The Degradation of TTC: Analyzing Delays in Toronto's Subway Service*

Anannay Sood

2024-01-25

This paper presents a data-driven analysis of the Toronto Transit Commission's subway delays from 2018 to 2023. It highlights a significant increase in delay occurrences, particularly post-COVID-19, with the severity of delays remaining stable but some exceeding eight hours. The study focuses on the Yonge-University and Bloor-Danforth lines, identifying passenger disorderliness and crowding as common delay causes. It concludes with insights into potential overcrowding and infrastructural issues, especially on the Yonge-University line.

Introduction

In recent years, Canada has established itself globally as an advocate for sustainability. Although it has introduced many new initiatives around reducing pollution and adopting renewable resources, one area where the government's approach has been criticized has been public transportation. Specifically, articles criticizing the state of the Toronto Transit Commission (TTC) have been increasingly common. For example, this article by Stephen Spencer Davis highlights the many issues with the transit infrastructure in the city, with a particular focus on unreliability (Davis 2023).

This paper examines the validity of these claims through rigorous data analysis. Specifically, the paper uses R (R Core Team 2022), tidyverse (Wickham and RStudio 2023), dplyr (Wickham, François, et al. 2023), lubridate (Spinu et al. 2023), knitr (Xie [aut et al. 2023), and ggplot2 (Wickham, Chang, et al. 2023) alongside TTC subway delay data since 2018 to examine how the quality of public transportation in the city has changed over the five years. The data is sourced from opendatatoronto (Gelfand and Toronto 2022) and is officially published by the Toronto Transit Commission.

The general takeaway from the analysis is that there has been a marked increase in the number of delay occurrences post-COVID-19. The number of delays has almost doubled. However, the

*Code and data available at https://github.com/AnannayS/STA302_Paper1 severity of delays has remained relatively stable over the five years, though the worst delays have exceeded 8 hours in length. Furthermore, our analysis finds that the delay increase is present only for the main lines of Yonge-University and Bloor-Danforth, with the former seeing the biggest increase in delay frequency. Moreover, the most common sources of delay are due to passenger disorderliness and crowding. The analysis indicates that the subway system is overloaded; the Yonge-University line may be overcrowded. However, it is important to note that our data is biased towards subway traffic and does not account for other public transportation such as buses, GO trains, and streetcars, and thus may paint an incomplete image of the infrastructural issues.

Data

As mentioned, the dataset used for this paper is officially published by the Toronto Transit Commission and sourced from opendatatoronto. The data was initially organized into several .xlsx files and was split by year. These files were cleaned and combined into one data frame for analysis. Table 1 shows some sample entries.

Table 1: Sample entries from the cleaned TTC Subway dataset.

Date	Day	Station	Code	Min.Delay	Line
2021-06-17	Thursday	EGLINTON STATION	SUUT	7	YU
2020-12-25	Friday	SCARBOROUGH CTR STATIO	SRDP	0	SRT
2023-05-14	Sunday	ROYAL YORK STATION	EUSC	3	BD
2022-02-24	Thursday	KIPLING STATION	MUIS	0	BD
2020-05-19	Tuesday	LESLIE STATION	TUSC	5	SHP

Each row of the dataset corresponds to a measurement done by the TTC to check for delays in the schedule. These are often zero, which indicates no delay in scheduling. Non-zero values correspond to the number of minutes a particular vehicle was delayed. The raw dataset had references to vehicle numbers, but these were removed as they were not relevant. Moreover, the dataset also records the date of the occurrence, the line it occurred on, the station it occurred at and the code indicating the reason for the delay.

Analysis

The first data dimension we examine is the frequency of delays occurring in the TTC subway system. As a first step, Table 2 shows some relevant summary statistics for the data and is divided into monthly periods. Notably, it includes the average length of delays (excluding zero-valued entries, which indicate no delay at the time of observation by the TTC measurement system), the length (in minutes) of the worst delay, and the number of delays during the month.

Table 2: Sample entries from the summarized data containing statistics about delays.

Year	Month	AvgNonZeroDelay	MaxDelay	DelayCount
2022	1	8.141479	315	933
2023	5	8.811511	195	695
2020	1	7.038321	258	548
2019	6	7.457539	455	577
2023	12	7.648649	139	703

Using this data, we construct line graphs for the number of delays to better understand the subway system's operational efficiency. Figure 1 shows this graph.

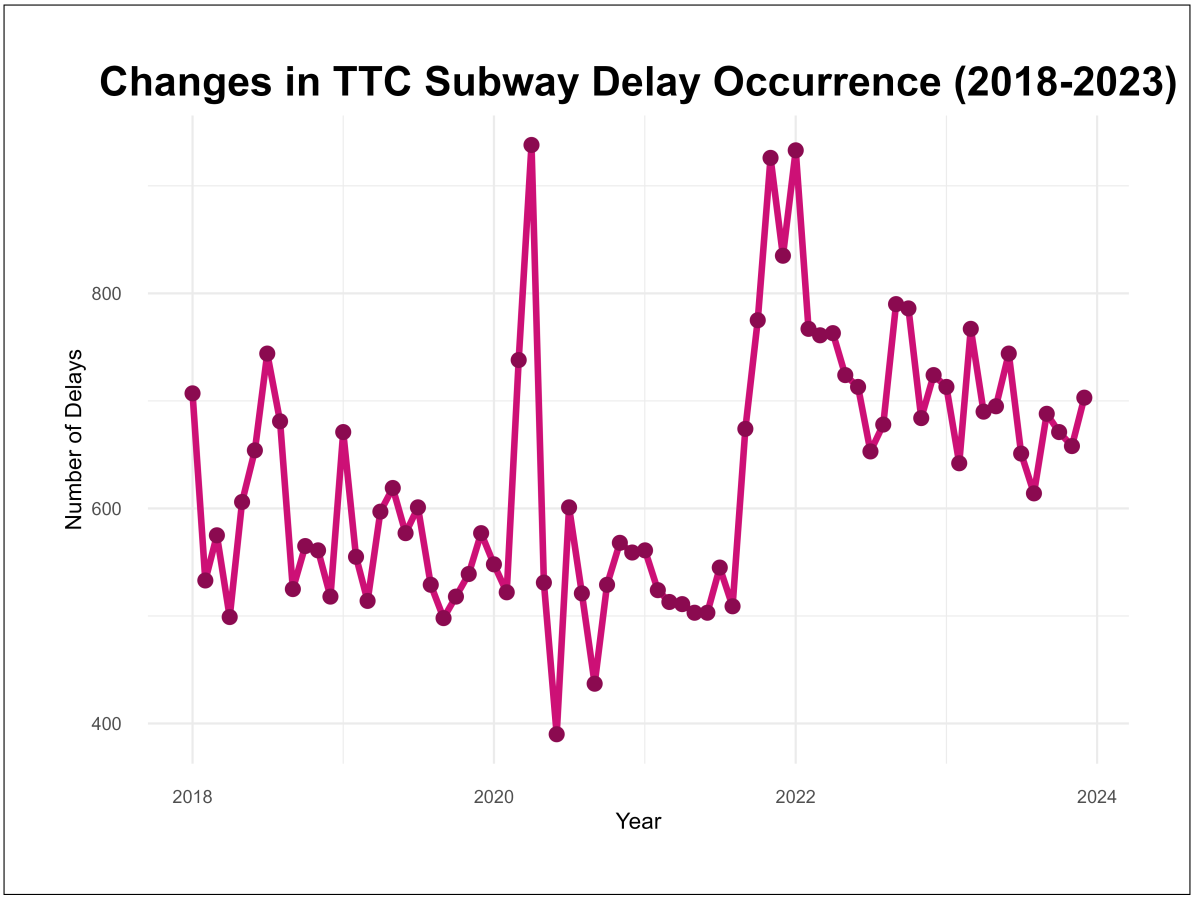

Figure 1: A line graph showing the changes in delay frequency over time.

Although the number of delays varies depending on the month, we can make two key observations. The first is that the two peaks observed in the data roughly coincided with introducing and suspending pandemic restrictions. Such drastic changes in expected traffic may have resulted in operational inefficiencies around these periods. Moreover, the valley between these peaks can also be attributed to the reduced traffic during lockdown. The second observation is that the post-pandemic delay frequency is noticeably higher than before the pandemic. This could be due to many reasons, such as an uptick in urban immigration, poor scheduling, etc. Despite this, this metric supports the assertion that the TTC's ability to manage its transportation infrastructure has worsened in recent years.

Another metric worth examining is the worst monthly delays in the service. Analyzing this metric can help illuminate whether the issues are related to the technology of the infrastructure itself. This is because multi-hour delays are likely due to issues with the tracks or the train itself and require technical expertise. Figure 2 shows a scatter plot using this metric.

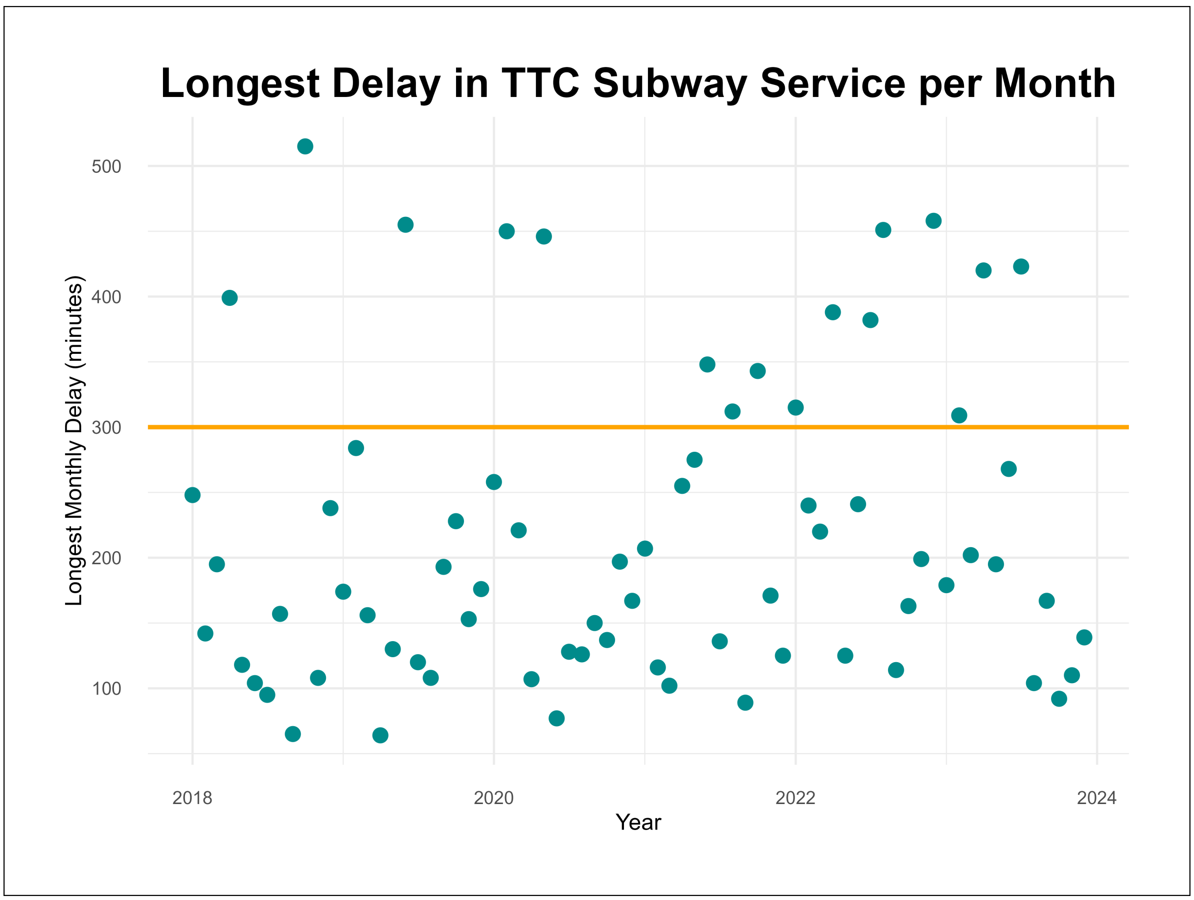

Figure 2: A scatter plot showing the worst delay every month between 2018 and 2023.

The trend here is not as clear as in Figure 1. In general, the longest delays seem to vary unpredictably over time, though there is a noticeable dip in the range of the worst delays during the lockdown period. Moreover, post-lockdown, especially severe delays, seems more common. This can be seen by comparing the delays exceeding five hours (300 minutes) between 2018-2020 and 2022-2024. The former saw delays of this length three times over the two years, whereas the latter had eight. As mentioned, long delays such as these may indicate infrastructural issues with the subway system.

Although the previous metrics provided a good overview of the state of the subway system, it is also important to explore the distribution of these delays across the different subway lines. This can help identify the subway system's most common points of failure. Figure 3 shows the main routes of the subway system.

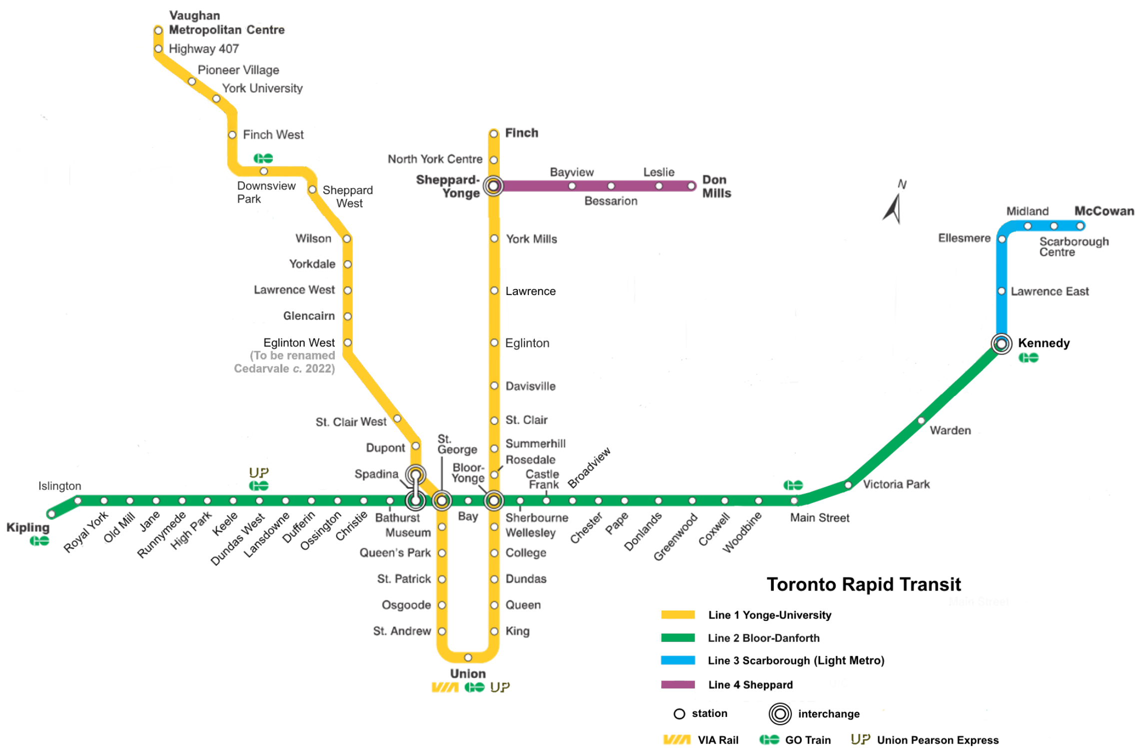

Figure 3: A map of the Toronto subway. Sourced from the TTC.

One caveat to this data is that as of January 2024, the SRT (blue) line is permanently closed and has been replaced with a bus service (TTC n.d.). The corresponds to the blue line on Figure 3. Figure 4 below shows the distribution of delays per station.

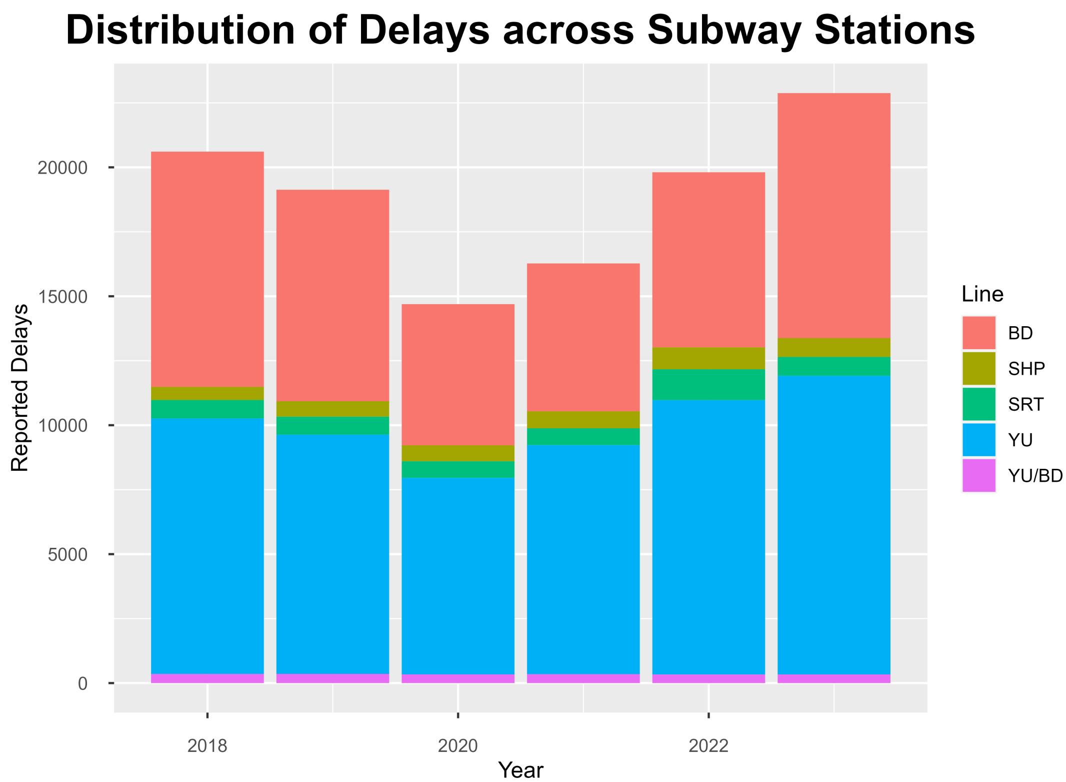

Figure 4: A stacked bar graph showing the distribution of delays.

Besides that, Figure 4 shows that a vast majority are present on the Yonge-University (YU) and Bloor-Danforth (BD) lines, which is expected given that they comprise the bulk of the subway system. However, when comparing 2018 and 2023, the number of delays on the YU line has increased significantly, while the delays on the BD line have remained roughly the same. This suggests that the YU line may be reaching its capacity limit as the backbone of the Toronto subway system.

Lastly, an additional dimension we can examine is the reason for the delay. The dataset provided by the TTC includes codes describing the type of delay and a key to interpret them. Using these codes and keys, we can see which delays are most common and how they may have changed over time. Table 3 shows the top three delay codes for each year. Additionally, Table 4 provides descriptions for all the codes in Table 3. The full list is available in the data section of the GitHub repository.

Table 3: The top 3 delay codes per year between 2018 and 2023.

Year	Code	Number.of.Incidents
2018	MUIR	479
2018	MUPAA	403
2018	SUDP	649
2019	MUATC	367
2019	MUIR	488
2019	SUDP	648
2020	MUATC	317
2020	MUNOA	1196
2020	SUDP	817
2021	MUATC	487
2021	SUDP	1050
2021	TUNOA	538
2022	MUATC	703
2022	PUOPO	730
2022	SUDP	943
2023	MUPAA	736
2023	PUOPO	664
2023	SUDP	1229

Table 4: The descriptions for the delay codes found in Table 3.

Delay Code	Description
MUIR	Injured or ill Customer (On Train) - Medical Aid Refused
MUPAA	Passenger Assistance Alarm Activated - No Trouble Found
SUDP	Disorderly Patron
MUATC	ATC Project
MUNOA	No Operator Immediately Available - Not E.S.A. Related
TUNOA	No Operator Immediately Available
PUOPO	OPTO (COMMS) Train Door Monitoring

A few interesting trends can be seen in Table 3. Firstly, the number of incidents resulting from disorderly patrons has roughly doubled over the last five years. This may indicate consumer frustrations or result from other municipal issues, such as homelessness in Toronto (Warmington 2023). Additionally, issues with train doors have become more common in the past two years. This could be a reference to overcrowded trains and may also explain the increase in disorderly patrons. Lastly, an interesting observation is the presence of the ATC (Automatic Train Control) project as a common reason for delays. Frequent delays in maintaining this system may be another sign pointing toward inadequate infrastructure.

Conclusion

Two main takeaways can be made from the data analysis above. Firstly, the increase in the number of severe delays alongside the number of delays needed to maintain the automatic train control suggests that the subway lines may have some underlying infrastructural issues. Additionally, the number of delays due to disorderliness, a driving force behind the overall increase in delays, suggests that the existing subway lines are inadequate to accommodate traffic. This is especially true for the Yonge-University line, as it has seen the biggest increase in delays. Though it would certainly be biased to make an overall conclusion about the TTC from this data, the organization's struggle to maintain its biggest and most funded services has shown its inability to respond to Toronto's growing need for public transportation.

References

Davis, Stephen S. 2023. "Who Broke the TTC? Inside Toronto's Public Transit Disaster." *Toronto Life*. https://torontolife.com/deep-dives/who-broke-the-ttc-inside-torontos-public-transit-disaster/.

Gelfand, Sharla, and City of Toronto. 2022. "Opendatatoronto: Access the City of Toronto Open Data Portal." https://cran.r-project.org/web/packages/opendatatoronto/index.html.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Spinu, Vitalie, Garrett Grolemund, Hadley Wickham, Davis Vaughan, Ian Lyttle, Imanuel Costigan, Jason Law, et al. 2023. "Lubridate: Make Dealing with Dates a Little Easier." https://cran.r-project.org/web/packages/lubridate/index.html.

TTC. n.d. "The Future of TTC's Line 3 Scarborough (SRT)." Accessed January 25, 2024. https://www.ttc.ca/about-the-ttc/projects-and-plans/Future-of-Line-3-Scarborough.

Warmington, Joe. 2023. "Toronto's Subway Platforms Have Become Homeless Shelters." *Torontosun*. https://torontosun.com/news/local-news/warmington-torontos-subway-platforms-have-become-homeless-shelters.

Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, et al. 2023. "Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics." https://cran.r-project.org/web/packages/ggplot2/index.html.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, Davis Vaughan, Posit Software, and PBC. 2023. "Dplyr: A Grammar of Data Manipulation." https://cran.r-project.org/web/packages/dplyr/index.html.

Wickham, Hadley, and RStudio. 2023. "Tidyverse: Easily Install and Load the 'Tidyverse'." https://cran.r-project.org/web/packages/tidyverse/index.html.

Xie [aut, Yihui, cre, Abhraneel Sarma, Adam Vogt, Alastair Andrew, Alex Zvoleff, Amar Al-Zubaidi, et al. 2023. "Knitr: A General-Purpose Package for Dynamic Report Generation in R." https://cran.r-project.org/web/packages/knitr/index.html.