0.161981	Park
1	Barking and Dagenham	51.554117	0.150504	Lara Grill	51.562445	0.147178	Turkish Restaurant
2	Barking and Dagenham	51.554117	0.150504	Asda	51.565770	0.143393	Supermarket
3	Barking and Dagenham	51.554117	0.150504	BP	51.549951	0.161963	Gas Station
4	Barking and Dagenham	51.554117	0.150504	Iceland	51.560578	0.147685	Grocery Store

3.2 Data Exploratory Analysis

The data frame obtained from the data cleaning process was then analysed to find out how many unique venue categories could be curated from all the returned venues.

The One-Hot encoding method was used to analyse each borough and the venue categories.

how many unique categories can be curated f Submit Notebook ... ned venues

[15]:
```
print('There are {} uniques categories.'.format(len(london_venues['Venue Category'].unique())))
```

```
There are 256 uniques categories.
```

Analyze Each Neighborhood

[16]:
```
# one hot encoding
london_onehot = pd.get_dummies(london_venues[['Venue Category']], prefix="", prefix_sep="")

# add borough column back to dataframe
london_onehot['Borough'] = london_venues['Borough']

# move neighborhood column to the first column
fixed_columns = [london_onehot.columns[-1]] + list(london_onehot.columns[:-1])
london_onehot = london_onehot[fixed_columns]

london_onehot.head()
```

Each borough was grouped in a row by taking into account the frequency of occurrence of each category using the python groupby function. The resulting data was sorted to find the top 10 venues for each borough

top 10 venues for each neighborhood

[22]:
```
num_top_venues = 10

indicators = ['st', 'nd', 'rd']

# create columns according to number of top venues
columns = ['Borough']
for ind in np.arange(num_top_venues):
    try:
        columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
    except:
        columns.append('{}th Most Common Venue'.format(ind+1))

# create a new dataframe
neighborhoods_venues_sorted = pd.DataFrame(columns=columns)
neighborhoods_venues_sorted['Borough'] = london_grouped['Borough']

for ind in np.arange(london_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(london_grouped.iloc[ind, :], num_top_venues)

neighborhoods_venues_sorted
```

[22]:

	Borough	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 (
0	Barking and Dagenham	Grocery Store	Gas Station	Supermarket	Park	Pub	Bar	Furniture / Home Store	Café	Cof
1	Barnet	Coffee Shop	Pub	Grocery Store	Pizza Place	Pharmacy	Golf Course	Café	Bookstore	Sandw

The K-means clustering algorithm was used to cluster the boroughs into 4 clusters.

Cluster Neighborhoods

run kmeans to cluster the boroughs into 4 clusters

```
[23]: # set number of clusters
kclusters = 4

london_grouped_clustering = london_grouped.drop('Borough', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(london_grouped_clustering)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
```

```
[23]: array([2, 2, 0, 0, 0, 1, 0, 1, 0, 1], dtype=int32)
```

create a new dataframe that includes the cluster as well as the top 10 venues for each borough

```
[24]: # add clustering labels
neighborhoods_venues_sorted.insert(0, 'Cluster Label', kmeans.labels_)

london_merged = df

# merge london_grouped with london_data to add latitude/longitude for each neighborhood
london_merged = london_merged.join(neighborhoods_venues_sorted.set_index('Borough'), on='Borough')

london_merged.head() # check the last columns!
```

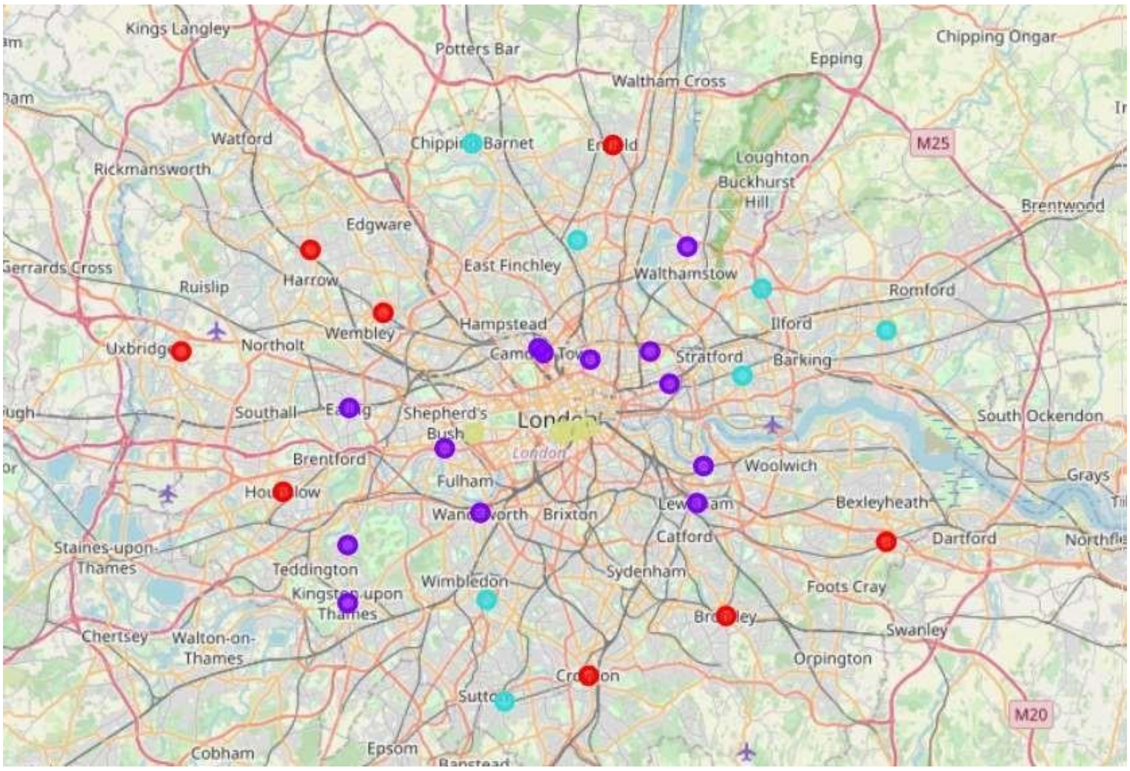
[24]:

	Borough	Local Authority	Population Estimate 2013	Latitude	Longitude	Cluster Label	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	Barking and Dagenham	Barking and Dagenham London Borough Council	194,352	51.554117	0.150504	2	Grocery Store	Gas Station	Supermarket	Park	Pub	Bar	Furniture / Home Store	Café	Coffee Shop	Shopping Mall
1	Barnet	Barnet London Borough Council	369,088	51.653090	-0.200226	2	Coffee Shop	Pub	Grocery Store	Pizza Place	Pharmacy	Golf Course	Café	Bookstore	Sandwich Place	Park
2	Bexley	Bexley London Borough Council	236,687	51.441679	0.150488	0	Pub	Fast Food Restaurant	Clothing Store	Chinese Restaurant	Hotel	Supermarket	Coffee Shop	Italian Restaurant	Greek Restaurant	Pharmacy

Each borough was defined according to what the most popular venue was, and what restarurants, if any, were in the 3 most common venue categories for each borough.

4. Results

Image showing cluster results



Cluster 0 (Red) : Pubs and coffee shops, followed by Indian restuarants in the top 2 most common venues in some boroughs
Cluster 1(Purple) : Most common venue is pubs,followed by coffee shops but 3rd most common venue in many boroughs is restaurants
Cluster 2(Blue): Grocery stores and Pubs, fast food and Italian restaurant in 2nd and 3rd most common venues for some boroughs
Cluster 3(Yellow): Hotels

5. Recommendations

Based on the results, there are several recommendations that can be made to the stakeholders. One of the most common venues in Cluster 0 and Cluster 1 is coffee shops. This indicates that coffee shops could be a good business to invest in.

Given the current world circumstances, restaurants are not coming up as the most common venue in any of the boroughs expect in Hounslow Borough.

There are some boroughs where restuarnats are in the 2nd most common venue. In Cluster 0, Indian Restaurants are the 2nd most popular venue in 2 of the boroughs. This suggests that Indian Restaurants can have potential as a business venture in these boroughs, or boroughs with similar demographics could be worth exploring.

Further analysis of boroughs in clusters 0 and 1 to see coffee shop density and also socioeconomic demographics can be carried out. This would allow for more detailed infromation on which boroughs and specific locations in the boroughs would be ideal to open up a coffee shop in, should the investors decide to want to go down the coffee shop route.

6. Conclusion

This project had given a brief insight on how data analysis can be used to scout for business opportunities.

To build the initial dataset I used various python libraries like beautfifulsoup to extract the web data. Geopy was used to obtain the co-ordinated for each London Borough. The Foursquare API to obtain the top venue.
Analysis was performed using Python and its libraries like Pandas and Scikit.
The data was visualised using Folium.
The results from the analysis has provided information on which boroughs are similar in terms of venues frequented, and has also identified popular venues that can be a good business opportunity.

7. References

https://en.wikipedia.org/wiki/London_boroughs

A link to the notebook showing the code and all the python libraries used in this project:
https://github.com/shahs1005/Battle-of-the-Neighbourhoods/blob/master/Capstone(1).ipynb