# A Data-Driven Approach to Enhancing Mass Transportation Utilization: A Case Study from Egypt

Manal A. Abdel-Fattah [1]
Ahmed Gamal Elish[2]
Nada Ayman Mohammad Aboelfadl[3]

## Abstract

This study investigates short-term forecasting models for predicting daily passenger demand in public transportation systems. Real passenger movement data collected from Mowasalat Misr (March-June 2022) was utilized to evaluate the effectiveness of various forecasting techniques. Three forecasting models (Facebook Prophet, SARIMA, and Holt-Winters) were evaluated. Evaluation revealed that SARIMA achieved the best performance among the established models by effectively capturing seasonality in passenger demand data. To further enhance forecasting accuracy, a novel hybrid model was developed by combining SARIMA with Holt-Winters. This hybrid approach yielded superior results, achieving a Mean Absolute Error (MAE) of 957, Mean Squared Error (MSE) of 3.2, and Root Mean Squared Error (RMSE) of 1776**.**

**Keywords:** Time-series Forecasting, Public Transport, Mass Transport, EDA, Prophet, ARIMA, SARIMA, Holt-Winters**.**

[1] Professor of Information Systems and Vice Dean for Education and Students Affairs-Faculty of Computers and Artificial Intelligence – Helwan University
[2] Assistant Professor at the Department of Economics and Foreign Trade, Faculty of Commerce and Business Administration – Helwan University
[3] Master's Degree Student at Faculty of Commerce and Business Administration-Helwan University

# نهج قائم على البيانات لتعزيز استخدام وسائل النقل الجماعى : دراسة حالة من مصر

## الملخص

تبحث هذه الدراسة نماذج التنبؤ على المدى القصير لتوقع الطلب اليومي على الركاب في أنظمة النقل العام. تم استخدام بيانات حركة الركاب الحقيقية التي تم جمعها من شركة مواصلات مصر (مارس–يونيو 2022) لتقييم فعالية تقنيات التنبؤ المختلفة. تم تقييم ثلاثة نماذج للتنبؤ Facebook Prophet, SARIMA Holt–Winters. كشف التقييم أن نموذج SARIMA حقق أفضل أداء بين النماذج المستقرة من خلال التقاطع الفعال للتقلبات الموسمية في بيانات طلب الركاب. لمزيد من تحسين دقة التنبؤ، تم تطوير نموذج هجين جديد من خلال الجمع بينSARIMA Holt–Winters . حقق هذا النهج الهجين نتائج متفوقة، حيث حقق متوسط الخطأ المطلق 957 (MAE) ومتوسط الخطأ التربيعي (MSE) 3.2وجذر متوسط الخطأ التربيعي .1776 (RMSE) تسلط هذه النتائج الضوء على الفوائد المحتملة لدمج تقنيات التنبؤ لتحسين توقع طلب الركاب في أنظمة النقل العام.

**الكلمات المفتاحية:** التنبؤ بالمتسلسلات الزمنية، النقل العام، النقل الجماعي، التحليل الإستكشافي للبيانات، EDA, Prophet, ARIMA, SARIMA, Holt–Winters.

# 1. Introduction:

Traffic congestion in the 21st century wastes not only valuable time but also crucial energy resources [1]. This highlights the need for innovative solutions within the transportation sector, such as those encompassed within the concept of 'smart' transportation. Smart transportation utilizes advancements like electric vehicles, intelligent vehicle controls, and alternative business models like car-sharing to achieve multiple goals, including reducing pollution, congestion, and travel costs, while also enhancing safety and improving transfer speeds [2]. Efficient operation of mass transportation systems relies heavily on the ability to accurately predict passenger demand. This allows for optimized resource allocation, scheduling adjustments, and improved service planning, ultimately leading to a more efficient and user-friendly transportation system. However, demand patterns can exhibit complex seasonality, trends, and external influences, posing challenges for forecasting models. This study examines the effectiveness of various short-term forecasting models, including Facebook Prophet, ARIMA, SARIMA, and Holt-Winters, in capturing daily passenger demand, with a particular focus on the unique challenge of Fridays having lower demand compared to other weekdays. By investigating this issue, we aim to contribute to the development of more accurate forecasting models for improved public transportation planning and operation.

## 2. Literature Review:

From predicting weather patterns to managing product demand and tackling public health challenges, time series forecasting plays a crucial role in various sectors. This review delves into several studies that explore the effectiveness of diverse forecasting models across different industries and applications. In weather forecasting, [1] & [2] explored the effectiveness of Prophet for predicting temperature. Both studies demonstrated promising results, with Zar Zar Oo achieving accurate temperature prediction and Jungang finding a new proposed Prophet-LSTM to outperform other models. Moving to manufacturing,[3] investigated product demand forecasting, comparing several models including Prophet-SVR, Holt-Winters, SARIMA, and others. Their research indicated that the combined model Prophet-SVR achieved the best results, suggesting its effectiveness in this domain. Within transportation, diverse forecasting tasks were explored, [4]focused on passenger traffic, utilizing SARIMA-SVR combined model. Their model achieved superior accuracy compared to others, highlighting its potential in this area. Similarly, [5] investigated railway freight volume, finding that Prophet-DeepAR outperformed other models, demonstrating its suitability for such tasks. [6]applied SARIMA to forecast road accidents, while [7] used ARIMA models for short-term traffic flow forecasting, both achieving promising results. In the field of medicine, [8] explored COVID-19 case forecasting. Their stacked LSTM-GRU model achieved superior performance compared to other models, indicating its potential in this critical area. To sum up, the reviewed studies [**Table 1**] showcase a wide range of time series forecasting models employed in diverse industries and applications. Prophet emerged as a strong performer in weather and sales forecasting, while hybrid models like Prophet-SVR and SARIMA-SVR demonstrated success in manufacturing and transportation. Notably, the stacked LSTM-GRU model stood out in COVID-19 case prediction. The choice of the most effective model depends on the specific context, data characteristics, and desired forecasting accuracy.

# 3. Data Description:

The dataset was collected from one of the largest mass transportation providers in Egypt, which utilizes a smaty ticketing system for data collection. It is real passenger movement data for a period of four months (March-June 2022). The dataset initially contains 1,309,482 rows with 16 columns. After cleaning and processing steps, the final dataset contains 1,178,410 rows with 13 columns. A confidentiality and non-disclosure agreement was signed by the researcher to ensure data privacy.

### 3.1 Data Integration and Preprocessing:

The raw dataset consisted of 1,309,482 records with 16 columns. To ensure data quality and prepare it for further analysis, the following preprocessing steps were taken:

#### 3.1.1. Data Cleaning and Feature Selection:

Irrelevant columns were identified and removed, reducing the feature space to focus on pertinent information.

#### 3.1.2. Feature Engineering:

To fill in missing information about passenger stops, the most frequent stop for each combination of line, vehicle, date, and time was identified. This reduced missing values from 37% to 3%. The remaining 3% were removed.

The date information was converted to a format that allowed them to extract additional details like year, month, day of the week, and hour. These details help identify trends and patterns in passenger demand.

#### 3.1.2. Outlier Detection and Removal:

Techniques suitable for the data types and analysis objectives were employed to identify and remove outliers.

#### 3.1.3. Final Data Shape:

After preprocessing, the resulting dataset contained 1,178,410 rows and 13 columns.

## 4. Models Evaluation Metrics:

This section provides an overview of various evaluation metrics commonly used in time-series forecasting studies. These metrics assess the accuracy and performance of different forecasting models in predicting passenger demand [**Table 2].**

- **MAE (**Mean Absolute Error)**:** MAE is a measurement tool used to assess how accurate time series forecasts are. It calculates the average of the absolute discrepancies between predicted and actual values [9][10][11] [12].
- **MSE** (Mean Squared Error): MSE is another way to measure how accurate a forecasting model is for time series data. Unlike MAE which uses absolute differences, MSE focuses on the squared differences between predicted and actual values. This means larger errors have a greater impact on the overall score [9][10] [13].
- **RMSE** (Root Mean Squared Error): The RMSE is the square root of the MSE [9][10].

**TABLE 1: Previous Studies In Time-Series Forecasting**

| **Reference** | **Industry** | **Model** | **Evaluation Metrics** | | |
|---|---|---|---|---|---|
| | | | **MAE** | **MSE** | **RMSE** |
| **Zar Zar Oo (2020) [1]** | Weather: Temperature | Prophet | | | 5.7573 |
| **Jungang (2023) [2]** | Weather: Max. Temperature | Prophet-LSTM | | | 0.11546 |
| | | Prophet: | | | 0.12130 |
| | | LSTM | | | 0.13657 |
| **Guo et al. (2021) [3]** | Manufacturing: Product Demand | Holt-winters | $1.62 \times 10^{+5}$ | $4.06 \times 10^{+10}$ | $2.02 \times 10^{+5}$ |
| | | Sarima | $1.40 \times 10^{+5}$ | $2.83 \times 10^{+10}$ | $1.68 \times 10^{+5}$ |
| | | Prophet | $1.41 \times 10^{+5}$ | $2.35 \times 10^{+10}$ | $1.53 \times 10^{+5}$ |
| | | LSTM | $1.28 \times 10^{+5}$ | $1.97 \times 10^{+10}$ | $1.40 \times 10^{+5}$ |
| | | SVR | $1.14 \times 10^{+5}$ | $1.79 \times 10^{+10}$ | $1.34 \times 10^{+5}$ |
| | | Sarima-SVR | $1.39 \times 10^{+5}$ | $2.82 \times 10^{+10}$ | $1.68 \times 10^{+5}$ |
| | | Holt-winters-SVR | $1.42 \times 10^{+5}$ | $3.53 \times 10^{+10}$ | $1.88 \times 10^{+5}$ |

| | | | | | |
|---|---|---|---|---|---|
| | | Prophet-SVR | 1.02 × 10+5 | 1.24 × 10+10 | 1.11 × 10+5 |
| **Xu et al. (2019)[4]** | Transportation: Passenger Traffic | SARIMA | 166.16 | | 200.45 |
| | | SARIMA_SVR 1 | 103.20 | | 123.72 |
| | | SARIMA_SVR 2 | 116.9 | | 147.48 |
| | | **SARIMA_SVR 3** | **102.80** | | **131.14** |
| | | SARIMA_SVR 4 | 176.22 | | 214.66 |
| **Zhao et al. (2023)[5]** | Transportation: Railway Freight | Prophet | 153.134 | | 200.011 |
| | | DeepAR | 158.323 | | 201.74 |
| | | Prophet-DeepAR | 112.321 | | 155.435 |
| | | LSTM | 155.54 | | 211.525 |
| | | Prophet-LSTM | 161.197 | | 209.52 |
| | | Wavelet | 274.526 | | 353.876 |
| | | Prophet-Wavelet | 236.215 | | 294.336 |
| | | Bilstm | 164.66 | | 216.714 |
| | | Prophet-Bilstm | 151.532 | | 206.14 |
| **Agyemang et al. (2023) [6]** | Transportation: Road Accidents | SARIMA | 6.5000 | 82.1667 | 9.0646 |
| **Sah et al. (2022)[8]** | Medicine: COVID-19 Cases | RNN | | | 120.35 |
| | | GRU | | | 94.558 |
| | | LSTM | | | 134.505 |
| | | Linear Regression | | | 284809.4 |
| | | Polynomial Regression | | | 149117.8 |
| | | ARIMA | | | 1260 |
| | | Prophet | | | 568.58 |
| | | LSTM-GRU | | | 69.9 |
| **Saeed et al. (2023)[14]** | Transportation: Container Rates | Prophet | 275.0511 | | 348.9746 |
| | | Prophet-Event | 259.5971 | | 306.4438 |
| | | SARIMA | 240.2749 | | 305.3528 |
| **Nurhamida** | Transportation | Holt-Winters | 19390 | 72179449 | |

| | | | | | |
|---|---|---|---|---|---|
| **h et al. (2020)[15]** | : Seasonal Time Series | | | 9 | |
| **Jha & Pande (2021)[16]** | Sales: Supermarket Sales | ARIMA | 22993.57 | | 151.64 |
| | | Holt-Winters | 7344.49 | | 85.7 |
| | | Prophet | 4329.64 | | 65.8 |

**TABLE 2: Performance Evaluation Metrics**

| Metric | Name | Equation |
|---|---|---|
| **MAE** | Mean Absolute Error | $\frac{1}{n}\sum_{i=1}^{n} \lvert \hat{y}i - yi \rvert^2$ |
| **MSE** | Root Mean Square Error | $\frac{1}{n}\sum\sum_{i=1}^{n}(\hat{y}i - yi)^2$ |
| **RMSE** | Root Mean Square Error | $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}i - yi)^2}$ |

Where:

- **MAE:** measures the average absolute difference between the predicted values ($\hat{y}i$) and the actual values ($yi$)
- **MSE:** Measures the average squared difference between predicted ($\hat{y}i$) and actual ($yi$) values.
- **RMSE**: Measures the average magnitude of the error, in the same units as the original data.

# 5. Model Selection and Methodology:

This study introduces a novel hybrid "smart model" for short-term demand forecasting, the proposed model framework is illustrated in [Figure 1]:

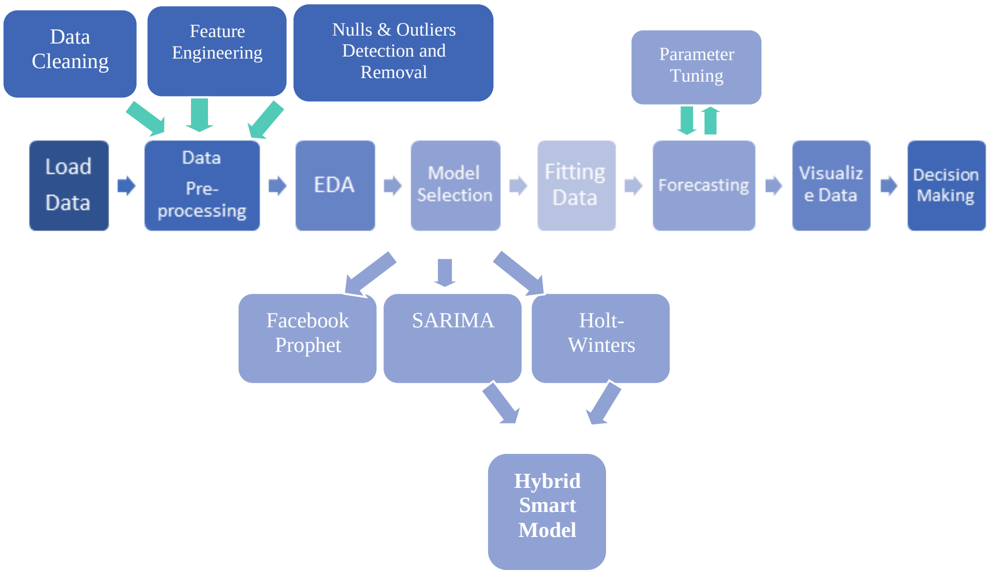


**FIGURE 1: The Proposed Smart Model Framework**

### 5.1 Acknowledging limitations and rationale:

Due to the seasonal nature of the data and the limited timeframe available, we focused on short-term forecasting models to capture the relevant patterns effectively.

To leverage the strengths of different approaches, we employed three widely used models and further explored the creation of a hybrid model for potentially improved forecasting performance.

- **Prophet:**

The Prophet forecasting model was investigated for its potential in predicting passenger demand. This model, known for its simplicity and high accuracy, automatically detects trends and incorporates seasonality and noise within the data to generate forecasts [3]. A key strength of Prophet is its ability to model trends at various granularities, making it suitable for capturing daily, weekly, monthly, and yearly patterns in passenger demand [4]. Although offering benefits like efficient fitting and adaptability to custom seasonal patterns, Prophet may not be ideal for highly complex time series or situations with limited historical data [5].

This model, available in Python and R, comprises four components [3] which is mathematically can be represented as:

$$\mathbf{y(t) = g(t) + s(t) + h(t) + \varepsilon t} \qquad (1)$$

- **Trend Function** g(t): Captures non-periodic changes in the time series.
- **Periodic Variation** s(t): Represents seasonal fluctuations, typically on a weekly or yearly basis.
- **Holiday Effect** h(t): Incorporates the impact of specific days, such as holidays, on the time series.
- **Error Term** (ε): Accounts for residual variability not explained by the model.

- **ARIMA and SARIMA:**

**ARIMA:** This study evaluated the ARIMA (Autoregressive Integrated Moving Average) model, a popular statistical method for time series analysis. ARIMA incorporates three components:

- **Autoregression (AR):** Uses past values of the data to predict future values.
- **Integration (I):** Makes the data stationary (constant mean and variance) by differencing it if needed.
- **Moving Average (MA):** Considers past forecast errors to improve model accuracy.

**SARIMA:** For time series with seasonality (e.g., daily or monthly patterns), the Seasonal ARIMA (SARIMA) model was investigated. SARIMA addresses seasonality by differencing the data based on the seasonal period.

- **Holt-Winters:**

This study explored the Holt-Winters exponential smoothing method for its effectiveness in capturing trends and seasonality in passenger demand data. Here's a breakdown:

- ***Strengths:***
  - Handles both trends and seasonality
  - Works with non-stationary data
  - Flexible for various forecasting tasks

- **Holt-Winters Additive Model:** This specific variant is suitable for time series with consistent seasonal amplitude regardless of the average level (e.g., daily passenger variations remain similar throughout the year).
- Applicable when there's no clear trend or the seasonal pattern doesn't depend on data size.

The core of the model involves calculating exponential smoothing for level and trend components, along with seasonal indices.

**Here's the equation for level smoothing:**

$$S_t = \alpha(x_t - I_{t-m}) + (1-\alpha)(S_{t-m} + b_{t-1})$$

where:

- $\boldsymbol{S_t}$: Exponential smoothing in time t
- $\boldsymbol{S_{t-m}}$: Exponential smoothing in time $t-m$
- $\boldsymbol{b_t}$=Smoothing trend elements in time t
- $\boldsymbol{b_{t-1}}$: Element smoothing trend on year $t-1$
- $\boldsymbol{x_t}$: Actual observation at time *t*
- $\boldsymbol{\alpha}$ : Exponential smoothing parameter for data ($0 < \alpha < 1$)
- $\boldsymbol{I_{t-m}}$: Seasonal index at time $t-m$ (where $m$ is the seasonal period).

The choice of these models balanced interpretability, statistical rigor, and the ability to adapt to the identified Friday demand pattern.

## 6. Model Configuration and Parameter Tuning:

Each model was configured and optimized to best suit the specific characteristics of the data.

- **For Prophet:** It requires the variable names y (target) and ds (dates) in the time series. no explicit parameter tuning was required due to its inherent flexibility.
- **For SARIMA:** Grid search was employed to find the optimal model order (p, d, q) and seasonal order (P, D, Q, s).

  **SARIMA:** An extension of ARIMA incorporating seasonality. The SARIMA model is denoted as SARIMA (p, d, q) (P, D, Q)s, where:
  - **(p, d, q)**: Non-seasonal ARIMA parameters (same as in the original model).
  - **(P, D, Q)**: Seasonal ARIMA parameters:
  - **P**: Order of the seasonal autoregression (number of lagged seasonal terms).
  - **D**: Order of seasonal differencing (number of times seasonal differencing is applied).
  - **Q**: Order of the seasonal moving average (number of lagged seasonal forecast errors).

- **s**: Seasonality (e.g., 12 for monthly, 4 for quarterly).

The presence of (3,0,1) [7] indicates seasonality with P=3, D=0, Q=1 and s=7 (meaning weekly seasonality) **[Table 3].**

**TABLE 3: Possible Tentative Arima Models Extracted By Auto.Arima Function**

| Model | AIC |
|---|---|
| ARIMA(1,0,1)(1,0,1)[7] | 1749.477 |
| ARIMA(0,0,0)(0,0,0)[7] | 1768.800 |
| ARIMA(1,0,0)(1,0,0)[7] | 1749.030 |
| ARIMA(0,0,1)(0,0,1)[7] | 175 078 |
| ARIMA(0,0,0)(0,0,0)[7] | 1897.762 |
| ARIMA(1,0,0)(0,0,0)[7] | 1759.071 |
| ARIMA(1,0,0)(2,0,0)[7] | 175 805 |
| ARIMA(1,0,0)(1,0,1)[7] | 1749.306 |
| ARIMA(1,0,0)(0,0,1)[7] | 1749.083 |
| ARIMA(1,0,0)(2,0,1)[7] | 1747.565 |
| ARIMA(1,0,0)(3,0,1)[7] | 1743.722 |
| ARIMA(1,0,0)(3,0,0)[7] | 1752.716 |
| ARIMA(1,0,0)(3,0,2)[7] | 1745.348 |
| ARIMA(1,0,0)(2,0,2)[7] | inf |
| ARIMA(0,0,0)(3,0,1)[7] | 1753.694 |
| ARIMA(2,0,0)(3,0,1)[7] | 1745.650 |
| ARIMA(1,0,1)(3,0,1)[7] | 1744.366 |
| ARIMA(0,0,1)(3,0,1)[7] | 1744.302 |
| ARIMA(2,0,1)(3,0,1)[7] | 1743.463 |
| ARIMA(2,0,1)(2,0,1)[7] | 1748.042 |
| ARIMA(2,0,1)(3,0,0)[7] | inf |
| ARIMA(2,0,1)(3,0,2)[7] | 1745.422 |
| ARIMA(2,0,1)(2,0,0)[7] | 1743.554 |
| ARIMA(2,0,1)(2,0,2)[7] | inf |
| ARIMA(3,0,1)(3,0,1)[7] | 1744.684 |
| ARIMA(2,0,2)(3,0,1)[7] | 1745.638 |
| ARIMA(1,0,2)(3,0,1)[7] | 1743.862 |
| ARIMA(3,0,0)(3,0,1)[7] | 1747.893 |
| ARIMA(3,0,2)(3,0,1)[7] | 1745.963 |
| **ARIMA(2,0,1)(3,0,1)[7]** | **inf** |

**Best model: ARIMA(2,0,1)(3,0,1)[7]**

- **For Holt-Winters:** Nelder-Mead optimization was used to identify the best smoothing parameters.

### 6.1 Addressing Holiday Demand Pattern:

Initially, most of the models struggled to accurately capture the distinct lower demand on Fridays, consistently exhibiting higher errors compared to other days **[fig 17].** This highlighted the need for a specific approach to address this unique pattern. The inclusion of the "day_of_week" feature as an exogenous variable proved highly effective in **SARIMA [Table 4].** This allowed the model to explicitly consider the weekly cycle and significantly improved its ability to predict Friday demand, as evidenced by the substantial reduction in both MAE and RMSE compared to the pre-tuning version **[fig 18].**

### 6.2 Hybrid Model

- **Individual Model Training:** We first train independent SARIMA and Holt-Winters models on the training data. The optimal parameters for each model are determined through appropriate methods, such as grid search or information criteria selection **[Table 3]**.

- **Weighted Average Combination:** To create the hybrid model, we combine the forecasts from the individual models using a weighted average approach. The weights are assigned based on the relative performance of each model on a validation set or through expert knowledge.

In this paper, we explore different weight combinations to investigate their impact on the overall performance.

## 7. Evaluation and Results:

The performance of each model was evaluated using various metrics, including MAE, MSE, RMSE **[Table 4].** While SARIMA emerged as the top-performing individual model **[fig 5]**, demonstrating the most accurate forecasts with the lowest overall MAE, MSE, and RMSE, the hybrid model achieved even lower error metrics **[fig 7].**

We investigated weighted average combinations of the SARIMA and Holt-Winters forecasts, as detailed in **[Section 7]**. The hybrid model exhibited superior performance compared to the individual models, as illustrated in **[Table 4]**. This suggests the potential benefits of combining complementary models for improved forecasting accuracy.

As illustrated in **[Figures 2 and 3],** Prophet effectively forecasts daily passenger. This figure highlights the model's proficiency in learning and replicating the cyclical nature (periodicity) and general trends (tendency) observed in the historical data. While the predictions closely match the actual values, minor discrepancies exist. Further analysis into these discrepancies, particularly in relation to potential external factors like seasonality or holidays, could enhance the model's overall accuracy.

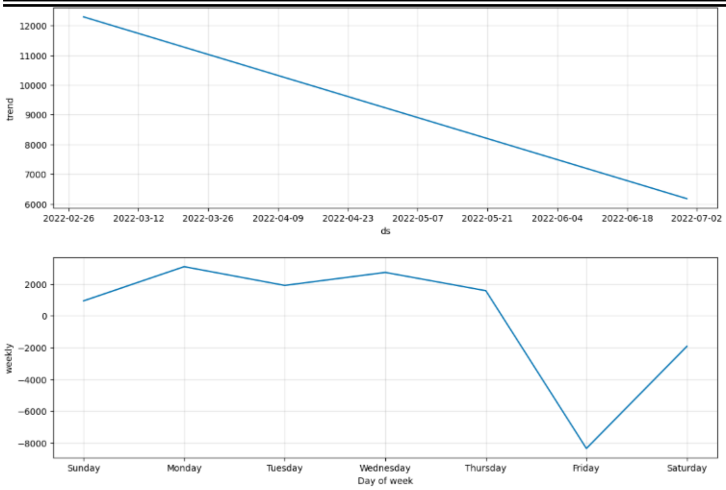


**Figure 2: Prophet Components**

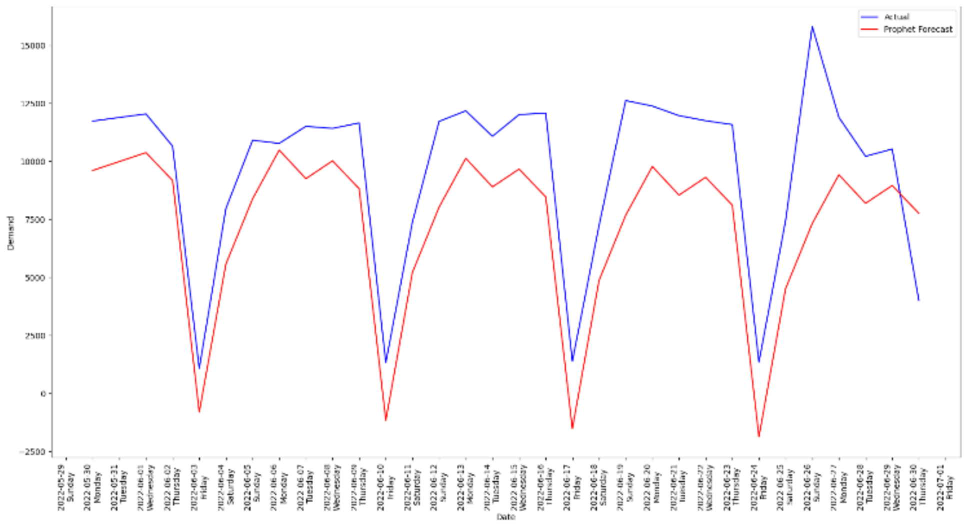


**Figure 3: Prophet Forecast Vs Actual**

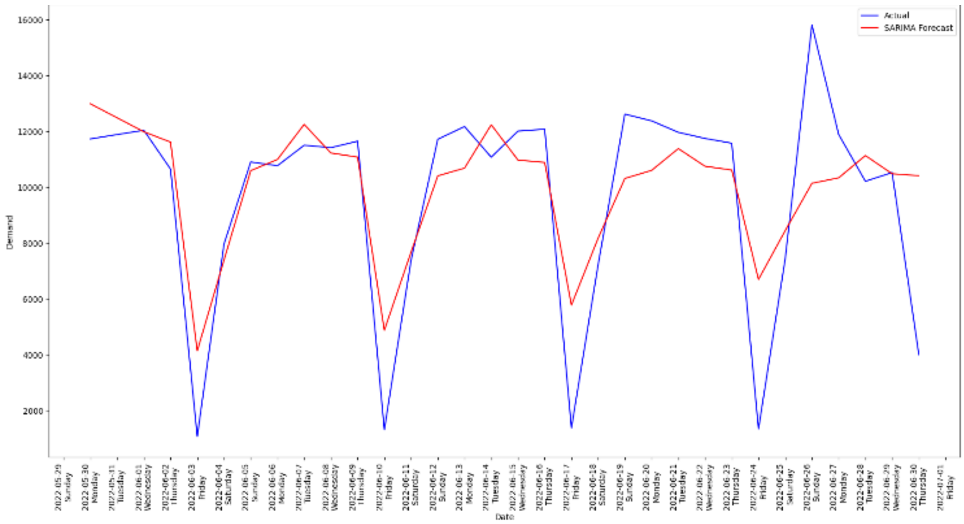


**Figure 4: Sarima Forecast Vs Actual (Before Tuning)**

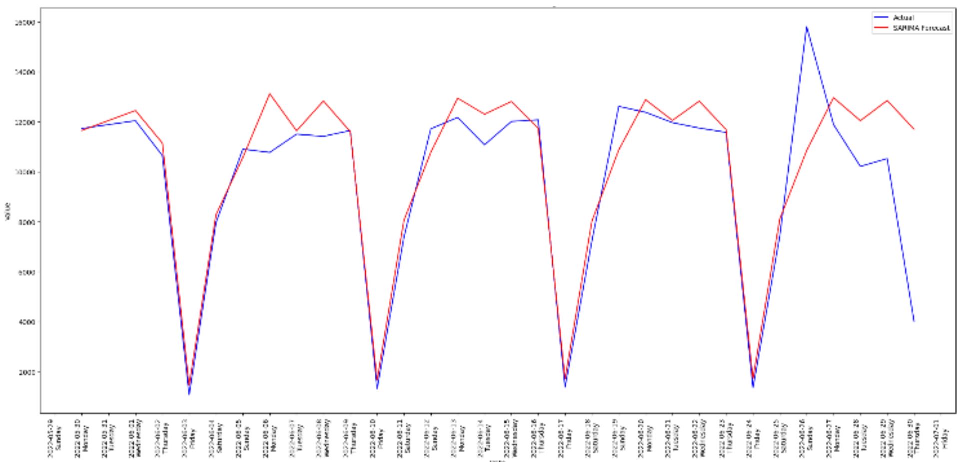

**Figure 5: Sarima Forecast Vs Actual - After Tuning**

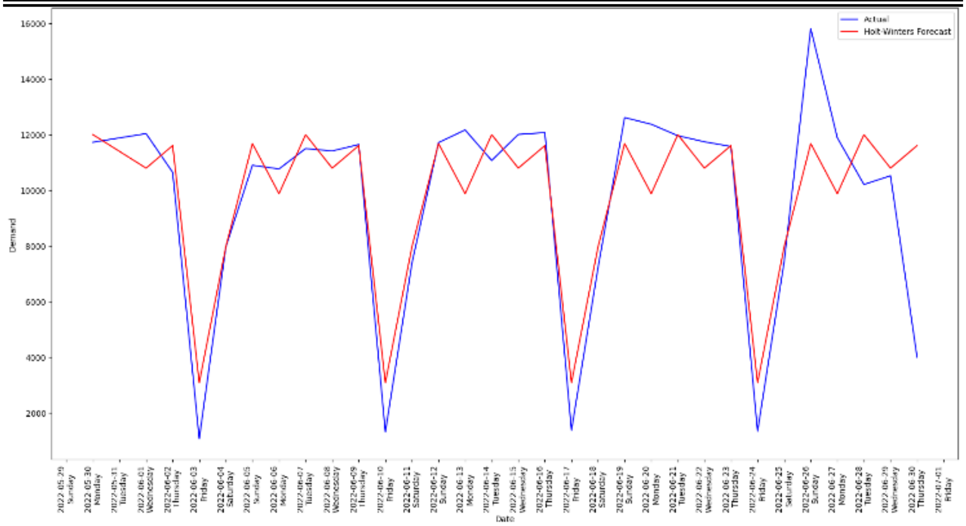


**Figure 6: Holt-Winters Forecast Vs Actual**

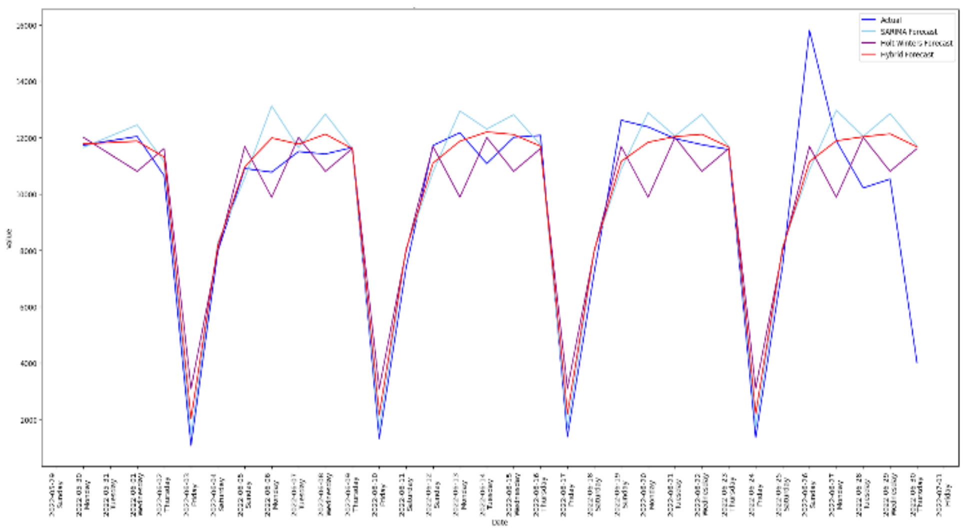


**Figure 7: Hybrid Model Forecast Vs Actual**

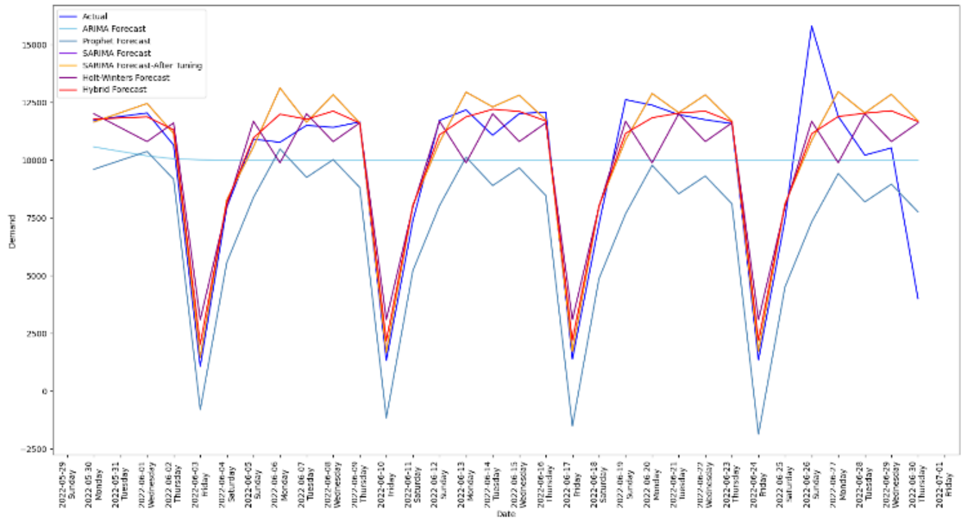


**Figure 8: Comparison On All Models**

**Table 4: Results Of Models Performance**

| Model | Mae | Mse (M) | Rmse |
|---|---|---|---|
| **Prophet [Fig 3]** | 2,709 | 9,2 | 3,032 |
| **SARIMA (Before Tuning) [Fig 4]** | 1,643 | 5,5 | 2,348 |
| **SARIMA (After Tuning) [Fig 5]** | 1,117 | 3.6 | 1,902 |
| **Holt-Winters [Fig 6]** | 1277 | 3.8 | 1942 |
| **SARIMA/Holt-Winters Hybrid Model [Fig 7]** | **957** | **3.2** | **1776** |

# 8. Conclusion:

This study explored the application of various short-term forecasting models to predict daily passenger demand, with a focus on overcoming the challenge of capturing the lower demand observed on holidays. By incorporating the "day_of_week" feature and through parameter optimization, SARIMA emerged as a leading contender for addressing this challenge. However, the study further investigated the potential of a hybrid model combining the strengths of SARIMA and Holt-Winters models through a weighted average approach. As detailed in **[Section 5.2],** this hybrid model achieved superior performance compared to both individual models and the base SARIMA model (2,0,1)x(3,0,1)[7], as illustrated in **[Table 4]**. This finding suggests that combining complementary models can further enhance forecasting accuracy by leveraging their diverse functionalities. Therefore, while the SARIMA (2,0,1)x(3,0,1)[7] model proved effective as an individual solution, this study highlights the potential benefits of exploring hybrid models to potentially achieve even greater accuracy in short-term passenger demand forecasting, particularly when dealing with complex data patterns like holiday effects.

## Further Research:

This study offers several promising avenues for further exploration and refinement in the domain of passenger demand forecasting, future studies could explore:

- **Spatiotemporal analysis:** Incorporating spatial data (e.g., bus stops, routes) to investigate demand variations across different regions within the city.
- **External factors:** Investigating the impact of external factors (e.g., weather, special events) on passenger demand.

- **Real-time forecasting:** Developing models that incorporate real-time data (e.g., traffic information, passenger arrivals) for more dynamic predictions.

## References:

[1] Z. Z. OO and S. PHYU, “Time Series Prediction Based on Facebook Prophet: A Case Study, Temperature Forecasting in Myintkyina,” *International Journal of Applied Mathematics Electronics and Computers*, vol. 8, no. 4, pp. 263–267, Dec. 2020, doi: 10.18100/IJAMEC.816894.

[2] G. Jun-Gang, “Time-Series Prediction Research Based on Combined Prophet-LSTM Models.” EasyChair, Mar. 07, 2023.

[3] L. Guo, W. Fang, Q. Zhao, and X. Wang, “The hybrid PROPHET-SVR approach for forecasting product time series demand with seasonality,” *Comput Ind Eng*, vol. 161, p. 107598, Nov. 2021, doi: 10.1016/J.CIE.2021.107598.

[4] S. Xu, H. K. Chan, and T. Zhang, “Forecasting the demand of the aviation industry using hybrid time series SARIMA-SVR approach,” *Transp Res E Logist Transp Rev*, vol. 122, pp. 169–180, Feb. 2019, doi: 10.1016/J.TRE.2018.12.005.

[5] F. Zhao, B. Dong, Y. Sun, and S. Huang, “Analysis and Forecast of Railway Freight Volume based on Prophet-Deep AR Model,” *Tehnički vjesnik*, vol. 30, no. 4, pp. 1126–1134, Jun. 2023, doi: 10.17559/TV-20230226000383.

[6] E. F. Agyemang, J. A. Mensah, E. Ocran, E. Opoku, and E. N. N. Nortey, “Time series based road traffic accidents forecasting via SARIMA and Facebook Prophet model with potential changepoints,” *Heliyon*, vol. 9, no. 12, p. e22544, Dec. 2023, doi: 10.1016/J.HELIYON.2023.E22544.

[7] M. Lippi, M. Bertini, and P. Frasconi, “Short-term traffic flow forecasting: An experimental comparison of time-series analysis and supervised learning,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 871–882, 2013, doi: 10.1109/TITS.2013.2247040.

[8] S. Sah, B. Surendiran, R. Dhanalakshmi, S. N. Mohanty, F. Alenezi, and K. Polat, “Forecasting COVID-19 Pandemic Using Prophet,

ARIMA, and Hybrid Stacked LSTM-GRU Models in India," *Comput Math Methods Med*, vol. 2022, 2022, doi: 10.1155/2022/1556025.

[9] R. Adhikari and R. K. Agrawal, "An Introductory Study on Time Series Modeling and Forecasting," Feb. 2013, Accessed: Mar. 27, 2024. [Online]. Available: https://arxiv.org/abs/1302.6613v1

[10] C. Hamzaçebi, "Improving artificial neural networks' performance in seasonal time series forecasting," *Inf Sci (N Y)*, vol. 178, no. 23, pp. 4550–4559, Dec. 2008, doi: 10.1016/J.INS.2008.07.024.

[11] G. P. Zhang, "A neural network ensemble method with jittered training data for time series forecasting," *Inf Sci (N Y)*, vol. 177, no. 23, pp. 5329–5346, Dec. 2007, doi: 10.1016/J.INS.2007.06.015.

[12] L. J. Cao and F. E. H. Tay, "Support vector machine with adaptive parameters in financial time series forecasting," *IEEE Trans Neural Netw*, vol. 14, no. 6, pp. 1506–1518, Nov. 2003, doi: 10.1109/TNN.2003.820556.

[13] P. G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, Jan. 2003, doi: 10.1016/S0925-2312(01)00702-0.

[14] N. Saeed, S. Nguyen, K. Cullinane, V. Gekara, and P. Chhetri, "Forecasting container freight rates using the Prophet forecasting method," *Transp Policy (Oxf)*, vol. 133, pp. 86–107, Mar. 2023, doi: 10.1016/J.TRANPOL.2023.01.012.

[15] N. Nurhamidah, N. Nusyirwan, and A. Faisol, "FORECASTING SEASONAL TIME SERIES DATA USING THE HOLT-WINTERS EXPONENTIAL SMOOTHING METHOD OF ADDITIVE MODELS," *Jurnal Matematika Integratif*, vol. 16, no. 2, p. 151, Dec. 2020, doi: 10.24198/JMI.V16.N2.29293.151-157.

[16] B. Kumar Jha and S. Pande, "Time Series Forecasting Model for Supermarket Sales using FB-Prophet," *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, pp. 547–554, Apr. 2021, doi: 10.1109/ICCMC51019.2021.9418033.