.**Introduction**

If you're a Torontonian you know how much the city of Toronto needs to update their metro system. Despite being one of the largest and fastest growing cities in the world there is very little infrastructure available for its citizens. This report will aim to show a visualization of the different neighborhood’s populations and the corresponding proximity to subway stations. There will also be an analysis of distance to subway and whether population or income are predictors of distance from the subway. This may help citizens decide where there is likely to be congestion on their daily commute or where they can choose to live so as to be close to the Toronto subway system.

.**Data**

I will be leveraging on the data used in Week 3 of the Capstone Project which contains data for the postal code, borough name, latitude and longitude data for specific neighborhoods in Toronto. On top of this I will get population and income data from stats Canada which has population data by postal code. From Foursquare I will acquire the closest subway stations latitude and longitude and plot them on a Folium map to show how well the metro infrastructure service the populace of this city.

**Data sources -**

**Postal codes in Canada that start with “M”**

<https://en.wikipedia.org/w/index.php?title=List_of_postal_codes_of_Canada:_M&oldid=945633050.>

**Geospatial data for latitude and longitude for postal codes.**

<http://cocl.us/Geospatial_data>

**Population and Income data for all postal codes**

<https://www.canada.ca/content/dam/cra-arc/prog-policy/stats/individual-tax-stats-fsa/2015-tax-year/tbl1a-en.csv>

**Methodology**

The main methodology was to geospatially visualize the different neighbourhoods in Toronto versus the location of subway stations. The interactive map shows cluster of neighbourhoods around the city and the location of the subway line. The other analysis was using modeling (Linear Regression, Multiple Linear Regression and Polynomial Regression) to see if any variables like Income, Income per capita or populations was predictive of proximity to a subway station.

**Results**

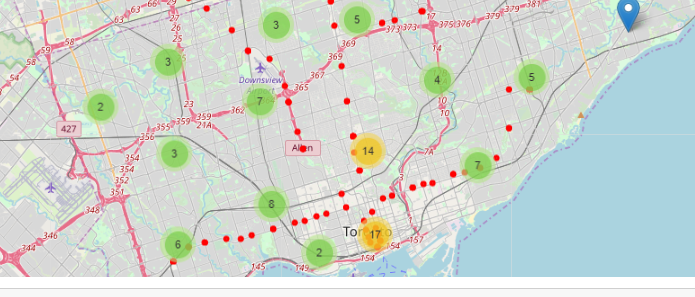


Figure 1 – Clusters of neighbourhoods and the red subway line

* Neighbourhood clusters at the bottom are well served by the subway line but the north east and North wets are not

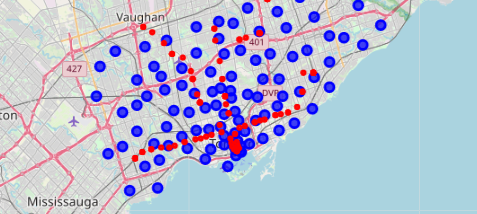


Figure 2 – Clusters of neighbourhoods and the red subway line

* Several neighbourhoods are not close at all to the subway line.

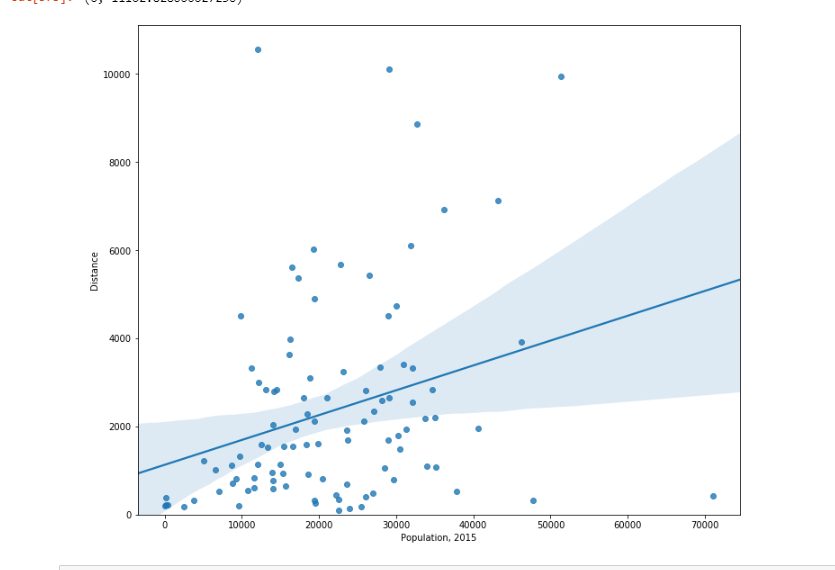


Figure 3 – Weak association of Populations of postal code versus proximity/ distance to subway

* Population does not have a great effect on the distance to subway

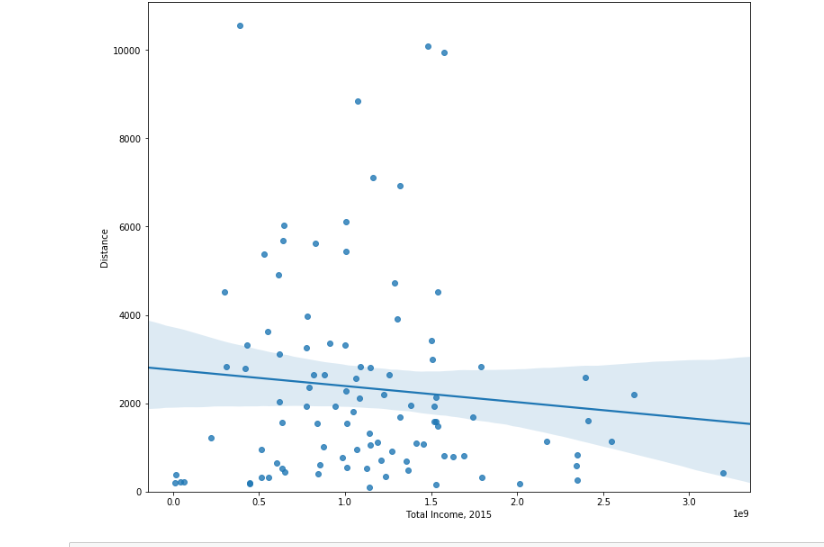


Figure 4 – Weak association of total income of postal code versus proximity/ distance to subway

* Total Income does not have a great effect on the distance to subway

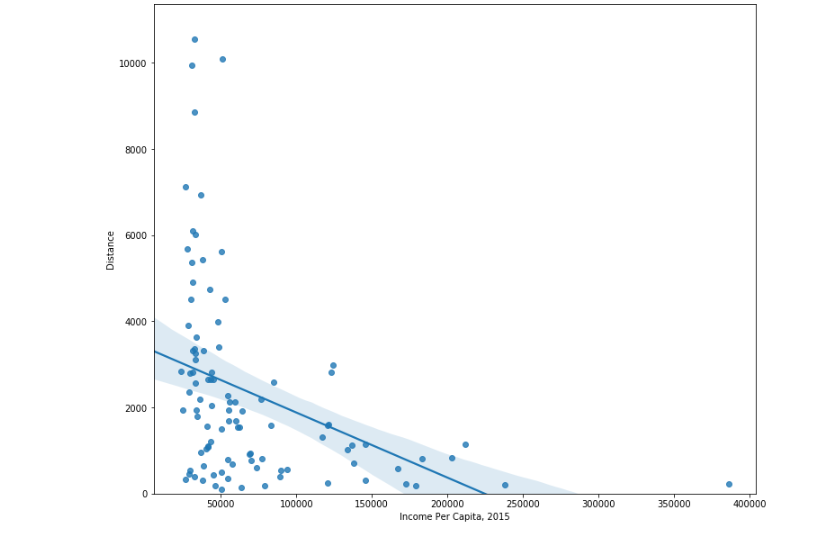


Figure 5 – Weak association of Per Capita Income of postal code versus proximity/ distance to subway

* Income Per Capita does not have a great effect on the distance to subway

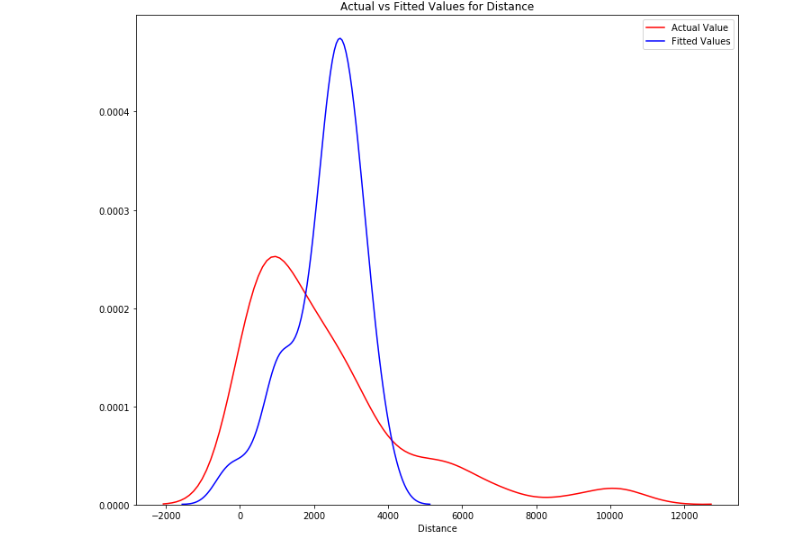


Figure 6 – Fit of Multiple Linear Regression Versus the actual values

* The multiple linear regression model of using Total income, population and Income per capita of the postal code matched the data pretty well.

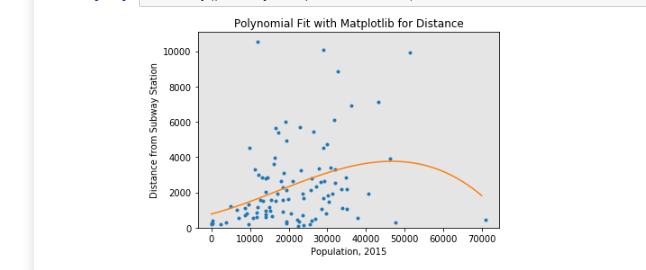


Figure 7 – Polynomial Model of Population versus distance to Subway

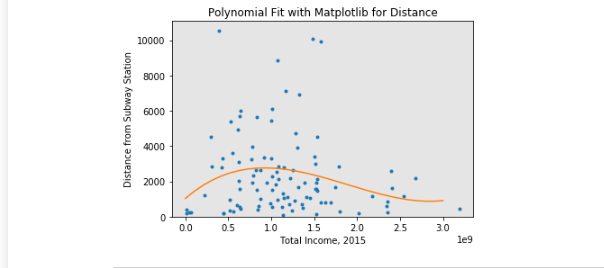


Figure 8 – Polynomial Model of Total Income versus distance to Subway

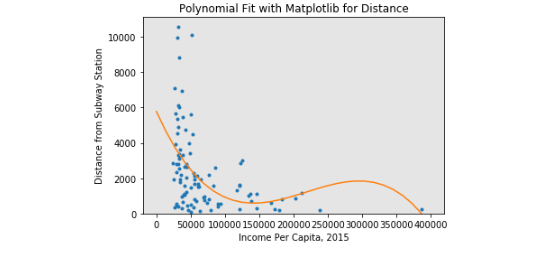


Figure 9 – Polynomial Model of Income Per Capita versus distance to Subway

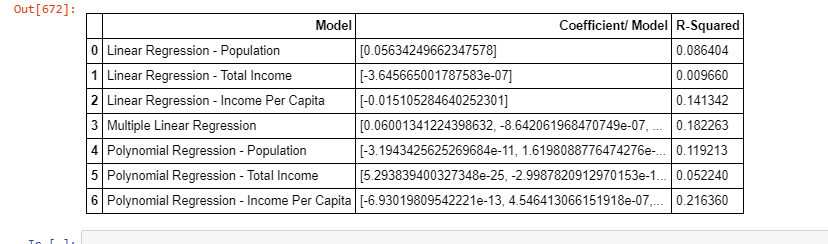


Figure 10 – Stats from the modelling

* Most of the variables were not very good which is maybe expected unless one was expecting income to be a driver of more subway stations close to them which is not the case in this data as our most predictive variable is Income per capita with and R-Squared value of 21.6% meaning our model can explain 21.6% of the data. There are no doubt much more variables needed for predictiveness. What’s disappointing but probably not surprising for Torontonians is how poorly predictive the Population data was of distance to subway. Ideally, the subway would serve as many people as possible but this data shows the locations of the subway stations do not serve the populations well.

**Discussion**

The Toronto subway system needs to serve more people better. Ideally, you would like to se the north east and north west Toronto neighbourhoods served better by the subway line as shown by the folium visualizations.

**Conclusion**

We can see from the data that more stations are needed to be installed strategically in order to serve more Torontonians.