NYC YELLOW TAXI DEMAND PREDICTION & ANALYSIS

DSO 522 Final Report

Abstract

In this project, we embark on the task of forecasting the daily demand for Yellow Taxis in the dynamic urban landscape of New York City (NYC). The forecasting analysis involves a strategic blend of classical and advanced forecasting techniques. Our methodology integrates these techniques with a wealth of historical taxi trip data, coupled with other important features, to comprehensively predict taxi demand at diverse time intervals and locations throughout NYC. The holistic analysis seeks to discover intricate patterns, decipher seasonality, and identify trends embedded within the taxi usage data, with a keen focus on understanding the multifaceted factors that influence demand variations. Beyond the technical intricacies, the project underscores a practical application of forecasting analysis aimed at elevating the operational efficiency of taxi services in NYC. By peeling back layers of data, we aspire to contribute actionable insights to inform and optimize urban transportation planning. This venture aligns with a broader goal of not only enhancing the precision of demand predictions but also fostering a nuanced understanding of the complex interplay between diverse variables affecting taxi services in one of the world's most dynamic and densely populated cities. The synthesis of historical data, forecasting techniques, and urban planning considerations positions this project at the intersection of data science and practical solutions for optimizing taxi services and advancing the broader discourse on sustainable and efficient urban transportation in NYC.

Project Introduction

At its core, this project delves into the iconic yellow taxi industry of New York City, a symbol deeply ingrained in the fabric of the city's identity. Beyond merely serving as a means of transportation, these yellow taxis embody a fundamental element of the urban experience, meeting the diverse and ever-changing mobility needs of the city's inhabitants. Recognizing the integral role these taxis play in shaping the urban landscape, this initiative aims to confront challenges related to efficiency, accessibility, and responsiveness within the yellow taxi system.

In acknowledging the dynamic nature of New York City's transportation demands, the project adopts a forward-thinking approach by leveraging data-driven insights and advanced analytics. By doing so, it aspires to enhance the adaptability of the yellow taxi industry, ensuring its ability to seamlessly navigate and respond to the intricacies of the city's evolving transport landscape. Through the fusion of cutting-edge technology and empirical analysis, the overarching objective is to contribute meaningfully to the ongoing conversation about sustainable and efficient transportation solutions within one of the world's most vibrant and dynamic cities.

In essence, this project not only recognizes the yellow taxi industry as a vital component of New York City's identity but also endeavors to propel it into the future by addressing contemporary challenges. By embracing a holistic perspective that intertwines technological innovation with empirical understanding, the aim is to foster a transportation ecosystem that not only meets the immediate needs of the city's residents but also aligns with the broader goals of sustainability and efficiency in urban mobility. This project, situated at the nexus of tradition and progress, seeks to play a pivotal role in shaping the trajectory of the yellow taxi industry in the ever-evolving landscape of New York City.

Motivation and Goals

In our comprehensive project, we center our efforts on the vital task of forecasting the daily demand for Yellow Taxis in New York City over the next 5-6 months. Our primary objectives revolve around uncovering the subtle trends, seasonality, and patterns inherent in taxi demand, particularly during peak hours. The driving force behind this endeavor is rooted in the motivations to implement dynamic pricing strategies, optimize driver schedules, enhance traffic management, and ultimately reduce waiting times. Through a multifaceted approach utilizing classical decomposition, smoothing techniques, Holt Winters, Seasonal ARIMA, LSTMs, and FB Prophet, our goal is to extract valuable insights into the intricate transportation dynamics of the city. The overarching aim is not only to bolster the operational efficiency of taxi services but also to contribute to the broader optimization of New York City's urban transportation system. The real-time applications of taxi demand forecasting extend far beyond immediate operational benefits. They empower service providers to allocate resources judiciously, implement responsive pricing strategies, and optimize driver schedules. Moreover, this approach supports effective traffic management, curtails instances of empty cruising, and promotes environmental sustainability. For passengers, the outcomes are tangible – reduced waiting times and an overall improvement in convenience. Our project is poised to address pivotal questions related to taxi service and passenger behavior, seeking to determine peak daily demand variations and explore potential day-to-day fluctuations. Additionally, we aim to identify the busiest time, day and meticulously track demand trends within this timeframe, providing a comprehensive understanding of the dynamic interplay between taxi services and the city's transportation landscape. The NYC taxi time series analysis offers vital insights for strategic decision-making. This includes identifying peak hours for potential surge pricing, which informs resource allocation and pricing strategies. Demand forecasting ensures efficient driver and vehicle deployment, upholding service quality.

Additionally, operational efficiency can be improved by identifying and addressing inefficiencies through responsive measures.

Data source and variables

The dataset employed in this project is sourced directly from the official website of the City of New York and has been made available through authorized technology providers participating in the Taxicab & Livery Passenger Enhancement Programs, in collaboration with the NYC Taxi and Limousine Commission (TLC). This robust dataset serves as the cornerstone of our analysis, providing comprehensive insights into the operational dynamics of the city's taxi services.

Our key variables encompass essential aspects of taxi journeys, including the total fare amount, pickup, and drop-off times, the number of rides, trip distance, geographical locations, and passenger count. Recognizing the need for a nuanced understanding of the intricacies within the dataset, we have undertaken feature engineering. This process involves creating additional relevant features that prove instrumental in conducting a thorough analysis and deriving key performance indicators.

Among the crafted features are variables such as trip duration, pickup, and drop-off hours, as well as pickup and drop-off days and months. These augmented features serve as crucial elements in unraveling deeper patterns and trends within the taxi trip data. Trip duration, for instance, provides insights into the temporal aspect of each journey, while pickup and drop-off hours allow for a granular examination of time-specific patterns. Similarly, pickup and drop-off days and months offer a broader temporal context, enabling a comprehensive exploration of seasonal and day-of-week variations in taxi demand and usage.

By incorporating these meticulously engineered features, our analysis aims not only to paint a detailed picture of past taxi usage but also to lay the foundation for accurate forecasting and insightful trend analysis. The richness of the dataset, coupled with these refined features, positions this project to offer a nuanced and comprehensive understanding of the Yellow Taxi dynamics in New York City, contributing to the ongoing efforts to optimize urban transportation and enhance the overall efficiency of taxi services in this bustling metropolis.

Exploratory Data Analysis

In this section, we embark on a detailed examination of the Yellow Taxi Demand dataset, aiming to uncover actionable insights that can inform strategic decisions for the taxi service. Our goal is to uncover patterns and trends in taxi demand from 2021 to August 2023. Starting with a close look at daily and monthly demand time series, we aim to identify key temporal patterns, helping us understand when and how demand fluctuates over the 2-to-3-year period. We will explore variations in demand across different days of the week to inform scheduling decisions and delve into hourly demand patterns, linking them with average fare quotes to gain insights into pricing dynamics. Beyond this, we will analyze traffic patterns by hour to pinpoint rush hours and assess revenue metrics like revenue per minute and revenue per mile to gauge the company's financial performance during specific hours. Finally, we will break down demand patterns based on workdays and weekends, offering a nuanced view that can guide optimized scheduling strategies. This EDA is a crucial step in our journey toward providing practical, data-driven recommendations to enhance the efficiency and profitability of the yellow taxi service.

1. Analysis of Total Daily and Monthly Taxi Demand (*Figure 1 Appendix*)

Our initial investigation into the Yellow Taxi Demand dataset, covering the period from January 2021 to August 31, 2023, centered around the comprehensive analysis of total daily and monthly time series demand. A striking trend becomes apparent in the daily demand data, revealing a notable uptick from January 2021 through the end of that year. This surge aligns with the broader economic reopening post-COVID-19 lockdown, suggesting a correlation between economic activities and increased demand for taxi services in NYC during this period. Furthermore, our scrutiny of the daily taxi demand time series unveils distinct seasonality patterns, characterized by a consistent cycle approximately every seven days—highlighting a weekly demand rhythm. Post-COVID recovery, the demand levels out, indicating a restored equilibrium in taxi demand. However, amid this stabilization, we observe considerable noise contributing to random fluctuations in both daily and monthly demand time series. These fluctuations add a layer of complexity to the analysis, prompting a deeper dive to identify and understand the contributing factors behind this variability. This analysis sets the stage for a more detailed exploration, where we aim to dissect the nuances of demand patterns, considering both the economic landscape and potential external influences, to better inform strategic decision-making within the taxi service sector.

2. Analysis of Daily Ride Distribution Across Days of the Week (*Figure 2 Appendix*)

Our investigation extended to understanding the distribution of the average number of daily rides aggregated by each day of the week. The utilization of a boxplot allowed us to discern patterns in demand variations throughout different days of the week. Notably, the data reveals that the average daily rides in 2021 were comparatively lower than those in subsequent years (2022 and 2023), indicating a recovery in demand post-pandemic. In specific terms, the average demand for each day of the week in 2023 appears lower than that of 2022. It's essential to note that the 2023 dataset concludes in August, potentially influencing this discrepancy. Despite these variations, consistent patterns emerge across all years. Wednesdays, Thursdays, and Fridays consistently exhibit the highest average daily demand, aligning with the heightened activity on these mid-week workdays. These findings reflect the bustling nature of these weekdays, as individuals tend to commute to offices, contributing to increased taxi demand. Conversely, lower ride demand is observed on Mondays and Sundays. This trend suggests a collective preference for relaxation and preparation at home on Sundays, anticipating the workweek ahead. On Mondays, a lower demand could be attributed to a decreased inclination to commute to offices, with individuals choosing to work remotely or start their workweek from the comfort of their homes. These insights into daily demand fluctuations contribute valuable information for optimizing resource allocation and scheduling, aligning taxi services with the nuanced preferences and behaviors of riders throughout the week.

3. Hourly Demand and Pricing Analysis (*Figure 3 Appendix*)

Our third analytical focus centered on dissecting the demand patterns by hour of the day, further categorized into four ride duration groups: short rides (2-5 mins), average rides (5-15 mins), above-average rides (15-30 mins), and longer rides beyond 30 mins. Simultaneously, we examined the average fare amounts per hour to discern potential correlations between demand and pricing dynamics. Our findings reveal that the majority of rides fall within the 5 to 30-minute duration range, with very short or very long rides constituting only around 20% of the total. The demand landscape exhibits distinctive peaks and troughs, with a sharp rise at 6 AM, reaching its zenith around 6 PM, and gradually tapering off to its lowest point at 4 AM. These patterns not only mirror the daily travel habits of New Yorkers but also highlight high-demand hours. Contrary to the conventional understanding of high demand correlating with high prices, we uncovered intriguing nuances in the relationship between demand and average fare amounts. In the early morning hours with lower demand, the average taxi fare is paradoxically the highest. This apparent discrepancy is rationalized by the fact that although demand is lower, the fewer rides taken during these hours are typically long-distance and more expensive. Additionally, the reduced number of taxi drivers operating during nighttime contributes to a lower supply, further influencing higher average fare prices. As the day progresses toward its peak demand hours, there is a subtle increase in the average fare amount. This nuanced fluctuation underscores the complex interplay between demand, supply, and pricing, providing insights into how these factors collectively shape the economic dynamics of the taxi service industry throughout the day.

4. Revenue Generation Metrics and Traffic Dynamics (*Figure 4 Appendix*)

In our continued exploration, we delved into key revenue generation metrics—average fare per minute and average fare per mile—aiming to comprehensively understand the taxi's performance at each hour of the day. Simultaneously, we examined average trip speed as a proxy for traffic conditions, shedding light on how traffic dynamics may impact operational efficiency. Our findings uncovered a noteworthy phenomenon: traffic conditions exhibit a significant variation throughout the day. During nighttime, particularly in the early morning hours around 4 and 5 AM, we observed the lightest traffic, with an impressive average trip speed of 20 MPH. Conversely, during the daytime from 9 AM to 6 PM, the speed diminishes substantially to approximately 10 MPH, indicative of heavy traffic congestion. This variation in traffic conditions plays a pivotal role in shaping revenue metrics. Surprisingly, our analysis revealed that traffic is more favorable during nighttime, allowing drivers to travel significantly faster. This observation is crucial, as it directly influences revenue metrics. Specifically, a night ride is approximately 23% cheaper, and drivers have the potential to earn between 25-40% more per minute during the early morning hours when traffic is light, compared to the heavy traffic hours during the day. This insight serves as a compelling incentive for riders to opt for longer trips during early morning hours, potentially contributing to a strategic boost in demand during these periods. Understanding these variations in traffic dynamics and their impact on revenue metrics provides a nuanced perspective for the company to consider when optimizing pricing strategies and incentivizing ridership during specific hours, ultimately contributing to more informed decision-making in the competitive taxi service landscape.

5. Hourly Demand Distribution Pattern Among Workdays and Weekends (*Figure 5&6 Appendix*)

In our final analysis, we examined the distribution of average taxi ride demand by hour, further dissecting the data across individual days of the week to observe variations in hourly demand patterns. Notably, the hourly demand on workdays follows a consistent trend, characterized by a minimal nighttime demand. A distinct pattern emerges on workdays, where demand experiences a rapid surge between 5 to 8 AM, reaching its zenith at 6 PM. Subsequently, demand diminishes swiftly into the night, indicating a widespread trend of individuals returning home to rest for the upcoming workday. However, on Fridays, evening-hour demand remains notably higher than other workdays, suggesting increased socializing activities after the workweek.

During weekends, the nighttime demand surpasses workdays, particularly on Saturday nights, showcasing heightened nighttime recreational activities in the city. Nevertheless, weekend daytime demand generally lags behind weekdays, with peak demand exhibiting a more gradual incline, indicative of a sustained demand level during daytime hours, contrasting the sharper peaks observed on workdays. This insight into variations in hourly demand, both across workdays and weekends, offers a valuable perspective for simulating and optimizing operational strategies and resource allocation, tailoring services to align with the distinct behavioral patterns of riders throughout the week.

Time Series Modeling

In this section, our primary objective is to carefully assess and compare a range of time series forecasting models as they are applied to the yellow taxi daily demand dataset. The dataset under consideration encloses the time frame from January 2021 to July 2023 for training purposes, while the validation set extends to cover the month of August 2023. Also, the forecasting horizon is considered to be the rest of year. The report delves into a diverse array of forecasting methodologies, including the Classical Decomposition Method, Seasonal Naïve Analysis, various Smoothing Techniques, the Holt Winter's Model, the Seasonal ARIMA model, and the FB Prophet Model. By examining these distinct approaches, we aim to provide a comprehensive understanding of their respective strengths, weaknesses, and overall performance in predicting the daily demand patterns within the context of yellow taxi services. This evaluation serves as a valuable resource for making informed decisions regarding the selection and application of time series forecasting models in similar transportation and demand prediction scenarios.

Classical Decomposition:

The Classical Decomposition Method is a widely used technique in time series forecasting, designed to separate a time series into its fundamental components: trend, seasonal, and residual components. This method offers distinct advantages, notably in terms of interpretability, as it produces components that are easily understandable, aiding in the comprehension of underlying patterns within the time series. Moreover, its simplicity in implementation makes it accessible to a broad audience, making it an attractive option for those without advanced statistical expertise. It is applicable to time series data exhibiting clear trends and seasonal patterns.

The decomposition process involves three main components: trend (Tt), seasonal (St), and residual (Rt). The trend component is typically derived using a moving average or smoothing technique, representing the long-term movement or direction of the time series. The seasonal component captures regular patterns that repeat over a specific period, such as a year. The residual component accounts for unexplained variability after removing the trend and seasonal components. The reconstruction of the original time series involves combining these components ($Y_t=T_t+S_t+R_t$), providing a comprehensive framework for understanding and modeling time series data with identifiable trends and seasonality.

Confining to the business case, the model has been considered as a baseline performance model. This is because, the Classical Decomposition Method also comes with certain limitations. One major assumption is the additivity of the components, which may not always hold in real-world scenarios. Additionally, the method is sensitive to outliers in the data, potentially impacting the accuracy of the decomposition. Its limited flexibility might pose challenges in capturing intricate relationships or irregular patterns within the time series.

Within classical decomposition, both additive and multiplicative models were employed for this business case. The additive model yielded an RMSE of 30,126.24 and MAPE of 31.63, while the multiplicative model produced higher errors with an RMSE of 91,387.14 and MAPE of 99.98.

Seasonal Naive:

Employing the Seasonal Naive Model for determining daily taxi demand in NYC presents several advantages. Its simplicity and ease of implementation make it a practical choice for quick analyses, requiring minimal computational resources. The absence of a training period is another notable advantage, making it suitable for scenarios where rapid predictions are needed. Additionally, the model excels at capturing seasonal patterns, a crucial aspect for industries like taxi services, where demand can vary significantly based on factors such as weather, holidays, and events.

However, the Seasonal Naive Model comes with inherent limitations. Its reliance on the assumption that the current season will resemble the same season in the previous year may lead to reduced accuracy, especially when facing significant changes in underlying patterns or unforeseen events. The model's inability to adapt to shifts in trends and its lack of consideration for external variables, such as economic factors or regulatory changes, further restrict its applicability. While effective in certain contexts, the Seasonal Naive Model may prove less suitable for dynamic environments where demand patterns are subject to short-term fluctuations or structural changes.

The model produced an RMSE of 15,366.06 and a MAPE of 14.99, showcasing its limitations in accurately capturing the nuances of the yellow taxi daily demand time series. Overall, Seasonal Naive serves as a basic and quick forecasting tool, but its simplicity may compromise accuracy, especially in scenarios with intricate temporal patterns.

Rolling Forward Smoothing Technique (Window Size 3):

This technique facilitates the smoothing of variability in the data, diminishing short-term fluctuations and noise. It is particularly beneficial for identifying underlying trends in taxi demand by reducing the impact of random spikes, providing a clearer picture of the evolving patterns. Moreover, the adaptability of the technique to gradual changes in demand trends makes it suitable for scenarios where a consistent evolution in taxi demand is observed. Its ease of interpretation further enhances its accessibility, allowing for quick analysis and decision-making.

However, the Rolling Forward Smoothing Technique also comes with inherent limitations. A significant drawback is the lag in response to sudden changes, as the averaging process over a fixed window size may not immediately reflect abrupt shifts in demand. The sensitivity of the technique to the chosen window size introduces another challenge, with different sizes yielding varied results and impacting the model's responsiveness. Additionally, the technique lacks explicit consideration for capturing seasonal variations, a crucial factor in industries like taxi services where demand is influenced by recurring events. Furthermore, its sensitivity to outliers can distort predictions, making it important to carefully select and assess the appropriate window size based on the specific characteristics of the taxi demand data. Despite these drawbacks, the model demonstrated promising performance with an MSE of 9,262.07 and a MAPE of 9.45, showcasing its effectiveness in capturing the underlying patterns in the taxi daily demand data.

Holt-Winters Model:

Notably, the method's capacity to incorporate both trend and seasonality makes it well-suited for scenarios where these factors significantly influence the data, as is often the case in the taxi service industry. Its adaptability to changes in underlying patterns and flexibility in predicting various horizons further enhance its utility. By giving more weight to recent observations, the technique proves responsive to dynamic shifts in demand, providing a robust tool for forecasting both short-term fluctuations and long-term trends.

There are some disadvantages to this model that makes us think twice before implementation. The complexity of implementation, particularly in comparison to simpler models, may pose a barrier for users without a strong statistical background. The method's sensitivity to initial conditions adds another layer of consideration during model initialization, demanding careful handling. Interpreting the parameters of the model, such as smoothing constants, can be intricate, requiring a deeper understanding of the underlying mathematical concepts. Additionally, the assumption of stationarity may limit the model's applicability in situations where the statistical properties of the data are not constant over time. As such, the adoption of Holt-Winters Exponential Smoothing should be guided by a nuanced understanding of the forecasting task's requirements and the user's proficiency in navigating more sophisticated models. However, even the Holt-Winters Model demonstrated competitive performance, yielding an RMSE of 9,746.41 and a MAPE of 8.45. Its ability to capture both trend and seasonality makes it a valuable choice for forecasting yellow taxi daily demand, with the trade-off of sensitivity to initial values.

Seasonal ARIMA:

Seasonal ARIMA, an extension of the traditional ARIMA model, addresses the complexities of time series forecasting by incorporating both non-seasonal and seasonal components. Represented as ARIMA(p, d, q)(P, D, Q)s, the model parameters (p, d, q) handle the non-seasonal component, while (P, D, Q, s) govern the seasonal aspects. The determination of these parameters involves a meticulous process known as model identification, where autocorrelation and partial autocorrelation functions are analyzed to guide the selection of suitable values for p, q, P, and Q. The integration orders (d and D) are determined based on assessments of the stationarity of the time series through differencing. This method provides a robust framework for capturing and modeling both non-seasonal and seasonal trends.

Seasonal ARIMA offers several advantages. First, its ability to decompose time series data allows for a comprehensive understanding of underlying patterns, making it suitable for datasets with intricate structures. Second, the model's flexibility enables it to handle both non-seasonal and seasonal trends, making it adaptable to a wide range of time series data. Additionally, Seasonal ARIMA is particularly adept at providing accurate predictions for time series with distinct seasonal variations, contributing to its effectiveness in forecasting applications.

However, the model comes with certain challenges. The complexity in selecting appropriate parameter values (p, d, q, P, D, Q) can be a daunting task, demanding expertise in time series analysis. Sensitivity to outliers and the assumption of stationarity are potential pitfalls and addressing these issues may require additional preprocessing steps. Moreover, the computational intensity of estimating parameters and making predictions with Seasonal ARIMA can be a consideration, especially for large datasets. Despite these challenges, Seasonal ARIMA stands as a powerful tool in time series forecasting, offering a comprehensive approach to capturing the nuanced patterns inherent in real-world temporal data.

In this specific application to yellow taxi daily demand forecasting, surprisingly Seasonal ARIMA exhibited a Mean Squared Error (MSE) of 13,548.78 and a Mean Absolute Percentage Error (MAPE) of 12.75. The identified trend and seasonal orders, (1,0,1) and (0,1,0) respectively, were deduced from the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) graphs and hyper parameter tuning.

FB Prophet:

The Facebook Prophet model is a robust time series forecasting tool developed by Facebook's Core Data Science team. It excels in predicting daily demand for services like NYC taxi rides by incorporating key components such as seasonality, holidays, and trend. The model's versatility lies in its ability to handle missing data, outliers, and large datasets efficiently. With a user-friendly interface and minimal hyperparameter tuning requirements, Prophet is accessible to users with varying levels of expertise in time series forecasting. Its interpretability allows users to understand the impact of different factors on the forecast, including trend, seasonality, and holidays.

One notable advantage of Prophet is its scalability, making it well-suited for applications with substantial historical data, such as daily NYC taxi demand. The model provides uncertainty intervals for forecasts, aiding in risk assessment, and automatically detects and incorporates holidays to capture unusual demand patterns. Additionally, users can introduce custom seasonality components to address unique patterns in the data. The simplicity and interpretability of Prophet, along with features like automatic holiday detection and uncertainty intervals, contribute to its popularity for predicting NYC taxi daily demand, providing valuable insights for decision-making.

In practice, using Prophet for NYC taxi daily demand involves training the model with historical data on daily taxi rides and relevant holiday/event information. The resulting forecast allows users to visualize trends, seasonality, and predicted values, offering a comprehensive understanding of expected daily demand patterns. Overall, Facebook Prophet stands out as a user-friendly and powerful tool, making it a preferred choice for time series forecasting tasks, particularly in applications like predicting the daily demand for NYC taxis. With an RMSE of 4,987.14 and a MAPE of 4.14, Prophet significantly outperformed other models in terms of accuracy. Its ability to capture both trend and seasonality, coupled with the automatic incorporation of seasonal patterns and holidays, makes Facebook Prophet a compelling choice for this forecasting task.

Results

In the context of predicting the daily demand for NYC yellow taxis, our comprehensive analysis of various time series forecasting models has identified the Facebook Prophet model as the optimal choice. The Prophet model exhibited superior accuracy metrics, boasting a remarkably low Root Mean Squared Error (RMSE) of 4987.14 and a Mean Absolute Percentage Error (MAPE) of 4.14. These impressive results reinforce our decision to adopt the Prophet model for its exceptional predictive performance.

The key strength of the Prophet model lies in its ability to adeptly capture both trend, seasonality, and holiday patterns within the NYC yellow taxi daily demand dataset. This crucial capability ensures a nuanced understanding of the underlying factors influencing demand, making it the ideal forecasting tool for our specific application. In contrast, alternative models, including Classical Decomposition (Additive and Multiplicative), Rolling Forward (W=3), Seasonal Naive, Holt-Winters, and (S)ARIMA, yielded higher RMSE and MAPE values. For instance, the multiplicative model under Classical Decomposition recorded an RMSE of 91387.14 and a MAPE of 99.98, further highlighting the significant advantages of the Prophet model for predicting daily demand in this specific transportation context.

In summary, the Facebook Prophet model stands out not only for its superior accuracy but also for its tailored suitability to capture the intricate patterns inherent in the daily demand of NYC yellow taxis. The decision to proceed with the Prophet model is validated by its exceptional performance, making it the preferred forecasting solution for our application requirements in predicting the daily demand for NYC yellow taxis.

Business Implications

The successful forecasting of daily demand for New York City's iconic yellow taxis has far-reaching implications for the industry and the urban landscape. Our project leveraged advanced data analytics and the Prophet model to uncover vital insights, enabling more efficient and responsive operations within the yellow taxi sector. These insights translate into tangible benefits for both taxi companies and passengers:

1. Dynamic Pricing Strategy:

The forecasting model empowers the implementation of a dynamic pricing strategy. By predicting demand peaks, the yellow taxi industry can dynamically adjust prices, striking a balance between supply and demand. This not only maximizes profitability but also offers competitive rates to customers, enhancing overall market competitiveness. Implementing dynamic pricing based on demand forecasts can increase revenue while also avoiding the inefficiencies of overcharging or undercharging. This pricing strategy might boost profits by around 5-10%, considering the balance between demand and price elasticity.

2. Optimized Driver Schedules:

With precise demand forecasts at their disposal, taxi company can optimize driver schedules effectively. This optimization leads to increased earnings for drivers, reduces idle time, saving costs related to fuel, maintenance, and driver wages, and ensures better service availability for customers. In essence, it's a win-win scenario for both drivers and passengers. A conservative estimate could suggest a cost reduction of approximately 10-15% in fleet management expenses.

3. Enhanced Traffic Management:

Accurate demand forecasting plays a pivotal role in improving citywide traffic management. By anticipating busy areas and peak hours, the yellow taxi industry contributes to alleviating traffic congestion in New York City. This not only benefits taxi operations but also enhances the overall efficiency of the city's transportation network.

4. Customer Satisfaction:

Reduced waiting times and improved service reliability directly impact customer satisfaction. The convenience of knowing that a taxi will be available when needed strengthens the brand image of the yellow taxi industry, particularly in a highly competitive transportation market. Enhanced customer satisfaction and brand image can translate into higher customer retention and acquisition, indirectly boosting revenue. This aspect, though not a direct cost-saving measure, contributes to the overall financial health of the industry.

5. Data-Driven Decision Making:

The use of advanced analytics, such as the Prophet model, empowers decision-makers within the industry. It equips them with actionable insights, facilitating informed strategic choices for the industry's growth and adaptation to evolving urban demands. The integration of data-driven decision-making processes will not only enhance operational efficiency, but it also leads to a substantial overall cost reduction. This cost-saving impact underscores the tangible benefits that advanced analytics can bring to the financial viability of businesses within the yellow taxi industry.

Future Recommendations:

As we look ahead, there are several avenues for further improvement and development:

Integration with Real-Time Data:

- Advanced Data Sources: Exploring and incorporating additional real-time data sources beyond traffic and weather, such as events, road closures, and public transportation schedules. These sources can provide a more comprehensive view of factors impacting demand.
- Machine Learning Integration: We might consider leveraging machine learning algorithms for real-time data processing and pattern recognition. Tools like XGBoost or Random Forest can enhance the ability to respond swiftly to unforeseen demand fluctuations.

Custom Seasonality Patterns:

- Event-Specific Models: Developing specialized forecasting models tailored to specific events or holidays. Implementing models that can detect and respond to unique seasonality patterns associated with major occasions, festivals, or tourist influxes.
- Social Media Data: Utilization of sentiment analysis and social media data to gauge public sentiment and predict demand shifts related to events, festivals, or major news in the city.

Advanced Analytics Integration:

- Deep Learning Networks: Investigating the potential of deep learning architectures, such as Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU), for time series forecasting. These models can capture complex temporal dependencies and further improve accuracy.
- Ensemble Models: Implementing ensemble forecasting techniques that combine the strengths of multiple models. Combining methods like ARIMA, Exponential Smoothing, and neural networks can lead to more robust predictions.

Identification of External Factors:

- Economic Indicators: Incorporating economic indicators such as GDP growth, inflation rates, and unemployment rates into the forecasting model. These indicators can help predict demand fluctuations influenced by changes in the economic situation.
- Weather Changes: Enhancing the model's ability to incorporate weather data by considering not only current weather conditions but also longer-term weather forecasts. Severe weather events can have a significant impact on demand patterns.

Conclusion:

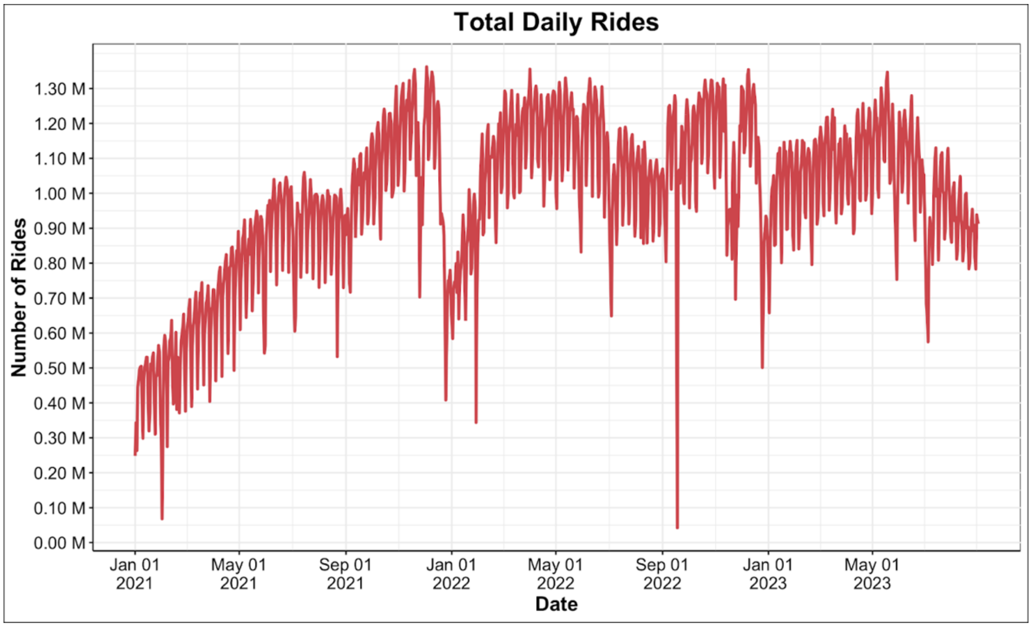
In conclusion, our project not only showcases the successful application of time series forecasting techniques, particularly with tools like Prophet, but it also highlights the substantial benefits they bring to traditional industries like taxi services. By accurately predicting daily demand and assisting in strategic planning, we have demonstrated the transformative power of data-driven forecasting in enhancing the operational efficiency of New York City's yellow taxi industry.

This project underscores the immense potential of time-series data-driven approaches to transform traditional sectors. By harnessing predictive analytics for decision-making and efficiency improvements, we pave the way for industries to adapt and thrive in an increasingly data-centric world. The yellow taxi industry is a prime example of how innovative applications of data science can rejuvenate and revitalize longstanding sectors, ensuring their continued relevance and competitiveness.

APPENDIX

EDA Visualizations

Figure 1. Analysis of Total Daily and Monthly Taxi Demand



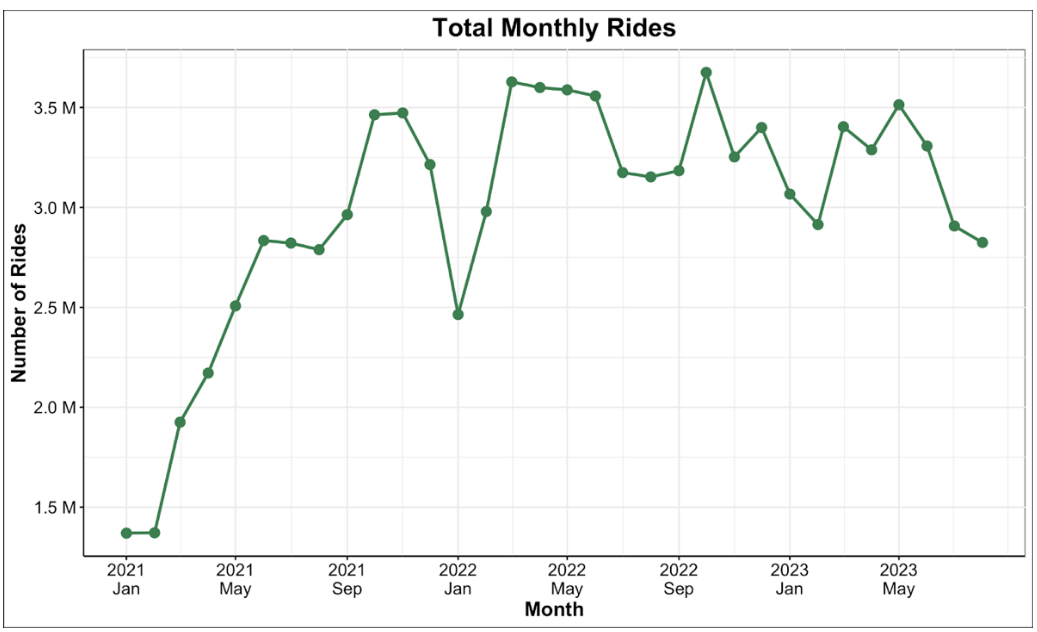


Figure 2. Analysis of Daily Ride Distribution Across Days of the Week

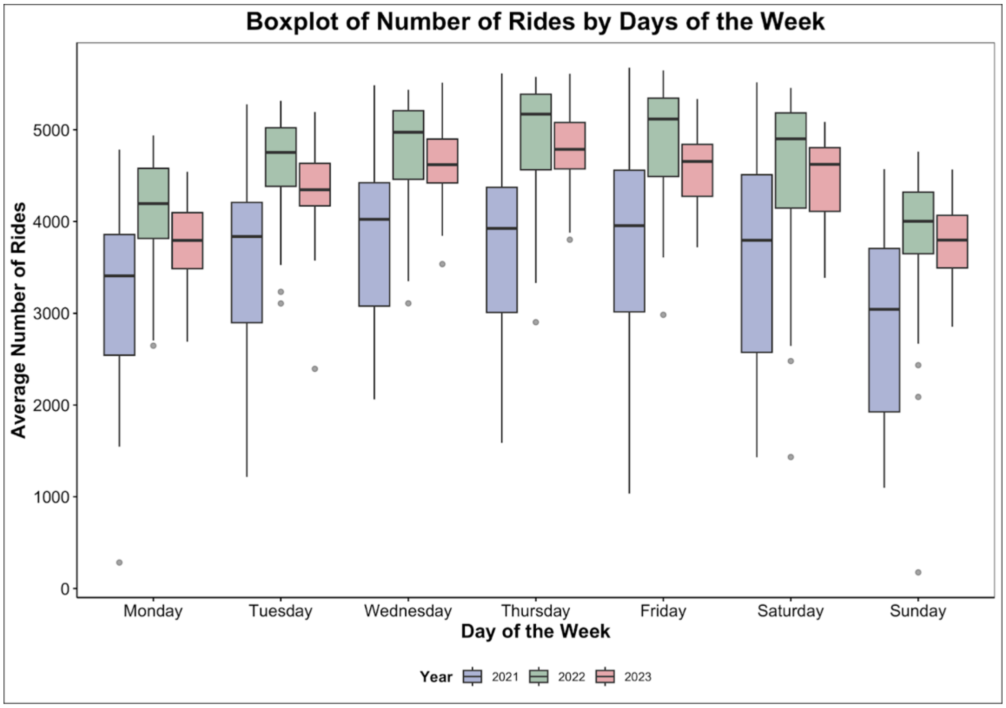


Figure 3. Hourly Demand and Pricing Analysis

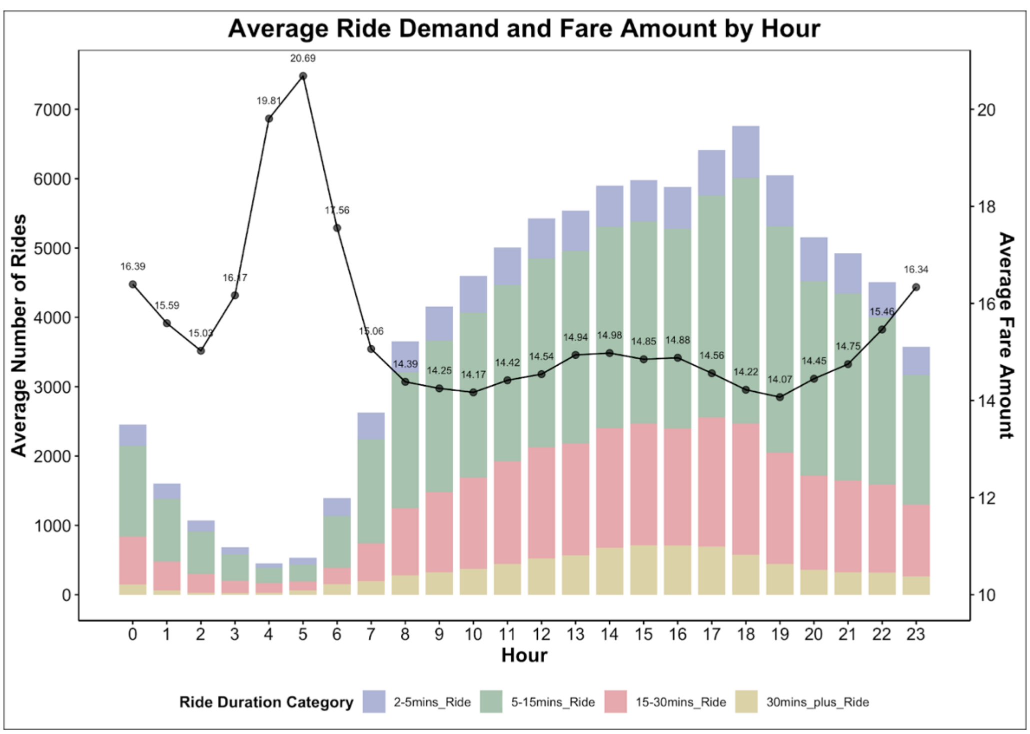


Figure 4. Revenue Generation Metrics and Traffic Dynamics

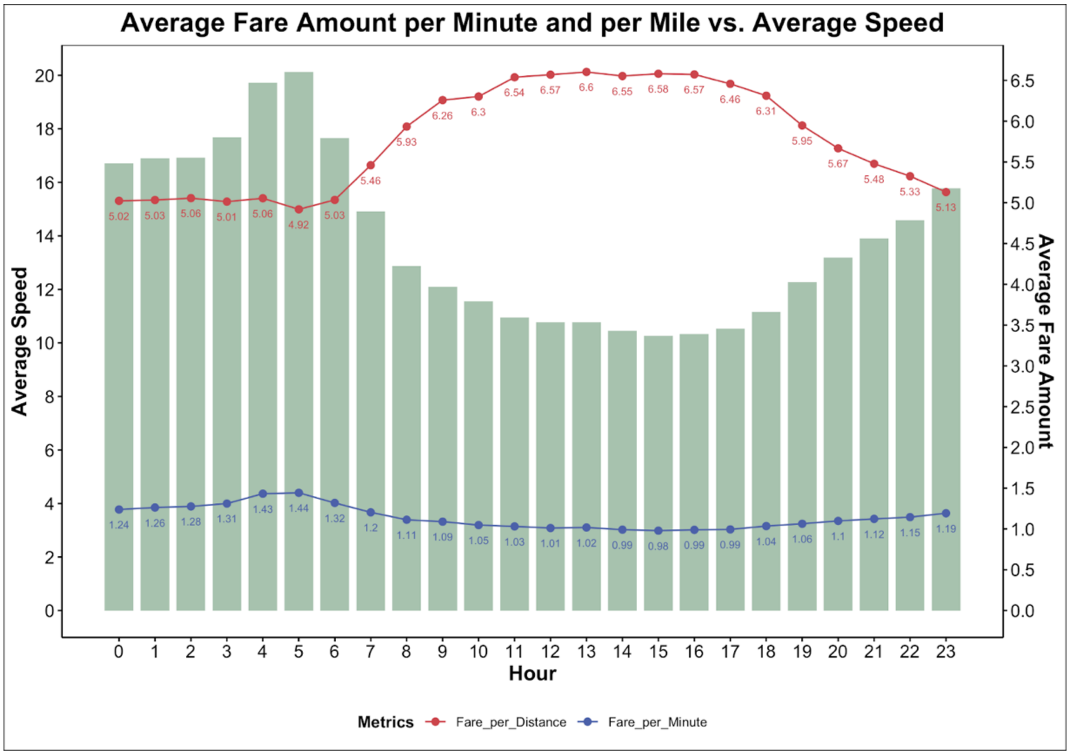
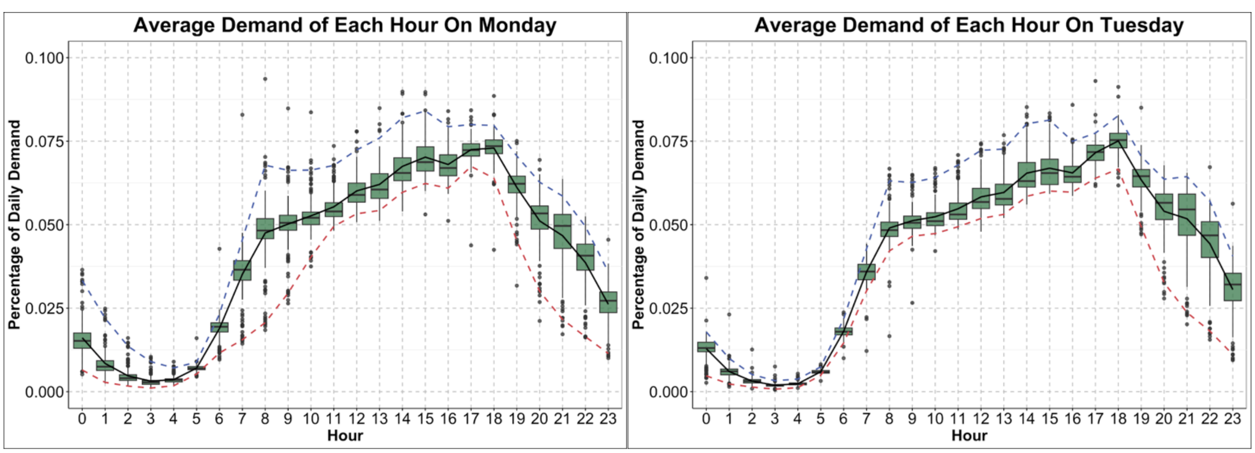


Figure 5. Hourly Demand Distribution Pattern Among Workdays



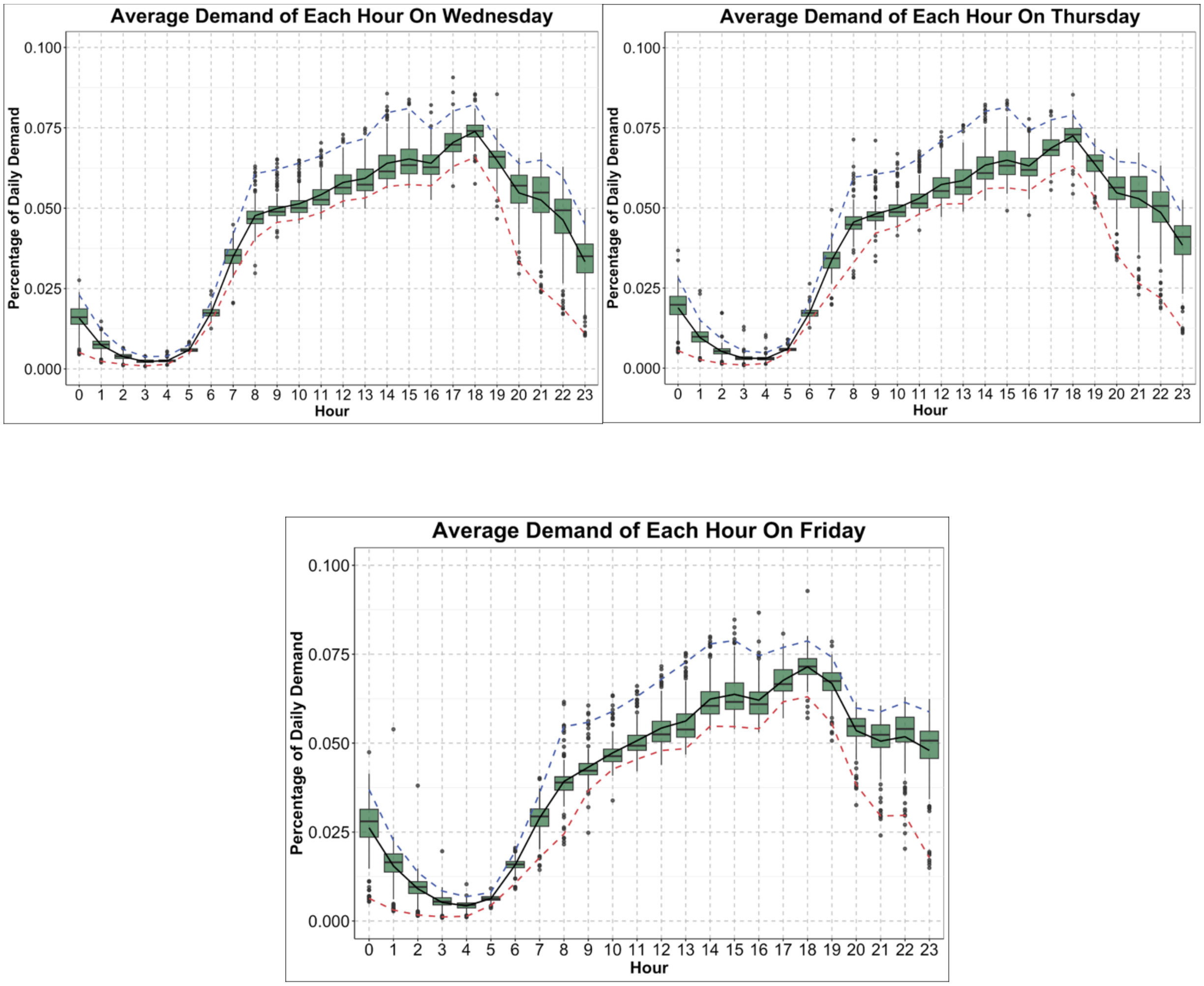
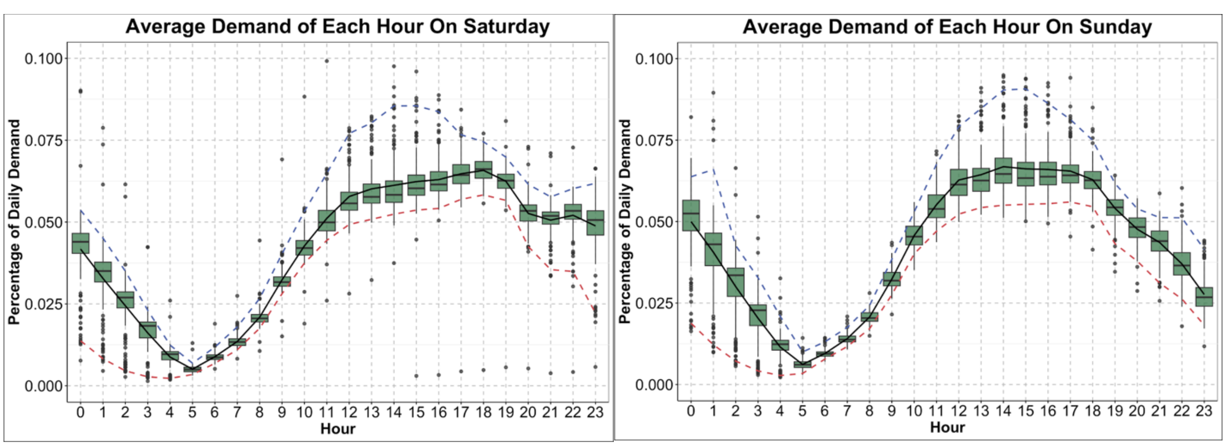


Figure 6. Hourly Demand Distribution Pattern Among Weekends



MODELING

Figure 7. CLASSICAL DECOMPOSITION (ADDITIVE MODEL)

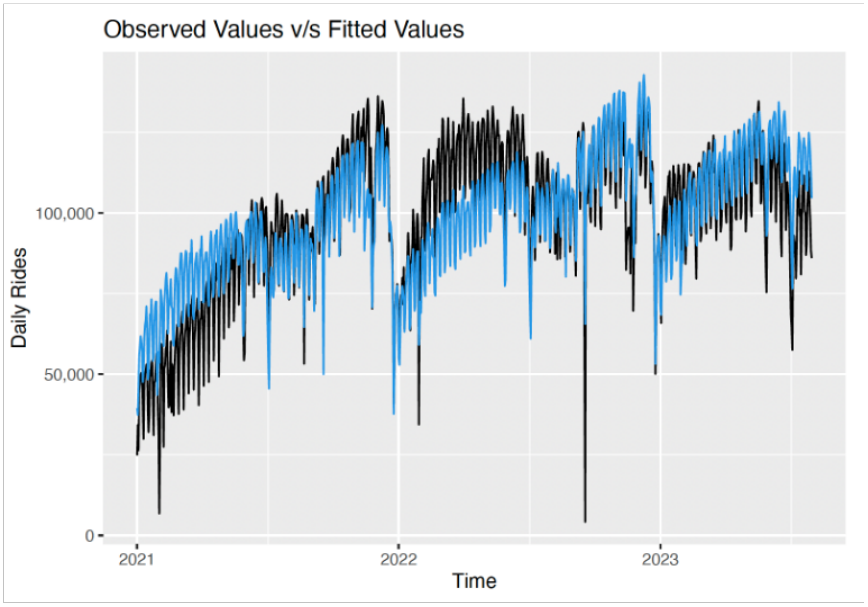


Figure 8. CLASSICAL DECOMPOSITION (MULTIPLICATIVE)

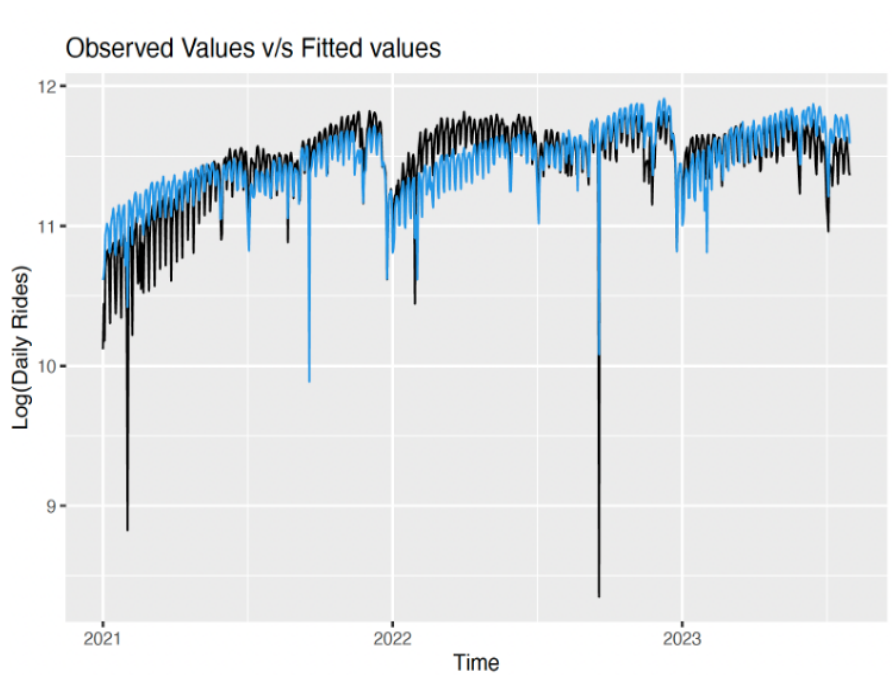


Figure 9. SEASONAL NAÏVE MODEL

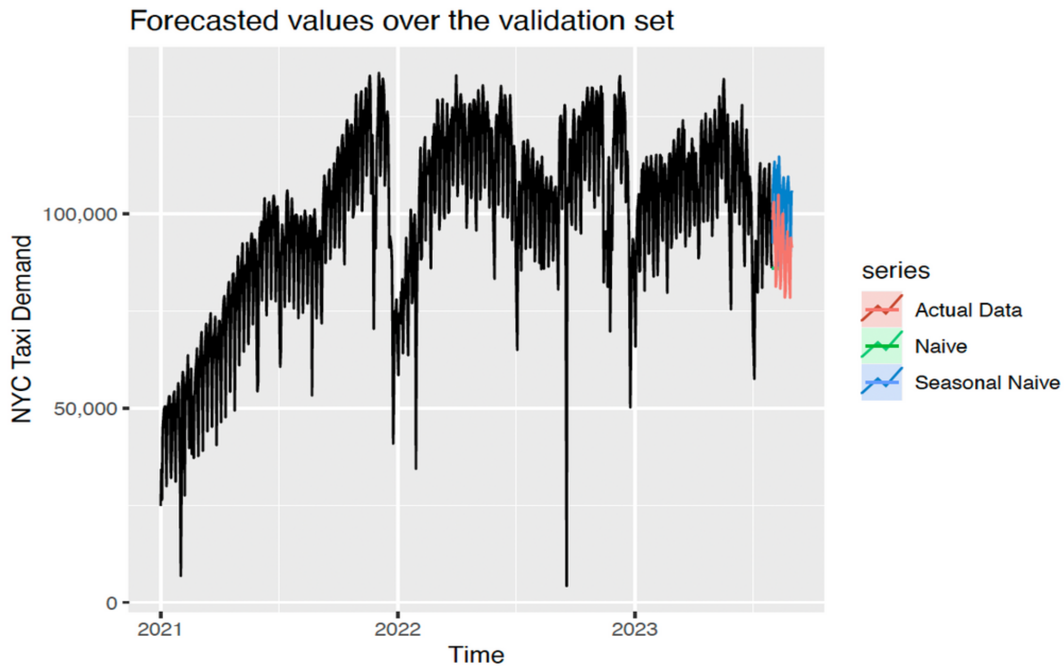


Figure 10. SMOOTHING TECHNIQUES (ROLLING FORWARD SMOOTHING WINDOW WITH W=3)

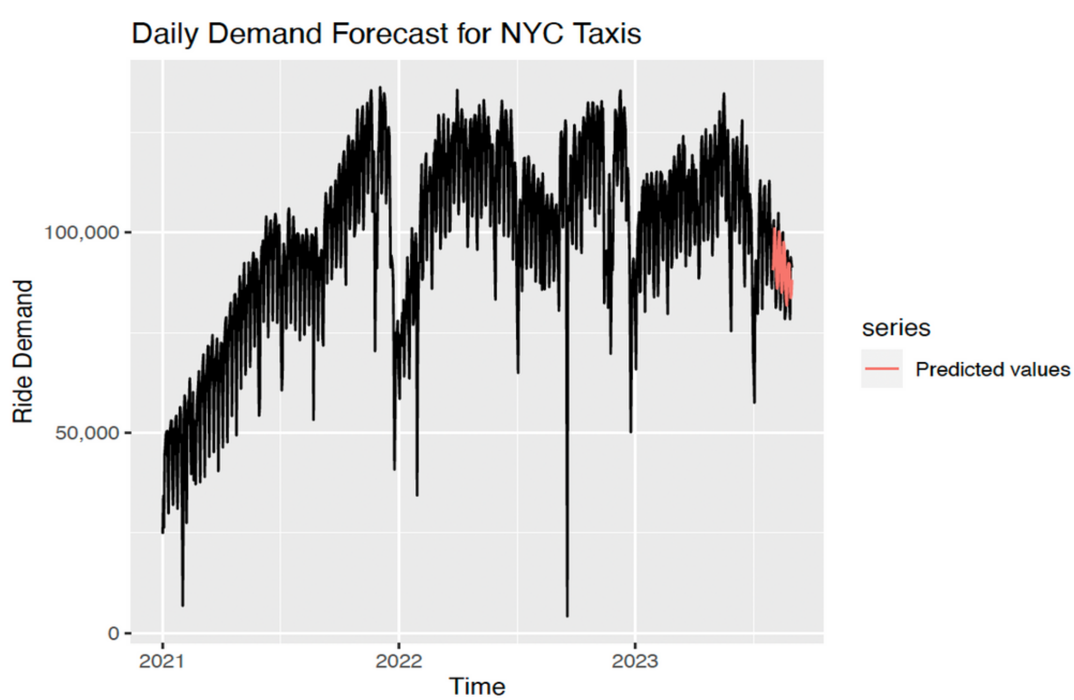


Figure 11. HOLT WINTER'S MODEL

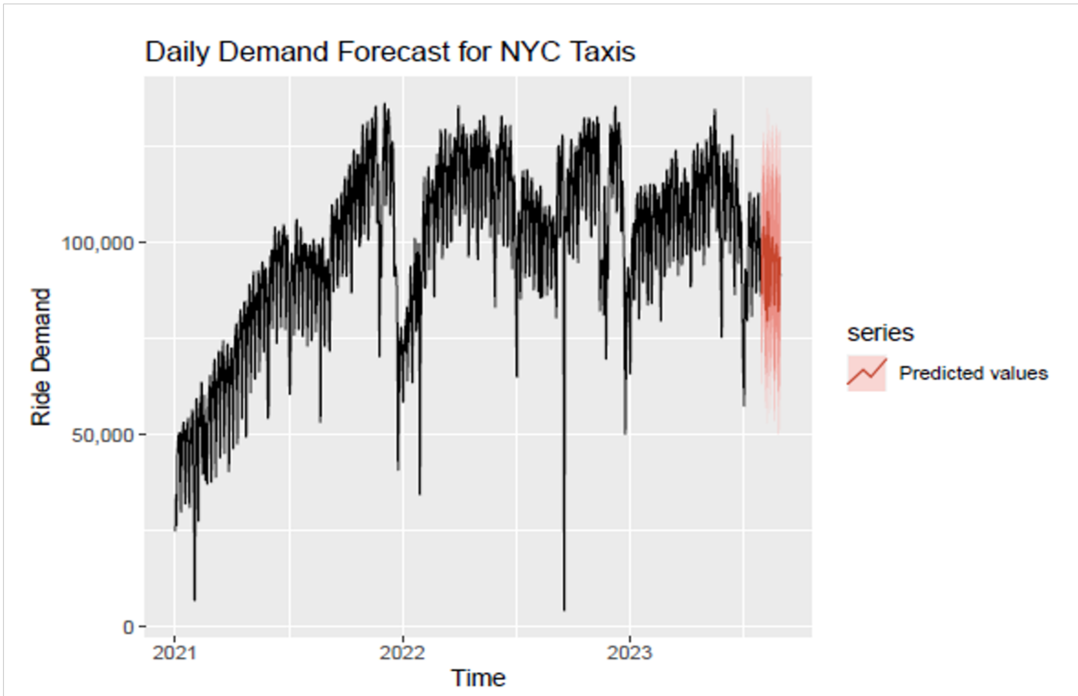
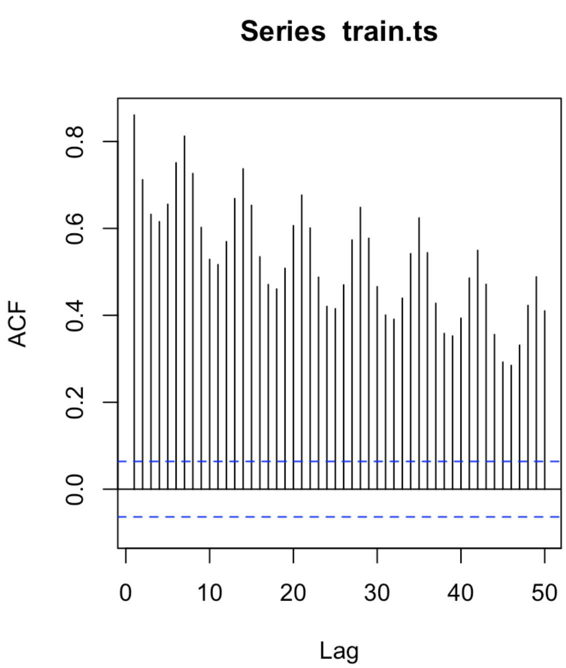
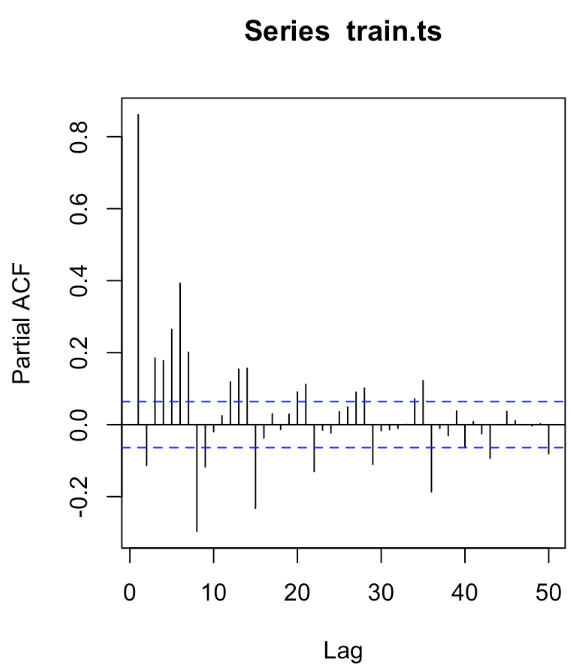


Figure 12,13. SEASONAL ARIMA MODEL





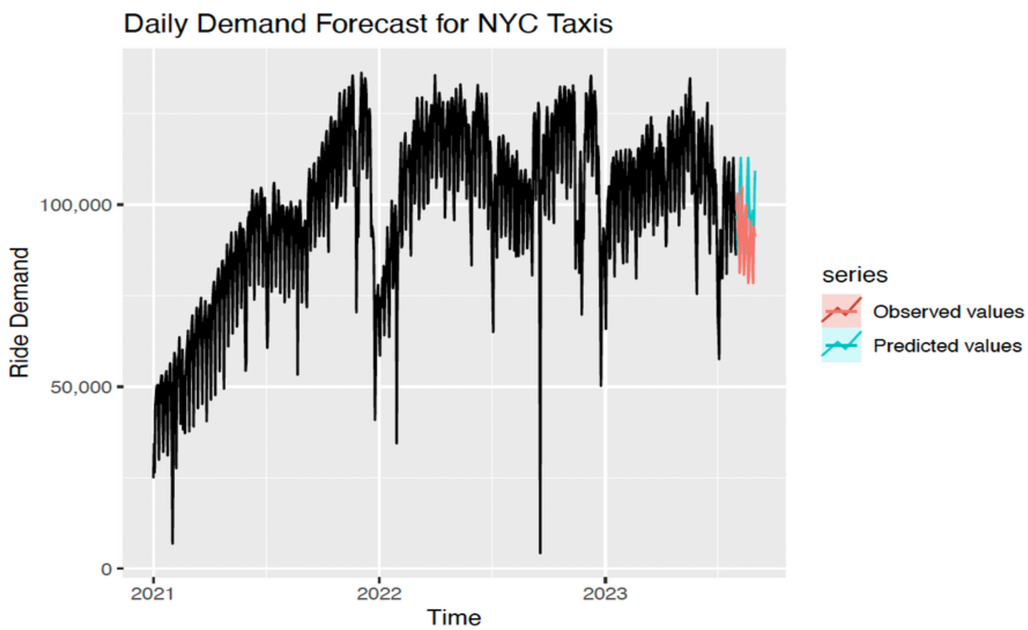


Figure 14. FB PROPHET MODEL

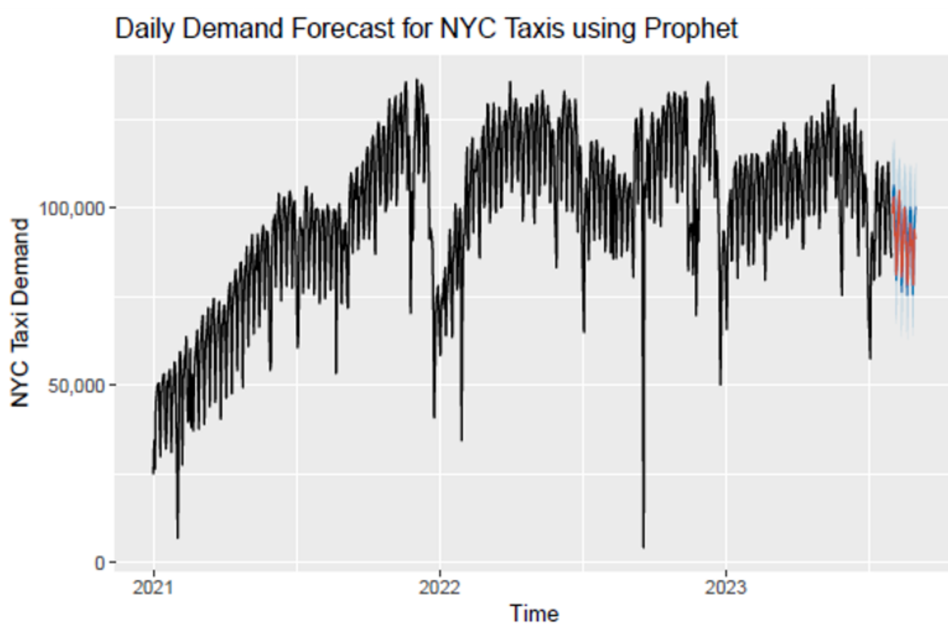
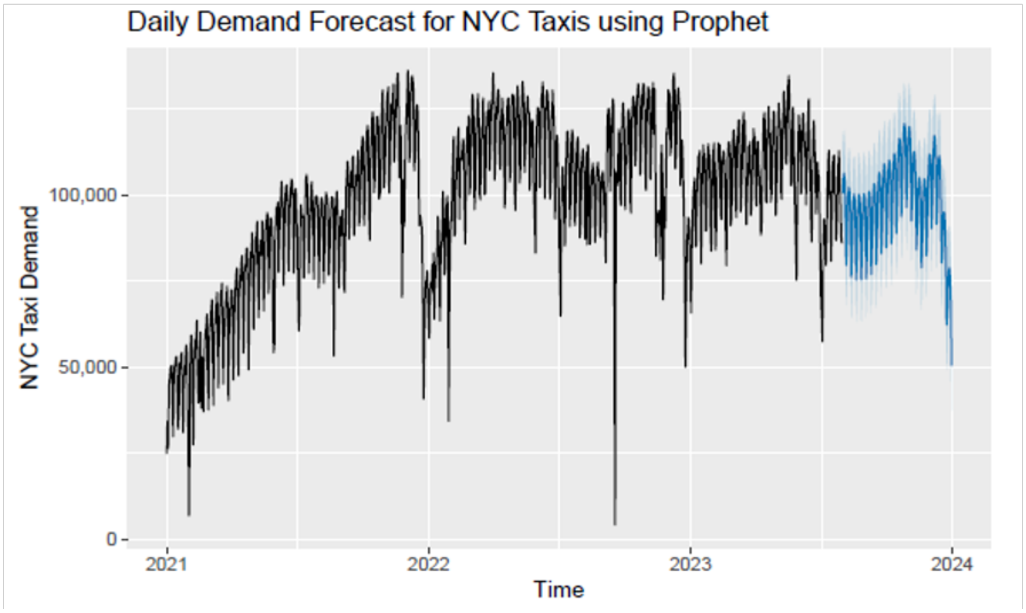


Figure 15. PREDICTION OF DAILY NYC TAXI DEMAND FOR THE REST OF 2023 USING PROPHET



PERFORMANCE METRICS

Table-1

Model Name	**Additive Model**	**Multiplicative Model**	**Seasonal Naive**	**Holt Winters**	**Seasonal ARIMA**	**Rolling Forward Smoothing (W=3)**	**FB Prophet**
RMSE	30126.24	91387.14	15366.066	9746.41	13548.78	9262.07	4987.14
MAPE	31.63	99.98	14.99	8.45	12.75	9.45	4.14