Project Progress Report 4

# Team members:

|  |  |
| --- | --- |
| NAME | INDIVIDUAL CONTRIBUTION |
| Giacomo Bizzotto | Answered to questions 1, 2, 4, and 5. Paginated this report and gave the structure to the Power BI report, assigned the roles to everybody |
| Rishitha Thatipally | Gave part of the recommendation to the company |
| Irfan Shaik | Answered to question 3 |
| Errolla Vivek | Gave part of the recommendation to the company |
| Shraddha Shakya | Worked on the dashboard to make it look more professional |
| Sampath Sai Raghav Allaboyina | Gave part of the recommendation to the company |

# Question 1: *Project the number of passengers for the bus terminal from 2025 to 2030. Port Authority will be building temporary staging facilities, and we need to know the number of people who will be using these facilities.*

To project passenger volumes from 2025 to 2030, we implemented an ETS (Error, Trend, Seasonality) model using R, chosen for its ability to model additive trends and seasonality while maintaining flexibility in the presence of structural changes in the time series. This was particularly important given the abrupt shock to ridership during the COVID-19 pandemic. As outlined in Project Report 3, we excluded the months March–June, November, and December of 2020 from the training phase, as these periods represented statistical outliers due to lockdowns and travel restrictions that would otherwise bias the model.

The dataset used to train the ETS model was primarily drawn from the “TBT\_Traffic\_Database\_PA-1”, which contains more than 5 million records related to vehicle and passenger activity across multiple facilities. This dataset was enriched with weather-related features obtained from the “Weather Database-1”, which includes variables such as average wind speed (AWND), precipitation (PRCP), snowfall (SNOW), snow depth (SNWD), and temperature extremes (TMAX, TMIN). These features were included to assess their marginal impact on monthly fluctuations, though long-term trends remained primarily driven by temporal variables.

Data cleaning, transformation, and integration were conducted in SQL Server Management Studio (SSMS), Excel, and RStudio. We used custom SQL queries to join bus departure and passenger tables (MBT\_Passenger\_Departures and MBT\_Bus\_Departures) and calculated monthly averages of passengers per bus. These averages were then multiplied by the forecasted number of buses per facility to estimate total passenger volumes.

The final forecast indicates a stabilization of total annual bus passenger volume around 155 million passengers from 2025 to 2030. This recovery, while modest compared to pre-2020 levels (~205 million in 2019), represents a steady return to normalized demand. However, the real insight from our ETS-based forecasting lies not in the system-wide total alone, but in its granular breakdown by individual facility, which reveals substantial variation in projected traffic volumes.

This facility-level analysis is particularly important for the Port Authority’s plans to construct temporary staging facilities to manage traffic rerouting during upcoming infrastructure projects. Our Power BI dashboard, fed by the “PassengersPerFacility.csv” dataset, offers interactive visualizations that allow stakeholders to filter traffic forecasts by facility, year, and month, revealing where congestion hotspots may arise. For example, Bayonne and GWB Lower consistently emerge as high-demand nodes across all projection years, suggesting the need for expanded staging capacity, additional personnel, and optimized bus scheduling. Conversely, facilities such as Outerbridge and GWB PIP show significantly lower ridership trends, making them ideal for more modest, cost-effective staging interventions.

By tailoring temporary infrastructure deployment based on facility-level forecasts rather than system-wide averages, the Port Authority can avoid the pitfalls of overbuilding in low-volume zones and underbuilding in high-demand locations, thereby optimizing resource allocation and improving overall commuter experience. In summary, the ETS-driven passenger forecasts offer a data-backed blueprint for infrastructure adaptation, aligned with the strategic goals outlined in the Port Authority’s modernization plan.

# Question 2: *What are the most important factors in predicting the number of passengers for the bus terminal?*

To thoroughly evaluate the impact of different variables on our forecasting model, we conducted a multi-phase feature assessment using the ETS (Error, Trend, Seasonality) modeling approach in R. Our objective was to identify which variables most significantly improved forecast accuracy, both for short-term (monthly) operational planning and long-term (annual) strategic forecasting. This was accomplished through a process of incremental variable inclusion, where we began with a baseline model using only time-based features, and then gradually added external variables, carefully monitoring performance metrics such as Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).  
Across all model variations, the historical ridership variables—specifically year, month, and lagged passenger volumes—emerged as the most powerful predictors. These temporal features captured consistent patterns of commuter behavior, including annual seasonality, post-COVID recovery trajectories, and long-term usage trends that persisted despite short-term disruptions. Models trained exclusively on these variables already produced stable, high-performing forecasts, reinforcing the reliability of internal system dynamics in determining future demand.

To enhance the model’s explanatory and predictive capabilities, we integrated external weather-related features sourced from the “Weather Database-1”. These included average wind speed (AWND), precipitation (PRCP), snowfall (SNOW), snow depth (SNWD), and temperature extremes (TMAX and TMIN). While these variables introduced minor improvements to month-level forecast accuracy—particularly during extreme weather seasons—their influence on aggregated annual totals was limited. For example, snowfall and low temperatures were associated with modest declines in ridership during winter months like January and February, likely due to deteriorated road conditions and reduced commuter willingness. However, such impacts were short-lived and often disappeared when analyzed over quarterly or yearly scales.

Nevertheless, integrating weather-related features within our ETS modeling framework—implemented through an ETS+X strategy using external regressors—yielded meaningful explanatory gains in select terminals. Specifically, the adjusted R² obtained from supporting TSLM regressions reached 0.83 for Holland Tunnel, 0.68 for Outerbridge Crossing, and 0.55 for Lincoln Tunnel, demonstrating that weather variables explained a substantial portion of variance in those cases. The following R code was used to assess their statistical contribution:

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

weather\_vars <- facility\_data %>% select(AWND, PRCP, SNOW, SNWD, TMAX, TMIN)

weather\_vars[is.na(weather\_vars)] <- 0

model\_tslm <- tslm(ts\_data ~ trend + AWND + PRCP + SNOW + SNWD + TMAX + TMIN,

data = cbind(facility\_data, trend = 1:length(ts\_data)))

summary(model\_tslm)$adj.r.squared

This approach allowed us to quantify the marginal contribution of environmental factors and identify the specific terminals where their inclusion enhanced model robustness. From a planning perspective, weather features proved valuable in operational contexts—such as scheduling around snowstorms or predicting drops during severe weather. However, for strategic long-range forecasting and infrastructure scaling, these variables added limited predictive value and introduced unnecessary volatility.

Overall, while weather data offered situational insights and context-specific accuracy gains, the most important and consistently reliable predictors of passenger volumes remained the core time-based features—year, month, and historical demand. These elements captured the underlying behavioral and seasonal dynamics of transit usage, making them essential for any high-confidence ridership forecast.

# Question 3: *Project the prediction results by the individual carrier.*

We applied the Prophet model to weekly ridership data segmented by individual carriers, using the comprehensive “PassangersByCarrier.csv” dataset, which spans from January 2020 to projected values through December 2030. This dataset includes granular time-series data for key bus operators including NJ Transit, Greyhound, Academy, Peter Pan, Martz, TransBridge, Coach USA, and others. Prophet, developed by Facebook, is particularly suited for this forecasting task due to its robust ability to model non-linear growth patterns, multi-level seasonality, and holiday effects, which are common in transit ridership data.

The forecasts generated by Prophet highlighted markedly different trends for each carrier—insights that are critical for planning temporary staging infrastructure during the Port Authority's phased construction projects. Among all carriers, NJ Transit exhibits the highest projected weekly ridership, driven by consistent weekday commuter patterns and a relatively quick rebound following the COVID-19 disruption. As one of the backbone services of the regional transit system, NJ Transit’s strong and steady demand necessitates robust and prioritized staging support, especially at high-traffic terminals like Lincoln Tunnel and GWB.

Coach USA, the second highest in projected volumes, shows a distinct seasonal pattern, with peaks recurring during major holidays (Thanksgiving, Christmas, July 4th) and summer vacation months. Its upward trend in ridership indicates a resurgence in long-distance intercity travel, making it essential for the Port Authority to allocate additional staging capacity during these peak intervals.

Academy Bus forecasts show moderate but steady growth, likely linked to contractual routes for corporate or institutional clients. Though not as volatile as other carriers, its dependable ridership justifies scaled but reliable staging, particularly in areas servicing schools, universities, or business districts.

In contrast, Martz and Peter Pan/Bonanza display flat or declining post-pandemic recovery curves. These carriers, which once thrived on budget-conscious, high-frequency urban routes, appear to have lost a share of ridership to remote work trends and alternative transport options. Given their current trajectory, limited staging adjustments may suffice unless future demand rebounds due to policy or pricing shifts.

C&J Bus Lines presents stable but niche ridership volumes, generally servicing specific regional corridors. Although, its impact across localized zones still requires flexible, decentralized staging solutions.

All these insights are integrated and visualized through our interactive Power BI dashboard, where weekly ridership trends are plotted with filters by carrier, year, and season. This allows decision-makers to explore comparative carrier performance, identify overlapping demand peaks, and make real-time planning decisions about space allocation, scheduling, and personnel deployment.

Overall, the strategy for temporary infrastructure during construction periods should prioritize NJ Transit, Coach USA, and Greyhound for heavy investment in capacity and flow management, while maintaining scalable and flexible support for Academy, Peter Pan, and the other operators. This differentiated planning ensures both operational efficiency and cost-effectiveness based on actual and forecasted demand dynamics.

# Question 4: *What would be the busiest time for the bus terminal staging facilities between 2025 and 2030? Visualize by week, month, and year.*

To determine the busiest periods for the Port Authority’s staging facilities between 2025 and 2030, we conducted a detailed analysis using time series forecasting powered by the ARIMA (Autoregressive Integrated Moving Average) model. The forecasting was applied to traffic data that includes buses, trucks, and cars, and was aggregated by week, month, and year across all facilities. The underlying dataset used for this projection was derived from historical vehicle flow patterns from 2015 to 2024, which we compiled, cleaned, and extended into the “Busiest times.xlsx” file.

Our modeling approach began with data augmentation—where missing 2024 values were estimated by applying a 5% year-over-year increase to comparable months from 2023, as outlined in Project Report 3. We then applied the ARIMA model in R to forecast traffic volumes across weekly intervals for each facility, projecting trends through the end of 2030. The results were then visualized using Power BI, offering stakeholders interactive, filterable dashboards that enabled exploration by facility, and time interval.

The analysis revealed that Week 14 (early April) consistently stands out as the busiest week, with an estimated traffic volume approaching 13 million vehicles across all facilities. This aligns with seasonal springtime travel surges, which typically follow the end of winter-related slowdowns and precede summer vacations. At the monthly level, December ranks as the most heavily trafficked month, with an average of 56 million vehicles, likely due to holiday travel and increased commercial freight movement in the lead-up to year-end. July and March also demonstrate significant spikes, reflecting vacation-related travel and spring break periods.

On an annual basis, the ARIMA model shows a gradual increase in total traffic volume year over year, culminating in 2030, which is projected to be the busiest year, with 108 million vehicles expected. This steady rise is likely influenced by a combination of population growth, infrastructure recovery post-COVID, and a behavioral shift toward private vehicle usage, as supported by trends observed in other parts of our analysis.

The Power BI dashboard, built from this ARIMA-driven forecast, includes filterable views that allow planners to examine each facility in isolation or in comparison to the network average, offering a dynamic tool for traffic trend simulation and capacity planning.

Furthermore, the insights from this analysis are directly applicable to the deployment of temporary staging infrastructure during high-demand periods. By aligning facility-level resource allocation with forecasted peaks at the weekly, monthly, and yearly levels, the Port Authority can ensure that congestion is mitigated, operational disruptions are minimized, and passenger experience remains smooth throughout the construction period and beyond.

# Question 5: *How does our current usage compare to 2019 which is the last year before covid?*

When comparing our current passenger volume projections to 2019, the last full year prior to the onset of the COVID-19 pandemic, a distinct shift in transportation behavior becomes evident. Using data from the “PassengersPerFacility.csv” and “Busiest times.xlsx ” datasets, we analyzed trends in both average passengers per bus and total vehicle flow, offering a comprehensive view of system usage before and after the pandemic.

In 2019, the Port Authority system saw peak ridership levels, with buses consistently carrying higher passenger loads per trip, indicating strong public transit adoption and a reliance on shared modes of transportation. However, by 2020, this sharply declined—ridership dropped to approximately 132 million, and the passenger-per-bus ratio plummeted due to public health concerns, lockdowns, and the broader shift to remote work. While our ETS and Prophet models forecast a gradual recovery in ridership, with total passengers stabilizing around 145–165 million annually by 2025–2030, this still represents a shortfall of nearly 50 million passengers compared to 2019 peak levels, as visualized in our Power BI dashboard.

Interestingly, this decline in per-bus occupancy is occurring in parallel with an increase in total vehicle traffic, including cars, trucks, and buses, which is projected to reach a system-wide high of 108 million vehicles annually by 2030, based on ARIMA model outputs. This growing divergence—decreasing public transit utilization amid rising overall traffic—suggests a persistent post-COVID behavioral shift. Many former bus users appear to have permanently transitioned to private transportation, driven by a desire for personal space, heightened hygiene awareness, and perhaps the flexibility afforded by hybrid and remote work models.

Our dashboard allows users to observe this phenomenon in detail: side-by-side visualizations of passenger-per-bus metrics and vehicle count trends make clear that public transportation recovery is lagging overall travel activity. This underscores a key insight—congestion is worsening despite stagnant transit ridership, suggesting that road infrastructure is under greater pressure, even as bus systems operate below historical efficiency levels. This trend is likely a lingering effect of the COVID-19 pandemic, which accelerated the shift toward remote work. As a result, fewer people are commuting to offices, and more are relying on flexible, individual travel options.

From a planning perspective, this trend has profound implications. The Port Authority must accommodate increasing traffic volumes through smarter staging, routing, and congestion mitigation strategies, even if public transit demand does not return to pre-pandemic levels. Simultaneously, efforts to rebuild confidence in public transportation—through enhanced cleanliness, real-time service updates, and targeted communication—will be essential to reversing the modal shift toward private vehicles.

In sum, the data paints a picture of a new normal for mobility in the region: one where private vehicle reliance is higher, roadway congestion is intensifying - but never as much as the pre pandemic levels - and bus ridership, while recovering, may plateau below historical highs. A forward-looking infrastructure strategy must embrace this reality, investing in both traffic management systems and transit incentives to balance demand and ensure a resilient, efficient transportation network for the next decade.

# Recommendations to the company.

Based on the longitudinal analysis of passenger data trends between 2015 and 2030, the Port Authority of New York and New Jersey is navigating a critical transformational phase in urban mobility planning. One of the most striking findings from our ETS and Prophet model forecasts is the precipitous decline in passenger volume in 2020, where ridership plummeted to approximately 132 million—a dramatic drop from the pre-pandemic peak of 205 million in 2019. This collapse, captured in both the raw data and our Power BI dashboard visualizations, reflects the widespread impact of COVID-19-related lockdowns, remote work adoption, and shifting commuter priorities.

From 2021 to 2024, passenger volumes have shown a gradual but incomplete recovery, with projections indicating stabilization around 145–165 million annual passengers during the 2025–2030 window. While this represents a meaningful rebound, it remains approximately 25% below the 2019 benchmark, suggesting that pandemic-induced behavioral shifts—such as hybrid work schedules, altered peak-hour demand, and rising preference for private vehicles—may have a lasting influence on transit patterns.

These insights underline the need for the Port Authority to proactively evolve its operational strategy. The legacy model of uniform service delivery must be replaced with agile, data-driven frameworks that align transit offerings with real-time demand dynamics. For example, routes and facilities with sustained underperformance post-COVID—such as those servicing low-density residential areas—should be reassessed for route frequency, fleet allocation, and staffing efficiency. Conversely, high-performing terminals like Lincoln Tunnel and GWB, which our projections flag as enduring traffic pressure points, should be prioritized for capacity expansion, dedicated staging zones, and enhanced service frequency during peak travel windows.

A central recommendation is to invest in smart mobility infrastructure. This includes the deployment of:

* Sensor-based passenger counters for real-time load tracking,
* Mobile app integrations to communicate delays, occupancy levels, and service changes,
* Predictive maintenance systems to reduce downtime and costs.

Such technologies will enable the Port Authority to move from reactive planning to anticipatory operations, allowing for mid-day rescheduling, on-demand dispatching, and real-time service optimization.

To support mid- and long-term planning, additional commuter intelligence must be collected and analyzed. The agency should implement:

* Travel behavior surveys to understand trip purposes (commute, leisure, education),
* Mobile ticketing data mining for trip frequency and loyalty trends,
* Wi-Fi and Bluetooth-based motion tracking for fine-grained foot traffic insights.

These data streams will help build dynamic traveler personas, guiding both macro-level investment (e.g., infrastructure upgrades) and micro-level service adjustments (e.g., targeted discounts or rerouted services).

Beyond internal operations, the Port Authority must remain responsive to external pressures and systemic shifts. These include:

* Macroeconomic variables, such as rising fuel prices and inflation, which may either encourage or discourage transit usage;
* Policy changes, including federal infrastructure funding, congestion pricing mandates, and urban mobility incentives;
* Urban development projects, particularly transit-oriented developments (TODs) that cluster residential and commercial activity around hubs;
* Climate-related disruptions, such as flooding, hurricanes, or extreme heat events, which threaten infrastructure, especially in coastal facilities like Bayonne and GWB PIP.

As part of its resilience and sustainability strategy, the Port Authority should expand its role in public-private partnerships (PPPs). Collaborations with corporations, universities, real estate developers, and regional governments can facilitate:

* Subsidized fare programs for employees and students,
* Shuttle integrations with residential complexes,
* Green mobility incentives that position public transit as an environmentally responsible option.

Finally, with the anticipated slow ridership recovery, the next five years offer a window of opportunity to reimagine and reposition transit in the eyes of the public. The Port Authority should double down on efforts to market transit as a modern, reliable, and climate-conscious alternative to private cars. Enhanced cleanliness, digital service transparency, and personalized route planning can rebuild trust and restore ridership, especially among occasional users and new residents.

In essence, this is not just a time to recover—it is a time to reinvent the transit experience, making it safer, smarter, and fully aligned with the evolving needs and expectations of New Yorkers and regional commuters in a post-pandemic era.

# Visuals from our Dashboard:

A screen shot of a graph

AI-generated content may be incorrect.

A graph on a yellow background

AI-generated content may be incorrect.

A screenshot of a graph

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.