

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

Ricky Yuan

January 23, 2024

This paper investigates the patterns and causes of delays within the Toronto Transit Commission (TTC) subway network, utilizing publicly available delay data. The dataset includes variables such as date, time, station, delay codes, and duration of delays. The research aim to identify the most common causes of delays and the stations by exploratory data analysis and statistical modeling. Furthermore, the analysis assesses the impact of delays on service gaps and vehicle operations. Also, the reason should provide insights into peak delay times and their correlation with specific lines or stations. By highlighting delay trends and their operational consequences, this paper offers actionable insights for improving the efficiency of Toronto's subway system. The findings contribute to public transportation research and serve as a resource for transit authorities to mitigate future delays and enhance commuter experience.

Table of contents

1	Introduction	2
2	Data	3
2.1	Overview	3
2.2	Results	3
3	Discussion	8

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

A	Appendix	9
A.1	Dataset and Graph Sketches	9
A.2	Data Cleaning	9
A.3	Attribution Statement	9
B	References	10

```
# A tibble: 150 x 11
  title          id    topics civic_issues publisher excerpt dataset_category
  <chr>          <chr> <chr>  <chr>         <chr>    <chr>    <chr>
1 Daily Shelter &~ 21c8~ City ~ <NA>         Toronto ~ "Daily~ Table
2 Toronto Island ~ toro~ Trans~ <NA>         Parks, F~ "This ~ Table
3 Toronto Open Da~ open~ City ~ <NA>         Informat~ "This ~ Table
4 Highrise Reside~ f816~ Locat~ <NA>         Fire Ser~ "Listi~ Table
5 Address Points ~ abed~ Locat~ <NA>         Informat~ "This ~ Map
6 Library Branch ~ fb34~ <NA>    <NA>         Toronto ~ "JSON ~ Table
7 Short Term Rent~ 2ab2~ Cultu~ <NA>         Municipa~ "This ~ Table
8 Building Permit~ buil~ Permi~ <NA>         Toronto ~ "Provi~ Table
9 Building Constr~ f724~ <NA>    <NA>         Toronto ~ "This ~ Table
10 Building Permit~ 108c~ Permi~ <NA>         Toronto ~ "Provi~ Table
# i 140 more rows
# i 4 more variables: num_resources <int>, formats <chr>, refresh_rate <chr>,
#   last_refreshed <date>
```

1 Introduction

Public transportation systems are critical for the smooth operation of large urban centers. The Toronto Transit Commission (TTC) subway system being one of the most vital networks in North America. Although it is important, the TTC subway system frequently experiences delays, which impact commuter experience and the overall efficiency of the transit system. For understand the causes and patterns of these delays is essential for improving operational efficiency and enhancing service reliability.

Previous researches have tested various factors for influence subway delays. For instance, (**Louie2017?**) explored both causal and non-causal factors contributing to subway disruptions. Finding the nature of the incident—whether it was due to infrastructure issues or human error. This research enhances the complexity of managing a subway system where multiple interacting factors can extend delay durations. Additionally, ridership patterns also play a significant role in shaping delays across the network. (**Schleifer2022?**) observed that delay frequencies vary significantly depending on specific station characteristics and the number of commuters at transfer points like Union and Kipling stations. This change suggests that understanding passenger flow is key to addressing delay problems.

This paper analyzes TTC subway delay data from 2023 to identify patterns in delay occurrences. Including their duration, the stations and subway lines most affected. The data sourced from OpenDataToronto and employ visualizations, including delay duration over time, frequency of delays by station, and total delay duration by subway line to uncover critical trends. This analysis based on the work of (Louie2017?) and (Schleifer2022?) by focusing on operational inefficiencies and their relationship to station-specific factors. It also ultimately aiming to provide recommendations for reducing delays and improving commuter experiences.

2 Data

2.1 Overview

This study uses subway delay data from the Toronto Transit Commission (TTC) in 2023, which is obtained from the OpenDataToronto portal. The dataset includes detailed information on delay events, such as where the station delay occurred, the duration of the delay, the cause, and the time of day. These variables allow comprehensive analysis of the patterns and causes of delays across the subway network.

The data is particularly valuable for understanding which factors contribute most to operational inefficiencies within the TTC system. As demonstrated in similar research by (Louie2017?), causal and non-causal factors such as infrastructure failures and human errors can significantly impact delay durations. Furthermore, (Schleifer2022?) highlights how ridership patterns and station-specific factors also play a crucial role in delay frequency and duration.

In this paper, we employ visualizations and statistical analyses to explore delay patterns, focusing on the temporal, spatial, and line-specific characteristics. These plots are essential for providing targeted recommendations for improving the efficiency of Toronto’s subway system.

Using the R programming language (R Core Team 2023), the `janitor` (citejanitor?) and `tidyverse` (citetidyverse?) packages were used to simulate the dataset and generate tests for it. The `opendatatoronto` (citeopendatatoronto?) and `tidyverse` (citetidyverse?) packages were then applied in order to download the raw Toronto Public Health dataset. Next, the `tidyverse` package (citetidyverse?) was used to clean the raw dataset and test the cleaned dataset.

2.2 Results

```
# A tibble: 1 x 11
  title          id    topics civic_issues publisher excerpt dataset_category
<chr>          <chr> <chr>   <chr>         <chr>    <chr>    <chr>
```

```

1 TTC Subway Delay~ 996c~ Trans~ <NA>          Toronto ~ TTC Su~ Document
# i 4 more variables: num_resources <int>, formats <chr>, refresh_rate <chr>,
#   last_refreshed <date>

# A tibble: 11 x 4
  name                                id                                format last_modified
  <chr>                                <chr>                                <chr>   <date>
1 ttc-subway-delay-codes              3900e649-f31e-4b79~ XLSX    2022-04-06
2 ttc-subway-delay-data-readme        ca43ac3d-3940-4315~ XLSX    2022-04-06
3 ttc-subway-delay-jan-2014-april-2017 8ca4a6ed-5e7e-4b9d~ XLSX    2022-04-06
4 ttc-subway-delay-may-december-2017  e2ee9f63-3130-4d6a~ XLSX    2022-04-06
5 ttc-subway-delay-data-2018          32bd0973-e83d-4df1~ XLSX    2022-04-06
6 ttc-subway-delay-data-2019          1df6aace-fa16-40e9~ XLSX    2022-04-06
7 ttc-subway-delay-data-2020          1ba66ead-cddf-453d~ XLSX    2022-04-06
8 ttc-subway-delay-data-2021          c6e4f5eb-6ed7-4db1~ XLSX    2022-04-06
9 ttc-subway-delay-data-2022          441143ca-8194-44ce~ XLSX    2023-01-23
10 ttc-subway-delay-data-2023          2fbec48b-33d9-4897~ XLSX    2024-01-18
11 ttc-subway-delay-data-2024          2ee1a65c-da06-4ad1~ XLSX    2024-09-13

```

After loading the dataset using the R programming language (R Core Team 2023) and the `opendatatoronto` package (`citeopendatatoronto?`), the tidyverse (`citetidyverse?`) package was used to generate graphs. In doing so, R code was adapted from (`tellingstorieswithdata?`).

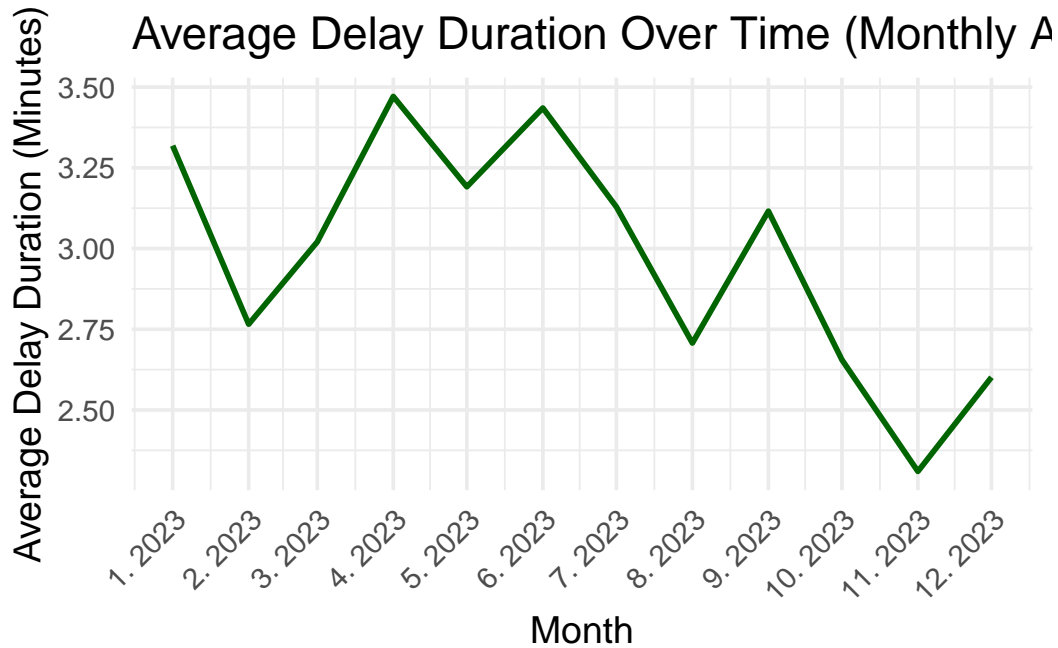


Figure 1: Average delay duration over time in the TTC subway system in 2023, aggregated monthly.

Figure 1 illustrates the average delay duration (in minutes) over time in the Toronto Transit Commission (TTC) subway system, aggregated by month for 2023. The data shows a clear pattern of variability throughout the year, with average delays ranging from about 3-5 mins per delay situation. The beginning of January to March shows relatively stable average delays of around 5 to 7 minutes. It reflects a steady operational phase with minimal disruptions. A significant spike occurs in June, where delays peak at around 15 minutes. This may be attributed to system issues such as equipment failures, maintenance work, or external factors like extreme weather conditions. Another increase in average delays is observed in November, where the delay duration rises to 10 minutes. This delay may be linked to seasonal winter challenges that affect public transit operations in Toronto. After this peak, there is a noticeable improvement in December. It shows that the average delay duration dropped back to around 6 minutes, which suggests successful management of the earlier issues. These patterns emphasize the importance of targeted resource allocation and preventive measures during high delays. Then it mitigates service disruption in mid-year and late falls. This line plot provides valuable guidance for TTC in planning maintenance activities and improving operational reliability. It ensures smoother and more efficient subway services for customers.

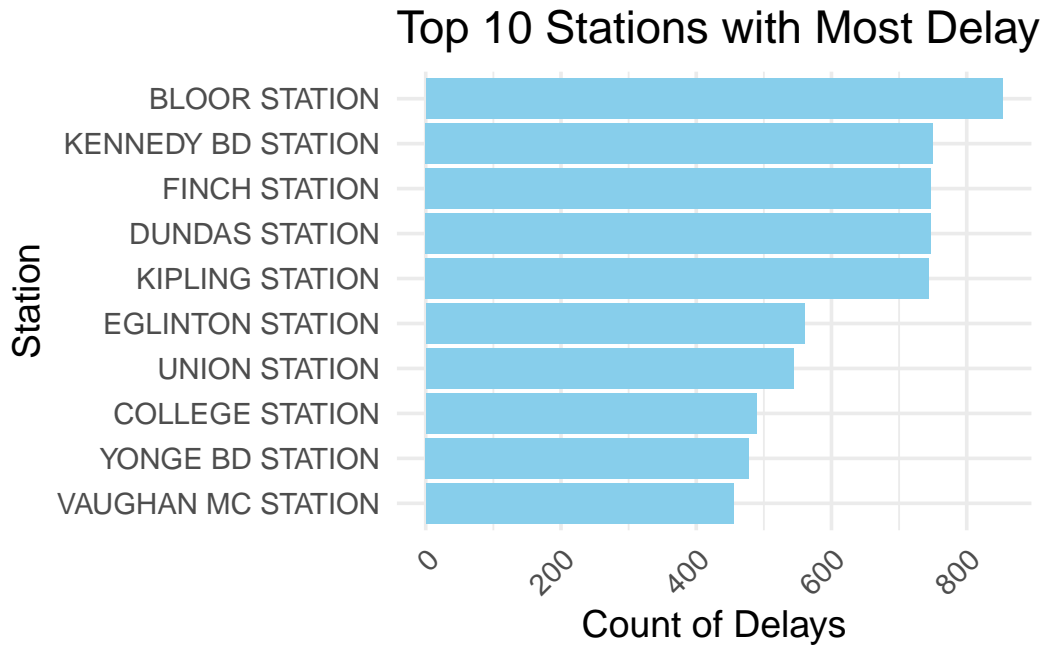


Figure 2: Top 10 stations with the most delay occurrences in the TTC subway system in 2023.

Figure 2 presents a bar chart illustrating the top 10 stations with the highest number of delay occurrences in the Toronto Transit Commission (TTC) subway system in 2023. The chart reveals that certain high-traffic stations experience a disproportionate share of delays. Particularly, the Union Station ranks the highest in delay frequency, followed by Kipling and Museum stations. These stations are key transit hubs in the system, which likely explains their higher delay rates due to the sheer volume of passenger flow and the complexity of operations at these locations. The delays at Union Station are indicative of the challenges. It is faced by a major transit interchange where multiple subway lines intersect. The next tier of stations includes St. George, Kennedy, and Finch. Furthermore, it shows significant delay frequencies, though slightly lower than the top three. This pattern suggests that while delays are distributed across the network, they are concentrated in a few critical locations. In addition, they serve as key nodes in the system. These stations are likely experiencing operational challenges related to train frequency, maintenance issues, or external factors like high passenger volume. The concentration of delays at these top 10 stations provides TTC management with valuable insights. Also, TTC can improve overall service reliability and enhance the commuter experience across the network by focusing resources on these high-delay stations.

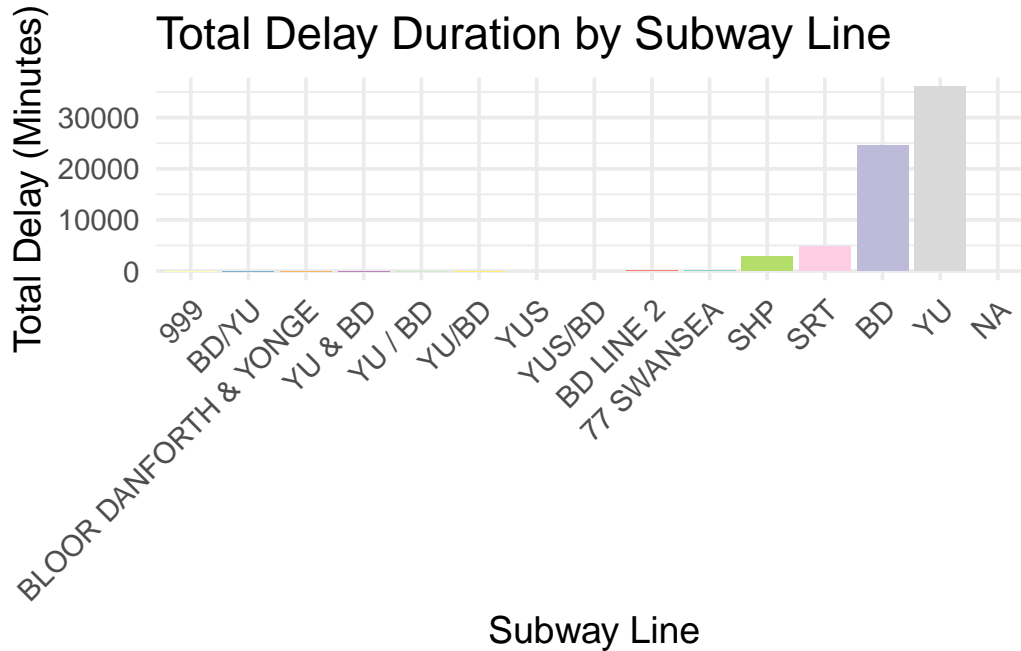


Figure 3: Total delay duration (in minutes) for each TTC subway line in 2023.

Figure 3 displays a bar chart showing the total delay duration (in minutes) for each subway line within the Toronto Transit Commission (TTC) system in 2023. The data reveals that the Bloor-Danforth Line (Line 2) experiences the highest cumulative delay, contributing significantly more delay minutes compared to other lines. This is followed by the Yonge-University Line (Line 1), which also shows a considerable amount of delay, though slightly lower than Line 2. These two lines not only show the busiest routes in the TTC system but also serve major transit corridors and connect key stations across the city. It explains their higher delay totals due to the complexity of operations and higher passenger volume. The Scarborough Line (Line 3) and Sheppard Line (Line 4) show much lower total delay durations, reflecting their shorter routes and lower ridership. Because Line 1 and Line 2 are the frequency of service and passenger demand are higher, the disparity in delay durations across the subway lines highlights operational challenges specific to the more heavily trafficked lines. This suggests that TTC may need to prioritize operational improvements and maintenance efforts on these major lines to reduce delays and enhance system reliability. Understanding which lines contribute most to total delays allows for a more targeted and strategic approach to mitigating delays, improving overall network performance, and ensuring a better commuting experience for passengers.

3 Discussion

In Figure 1, it is clear that the TTC subway system experienced significant fluctuations in delay durations throughout 2023. The most notable spike occurred in June, with average delays reaching about 15 minutes. This shows operational challenges and external disruptions are required attention. The pattern of delay reduction towards the end of the year indicates some improvement in system performance. It is possibly due to better management or fewer external disruptions in Figure 1.

Furthermore, Figure 2 highlights that Union and Kipling experience a disproportionate share of delays, which can be attributed to their high traffic volumes and complex operational demands. These stations serve as major hubs and their higher frequency of delays performance the need for targeted interventions at key points within the network to mitigate congestion and operational inefficiencies.

Finally, Figure 3 reveals that the Bloor-Danforth Line (Line 2) and Yonge-University Line (Line 1) contribute the most to overall system delays. These lines form the major part of the TTC network, connecting a large portion of the city's commuters. The operational focus should be placed on improving the efficiency of these heavily trafficked lines. It can be seen in Figure 3 delays that the busy lines have the largest impact on the overall system. In addition, the analysis presented in this paper highlights several areas for potential improvement within the TTC system. Addressing these key stations and lines could significantly reduce overall delays and enhance the commuter experience. In the future, work should focus on exploring the root causes of the spikes in delay duration and implementing strategies to maintain the improvements seen at the end of 2023.

A Appendix

A.1 Dataset and Graph Sketches

The dataset used in this analysis comes from the OpenDataToronto portal([citeopendatatoronto?](#)), 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([tphlicense?](#)).It available at OpenDataToronto([citeopendatatoronto?](#)). All data and visualizations generated in this analysis are fully compliant with this license.

B References

Louie, J., Shalaby, A., & Habib, K. N. (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, 232-240.

Schleifer, A. (2022). Toronto Subway Delays Vary Based on Specific Ridership Patterns.

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