TTC bus delays vary considerably based on time of day*

How long you wait depends on the type of delay, what route you take, and when you take it

Hudson Yuen

27 April 2022

Abstract

Despite ranking amongst North America's best transit systems, the TTC is often derided for irritating delays and service interruptions. This report examines TTC bus delay data, aiming to identify the factors influencing such delays and how improvements can be made. Certain incident types and days of service see disproportionately long delays, creating opportunities to address the most unpleasant aspects of the rider experience in a targeted manner. Doing so will not only strengthen the city's ability to move people, but to also facilitate sustainable development. **Keywords:** public transit, transit efficiency, urban mobility, sustainble transportation, transportation equity

Contents

*All data and scripts available at https://github.com/hudyu17/transit-delays

1 Introduction

A strong transit system is critical to the vitality and sustainability of urban centres, yet the Toronto Transit Commission (TTC) is still seen as sub-par by many city residents (Vella 2018). A key driver of this dissatisfaction stems from chronic delays and service cancellations, creating unpleasant knock-on effects such as overcrowding. The TTC has been named North America's transit system of the year as recently as 2017 (Muriel Draaisma 2017), yet this may reflect more on state of public transportation on the subcontinent than the merits of Toronto's network.

As the TTC breaks ground on ambitious projects like the Ontario Relief Line (Draaisma 2022), there remains plenty of opportunity to improve how existing services are run. Understanding the causes behind the most severe transit delays can help address issues in a targeted manner, relieving stress throughout the system. While some delays are implemented by design, such as weekend diversions to facilitate larger construction operations, the TTC should seek to mitigate their negative effects.

The future competitiveness and environmental sustainability of Toronto is intrinsically tied to the strength of the TTC. An inadequate transit system will hamstring growth and opportunity, both on individual scales and beyond (Patrick Miller 2016). Buses will remain a critical aspect of our transit network due to their relatively low cost of implementation and high operational flexibility, especially in the sprawling Greater Toronto Area. Understanding where delays are the most severe can not only help address current limitations, but also aid prediction and prevention of future bottlenecks as projects like the Ontario Line begin construction through the downtown core.

The remainder of this paper is structured as follows: Section 2 covers the data source and analysis methodology employed. Section 3 contains the statistical models used to derive the findings in this paper, with section 4 covering the results. Section 5 then examines key dimensions of the data, proposing conclusions and areas of further exploration.

2 Data

2.1 Data Source

This report examines TTC bus delay data from 2014 to 2021. This dataset was obtained from the City of Toronto Open Data Portal, available to the public on an ongoing basis since 2014. The data is refreshed monthly and was last updated in April 2022; however, this report only analyses data up until the end of 2021 for the sake of having complete annual data. The R package `opendatatoronto` (Gelfand 2020) was used to obtain all relevant data.

2.2 Methodology and Data Collection

The data collection methodology of original dataset is not fully clear from the source, and does not currently have a data quality score. Most features are of an automated nature and take advantage of the bus fleet's innate location-tracking capabilities, yet the nature of each delay - the "incident" feature - relies on self-reporting. Some data fields are empty or miscast as a result, necessitating data cleaning before detailed analysis.

Each delay event is associated with the following features: date, route, time, day (of week), location of the causal incident, delay duration from regular schedule in minutes, direction of bus route, and vehicle identification number.

`R` (R Core Team 2021) was the language and environment used for the bulk of this analysis, alongside the `tidyverse` (Wickham et al. 2019), `janitor` (Firke et al. 2021), `kableExtra` (Zhu et al. 2021), `dplyr` (Wickham et al. 2021), `AICcmodavg` (Mazerolle 2020), `Rmisc` (Hope 2013), and `data.table` (Dowle et al. 2021) packages.

2.3 Data Processing

The workflow begins with the `data-download.Rmd` script; data from each year was downloaded using the `opendatatoronto` package. The `data-cleaning.Rmd` script was then used to join each year's data into a single unified dataset, before removing anomalies to prepare the dataset for further analysis. This comprised the following steps:

- Renaming columns using `janitor` for consistency e.g. "Min Delay" to "min_delay"
- Recasting the time field from a string to a time type
- Dropping columns with limited data e.g. `incident_id` only had values for 5k/517k total rows
- Coalescing equivalent but differently-named columns e.g. `delay` and `min_delay`
- Dropping remaining rows with N/A values - 17.5k/517k total rows
- Removing rows with delay durations beyond the 99th percentile (235 minutes); those beyond were deemed outliers that had little interpretable business benefit

The resulting dataset was saved as `bus-clean.RDS` in the `inputs` directory, allowing it to be quickly loaded into other scripts with ease. This dataset was then brought into the `data-exploration.Rmd` script for exploratory analysis, creating the initial plots seen below.

2.4 Data Characteristics

Exploratory analysis revealed a potentially concerning trend; the average duration of delays appear to be steadily increasing since 2014, despite what seems like a downward trend in the number of delays. The downward trend in count also needs to be considered in the context of COVID, when delays would have decreased alongside the significant drop in ridership.

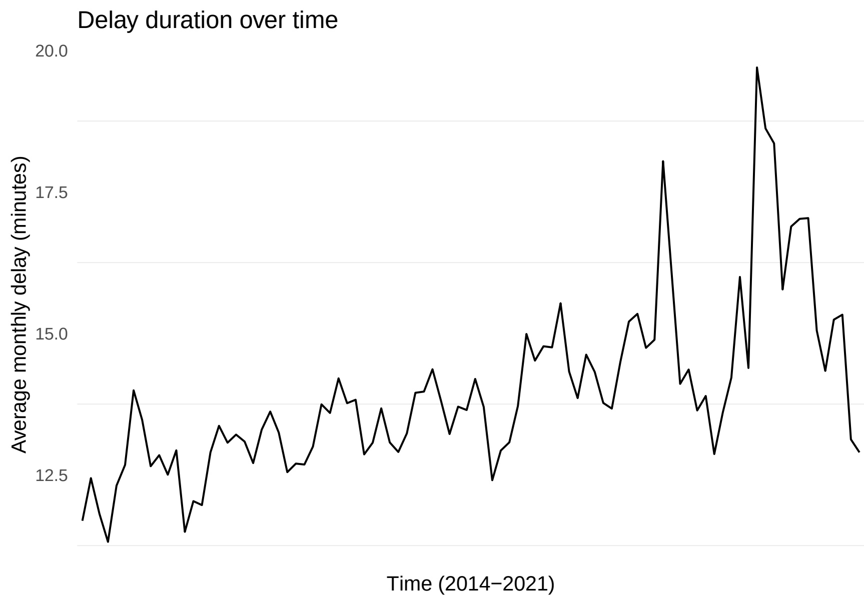

Figure 1: Delay duration over time

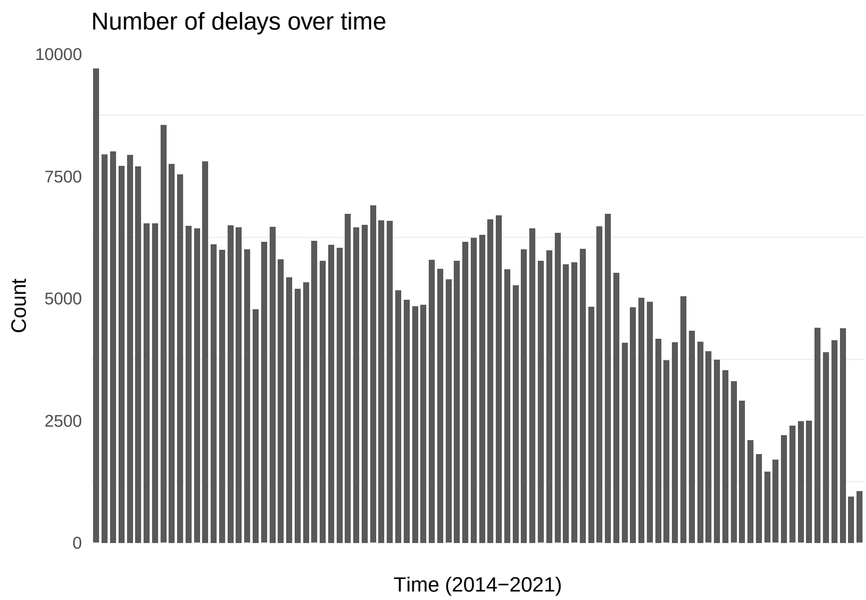

Figure 2: Number of delays over time

Segmenting the data by day also revealed interesting differences in delay duration. Perhaps similar to general trends, delay duration appears to be higher on the weekends despite fewer delays happening in the first place.

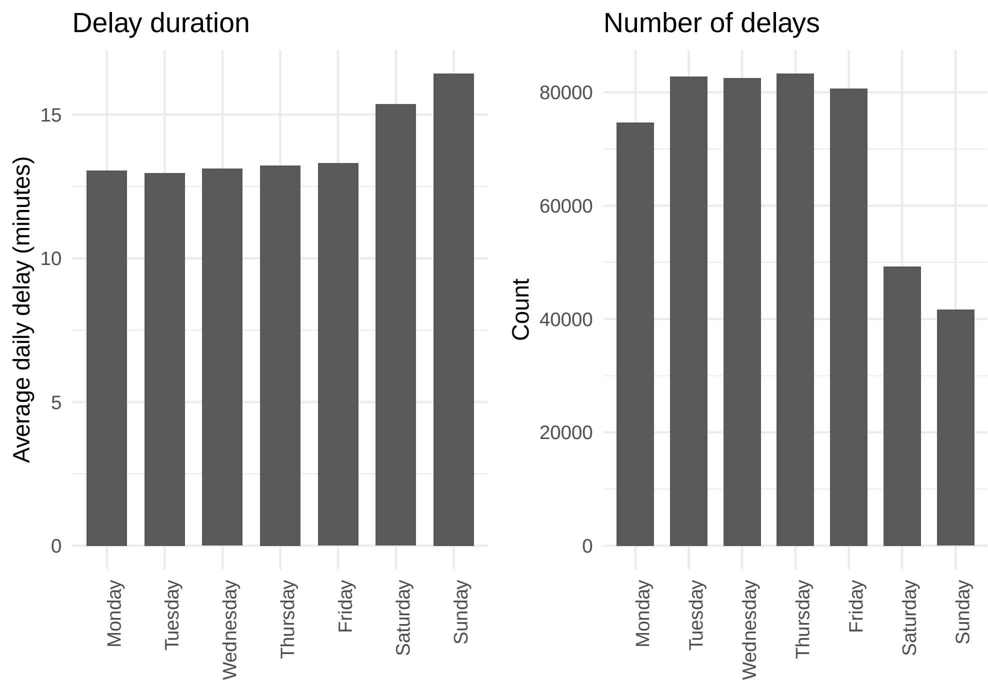

Figure 3: Delay behaviour by day of week

Finally, there are clearly incidents that occur more often, whilst others occur with much greater severity. For example, road blocks result in delays of over 50 minutes on average, but mechanical incidents happen with much more frequency.

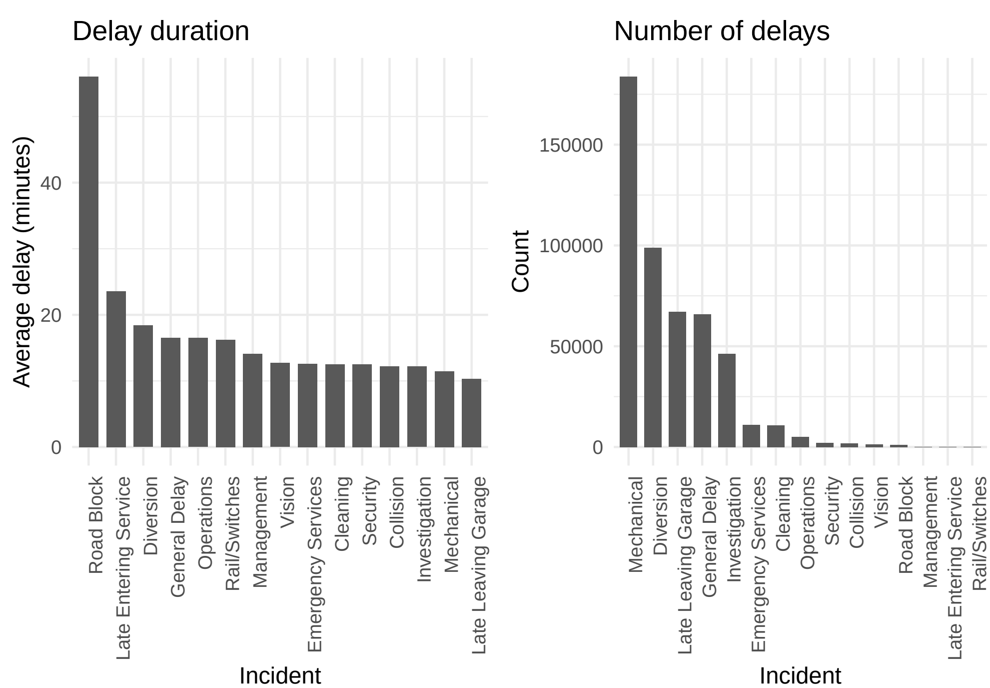

Figure 4: Delay behaviour by incident type

3 Model

3.1 Goals

The primary goal behind modelling was to deduce the causal factors behind delays, if any existed, to form the basis for further analysis in exploring how delays could be mitigated. Of the features in the dataset, the most relevant for modelling were: day, incident type, and a time-based metric (such as hour or minute). Location and route data provided valuable geographical context, yet they would be too granular without further aggregation; this grouping was unfortunately difficult to execute, as many locations were manually typed in with inconsistent syntax. This paper does isolate specific "worst-offender" routes for detailed deep-dives into why some routes consistently perform worst than others.

3.2 ANOVA

I set out to fit a model that best predicted delay duration based on some combination of these selected features. It was prudent to also determine whether these features had a true effect on delay duration; both these goals were achieved by fitting multiple ANOVAs and evaluating them based on a measure of prediction error, such as AIC. Although not exhaustive, 9 potential models were fitted to evaluate factors that influenced delay duration:

- The two-way impact of day, incident type, and hour = 3 models (a two-way ANOVA also examines the impact of individual variables)

  $Y_1 = \mu + \beta_0 + \beta_1$

- The interactions between 2 variables = 3 models

  $Y_1 = \mu + \beta_0 * \beta_1$

- The aforementioned interactions with the additive impact of the remaining feature = 3 models

  $Y_1 = \mu + \beta_0 * \beta_1 + \beta_2$

An ANOVA (Analysis of Variance) test evaluates whether certain features cause statistically significant - as opposed to random - differences between the means of more than 2 groups. This is done by examining the ratio of variance between groups to variance within groups; if this ratio is high, then it is more likely that there are true differences between groups.

3.2.1 Data Transformation

ANOVA tests assume that the data is normally distributed and homoscedastic, requiring further cleaning of the original data. The original distribution can be seen below and is heavily skewed right; a log transformation was applied, resulting in a more normal distribution.

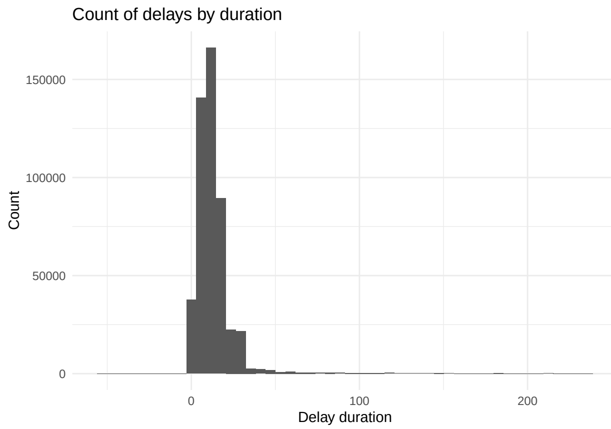

Figure 5: Histogram of delay durations

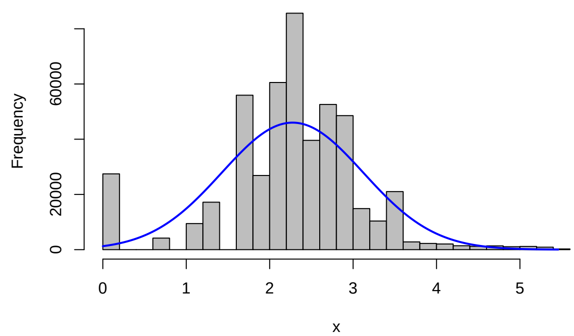

Figure 6: Histogram of delays after log transformation

3.2.2 Model and Feature Selection

The Akaike information criterion (AIC) was used to evaluate the quality of all models. This technique calculates a model's log-likelihood of existing given the provided data, resulting in an interpretable metric where a lower AIC represents a better fitting model. The `Delta_AICc` score shows the difference in AIC between the model in question and the best-fitting model.

Table 1: AIC scores for model evaluation

Modnames	AICc	Delta_AICc	LL
combine_day_hour_plus_incident	1192070	0.00000	-595852.1
combine_incident_hour_plus_day	1192121	50.94143	-595739.4
combine_day_incident_plus_hour	1194359	2288.80139	-597055.5
interaction_incident_hour	1194619	2548.97905	-596994.4
two_way_incident_hour	1197627	5556.26875	-598774.3
interaction_day_incident	1201802	9731.72083	-600800.0
two_way_day_incident	1202431	10360.83323	-601193.6
interaction_day_hour	1211305	19234.80507	-605483.5
two_way_day_hour	1213962	21891.52408	-606949.9

2 models were close in accuracy; the 2nd place model featured a `Delta_AICc` score of 50, compared to a score of 2288 and 2548 for the 3rd and 4th place models respectively. It was therefore determined that the first 2 models were effectively equal in accuracy, and the decision was made to use `combine_incident_hour_plus_day` as the model of choice; this followed the following:

$delay = day + incident * hour$

3.2.3 Post-hoc Tests

With our selected model demonstrating significant differences between groups, further analysis was performed to identify which specific groups differed from each other. The Tukey Honestly Significant Difference test (Tukey's HSD) was used to compare every possible treatment, a technique that improves upon multiple t-tests through utilising a pooled variance estimate and adjusted confidence intervals for multiple tests.

3.2.4 Follow up: Trends Over Time

After examining specific features, I examined whether delays changed significantly from year to year. Exploratory plots revealed a growing trend, so I aggregated delay duration by year and fit a linear regression.

$delay = \mu + \beta_0 X$

3.3 Limitations

There are several limitations to the modelling approach described above, chiefly regarding the natural distribution of data and preexisting outliers. The most egregious outliers beyond the 99th percentile were eliminated during the cleaning process, but over 37k delays (~7%) were deemed "outliers" beyond the upper IQR. It was difficult to justify removal of these points; when viewed on a histogram they appear to fit the natural distribution of the data.

The heavily right-skewed nature of the data also created issues for modelling. Despite a log transformation, diagnostic residual plots do not fully exhibit the desired trends. The resulting large log likelihood statistics is an effect of this.

Despite these shortfalls, I believe that this modelling still reveals key directional insights surrounding the causes of bus delays. For example, the estimated magnitude of impact for a mechanical failure at 11pm may not be accurate in isolation, but the helpful takeaway is that delays are frequently exacerbated at certain times of day. A careful balance should therefore be struck between interpreting model results as specific or more relationship-focused.

4 Results

The Tukey HSD test revealed that the following insights regarding significant differences in delay time:

- Days can be categorised into early, mid-week, and weekend; delay duration differs between groups, but less so within groups
- The `management, investigation, cleaning, and rail/switches` incidents are not a strong indicator of delay duration
- Delays are similar for certain time brackets; the 8-11pm and 1-4am time frames see similar delay durations in particular

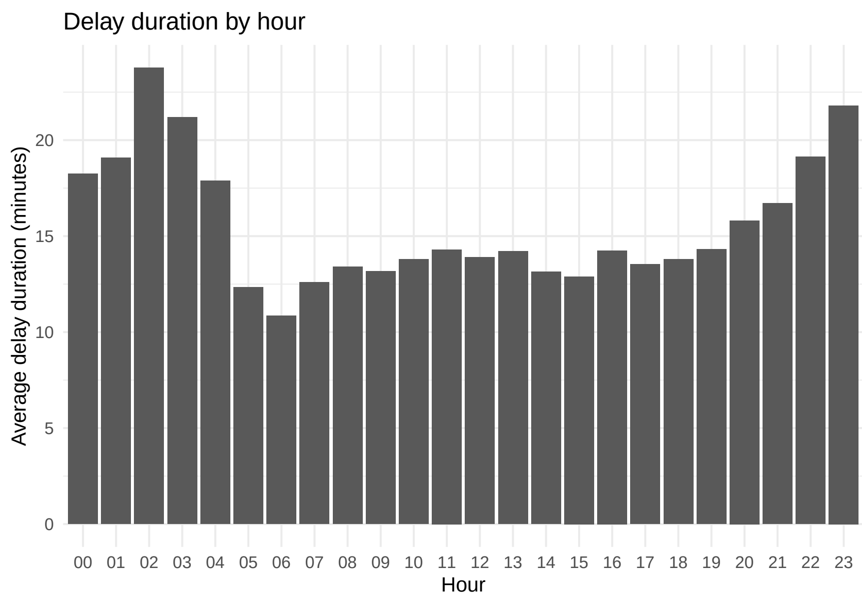

Figure 7: Average delay duration by hour

Due to the large number of categorical levels across the 3 factors, for example one level for each of the 24 hours, plotting every comparison would be impractical and difficult to interpret. Instead, boxplots of the individual features can be seen below.

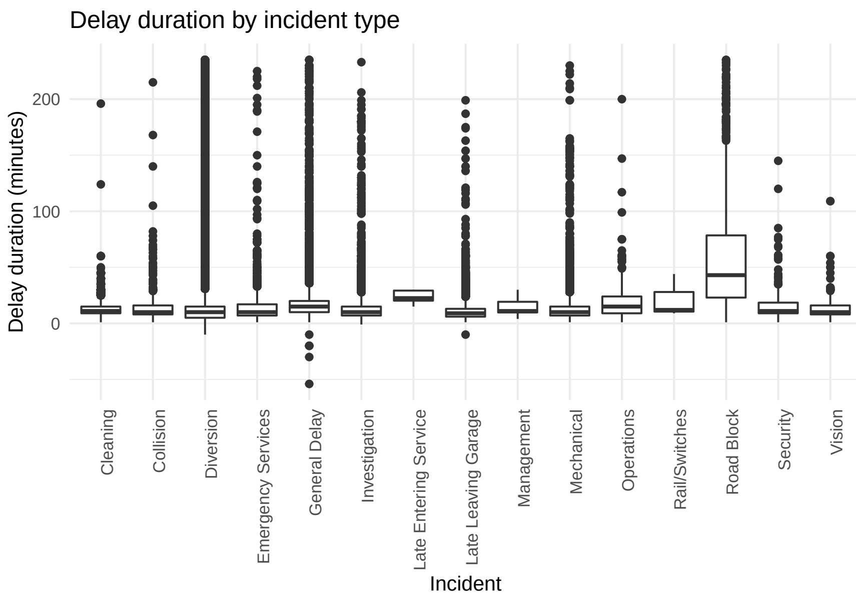

Figure 8: Delay duration by incident

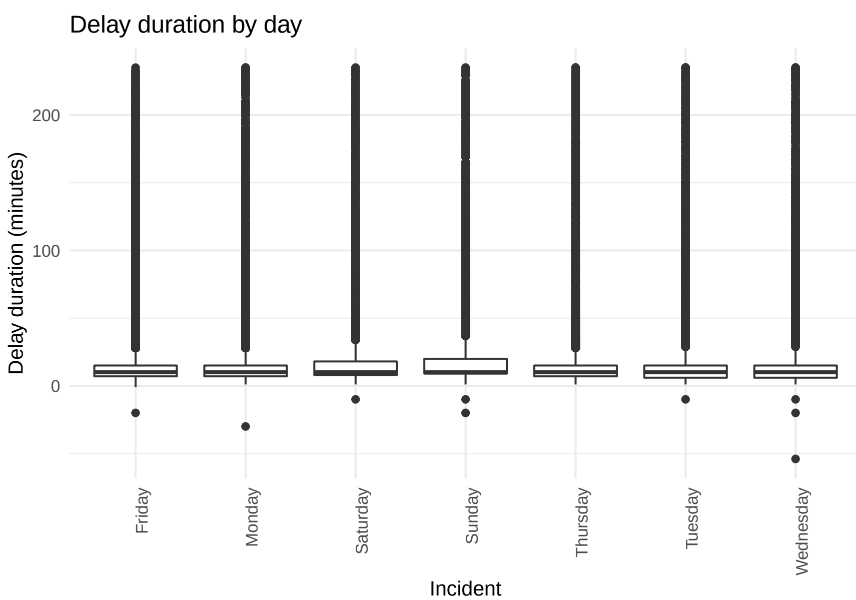

Figure 9: Delay duration by day of week

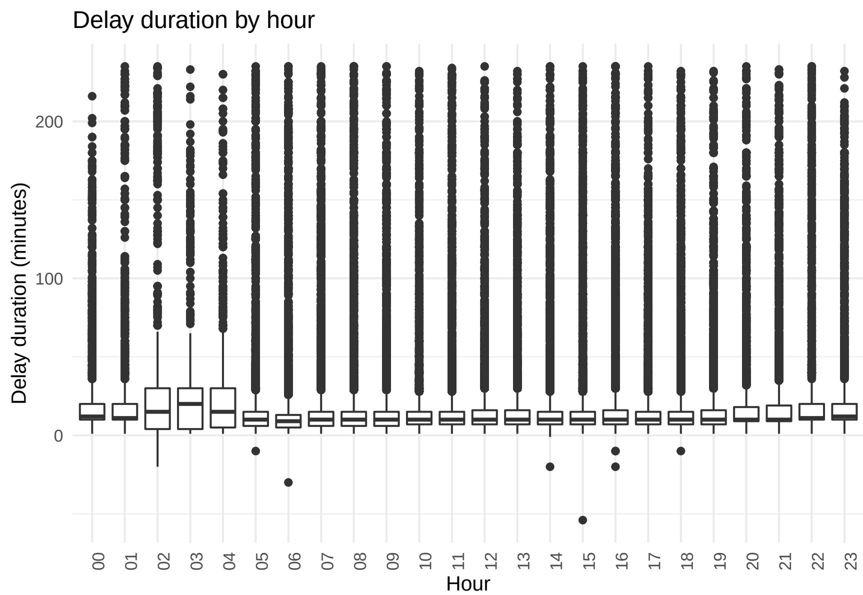

Figure 10: Delay duration by hour

The followup linear model showing delay duration over time can be seen below, with further splits by certain notable incident types.

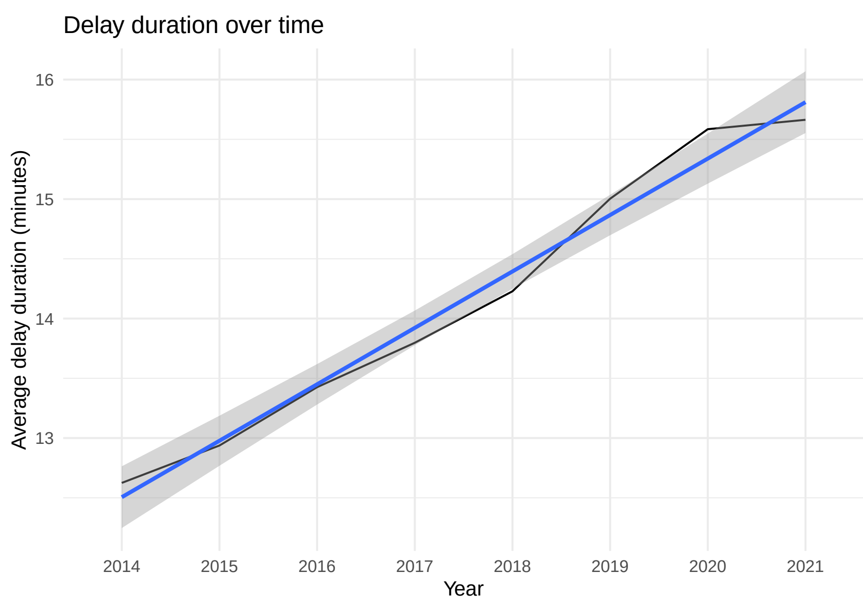

Figure 11: Delay duration over time

The 2 worst-performing routes see average delay durations over 50% worse than the next most delayed routes. Of the top 30 longest delayed routes, 26 are night buses (route numbers in the -300 range).

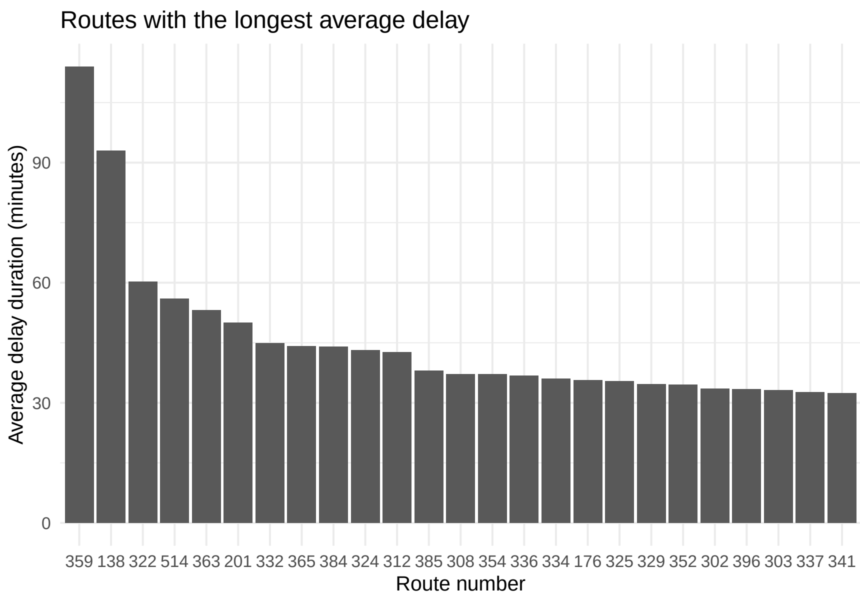

Figure 12: Most severely delayed routes

5 Discussion

5.1 The impacts of incident, hour, and day

The effectiveness of the incident:hour interaction suggests that these features compound on each other to exacerbate the duration of delays. For example, a mechanical incident at 3am is expected to be considerably worse than at 10am. This conclusion is unsurprising, but the modelling appears to reveal 2 windows - one late at night and the other early in the morning - that see much longer delays. It is likely that there are fewer resources and staff to handle delays during the entire night window, so these spikes could be a result of increased ridership placing undue pressure on the network; the 10pm rush may correspond with riders returning home before the final subway service runs for the night, and the 2am window with late-night revellers finally making the trip home. Given the unpleasant and potentially dangerous nature of having riders wait through extended delays at night, this presents an opportunity for the TTC to decrease delay times with potentially outsized returns in rider satisfaction and safety.

The day of the week also influences delay duration, with delays on the weekend seeing significantly longer durations. As mentioned in the model limitations, the specific magnitude of this influence is difficult to ascertain, but it is clear that weekends are a worse time to be stuck on a bus route. However, this interaction is likely by design; the TTC regularly schedules route diversions and upgrade works to occur on the weekend to avoid severe disruptions to general work schedules. This feature may therefore not reveal a particular opportunity for improvement; rather, it only serves to quantify the worsened weekend experience felt by many.

5.2 Delays are not getting shorter

Perhaps more concerning is how delay duration appears to be flat or increasing since 2014, despite continued public concern and greater recognition of the need for effective transit. This would not be as pressing an issue if TTC bus service had expanded accordingly, but service has remained largely equal from a frequency and coverage standpoint.

The reliability of night buses is also an area worthy of closer examination, given that the longest delays by far are found on these routes. It is unlikely that the nature of operating at night results in greater exposure to the severe incidents of a mechanical or diversion nature, suggesting that long delays are a result of an inadequate delay response.

5.3 Weaknesses

The quality of the dataset was a key blocker in carrying out a more rigorous analysis, resulting in limitations outlined earlier with regards to model accuracy. However, a structural limitation to this analysis is the limited feature set in the original dataset. The features included in the model were - unsurprisingly - the only features that could be aggregated or generalised: incident type, day of week, and datetime-based metrics (of which hour was created). Location-based analysis with route data would have added to the depth of this exploration, yet this would require extensive and potentially futile cleaning of various text fields.

6 Conclusions

The duration of bus delays are heavily influenced by the type of incident that occurs, the day of the week, and the time of day. Night services in particular appear to see the most severe delays, creating an opportunity to drastically improve the quality of bus service offered in a targeted manner. Other immediate findings reveal a delay situation that has generally worsened in severity since 2014, even if the total number of delays may be decreasing.

Going forward, more granular analysis on incident types and poorly performing routes will help determine specific areas to target and inform subsequent strategies. The lack of location-based analysis is a current weakness, for example, but such detail can greatly strengthen the sweeping findings of this initial analysis. The TTC has an opportunity to use such data to address past deficiencies before ridership resumes to pre-pandemic levels, creating a stronger and more resilient city for the future.

6.1 Enhancements

An additional datasheet was prepard for the dataset used in this analysis (link here)

References

Dowle, Matt, Arun Srinivasan, Jan Gorecki, Michael Chirico, Pasha Stetsenko, Tom Short, Steve Lianoglou, et al. 2021. *Data.table: Extension of 'Data.frame'.* https://cran.r-project.org/package=data.table.

Draaisma, Muriel. 2022. "Ontario Premier Says Construction Underway on New Toronto Subway Line." *Canadian Broadcasting Corporation (CBC).* https://www.cbc.ca/news/canada/toronto/ontario-line-official-breaking-ceremony-toronto-doug-ford-john-tory-1.6399282.

Firke, Sam, Bill Denney, Chris Haid, Ryan Knight, Malte Grosser, and Jonathan Zadra. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://cran.r-project.org/package=janitor.

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal.* https://cran.r-project.org/package=opendatatoronto.

Hope, Ryan M. 2013. *Rmisc: Ryan Miscellaneous.* https://cran.r-project.org/package=Rmisc.

Mazerolle, Marc J. 2020. *AICcmodavg: Model Selection and Multimodel Inference Based on (q)AIC(c).* https://cran.r-project.org/package=AICcmodavg.

Muriel Draaisma, Lisa Naccarato. 2017. "'No One Saying It's Perfect': TTC Named Best Transit System in North America." *Canadian Broadcasting Corporation (CBC).* https://www.cbc.ca/news/canada/toronto/programs/metromorning/ttc-named-best-transit-system-north-america-1.4178574.

Patrick Miller, Lina Kattan, Alexandre G. de Barros. 2016. *Public Transportation and Sustainability: A Review. KSCE Journal of Civil Engineering.* https://link.springer.com/article/10.1007/s12205-016-0705-0.

R Core Team. 2021. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Vella, Erica. 2018. "Frequent TTC Riders Dissatisfied with Service, Concerned with Overcrowding." *Global News.* https://globalnews.ca/news/4018448/frequent-ttc-riders-dissatisfied-overcrowding-poll/.

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.* Vol. 4. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Zhu, Hao, Thomas Travison, Timothy Tsai, Will Beasley, Yihui Xie, GuangChuang Yu, Stéphane Laurent, et al. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* https://CRAN.R-project.org/package=dplyr.