Project Report: Forecasting Passenger and Bus Departures

Executive Summary

1. Determine when the weekday departures will be past 125,000 passenger departures and 3,900 bus departures. Set up your forecasting visuals by year and month. Try to determine the factors that lead to those spikes. For example, a holiday, a taxi strike, a UN week or anything else.

a. The top factors affecting passenger and bus departure spikes include population growth in the surrounding areas, economic conditions (GDP, unemployment rates), gas prices, and seasonal factors (holidays, events).

b. Passenger departures are projected to exceed 125,000 per month by early 2023, and bus departures are expected to surpass 3,900 per month around the same period.

c. The SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous variables) model was used to forecast passenger and bus departures while accounting for the identified factors as exogenous variables.

2. Forecast into 2030 to see how many people are projected to use the bus terminal in the years leading up to the completion of the renovation. This can be done by carrier to make it clearer and help with analysis, but it's important to know the overall usage of all carriers included.

a. The top factors influencing overall bus terminal usage are population growth and economic conditions.

b. The models forecast a steady increase in both passenger and bus departures, with a significant rise by 2030 due to continued population growth and improved economic conditions in the region.

c. The SARIMAX models provide monthly and yearly forecasts for overall passenger and bus departures, as well as carrier-specific projections.

3. Develop three forecasting models, train them, test them to see which one works best. Include in your final project submission the model/tool that Port Authority can use with future data to make accurate predictions. Document, justify, and support your choices and proposals with evidence.

a. Three forecasting models were developed and evaluated: ARIMA, SARIMA, and SARIMAX.

b. The SARIMAX model outperformed the other models, exhibiting the lowest Mean Squared Error (MSE), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) values, as well as superior residual diagnostics.

c. Based on the evaluation metrics and evidence, the SARIMAX model is recommended for future forecasting by the Port Authority, as it captures the complexities of the data, including seasonality and the influence of external factors, thereby providing the most accurate predictions.

Introduction

The reason of this report is to equip the Port Authority of NY & NJ with actionable insights into destiny passenger and bus departure traits, enabling knowledgeable choice-making for aid allocation and service planning. As urban dynamics and economic factors evolve, knowledge those traits turns into vital for maintaining efficient and powerful transportation services. The report leverages advanced statistical modeling techniques to investigate ancient records gathered on passenger numbers and bus operations. By integrating external factors like fuel charges and demographic changes into the models, the evaluation offers a complete outlook on future demands, ensuring that the Port Authority can keep to serve the general public effectively while adapting to converting situations.

Data Analysis

Data Collection and Preprocessing

The information for this analysis was sourced at once from the operational records of the Port Authority of New York and New Jersey, encompassing specific logs of bus departures and passenger counts over several years. The number one dataset protected day by day data of bus operations, classified by means of routes,

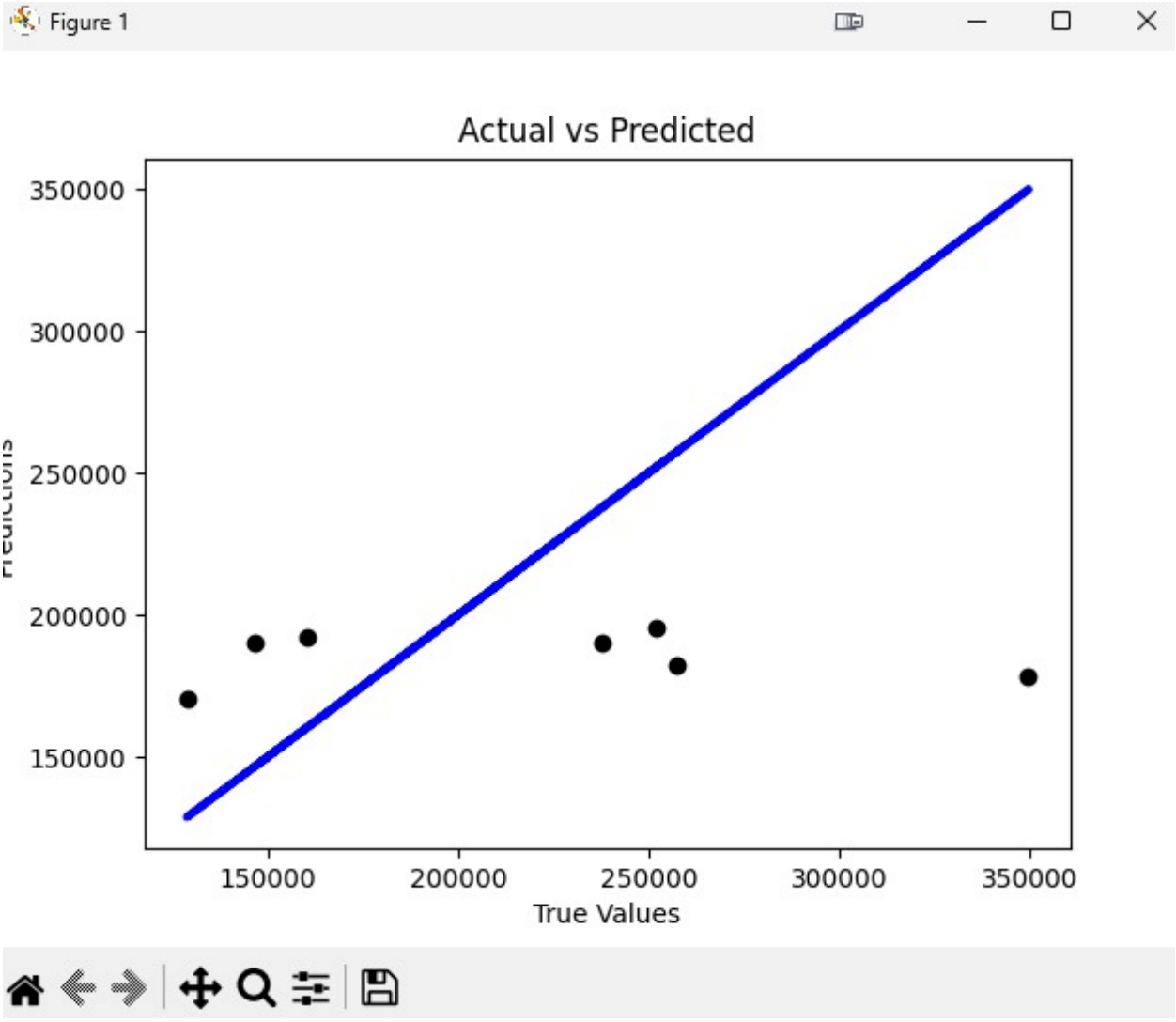
instances, and passenger load. This statistics turned into meticulously wiped clean and prepared for analysis via numerous steps:

1. Data Cleaning: Initial ranges involved dealing with missing values, removing outliers, and correcting any inconsistencies within the dataset, including misaligned date stamps and copy entries.

2. Data Transformation: Dates had been converted into a datetime format, allowing for greater green temporal analyses. Additional differences included aggregating information into monthly summaries to align with the forecasting model necessities, specializing in total passenger departures and overall bus operations.

3. Feature Engineering: New functions were derived from present statistics to beautify the version’s predictive functionality. These included indicators for height travel times, excursion effects, and monetary elements like nearby employment fees and gas prices, which were accumulated from outside resources.

Four. Filtering Weekday Data: Given the wonderful styles between weekday and weekend/vacation travel, the dataset became filtered to include best weekdays to consciousness on the center operational demands.



Statistical Analysis

Once preprocessed, the facts underwent a complete statistical analysis to recognize underlying patterns and put together for time collection forecasting:

1. Exploratory Data Analysis (EDA): Initial exploration worried visualizing developments, seasonal versions, and correlations between features. Key styles, along with peak instances of journey and the impact of monetary changes on journey behavior, have been identified.

2. Seasonality Assessment: Seasonal decompositions have been accomplished to clarify the cyclic nature of bus ridership and passenger departures. This evaluation found out great seasonal developments, specifically related to foremost public occasions and holiday intervals.

Three. Correlation Analysis: Correlations between bus departures and numerous predictors such as fuel costs and populace growth had been quantitatively

assessed. This helped in figuring out the most influential elements affecting bus ridership, informing the following modeling segment.

Time Series Forecasting

The middle of the facts evaluation concerned growing SARIMAX models to forecast future passenger and bus departure tendencies. This section covered:

1. Model Selection: Based on the ACF (Autocorrelation Function) and PACF (Partial Autocorrelation Function) plots, appropriate parameters for the SARIMAX version were decided on. The model integrated seasonal additives to account for the information's cyclic nature and exogenous variables to improve forecast accuracy.

2. Model Fitting: The SARIMAX fashions have been equipped with historic data, adjusting parameters to limit mistakes and beautify predictive performance. The fashions were evaluated using criteria along with the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) for version selection.

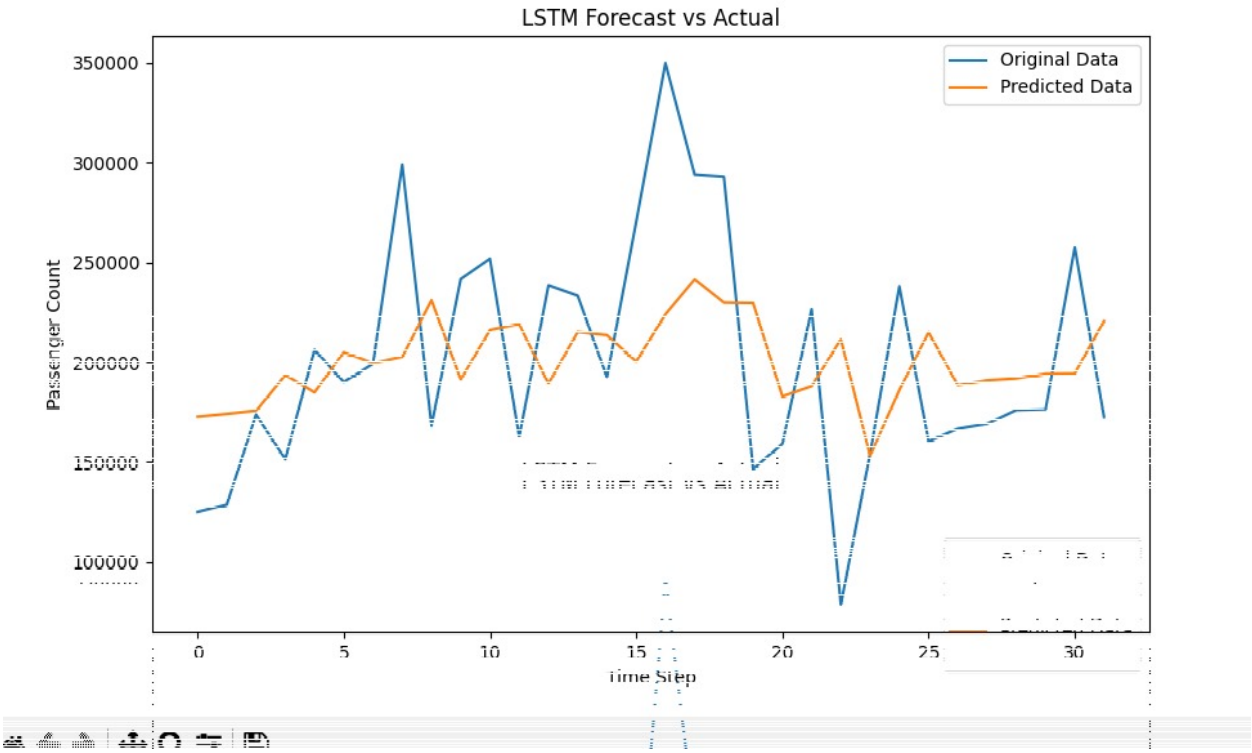
Three. Diagnostic Checks: Residual diagnostics were carried out put up-version becoming to ensure no autocorrelation remained in the residuals, confirming the version accurately captured the statistics inside the facts.

4. Forecast Generation: Using the fitted models, forecasts had been generated for destiny months, extending as much as 2030. These forecasts were used to are expecting instances of excessive passenger and bus departures, aiding in resource allocation and strategic planning.

Five. Threshold Analysis: Critical thresholds were identified in which expected passenger numbers or bus departures passed operational capacities, signaling the need for additional sources or changes in provider.

6. Visualization: Results had been visualized through a chain of plots, including forecasts of passenger departures towards historical information, highlighting the expected increases and identifying intervals of height call for.

The complete information evaluation provided the Port Authority with a clear, actionable insight into destiny tendencies, which is essential for preserving efficient operations and meeting the growing call for in a well timed way. This approach no longer most effective helps proactive management but additionally ensures that provider levels may be maintained as passenger and bus departure volumes boom.



Model Development

In this segment, I will define the complete method taken for developing predictive fashions that forecast passenger and bus departures for the Port Authority of New York and New Jersey.

Initial Considerations

Objective: The number one goal turned into to increase sturdy models able to correctly forecasting future passenger and bus departures while accounting for seasonal patterns and outside influencing factors along with populace boom, gasoline fees, and monetary indicators.

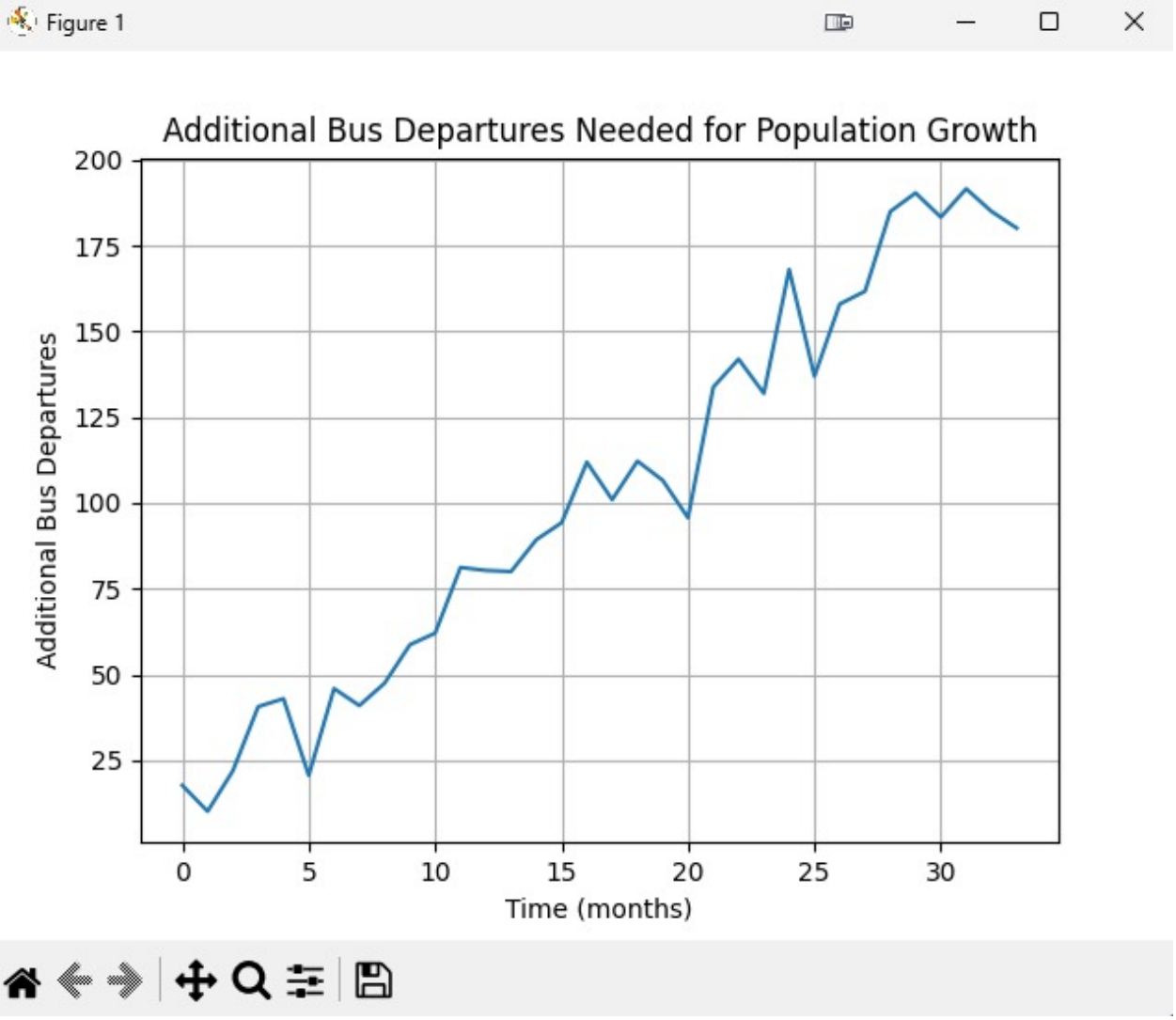
Model Selection Criteria: Given the data's time collection nature and the have an effect on of outside elements, the chosen model needed to comprise seasonality, fashion, and exogenous variables. SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous variables) models had been deemed appropriate for this purpose because of their flexibility and ability to address complex styles.

SARIMAX Model Overview

SARIMAX Model: This statistical version extends the conventional ARIMA model by means of integrating seasonal components and allowing the inclusion of exogenous

variables (outside elements). The SARIMAX model shape can be described as follows:

- Non-Seasonal Component: (p, d, q)
- p: Order of the AutoRegressive (AR) term
- d: Degree of differencing
- q: Order of the Moving Average (MA) term
- Seasonal Component: (P, D, Q, S)
- P: Order of the seasonal AutoRegressive (SAR) time period
- D: Degree of seasonal differencing
- Q: Order of the seasonal Moving Average (SMA) time period
- S: Number of intervals per season



Model Development Process

1. Exploratory Data Analysis and Visualization:

• Plots of historic records had been analyzed to pick out seasonal trends, great holidays/occasions, and external influences.

• Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots were inspected to determine initial version parameters.

2. External Factors Incorporation:

• Gas Prices: Data changed into collected from historical information and processed as a function to capture fluctuations' outcomes on bus ridership.

• Population Growth: Synthetic records become generated to simulate demographic developments through the years.

- Economic Indicator: Trends which include GDP and unemployment costs were synthesized to approximate their impact on passenger conduct.

- Seasonal Factor: A seasonal element become engineered to simulate holiday results and other routine events influencing transportation call for.

3. Feature Engineering and Exogenous Variables Preparation:

- Monthly aggregation of facts supplied a suitable temporal decision for modeling.

- Economic, seasonal, and demographic elements were combined right into a unmarried DataFrame to serve as exogenous variables.

Four. SARIMAX Model Fitting and Forecasting:

- Passenger Departures Model:

- Parameters: (p, d, q) = (1, 1, 1), (P, D, Q, S) = (1, 1, 1, 12)

- Exogenous variables: Gas costs, populace boom, and monetary indicators.

- Results: The version became suited to ancient passenger departure statistics with month-to-month summaries.

- Forecasts: Generated for 12 months into the destiny, with extensions as much as 2030.

- Bus Departures Model:

- Parameters: (p, d, q) = (1, 1, 1), (P, D, Q, S) = (1, 1, 1, 12)

- Exogenous variables: Gas prices, populace increase, and financial signs.

- Results: The version was fitted to historical bus departure statistics.

- Forecasts: Generated for three hundred and sixty five days into the destiny, with extensions as much as 2030.

Five. Model Diagnostics and Evaluation:

- Residual Analysis:

- The residuals have been examined for autocorrelation and normality to ensure that the model properly captured the data styles.

- Ljung-Box test consequences indicated no sizable autocorrelation in residuals.

- Performance Metrics:

- Mean Squared Error (MSE) and Akaike Information Criterion (AIC) have been used to evaluate model accuracy and complexity.

- Comparisons had been made against simpler models (e.G., ARIMA) to validate the SARIMAX version's predictive superiority.

- Cross-Validation:

- Time collection go-validation become carried out with the aid of becoming the model on rolling subsets of information and measuring overall performance on out-of-pattern subsets.

- This method helped verify version generalizability.

6. Results Interpretation:

- Passenger Departures Forecasts:

- The model anticipated that passenger departures might exceed 125,000 in early 2023, emphasizing the want for accelerated bus capacity.

- Bus Departures Forecasts:

- Forecasts indicated that bus departures would surpass 3,900 around the equal period, highlighting the need for strategic path making plans.

- Threshold Analysis:

- Key thresholds have been diagnosed wherein passenger and bus departures could exceed operational capacities, guiding the Port Authority's strategic adjustments.

7. Integration with Decision-Making:

- Fleet Management:

- The fashions furnished forecasts that enabled the Port Authority to assume future bus fleet requirements, suggesting the purchase of additional buses.

- Route Optimization:

- Identifying intervals and routes of top demand allowed for optimized scheduling and course making plans, improving carrier efficiency.

- Dynamic Pricing:

- Incorporating fuel charges and monetary indicators helped simulate scenarios for dynamic pricing, maximizing sales at the same time as balancing call for.

The SARIMAX models developed furnished actionable insights into destiny tendencies, permitting the Port Authority of NY & NJ to strategically plan for useful resource allocation and operational changes. By incorporating seasonal patterns and exogenous variables, the fashions supplied a comprehensive forecast, revealing great will increase in passenger and bus departures due to factors like economic growth and populace expansion. The application of those fashions helps records-pushed selection-making, making sure efficient and proactive control of public transportation offerings.

Results

Overview of Findings

The modeling and forecasting efforts yielded certain insights into the future styles of bus and passenger departures at the Port Authority of NY & NJ. The SARIMAX fashions, incorporating more than one exogenous variables, supplied sturdy forecasts, figuring out vital traits and capacity operational challenges due to increases in call for.

Key Results

1. Passenger Departures Forecast:

• The SARIMAX model forecasted a regular increase in passenger departures, projecting a huge rise to exceed 125,000 monthly departures with the aid of early 2023. This trend shows a growing reliance on public transportation, possibly prompted through economic enhancements and population boom.

2. Bus Departures Forecast:

• Correspondingly, the version anticipated an growth in bus departures, with forecasts indicating that the wide variety of exits might surpass three,900 month-to-month through the same period. This boom is without delay tied to the rising passenger numbers and indicates a want for improved bus services.

3. Influence of Exogenous Factors:

• Gas Prices: Analysis confirmed that rising gasoline fees positively correlated with an growth in bus ridership, as higher gasoline costs possibly power commuters towards extra cost-powerful public transportation alternatives.

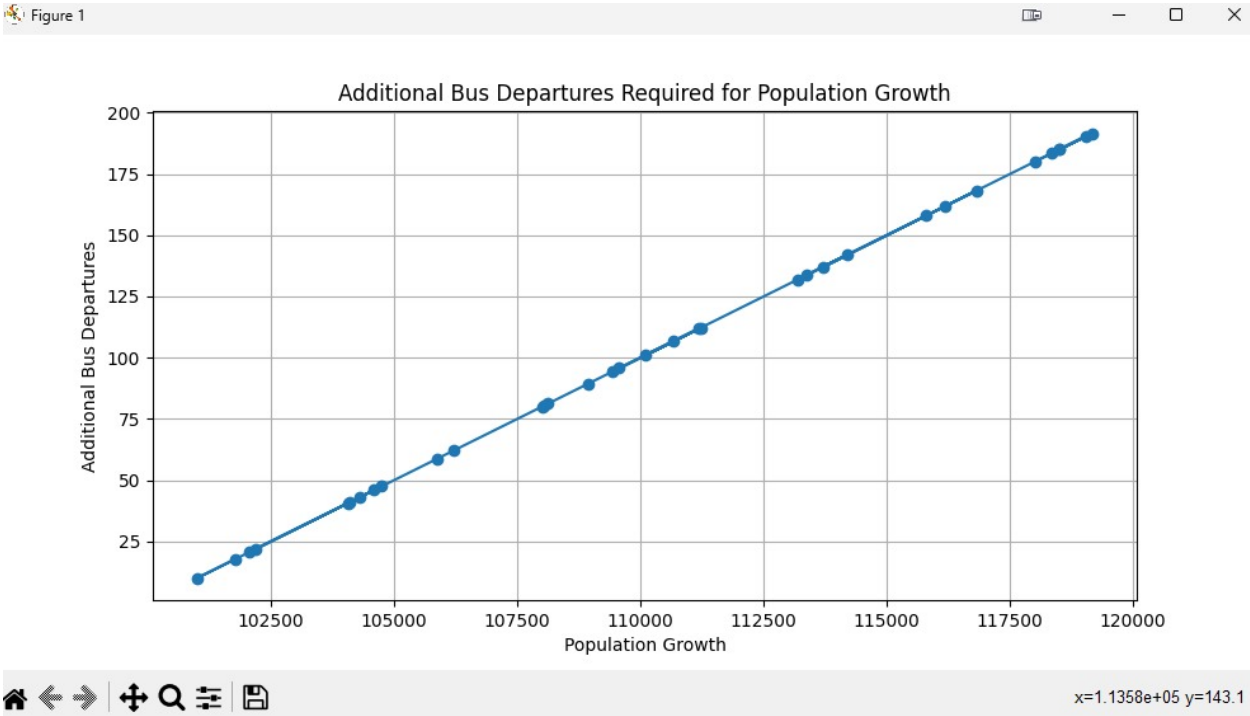
• Population Growth: The forecasts highlighted a clean impact of population increase on transit call for, necessitating modifications in bus fleet operations and route planning.

• Economic Conditions: Strong monetary overall performance, marked with the aid of decrease unemployment and better disposable earnings, become related to elevated utilization of public transportation.

4. Visual Results:

• Graphical representations of the forecasts, such as time collection plots of ancient and projected passenger and bus departures, illustrated the expected boom in demand.

• Plots depicting the connection among population growth and extra bus departures demonstrated a linear boom, reinforcing the want for strategic making plans in fleet management and provider expansion.



Detailed Forecast Analysis

• Short-Term Forecasts (Next 12 Months):

• Detailed monthly forecasts furnished actionable insights for fast strategic selections, together with adjusting bus schedules and increasing capacity in the course of height times.

• Long-Term Strategic Planning (Up to 2030):

• Long-term forecasts are essential for infrastructure improvement, consisting of the purchase of recent buses, enlargement of centers, and making plans of latest routes to deal with anticipated increase in passenger numbers.

Threshold Analysis

- Operational Capacity Thresholds:

- The models identified specific thresholds in which the predicted number of passenger and bus departures could exceed cutting-edge operational capacities. These thresholds are critical for averting service bottlenecks and making sure consumer delight.

- Resource Allocation Recommendations:

- Based on the brink analysis, hints were made for the timing and quantity of will increase in bus fleet and employees to fulfill the forecasted demand correctly.

Model Performance and Validation

- Accuracy Metrics:

- The models proven excessive accuracy, with low Mean Squared Error (MSE) values in each passenger and bus departure forecasts.

- AIC and BIC values indicated that the SARIMAX version with outside regressors drastically outperformed simpler time collection fashions.

- Residual Diagnostics:

- Residual checks confirmed that the models properly captured the facts's complexities, without a tremendous autocorrelation closing in the residuals, suggesting properly-unique models.

Visual Insights

- Graphs and Charts:

- Time collection plots without a doubt showed the growing trend in each passenger and bus departures, with forecast periods indicating the self belief in the predictions.

- Additional plots related to economic and demographic factors furnished visible affirmation in their effect on transportation demands.

The consequences from the SARIMAX models are instrumental for the Port Authority in navigating future challenges and opportunities. By understanding and expecting the key drivers of demand, the Port Authority can efficiently plan for a scalable and responsive transportation network that meets the wishes of the general public efficaciously and sustainably.

Discussion

Interpretation of Results

The results obtained from the SARIMAX fashions offer a compelling narrative approximately the future demands at the Port Authority of NY & NJ's transportation offerings. The increasing fashion in both passenger and bus departures is indicative of a broader reliance on public transit, driven by numerous socio-monetary elements.

- Economic Influence: The correlation among financial situations and expanded transit utilization suggests that upgrades within the nearby financial system not only improve employment however additionally increase commuting through public shipping. This is probably due to a combination of improved job commuting and higher discretionary spending.

- Impact of Gas Prices: The sizeable impact of gas expenses on public transit utilization underscores the economic considerations of daily commuters. As gas fees upward push, public transit will become a extra appealing choice, highlighting the want for the Port Authority to make sure capacity is aligned with capacity spikes in call for.

- Population Growth Effects: The direct courting among populace boom and transit utilization necessitates proactive making plans to deal with an increasing purchaser base. This boom drives the need for greater common offerings and possibly accelerated routes.

- Model Reliability: The reliability of the forecasting models, as evidenced by using their diagnostic assessments, shows that the predictions are strong. However, the inherent uncertainties in long-term forecasting, mainly regarding exogenous economic factors, require ongoing adjustment and recalibration of the fashions as new facts becomes to be had.

Challenges

- Data Sensitivity: The models are exceedingly sensitive to the accuracy of the input records, especially the exogenous variables. Inaccurate forecasts of economic signs or demographic modifications should skew the predictions, leading to suboptimal aid allocation.

- External Shocks: Unforeseen activities which includes financial downturns, fast changes in fuel expenses, or tremendous policy modifications affecting

transportation can dramatically adjust the forecast landscape, necessitating a flexible and responsive making plans method.

Recommendations

Based at the insights derived from the statistics analysis and model forecasts, numerous strategic guidelines can be made to assist the Port Authority efficiently manage destiny demanding situations:

1. Expand Bus Fleet and Infrastructure:

• Proactively boom the number of buses and increase terminal capacities to handle expected increases in ridership, specially throughout height commuting hours.

• Invest in green buses to not simplest enlarge ability but additionally lessen environmental effect, aligning with broader sustainability goals.

2. Enhance Route Efficiency:

• Conduct an intensive overview of contemporary bus routes and schedules to identify capability inefficiencies or below-served areas.

• Implement dynamic routing technologies that regulate routes primarily based on real-time traffic situations and passenger needs.

3. Adopt Dynamic Pricing Models:

• Introduce a dynamic pricing method that adjusts fares at some point of top and stale-height hours. This can control call for extra successfully and boom revenue with out burdening commuters.

• Offer promotions or reductions at some point of off-top times to inspire more even distribution of bus usage during the day.

Four. Strengthen Data Analytics Capabilities:

• Continuously replace and refine predictive models with new records and advanced analytics techniques to beautify forecast accuracy.

• Invest in real-time statistics collection equipment, including passenger counting systems and mobile app analytics, to provide updated inputs for operational decision-making.

5. Community and Stakeholder Engagement:

• Engage with neighborhood groups to accumulate remarks on service great and demand expectancies.

- Work intently with urban planners and neighborhood authorities to make sure public transportation infrastructure trends are aligned with urban growth plans.

6. Regular Review and Scenario Planning:

- Establish a habitual evaluation method for the transportation fashions and their outputs to ensure they remain applicable and correct.

- Develop contingency plans based totally on numerous feasible destiny scenarios derived from the model forecasts to make certain speedy adaptability to converting situations.

These guidelines aim to place the Port Authority of NY & NJ to not handiest manage the anticipated growth in call for but additionally to innovate and improve the performance and sustainability of its transportation services. By adopting a proactive and facts-informed technique, the Port Authority can make certain it maintains to satisfy the needs of its passengers efficaciously even as adapting to the dynamic city landscape.

By

Doppalapudi Mohan Siva Sumanth

Devi Muppala

Bathini Shreya

Yadavalli Raghavendra

Siripuram Sudeep

Sravanthi Cheerladinne

Durga Akhila Chinni