An Analysis of Subway Delays in Toronto: Insights from 2023 TTC Data*

Diving deep into subway delays

Charlie Zhang

September 25, 2024

This report presents an investigation into subway delays in Toronto using data from the Toronto Transit Commission (TTC) for the year 2023. The analysis examines patterns of delays across various subway stations, lines, and days of the week, with the goal of identifying critical factors that contribute to service disruptions. Key findings show that Monday and Thursday experienced the highest number of delays, with the Bloor-Danforth (BD) line being the most affected. These insights provide a data-driven foundation for improving operational efficiency and minimizing service interruptions in the future.

1 Introduction

Public transit plays a vital role in the daily lives of millions of people in urban centers around the world. In Toronto, the subway system, operated by the Toronto Transit Commission (TTC), serves as a backbone for many commuters. However, delays in subway services can cause significant inconvenience and disrupt the city's flow of traffic. In 2023, Toronto's subway system faced numerous delays, which are the focus of this report.

This analysis uses a comprehensive dataset of subway delays from the TTC to investigate the nature and extent of service disruptions. Specifically, the report aims to answer the following questions:

- Which days of the week are most prone to delays?
- What are the most affected subway lines and stations?
- What times of day experience the most delays?

---

*Code and data are available at: https://github.com/Zqyyk11/Term-Paper-1.

By answering these questions, the report seeks to provide actionable insights for improving the reliability of Toronto's subway system, reducing the frequency and duration of delays, and enhancing the overall commuting experience.

The findings in this report are based on data-driven analyses and visualizations, which highlight delay trends throughout the year. This investigation serves as a foundation for further research into the causes of delays and offers guidance for policy makers and transit authorities in addressing key operational challenges.

2 Data

The data utilized in this report is sourced from Open Data Toronto and was imported using the opendatatoronto library (Gelfand 2022). The specific dataset used for analyzing TTC subway delays in Toronto is the "TTC Subway & SRT Train Service Delay Data" dataset [Toronto (2024). All data analysis was conducted using R (R Core Team 2023), with the assistance of several key packages, including tidyverse (Wickham et al. 2019), dplyr (Wickham 2023),here (Müller and Wickham 2020) janitor (Firke 2023), ggplot2 (Wickham 2016), and knitr (Xie 2014).

To transform the raw data into a clean, analysis-ready format, several preprocessing steps were carried out using R. First, unnecessary columns that did not contribute to the analysis, such as metadata or irrelevant identifiers, were removed to streamline the dataset. Using the dplyr package, specific columns were selected based on relevance to subway delays, such as date, time, station, line, and delay duration. The data was further cleaned by handling any missing or inconsistent values with the janitor package to ensure data integrity. Any outliers or erroneous entries were also identified and either corrected or removed as appropriate. These steps allowed for the creation of a clean dataset, which provided a solid foundation for the subsequent analysis. The entire process was conducted through efficient R code, leveraging the functionality of packages like dplyr for filtering and selection, and janitor for data cleaning. \ \

Total Delays by Day of the Week (in minutes)
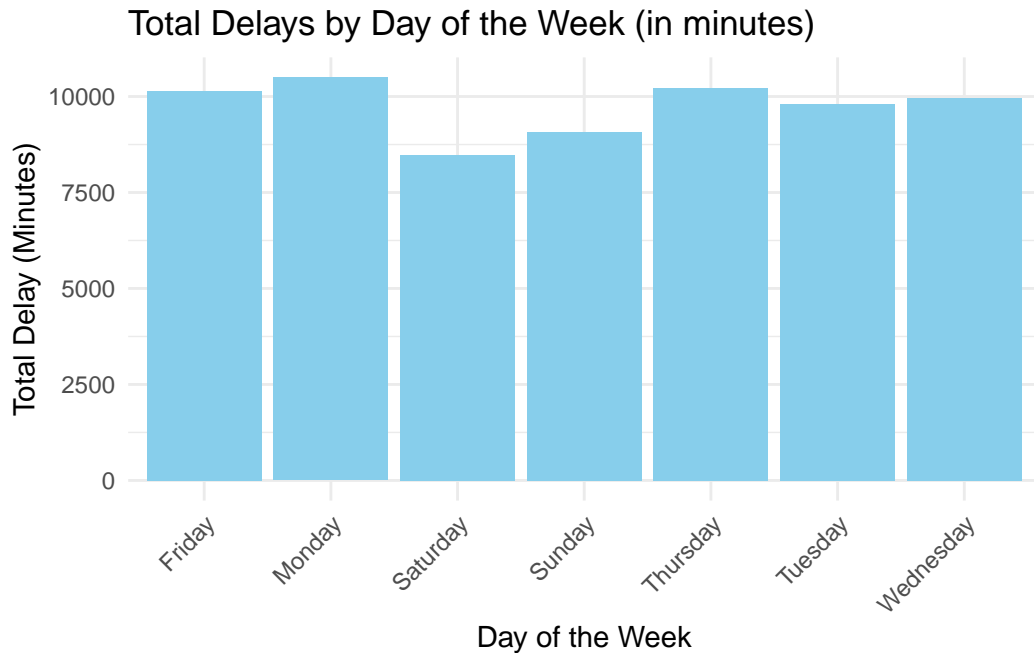

The bar chart shows the total minutes of delays across different days of the week, with Mondays experiencing the highest delay times at 10,494 minutes, followed closely by Thursday and Friday. These days likely face more operational challenges due to higher ridership at the start and end of the workweek. In contrast, weekends see fewer delays, possibly because of lower ridership and reduced service frequency, leading to less strain on the subway system. \ \

Total Delays by Subway Line (in minutes)
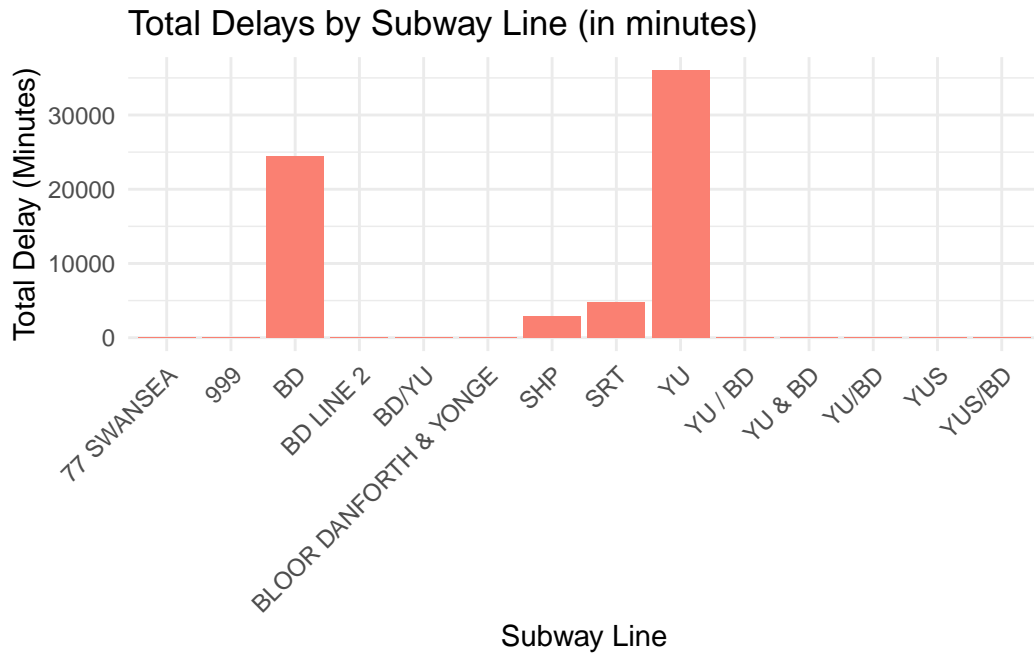
The bar chart shows the total delay minutes per subway line, with the Yonge-University (YU) line experiencing the most delays, followed by the Bloor-Danforth (BD) line. The high delay times on the YU Line are likely due to its status as one of the busiest lines, serving downtown Toronto and major commuter hubs. The BD Line also faces substantial delays, reflecting its heavy operational load. In contrast, other lines such as the Sheppard (SRT) experience significantly fewer delays, likely due to lower ridership or reduced service frequency. \ \

## Total Delays by Top 10 Stations (in minutes)
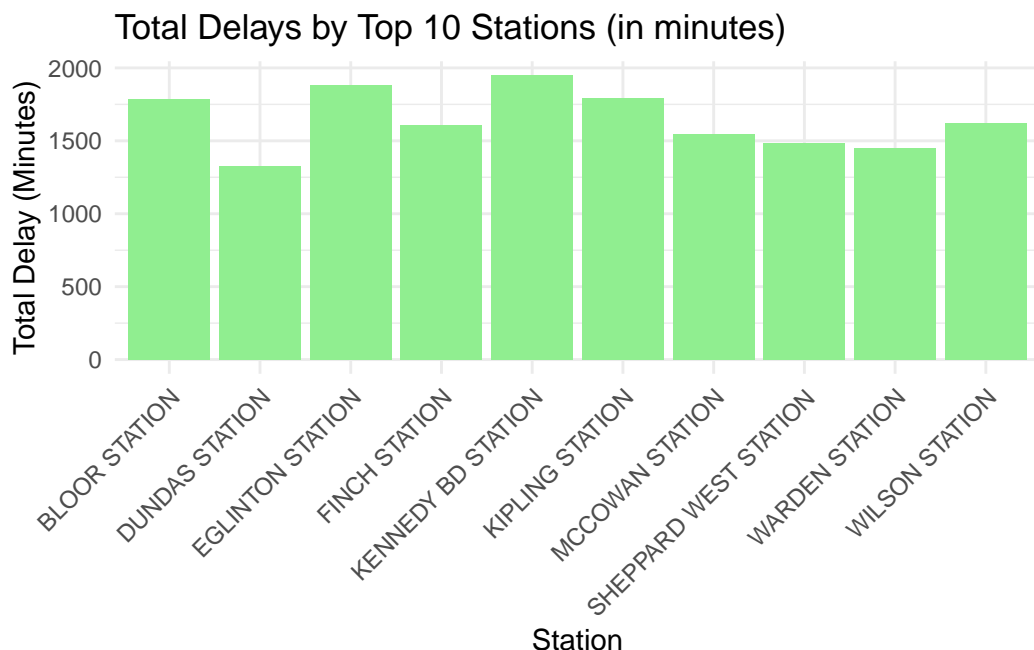
The bar chart highlights the top 10 stations with the highest delay minutes, with Kennedy BD Station leading, followed by Eglinton and Kipling stations. Kennedy, as a major terminal on the Bloor-Danforth line, likely faces operational challenges contributing to its delays. Eglinton Station, undergoing significant construction due to transit expansions, also experiences high delays. Key transfer hubs like Kipling and Bloor, which handle large passenger volumes, further contribute to increased delays at these stations.

3 Results

The analysis of subway delays in Toronto for 2023 reveals significant patterns across days of the week, subway lines, and specific stations. Mondays experienced the most delays, followed closely by Thursday and Friday, suggesting that the beginning and end of the workweek are the most challenging periods for subway operations. This pattern may be linked to increased ridership and operational pressures during these days. In terms of subway lines, the Yonge-University (YU) line recorded the highest delay minutes, likely due to its extensive coverage of key commuter areas, including downtown Toronto, where traffic and operational complexities

are higher. The Bloor-Danforth (BD) line also saw substantial delays, reflecting its heavy usage and operational load. Meanwhile, other lines, such as the Sheppard (SRT), experienced fewer delays, which can be attributed to lower ridership and less frequent service. At the station level, Kennedy BD Station led in delay minutes, followed by Eglinton and Kipling stations, all of which are either major transfer hubs or subject to ongoing construction, leading to operational disruptions. The results highlight the need for focused operational improvements at key lines and stations, particularly during peak periods at the start and end of the workweek. \
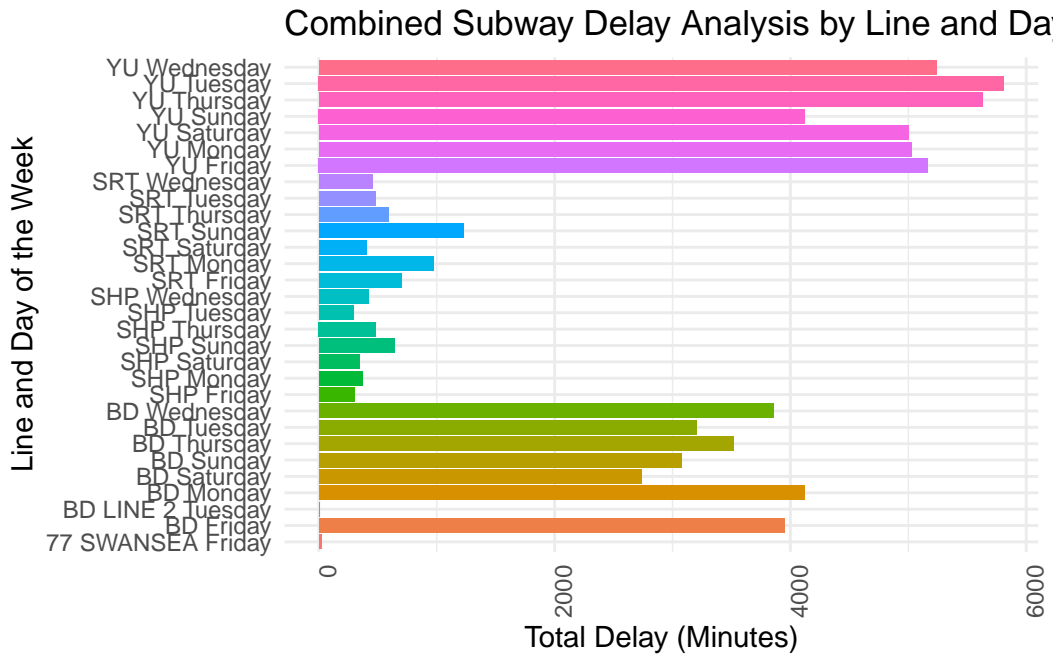
4 Discussion

The analysis of subway delays in Toronto for 2023 reveals several important insights into the performance of the city's subway system. Delays are more pronounced during the workweek, particularly on Mondays and Thursdays, suggesting that the operational strain is highest at the start and end of the week, likely due to increased commuter traffic. The Yonge-University (YU) and Bloor-Danforth (BD) lines experience the most delays, reflecting their roles as major transit arteries with high passenger volumes and frequent service demands. Stations like Kennedy and Eglinton, which are key transfer points and hubs of construction activity, are more prone to delays, pointing to infrastructure and operational challenges that exacerbate service interruptions.

The removal of zero-delay entries further emphasizes that certain stations and lines consistently face challenges, while others see minimal delays. This highlights potential areas for targeted improvements, particularly on high-traffic lines and during peak commuting days. To reduce

delays, strategic investments in infrastructure and more efficient scheduling could mitigate the operational pressures observed on busy lines and at critical stations. In addition, proactive measures, such as real-time monitoring and passenger load management, could further improve service reliability. Overall, the data points to a need for a more flexible and robust transit system capable of handling fluctuations in demand while minimizing disruptions.

References

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

Gelfand, A. 2022. *Opendatatoronto: Access Toronto Open Data.* https://cran.r-project.org/package=opendatatoronto.

Müller, Kirill, and Hadley Wickham. 2020. *Here: A Simpler Way to Find Your Files.* https://CRAN.R-project.org/package=here.

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

Toronto. 2024. *Non Regulated Lead Sample.* https://open.toronto.ca/.

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

———. 2023. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Mara Averick, Jennifer Bryan, et al. 2019. "Welcome to the Tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.