

A Comprehensive Analysis of Delay Patterns and Operational Efficiency in the Toronto Transit Commission (TTC) Subway System*

Ricky Yuan

September 26, 2024

This paper uses publicly available data to explore the patterns and causes of delays in the Toronto Transit Commission (TTC) subway network. Through exploratory data analysis and statistical modeling, the study identifies the most common causes of delays, stations with frequent delays, and the operational impact of these delays on service gaps and vehicle performance. Not only that, this paper also analyzes peak delay times and their correlation with specific subway lines and stations. These findings provide transit operators with solutions to reduce delays and improve the passenger experience. Therefore, this improvement is critical to increasing operational efficiency. This study also provides practical recommendations for reducing future disruptions to the TTC subway system by addressing these issues.

Table of contents

1	Introduction	2
2	Data	3
2.1	Overview	3
2.2	Results	4
2.3	Measurement	4
3	Discussion	9

*A GitHub Repository containing all data, R code, and other files used in this investigation is located here:
<https://github.com/RickyYuan666/TTC-delay>

A Appendix	10
A.1 Dataset and Graph Sketches	10
A.2 Data Cleaning	10
A.3 Attribution Statement	10
References	11

1 Introduction

Public transportation systems are critical to the efficient operation of large urban centers, providing reliable transportation for millions of daily commuters. The Toronto Transit Commission (TTC) subway system is one of the most critical networks in North America and is the primary mode of transportation for the city’s residents. However, the Toronto Transit Commission (TTC) subway system suffers from frequent delays that reduce the overall efficiency of the subway system. As the city grows and the demand for efficient public transportation increases, understanding the causes and patterns of these delays is critical to improving efficiency and transit system reliability.

Previous studies have examined the various factors affecting subway delays. For example, Louie, Shalaby, and Habib (2017) explored causal (e.g., infrastructure problems) and non-causal (e.g., human error) factors contributing to subway disruptions. Their research emphasizes the complexity of managing the subway system. The reason is multiple interacting factors can prolong delays in a subway system. Additionally, Schleifer (2022) demonstrated that delay frequency varies significantly by ridership patterns, especially at major transfer points such as Union and Kipling stations. These findings suggest that understanding the specific characteristics of passenger flows and stations is critical to addressing delays within the system.

This paper aims to analyze TTC subway delay data for the year 2023 to reveal patterns of delay occurrence, duration, and the most affected stations and subway lines. This study uses data obtained from OpenDataToronto, visual modeling analysis, and statistical methods to examine trends. Examples include delays over time, the frequency at each station, and the total duration of delays on each subway line. Building on research by Louie, Shalaby, and Habib (2017) and Schleifer (2022), this analysis focuses on operational inefficiencies and their relationship to specific stations and lines. The analysis results provide actionable insights and recommendations for reducing delays and improving the rider experience within the Toronto subway system.

2 Data

2.1 Overview

This study uses Toronto Transit Commission (TTC) data from the OpenDataToronto portal for the year 2023. This dataset provides key information on delay events such as the station where the delay occurred, the cause of the delay (e.g., mechanical failure, passenger-related disruptions), the duration of the delay (in minutes), and the time of day. This clear data enables us to distinguish between peak and off-peak hours. These variables allow us to comprehensively understand the patterns and causes of delays on the Toronto subway network.

In addition to the basic structure of the dataset, we constructed several variables during the cleaning process to enhance the analysis. For example, delays were categorized based on their duration, with 0-5 minutes being short, 5-15 minutes being medium, and more than 15 minutes being long. In addition, the dataset was subdivided to differentiate between delays occurring during peak and off-peak hours, which provides a more nuanced understanding of the impact of the time of day when delays occur on the incidence of delays.

This dataset is particularly important for identifying operational inefficiencies within the TTC system. As previous research has shown, causal factors such as infrastructure failures and non-causal factors such as human error can significantly impact subway delay duration ((Louie2017?)). In addition, ridership levels and station-specific characteristics have also played a crucial role in delays, especially at major transfer points with high concentrations of riders such as Union and Kipling stations (Schleifer (2022)). By examining these variables, this study aims to gain insight into the most common causes and patterns of delays, which can inform strategies to improve system efficiency.

The analysis employs a variety of visualization and statistical methods to explore temporal, spatial, and route-specific delay patterns. Peak periods of inefficiency are identified by mapping delay durations over time, while spatial visualizations highlight the stations with the most frequent delays. In addition, bar charts can be used to compare the total delay duration of different subway lines, revealing which lines are most prone to failure. By using these methods, the causes of TTC delays can be identified, and solutions can be provided to improve transit system performance.

The analysis results show that the average delay across the network is about 7 minutes. However, delays are significantly longer at Union and Kipling stations, especially during peak hours. This phenomenon has significant scope for operational improvements at specific stations and periods.

The data was cleaned and organized using the janitor (Firke 2023) and tidyverse (Wickham et al. 2019) packages in R (R Core Team 2023), and the opendatatoronto (Gelfand 2022) package from the OpenDataToronto portal to obtain raw data. After cleaning, the dataset was tested for accuracy and the cleaned dataset was used for further analysis.

2.2 Results

After loading the dataset using the R programming language (R Core Team 2023) and the `opendatatatoronto` package (Gelfand 2022), the `tidyverse` (Wickham et al. 2019) package was used to generate graphs. In doing so, R code was adapted from Alexander (2023).

2.3 Measurement

The dataset used in this study is from the OpenDataToronto portal and focuses on Toronto Transit Commission (TTC) subway delay data for the year 2023. These data are collected by the Toronto Transit Commission (TTC) during its daily operations, using a combination of automated record-keeping systems and manual reporting to track delays across subway lines and stations. Critical variables in the dataset include the station where the delay occurred, the exact date and time of the delay, the type of delay (e.g., “mechanical failure,” “signaling problems,” “passenger assistance,” or “weather conditions”), and the duration of the delay in minutes. Also, station variables are recorded manually or automatically, which can provide spatial information about where the delay occurred. Also, dates and times are recorded using the TTC’s internal time tracking system for time analysis. Types of delays are reported by staff or maintenance teams, depending on the event, and the TTC calculates the duration of delays from the time of disruption to the return to regular service using either the automated system or manual recording. These measurements provide a comprehensive understanding of delay patterns in the subway system, documenting when and where delays occur and the causes and durations of delays. Through this robust data collection process, this study can effectively identify spatial and temporal trends in delays.

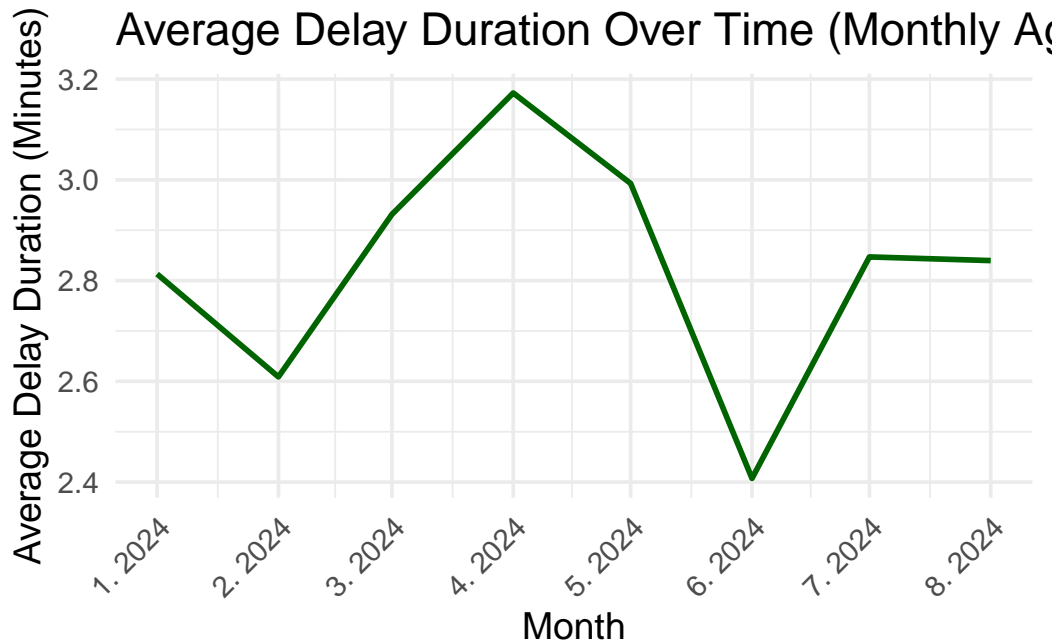


Figure 1: Average delay duration over time in the TTC subway system in 2023, aggregated monthly.

Figure 1 shows the average delay duration (in minutes) for the Toronto Transit Commission (TTC) subway system in 2023, summarized by month. The data shows the variation in delay duration over a year, with the average delay per delay event ranging from 2.75 minutes to 3.5 minutes.

The chart shows that from January through February, the average delay duration decreased slightly to just under 3 minutes. However, from March through June, the data shows an upward trend, peaking in June at approximately 3.5 minutes. This increase may be related to external factors such as infrastructure maintenance, system upgrades, or weather challenges. This suggests that the TTC may have experienced operational inefficiencies during this period. From June through October, the situation improved significantly, with average delays decreasing, reaching a low of approximately 2.6 minutes in October. This decrease may indicate that the tunnel company made operational adjustments to minimize delays during this period.¹¹ In November, there was again a slight increase in the duration of delays, which may be attributed to the inconvenience caused by increased ridership and the early winter weather in Toronto. However, the trend improved in December, with the average delay reduced to nearly three minutes, suggesting that operational efficiencies are picking up as the TTC moves with the warmer weather.

The overall trend underscores the importance of delays to the entire transit system, especially during seasonal shifts. The TTC should reduce delays with timely maintenance and proper

resource allocation. The data suggests that targeted summer and late fall interventions can be particularly effective in controlling operational disruptions. This analysis informs the TTC's planning for improvements to ensure smoother, more efficient service throughout the year.

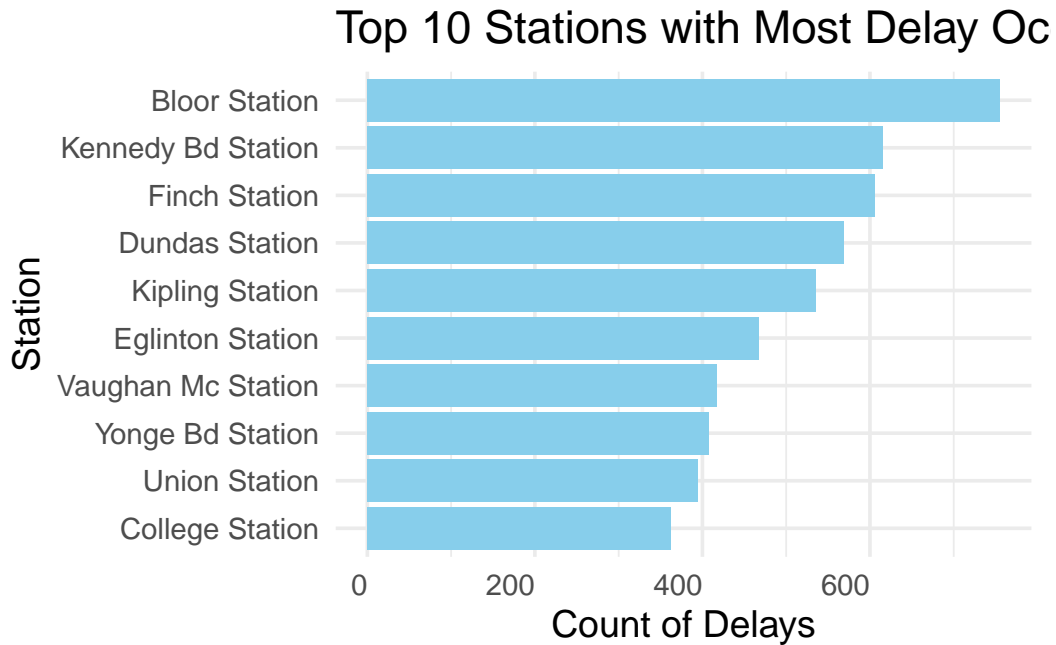


Figure 2: Top 10 stations with the most delay occurrences in the TTC subway system in 2023.

The Figure 2 shows the top 10 stations in the Toronto Transit Commission (TTC) subway system experiencing the highest number of delays in 2023. The bar chart shows that Bloor Station has the highest number of delays at over 800. Other high-traffic stations, such as Kennedy BD Station and Finch Station, followed closely behind, with more than 700 delays.

These stations are all important transportation hubs in the system and are likely to experience delays due to high passenger volumes, complex operations, and frequent train movements. As a key interchange station for multiple subway lines, Union Station ranked seventh with over 500 delays. This suggests that while Union Station is a major transit hub, it does not have the highest frequency of delays, contrary to what would be expected from a ridership level departure. Stations such as Kipling Station and Dundas Station are terminal or transfer points. The concentration of delays at these stations suggests that operational inefficiencies may be due to train scheduling, maintenance programs, or platform congestion issues.

The pattern at the first 10 stations highlights how delays are disproportionately distributed across the Metro network. While delays are not confined to a single line, they are concentrated at major interchanges or terminals, and the delayed stations are the most complex in terms of operations and passenger flow. This CASE allows for a more rational allocation of resources and maintenance planning or adjustments. Targeting improvements to high delay stations can significantly improve overall service reliability and reduce delays across the system.

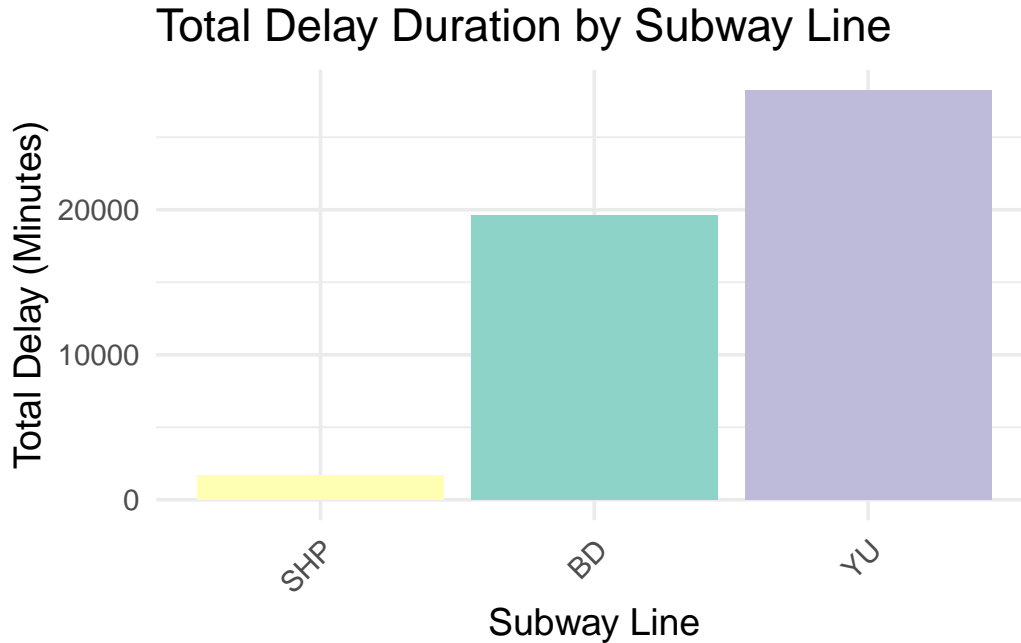


Figure 3: Total delay duration (in minutes) for each TTC subway line in 2023.

Figure 3 shows the total delays (in minutes) for each subway line in the Toronto Transit Commission (TTC) system in 2023. The data shows that the Bloor-Danforth Line (Line 2) has the most extended cumulative delays, totaling over 30,000 minutes. It is closely followed by the Yonge-University line (Line 1), which has considerable delays but slightly less than Line 2. These two lines are Toronto’s main bus routes, connecting major transit corridors and almost exclusively the city’s most heavily traveled areas. In contrast, the Scarborough Line (Line 3) and Sheppard Line (Line 4) have much shorter total delays. The shorter routes and lower ridership on these lines compared to Lines 1 and 2 may account for their fewer delays. These lines have less frequent service and fewer transfer points. As a result, there are fewer operational disruptions.

The data shows significant variation in the length of delays across the metro network, with the highest number of delays concentrated on the busiest and most complex lines (Lines 1 and 2). This suggests that these significant lines have more pronounced operational inefficiencies and maintenance issues. These lines may be more stressed due to frequent service demands and greater ridership. In contrast, smaller lines have fewer operational disruptions, likely due to their more straightforward routes and lower ridership.

Finding which routes impact total delay rates remarkably allows the TTC to take a more targeted approach to addressing operational inefficiencies. The TTC could prioritize maintenance and resource allocation on the Bloor-Danforth and Wynn-University routes, which could significantly reduce delays and, ultimately, improve the efficiency of the overall transit system.

The TTC could also prioritize maintenance and resource allocation on the Wynn-Danforth and Wynn-University routes. The study results indicate that focusing on these high-delay lines would significantly increase service reliability and improve the rider experience.

3 Discussion

As shown in Figure 1, the TTC subway system has experienced significant fluctuations in delay duration throughout 2023. The most notable peak occurs in June when the average delay duration reaches approximately 3.5 minutes. This shows that operational challenges and external disturbances require more attention during this period. Towards the end of the year, the delays decreased, indicating improved system performance, which could be due to better management practices or reduced external interference.

Additionally, Figure 2 shows a disproportionate number of delays at Union and Kipling stations, which could be attributed to the high volume of passengers and complex operational demands. As significant hubs in the Metro network, these two stations have a high frequency of delays and, therefore, require targeted interventions to alleviate congestion and reduce operational efficiency.

Finally, Figure 3 shows that the Bloor-Danforth Line (Line 2) and the Yonge-University Line (Line 1) impact system-wide delays the most. These two lines form the TTC's main routes and connect most of the city's riders. Therefore, given the system-wide impact of these two heavily traveled lines, the TTC should prioritize improving its operations. The analysis in this paper emphasizes the importance of addressing these key stations and lines to reduce overall delays and enhance the rider experience.

A Appendix

A.1 Dataset and Graph Sketches

The dataset used in this analysis comes from the OpenDataToronto portal(Gelfand 2022), specifically focusing on the Toronto Transit Commission (TTC) subway delay data. This dataset contains detailed records of delay events, including the station, time, date, delay caused, and delay duration. The graphs generated in this analysis include visual representations of delay patterns over time, the frequency of delays by station, and the total delays by subway line, which help provide insights into the operational challenges faced by the TTC in 2023. All visualizations were produced using the ggplot2 library(**citetggplot2?**) in R(R Core Team 2023).

A.2 Data Cleaning

The process of data cleaning involved filtering the relevant data for the TTC subway delays. The dataset was cleaned by renaming some columns to make them more readable, removing unnecessary columns, and handling missing values. Therefore, the cleaned dataset is used for analysis, ensuring accuracy and clarity in the visualizations and statistical evaluations.

A.3 Attribution Statement

This research uses the TTC subway delay data which is licensed under the Open Government Licence of Toronto(**tplicense?**).It available at OpenDataToronto(Gelfand 2022). All data and visualizations generated in this analysis are fully compliant with this license.

References

- Alexander, Rohan. 2023. *Telling Stories with Data*. Boca Raton: CRC Press. <https://tellingstorieswithdata.com/>.
- Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- Louie, Jacob, Amer Shalaby, and Khandker Nurul Habib. 2017. “Modelling the Impact of Causal and Non-Causal Factors on Disruption Duration for Toronto’s Subway System: An Exploratory Investigation Using Hazard Modelling.” *Accident Analysis & Prevention* 98: 232–40.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Schleifer, Alyssa. 2022. “Toronto Subway Delays Vary Based on Specific Ridership Patterns.”
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.