Uber Demand Forecast in New York City Project Report

Davide Christian Mancosu Bustos †, Mehran Farajinegarestan†, Karim Eugenio Hamdar †

***Abstract*—Ridesharing services like Uber have become integral to urban transportation systems, necessitating accurate demand forecasting to enhance operational efficiency and customer satisfaction. This study aims to forecast hourly Uber pickups in New York City using a dataset sourced from the NYC Taxi & Limousine Commission (TLC), obtained via a Freedom of Information Law request. Various time series and regression models, including Time Series Linear Models (TSLM), ARIMA, ARIMAX, Generalized Additive Models (GAMs), Gradient Boosting, Exponential Smoothing, and Local Regression, were applied to capture trends, seasonality, and the influence of exogenous variables such as weather conditions, gas prices, and wind speed. Model performance was evaluated using metrics such as the coefficient of determination (R^2), Akaike Information Criterion (AIC), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The findings highlight the potential for improved demand prediction, leading to reduced surge pricing events, optimized driver allocation, and enhanced urban transportation planning.**

***Index Terms*—Time Series Analysis, Forecasting, Demand Prediction**

I. Introduction

Ridesharing services, including Uber and Lyft, have revolutionized urban transportation by offering convenient, cost-effective, and flexible mobility solutions. Accurate demand forecasting is critical for such platforms to maintain operational efficiency, minimize surge pricing, and optimize resource allocation. Furthermore, precise predictions can assist drivers in identifying high-demand periods and help city planners manage traffic flow more effectively.

This project focuses on forecasting hourly Uber pickups in New York City, leveraging a rich dataset obtained from the NYC Taxi & Limousine Commission (TLC) through a Freedom of Information Law request. The dataset spans six months and includes hourly observations of the number of pickups. Furthermore, various explanatory variables such as weather conditions (temperature, precipitation, wind speed), gas prices are gathered from other sources as exogenous variables to enrich our dataset. Also temporal factors (hour of the day, day of the week, and seasonality) are extracted based on the date-time of each recording.

To capture the complex dynamics of ridesharing demand, we applied a diverse range of modeling techniques, including statistical models (TSLM, ARIMA, ARIMAX, Exponential Smoothing), machine learning approaches (Gradient Boosting), and non-linear models (Generalized Additive Models and Local Regression). These methods were selected to explore different facets of the data, from linear relationships to non-linear trends and interactions.

Model performance was assessed using rigorous evaluation metrics, including the coefficient of determination (R^2) , Akaike Information Criterion (AIC), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The results of this study demonstrate the feasibility of accurately forecasting Uber demand and its potential applications in enhancing ridesharing services, improving customer satisfaction, and supporting urban planning initiatives.

The remaining is structured as follows:

-Exploratory Data Analysis (Section 2): Analyzes temporal trends, seasonality, outliers, and relationships between predictors and Uber pickups, supported by visualizations.

-Modeling (Section 3): Describes the application and evaluation of different models, including TSLM, ARIMA, Exponential Smoothing, GAMs, and Gradient Boosting, with a focus on methodology and performance.

-Results (Section 4): Summarizes model performance using metrics like MSE, AIC, and Adjusted R^2, and compares actual vs. predicted values with residual diagnostics.

-Conclusion (Section 5): Highlights key findings, discusses model limitations, and provides recommendations for improving demand forecasting through advanced techniques and additional predictors.

-References (Section 6): Lists the sources and tools used in the project.

II. Exploratory Data Analysis

Exploratory Data Analysis (EDA) was performed to examine the temporal trends and variability in the Uber pickups dataset. The dataset spans six months, from January 1st, 2015, to June 30th, 2015, and contains 4,334 hourly observations of Uber pickups. Figure 1 provides a detailed time series plot of hourly pickups over the six-month period.

A. Time Series Overview

The time series plot (Figure 1) reveals the following key patterns:

- **Overall Trend:** An increasing trend in the number of pickups is observed over time, reflecting the growing popularity and usage of Uber during this period.

†Department of Mathematics, University of Padova, email: {davidechristian.mancosubustos}@studenti.unipd.it
†Department of Mathematics, University of Padova, email: {mehran.farajinegarestan}@studenti.unipd.it
†Department of Mathematics, University of Padova, email: {karimeugenio.hamdar}@studenti.unipd.it

- **Seasonality:** Periodic fluctuations in pickups suggest the presence of daily and weekly seasonal patterns. Higher demand is likely on weekends or specific times of the day.
- **Outliers:** Several sharp spikes in the number of pickups indicate periods of exceptionally high demand, potentially caused by special events, holidays, or weather anomalies.

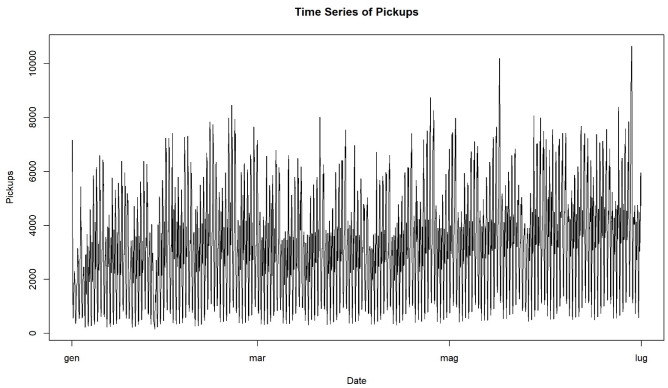


Fig. 1: Time series of hourly Uber pickups from January to June 2015. The plot shows a general upward trend, with periodic fluctuations and occasional sharp spikes.

B. Outlier Detection and Handling

The presence of outliers in the `pickups` time series was identified through a combination of visual inspection using boxplots (Figure 2) and statistical analysis. Outliers were detected as data points that deviated significantly from the interquartile range (IQR), indicating unusually high or low pickup values that might distort the analysis.

To address the impact of these outliers, the following steps were taken:

- **Identification:** Outliers were identified using robust statistical techniques based on IQR thresholds.
- **Replacement:** Detected outliers were replaced with interpolated values derived from neighboring observations. This approach preserves the overall temporal structure of the data while mitigating the influence of extreme values.

This preprocessing step ensures that the integrity of the time series is maintained, facilitating accurate and unbiased forecasting in subsequent analyses. The handling of outliers is particularly critical in demand forecasting, where extreme values can significantly impact model performance and insights.

C. Trend and Seasonality Analysis

Building on the observations from the time series plot, we further investigate the presence of seasonality in Uber pickups. A preliminary analysis was conducted by examining the average number of pickups across different intervals of the day. Figure 3 shows that Uber pickups follow a clear daily

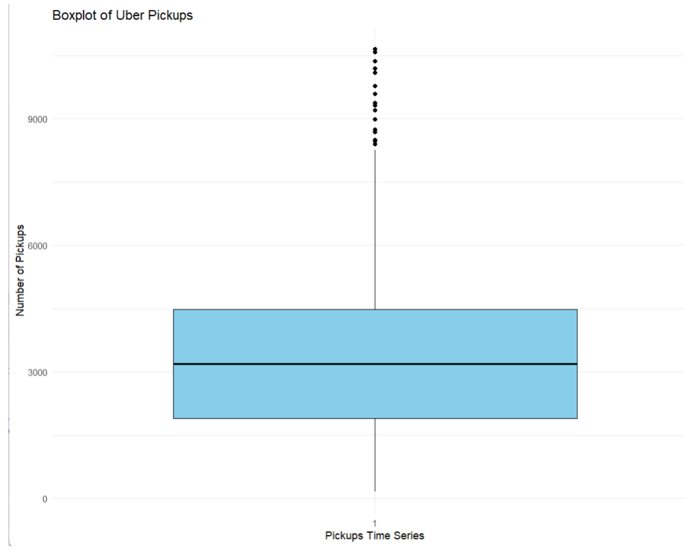


Fig. 2: Boxplot of Uber Pickups: The boxplot highlights the distribution of pickup values, with dots representing detected outliers.

pattern, with demand peaking in the evening and declining sharply during the night. This observation strongly suggests the presence of daily seasonality in the data

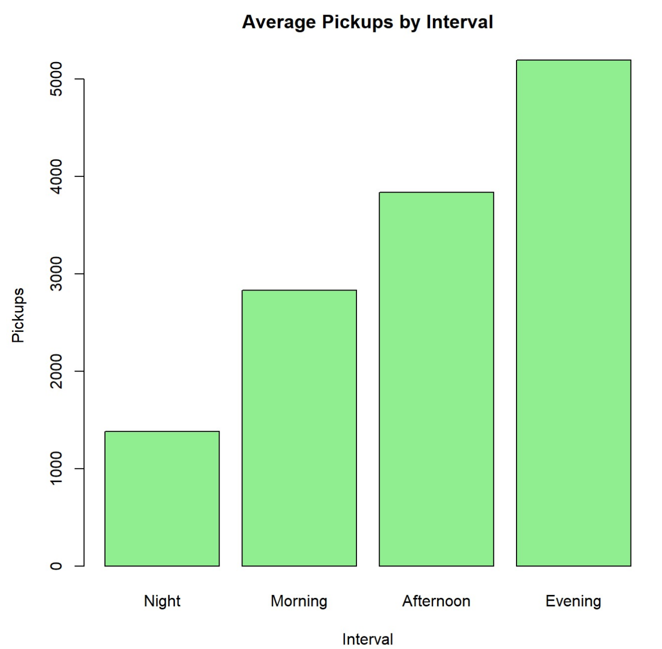


Fig. 3: Average Uber pickups by daily interval. The plot suggests a daily seasonality pattern, with the highest demand occurring in the evening and the lowest during the night.

To confirm this hypothesis and delve deeper into the patterns of seasonality and trend, we applied STL (Seasonal and Trend decomposition using Loess) decomposition to the time series. The decomposition separates the observed data into trend, seasonal, and residual components, as shown in Figure 4.Figure 4 illustrates the decomposition, highlighting the distinct roles of these components.

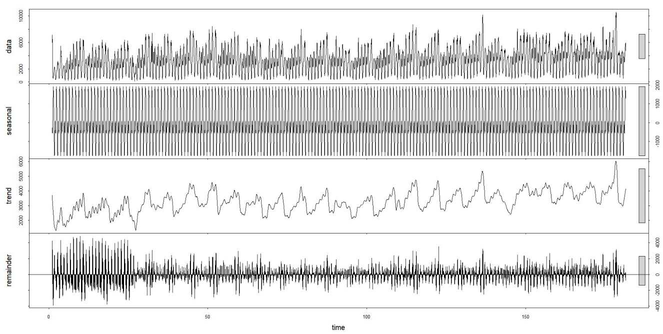


Fig. 4: STL Decomposition of Uber Pickups Time Series: The observed time series is decomposed into seasonal, trend, and remainder components.

The results confirm and extend the hypotheses proposed in the **Time Series Overview** subsection:

- **Trend Confirmation:** The gradual upward trend observed in the time series plot is validated by the decomposition, with the trend component showing a consistent increase over time.
- **Seasonality Confirmation:** The periodic fluctuations noted earlier are further validated, with the seasonal component capturing regular daily and weekly cycles.
- **Outliers and Residuals:** The decomposition also highlights the presence of significant variability in the residual component, which likely corresponds to the previously noted sharp spikes in demand caused by external factors.

To quantify the importance of the trend and seasonal components, we calculated their respective strengths using statistical measures:

- **Trend Strength:** 0.347 (approximately 35%)
- **Seasonality Strength:** 0.523 (approximately 52%)

a) Interpretation::

- **Seasonality Dominance:** A seasonality strength of 0.523 indicates that recurring patterns, such as daily and weekly cycles, explain around 52% of the variability in Uber pickups. This aligns with expectations for urban mobility patterns, which are often influenced by daily routines and weekend effects.
- **Moderate Trend Contribution:** The trend strength of 0.347 suggests that the long-term increase in Uber pickups accounts for approximately 35% of the variance. While significant, this indicates that the trend plays a secondary role compared to seasonality.
- **Residual Variability:** The remainder component accounts for the remaining variability, highlighting the presence of random noise or factors not explained by trend or seasonality. These may include special events, holidays, or external variables like weather and gas prices.

b) Modeling Implications:: The results underscore the importance of capturing both seasonality and trend in the modeling process:

- Seasonal patterns should be explicitly modeled to account for the dominant periodic fluctuations in the data.
- The trend component should be included to reflect the gradual increase in Uber demand over time.
- Residual variability may be addressed by incorporating additional predictors or using advanced forecasting techniques.

This extended analysis validates and strengthens the hypotheses proposed earlier, confirming that seasonality is the dominant component in the time series, while the trend contributes moderately. These findings emphasize the need for models that effectively capture temporal dependencies and cyclical behaviors to improve forecasting accuracy.

D. Temporal Features and Correlation Analysis

To further analyze the temporal dynamics, features such as the hour of the day, day of the week, and weekend indicators were extracted from the dataset. Additionally, weather-related variables such as temperature, precipitation, and wind speed were integrated to explore their relationship with pickup demand.

Autocorrelation (ACF) and partial autocorrelation (PACF) plots were generated to investigate the time series' dependence structure. The ACF plot (Figure 5) highlights significant positive correlations at short lags, reflecting strong temporal dependencies in hourly demand. Peaks at specific lag intervals further suggest the presence of daily and weekly seasonal patterns, while the diminishing correlations at longer lags point to gradual fading of influence from prior time points.

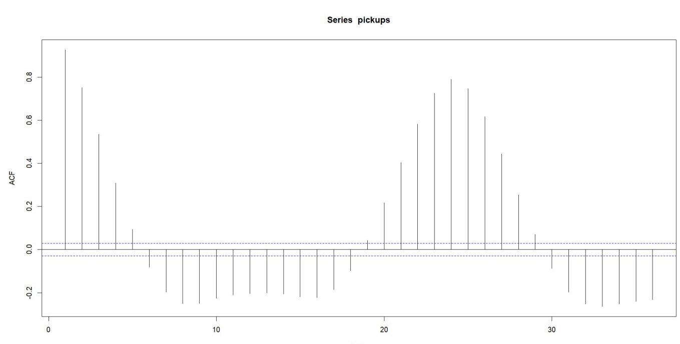


Fig. 5: Autocorrelation Function (ACF) plot of hourly Uber pickups. The plot reveals strong positive correlations at shorter lags and periodic peaks corresponding to daily and weekly seasonal patterns.

The partial autocorrelation function (PACF) plot (Figure 6) complements the ACF analysis by isolating the direct correlations between the time series and lagged values. It exhibits significant spikes at shorter lags, suggesting the need for models that account for both immediate and seasonal dependencies. The PACF's rapid decay also confirms the presence of

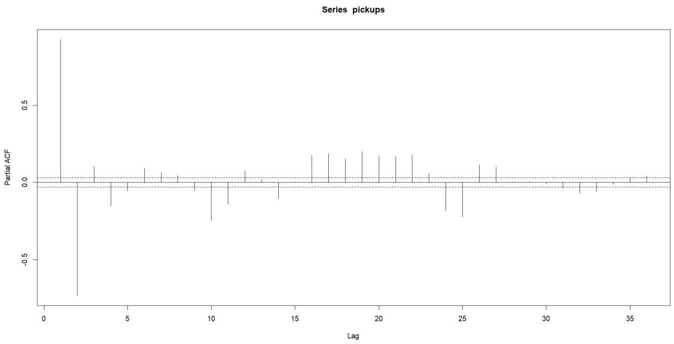


Fig. 6: Partial Autocorrelation Function (PACF) plot of hourly Uber pickups. The plot highlights significant spikes at shorter lags, indicating direct correlations and reinforcing the evidence of seasonality in the data.

underlying seasonal patterns, further supporting the temporal dependencies identified in the ACF plot.

E. Insights for Modeling

The insights gained from EDA highlight the importance of accounting for both temporal factors and exogenous variables (e.g., weather and gas prices) when modeling Uber pickup demand. The observed seasonality and periodic peaks strongly suggest the need for models capable of capturing temporal dependencies and cyclical patterns. Additionally, external factors such as temperature and precipitation are expected to significantly influence demand fluctuations. These findings guided the selection of appropriate statistical and machine learning models for forecasting.

F. Relationship Between Predictors and Uber Pickups

The dataset includes several quantitative variables that exhibit diverse time series behaviors, providing valuable insights into their potential relationships with Uber demand. Figure 7 presents a comparison of the temporal patterns of these variables.

The time series of `pickups`, representing Uber demand, reveals a general upward trend and exhibits clear daily and weekly seasonality. `Temperature` shows a gradual increase over the six-month period, aligned with seasonal changes, and displays pronounced daily cycles. `Gas price` remains relatively stable, with only minor variations over time. In contrast, `total precipitation` is characterized by sporadic spikes, indicating periods of heavy rainfall. The variables `u_wind` and `v_wind`, which represent wind components, fluctuate around zero, reflecting directional and intensity changes over time.

Among these variables, temperature shows notable variability and trends that make it a compelling factor for further exploration. Its seasonal and daily patterns align with human mobility behaviors, suggesting it could significantly influence Uber demand. This motivates further investigation into temperature as a possible explanatory variable while remaining cautious about its influence.

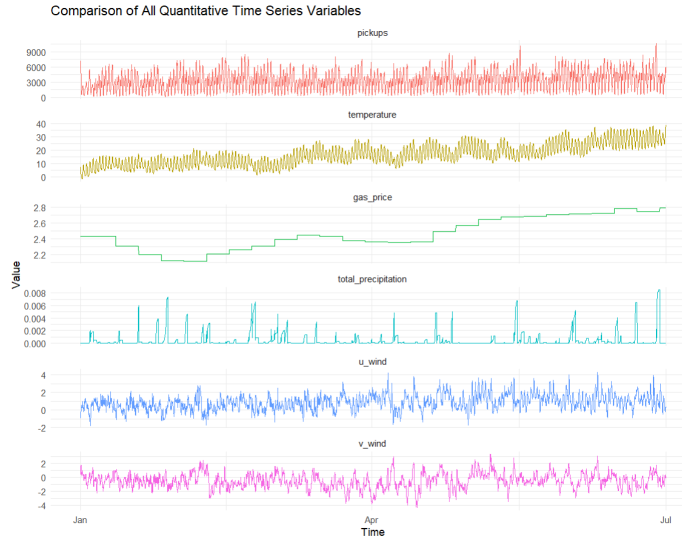


Fig. 7: Comparison of all quantitative time series variables in the dataset. Each variable exhibits distinct temporal patterns that provide insights into their relationship with Uber pickups.

G. Temperature and Its Relationship with Uber Pickups

Temperature is a key factor influencing transportation patterns and user preferences, making it an important variable in understanding Uber demand. Over the six-month period, the temperature time series reveals distinct seasonal and daily patterns, as shown in Figure 8.

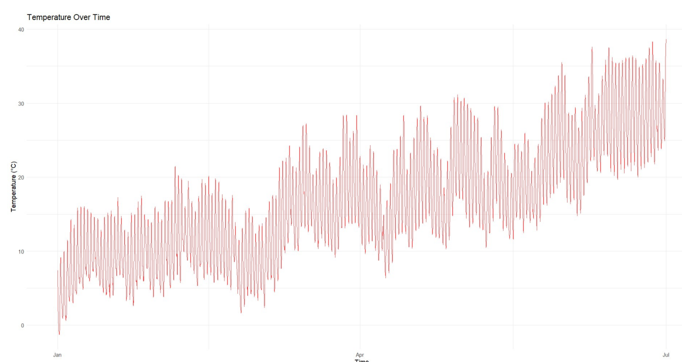


Fig. 8: Temperature Over Time: The plot shows an increasing trend in temperature from January to June, with noticeable daily fluctuations. This pattern reflects seasonal changes and cyclical temperature variations.

The temperature trends upward from January's colder months to the warmer summer months in June, reflecting expected seasonal transitions. Daily fluctuations are evident, with daytime temperatures generally higher and nighttime temperatures cooler. These variations are closely tied to human mobility patterns, where warmer temperatures may encourage outdoor activities and increase ride-sharing demand, while colder conditions might prompt greater reliance on Uber to avoid walking or using public transport.

To examine the potential relationship between temperature and Uber pickups, a scatter plot was created (Figure 9).

The plot highlights a slight upward trend, suggesting a potential positive association between higher temperatures and increased Uber demand.

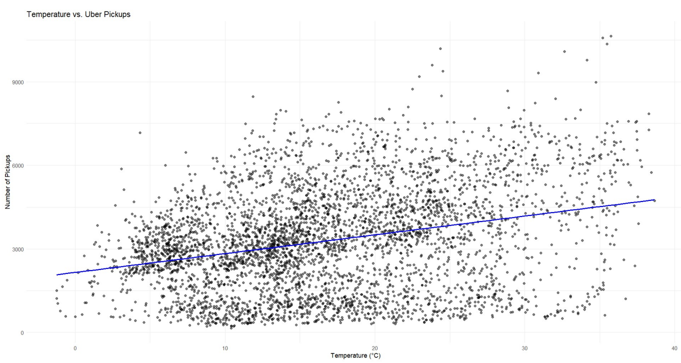


Fig. 9: Scatter plot of temperature vs. Uber pickups, with a fitted regression line. The plot indicates a slight positive correlation, where higher temperatures are associated with increased demand.

The the plot indicates a slight upward trend, suggests a slightly positive correlation as temperatures rise, Uber pickups tend to increase. At higher temperatures, greater variability in the number of pickups is also observed, potentially reflecting other additional or interactions as special events, holidays, or weather anomalies. This analysis highlights the importance of temperature as a predictive variable for demand forecasting, given its strong seasonal and daily patterns and its influence on user behavior.

By incorporating temperature into predictive models, it becomes possible to improve the accuracy of demand forecasts, particularly during periods of seasonal transitions or extreme weather conditions. The focus on temperature is motivated by its consistent variability and evident relationship with Uber demand, making it a significant factor for further exploration.

H. *Correlation Analysis*

To further explore the relationships between the variables in the dataset, a correlation heatmap was generated (Figure 10). The heatmap visualizes the pairwise Pearson correlation coefficients among the variables: `pickups`, `total_precipitation`, `temperature`, `u_wind`, `v_wind`, and `gas_price`.

a) Key Observations::

- **Temperature and Pickups:** A weak positive correlation is observed between `temperature` and `pickups`, indicating that warmer temperatures are slightly associated with increased Uber demand.
- **Gas Price and Pickups:** A weak positive correlation exists between `gas_price` and `pickups`, suggesting a minor increase in Uber demand as gas prices rise.
- **Total Precipitation:** `total_precipitation` shows a negligible correlation with `pickups`, indicating limited impact on Uber demand.

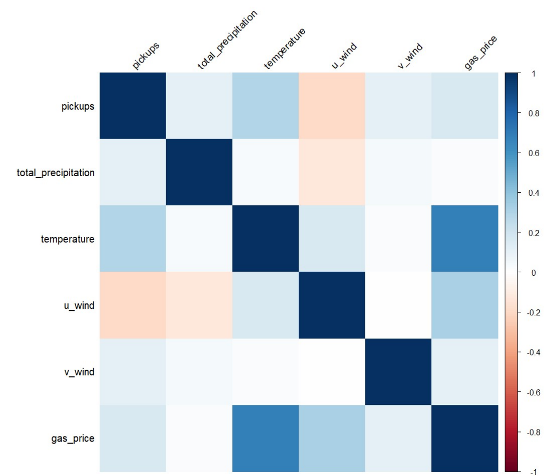


Fig. 10: Correlation heatmap of the variables in the dataset. The color intensity represents the strength of the correlation, ranging from -1 (negative correlation) to 1 (positive correlation).

- **Wind Components:** The correlations involving `u_wind` and `v_wind` are weak across all variables, reflecting minimal influence on Uber demand or the other variables.

This analysis underscores the complexity of modeling Uber demand, as it may be influenced by a combination of these variables or other unobserved factors. The weak correlations suggest that while these predictors can provide some insights, more sophisticated models may be required to capture the interplay of variables and accurately forecast demand.

b) Collinearity Analysis:: There is no evidence of significant collinearity among the variables. The correlation coefficients remain below thresholds that typically indicate collinearity concerns (e.g., $|r| > 0.7$). This ensures that the variables are suitable for use in regression and machine learning models without requiring extensive preprocessing to address multicollinearity.

I. *Multicollinearity Analysis*

To assess the potential multicollinearity between explanatory variables, we calculated the Variance Inflation Factor (VIF) for the variables `gas_price` and `temperature` included in the regression model. The VIF values were:

- `gas_price`: 1.88
- `temperature`: 1.88

Both values are well below the commonly used threshold of 5, indicating no significant multicollinearity between these variables. As such, both variables were retained for subsequent modeling without further adjustment.

III. Modelling

A. *TSLM*

A Time Series Linear Model (TSLM) was developed to predict the number of Uber pickups in New York City using a combination of trend and seasonal components as predictors. For seasonality, both daily seasonality and weekly seasonality are used and compared. While we acknowledge that the data is not purely linear and that the TSLM model might not capture all the complexity of the dataset, this approach serves as an exploratory step to assess the significance of trends and seasonality. The model with daily seasonality is represented as:

$$Y_t = \beta_0 + \beta_1 \cdot \text{trend}_t + \sum_{i=2}^{24} \beta_i \cdot \text{season}_i + \epsilon_t,$$

And the model with weekly seasonality is represented as:

$$Y_t = \beta_0 + \beta_1 \cdot \text{trend}_t + \sum_{i=2}^{168} \beta_i \cdot \text{season}_i + \epsilon_t,$$

where β_0 is the intercept, β_1 captures the linear trend, $\beta_2, \ldots, \beta_{168}$ are the seasonal coefficients, and ϵ_t represents the residual errors.

1) Performance Performance: The performance metrics for the TSLM models with daily and weekly seasonality are summarized in Table **??**.

TABLE 1: Performance Metrics for TSLM Models

TSLM Model	R^2	**RMSE**	**AIC**	**BIC**
Daily Seasonality	0.45	1465	59951	60116
Weekly Seasonality	0.60	989	58755	59831

For the TSLM model with daily seasonality, the Adjusted R^2 of 0.45 indicates that the model explains approximately 45% of the variation in the data, which reflects some limitations in capturing the underlying patterns. The trend coefficient was highly significant (p-value $< 2 \times 10^{-16}$), confirming the presence of a positive linear trend. Additionally, several seasonal coefficients were significant, suggesting that the model captures certain aspects of daily seasonality.

However, the RMSE of 1465 highlights substantial room for improvement, especially in capturing finer seasonal variations. The AIC (59951) and BIC (60116) values are relatively high, indicating potential overfitting and a need for greater model complexity or improved feature representation.

For the TSLM model with weekly seasonality, the Adjusted R^2 of 0.60 suggests a notable improvement over the daily seasonality model, with approximately 60% of the variation explained. The trend coefficient remained highly significant (p-value $< 2 \times 10^{-16}$), and key weekly seasonal coefficients were also statistically significant. This supports the presence of strong weekly patterns in the data.

The RMSE of 989 indicates better predictive accuracy and reduced unexplained variance compared to the daily model. Moreover, the lower AIC (58755) and BIC (59831) values further confirm the superiority of the weekly seasonality model, as it strikes a better balance between fit and complexity.

2) Residual Analysis Using ACF and PACF: The ACF and PACF plots for the residuals of the TSLM model with daily seasonality are shown in Figure 11.

- The ACF plot exhibits significant autocorrelations at multiple lags, indicating that the residuals are not purely random. This suggests that the model does not fully capture all underlying temporal patterns.
- The PACF plot also shows significant spikes at certain lags, reinforcing the observation that the model leaves room for improvement, particularly in addressing autocorrelation within the residuals.

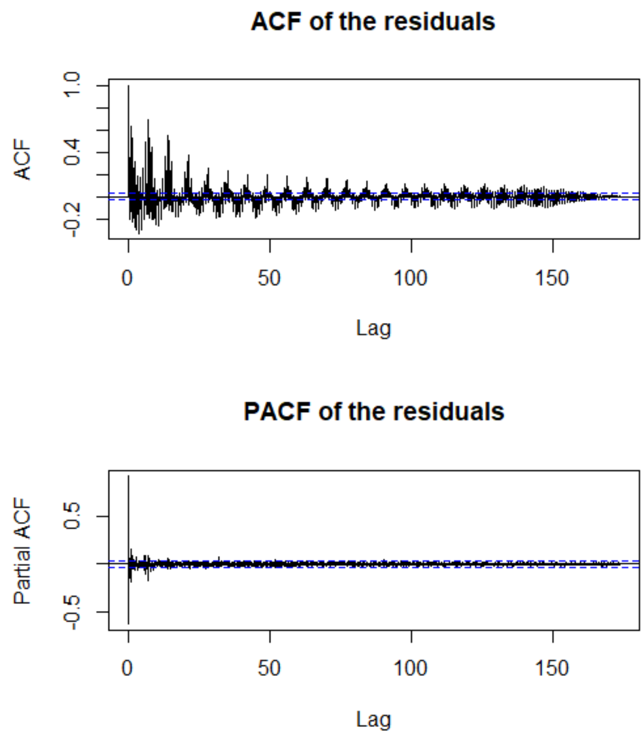


Fig. 11: ACF & PACF plot for TSLM model with daily seasonality

The ACF and PACF plots for the residuals of the TSLM model with weekly seasonality are shown in Figure 12.

- While the ACF plot displays some persistent autocorrelations, the magnitude and frequency of significant spikes are notably reduced compared to the daily model. This indicates that the weekly seasonality model accounts for a larger proportion of the autocorrelated structure in the data.
- The PACF plot shows fewer significant spikes, suggesting better overall performance in capturing the underlying patterns.

B. ARIMA Model

To capture the complex dynamics of Uber pickups in New York City, we fit an ARIMA model to the time series data using the `auto.arima()` function. The selected model, based on AIC minimization, was ARIMA(4,0,2)(2,1,0)[24], incorporating daily seasonality ($m = 24$).

The model's residual variance, σ^2, was estimated at 157098. The AIC and BIC values were 61300.25 and 61357.21, respectively, reflecting a good balance between model complexity and fit.

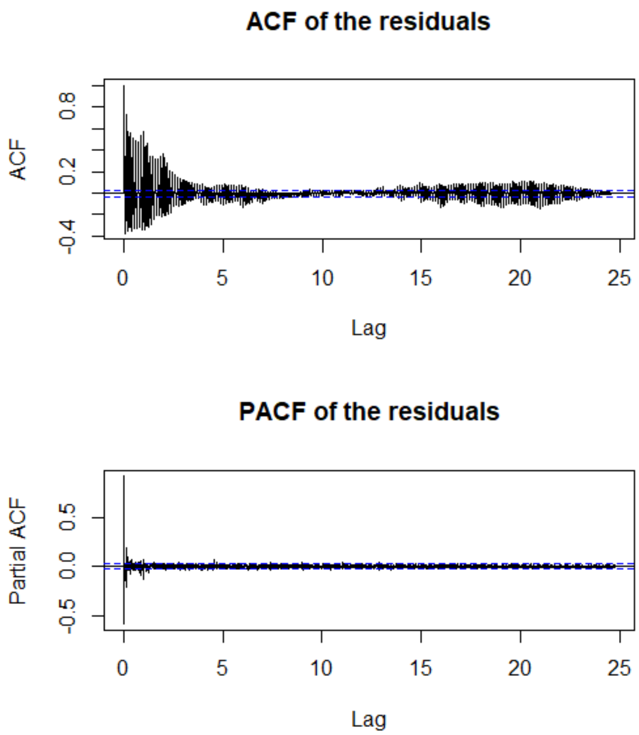


Fig. 12: ACF & PACF plot for TSLM model with weekly seasonality

Model Performance Summary: Table 2 provides a summary of key performance metrics for the ARIMA model. These include the RMSE, MAPE, AIC, and BIC values. The results highlight the model's capability to capture much of the variability in the training data, outperforming the TSLM models discussed earlier.

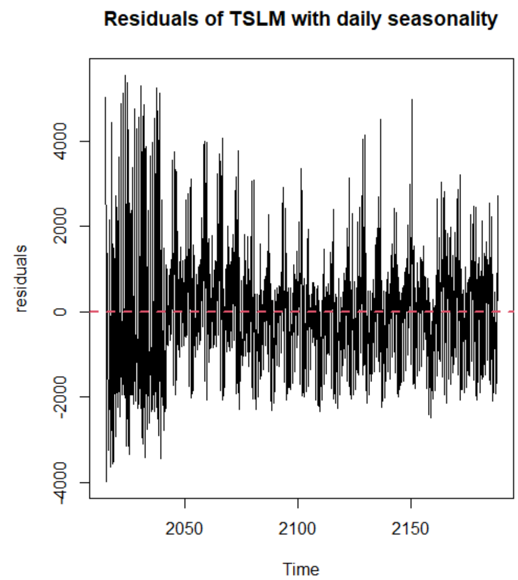
TABLE 2: Performance Metrics for ARIMA Model

Model	RMSE	MAPE (%)	AIC	BIC
ARIMA	394.83	10.80	61300.25	61357.21

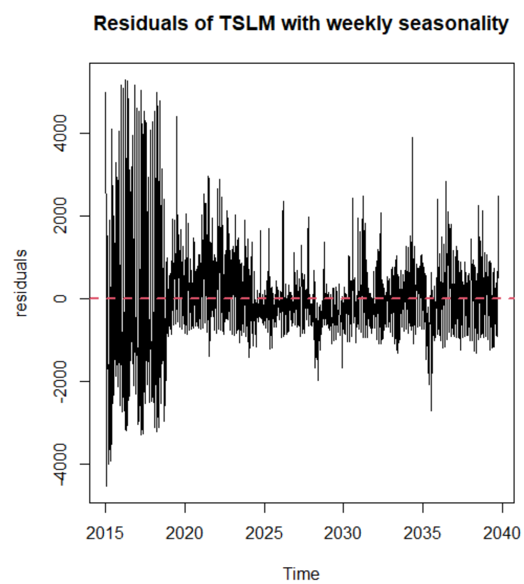
The ARIMA model achieved an RMSE of 394.83 and a MAPE of 10.80%, demonstrating its effectiveness in modeling the training data. Compared to the TSLM models, this represents a substantial improvement in predictive accuracy. However, residual diagnostics reveal areas for refinement, as discussed below.

Residual Analysis: Residual diagnostics were performed to evaluate the adequacy of the ARIMA model. The Ljung-Box test yielded a Q^* statistic of 535.82 with 40 degrees of freedom, corresponding to a p-value less than 2.2e-16. This indicates significant autocorrelation in the residuals, suggesting that the model does not fully capture the underlying data structure.

Figure 14 presents the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots for the residuals, which reveal significant autocorrelations at various lags. This implies that additional model refinements, such as incorporating external regressors or exploring alternative seasonal structures, may further improve the fit.



(a) Residuals of TSLM model with daily seasonality



(b) Residuals of TSLM model with weekly seasonality

Fig. 13: Residuals of TSLM models

Forecasting Performance: The ARIMA model was used to forecast Uber pickups on the test dataset. Figure 15 compares the forecasted values with the actual data. The forecast successfully captures the general trend, though discrepancies are evident in specific periods, likely due to unmodeled complexities.

Conclusion: The ARIMA model provides a solid baseline for forecasting daily Uber pickups, with relatively low RMSE and MAPE values. While the model demonstrates good pre-

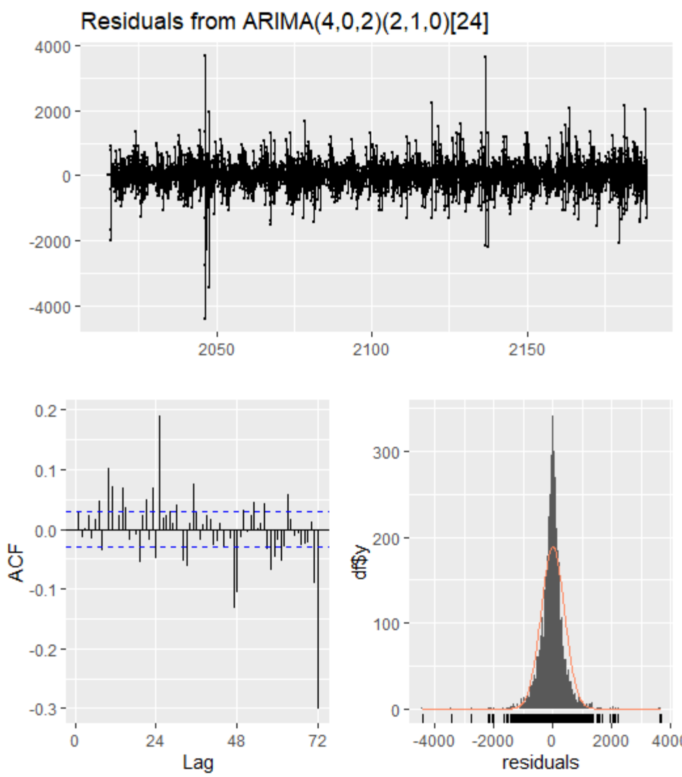


Fig. 14: ACF & PACF plots for ARIMA model residuals with daily seasonality.

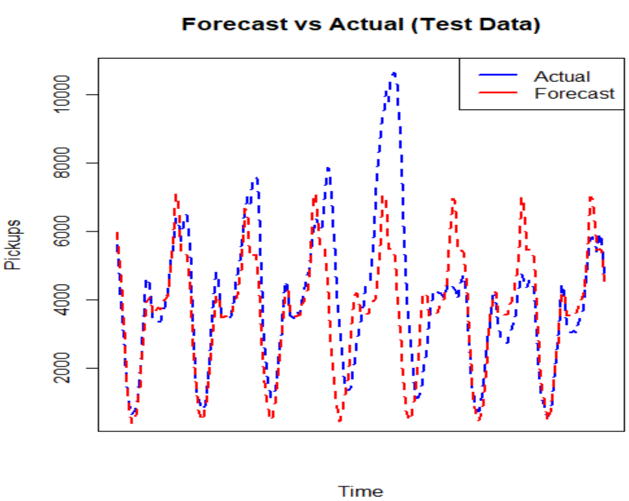


Fig. 15: Forecast vs. Actual Uber pickups for the test dataset using the ARIMA model.

dictive accuracy, residual diagnostics and the Ljung-Box test suggest the need for further refinement to address residual autocorrelations and improve forecast precision.

C. Exponential Smoothing

The Holt-Winters' seasonal method was applied to the training dataset to model both trend and seasonality. Two seasonal approaches were tested: additive and multiplicative. The additive method assumes that seasonal variations remain roughly constant, while the multiplicative method assumes that seasonal variations change proportionally to the level of the series. Based on the observed data, the additive method was hypothesized to perform better.

Model Performance Summary: Table 3 summarizes the performance metrics for the Holt-Winters additive and multiplicative methods. These metrics include the RMSE and MAPE, which were used to evaluate model accuracy on the training dataset.

TABLE 3: Performance Metrics for Holt-Winters Methods

Model	**RMSE**	**MAPE (%)**
Holt-Winters Additive Method	533.09	17.42
Holt-Winters Multiplicative Method	893.42	27.67

The Holt-Winters additive method significantly outperformed the multiplicative method, achieving a much lower RMSE of 533.09 and a MAPE of 17.42%. These results suggest that the additive method is more suitable for this dataset, where seasonality appears constant over time. In contrast, the multiplicative method produced higher errors, likely due to its assumption of proportional seasonal variation, which does not align with the dataset's characteristics.

Forecasting Performance: Both methods were applied to forecast Uber pickups over the test dataset (next 7 days). Figures 16 and 17 illustrate the predictions generated by the additive and multiplicative methods, respectively.

The additive method closely aligns with the actual values, demonstrating its ability to capture the underlying trend and seasonality effectively. On the other hand, the multiplicative method shows larger deviations, particularly during periods of high variability, further reinforcing the conclusion that the additive approach is more appropriate for this dataset.

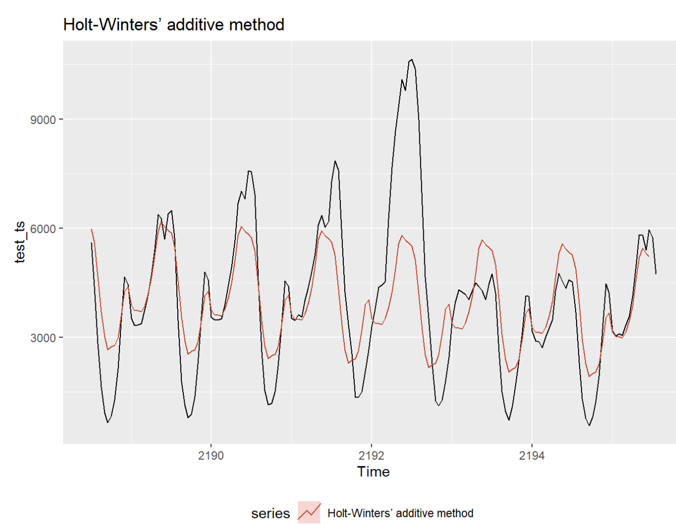


Fig. 16: Holt-Winters Additive Method Forecast vs. Actual Uber Pickups for the Test Dataset.

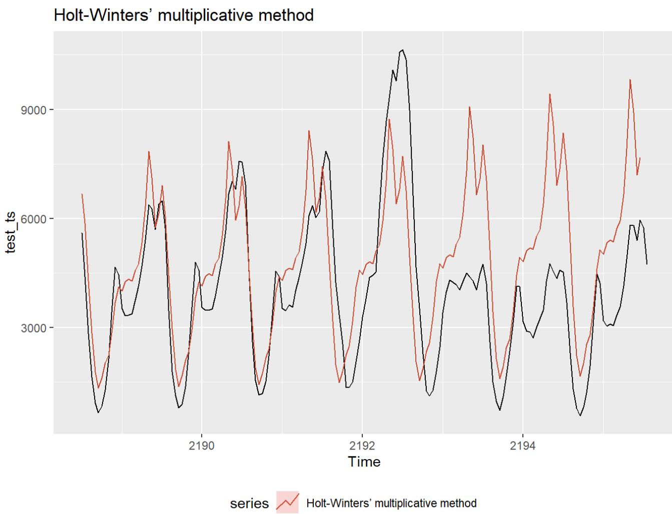


Fig. 17: Holt-Winters Multiplicative Method Forecast vs. Actual Uber Pickups for the Test Dataset.

Discussion: The Holt-Winters additive method emerged as the superior model, producing lower RMSE and MAPE values. This aligns with the observation that seasonality in the dataset is constant rather than proportional to the level of the series. However, residual analysis reveals some remaining patterns, such as significant autocorrelation at lag 1 (ACF1 $= 0.549$), suggesting that further refinements could improve the model's performance.

The results demonstrate that the additive method provides a reliable forecast, making it a suitable choice for predicting Uber pickups. However, addressing residual autocorrelations may further enhance predictive accuracy.

D. Generalized Additive Models (GAMs)

Generalized Additive Models (GAMs) provide a flexible approach for modeling nonlinear relationships by combining smooth functions of predictors. For this analysis, three distinct GAM models were implemented, each with different combinations of smoothing and regression techniques. These models aimed to forecast Uber pickups using temperature, gas price, precipitation, and wind components (`u_wind`, `v_wind`) as predictors.

1) GAM Models Implemented:

- **Splines Model:** This model utilized smoothing splines for all predictors, allowing for flexible, nonlinear relationships. Each predictor (`temperature`, `gas_price`, `total_precipitation`, `u_wind`, and `v_wind`) was modeled using a smooth function (`s()`), enabling the model to capture detailed variations in Uber demand influenced by these variables.
- **Local Regressors Model:** This model applied local regression (`lo()`) to all predictors, with specific spans defined for each variable to control the degree of smoothness. Local regression is particularly useful for capturing fine-scale patterns in the data.
- **Mixed Model:** The mixed approach combined smoothing splines (`s()`) and local regression (`lo()`). Selected predictors, such as `gas_price` and `u_wind`, were modeled with local regression to capture localized effects, while the remaining predictors used smoothing splines for broader trends.

Each model was trained on 96.1% of the dataset, reserving 3.9% for testing. Evaluation metrics included Adjusted R^2, Akaike Information Criterion (AIC), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE), summarized in Table 4.

Model	MSE	AIC	Adjusted R^2
Splines	26,161,129	73,174.4	0.241
Local Regressors	3,571,397	73,262.34	0.213
Mixed	26,633,666	73,177.33	0.240

TABLE 4: Performance Metrics for GAM Models

Model	RMSE	MAPE (%)
Splines	5,114.80	97.42
Local Regressors	1,889.81	63.99
Mixed	5,160.78	98.47

TABLE 5: Additional Metrics (RMSE and MAPE) for GAM Models

2) Model Evaluation:

- **Splines Model:** This model was able to capture general smooth trends in Uber demand but demonstrated limited ability to address finer-scale fluctuations, as reflected in its moderate Adjusted R^2 of 0.241 and relatively high MSE (26,161,129). The RMSE of 5,114.80 and a very high MAPE of 97.42% indicate significant errors in absolute terms and poor relative accuracy. While the model offers insights into broader patterns, its predictive accuracy leaves room for improvement.
- **Local Regressors Model:** Despite achieving the lowest MSE (3,571,397) and RMSE (1,889.81), this model's overall performance, as indicated by an Adjusted R^2 of 0.213, was not significantly better in explaining variance. The MAPE of 63.99% is lower than that of the other models, suggesting better relative accuracy in percentage terms. However, its reliance on `lo()` functions increases computational complexity without delivering substantial gains in overall accuracy. This highlights the potential trade-off between capturing localized effects and maintaining model simplicity and interpretability.
- **Mixed Model:** By combining smoothing splines and local regressors, this model aimed to strike a balance between capturing broad trends and localized patterns. However, its performance metrics, including an Adjusted R^2 of 0.240, a high MSE (26,633,666), and an RMSE of 5,160.78, indicate that the approach did not significantly outperform the simpler models. Furthermore, its MAPE of 98.47% suggests poor relative predictive accuracy.

This indicates that combining techniques in this case did not provide significant benefits and may have introduced additional model complexity without improving results.

3) Visualization of Results: Figure 18 presents the alignment between predicted and actual Uber pickups for the three implemented GAM models.

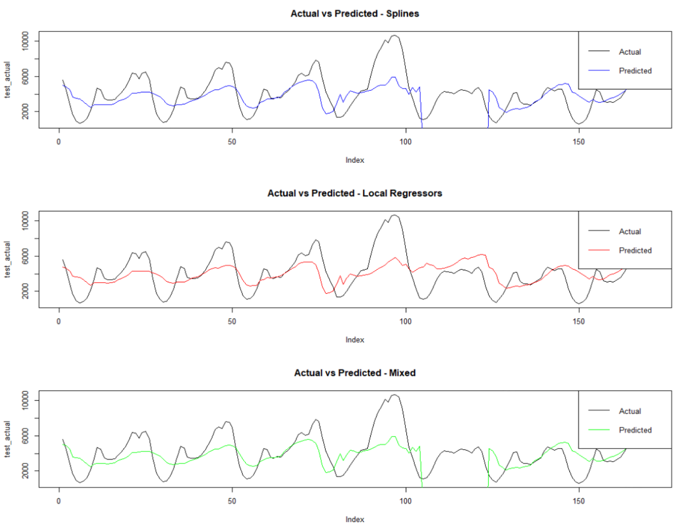


Fig. 18: Actual vs. Predicted Uber Pickups for GAM Models: (a) Splines Model, (b) Local Regressors Model, (c) Mixed Model.

In panel (a), corresponding to the Splines Model, the predictions capture the general trend of the observed data but fail to fully replicate the magnitude of some peaks and troughs, indicating that while broader patterns are modeled effectively, finer-scale details are missed.

Panel (b), showcasing the Local Regressors Model, demonstrates a closer fit to the actual data in terms of both trend and scale, especially in regions with sharp variations. This indicates the model's ability to capture localized effects, although some deviations remain evident in areas of rapid change.

Panel (c), for the Mixed Model, shows a balance between the smoothness of the Splines Model and the localized patterns of the Local Regressors Model. While it performs slightly better than the Splines Model in following observed peaks, it does not match the Local Regressors Model in fully capturing smaller fluctuations.

Overall, the Local Regressors Model achieves the closest alignment to the observed data, but all models display some limitations in replicating the extreme values.

4) Residual Analysis: Residual diagnostics were conducted for the Splines Model, Local Regressors Model, and Mixed Model to evaluate the goodness-of-fit and identify any patterns in the residuals. Each figure includes a time series plot of the residuals, the Autocorrelation Function (ACF), and the Partial Autocorrelation Function (PACF). These visualizations help assess whether the residuals exhibit systematic patterns or significant autocorrelations that might indicate model inadequacies.

These diagnostics indicate that the residuals of all three models exhibit significant autocorrelation, implying that the models are not fitting the data effectively and fail to capture important patterns and dependencies within the Uber pickup data. Further refinement or alternative modeling approaches may be necessary to improve performance

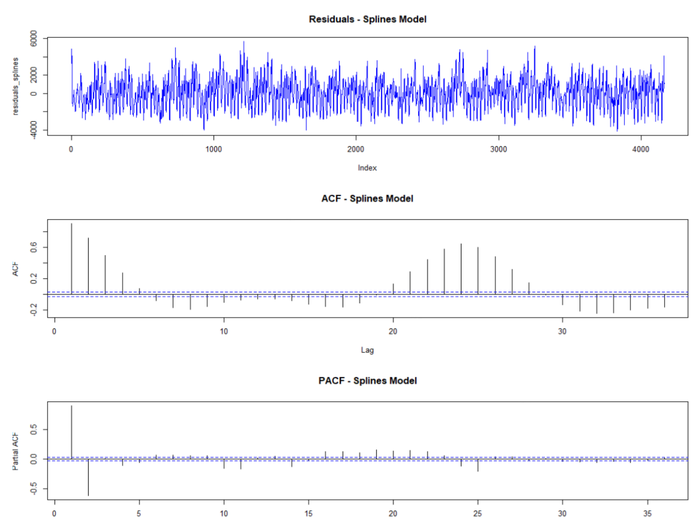


Fig. 19: Residual Diagnostics for the Splines Model: The time series plot reveals consistent residual variability over time. However, the ACF plot shows significant autocorrelation in the lower lags, and the PACF plot highlights notable partial autocorrelations, suggesting the model fails to capture all underlying patterns in the data.

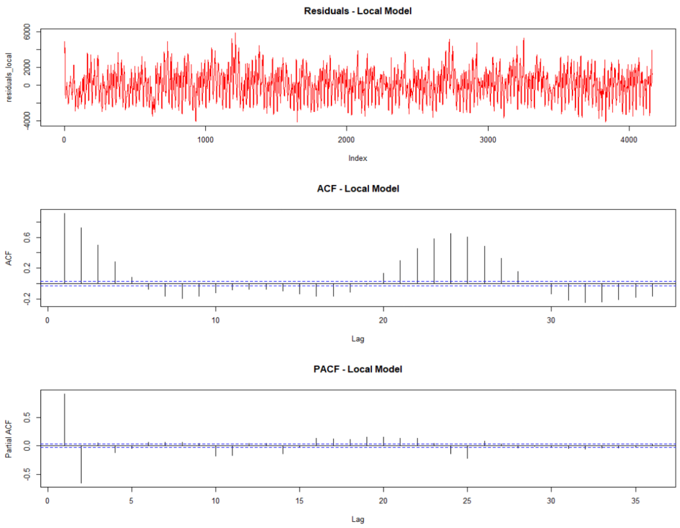


Fig. 20: Residual Diagnostics for the Local Regressors Model: The time series plot suggests residuals with higher variability. The ACF plot shows persistent autocorrelations across multiple lags, and the PACF plot confirms significant partial autocorrelation, indicating that the model does not adequately fit the data or capture all relevant patterns.

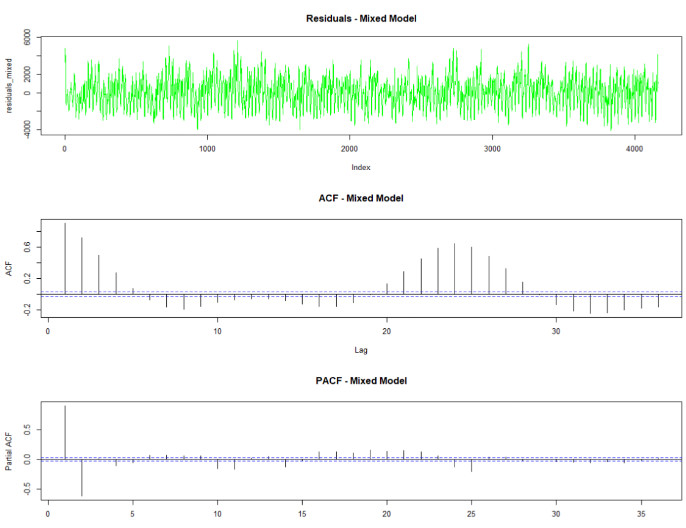


Fig. 21: Residual Diagnostics for the Mixed Model: The time series plot shows residuals with similar variability to the Splines Model. The ACF plot indicates noticeable autocorrelations, and the PACF plot reveals persistent partial autocorrelations, suggesting the model fails to fully account for patterns in the data.

These diagnostics indicate that the residuals of all three models exhibit significant autocorrelation, implying that the models are not fitting the data effectively and fail to capture important patterns and dependencies within the Uber pickup data. Further refinement or alternative modeling approaches may be necessary to improve performance.

5) Conclusion: The analysis of Generalized Additive Models (GAMs) highlights significant limitations in the models' ability to fit the Uber pickup data. While the Splines, Local Regressors, and Mixed Models offer different approaches to balancing flexibility and interpretability, residual diagnostics reveal that none of the models adequately capture the underlying patterns in the data.

The persistent autocorrelation observed in the residuals for all three models indicates that key dependencies and structures in the data remain unmodeled. This suggests that the chosen predictors and smoothing techniques, while capable of capturing general trends, fail to account for finer-scale variations and interactions between predictors. Moreover, the relatively high Mean Squared Error (MSE) values and moderate Adjusted R^2 scores further underline the models' limited predictive accuracy.

These findings underscore the need for alternative modeling approaches or additional predictors to better capture the complexity of Uber demand. Possible next steps could include:

- Incorporating additional relevant predictors, such as traffic conditions, special events, or socioeconomic variables, to better explain variations in Uber demand.
- Exploring alternative modeling techniques, such as machine learning models (e.g., Random Forests, Gradient Boosting, or Neural Networks), which may be better suited for capturing complex, nonlinear interactions and patterns in the data.
- Considering temporal models, such as ARIMA or state-space models, to explicitly address the autocorrelation observed in the residuals.

In conclusion, while GAMs offer valuable insights into the relationships between predictors and Uber pickups, their current implementation falls short of providing an accurate and reliable forecasting tool. Further refinement and experimentation with alternative approaches are essential for improving model performance and capturing the complexity of Uber demand dynamics.

E. Gradient Boosting Analysis

Gradient Boosting (GB) is an ensemble machine learning method that builds models iteratively by combining weak learners, typically shallow decision trees, to minimize prediction errors. This technique is particularly effective for modeling complex relationships and is well-suited for forecasting Uber demand, where external factors such as weather conditions, gas prices, and wind speeds play a significant role.

1) Methodology: Three Gradient Boosting configurations were implemented to model the relationship between Uber pickups and predictors such as temperature, gas price, total precipitation, and wind components. The configurations differ in their hyperparameter settings:

- **Model 1 (Default Parameters):** Used shallow trees with a depth of 1 and 5000 boosting iterations to establish a baseline.

Model	Minimum Test Error	Iteration at Minimum Error	Overfitting Observed
Model 1 (Default)	2,901,429	2,394	No
Model 2 (Deeper Trees)	3,054,960	94	Yes
Model 3 (Smaller Learning Rate)	3,058,184	3,138	Minimal

TABLE 6: Gradient Boosting models' test errors and behavior analysis.

- **Model 2 (Deeper Trees):** Increased tree depth to 4 to better capture complex interactions at the cost of higher variance.
- **Model 3 (Smaller Learning Rate):** Applied a learning rate of 0.01, requiring more iterations (up to 5000) to achieve convergence, improving the model's precision and stability.

2) Results and Analysis: The performance of the Gradient Boosting models was evaluated by comparing train and test errors over the iterations. Figures 22, 23, and 24 illustrate the train vs. test errors for each model configuration.

- **Model 1 (Default):** Achieved the minimum test error of **2,901,429** at iteration **2394**. The shallow depth of trees limited the model's capacity, capturing only simple patterns in the data.
- **Model 2 (Deeper Trees):** Reached a minimum test error of **3,054,960** at iteration **94**. While the deeper trees enabled the model to better fit complex interactions, overfitting was observed at higher iterations as the test error began to increase.
- **Model 3 (Smaller Learning Rate):** Obtained a minimum test error of **3,058,184** at iteration **3138**. The smaller learning rate allowed the model to converge gradually, ensuring stable performance even at higher iterations.

Figures 25, 26, and 27 present the relative importance of variables for each model configuration. These plots highlight the contributions of each feature to the predictions made by the model:

- **Model 1 (Default Parameters):** Temperature and v_wind were the most significant predictors, followed by total_precipitation and u_wind.
- **Model 2 (Deeper Trees):** Total_precipitation gained more importance, dominating the feature contributions, with temperature and wind components also playing significant roles.
- **Model 3 (Smaller Learning Rate):** Total_precipitation and temperature were the most critical predictors, with stable contributions from u_wind and v_wind.

3) Discussion: Figure 22 (Model 1) demonstrates a slower reduction in training error due to shallow trees, limiting the model's ability to capture complex relationships. In contrast, Figures 23 (Model 2) and 24 (Model 3) illustrate faster convergence, but with differences in overfitting behavior. Model 2 overfits rapidly due to its deeper trees, while Model 3 shows a more gradual increase in test error, reflecting the effectiveness of a smaller learning rate.

4) Conclusion: Gradient Boosting provided valuable insights into Uber demand forecasting:

- Shallow trees (Model 1) captured simple relationships effectively, but lacked capacity for complex patterns.
- Deeper trees (Model 2) performed better initially but were prone to overfitting at later iterations.
- Smaller learning rates (Model 3) offered the most stable results, balancing precision and robustness.

The results suggest that while Gradient Boosting is a powerful tool for forecasting, careful tuning of hyperparameters such as tree depth, learning rate, and the number of iterations is essential to achieve optimal performance. The performance of the Gradient Boosting models is summarized in Table 6.

Fig. 22: Train vs Test Error for Model 1 (Default Parameters). Minimum test error at iteration 2394.

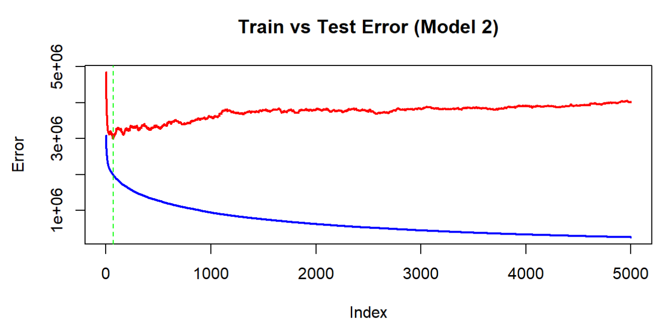


Fig. 23: Train vs Test Error for Model 2 (Deeper Trees). Minimum test error at iteration 94.

IV. Results

The performance of different models for forecasting Uber demand in New York City is summarized in this section.

TABLE 7: Summary of Model Performance

Model	RMSE	MAPE (%)	Adjusted R^2	AIC/BIC
TSLM (Daily)	1465	38.43143	0.45	59951/60116
TSLM (Weekly)	989	23.00659	0.60	58755/59831
ARIMA(4,0,2)(2,1,0)[24]	394.83	10.80	-	61300.25/61357.21
Holt-Winters Additive	533.09	17.42	-	-
Holt-Winters Multiplicative	893.42	27.67	-	-
GAM (Splines)	-	-	0.241	73174.4
GAM (Local Regressors)	-	-	0.213	73262.34
GAM (Mixed)	-	-	0.240	73177.33
Gradient Boosting (Shallow Trees)	1699.337	58.23	-	-
Gradient Boosting (Deeper Trees)	1713.111	48.9	-	-
Gradient Boosting (Small Learning Rate)	1751.533	56.11	-	-

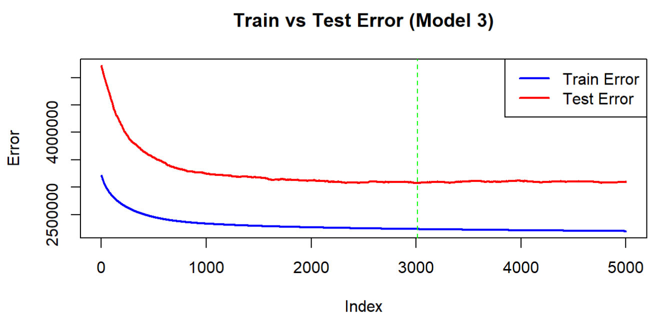


Fig. 24: Train vs Test Error for Model 3 (Smaller Learning Rate). Minimum test error at iteration 3138.

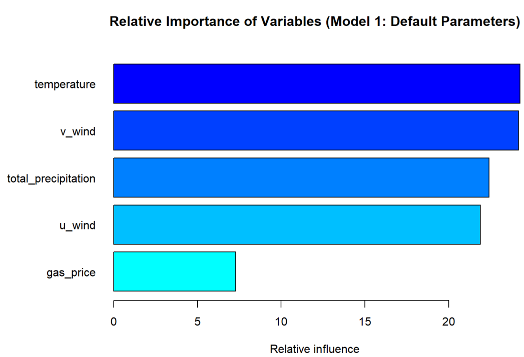


Fig. 25: Relative Importance of Variables for Model 1 (Default Parameters).

Each model demonstrated varying levels of accuracy and capabilities in capturing trends, seasonality, and external dependencies.

Table 7 provides an overview of the results obtained from each model, including key performance metrics such as RMSE, MAPE, Adjusted R^2, and AIC/BIC where applicable.

The following observations were noted for each model based on the results:

- **TSLM (Daily and Weekly):** These models effectively captured trend and seasonality. The weekly model demonstrated higher accuracy (RMSE 989) and Adjusted R^2 but showed residual autocorrelation, indicating room for further refinement.

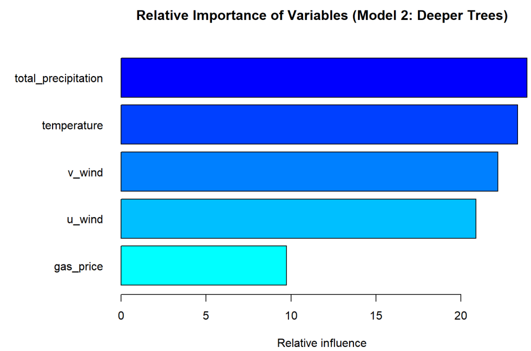


Fig. 26: Relative Importance of Variables for Model 2 (Deeper Trees).

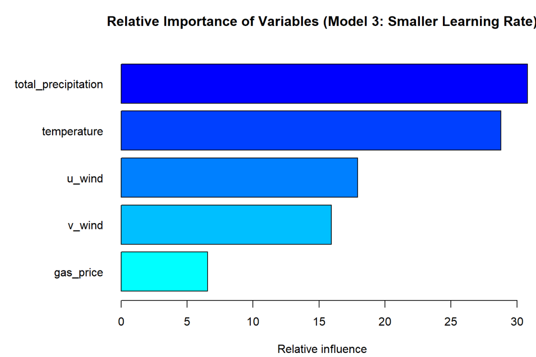


Fig. 27: Relative Importance of Variables for Model 3 (Smaller Learning Rate).

- **ARIMA(4,0,2)(2,1,0)[24]:** This model achieved the best accuracy among statistical methods with an RMSE of 394.83 and a MAPE of 10.80%. However, residual autocorrelations persisted, suggesting that additional external predictors or higher-order terms might be necessary.

- **Holt-Winters Models:** The additive variant aligned well with the assumption of constant seasonality (RMSE 533.09, MAPE 17.42%), while the multiplicative version underperformed (RMSE 893.42, MAPE 27.67%), likely due to inappropriate proportional seasonality assumptions.

- **GAM Models:** The GAM with splines captured smooth trends effectively but struggled with finer-scale details. The local regressor variant introduced better localized effects but was prone to overfitting. The mixed approach balanced these issues but did not outperform the simpler models.

- **Gradient Boosting:** The inclusion of RMSE and MAPE metrics revealed that while Gradient Boosting models demonstrated the potential to capture complex patterns, they often struggled with interpretability and stability. The shallow-tree variant had an RMSE of 1699.34 and MAPE of 58.23%, reflecting its simplicity. Deeper trees led to marginally worse performance (RMSE 1713.11, MAPE 48.9%), likely due to overfitting. The small learning rate variant had an RMSE of 1751.53 and MAPE of 56.11%, achieving slightly more stability but without significant accuracy improvements.

Overall, while simpler models offered interpretability and faster computation, more complex methods such as ARIMA and Gradient Boosting achieved greater accuracy at the cost of additional computational complexity and susceptibility to overfitting. Residual analysis across all models highlights opportunities for incorporating additional features or exploring advanced machine learning architectures to address observed limitations.

V. Conclusion

In conclusion, this study successfully demonstrated the application of various statistical and machine learning models for forecasting Uber demand in New York City. Among the models evaluated, ARIMA achieved the highest predictive accuracy, highlighting its suitability for capturing temporal patterns and seasonality in hourly ridesharing demand. While the Holt-Winters additive model and generalized additive models (GAMs) offered valuable insights into trends and exogenous factors, they faced challenges in addressing residual autocorrelations and finer-scale fluctuations. Gradient Boosting, with careful tuning, showed promise for incorporating complex, non-linear relationships but required balance to avoid overfitting. These findings underscore the critical role of both temporal dependencies and external predictors, such as weather and economic factors, in achieving robust demand forecasts. Future efforts should explore hybrid models that combine the interpretability of traditional methods with the flexibility of machine learning techniques, alongside incorporating additional predictors to further refine forecasting accuracy and utility for operational decision-making in urban transportation systems.

References