**Is Toronto a reasonable place to open a coffee shop?**

*Data Science Capstone - IBM Data Science Professional Certificate on Coursera*

**Introduction**

In order to determine if a city is a reasonable place to open a coffee shop, it makes sense to compare it to other cities as well as to look at the current location and distribution of current coffee businesses. In order to assess the initial possibility of locating a shop in Toronto, the number and density of coffee shops in the city were assessed in addition to making comparisons to other cities.

**Data section**

I will use the FourSquare API to collect data about locations of coffee stores in 5 major cities: Toronto, ON, New York, NY, San Francisco, CA, Boston, MA and Portland, OR. These are some of the most populated cities and each have a strong “coffee culture.”

**Methodology**

My main target here is to asses which city would have the highest coffee shop density. FourSquare API venues channel was used to query for venues in each of the cities. The CategoryID was set to show only coffee shops.

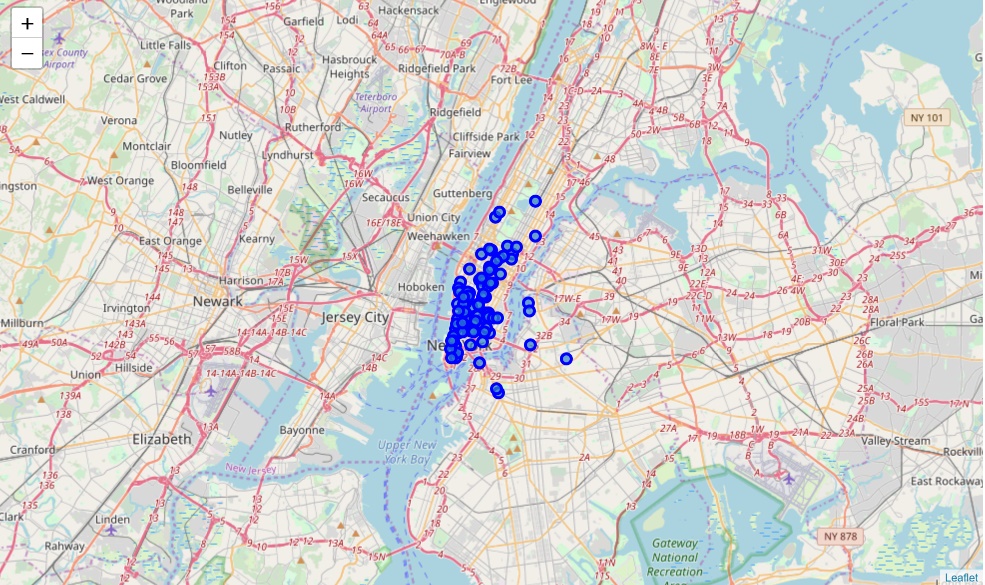
FourSquare limits requests to maximum of 100 venues per query. The request for venues was repeated for all studied cities. Name and coordinate data were saved from the results and plotted them on the map for visual inspection.

Next, to get an indicator of the density of coffee shops, a center coordinate of the venues was calculated to get the mean longitude and latitude values. Then the mean Euclidean distance from each venue to the mean lat/long was calculated. That result was the mean distance to the mean coordinate.

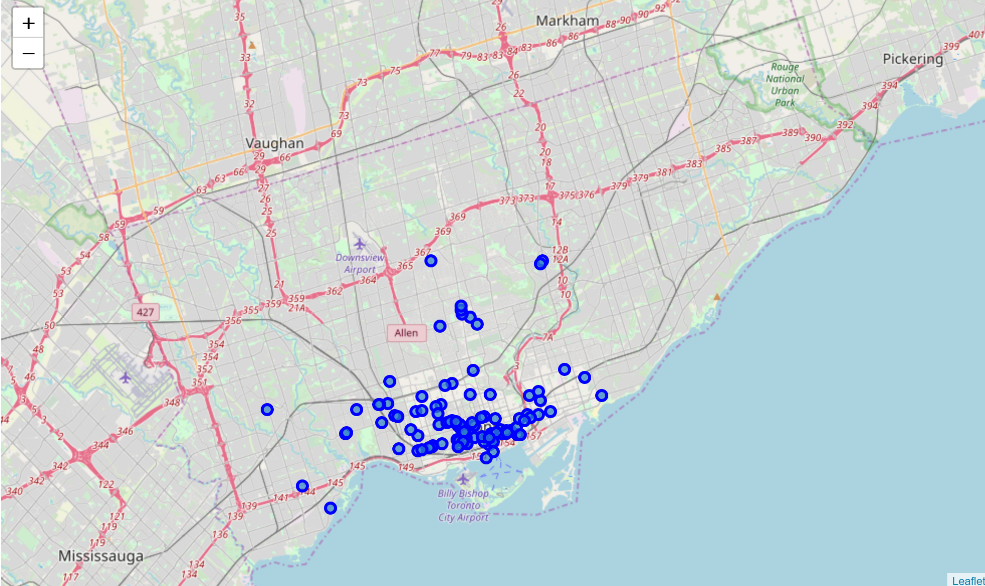
**Results**

For our initial visual inspection we see that they all have multiple coffee shops and often more than Foursquare would like to supply us. The following geoplots were generated with folium:

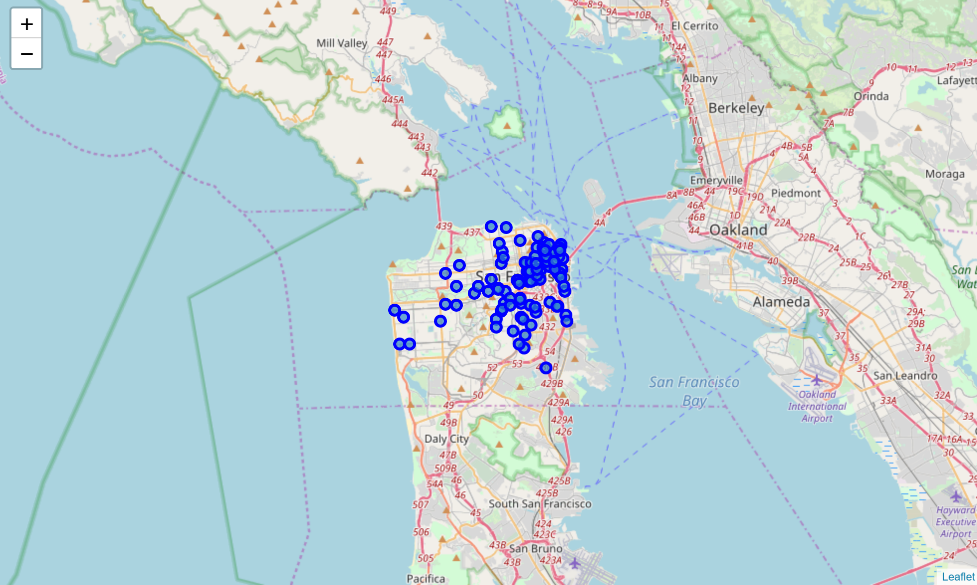
**New York:**

****

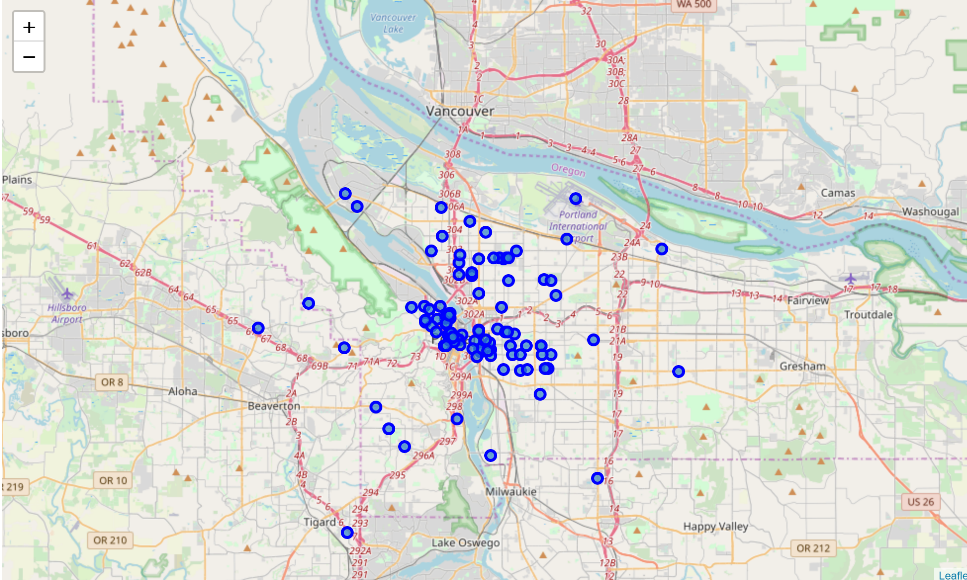
**Toronto:**

****

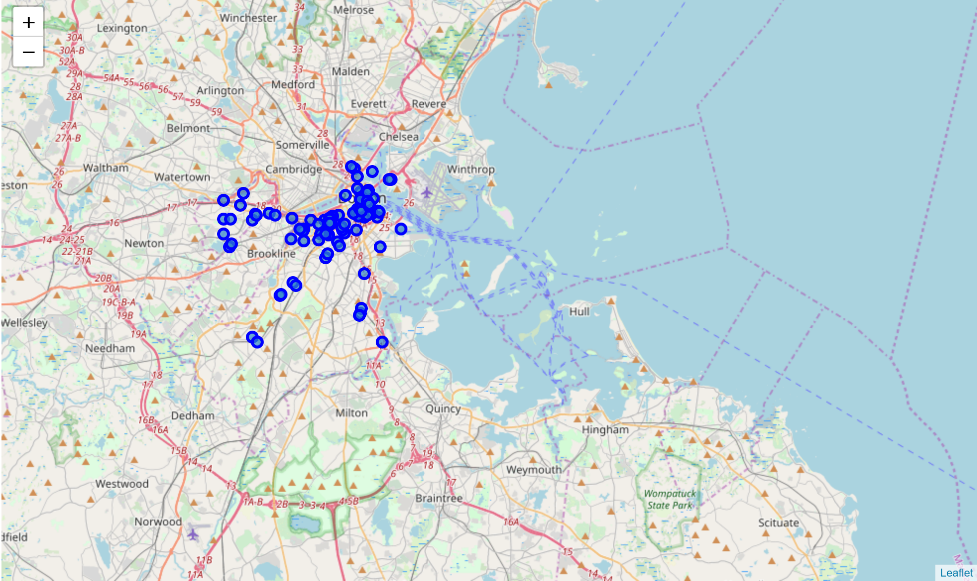
**San Francisco:**

****

**Portland:**

****

**Boston:**

****

Upon first inspection we see that New York and San Francisco are the most densely cities. Net, the mean coordinate and the mean distance to mean coordinate, i.e. density, were calculated.

New York, NY

Mean Distance from Mean coordinates

0.019933707590482175

Toronto, ON

Mean Distance from Mean coordinates

0.03580848539466422

San Francisco, CA

Mean Distance from Mean coordinates

0.026116104528479823

Portland, OR

Mean Distance from Mean coordinates

0.04096467122566089

Boston, MA

Mean Distance from Mean coordinates

0.02959977479772757

Therefore the results, from most dense to least, are :

1. New York
2. San Francisco
3. Boston
4. Toronto
5. Portland

**Discussion:**

Since some cities had what appeared to be outliers (shops far away from the core densities), it was determined that removal of these outliers may help get a more realistic picture of the true density in the hearts of the cities. For example, Portland seemed to have many outliers. When 10% of the furthest shops were removed from the calculations, Portland moved from #5 to #3. One consideration to do further work on is to move the location of the Foursquare API query until all the coffee shops in each city and do the calculations again.

**Conclusion:**

It appears that with more research, Toronto could be a reasonable location for a new coffee shop. With lower densities than similar cities, there is likely a place that could serve as a niche for a new shop.