**Port Authority of NY & NJ Passenger and Traffic Forecasting Report.**

Final Project Report

**1. Project the number of passengers for the bus terminal from 2025 to 2030.**

1. **Most important variables + dependent variable:** The dependent variable is the monthly total passenger volume at each terminal (GWBBS and MBT). Key variables include time (year, month), seasonal trends, and changepoints (e.g., COVID-19 impacts) to capture patterns in historical data and accurately project future demand.
2. **Recommendations:** Ensure historical passenger records are complete and consistently formatted. Including external shocks (like economic trends or policy changes) and updated holiday/event calendars would further strengthen projections.
3. **Models and techniques:** We used the Prophet time series forecasting model, leveraging its strength in handling seasonality, trends, and changepoints.
4. **Tools used:** Python (Google Colab) with Prophet library.

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

1. **Most important variables + dependent variable:** The dependent variable is total passenger volume. Key influencing variables include weather conditions (temperature, precipitation, snow), time-based features (month, day of week), and special indicators (holidays, weekends) that affect transit demand.
2. **Recommendations:** Ensure the weather dataset is complete and aligned with passenger records. Incorporating additional context like fuel prices, local events, or competing transportation services could improve model precision.

**3. Project the prediction results by individual carrier.**

1. **Most important variables + dependent variable:** The dependent variable is the monthly or hourly passenger count per carrier. Key variables include time (month, hour), weather data, and carrier-specific trends, as these directly affect demand per operator.
2. **Recommendations:** Maintain accurate and detailed carrier-level data, ensuring no gaps or inconsistencies. Adding variables like carrier pricing, route changes, and customer satisfaction metrics would enhance predictions.
3. **Models and techniques:** We used XGBoost regression models tailored for each carrier, along with outlier detection and feature engineering.
4. **Tools used:** Python (Google Colab) with XGBoost and pandas for data handling.

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

1. **Most important variables + dependent variable:** The dependent variable is total vehicle count or passenger volume by hour. Key variables include hour of the day, day of the week, month, and clustering patterns to classify peak periods.
2. **Recommendations:** Ensure complete hourly traffic data and consider integrating public event schedules and traffic congestion data. Include weekend vs. weekday differentiation to capture realistic demand peaks.
3. **Models and techniques:** We applied K-Means clustering for traffic level classification and time-based visualization across week, month, and year.
4. **Tools used:** Python (Google Colab) with scikit-learn, matplotlib, and seaborn for visualization.

**5. How does our current usage compare to 2019, the last year before COVID?**

1. **Most important variables + dependent variable:** The dependent variable is the monthly or yearly total passenger volume per terminal. Key variables are year, month, and terminal identifiers, allowing comparison between 2019 and recent (2024) trends.
2. **Recommendations:** Ensure that data is consistently aligned across time, and account for any COVID-related reporting anomalies. Including socioeconomic recovery indicators like vaccination rates or travel restrictions could contextualize differences.
3. **Models and techniques:** We used Prophet forecasts and comparative time series visualizations to benchmark 2025 projections against 2019 levels.
4. **Tools used:** Python (Google Colab) with Prophet and Power BI for visual comparison.