

# Capstone Project

Battle of Neighborhood (Toronto Non-Private Schools)

*July 2020*

# Contents

- Introduction
- Data
- Methodology
- Result
- Discussion
- Conclusions

# Introduction

- Early Learners Inc is interested in exploring the location and distribution of non-private schools in Toronto, as part of an ongoing product development effort.
- This study therefore seeks to explore the distribution of non-private schools across Toronto region. It analyses each neighborhood and identifies the leading location in terms of count and how these helps to group the schools for the purpose of further research works.

# Data

- **School Data:** Obtained from the Toronto open data web page. It is a geographical file covering the City of Toronto that contains all the point location of schools, including private schools. The data was last refreshed on Feb 4, 2019.
- **Postal Code:** I got the Toronto Postal Codes from Wikipedia.
- **Venue Data:** I used the Foursquare API to get the common venues based on the borough and neighborhood.

# Methodology

- Programming tools: Python program and modules
- Change management tool: GitHub
- I got the list of Postal Codes from Wikipedia and cleaned it up

```
List_of_postcode.head()
```

	PostalCode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government

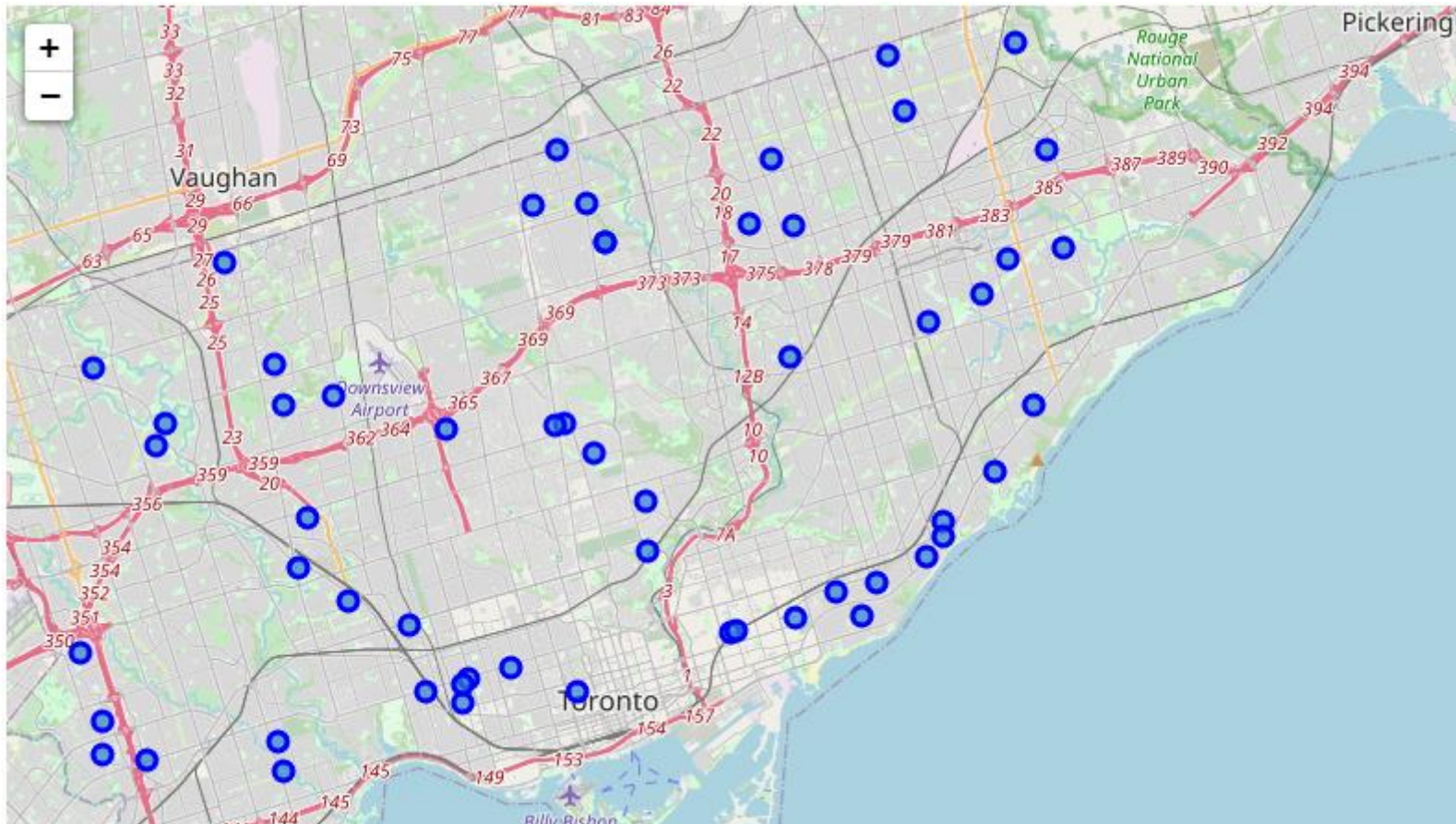
Number of schools per Municipalities in Toronto region (Total: 99).

```
school_type.sort_values(by=['municipality','count_of_schools'], ascending = False)
```

	municipality	school_type_desc	count_of_schools
8	former Toronto	Non Private	15
9	former Toronto	Priv	12
6	York	Non Private	2
7	York	Priv	2
4	Scarborough	Non Private	16
5	Scarborough	Priv	5
3	North York	Priv	23
2	North York	Non Private	13
1	Etobicoke	Non Private	9
0	East York	Non Private	2

# Methodology contd.

Visualized distribution of non-Private Schools only



# Methodology contd.

## Exploring the Top Venues:

I extracted the top 5 venues per neighborhood and their respective frequencies for the purpose of in-depth review, (see the excerpt below).

```
top5_venues_df[:5]
```

	hood	TopVenue1 freq	TopVenue2 freq	TopVenue3 freq	TopVenue4 freq	TopVenue5 freq
0	Bayview Village	Café 0.1	Coffee Shop 0.06	Clothing Store 0.06	Restaurant 0.04	Shopping Mall 0.04
1	Bedford Park, Lawrence Manor East	Sushi Restaurant 0.07	Coffee Shop 0.06	Bakery 0.06	Italian Restaurant 0.06	Sandwich Place 0.04
2	Birch Cliff, Cliffside West	Coffee Shop 0.08	Park 0.08	Gym 0.04	Pizza Place 0.04	Grocery Store 0.04
3	Cedarbrae	Indian Restaurant 0.11	Coffee Shop 0.11	Bank 0.08	Bakery 0.08	Gas Station 0.05
4	Christie	Café 0.11	Korean Restaurant 0.1	Coffee Shop 0.05	Grocery Store 0.04	Cocktail Bar 0.03

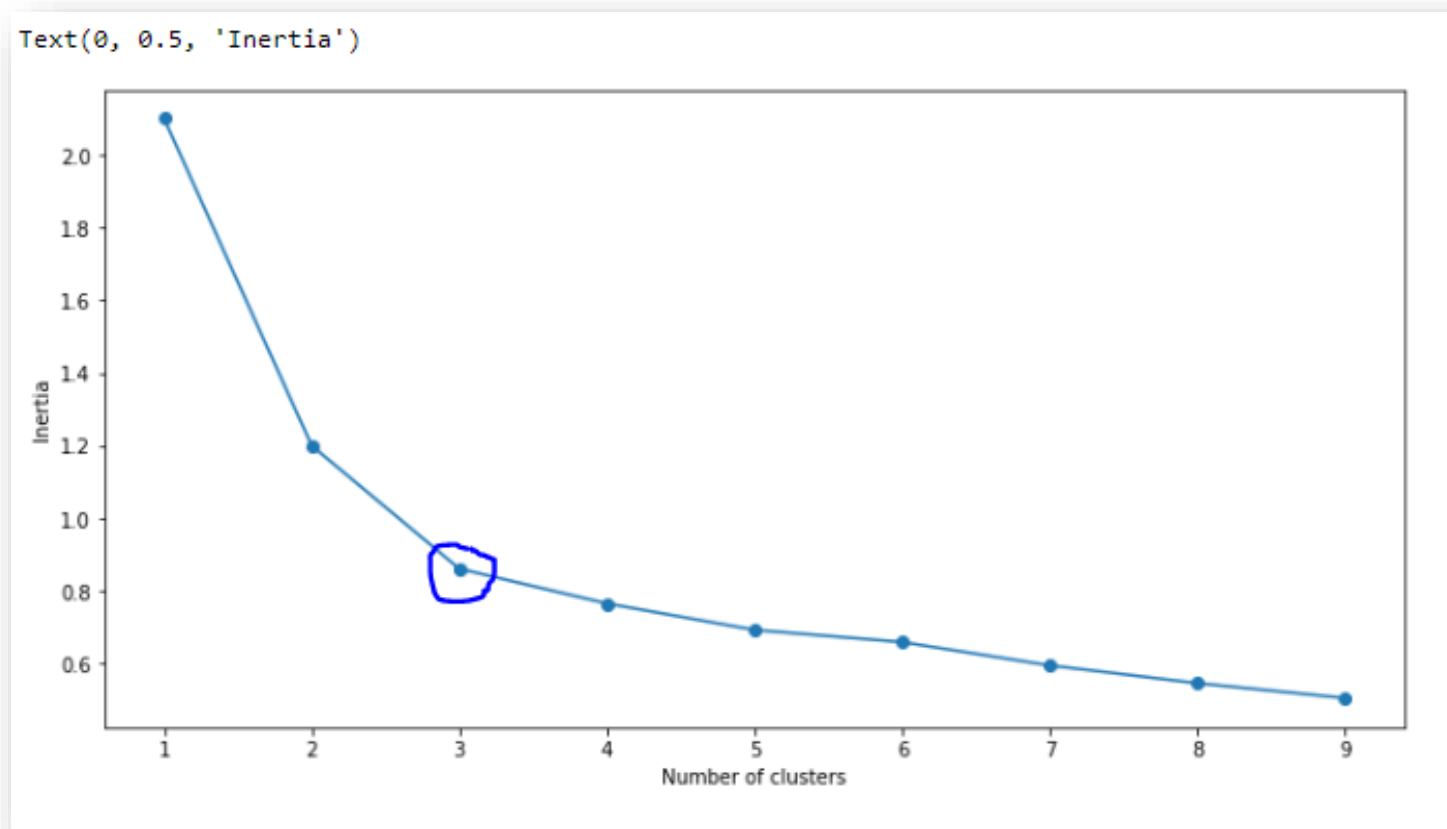
# Methodology contd.

- Clustering: From the sklearn module, I used the k-Means clustering functionality, which adopts an unsupervised learning approach, to detect and present clusters of schools based on the characteristics of their location or neighborhood.
- So, using the merged dataset of the schools and neighborhoods, we now have some location features for these schools that could be used for the clustering exercise. I selected 3 as my preset number of clusters (k), based on elbow method. (see the next slide)



# Methodology contd.

Elbow Method adopted to support the selection of ideal number of clusters (K)



# Results

- The map below is the visualization of the clusters of schools based on the k-Mean clustering algorithm. I have also shown the list of the schools that belong in each cluster.



# Discussion

- It is pleasing to know that most of the schools were grouped into one large category which contains 55 schools.
- Only two schools (Blessed Pier Catholic Elementary School and Ben Heppner Vocal Music Academy) were alone in two remaining groups respectively. These two schools will be studied on their own merit to understand the specific features that make them unique.

# Conclusions

- The categorization does not reveal any major divergence among the non-private schools in Toronto region, based on neighborhood characteristics.
- Overall, this preliminary analysis provides a very useful and strong support to the ongoing research at Early Learners Inc, Canada.
- After a critical review of the results, we should be able to make a better decision regarding the new product development and delivery channels in the most optimal way to serve these no-private schools.

Thank you