ASSIGNMENT 1

Jinlun Zhang	**20055835**
Wenqi Tang	**20093622**
Eissa Khan	**20082302**

Used Software: Google Colab, VSCode, Power BI, Jupyter Notebook

Used Package[1] (Python): Pandas, Gmplot, Matplotlib, Googlemaps, Geopandas, Seaborn

Used Dataset[2]: Transit Data - October.csv, transit-gtfs-routes.csv, transit-gtfs-stops.csv, civic-addresses.csv

Description of the analytics process used: We start by analyzing the 'Transit Data – October' data collectively to obtain a general understanding about the transit data and how to address the provided 4 questions. Through some simple data cleaning and pre-processing of the 'Transit Data – October' data, each team member uses Python or other visualization tools to generate figures related to their assigned questions. Moreover, to further support the answers we arrived, we have researched and analyzed other datasets that contain more geographic information about the bus routes and Kingston resident addresses. Then, by using the Google Map API and Geopandas Python package, we manage to visualize those geographical data interactively.

Bus Service Accessibility (Eissa):

The data used was obtained from the City of Kingston's official data catalog. This data was parsed into pandas for location name and longitude latitude for each bus stop. Next, the bus routes were obtained from the same source as a geojson format file. The file was loaded into pandas and the routes were extracted as paths with latitude and longitudes. Finally, all civic addresses in Kingston were obtained in csv format. This data also contained the geolocations of each address as a geo point. This data was parsed into pandas as latitudes and longitudes. Once again, this data was obtained from the City of Kingston's official data catalog.

Matplotlib was then used to create a nice presentable plot for exploration and interpretation. The bus routes were plotted using a scatter plot and colored orange. Civic addresses were plotted with a low alpha value in blue. This implicitly gives us a type of heat map with dense areas being opqaue blue, and less dense as transparent blue.

[1] The packages are all supported in Python, which can be installed through command: pip install * (as demonstrated in the submitted code)

[2] Datasets can be retrieved at: https://github.com/JinL-Zhang/CISC451-/tree/main

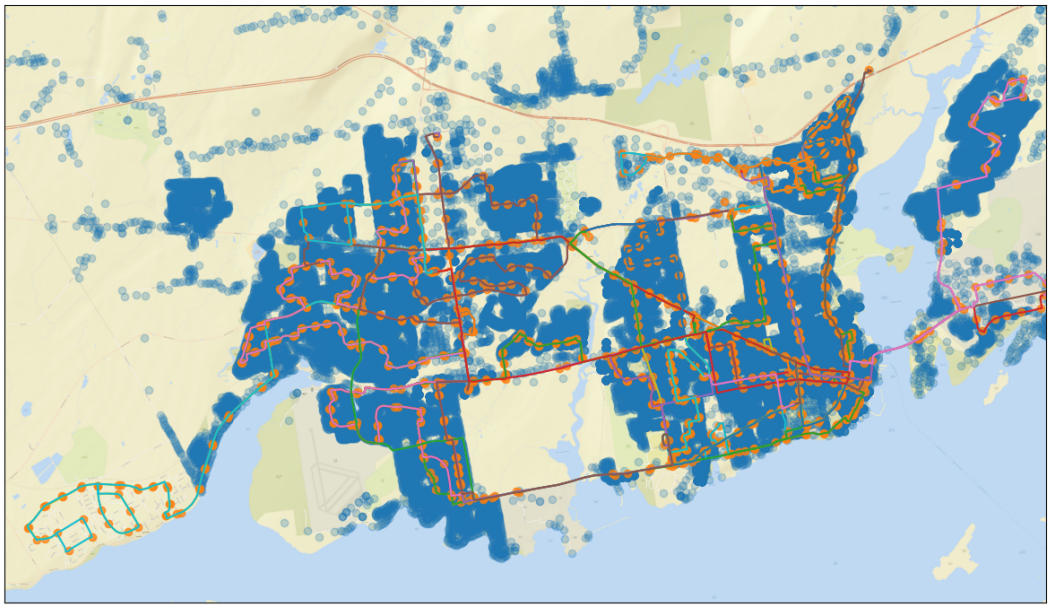

Figure 1: Bus routes and civic addresses

From the final plot one can make many interpretations and conclusions about how well/poorly the citizens of Kingston are being served by their public bus service. The most striking observation is that the rural population is poorly served, with little to no bus routes within these communities. In particular, north of the Macdonald Cartier Freeway has no bus routes going in or out. Moving East ward, CFB Kingston has only 1 bus route, in-spite of having a relatively dense population.

The bus routes in Kingston mostly serve the urban population well. Moreover, the bus route themselves in these areas are fairly convenient, in the sense that they do connect neighborhoods to neighborhoods, unlike other cities where all routes only flow in and out of outskirts of a city to the core down town, and aren't very well connected to the surrounding areas.

Redundant Bus Route (Jinlun):

After pre-processing the 'Transit Data – October' data, I utilize the Seaborn Python library to generate the column chart displaying the counts of buses (y-axis) by its routes (x-axis) and classes (Legend) as shown below, which provides me an idea about which route may be redundant.

As shown illustrated below, though Routes 8 and 13 are used by the extra buses to help service peak demand in certain areas of the city, they are not heavily used by the students or the citizens, while for the route 17, which is primarily used by Queens students and is not designed for the extra buses, has a relatively low total demand comparing to the other routes such as route 2 that are used by Queen's students for more than 40000 times in a month. Thus, from this one graph, we may conclude that it will be cost-efficient to discontinue the route 17 bus service as it appears to be redundant, especially during the summer time when the majority of Queen's students have returned to their homes.

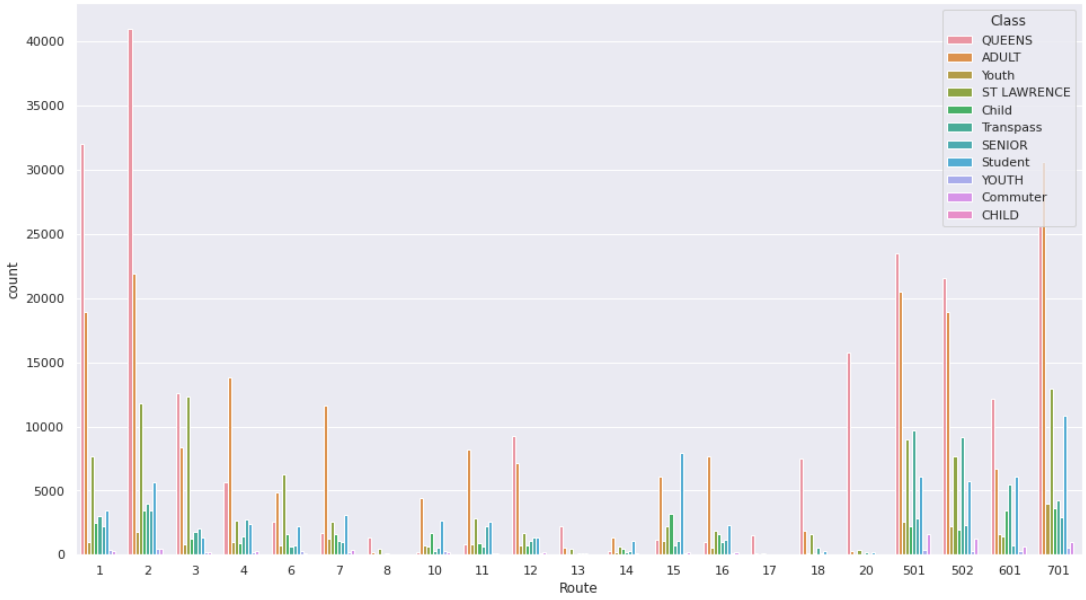

Figure 1: Column chart displaying the counts of buses by routes and classes

To further support my conclusion, I have employed Power BI to produce more comprehensive and interactive graphs as shown below, which includes 4 charts illustrating the 'Transit Data – October' dataset. As shown in the 'Count of Class by Class' graph (the first graph, column chart), we realize that the students in general are the users of the Kingston Bus service, and among those students, Queen's students are the main customers of the bus service (which can be partially attributed to the bus fees that are included in the Queen's tuition). However, when we filter by route 17, we notice that there are very few passengers taking the bus running on the route 17. Furthermore, moving to the third graph (Count Route by Route), which has already been filtered to highlight the route 17 portion only, the pie for route 17 is not only 'invisible' to view but also account for less than 0.5% of the total buses operated in the month (as only 1.71k buses used the route 17, while 24k buses only account for 3.4% of the total buses operated in the month). Thus, I can affirm the conclusion I draw before that the demand for the route 17 is low. Moreover, while this route is primarily used by the Queen's students as indicated by Figure 1, the Queen's students' demand for the route 17 is very low comparing with the demand for the other routes, implying a redundant bus route.

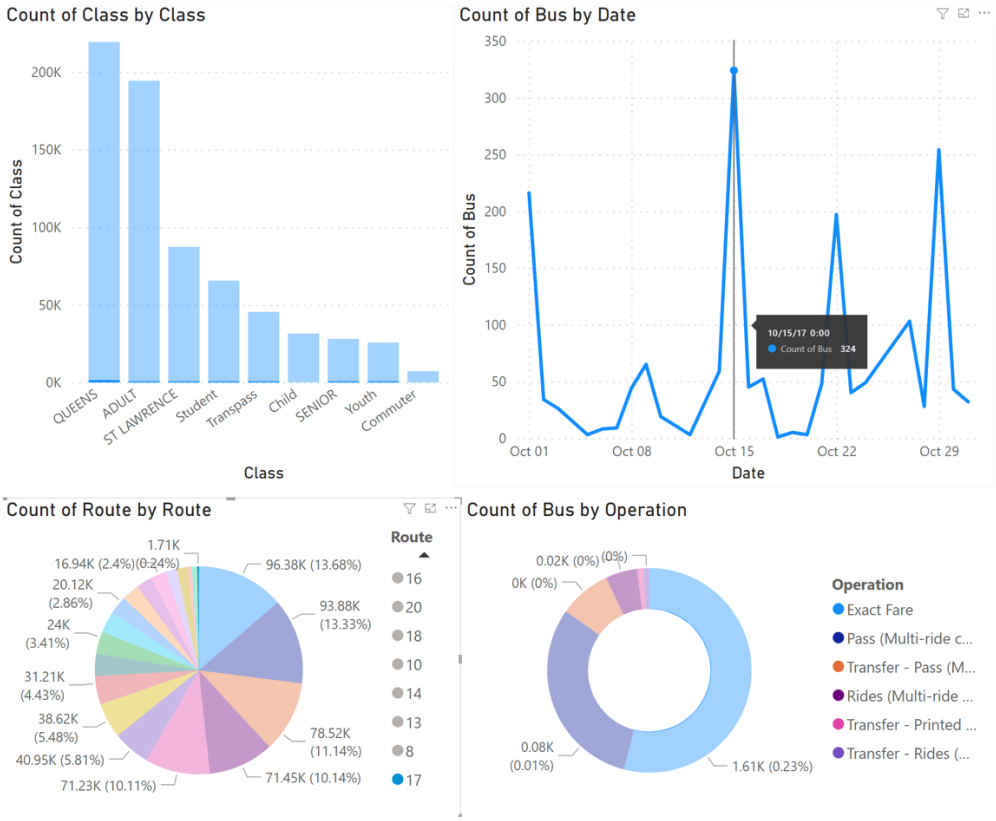

Figure 2: Interactive Power BI Figures Specifically Focusing on the Route 17 Data

In addition, through using the visualization provided in 'Open Data Kingston' website and the IPython Python Library, I manage to display the route 17 that are used by bus '17A', '17D', '17P', '17W' as shown below (indicated by black lines). Through observing the figure, we can easily see that the route 17 is mainly covering the Queen's Campus district and the length of the route 17 is relatively short comparing with the other routes such as the route 2 (represented in the figure as yellow line), which explains why the route 17 bus service is nearly exclusively used by Queen's students, rather than the Kingston citizens.

Therefore, as the bus services are mainly used by students, and the route 17 is mainly used by Queen's students, while the Queen's students' demand for this bus route is low comparing with the other routes due to its small coverage, we may consider to remove this route if the maintenance cost of or the operation cost of this route is high, and the limited contributions from Queen's students to this bus route is not sufficient to cover the expenses, meaning that it may be economically feasible to remove this route as it appears to be redundant for the city transit as whole.

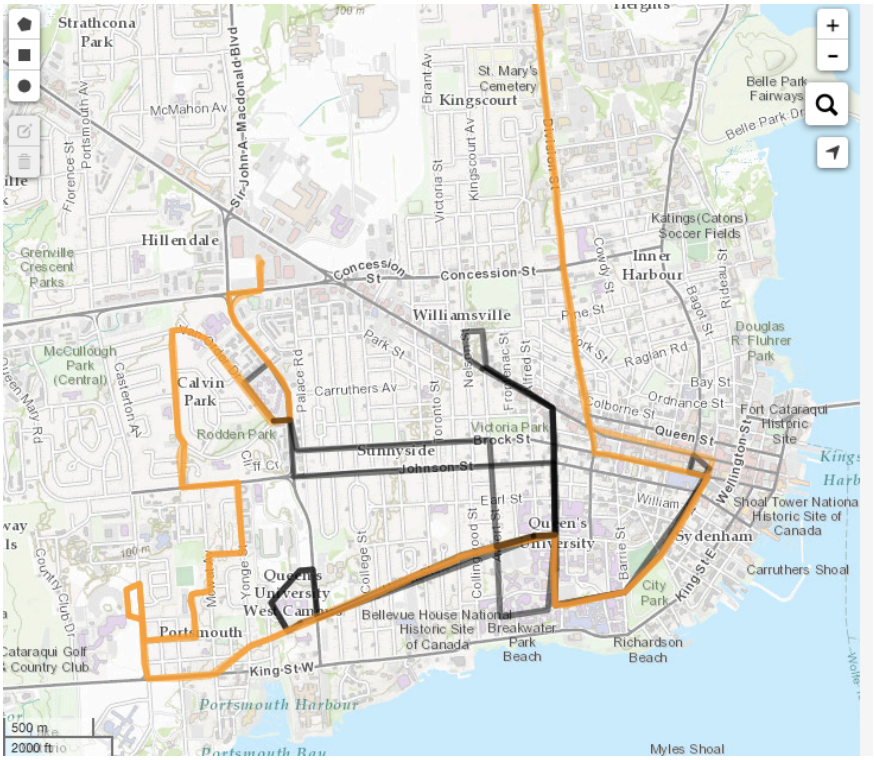

Figure 3: Geographical illustration of the transit bus route 17 (black) and the route 2 (yellow)

New Bus Routes to Make People Leave Cars at Home and Reduce the Gas Emissions (Wenqi):

We are trying to analyze the number of bus routes that need to be added to make sure that people can leave their cars by trying to see the relationship between the existing bus route and the driveways. Extracting data from the 'transit-gtfs-routes.geojson' file, we can get a bus route figure (figure 1) as shown below in blue lines.

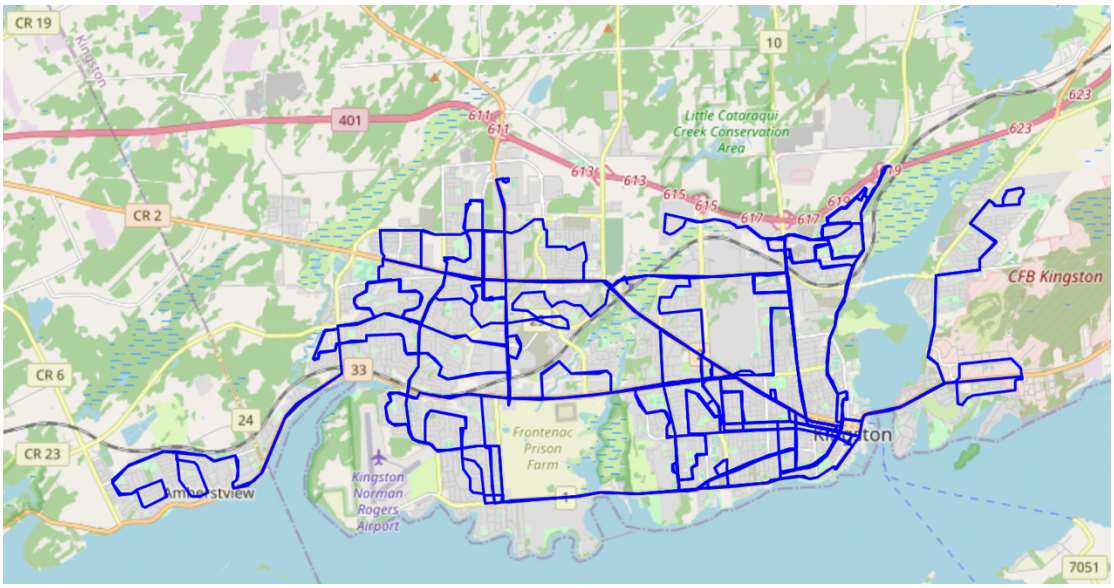

Figure 1: Bus Route Illustration

Basically, we can see from figure 1 that the bus route is kind of sufficient. Citizens can just walk to a nearby bus stop to get on a bus. However, that is not enough evidence that the route is enough so we also obtain data from 'driveways.geojson' and get figure 2 which shows driveways in yellow spots.

Figure 2: Distribution of Driveways (Yellow Lines)

Although the basic need of the majority of residents is satisfied by the existing bus routes, some areas are not covered by the bus route that has plenty of driveways. Hence, it is a good idea to add 3 extra bus routes. The first one can be added at Lappan's Ln, going from the Novelis area to downtown Kingston. The second one can be added at Sydenham Rd, from Lakeshore school to downtown Kingston. So this line not only makes sure that the students in Kingston can go to school easily, but also the people who live in that area can get to downtown Kingston. Last but maybe not least, one can be added at the

further way of Princess St, from Turtle Crossing to downtown Kingston, because this area also has many residents living there.

So now moving on to the gas emissions, let's first assume that all the people who live in Kingston work in downtown Kingston. Assuming that all the people in Kingston choose to take a bus instead of staying at home. Assuming that one-eighth of the population in Kingston needs to drive to work before deciding to stop driving. The emission of one mile is about 0.404[3] kilogram. So it is one kilometre is about 0.000650036 tons. The population in Kingston is about 117,660[4], so the driving population is about 14707. Assume that the average driving distance per day is 8 kilometres. Then per year is about 12*22*8 = 2112 kilometres. Thus, the total emission that is saved is 0.000650036*14707*2112 = 20190.7 tons.

Interesting in the data that can help planning the bus routes (Jinlun):

When we combine the 'transit-gtfs-stops' dataset with the 'civic-addresses' dataset (plotting the bus stops on the embedded heatmap of the Kingston resident addresses), we realize that there is a civic region, as indicated in the black circle of the below figure, where no bus stops at, while the region is still considered to be within the city of Kingston. Moreover, as shown on the bottom left of the below picture, there are bus stops (blue dots) in the area where the residential information is not included in the 'civic-addresses' dataset, which may imply that the Kingston Buses are serving the people living outside of Kingston. Thus, we may want to add a new bus route to connect the citizens living on the far west of Princess Street and reduce the bus operating on the Amherstview area.

[3] https://www.epa.gov/greenvehicles/greenhouse-gas-emissions-typical-passenger-vehicle

[4] https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=POPC&Code1=0415&Geo2=PR&Code2=35&Data=Count&SearchText=Kingston&SearchType=Begins&SearchPR=01&B1=All

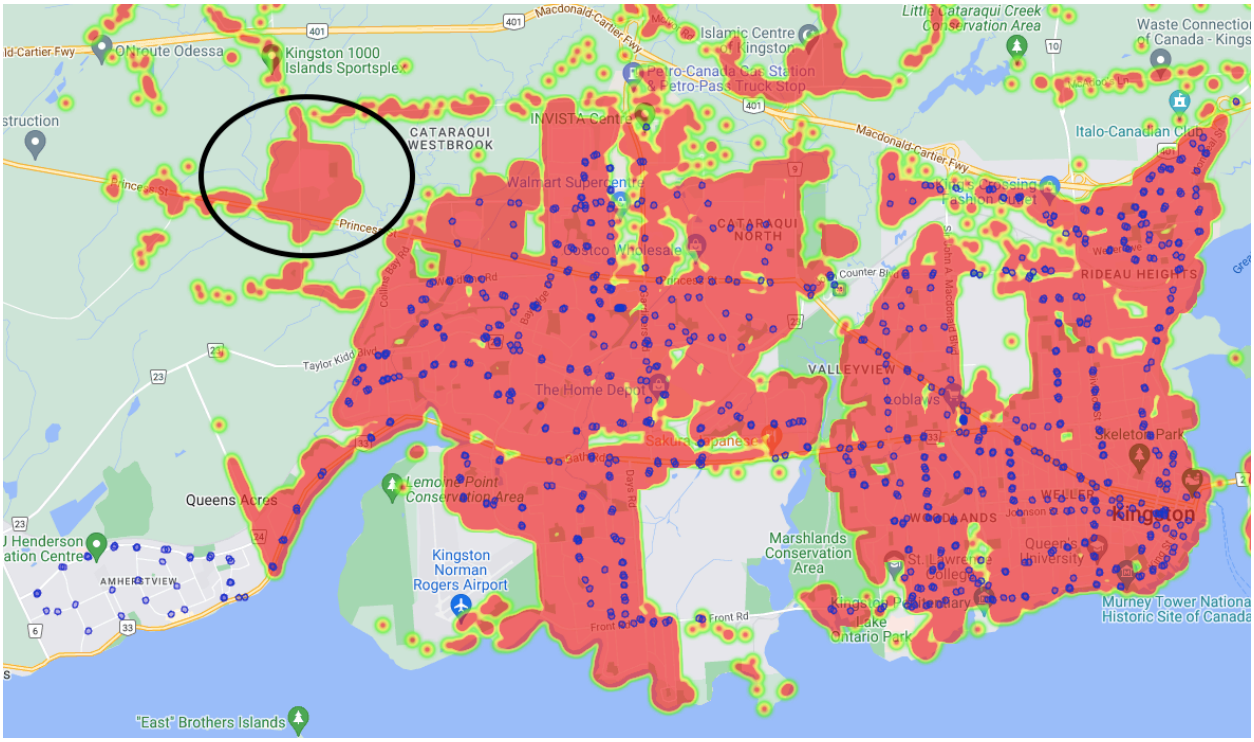

Figure 1: Bus stops (Blue Dots) on the heatmap of the Kingston resident addresses

In addition, an interesting fact I discovered in the 'Transit Data – October' dataset, after visualizing the distribution of the count of the bus using the route 17 against the time in a day, is that the bus on the route is essentially operating during night period to serve the students living in the Queen's Campus district, which further explains why the demand for this route is very low[5] and is nearly exclusively used by Queen's students. In this sense, the conclusion I proposed before may be invalid as this route is not aimed to generate economic benefits, but to provide cares for students who come home late for various reasons (such as studying late for exams or enjoying parties). However, it may still be desirable to discontinue the route 17 bus service during the summer time when the demand for the route is significantly diminished as majority of Queen's students are not living in Kingston at this time period.

[5] Only serving the people who are living in the Queen's University area and have returned home late

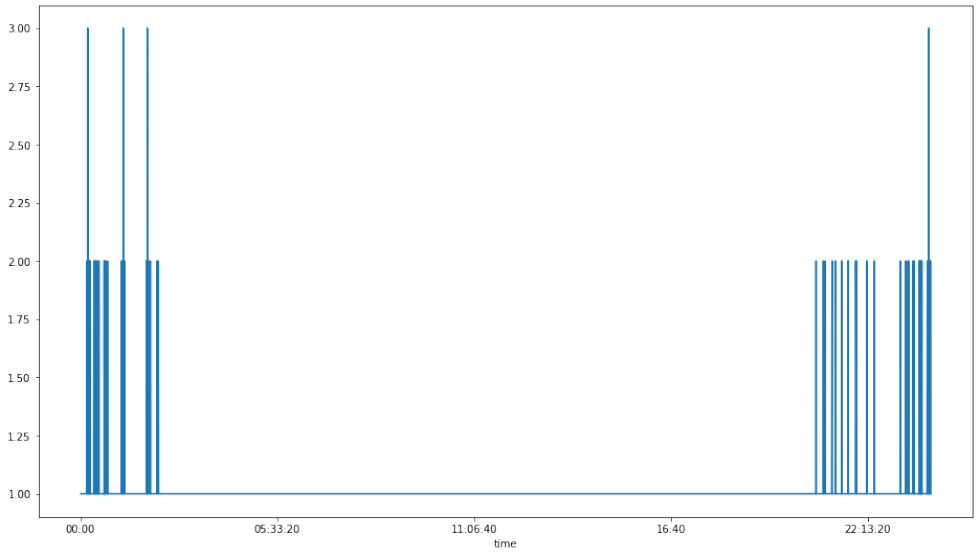

Figure 2: Bar chart of the count of the bus using the route 17 against the time in a day

Interesting in the data that can help planning the bus routes (Wenqi):

It is interesting to see that most of the cycling path is on the way where most of the bus route is covered. From question 3 we suppose that there is one bus route that starts from Turtle Crossing. However, there are many assumptions that we made, for example, people here work in downtown Kingston and want to take a bus. But noticed that that area is far away from downtown Kingston and that bus route is added from our opinion caused by the number of driveways. But that starting point can include working opportunities, as well as the people here, might like to ride a bike instead of taking a bus. And also, figure 3 used the dataset 'cycling-facilities.geojson' just represent parts of the possible paths for bikes and there might be another area that has similar problems in Kingston. So the bus route might not need that much.

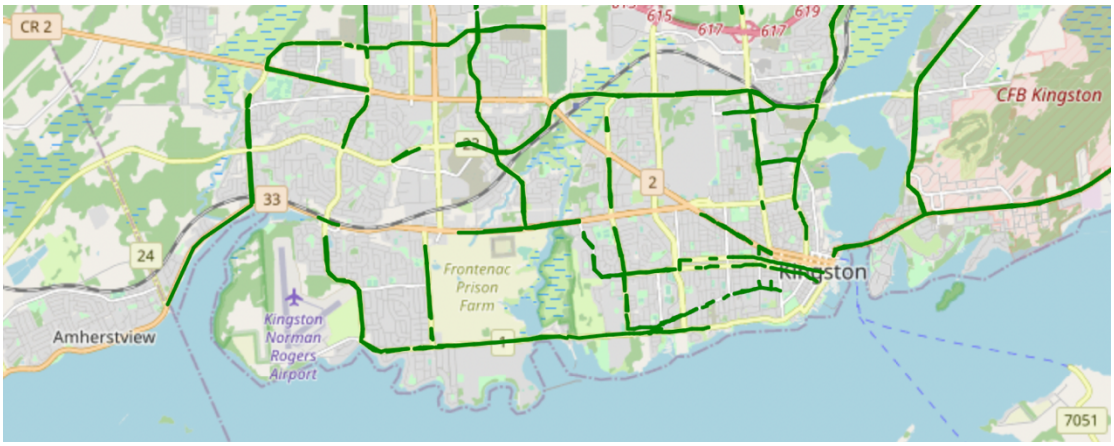

Figure 1: Bicycle Trace (Green Line)