

What’s Delaying Toronto’s Subways?*

Revealing Peak Hour Congestion, Infrastructure Failures, and High-traffic Hubs
as Major Contributors

Weiyang Li

September 29, 2024

This paper explores the underlying patterns of subway delays in Toronto, finding that subway delays are more frequent during peak hours and tend to be concentrated around major stations and high-demand subway lines. Notably, while most delays are brief, the data also highlights some extreme cases of disruptions lasting several hours. The analysis shows that facility failures and extreme weather can lead to longer delays. Targeted improvements to key stations and infrastructure could significantly reduce overall delay times and improve service reliability.

1 Introduction

Urban public transportation systems are an important part of urban construction and development, connecting millions of commuters to their daily destinations and underpinning the vitality of city life. Compared to traditional bus systems, subway systems are emerging as one of the most efficient transportation networks, thanks to advantages such as shorter commute times across regions, relief from surface traffic, and independence from the weather. Yet, even in cities like Toronto, subway systems face frequent delays. Based on feedback from a random sample of 1,936 transit users surveyed by Forum Poll, about one-third of TTC (Toronto Transit Commission) regulars report being dissatisfied with its service, with almost half of riders mentioning that they often experience unexpected delays while riding the TTC (Forum Research 2019).

Recent analyses show that the TTC’s ridership recovery after the pandemic remains only at 70% of pre-pandemic levels, exacerbating financial strain and leading to increased delays, particularly during peak hours (Matt Elliott 2023). This “transportation death spiral” – reduced ridership leading to fewer resources and thus more delays – has placed the system in a critical

*Code and data are available at: <https://github.com/doravmony/TTC-Delay>

state, threatening the reliability of the TTC network (Matt Elliott 2023). Therefore, an in-depth understanding of the causes and patterns of delays is important for improving service reliability. By determining which factors cause more frequent or longer delays, and whether certain stations or lines are more prone to disruptions, urban transit managers can target operational improvements to reduce unnecessary delays and improve the overall passenger experience.

Although previous research has identified commuter dissatisfaction and financial strain related to public transit delays (Matt Elliott 2023), there has been little rigorous analysis to determine specific patterns in the duration and frequency of these delays. To address this gap, we conducted an in-depth analysis of the 2024 TTC subway delay dataset, focusing on delay trends at the station and line level, as well as across different dates. Through exploratory data analysis and visualization, our study examines which stations and lines experience the most delays, the times of day when issues are most severe, and the primary causes behind the longest service disruptions. Our findings reveal that transfer stations and high-traffic lines are disproportionately affected by delays, and that disruptions caused by mechanical failures tend to last significantly longer than those caused by other factors. These insights provide a foundation for policy changes, resource allocation, and targeted interventions to improve the reliability of TTC services.

In this paper, we analyze the 2024 TTC subway delay dataset, made publicly available through Open Data Toronto, to examine the frequency, duration, and causes of delays across various subway stations and lines. Our analysis investigates patterns based on factors such as time of day, location, and specific causes of delays. We also focus on identifying which stations or lines are more prone to disruptions and what factors contribute to longer or more frequent delays.

The remainder of the paper is structured as follows: Section 2 introduces the raw dataset and describes the cleaned datasets, along with a preliminary analysis through numerical summaries and visualizations. Section 3 explores key findings from our analysis. Lastly, Section 4 addresses the limitations of the analysis and offers recommendations for reducing delays and improving subway service reliability.

2 Data

2.1 Raw Data

The dataset used for this analysis is the 2024 TTC Subway Delay Data, accessed from the Open Data Toronto portal and read into this paper via the `opendatatatoronto` library (Sharla Gelfand 2022). This dataset, released by the Toronto Transit Commission, comprehensively documents 17,517 subway delays in Toronto’s public transit system during the year 2024 (City of Toronto 2024). Each entry in the dataset documents a delay event. Data is updated monthly and is as of August 31, 2024 for use in this paper. All the data analysis was conducted using R

(R Core Team 2023) with the support of the following packages: `tidyverse` (Hadley Wickham and Mara Averick and Jennifer Bryan and Winston Chang and Lucy D’Agostino McGowan and Romain François and Garrett Golemund and Alex Hayes and Lionel Henry and Jim Hester and Max Kuhn and Thomas Lin Pedersen and Evan Miller and Stephan Milton Bache and Kirill Müller and Jeroen Ooms and David Robinson and Dana Paige Seidel and Vitalie Spinu and Kohske Takahashi and Davis Vaughan and Claus Wilke and Kara Woo and Hiroaki Yutani 2019), `readxl` (Hadley Wickham and Jennifer Bryan 2023), `janitor` (Sam Firke 2023), `knitr` (Yihui Xie 2023), `kableExtra` (Hao Zhu 2024), and `gridExtra` (Baptiste Auguie 2017).

Subway delays stem from factors like mechanical failures, weather, or accidents. Each delay is logged by the TTC with details on location, time, and cause. This information forms the TTC subway delay dataset, where each entry is a specific event. In the data section, we track how real-world delays are recorded and transformed into structured data for analysis. In this way, the TTC dataset is well suited for our analysis and specificity allows for a thorough analysis of Toronto’s transportation challenges, whereas other datasets such as New York (MTA) or London (Underground) may not be as feasible. The essential details included in the dataset that allow for an in-depth examination of the temporal and spatial distribution of delays, as well as an exploration of the underlying causes contributing to service interruptions. The key variables are as follows:

Variable	Description
Date	The date when the delay occurred.
Time	The time when the delay occurred (24-hour format).
Station	The specific subway station affected by the delay.
Line	The subway line involved in the delay (e.g., Yonge-University).
Bound	The direction of travel (e.g., Northbound,).
Code	The cause of the delay, categorized into different codes (e.g., mechanical failure).
Min Delay	The duration of the delay in minutes.
Min Gap	The time gap between trains during the delay event, in minutes.
Vehicle	The train number associated with the delayed subway car.

Figure 1: Key Variables in the TTC Subway Delay Data

2.2 Cleaned Data

In this paper, we use two cleaned datasets for analysis. In the first dataset ([ttc dataset](#)), missing values in the station and line variables were removed, as the focus is on analyzing delays at specific stations and on particular lines. In the second dataset ([delay dataset](#)), we retained only records with actual delays, ensuring that our analysis is centered on significant delay events. Detailed data cleaning is included in the Appendix [A.1](#).

Statistic	Delay Duration	Time Gap
Min	0.00	0.00
1st Qu.	0.00	0.00
Median	0.00	0.00
Mean	2.84	4.20
3rd Qu.	4.00	8.00
Max	716.00	719.00
Standard Deviation	11.06	11.80
Variance	122.42	139.22

Figure 2: Summary Statistics for Delay Duration and Time Gap

We analyzed the cleaned TTC dataset containing complete records, showing an average delay of 2.84 minutes, with most delays being brief. Over 50% of events had no delay (median = 0), and 75% lasted under 4 minutes. Although most delays were minor, some were significantly longer, including a 716-minute outlier from an exceptional event. Upon closer inspection, it was determined that this outlier should not be discarded, as it resulted from an exceptional event rather than a data entry error.

The average train gap during delays was 4.20 minutes, with a median gap of 0. Most gaps were under 8 minutes, indicating minimal disruption to the subway schedule. The maximum gap recorded is 719 minutes, an extreme case tied to rare incidents causing major disruptions. However, 75% of gaps are under 8 minutes, indicating that while delays happen, they rarely cause significant service interruptions.

2.3 Exploratory Data Analysis

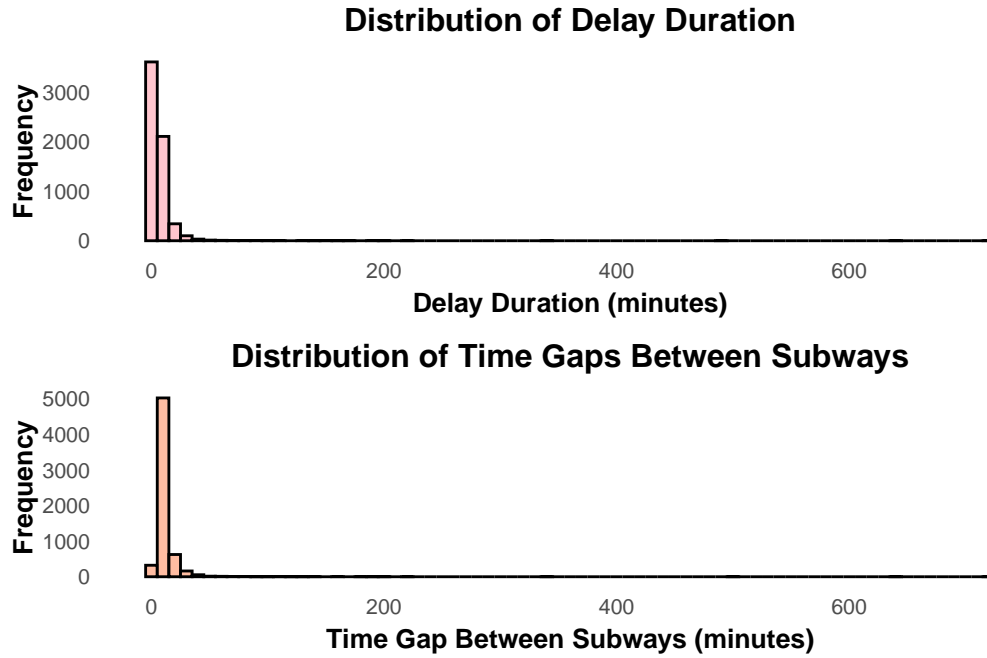


Figure 3: Delay Duration and Time Gap

Figure 3 for delay duration and time gaps for actual delay dataset align with our analysis from the Figure 2. The vast majority of delays and time gaps are concentrated around very short durations, with most data points clustered near small values. Both distributions exhibit extreme outliers, resulting in long right tails that reflect rare but significant disruptions. While these outlier events are infrequent, their magnitude is substantial, as highlighted by the extended tails of the distributions.

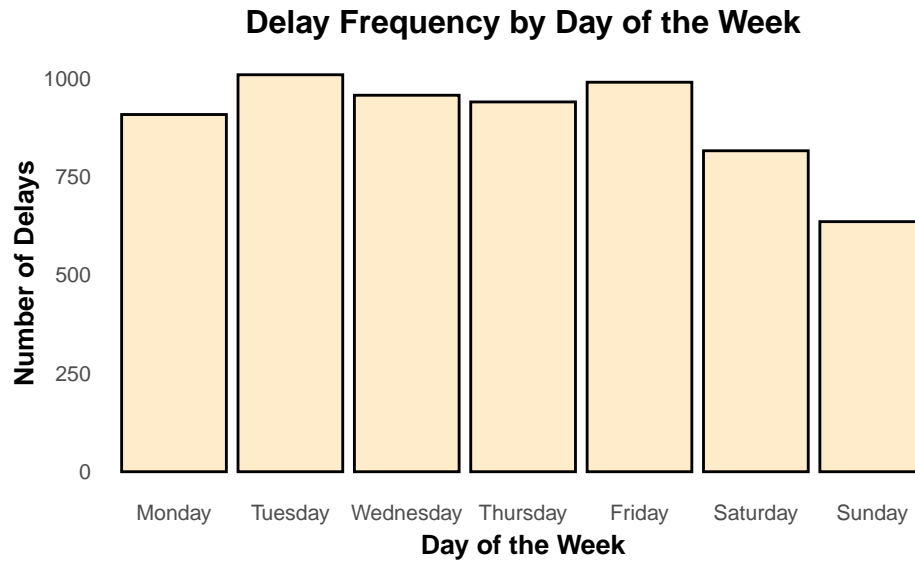


Figure 4: Delay Frequency by Day of the Week

Figure 4 illustrates the frequency of subway delays by day of the week, considering only delays where the duration exceeds zero. The data reveals that delays are most frequent on Tuesdays, followed closely by Fridays and Wednesdays, indicating a peak in delays during the middle of the workweek. In contrast, Sundays experience the fewest delays, likely due to reduced subway usage and service during weekends. Overall, the pattern suggests that the subway system experiences more frequent delays during weekdays, particularly in the middle of the week, when commuter traffic and operational demands are higher.

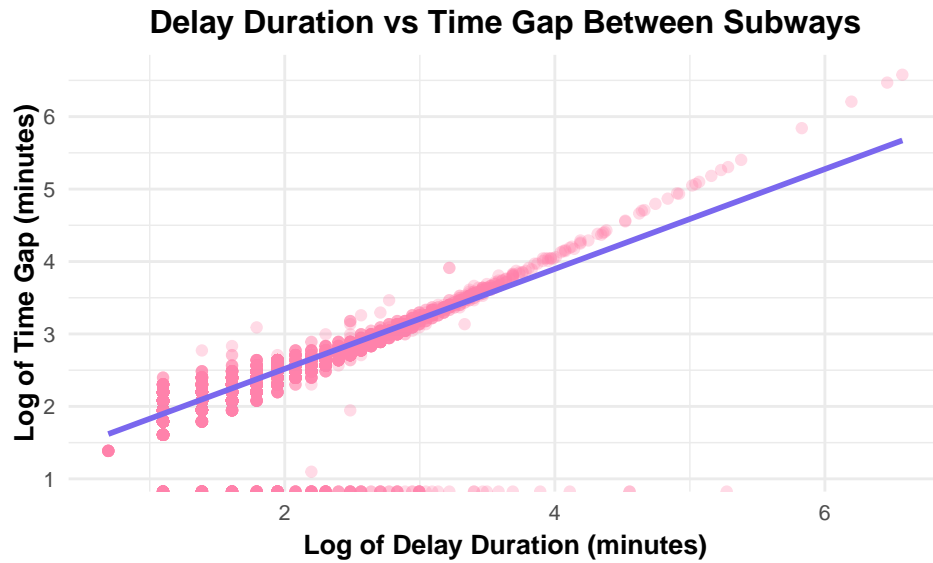


Figure 5: Relationship between Delay Duration and Time Gap Between Subways

Figure 5 shows a strong positive linear relationship between the logarithm of delay duration and the logarithm of the time gap between subways. As the delay duration increases, the time gap between trains also rises proportionally. Longer delays directly lead to larger gaps between subway services, likely causing significant disruptions. This relationship highlights the operational challenges of managing longer delays, as they not only prolong the delay duration but also increase the intervals between trains, contributing to greater service inefficiencies.

3 Results

3.1 Do longer and more frequent subway delays occur at specific times of the day?

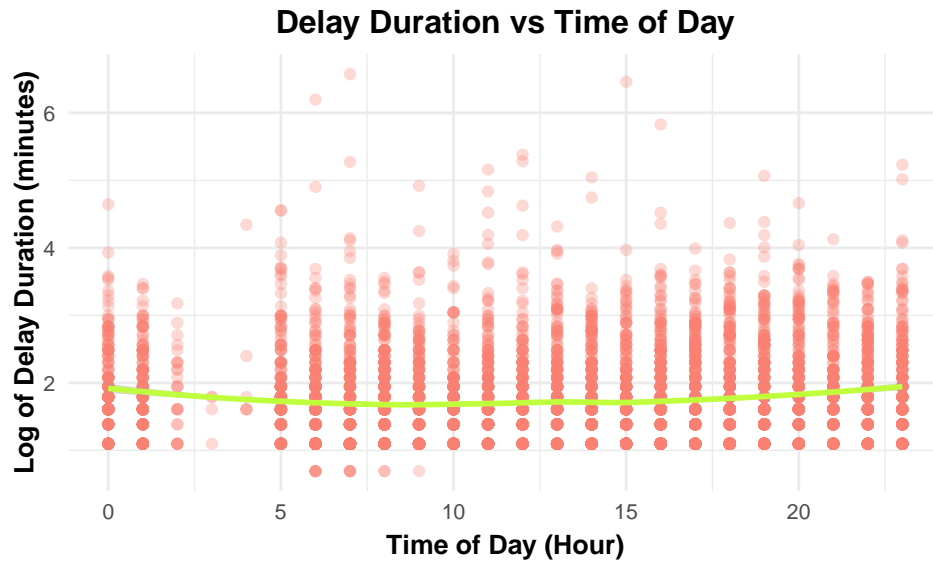


Figure 6: Relationship between Delay Duration and Time of Day

Figure 6 reveals that the majority of delays are clustered near zero, indicating that most recorded delays are relatively short and have minimal impact on service. These brief delays occur consistently throughout the day, with no significant variation during specific times, such as the morning or evening rush hours. However, the plot also highlights a few extreme outliers, where delay durations exceed 200 minutes, and in some cases, even 600 minutes. These outliers are scattered across different times of the day, suggesting that such long delays are likely caused by special or rare events, rather than being tied to a particular time of day.

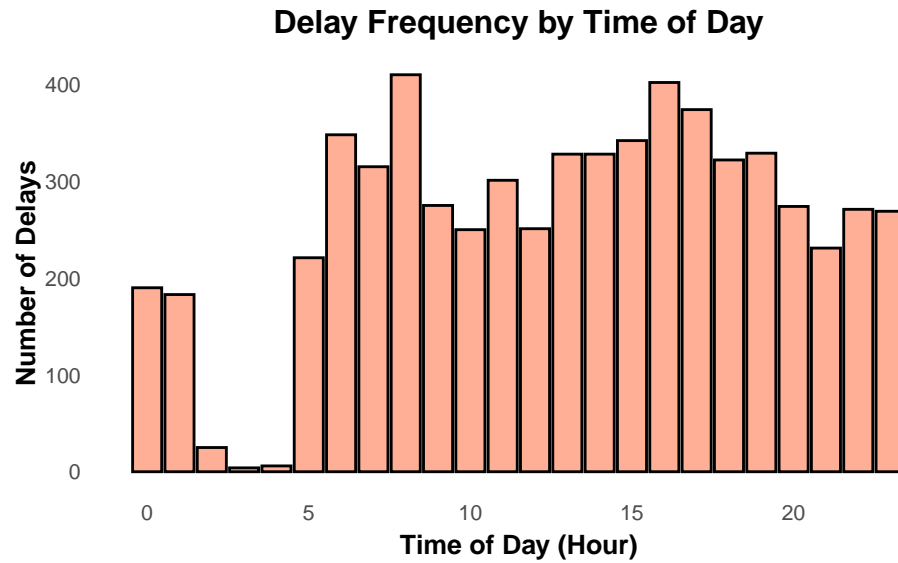


Figure 7: Delay Frequency by Time of Day

Figure 7 illustrates the frequency of delays across different hours of the day, filtered to include only events where the delay duration is greater than zero. The data shows that delays are concentrated during the busiest times of the day. Delays spike significantly from 7 AM to 10 AM, aligning with the morning rush hour, and again between 4 PM and 6 PM during the afternoon rush. In contrast, delays drop off sharply late at night and early in the morning, particularly between 12 AM and 4 AM, likely when fewer trains are running. This highlights that delays are more common when the system is under greater pressure, coinciding with high commuter demand during peak hours.

3.2 Do certain stations or lines experience more frequent delays?

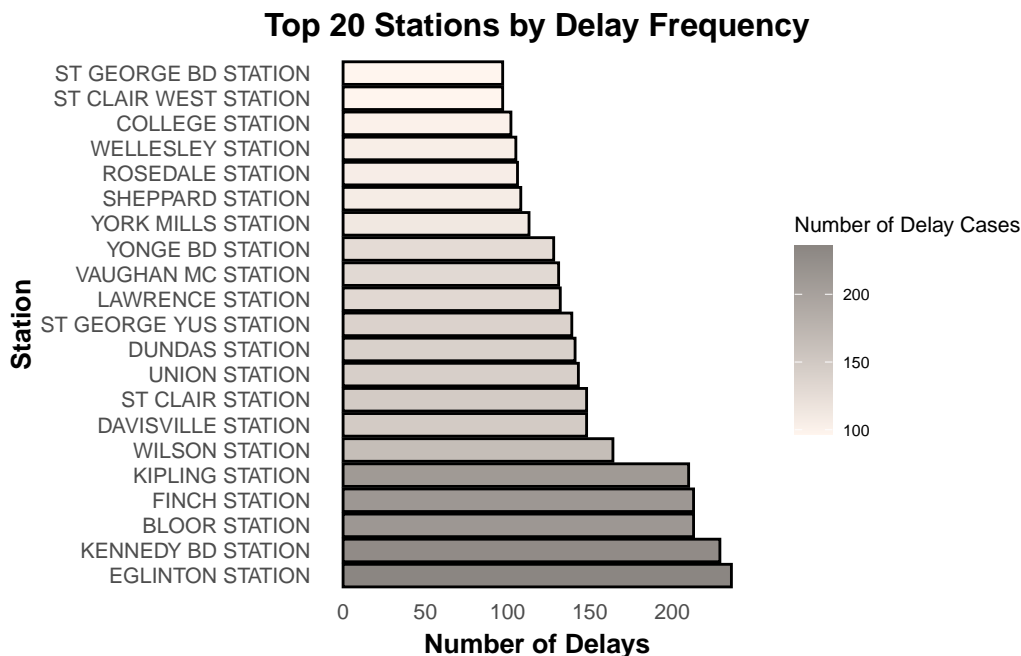


Figure 8: Delay Frequency by Top 20 Stations

Figure 8 highlights the 20 subway stations with the most frequent delays. Eglinton and Kennedy BD stations top the list with over 200 delays, followed by Bloor, Finch, and Kipling stations, each recording more than 150 delays. Eglinton Station faces frequent delays largely due to ongoing construction for the Eglinton Crosstown LRT project, which causes regular disruptions. Kennedy BD Station, on the other hand, is a major transfer hub connecting multiple subway and bus lines, which results in high passenger traffic and operational complexity. These factors create significant pressure on both stations, making them prone to delays. Furthermore, delays are more common in central, high-traffic stations such as Union and St. George YUS, which are major transfer hubs. This suggests that stations serving multiple lines or facing heavy passenger flow are more prone to delays, likely due to operational challenges, maintenance, or increased passenger demand.

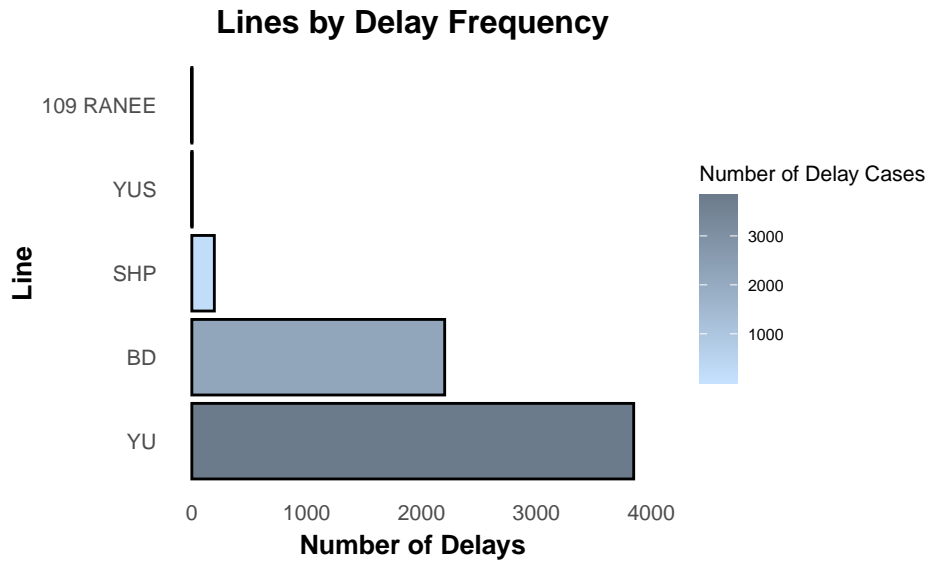


Figure 9: Delay Frequency by Lines

Figure 9 illustrates the delay frequency across different subway lines. The YU (Yonge-University) line experiences the highest number of delays, with nearly 4,000 recorded instances. This is followed by the BD (Bloor-Danforth) line, which also has a significant number of delays, close to 3,000. This distribution highlights that the most heavily used lines, particularly the YU and BD lines, tend to experience more frequent delays, likely due to their higher passenger volume and operational load. These lines serve as the backbone of the Toronto subway system, connecting major hubs and handling a significant proportion of daily commuters.

3.3 Do certain delay causes lead to longer delays than others?

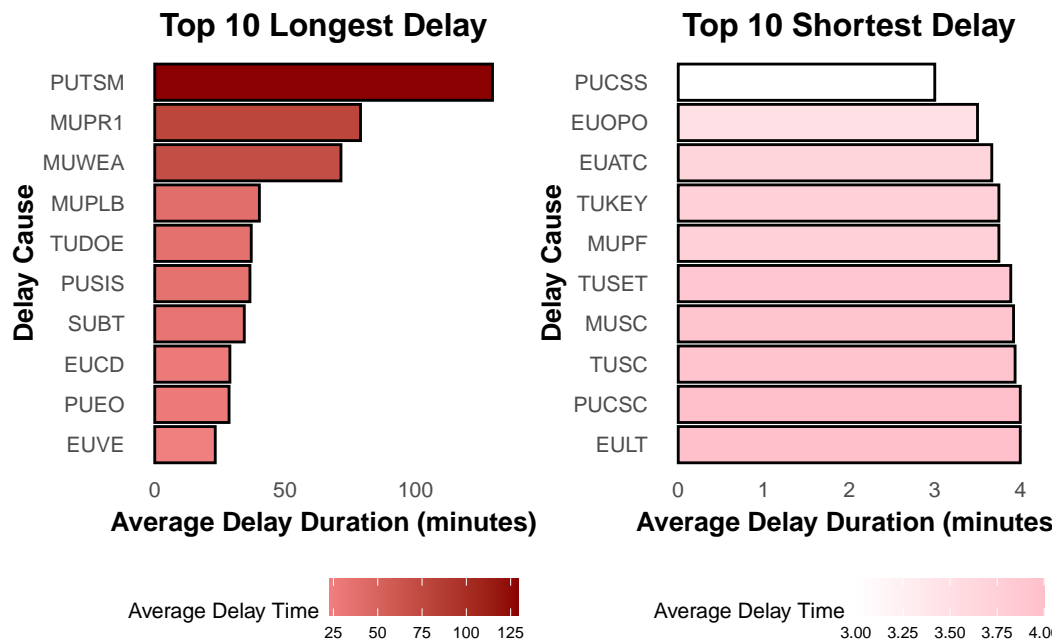


Figure 10: Top 10 Longest and Top 10 Shortest Delay Causes by Average Duration

Figure 10 presents the top 10 causes of both the longest and shortest subway delays, revealing the wide range of impacts that different issues can have on service. (Note: the detailed explanations for each delay cause code are provided in the [codebook](#).) The longest delays, led by PUTSM (Track Switch Failure), with an average delay over 125 minutes, stem from critical infrastructure problems requiring complex repairs. Other major causes, like MUPR1 (Train in Contact with Person) and MUWEA (Weather Events), also lead to significant delays due to safety protocols or external weather conditions. These delays often involve multiple teams and external coordination, extending the service interruptions.

On the other hand, the shortest delays are related to less severe issues like PUCSS (Central Office Signaling System) and EUOPO (Non-Train Door Monitoring), which average only 3 to 4 minutes. These minor problems can be resolved quickly, causing minimal disruption to subway operations. This contrast highlights the importance of addressing critical operational failures to minimize lengthy delays, while the more frequent short delays require efficient management to keep subway services running smoothly.

4 Discussion

This study examines the patterns and causes of delays in the 2024 TTC subway system, focusing specifically on instances where actual delays occurred (i.e., when delay minutes exceeded zero). Through exploratory data analysis (Section 2.3), we observed that while most delays are brief and fall within an acceptable range, there are occasional major disruptions that significantly impact overall subway operations. Furthermore, the analysis reveals that delays are more frequent on weekdays compared to weekends, likely due to higher commuter traffic and increased operational demand during the workweek. Additionally, the data suggests that longer delays can increase the spacing between trains, resulting in a cascade of subsequent delays and further reducing service efficiency.

In the results section (Section 3), the analysis indicates that delays occur more frequently during peak hours, though there is no evidence suggesting that these delays are longer in duration compared to non-peak hours. This suggests that while delays are more common during busy periods, their severity remains consistent throughout the day. Additionally, the data reveals a significantly higher number of delays at key transit hubs, likely due to their role as major interchange points in the TTC network, where high traffic volumes and complex operations increase the likelihood of disruptions. The analysis also shows that the most heavily utilized lines experience more frequent delays, driven by higher ridership and operational demands. Moreover, mechanical failures are identified as a leading cause of major delays, in contrast to shorter delays caused by minor issues. Addressing these large-scale disruptions requires prioritizing investments in infrastructure and operational improvements to reduce the frequency and impact of prolonged delays.

4.1 Weaknesses and next steps

Our analysis has certain limitations, as it focuses solely on a snapshot of TTC operations from January to August 2024, which may not capture long-term trends in operational efficiency. In the future, we could incorporate the full set of TTC subway delay data from 2017 to 2023, available through Open Data Toronto, to examine whether subway delays have improved over a longer period, particularly around the pre- and post-pandemic periods (before and after 2019). With more comprehensive data, we could also further investigate whether large-scale delays caused by significant events have been mitigated through infrastructure upgrades or other measures. Additionally, a comparison of delays between the subway and bus systems could provide insights, such as whether subway delays are less severe than bus delays during extreme weather conditions.

Overall, addressing subway delays can enhance the reliability of TTC services. By focusing on infrastructure improvements and better management of peak-hour traffic, TTC can provide passengers with a more consistent and satisfactory experience.

A Appendix

A.1 Data cleaning

In the data cleaning process, we standardized the coding format of the station, line, and bound variables to ensure consistency for comparing delay events across stations and subway lines. We removed missing or invalid values from key fields such as station and line, as one of the key focuses of our analysis was to look at trends in delays at different stations and lines. Similarly, we removed entries with missing delay codes describing the cause of delays to ensure the reliability of our analysis of the causes of subway disruptions. We normalized the date and time columns to facilitate time-based analyses, such as identifying peak hours of delay. While we allowed for missing values in the constraint column (direction of travel) because it was not important to our investigation, we ensured the completeness and consistency of all other key information to enable robust exploration of the dataset.

References

- Baptiste Auguie. 2017. *gridExtra: Miscellaneous Functions for "Grid" Graphics*. <https://CRAN.R-project.org/package=gridExtra>.
- City of Toronto. 2024. "TTC Subway Delay Data." <https://open.toronto.ca/dataset/ttc-subway-delay-data/>.
- Forum Research. 2019. "Toronto Transit Poll, May 2019." <https://poll.forumresearch.com/post/2958/toronto-transit-may-2019/>.
- Hadley Wickham and Jennifer Bryan. 2023. *readxl: Read Excel Files*. <https://CRAN.R-project.org/package=readxl>.
- Hadley Wickham and Mara Averick and Jennifer Bryan and Winston Chang and Lucy D'Agostino McGowan and Romain François and Garrett Grolemund and Alex Hayes and Lionel Henry and Jim Hester and Max Kuhn and Thomas Lin Pedersen and Evan Miller and Stephan Milton Bache and Kirill Müller and Jeroen Ooms and David Robinson and Dana Paige Seidel and Vitalie Spinu and Kohske Takahashi and Davis Vaughan and Claus Wilke and Kara Woo and Hiroaki Yutani. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Hao Zhu. 2024. *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.
- Matt Elliott. 2023. "The Rise of TTC Delays, Minute by Minute." <https://toronto.cityhallwatcher.com/p/chw257>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Sam Firke. 2023. *janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Sharla Gelfand. 2022. *opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- Yihui Xie. 2023. *knitr: A General-Purpose Package for Dynamic Report Generation in R*. <https://yihui.org/knitr/>.