TTC Delay and Traffic Collision Analysis Report: Insights by Membership Type

Prepared by: Nihith Mekala
Tools Used: SQLite, DBeaver, Tableau Public

---

1. Executive Summary

This report analyzes the relationship between traffic collisions and subway delays within Toronto in 2023. The public often assumes a strong correlation between these two issues, particularly during peak hours. This analysis uncovers:

- **No Strong Correlation:** Subway delays are not driven by traffic collisions; internal operational issues dominate.
- **Hourly Patterns:** Subway delays peak during early mornings (6 AM) and late evenings (10 PM), while collisions peak during rush hours (8 AM, 5 PM).
- **Monthly Trends:** Subway delays remain steady year-round despite minor fluctuations in collisions.

The study uses SQLite for data preparation, DBeaver as the database client, and Tableau Public for creating interactive visualizations.

---

2. Data Preparation and SQL Workflow

2.1 Data Source

- **Traffic Collision Data:** Counts and severity of collisions (hourly, monthly).
- **Subway Delay Data:** Delay frequency, average durations (hourly, monthly).

Both datasets were originally provided as **12 Excel files**—one for each month.

2.2 SQL Workflow for Data Preparation

Step 1: Import Excel Tables
All 12 Excel files were imported into SQLite as separate tables: collisions_jan, delays_jan, etc.

Step 2: Combine Tables into Unified Tables

The following query merges all monthly tables into two unified tables: all_collisions and all_delays.

sql
Copy code
```
-- Combine all collision data
CREATE TABLE all_collisions AS
SELECT * FROM collisions_jan
UNION ALL
SELECT * FROM collisions_feb
UNION ALL
...
SELECT * FROM collisions_dec;

-- Combine all delay data
CREATE TABLE all_delays AS
SELECT * FROM delays_jan
UNION ALL
SELECT * FROM delays_feb
...
SELECT * FROM delays_dec;
```

Step 3: Enrich Data with Additional Columns

Using **JULIANDAY** and **strftime** functions, additional analysis columns were added:

- **Ride Duration in Minutes (Subway Delays):**

sql
Copy code
```
ALTER TABLE all_delays ADD COLUMN delay_duration_minutes REAL;
UPDATE all_delays
SET delay_duration_minutes = (JULIANDAY(end_time) - JULIANDAY(start_time)) * 1440;
```

- **Hour of Day:** Extracted from timestamps.

sql
Copy code
```
ALTER TABLE all_delays ADD COLUMN hour_of_day INTEGER;
UPDATE all_delays
SET hour_of_day = strftime('%H', start_time);
```

- **Day of Week and Month-Year:**

sql
Copy code
```
ALTER TABLE all_collisions ADD COLUMN day_of_week TEXT;
UPDATE all_collisions
SET day_of_week = strftime('%w', collision_time);

ALTER TABLE all_collisions ADD COLUMN month_year TEXT;
UPDATE all_collisions
SET month_year = strftime('%m-%Y', collision_time);
```

Step 4: Remove Erroneous Data
Outliers, such as negative durations, were filtered:

---

3. Analytical Approach

- **Step 1:** Aggregate collision and delay data by hour and month.
- **Step 2:** Identify trends using SQL queries.
- **Step 3:** Use Tableau Public for visual analysis and insights.

---

4. Visualizations and Insights

4.1 Hourly Trends: Traffic Collisions vs. Subway Delays

Visualization: Line Chart

Key Observations:

- **Collision Peaks:** 8 AM and 5 PM (rush hours).
- **Subway Delays:** Peak at 6 AM and 10 PM, indicating operational issues unrelated to collisions.

Insight: Delays during early mornings and late nights are likely caused by internal factors like train startups, maintenance, or staff shortages—not external traffic incidents.

4.2 Average Delay Duration by Hour

Visualization: Horizontal Bar Chart

Hour	Average Delay Duration
6 AM	4.3 minutes
10 PM	3.8 minutes

Key Observations:

Delays are longest at 6 AM, suggesting startup issues like incomplete maintenance or equipment failures.

- Delays in late evenings (10 PM) align with maintenance scheduling or reduced workforce.

Insight: Internal operational inefficiencies significantly impact service reliability during critical time windows.

4.3 Monthly Trends: Collisions vs. Delays

Visualization: Combined Line and Bar Chart

Key Observations:

- Traffic collisions peak slightly in March and October but remain stable overall.
- Subway delays show no significant month-to-month variations.

Insight: Seasonal changes in traffic do not translate into significant disruptions in subway operations. This further disproves the assumption of a direct correlation.

4.4 Hourly Correlation Between Collisions and Delays

Visualization: Scatter Plot

Key Observations:

- Hours with the highest collision counts (8 AM, 5 PM) do not coincide with prolonged delays.
- Early morning delays occur when collision counts are minimal.

Insight: Subway delays are decoupled from surface-level traffic trends. Operational challenges (e.g., signal failures, crew rotations) are likely responsible.

---

5. Key Analytical Insights

1. **No Strong Correlation:**
   Traffic collisions have minimal impact on subway delays.
2. **Early Morning Challenges:**
   Operational inefficiencies cause the longest delays at 6 AM.
3. **Steady Monthly Trends:**
   Subway delays remain stable year-round despite fluctuations in collision rates.

---

6. Recommendations

1. **Optimize Early Morning Operations:**
    - Improve startup protocols to reduce early delays.
    - Assign maintenance crews specifically to address critical early-hour bottlenecks.
2. **Invest in Predictive Maintenance:**
    - Use data-driven solutions to anticipate equipment failures and minimize disruptions.
3. **Enhance Public Communication:**
    - Educate commuters on delay causes to reduce frustration and manage expectations.
4. **Resource Allocation:**
    - Focus staffing and technical resources during critical delay windows (6 AM, 10 PM).
5. **Deeper Analysis:**
    - Explore geographic delay trends across subway lines to pinpoint recurring issues.

---

7. Tools and Methodology

- **SQLite:** For data preparation, combining tables, and performing calculations.
- **DBeaver:** Used as the database client for managing queries.
- **Tableau Public:** Visualized hourly, monthly, and correlation trends to uncover insights.

---

8. Conclusion

This analysis reveals that subway delays in Toronto are largely driven by internal operational inefficiencies, particularly during early mornings and late evenings. Contrary to public perception, traffic collisions have minimal influence on delay patterns. By focusing on maintenance improvements, optimized resource allocation, and transparent public communication, the TTC can enhance service reliability and commuter satisfaction.

Appendix: SQL Code Summary

1. **Combining Monthly Tables:**
   UNION ALL query to merge all data.
2. **Adding Analysis Columns:**
    - Ride durations using **JULIANDAY**
    - Hours, days, and months using **strftime**
3. **Data Cleaning:**
   Removing outliers and negative durations.