Investigation of Toronto Bike Share service in December 2021*

Bike Share Toronto Ridership for December 2021

Sampson Zhao

06 February 2022

Abstract

This report uses data collected by Bike Share Toronto to log the usage of the service throughout the city. This information can be helpful in investigating the division of resources and understanding how to alter service during off-peak seasons, like December. Results show that certain stations see less pickup traffic due to their surrounding areas, which can be an indication that these lower-use stations are located in areas that cater a population that require the service less, such as office spaces or highly residental based neighorhoods.In contrast, high foot traffic areas with dense population, like near the University of Toronto and Hospitals, see increased use. The collected information can be informative to the Bike share service, as it can use processed data to understand the movement of their bikes, and gauge out the typical flow of the service, helping them optimize when bikes may need to be moved from one station to another to cater to the demand.

Contents

*Code and data are available at: https://github.com/Sampsonzhao/bikesharetorontodec2021.

1 Introduction

Bike Share Toronto is a municipally managed bike sharing program that provides short term bike rentals to users to do short trips around Toronto. There are 625 stations around the city for users to pick up a bike either through a one-time use pass or an annual pass. Depending on the type of pass purchased by the user, rides usually last around 30-45 minutes, with varying fees after the allotted time. Users of the Annual pass are allowed unlimited 30 minute rides. Bike Share Toronto released it's ridership data as a form of sharing the overall usage and usability of their service, and ultimately striving to increase the public interest in data relating to government related programs.

This data has been shared since 2014, when the Bike Share Toronto system was transferred from a private company to the Toronto Parking Authority, which lead to the further expansion of their system, which increased the accessibility of their services overall in the city. Overall, a general decrease in the use of this system by non-pass holders (i.e. non-annual pass users) has been seen as the service has now been considered a part of the common mode of public transportation(Kalinowski 2014), as this was their goal back in 2014.

This report will discuss the granularity of the data, looking at the collection of this data. It will then address the concerns with privacy of information, as information uniquely tied to users were shared in the data set. We can then look into processing the data, while trimming the data to understand the overall usage of the Bike rental system, by randomly selecting a percentage of the stations and looking at the usage of the station.Finally, we would look at how this data can be used in the allocation of labor and talk about the implication of how this data can be used in the form of expansion.

2 Data

2.1 Acknowledgement

This analysis for this project uses the 'R' programming language (R Core Team 2020), and the 'R-Studio' Integrated development environment (RStudio Team 2020). The 'Tidyverse' package was used for manipulation of the datasets (Wickham et al. 2019). The 'Here' package was used to locate files produced from R scripts (Müller 2020). The data was retrieved from Open Data Toronto through the 'opendatatoronto' package (Gelfand 2020). 'lubridate' package is used to format time and dates(Grolemund and Wickham 2011), 'KableExtra' package was used to enhance the overall appearance of tables(Zhu 2021). 'knitr' was used to tie all of the information on this document into one digestible PDF Xie (2014). 'ggplot2' was used to create the bar plots in this document. 'reshape2' was used to modify and combine dataframes for the construction of the bar plots (Wickham 2007).

2.2 Where this data comes from

The data that is collected for the Bike Share Toronto was created and shared by the Toronto Parking Authority, which is then hosted on the Open Data Toronto (ODT) portal. This allows for free public access to different data sets relating to data collected by the City of Toronto. The Bike Share Toronto data has been collected ever since the beginning of the service in 2011, when the service was owned by BIXI Toronto (Tse 2011). The previously collected usage data was not shared by the private company, but when it was acquired by the City of Toronto under the Toronto Parking Authority after the company went bankrupt, this data was not shared in its full capacity through Open Data Toronto, leading to the lack of information before 2014. Since then, the Bike Share Toronto has expanded their services and updated their system to allow them to share all of the trip data anonymously, through ODT.

The data is formatted as the following:

- Trip ID:

  - Unique for every trip, used for cross-referencing data between start and end stations
- Trip Duration:
  - Summarized data from start and end time in seconds.
- Start and End Station ID:
  - Unique for every station for determining travel time between start and end stations, as well as other cross-referencing data points.
- Start and End Time:
  - Formatted in MM/DD/YYYY HH:MM
- Start and End Station Name:
  - actual location of the station, formatted as either the address or the intersection.
- Bike ID:
  - Unique to each bike, may be used to track relative use of bike to keep track of wear and tear of the bike. Can be useful for managing the repair of bikes, when a certain level of use is achieved.
- User Type:
  - defined as either a casual user (Pay per use either single trips, 24 or 72 Hour Passes) or a Annual User (Pay for access for a year, limiting each trip to 30 or 45 minutes.)

2.3 Concerns

When we consider the format of the data, the first thing that came into concern, was the validity of their statement of the data being completely anonymous, as most of the identifying information was removed before the data was shared on ODT. But it is still a concern that there is some factor of tracking, as there were specific and unique Trip IDs that were shared in the data, making this information less then secure in terms of anonymity. One such concern is if there was a data breach of the centalized data, where information of annual pass user's and their past usage information would be leaked, this information can then be cross-referenced with other past trip data, creating a profile for people to use to track these user's based on their routines. The biggest issue here is less so related to how this data can be manipulated to aid crimes like stalking. So there is a concern, ethically, where the trip ID information can be a point of entry when considering the anonymity of the users.

2.4 Data Selection

In this report, due to the expansive amount of stations in Toronto (around 625 within the downtown core), the data was trimmed into looking at a subset of the data, but this may cause bias in the data, if data was chosen by hand. As such, a random number generator was used to generate 6 different station IDs, to be used as a uncontrolled variable, and these numbers were obtained from random.org (Haahr 2019). The number of stations was chosen as it represented 1% of all the stations in Toronto, and allowed for a breadth of representation of the data set. In theory, The utilization of a pseudo-random number generator ensures that the numbers generated in succession have no common variable between them, allowing for the values that were chosen to not be influenced by the user or other readers, and removing a source of bias (Haahr 2019). Though this may be one area that can be controlled, other factors like the skewing of the data set itself are uncontrolled bias. As the surrounding area of each station is highly dependent on the particular area of Toronto, certain groups of stations may see significantly higher proportions of usage, due to the immediate surroundings of the station. Figure 1 represents all of the ridership data of bikes being pickup from stations, and we can see that there are select few stations that have significantly higher station use, as they are typically closer to residential areas and university buildings.

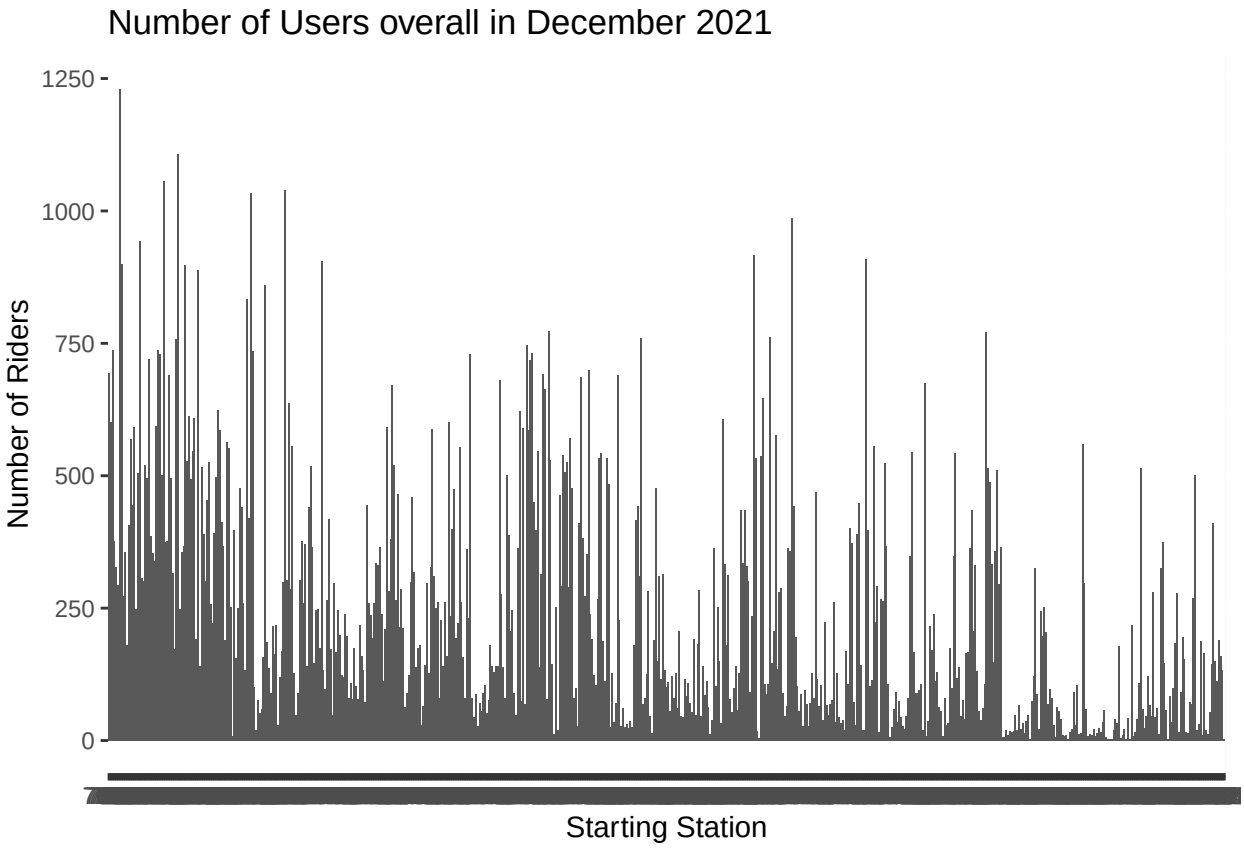

Figure 1: Overall Usage of Bike Share Toronto in December 2021

2.5 Factors that affect the data

The data also shows certain stations with close to no traffic. Tables 1 and 2 shows the disparity between the two types of stations. Understanding that the data set represents an off-peak season, where there is only a small subset of people who would be using Bike Share service as their mode of transportation, it is within reason that some stations, like the ones represented in Table 1 have little to no pick-up usage. Other factors like the weather and construction may have physically prevented users from accessing the station to pick up or drop off the rented bikes.There are other factors that can explain why some stations may be seeing increased traffic as well. For example, stations nearby may be physically impeded or were unavailable, causing users to visit other stations, thus leading to increased traffic to these stations. The number of factors that can affect the use of a particular station is multifactorial, resulting in the possibility of the data being skewed.

Table 1: Top 6 Least used Starting stations of Bike Share Toronto in December 2021

Starting Station ID	Number of Users
7651	3
7688	3
7613	2
7614	2
7615	2
7626	2

Table 2: Top 6 Most used Starting stations of Bike Share Toronto in December 2021

Starting Station ID	Number of Users
7006	1231
7038	1107
7030	1056
7100	1039
7078	1033
7418	986

2.6 Comparing activities between stations

As mentioned before, 6 random station IDs were chosen as a subset to approximate the relative activity that each station would see. The following 6 stations were randomly generated and is represented by Figure 2 and 3:

- 7107 - Cherry St / Distillery Ln
- 7249 - Dovercourt Rd / Harrison St (Green P) - SMART
- 7293 - College St / McCaul St
- 7338 - Logan Ave / Bain Ave
- 7428 - Woodbine Ave / Lake Shore Blvd E
- 7662 - Beaty Ave / Queen St W

Figure 2: Map of 6 Randomly Chosen Locations

In the case of Figure 3, we can see that there are parts of the two extremes that were mentioned in Section 2.4 and 2.5 are seen in this set of stations. While each station have similar pick-up and drop-off events at each station, there are discrepancies in terms of the higher traffic station located at College st. and McCaul St. This is most likely due to the area that this station is located, as it is located next to the University of Toronto, a junior public school as well as Mt. Sinai Hospital, this station would see more traffic due to the area having buildings with high foot traffic (the university and the public school), as well as workplaces (the hospital). As well this station is surrounded by an residential area, as well as recreational spaces, ultimately resulting in the immediate area around the station to be high in overall foot traffic. If we also take a look at the lowest scoring station, located at Woodbine Ave and Lake Shore Blvd, we can see that this area is mostly residential, with a beach park next to the station, resulting in less people who may be interested in renting a bike for use, As well there are multiple stations located nearby, as well as being closer to the main entrances

of Woodbine Park, and other high traffic buildings, which may explain the lack of usage compared to the other stations. This trend of residential areas and low foot traffic amenities next to low-usage stations are seen throughout the entire data set, as these areas seem to have less of a need for these rental bikes. These stations are also located a bit further from the central part of Downtown Toronto, where the population is most dense, which may also play a factor in less users in those parts of the city. At the same time, high-usage stations are located near high foot traffic buildings like the University of Toronto, Public schools, hospitals, and other similar buildings, hence leading to the need for people to travel to and from that particular station, when going to and from work.

Google Maps

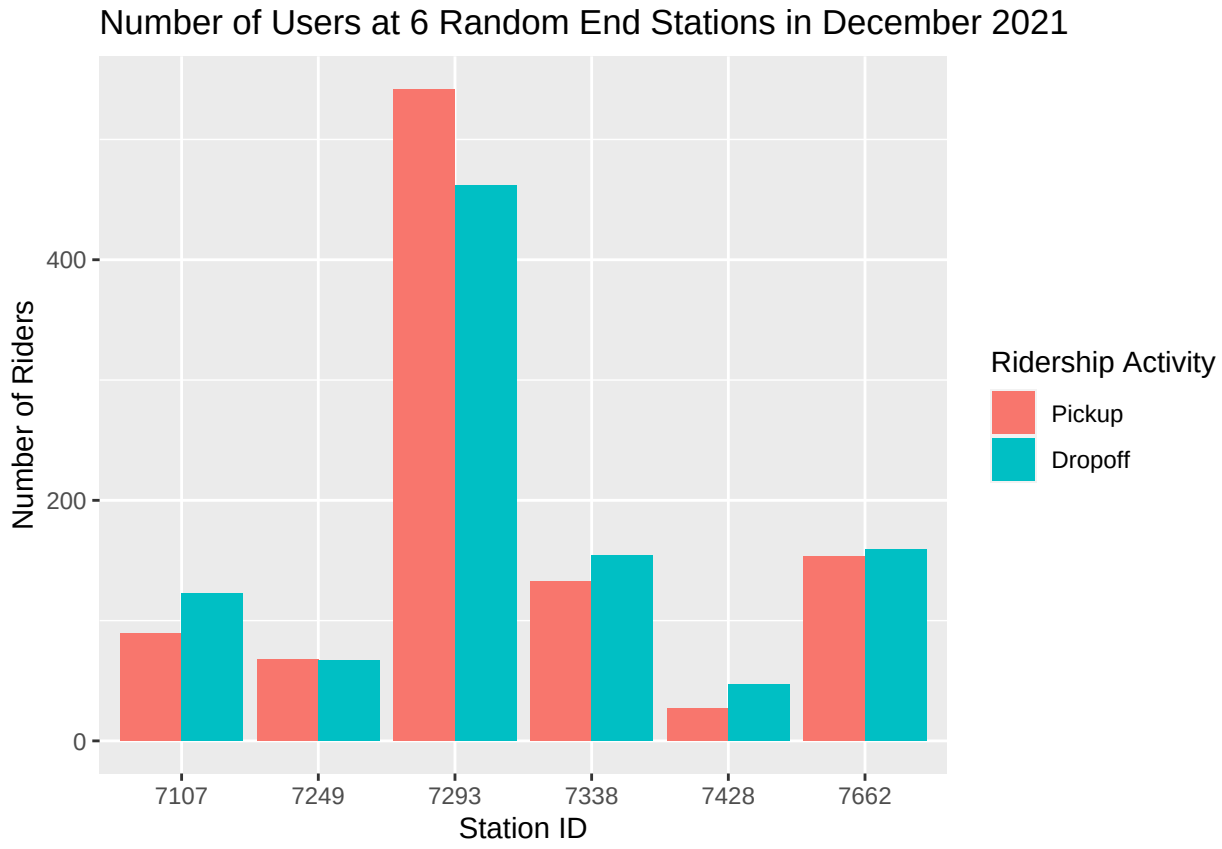

Figure 3: Overall Usage of Bike Share Toronto in December 2021

3 Discussion

3.1 How can this information be helpful

In areas with the increased foot traffic and significantly more ridership can be a sign of the station being used regularly enough that the station may end up with no bikes at the station. This may result in the supply of rental bikes not satisfying the demand for these bikes at those particular stations. As a result some people may choose to ignore the service when there are a lack of bikes during those peak usage times. Take for example February 6th, 2022 at 9:50PM, as live station information is shared on Google Maps. We can see that even during the evening of a weekend that the station at this time only has 4/14 bikes available.

This can be an indicator for the Bike Share Toronto staff to focus more on these high-usage stations, by understanding the on- and off-peak trends of these stations, and timing restocking of the rental bikes, from lower-traffic stations to higher-traffic stations. These can ensure that bikes available to users at any point of time, even at the high-usage stations. This information can also help modify station placements relating to lower-use stations. Depending on how little use a station may see during on- and off-peak months, this can help them identify which stations can be considered for removal, hence having less stations to service in low-usage settings.

3.2 Data that wasn't used

Within the data set, there are variables that are less used, due to the fact that the data itself is not looked into by the general public. In the case of this investigation, information related to the exact time and date that the trip was initiated, or the specific Trip ID of each trip, as well as the specific ID of the bikes themselves were not used, due that data variable could be used in another disucssion looking at traffic flow. Time and date information could be tracked over the frame of a day or a week to track on and off-peak hours for the particular station, or what days of the week have the most use. These data points could be an interesting continuation into looking at the overall movement of the bikes. As for data relating to particular bike IDs, that information can be used to track the wear-and-tear of specific bikes, and could be important factor to trace when the bike should be brought in for annual repairs, or even when bikes should be replaced. These sets of information can be helpful in the everyday monitoring of the assets of the company, but in the case of this investigation, this information would not be as useful into looking at the overall traffic and use of the system itself, not the specific use of a particular bike. As for the time element, because of this data is based off of an off-peak season, a majority of the stations don't see enough traffic to properly depict use over a day to represent which hours of the day are statistically busier then others. It should be a point of interest to look at On-peak seasons to investigate the on- and off-peak hours of the Service, be it the entire system or a particular station.

4 Conclusion

The findings from this paper ultimately can help Bike Share Toronto refocus some of the downfalls of this type of system, where stations are involved in the pickup and drop-off of rental bikes. There were other companies who took this weakness and improved on the bike rental system, for example, having electronic locks on the bikes themselves, and using an app interface to lock and unlock these bikes. Though those are future actions that Bike Share Toronto can have, for the mean time, understanding the flow of traffic in the 625 stations that they manage can optimize the amount of availability there is at higher traffic stations, while keeping ensuring that lower-traffic stations aren't overflowed with rental bikes. This system allows them to collect the necessary information that can help with not only the movement of bikes, but also future planning of station placement, where lower-use stations may be moved to other areas in the city, where either access is limited or not enough are available for use.

References

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

Haahr, Mads. 2019. "RANDOM.ORG - True Random Number Service." *Random.org*. https://www.random.org/.

Kalinowski, Tess. 2014. "Bike Share Toronto Sponsor Means 20 New Stations Next Year." *The Toronto Star*, December. https://www.thestar.com/news/gta/2014/12/09/bike_share_toronto_sponsor_means_20_new_stations_next_year.html.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

RStudio Team. 2020. *RStudio: Integrated Development Environment for r*. Boston, MA: RStudio, PBC. http://www.rstudio.com/.

Tse, Jennifer. 2011. "BIXI Bike-Sharing Officially Launches in Toronto." *Www.blogto.com*, May. https://www.blogto.com/city/2011/05/bixi_bike-sharing_officially_launches_in_toronto/.

Wickham, Hadley. 2007. "Reshaping Data with the reshape Package." *Journal of Statistical Software* 21 (12): 1–20. http://www.jstatsoft.org/v21/i12/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

———. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. https://yihui.org/knitr/.

———. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.