

Toronto Subway Analysis

by Snehil Vishwakarma

1. Introduction: Business Problem

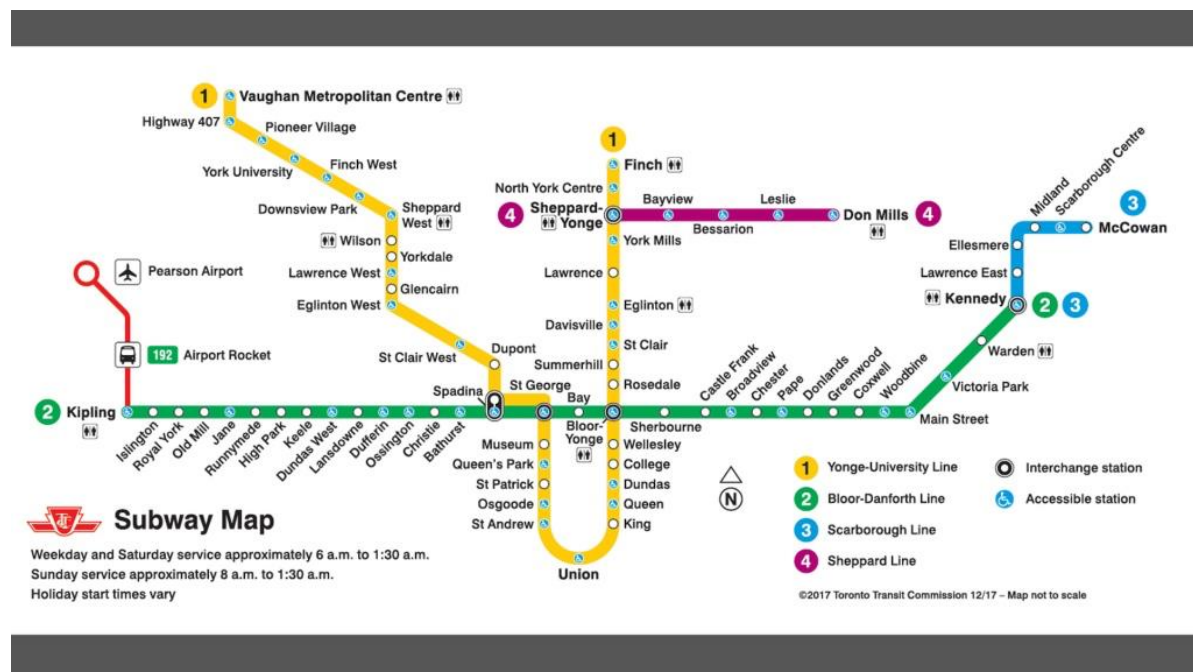
In this project, we have analyzed the **Subway Stations of Toronto**, Canada on the basis of **Venues** around them. Toronto is Canada's largest city and a world leader in such areas as business, finance, technology, entertainment and culture. With a recorded population of 2,731,571 in 2016, it is the most populous city in Canada and the fourth most populous city in North America. The Toronto subway is a rapid transit system serving Toronto and the neighboring city of Vaughan in Ontario, Canada, operated by the Toronto Transit Commission (TTC)

This report will be helpful to people who are new to Toronto - Immigrants, people moved from different cities.

The Toronto subway has 75 stations. As there are a lot of venues around any subway station and some stations are at short distance from each other, we have focused on top 100 venues within a radius of 300 meters.

Any resident in Toronto uses subway frequently and people who new to the city don't know what kind of places they can't expect or find around a subway station. Moreover, such information is not readily available.

When we consider all these, we used data science tools and machine learning algorithms to create a map and chart where all subway station is clustered according to the venues around them.



2. Data

In this report we have used following datasets

- list of all subway stations in Toronto – we have got the list from **Wikipedia**
- Geographic coordinates of all subway stations - using **Nominatim API**
- Popular venues arounds stations - using **FourSquare API**

2.1 Snapshot of Data after preprocessing

Subway Station	Station Latitude	Station Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Finch, Toronto	43.781297	-79.415899	Pyung Won House	43.779386	-79.415511	Korean Restaurant
Finch, Toronto	43.781297	-79.415899	Burrito Place	43.781258	-79.415801	Burrito Place
Finch, Toronto	43.781297	-79.415899	Jungsoo Nae 정수네	43.783590	-79.416480	Korean Restaurant
Finch, Toronto	43.781297	-79.415899	Toronto a la Cart - Korean	43.780563	-79.415801	Food Stand
Finch, Toronto	43.781297	-79.415899	Huh Ga Ne 허가네	43.779362	-79.417366	Korean Restaurant

3. Methodology

The main goal of this project is finding a mechanism to group the subway stations together. One way of doing that is finding venues around them and see whether which subway stations have similar venues around them.

Firstly, we have prepared **dataset having list of subway stations with geographic coordinates**. Secondly, we have used **FourSquare API** to get list of top 100 venues around any subway stations in 300 meters radius.

Finally, we have analyzed a combined dataset having subway stations coordinates and venues around them. We have checked number of venues around subway stations in 300 meters radius using different graphs. We have also checked what are the most common venues. Once we have done that, we have used **k-means clustering** to group the subway stations.

About K-means algorithm: It identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

This will help people to know what kind of venues they can expect any subway station and which subway stations share common features in terms of venue.

4. Analysis

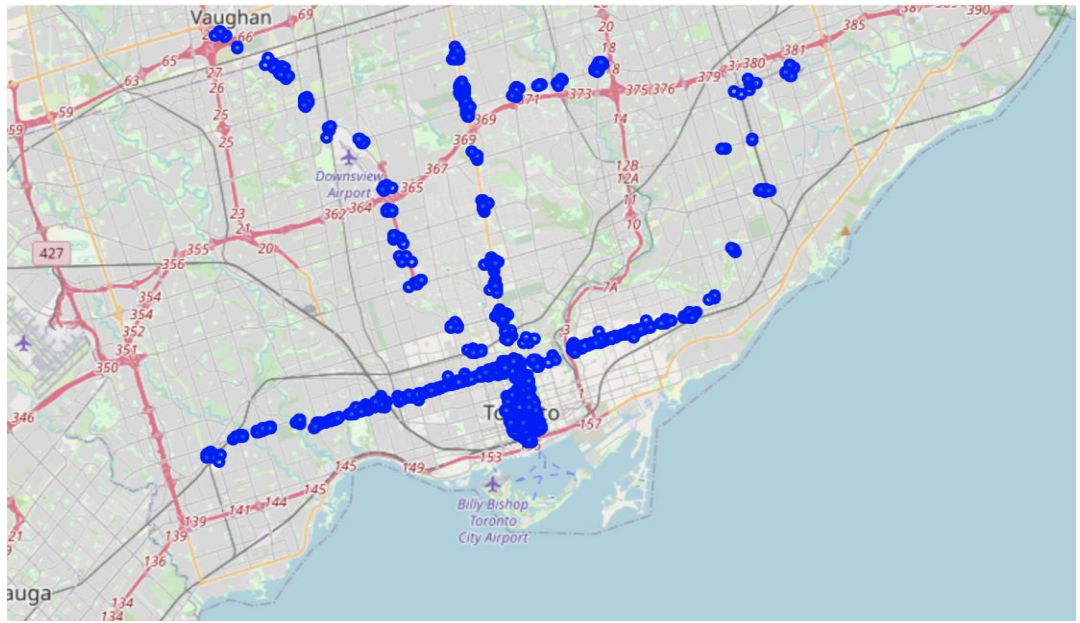
Once we have prepared the dataset in required format, we did some analysis on venues around subway stations

4.1 Concentration of Venues

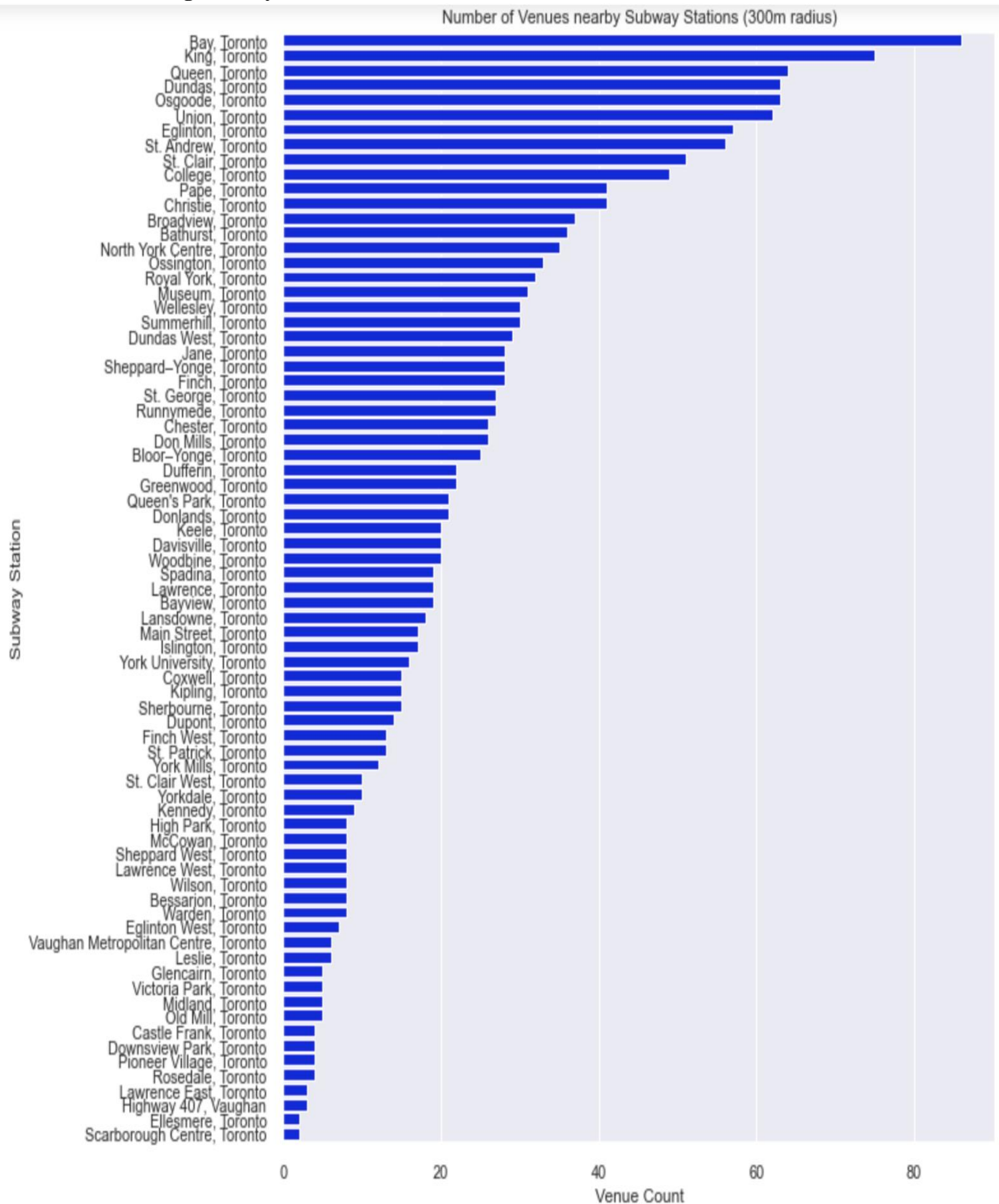
Folium map showing subway stations



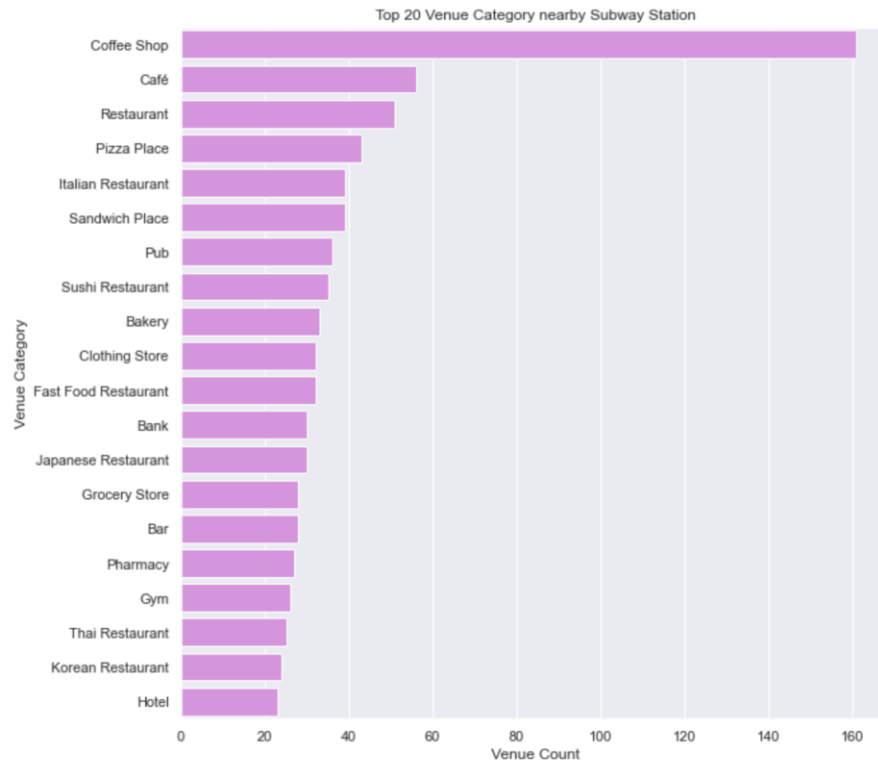
Folium map showing concentration of venues.



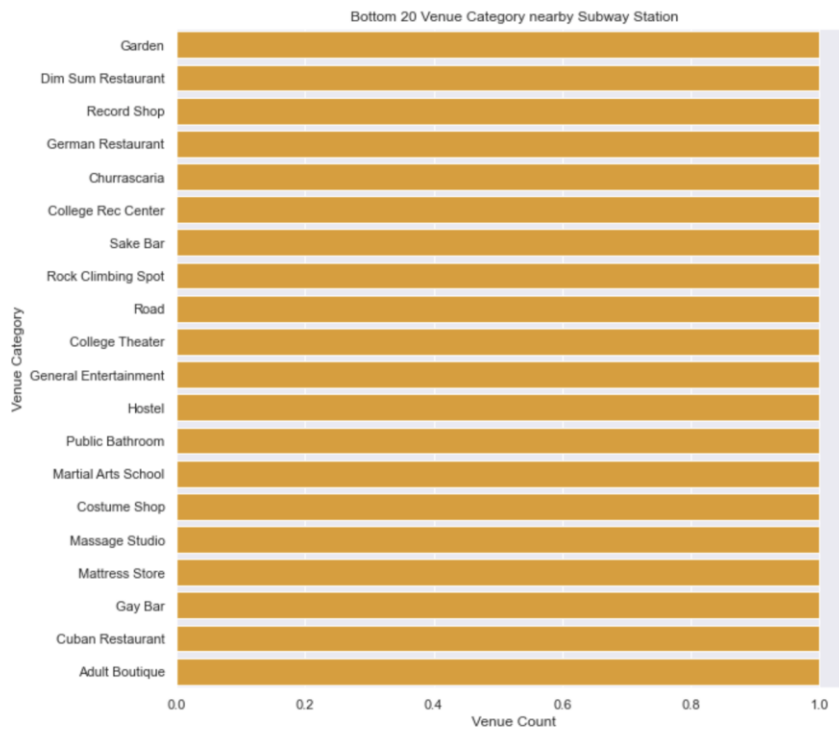
4.2 Bar chart showing Sunways vs Number of Venues in 300 meters



4.3 Top 20 venues around Subway Stations



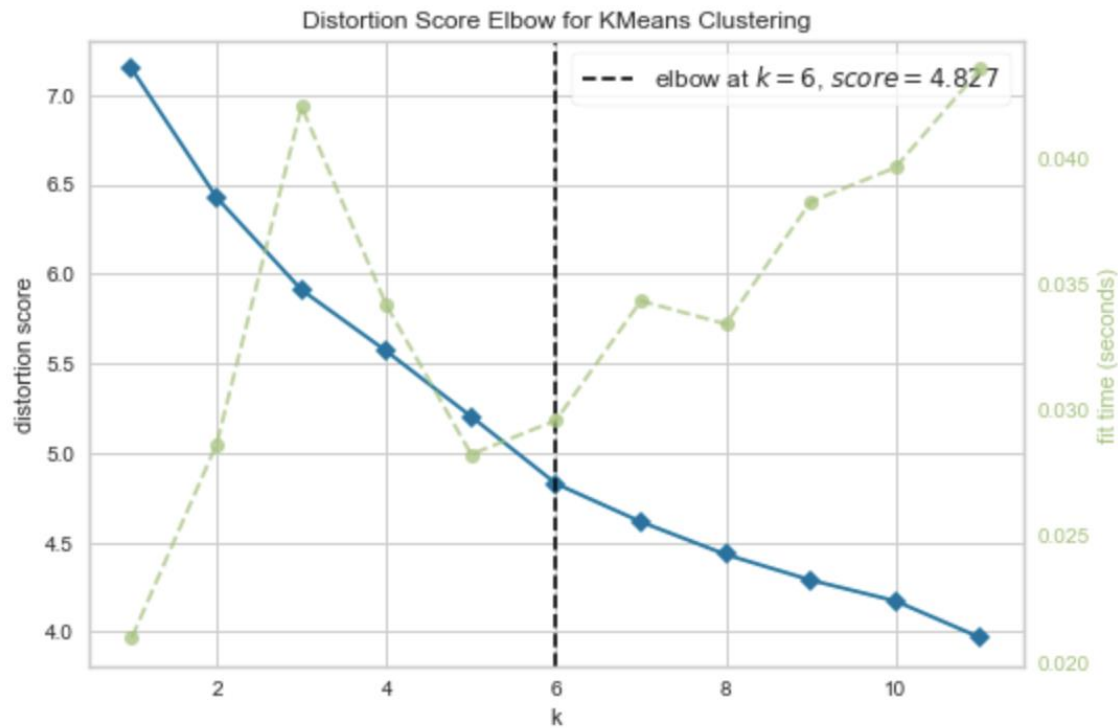
4.4 Bottom 20 venues around Subway Stations



4.5 Snapshot of dataset showing Subway Stations and common venues around them

	Subway Station	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	Bathurst, Toronto	Korean Restaurant	Ice Cream Shop	Eastern European Restaurant	Mexican Restaurant	Health Food Store	Restaurant	Indie Movie Theater	Lingerie Store	School	Sandwich Place
1	Bay, Toronto	Italian Restaurant	Boutique	Coffee Shop	Café	Spa	French Restaurant	Clothing Store	Restaurant	Hotel	Pizza Place
2	Bayview, Toronto	Café	Coffee Shop	Indian Restaurant	Toy / Game Store	Supermarket	Restaurant	Spa	Burger Joint	Italian Restaurant	Thai Restaurant
3	Bessarion, Toronto	Pizza Place	Outdoor Supply Store	Gas Station	Sandwich Place	Fish Market	Sporting Goods Shop	Breakfast Spot	Persian Restaurant	Flea Market	Electronics Store
4	Bloor-Yonge, Toronto	Café	Japanese Restaurant	Yoga Studio	Mediterranean Restaurant	Salad Place	Restaurant	Caribbean Restaurant	Coffee Shop	Comic Shop	Music Venue

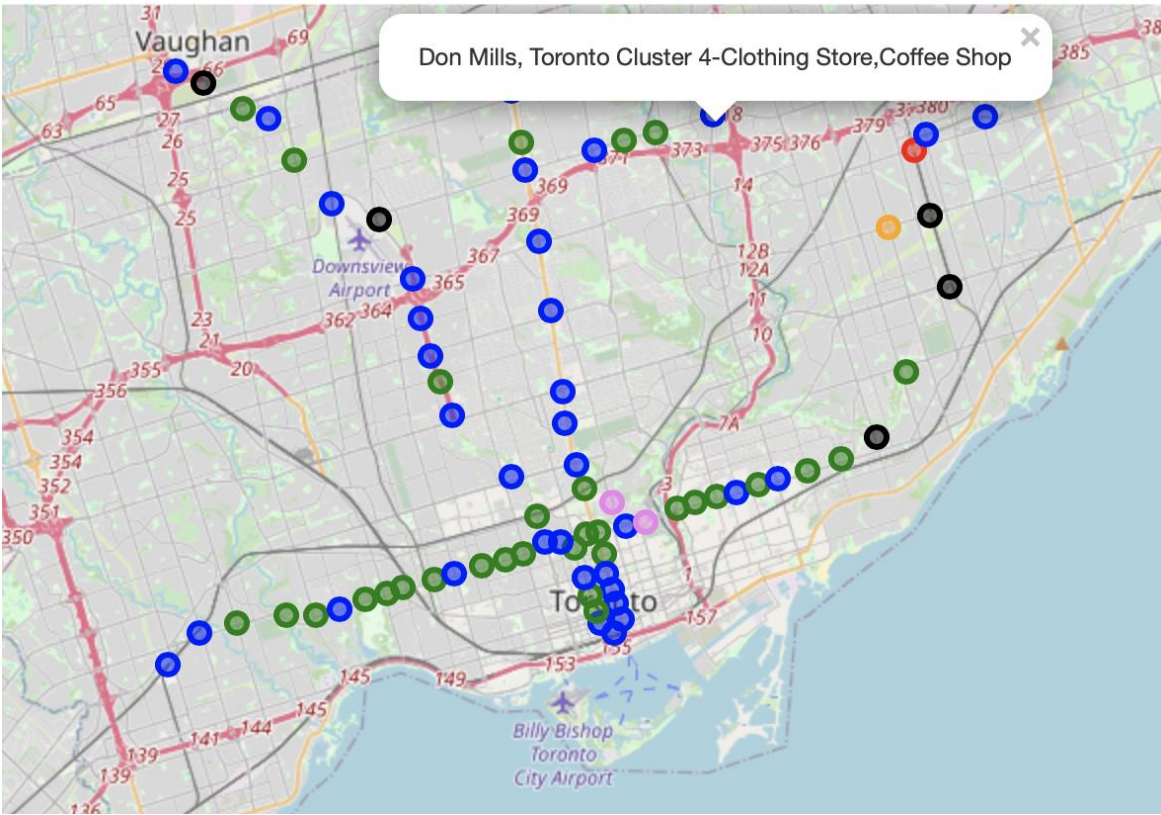
4.6 Distortion score Elbow for K-means Clustering. We can see that elbow is at 6, which means we can group subway stations in **6 clusters**.



4.7 Snapshot of data after we have performed K-Means Clustering. Subway stations have been labelled as Cluster 0 to 6.

Subway Station	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
Finch, Toronto	43.781297	-79.415899	4	Korean Restaurant	Coffee Shop	Bus Station	Sandwich Place	Café	Restaurant	Japanese Restaurant	Thai Restaurant	Bank	Sake Bar
North York Centre, Toronto	43.768679	-79.412630	0	Coffee Shop	Shopping Mall	Restaurant	Café	Sushi Restaurant	Ramen Restaurant	Japanese Restaurant	Juice Bar	Bubble Tea Shop	Sandwich Place
Sheppard-Yonge, Toronto	43.761452	-79.410915	4	Coffee Shop	Japanese Restaurant	Burger Joint	Restaurant	Grocery Store	Bar	Fast Food Restaurant	Fried Chicken Joint	Breakfast Spot	Ramen Restaurant
York Mills, Toronto	43.744039	-79.406657	4	Restaurant	Coffee Shop	Thai Restaurant	Gym	Burrito Place	Sandwich Place	Intersection	Food Court	Optical Shop	Park
Lawrence, Toronto	43.726348	-79.402474	4	Coffee Shop	Seafood Restaurant	Sushi Restaurant	Thai Restaurant	Bank	Pool	Gastropub	Ice Cream Shop	Supermarket	Pizza Place

4.8 Folium map showing subway stations in clusters. Subway stations have been grouped in 6 clusters. Cluster 0 (Green) and Cluster 4 (Blue) have the most subway stations.



5. Results and Discussion

From the analysis, we can see that subway stations in downtown area have a greater number of venues around them. **Bay, King and Queen** subway stations have highest number of venues while **Scarborough center and Ellesmere** have the least.

On further analysis, we found that **Coffee shops and Cafe** are the most common venue around subway while **Garden and Hostel** are some of the least common spots. After doing the cluster analysis of venues around subway stations, we gathered that the elbow is for **k=6** which means we can group the subway stations in **6 Clusters** namely - Cluster 0, Cluster 1, Cluster 2, Cluster 3, Cluster 4 and Cluster 5. Most of the subway stations are grouped into either **Cluster 0 and Cluster 4**.

On checking the most common venues in Clusters, we found that the following

Cluster 0 (Green) - Pizza place, Sandwich, Coffee shops, Cafe and Bar place are most common venues.

Cluster 4 (Blue) - Coffee shops and Clothing Store are most common venues.

Cluster 1 (Violet) - Park is the most common venue. Rosedale and Castle Frank subways are there.

Cluster 2 (Black) - It has 5 subway stations. Bus station is the most common venue. Baseball field and Hockey arena are also there.

Cluster 3 (Red) - Gym / Fitness Centre is the most common venue. Only Ellesmere subway is there.

Cluster 5 (Orange) - Hookah Bar is the most common venue. Only Scarborough Centre is there.

6. Conclusion

Subway is the most common mode of transport in Toronto. From our analysis, we have come to conclusion that subway stations can be grouped into 6 categories based on venues around them. People who are looking for **Pizza place and Sandwich** venues can visit explore venues around Subway in Cluster 0. People who are looking for **Clothing stores** can visit explore venues around Subway in Cluster 4. Anyone looking for **Coffee shops** can find them around subways in Cluster 0 and Cluster 4. There are only **9 Subway stations out of 75**, which do not fall under Cluster 0 or Cluster 4. Cluster 1, Rosedale and Castle Frank subways, have **Park** as the most common venue around them while Cluster 3, Ellesmere Subway, has **Gym/Fitness Centre** as the most common venue. Cluster 2 has 5 subway stations; **Bus station, Baseball field and Hockey arena** are common venues. **Hookah Bar** is the most common venue around Scarborough Centre, which is in Cluster 5. These clusters provide insights into what want kind of venues you can expect around any subway stations. Anyone new to Toronto can use these insights to explore the venues easily without going too far from any subway station.

7. References

- a. List of Toronto subway stations

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

- b. Four Square API

<https://developer.foursquare.com>