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. Datasets are collected based on months of usage, and not throughout a fiscal year, creating some limitations, but this segregation 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, 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, etc. 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

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.

---

*Code and data are available at: https://github.com/senpaiz477/opendatatoronto-bikeshare-dec2021.

This report will discuss the granularity of the data, looking at the importance of the how the data is collected. Then we would look at how this data can be used in the allocation of labor and finally we can talk about the implication of how this data can be used in the form of expansion.

2 Data

This analysis for this project uses the R programing 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 enchance the overall appearance of tables(Zhu 2021).

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.)

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

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

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.

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). 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. 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.

The data also shows certain stations with close to no traffic. Figure

Based on these

BikeShare 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. Here we obtained data from the first 6 bikeshare stations in Toronto, during the month of December 2021.

Open Data Toronto shares this information through their website https://open.toronto.ca/. We downloaded the data, then cleaned the dataset using R and R Studio.
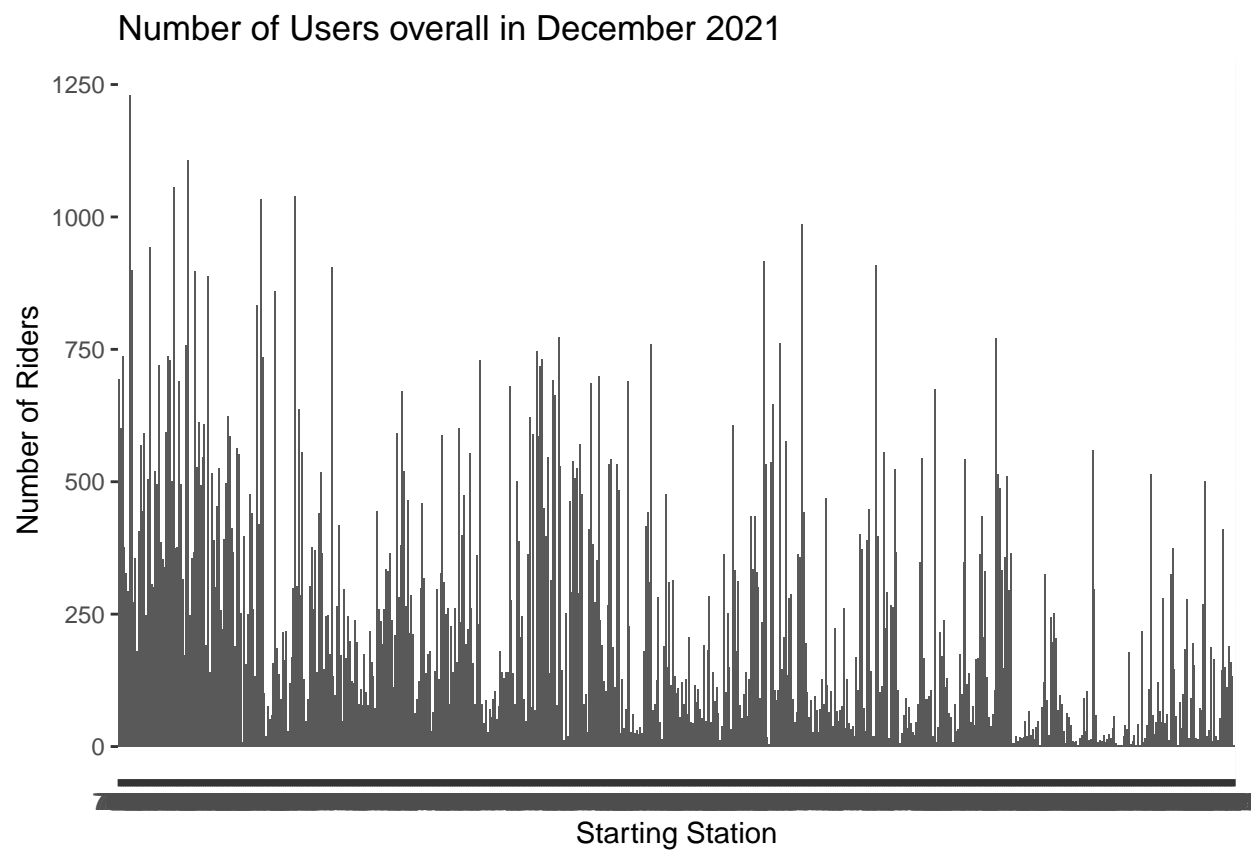
Number of Users overall in December 2021

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

We found that the density of the area around particular stations have more riderships, even during the month of December, as seen by the higher density in the area of the station was highly related to the number of rides seen at those areas, such as the St. George Station, where there is a high density of residence buildings, where as closer to King St. West Station is mainly composed of corporate offices and Mall Spaces.

These findings can support the focusing of the placement of the bikes, as higher density areas will more commonly run out of rental bikes during peak hours, hence by looking at the higher number of riderships at particular stations, predictions can be made to supplement peak hours, and moving bikes from lower density areas to higher density areas will help with the supply and demand of the bikes on a ridership level. This variable can also help with the prediction of construction and removal of stations, as this may help decide the popularity of the stations, as well as the usage, which will help decide whether a new station near by would be beneficial to the area.

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, 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.

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