**TTC BUS DELAY 2022**

This report analyzes the Toronto Bus Delay dataset for 2022 to identify patterns and causes of bus delays and suggest ways to reduce them. The findings reveal that most routes experience delays, mainly due to disinfection, and most delays exceed the 15-minute target wait time. The report provides valuable insights for the TTC and stakeholders to improve the efficiency and effectiveness of the transit system.

* Introduction: This report analyzes the Toronto Bus Delay dataset for the year 2022. The report's objectives are to provide the Toronto Transit Commission (TTC) and the public with insights into the patterns and causes of bus delays in Toronto and suggest ways to reduce delays.
* Background: The Toronto Transit Commission (TTC) is the public transit agency that operates buses, subways, and streetcars in Toronto. The TTC collects data on bus delays to monitor performance and identify areas for improvement. This report was initiated by a group of data analysts interested in analyzing the TTC's bus delay data to identify patterns and trends. The dataset includes information such as the date and time of the delay, the bus route and number, the reason for the delay, and the duration of the delay. The need for this report arises from the importance of public transit in Toronto, where TTC buses are a critical mode of transportation for many residents. By analyzing the delays experienced by TTC buses, the report provides valuable information for stakeholders, including transit authorities, policymakers, and commuters.
* Purpose of the Analysis: The analysis is being done to provide insights and recommendations to improve Toronto's bus transportation system and reduce passenger delays. The findings can be used by transportation planners and policymakers to make data-driven decisions to improve the efficiency and effectiveness of the city's transit system. This analysis involves regular data analysis, where visualizations are used to gain insights into the data.
* Assumptions and Limitations: One assumption of the analysis is that the data in the dataset is representative of all bus delays on the TTC system. However, there may be delays that are not captured in the dataset. One of the limitations is that the analysis is based on data from 2022, which may not represent current performance due to changes in the system or external factors such as the COVID-19 pandemic.
* Hypothesis: Hypothesis: Ho: µ = µo

Ha: µ ≠ µo

The null hypothesis is that the average delay time is 15 minutes, while the alternative hypothesis suggests that the average delay time differs from 15 minutes.

Since we are evaluating if the average minute of delay is 15 minutes, we have the following setup: Ho: µ = 15 Ha: µ ≠ 15 with a confidence interval of 95%.

* Questions: The data analysis answers the following questions:
* What are the most common reasons for bus delays?
* Which routes have the highest delays?
* What is the average delay time?
* What are the busiest months for delays?
* What were the busiest days for delays according to the analysis?
* How long did TTC set the maximum wait time target for its transit services?
* How does the delay time vary by location and day of the week?
* Are there any patterns or trends in the data that suggest ways to reduce delays and improve the rider experience?
* Key Documents:
* Data source: <https://www.kaggle.com/datasets/reihanenamdari/toronto-bus-delay-2022>
* All visualizations are included in the appendix below.

1. Method

Our dataset was sourced from Kaggle and consisted of date, route number, time, day of the week, location, incident, minimum delay, minimum gap, direction, and vehicle. With a little bit of research, we found that TTC promotes 15 minutes maximum for transit waits. Therefore, this was our benchmark for our analyses. The data was already cleaned and did not end up having any missing or null values that had to be filtered. We used PowerBI for our visualizations where we can see that there are certain days and months that tend to have more delays overall. We can also see that the majority of routes have some delay from the target wait time of 15 minutes, with the majority of delays reported as being because of disinfection.

1. Results
   1. Quality of the Data

As the data was collected for company needs, the data was of very good quality and easy to work with, needing little manipulation for cleaning purposes. The dataset had no null or missing values that had to be cleaned and accounted for. However, the findings of the data pose serious improvements needed for many routes of the TTC.

* 1. Findings

As our visualizations show, there are many delays that need improvement upon, the main reason for which was disinfection. We also found that most routes experienced some delay, which could be up to 30 minutes of wait time, 15 minutes longer than the 15-minute maximum target wait time. We found that of the 6 months observed, January, March, and April experienced the highest wait times. The daily breakdown of wait time found that Sunday, followed by Saturday, were the busiest days with the longest minimum wait time at 15.04 minutes.

1. Appendix

* Data source: <https://www.kaggle.com/datasets/reihanenamdari/toronto-bus-delay-2022>
* Visualizations:









