Final Report: The Port Authority Tunnels and Bridges

Team 8

Anthony Denis	Report 2 ; Datasets ; Python Code ; Models ; Dashboards ; Recommendation ; PowerPoint
Pavana Sree Prathyusha Kollu	Report 1 ; Python Code ; Report 3 ; Models ; Recommendations ; Dashboard ; Report 4.

Question 1: What are the factors that affect the usage of bridges and terminals?

- **Important Variables & Dependent Variable:**
    - The company should consider **environmental**, **temporal**, and **facility-based factors**. The most influential variables identified were:
        - **Snow Days**: Strong negative correlation with usage.
        - **Time of Day**: Higher traffic volume observed between **8 AM and 4 PM**.
        - **Temperature (TMAX)** and **Precipitation (PRCP)**: Warmer and drier days tend to increase facility usage.
        - **Facility Code**: Captures structural and geographic differences between facilities.
    - The **dependent variable** used was **EZPASSCount**, which reflects real-time facility usage and is highly reliable for modeling daily and hourly patterns. Its continuous nature makes it well-suited for regression analysis.
- **Recommendations to the Port Authority (How, Usage, Benefit):**

1. **Weather Preparedness**
    - **How:** Integrate real-time weather feeds into traffic forecasting systems and develop contingency plans for snow/rain days.
    - **Usage:** Use forecasts to adjust traffic signage, lane management, and public advisories proactively.

- **Benefit:** Reduces congestion and improves safety, while maintaining consistent traffic flow during inclement weather.

2. **Operational Timing Optimization**

   - **How:** Align staffing, toll operations, and maintenance schedules with peak traffic hours (8 AM – 4 PM).
   - **Usage:** Implement dynamic shift scheduling and maintenance during off-peak hours.
   - **Benefit:** Enhances service efficiency, reduces wait times, and minimizes operational disruptions.

3. **Data Enrichment for Deeper Insight**

   - **How:** Incorporate NYC event calendars, congestion indexes, and real-time alerts into the data pipeline.
   - **Usage:** Use enriched data in predictive models for demand forecasting and scenario analysis.
   - **Benefit:** Improves accuracy of traffic predictions and enables smarter infrastructure planning.

4. **Live Data Integration**

   - **How:** Implement APIs to pull live traffic and weather data into dashboards and models.
   - **Usage:** Deploy real-time analytics for dynamic decision-making.
   - **Benefit:** Allows for immediate operational adjustments and improves responsiveness to sudden demand changes.

- **Models & Performance:**

  - **Multiple Linear Regression (MLR)** using **Azure AutoML** was the primary model used.
  - It achieved an excellent **$R^2 = 0.99$**, indicating a near-perfect fit between predictors and EZPass usage.

- **Tools Used:**
  - Microsoft Access (data joins and extraction)
  - Python & Excel (cleaning and preprocessing)
  - Azure AutoML (modeling)
  - Looker Studio (dashboard visualization)
  - PowerPoint (presentation)

Question 2: How many toll violators are there, and where/when do they occur?

- **Important Variables & Dependent Variable:**
    - This analysis required a classification approach to identify **high-risk violation scenarios**. Key variables included:
        - **Facility Code**: Certain bridges, especially **GWB Lower Span**, saw the highest violation rates.
        - **Time Block (Hour of Day)**: Violations spike during **late night and early morning hours**.
        - **EZPASSCount**: High traffic volumes are correlated with more violations.
        - **Day of Week**: Mondays and Fridays had more violations, possibly linked to commuter behavior.
    - The **dependent variable** was a **binary label "ViolationLevel" (High vs. Low)**, generated using a threshold-based classification on violation counts.

Recommendations to the Port Authority (How, Usage, Benefit):

1. **Automated Enforcement via ALPR**
    - **How:** Install high-resolution cameras with ALPR at high-violation facilities like GWB Lower Span.
    - **Usage:** Continuously scan and record plate numbers during violation-prone hours (late night/early morning).
    - **Benefit:** Boosts toll collection rates, deters violations, and reduces manual enforcement costs.

2. **Dynamic Fine System**
    - **How:** Develop a tiered fine structure using ML-based violation risk scores (e.g., higher fines for frequent violators).
    - **Usage:** Integrate into billing and enforcement platforms.
    - **Benefit:** Encourages compliance, targets chronic violators, and maximizes fine revenue.

3. **Driver/Vehicle Profiling**
    - **How:** Enrich models with vehicle type, license plate state, and historical violation behavior.
    - **Usage:** Focus enforcement efforts on high-risk groups or out-of-state drivers.
    - **Benefit:** Improves detection accuracy and optimizes enforcement resource allocation.

- **Models & Performance:**
    - Classification models tested included **Random Forest**, **Gradient Boosting**, and **Logistic Regression** via Azure AutoML.
    - The **Gradient Boosting model** performed best, achieving **accuracy = 97.24%** and **AUC = 0.997**, suggesting exceptional discrimination power between high and low violation instances.

- **Tools Used:**
    - Microsoft Access, Excel, Python, Azure AutoML, Looker Studio, PowerPoint

Question 3: What are the busiest times of the year and how are traffic and speeds affected by seasonality, weather, holidays, and NYC events?

- **Important Variables & Dependent Variable:**
    - This time series problem focused on identifying **seasonal and event-driven traffic patterns**. The most significant features were:
        - **EZPASSCount** (target variable)
        - **Traffic Date** and derived
        - **Weather Indicators** (TMAX, PRCP, SNOW)
        - **NYC Event/Holiday Flags** (manually added from city event calendar)
    - **EZPASSCount** was again the **dependent variable**, due to its capacity to reflect fluctuations in traffic volume across time and season.

Recommendations to the Port Authority (How, Usage, Benefit):

1. **Holiday & Event Traffic Planning**
    - **How:** Use calendar-based forecasting (Prophet/ARIMA) to predict spikes during known holidays and city events.
    - **Usage:** Plan toll booth staffing, traffic signage, and alternate routing accordingly.
    - **Benefit:** Minimizes congestion and enhances traveler experience during high-volume periods.

2. **Adjust to Remote Work Patterns**
    - **How:** Monitor weekday vs. weekend usage shifts and adjust operational models post-pandemic.

- **Usage:** Reallocate resources (e.g., reduced weekday staffing) and modify forecasting algorithms.
- **Benefit:** Aligns resources with new commuter behavior, reducing costs and improving efficiency.

3. **Include Real-Time Speed & Congestion Data**
   - **How:** Integrate sensors or partner with GPS traffic providers (e.g., Waze, INRIX).
   - **Usage:** Adjust dynamic tolling or open extra lanes when congestion is high.
   - **Benefit:** Enhances travel time reliability and supports smarter demand management.

4. **Public Transport Disruption Monitoring**
   - **How:** Set up alerts for MTA/NYC transit service changes.
   - **Usage:** Predict spillover demand on bridges and proactively communicate alternate options.
   - **Benefit:** Prevents unexpected surges and enhances multimodal coordination.

- **Models & Performance:**
  - Time series models applied included:
    - ARIMA (AutoRegressive Integrated Moving Average)
    - Prophet Time Series
  - The Prophet model performed well, with **R^2 = 96.93%** and **MAE = 833**, effectively capturing trend, seasonality, and event-driven spikes.

- **Tools Used:**

  - Microsoft Access, Excel, Python, Azure AutoML, Looker Studio, PowerPoint

Project Dashboard:
https://lookerstudio.google.com/reporting/f7ef7580-3c10-40c1-9b0f-df0d3528b385

Datasets attached :

- **Data_1 → Dashboard 1**
- **Data_2 → Dashboard 2**
- **Data_3 → Dashboard 3**