

● ● ● Faculty Mentor: **Satayaki Das**

TTC Subway Delay Data Analysis

Data Analytics Capstone
Project | Group 17



Operational Context



Public transit reliability is critical for Toronto's economy; in 2025, subway delays resulted in approximately **1,179 lost service hours**. The lack of a structured framework at the TTC to identify root causes by line, station, or time has led to inefficient allocation of maintenance and workforce resources. This project aims to transform raw delay logs into an **interactive dashboard** to enable data-driven maintenance scheduling and workforce planning.

Context & Problem





Data Pipeline & Cleaning



Data was sourced from the City of Toronto Open Data Portal covering Jan–Dec 2025 with 25,713 rows. Cleaning included replacing 36.8% missing "Bound" (direction) values and standardizing over 20 inconsistent line names down to six canonical codes. Feature engineering added five columns — Month, Hour, Delay_Category, Time_Period, and Is_Delayed — to support temporal and severity analysis.

Pipeline overview

Key Performance Indicators

35.37%

Delay Rate stood at 35.37%, representing the significant share of incidents that caused actual service disruption across the entire network system.

42.2%

AM Peak Delay Rate is notably higher at 42.2%, highlighting a pronounced and significant impact on morning commuters during their daily travels.

70,754
minutes

Aggregate Total Delay Minutes reached 70,754 minutes, which equates to roughly 1,179 hours of operational loss and quantifies the scale of the problem.

Major Data Observations



Line 1 (YU Line) is the network's weakest performer with a delay rate of 38.7%, indicating specific and targeted operational issues on that particular line.

Line performance



Seasonality analysis shows February as the worst month with 917 incidents, likely driven by harsh winter weather conditions affecting operations.

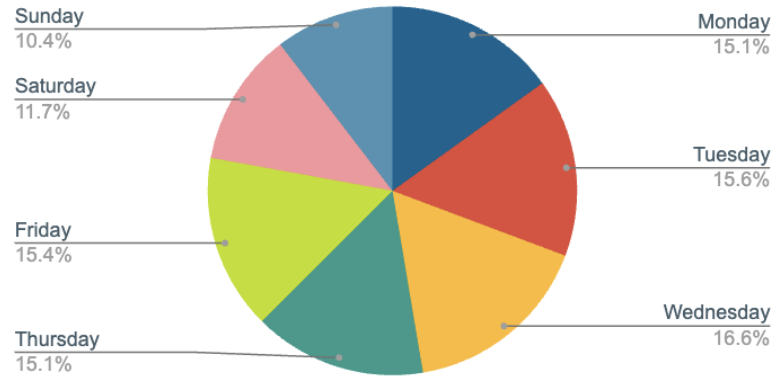
Seasonality



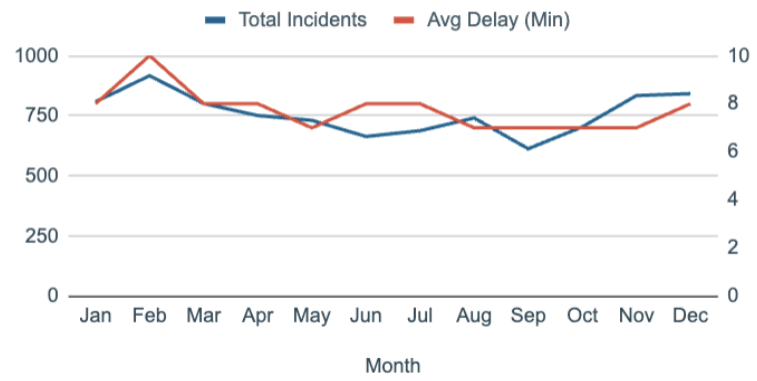
Sundays have the fewest incidents (946) but the highest average delay (10 minutes), likely due to reduced staffing or slower recovery on low-service days.

Day-of-week pattern

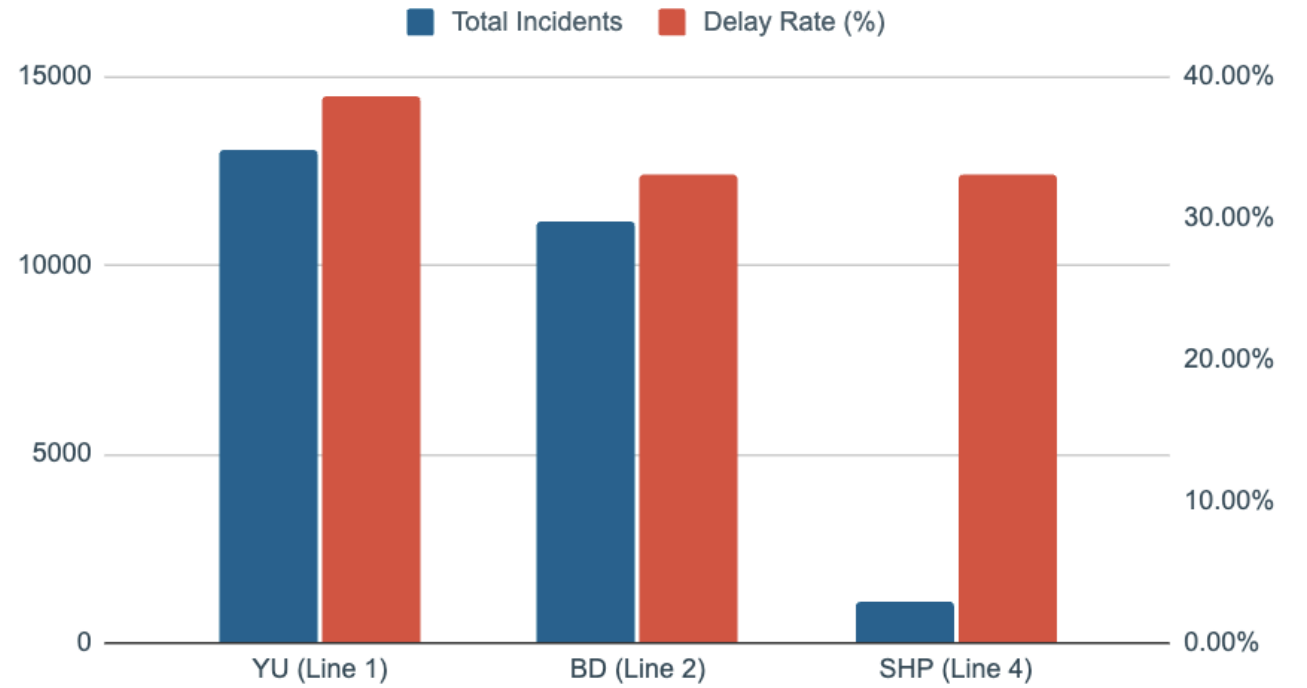
DAY OF WEEK (Avg Delay)



MONTHLY TREND (SEASONALITY)

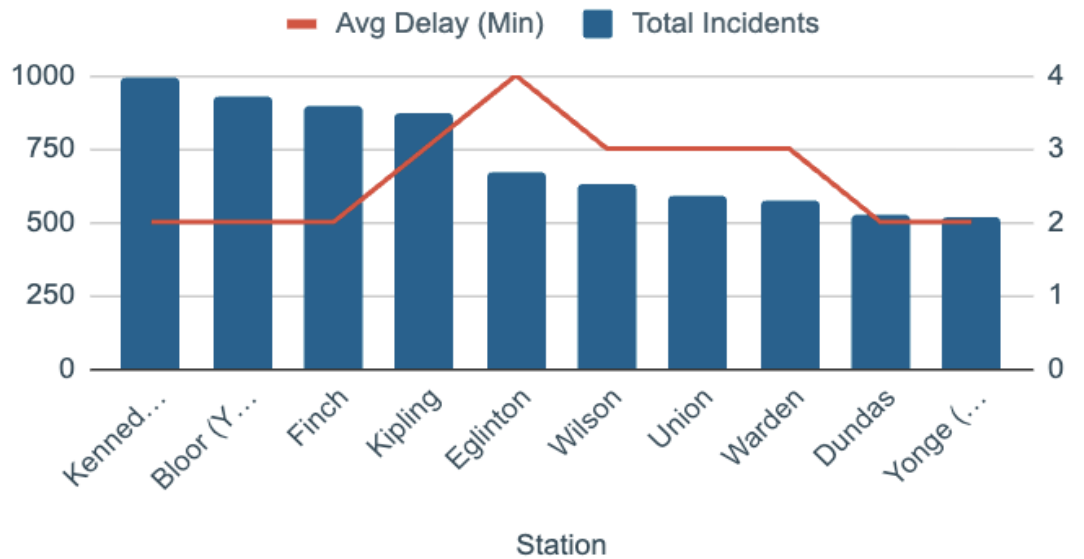


LINE PERFORMANCE



Segmentation & Severity

TOP 10 WORST STATIONS



Pareto analysis shows top 10 stations (12% of network) cause 27% of incidents, highlighting intervention hotspots. 64.6% of logs are 'No Delay', but 2.3% 'Major' (over 16 mins) have big impacts. Stations like Kennedy and Bloor, with severe delays, need priority fixes.

Pareto hotspots & severity

Operational Dashboard



The proposed dashboard provides an executive view with top-row KPI cards displaying Total Incidents and Delay Rates for quick decision-making. Visualizations include a "Top 10 Worst Stations" bar chart to spotlight hotspots and a "Monthly Seasonality" line chart to track trends and identify peak months. These components together enable managers to prioritize short-term operational fixes and long-term capital investments.

Dashboard overview

TTC Subway Delay Analysis 2025

This project contains a comprehensive analysis of Toronto Transit Commission (TTC) subway delay data for the year 2025. The dataset includes detailed information about delay incidents across the TTC subway network, providing insights into operational patterns, peak delay periods, and affected stations.

