Battle of Neighborhoods between London and Toronto

[People Interests and localities Analysis]

Akhil Patil

December 3rd, 2019

1. Introduction

1.1 Background

London is the capital and largest city of England and the United Kingdom. Standing on the River Thames in the south-east of England, at the head of its 50-mile (80 km) estuary leading to the North Sea, London has been a major settlement for two millennia. 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, the most populous of any city in the European Union and accounting for 13.4% of the UK population.

Toronto is the provincial capital of Ontario and the most populous city in Canada, with a population of 2,731,571 as of 2016. Current to 2016, the Toronto census metropolitan area (CMA), of which the majority is within the Greater Toronto Area (GTA), held a population of 5,928,040, making it Canada's most populous CMA. The diverse population of Toronto reflects its current and historical role as an important destination for immigrants to Canada. More than 50 percent of residents belong to a visible minority population group, and over 200 distinct ethnic origins are represented among its inhabitants. While the majority of Torontonians speak English as their primary language, over 160 languages are spoken in the city.

1.2 Problem

Both the cities are financial capitals of respective countries and are diverse in cultures, as well as these cities have many visitors every year. So, aim of the project is to analyze the similar and dissimilar interests of the people living in both the cities and determine the localities that are popular for some specific venue categories.

1.3 Interested Audience

I am a Traveller myself, I like to visit places with significance importance and explore their beauty, these both cities are very popular and have been experiencing increase in visitors overall over years. Those who are planning or interested in visiting these cities might be interested what are good localities and local people interest overall in the both cities.

2. Data Acquisition and Cleaning

2.1 Data Sources

For our analysis we will require data of neighborhoods of both the cities which I have taken from Wikipedia. We will require attributes such as name of neighborhoods, boroughs, postal codes of

the neighborhoods. These all attributes are provided in the form of table in Wikipedia. We will scrape these table using beautiful soup library in python.

After scraping London city data from Wikipedia link here our table looks like below:

	Location	borough	town	Postcode_district	Dial_code	OS_grid_ref
0	Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2	20	TQ465785
1	Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4	20	TQ205805
2	Addington	Croydon[8]	CROYDON	CR0	20	TQ375645
3	Addiscombe	Croydon[8]	CROYDON	CR0	20	TQ345665
4	Albany Park	Bexley	BEXLEY, SIDCUP	DA5, DA14	20	TQ478728

After scraping London city data from Wikipedia link here our table looks like below:

	PostalCode	Borough	Neighborhood
0	M1A	Not assigned	Not assigned
1	M2A	Not assigned	Not assigned
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront

2.2 Data Cleaning

Since our data was scraped from Wikipedia, we will not have much work dealing with Null values. But there are some "Not Assigned" values in Boroughs and Neighborhoods column. So, to deal with those Firstly, I have dropped the "Not Assigned" Boroughs and for Neighborhoods, I have assigned borough name to Not Assigned values of Neighborhoods. Also, we have multiple neighborhoods for a single borough. So, every row must have a single borough for that I have combined multiple neighborhoods that have same boroughs to one record.

We have multiple towns in London and same in Toronto I have sliced both the data frames to focus more on city centers rather than the outskirts of the cities. So, in Toronto I have just kept boroughs that have name Toronto in it i.e. Central, East, West and Downtown Toronto. Similarly, in London I have just kept rows that have town column as London.

Finally, we have different names for the columns for neighborhoods like location in London data frame. So, just to be clear I have renamed all the attributes and simplified them to "Neighborhood", "Postcode" and "Borough".

Now further there are some other attributes than that required for analysis, like in London scraped has dial_code, OS_grid_ref and town which we don't require for analysis. So, I have just selected "Postal code", "Borough" and Neighborhood columns in both the data frames.

After Data Cleaning your resulting datasets looks like below:

London:

	Postcode	Borough	Neighborhoods
0	DA5	Dartford	Dartford
1	E1	Tower Hamlets	Mile End,Ratcliff,Shadwell,Spitalfields,Stepne...
2	E10	Hackney	Lea Bridge
3	E10, E15	Waltham Forest	Leyton
4	E11	Redbridge	Wanstead

Toronto:

	PostalCode	Borough	Neighborhood
2	M3A	North York	Parkwoods
3	M4A	North York	Victoria Village
4	M5A	Downtown Toronto	Harbourfront
5	M6A	North York	Lawrence Heights
6	M6A	North York	Lawrence Manor

3. Exploratory Data Analysis

3.1 Getting the Latitude and Longitude for each Neighborhood.

I am using the geocoder library in python to perform this task. Once we get all the latitude and longitude we then proceed for main part of our project.

3.2 Getting venues information for each neighborhood in both the cities.

Further based on the postal codes and neighborhoods, I got Latitude and Longitude from geocoder. Once, I have latitude and longitude for each postal code in both the cities we can proceed to get the venues data that will enable us to get the popular venues i.e. we will term as people's interests in the cities. Foursquare api will give us categories of the venues which is easy to work with.
Final data frame which will be ready for analysis will look like this for both the cities:

London :

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Leyton	51.55885	-0.00733	Leyton Technical	51.558781	-0.007328	Pub
1	Leyton	51.55885	-0.00733	Deeney's to go	51.561253	-0.009889	Sandwich Place
2	Leyton	51.55885	-0.00733	Coronation Gardens	51.559517	-0.010643	Park
3	Leyton	51.55885	-0.00733	Oceano Restaurant	51.561250	-0.010059	Mediterranean Restaurant
4	Leyton	51.55885	-0.00733	Leyton Orient Supporters Club	51.559931	-0.013497	Sports Bar

Toronto:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	The Beaches	43.676357	-79.293031	Glen Manor Ravine	43.676821	-79.293942	Trail
1	The Beaches	43.676357	-79.293031	The Big Carrot Natural Food Market	43.678879	-79.297734	Health Food Store
2	The Beaches	43.676357	-79.293031	Grover Pub and Grub	43.679181	-79.297215	Pub
3	The Beaches	43.676357	-79.293031	Upper Beaches	43.680563	-79.292869	Neighborhood
4	The Danforth West,Riverdale	43.679557	-79.352188	Pantheon	43.677621	-79.351434	Greek Restaurant

We get the venues and their categories that are popular in that neighborhoods.

3.3 Generating word cloud for checking the most popular categories in cities.

Result is as follows :

London:

From the generated word cloud, we can see that People in London have pretty outgoing life as Pubs, Café, Coffee Shops, Restaurants are very popular.

Toronto:

Looking at Toronto people's interests indicates that Restaurants, Coffee shops, Stores are more popular. Analyzing both the images we can say that most of the people have similar interests in both the cities. Mainly **Italian Restaurants** are popular in both the cities. Also, people drink lot of **coffee** and eat out are **restaurants** in both the cities.

3.3 We can find out the busiest or Popular neighborhoods to visit in both the cities.

We have the data of trending multiple venues for each neighborhoods. We can get the frequency of each neighborhoods with multiple venues that will give us the desired result of finding out the popular neighborhoods that are usually crowded.

For **London:**

We can observe here that most popular neighborhoods are **Temple** and **South Kensington** which have more than 100 venues. Since I have limited 100 venues for each neighborhood it shows maximum as 100.

For **Toronto:**

Busiest Neighborhoods of Toronto

We can observe here that most popular neighborhoods are **Adelaide, Toronto Dominion Centre**, **First Canadian Place**, **Harbourfront East**, **Garden District** and **St. James Town** which have more than 100 venues. Since I have limited 100 venues for each neighborhood it shows maximum as 100.

4. Predictive Modelling [K-means]

4.1 Performing K-means on our data.

Since we are dealing with spatial data, I have decided to use k-means to form the clusters that will give us more clarity about these two cities.
We will k-means to form clusters and plot those clusters on map for understanding the areas more accurately.

Since cities are covered over vast areas, I have decided to go for 5 clusters for each city.

This is my final data frame after clustering with cluster label for each Neighborhood with its 10 most common venue categories.
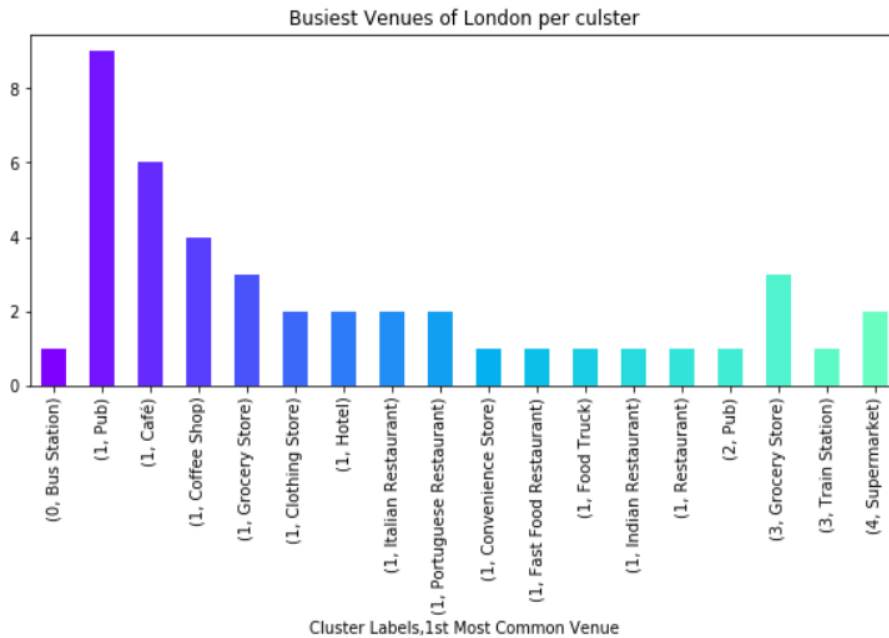
London:

	Postcode	Borough	Neighborhoods	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	
3	E10, E15	Waltham Forest	Leyton	51.55885	-0.00733	1	Fast Food Restaurant	Platform	Grocery Store	Fried Chicken Joint	Café	Clothing Store	Sandwich Place	Coffee Shop	
10	E13, E15	Newham	West Ham	51.52653	0.02876	0	Bus Station	Café	Gym	Yoga Studio	Fruit & Vegetable Store	French Restaurant	Food Truck	Food Stand	
27	E5, E8, E9, N1, N16	Hackney	Hackney	51.54505	-0.05532	1	Pub	Coffee Shop	Bakery	Cocktail Bar	Café	Grocery Store	Hotel	Brewery	V
29	E6, E13	Newham	Upton Park	51.53292	0.05461	1	Clothing Store	Bakery	Electronics Store	Park	Grocery Store	Coffee Shop	Gym Pool	Café	
32	E8, N16	Hackney	Shacklewell	51.54505	-0.05532	1	Pub	Coffee Shop	Bakery	Cocktail Bar	Café	Grocery Store	Hotel	Brewery	V

Toronto:

PostalCode	Borough	Neighborhood	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
M4E	East Toronto	The Beaches	43.676357	-79.293031	0	Health Food Store	Trail	Pub	Wings Joint	Dessert Shop	Event Space	Ethiopian Restaurant	Electronics Store
M4K	East Toronto	The Danforth West,Riverdale	43.679557	-79.352188	2	Greek Restaurant	Coffee Shop	Ice Cream Shop	Italian Restaurant	Restaurant	Bookstore	Furniture / Home Store	Fruit & Vegetable Store
M4L	East Toronto	The Beaches West,India Bazaar	43.668999	-79.315572	2	Park	Pizza Place	Pet Store	Movie Theater	Sushi Restaurant	Italian Restaurant	Fish & Chips Shop	Ice Cream Shop
M4M	East Toronto	Studio District	43.659526	-79.340923	2	Café	Coffee Shop	Bakery	Italian Restaurant	American Restaurant	Bookstore	Convenience Store	Latin American Restaurant
M4N	Central Toronto	Lawrence Park	43.728020	-79.388790	3	Park	Swim School	Bus Line	Wings Joint	Dessert Shop	Ethiopian Restaurant	Electronics Store	Eastern European Restaurant

4.2 Finding the appropriate names for clusters w.r.t. their venues

Here I have plotted bar graph of the frequencies each venue categories as per each cluster. This helps to decide that should be the name of cluster has we will clearly have the most poular category in the cluster.

Below are the results of the bar graph.

Busiest Venues of London per culster

Cluster Labels,1st Most Common Venue

London:

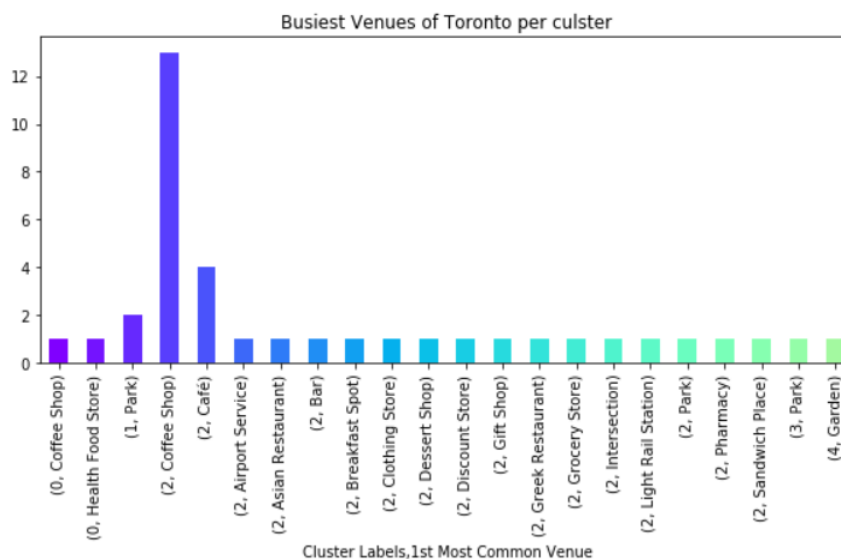Here we get the busiest venues in the clusters.so naming them accordingly:

Cluster1: Bus Stations.

Cluster2 : Pubs

Cluster 3: Pubs

Cluster 4: Grocery Stores

Cluster 5: Supermarkets

Busiest Venues of Toronto per culster

Cluster Labels,1st Most Common Venue

Toronto:

Here we get the busiest venues in the clusters.so naming them accordingly:

Cluster1: Health Food Stores

Cluster2 : Parks

Cluster 3: Coffee Shops

Cluster 4: Parks

Cluster 5: Gardens

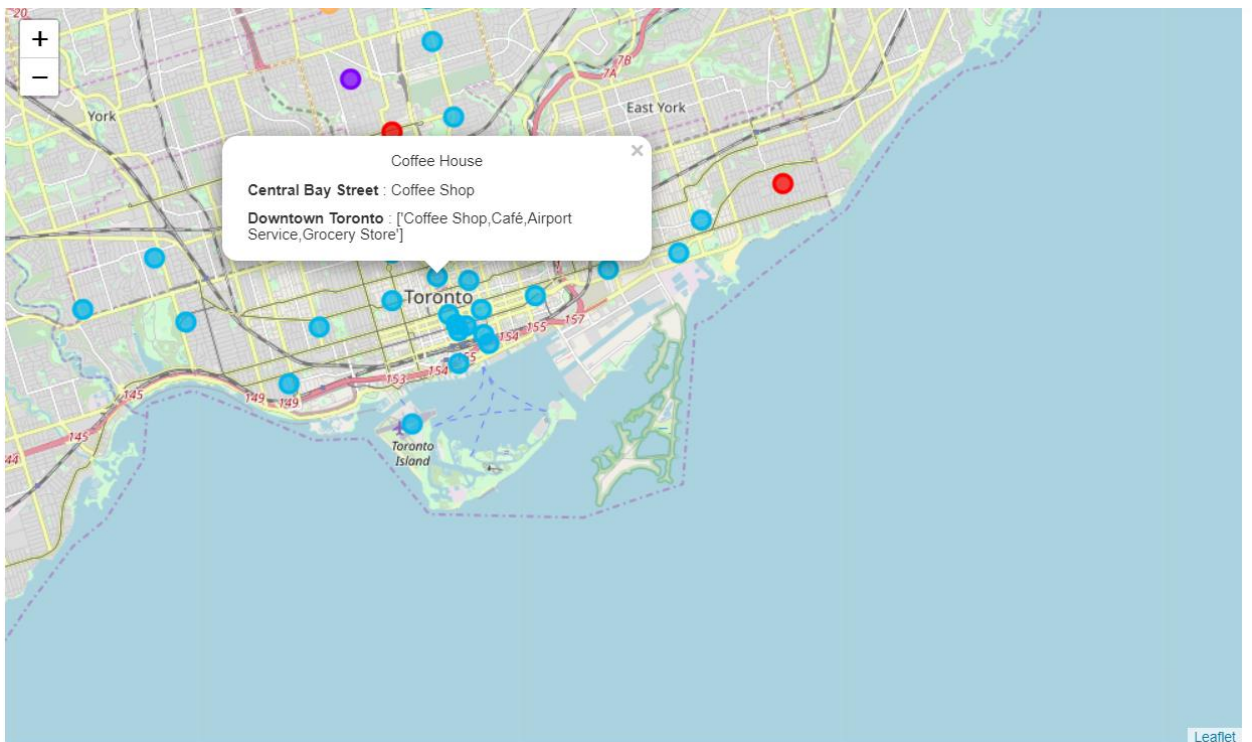## 4.3 plotting above obtained data on Map using Folium

You can see the plot of these clusters and venues on the pop-ups.

London:

Pubs

Holland Park : Italian Restaurant

Kensington and Chelsea : ['Hotel,Café,Italian Restaurant']

Toronto:

Coffee House

Central Bay Street : Coffee Shop

Downtown Toronto : ['Coffee Shop,Café,Airport Service,Grocery Store']
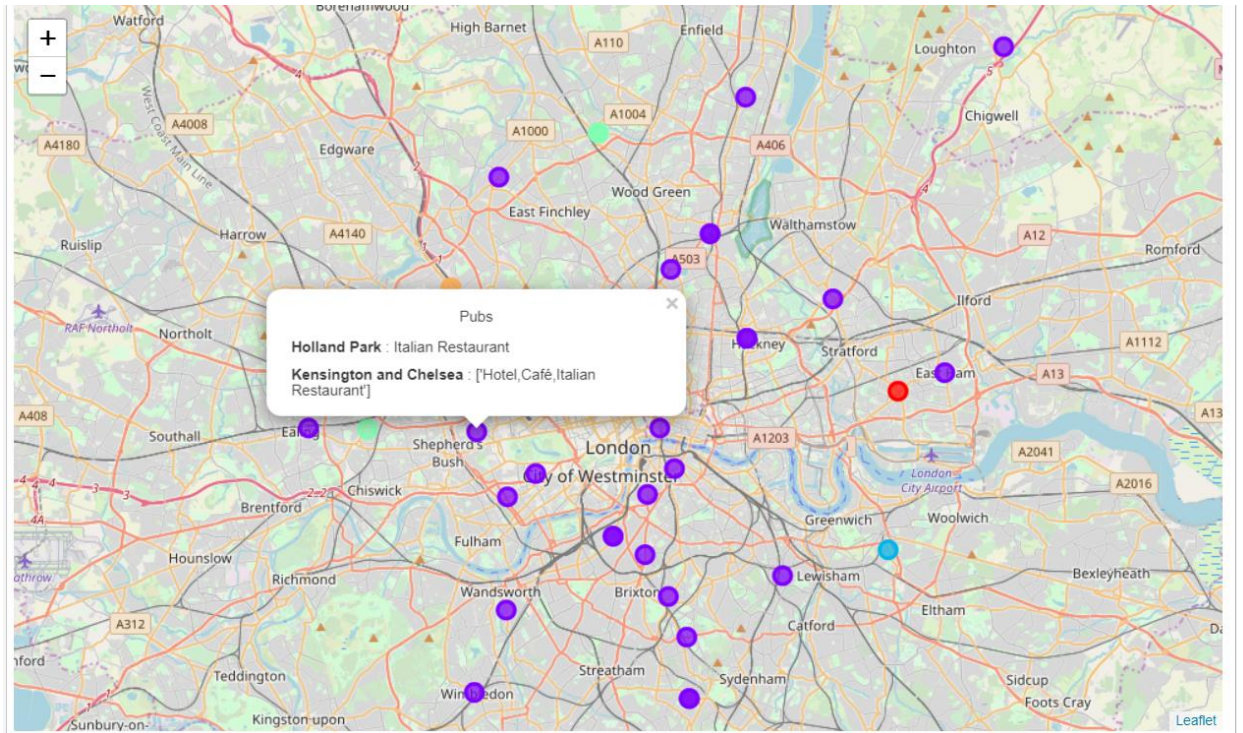
To make pop-up more interesting and map to be more useful, I have made popup that carries some meaningful information. Title is the name of Cluster that is determined by the bar graph which will be discussed in above section. After title the 1st Highlighted name is the name of Neighborhood with its most popular venue category. 2nd Highlighted Title is the name of Borough in which neighborhood relies with its most popular i.e. "1st most common venue" column from out data frame.

5. Discussions:

As, I have discussed both cities are very diverse and have very dense population. In the whole analysis, I have bought to you what are similarities and dissimilarities between these two cities we will discuss in detail below. Also, we have visualized neighborhoods on map where one can easily know which is the go-to venue in each neighborhood and some venues that are popular nearby in that borough.

Also, we can say that some of the neighborhoods that we analyzed using the bar graphs are busiest and crowded with multiple popular venues. We could have analyzed this venue further if I had access to foursquare premium calls, we could have analyzed the age breakdown of public with venues, number of checking's in venues, which hours of the day venues are crowded, etc.

I have used k-means clustering for the analysis of clusters and even it is basic more rigorous classifiers can be used in this scenario and could have got more insights. Tough I have used k=5 for clustering one of the clusters is very big in both the city i.e. down town area in both the cites.

I have ended by visualizing the clusters on the Maps of both cities with some useful information on popups. This data can be further used to analyze these venues accurately.

6. Conclusion:

Firstly, we have determined the busiest Neighborhoods in both the cities, also some venues popular in these neighborhoods.

Similarities between peoples interests in Toronto and London:
1. Many people eat out in restaurants and mainly Italian Restaurants are popular in both the cities.
2. People like to grab a coffee at cafes and coffee shops.
3. Both cities have Parks and Gardens and are popular in some neighborhoods that are away from Downtowns of cities.
4. Little outskirts of the city are well equipped with public transits like busses.

Dissimilarities between peoples interests in Toronto and London:
1. Most People go to Gyms in London while its not the case with Toronto.
2. Pubs are very popular in London while Bars are there in Toronto but not many people visit those.

In conclusion, I can say that people's interests are not divided by the diversity or different continents. If one is living in financial capitals like Toronto or London, he/she will tend to got out to

dine in restaurants, visit Pubs/Bars in weekends, grab coffee's frequently and to take some time out from busy life go to parks/gardens and gyms.

To the Future,

Akhil Patil

Link to full Notebook : **Battle of Neighborhoods between London and Toronto.**

G. references:

- [1] London Neighborhoods Wikipedia
- [2] Toronto Neighborhoods Wikipedia
- [3] Foursquare API