Avoiding Delays on the TTC Streetcar*

A Data-Driven Guide for Smarter Travel

Khushaal Nandwani

September 27, 2024

This study examines delays in the Toronto Transit Commission (TTC) streetcar system, analyzing patterns based on time, day, direction, and streetcar line. The data reveals that eastbound and westbound TTC streetcars face more delays, particularly during early mornings and late evenings. These findings can assist riders in planning routes to avoid delays and help improve transit scheduling.

1 Introduction

The Toronto Transit Commission (TTC) streetcar system is one of the city's most iconic and frequently used public transit services. However, streetcar delays have been a persistent issue for riders, often resulting in inconvenience and unpredictable travel times. There is a messaging service available, where users can text a number to receive information about the next arriving TTC streetcar. However, this is not always accurate. These delays can be influenced by various factors, including the time of day, the direction of travel, and external circumstances like construction or traffic congestion. Understanding the patterns and causes of these delays can help both daily commuters and occasional travelers optimize their routes and avoid unnecessary disruptions.

In this paper, we explore the patterns of delays in the TTC streetcar network by analyzing data provided by Open Data Toronto (Gelfand (2022a)). We analyze the streetcar delays, examining factors like time of day, day of the week, travel direction, and specific streetcar lines. Notably, delays tend to be longer on eastbound and westbound routes, particularly during early morning and late evening periods. The analysis reveals trends such as Line 501 and 506 experiencing the longest delays. By identifying these patterns, the study provides practical insights for riders to plan their trips and for transit authorities to optimize operations.

The structure of the paper is as follows: First, it outlines the dataset used (Section 2). Next, the results of the analysis are presented (Section 3). Finally, the paper addresses the practical

*Code and data are available at: https://github.com/khushaal-nandwani/ttc-avoiding-delays implications of the findings for commuters, transit planners, and city infrastructure management (Section 4), along with the study's limitations and potential areas for future research (Section 5). This structure is inspired by the work of Rohan (Alexander (n.d.)).

2 Data

The raw data used in this paper, was from Open Data Toronto and was extracted using R programming language by R Core Team (2023) and its packages including `tidyverse` by Wickham et al. (2019), `deplyr` by Wickham et al. (2023) and `opendatatoronto` by Gelfand (2022b).

The dataset contained information about delays in the TTC streetcar system. Each row of the data represents a delay in the TTC streetcar system. The columns in the dataset along with their description are given in Table 1. [1]

Table 1: Description of variables in the dataset

Variable	Description	Example
Date	The date when the delay was caused.	2023-01-01T00:00:00Z
Line	The line number or route of the streetcar	501
Time	The time when the delay was caused.	02:37:00
Day	The day of the week when the delay was caused.	Monday
Location	The location where the delay was caused.	Queen and Spadina Ave.
Incident	The reason for the delay.	Diversion
Min Delay	The length of the delay in minutes.	5
Min Gap	The time gap with the next streetcar.	10
Bound	The direction in which the streetcar is heading.	W
Vehicle	The vehicle number of the streetcar.	4001

In Figure 1, we show a sample of the raw data from Open Data Toronto.

2.1 Measurement

Even though OpenDataToronto does not provide the exact methodology for measuring delays, it is likely that delays are measured from the scheduled arrival time to the actual arrival time. We know that the TTC keeps track of each streetcar's arrival and departure times at each stop, as all data is sent to GTFS for updating Google Maps (Transitland (2024)). So, calculating the delay would be a simple subtraction of the actual arrival time from the scheduled arrival time.

[1] Every variable was used in analysis except Location, Vehicle and Min Gap.

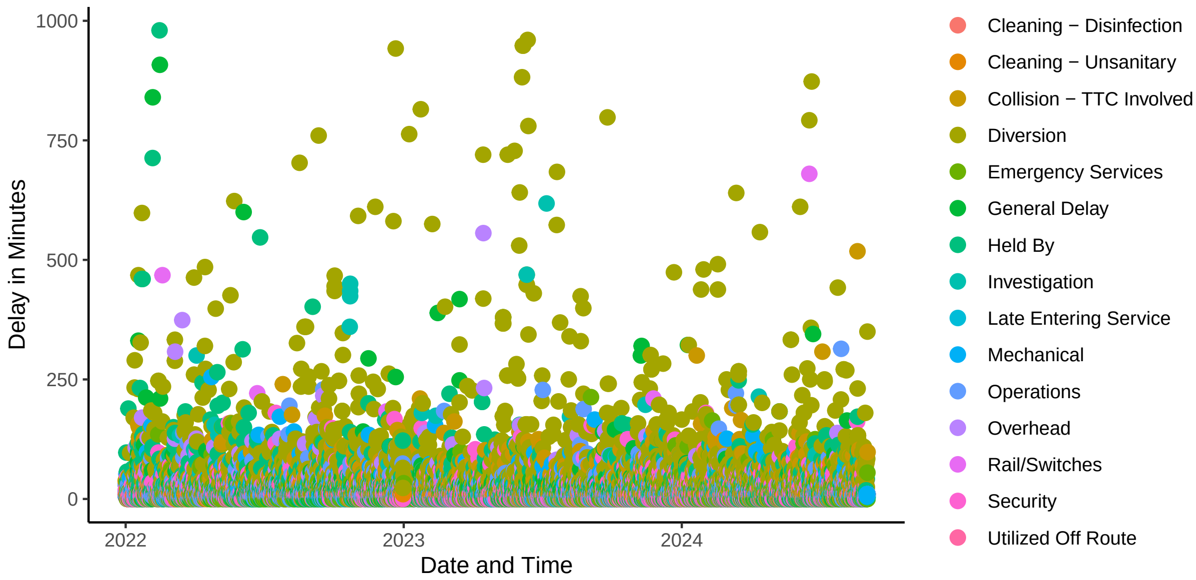

Figure 1: Sample of the raw data from Open Data Toronto showing delays (minutes) in the TTC streetcar system

2.2 Similar Datasets

Other datasets available included information on TTC subways and buses. In the Section 5, we discuss how combining these datasets could provide a more comprehensive view of transit delays in Toronto.

3 Results

After thorough analysis, we observed that the average delay was highest for Eastbound streetcars, while Southbound streetcars had the shortest delays. Overall, streetcars traveling in the North and South directions experienced lower average delays compared to those traveling East and West, as illustrated in Figure 2. This is consistent with expectations, as Toronto's nearby suburbs, such as Oshawa, Hamilton, Scarborough, Mississauga, and Brampton, require East-West travel, making those routes busier and therefore more prone to delays. This information is also cross verifed when we find in Section 4.4 that Line 501 and 506 have the highest average delays which is an East-West bound streetcar.

When we examine the average delay of streetcars by line numbers in Figure 3, we find that the 501 line has the longest delays, while the 503 line has the shortest average delays. The significant variation in delays across different streetcar lines makes it valuable for riders to consider when making travel decisions. Often, riders have the option to choose between streetcar numbers when they are between two or three routes or if the stops are not too far apart.

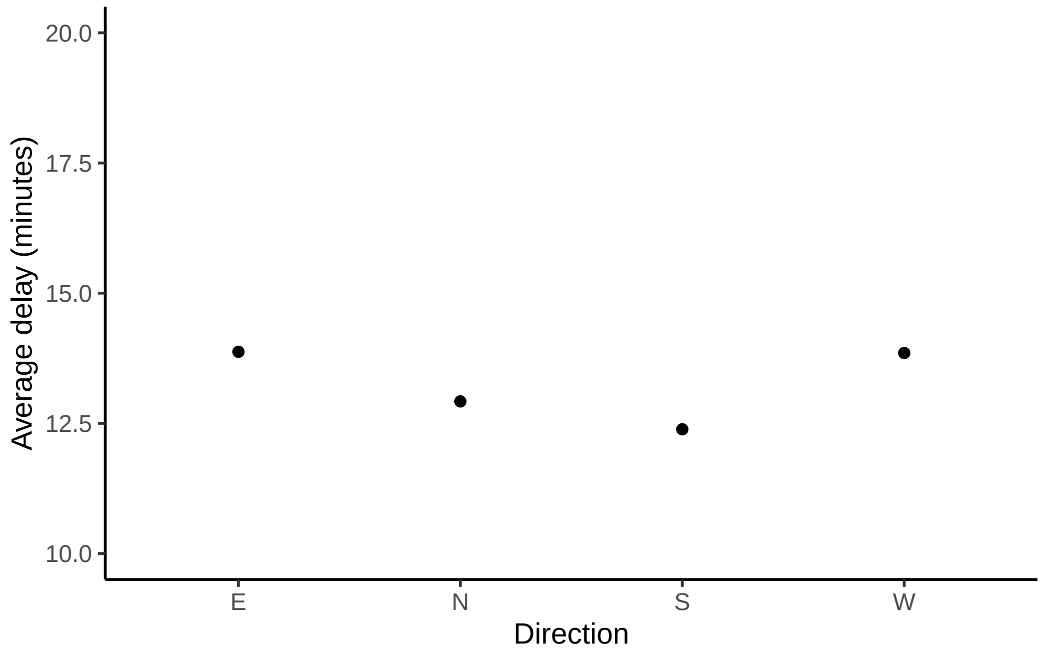

Figure 2: Average delay by direction of TTC streetcars

However, Google Maps typically displays the closest streetcar, not necessarily the one with the shortest delay. On average, Line 506 and 509 had the highest average delay length, while Line 507 had the lowest.

The bars represent the number of delays, which could be influenced by either a lower frequency of streetcars on these lines or fewer opportunities for delays. Since we do not know the actual number of streetcar operations, we cannot present an adjusted value, making frequency-based estimates potentially inaccurate. However, being aware of this limitation is still useful for interpreting the data.

When we examine the delays by time of day in Figure 4, we notice a spike at 4 AM and 8 PM. This is likely due to driver shift changes. This information is valuable for riders, as they can anticipate potential delays if traveling around these times. During peak office hours, there is a slight increase in delays, but it is not significant. This is likely due to the increased frequency of streetcars during these periods, which helps to mitigate delays.

When examining the average delay length by month in Figure 5, we observe spikes in March, August, and November. Explaining these trends could be quite interesting.

We can also examine the average delay length by day of the week in Figure 6. Surprisingly, there isn't a clear trend across different days. This could be because the primary causes of delays are not traffic-related but due to other factors. Upon investigating the reasons for delays

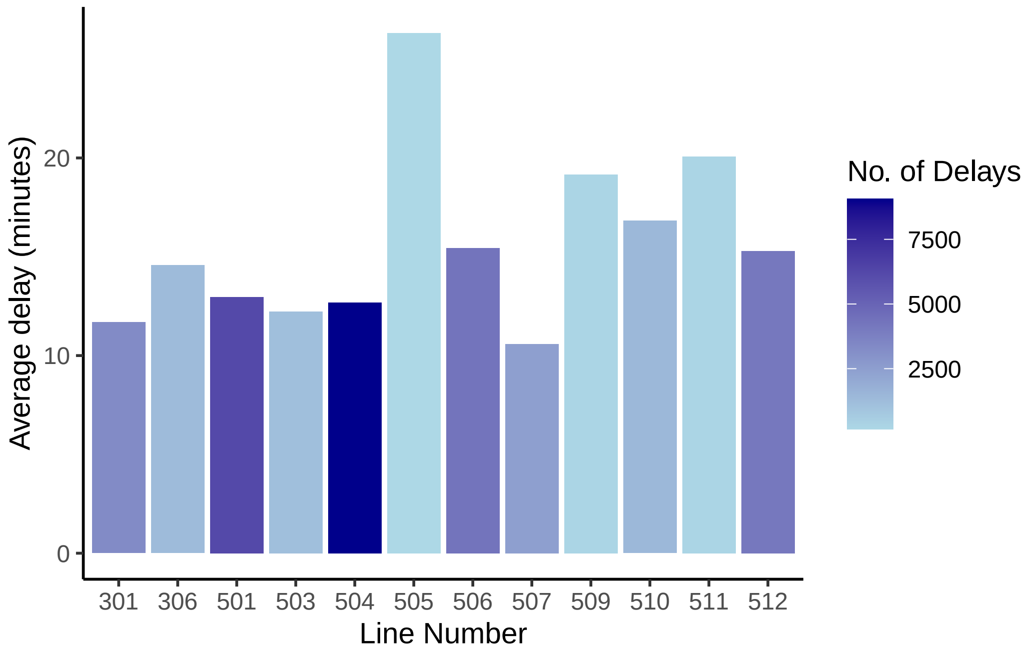

Figure 3: Average delay of streetcars by Line number along with the number of delays caused on each line

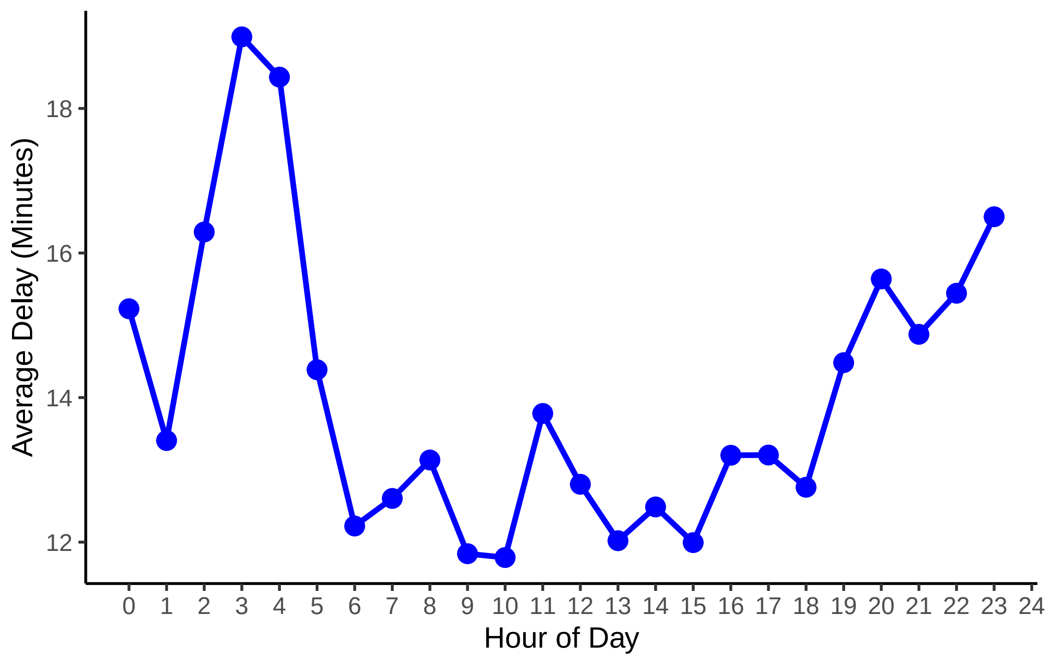

Figure 4: Average delay by hour of the day

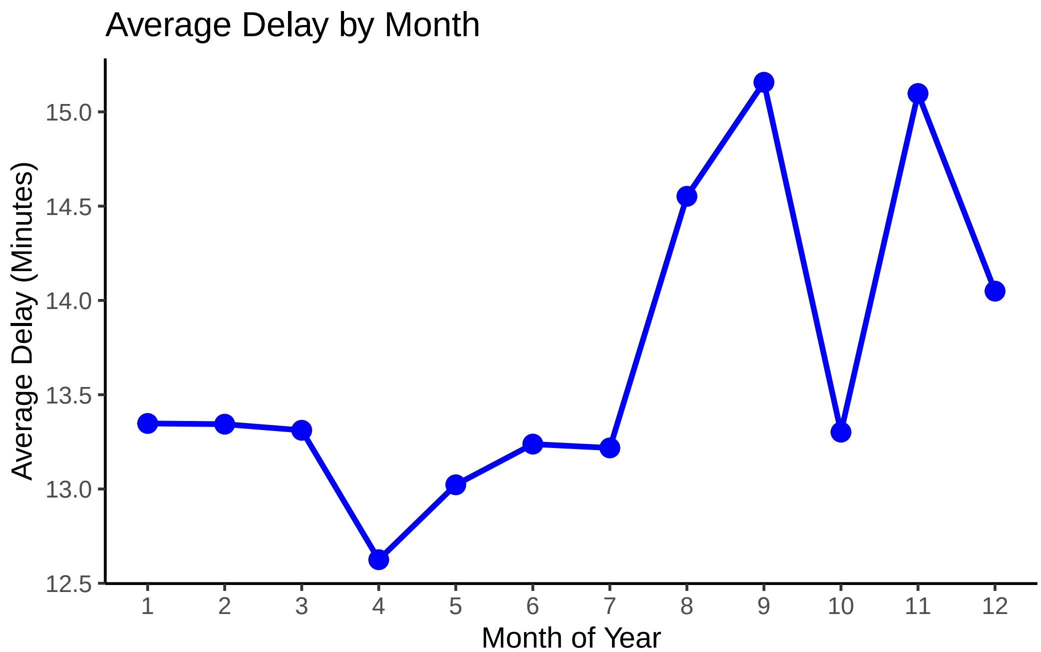

Figure 5: Average delay by month of the year

in Figure 7, this becomes clearer. The most common reasons for delays are "Diversion," "Overhead," "Collision," and "Held By," none of which are directly related to traffic. Interestingly, "General Delay" is one of the least common reasons, as shown in the graph below.

3.1 Application

As shown in Figure 8, if a person is at position X_1 and wants to go to X_2. Based on our analysis, they should take the 501 and 504 lines, as these show fewer delays compared to the 511 and 510 lines.

4 Discussion

The findings from our analysis provide significant insights into the patterns of delays in the TTC streetcar system. These insights can be broken down into a few key areas: the influence of direction, time of day, day of the week, and specific streetcar lines. In this discussion, we will interpret these results in the context of practical use for daily commuters, transit planners, and city infrastructure management.

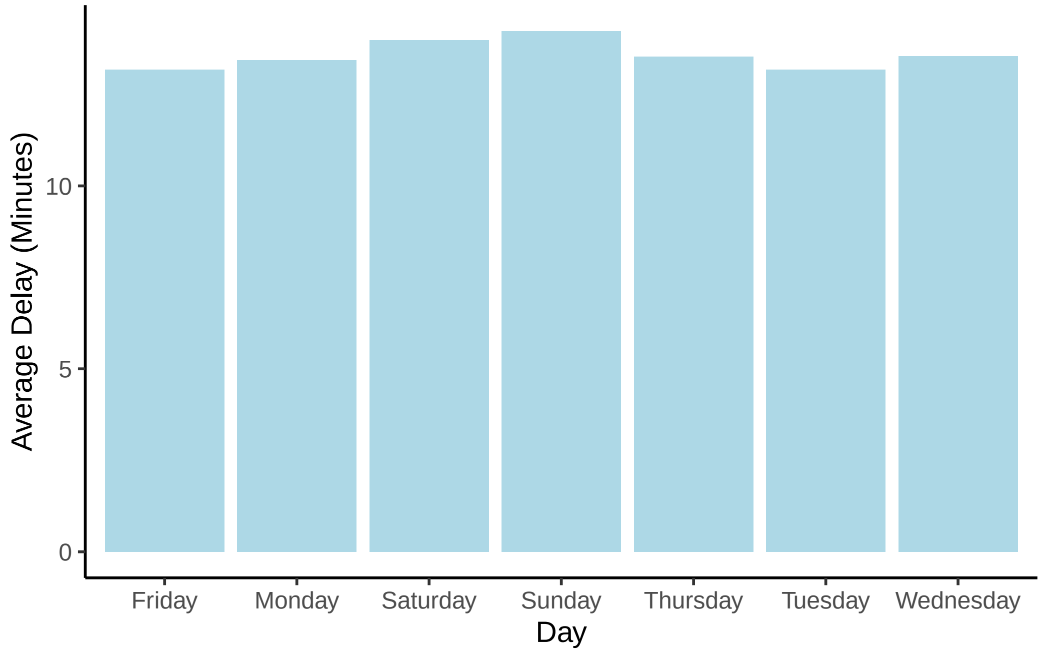

Figure 6: Average delay (minutes) by day of the week in TTC Streetcars

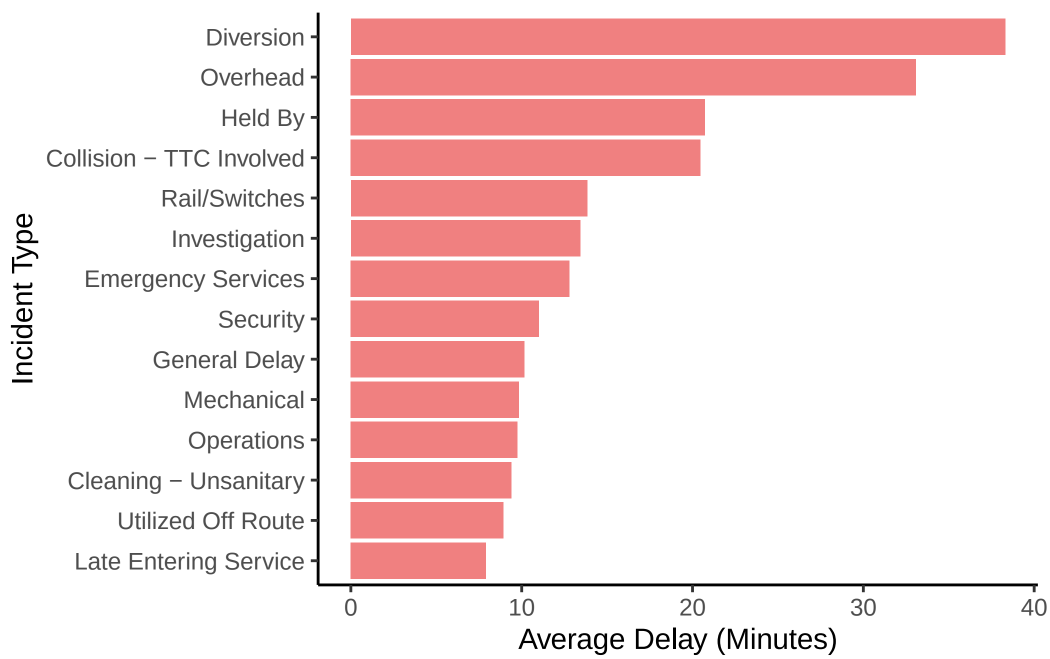

Figure 7: Reasons for delays in TTC Streetcars

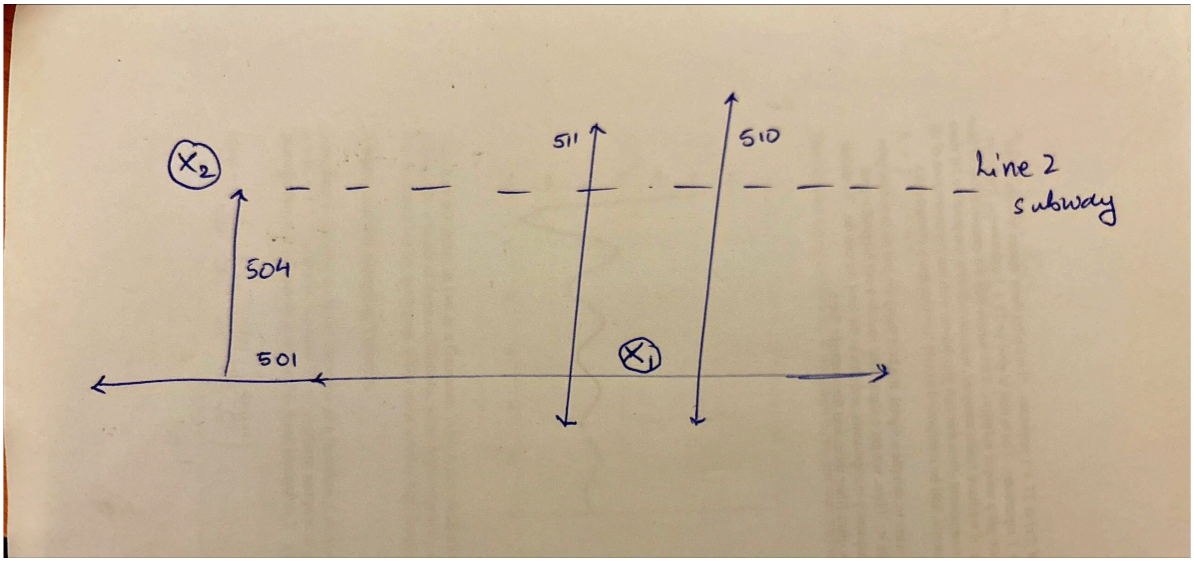

Figure 8: An example problem where a person at X_1 wants to go to X_2

4.1 Directional Influence on Delays

The analysis reveals that streetcars traveling in the eastbound and westbound directions experience longer delays compared to those traveling north and south. This could be due to the geographical layout of Toronto, where many busy corridors, such as Queen Street and King Street, run in the east-west direction, connecting major suburbs and districts. These routes are more prone to traffic congestion, roadwork, and external events like construction or parades. Additionally, these routes do not have dedicated lanes for streetcars, unlike Spadina, which is a north-south bound streetcar route. Conversely, north-south routes may encounter less traffic due to fewer major cross-city connections in those directions.

From a commuter's perspective, understanding this directional influence can help optimize travel plans. For instance, If a commuter can choose between an eastbound or northbound route, the northbound option may provide a more reliable and punctual journey during peak hours.

4.2 Time of Day and Peak Periods

The data indicates that delays spike during specific times of the day, particularly around 4 AM and 8 PM. These times are likely influenced by shift changes or maintenance schedules within the TTC. However, it is worth noting that the expected increase in delays during traditional rush hours (8 AM to 10 AM, 4 PM to 6 PM) was not as pronounced as one might expect. The higher frequency of streetcars during peak hours likely mitigates the overall impact of individual delays.

For transit riders, avoiding travel during shift change periods may reduce the likelihood of encountering long delays. Meanwhile, the TTC could use this information to reassess staffing or scheduling to minimize disruptions during these critical times.

4.3 Day of the Week and Seasonality

The analysis did not reveal a significant trend in delays across different days of the week. This suggests that the factors causing delays—such as collisions, diversions, or equipment issues—are not strongly tied to weekday traffic patterns. We observed a notable spike in delays during the summer months, particularly in June, likely due to seasonal construction and increased tourism during the summer festival season.

For city planners, the summer spike in delays highlights the importance of coordinating roadwork schedules and traffic management systems to minimize the impact on transit. For TTC users, planning trips around these periods or choosing alternate routes could help avoid unnecessary delays.

4.4 Delays Specific to Streetcar Lines

Certain streetcar lines, such as the 501 Queen and 506 Carlton, exhibited significantly higher delays compared to others. This disparity could be attributed to various factors, including the length of the route, the traffic conditions in areas serviced by these lines, and operational challenges unique to each line.

Commuters familiar with the TTC may benefit from being selective about which streetcar line they choose, especially if they are traveling during periods when delays are more likely. Additionally, transit planners can use this data to focus interventions, such as priority traffic signaling or dedicated streetcar lanes, on the most delay-prone routes to improve overall service reliability.

5 Limitations and Further Research

While this study provides valuable insights into TTC streetcar delays, there are a few limitations that must be acknowledged. First, the dataset does not include the total number of streetcars that ran during the period, which prevents us from calculating an adjusted delay frequency relative to the service volume. Furthermore, the analysis does not account for external factors such as weather conditions, which may also influence delay patterns.

In addition, datasets for other TTC transportation modes, such as subways and buses, are also available. Cross-comparing these datasets could provide a more comprehensive view of transit delays, helping commuters make more informed decisions about their travel options. Since

TTC tickets are valid for two hours regardless of the transportation type used, this analysis could significantly improve commuters' ability to choose the optimal combination of public transport, leading to a more efficient travel experience.

Future research could expand on this analysis by integrating additional datasets, such as weather reports, traffic data, or real-time transit schedules. A more granular understanding of how these external factors interact with the internal workings of the TTC system could help in developing more targeted solutions to reduce delays.

In summary, our findings suggest that TTC riders can minimize delays by considering the time of travel, direction, and streetcar line. For city and transit authorities, the insights can inform operational improvements and better scheduling practices.

6 Appendix

6.1 Cleaning Data

The data required some basic cleaning, such as dropping empty values. In addition, certain line numbers had very few values. Line numbers with fewer than 100 records were dropped, as they were likely temporary or emergency lines. Despite this, we were still able to cover most of the lines mentioned on the TTC website City of Toronto (n.d.). Records where the delay was greater than the gap were also removed, as it would not make sense for a streetcar to be delayed while the gap from the next streetcar was smaller.

Upon reviewing the data, we also found several anomalies where the delay exceeded 1,000 minutes. This could either be an error in logging, or an extreme case. Regardless, I have decided to exclude all observations with delays greater than 300 minutes, as it is highly unlikely that a streetcar would be delayed for more than 5 hours on a regular day, and including such outliers could skew the overall analysis.

References

Alexander, Rohan. n.d. *Telling Stories with DataScience*. https://tellingstorieswithdata.com.

City of Toronto. n.d. "Toronto Transit Commission." https://www.ttc.ca/routes-and-schedules/listroutes/streetcar.

Gelfand, Sharla. 2022a. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

———. 2022b. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

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

Transitland. 2024. https://www.transit.land/feeds/f-dpz8-ttc.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.