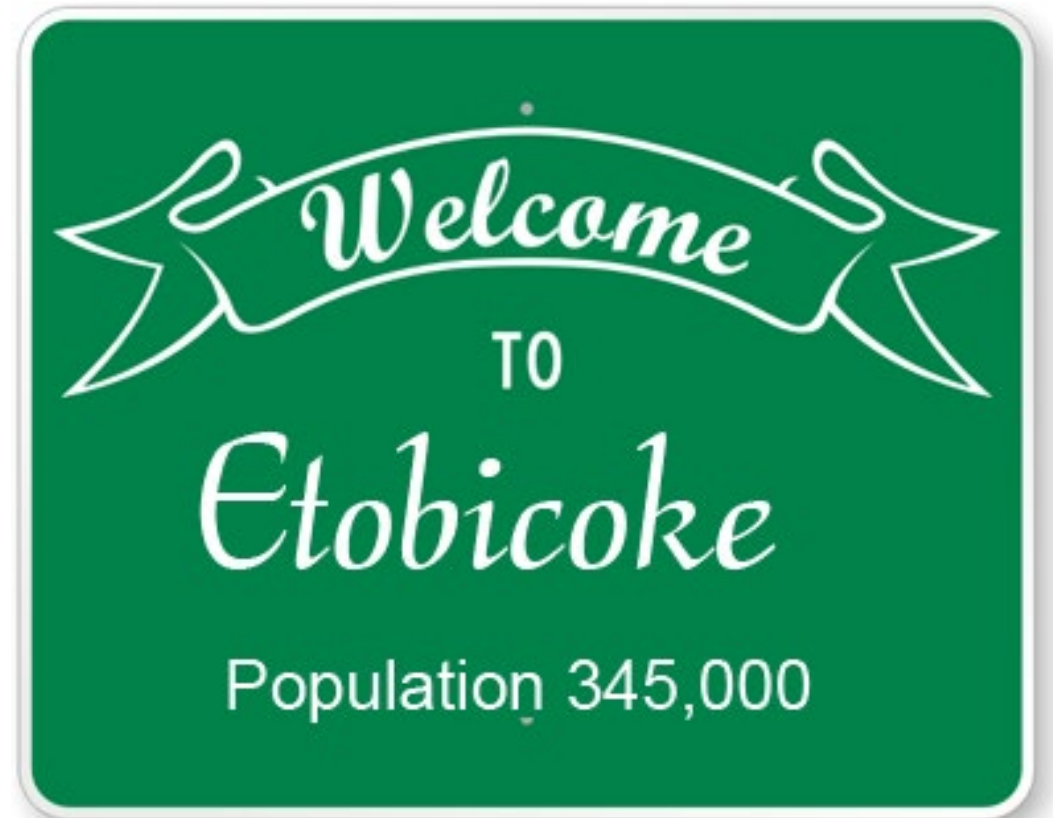


IBM Coursera Capstone Project

The Battle  
of  
Neighborhoods  
**ETOBICOKE**



# Introduction

This report is an analysis of the neighborhoods of Etobicoke in order to help Jay to make his decision.

Jay is 40 years old who born in an Indian family. He came to Canada with his parents in 1990. He has been living in Etobicoke since then but his problem is that since he came here to now, he couldn't find the one stop place where he can eat local Indian food, shop Indian groceries or other things such as buy sweets. He is a business man who wants to help other people, not just Indian, but the people from the neighbour countries such as Pakistan, Nepal, Bangladesh, Sri Lanka who are homesick or who can find themselves at home and find things they want from one place.

The information gathered from FourSquare in combination with Data Science methods are a good basis to derive data driven decisions regarding neighbourhoods that best fit the specific needs at hand. We are also going to use K-Means clustering algorithm to achieve the task. Folium visualization library will be used to visualize the clusters superimposed on the map of Etobicoke. These clusters can be analyzed to help Jay to make his decision

As we know already, Target Audience would be Jay who wants to open a South Asian Plaza in Etobicoke, but any stake holders can use this approach whether they want to open any shop, restaurants or anything.

# Data Requirements

All the postal codes are available on the Wikipedia page so we used following Wikipedia page to scrape out postal codes of the Toronto area, their respective borough and their neighborhoods. We focused on Etobicoke borough only.

- [https://en.wikipedia.org/wiki/List\\_of\\_postal\\_codes\\_of\\_Canada:\\_N](https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_N)

After getting postal codes, borough and neighborhoods from the above Wikipedia page, we used Google maps Geocoding API to get the latitude and the longitude coordinates of each neighborhood. For simplicity of this project I gathered all the information in the CSV file. Here is a link to a csv file that has the geographical coordinates of each postal code:

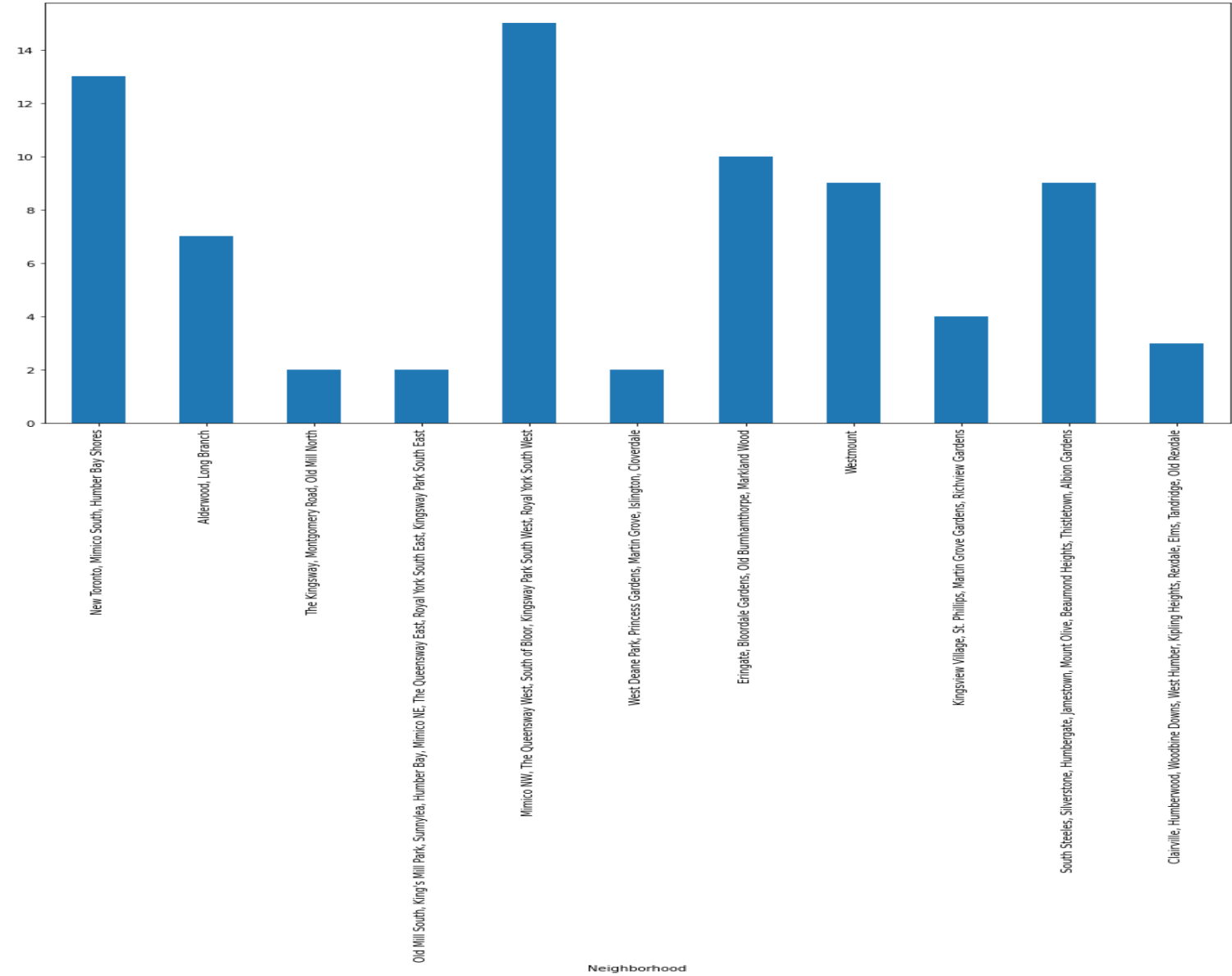
- [https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs\\_v1/Geospatial\\_Coordinates.csv](https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs_v1/Geospatial_Coordinates.csv)

There are a total of 12 neighborhoods. The details of the venues in 500 meters radius of these neighborhoods are obtained from Foursquare API. Those details are Venue, Venue Latitude, Venue Longitude and Venue Category. A total of 76 venues data have been obtained from Foursquare.

# Methodology

- Now, we have 12 neighborhoods data of Etobicoke. We also have the most popular venues in each neighborhood obtained using Foursquare API. A total of 76 venues have been obtained in the whole city and 43 unique categories. Here we have only 12 neighborhoods so we will consider and analyze all.
- We can perform one hot encoding on the obtained data set and use it to find the 10 most common venue category in each neighborhood. Then clustering can be performed on the dataset. Here K-Means clustering technique have been used.
- The clusters obtained can be analyzed to find the major type of venue categories in each cluster. This data can be used to suggest Jay some suitable locations based on the category.

# Analysis

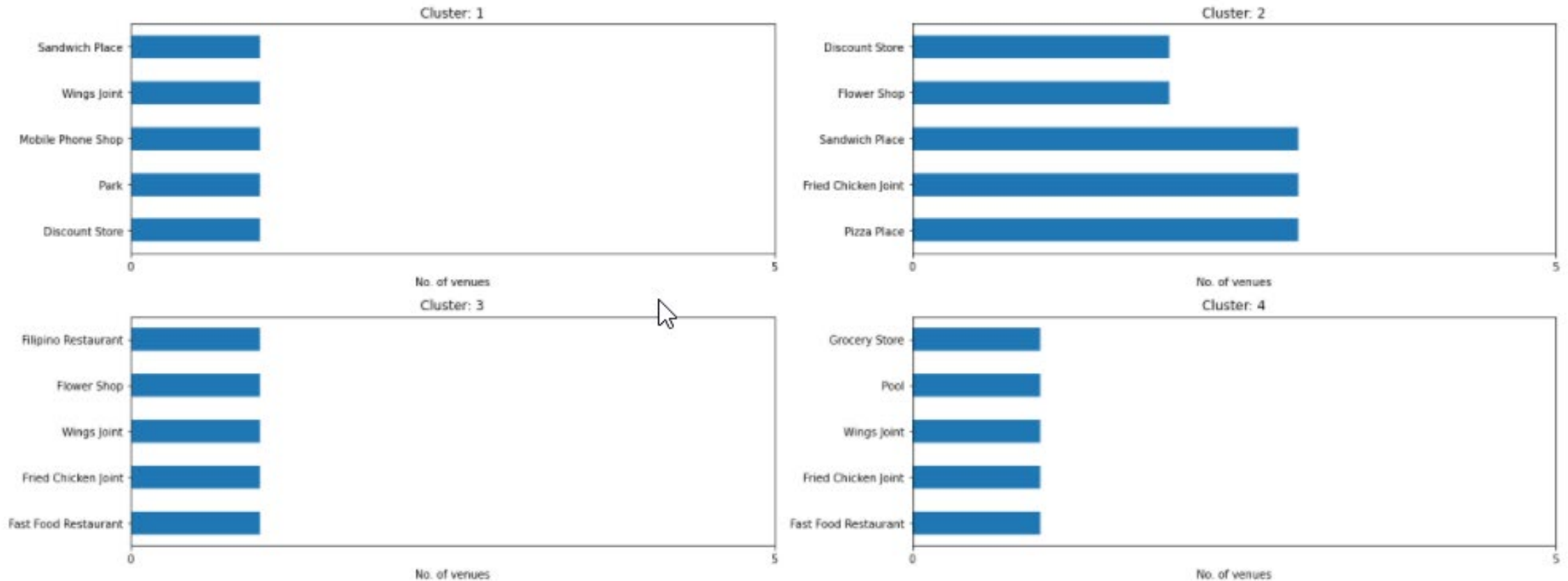


# Analysis

- As per given in the last chart, we found that there were few neighborhoods with as low as 2 venues. But we didn't remove those because we have only 12 neighborhoods, so we kept those.
- One hot encoding was performed on the data to obtain venue categories in each neighborhood. Then grouped data by neighborhood and take the mean value of the frequency of occurrences of each category.
- This was used to obtain top 10 most common venues in each neighborhood.
- The resultant dataset can be used for the clustering algorithm. So, here the K-Means clustering algorithm was used.
- We also did some data wrangling such as replacing all NaN values to 0 and add 1 to all cluster labels. So 0 became 1, 1 became 2 and so on.
- Total 7 clusters were created for better segmentation.
- In addition to this, we also visualize those clusters in map

# Results and Discussion

- Visualization of the top 5 most common venue categories in each cluster
- Cluster 1-4:



# Results and Discussion

- Visualization of the top 5 most common venue categories in each cluster
- Cluster 5-8:

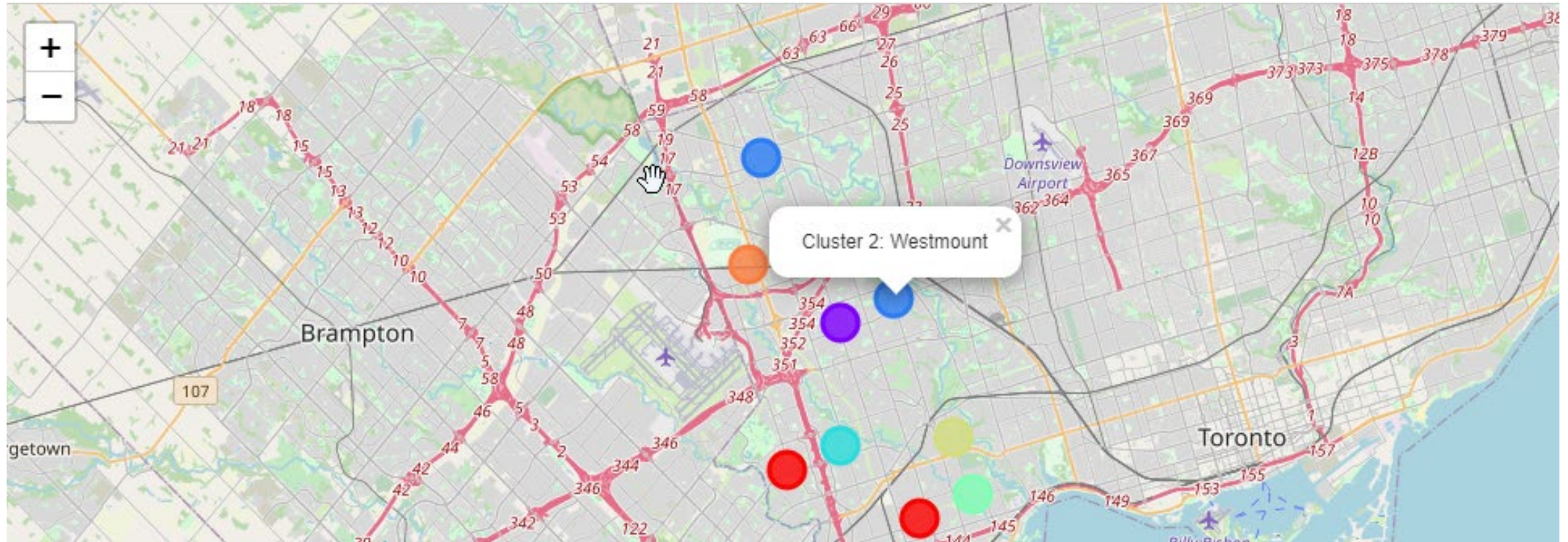


Note: Cluster 8 was the mimic of cluster 1 in order to create a better visualization so that's why cluster 8 has been blurred.



# Results and Discussion

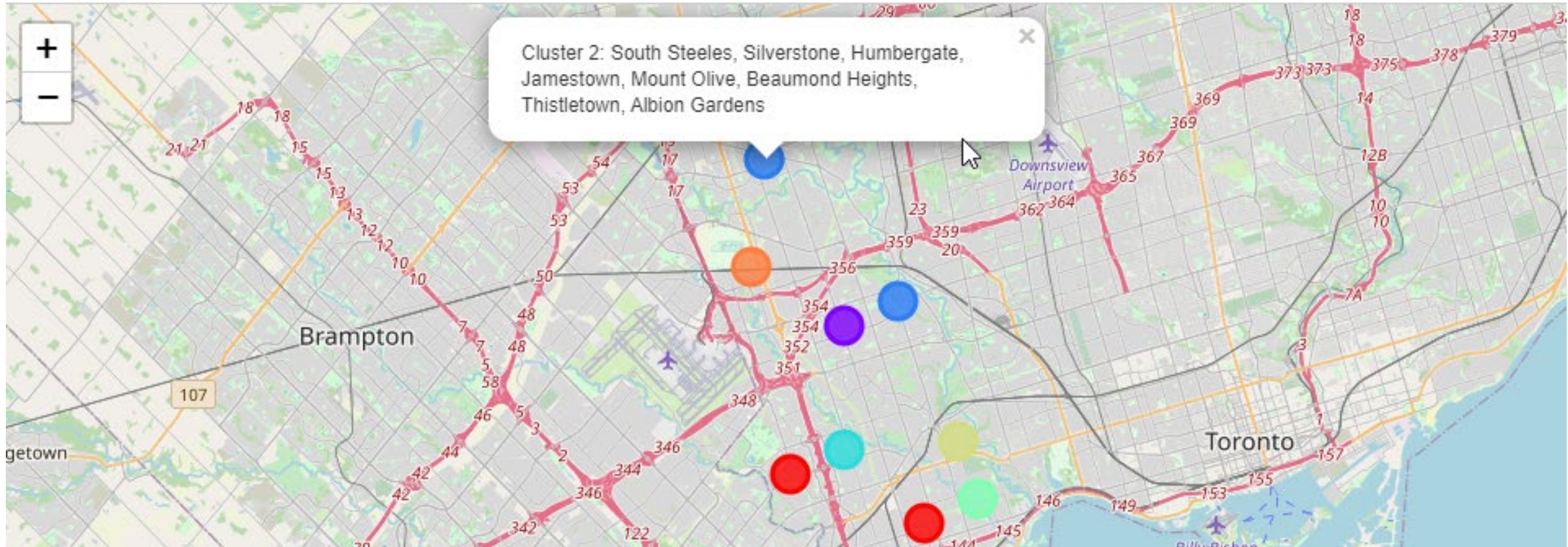
- Visualization of the cluster 2 in the map





# Results and Discussion

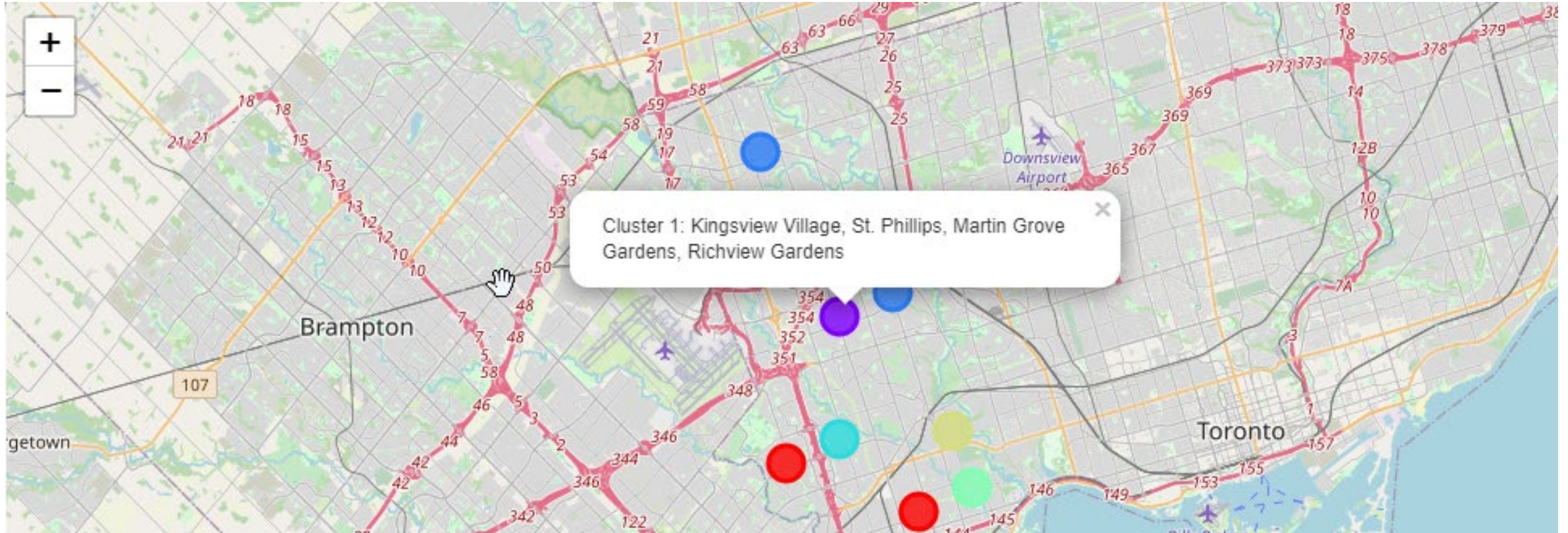
- Visualization of the cluster 2 in the map





# Results and Discussion

- Visualization of the cluster 1 in the map



# Results and Discussion

- The above slides of bar charts and maps can be used to suggest valuable information to Jay. Let's discuss them below.

## Cluster 2:

- It's very clear that there are many venues belong to this cluster such as few pizza places, some asian restaurants, fast food restaurants, playground, gym, etc. so it clearly gives an advantage to open any business in these neighborhoods. Jay can open his plaza in either Westmount neighborhood or near Humbergate. Let's discuss few advantages of these two neighborhoods.
- **Westmount:** As it shows in the map, it's very near to highway 401 which can attract daily commuters to this plaza. Moreover, this neighborhood is in residence area so it's possible that this plaza would attract many inhabitants and not just South Asian community.
- **Humbergate:** As it is depicted in the above map that this area is very near to Humber College and also to very popular Hindu temple (BAPS temple) which can attract many students and devotees.

## Cluster 1:

- **Kingsview village** in this cluster is very similar to the Westmount area of cluster 2. In addition to that, it has Bus line too so it may attract some people who has no vehicles such as students.

# Conclusion

- In this project, the purpose was to analyze all the neighborhoods located in Etobicoke city and cluster them using k-Means clustering techniques. We had to provide suggestion to Jay so he could open his plaza in suggested neighborhoods.
- We used Foursquare API to get necessary information of venues located near to each neighborhoods and their respective geospatial data. As we saw in previous visualization in discussion section there are not many popular neighborhoods in Etobicoke except some
- So, we took few neighborhoods into consideration and we provided suggestion that Jay can open his plaza either in Westmount or near to Humbergate.
- Kingsview village is very similar to Westmount neighborhood so we can remove Kingsview village from the consideration.
- We strongly recommend that Jay opens his plaza near to Humbergate area because it would boost his business. This place is very near to Humber college and some popular religious places nearby

