**Coursera Capstone Project**

**The Battle of Neighbourhoods**

**Week 1 and 2**

As part of the final project, I have chosen to explore suitable new store Locations for a Fashion Retail chain in the city of Toronto.

**Introduction and Business Objective**

Business objective of this project is to assist Fashion retailer to make data-driven decisions on the new locations that are most suitable for their new stores in Toronto. This will be a major part of their decision-making process, the other being on the ground qualitative analysis of districts once this data and report are reviewed and studied.

Toronto, the capital of the province of Ontario, is a major Canadian city along Lake Ontario’s north western shore. It's a dynamic metropolis with a core of soaring skyscrapers, all dwarfed by the iconic, free-standing CN Tower. Toronto also has many green spaces, from the orderly oval of Queen’s Park to 400-acre High Park and its trails, sports facilities and zoo.

Fashion retailer planning to explore locations in Toronto city’s four Boroughs - East Toronto, Central Toronto, Downtown Toronto, West Toronto.

Also, as study suggests locations with coffee shops, café and restaurants provides a good exposure to store. customers will choose to buy product when they see Fashion retailer sign. Therefore, the more people that drive or walk by store the more customers likely to attract.

Another point to be considered is to place store at a point which has businesses around that can eventually help drive traffic into new store. Any business or category of product that is closely related can help new store business.

So, in summary we will be looking for new store locations with below criteria:

* Location with coffee shops or café or restaurants.
* Location must have Fashion retail store / clothing store.
* Location must be from one of the four Boroughs - East Toronto, Central Toronto, Downtown Toronto, West Toronto.

Please note the analysis and recommendations for new store locations will focus on neighbourhoods with these establishments, not on specific store addresses. Narrowing down the best options derived from analysis allows for either further research to be conducted, advising agents of the chosen neighbourhoods, or on the ground searching for specific sites by the company's personnel.

**Data**

For this project we need following data –

* Toronto City data that contains list Boroughs, Neighbourhoods along with their Postcodes.

Data source - <https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&direction=next&oldid=942655364>

* GeoSpace data that’s contains postcodes along with their latitude and longitude.

Data Source - <http://cocl.us/Geospatial_data>

* Venues in each neighbourhood of Toronto city.

Data Source - Foursquare API

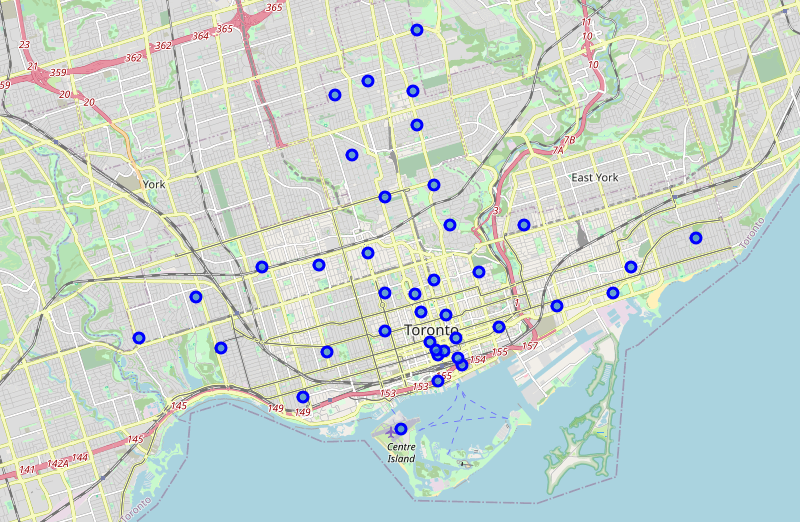
**Methodology**

* After importing, wrangling and cleaning the data, master data frame has all the required components Postal Code, Borough, Neighbourhood, Latitude, Longitude for four Toronto city Boroughs - East Toronto, Central Toronto, Downtown Toronto, West Toronto.

Below is snapshot of few records from same.



* I used python folium library to visualize geographic details of boroughs and its neighbourhood. I have created a map of Toronto by using latitude and longitude values to get the visual as below:



* I utilized the Foursquare API to explore the venues in each neighbourhood from master data frame. I designed the limit as 100 venue (Foursquare limitation) and the radius 500 meter for each neighbourhood from their given latitude and longitude information. Below are the few records from data frame having Venues name, category, latitude and longitude information from Foursquare API.



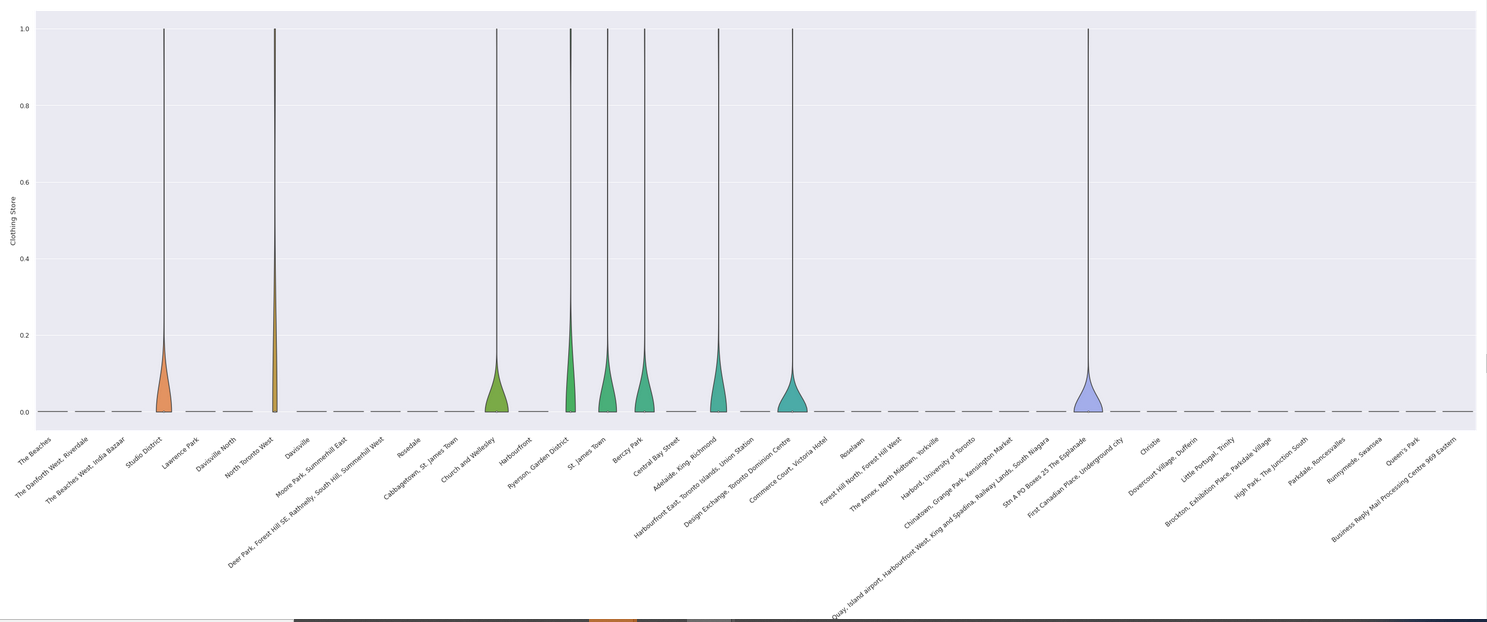
* In summary, 231 unique venue categories were returned by Foursquare.

Below are the few records from data frame which is having top 10 venues for each neighbourhood.

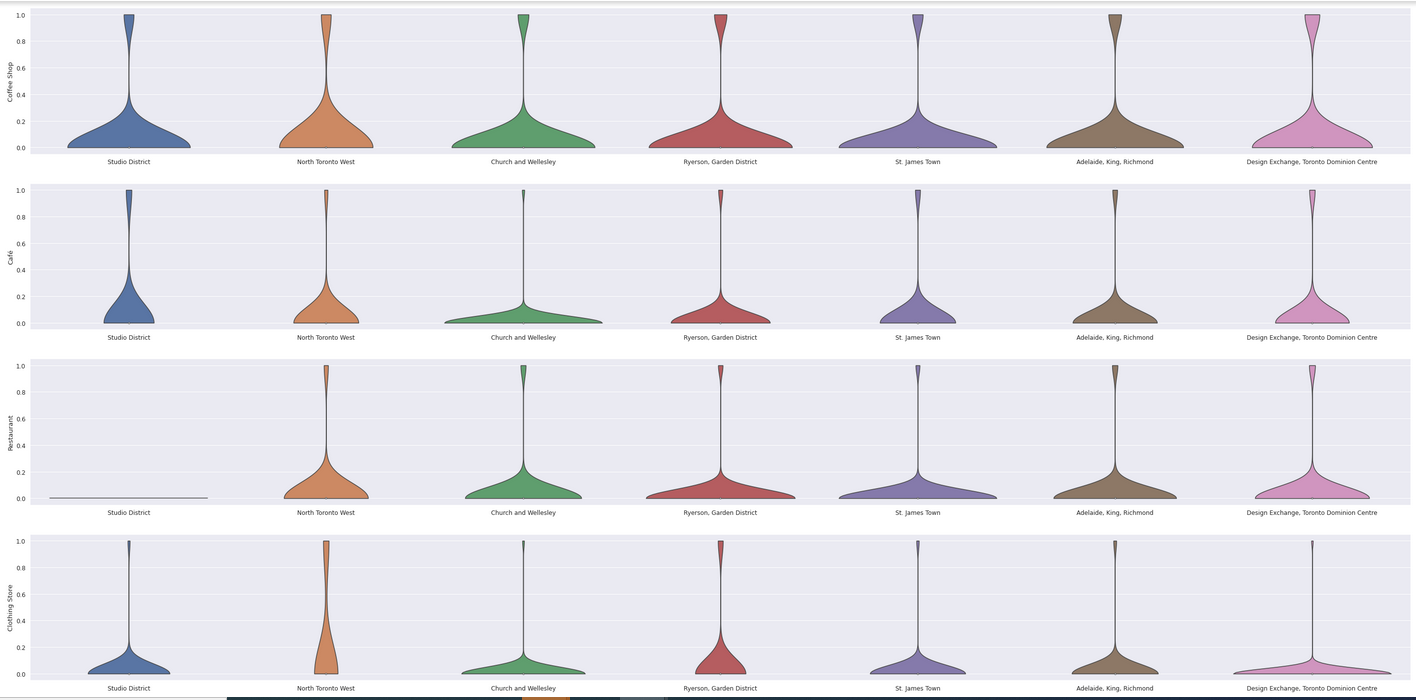


This is a very useful results table that can provide at a glance information for all the neighbourhood. Even once any conclusions are drawn further into the data workflow, we can refer to this table for meaningful insights about the top categories of businesses in all the neighbourhoods. Even without actual counts and numbers, it makes a great reference table.

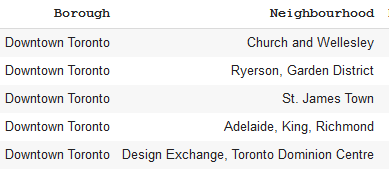
* I have visualized frequency of occurrence of clothing store in each neighbourhood as per client criteria. I've used a violin plot from the seaborn library - it is a great way to visualise frequency distribution datasets, they display a density estimation of the underlying distribution.



* From above I found that only below neighbourhoods are having clothing store so for further analysis we will focus on these neighbourhoods only.
* Studio District
* North Toronto West
* Church and Wellesley
* Ryerson, Garden District
* St. James Town
* Adelaide, King, Richmond
* Design Exchange, Toronto Dominion Centre
* As another client criteria for new store location was to have coffee shops, café and restaurants nearby. so, I have further visualized frequency of occurrence of coffee shops, café, restaurants and clothing store in above neighbourhoods.



* So, as we can see from the analysis below neighbourhoods suitable to open new store - according to the criteria that they have the specified venues (coffee shops, café, restaurants and clothing store) in a great frequency.

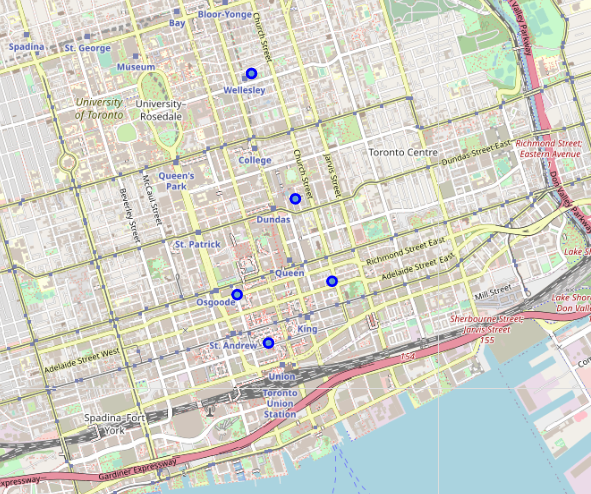


**Result**

Above 5 neighbourhoods from Borough - Downtown Toronto are potential location for Client's new store. With the help of data analysis and domain knowledge of retail & marketing, I was able to narrow it down to above.

These neighbours have met the criteria set by client as they are lively with Coffee shops, Restaurants, Cafés and Clothing Stores. this will provide a good exposure and eventually help drive traffic into new store.

I have visualized geographic details of shortlisted neighbourhoods. I have created a map of Toronto by using latitude and longitude values to get the visual as below:



**Discussion**

It is obvious these neighbourhoods are all very centrally located, locations fitting the criteria for popular venues would normally be in central locations in many cities of the world.

From this visualisation it is clear that on a practical level, with no data to base decisions on, researching and then visiting all the neighbourhoods would be a daunting and time-consuming task. We have narrowed the search area down significantly from 39 neighbourhoods to 5 that should suit the client's retail business.

I have made inferences from the data in making the location recommendations, but that is exactly the point. There is no right or wrong answer or conclusion for the task at hand. The job of data analysis here is to steer a course for the location selection of new stores (i) to meet the criteria of being in neighbourhoods that are lively and (ii) to narrow the search down to just a few of the main areas that are best suited to match the criteria.

**Conclusion**

There are many ways this analysis could have been performed based on different methodology and perhaps different data sources. I chose the method I selected as it was a straightforward way to narrow down the options, not complicating what is simple in many ways – meeting the criteria for the surrounding venues, and in my case, domain knowledge I have on the subject. I originally intended to use the clustering algorithms to cluster the data, but as it progressed it became obvious that this only complicated the task at hand. The analysis and results are not an end point, but rather a starting point that will guide the next part of the process to find specific store locations. The next part will involve domain knowledge of the industry, and perhaps, of the city itself. But the data analysis and resulting recommendations have greatly narrowed down the best options based on data and what we can infer from it.

Without leveraging data to make focussed decisions, the process could have been drawn out and resulted in new stores opening in sub-standard areas for this retailer. Data has helped to provide a better strategy and way forward, these data-driven decisions will lead to a better solution in the end.

**Limitation**

The accuracy of data depends purely depends on the data provided by FourSquare.