Problem Statement - Web Based Population Clustering

Connor Barthen, Nathan Barthen, & Cole Cassano

Bus routing in school districts is not always very effective or efficient. Finding a way to ensure effective bus routing for school districts is essential to get students to school on time, use less fossil fuels, and to reduce overall costs. Maintenance costs for the vehicles and their fuel can become issues without proper routing. It is also integral to get the students picked up on time so they can arrive at the school for classes. Some of the other main issues with bus routing include how many students can occupy each bus as well as how long each student will have to ride on the bus. Similarly, these buses will need to be aware of district lines to know the scope of their routes.

To solve this problem, it will require a multi-step process of data acquisition, formatting, implementation, and user intractability. The two major components of data will need to be acquired through the US Census Bureau and Google Maps APIs. From the US Census Bureau, we will need information regarding the tiger files for geographic separation of the school districts, road networks, general addresses of students. However, the general address will be acquired though another site, openaddresses.io. With the general addresses we will need to do clustering to determine bus routes and number of busses needed. We also need to attain the location address of the actual elementary and high schools in the district and convert the address to longitude and latitude. After acquiring data from these sites, we will need to design a framework on QGIS and Eclipse that will take these datasets and generate the bus routing information for each school in a given district in Pennsylvania. This will require extensive testing and debugging to effectively implement this solution. Lastly, it will need to be user interactable on a webpage that will allow the user to input a specified school district to generate the bus routing.

By the end of this project, we hope to have created a software that will load a webpage with an UI that a user can use to look up a specific school district. The software will then load data from the US Census Bureau about that school district and cluster student houses with that information. Then it will use this data for the clusters while keeping in mind constraints such as time and number of seats. This tool would allow school districts to look up and generate rough layouts for bus routes and give a good estimate on the number of buses needed for each school. If schools are informed of this information, it would make planning and resource allocation more efficient through the insights provided by this web-based population clustering tool.