

Capstone Project - Toronto Subway Analysis

Applied Data Science Capstone by IBM/Coursera

By – Snehil Vishwakarma

Outline

- Introduction: Business Problem
- Data
- Methodology
- Analysis
- Results and Discussion
- Conclusion

Introduction: Business Problem

- People who new to Toronto don't know what kind of places are around a subway station
- Such information is not readily available
- Analyze Toronto Subway Stations on basis of Venues around them

Data

- List of Toronto Subway Stations
- Subway Stations Geographical Coordinates
- List of venues around any subway station in 300 meter radius

Data

Snapshot of data after preprocessing

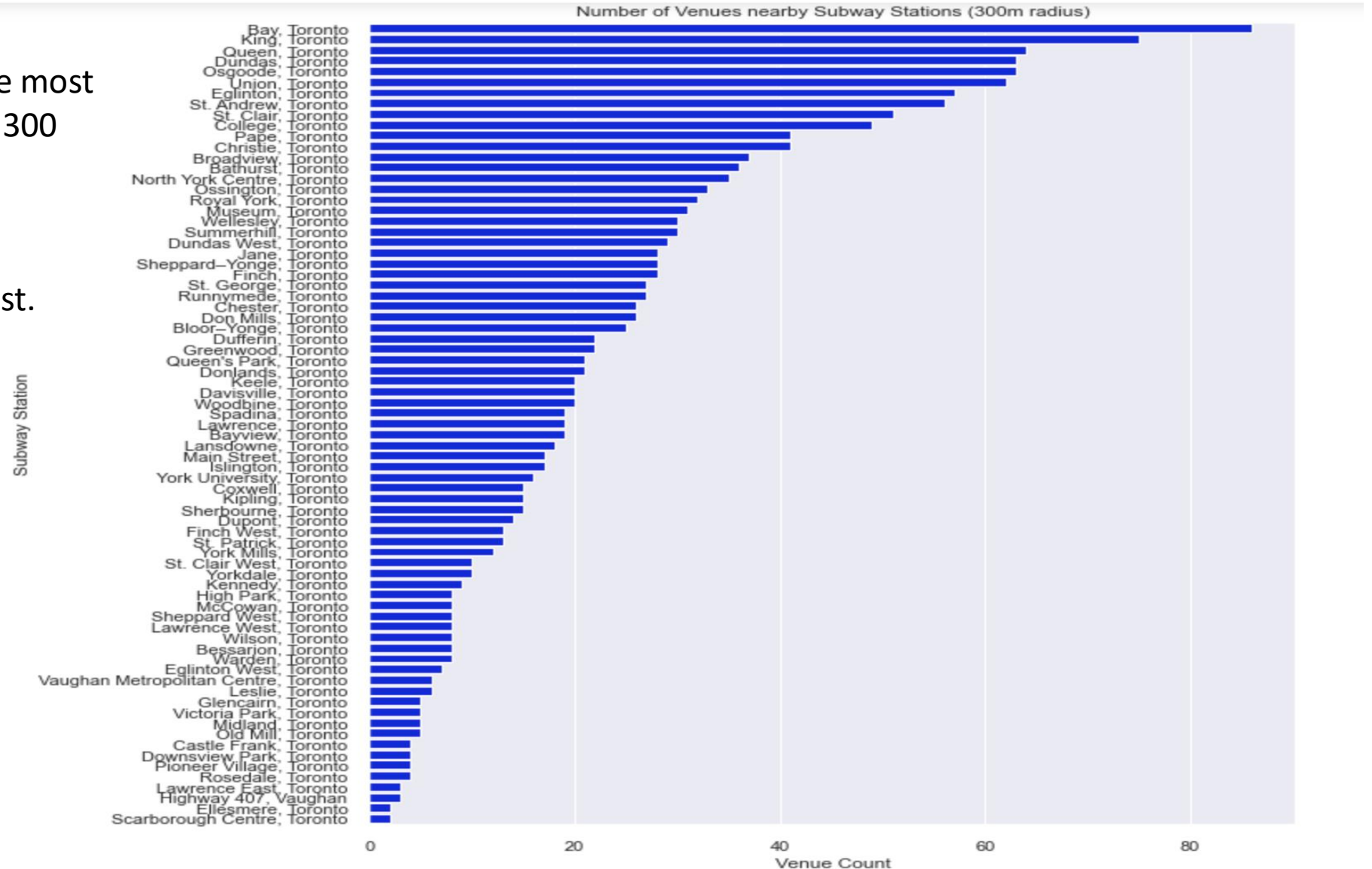
Subway Station	Station Latitude	Station Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Finch, Toronto	43.781297	-79.415899	Pyung Won House	43.779386	-79.415511	Korean Restaurant
Finch, Toronto	43.781297	-79.415899	Burrito Place	43.781258	-79.415801	Burrito Place
Finch, Toronto	43.781297	-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

Methodology

- Prepared subway stations dataset with geographic coordinates.
- Used FourSquare API to get top 100 venues around any subway stations in 300 meters radius
- Have utilized K-means clustering algorithms to find Elbow for number of clusters
- **K-means algorithm** identifies k number of centroids, and then allocates every data point to nearest cluster, while keeping centroids as small as possible.
- Once we know number of clusters, again utilized K-means clustering to group subway stations
- Plot subway clusters using Folium maps

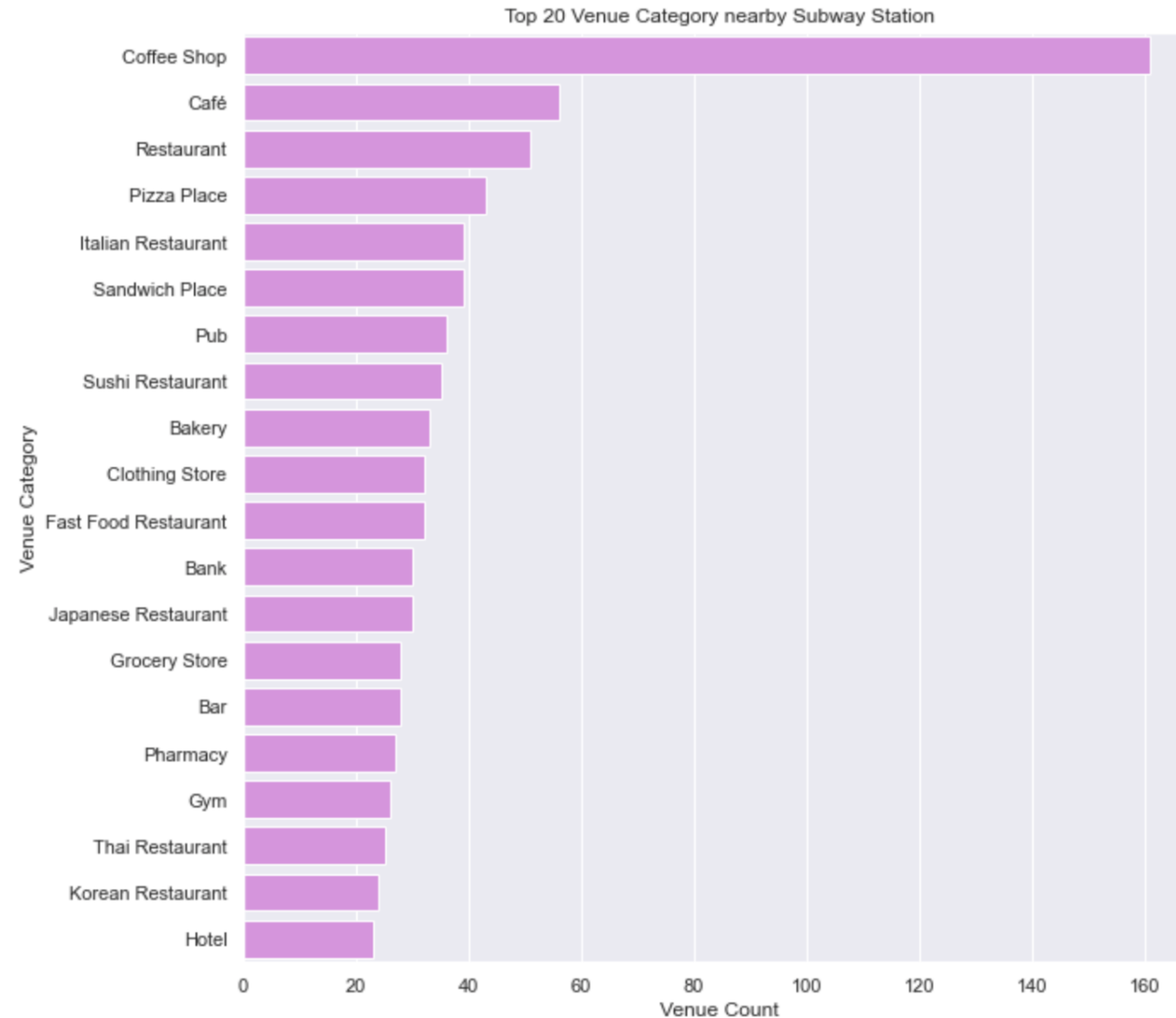
Analysis

1. **Bay Station** has the most number of venues in 300 meters radius while **Ellesmere & Scarborough Centre Stations** have the least.



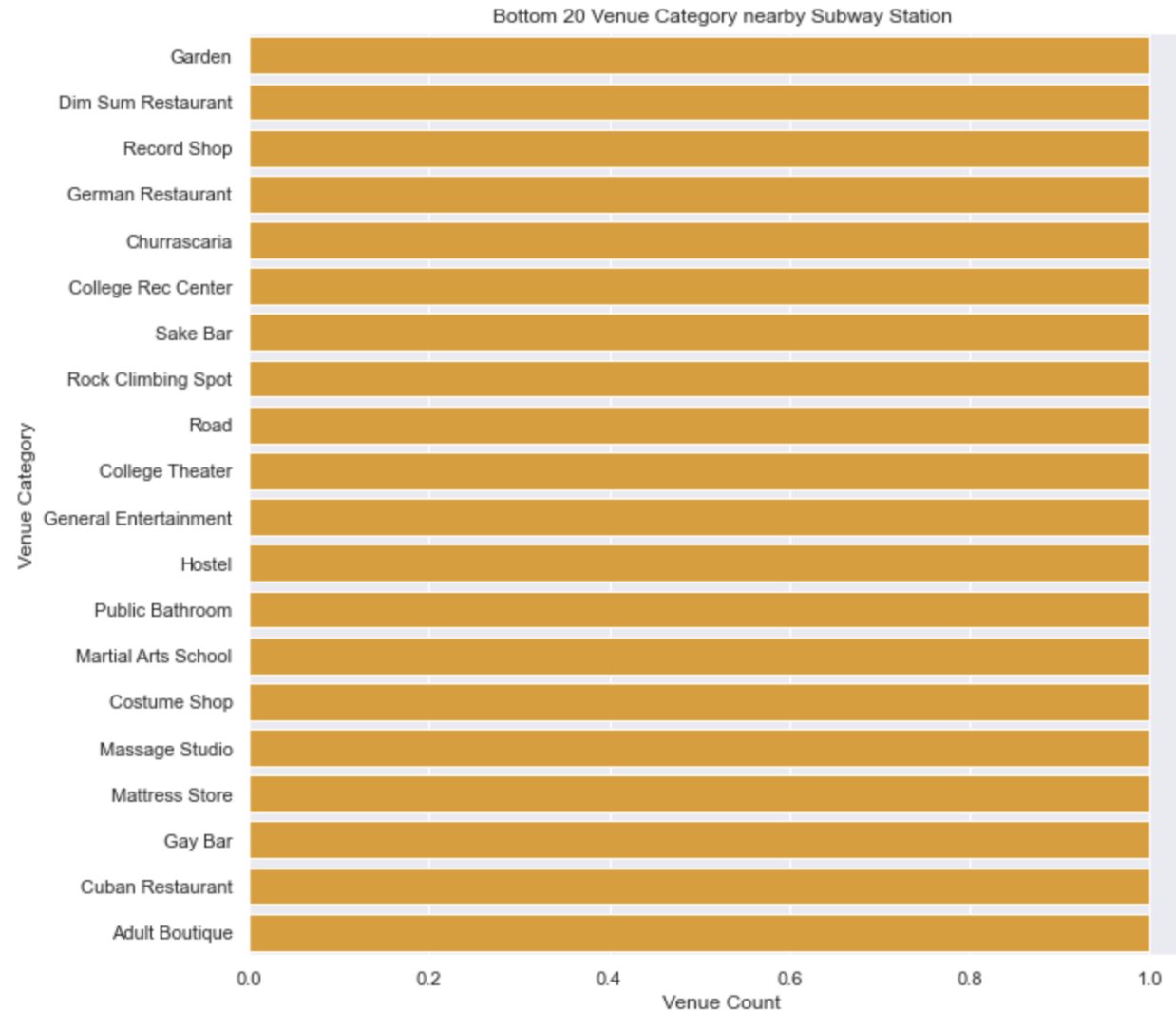
Analysis

2. Coffee Shop & Cafe are the most common venue across subway stations.



Analysis

3. Garden, Hostel, German restaurant are some of the least common venues.



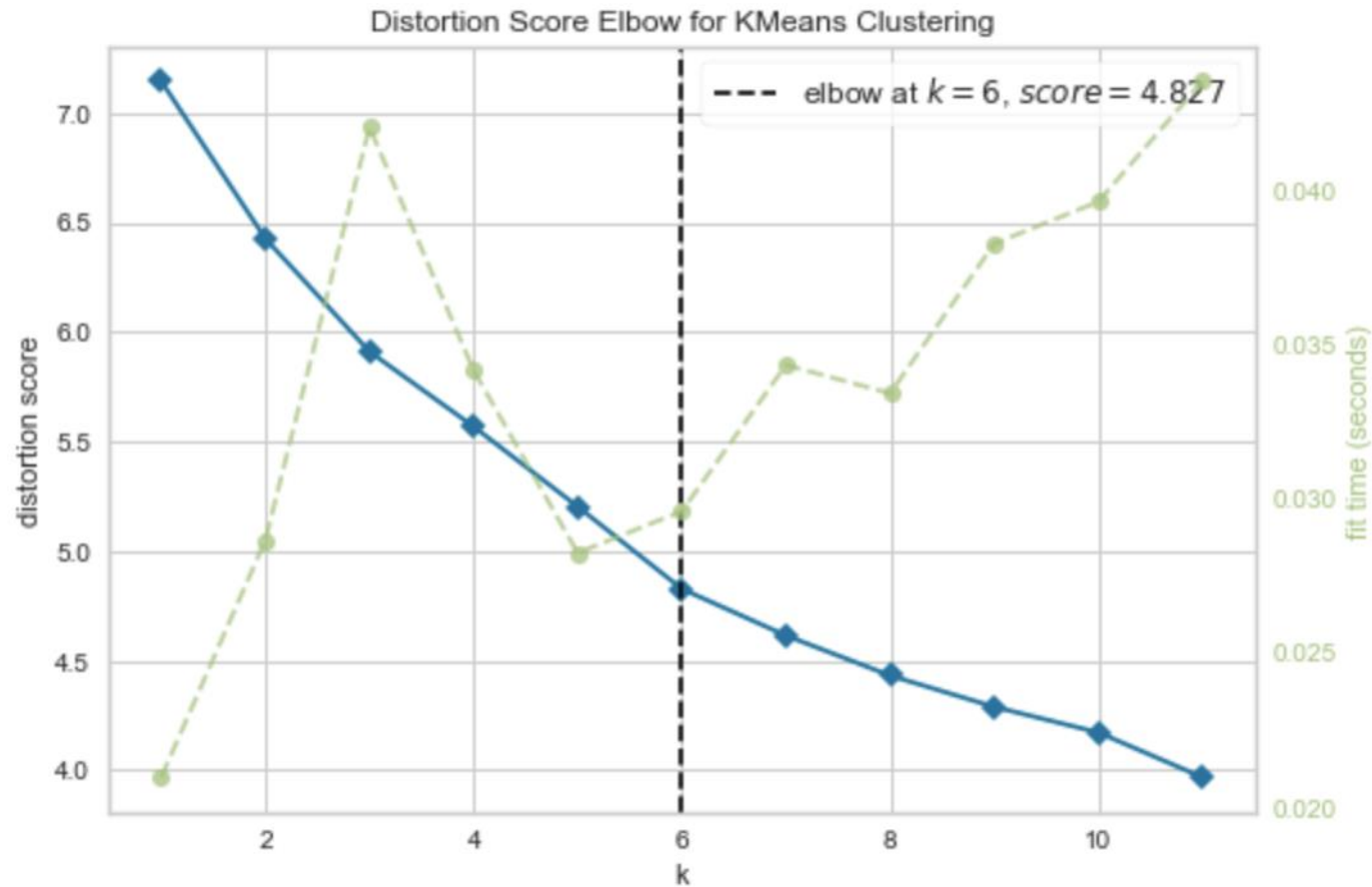
Analysis

4. Subway stations with 10 most 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

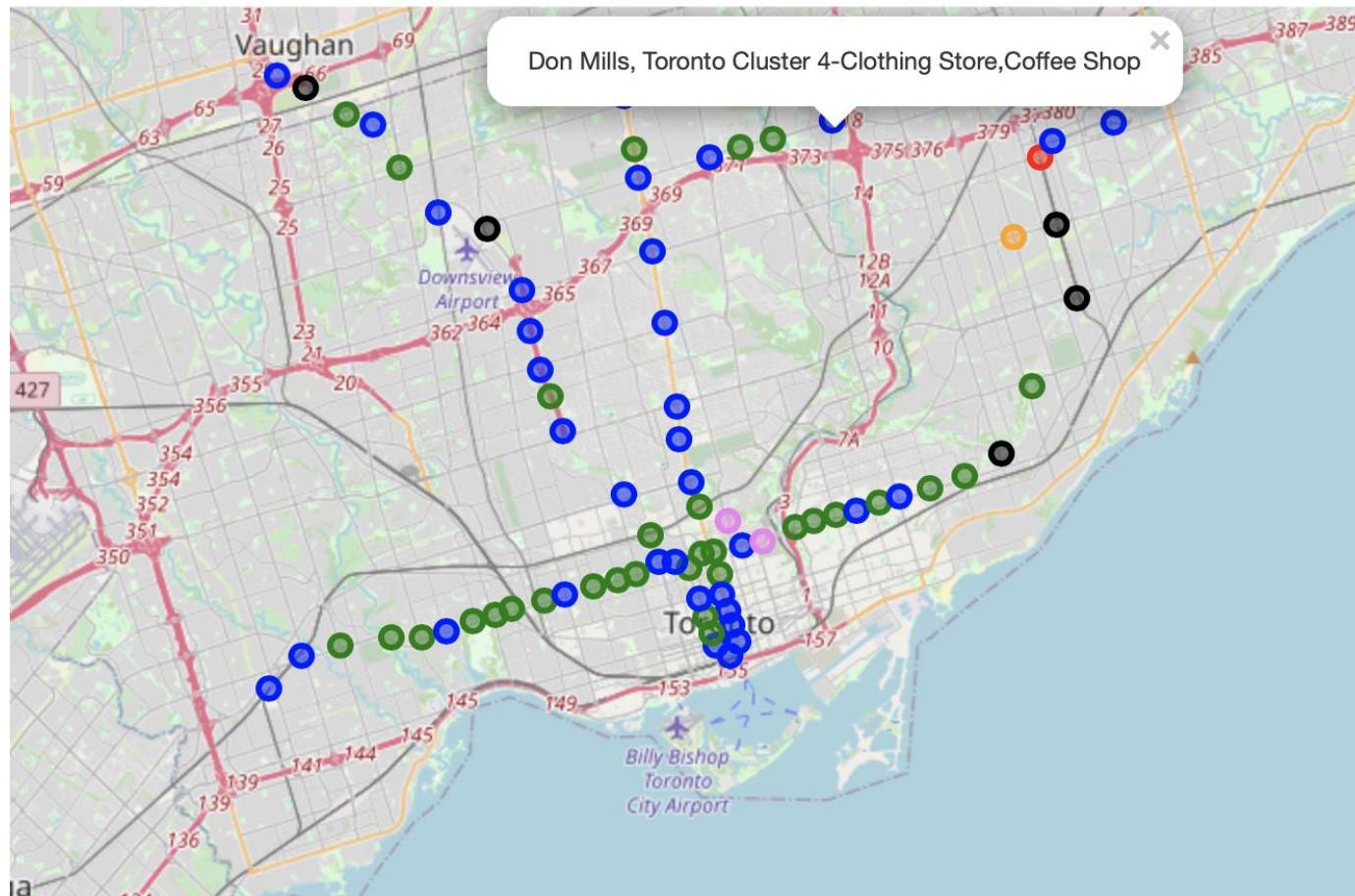
Analysis

5. Elbow at $k=6$ for K-means Clustering which means we can have **6 clusters** of subway stations.



Analysis

6. Folium map showing Toronto subway stations in 6 Clusters and 2 most popular venues around them



Results and Discussion

- 66 out of 75 Subway stations - fall under Cluster 0 or Cluster 4
- **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.

Conclusion

- Clusters provides insights into kind of venues to expect around any subway station
- People looking for Pizza place and Sandwich venues can explore venues around Subway in Cluster 0.
- People for Clothing stores can explore venues around Subway in Cluster 4.
- Anyone looking for Coffee shops can find them around subways in Cluster 0 and Cluster 4.
- Cluster 1, Rosedale and Castle Frank subways, have Park around them
- Cluster 3, Ellesmere Subway, has Gym/Fitness Centre as 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.