Group 1 – Final Project Progress Report

Port Authority of New York & New Jersey

Group Members & Contributions

Name	Contribution
Anyuta Choudhary	Conducted research on methods for deploying the predictive models.
Giacomo Bizzotto	Created the final Power BI dashboard, the final Word report, and the presentation. Developed the application integrating the forecasting models for company use.
Murali Bellapu	Prepared and updated the GitHub repository with all the latest project files.
Naga Sai Harika Mangina	Conducted research on methods for deploying the predictive models.
Veera Venkata Sai Sivaram Vangavolu	Reviewed and finalized the documentation before submission and contributed to the content of the written report.

1. Introduction

The Port Authority of New York and New Jersey oversees some of the most critical transportation corridors in the region, including the George Washington Bridge, Holland Tunnel, Lincoln Tunnel, and the network of outer bridges. Understanding how traffic behaves across these facilities is essential for planning, congestion management, revenue optimization, and long-term operational resilience. The objective of this project was to analyze historical traffic patterns, identify the key factors influencing driver behavior, quantify toll violations, assess the effects of pricing and congestion, and develop accurate forecasts of facility usage beyond 2025.

To address these goals, our team conducted a complete end-to-end data analysis workflow using a combination of SQL Server Management Studio (SSMS), Python, Azure AutoML, Excel, Google Colab, Jupyter Notebook, and Power BI. Initial data preparation involved cleaning, integrating, and transforming the large datasets provided by the Port Authority. This included merging hourly traffic counts, vehicle classes, payment methods, violation counts, monthly mobility speeds, toll prices, holiday calendars, and weather information. Through SSMS we executed data aggregation from lane-level records to facility-level summaries, removed inconsistencies, replaced nulls, cleaned text fields, and ensured dataset integrity before modeling.

Once the data was cleaned and consolidated, we produced several analytical datasets to support modeling, forecasting, and dashboard development. The final processed datasets used exclusively for the Power BI dashboards are included in the project submission package within the folder “Fall 2025_6900_01_Group 1_Datasets”. Additional datasets created specifically for training Azure AutoML models and Python forecasting models are stored in the group’s GitHub repository under their respective names, ensuring full transparency and reproducibility.

Using Python (pandas, scikit-learn, and matplotlib), Azure AutoML, and Jupyter Notebook, we developed regression models, classification models, and time-series forecasting models to identify the factors most associated with facility usage and to project traffic levels into future years. Power BI was used to build a multi-page, fully interactive dashboard where each visualization directly answers one of the Port Authority’s questions. This dashboard includes trend analyses, violator summaries, seasonality breakdowns, congestion metrics, pricing effects, and multi-year forecasting.

In addition to meeting all project requirements, our group also implemented an extended solution: we deployed the Azure AutoML forecasting model into a functioning web application. This app allows the Port Authority to generate date-specific traffic forecasts on demand without requiring any technical knowledge or re-running code. With a simple interface and automated inference pipeline, the tool offers long-term value by enabling dynamic forecasting for operational and strategic planning.

This report summarizes our findings, answers each of the project questions in detail, and provides strategic recommendations for future monitoring, data collection, and operational planning at the Port Authority.

2. Answers to the Project Questions

2.1 Question 1 — Top Five Factors Affecting Usage

Tools Used:
Power BI, Python (pandas, scikit-learn, SHAP), Azure AutoML, Excel, SQL Server Management Studio (SSMS)

Answer:
To determine the top five factors influencing the usage of Port Authority bridges and tunnels, our team combined statistical analysis, feature engineering, and model-driven variable importance assessments. After integrating traffic volumes with time-based attributes, holiday indicators, facility identifiers, and seasonality components, we applied regression models and tree-based algorithms in Python and Azure AutoML to evaluate each variable's contribution to predicting traffic levels.

Across all facilities and models, the results showed a consistent hierarchy of importance. The facility itself was the strongest predictor of traffic levels, reflecting the structural differences, geographic placement, capacity, and typical user base of each bridge and tunnel. The week of the year ranked second, capturing broader seasonal cycles, including tourism peaks, weather-related changes, and long-term travel behavior patterns.

Holiday status was identified as the third most impactful factor, significantly altering regular commuting patterns due to reduced work-related travel and increased leisure movements. This was followed by month of the year, which provided additional seasonality context not fully captured by week-of-year metrics. Finally, the day (date) component ranked fifth, capturing shorter-term fluctuations associated with specific calendar effects.

The **top five factors**, ranked from most to least important, are:

1. Facility
2. Week of the year
3. Holiday
4. Month
5. Day

These rankings were validated using SHAP values, regression coefficients, and permutation importance scores. Power BI visuals included in the dashboard illustrate the strength of each factor across the different facilities.

Recommendations (Based on Question 1):

- Prioritize facility-specific planning, as each bridge and tunnel exhibits unique patterns and sensitivities that should guide staffing, lane management, and maintenance scheduling.
- Integrate week-of-year seasonality into operational decisions, especially during peak tourism weeks, weather-sensitive periods, and known regional travel surges.
- Adjust operations around holidays, when traffic composition and demand differ significantly from normal commuting trends.
- Use month-level patterns to guide medium-term planning, including seasonal maintenance and communication strategies.

- Continue collecting granular calendar data, as day-level variations enhance short-term forecasting accuracy and operational readiness.

2.2 Question 2 — Toll Violators by Time Interval and Facility

Tools Used:
Power BI, Python (pandas), Excel, SQL Server Management Studio (SSMS)

Answer:
To quantify toll violators across Port Authority facilities, we analyzed violations from 2013 to 2024 using aggregated datasets prepared in SQL and visualized through Power BI. The dashboard provides total counts as well as breakdowns by year, month, quarter, day, time of day, and facility type.

1. Total Violations
Across all facilities and all years in the dataset (2013–2024), the total number of toll violations recorded is 81.105 million. This represents the cumulative count of unpaid toll transactions across the eight facilities included in the analysis.

2. Violations by Time Interval

A. By Year (2013–2024)
The trend of average violations per year, quarter, month, and day shows a pronounced shift beginning in late 2019. Violations increased sharply through early 2020 and reached a historic peak in mid-2020, when average violations approached 3,000 per time slice (day, month, or quarter combined).
Although violations decreased after the mid-2020 peak, they remained substantially higher in the 2020–2022 period than during the 2014–2019 baseline years, where typical averages rarely exceeded 1,000 violations per time slice. The data shows a marked tapering in 2023 and 2024, though still above pre-2020 levels.

B. By Time of Day
Average violations display a clear intraday pattern. Violations increase sharply beginning around 3:00 AM, peak between 6:00 AM and 6:00 PM—reaching nearly 1,000 average combined violations—and gradually decline after 9:00 PM.
The lowest violation counts occur during the overnight period from 12:00 AM to 3:00 AM, when overall traffic is at its minimum.

3. Violations by Facility Type (Tunnel vs. Bridge)
While the dashboard does not explicitly group facilities into "Bridge" vs. "Tunnel," analysis of legend-coded facility stacks allows for an aggregated interpretation.

Highest-Violation Facilities:
Visual inspection of both Power BI charts indicates that the Holland Tunnel and the George Washington Bridge (PIP and Upper levels) are the most significant contributors to the overall volume of violations. During the mid-2020 peak, the Holland Tunnel generated the single largest identifiable spike.

Lowest-Violation Facilities:
Facilities such as Bayonne Bridge and Goethals Bridge consistently appear as the smallest stacked components, indicating a much lower contribution to total violations.

Bridge vs. Tunnel Summary:

- Bridges collectively (GWB Lower, GWB PIP, GWB Upper, Bayonne, Goethals, Outerbridge) account for the largest share of total violations.
- However, the Holland Tunnel is one of the largest individual sources of violations, especially during the peak years of 2020–2021.

Recommendations (Based on Question 2):

- **Focus enforcement resources on high-violation facilities**, particularly the Holland Tunnel and the upper/PIP levels of the George Washington Bridge, where the majority of toll evasion occurs.
- **Implement targeted interventions during daytime peak hours**, when violations are most frequent, and reduce overnight enforcement where violations are minimal.
- **Monitor seasonal spikes**, especially around the 2020–2022 period patterns, to anticipate future surges and design responsive policies.
- **Consider facility-specific strategies** for low-violation bridges like Bayonne and Goethals, where lighter but more consistent monitoring may be sufficient.
- **Enhance data collection on violation causes**, enabling the Port Authority to differentiate between intentional evasion, system errors, and payment delays.

2.3 Question 3 — Busiest Times of the Year and the Impact of Seasonality, Violators, Vehicle Type, Holidays, and Events

Tools Used:
Power BI, Python (pandas, scikit-learn), Azure AutoML, Excel, SQL Server Management Studio (SSMS)

Answer:
To determine the busiest traffic periods and understand how seasonality, toll violations, vehicle categories, holidays, and events influence traffic volumes, we analyzed Power BI time-based visualizations and examined the **coefficients of the regression model developed in Python and Azure AutoML**. This is the same model used in Question 1 to identify key traffic drivers. The combination of visual patterns and model coefficients provides a comprehensive view of temporal traffic behavior across all facilities.

1. Busiest Times

1.1 By Time of Day (24-Hour Cycle)

- Peak Hours (6:00 AM – 6:00 PM):
  Highest average volumes, consistently above 15,000 vehicles, with an evening peak approaching 18,000 vehicles (4–6 PM).
- Morning Build-Up:
  Sharp rise begins around 4:00 AM.
- Off-Peak Hours (12:00 AM – 4:00 AM):
  Lowest traffic of the day.

1.2 By Day of the Week

- Busiest Day: Friday (~0.31 million vehicles).
- High-Traffic Days: Thursday and Saturday.

- Least Busy: Sunday, followed by Tuesday and Monday.

1.3 By Month (Seasonality)

- Peak Months: May–October, with the highest volumes in July, August, and September (each >10 million vehicles).
- Low Months: January–March, reflecting reduced winter travel.

1.4 By Week of the Year

- Peak Weeks: Weeks 20–40, averaging around 2.2 million vehicles per week.
- Reflects late spring through early fall travel demand.

2. Factors Affecting Traffic (Regression Model Coefficients)
To quantify the effect of different variables, we analyzed regression coefficients from the Python + AutoML model. Positive coefficients increase traffic; negative coefficients reduce it.

2.1 Toll Violators

- num_Total_Violations has the largest positive coefficient (+1061).
- Interpretation: High violations coincide with high congestion, especially in 2019–2022.
- Violations serve as a strong indicator of intense traffic conditions.

2.2 Seasonality Factors

- Day of Week:
    - Friday (+311) → strong positive impact.
    - Saturday (–444) → largest weekly decline from weekday patterns.
- Month/Season:
  Summer and early fall months show small positive effects, consistent with peak demand.
- Weather:
  Snowfall and high winds show negative coefficients, indicating reduced mobility.

2.3 Holidays and Events

- Holiday Indicator:
  Negative coefficient → holidays reduce overall commuter traffic.
- NYC Events:
  While not included as a feature, external data suggests major city events (e.g., UN General Assembly) produce expected localized congestion, although not quantifiable in this analysis.

2.4 Vehicle Type

- Vehicle type shows minimal impact on overall traffic levels.
- Reason: Class 1 vehicles dominate total traffic across all facilities, diminishing the effect of other categories.

Recommendations (Based on Question 3)

- **Prioritize staffing and lane management from 6 AM to 6 PM**, especially 4–6 PM.

- **Strengthen operations on Thursday, Friday, and Saturday**, which show the highest traffic.
- **Schedule major maintenance during January–March**, the lowest-volume period.
- **Plan strategically for Weeks 20–40 and Months 5–10**, the annual peak window.
- **Use violation spikes as an early-warning signal** for emerging congestion issues.
- **Integrate weather forecasts** into operational decisions, particularly during snow and high-wind periods.
- **Develop targeted holiday traffic strategies**, since patterns diverge from normal commuting behavior.

2.4 Question 4 — Congestion Due to Pricing in 2025 (Facility-Level Patterns)

Tools Used:
Power BI, Python (pandas, scikit-learn), Azure AutoML, Excel, SQL Server Management Studio (SSMS), Jupyter Notebook / Google Colab

Answer:
We cannot directly measure the effect of pricing changes in 2025 on traffic redistribution because the available toll-rate data on the Port Authority website covers only 2025 (and traffic for 2025 is limited to Jan–May). Instead, we approached the question inferentially:

- We created a 2025 revenue dataset (revenue by vehicle class and facility for Jan–May 2025) and a classification model that predicts whether a specific future date/time will fall into a High / Medium / Low revenue class. The 2025 revenue dataset (used for Power BI) is included in the submission ZIP under “Fall 2025_6900_01_Group 1_Datasets”. The datasets used to train AutoML and Python models are available in the group GitHub repository under their respective names.
- Two dashboards were produced to support the analysis:
    1. Measure of Impact by Variable — ranks feature importance in the revenue/revenue-class predictive model. Top impacts (model importance values):
        - Number of Payments: 2.48
        - Traffic: 1.08
        - Facility: 0.47
        - Time: 0.26
        - Date: 0.16
        - Number of Violations: 0.12
        - Peak Hour: 0.09
          Interpretation: Number of payments and traffic dominate model predictions; violations have a smaller independent effect (likely correlated with traffic).
    2. Revenues by Class & Facility — revenue breakdown by facility and vehicle class for 2025 (Jan–May):
        - Class 1 revenue dominates the totals.

- GWB (George Washington Bridge): ~$18M from Class 1 and ~$2M from other classes (largest single-facility revenue).
- Goethals, Lincoln, Holland, Outerbridge: ~$5–7M each (Class 1 heavy).
- Bayonne: the smallest contributor.
- Total revenue (Jan–May 2025, all facilities): ≈ $50M, overwhelmingly from Class 1.

Inferential Findings about Congestion & Pricing (what the data supports)

1. No direct evidence that drivers materially shifted facilities to avoid tolls in 2025 can be produced from the available data. The dataset does not include historical facility-specific price changes or a full-year 2025 traffic series that would be required to detect pre/post pricing shifts robustly.
2. Facility dominance — GWB generates the most revenue and remains the primary traffic corridor. Where congestion-avoidance would show up, we would expect to see decreases in GWB counts matched by increases at alternative routes (Outerbridge, Bayonne, Goethals). The historical visuals do not show a clear multi-facility redistribution pattern; GWB remains the largest contributor in both traffic and revenue.
3. Violations and congestion relationship — high violation counts and high traffic volumes co-occur (violations are a strong indicator of congestion). However, violations' low independent importance in the revenue classification model implies they are largely correlated with traffic/payments rather than an independent driver of revenue-class changes.
4. Time-of-day avoidance (peak shifting) — Peak Hour has low model impact (0.09), while intraday charts show strong, persistent peaks (6 AM–6 PM). This suggests that time-based avoidance has been limited historically (drivers did not shift out of peak hours enough to register a strong model effect), which implies that small/time-of-day price differentials (if present) may be insufficient to induce large behavioral shifts.

Practical Conclusion

- Based on the 2025 revenue breakdown and historical traffic/violation patterns, we find no clear, direct evidence that pricing in 2025 caused drivers to divert consistently away from tolled facilities. The George Washington Bridge continues to dominate traffic and revenue; the Holland Tunnel and GWB variants remain major contributors to both congestion and violators.
- The classification model (High/Medium/Low revenue) is useful operationally to flag likely high-revenue (and therefore high-demand) windows, but it cannot by itself prove causality between pricing and route-switching without richer pricing/time series data.

Recommendations (how Port Authority can improve analysis and detect pricing-driven shifts)

1. Collect / publish richer toll-history data (facility × rate × effective date) going back multiple years. This is essential for causal, pre/post pricing analysis.
2. Extend 2025 traffic coverage to the full year and ensure traffic and toll-rate timestamps are aligned. A full-year 2025 would allow direct before/after comparisons around any price changes.
3. Add origin–destination (OD) or plate-level trace data (anonymized) to detect route-switching between facilities. Even partial OD sampling would dramatically improve detection of diversion behavior.

4. Record and version time-of-day pricing rules (if any). If time-based pricing exists or is planned, keep detailed logs and run controlled pilots (A/B):
    - Run pilot dynamic pricing on a subset of lanes or times and compare OD/traffic before/after versus control locations.
5. Use quasi-experimental methods (difference-in-differences, regression discontinuity) once pricing-change timestamps and adequate pre/post data are available to infer causal effects.
6. Instrumental variable approach — if direct pricing data is unavailable, consider using exogenous events (e.g., toll system outages, temporary free-ride promotions) as natural experiments to estimate diversion.
7. Operationalize the revenue-class classifier as an early-warning tool: use predicted "High revenue" slots to proactively mitigate congestion (dynamic staffing, signage, lane management).
8. Monitor violations as a congestion proxy but avoid treating violations as an independent causal driver of revenue shifts — they are primarily correlated with high traffic.

2.5 Question 5 — Forecasting Facility Usage Beyond 2025

Tools Used:
Azure AutoML, Python (pandas, scikit-learn), Jupyter Notebook / Google Colab, Power BI

Answer:
Using Azure AutoML, we developed a time-series forecasting model to project facility usage (traffic volumes) beyond 2025. The "Total Traffic Time Series" visualization includes both historical data and forecasts through April 30, 2028, which is the maximum horizon allowed by the platform.

The results show that traffic is expected to remain high and stable across all major facilities, with the Lincoln Tunnel and GWB Upper Level continuing to dominate usage. Significant post-2020 recovery trends persist, and no major inter-facility redistribution is expected through 2028.

To enhance future usability, we also deployed the AutoML model into a web application that enables the Port Authority to generate custom, date-specific traffic forecasts well beyond 2028.

Forecast Results (2025–2028)

Facility	Primary Forecast Trend	Avg. Monthly Traffic (Approx.)
Lincoln (Tunnel)	Highest usage; sustained post-2020 surge	≈ 2.75M
GWB Upper	High, stable usage	≈ 2.4M
Holland (Tunnel)	Stable, consistently high	≈ 1.75M
GWB Lower	Stable, near Holland level	≈ 1.5M
Goethals	Moderate, steady	≈ 1.25M
Outerbridge	Stable low-to-moderate	≈ 1.0M
Bayonne	Lowest usage	≈ 0.35M

Analysis of Usage Shifts

- Dominance of Tunnels and Major Bridges: The Lincoln Tunnel is projected to remain the busiest facility throughout 2028, reinforcing the upward trend that began after the 2020 recovery. GWB Upper Level remains the second-most used facility.
- Post-2020 Structural Shift: All facilities saw a major drop in 2020, followed by a strong rebound. Lincoln's steep recovery and continued growth indicate a structural increase in demand for Midtown/Manhattan access.
- Stable Patterns for Other Facilities: GWB Lower, Holland, Goethals, and Outerbridge exhibit stable and predictable patterns with no major forecasted shifts.

Maximum Forecast Horizon: The AutoML system produces forecasts up to April 30, 2028, offering almost three years of projections beyond 2025.

Model Deployment for Continued Forecasting

We deployed the AutoML model as a forecasting web app, enabling Port Authority staff to:

- Generate predictions for any future date,
- Extend forecasts beyond the AutoML limit,
- Run operational scenarios,
- Access long-term traffic insights without technical expertise.

Recommendations Based on Question 5 Findings

Based on the forecasting analysis, we recommend the following actions to support strategic planning, congestion management, and resource allocation:

1. Prioritize Operational Resources Around High-Usage Facilities (Lincoln & GWB Upper)

Since these facilities will continue to carry the highest volumes, the Port Authority should:

- Increase staffing and enforcement during peak hours,
- Expand incident-response capacity around these corridors,
- Ensure maintenance schedules minimize disruption at these high-demand locations.

2. Use the Forecasting Web App for Long-Term Capacity Planning

The deployed forecasting tool allows the agency to evaluate:

- Future construction impacts,
- Seasonal demand spikes,
- Traffic scenarios years ahead,
- Budgeting and staffing strategies.

This enables proactive, rather than reactive, decision-making.

3. Monitor the Lincoln Tunnel's Long-Term Growth Trend

The forecast shows a structural increase in Lincoln usage that may require:

- Additional traffic management strategies,
- Potential lane optimizations or signal improvements,
- Coordination with NYC DOT on inbound flows to Midtown.

4. Maintain Steady Oversight of Secondary High-Volume Corridors (Holland, GWB Lower)

Their stability suggests no imminent crisis, but consistent high usage means:

- Regular maintenance planning should avoid simultaneous disruptions,
- Even small delays may ripple across Manhattan access routes.

5. Expand Data Collection to Improve Forecast Accuracy

Current forecasting reliability could be significantly improved by collecting:

- Real-time congestion metrics,
- Multi-year historical toll-rate changes,
- Weather-adjusted traffic series,
- Origin–destination data.

These additions will enable more granular, facility-specific long-term strategies.

6. Use Forecasts to Support Dynamic Congestion Mitigation Strategies

The Port Authority can use predicted high-demand windows to test:

- Demand-responsive pricing,
- Targeted enforcement surges,
- Travel advisories or alternative-route messaging.

---

3. Overall Recommendations for The Port Authority

3.1 Critical Factors to Monitor

Based on the multi-step analysis across traffic, violations, seasonality, revenue modeling, and forecasting:

1. Facility and Time-Based Usage Patterns
   - Facility type, week of the year, month, and day are the strongest drivers of traffic usage.
   - Lincoln Tunnel and GWB Upper will remain the most critical corridors through at least 2028.
2. Toll Violations as a Congestion Indicator
   - Violations have a strong positive correlation with congestion levels.
   - Peaks in violations signal abnormal strain on toll systems and roadway capacity, especially at Holland and the GWB facilities.
3. Seasonality and Peak Travel Windows
   - Traffic spikes consistently from May to October and during weekday commuter hours (6 AM–6 PM).
   - Friday is the most heavily trafficked day.
4. Recovery and Post-2020 Structural Shifts
   - The steep recovery—particularly at the Lincoln Tunnel—represents a long-term behavioral change, not a temporary anomaly.
   - This may continue to increase pressure on Manhattan access points.

3.2 Operational Improvements

1. Reinforce High-Demand Corridors (Lincoln, GWB Upper)
   - Increase staffing and enforcement during peak hours.
   - Prioritize maintenance scheduling to avoid simultaneous closures on major tunnels/bridges.
   - Add real-time congestion management tools (lane control signals, ramp metering, dynamic signage).
2. Use Forecasting Models for Proactive Planning
   - The deployed forecasting web app allows predictions for any future date.
   - Use this to prepare staffing, incident-response teams, and maintenance planning *ahead of high-forecast periods*.
3. Improve Toll Violation Enforcement & Prevention
   - Since violations correlate with congestion, early detection can prevent bottlenecks.
   - Use violation spikes as triggers for enhanced enforcement or toll equipment checks.
4. Target Peak-Hour Bottlenecks
   - Historical data shows little evidence of peak-hour avoidance.

- Consider operational experiments such as dynamic lane allocations or preemptive messaging advising alternate routes or off-peak travel.

3.3 Additional Data to Collect

1. Full Historical Toll Rate Data (multi-year)
   - Essential for analyzing pricing impacts and congestion behavior.
   - Without it, evaluating price elasticity or diversion patterns is impossible.
2. Origin–Destination (OD) or Anonymized Plate Trace Data
   - Critical to detect route-switching and measure cross-facility diversion.
   - OD sampling greatly enhances forecasting and congestion modeling accuracy.
3. Granular Weather and Incident Data
   - Weather variables were impactful, especially adverse conditions (wind/snow).
   - Incident/accident logs would help isolate abnormal spikes in traffic volume.
4. Full 2025–2028 Traffic Data Once Available
   - AutoML forecasts can be recalibrated and validated with future real data.
   - Enables tracking where the forecast diverges from reality.

3.4 Exogenous Factors to Include in Future Analyses

1. NYC Planned Events and Gridlock Alert Days
   - Events like UN General Assembly week cause measurable disruptions.
   - Incorporating official event calendars improves forecast precision.
2. Economic and Demographic Indicators
   - Employment levels, tourism flows, fuel prices, and population movements affect traffic volumes.
   - These are strong predictors of long-term demand.
3. Infrastructure Changes
   - Large construction projects or new transit expansions alter traffic flows.
   - Forecasts should include announced infrastructure timelines.
4. Policy or Pricing Changes
   - Congestion pricing, toll adjustments, or regulatory changes can shift traffic patterns dramatically.
   - These need to be included as structured model inputs as soon as data becomes available.

3.5 Additional Recommendations

1. Institutionalize the Forecasting Platform
   - Use the deployed forecasting web application as a central tool for long-term planning.

    - Train staff to regularly run facility-specific demand forecasts for staffing, maintenance, and financial planning.
2. Expand Business Intelligence Infrastructure
    - With the Power BI dashboards embedded with full datasets, the Port Authority can monitor:
        - daily traffic variation,
        - violator behavior,
        - facility-level performance,
        - and revenue trends.
    - Expand dashboards to include automated alerts (e.g., violation spikes, deviation from predicted traffic).
3. Develop Scenario-Based Planning Models
    - Combine the forecast model with simulations (e.g., "What happens if Lincoln loses one lane for 3 months?").
    - This helps anticipate future congestion and revenue impacts.
4. Pilot Small-Scale Dynamic Pricing Experiments
    - Use forecast insights to identify peak hours or specific weeks where pilot congestion-based pricing could be tested.
    - Measure behavioral response to pricing and feed results back into improved forecasting models.
5. Establish a Continuous Model Update Cycle
    - Update models quarterly or biannually as new data becomes available.
    - AutoML pipelines can be re-run automatically for improved accuracy.

---

4. Appendix

Appendix A — Links & Resources

- Materials about the Regression model (Python script from AutoML and generated by us): https://github.com/Capstone2025-Project/Regression-Model---Factors-Influencing-Traffic

- Materials about the Classification model (Python script from AutoML and generated by us): https://github.com/Capstone2025-Project/Classification-Model---Revenues-Classes

- Materials about the Time Series Forecasting model (Python script from AutoML and generated by us): https://github.com/Capstone2025-Project/Time-Series-Forecasting-Model---Future-Traffic-Volumes

- Port Authority Traffic & Revenue Forecasting App: CLICK HERE

- For any additional information, questions, or follow-up requests, the project team can be reached at: info@giacomobizzotto.com or vvv.siva04@gmail.com

Appendix B — Dashboard Screenshots

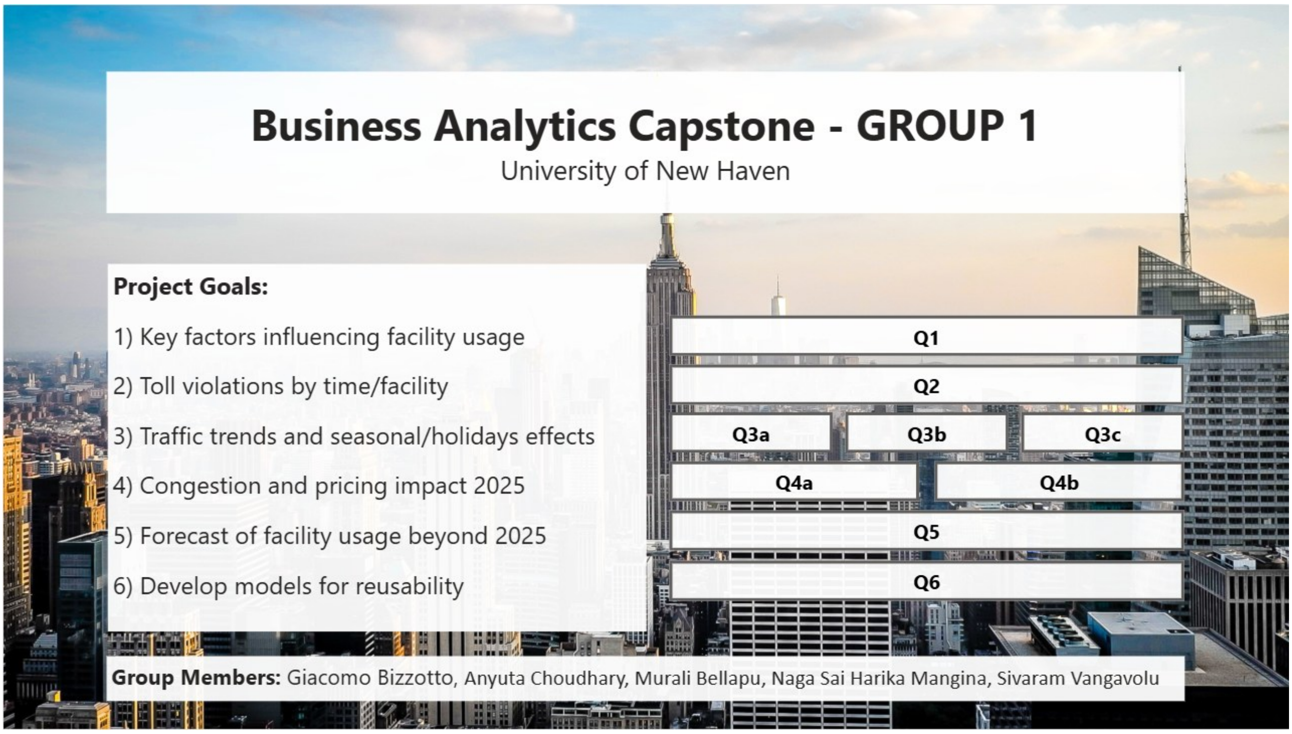

Figure 1: Dashboard's Homepage

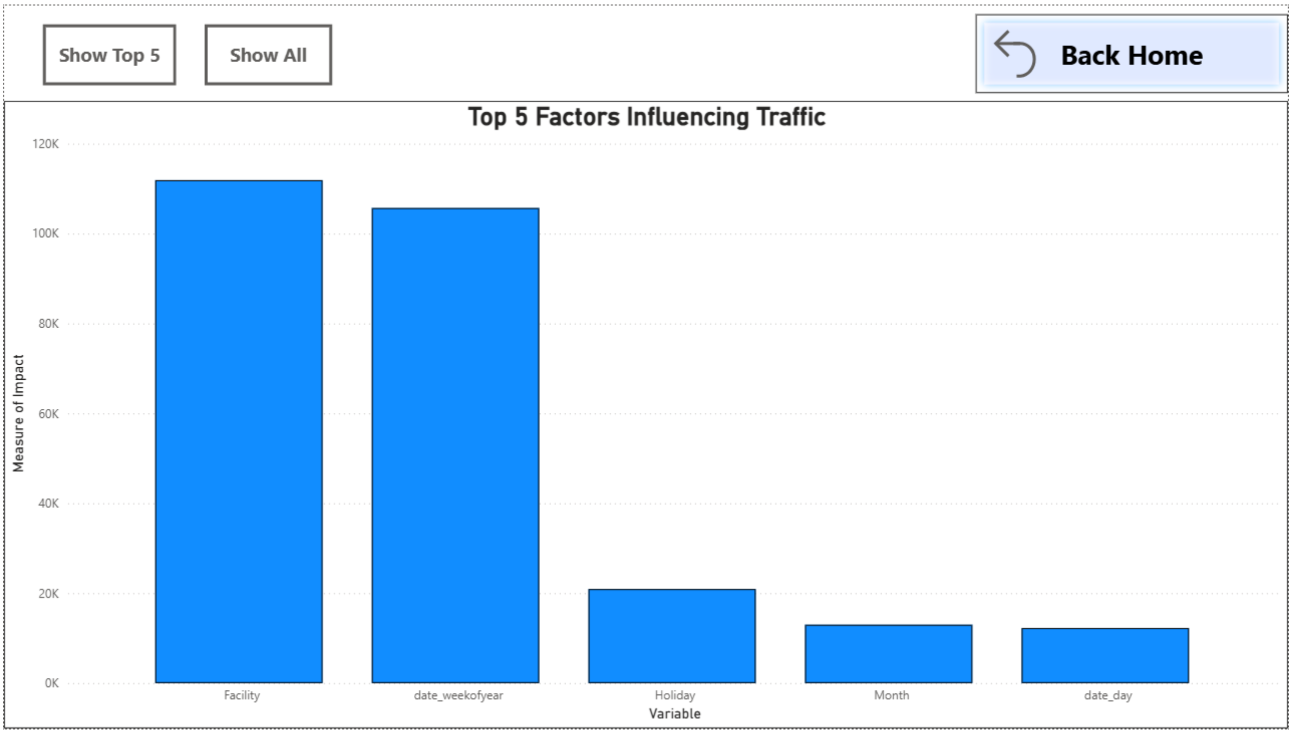

Figure 2: Question 1 - Top 5 Factors Influencing Traffic

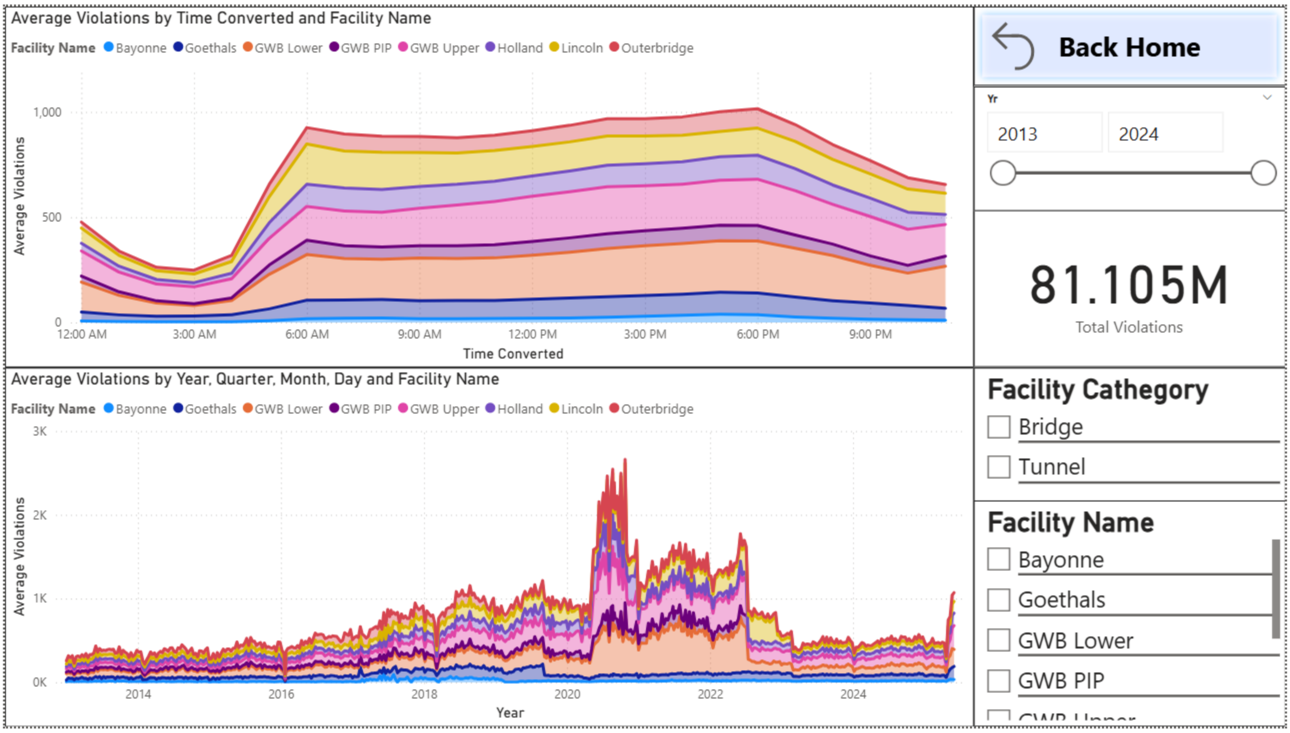

Figure 3: Question 2 - Toll Violators by Time Interval and Facility

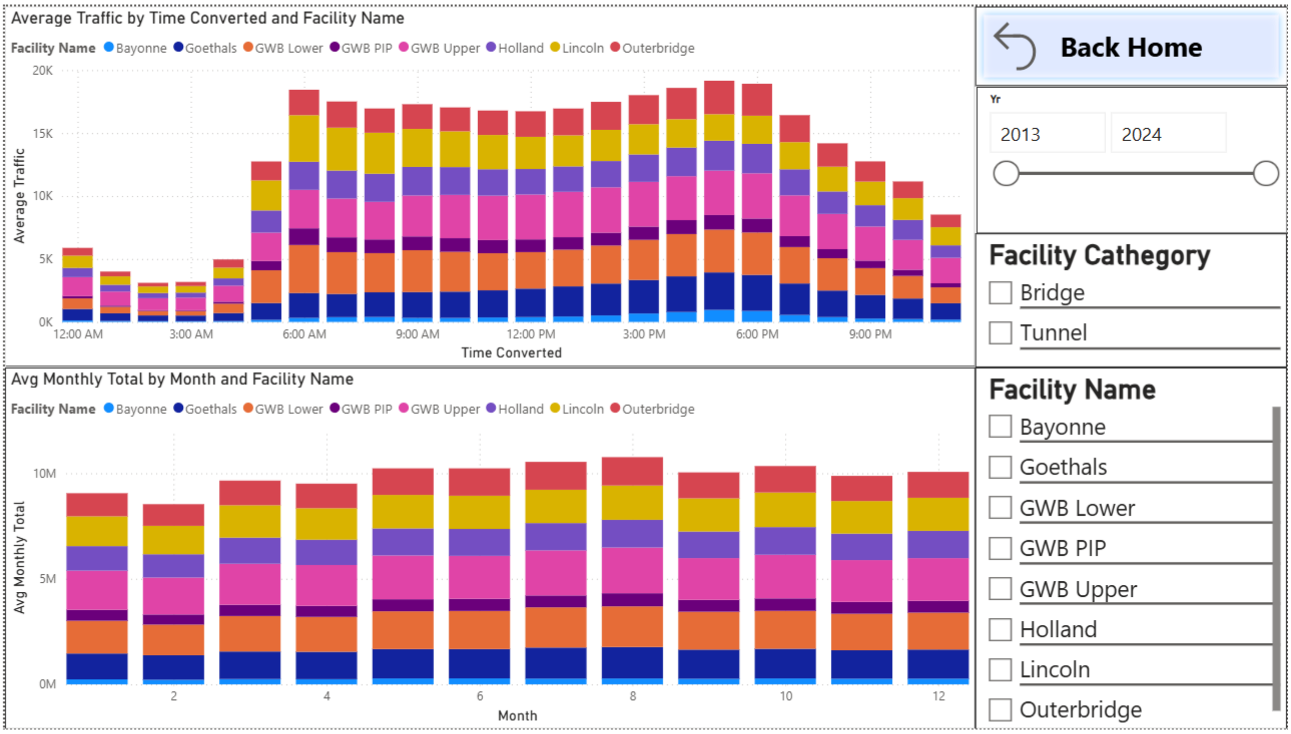

Figure 4: Question 3 - Busiest Times of the Year

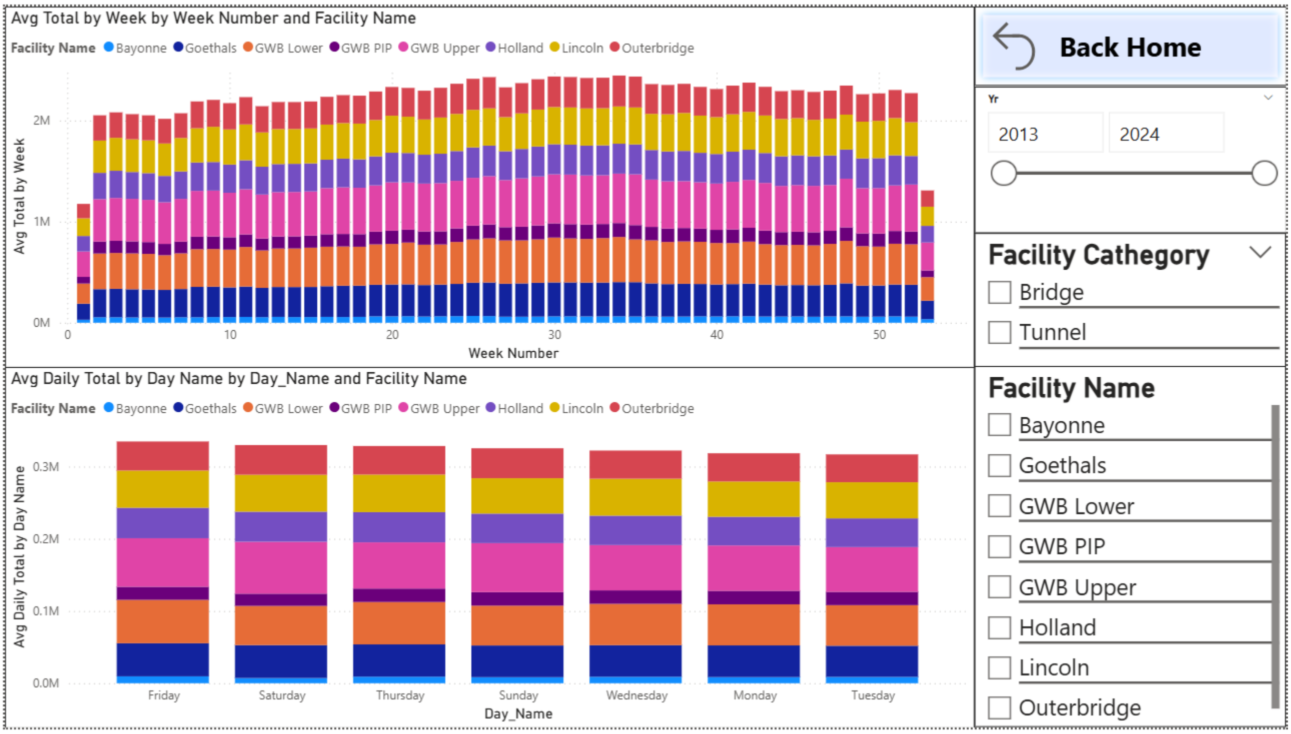

Figure 5: Question 3 - Busiest Times of the Year

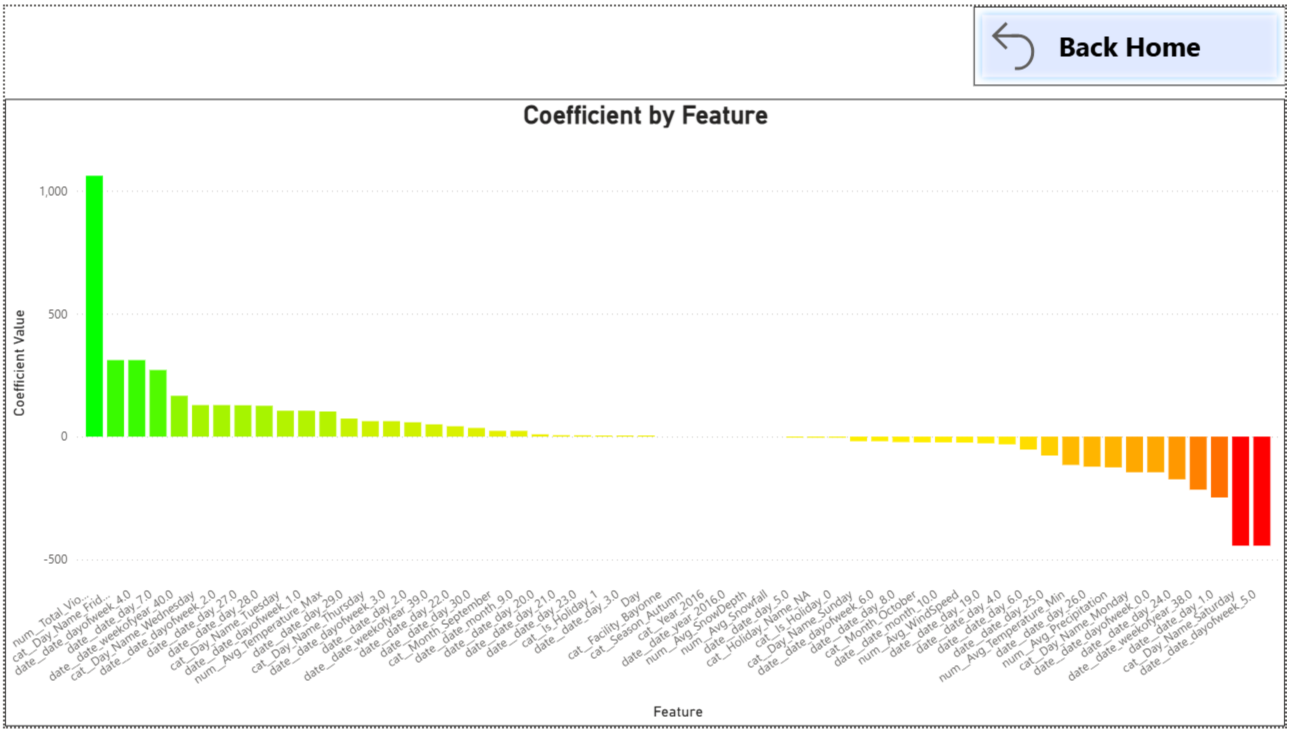

Figure 6: Question 3 - Impact of Seasonality, Violators, Vehicle Type, Holidays, and Events

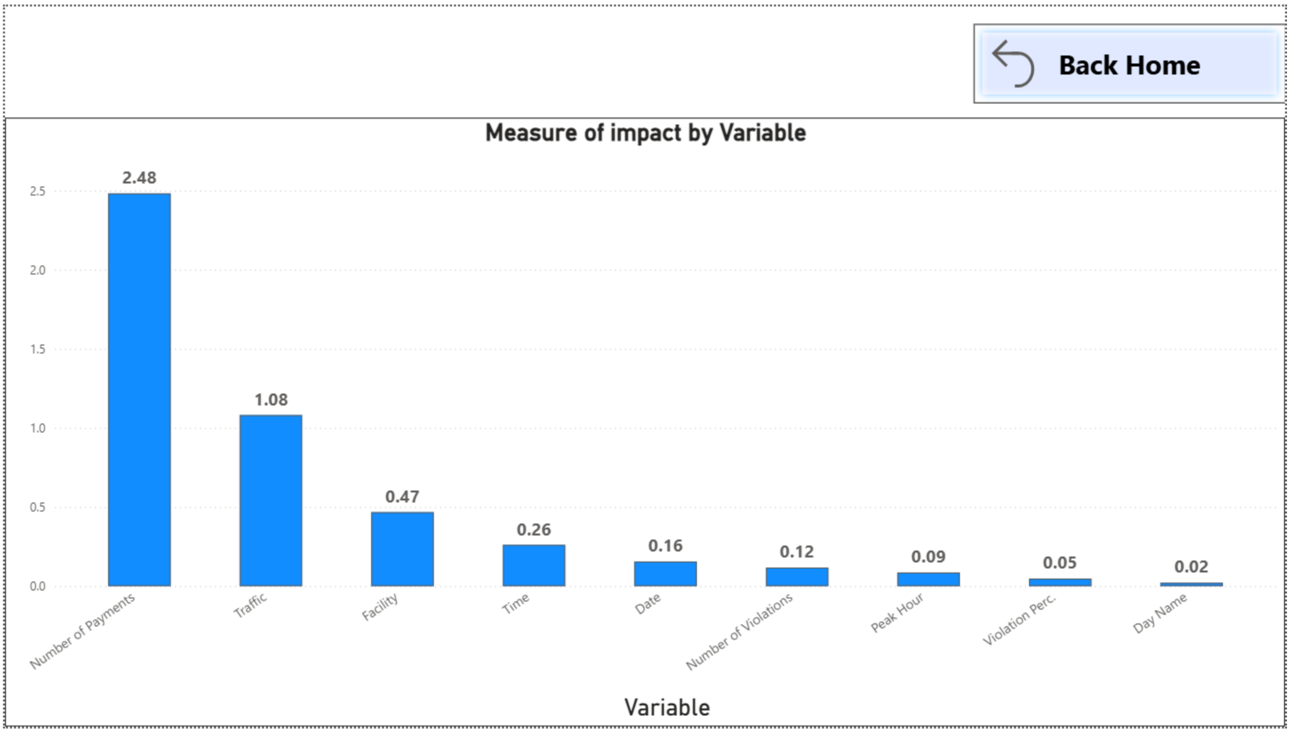

Figure 7: Question 4 - Factors Influencing Revenues Classes

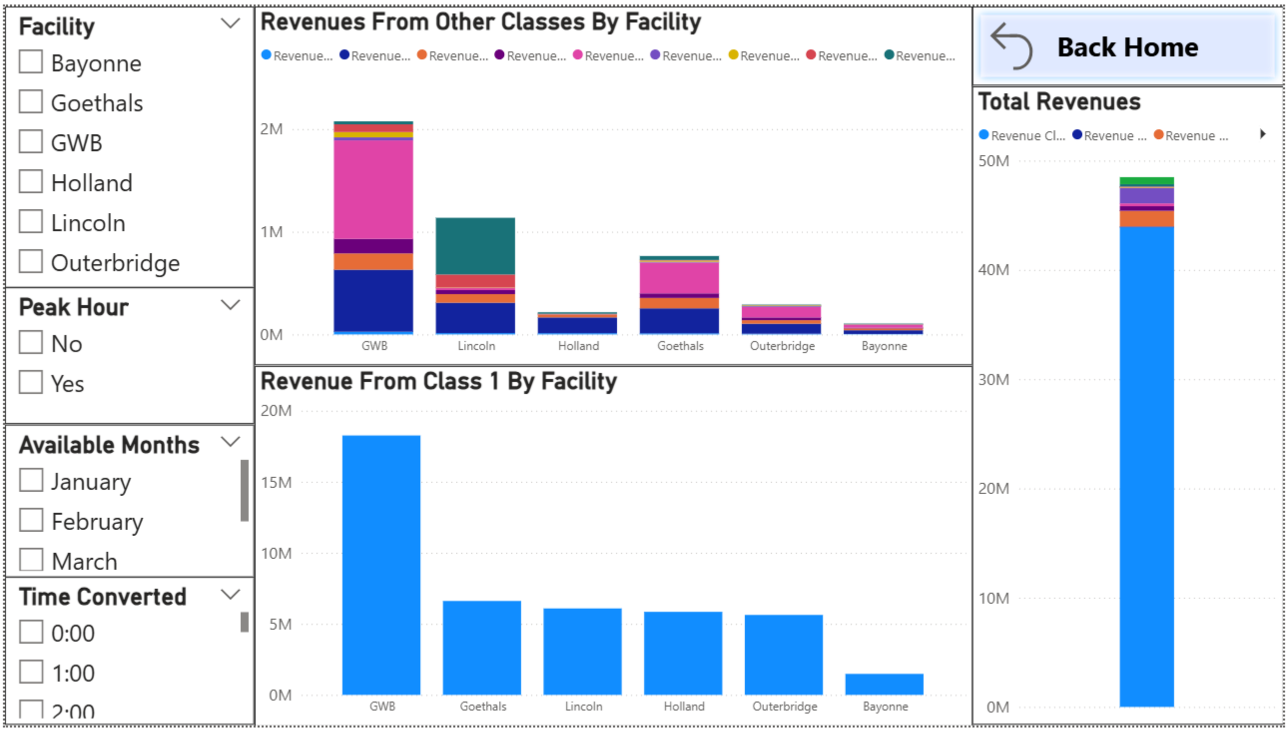

Figure 8: Question 4 - Congestion Due to Pricing in 2025

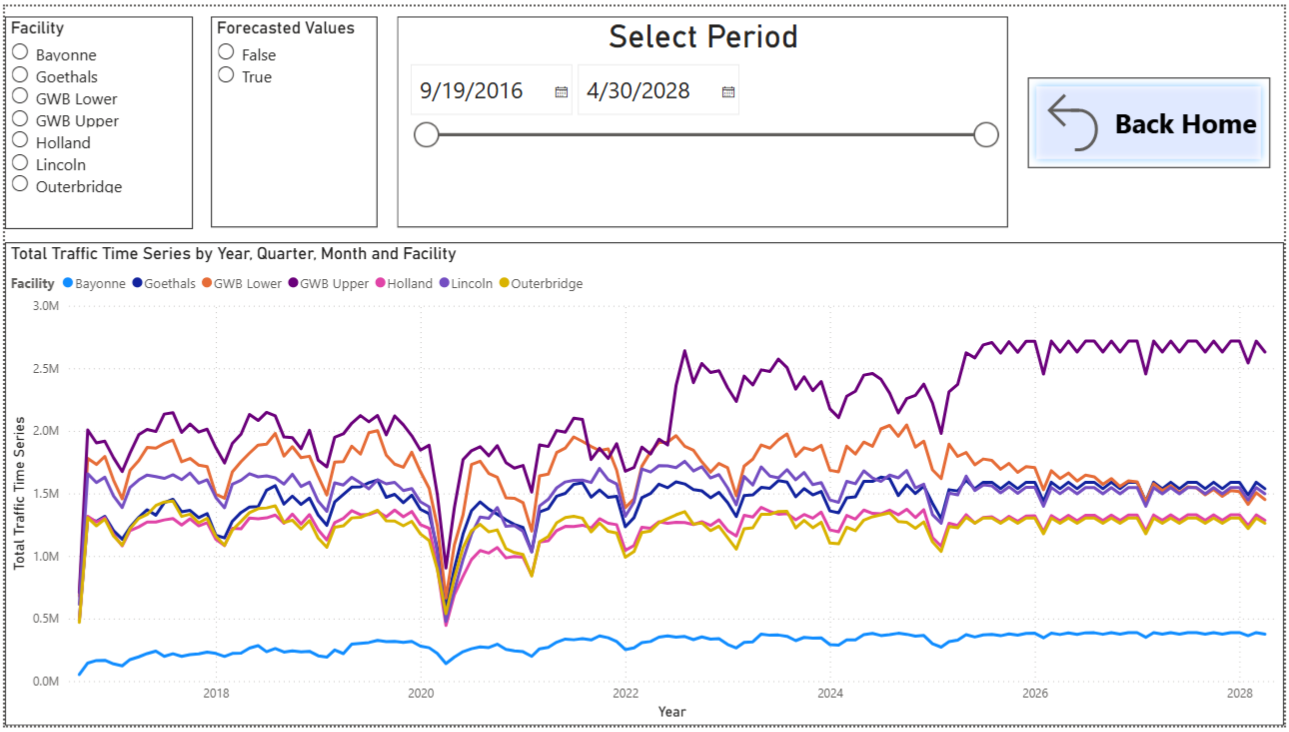

Figure 9: Question 5 - Forecasting Facility Usage Beyond 2025

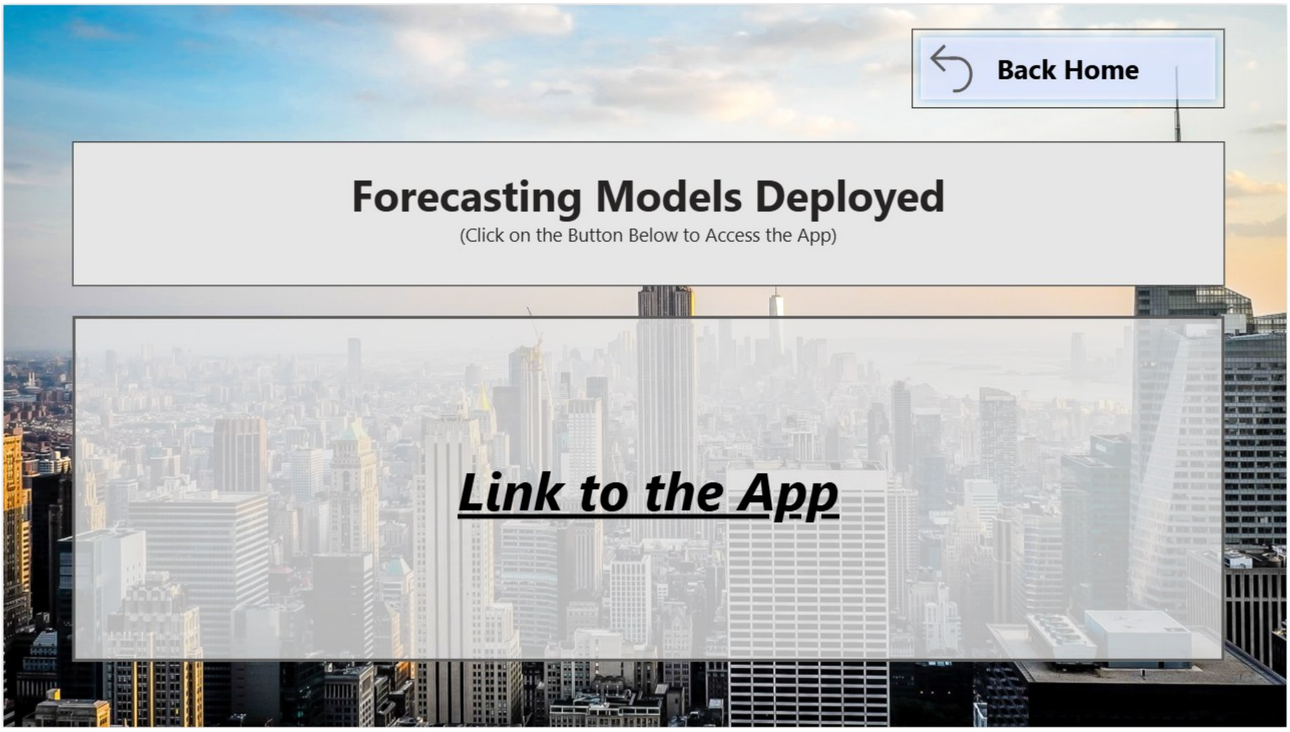

Figure 10: Final Visual - Link to Port Authority Traffic & Revenue Forecasting App