

Edmonton Metro line project

By Jayden Geisler

Introduction

I live in Edmonton Alberta Canada and if you live here too you will know that the LRT system (Light Rail Transit) is not very good for people who live on the east and west side of the city. In this project I attempt to use data of the population and the city to try and plot out the most efficient new LRT line. A new efficient line would value the number of people it reaches and if they will use it as well as the value of the destinations it leads to. The city would have interest in this problem if they were looking to expand the LRT line

Data

I will do this by making 2 data sets. One will be the population of the neighborhoods along with income and education percentage which was attained from the city of Edmonton. The other data set will be venues visited within the neighborhood; this will be obtained through foursquare API. the source of the data is the city of Edmonton data library

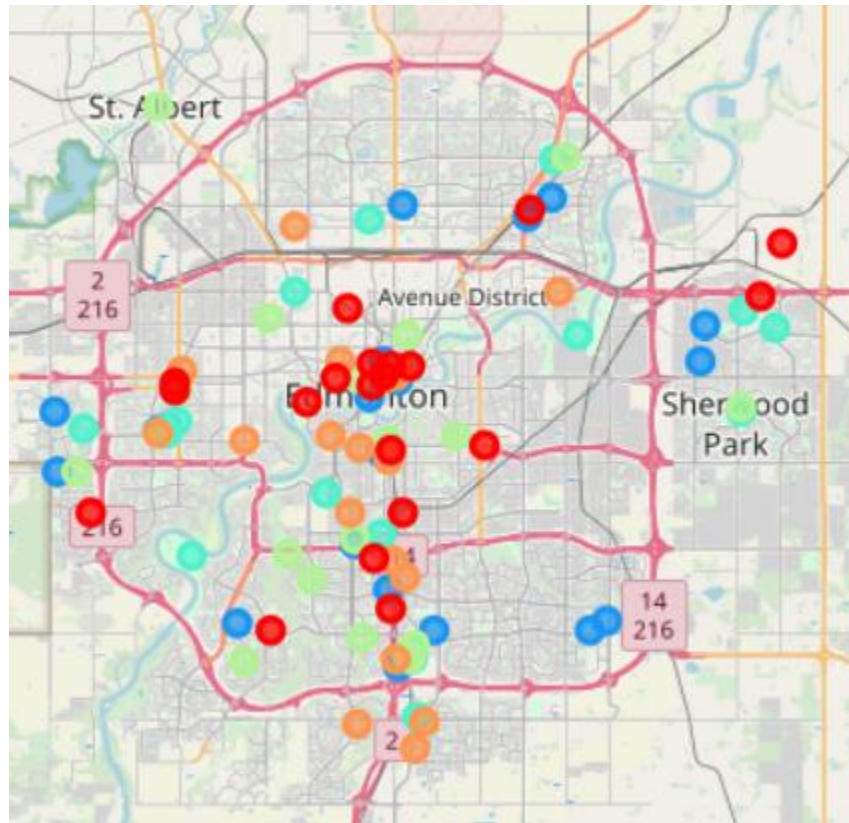
Methodology

To solve the problem, I will do two k means on each of the 2 data set. On the venue data set there will be between 3 to 5 clusters and the highest valued cluster will be used as “endpoints” for the new line. For the second data set there will have more clusters, between 8-12 so there are many different values for each neighborhood. Reading a paper for 2017 from the American Public Transportation Association 51% of the people who ride transit hold a bachelor’s degree, for income low - middle income people uses transit the most. After the 2 clusters have been mapped out the center points of the data will be connected through the use of a A-star algorithm that uses both cluster values as a heuristic.

American Public Transportation Association article: <https://www.apta.com/wp-content/uploads/Resources/resources/reportsandpublications/Documents/APTA-Who-Rides-Public-Transportation-2017.pdf>

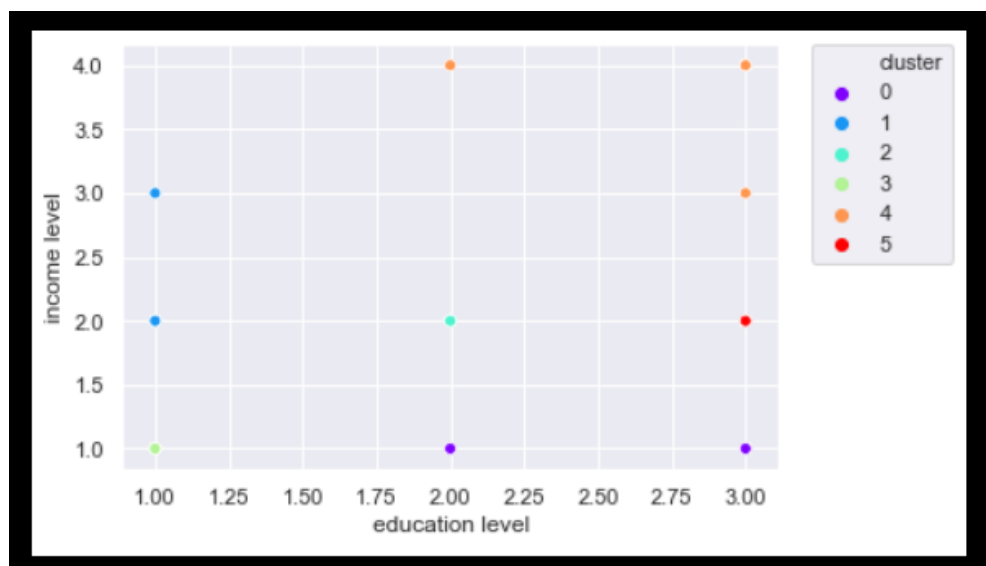
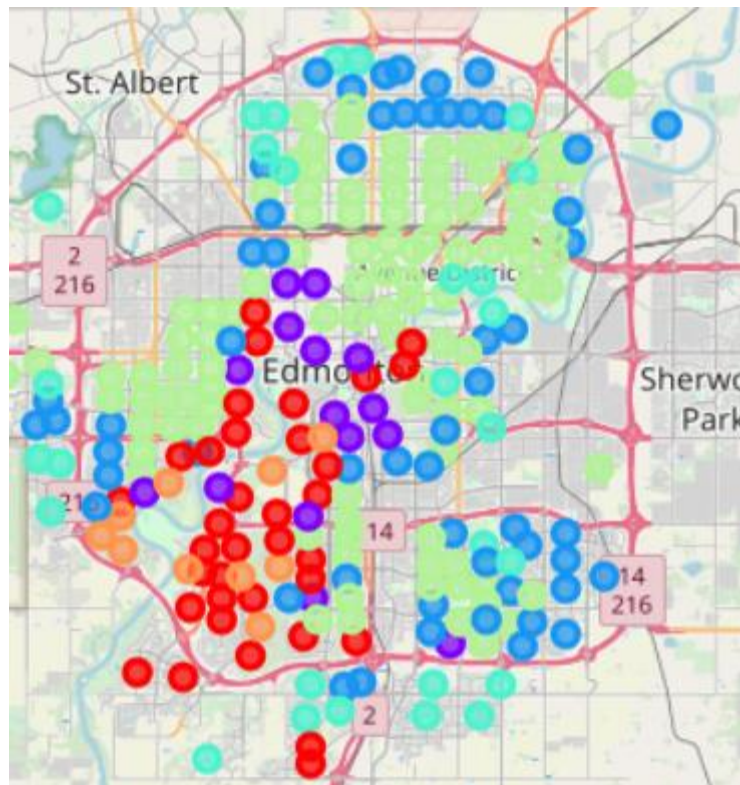
Most Popular locations

a Map of the most popular location within Edmonton where blue is the highest rank and red is the lowest



Most popular locations are down town and south, around the current line but does not go as far south. Also cluster to the west is west Edmonton mall which is a very popular spot.

Neighborhood clusters and charts



You can see how the close to the river in and in the south is where the high income high educated populations are

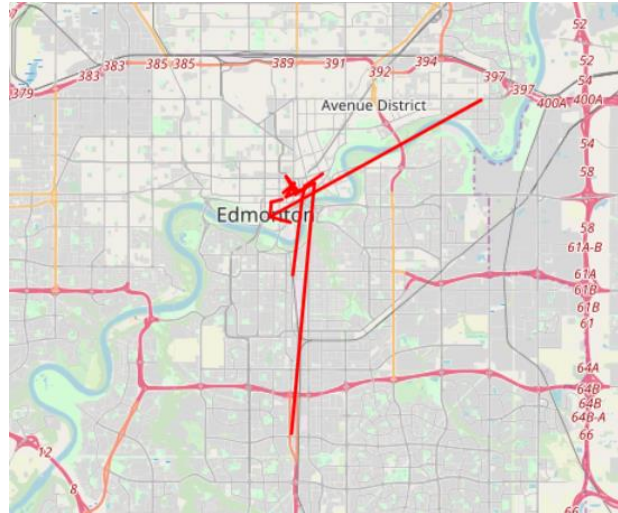
Making a Heuristic

to find the best path we will have to calculate a heuristic. this heuristic will be calculated in 2 different categories.

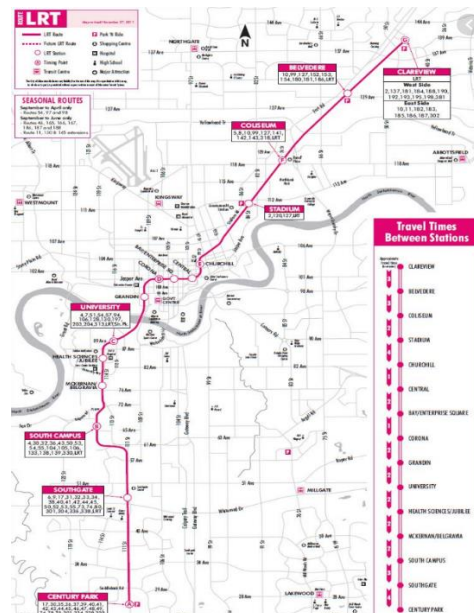
1. the second will be locations in the city, these will be classified as possible endpoints so the popularity. the heuristic will be divided up by rank = 1, $h = 1.00$, rank = 2, $h = 0.75$ rank = 3, $h = 0.50$, rank = 4, $h = 0.25$
2. the first will be the neighborhoods or the type 1 and will be based on the clustering and population cluster values are only 0-5 but population values vary a lot will use the normalized value key demographics are high educated low income areas so clusters 0,2,5 are most valued at 1.0 low income and low education clusters 1 & 3 will be valued the same at 0.75 high income is not valued as much so cluster 4 will be valued at 0.25 The total heuristic will be made by adding the population and cluster heuristic

Testing the path algorithm

when the start point is placed in the middle of the city the line is scattered in the center but shoots out south and north east this is similar to the actual line that is in the city



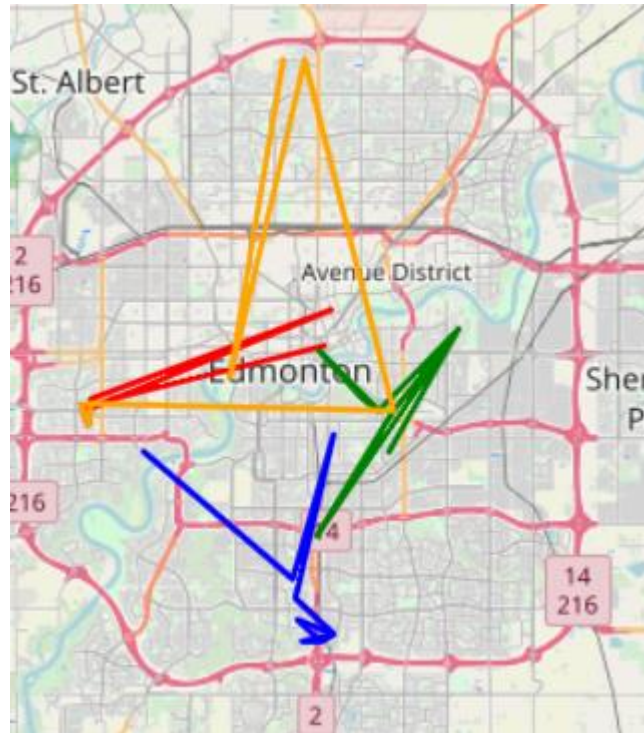
and the actual line that is in the city



as you can see they are pretty close and my prediction is that they would be closer if foursquare picked up the University of Alberta but currently in the data set that is not a location it would definitely be popular enough to spend the line more west

Results & Discussion

The final results look more like abstract art than it does a metro line lay out, but there is still some interesting take away from this result



The first thing to note is even though the red line started at west Edmonton mall the north and south lines both make their way toward it, making it a popular location and one that should be considered for the new line. The above image is not that of a best new line but that of the best line for each side of the city, for people on the west side the best line would be that of one north of the river that goes downtown. and people on the east side would benefit most for a line that then goes south and toward downtown. Starting in the south the best most popular routes go north and toward west Edmonton mall. The north line splitting at the top shows that picking up greater amount of people is more valuable than getting to a location that is directly downtown.

Conclusion

The purpose of this project was to see where the best new LRT line would be, to help the city planners plan the next line in the future. By calculating and clustering the neighborhood demographics and then creating a pathing algorithm we can show the general path of people of where they live to where they want to be. There are multiple improvements and changes that I would like to implement in the future these are:

The path algorithm still needs some work to smooth out the lines, but the jagged lines still tell a story about where the population is and where they want to go.

More data would also be a massive benefit foursquare is not used that much in Edmonton and the city actually has much better datasets.

many more scenarios with start points, even use regression to find optimal start position

Thank you for taking the time to read my analysis of the Edmonton LRT system