1 Multi-year Data Analysis of TraffiDent

We analyze the weekday and weekend daily trending variation in 3 months (January, June and December) based on all sensors, excluding those not deployed in 2022. Also, we compared the average flow on the hub road and fringe road during weekdays and weekends. Year-on-Year Trending on Weekday and Weekday in January, June, and December. Through year-on-year change analysis and observation, we identified unusual variations in traffic across different years. We divided traffic data from all sensors into two groups: weekdays and weekends. We then calculated the average traffic for each group, resulting in two distinct traffic change trends. To enhance visualization, we normalized the data by using three types of traffic as baselines and dividing the data from other years by these baselines. The analysis results are illustrated in Figure 1.

We observe that traffic patterns exhibit similar temporal variations across different months. Specifically, on weekdays, both traffic flow and speed in 2023 are consistently lower compared to 2022 and 2024. In contrast, weekend traffic shows a steady increase year over year. These observations suggest that, despite fluctuations across years, traffic characteristics tend to stabilize into comparable patterns within the same month across different years.

(a) January - Weekday	(b) June - Weekday	(c) December - Weekday

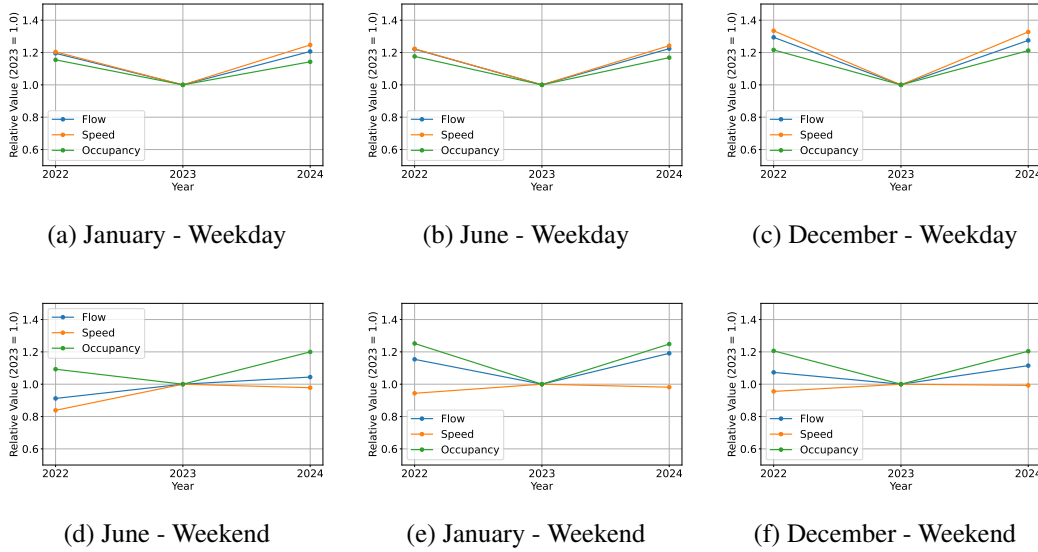| (d) June - Weekend	(e) January - Weekend	(f) December - Weekend

Figure 1: Multi-year analysis of TraffiDent. The up figures (a)(b)(c) show the average traffic on Weekdays and the down figures (d)(e)(f) show the average traffic on Weekends.

2 Discussion of Limitations

While the TraffiDent dataset represents a significant advancement as a multi-year, large-scale traffic and incident dataset enriched with meta features, there remain areas for potential improvement.

First, although we discuss the data imputation methods employed in Section 3 of the paper, a more comprehensive evaluation of how different imputation strategies affect model performance across various experiments is warranted. Incorporating these experimental results and the corresponding insights into the dataset documentation would enhance the utility of the dataset for future research.

Second, while Section 4 outlines the approach for associating incidents with sensors, further refinement could involve introducing these matched results as new incident-related features within the dataset. This enhancement could provide additional context for incident analysis and support the development of more advanced traffic forecasting models. Future work will focus on exploring this direction and integrating the findings into the dataset.