Data Understanding Competition

Andrew Grebenisan, Tayyab Ahmad

October 2019

1 Introduction

Towards the end of July this year, an article was published by CBC titled "Cities turning to Kingston for transit advice". Municipalities across Canada were turning to Kingston due to Kingston transit having the fastest increase in ridership in the country. According to the article, in 2014, Kingston transit started more frequent service and express stops for key destinations, such as the downtown transfer point. Additionally, they advised cities such as Guelph and Halifax that they should not be concerned about trying to maximize reliability and availability for everyone in the city, but to concentrate on the needs of the masses. Their commitment to exceptional service increased their ridership numbers from four million in 2014 to more than six million in 2017, an increase of over 50% in just three years.

Although Kingston appears to have made significant improvements this decade, the role of this competition is to determine whether there are any pitfalls with the service. With our discoveries, we hope to make pragmatic suggestions to Kingston transportation on how to further improve their service to the community.

2 Data

Kingston generously gave us their bus route data for the month of October in 2017. The data is comprised of 8 attributes, and contains 704539 instances of people taking the bus. 55753 records have geolocations of 0 for both latitude and longitude, which means that the GPS system of the bus was not working for those records. Interestingly, the only day that the GPS was being problematic was on October 1^{st}, so it could potentially be useful to ignore the first day of records. Additionally, there were some locations in the data set that were not valid Below is a screenshot of the first five records of data:

	Date	Time	Class	Operation	Bus	Route	Latitude	Longitude
0	2017-10-01	1900-01-01 01:43:03	QUEENS	Exact Fare	620	17	44.227860	-76.496938
1	2017-10-01	1900-01-01 01:43:08	QUEENS	Exact Fare	620	17	44.227860	-76.496938
2	2017-10-01	1900-01-01 01:58:13	ADULT	Pass (Multi-ride card)	620	17	44.232035	-76.491397
3	2017-10-01	1900-01-01 01:58:25	QUEENS	Exact Fare	620	17	44.232035	-76.491397
4	2017-10-01	1900-01-01 01:58:27	QUEENS	Exact Fare	620	17	44.232035	-76.491397

Figure 1: First five rows of data in "Transit Data - October.xlsx" dataset

3 Questions

In this section, we would like to pose some questions we hope to answer using the dataset. Answering these questions should allow us to provide reasonable advice to Kingston transportation. The questions are

1. How accessible are the bus stops?

2. How many new bus routes are needed to make people leave their cars at home?

3. How redundant are the bus routes?

4. Are there any days of the week in which ridership is lower and more gas is being used than normal? Should the schedule be reconsidered on those days?

4 Is the Kingston transit system accessible to all residents of the city?

The goal here was to find if anyone in the city is viably able to take the bus, either by walking, biking, or driving and parking at the nearest bus stop. We decided that the definition of viability is as follows. The nearest bus stop to any address is:

1. No more than a 5-7 minute walk (500 meters)

2. No more than a 10 minute bike ride (2.5 kilometers)

3. No more than a 10 minute drive to a parking area that is within 500 meters of a bus stop.

After defining accessibility to a bus stop, we had to decide how to find out if the transit system is accessible to all. Using a Google Maps API, we plotted every bus stop and all of the addresses in the city (using the "civic-addresses" data set supplied by Open Data Kingston). We also obtained the "parking-areas" data set, which represents public and private areas where one can park their car. With this information, we can analyze the scatter plot of addresses, stops, and parking areas to decide whether anyone can either walk, bike, or ride and park to the nearest bus stop.

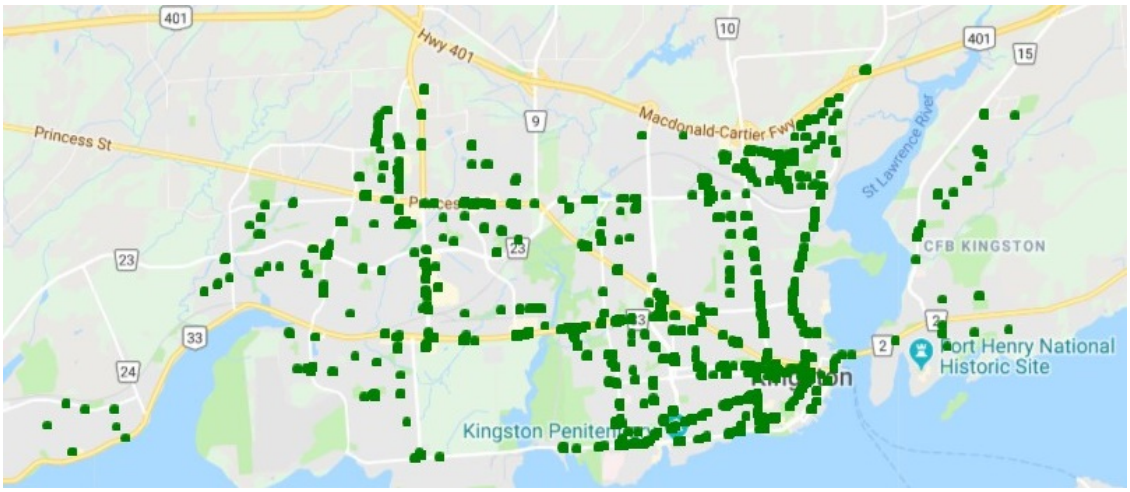

Figure 2: Bus stops in Kingston

We were able to overlay civic addresses onto the plot of bus stops. Since plotting individual addresses using the Google Maps API is extremely computationally expensive, we divided up addresses by electoral district and plotted them in groups. With this we can directly observe and analyze for whom it is feasible to walk, bike, or drive to a bus stop.

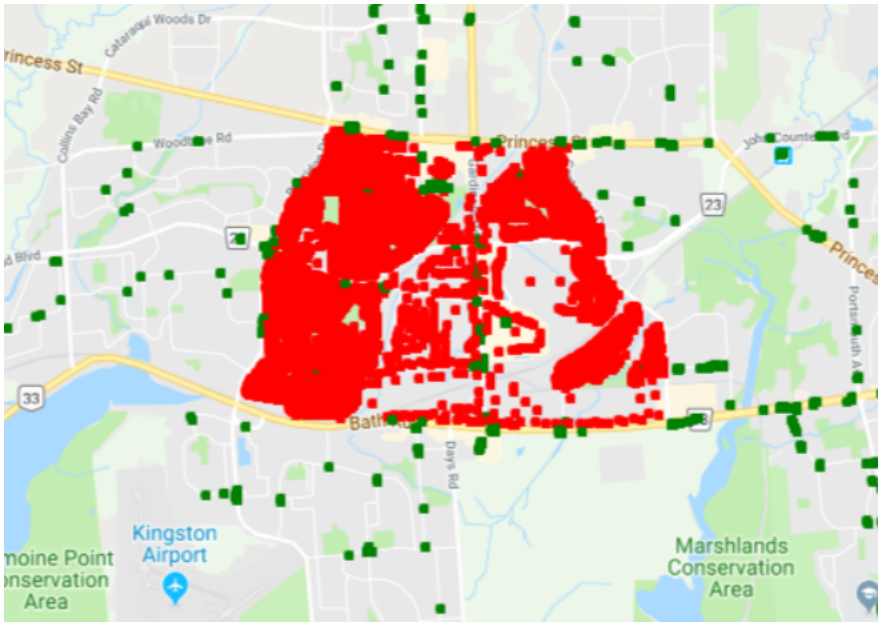

Figure 3: Civic addresses in the Trillium electoral district plotted with the bus stops.

In this case, for the Trillium district, the transit system is accessible to everyone. There is an abundance of bus stops all over the area, therefore there is a good amount of people that could walk to a bus stop. Moreover, considering that the Trillium district is 3km across, anyone in this area can bike to the nearest stop. As there are several stops along the middle of the district, one would have to cycle no more than 1.5km to get to a bus stop. Lastly, for anyone that may not want to or may not be able to bike, the numerous parking areas on the major road that cuts through the district would allow them to drive and park, and take the bus from there.

We repeated a similar analysis for every district in the city of Kingston and visualized our observations below.

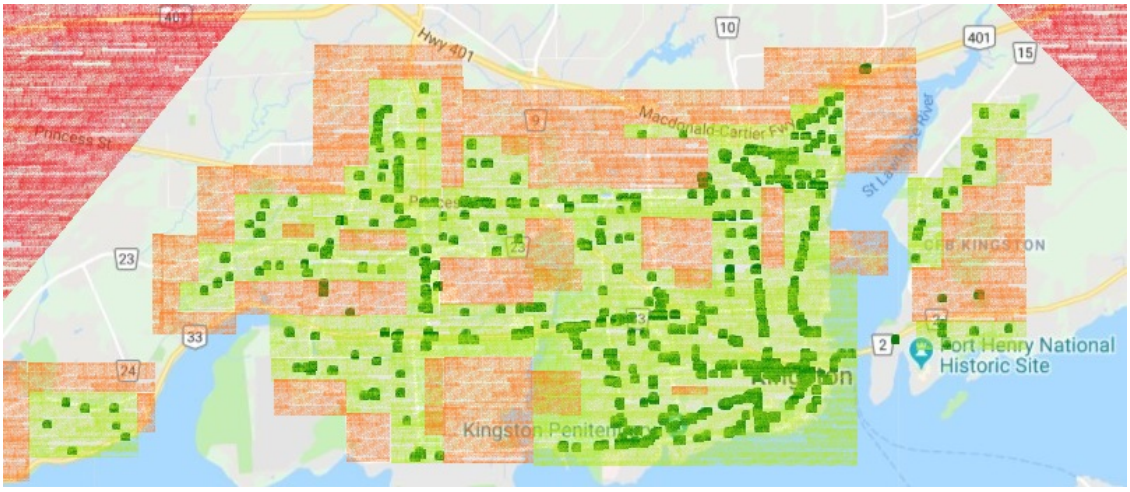

Figure 4: Accessibility to bus stops for residents of Kingston. Green means people within this area can feasibly walk to the nearest bus stop. Orange means they can bike. The red zone means that the transit system is not accessible to the people in this area.

In conclusion, although Kingston Transit does a good job at making busses accessible to those who live closer to the city center, there are some larger areas towards the outskirts of the city that do not have feasible

access to the transit system. However, considering the low population of these areas, and the fact that most of these people have cars due to the remote location of where they live, this doesn't pose a huge problem.

5 How many new bus routes are needed to make people leave their cars at home?

Assumption: the question is saying that all people that currently are working adults will now work in Downtown Kingston.

If we want everyone to leave their cars at home, that means we must make the transit system accessible to everyone by either walking or biking. For this, we can revisit the earlier analysis of accessibility of the transit system and design routes for those who were only able to access the bus through ride and park, and for those who weren't able to access it at all. Secondly, if we want everyone to take the bus instead of driving, we would need to account for the sheer amount of people that would take the bus that otherwise wouldn't.

The first part of solving this problem is to extend the transit system to reach people who usually have to drive to the nearest bus stop. This would be the areas that are not shaded in the figure below.

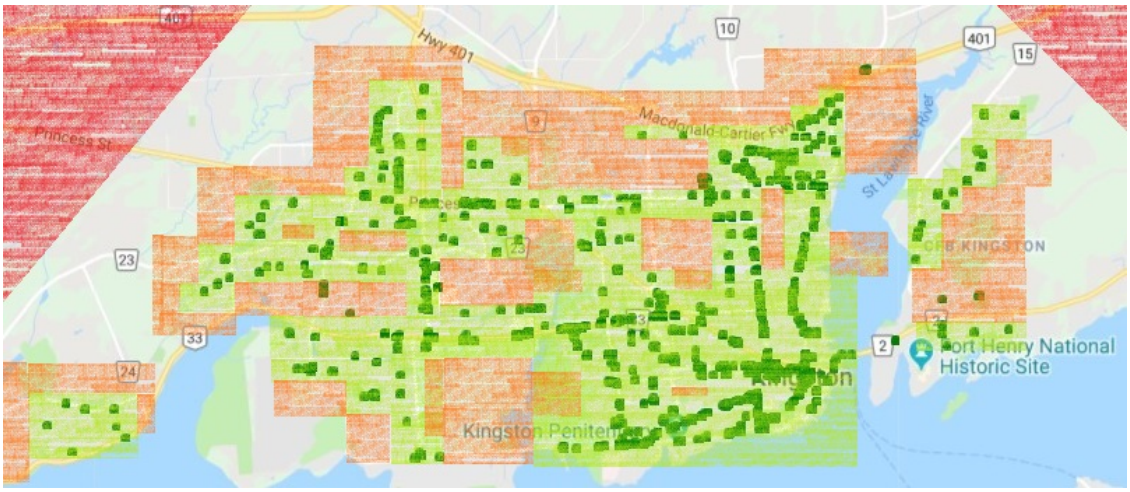

Figure 5: In this scenario, we want to increase accessibility to those who live in the non-shaded and the red-shaded areas.

We make a reasonable assumption that we would need to either;

1. Add 10 bus routes that stop at central locations in these areas and take them closer to downtown where they can get directly to work or transfer to another bus that can get them there.

2. Create 5 new bus routes that travel more robustly within these regions, and also extend some existing bus routes that come from the city center and connect them to the new routes. Therefore, people living in these districts can easily travel within their own district but also efficiently take the bus all the way downtown.

We believe that the second option would be better for the population. These number of routes can sufficiently handle the amount of people that live in these areas and also feasibly allow them to travel downtown to work, but also allow them to have more freedom in travelling around their own area. This method would also reduce congestion during rush hours, since people would have more options in which route to take to get home/to work. Moreover, since these neighborhoods are very sparsely populated, so the busses would not need to cover every road like the bus routes in the city center do, and not need to have as many stops, so it could actually be a viable solution.

The second part of this problem is that we need to account for the increase in ridership in the areas closer to downtown. There are many people that could feasibly take the bus right now bus still choose not

to. This is likely from the fact that don't have the time or energy to walk or bike to the nearest stop. The solution to this would be to have busses travel through more neighborhoods and closer to people's homes so that they would be more likely to take the bus. Possible solutions could include the following;

1. Alter current routes to travel more robustly within neighborhoods.
2. Add new routes to travel within neighborhoods and connect them to transfer points.

We think that the first option would be better for the transit system, as altering routes would be logistically easier than planning new routes. This would be the most efficient way to increase accessibility and therefore ridership on the transit system.

With the new and altered routes to reach more people in Kingston, we believe that the above solutions would be able to cut gas emissions in half. We believe that given the option, people would rather ride the bus since it requires less effort and expense on the rider's part, and if the routes reach close enough to their houses they may seriously consider this option

6 How redundant are the bus routes?

From the Kingston transportation service website, we found that Monday to Saturday has the exact same bus schedule, and Sunday has its own schedule. Given that the latitudes and longitudes are defined for the bus stops, my first objective was to take an arbitrary weekday and find the overlap between stops from each bus route to each other bus route. This would give me an idea of which routes are completely disjoint, and allow me to hone in on the potential redundancy of the overlapping routes.

Below is a bar plot representing the number of common stops between bus routes. If two routes did not have a significant overlap (arbitrarily defined as 50 people getting on a bus from either route at the same stop), they were not included in the plot.

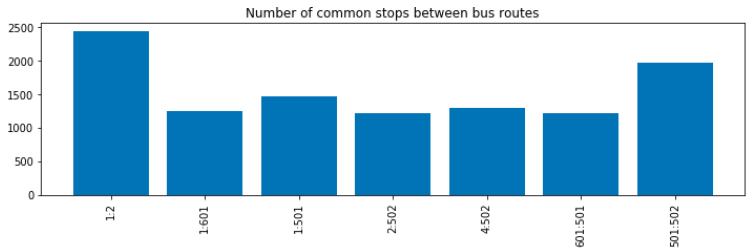

Figure 6: Common stops between bus routes.

Now that we found which routes had common stops, we were interested in figuring out which stops exactly were the most common. Given that Kingston transit was trying to make key spots in the city have buses coming through often, then we hypothesize that the redundancy that we will find is actually desirable in the eyes of the passenger.

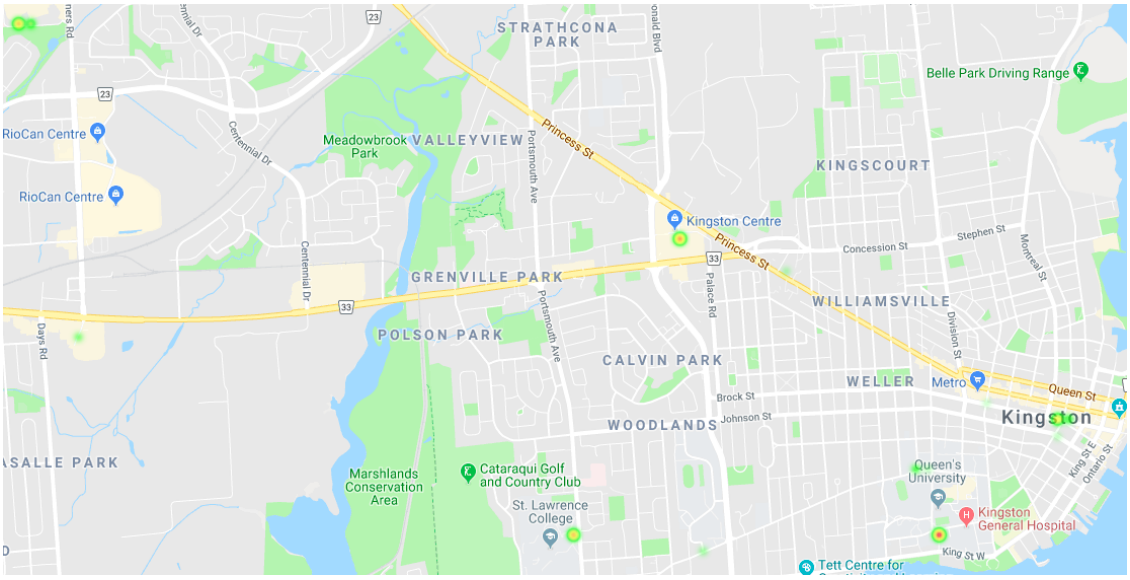

Figure 7: Three locations with redundancy as seen in the top left, around the downtown transfer point, and a significant amount of traffic around Queen's university.

It can be seen that there appears to be quite a bit of redundancy in the bus route around Queen's university on Stuart Street; however, from our own experience, there is always a large group of people waiting for the bus there suggesting that this redundancy is necessary. Additionally, there is a decent amount of redundancy around the downtown transfer point, but the redundancy is there because people are transferring from one bus to another. You can also see common stops at the Kingston center, which is another important transfer point. Therefore, from this simple visualization, we determined that there is redundancy, but it is necessary.

7 Are there any days in which Kingston should change their bus routes?

To answer this question, I was interested in seeing whether there were any days with ridership significantly lower than the mean. Below is a plot of the amount of ridership per day:

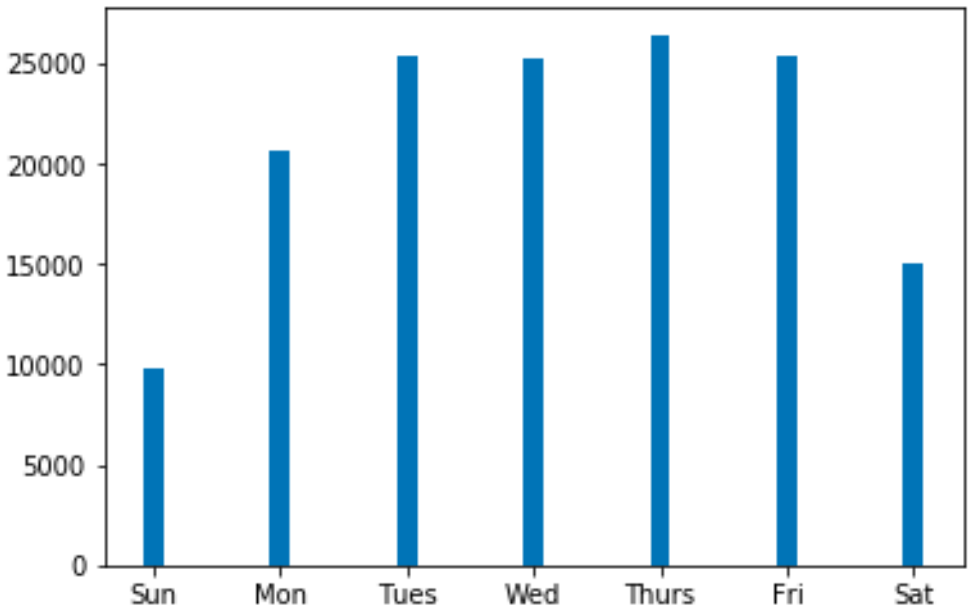

Figure 8: Bar plot of people taking the bus as a function of day of the week

It can be seen that both Saturday and Sunday have much less ridership than the weekdays, however Sunday also has a much more limited bus schedule. Since Saturday has the same bus schedule as the weekdays, then the Saturday schedule wastes much more gas, and a modified schedule should be suggested. Below is a plot of a large 1000-size sample of stops across numerous Saturdays:

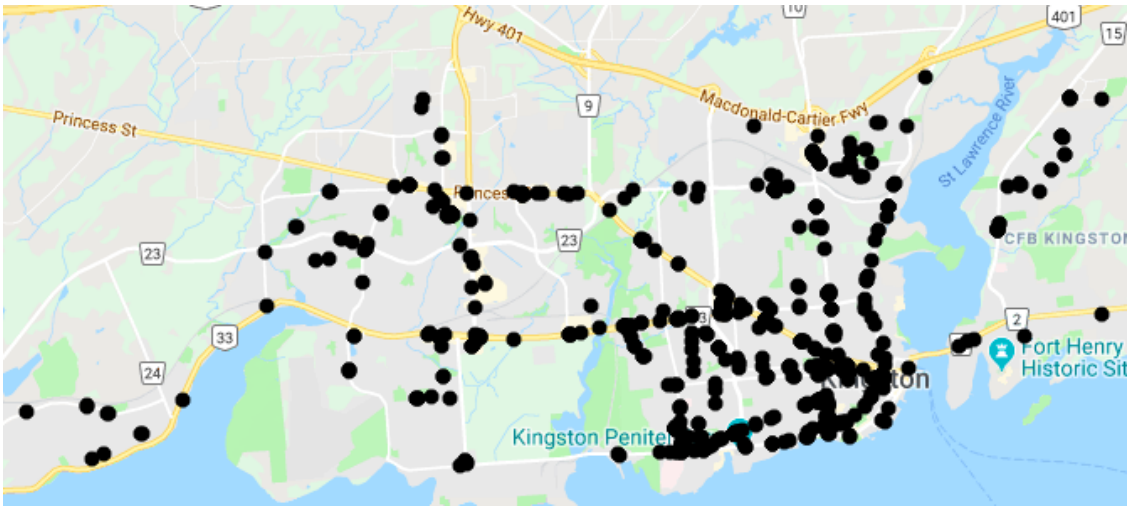

Figure 9: Sample of stops for an arbitrary Saturday in the month of October

It can be seen that on the outskirts of the city, the density of bus stops is very low, and there are large distances that need to be covered to get from those stops towards the more central areas of the city. This could potentially mean that bus routes on the outskirts should also be more limited on Saturdays given that less people are taking these buses.

8 Conclusion

Overall, it seems like Kingston Transportation got most things right. There are only a few areas of high redundancy, and you can clearly see that those areas are important transfer points in the city. Kingston could do some work on finding a better bus route for Saturdays since much fewer people are taking the bus, yet the route is identical to the routes on busy weekdays. Additionally, five new bus routes should be added to reach the rural areas to give them the accessibility to reach downtown Kingston.