A Comprehensive Analysis of Trends, Causes, and Recommendations for Toronto's streetcar's delay*

YITONG WANG

September 27, 2024

This paper analyze the delay patern of streetcar in Toronto in 2023 years based on the data from opendatatoronto. The result of analysis figure out that most streetcars' delays are concentrate in peak hour and occurs in specific intersection which located at prosperous square. Overall delays are short and within 10 minutes, extreme cases occasionaly happen. The main cause of delays are operational issue and mechanical problems. Delay patter and related incidents are siginificated insights to improve the efficiency of public transportation, and it is highly related with citizen's life quality, the finding provides an instructive direction which public transit should consider.

1 Introduction

In the metropolis like Toronto, public transportation is critical and essential for citizens' mobility, the streetcar runs by TTC is one common way that people use in their life. Public transportation originally aims to provide a more convenient way for people's mobility, and reduce the traffic congestion, but the frequent delays can cause an significant disruption for living quality and business.

It is important to understand the pattern of delays, which can potentially improve the public transportation system. The reason of delays can be widely varied with time, location and incident. The data analysis notice that delay of streetcar has a high possibility occur during rush hour in some prosperous region such as Dundas square. The statistic analysis show that overall delay's time within 1 hour, and concentrate in 10 minutes, some extreme cases also

*Code and data are available at: https://github.com/demainwang/streetcar_delay_analysis

happen but with a less possibility. Due to the construction operating route of streetcar, most lines are east-west orientation, most delays occur in east-west bound.

The paper focus on the 2023 streetcar's delay data obtained from opendatatoronto (Gelfand (2022)), based on R (R Core Team (2023)) to do the data computing, and use package 'tidyverse' (Wickham et al. (2019)) to create visualization and plotting, especially 'ggplot2' (Wickham (2016)), most plots are created through this package. The usage of 'knitr' (Xie (2023)) is for establishing the summary table. We also use 'hms' (Müller (2023)) package to ensure data about time work correctly. The paper aims to figure out the delay's pattern, and identify the variables contribute to delays of streetcar, such as location and time period, especially which incidents mainly affect the normal operation of streetcar. The rest of paper will discuss the source data and introduce the variables in Section 2, the following part is visualization which contains some bar chats to analyze the frequency of streetcar's delay pattern, and summary table to conclude the paper.

2 Data

2.1 Data bacground and measurements

The data we use in this paper comes from opendatatoronto (Gelfand (2022)), the dataset 'TTC Streetcar Delay Data' provides detailed information on streetcar delays in Toronto in 2023, which published by Toronto Transit Commission. This is the official data of the recording of streetcar's delay, all data comes from TTC's original internal records, so the likelihood of bias and errors is very small. The raw data contains the following variables, 'Date' for date and 'Time' for exact time when delay happens, in form "HH:MM", the 'Line' for route, the 'Location' for location where delay occurred, the 'Incident' about reason for the delay, 'Min Delay' the duration of the delay. 'Min Gap' for the time of streetcar ahead of the following streetcar.

2.2 Data Cleaning

We mainly focus on the pattern of occurred time, duration, location and direction of streetcar to figure out the pattern of occurrence of delay, therefore we do not need any information about route number and vehicle number, and 'Min Gap' is not our focus point. Since the original document is xlsx, the process of data cleaning as following steps:

1. The original data mostly looks good, but some values do not present correctly. We firstly use trimws() to delete any possible whitespace nearby the data, then use as.numeric() to make sure data store in 'Min Gap' column as numeric value.
2. During this process, some blank or abnormal data will become to N/A value, we simply use na.omit() to ignore any rows contain N/A values.
3. We use sum to glimpse the delay values interval, and figure out that over 99 percent's data within one our, some extreme case may affect the the performance of overall data, so we filter the 'Delay Min" within one hour, which also omit some 0 minutes delay.

2.3 Ethical and social consideration

From the ethical perspective, this dataset does not contain any personal information, so there are fewer privacy issues. However, we need to consider the social context that drives data collection. Toronto's public transportation system, especially streetcar services, has always been a concern due to its reliability issues. The collection of this data aims to improve transparency and provide a foundation for data based policy making to improve the efficiency of public transportation.

2.4 Visualization

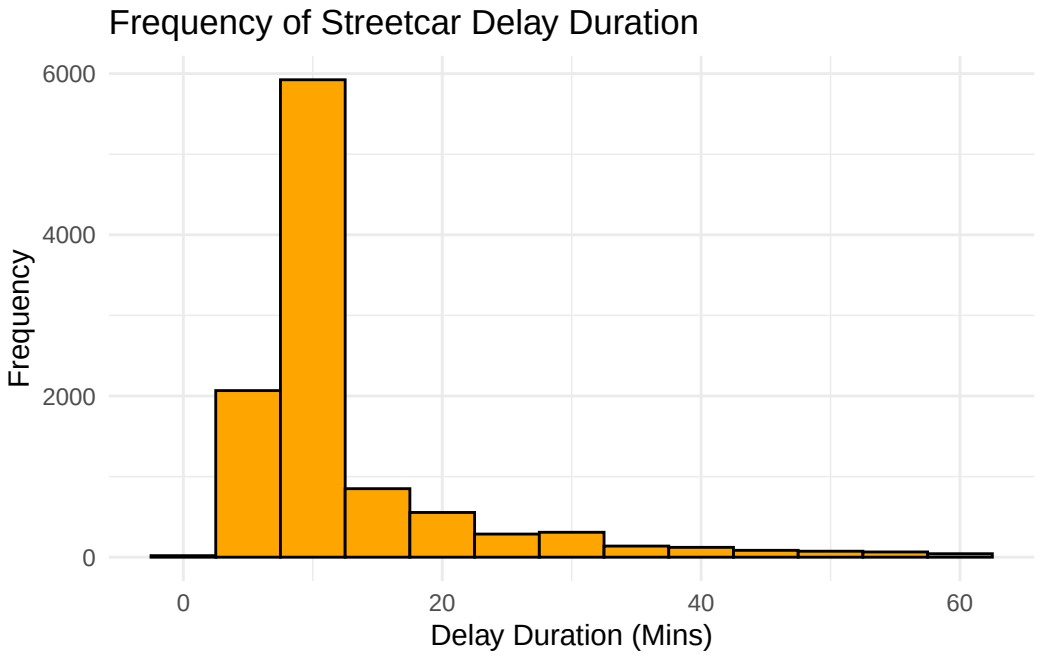

Figure 1: Distribution of streetcar's delay time frequency

Figure 1 shows a clear right skewed distribution of the frequency of delay time period, most common duration is around 10 minutes, which occupy 60 percents of total delay times. It presents in most situation the streetcar just typically delay for short time, and does not make a serious disruption for citizen's convenience.

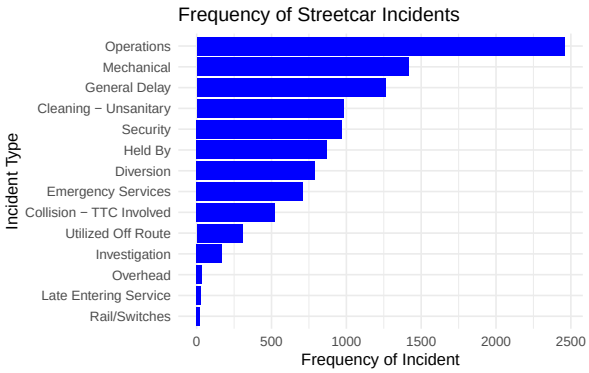

Figure 2: Frequecy of incidents and incidents causing delay time

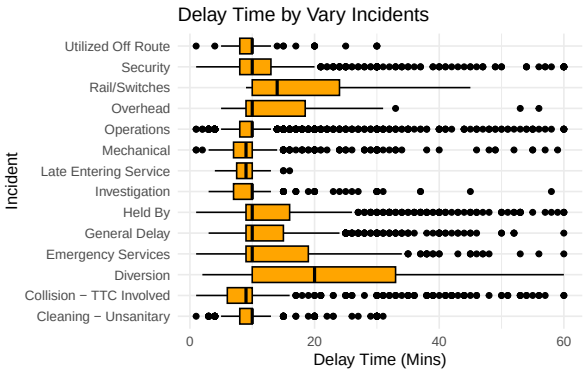

Figure 3: Frequecy of incidents and incidents causing delay time

Figure 2 The bar plot presents operational issue and mechanical problem are the top 2 main cause of delay, and the responsibility for this 2 factors belong to TTC, it is obvious TTC company should improve their internal training and have a better inspection before the operation of streetcar.

Figure 3 Fortunately, the box plot shows these 2 main factors of delay do not result in a hard disruption of citizen's transportation, they usually lead to a 10 minutes delay.

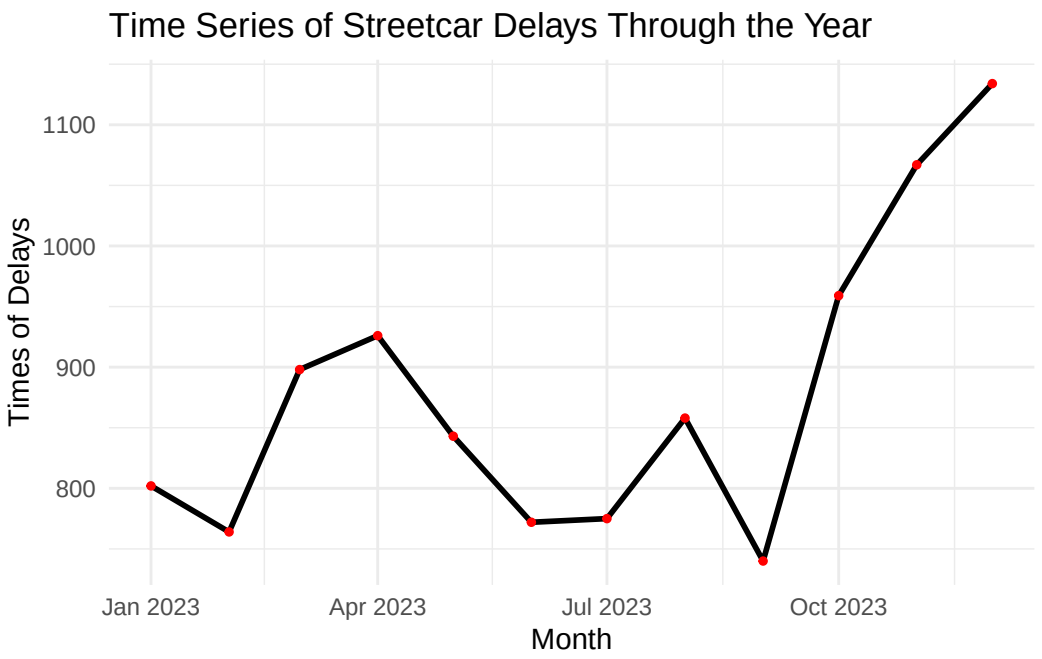

Figure 4: Trend of delay times by month

Figure 4 The line plot shows a increase trend of the frequency of delays, before September the monthly delay has fluctuation between 750 to 900, and has a significant increase from September to October, then keep going up. The increase trend shows the delay frequency becoming increasingly serve alone time.

Table 1: Summary of Streetcar Delays by Location

Location	Total Delays
DUNDAS WEST STATION	192
SPADINA STATION	166
QUEEN AND DUFFERIN	128
WOLSELEY LOOP	114
QUEEN AND SPADINA	109
QUEEN AND RONCESVALLES	108
QUEEN AND BROADVIEW	103
QUEEN AND KINGSTON	102
SPADINA AND QUEEN	102
UNION STATION	99
ST CLAIR STATION	97
KING AND BATHURST	90
VICTORIA PARK STATION	88
SPADINA AND KING	85
KING AND SPADINA	78

Figure 5: Trend of delay times by month

Figure 5 This table summary the total delays times by different location, and most of the top locations are busy intersection nearby the business district, which has a higher demand for public transportation, the TTC company should consider this situation and find the solution to reduce delay.

Table 2: Monthly Summary of Streetcar Delays

Month	Average Delay (Mins)	Total Delay Times	Max Delay (Mins)	Most Frequent Delay Location	Most Frequent Bound
2023-01	13.23691	802	60	LAKE SHORE AND ISLINGT	S
2023-02	12.94241	764	60	RONCESVALLES AND QUEEN	N
2023-03	12.70601	898	57	QUEEN AND DUNN	W
2023-04	12.61771	926	60	GERRARD AND COXWELL	N
2023-05	12.17556	843	60	ST CLAIR AND EARLSCOUR	W
2023-06	12.76684	772	60	SHAW AND ADELAIDE	W
2023-07	12.10581	775	59	FLEET AND STRACHAN	S
2023-08	12.24009	858	60	BINGHAM AND KINGSTON	W
2023-09	11.48919	740	57	GERRARD AND JONES	W
2023-10	12.10845	959	60	EXHIBITION LOOP	E
2023-11	12.89410	1067	60	ST CLAIR AVE W AND KEE	W
2023-12	12.60935	1134	60	ST CLAIR AND OLD WESTO	E

Figure 6: Monthly summary

Figure 6 The summary table by month presents the mean delay times for each month around 11.49 to 13.24, which fits the Figure 1 as most delay times are around 10 minutes, that may slightly disrupt citizen's plan. However, some extreme cases up to 60 minutes delay can totally destroy the citizen's reliability of public transit, TTC company should try best to avoid the occurrence of extreme case.

Based on table, there is a increase trend of delay frequency through the year, public transit system should notice this situation and control the possibility of delay. West bound is the most serve direction of delay problem, TTC could consider optimize the route or timetable of the streetcar to reduce the possible delay. For most frequent delay's location, TTC should rewrite the estimate timetable based on actual situation, in order to make the estimate arrival time more exact, and increase the reliability of public transit.

References

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

Müller, Kirill. 2023. *Hms: Pretty Time of Day*. https://CRAN.R-project.org/package=hms.

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

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. 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.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://CRAN.R-project.org/package=knitr.