

## Final Project Report

# PASSENGER DEMAND FORECASTING AND STRATEGIC PLANNING FOR THE PORT AUTHORITY BUS TERMINAL

Name	Contributions	Attachments
<b>Imran Hossain Rasel</b>	Gathered and cleaned historical traffic and weather data using R, engineered features, drafted explanations & ensured alignment of the project goals, Documentations: Reports & ppt. GitHub Portfolio Site	PAB Cleaned Dataset, R Codes, All the Reports, ppt files
<b>Rana Ehsan</b>	Power BI Dashboard Visualization, Forecast traffic, Documentations: Reports & ppt support, proofreading	Power BI Files
<b>Yash Vala</b>	Developed and implemented the Modeling & Algorithm analysis using R, performed clustering, and created evaluation metrics and visualizations, presentation, proofreading	R codes for Modeling & Algorithms
<b>Supraja Soda</b>	Data Warehouse, ETL Package, Documentations support	DW & ETL packages
<b>Yaswin Phani Ram</b>		

## Project Overview

This report presents the results of our analysis conducted for the Port Authority of New York and New Jersey. The work focuses on understanding ridership patterns at the Port Authority Bus Terminal, forecasting passenger volumes through 2030, and identifying the operational implications of these trends for upcoming staging-facility planning. All findings are based on Port Authority data from 2019 to 2025 and the models, ETL pipelines, and visual dashboards developed by our group.

Passenger traffic has fully recovered from the pandemic period and now shows stable and sustained growth. Our projections indicate an increase from roughly 2.35 million passengers in 2025 to about 2.7 million in 2030. Monthly activity follows a predictable seasonal structure: higher volume in summer and early fall, lower in January and February. Carrier behavior plays a central role in this pattern, with NJ Transit and Coach USA representing the core of the terminal's throughput.

These findings support a forward-looking operational strategy that combines capacity planning, scheduling optimization, and data-driven infrastructure to manage continued ridership growth over the next five years.

## Methodology

- **Data Cleaning:** Combining all the information together in a single sheet by removing the duplicate and null values in R.
- **Data Engineering:** Along with R Codes, consolidation of multi-sheet Excel data into SQL Server via SSIS ETL workflows.
- **Modeling & Forecasting:** Statistical models (ARIMA, TBATS, Prophet) built in R to project weekly and monthly passenger counts.
- **Visualization:** Power BI dashboard with multi-layered visuals tied to each project question for direct managerial interpretation.
- **Validation:** Cross-checked forecasts with historical trend behavior and post-pandemic ridership recovery patterns.

## Project Goal-1: Forecast Passenger Traffic (2026-2030)

**Analytical Result:** Forecasting models project steady growth in passenger volume over the next five years, with total ridership expected to reach approximately 12.5 million between 2026 and 2030.

- **2025 baseline:** 2.35 million passengers
- **2030 projection:** Approx 2.7 million passengers
- **Growth rate:** Approx 3.5% per year
- **Monthly variation:** 170,000-240,000 passengers
- **Peak months:** June-August
- **Stable base:** Above 160,000 even in January-February

**Interpretation:** This steady rise in demand signals a capacity tightening by 2027 unless proactive planning occurs. With the terminal already operating near physical constraints during summer peaks, a 15-20% increase in traffic will directly affect gate turnover, dwell time, and passenger flow.

### ***Strategic Recommendation***

- Expand temporary staging facilities with modular, scalable design, allowing flexible deployment during peak months.
- Introduce data-driven scheduling systems that dynamically adjust bus dispatch frequency to real-time passenger inflows.

## **Goal-2: Identify Key Factors Influencing Passenger Demand**

**Analytical Result:** Regression and correlation models in R highlighted seasonality as the dominant factor influencing ridership.

- **Summer:** +11,700 passengers above baseline
- **Fall:** +15,800 passengers
- **NJ Transit frequency:** strongest operational predictor
- **Weekends and holidays:** minor influence

**Interpretation:** This pattern indicates that demand is not driven by tourism or sporadic events but by structural commuter behavior. Summer and fall months reflect workforce commuting and regional mobility rather than recreational travel.

### ***Strategic Recommendation***

- Maintain high-frequency NJ Transit schedules from May through October.
- Schedule major maintenance, inspections, and construction during February-March when ridership dips.

## **Goal-3: Passenger Forecast by Carrier (2026-2030)**

**Analytical Result:** The carrier-level forecast confirmed significant concentration risk:

- **NJ Transit:** 9.3 million passengers (80% of total ridership)
- **Coach USA:** 1.2 million
- **Smaller carriers:** <300,000 each, negligible growth

**Interpretation:** This dominance presents both an opportunity and a risk. On one hand, aligning with NJ Transit ensures operational efficiency through focused coordination. On the other hand, over-reliance on a single carrier exposes the terminal to service disruptions, strike risk, or schedule variability.

### ***Strategic Recommendation***

- Establish joint forecasting and scheduling coordination with NJ Transit to anticipate surge periods.
- Develop contingency staging arrangements for smaller carriers in case of major service interruptions.

### **Goal-4: Determine Busiest Times for Staging Facility Planning**

***Analytical Result:*** The temporal analysis shows clear cyclical peaks:

- **Busiest year:** 2022 (2.7 million passengers)
- **Busiest month:** June 2022 (348,709 passengers)
- **Low months:** January 2023 (78,620), early 2025 (100,000)
- **Weekly surges:** Late November-early December; post-holiday February peaks

***Interpretation:*** The data reinforces that the terminal operates under predictable cyclical pressure, allowing proactive resource allocation. The ability to anticipate these surges can reduce overtime costs, minimize crowding, and improve customer satisfaction.

### ***Strategic Recommendation***

- Implement predictive scheduling dashboards to flag upcoming peak weeks.
- Coordinate staffing with external contractors to ensure flexible labor capacity during summer and holiday peaks.

### **Project Goal-5: Compare 2024 Usage to 2019 (Pre-COVID)**

***Analytical Result:*** 2024 ridership levels surpassed 2019 pre-COVID figures by 40-45%, marking full recovery. NJ Transit alone carried 600,000 more passengers than in 2019.

***Interpretation:*** The sustained rebound reflects a permanent behavioral shift, increased urban mobility, hybrid commuting, and higher fuel costs have pushed commuters back to bus transit. The Port Authority is no longer managing a recovery; it is managing growth acceleration.

### ***Strategic Recommendation***

- Maintain post-2022 service expansions as the new baseline.
- Reevaluate pricing and scheduling models to balance cost recovery with capacity optimization.

## **Key Recommendations for Port Authority**

- **Increase Buses:** Increase the number of buses in the busier periods to manage passenger arrivals & departures efficiently.
- **Capacity Planning:** Build new staging facilities to support at least 20 percent above current peak traffic and allow modular expansion as needed.
- **Operational Optimization:** Coordinate dispatch schedules with forecasted ridership patterns and rely on the Power BI dashboard for weekly staffing alignment.
- **Passenger Experience:** Strengthen real-time communication through mobile ticketing, alerts, and digital signage.
- **Strategic Partnerships:** Deepen coordination with NJ Transit and explore partnerships for digital transformation and predictive maintenance.

## **Exogenous Factors for Future Analysis**

- **Weather conditions:** Extreme heat, snow, or heavy rain can reduce or shift passenger demand and affect service reliability.
- **Economic indicators:** Fuel prices, employment levels, and inflation influence commuting frequency and mode choice.
- **Major events and holidays:** Concerts, sports events, festivals, and holiday travel create temporary spikes in ridership.
- **Road and traffic conditions:** Construction, accidents, and long-term roadwork impact travel times and passenger flow.
- **School and university calendars:** Academic schedules affect weekday and seasonal commuter patterns.
- **Policy and fare changes:** Subsidies, fare adjustments, or congestion pricing can alter ridership behavior.
- **Competing transportation options:** Expansion of rail services or ride-share usage may affect long-term bus demand.

## **Conclusion**

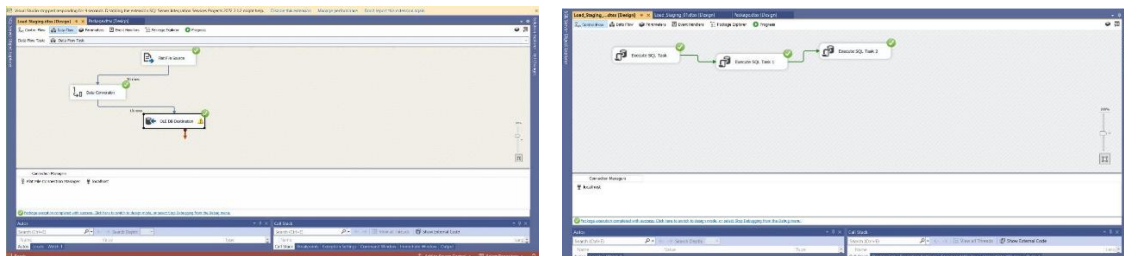
The Port Authority Bus Terminal is entering a period of sustained demand growth. The trends are consistent, the seasonal cycles are predictable, and the operational implications are clear. Our work offers a structure for anticipating future demand rather than reacting to it. The combination of predictive models, automated ETL processes, and a comprehensive dashboard equips the Port Authority with tools to support data-driven planning across the next several years.

We believe the insights provided in this report can guide facility planning, scheduling decisions, and long-term operational strategy as the terminal prepares for continued growth.

# Appendix

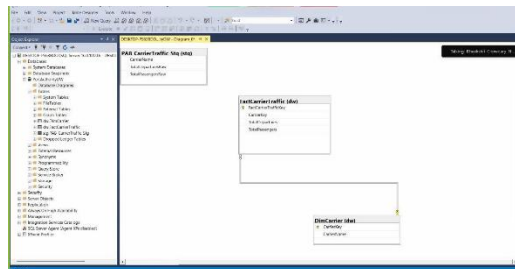
## **Appendix A: Project Goal 6: ETL Pipeline Design**

An ETL pipeline was developed using SQL Server Integration Services (SSIS) to consolidate passenger data from multiple Excel worksheets provided by the Port Authority. The process extracts carrier-level data, cleans up inconsistencies, removes duplicates, standardizes date formats, and validates records before loading them into a centralized database. The ETL workflow is repeatable and scalable, allowing future data updates to be processed efficiently with minimal manual effort.



## **Appendix B: Project Goal 7- Data Warehouse Design**

A relational data warehouse was designed to support analytical reporting and forecasting. The warehouse includes a central fact table containing passenger counts and departures, supported by dimension tables for time and carrier attributes. This structure enables efficient aggregation by year, month, week, and carrier and serves as the single source of data for the Power BI dashboard and forecasting models.



## **Appendix C: Project Goal 8: GitHub Project Portfolio**

A GitHub repository was created to store and document all project artifacts, including cleaned datasets, R scripts, ETL packages, Power BI dashboards, and report files. The repository supports version control, collaboration, and transparency across the team.

### **GitHub Repository:**

Database-Project-6430\_Group-02-Fall-2025

**Link:** [https://github.com/ImranRasel39/Database-Project-6430\\_Group-02-Fall-2025](https://github.com/ImranRasel39/Database-Project-6430_Group-02-Fall-2025)