# OPTIMIZING AIR TRAVEL

Open Projects 2025 – Analytics Track"

Sahil Hussain 23124031

# PROBLEM STATEMENT & PROJECT OBJECTIVES

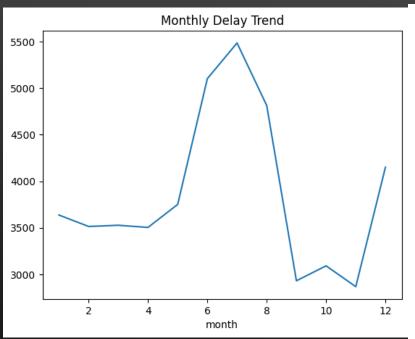
#### • **7** Problem:

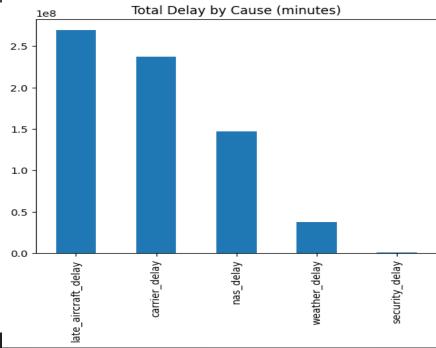
Flight delays cause major inconvenience to passengers and cost airlines millions. Understanding and anticipating delays is critical for better planning.

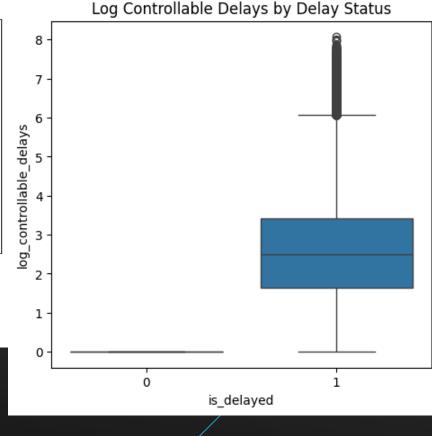
#### • Objectives:

- Analyze historical flight data to uncover delay patterns.
- Build predictive models for:
  - Delay occurrence (Yes/No)
  - Delay duration (in minutes)
- Interpret model decisions using SHAP
- Analyze controllable and external cause of delay.
- Provide actionable recommendations with operational focus.

# EXPLORATORY DATA ANALYSIS – DELAY TRENDS & CAUSES







Monthly Delay Patterns:
Delays peak during summer
(June–August) and December,
aligning with travel demand
surges.

**Lowest delays** occur in September–November.

#### **Delay Causes:**

Late Aircraft and Carrier-related issues are the dominant contributors.

NAS (Airspace) and Weather also significant, but secondary.

Security delays are negligible.

#### **Controllability Analysis:**

Controllable delays (carrier/aircraft) show much higher impact on delayed flights.

#### FEATURE ENGINEERING & MODEL PERFORMANCE

#### **New Features Created** (all based on safe, preflight data):

- •carrier\_delay\_rate: Historical average delay % for each airline
- •airport delay rate: Historical delay % for each airport
- •route\_delay\_rate: Combined airport-airline route delay %
- •season: One-hot encoded seasonal labels from month
- •carrier\_name, airport\_name: One-hot encoded categorical features

### Leaky Features Dropped (cannot be used for real-time prediction):

arr\_delay, arr\_del15: Targets
carrier\_delay, late\_aircraft\_delay, nas\_delay, weather\_delay, security\_delay: Post-flight known causes (would leak label info)
arr cancelled, arr diverted, delay counts: Outcomes, not inputs

#### Why?

These features reflect what happened **after** the flight, not what would be known **before** — using them would cause data leakage and make the model unrealistic.

#### Classification Model

"We ensured the model uses only preflight features, dropping all delay cause columns that reflect outcomes.

Then, we engineered new features like historical delay rates per carrier, airport, and route.

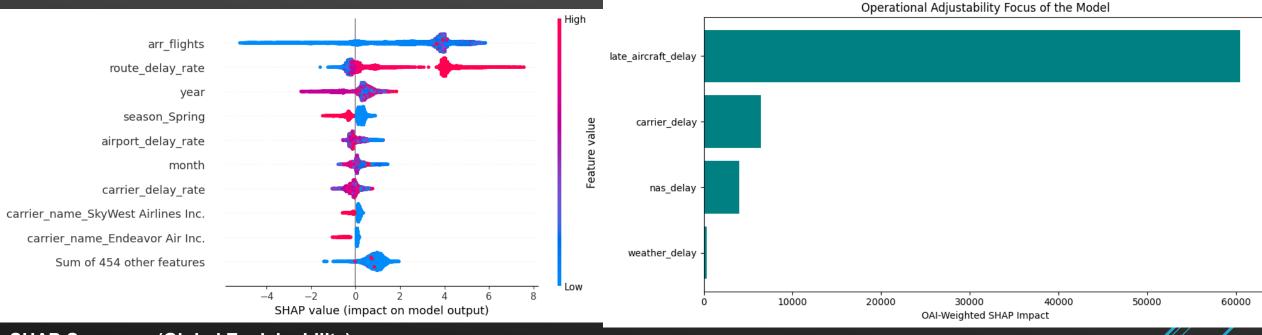
Among three classifiers, **XGBoost** performed best — delivering a 95% accuracy and a strong F1-score of 0.97."

#### Regression Model

#### Regression (Gradient Boosting):

MAE: ~10.5 minutes RMSE: ~18.5 minutes R square score- 0.74

# INTERPRETING THE MODEL: SHAP & OPERATIONAL FOCUS



#### **SHAP Summary (Global Explainability)**

- •Used SHAP to understand which features most influenced predictions
- •Top Influential Features (from SHAP beeswarm plot):
  - 1.arr flights Number of arrivals for airport
  - 2.route delay rate Historical route delay likelihood
  - 3.season\_Spring, airport\_delay\_rate, carrier\_delay\_rate

#### Insight:

The model heavily relies on **historical delay patterns**, seasonal context, and traffic volume — all available pre-flight.

#### **Conclusion:**

Model learns from operational patterns, not leaked outcomes

#### OAI SHAP Analysis (Operational Focus)

Used OAI to measure how much of the model's attention targets controllable delays.

**Top Contributors** 

Late aircraft delay: Highest impact (controllable)

Carrier delay: Strong secondary factor

OAI Focus Score: 93.91%

Insight:

"Over 93% of the model's weighted attention is on controllable delays—making it highly actionable for airline operations."

#### ACTIONABLE RECOMMENDATIONS: REDUCING DELAYS EFFECTIVELY Controllable Delays

#### ☐ Primary Delay Sources (Controllable):

- Late Aircraft Turnarounds
- Carrier-Driven Operational Issues (e.g., crew, maintenance, gate delays)
  - ☐ Late Aircraft Delays Streamline turnaround and gate operations at key hubs. Reduce reliance on tight aircraft reuse.
  - □ Carrier Delays

Enhance crew scheduling and maintenance planning. Focus on carriers with high SHAPdriven delay impact.

□ NAS Delays Mitigate through smarter routing and flight timing, guided by SHAP-identified congestion nattorns

#### External Delay

- ☐ External Delay Sources:
- Weather
- Airspace/NAS
- Security
- ☐ Recommendations:
- ☐ Leverage predictive insights to inform passengers of potential risk in advance
- ☐ Re-route or reschedule proactively when high-risk weather conditions are predicted
- ☐ Advocate with regulators for priority slots on historically delayed NAS corridors
- ☐ Use airport historical data to avoid congestion-prone time slots

#### Why?

While not directly controllable, these delays can be anticipated and communicated to reduce passenger impact.

## THANK YOU