

Final Report

Abstract/Intro

Natick Public Schools are experiencing heavy traffic congestion and have observed that one of the major factors could be because many students in Natick are not taking the bus. Parents report that poor crosswalks and long bus rides are the main deterrents. Our main focus on this project is to analyze data provided by the school and determine factors motivating students to opt to drive or walk to school rather than riding the bus. Additionally, we will propose possible solutions that would incentivize more students to ride the bus.

Data Sets

The data sets used were as follows:

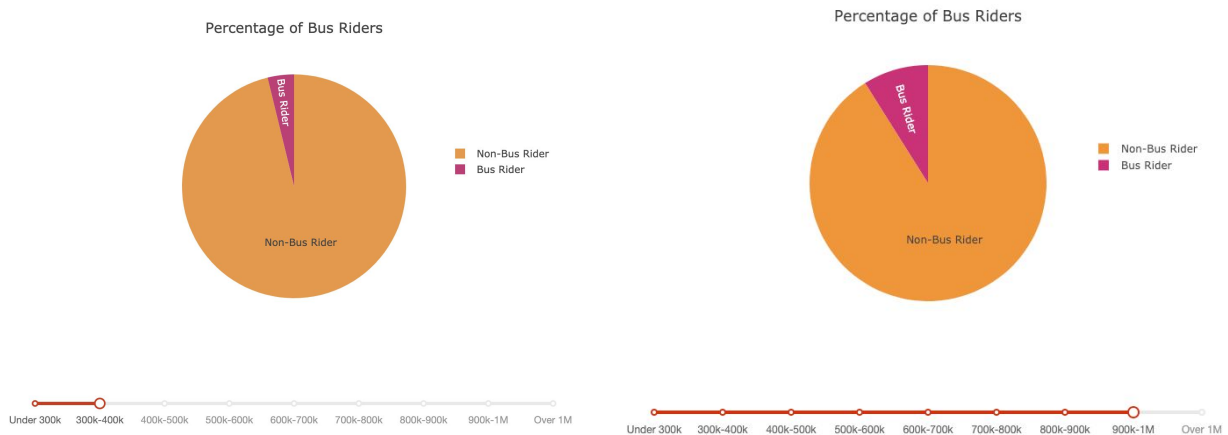
- The transportation study which gave us the addresses of students who take the bus.
- The Spark! property data set which gave us the total assessed value associated with each home in Natick.
- The ASAP student address data set which gave us the addresses of students who do not take the bus
- The Student Address data set which gave us all students and their addresses

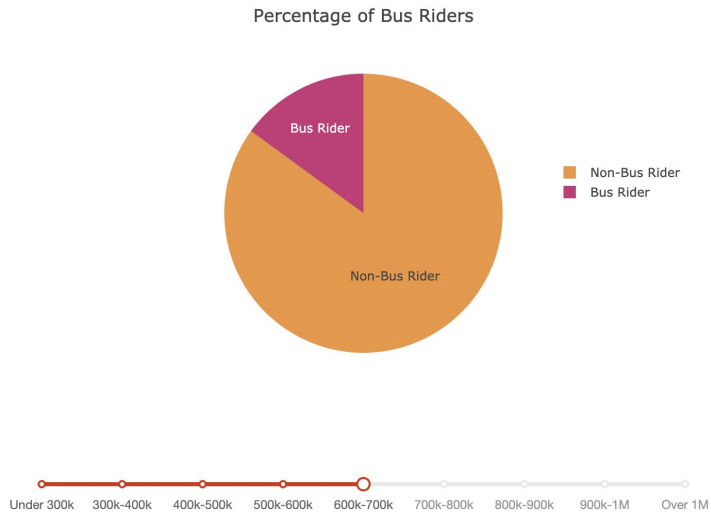
Using these data sets, we were able to find a correlation between the assessed value associated with a home and whether students take the bus. We were also able to pull all the bus stops from the transportation study as it contained the bus stops students would go to. Using this data, we were able to find more optimal positions for these bus stops. We were given other data sets corresponding to the status of sidewalks. However, we were unable to find a correlation between students who take the bus and damaged sidewalks.

Property Value Analysis

Property value is a factor we considered that could determine whether a family decides not to use the bus as their primary mode of transportation for their children. In order to find a correlation between income and whether or not students take the bus, we analyzed the “Property value of enrolled students” data and compared it with “Addresses that pay for bus services”. We found that families with extraordinarily low or high associated assessed values were less likely to take the bus. This may be for a variety of reasons, for example, the reduced pricing for taking the bus may still be too expensive for students with property values under 400k. And houses with higher property values can afford to buy their children their own cars. Although we didn’t pinpoint an exact reason, we did find that there is a correlation between assessed value and whether a student takes the bus. Given a household with an assessed value less than 400k, a student is 80% chance that they don’t take the bus.

Bus Riders vs Non-bus Riders Based on Property Value



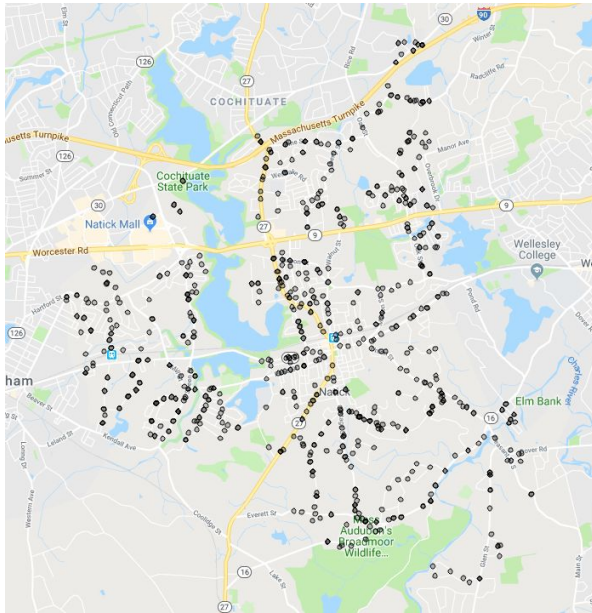


For our statistical analysis, we calculated the correlation between household value and if they take the bus or not. This will tell us the impact property value has on the likeliness of a student taking the bus. Using proportion agreement (accuracy) we had an kappa value of ~80%. We used binary variables to represent whether or not a student takes the bus, and whether or not their house value is worth over an average value of 400k.

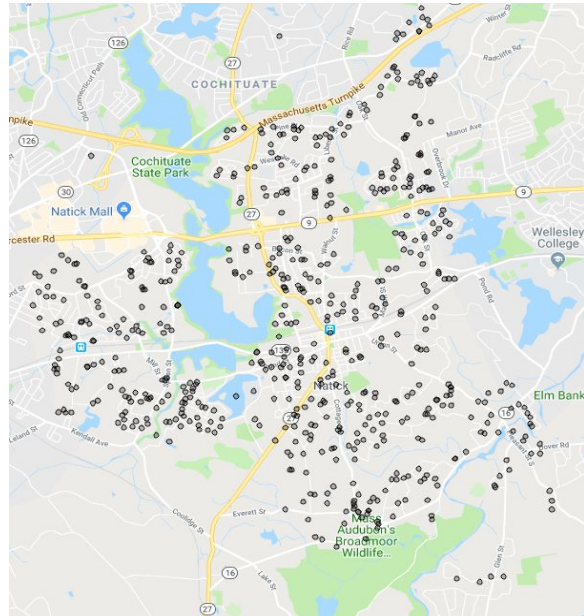
Bus Stop Relocation

Relocating the bus stops in order to decrease the distance needed for a student to get there would be a great incentive for more parents to let their kids use the bus. To address this problem we decided to use k-means to minimize the walking time between houses and stops using the Google maps API. Using the current bus stops as our centroids and the students' addresses as the points, we ran k-means to find the optimal positions for each stop, based on the distance to each student's house. We calculated the average travel times between students walking to old bus stops and compared them to the average travel time it would be if new stops were implemented. Our implementation of k-means algorithm yielded smaller travel times on average by the students. Our results found that it is possible to have more efficiently placed bus stops, however, some of our new stops are placed inside homes, bodies of water or places without roads. This is solved by placing the bus stop onto the nearest road.

BUS STOPS BEFORE



BUS STOPS AFTER



Conclusion & Future Work

In order to incentivize more students to take the bus across the nation, we must optimize bus stop placements and bus routes. K-means is a great resource to use for this problem. As we have seen from the household analysis, there isn't a clear reason as to why more students are not taking the bus. Therefore, we need to analyze more data in order to pinpoint the main factors contributing to this situation. For the future, we would like to do a comparison of traffic congestion caused by cars that parents use to drive their kids to school with the traffic congestion caused by buses. We believe that congestion would be much lower with school buses and we can use that data in order to incentivize more parents to let their kids take the bus.