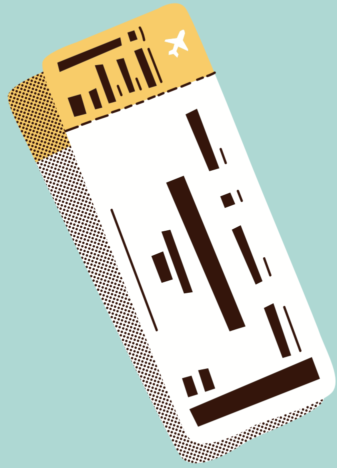
OPTIMIZING AIR TRAVEL DATA DRIVEN ANALYSIS

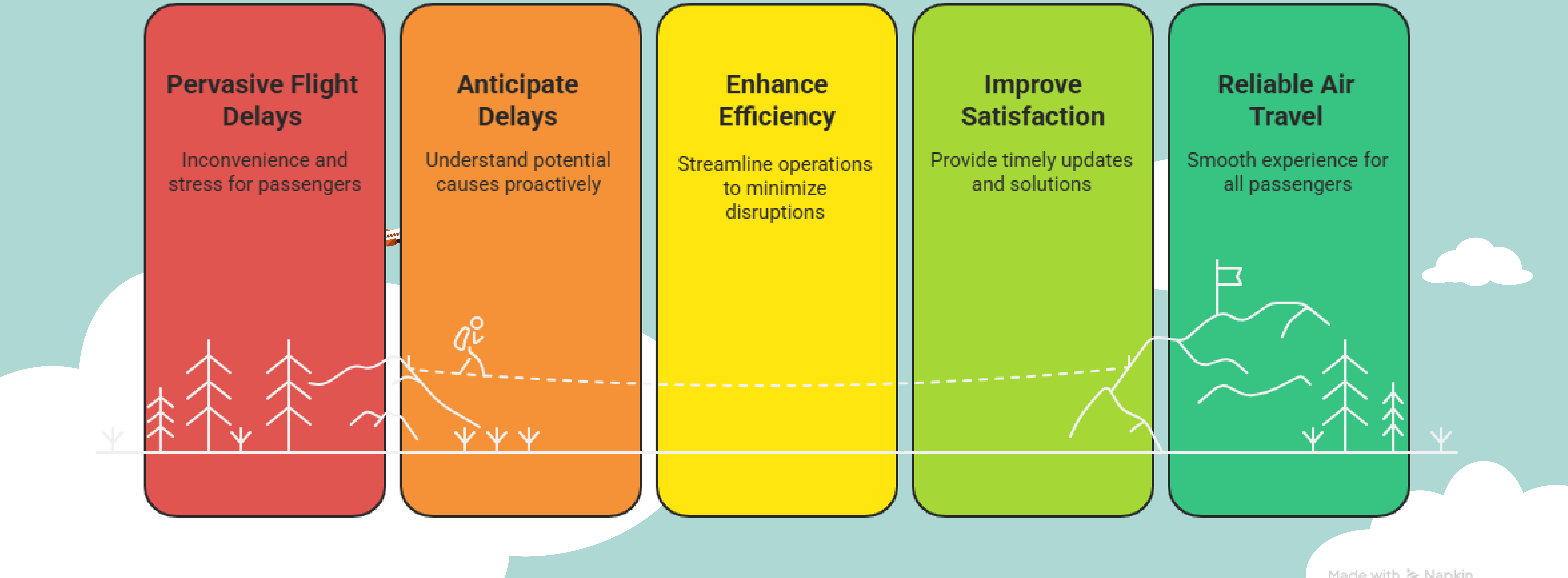


Leveraging Data for a Smoother Travel Experience

Chaman Singhal
4th year (Meta)
22118019

The Challenge of Flight Delays

From Flight Delays to Smooth Air Travel



OBJECTIVES AND METHODOLOGY

1. Understanding Delay Patterns:

Analyzing historical flight data to uncover trends, pinpoint key delay contributors (such as carrier, weather, National Airspace System issues), and identify seasonal impacts.

2. Predicting Delay Risk (Classification):

Developing a binary classification model to accurately determine if a flight is likely to be delayed.

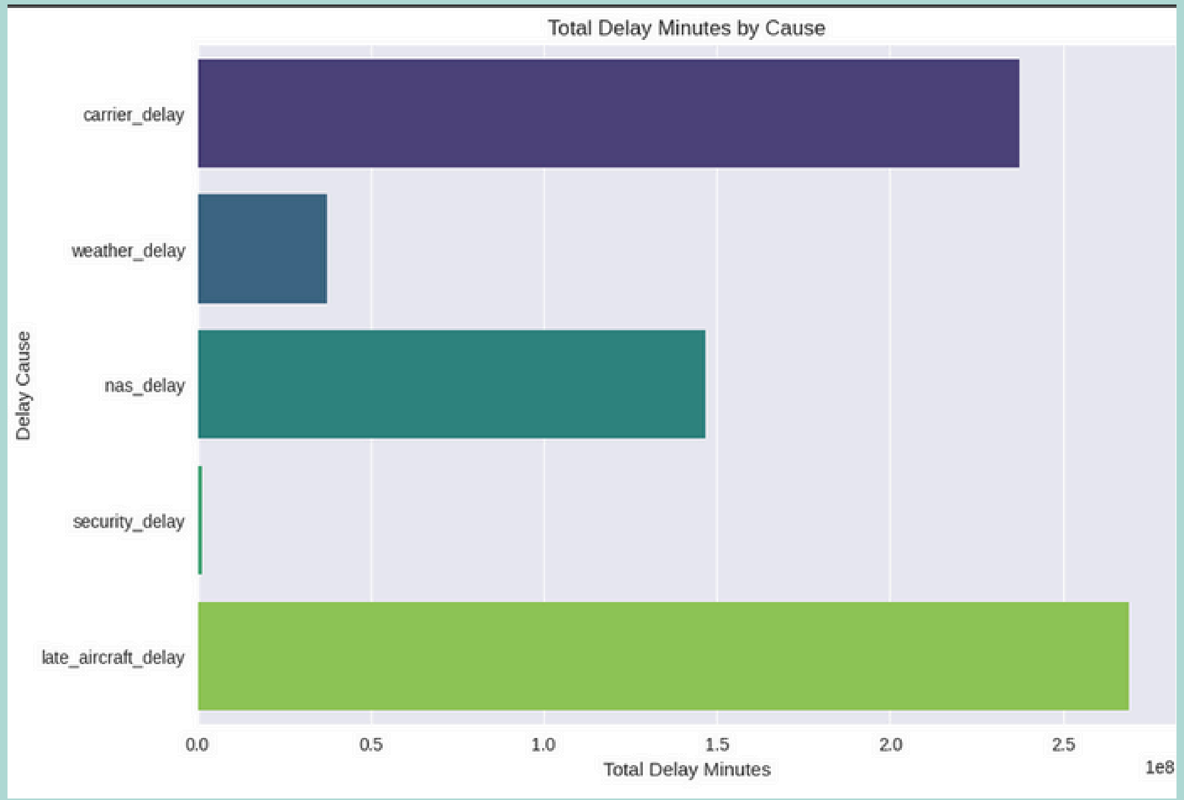
3. Estimating Delay Duration (Regression):

Creating a regression model to forecast the expected arrival delay in minutes, utilizing features like flight volume, historical delay data, and operational indicators.

4. Focusing on Controllable Delays:

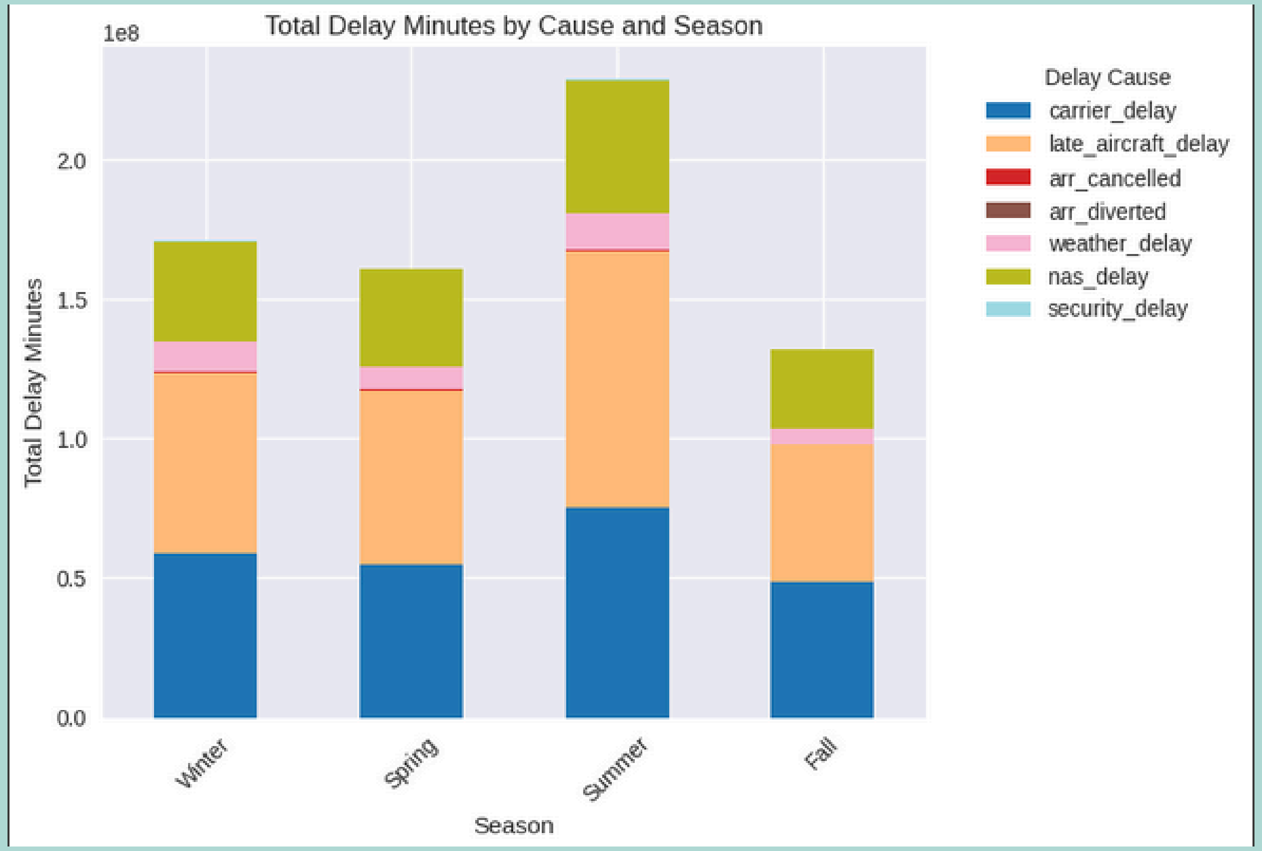
Introducing and applying an Operational Adjustability Index (OAI) to assign more weight to delays that airlines can directly control (e.g., carrier-specific issues, late aircraft arrivals), enabling targeted interventions.

EDA Insights



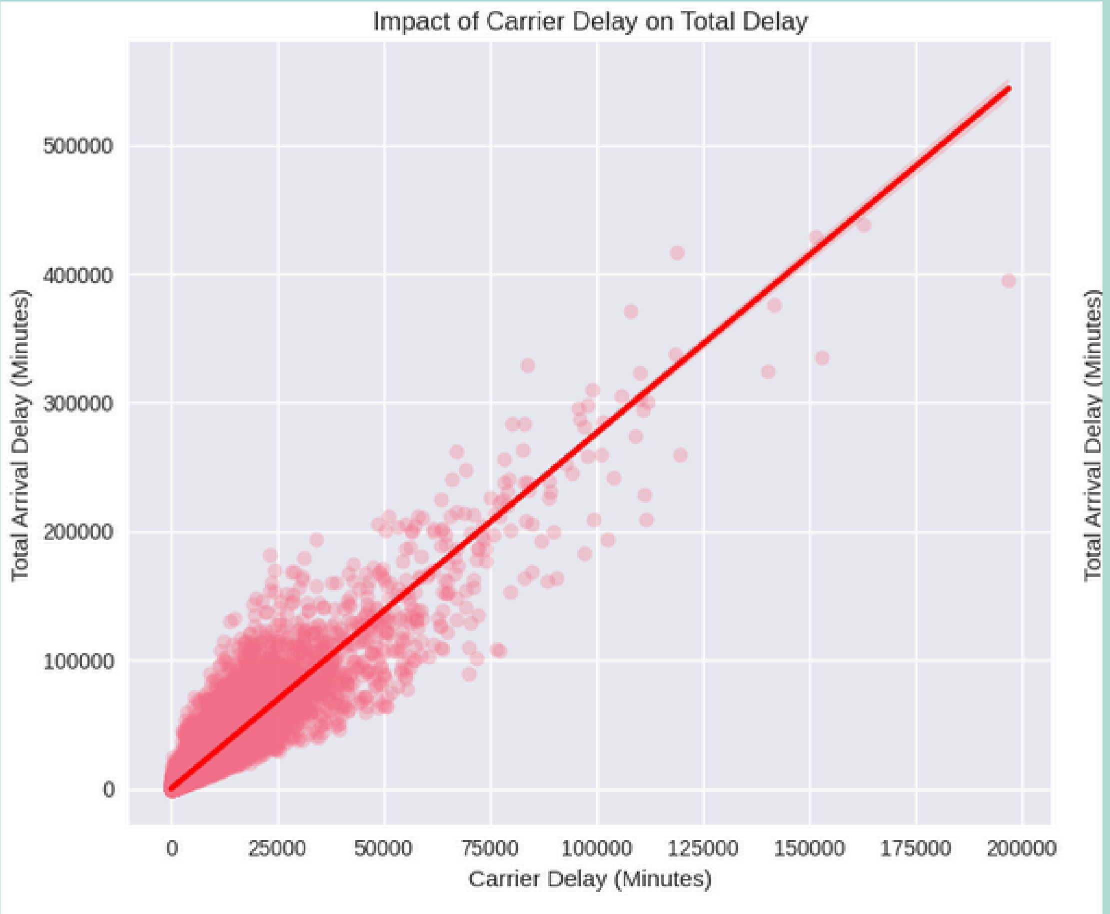
Total Delay Minutes by Cause

Late aircraft delays and carrier delays are the leading contributors to total delay minutes, while security delays have minimal impact.



Total Delay Minutes by Cause and Season

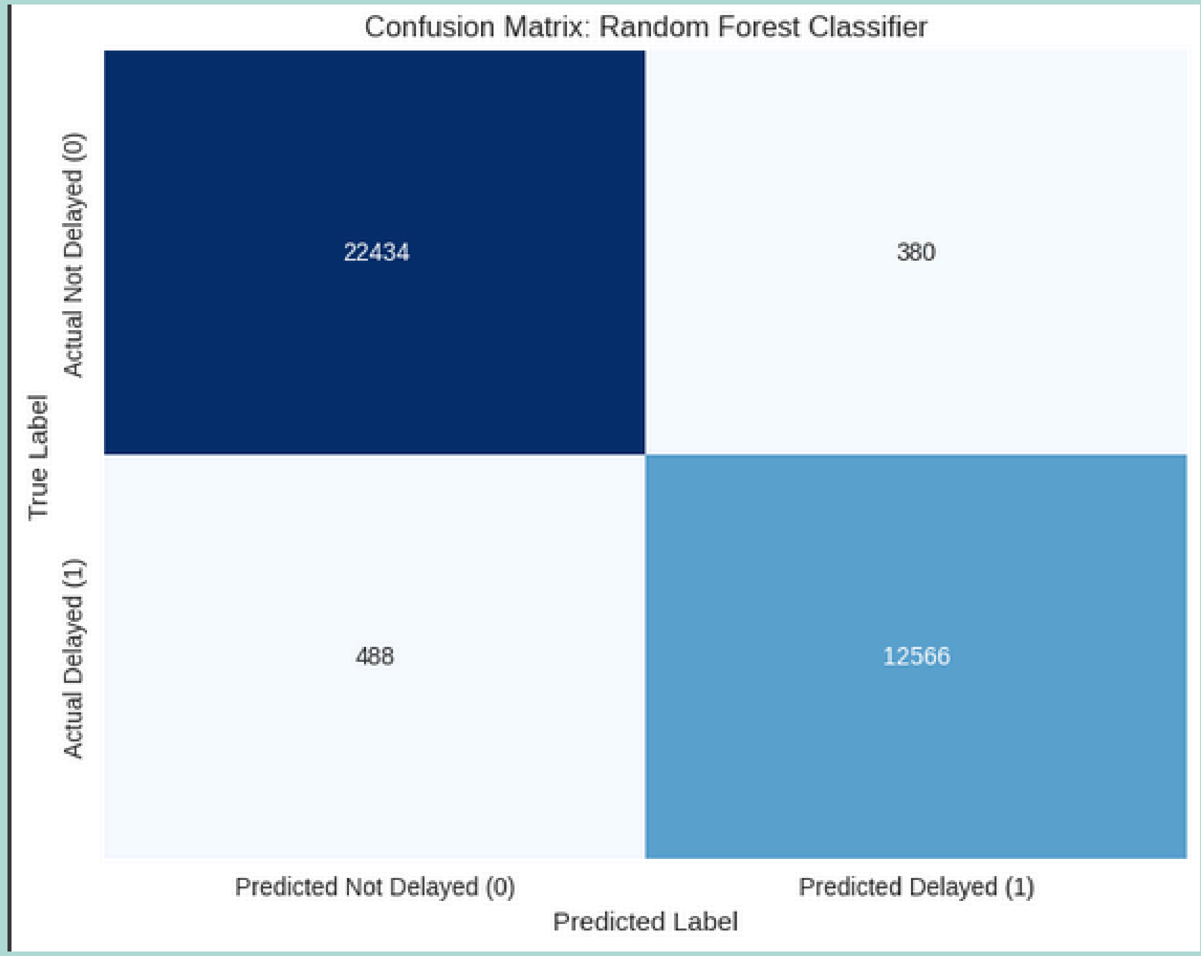
Summer shows the highest total delay, primarily due to late aircraft and carrier delays.
Fall experiences the least overall delays, indicating potentially smoother operations.

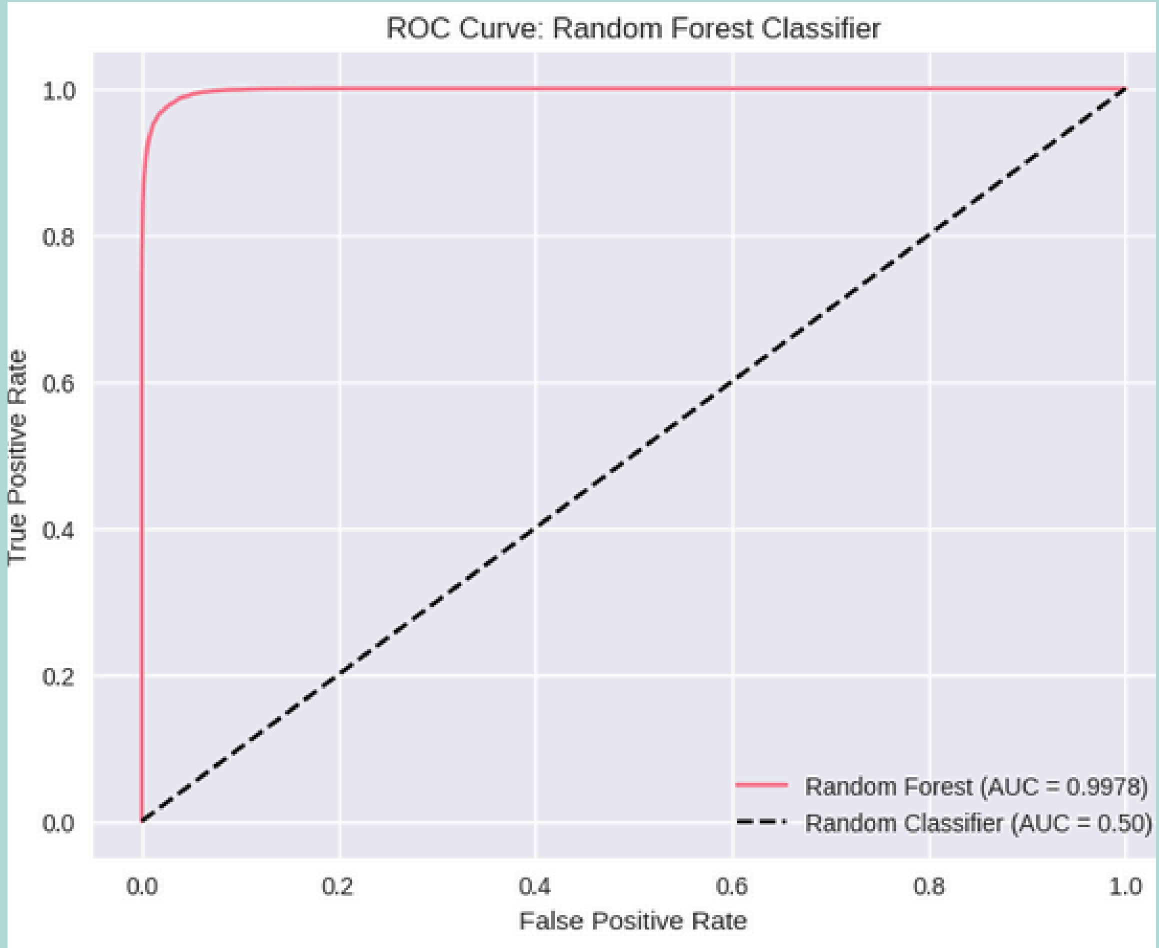


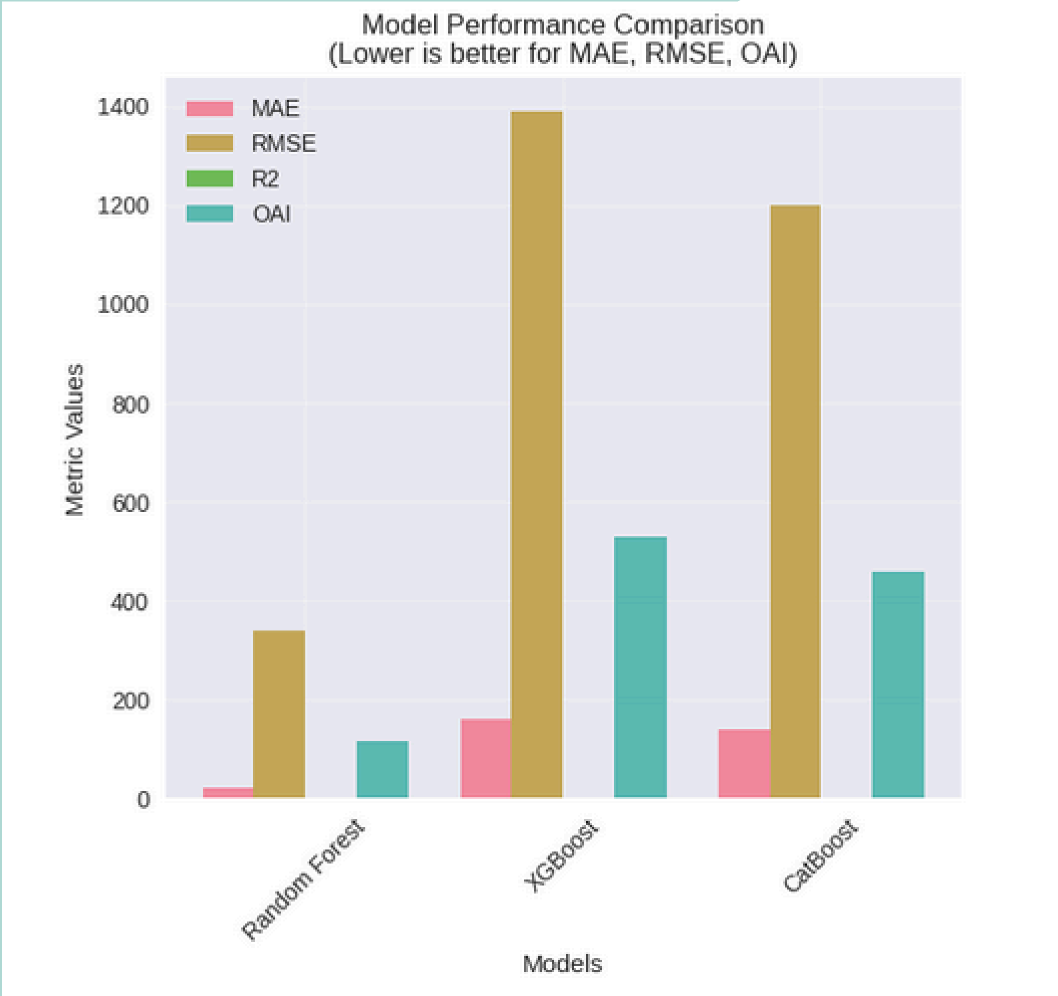
Impact of Carrier Delay on Total Delay

There is a strong positive linear correlation between carrier delay and total arrival delay, indicating that carrier-related issues are a major driver of overall delays.

MODEL PERFORMANCE







Classification (Random Forest Classifier):

Achieved 97.6% accuracy with an AUC of 0.9978, indicating excellent classification capability. The confusion matrix shows very low false positives and false negatives, reflecting high reliability in predicting delays → Best Model for Classification.

Regression (Random Forest Regressor):

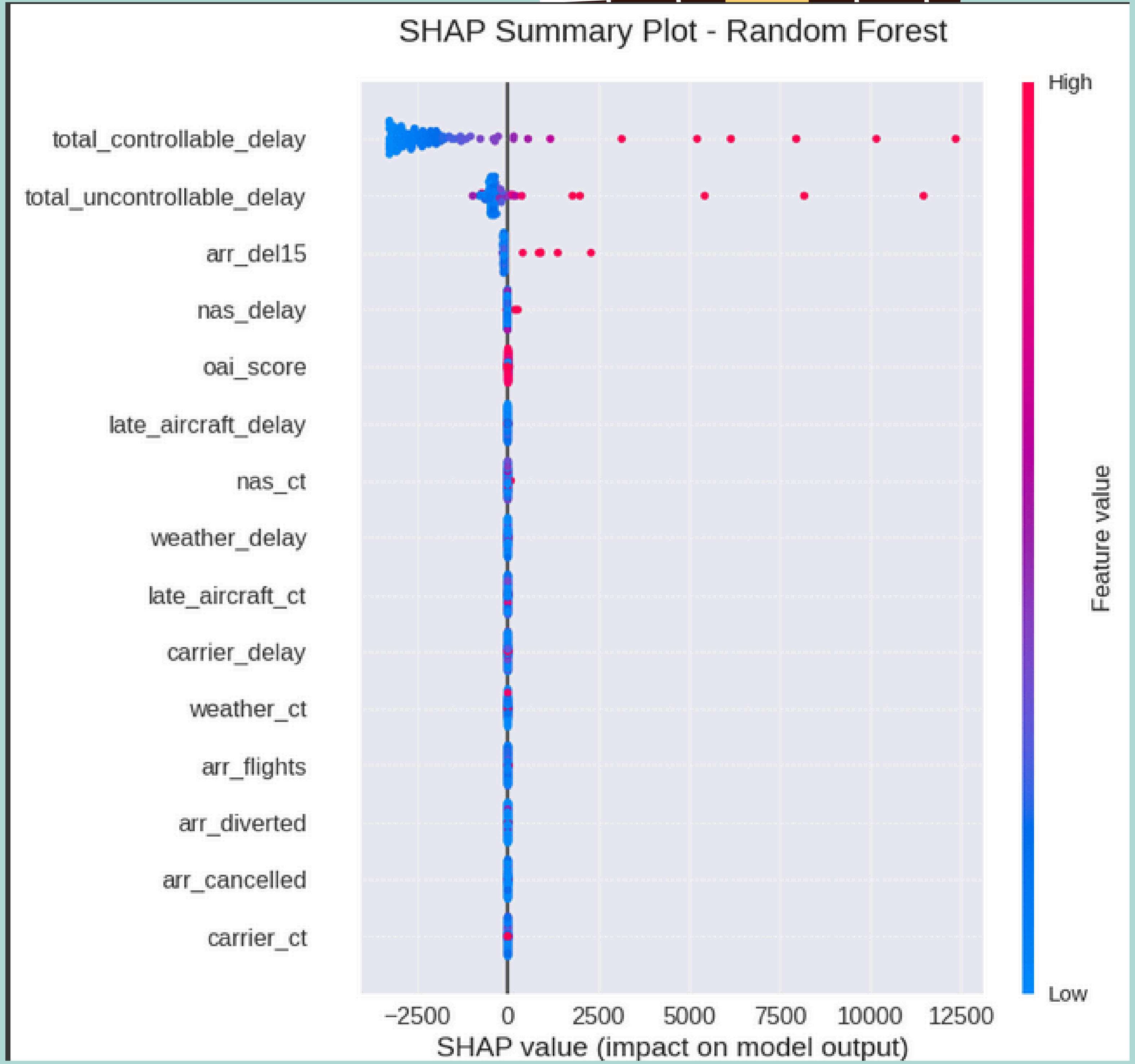
Outperformed Linear and Gradient Boosting models in MAE, RMSE, and R^2, showcasing superior predictive accuracy for estimating delay durations. Its robustness and ability to capture non-linear patterns make it ideal for complex delay scenarios.

SHAP ANALYSIS AND OAI

Key Takeaways from SHAP Analysis (Random Forest):

Top Influential Features are
total_controllable_delay dominates model output influence
total_uncontrollable_delay, arr_del15, and nas_delay follow

OAI-Weighted SHAP refines importance by emphasizing controllability (e.g., carrier & late aircraft delays).



OAI (Operational Adjustability Index) Highlights:

Avg. OAI Score: 0.85 → Most delays are highly controllable.
89% of flights are in the High Controllability zone (OAI > 0.7)

Reducing 50% of controllable delays could save ~251M minutes/year
Estimated cost savings: $25.18 billion/year

ACTIONABLE RECOMMENDATIONS

Key Delay Drivers:

SHAP analysis shows that late aircraft delays, carrier delays, and the OAI score (a measure of airline efficiency) are strong predictors of overall delays—especially when weighted for controllability.

EDA vs. Model Insights:

While controllable delays are more common (as seen in EDA), the model reveals that uncontrollable factors like weather and NAS (air traffic delays) have a greater impact when they do happen.

High-Impact Uncontrollable Events:

Though less frequent, delays from weather and NAS issues are often severe, particularly during peak traffic times, causing major disruptions to flight schedules.

Operational Bottlenecks & Recommendations

Traffic Bottlenecks Identified

Major hubs like LAX and Chicago O'Hare showed high volumes of flights and delays, especially during peak daytime hours—indicating capacity strain.

Delay Risk Increases with Load

More flights per day led to a higher likelihood of delays, especially where operations cluster around specific time windows.

Carrier & Route Patterns

Certain airlines (e.g., Southwest, Spirit) and routes were more prone to controllable delays, showing variation in operational resilience.

THANK YOU