Analyzing TTC Delay Data to Improve Future Efficiency*

Timothius Prajogi

January 23, 2024

TTC delays have been increasing over time, and especially now in a period of overhaul in many parts of the system, it becomes important to analyze the effects of subway and bus delays in relation with each other as well as commuter loads. Using data published by the TTC and accessed through `opendatatoronto` we find that the greatest causes of delays are date and time, and equipment quality. Date and time factors are out of anyones control, but must be accounted for to improve future operational efficieny, while equipment quality has also been shown to be important, with mechanical failures heavily decreasing the reliability of buses.

Table of contents

*Code and data supporting this analysis is available at: https://github.com/prajogt/ttc_analysis.git

1 Introduction

The main source public transportation of public transportation in the city of Toronto is the TTC (Toronto Transit Commission). The most common methods of this public transportation is subways and buses. For people who frequently use these methods, delays are common and expected (Westoll 2023). So much so that leaving early is leaving on time. With the delays getting worse and service becoming more inconsistent (Katie Daubs 2023), it is important to analyze potential causes, which would therefore enable the TTC to explore potential solutions. One of such possible factors is the tight coupling of the subway and bus systems. When issues occur with the subway, the overflow is sent to buses, either from detoured commuter traffic, or resource reallocation in the form of converting regular route buses to shuttle buses (Larsen 2023). Whether you, the reader, may be a commuter or not, having a good and reliable source of public transportation is crucial to reducing travel time for all city goers, not to mention its additional benefits for the environment.

Within this paper we will analyze the correlations between the delays occurring on the subway and the delays that occur for buses. In particular, the correlation between mechanical or security issues on the subway, which tend to shut down a portion of the track, and similar equipment and now operator failures that occur for buses. We will see the general affects of the commuter load on the transit system and how the system can be better optimized or at least supported with redundancies to prevent such frequent delays. The paper will show the trends of the amount of delays for both systems as a function of day and time. It was found that work days put additional load on both systems, the total amount of delays during weekdays being significantly higher than on other days. It is found that the cycle of the traditional work week and work day, the environment in which each exists in with regards to the facilities, equipment, and workers, have a strong correlation with delays in the TTC.

The next section will cover the content of the data and analyses on the results that were obtained from the datasets.

2 Data

The datasets used for this report are the latest delay statistics from the 2023 year, `TTC SUbway Delay Data` (Data 2023b) and `TTC Bus Delay Data` (Data 2023a), both of which were published by the TTC.

Data was downloaded, cleaned, parsed, analyzed, and visualized using R (R Core Team 2023), a statistical programming language, with package support from `tidyverse` (Wickham et al. 2019), a collection of libraries which included the following packages that were utilized:

- `ggplot2` (Wickham 2016)
- `dplyr` (Wickham et al. 2023)
- `readr` (Wickham, Hester, and Bryan 2023)

- `tibble` (Müller and Wickham 2023)

For additional assistance with cleaning, the `janitor` (Firke 2023) package was used. For additional assistance with report generation the `knitr` (Xie 2023) package was used.

2.1 Overview of Datasets

Table 1: Sample of Bus Delay Statistics

Date	Time	Day	Incident	Min Delay	Min Gap
2023-01-01	02:30:00	Sunday	Miscellaneous	81	111
2023-01-01	02:34:00	Sunday	Security/Safety	22	44
2023-01-01	03:06:00	Sunday	Security/Safety	30	60
2023-01-01	03:14:00	Sunday	Security/Safety	17	17
2023-01-01	03:43:00	Sunday	Security/Safety	1	1

Table 2: Sample of Subway Delay Statistics

Date	Time	Day	Incident	Min Delay	Min Gap	Line
2023-01-01	01:20:00	Sunday	Operator	0	0	Yonge-University
2023-01-01	01:54:00	Sunday	Security/Safety	0	0	Yonge-University
2023-01-01	02:22:00	Sunday	Security/Safety	3	9	Yonge-University
2023-01-01	02:30:00	Sunday	Security/Safety	0	0	Bloor-Danforth
2023-01-01	02:33:00	Sunday	Security/Safety	0	0	Bloor-Danforth

Both Table 1 and Table 2 contain columns "Date" and "Time" in which the delay had occurred, and an "Incident" in which the type of incident that had occurred out of these 4 general types: "Operator" (for driver/operator related issues), "Security/Safety" (for fire, police, trespassing, etc.), "Equipment/Mechanical" (for equipment, track, vehicle, etc., failures), and "Miscellaneous" (for diversions and other external issues).

For Table 1 and Table 2 the "Min Delay" and "Min Gap" are in minutes. Where "Min Delay" represents the delay in the schedule for the next bus and "Min Gap" is the scheduled time from the bus ahead of the following bus.

For Table 2 the additional parameter "Line" represents the subway line in which the delay took place in. For those not familiar, these are the names of the different subway tracks the TTC utilizes.

In total Table 1 contains 51,406 entries, whilst Table 2 contains 22,465 entries, a significant difference.

Both datasets went through basic cleaning for readability and irrelevant or unusable columns were removed and not considered in this report.

Originally, the subway data reported the incident causing the delay as a unique code rather than an understandable description. This description was included instead in a separate table from the same package (Data 2023b). Using `merge` from dplyr (Wickham et al. 2023), the incident was mapped to a generalized version of the description to match with the same set of generalized descriptions in the bus delay data. This selection into sections based on the description important to note due to the potential subjectivity of the sorting. There are incident descriptions in which the record may fall in one or the other. Please see the data cleaning scripts for more details.

2.1.1 Issues with Dataset

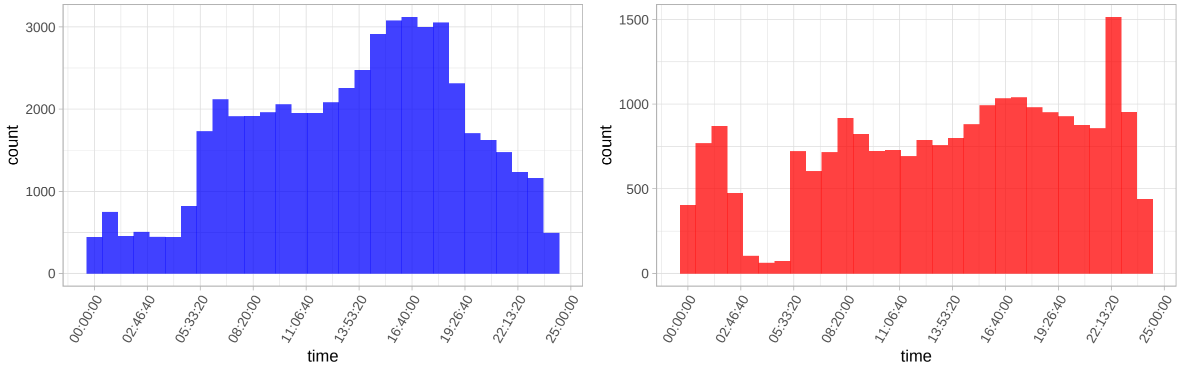
The vast majority of entries within the subway delay dataset had a recorded delay of 0 minutes. It seems as if a delay of 0 minutes really meant a delay of strictly less than 1 minute which is reasonable for a subway system. However, since this is an assumption, I have provided the filtered dataset for reference in the source repository, which only contains delay times greater than 0 minutes.

Note that this filtered dataset has been reduced to only 8,236 entries, comprising only 36.66% of the original dataset.

2.2 By Time

Figure 1 shows the difference in the trends of delays over time. Whilst bus delays heavily spike during rush hours where traffic alongside the buses has an impact on their operation, subway delays show a less significant increase in delays. Something important to note is the difference in scale that is present for all the graphs presented in this report, with the peak count of delays for Figure 1b being only around half of the peak count of delays for Figure 1a.

Regardless of the scale / degree of effect that the work cycle has on the TTC, it is evident that there is a strong correlation between the traditional schedule of the work day has with the amount of delays that occur. In a similar manner, the work week cycle has a similar correlation.

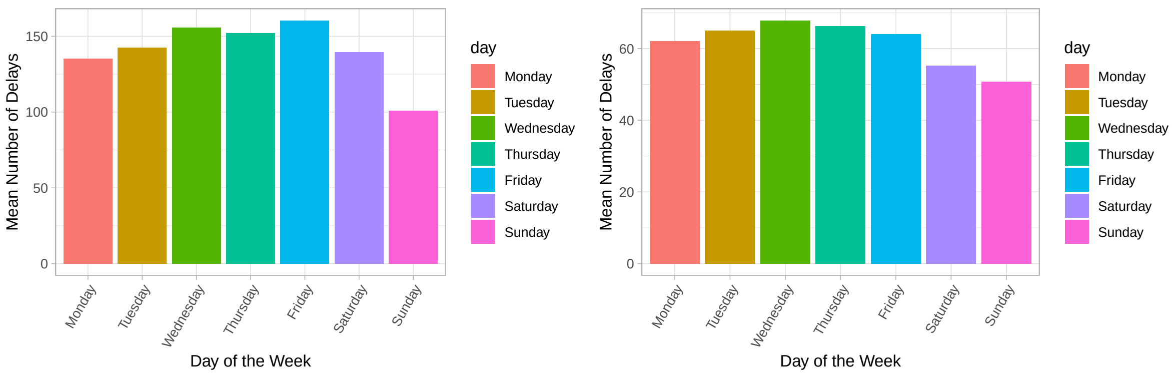


(a) Bus Delays

(b) Subway Delays

Figure 1: The amount of delays with respect to the time of day

2.3 By Day



(a) Bus Delays

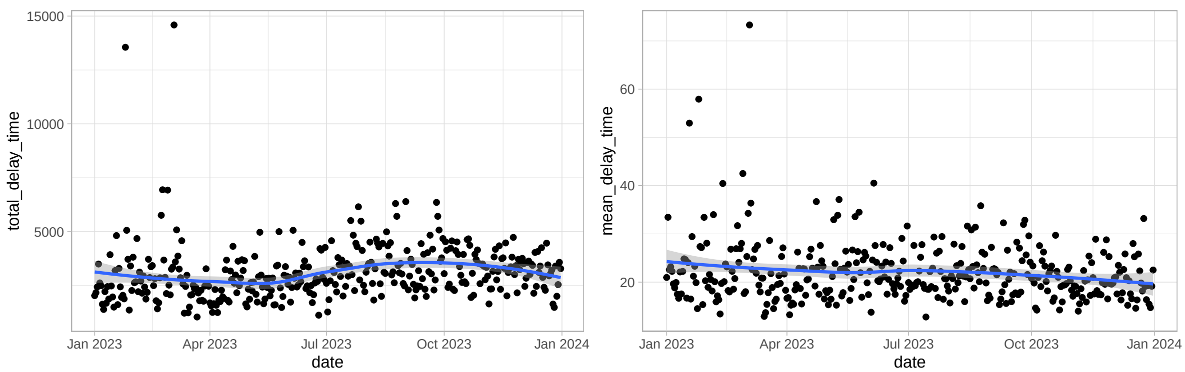
(b) Subway Delays

Figure 2: The average number of delays occurred on a given day of the week

Figure 2 shows the similar trends that are present in the average number of delays that occur on certain days of the week. On average Sundays in 2023 produce the lowest amount of delays. Buses and subways average 101.00 and 50.75 delays respectively on Sundays. It seems that the added commuter load or the assumption of an increased commuter load causes more delays. We cannot say for certain which of the two it is, as it is equally likely that the reason that more delays occur on certain days is because more buses and trains are scheduled more frequently to accommodate the load of working commuters. Equally, it is possible that the increase in delays is caused by a human actor. As shown by Figure 4, a large part of the delays for subways are security and safety issues which is almost entirely comprised of human effects. What we can say now however, is that more preparation and further analysis into commuter effects must be done by the TTC to lower the large amount of delays caused.

Table 3: The total and average time in minutes wasted on delays

Date	Day	Total Delay Time	Delay Count	Average Delay Time
2023-08-11	Friday	4293	202	21.25248
2023-02-27	Monday	6929	163	42.50920
2023-11-10	Friday	4183	207	20.20773
2023-07-19	Wednesday	3411	149	22.89262
2023-07-04	Tuesday	2722	143	19.03496



(a) Total Delay Time

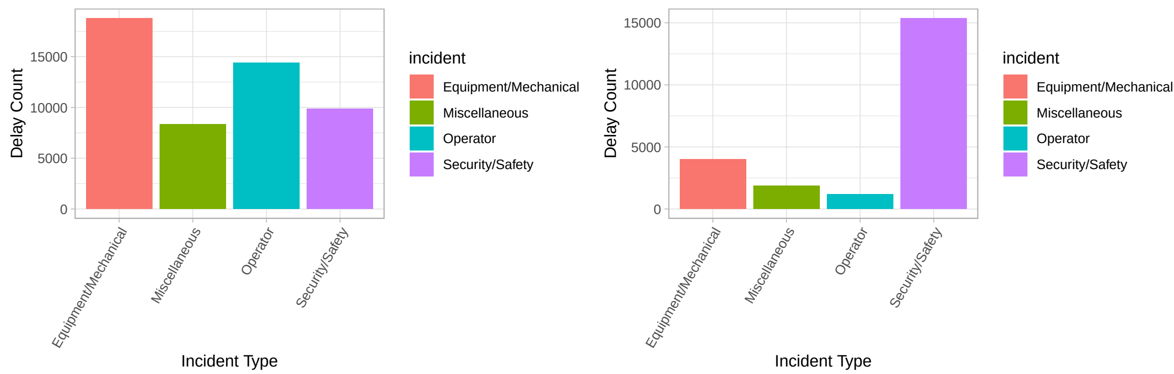
(b) Average Delay Time

Figure 3: Bus Total and Average Delays in Time (Minutes)

Here Table 3 is a random sampling of dates in 2023 and the total amount both in time (minutes) and quantity of delays that occur on that particular date. So much time is wasted on delays for commuters, supporting the claim that City News (Westoll 2023) made about the quality of service. Seeing this further in Figure 3, we see that the average bus commuter spends around 20-25 minutes every day waiting for delays, and the TTC overall has wasted around 30,000 minutes (500 hours) of commuters time as a whole.

2.4 By Incident

From Figure 4b we can see that the leading cause of delay on the subways is “Security/Safety” incidents with 15,367 delays caused by such issues, with the next leading incident being “Equipment/Mechanical” issues with 4,005 delays occurred. “Security/Safety” covers track invasions, unsanitary trains, fire, etc. and “Equipment/Mechanical” covers any problems occurring with the functionality of the trains, the station, and the tracks. From personal experience, these results are expected. Comparatively, bus delays in Figure 4a are majorly caused by “Equipment/Mechanical” issues with 18,786 caused. The second leading incident type is operator



(a) Bus Delays (b) Subway Delays

Figure 4: The average number of delays occurred due to incident types

issues which covers any problems with the driver of the buses. These incidents account for 14,400 delays. These results pose an interesting connection.

From the TTC website, we know that they have been moving towards implementing automation in the operation of subway services since 2022("TTC's Line 1 Now Running on an ATC Signalling System" 2023). This has not been possible for buses which is why operator error is still high on the list. This could also be explained by the ratio of operators to commuters. Whilst each bus requires an operator, multiple cars on a subway only require 1-2 operators especially with the addition of automated signalling. Similarly, the lack of employee oversight within stations and subways can explain the difference in security and safety issues. Buses don't have much to oversee, however since stations are large and employees are only stationed within the entrance floor, prevention of disruptive behaviour is more commonplace as displayed by Figure 4b above.

2.5 By Line

From Figure 5 we see an expected result, where the longest (and oldest) lines with Yonge-University having 38 stations and Bloor-Danforth having 31 compared to Sheppard's 5 and Scarborough's 6 (previously existing) stations. With the load that these older lines are experiencing on a daily basis, the result from our analysis on delay causes is further supported. Constant maintenance and repair is required for these lines, and increased ridership also puts extra stress on the systems. Further engineering analysis is required to definitively know the extent of the effects of stress loads on these lines, and if those are the cause for the increase in delays.

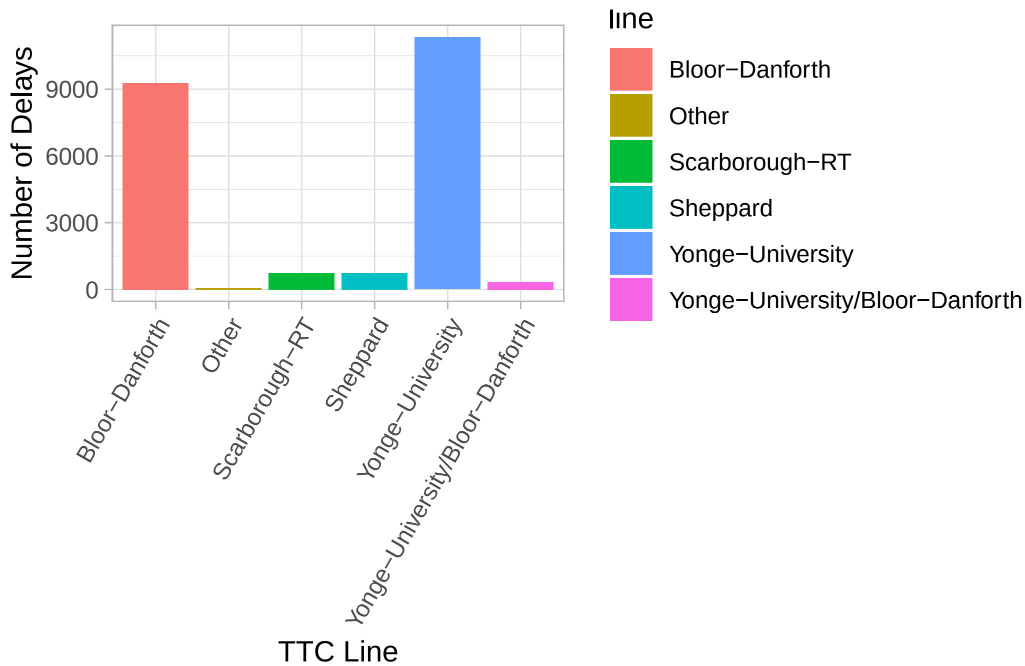


Figure 5: The number of delays on each TTC Line

3 Conclusion

In conclusion, the analysis of TTC delays has revealed critical factors impacting the efficiency and reliability of Toronto's major transit system. Analyzing data obtained from the TTC through `opendatatoronto`, we identified date and time, equipment quality, and operation failure as the major contributors to delays. Although we cannot control date and time factors, acknowledging them and their effects on the system is vital for improving future operational efficiency. Recognizing the impact of equipment quality, particularly in relation to bus mechanical failures, highlights the need for specific upgrades and maintenance in transit infrastructure. The insights from this analysis can guide future spending and research, in an attempt to be a part of the effort to enhance the TTC system for the benefit of both commuters and the city.

Reference

Data, Toronto Open. 2023a. "TTC Bus Delay Data, 2023." https://open.toronto.ca/dataset/ttc-bus-delay-data/.

———. 2023b. "TTC Subway Delay Data, 2023." https://open.toronto.ca/dataset/ttc-subway-delay-data/.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Katie Daubs, Andrew Bailey. 2023. "Attention TTC Passengers: The Delay You're Experiencing on the Subway Is Longer Than Ever. Here's Why." https://www.thestar.com/news/gta/attention-ttc-passengers-the-delay-you-re-experiencing-on-the-subway-is-longer-than-ever/article_ca54f6e9-db07-5f8b-800e-3cdfe616257b.html.

Larsen, Christian. 2023. "Scarborough Commuters Say New Bus Shuttle Service Is Causing Delays, Hassles: Decommission of Scarborough RT Line Has Affected Tens of Thousands of Riders." https://torontoobserver.ca/2023/09/22/scarborough-shuttle-bus-rt/.

Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

"TTC's Line 1 Now Running on an ATC Signalling System." 2023. https://www.ttc.ca/news/2022/September/TTCs-Line-1-now-running-on-an-ATC-signalling-system#:~:text=The%20TTC\XeTeXglyph\numexpr\XeTeXcharglyph"0027\relax{}s%20modernized%20Automatic%20Train,modernize%20our%20current%20transit%20system.

Westoll, Nick. 2023. "TTC Receives 'b' Grade for Service and Reliability, but How Accurate Is the Ranking?" https://toronto.citynews.ca/2023/07/06/ttc-buses-streetcars-subway-reliability/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

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.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.