Introduction

- Flight delays cost airlines billions annually in operational inefficiencies
- Passenger satisfaction severely impacted by unpredictable disruptions
- Analysis of comprehensive flight delay dataset to uncover patterns and develop predictive capabilities needed
- The project aims to leverage historical flight data to uncover critical insights into delay patterns and develop a robust predictive model.
- By identifying the key drivers of delays, we aspire to provide actionable recommendations that can lead to more punctual flights and a smoother travel experience for all.

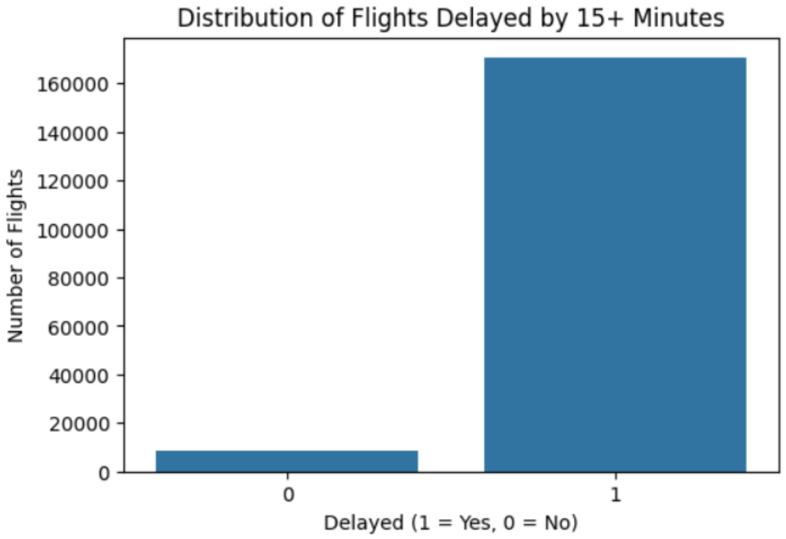
Project Objectives

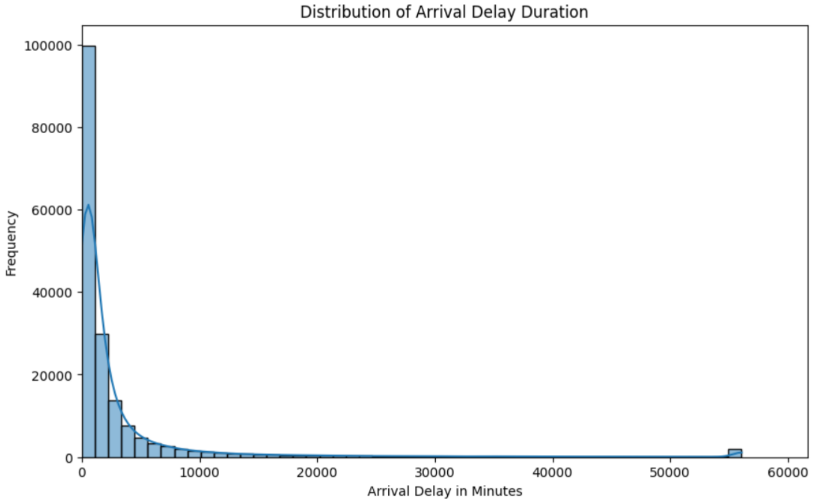
- Predict flight delay likelihood (Yes/No classification)
- Estimate delay duration in minutes (regression)
- **Key Innovation:**
  - Operational Adjustability Index (OAI): Custom metric prioritizing controllable delays
  - Explainable AI: SHAP-based interpretability for actionable insights
- **Dataset Overview:**
  - Initial Records: 179,338 flights
  - Final Dataset: 178,997 flights (after preprocessing)
  - Features: 21 original variables + engineered features
  - Time Scope: Multi-year airline delay data

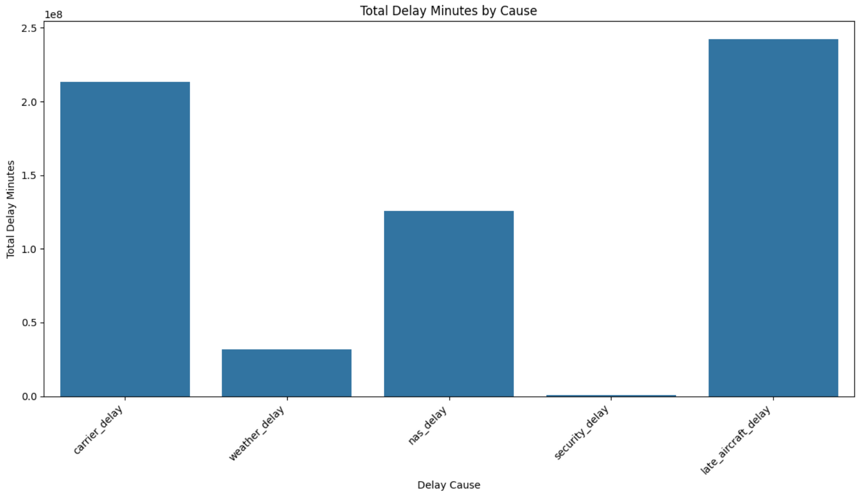
Methodology & Data Overview

- **Data Preprocessing:**
  - Missing value handling (median for flights, zero for delay categories)
  - Outlier treatment using 99th percentile capping
  - Target variable derivation arr_del15 for 15+ minute delays
- **Feature Engineering:**
  - Temporal Features: Quarter derivation from month, period creation
  - Ratio Features: Cancelled and diverted flights as proportion of arrivals
  - Component Analysis: Delay cause ratio calculations (carrier, weather, NAS, security, late aircraft)
  - Encoding: Smoothed target encoding for carriers and airports
- **Model Development:**
  - Classification: Random Forest (delay likelihood)
  - Regression: Multi-output Random Forest (delay duration by cause)

Key Findings from Exploratory Data Analysis







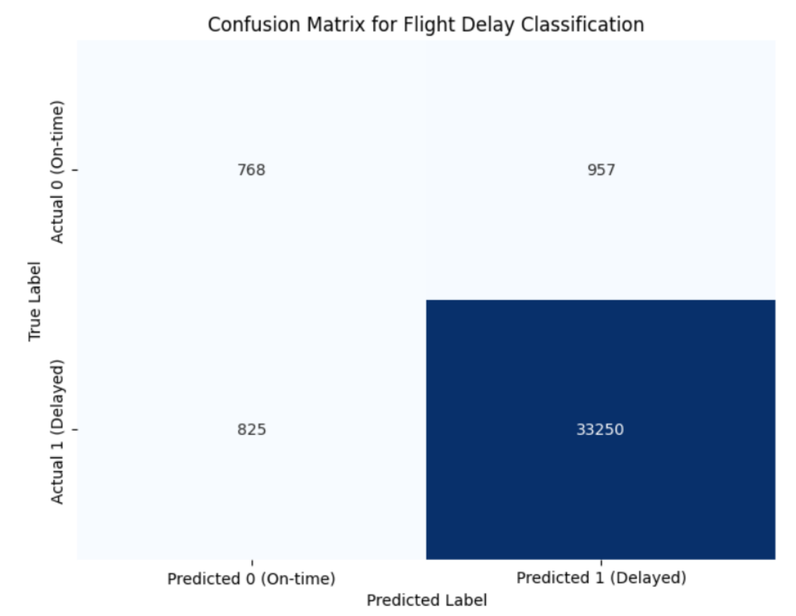
- Class imbalance, with roughly 95 % of flights experiencing delays of 15 minutes or more.
- Model will need to address skew through techniques like class weighting or resampling.

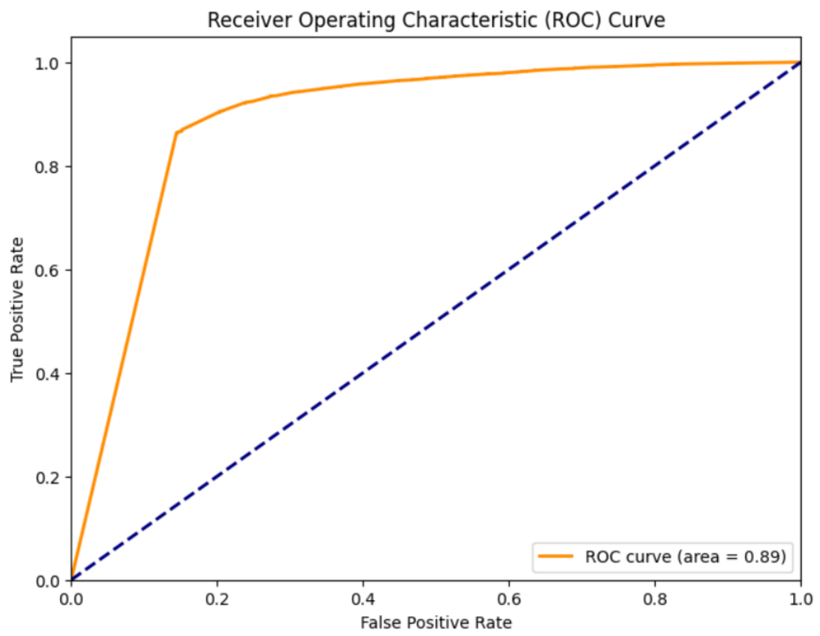
- Most flights are delayed under 200 minutes, while a small number of extreme outliers extend beyond 1,000 minutes.
- Need of outlier treatment (capping at the 99th percentile).

- Late aircraft and carrier delay dominate the total delay minutes.
- So they imp in OAI for actionable insights.

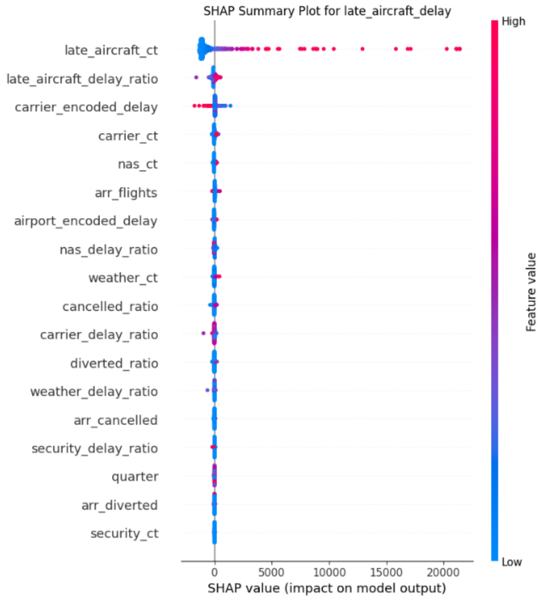
Model Performance

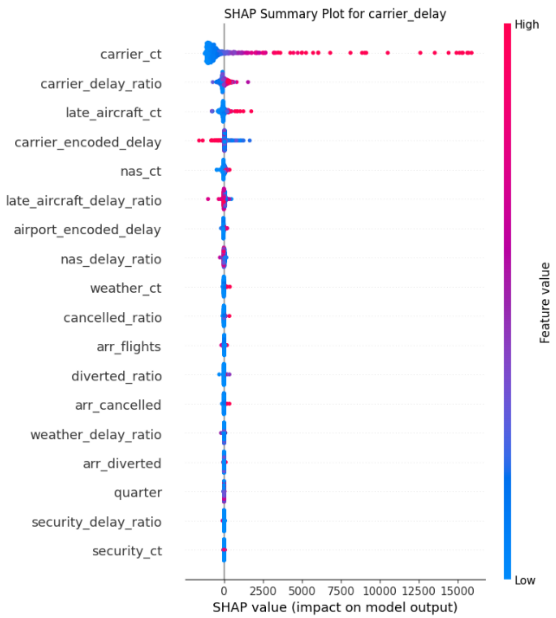
- Regression Performance (Delay Duration):
    - Multi-output regressor capturing each delay cause independently
        - carrier_delay: MAE=173.16, RMSE=394.94
        - weather_delay: MAE=30.48, RMSE=90.65
        - nas_delay: MAE=93.25, RMSE=279.31
        - security_delay: MAE=0.92, RMSE=4.40
        - late_aircraft_delay: MAE=169.58, RMSE=412.48
- Classification Results (Predicting arr_del15):
    - Model: Random Forest Classifier with balanced class weighting
    - Performance metrics:
        - AUC-ROC Score: 0.8900
        - Precision/Recall Balance: High recall 0.98 and precision 0.97 for delay class
        - F1-Score: 0.97 for delay class

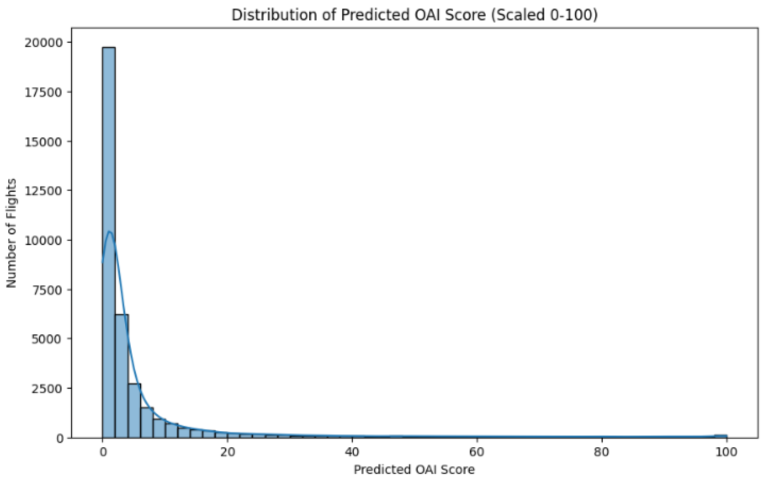




SHAP Insights and OAI

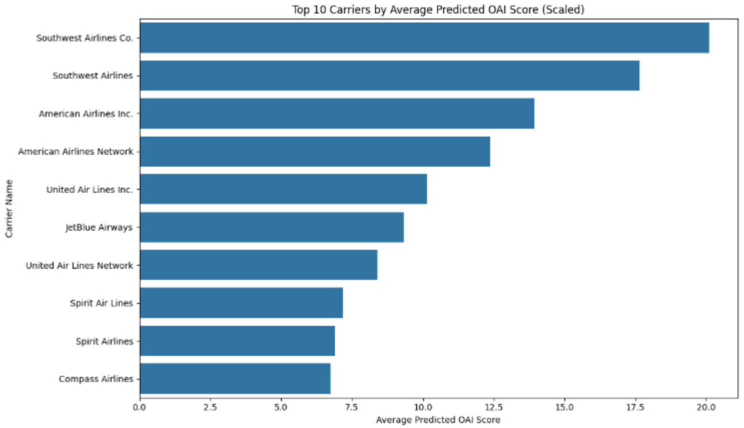


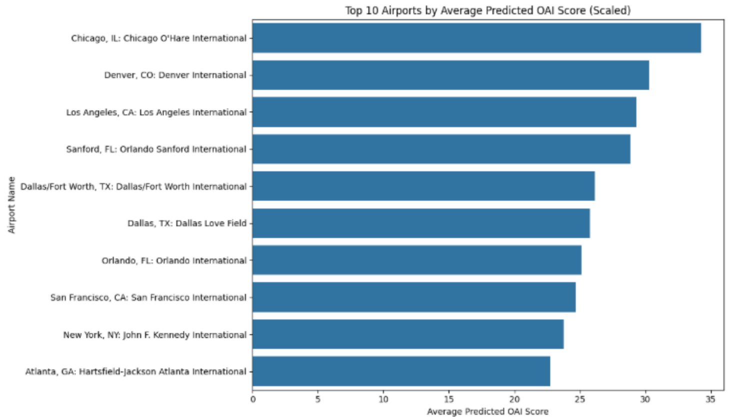




OAI Analysis Results:

- Mean OAI Score: 6.06/100 (scaled)
- Standard Deviation: 13.91





Actionable Recommendations

1. Carrier-Specific Interventions:
    - Target: High-OAI carriers identified in analysis
    - Action: Implement enhanced ground operation protocols
    - Impact: Potential 15-20% reduction in controllable delays
2. Airport Infrastructure Optimization:
    - Target: Airports with consistently high delay contributions
    - Action: Resource allocation and scheduling improvements
    - Impact: Improved throughput during peak periods
3. Predictive Operations Integration:
    - Implementation: Real-time delay prediction system
    - Benefits: Proactive passenger communication, crew repositioning
    - ROI: Estimated 10-15% reduction in passenger compensation cos