CISC 451 - Assignment 1

Will Murray: 20054564

Emma Ritcey: 20007918

Rylen Sampson: 20051918

Preliminaries

To perform the analysis of the Kingston transit data our group decided to use Python, supplemented by the numerous data analytics libraries available for use. Python itself is freely available on the programming language's website. For some visualizations, jupyter notebooks was used. This is a platform that allows you to execute certain blocks of code at one time. This way, visualizations could be displayed one at a time in real time. When you download Python it comes with many of the packages we used pre-installed. For the libraries that do not come with the basic Python installation, one can use the built-in Pip installer to access those libraries. More regarding the libraries we used are discussed below.

Additional datasets included the geographic data for the City of Kingston bus transit routes and locations of driveways throughout the Kingston region. These can be downloaded from Kingston's open data library.

Libraries

- os: The os library allows you to change which directory you're accessing on your computer. This helps keep an organized project folder.
- pandas: Pandas is a fundamental data analytics library. It provides a *dataframe* data structure perfect for manipulating data for analysis.
- numpy: Similar to pandas, Numpy provides data structures and algorithms designed for data analytics specifically.
- plotly: Plotly is a data visualization library in Python with data analytics in mind.
- datetime: Datetime is a Python library, capable of handling data in various date formats with further functions to support *datetime* objects.
    - Mapbox: Mapbox is a supporting component of Plotly which allows for the plotting of data over high-resolution maps.
- folium: Folium is a package that allows easy visualization of data on an interactive leaflet map. Folium plugins must also be imported in order to visualize and manipulate heatmaps
- branca: a spin off package of folium that allows the user to alter non-map specific features of the figure, such as the height and width

Analytics Process

Every team or group goes about analyzing data in a different way, and even each team member may approach a problem differently. Our situation was no different and therefore we provide a general overview of the steps we took as a group and finer details are explained in the paragraphs relevant to each figure.
Each script of code began with the dataset being loaded in using **pandas**. This gives us an easy to manipulate data structure making the task of visualizing and manipulating the data much easier. As mentioned in the assignment description, there are numerous recordings where the Latitude and Longitude features are 0. There are also recordings with latitude and longitude

values that are not accurate and are outside of the region where Kingston's bus system runs, as well as datasets that contain NaNs. Therefore, we dropped recordings that fell under any of these categories before visualizing the data in a geographic manner as they are considered erroneous data.

Question 1)

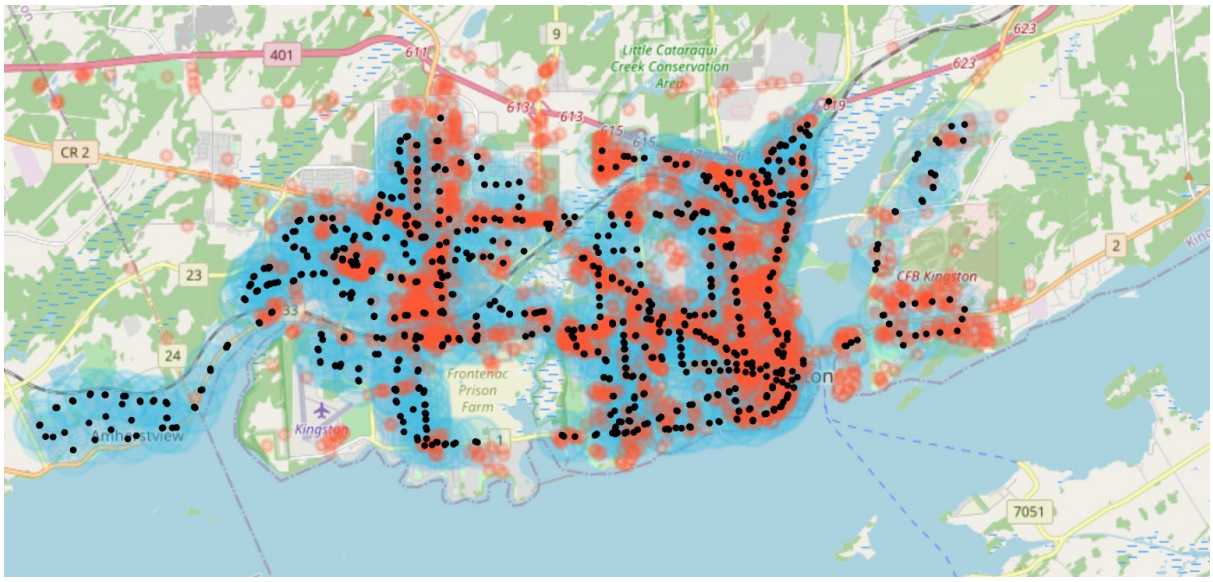

Figure 1: Accessibility of the bus transit system through walkability and carpooling locations

Figure 1 shows the accessibility of the transit system. The blue regions are constructed with blue circles with radii of 500m and whose centers are each at a bus stop, shown by the black circles. This shows the area of Kingston that can reach a bus stop by walking a distance less than 500m. The darker blue regions mean there are more stops within this distance, and therefore even more accessibility to various bus routes and different parts of the city. The red circles are each located at a parking area and have a radius of 100m. We assumed that if someone has to drive to a bus stop, they will want a short walk from their car to the bus stop. Any bus stop that is not located in, or close to, any of the red circles is deemed to be inaccessible by car.

Findings

- The blue covers the majority of Kingston, and therefore the bus system is relatively accessible to those who live directly in Kingston.
- For the most part, there is adequate parking accessibility, however, the west end has very few parking areas, and Amherstview has exactly zero.
- Downtown Kingston has the most parking

Question 1 Summary

For those who live in town, the bus system is relatively accessible by foot, however, for those that live on the outside of town, and can't reach a bus stop by a reasonably length walk, there needs to be another option, such as driving to a bus stop and parking. Because of the lack of parking on the far west end of Kingston, people from this area, Amherstview, or outside of the city altogether have to drive further into town to reach the bus, or drive directly to their destination. This defeats the purpose of a transit system and means that more people are using their cars to get into busy parts of the city. Downtown Kingston has the most parking, which is needed as there are always people visiting for the local restaurants, concerts, and waterfront view, and parking always seems to be full. On the other hand, this likely entices people to drive downtown and use up parking spaces instead of taking advantage of the bus system. Getting more people to use the transit system to go downtown would decrease traffic in that area and solve the problem of there never being enough parking. For people coming into the city from the east or north, there is ample parking to use to access a bus going into town.

Question 2)

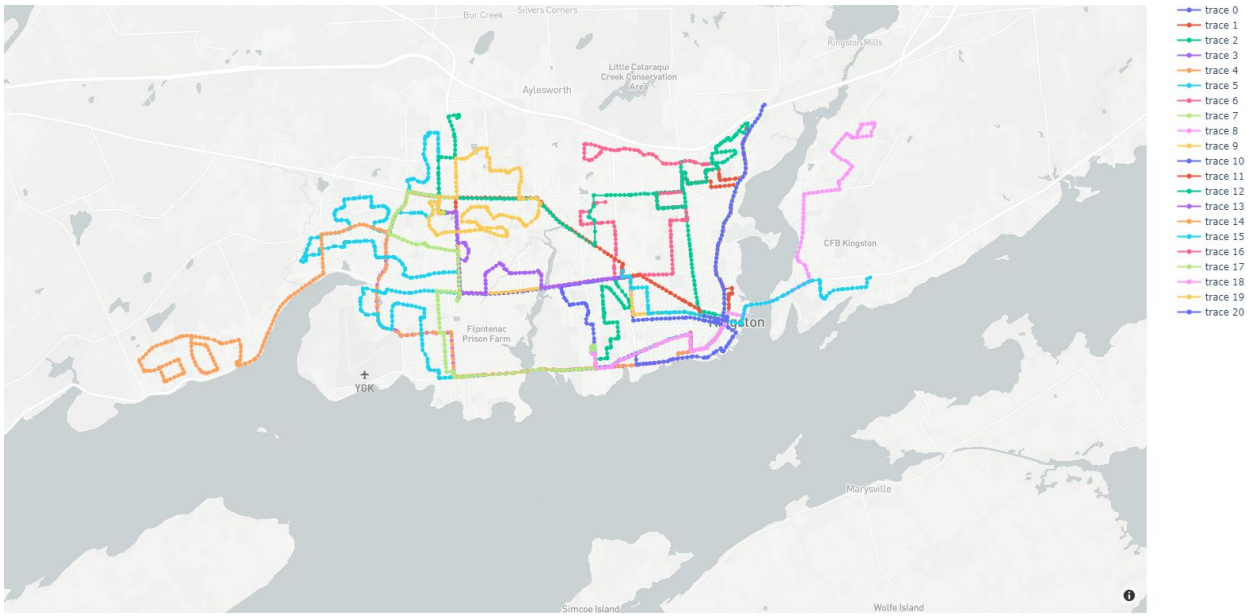

Figure 2: Plots the bus routes on top of a map of Kingston.

Shown in *Figure 2* is a map of the bus routes which makeup the bus transit network for Kingston. To obtain the Kingston background map we used **Plotly** and **Mapbox**. **Mapbox** requires you to make a free public key for their services which is then included in the **Plotly** code. The other steps involved with generating this plot were downloading and processing the route data and the locations of stops. Once again, a **pandas** dataframe was used to store the data prior to plotting.

Findings

- Evidently, the west north and east extremities lack accessible bus transit the most but the argument can be made that less citizens out in those areas require bus transit as they may not necessarily have jobs in Kingston and their livelihood is less attached to Kingston as a whole.
- Kingston's bus transit network can nearly be partitioned into two main portions. The first being the west side of the city and the second being the east side where these two portions are split via the north-south running river (Right of Frontenac Prison Farm) in the middle of the city.
- Despite whether this is the most optimal layout of routes available, the City of Kingston does a good job of providing access to all immediate surrounding neighbourhoods.
- Lastly, it can be seen that any route only requires one or two route changes to get anywhere else via the bus network. Given that you're typically on a bus (in Kingston) for approximately fifteen minutes, that means in under an hour you are able to get within a close proximity of wherever you need to go.

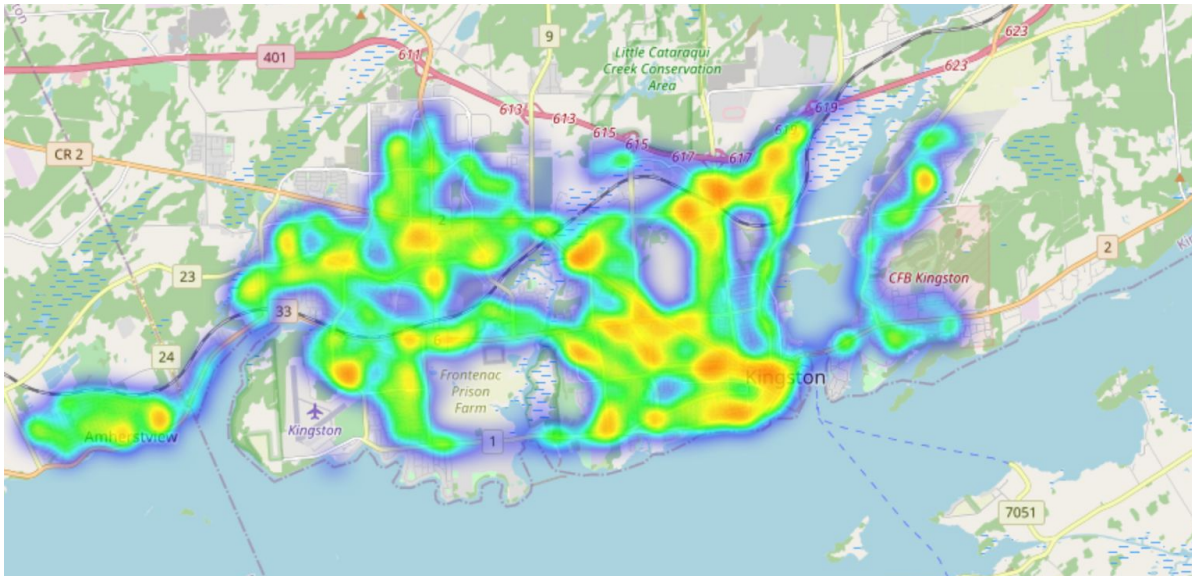

Figure 3: The density of boarding locations for the time period of the dataset.

Figure 3 shows a heat map of all of the bus passengers boarding locations in October 2017, from the transit_data dataset. The red regions indicate regions where more passengers are boarding buses, while blue and green indicate fewer passengers boarding.

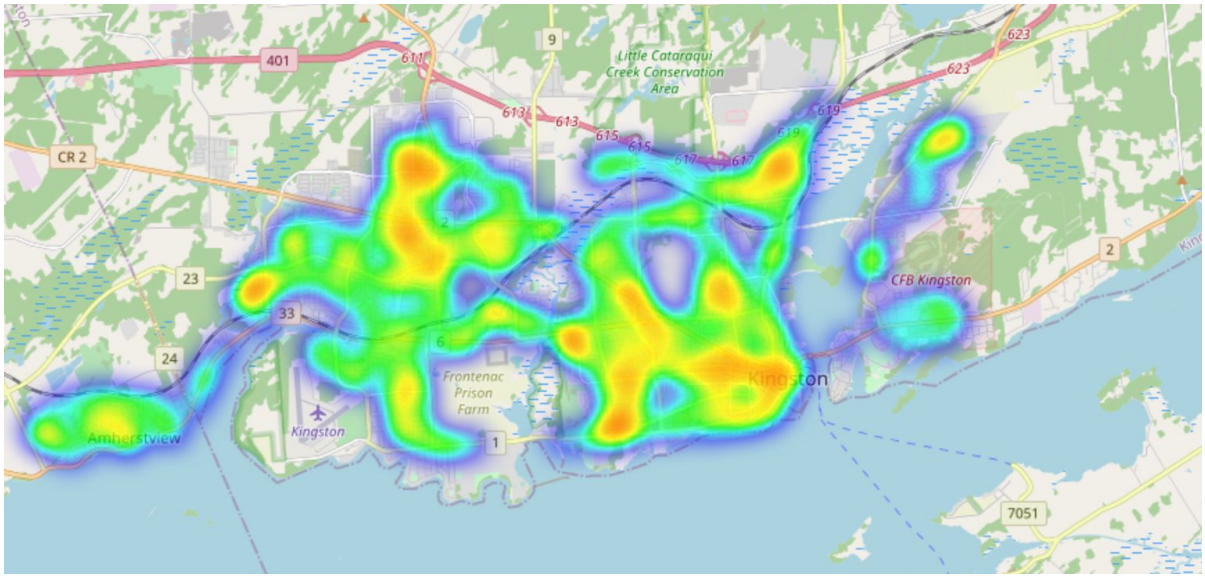

Figure 4: The density of bus stops for the City of Kingston bus transit network.

Figure 4 shows the bus stops as a heatmap, which displays the areas that contain the most bus stops. Ideally, one would want the red regions of this map to correspond to the red regions of *Figure 3*, as this would mean that the bus stop locations have been chosen in a way that accommodates the locations with increased passenger traffic.

Findings

- Downtown and Princess Street is one of the busiest locations for boarding buses, which is to be expected, along with northern Kingston by Division street and highway 401.
- Comparing *Figures 3 & 4*, they mostly match each other, however, the top left red region in *Figure 4* along Gardiners Road does not appear in *Figure 3*. This means that there are likely too many bus stops in this area of town given the number of people that boarded at these bus stops.

Question 2 Summary

As noted in *Figure 2*, Kingston's bus transit system does a great job of covering the entire city and to a degree the extremes. Although there is overlap between bus routes, this is to be expected with any transit system and in fact is a positive rather than a negative. Removing a route is easier said than done as it has a drastic effect on the remaining routes, influencing the number of people on those routes at certain segments and how long those on the bus must stay on to get their destination.

If we were to recommend a route to be removed it would have to be route 7. This is because it is just about completely covered by other routes such as route 20, and route 1 except these routes then go further beyond what route 7 covers. In *Figure 2*, route 7 is in fact not visible (located in the downtown Kingston area). Our recommendation comes with the condition that a deeper consideration of how removing this route would affect the above mentioned routes that

are in close proximity. Another possibility would be to reduce the number of stops that the routes going through the Gardiners road region have, or altering the course of one completely so that it doesn't go through that area anymore. This is likely an easy decision because the Gardiners Road region seems to have too many bus stops given the number of people that board buses in that area, however, before making this decision it would need to be decided if this would negatively impact the flow of peoples' travel or the bus system in general.

Question 3

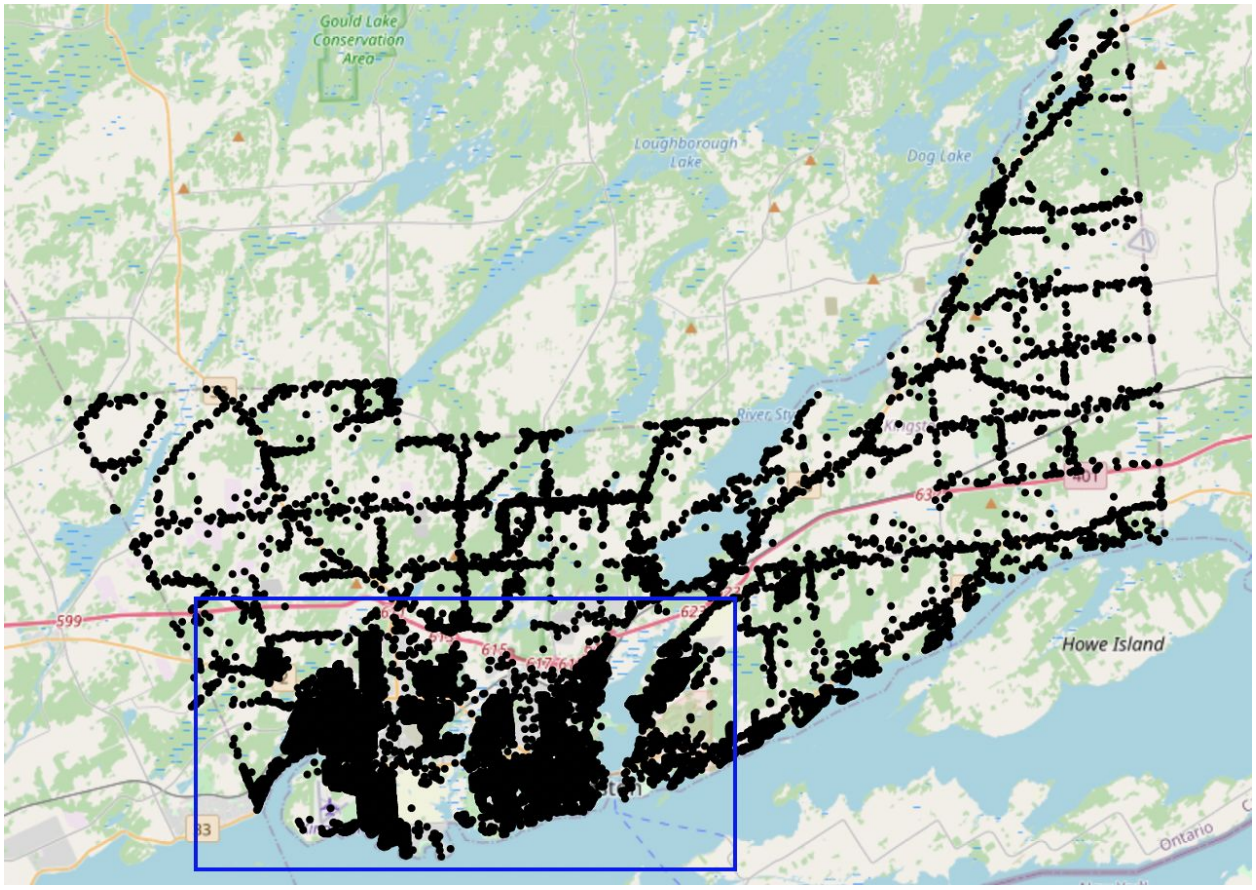

Figure 5: Location of driveways throughout Kingston.

Figure 5 shows all of the recorded driveways in Kingston. This dataset included data for the entire city of Kingston, not just the inner region we have been showing in previous plots. For reference, the blue rectangle is the region that has previously been shown and analyzed.

Findings

- The majority of people who live within this rectangle would be able to reach a bus stop that is within a relatively short walk, although perhaps a little more than the 500m distance we used for *Figure 1*.

- Those in the top left corner of the rectangle cannot access a bus route by foot, which can be seen in *Figure A*
- For everyone outside of this rectangle, new buses would need to be added to accommodate them.

Question 3 Summary

If every person in Kingston worked downtown, people who live in the top left corner of the rectangle would need a new bus route or a current bus route to be extended so that they can access the system, as they currently cannot access a bus by walking. For those who live outside of the rectangle, because of the distance, there will likely need to be at least 5 buses added so that they don't take an unreasonable amount of time to reach downtown. There are 45,295 driveways in Kingston. Assuming that there are two cars in each driveway and everyone works in downtown Kingston and travels separately, then there would be 90,590 cars that would be travelling to downtown. Assuming that Canadians, on average, use approximately 2000 liters of gas for their cars per year, then they use a little over 5 litres per day. If all of Kingston completely stops using their cars, then a total of 90,590*5 = 452,950 litres of gas per day would be saved.

Question 4)

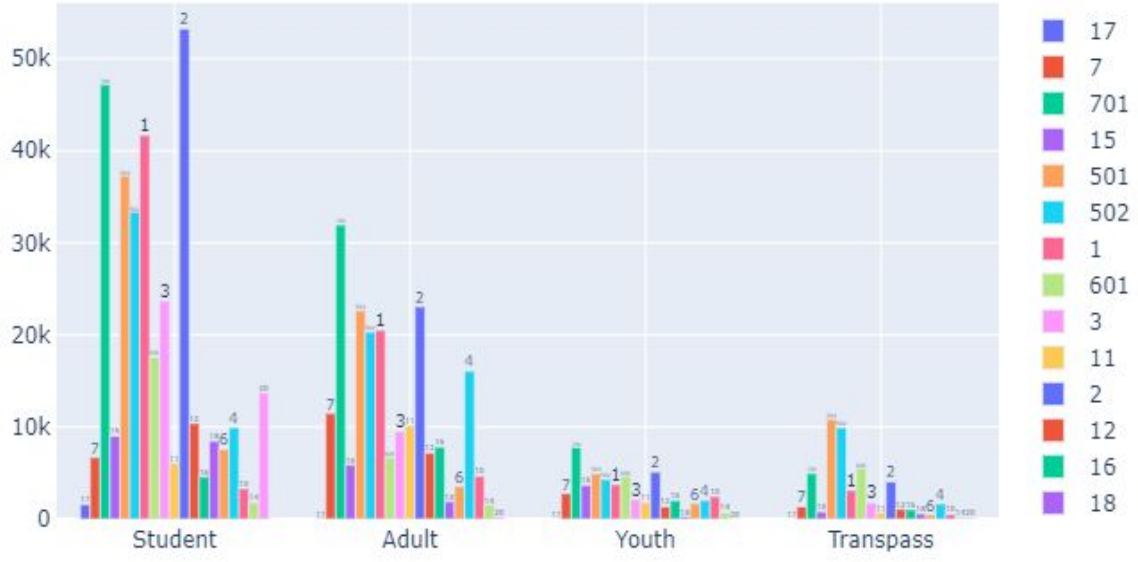

Figure 6: Plots how many of each rider label takes each route.

Plotted in *Figure 6* is a breakdown of the types of people utilizing bus services and which routes they take. To obtain this plot we first had to handle the categorical labels in the *Class* column of the Kingston transit dataset. This column contained repeated labels (I.e. "Adult" vs. "ADULT") which would have resulted in more labels than necessary. Furthermore, there existed labels

which could be easily grouped such as “Queens” and “St Lawrence”. By eliminating the duplicate labels and grouping ones describing similar demographics we were able to achieve four overarching rider class labels; “Student”, “Adult”, “Youth” and “Transpass”. The “Transpass” label refers to people who work at organizations in and around Kingston that supply bus passes to their employees.
For the purpose of clarity, we dropped routes 8 and 13 as they are not regularly occurring bus routes and likely would tailor to one of the four class labels much more than the others.

Findings

- Students (or those who have a school affiliated bus pass) make up the overwhelming majority of the Kingston and surrounding areas bus usage.
- For all rider labels, route 17 is very underused.
- Despite adults using bus services less than students, adults use route 4 to a higher degree
- Route 2, across all rider types, is one of the most used route

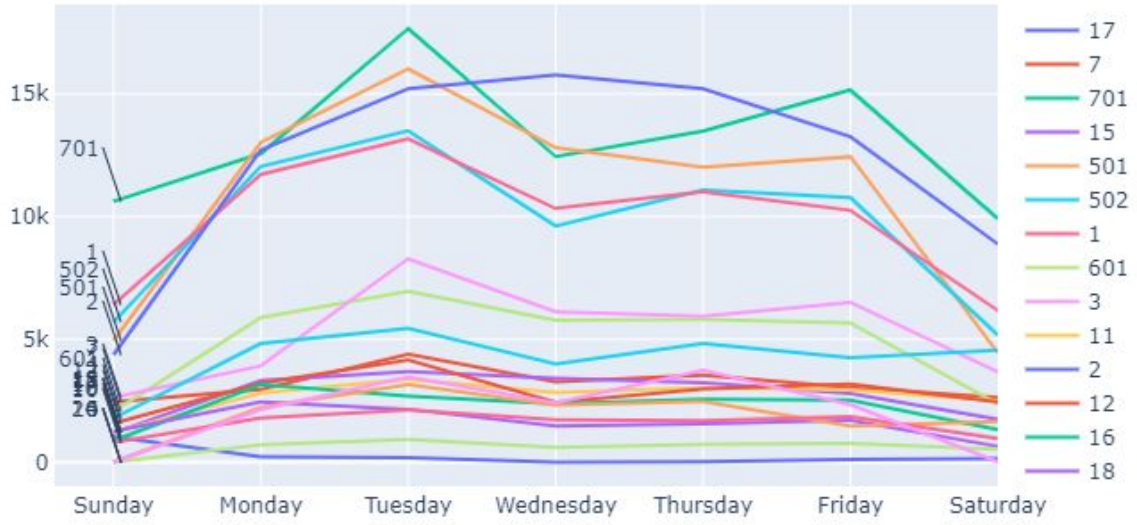

Figure 7: Plots the usage of each route aggregated over the entire month of October given in the dataset.

In this figure, (*Figure 7*) we’ve plotted the number of times each route was taken on a given week day. I.e. Route 701 on a Monday is the busiest. As the data does not readily give which day of the week a recording is for, we had to figure that out ourselves. The **datetime** library is a tremendous resource when it comes to handling date features in data science. The first step in processing these dates was to parse the string containing them. Through the **datetime** library we could then create *datetime* objects. Lastly we were able to access the *weekday()* property of the formerly created *datetime* objects to get the day of the week day.
As done in the previous figure, we dropped routes 8 and 13 as they are not regularly occurring bus routes.

Findings

- Routes 1, 2, 3, 501, 502 and 701 tend to fluctuate the most drastically in usage throughout the week with peak days appearing on Tuesday and Friday. Route 2 is an exception as its usage for a given weekday appears to follow a parabolic arch.
- The remainder of the bus routes are used on a more consistent basis.
- As expected, weekends have the lowest usage of bus services in Kingston.

Question 4 Summary

Using *Figure 6 & 7*, rather than address bus routes we are able to address the scheduling of bus routes. Using *Figure 6* it is clear that routes frequented by all rider classes could have a greater number of buses in circulation whereas routes unique to a subset of rider classes could be explored in further detail. For example route 11 is largely used by adults, figuring out when this route is in peak demand for adults could help optimize the number of buses needed at different points during the day.
Continuing on with this train of thought. We may want to suggest to the City of Kingston they look into optimizing the number of buses on a given day. Clearly Tuesdays and Fridays require more buses to handle the increased usage of bus transit compared to a Sunday where usage heavily drops off.

Rather than looking at the physical layout of routes but rather how and when they are being used and by who gives additional insight that would have otherwise remained hidden. This gives actionable knowledge that can be used for addressing issues around scheduling, arrival frequency of buses driving a particular route and more. This route information is as important as the physical aspect as hand-in-hand they contribute to a successful transit system.