

Data analysis for 2023 streetcar's delay pattern in Toronto*

YITONG WANG

September 25, 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 occasionally happen. The main cause of delays are operational issue and mechanical problems, the finding provides an instructive direction which public transit should improve..

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 vary 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 dunds square. The statistic analysis show that overall delay's time within 1 hour, and concentrate in 10 minutes, some extreme cases also happen but with a less possibility. Due to the construction operating route of streetcar, most lines are east-west orientation, most delays occure in east-west bound.

The paper focus on the 2023 streetcar's delay data 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, we use 'knitr' (Xie (2023)). The

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

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 visualization of data and tabular summary also present the important trend of streetcar delay data.

2 Data

2.1 Data bacground

The data we use in this paper comes from opendatatoronto, the dataset ‘TTC Streetcar Delay Data’ we use provides detailed information on streetcar delays in Toronto in 2023. Each record contains the date and time of the delay, the route, the location where the delay occurred, the reason for the delay, and the duration of the delay. We select some interest column to figure out the pattern of occurrence of delay, summary the delay times and frequency for different variables. We cleaned the data, which removes row contain and abnormal values, such as N/A or 0, and limit the delay time interval within 1 hour interval to exclude the outlier, since most delay time does not over 1 hour. We ignore the ‘route’, ‘Line’, ‘Vehicle’ and ‘Min Gap’ in raw data since it does not contribute to our research.

2.2 Ethical focus

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.

3 Visualization

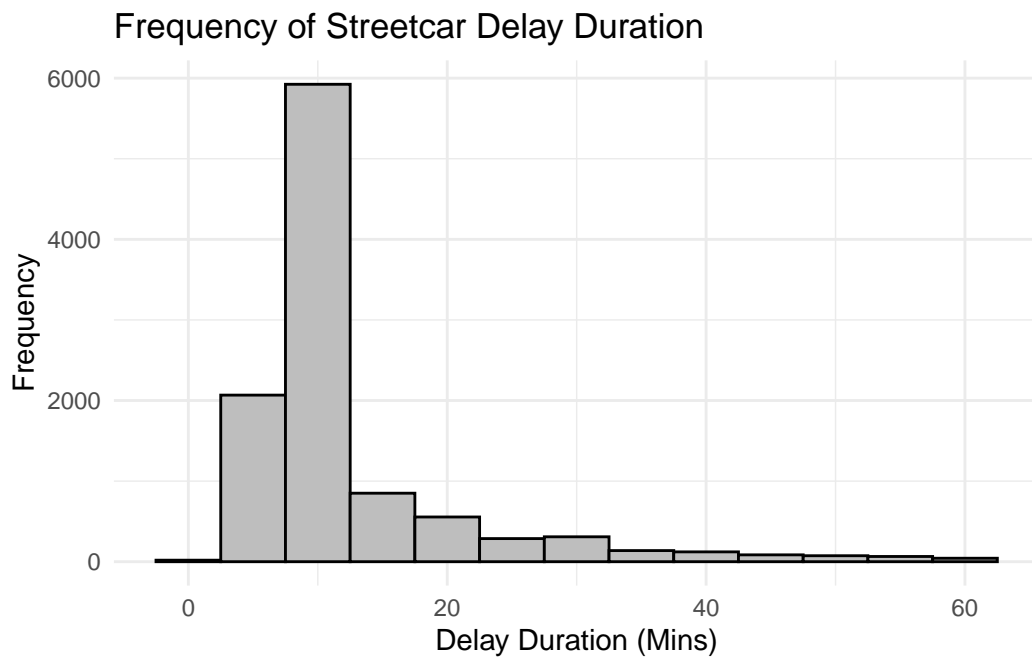


Figure 1: Distribution of streetcar's delay time frequency

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.

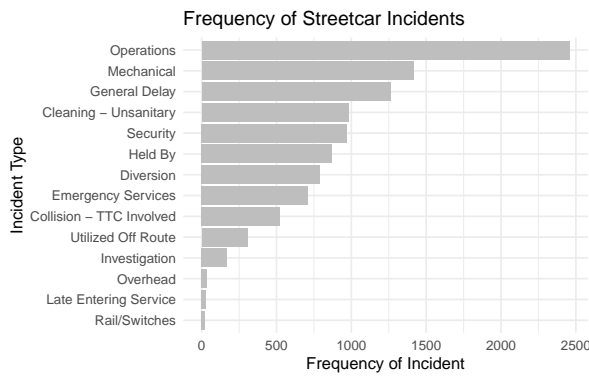


Figure 2: Frequency of incidents and incidents causing delay time

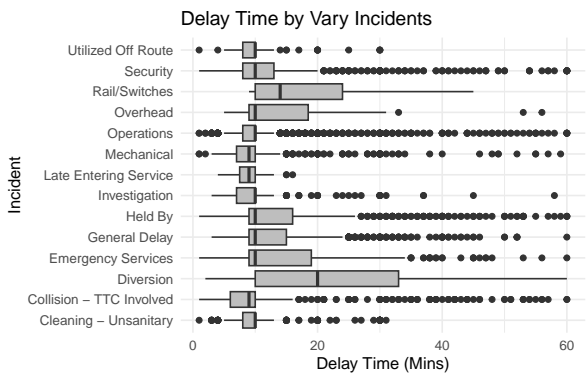


Figure 3: Frequency of incidents and incidents causing delay time

Figure 2 The bar plots 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.

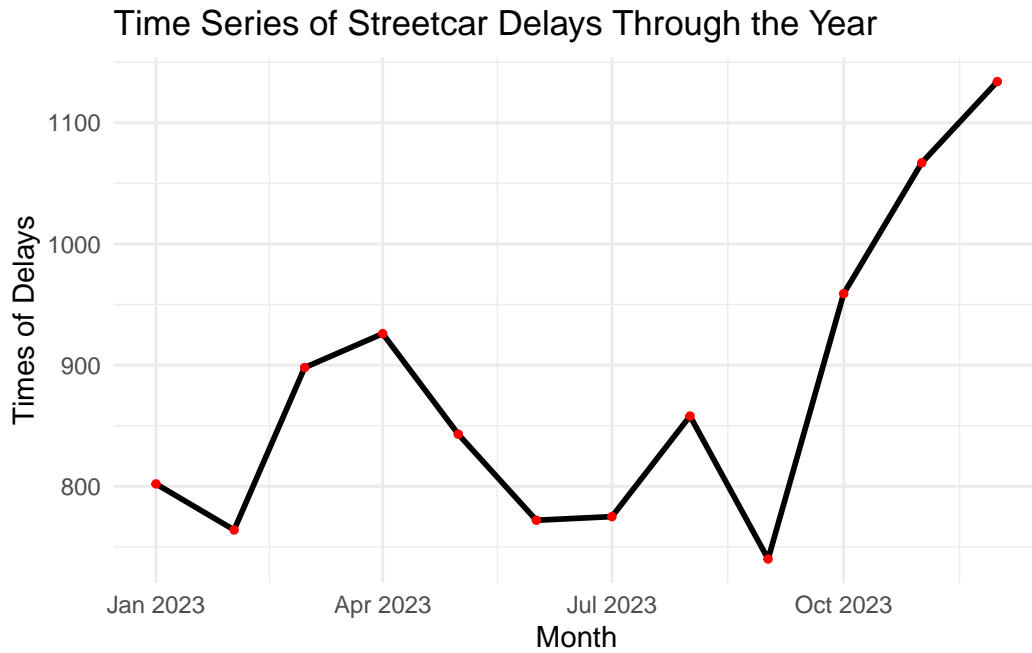


Figure 4: Trend of delay times by month

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.

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/>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Golemund, 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>.