Proceedings of CONF-MLA 2026 Symposium: Learning and Decision Making in Multi Agent Software Systems
DOI: 10.54254/2755-2721/2026.BA36256

Analysis of Temporal and Spatial Patterns of TTC Subway Delays in Toronto Based on Delay Data

Yifei Liu

Faculty of Arts & Science, University of Toronto, Toronto, Canada
felixyf.liu@mail.utoronto.ca

Abstract. This study examines the temporal and spatial patterns of Toronto Transit Commission (TTC) subway delays in Toronto using the TTC Subway Delay Data. The dataset contains 28,191 delay records from January 1, 2025, to January 31, 2026, and the 2025 records are selected for analysis. Descriptive statistics and data visualization are used to examine hourly and station-level delay patterns. A random forest classification model is developed to predict whether at least one delay incident will be recorded during a specific date-hour period at Union Station. The results show that TTC subway delays are unevenly distributed across time and stations. Delay incidents are more frequent around 22:00, while several major stations, including the Kennedy Bloor-Danforth Line (BD), Bloor, Finch, Kipling, and Union Station, record relatively high numbers of delay incidents. The classification model achieves relatively high recall but low precision, indicating that it identifies many date-hour periods containing recorded delay incidents.

Keywords: TTC subway, delay pattern analysis, delay-risk classification, random forest classification

1. Introduction

The Toronto Transit Commission (TTC) is one of the major public transportation systems in Toronto and plays an important role in supporting daily commuting and urban mobility. TTC has long been regarded as a highly integrated transit system in North America [1]. As subway services connect residential areas, employment centers, schools, and major transfer hubs, their reliability directly affects passengers' travel efficiency and transfer experience. Subway delays may cause missed connections, longer waiting times, and uncertainty in route planning. Therefore, analyzing TTC subway delay records is meaningful for understanding when and where delays are more likely to occur and for providing data-based support for improving service reliability.

Previous studies have shown that data-driven methods have been widely applied to train and urban rail delay prediction. Tiong et al. reviewed data-driven approaches to train delay prediction and emphasized that accurate prediction is important for railway planning, operation management, and passenger service quality [2]. Spanninger et al. summarized train delay prediction approaches and classified existing studies according to different modelling paradigms, including data-driven, event-driven, and mathematical methods [3]. In terms of model selection, Sarhani and Voß showed that open data sources can support rail transit delay prediction with machine learning models [4].

DOI: 10.54254/2755-2721/2026.BA36256

Lapamonpinyo et al. applied random forest, gradient boosting machine, and multilayer perceptron models to real-time passenger train delay prediction, while Shi et al. used XGBoost and Bayesian optimization to predict train arrival delay [5, 6]. Chen et al. compared statistical models and machine learning models for incident delay prediction in urban railway systems, showing that both types of methods can be useful depending on the prediction target and data structure [7]. In addition, Klumpenhouwer and Shalaby used railway delay logs and machine learning to support passenger railway operations, while Wang et al. applied statistical learning methods to analyze train delays and external climate influences [8, 9].

In summary, existing studies have made significant progress in railway delay prediction, especially in the use of open data, statistical models, machine learning models, delay logs, and operational information. Previous studies provide useful methodological references for analyzing TTC subway delay records. However, most existing research focuses on general railway systems, passenger train routes, or incident delay duration. More research is required for the TTC system in Toronto, as there is a research gap.

Based on this research background, this study examines TTC subway delays in 2025 through temporal and spatial analysis. It identifies delay patterns in 2025 and uses Union Station as a case to analyze station-level delay-risk classification. The purpose of this research is to help passengers understand the delay patterns in order to arrange their daily commute plan.

2. Method

2.1. Data source and description

This project uses the TTC Subway Delay Data as the main dataset, which is published by the Toronto Transit Commission through the City of Toronto Open Data Portal [10]. It records subway delay incidents in the TTC system. The original dataset used in this study contains 28,191 delay records from January 1, 2025, to January 31, 2026. This project uses the annual data of 2025 for analyzing various patterns of the TTC subway.

Each record in the dataset represents one subway delay incident. The dataset contains several types of information, including the time, station, cause of the delay, and delay duration. Among these variables, Min Delay is the main measurement of delay duration, while Min Gap reflects the time gap between trains after a delay incident. The Code variable is a categorical variable used to indicate the cause or type of each delay incident. For example, different delay codes may represent signal problems, security incidents, mechanical issues, operational problems, or other service interruptions (Table 1).

Table 1. Sample subway delay data

Field name	Data type	Description	Example
Date	object	Date	2025/1/1
Time	object	Time (24 hours)	2:10
Station	object	Station name	BATHURST
Code	object	Delay code	MUSAN
Min Delay	int	Delay (min)	5

DOI: 10.54254/2755-2721/2026.BA36256

2.2. Introduction to method

2.2.1. Data preprocessing

The data preprocessing process includes five main steps. Firstly, the original dataset is filtered by date, and only the 2025 records are retained. Second, the Date and Time variables are converted into datetime format, which makes it possible to extract additional time-related features, such as hour, month, and day of week. Third, categorical variables like Station, Line, Bound, and Code are checked and cleaned to make the formatting consistent. Next, the Min Delay variable is used to classify incidents of different delay times. For instance, delays of 0 minutes, 1-5 minutes, 6-15 minutes, and over 15 minutes are divided into different groups of severity levels, which helps compare the distribution of minor and serious delays.

For the Union Station classification task, the original data could not be used directly to create two classes because every row represented a delay that had already occurred. The data were reorganized so that each observation represented one hour on a specific date. All date-hour combinations for Union Station during the study period were included, even when no delay was recorded. A binary variable called Delay Occurred was then created. It was coded as 1 when at least one delay record appeared within a date-hour period and as 0 when no delay was recorded. This variable only shows whether a recorded delay occurred during that period. It does not measure the proportion of delayed trains among all trains.

2.2.2. Exploratory data analysis

The exploratory analysis focused mainly on time and station patterns. For the temporal analysis, the 2025 records were grouped by hour to examine which parts of the day had more delayed incidents. The records were also summarized by day of week to compare the distribution of delays across the seven days. For the station-level analysis, the number of delay records at each station was counted and ranked. Stations with higher numbers of recorded incidents were then shown in a chart. The same counting method was also used to compare delay frequency across TTC subway lines. Delay codes and delay duration were included as supporting information. The frequency of each code was used to identify common types of delay incidents, while the grouped values of *Min Delay* were used to compare different levels of delay severity. Together, these summaries describe the main features of the 2025 TTC delay records in terms of time, location, type, and duration.

2.2.3. Machine learning modeling

After the descriptive analysis, a classification model was built using Union Station as a case study. Union Station was selected because it is a major transfer hub in Toronto and has a relatively large number of delay records in the dataset. The model does not predict the actual proportion of delayed trains. The TTC dataset lists recorded delay incidents but does not provide complete information about trains that operated normally during each period. Given this limitation, the model estimates whether at least one delay record is likely to appear at Union Station within a particular date-hour period. A random forest classifier was used for this task. It was chosen because the relationship between time-related features and delay occurrence may not follow a simple linear pattern. Random forest builds several decision trees using different samples and combinations of features, then combines their predictions. Compared with a single decision tree, this approach is less dependent on one particular split of the data and can reduce the effects of noise and overfitting.

DOI: 10.54254/2755-2721/2026.BA36256

Since the task is a classification problem, the final prediction is based on the combined decisions of the individual trees. Each date-hour period is classified as either a delay period or a non-delay period. Random forest can also handle more complex relationships between the input variables and provide feature-importance results, which can be used to compare the role of different time-related features.

In this study, the machine learning task is defined as a binary classification problem with the target variable *Delay Occurred*, which records whether there is a delay incident taking place at a certain time slot on a certain day of the week. The model therefore predicts the occurrence of a recorded delay event rather than the delay duration or the true operational delay rate of all subway trains. The input variables include day of week, hour of day, weekend status, peak-hour status, and cyclical hour features. The cyclical features are created by transforming the hour variable using sine and cosine functions so that the model can recognize the continuous relationship between late-night and early-morning hours. The observations are arranged chronologically and divided into training and testing sets. The earlier 80% of the observations are used to train the model, while the remaining 20% are used for out-of-sample testing. This chronological division reduces the risk of using later observations to predict earlier periods.

This study uses a random forest classifier to estimate the probability that each date-hour observation belongs to the positive class. The random forest combines the results of multiple decision trees, so its final prediction result does not overly rely on a specific data partition rule compared to a single decision tree. The periods classified as positive by the model are called high-risk periods, while the periods classified as negative are called low-risk periods. However, these categories are only the risk classification results generated by the model and should not be understood as the direct measurement result of the proportion of delayed trains among all subway trains. This study evaluates the model's ability to identify delay records using accuracy, precision, recall, and F1 score. Accuracy represents the proportion of correct predictions in all predicted results of the model. Precision represents the proportion of periods predicted as high-risk by the model that actually contain delay records. Recall represents the proportion of periods with actual delay records among all periods that are successfully identified by the model. The F1 score integrates the precision and recall rates and evaluates the overall performance of the model in these two indicators.

3. Results & analysis

3.1. Analysis and visualization of delay data

In order to study how the TTC delay pattern varies during different time slots of the day, this study classified all the delay incident records according to different hours of the day. Figure 1 shows a significant difference in different time slots, where delay incidents at 8 am and 10 pm took up the largest proportion. The peak that happens at night may be related to the operation arrangement, as TTC announces that the gap between peak hours is usually 2 to 3 minutes, whereas this time increased to 4 to 5 minutes during other times of the day [11]. So, it explains why there are more delay incidents after peak commuting hours.

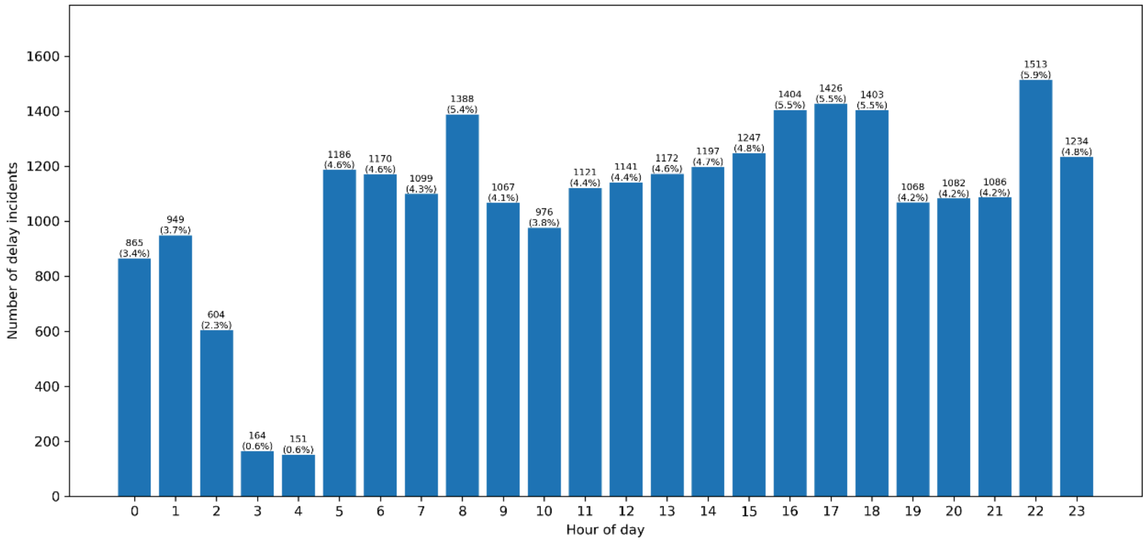

Figure 1. Hourly distribution of TTC subway delays in 2025 (data source: TTC Subway Delay Data [10]; figure credit: original)

Further analysis of the stations with the most delay incident records is then conducted. Figure 2 shows the top 10 TTC stations by delay incidents in 2025, with Kennedy BD Station recording the largest number of delay incidents, followed by Bloor, Finch, and other stations.

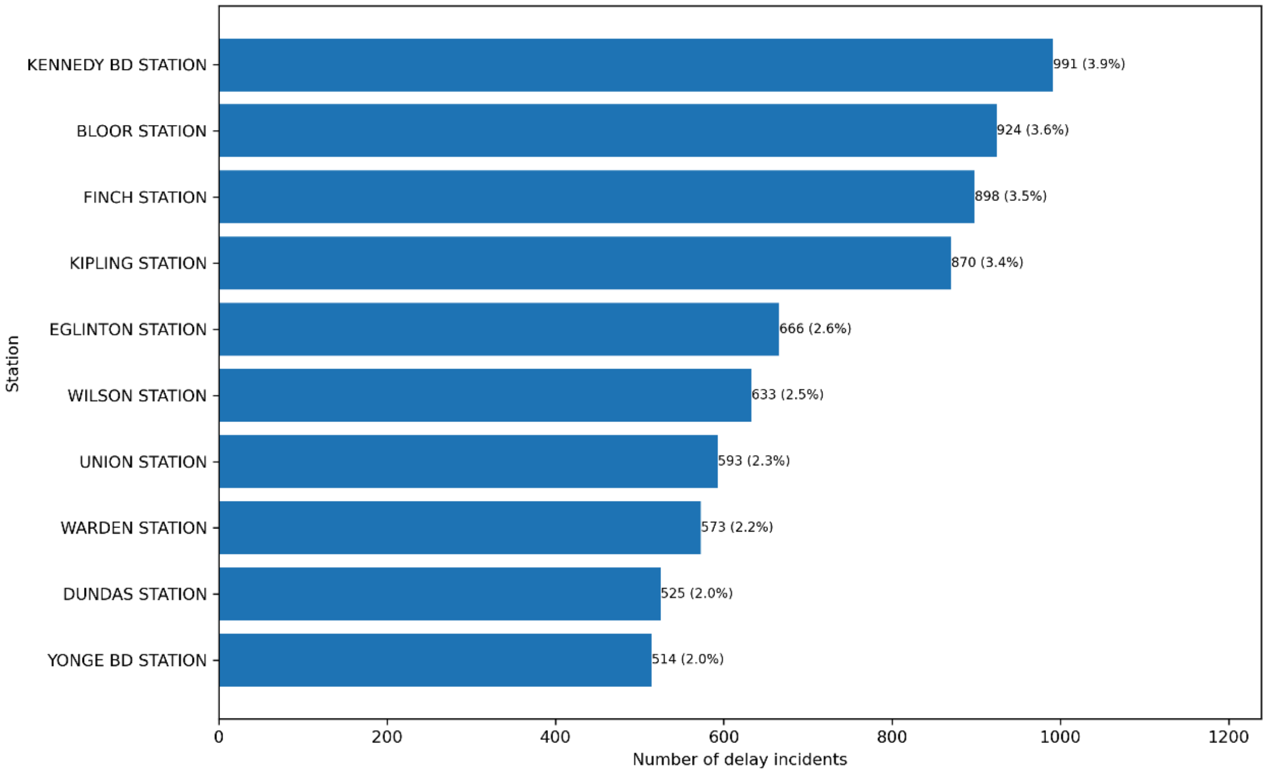

Figure 2. Top 10 TTC stations by delay incidents in 2025 (data source: TTC Subway Delay Data [10]; figure credit: original)

DOI: 10.54254/2755-2721/2026.BA36256

The frequent records of delay incidents at these stations may be related to station function and network positions since some of the top stations are terminal stations, transfer stations, or major boarding and alighting locations. Union Station is also included in the top 10, which supports its selection as a station-level prediction case.

3.2. Delay-risk classification of Union Station

To further examine delay patterns, Union Station is selected as a classification case because it is a major transfer station and contains a large number of delay records for station-level analysis. The random forest model estimates the probability that at least one delay incident will be recorded during each date-hour observation in the testing set. Figure 3 presents the average predicted probability of a recorded delay incident at Union Station by hour of day, then grouped by hour and averaged to show the general hourly pattern.

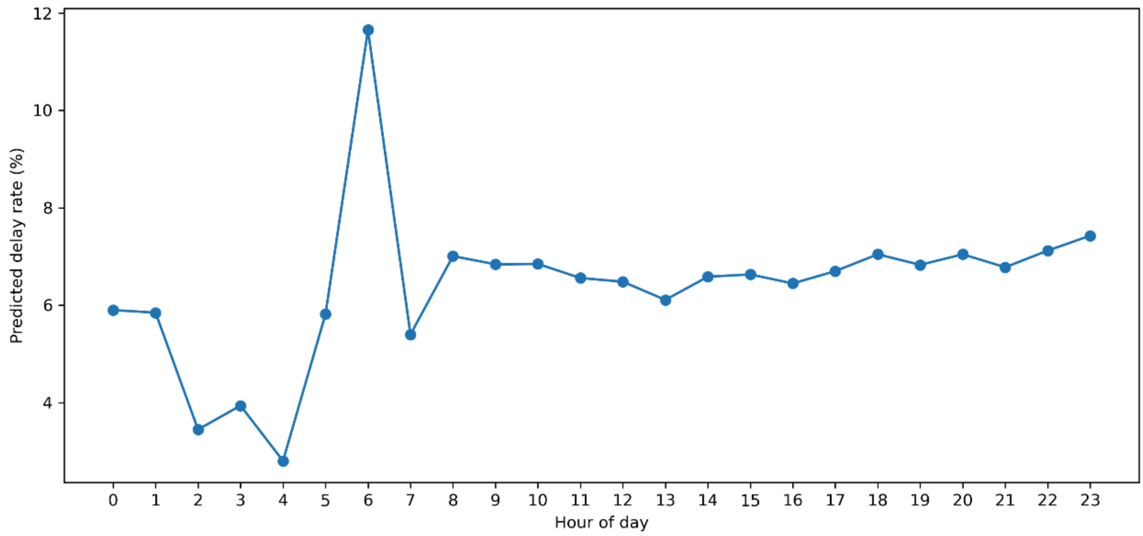

Figure 3. Delay rate by hours of union station (data source: TTC Subway Delay Data [10]; figure credit: original)

The result in Figure 3 shows that the predicted delay rate varies as time goes on, with 6 am being the most likely delay hour and 4 am the least likely to be delayed. Most time slots of the day have a similar rate of delay. Besides, the database the model used mainly recorded historical delay incidents and fundamental time properties, so Figure 3 mainly showed the probability that a random delay will take place for the next coming train during this time slot instead of the rate of delay incidents for all trains.

Figure 4 shows the average predicted probability of a recorded delay incident at Union Station by day of week. The probabilities generated for individual date-hour observations in the testing set are grouped and averaged by day of week. The differences between the seven days are relatively small. Monday has the highest average predicted probability, while Sunday has the lowest. However, the results do not show a consistent separation between weekdays and weekends, because the predicted probability for Saturday remains similar to that of several weekdays. Therefore, the observed differences should be interpreted as general historical patterns rather than clear evidence that weekday commuting demand or weekend service arrangements directly cause changes in delay occurrence.

DOI: 10.54254/2755-2721/2026.BA36256

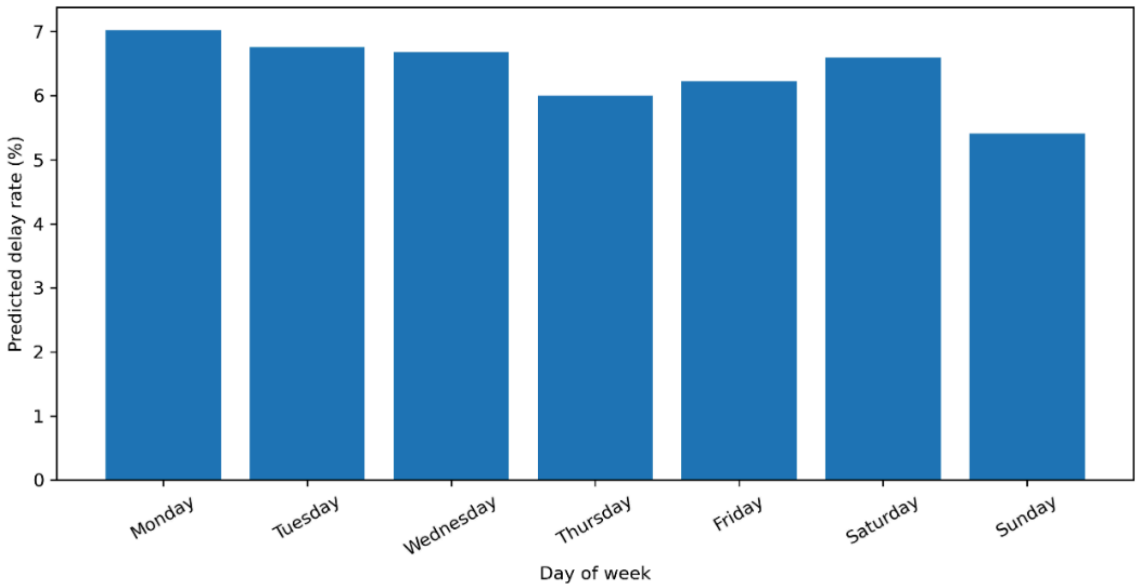

Figure 4. Average predicted probability of a recorded delay incident at union station by day of week (data source: TTC Subway Delay Data [10]; figure credit: original)

For the classification task, Accuracy, Precision, Recall, and F1 Score were used to evaluate the model's ability to identify high-risk delay time slots. The evaluation results are shown in Table 2.

Table 2. Classification performance of the union station delay-risk model

Metric	Value
Accuracy	0.764035
Precision	0.197717
Recall	0.774597
F1 Score	0.315024

The model achieved relatively high recall but low precision. A recall of 0.775 indicates that the model successfully identified approximately 77.5% of the date-hour periods in which at least one delay incident was recorded. However, the precision of 0.198 indicates that only about 19.8% of the periods predicted as positive actually contained a recorded delay incident. The model therefore detects many actual delay periods but also produces a large number of false-positive predictions. Its results are more suitable for preliminary screening than for precise prediction of individual delay periods.

4. Discussion

The relatively high recall indicates that this model is able to capture time slots containing delay incidents of a relatively large proportion. However, the low precision and F1 Score show that many periods predicted as positive do not actually contain a recorded delay, which limits the practical reliability of the model. Therefore, there are still restrictions on this model, as it can be considered an exploratory classification tool but not an accurate system predicting real incidents.

DOI: 10.54254/2755-2721/2026.BA36256

Another limitation comes from the TTC Subway Delay Data itself since the dataset records only delay incidents but not normal train operations. Therefore, the predicted category reflects whether a delay event is likely to be recorded during a date-hour period rather than the true operational delay rate of all subway trains. What's more, the current model mainly uses basic temporal and station-related features, while other factors such as extreme weather, maintenance activities, and network interactions are not included. Previous studies have shown that variables such as historical delay profiles, ridership, day of week, geography, weather, and network conditions can contribute to railway delay prediction [5, 8, 9]. In addition, the model relies on historical delay frequency, which cannot fully capture unexpected disruptions or special events. Previous research indicates that delay-prediction performance depends strongly on data quality, model validity, and model selection [2, 7].

5. Conclusion

This study analyzed TTC Subway Delay Data to explore the temporal and spatial patterns of subway delays in Toronto and further examined Union Station as a station-level delay-risk classification case.

In addition to descriptive analysis, this study transformed incident-level delay records into complete date-hour observations for Union Station and explored delay-risk classification for Union Station. The classification results show that historical delay records can provide preliminary risk-screening information, but the model performance is limited by the structure of the available data.

Overall, the results highlight that station-level delay-risk classification is challenging with limited open data. While the model can identify general high-risk patterns, future work should integrate richer data sources to improve accuracy and practical usefulness.

References

[1] Soberman, R. M. (1997). *The track ahead: Organization of the TTC under the new amalgamated city of Toronto*. Toronto Transit Commission.

[2] Tiong, K. Y., Ma, Z., & Palmqvist, C. W. (2023). A review of data-driven approaches to predict train delays. *Transportation Research Part C: Emerging Technologies, 148*, 104027.

[3] Spanninger, T., Trivella, A., Büchel, B., & Corman, F. (2022). A review of train delay prediction approaches. *Journal of Rail Transport Planning & Management, 22*, 100312.

[4] Sarhani, M., & Voß, S. (2024). Prediction of rail transit delays with machine learning: How to exploit open data sources. *Multimodal Transportation, 3*(2), 100120.

[5] Lapamonpinyo, P., Derrible, S., & Corman, F. (2022). Real-time passenger train delay prediction using machine learning: A case study with Amtrak passenger train routes. *IEEE Open Journal of Intelligent Transportation Systems, 3*, 539–550.

[6] Shi, R., Xu, X., Li, J., & Li, Y. (2021). Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization. *Applied Soft Computing, 109*, 107538.

[7] Chen, X., Ma, Z., & Sun, W. (2024). Incident delay prediction in urban railway systems: Methodology review and exploratory comparative analysis. *Transportation Research Record, 2678*(12), 1629–1638.

[8] Klumpenhouwer, W., & Shalaby, A. (2022). Using delay logs and machine learning to support passenger railway operations. *Transportation Research Record: Journal of the Transportation Research Board, 2676*(9), 134–147.

[9] Wang, J., Mantas-Nakhai, R., & Yu, J. (2023). Statistical learning for train delays and influence of winter climate and atmospheric icing. *Journal of Rail Transport Planning & Management, 26*, 100388. https://doi.org/10.1016/j.jrtpm.2023.100388

[10] City of Toronto Open Data Portal. (n.d.). TTC subway delay data. *Toronto Open Data*. Retrieved from https://open.toronto.ca

[11] Toronto Transit Commission. (n.d.). Subway Yonge-University Line Union Station. *TTC Routes and Schedules*. Retrieved July 20, 2026, from https://www.ttc.ca/routes-and-schedules/1/1/13816