**Determining Best areas to live in.**

**in Toronto by using schools count**

1. Introduction

1.1 Background

Toronto is a very nice area to live in in Canada. All accommodation facilities are there and man families like to stay in.

1.2 Problem

If a new family is like to relocate to Toronto, there is no any measure to get best areas to live in.

1.3 Interest

solving this dilemma will help families in their relocation, and also will encourage other area responsible/government/municipality..etc to provide same accommodation facilities to increase probability for a family to live there.

2. Data Acquisition and cleaning

2.1 Data sources

We will use the Wikipedia Toronto Postal code source which will give the area postal code name, and area name. Also we will use the conversion sources from postal code to longitude and latitude, this could be a google function but for better and reliable use we have a prepared file for this activity. Also we will have to get the number of schools in each area by use of the Foursquare API.. we could get the average rating too but it needs a paid subscription

2.2 Data Cleaning

We prepared the data so that every postal code belongs to a named area .. no dummy postal codes are there. So total postal codes there are 103

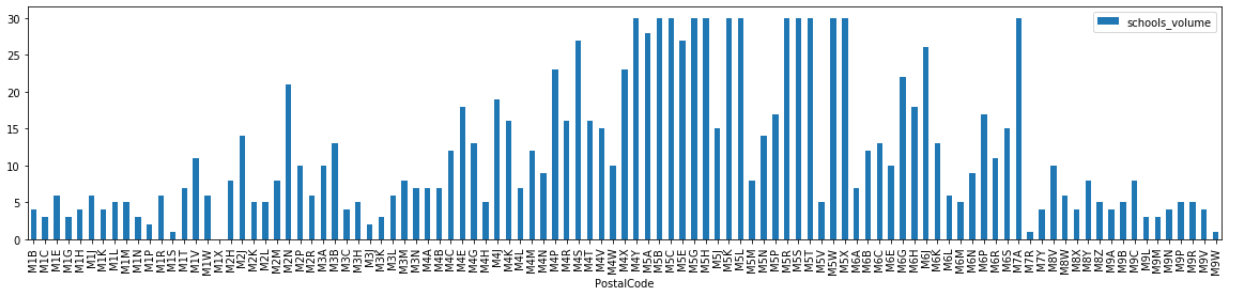
2.3 Feature selection

by using converting the postal code to longitude and latitude values and also by adding the number of schools found we are using three features. The schools value weight ofcourse are more than the location values (long. And lat.)

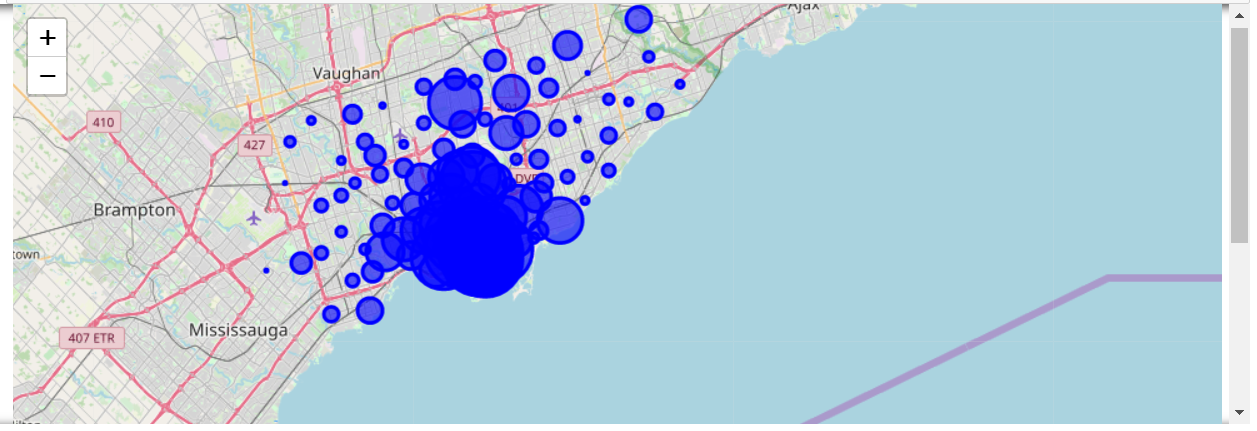


3. Exploratory Data Analysis

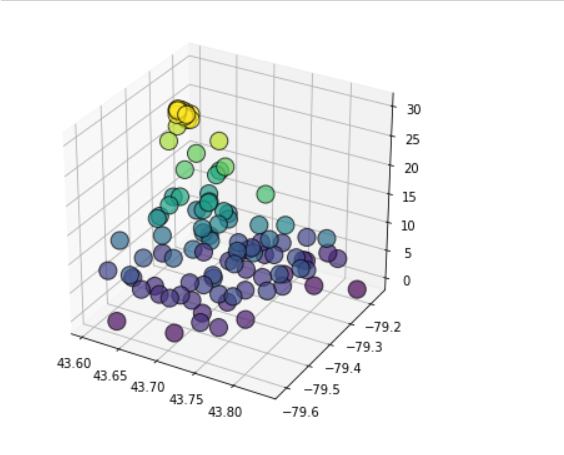
we examined the distribution of schools along the areas as in below figure, which showed how is schools are normally evenly distributed along most of the areas



This gives the idea that the Toronto region has a good distribution of schools along many areas



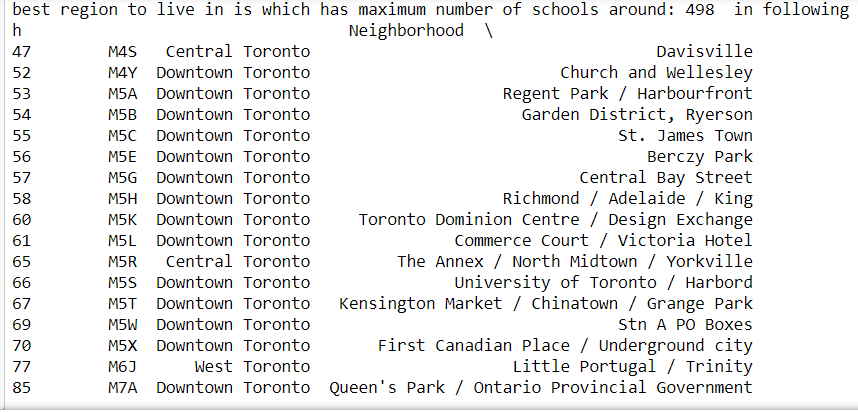
We calculated the scatter matrix between location and schools number which shows a kind of clustering in the areas as of below

where the height represents Schools volumes

4. Modeling

4.1 Cluster modeling

using the K-Mean method we calculated some the cluster areas which will lead to maximum number of schools around this areas. We are limiting the clusters to 5 which shows that there is a group of postal code regions which have the maximum schools around



5. Conclusion

we found an group of 13 areas in Toronto that best to live in for families.

6. Future directions

We can add in future many important features like ratings of schools, other accommodation facilities and populations that will give better results