Time Series Forecasting for Urban Traffic
Time Series Analysis and Forecasting Project

Rani Herbawi
Technion - Israel Institute of Technology
ranykhirbawi@campus.technion.ac.il

Contents

Abstract

This study investigates the application of time series analysis techniques to forecast traffic volume at a key urban junction. The dataset comprises hourly vehicle counts spanning multiple years, exhibiting clear seasonal and trend components. The analysis focuses on three primary forecasting models: SARIMA, Prophet, the Holt-Winters exponential smoothing method and LSTM. I evaluated the performance of these models using both classical and forward-chaining validation strategies, with Prophet demonstrating superior accuracy in the multivariate setting, and LSTM performing best when forecasting without exogenous variables.

To explore the impact of external factors on traffic patterns, we incorporate temperature as an exogenous variable. However, statistical analysis reveals only a modest correlation between temperature and traffic volume, limiting the variable's predictive utility. Additionally, change-point detection methods are employed to identify significant distributional shifts, which could be attributed to policy changes or external disruptions. These points are further analyzed using statistical and visual tools, providing context for observed deviations in the data.

Overall, the findings underscore the effectiveness of Prophet and LSTM models for short-term traffic forecasting, depending on whether exogenous variables are included. While SARIMAX benefited moderately from external temperature data, the results also highlight the challenges of incorporating weakly correlated or noisy exogenous variables. The methodological framework and insights presented here can support urban planners and policy analysts in crafting data-driven traffic management strategies.

1 Introduction

Urban traffic management remains a critical challenge for modern cities, especially in light of increasing vehicle volumes, environmental concerns, and the need for efficient transportation planning. Accurate forecasting of traffic at key junctions can inform decisions regarding infrastructure, traffic lights optimization, and policy-making, ultimately improving mobility and reducing congestion.

This project focuses on modeling and forecasting the number of vehicles passing through a central traffic junction using time series analysis techniques. The dataset utilized comprises hourly vehicle counts across several years. The primary goal is to compare different forecasting methods and evaluate their performance in predicting future traffic volumes.

I implemented and assessed three prominent forecasting models: Seasonal Autoregressive Integrated Moving Average (SARIMA), the Prophet model developed by Meta (formerly Facebook), the Holt-Winters exponential smoothing technique and also Long Short-Term Memory model (LSTM). These models were selected due to their proven effectiveness in capturing seasonal and trend-based dynamics commonly observed in time series data.

Beyond univariate forecasting, i explored the integration of an exogenous variable—ambient temperature—into the modeling process. This external factor is hypothesized to have an influence on traffic behavior, particularly during extreme weather conditions.

Finally, we employ change-point detection techniques to identify statistically significant shifts in the traffic pattern.

The following sections detail the methodology, present model comparisons, and discuss the implications of our findings for data-driven urban traffic planning.

1.1 Dataset

The dataset[1] used in this project consists of time series data from industrial IoT sensors, with each row representing sensor readings at a specific timestamp. Key columns include timestamp, sensor location, and temperature measurements (in Kelvin). The data was sourced from Kaggle, provided for time-series forecasting analysis using IoT sensor data.

1.2 Analysis Questions

This project aimed to answer the following key questions:

- Can we accurately forecast future values in the time series using statistical and ML models?
- How do different forecasting models compare in terms of predictive accuracy?
- Can external variables or structural change detection improve the forecast?

2 Exploratory Data Analysis

This section explores the temporal and statistical properties of hourly vehicle traffic data collected at multiple junctions. The primary focus is on Junction 1, which exhibits the most variability and volume, making it a suitable candidate for forecasting tasks.

2.1 Initial Visual Exploration

To gain a preliminary understanding of traffic dynamics across the city, we first examine trends across all four junctions, followed by a distributional analysis and a focused time series plot of Junction 1 on the first month of the year 2017.

[1] `https://www.kaggle.com/datasets/vetrirah/ml-iot/data`

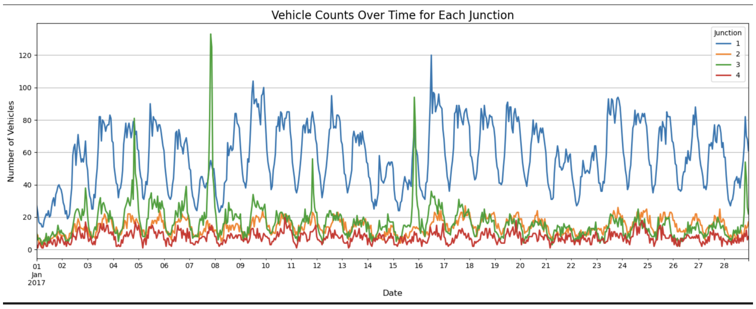

Figure 1: Vehicle counts over time for each junction.

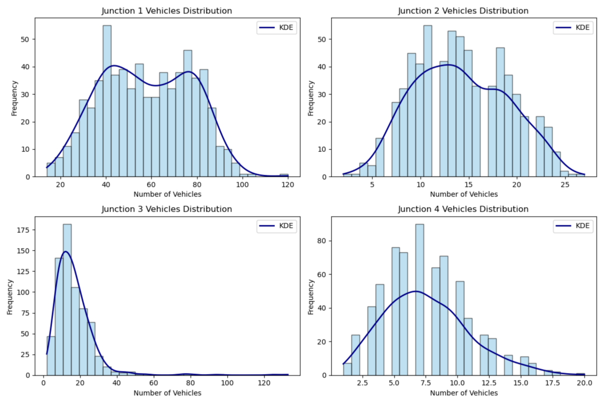

Figure 2: Vehicle count distributions.

The histograms depicted in Figure 2 illustrate the distribution of hourly vehicle counts at each junction. Junction 1 shows a broader range and higher frequency of vehicle counts, indicating greater traffic variability. Conversely, junctions 2, 3, and 4 display narrower distributions, reflecting more consistent and lower overall traffic volumes.

Figure 3 shows a detailed time-series plot of vehicle counts specifically for Junction 1. It clearly indicates daily fluctuations with discernible periodic patterns, suggesting pronounced hourly and possibly weekly seasonality.

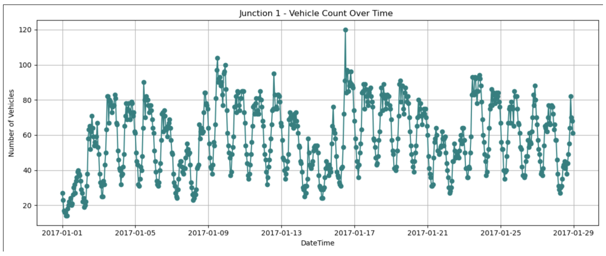

Figure 3: Vehicle count at Junction 1 over time.

2.2 Weekday and Hourly Analysis

The following visualizations uncover key cyclical traffic patterns. Weekdays show significantly higher traffic volumes than weekends, and peak traffic typically occurs during morning and evening rush hours.

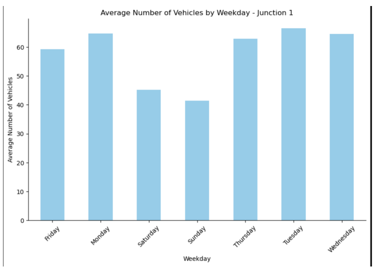

Figure 4: Average vehicles by weekday.

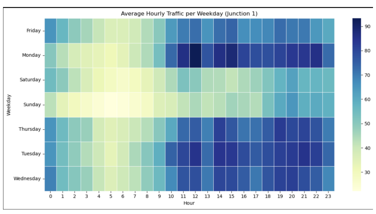

Figure 5: Hourly traffic per weekday.

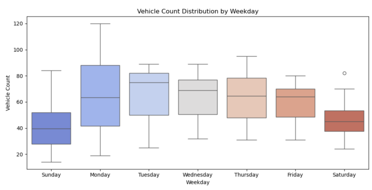

Figure 6: Boxplot of weekday traffic.

2.3 Autocorrelation and Stationarity

To support the decision for model types and differencing, we examine the autocorrelation structure and assess stationarity. Strong autocorrelations in the raw series confirm trend and seasonality. Differencing the series reduces autocorrelation and yields a mean-reverting signal.

The initial ACF and PACF plots of the raw series (Figure 7) exhibit strong autocorrelations at both short and seasonal lags. The ACF shows a slow decay and distinct peaks at lags 12, 24, and 36, confirming a strong seasonal component, while the PACF has significant spikes at lags 1 through 3 and near lag 12, indicating the presence of both short-term and seasonal autoregressive behavior.

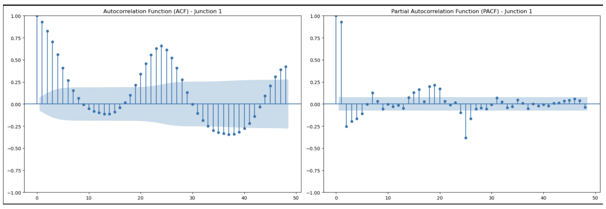

Figure 7: ACF and PACF for raw data.

After applying both regular and seasonal differencing, the updated plots (Figure 8) show reduced autocorrelation, with a prominent ACF spike at lag 1 and diminished seasonal structure. The PACF tapers off after lag 3, and most lags fall within the confidence bounds, indicating stationarity. These patterns justify the use of a SARIMAX(3,1,2)(1,1,1)[24] model, which captures the observed short-term AR(3) and MA(2) structure, along with seasonal AR(1) and MA(1) effects. The differencing orders $d = 1$ and $D = 1$ are essential for handling trend and seasonal non-stationarity.

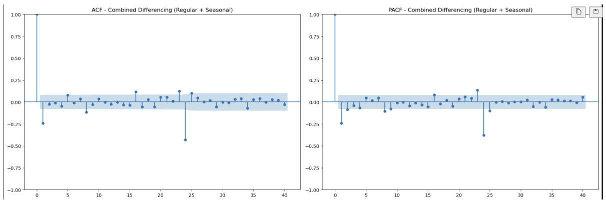

Figure 8: ACF and PACF after differencing.

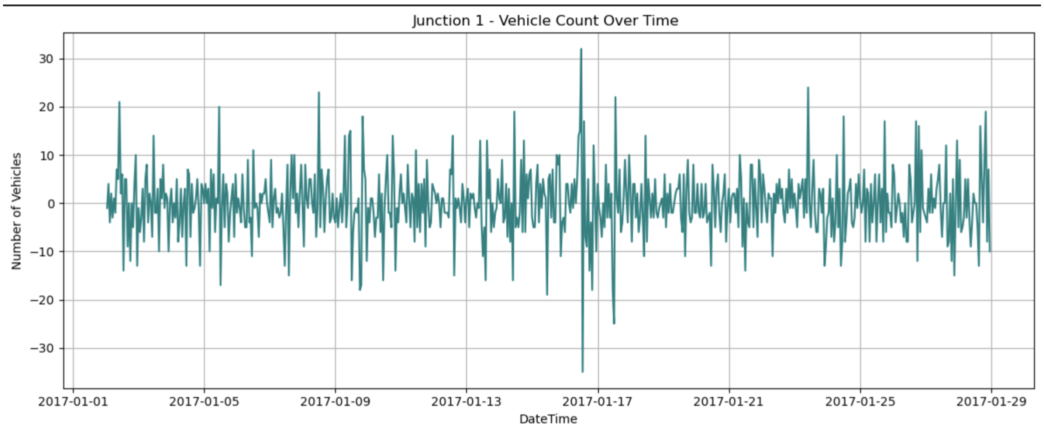

Figure 9: Differenced vehicle count series (Junction 1), which now fluctuates around a zero mean and lacks trend or seasonal structure, confirming stationarity.

2.4 Conclusion from EDA

This exploratory analysis reveals clear daily seasonality with a periodic pattern recurring every 24 hours, consistent with urban traffic flow dynamics. In addition to this intraday cyclicity, long-term trends and variations across weekdays are evident, underscoring the need for models capable of capturing both seasonal and trend components. The presence of strong autocorrelations at lag 24 further supports the choice of a seasonal SARIMAX model with a periodicity of 24 (i.e., SARIMAX with seasonal order s = 24). Consequently, our model selection—focusing on SARIMAX(3,1,2)(1,1,1) 24, Prophet, and Holt-Winters—reflects the data's structure and ensures the models can handle the detected regularities. The potential influence of external covariates, such as temperature, will be further examined through regression analysis and correlation tests in the next phase of the project.

3 Methodology and Results

In this section, we first outline our forecasting methodology and data partitioning in Section 3.1, followed by an overview of the selected models in Section 3.2. Section 3.3 examines the integration of temperature as an exogenous variable. In Section 3.4, we apply statistical tools to detect change-points in the series, and Section 3.5 presents the evaluation metrics used to compare model performance.

3.1 Forecasting Strategy

To simulate a realistic forecasting environment, the dataset was divided into a training and test set. The training set contained the first 600 hourly observations, while the last 72 hours (3 days) were reserved for testing. All models were trained on the same training data and evaluated on the test set using consistent metrics.

To ensure our models could capture sequential dependencies effectively, i used standard forecasting and a sliding-window approach (in the case of LSTM). SARIMA and ETS models directly forecasted the next 72 hours, while Prophet and LSTM were retrained iteratively for one-step-ahead forecasting.

3.2 Model Fitting

In this section, we evaluate three primary forecasting models: SARIMA, Prophet, and Exponential Smoothing (ETS), each chosen based on their theoretical suitability and practical strengths for modeling hourly vehicle counts at Junction 1.

Rationale for Model Selection

- **SARIMA** was selected for its robustness in capturing both seasonal and non-seasonal autoregressive patterns. Given the strong weekly and daily seasonality in our data, SARIMA provides a solid statistical foundation for baseline modeling.
- **Prophet**, developed by Meta, was employed due to its strong ability to decompose and model trend and multiple seasonality effects without intensive parameter tuning. It also includes changepoint detection and supports external regressors, making it ideal for dynamic real-world time series like traffic data.
- **Exponential Smoothing (ETS)** was used as a benchmark due to its effectiveness in modeling series with additive components. Its simplicity and interpretability offer practical utility and quick baseline comparisons.

Parameter Selection Summary

The parameters of each model were selected based on empirical testing and diagnostic metrics:

- **SARIMA**

  To validate the need for differencing, we performed a visual inspection using 24-hour rolling mean and standard deviation plots. These showed varying mean and volatility, indicating non-stationarity. We then applied the Augmented Dickey-Fuller (ADF) test to the original training series. The test returned a high p-value (p ¿ 0.05), supporting the hypothesis that the series was non-stationary. As a result, we proceeded with seasonal differencing (D=1, s=24) and non-seasonal differencing (d=1).

A subsequent ADF test on the differenced series confirmed stationarity with a significantly reduced p-value. We then generated ACF and PACF plots on the differenced series to guide model selection.

Multiple candidate SARIMA models were tested:

`SARIMA(0,1,1)(1,1,1)[24]`, `SARIMA(1,1,1)(1,1,1)[24]`, `SARIMA(1,1,0)(1,1,1)[24]`, and `SARIMA(3,1,2)(1,1,1)[24]`

Each configuration was fit using the SARIMAX implementation with constraints disabled (`enforce_stationarity=False, enforce_invertibility=False`). Evaluation was performed based on AIC, MAE, and RMSE. Among the tested models, `SARIMA(3,1,2)(1,1,1)[24]` yielded the lowest AIC of 3333.26 and provided the best forecast performance.

Final diagnostics confirmed good residual behavior with no strong autocorrelation. This best-performing model was used to generate the final forecasts.

- **Prophet**: Trained with an hourly frequency and daily seasonality enabled. Prophet required minimal hyperparameter tuning and successfully modeled multiple recurring patterns.
- **ETS**: An additive model was selected for both the trend and seasonality components, with a seasonal period set to 24 hours. The model was optimized using built-in grid search, resulting in an AIC of 1907.94.

Forecast Visualization and Comparison

Figure 10 provides a direct comparison of the test set forecasts generated by all three models. SARIMA exhibited smooth cyclic patterns, while Prophet most closely tracked the actual observed spikes and drops. ETS offered moderate alignment with actual values.

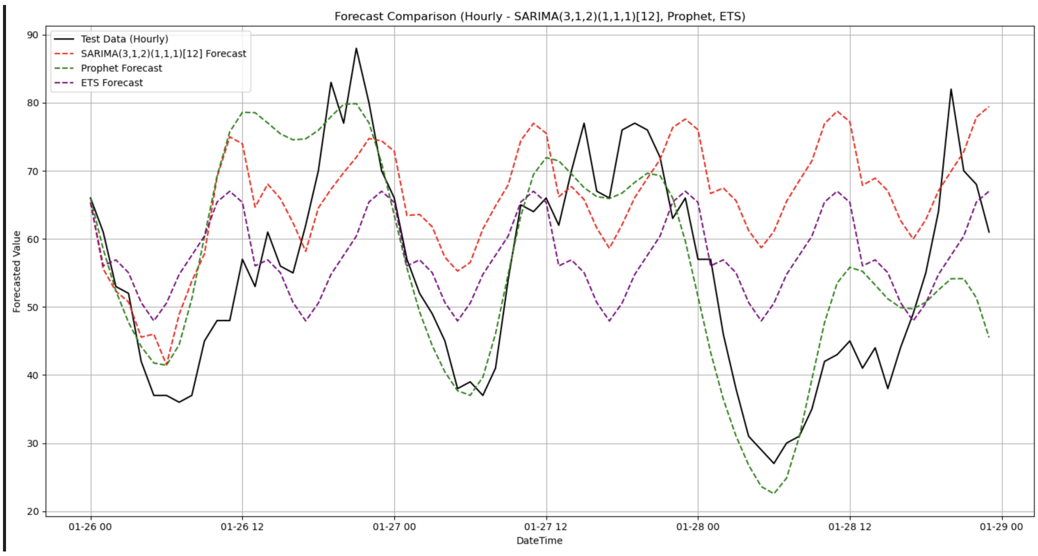

Figure 10: Forecast comparison on test data: SARIMA, Prophet, and ETS.

Additionally, Figure 11 shows the full forecast output from Prophet, including its uncertainty intervals. The forecast captures recurring weekly seasonality, with test points aligning well within the prediction bounds.

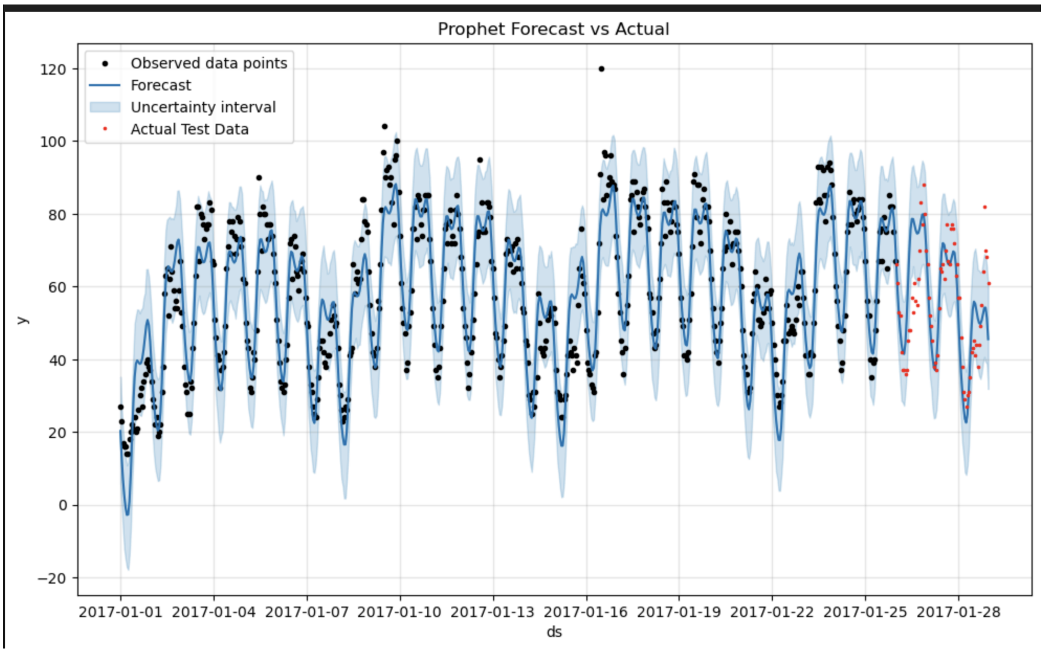

Figure 11: Prophet forecast with uncertainty intervals.

Quantitative Evaluation

Model	MAE	RMSE	AIC
Prophet	**7.84**	**10.39**	–
SARIMA	10.18	13.73	3333.26
ETS	10.44	14.06	1907.94

Table 1: Forecast error and model performance metrics.

Conclusion

Among all evaluated models, **Prophet** demonstrated the highest accuracy, with the lowest RMSE and MAE values. Its ability to handle complex seasonal structures and smooth forecast behavior made it the most suitable for our short-term traffic forecasting task. SARIMA and ETS provided reasonable baselines and offered interpretable statistical structure but lagged in predictive precision.

3.3 Incorporating an Exogenous Variable

To explore the impact of weather on traffic volume, we incorporated hourly temperature data from the **Tel Aviv district** as an exogenous regressor into the SARIMA, Prophet, and LSTM models. This integration aimed to assess whether incorporating external climate-related information could enhance the predictive accuracy of the traffic models.

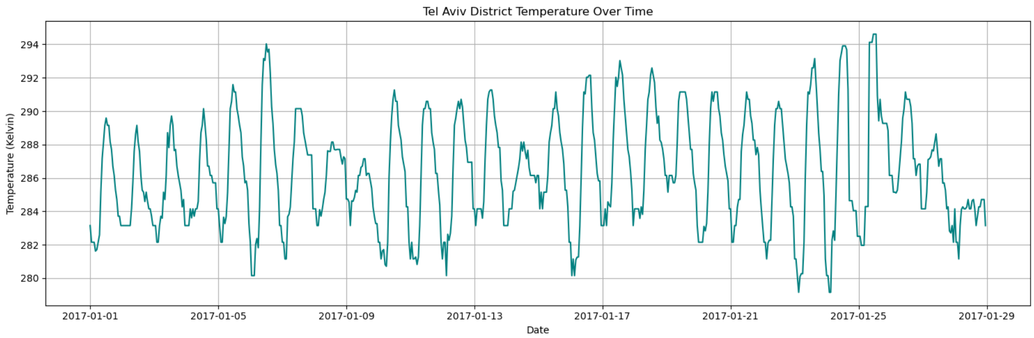

Figure 12: Tel Aviv District Temperature Over Time

SARIMAX

I extended our best SARIMA configuration—`SARIMA(3,1,2)(1,1,1)[24]`—by incorporating temperature as an exogenous input, resulting in a SARIMAX model. This integration was implemented using the `SARIMAX` class from the `statsmodels` library, aligning hourly temperature readings with the target time series.

- **Without Temperature Variable:** RMSE = 17.31
- **With Temperature Variable:** RMSE = **15.43**

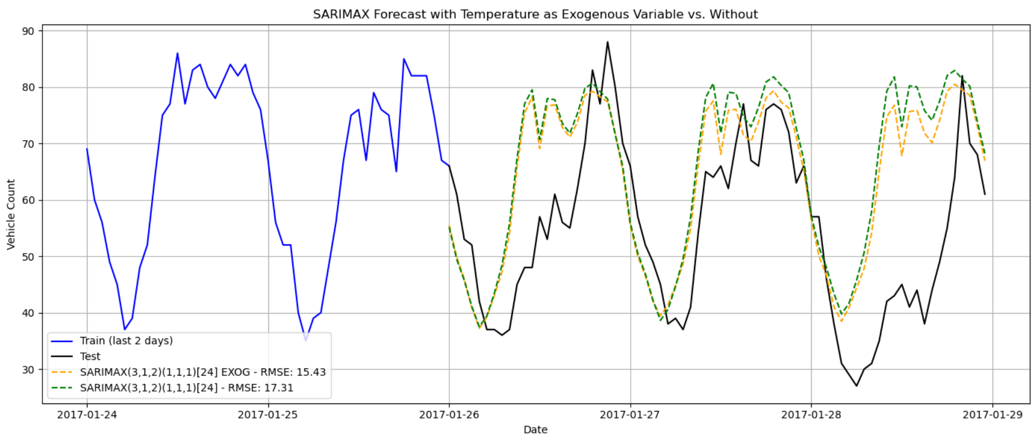

Figure 13: SARIMAX Forecast with Temperature as Exogenous Variable vs. Without

This improvement suggests that weather variations—captured via temperature—may influence short-term traffic patterns, potentially due to behavioral changes in commuting and travel decisions under varying climate conditions.

Prophet with External Regressor

Prophet allows for the incorporation of additional regressors directly into its framework. We added temperature as a custom regressor using the `add_regressor` method.

- **Without Temperature Variable:** RMSE = 10.40
- **With Temperature Variable:** RMSE = **10.29**

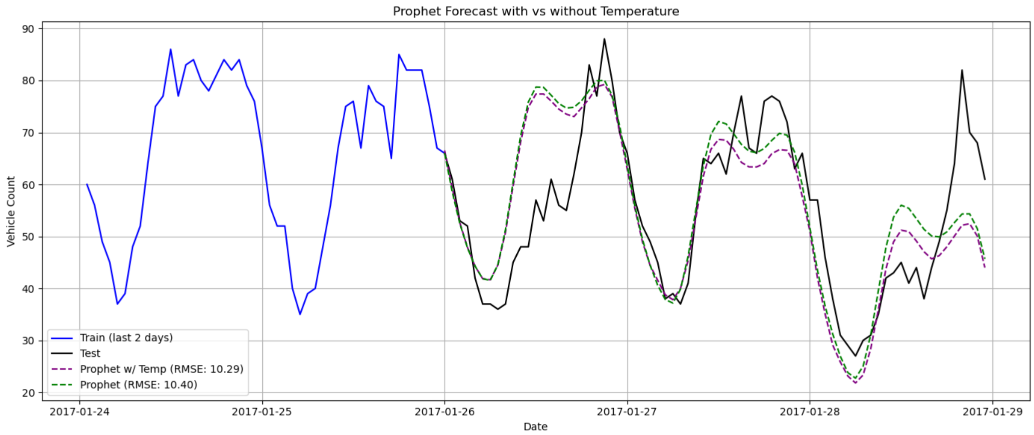

Figure 14: Prophet Forecast with vs. without Temperature

While the performance boost is modest, Prophet was able to refine its trend and seasonal component modeling when temperature data was introduced, particularly during lower-traffic periods.

LSTM – Univariate vs. Multivariate

To evaluate the impact of exogenous features on LSTM performance, we trained two models: a univariate LSTM using only traffic counts with a look-back window of 24, and a multivariate LSTM that also included temperature data. Interestingly, the univariate model outperformed the multivariate one, achieving an RMSE of **8.24** compared to **11.93** with temperature, suggesting that the added complexity and weak correlation of the exogenous variable may have hindered the model's effectiveness.

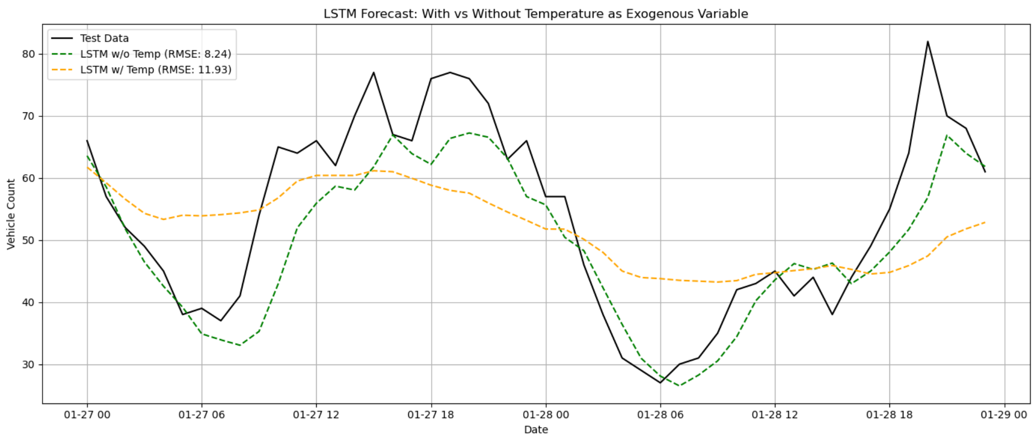

Figure 15: LSTM Forecast: With vs. Without Temperature as Exogenous Variable

Surprisingly, the multivariate LSTM underperformed. This is likely due to limited training data (600 hourly observations), which may not have been sufficient to justify the added complexity of a multivariate architecture. It highlights a crucial insight: more complex models are not always better, especially when the exogenous signal is noisy or only weakly correlated.

Conclusion

Incorporating temperature improved forecasting accuracy in **SARIMAX** and slightly in **Prophet**, confirming its moderate utility as an exogenous variable. However, its effect was **detrimental for LSTM**, underscoring the importance of model-specific validation when extending inputs.

A Pearson correlation test conducted between temperature and traffic volume revealed a moderate positive linear relationship, with a correlation coefficient of **0.45** and a statistically significant p-value of **5.84e-35**.

3.4 Change-Point Detection

Many time series exhibit changes in distribution due to external events (e.g., regulatory updates, unexpected shocks). In this part, we investigate whether such a change occurs in the vehicle count data at Junction 1 using a statistical approach based on **Interrupted Time Series Regression (ITSR)**.

We modeled the series using the following regression form:

$$Y_t = \beta_0 + \beta_1 \cdot \text{TimeIndex}_t + \beta_2 \cdot \text{PostChange}_t + \beta_3 \cdot \text{TimeSinceChange}_t + \varepsilon_t$$

- **TimeIndex** denotes the continuous time trend throughout the series.
- **PostChange** is a binary indicator (0 before January 16, 2017 at 12:00, 1 afterward).
- **TimeSinceChange** captures the slope change after the change-point.

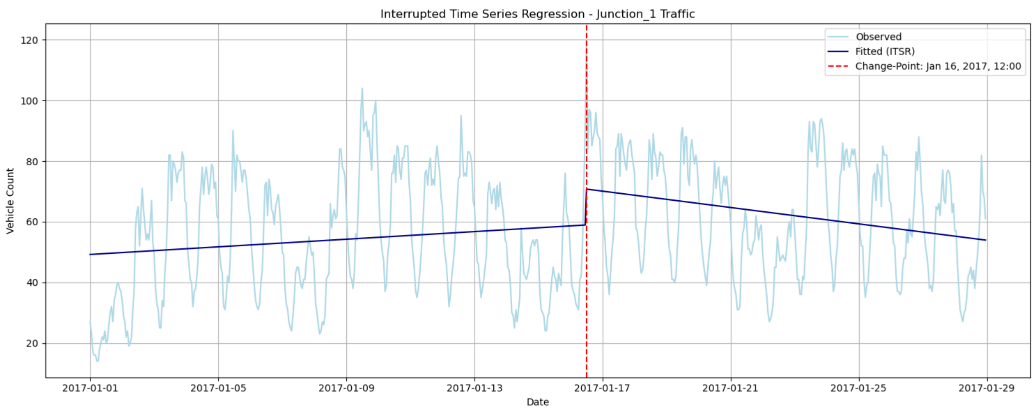

Figure 16: Interrupted Time Series Regression – Detected Change-Point at January 16, 2017, 12:00

Key Regression Results:

- **Pre-change trend** (β_1): **+0.026**, *p = 0.004* – indicates a slight increasing trend before the change.
- **Level shift** (β_2): **+11.82**, *p ¡ 0.001* – denotes a statistically significant increase of ∼12 vehicles/hour immediately after the change.
- **Post-change slope** (β_3): **−0.082**, *p ¡ 0.001* – reflects a negative trend following the change.

Interpretation: These results reveal a **statistically significant structural shift** in the series around mid-January. After a period of mild growth, traffic at Junction 1 experienced a sharp rise followed by a consistent downward trend. All detected effects are strongly supported by low p-values (< 0.01), confirming the presence of a distributional change in the underlying data.

3.5 Evaluation Metrics

For model selection and performance comparison, we used the following metrics:

- **Akaike Information Criterion (AIC)**: Used to compare SARIMA models.

- **Root Mean Squared Error (RMSE)**: The main metric to compare final model forecasts on the test set.
- **Mean Absolute Error (MAE)**: Used as a secondary error metric.

While AIC helped identify the optimal SARIMA configuration, RMSE and MAE provided robust comparisons across different models, especially for those like Prophet and LSTM where AIC is not applicable.

4 Conclusion and Discussion

This project presented a comprehensive time series analysis and forecasting study on hourly traffic volume data from Junction 1, incorporating multiple modeling techniques and external variables to enhance predictive accuracy.

I began with an exploratory data analysis that revealed strong daily and weekly seasonality, as well as potential structural shifts in the time series. These insights motivated our use of seasonal models like SARIMA and Prophet, as well as simpler interpretable models like Holt-Winters ETS.

Model Performance Summary

Three baseline models were evaluated: SARIMA, Prophet, and ETS. Among them, **Prophet** demonstrated the best predictive accuracy in the univariate setting, capturing seasonal and trend patterns effectively with minimal tuning. SARIMA provided robust baseline performance and interpretability, while ETS delivered competitive results, particularly in capturing smooth cyclical behavior.

Incorporating temperature as an exogenous regressor led to mixed outcomes. **SARIMAX** improved over its univariate counterpart, while **Prophet** benefited slightly. Surprisingly, **LSTM** performed best without the additional variable, suggesting that model complexity may not always lead to better results, especially in limited-data scenarios or with weakly correlated exogenous inputs.

Change-Point Detection

Using Interrupted Time Series Regression, i identified a statistically significant change in the distribution of the traffic series around January 16, 2017. This shift included an immediate increase in traffic followed by a gradual decline, emphasizing the importance of incorporating external events or policy changes in time series modeling.

Discussion

The results highlight several key takeaways:

- **Model selection must balance interpretability and complexity.** While Prophet and SARIMA yielded similar trends, Prophet provided more intuitive outputs, while SARIMA allowed detailed control over model structure.
- **External variables require validation.** Although temperature showed a moderate correlation with vehicle counts, its contribution to model accuracy varied significantly across modeling approaches.
- **Deep learning models like LSTM are sensitive to data size and structure.** Despite their potential, they may underperform in low-data regimes without sufficient signal or when multivariate inputs increase noise.
- **Change-point analysis is essential for understanding long-term shifts.** Incorporating domain knowledge and statistical testing helped detect significant behavioral changes, which could influence modeling decisions and policy recommendations.