CISC 839

Topics in Data Analytics

Competition 1

Data understanding

Sahil Shethiya

20186685

19ss55@queensu.ca

School of Computing

Kamal Andani

20188032

19kaa3@queenu.ca

Department of Electrical and Computer Engineering

Software Used:

Here we use Python 3.7 as the programming language, which can be downloaded from https://www.python.org/downloads/

We use Jupyter Notebook for writing code and visualization of results. It can be installed with following command.

- pip install notebook

Software Packages Used:

1. **Folium**
   Folium is a python library that can be used for visualizing geospatial data. It is a python wrapper for *leaf.js* which is opensource JavaScript library to plot maps. So here we use *Folium* to plot maps.
   It can be installed easily with the following command:
   - pip install folium
2. **Pandas**
   Pandas is a python library which is built on top of NumPy. It is designed to make data preprocessing in mind. It is a very popular library to do data analysis like data cleaning. It aims to be the fundamental high-level building block for doing practical, real-world data analysis in Python.
   It can be installed using following command:
   - pip install pandas

3. **Geopandas**
   GeoPandas is an open-source project to make working with geospatial data in python easier. GeoPandas extends the datatypes used by pandas to allow spatial operations on geometric types.
   It can be installed by following command
   - pip install geopandas

Dataset used

1. **Transit Bus Stops(GTFS-Static):**

Where to download? : https://opendatakingston.cityofkingston.ca/explore/dataset/transit-gtfs-stops/information/

Description: This dataset provides all bus stops which are currently being used by the Kingston transit. It contains the Unique ID, Name, Coordinates(Longitude & Latitude) of each bus stop.

2. **Transit Bus Routes(GTFS-Static):**

Where to download? : https://opendatakingston.cityofkingston.ca/explore/dataset/transit-gtfs-routes/information/

Description: This dataset contains all the routes throughout the city of Kingston. It contains Route ID, Bus Color, Shape which plots the whole route on Map of each bus.

3. **Cycling Facilities:**

Where to download? : https://opendatakingston.cityofkingston.ca/explore/dataset/cycling-facilities/information/

Description: This dataset provides information about bike lanes and paved shoulders in the City of Kingston. It contains cycle status which defines if a road is just a bike lane, bike lane shared with the car lane, Construction status, GeoJSON which plots the bike lane on a map.

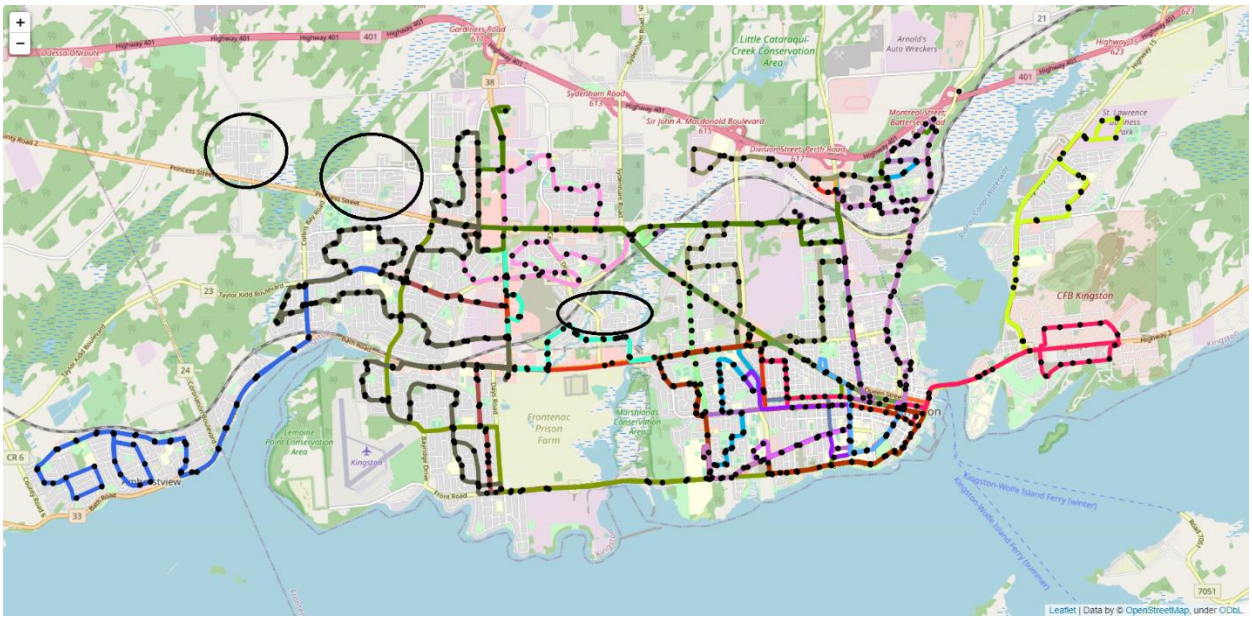
4. **Parking areas:**

Where to download? : https://opendatakingston.cityofkingston.ca/explore/dataset/parking-areas/information/

Description: This dataset contains information about the private and public parking areas in Kingston.

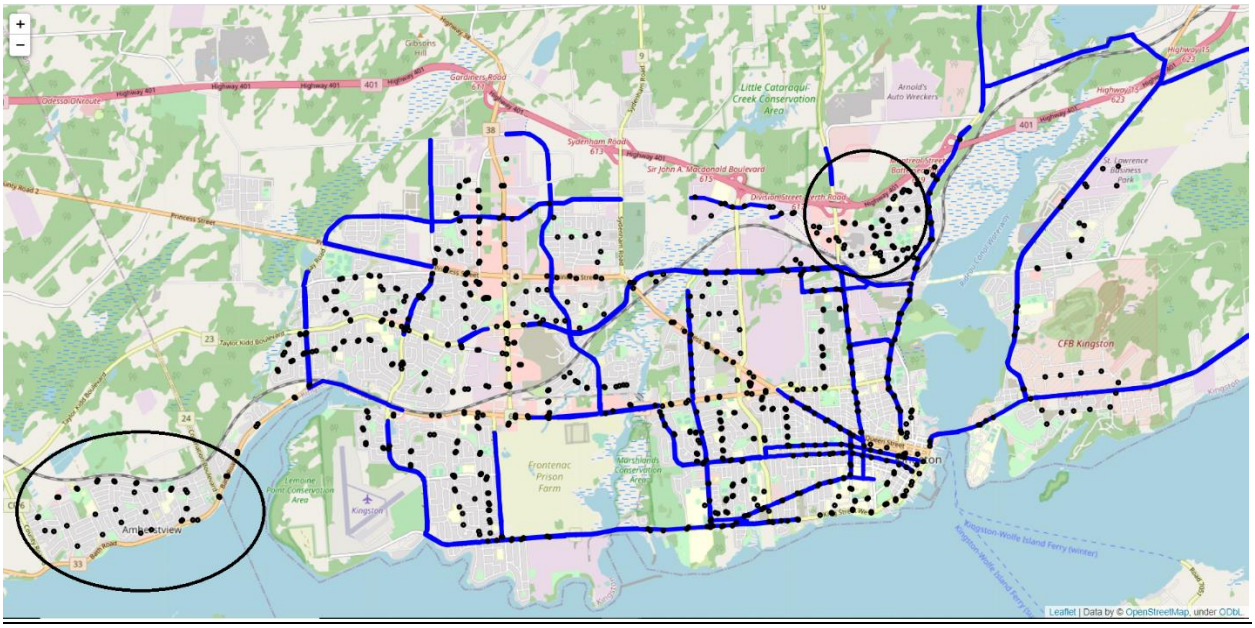
Task 1 : Is the bus service accessible to everyone in Kingston (walk, bike, park & ride to a bus stop)?

Walk:



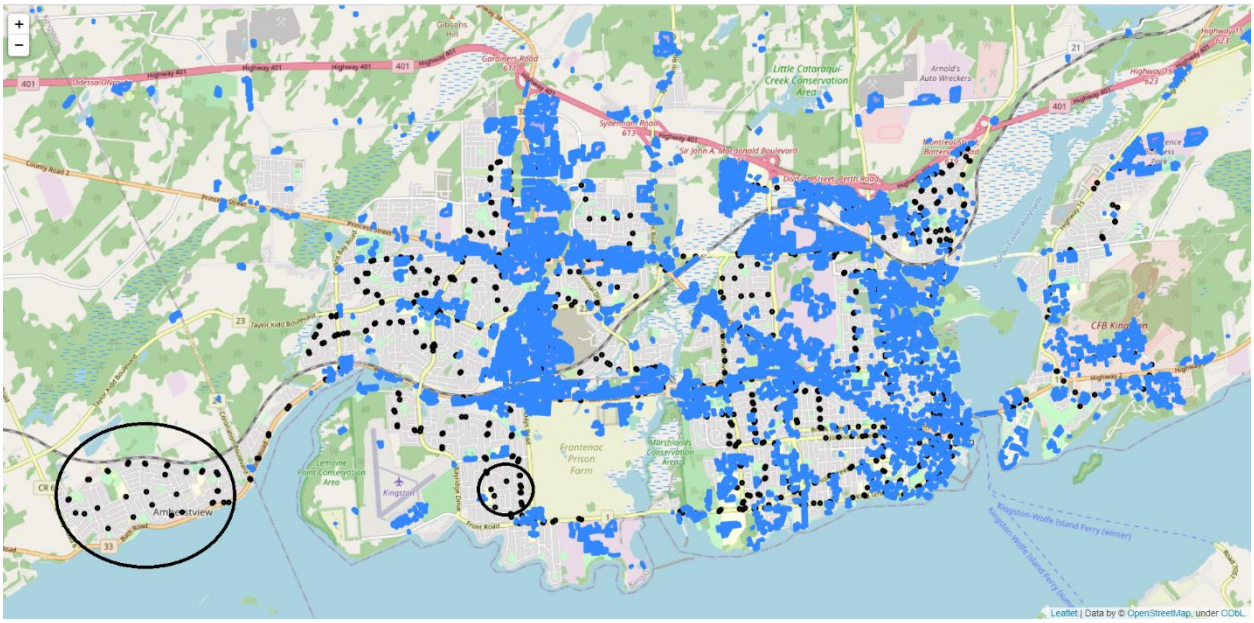
As we can see from the map, almost all the areas are accessible to the bus stop apart from the area which is circled on the map which is not within walking distance from any bus stop. Here, different coloured lines represent different routes in the City of Kingston and small black dots indicates bus stops.

Bike:



Visualizing from the map above, we can say that almost all the bus stops are accessible by bike. However, the bus stops at the Amherstview side and division street near 401 are not accessible by bike. Blue lines on the map indicate bike lanes or paved shoulders and black dots indicate bus stops.

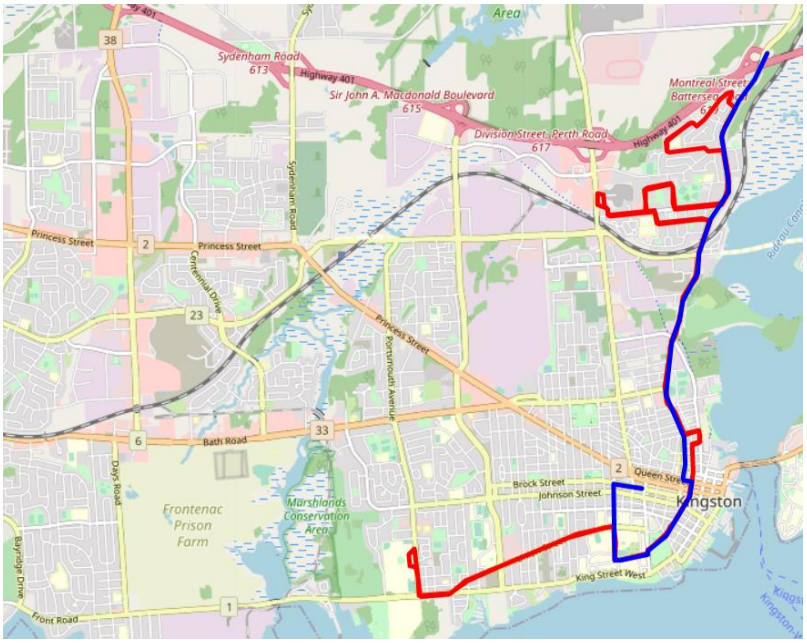
Park:



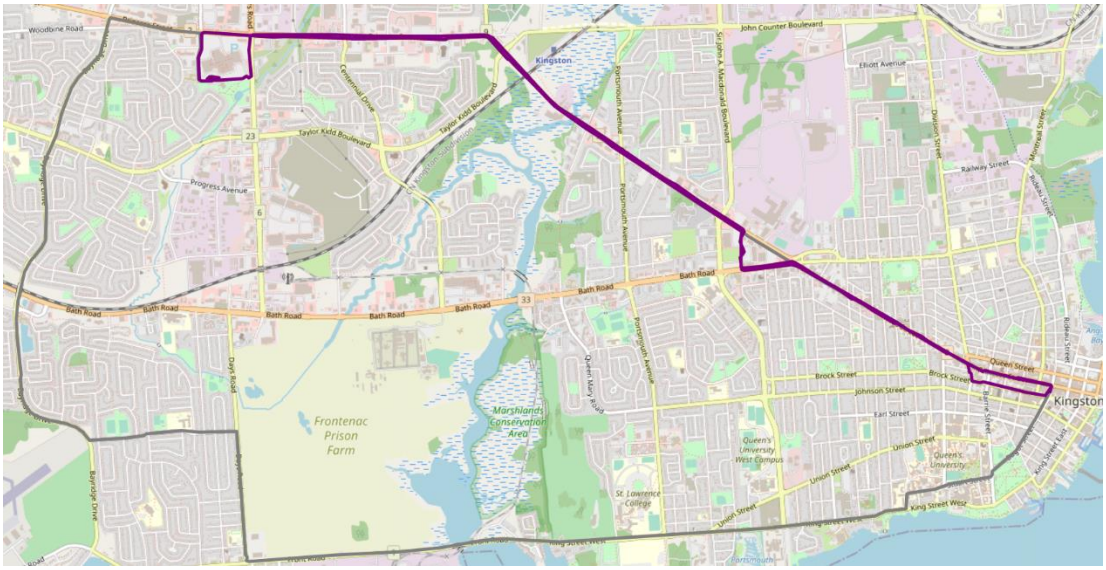
Almost all the bus stops apart from the Amherstview side and near the airport are accessible to parking lots. Parking lots are filled with blue color in the map and bus stops are indicated by black dots.

Task 2 : Are there some redundant bus routes that can be cancelled?

By seeing at the bus routes map, we discovered some bus routes that are redundant.



In the above map, blue line indicates route 801/802 and red line indicates route 1. We can see that most of the route for 801/802 is covered by route 1. So 801/802 can be cancelled.



In the above map, Route 501 is indicated by gray coloured line, and route 4 is indicated by purple line. By looking at this, we can say that route 4 is a part of route 501 itself except from few bus stops. So we can say that, route 4 is redundant. Hence, route 4 can be cancelled by modifying route 501 slightly.

Task 3: Assume that each driveway has two cars and that all people work in downtown Kingston, how many new bus routes are needed to make people leave their cars at home? How much would that reduce the gas emissions (make your own reasonable assumptions)?

According to Driveway data provided by City of Kingston (https://opendatakingston.cityofkingston.ca/explore/dataset/driveways/information/?disjunctive.material) , the city has around 45000 driveways. If we assume each driveway has 2 cars, then there might be approx. 90,000 cars in the city. If we assume every person works in downtown, then we can assume average travelling distance to and from home in the city would be around 11 kms for each person. Now, we can say that total average travelling distance would be 11*90000 = 9,90,000 kms. If we assume average car gives mileage of 12 kmpl, then approx. 82.5k Liters of fuel will be needed. As per (https://www.nrcan.gc.ca/sites/www.nrcan.gc.ca/files/oee/pdf/transportation/fuel-efficient-technologies/autosmart_factsheet_6_e.pdf) , 1 L fuel approximately emits 2.3 kg of CO2. That means 82.5k Liters of fuel will emit approx. **1,90,000 kg** of CO2.

Due to COVID-19 each bus has capacity of max 20 passengers. And there are 21 routes . If we assume every person travel during peak hours, we'll need around 90000/20 =4500 buses during peak hours. Now each bus travels approx. 11 kms. and it gives mileage of approx. 1.9 kmpl. So each bus uses approx. 5.8 Liters of fuel. So 4500 buses will use 26100 Liters of Fuel. Same amount of fuel will be needed for another peak time at evening. So total 26100*2=52,200 Liters of fuel will be required. Now during non-peak hours, suppose we will require half of the buses than in peak hours (That will be 2250 buses) so they will require approx. 13,000 Liters of Fuel. So total fuel requirement will be 52,200+13,000 = 65,200 per day. So it will emit approx. **1,50,000 kg** of CO2.

So As we can see, the usage of bus will reduce the emission of CO2 by **40,000 kg** approx. each day.

Task 4: Is there anything else you find interesting in the data that can help planning the bus routes?

Here we use Transit data of October,2017 which was provided on OnQ. This data contains around 7 Lacs of entries.

First of all, we sort the data according to 'Route' column. We, then, extract date and day of the week from 'Date' column. We then extract hour and minute from 'Time' column. We then remove original 'Date' ,'Time' and 'Bus' column from data as we don't require them.

```
date=df['Date']
```

```
weekday=[ i.strftime("%w") for i in date]
```

```
monthday=[i.strftime("%d") for i in date]
```

```
time=df['Time']
```

```
hour = [i.strftime("%H") for i in time]
```

```
minute=[i.strftime("%M") for i in time]
```

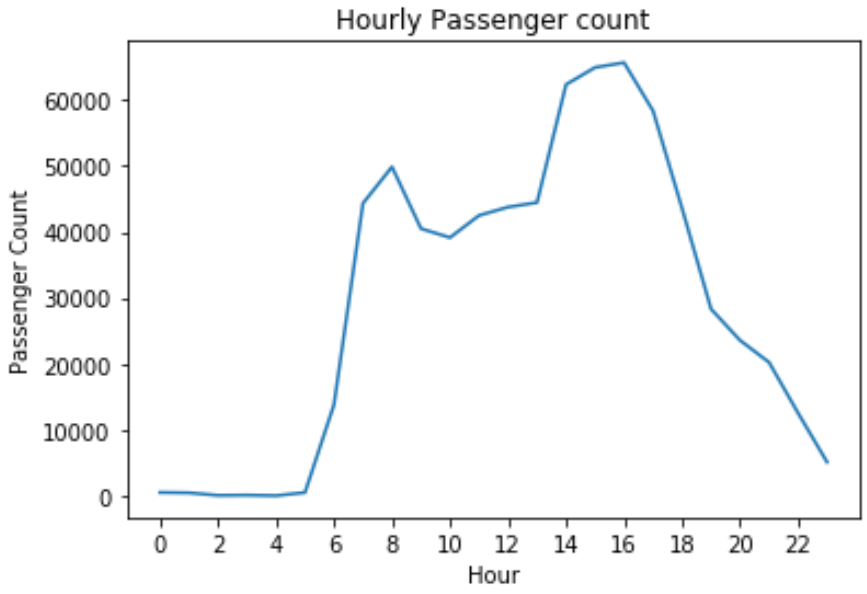
```
newdf=df.copy()
```

```
newdf['weekday']=weekday
newdf['monthday']=monthday
newdf['hour']=hour
newdf['minute']=minute
```

```
newdf.drop(['Bus','Date','Time'],axis=1, inplace=True)
```

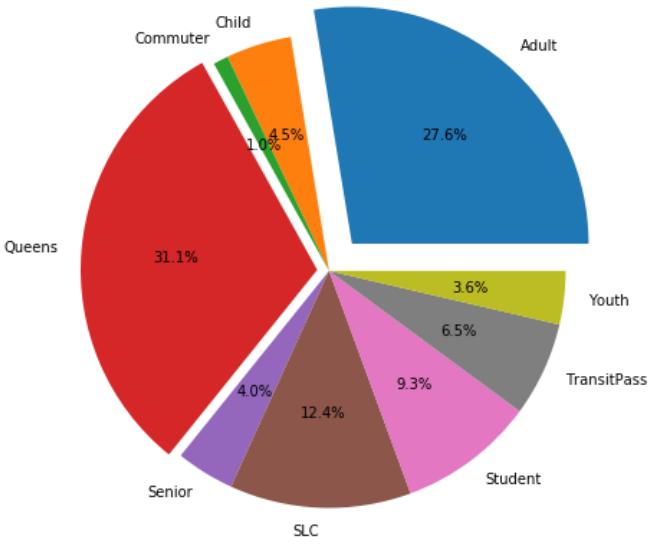
We group the data by 'weekday', ' hour', 'class' and ' route' and do the analysis.

I. Passenger count Per hour



As we can see in the above line chart, people travel more during peak hours (7am to 9am and 2pm to 5 pm) . So we can say that bus frequencies can be increased during these hours.

II. Class of Passengers

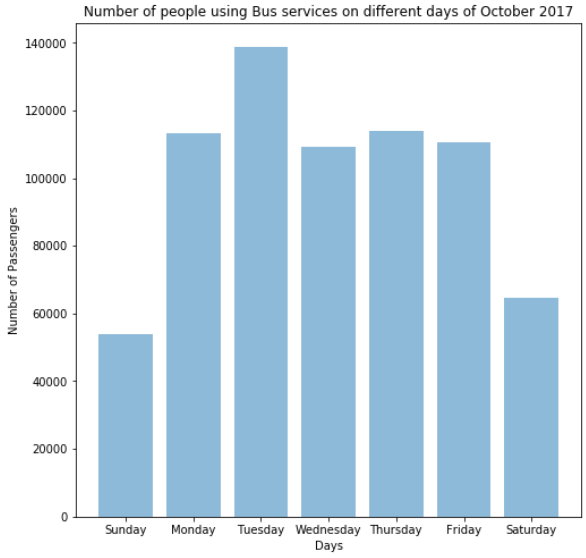


The above pie chart shows distribution of types of people travelled during October,2017.

We can see that majority of the users are students of Queen's University with 31.1% of total transit users. After that, 27.6% users used Adult Pass to travel to their destination. We can note that only 1% of the total users are commuters.

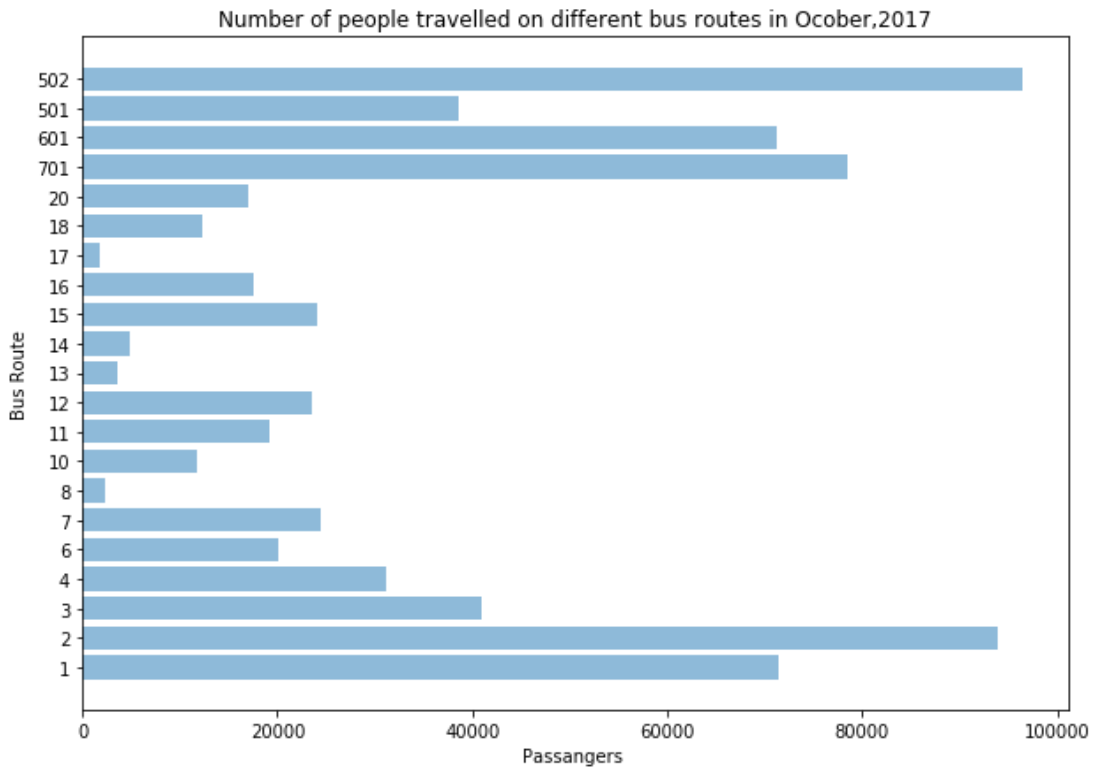
From this, we can conclude that more routes can be added covering Queen's University.

III. Number of passengers on different days of week



From the above bar chart, we can see that less people travel through transit on Weekends. So, frequencies of buses can be reduced on weekends, and can be increased on weekdays as more people use bus services on weekdays.

IV. Usage of different bus routes



The above bar chart shows that route 2 and 502 are the most popular route among the users of Kingston transit. We can also see that very few people travelled through route 17,13,14 and 8 compared to other routes. From this, we can say that frequencies of route 8,13,14,17 can be reduced and frequencies of route 502,601,2,701 can be increased if needed, or new routes similar to the later ones can be introduced in future.