Project Progress Report 3

Team members:

Giacomo Bizzotto, Rishitha Thatipally, Irfan Shaik, Errolla Vivek, Shraddha Shakya, Sampath Sai Raghav Allaboyina.

## 1 ) Executive Summary: Project Objectives and Model Selection Overview

The Port Authority of New York and New Jersey wants to plan better for the future at each of their facilities. They need to know how many people will be using each facility from 2025 to 2030. Things have changed a lot since COVID, so they need good guesses to help them decide how to run things. This report explains how we're using three tools to help them: first, we're figuring out what things really change how many people use each facility. Second, we're using past numbers to forecast how many people will be there in the future, both in total and for each carrier, at each facility. Third, we're looking at when each facility is busiest, so they can compare it to how busy it was before the pandemic. These tools will help the Port Authority make smart choices about how to use their space and help everyone get where they need to go.

This Project Progress Report 3 focuses on the model we used to address these goals. We will explain the methodology behind our approach and how it supports accurate forecasting and decision-making. Additionally, as a final result, we plan to present an interactive dashboard that will allow the company to explore the data dynamically. This dashboard will provide valuable insights into facility usage trends, peak hours, and projected demand, enabling the Port Authority to make data-driven decisions and optimize operations for the future.

## 2) Model 1: ETS (Error, Trend, Seasonality) model to forecast bus passengers from 2025 to 2030 in each facility

**Justification for Model Selection**

To forecast bus passengers at each facility from 2025 to 2030, we employed the ETS (Error, Trend, Seasonality) model. This time-series forecasting approach effectively captures long-term trends, seasonal variations, and fluctuations in transit usage, making it well-suited for public transportation demand forecasting.

The ETS model was chosen for several key reasons:

* **Trend Modeling**: Public transportation demand evolves over time due to changes in commuter behavior, infrastructure developments, and policy shifts. The ETS model effectively captures these gradual increases or decreases in ridership.
* **Seasonality Handling**: Transit usage fluctuates based on factors such as weather, holidays, and changes in travel patterns. The ETS model decomposes time-series data into error, trend, and seasonality components, ensuring accurate recognition of recurring patterns.
* **Adaptability to External Shocks**: The COVID-19 pandemic significantly impacted ridership trends, particularly in 2020. To prevent biases from pandemic-induced anomalies, we excluded data from March, April, May, June, November, and December 2020 when training the model. This approach helps the ETS model focus on long-term trends rather than short-term disruptions.

**ETS Model in Industry Applications**

The ETS model is widely utilized in various industries for predictive analytics and strategic decision-making:

* **Public Transportation & Urban Planning**: Transit authorities and city planners use ETS models to predict ridership trends, optimize schedules, and allocate resources efficiently. For example, Transport for London (TfL) applies time-series forecasting to assess demand variations and improve service reliability.
* **Retail Demand Forecasting**: Companies such as Walmart and Amazon use ETS models to forecast sales, optimize inventory, and prepare for seasonal demand shifts. By leveraging historical sales data, ETS helps in reducing stockouts and overstock issues.
* **Energy Sector & Load Forecasting**: Utilities use ETS to predict electricity demand, considering seasonal consumption patterns and long-term demand trends. This enables better grid management and load balancing.
* **Financial Services**: Banks and investment firms apply ETS models to forecast financial trends, such as stock prices or market demand for specific financial products. The ability to capture both trend and seasonality makes ETS valuable in portfolio risk assessment.

Given these industry applications, the ETS model’s ability to deliver accurate and interpretable forecasts makes it an optimal choice for predicting future bus ridership patterns and aiding transit system decision-making.

**Independent and Dependent Variables**

The dependent variable for our model is the number of buses operating at each facility per month. This choice is based on the assumption that bus supply is a key determinant of passenger flow, as transit agencies adjust fleet size to meet demand. To enhance the model’s predictive accuracy, we incorporated several independent variables that influence bus operations and ridership patterns:

* **Date (Year, Month)**: Essential for capturing long-term trends and seasonal variations.
* **AWND (Average Wind Speed in mph)**: High wind speeds can disrupt transit operations and affect passenger decisions.
* **PRCP (Precipitation in inches)**: Rain can discourage travel and influence ridership levels.
* **SNOW (Snowfall in inches) & SNWD (Snow Depth in inches)**: Snow accumulation affects bus accessibility and overall commuting behavior.
* **TMAX (Maximum Temperature in °F) & TMIN (Minimum Temperature in °F)**: Extreme temperatures can influence ridership, with cold weather discouraging travel and warm temperatures increasing outdoor mobility.
* **PassengersPerBus**: This variable, calculated separately using SQL, provides an estimate of the number of passengers per bus, which is crucial for translating bus forecasts into total passenger estimates.

**SQL Query: Calculating the Average Number of Passengers per Bus**

To obtain an accurate measure of passengers per bus, we executed the following SQL query after importing into SQL the MBT\_Passenger\_Departures table and the MBT\_Bus\_Departures table from the Unit\_571\_database:

SELECT

MONTH(p.column2) AS DepartureMonth,

AVG(p.column7 / NULLIF((b.column8 + b.column9 + b.column10 + b.column11 + b.column12 + b.column13 + b.column14), 0)) AS AvgPassengersPerBus

FROM

MBT\_Passenger\_Departures p

JOIN

MBT\_Bus\_Departures b ON p.column1 = b.column1

GROUP BY

MONTH(p.column2)

ORDER BY

MONTH(p.column2);

This query calculates the average number of passengers per bus for each month. The NULLIF() function prevents division by zero when summing up bus departures, ensuring robust data integrity. The results from this query were used to estimate total passenger volume for each facility by multiplying the forecasted number of buses by the average number of passengers per bus.

**R Code for Forecasting Using the ETS Model**

The forecasting process was implemented in R, where we structured the workflow into key steps: data preprocessing, model training, forecasting, and result aggregation.

# Load necessary libraries

library(dplyr)

library(lubridate)

library(forecast)

library(ggplot2)

# Step 1: Load and prepare data

bus\_data <- read.csv("BusesPerFacility.csv")

bus\_data <- bus\_data %>% mutate(Date = as.Date(paste(Year, Month, "01", sep = "-")))

# Step 2: Calculate monthly weather averages to be used as independent variables

weather\_avgs <- bus\_data %>%

group\_by(Facility, Month) %>%

summarise(

AWND = mean(AWND, na.rm = TRUE),

PRCP = mean(PRCP, na.rm = TRUE),

SNOW = mean(SNOW, na.rm = TRUE),

SNWD = mean(SNWD, na.rm = TRUE),

TMAX = mean(TMAX, na.rm = TRUE),

TMIN = mean(TMIN, na.rm = TRUE),

.groups = "drop"

)

# Step 3: Exclude COVID-19 months to prevent bias in forecasting

bus\_data\_forecast <- bus\_data %>%

filter(!(Year == 2020 & Month %in% c(3, 4, 5, 6, 11, 12)))

# Step 4: Train and apply the ETS model for each facility

facilities <- unique(bus\_data$Facility)

forecast\_tables <- list()

for (fac in facilities) {

facility\_data <- bus\_data\_forecast %>% filter(Facility == fac) %>% arrange(Date)

ts\_data <- ts(facility\_data$Buses, start = c(facility\_data$Year[1], facility\_data$Month[1]), frequency = 12)

if (min(ts\_data) <= 0) {

ts\_data <- ts\_data + abs(min(ts\_data)) + 1

}

model <- ets(ts\_data, model = "AAN") # AAN denotes additive error, additive trend, no seasonality

forecast\_horizon <- interval(ymd("2024-06-01"), ymd("2030-12-01")) %/% months(1) + 1

fc <- forecast(model, h = forecast\_horizon)

future\_dates <- seq.Date(from = as.Date("2024-06-01"), by = "month", length.out = forecast\_horizon)

future\_months <- month(future\_dates)

future\_years <- year(future\_dates)

weather\_future <- data.frame(Facility = fac, Month = future\_months) %>%

left\_join(weather\_avgs %>% filter(Facility == fac), by = c("Facility", "Month"))

forecasted\_buses <- as.integer(abs(fc$mean))

fc\_df <- data.frame(

Year = future\_years,

Month = future\_months,

Facility = fac,

AWND = weather\_future$AWND,

PRCP = weather\_future$PRCP,

SNOW = weather\_future$SNOW,

SNWD = weather\_future$SNWD,

TMAX = weather\_future$TMAX,

TMIN = weather\_future$TMIN,

Buses = forecasted\_buses

)

forecast\_tables[[as.character(fac)]] <- fc\_df

}

# Step 5: Combine forecasted results and estimate total passengers

final\_forecast <- bind\_rows(forecast\_tables)

passengers\_data <- read.csv("PassengersPerBus.csv")

final\_forecast <- final\_forecast %>% left\_join(passengers\_data, by = "Month")

final\_forecast <- final\_forecast %>% mutate(TotalPassengers = Buses \* PassangersPerBus)

# Step 6: Save the forecasted dataset

write.csv(final\_forecast, "Forecasted\_Buses\_With\_Passengers.csv", row.names = FALSE)

This process ensures that our forecasted bus counts incorporate meaningful external factors and that total passenger estimates align with historical occupancy patterns. The ETS model’s ability to adapt to post-pandemic shifts makes it a robust choice for predicting transit trends, supporting the Port Authority in optimizing resource allocation for 2025–2030.

## 3) Model 2: Prophet model to project the results by individual carrier

**Justification for Model Selection**

The Prophet model was selected due to its strong ability to handle seasonality and trends in time-series data. Prophet, developed by Facebook, is specifically designed for time series data that exhibit trends, seasonality, and holidays. Unlike traditional methods, Prophet automatically detects trends and seasonality components in data, making it especially suitable for data with irregular patterns, such as ridership numbers for different bus companies. The model's flexibility in handling holidays, different seasonality, and trend changes over time makes it particularly useful for transportation-related forecasting tasks, where external factors like weather or events can impact ridership.

**Prophet Model in Industry Applications**

The Prophet model is increasingly used in industries like transportation, retail, finance, and e-commerce for forecasting demand, sales, or traffic. In transportation, Prophet has been employed to forecast passenger ridership, vehicle traffic, and demand for various transport services. Its ability to handle multiple seasonal effects, holidays, and trend changepoints has made it a go-to tool for forecasting traffic data, particularly for systems with complex, cyclical patterns. In transportation, where the demand often fluctuates based on external factors like holidays, special events, and weather, Prophet allows organizations to make informed decisions regarding staffing, fleet management, and schedule planning.

**Independent and Dependent Variables**

In the forecasting process for bus ridership, the independent variable is the time component, which is represented as weekly intervals starting from January 2020. This weekly time series is common across all bus companies, serving as the reference for temporal trends.

The dependent variable, which we aim to forecast, is the ridership count for each bus carrier over time. Each bus company in the dataset—such as Academy, Greyhound, Martz, NJ Transit, Lakeland, Trailways, Coach USA, TransBridge, Peter Pan/Bonanza, DeCamp, and C&J Bus Lines—has its own time series representing the number of passengers using their services each week. The ridership counts are influenced by factors such as the seasonality of the year (e.g., holidays, summer, winter), special events (e.g., festivals, sporting events), and external factors like changes in transportation policies or the COVID-19 pandemic, which affect travel patterns.

**Description of the Dataset Used for Forecasting**

The dataset used for forecasting consists of weekly ridership data for several bus companies. The data includes ridership counts for various companies such as Academy, Greyhound, Martz, NJ Transit, Lakeland, Trailways, Coach USA, TransBridge, Peter Pan/Bonanza, DeCamp, and C&J Bus Lines. The dataset covers ridership numbers from the beginning of 2020, and forecasts are generated for the following 6 years (through 2030). The dataset is organized with time in the form of weekly time intervals, and each bus company has its ridership data (passenger counts) for each week.

**R Code for Forecasting Using the Prophet Model**

The following R code was used to fit the Prophet model to the ridership data for each bus company, forecast future ridership, and generate results for each bus company's forecasted ridership over the next 6 years.

library(readxl)

library(tsibble)

library(fable)

library(dplyr)

library(lubridate)

library(tidyr)

library(forecast)

library(prophet)

library(readr) # For exporting CSV

# Load the Excel data

file\_path <- "C:/Users/Irfan/Downloads/Data (1).xlsx"

df <- read\_excel(file\_path, sheet = "Sheet1")

# Extract each bus company's passenger data

bus\_data <- list(

"Academy" = df$Academy,

"Greyhound" = df$Greyhound,

"Martz" = df$Martz,

"NJ Transit" = df$`NJ Transit`,

"Lakeland" = df$Lakeland,

"Trailways" = df$Trailways,

"Coach USA" = df$`Coach USA`,

"TransBridge" = df$TransBridge,

"Peter Pan/Bonanza" = df$`Peter Pan/Bonanza`,

"DeCamp" = df$DeCamp,

"C & J Bus Lines" = df$`C & J Bus Lines`)

# Create an empty dataframe to store combined data

full\_data\_with\_forecast <- data.frame(ds = seq(as.Date("2020-01-05"), by = "week", length.out = nrow(df) + 312))

# Function to fit Prophet model and forecast

fit\_forecast\_merge\_prophet <- function(bus\_pass\_data) {

# Ensure numeric values

bus\_pass\_data <- as.numeric(bus\_pass\_data)

# Handle missing values (replace with median)

bus\_pass\_data[is.na(bus\_pass\_data)] <- median(bus\_pass\_data, na.rm = TRUE)

# Create a dataframe for Prophet

df\_prophet <- data.frame(

ds = seq(as.Date("2020-01-05"), by = "week", length.out = length(bus\_pass\_data)),

y = bus\_pass\_data

)

# Fit Prophet model (without visualization)

m\_prophet <- prophet(df\_prophet,

yearly.seasonality = TRUE,

weekly.seasonality = TRUE,

daily.seasonality = FALSE,

changepoint.prior.scale = 0.1)

# Create future dataframe (forecasting for 312 weeks)

future <- make\_future\_dataframe(m\_prophet, periods = 312, freq = "week")

# Generate forecast

forecast\_prophet <- predict(m\_prophet, future)

# Extract forecasted values

forecast\_values <- forecast\_prophet$yhat[(nrow(forecast\_prophet)-311):nrow(forecast\_prophet)]

# Merge historical and forecasted data

return(c(bus\_pass\_data, forecast\_values))

}

# Loop through each bus company and store merged data

for (bus\_name in names(bus\_data)) {

full\_data\_with\_forecast[[bus\_name]] <- fit\_forecast\_merge\_prophet(bus\_data[[bus\_name]])

}

# Save the merged dataset to CSV

csv\_path <- "C:/Users/Irfan/Downloads/Original\_And\_Forecasted\_Buses\_With\_Passengers\_ConsideringCovid.csv"

write\_csv(full\_data\_with\_forecast, csv\_path)

print(paste("Full data with forecasts saved at:", csv\_path))

VIZ CODE-

# Academy

ggplot(df, aes(x = ds, y = Academy)) +

geom\_line(color = "steelblue") +

geom\_point(color = "darkred") +

labs(title = "Academy - Passengers Over Time", x = "Date", y = "Passengers") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))+theme\_fivethirtyeight()

# Greyhound

ggplot(df, aes(x = ds, y = Greyhound)) +

geom\_line(color = "steelblue") +

geom\_point(color = "darkred") +

labs(title = "Greyhound - Passengers Over Time", x = "Date", y = "Passengers") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

# Martz

ggplot(df, aes(x = ds, y = Martz)) +

geom\_line(color = "steelblue") +

geom\_point(color = "darkred") +

labs(title = "Martz - Passengers Over Time", x = "Date", y = "Passengers") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

# NJ Transit

ggplot(df, aes(x = ds, y = `NJ Transit`)) +

geom\_line(color = "steelblue") +

geom\_point(color = "darkred") +

labs(title = "NJ Transit - Passengers Over Time", x = "Date", y = "Passengers") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

# Lakeland

ggplot(df, aes(x = ds, y = Lakeland)) +

geom\_line(color = "steelblue") +

geom\_point(color = "darkred") +

labs(title = "Lakeland - Passengers Over Time", x = "Date", y = "Passengers") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

# Trailways

ggplot(df, aes(x = ds, y = Trailways)) +

geom\_line(color = "steelblue") +

geom\_point(color = "darkred") +

labs(title = "Trailways - Passengers Over Time", x = "Date", y = "Passengers") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

# Coach USA

ggplot(df, aes(x = ds, y = `Coach USA`)) +

geom\_line(color = "steelblue") +

geom\_point(color = "darkred") +

labs(title = "Coach USA - Passengers Over Time", x = "Date", y = "Passengers") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

# TransBridge

ggplot(df, aes(x = ds, y = TransBridge)) +

geom\_line(color = "steelblue") +

geom\_point(color = "darkred") +

labs(title = "TransBridge - Passengers Over Time", x = "Date", y = "Passengers") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

# Peter Pan/Bonanza

ggplot(df, aes(x = ds, y = `Peter Pan/Bonanza`)) +

geom\_line(color = "steelblue") +

geom\_point(color = "darkred") +

labs(title = "Peter Pan/Bonanza - Passengers Over Time", x = "Date", y = "Passengers") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

# DeCampABS

ggplot(df, aes(x = ds, y = DeCampABS)) +

geom\_line(color = "steelblue") +

geom\_point(color = "darkred") +

labs(title = "DeCamp ABS - Passengers Over Time", x = "Date", y = "Passengers") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

# C & J Bus Lines ABS

ggplot(df, aes(x = ds, y = `C & J Bus Lines ABS`)) +

geom\_line(color = "steelblue") +

geom\_point(color = "darkred") +

labs(title = "C & J Bus Lines ABS - Passengers Over Time", x = "Date", y = "Passengers") +

theme\_minimal() +

theme(axis.text.x = element\_text(angle = 45, hjust = 1))

## 4) Model 3: ARIMA (Autoregressive Integrated Moving Average) model to forecast the busiest year, month and week for each facility

**Justification for Model Selection**

To forecast the busiest year, month, and week for each facility from 2025 to 2030, we employed the ARIMA (Autoregressive Integrated Moving Average) model. This time-series forecasting approach is well-suited for analyzing historical trends and making short- to medium-term predictions based on past patterns in facility usage.

The ARIMA model was selected for several key reasons:

* **Capturing Weekly, Monthly, and Yearly Trends:** Our dataset includes total vehicle traffic data for each facility from 2015 to 2024, segmented by week, month, and year. ARIMA is highly effective at detecting long-term patterns in such structured time-series data.
* **Predicting Traffic Based on Historical Data:** Since ARIMA relies on past observations to make forecasts, it is an ideal choice for our dataset, where historical traffic volumes provide strong indicators of future trends.
* **Ensuring Stability in Forecasts:** By applying different techniques, ARIMA helps stabilize the dataset, reducing the impact of irregular fluctuations while preserving meaningful trends.
* **Reliable Five-Year Forecasting:** With nearly a decade of historical data, ARIMA is well-suited for making predictions for 2025-2030, allowing for data-driven planning and optimization of facility capacity.

**ARIMA Model in Industry Applications**

The ARIMA model is widely used across various industries for predictive analytics and trend forecasting:

* **Transportation and Urban Planning:** Transit agencies often leverage ARIMA to forecast passenger demand and optimize resource allocation. For instance, public transit systems have used similar forecasting models to enhance scheduling and improve service efficiency.
* **Retail Demand Prediction:** Large retailers apply ARIMA-based methods to estimate sales trends and manage inventory levels, ensuring a balanced supply chain.
* **Energy Consumption Forecasting:** Power grid operators use ARIMA to predict electricity demand, aiding in efficient energy distribution and load balancing.
* **Financial Market Analysis:** Banks and investment firms employ ARIMA for time-series modeling stock price movements, helping to identify potential market trends.

These examples illustrate ARIMA’s adaptability in analyzing time-dependent data, making it a suitable choice for predicting peak facility usage.

**Independent and Dependent Variables**

The dependent variable for our model is TotalTraffic, representing the total number of buses using each facility per week. This metric serves as the foundation for identifying peak usage trends. The independent variables considered in the ARIMA forecasting model include:

* Date (Year, Month, Week): Used to structure time-series data for accurate trend analysis.
* TotalTraffic: The dependent variable that represents facility traffic over time.

By utilizing ARIMA, we generated forecasts for the busiest week, month, and year at each facility between 2025 and 2030, assisting transit planners in optimizing operations and capacity planning.

**R code for forecasting the busiest year, month and week for each facility**

library(ggplot2)

library(dplyr)

library(lubridate)

# Load the dataset

data = read.csv("C:/Users/acer/OneDrive/Desktop/Baruch/BusiestTimesDataset.csv")

head(data)

data$Date <- as.Date(data$Date)

head(data)

# Extract year from the Date column

data$Year <- year(data$Date)

# Separate existing 2024 data

data\_2024\_existing <- filter(data, Year == 2024)

# Extract 2023 data for April - December (to fill in 2024)

data\_2023\_missing\_months <- filter(data, Year == 2023 & Month %in% c("April", "May", "June", "July", "August", "September", "October", "November", "December"))

# Generate missing 2024 data (April-Dec) by increasing 2023 traffic by 5%

data\_2024\_new <- data\_2023\_missing\_months %>%

mutate(

TotalTraffic = round(TotalTraffic \* 1.05), # Apply 5% increase

Year = 2024,

Month\_Num = match(Month, month.name), # Convert month names to numbers

Date = as.Date(paste(2024, Month\_Num, day(Date), sep = "-"))

)

# Combine all years (2014-2024) + new 2024 rows (April- December rows)

final\_data <- bind\_rows(data, data\_2024\_new) %>%

select(-Month\_Num) # Remove temporary Month\_Num column

# Save the final dataset as an Excel file

write\_xlsx(final\_data, "Complete\_BusiestTimeDataset.xlsx")

# Confirm file saved successfully

print("Excel file saved: Complete\_BusiestTimeDataset.xlsx")

library(forecast)

library(dplyr)

library(tidyr)

# Convert Year, Month, and Week to a proper date format

data$Date <- as.Date(paste(data$Year, data$Month, data$Week, 1, sep = "-"), format = "%Y-%m-%d")

# Aggregate total traffic per Facility, Week, Month, and Year

agg\_data <- final\_data %>%

group\_by(Year, Month, Week,Facility) %>%

summarize(TotalTraffic = sum(TotalTraffic), .groups = "drop")

# Forecast function for each Facility (forecasting for 2025-2030)

forecast\_facility <- function(df) {

ts\_data <- ts(df$TotalTraffic, start = c(min(df$Year), min(df$Week)), frequency = 52) # Weekly frequency

# Fit ARIMA model

model <- auto.arima(ts\_data)

# Forecast for the period between 2025 and 2030

start\_year <- 2025

end\_year <- 2030

weeks\_in\_year <- 52

total\_weeks <- (end\_year - start\_year + 1) \* weeks\_in\_year

forecasted\_values <- forecast(model, h = total\_weeks)

# Prepare future weeks and years

future\_weeks <- seq(1, total\_weeks, by = 1)

future\_years <- rep(start\_year, total\_weeks)

# Adjust year and week for the forecast horizon

for (i in 1:total\_weeks) {

year\_shift <- floor((i - 1) / weeks\_in\_year)

future\_years[i] <- start\_year + year\_shift

future\_weeks[i] <- ((i - 1) %% weeks\_in\_year) + 1

}

# Create a Date column from the Year and Week (assuming Week starts on a Monday)

future\_dates <- as.Date(paste(future\_years, future\_weeks, 1, sep = "-"), format = "%Y-%U-%u")

# Extract Month Name from Date

future\_month\_names <- format(future\_dates, "%B")

return(data.frame(

Facility = df$Facility[1],

Year = future\_years,

Week = future\_weeks,

month= future\_month\_names,

Forecasted\_Traffic = as.numeric(forecasted\_values$mean)

))

}

# Apply forecasting function to each Facility

forecast\_results <- agg\_data %>%

group\_by(Facility) %>%

group\_split() %>%

lapply(forecast\_facility) %>%

bind\_rows()

# Load the package

library(openxlsx)

# Write the data to an Excel file

write.xlsx(forecast\_results, "busiesttime.xlsx")

# View forecasted values

print(forecast\_results)

## 5) Pragmatic Applications: Addressing Corporate Questions Directly

The insights generated from our forecasting models offer practical applications for the Port Authority of NY and NJ, enabling data-driven decision-making across multiple operational areas. First, by accurately predicting future ridership trends, the Port Authority can optimize bus scheduling and resource allocation, ensuring adequate capacity during peak hours while minimizing inefficiencies during off-peak times. Additionally, insights into seasonal fluctuations and weather-related impacts allow for proactive planning, such as adjusting routes or preparing contingency measures for extreme weather conditions.

The integration of our interactive dashboard will further enhance the usability of our findings. The Port Authority can dynamically explore data trends, compare pre- and post-pandemic usage patterns, and identify anomalies or emerging shifts in travel behavior. This tool is particularly valuable for long-term infrastructure planning, as it provides a clear visualization of future demand at each facility. Furthermore, understanding the busiest periods for each facility allows for strategic staffing decisions, improved crowd management, and enhanced passenger experience.

By leveraging the outputs from the ETS, Prophet, and ARIMA models, the Port Authority can address key corporate questions such as:

* How will ridership evolve at each facility in the coming years?
* What are the differences in demand across various carriers, and how should service adjustments be planned?
* Which years, months or weeks will experience the highest traffic volume, and how can operations be adjusted accordingly?

Through these applications, our forecasting models provide actionable insights, allowing the Port Authority to enhance operational efficiency, allocate resources strategically, and improve overall transit system reliability.

## 6) Individual Contributions: Detailed Breakdown of Work

**Giacomo Bizzotto:** Assisted in gathering the requirements and datasets for the Port Authority New Jersey-New York project. Collaborated with the team to collect relevant data on port activity, infrastructure, and regulations. Contributed to ensuring the project had the necessary information for informed decision-making and planning. Overviewed everything and assigned the roles to each member of the group that was present at the meetings. Designed the dataset using SQL and Excel used in the first forecasting model, created the latter and the visualization to gain insights and answers. Written the entire report apart from point 3 and 4.

**Rishitha Thatipally:** Assisted in gathering the requirements and datasets for the Port Authority New Jersey-New York project. Collaborated with the team to collect relevant data on port activity, infrastructure, and regulations. Contributed to ensuring the project had the necessary information for informed decision-making and planning. Worked in the visualization on Power BI to make it look better.

**Irfan Shaik:** Assisted in gathering the requirements and datasets for the Port Authority New Jersey-New York project. Collaborated with the team to collect relevant data on port activity, infrastructure, and regulations. Contributed to ensuring the project had the necessary information for informed decision-making and planning. Designed the dataset used in the second forecasting model, created the latter and the visualization to gain insights. Written the third point of this report.

**Errolla Vivek:** Assisted in gathering the requirements and datasets for the Port Authority New Jersey-New York project. Collaborated with the team to collect relevant data on port activity, infrastructure, and regulations. Contributed to ensuring the project had the necessary information for informed decision-making and planning. Helped to develop the data set for the second forecasting model.

**Shraddha Shakya:** Assisted in gathering the requirements and datasets for the Port Authority New Jersey-New York project. Collaborated with the team to collect relevant data on port activity, infrastructure, and regulations. Contributed to ensuring the project had the necessary information for informed decision-making and planning. Designed the dataset used in the third forecasting model, created the latter and the visualization to gain insights. Written the fourth point of this report.

**Sampath Sai Raghav Allaboyina:** Assisted in gathering the requirements and datasets for the Port Authority New Jersey-New York project. Collaborated with the team to collect relevant data on port activity, infrastructure, and regulations. Contributed to ensuring the project had the necessary information for informed decision-making and planning. Helped Giacomo in coordinating everything, supported the group in the development of the dataset for the first and the second model. Tried to develop some alternative forecasting model using phyton. Reviewed the Report before the submission.

## 7) Conclusion: Summary of Findings and Recommendations

This report outlines the methodologies employed to forecast bus passenger numbers and traffic patterns for the Port Authority. By utilizing the ETS, Prophet, and ARIMA models, we aim to provide actionable insights that will support strategic decision-making regarding resource allocation, operational efficiency, and infrastructure planning.

Our analysis, supported by the preliminary visualizations from our developing Power BI dashboard, reveals the following key findings:

* **Overall Passenger Trends:** The ETS model forecasts a stabilization in overall passenger numbers from 2025 to 2030, with projected figures reaching approximately 155 million passengers. This result suggests a growing demand for bus terminal facilities, necessitating proactive capacity planning.
* **Individual Carrier Variations:** The Prophet model enables a deeper understanding of passenger trends at the individual carrier level. Visualizations highlight diverse patterns among carriers, with some experiencing steady growth and others exhibiting more fluctuations. These differences in demand trajectories necessitate tailored service adjustments and resource allocation strategies for each carrier.
* **Busiest Times of Operation:** The ARIMA model helps identify peak periods of traffic volume. Initial visualizations indicate variations in traffic volume across weeks and months. Further refinement of this analysis will enable the Port Authority to anticipate and prepare for the busiest times, optimizing staffing and crowd management strategies.
* **Comparison to 2019:** The graph titled “Passengers by Year” shows a clear trend in bus terminal usage over time. From 2013 to 2019, passenger numbers remained stable at approximately 205 million per year. However, in 2020, coinciding with the onset of the COVID-19 pandemic, there was a sharp and dramatic decline to 132 million, followed by a further drop to 130 million in 2021. After this steep decline, the trend reverses slightly. From 2022 onward, the number of passengers begins to gradually increase, reaching 155 million by 2025, where it stabilizes and remains constant through 2030. While the data shows signs of recovery, it’s clear that bus terminal usage is not projected to return to pre-pandemic levels. Instead, it is expected to plateau at about 50 million passengers below the 2019 peak, suggesting a potential long-term shift in commuting patterns, travel behavior, or service structure.

**Recommendations:**

Based on these findings, we recommend the following:

* **Capacity Planning:** The Port Authority should consider the projected increase in overall passenger numbers when planning future facility capacity and infrastructure developments.
* **Carrier-Specific Strategies:** Tailored service adjustments and resource allocation strategies should be developed for individual carriers, taking into account their unique demand patterns.
* **Peak Period Management:** Staffing and crowd management strategies should be optimized to accommodate the busiest weeks and months, ensuring a smooth and efficient passenger experience.

By implementing these recommendations, the Port Authority can effectively leverage our forecasting models to enhance operational efficiency, optimize resource allocation, and improve the overall transit system for the benefit of both passengers and stakeholders.

## 8) Appendix: Supplementary Materials

A graph showing the growth of the year

AI-generated content may be incorrect.A graph showing the growth of passengers over time

AI-generated content may be incorrect.In this section, we will include images of the Power BI dashboard we are developing. This dashboard will be a key tool for the company, allowing for flexible data exploration and analysis. While the current design is still in its early stages, our goal for the next report is to present a solution that is not only highly functional but also professionally designed and visually polished.

A graph showing the arrival of passengers

AI-generated content may be incorrect.A graph showing the number of signals

AI-generated content may be incorrect.A graph showing the growth of passengers

AI-generated content may be incorrect.A graph showing the growth of passengers over time

AI-generated content may be incorrect.A graph showing the growth of passengers

AI-generated content may be incorrect.A graph showing the growth of passengers

AI-generated content may be incorrect.A graph showing the growth of passengers

AI-generated content may be incorrect.A graph showing a number of passengers

AI-generated content may be incorrect.A graph showing passengers over time

AI-generated content may be incorrect.A graph showing a graph of passengers over time

AI-generated content may be incorrect.

A graph of blue lines

AI-generated content may be incorrect.

A screenshot of a graph

AI-generated content may be incorrect.

A graph of blue bars

AI-generated content may be incorrect.