Capstone Project - The Battle of Neighborhoods

For the last nine months, I have been very busy pursuing a certificate in Professional Data Science, offered by [**IBM**](https://www.ibm.com/cloud/watson-studio) through [**Coursera**](https://coursera.org/).

As of today, I have completed my final project, and am posting it here.

This is a quick overview of what I'm posting here:

1. The assignment
2. What I chose to focus on, and why
3. What I learned and how I learned it
4. Conclusions

*... And we're on our way!*

* The **assignment:**

*Now that you have been equipped with the skills and the tools to use location data to explore a geographical location, over the course of two weeks, you will have the opportunity to be as creative as you want and come up with an idea to leverage the Foursquare location data to explore or compare neighborhoods or cities of your choice or to come up with a problem that you can use the Foursquare location data to solve*

* **What I chose to focus on, and why**

I live in Chestnut Hill, in the northwest corner of the city of Philadelphia, in Pennsylvania. It is a lovely place to live and work - leafy green, walkable, and historic. I wondered whether Toronto, Ontario, CA might offers someplace similar - should I ever want to move my business (and myself) there.

* **What I learned and how I learned it**

Since this was a data science class, the first order of business, of course, was to use data science tools. Thus, my methodology focused most heavily on the use of such tools. Here is a summary of the steps, and what resulted from each.

**Conduct a review of the relevant literature**, using resources available online. Topics include:

* Toronto's and Philadelphia's history and current state (geographic, demographic, economic, etc
* Business trends

**Results**

Both cities are located at the nexus of several major waterways, and have grown partially by virtue of trade. Both were originally inhabited by Indigenous peoples, and both cities were formed along what used to be Native American trails.

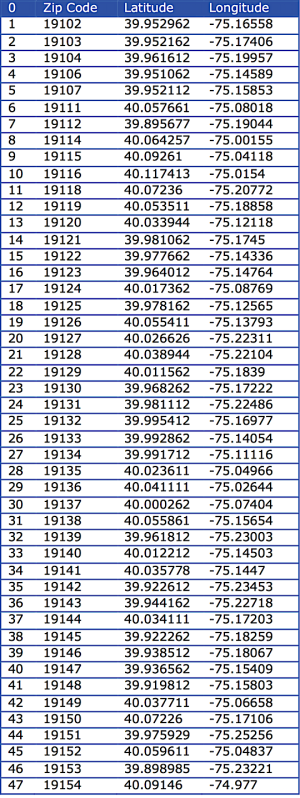
Their metro areas are almost the same size: Toronto's, as of 2016, was 6,417,516; Philadelphia's was, in 2017, 6,096,120. (Sources: US Census Bureau, Canadian Statistics Bureau).

Toronto and Philadelphia both have made names for themselves as leaders in technology innovation, although Toronto has done more in recent times and is beginning to be a technology/business hub of sufficient force to, someday soon, eclipse Silicon Valley.

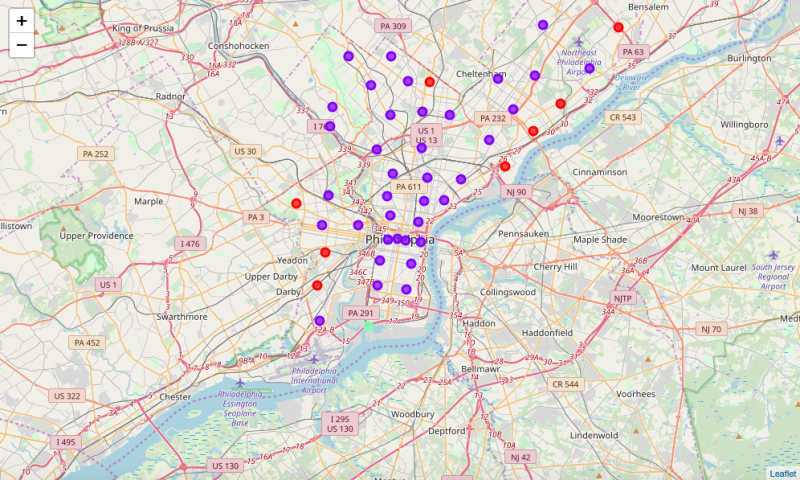
**Review Data specifications and availability**

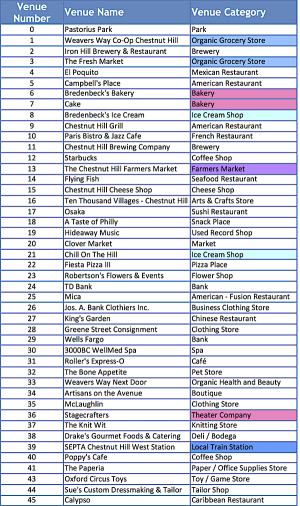
* + Locate Web sites offering Zip and or Postal Code information that can be readily scraped.
  + We will use python’s *beautifulsoup* library to extract postal code lists.
  + Then, we will get the geographical coordinates (latitude and longitude) so we can use them to query the Foursquare API database.**[[1]](https://contentfreeconsulting.typepad.com/blog/a-tale-of-two-neighborhoods-toronto-philadelphia.html" \l "_ftn1)** A *geocoder* will allow us to do so.
  + We will then be able to load this information into a pandas dataframe, then using *folium*, we will visualize each city’s neighbourhoods on the map.**[[2]](https://contentfreeconsulting.typepad.com/blog/a-tale-of-two-neighborhoods-toronto-philadelphia.html" \l "_ftn2)**
  + Load Foursquare data for all Zip Codes in Philadelphia and all Postal Codes in Toronto.
  + Using the Foursquare API, we will subsequently get the top 100 venues that are within a radius of 500 meters from the center point of each Zip or Postal Code. We do this by making API calls to Foursquare, passing the geographical coordinates until we are done via a Python loop. Foursquare then returns venue data to us in a JSON format, and we extract the venue name, category, latitude, and longitude. With these data, we will be able to check to see how many venues were returned for each neighbourhood and to tally up the number of (somewhat)[**[3]**](https://contentfreeconsulting.typepad.com/blog/a-tale-of-two-neighborhoods-toronto-philadelphia.html#_ftn1) unique categories can be curated from all the returned venues.

Results  
***As this table shows, after removing duplicates and P.O. boxes, we find that there are 47 Zip codes in Philadelphia, PA.***

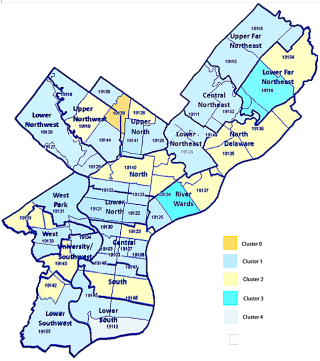
***[](https://contentfreeconsulting.typepad.com/.a/6a00d834681efa53ef0240a46fb982200c-popup)***

**Conduct k-means clustering**, using the mean frequency of occurrence of each venue category to create a centroid for each postal code. The k-means clustering algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as distinct as possible.

**Results**  
***This map depicts the five clusters identified by the analysis, on a map generated using Nominatim, and openstreetmap.org library.***  
  
**[](https://contentfreeconsulting.typepad.com/.a/6a00d834681efa53ef0240a498fea3200d-popup)**  
  
  
***When we look more closely at Chestnut Hill (Zip code 19118), this is the mix of venues we find.***

**[](https://contentfreeconsulting.typepad.com/.a/6a00d834681efa53ef0240a498fc73200d-popup)**

***... Which should not in any way imply that the clustering process I used should be determinative. See what these clusters look like on a map, below.***

***[](https://contentfreeconsulting.typepad.com/.a/6a00d834681efa53ef0240a498fc8b200d-popup)***

***Certainly, it would seem, there are neighborhoods in Toronto with a similar range of venues.***

***[](https://contentfreeconsulting.typepad.com/.a/6a00d834681efa53ef0240a498fd14200d-popup)***

Among the Toronto Postal Codes covered by our analysis of Foursquare venues, the Toronto neighbourhoods called Rosedale and Moore Park seem to have the qualities I would seek.

* [**Discussion**](https://contentfreeconsulting.typepad.com/blog/a-tale-of-two-neighborhoods-toronto-philadelphia.html#h_17671260391563558564692)

The primary purpose of this exercise was to determine whether we were able to use what we learned during the course of this Specialization, independently and without any Lab to provide explicit instructions. In that, this project was successful. I was able to run code that produced a coherent result. I also got to learn more about my own Philadelphia neighbourhood, since that was the topic upon which I chose to focus.

* [**Conclusions**](https://contentfreeconsulting.typepad.com/blog/a-tale-of-two-neighborhoods-toronto-philadelphia.html#h_17671260391563558564692)

Relocation analysis is serious business, and this data collection/analysis process is a good beginning. Going forward, I plan to use it as a jumping off point for looking at neighbourhoods, using units of measurement (e.g., the Census block group that are more stable and are linked with larger data sets like the Economic Census, as well as differences in governance, etc. - which are a function of differences between Canada and the US. One thing I know and like very much is Canada's approach to immigration (in other words, it is good for society, the economy, the wellbeing of all and should be encouraged).

Still, I did find a partial answer to my question of where I would want to live/work if I moved to Toronto. "Old Toronto" looks very attractive to me for so many reasons - not the least of which is the presence of the University nearby. Thriving educational institutions are essential to a good economy - especially if the type of work one does is cognitive in nature, as is the case for me. This is a pretty large area and includes a broad variety of neighbourhoods.

One of the analyses I conducted was to look at the mix of venues in my own neighbourhood, then sort the Toronto data to see which among the neighbourhoods covered in our class's work was most similar to my own. One thing I noticed: Chestnut Hill likes food, and parks! There are several ice cream parlours and bakeries a farmer's market, and two of the three grocery stores are organic. There are numerous parks, one of which is among the largest in the US, and there are two light commuter rail lines. Among the Toronto Postal Codes covered by our analysis of Foursquare venues, the Toronto neighbourhoods called Rosedale, Moore Park," seem to have the qualities I would seek. Yes, more research is required, but Data Science has given me more and better tools. This is just the beginning.