Incidents Causing TTC Bus Delays in 2024*

The time period covered is January 2024 to August 2024

Aliza Abbas Mithwani

September 27, 2024

This paper discusses the causes for TTC bus delays in 2024 from January until August. We find that although mechanical incidents cause the most frequent delays, diversions cause the most time-consuming delays. We also test the validity of rush hour riders claiming to face more delays. People seem to be frustrated by the TTC delays and this is reducing consumer satisfaction.

Table of contents

*Code and data are available at: https://github.com/alizamithwani/TTC-Delays-Analysis.git

1 Introduction

In 2019, the Global News reported that although the majority of TTC riders said they are satisfied with the company's service, half have complained of regular delays when using the service. This problem has led a reduction in satisfaction for these people. Riders at rush hour reported being more affected by unanticipated TTC delays. This data was collected by a poll conducted in May 2019, which sampled almost 2,000 riders in Toronto, Calgary, Edmonton, Vancouver and Montreal. (Vuong (2019))

The goal of this research paper is to discover which incidents are causing the most time-consuming TTC bus delays. The dataset used in this paper is "TTC Streetcar Delay Data" and specifically for 2024 (the file name is "ttc-streetcar_delay_data-2024.xlsx". This dataset was taken from City of Toronto's Open Data Portal and published by the Toronto Transit Commission (TTC). For the sake of simplicity, this paper only focuses on TTC buses as opposed and does not include or draw conclusions about TTC subways and streetcars. Type of incident is not featured in the TTC Subway Delay Data and TTC Streetcar Delay Data included a couple of additional incident types.

The rest of this paper is structured such that we will discuss the data (Section 2) next. More specifically, we will go over where this data comes from (the source) and give a more comprehensive context for this data. Several variables like time, incident, and the delay amount will be used in our analysis, answering key questions like whether rush hour is really facing more incidents, which incident type is causing most delays, and finally which incident type is wasting the most time for riders in Toronto. We will move to outlining the results we can draw from analyzing the several variables mentioned using different graphs and tables. Finally, we will move to a detailed discussion of some key points we can draw from the analysis (Section 3), along with outlining weaknesses (Section 3.4) in the paper and possible next steps

(Section 3.5) after considering this study. All the processes involved in simulating, cleaning and testing data are outlined in the Appendix at the end of this paper.

2 Data

2.1 Software and Packages

As mentioned earlier, the dataset was found on City of Toronto's Open Data Portal and titled "TTC Bus Delay Data". When simulating the data we used `tidyverse` (Wickham et al. (2019)), when downloading the data, we used `opendatatoronto` and `tidyverse` (Gelfand (2022) and Wickham et al. (2019)), and when cleaning and testing the data we used tidyverse (Wickham et al. (2019)). Furthermore, we used the statistical programming language R (R Core Team (2023)) to write any code associated with this paper. Figure 1 was constructed using `knitr` (Xie (2022)). I used `ggplot2` (Wickham (2016)) to make Figure 2.

2.2 Dataset and Measurement

The dataset obtained contained 11 datasets, each corresponding to one of the last 11 years. Each dataset has 10 variables: Report Date, Route, Time, Day, Location, Incident, Min Delay, Min Gap, Direction and Vehicle. One specific variable of interest is Incident, which provides a description of the incident that caused the delay. The second variable of interest is Min Delay, which is the delay to the schedule for the next bus (in minutes). Only the 2024 dataset (January-August) was used to analyze the most up-to-date situation of the TTC bus delays. There might be other contributing factors in older years, which would result in reduced applicability of conclusions to the current situation if those contributing factors have changed. The dataset was published by the Toronto Transit Commission (TTC), which is the company that provides the bus service in discussion.

2.3 Cause of Interest

The Global News reported that half of the riders they sampled said they were dissatisfied by the TTC delays (Vuong (2019)), sparking an interest in why these delays occur and is there a certain incident that is more responsible for these delays compared to rest the rest. Finding an answer to this question can lead to more focused investigation into the problem and more tailored solutions.

2.4 Analyzing the Data

2.4.1 Most Frequent Incidents

Analyzing how frequently each incident has occurred over the last year (Figure 1), we were able to see which incidents occurred more than others. The table shows that an overwhelming amount of incidents are mechanical. This means that issues with mechanical components of the TTC bus (engine, brakes, transmission, electrical systems etc.) cause the most frequent delays. This is useful information as if this is such a frequent issue compared to other issues that the TTC faces, the TTC could target investigating further and working to improve in that department. This can possibly be by hiring more experienced, skilled or a higher quantity of mechanics and engineers on their team. Since the TTC is so integral to Toronto, the government can also step in and invest in developing the infrastructure and system currently in use by the TTC. Although this table tells us a lot about which incidents the TTC buses are facing the most, the feelings of dissatisfaction from TTC delays likely have more to do with the amount of time wasted or the lateness of the rider in reaching their destination. SO, it is important to recognize that Figure 1 does not tell us anything about most time-consuming cause of TTC bus delays.

Incident Type	Number of Incidents
Cleaning - Unsanitary	1422
Collision - TTC	2385
Diversion	2380
Emergency Services	1607
General Delay	2411
Investigation	613
Mechanical	12041
Operations - Operator	5972
Road Blocked - NON-TTC Collision	117
Security	2550
Utilized Off Route	1703
Vision	1037

Figure 1: The total number of incidents in each category. There are several different types/catgories of incidents that cause TTC bus delays in 2024.

2.4.2 Most Time-Consuming Incidents

To investigate which incidents are causing the most time-costing bus delays, we find the total delay time due to each type of incident as displayed in Figure 2. Here we used the variable

Incident for the x-axis and the variable Min Delay for the y-axis; however we converted minutes to hours by dividing by 60. This makes it easier to fit the data in the graph as the sum of total delay time in minutes for each incident was too large.

As we can see, Figure 2 clearly tells a different story from Figure 1. In Figure 2 we can see that diversions are, in fact, the most time-consuming incident and have resulted in the most delay time overall in 2024 so far. Certain incidents cause longer delays than others and each diversion is bound to cause a large delay as it means that the TTC bus is taking a different route due to construction, accidents, events, weather, and traffic. Construction, specifically, can go on for a long time (weeks, months…), causing delays for a longer period of time and aggregating a large amount of total delay time.

Interestingly, the Transit app has had an update recently that allows riders access to live updates regarding TTC detours and diversions (Wilson (2024)). Since this update took place in September and this dataset only contains data from January till August, it will be interesting to see if this Transit app update will result in a better experience for riders as they will be able to become aware of diversions more efficiently and can avoid using the TTC bus service if they are in a hurry.

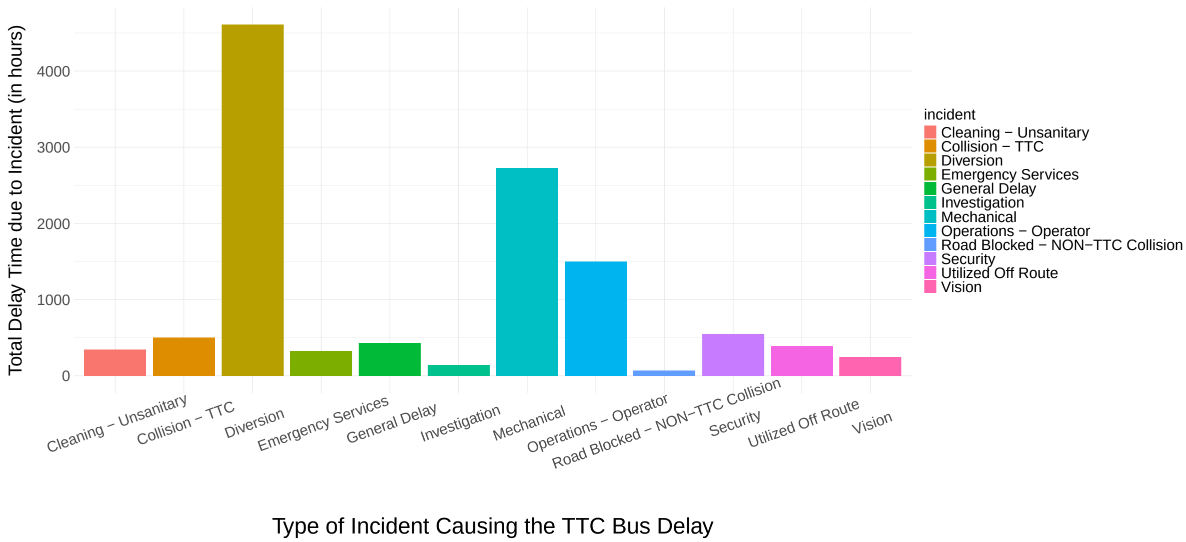

Figure 2: The total delay time for TTC buses by incident type. This total delay time is only for 2024 (Jan-Aug). The time is converted into hours from minutes for readability.

2.4.3 Rush Hour Incident Surge

As mentioned earlier, the Global News poll in 2019 revealed that riders at rush hour claimed to suffer more from TTC Delays (Vuong (2019)). To investigate whether the data supports their argument, we graphed the number of incidents and the corresponding hour those incident occurred, which gave us Figure 3. Bus services usually start around 6am and end around 1am with the exception of some routes that run all 24 hours. This is most likely contributing to the striking low number of incidents from 12am till 5am (0-5th hour). From 5am till 2pm (5-14th hour) the average number of incidents is approximately 1500. Then from 2pm till 7pm (14th to 19th hour), the number of incidents are quite high, reaching a maximum of 2500+ incidents. Then the number of incidents declines till 12am once again. If we assume 7-9 AM and 4-7 PM are rush hour, we can see that the number of incidents during morning rush hour doesn't deviate much from the number of incidents after morning rush hour. The surge at 5am is most likely due to the fact that most TTC buses resume services around that time. However, the peak that occurs around 3-6pm could be considered somewhat in support of what rush hour riders that participated in the 2019 Global News poll reported (Vuong (2019)).

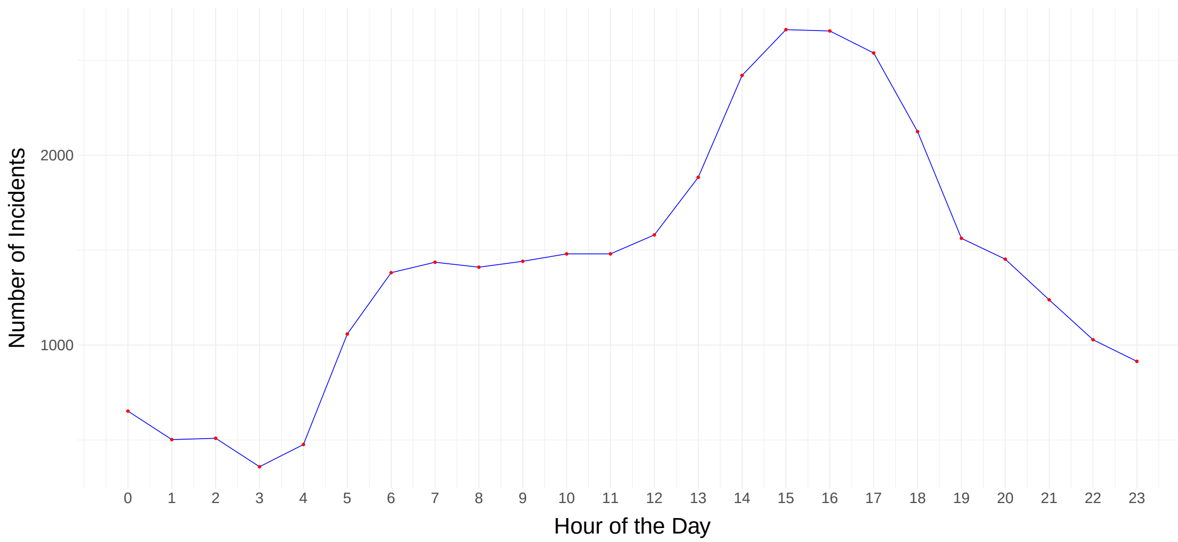

Figure 3: The number of incidents that occur in each hour of the day. This graph is used to analyze which portions of the day have a higher number of incidents and if these portions correspond to rush hours.

3 Discussion

3.1 Discrepancy Between Incident Frequency and Impact

The analysis reveals an important distinction between the frequency of incidents and their overall impact on delay times. While mechanical incidents occur far more frequently than any other type of issue, diversions result in significantly longer delays.

3.1.1 Interpretation

The high occurrence of mechanical incidents points to underlying, systemic challenges within the TTC's fleet management and maintenance practices. Frequent mechanical failures suggest that either the buses in use are aging, or there are gaps in the current maintenance protocols. Components like engines, transmissions, electrical systems, and brakes could be prone to malfunctioning more often than they should, likely due to wear and tear or inadequate preventative maintenance measures. This indicates that the TTC may need to invest in newer buses, improved maintenance scheduling, or more efficient parts management to prevent repeated breakdowns.

On the other hand, diversions, though less frequent, cause more prolonged disruptions to service. Diversions typically arise due to external factors, such as road closures, construction, traffic congestion, or emergencies. These factors are outside of the TTC's direct control, and each diversion forces buses to follow alternate routes, often resulting in longer travel times. Unlike mechanical issues, diversions can affect an entire route for an extended period, which accumulates into significant delays over time. The complexity of managing these diversions, particularly during high-traffic periods, means that a single diversion can have ripple effects, delaying multiple buses across the network.

3.1.2 Implications

The TTC faces a trade-off between addressing frequent but shorter mechanical delays and mitigating less common but far more time-consuming diversions. Targeting mechanical reliability through better maintenance strategies could reduce the sheer number of incidents. However, from a time-savings perspective, focusing on minimizing the impacts of diversions would likely yield greater overall improvements in service efficiency.

Given the large cumulative time delays caused by diversions, the TTC could work more closely with city authorities and event organizers to better plan for and mitigate these disruptions. By improving coordination and establishing protocols for real-time diversion management, the TTC could limit the impact of external disruptions on bus schedules, thus prioritizing resources toward reducing both types of delays efficiently.

3.2 Diversions as a Major Contributor to Delays

Although mechanical failures occur far more often, the data highlights diversions as a major contributor to total delay times. These longer, unplanned deviations in route cause significant disruptions to both bus operations and the rider experience.

3.2.1 Interpretation

Diversions often stem from conditions beyond the TTC's control. Examples include road construction, accidents, special events, or emergency situations requiring temporary route changes. When a bus must divert from its regular route, it is often forced to take a longer, less direct path to its destination, exacerbating the delay for all passengers on board. Additionally, such diversions often cause a cascading effect, delaying subsequent buses and impacting schedules throughout the system.

Moreover, the fact that some diversions are planned (such as those for long-term construction) means they accumulate significant delay time over weeks or months. The unpredictable nature of other diversions, like those caused by traffic accidents or weather events, adds to the complexity of managing them effectively. While planned diversions may offer the TTC an opportunity to communicate delays to riders in advance, unplanned ones often catch both drivers and passengers off guard, further complicating efforts to minimize their impact.

3.2.2 Implications

Delays caused by diversions, particularly during peak hours, can significantly disrupt rider schedules and increase frustration. Given the time-consuming nature of these incidents, riders can be delayed by substantial amounts of time, leading to a negative perception of the overall reliability of TTC services. This is particularly concerning for commuters who rely on the bus to get to work or appointments on time.

Improving communication and providing real-time updates regarding diversions could help manage rider expectations and allow them to adjust their routes or travel plans in advance. The recent update to the Transit app, which offers live updates on diversions, holds promise in this regard. If used effectively, such tools could mitigate rider dissatisfaction by helping them make informed decisions when diversions occur. Additionally, the TTC might explore ways to reduce the time impact of diversions, such as implementing dynamic rerouting strategies or prioritizing bus movement through traffic signals in congested areas affected by diversions.

3.3 Impact of Rush Hour on Incident Rates

The data shows a notable increase in incidents during the late afternoon and early evening hours, though this surge is not as pronounced during the morning rush hours as might be expected based on rider perceptions.

3.3.1 Interpretation

Incidents are distributed relatively evenly throughout the day, with a significant uptick occurring in the late afternoon. This aligns with the natural flow of traffic congestion and increased road use during these times. However, the morning rush hour does not exhibit the same spike in incidents, suggesting that either TTC services are better equipped to handle morning traffic or that external conditions (such as construction or weather-related issues) are less prevalent in the early hours of the day. The higher concentration of incidents in the late afternoon could be due to various factors, including fatigue among operators, increased vehicle congestion, or the cumulative effect of incidents earlier in the day that have not yet been fully resolved.

Interestingly, the surge in incidents aligns more closely with the general increase in city traffic and road congestion in the late afternoon, rather than being tied purely to passenger volume. This suggests that factors such as traffic density and external pressures on the road network may play a larger role in TTC bus delays than the simple number of passengers using the service.

3.3.2 Implications

The fact that incidents spike in the late afternoon indicates that managing delays during this period may be more complex than during morning rush hours. Addressing these late-afternoon delays might require a different approach, focusing on better traffic management and enhanced coordination with city authorities to ease congestion during these hours.

Additionally, operational stress on bus drivers and maintenance teams may increase later in the day due to accumulated issues, leading to a greater number of incidents. It could be beneficial for the TTC to investigate whether increased staffing or staggered schedules for maintenance and operations teams during the late afternoon could alleviate some of this pressure. This could help ensure smoother operations during a period when both external traffic and internal stress are at their peak.

3.4 Weaknesses

While this analysis provides valuable insights into the causes and impacts of TTC bus delays, there are several limitations that should be acknowledged. Firstly, this study focuses exclusively on the 2024 dataset, spanning from January to August. As such, any conclusions drawn may not fully account for seasonal variations in delay patterns, such as those caused by winter weather conditions in Toronto, which could significantly affect bus delays in the latter part of the year.

Additionally, the dataset does not include subway or streetcar delays, meaning that the findings here are limited to the bus system alone. Future studies could benefit from a broader analysis that incorporates all transit modes to offer a more comprehensive understanding of the TTC's overall performance. Expanding the analysis to include external factors, such as weather conditions, traffic patterns, or major events, could also provide a richer context for understanding the causes of delays.

3.5 Next Steps

To address these limitations, future research should extend the analysis to cover the entire calendar year and examine the impact of additional variables such as weather, traffic data, and major city events. Incorporating qualitative data, such as rider feedback and TTC staff reports, could also offer valuable insights into the subjective experiences of those affected by delays, adding depth to the quantitative findings.

Moreover, further analysis could focus on the effectiveness of the Transit app's live diversion updates and their impact on rider satisfaction. If these tools are found to reduce frustration and improve the overall experience, they could be a model for other public transit systems facing similar issues.

Appendix

A Data Simulation

Simulating data for this TTC bus delay study provides a structured approach to understand and analyze various incident types and their impacts on service delays. It enhances your ability to draw meaningful conclusions, supports model validation, and assists in effective decision-making based on simulated outcomes.

I simulated all the variables I used in this paper, namely Incident, Min Delay, and Time. I also simulated Date to use in future studies on the effect of weather on the types and frequency of incidents that cause TTC bus delays. In my simulation process for analyzing TTC bus delays, I first set up the R workspace and loaded the necessary library `tidyverse` (Wickham et al. (2019)). I defined the date range from January 1, 2024, to August 31, 2024, and decided to generate 100 random dates within this range. To simulate the time of day, I generated random hours and minutes, formatting them into a HH:MM string format.

Next, I created a vector of potential incident types that could affect bus delays, such as "Collision - TTC" and "Mechanical." Using these components, I constructed a tibble containing four main variables: random dates, incident types sampled from my defined list, simulated times of day, and a simulated delay duration measured in minutes generated from a Poisson distribution with a lambda of 10. Finally, I saved this simulated dataset as a CSV file.

B Data Cleaning

In my data cleaning process, I again loaded the necessary library `tidyverse` (Wickham et al. (2019)) for data manipulation. I then read the raw TTC data from a CSV file into a dataframe named raw_data. To clean the data, I began by standardizing the column names using the janitor::clean_names() function, which converts them to lowercase and replaces spaces with underscores. Next, I ensured that the min_delay column was treated as a numeric variable and the incident column was converted to a character type for consistency.

I then addressed formatting issues by removing the extraneous timestamp that appeared in the date column, specifically the "T00:00:00Z" portion. The time column was converted to a proper time format using strptime(), ensuring it was appropriately recognized for any subsequent analysis. After cleaning the data, I removed any rows that contained missing or NA values using tidyr::drop_na(). Finally, I saved the cleaned dataset as a new CSV file for further analysis, indicating that the data is now ready for your analytical work.

C Data Testing

In my data testing process, I once again loaded the necessary library `tidyverse` (Wickham et al. (2019)). I then read the cleaned analysis data from a CSV file into a dataframe named analysis_data. To ensure the integrity of the data, I conducted several tests. First, I verified that all entries in the date column are from the year 2024 by extracting the year and checking if it matches "2024," expecting the result to be TRUE. Next, I tested whether the min_delay column contains only numeric values and that all values are positive, again expecting a TRUE result. I also confirmed that all entries in the incident column are strings by checking if they are of character type, which should also return TRUE.

For the simulated data, I read it from another CSV file with simulated data (Section A) into a dataframe called simulated_data. I checked for negative values in the min_delay column, expecting this test to return FALSE, indicating that there should be no negative delays. Lastly, I tested for any missing values (NAs) across the min_delay, incident, time, and date columns in the simulated dataset, expecting all tests to return FALSE, indicating that there should be no missing values in these columns. This comprehensive testing process ensures the data's quality and reliability for subsequent analysis.

References

Gelfand, Sharla. 2022. *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/.

Vuong, Oriena. 2019. *Most Transit Commuters Satisfied with TTC, but Many Affected by Unexpected Delays: Poll*. https://globalnews.ca/news/5338014/ttc-service-toronto-transit-commuters-poll/.

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

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.

Wilson, Jermaine. 2024. *Starting Today, TTC Riders Are Getting a New Way to Avoid Commuting Headaches*. https://www.thestar.com/news/gta/starting-today-ttc-riders-are-getting-a-new-way-to-avoid-commuting-headaches/article_9a9cc6f6-6627-11ef-978a-ef7f079ba1ea.html.

Xie, Yihui. 2022. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.