**Project Problem and Hypothesis**

In 2012 Super storm Sandy damaged the Canarsie Tunnel, which carries the L train under the East River between Brooklyn and Manhattan. Due to the extensive damage the tunnel, would need to be repaired. Two options were presented, to do the repairs under a full 18-month tunnel closure or a one-track, three-year closure. After holding 4 large interactive community meetings and 11 community boards visits, the results from emails, social media and meetings are that 77% were in favor of the full, 18-month tunnel closure set to begin in 2019.

The impact of the tunnel closure will effect 225,000 riders who travel between Manhattan and Brooklyn, while 400,000 people ride along the entire line.

With this conundrum, my focus is on the real estate effected. More specifically, I believe that there will be a decrease in the residential prices in the Williamsburg region of Brooklyn. There will be no focus on commercial property. I want to focus on Williamsburg because of the socio economic boom that it has undergone over the past 10 years.

Machine learning will be used to predict the accuracy of my hypothesis that the average residential property value will decrease between 2% - 5%, while holding all other variables. I will also try to prove my hypothesis correct with historical data from when the R train was shut down. The biggest impact in predicting the decrease in value are the implications from other projects that predict there will be an increase in demand and property value by neighboring areas whose transit systems will not be closed. Such neighborhoods as Greenpoint and Bedford-Stuyvesant.

**Datasets**

The datasets will be used from the MTA real-time data feed and list of static data feeds from my MTA Feed API Key: f30756de00aa2ea9d90a26c34fdb32f3.

Additional datasets will be provided from NYU Furman Center, which “collects a broad array of data on demographics, neighborhood conditions, transportation, housing stock and other aspects of New York City’s housing and neighborhoods.”

**Domain knowledge**

I am new to the field of price prediction, in the real estate market, but this is a topic that is of interest for future endeavors. I am directly effected by this project, as I am a rider of the L train, which is another reason as to why I am conducting this research to test my hypothesis.

There are articles with projections on an increase/decrease of real estate prices effected by the L train, such as dnainfo.com, yet these articles don’t appear to provide any data sets. There are data visualizations but no raw data accessible. There is a similar study with the L train except, the focus is on alternate routes. The blog is https://carto.com/blog/looking-at-the-l/.

**Project Concerns**

My major concerns for my project of course, would be the coding aspect and to make sure that, not only are the codes entered correctly, but also, I want to make sure that I properly interpret the outcome. I want to focus more on the average pricing and to see the change in percentage of my hypothesis.

The assumptions are that prices within the Williamsburg area will go down. These assumptions have other factors that tie into the construction of the L train line. The belief is a migration to surrounding neighborhoods as many foreigners may find better real estate opportunities in areas with reliable transit systems. There are other assumptions not related to pricing as well.

I was looking forward to having data from other articles that have made the assumption and provided data visualizations. I would have loved to see the methods that were used and how they came about their findings.

My datasets already make the assumption that there will be a decrease in price. Also the historical data from when the R train was shut down, also supports my hypothesis.

The major risks of this project would be if my hypothesis was completely off. If my findings were to be published and used by real estate developers, tenants, investors, etc., and my predictions were wrong than that would be a loss of credibility. For now, this is a learning experience and timing is a major factor here. I am predicating future prices for the year 2019, so only time will tell. At the moment, my learning and projections are my benefits and outweighs the cost.

**Outcomes**

I expect the outcome to be that my hypothesis of residential real estate prices in Williamsburg, Brooklyn, will decrease by 2% - 5% as a direct result of the L train being shut down for 18 months in 2019. I believe my target audience will agree with me as well. I want my outcomes to be as simple as possible so that my target audience will understand the results with as little confusion as possible. Ideally, I would like to be as accurate as possible, but if the results lie 1% outside of my projection, I would still consider it to be a success. If it is more than 1% outside of my projection then it will be considered a bust, which is alright because it will give me the opportunity to go back to the drawing board and try other methods.