# A Data-Driven Approach to Flight Delay Prediction and Operational Optimization

Yash Arora





A Data-Driven Approach to Flight Delay Analysis and Prediction

### **Project Overview**

**Dataset:** 179,338 flight records (2015-2023)

**Objective:** Predict delays & provide actionable insights **Approach:** EDA + Machine Learning + SHAP Analysis



## **©** Project Objectives & Methodology



Comprehensive EDA to identify delay trends, causes, and correlations across 179k flight records from 2015-2023



Build robust ML models for delay occurrence (classification) and duration prediction (regression)

## **Generate Actionable Insights**

Provide data-backed recommendations using SHAP analysis to distinguish controllable vs. external factors

**6** Key Innovation: Operational Adjustability Index (OAI)

Custom evaluation metric prioritizing controllable delays (carrier & late aircraft) to focus interventions where airlines have direct operational control.

## **Explainable Al Approach**

SHAP (SHapley Additive exPlanations) provides transparency by showing exactly **why** each prediction was made, enabling targeted operational decisions.

## **Key EDA Findings**

73.2%

Controllable Delays

38.9%

Late Aircraft Impact





### Seasonal Patterns

**Summer Peak:** All delay types intensify during June-August

Cascading Effect: Late aircraft delays create ripple effects (r=0.97 correlation with total

delays)

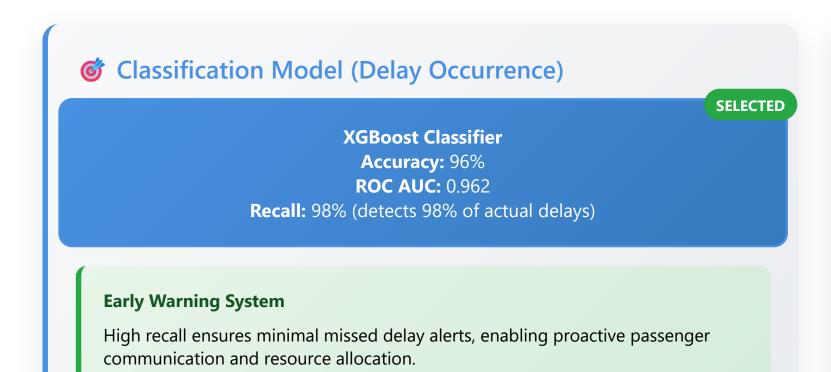
Airport Congestion: High arrival flight volumes significantly increase delay probability

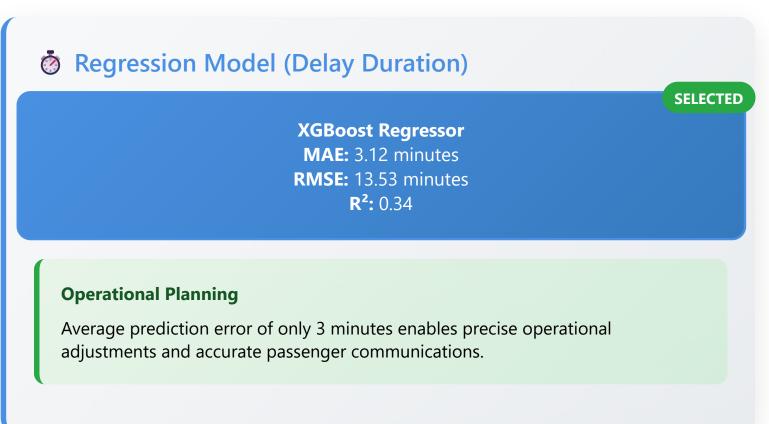


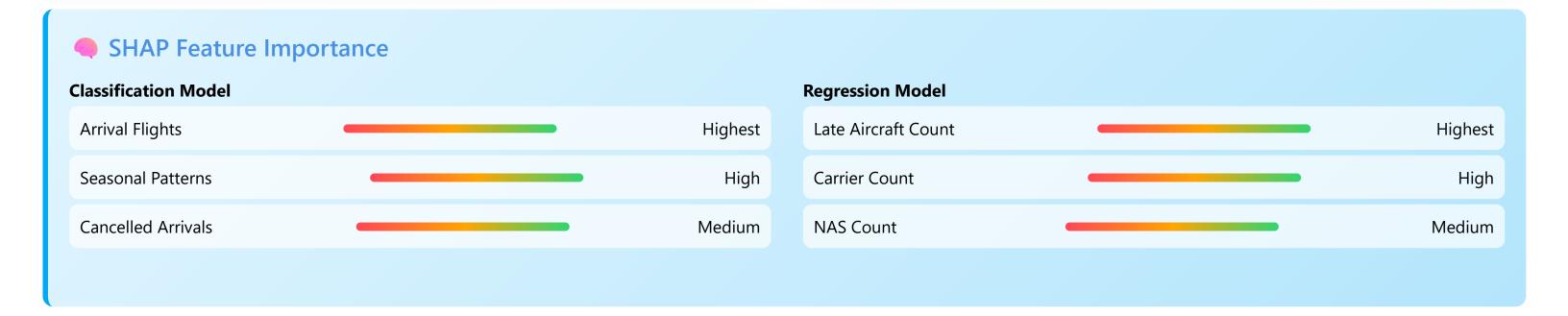
The majority of delays stem from internal operational issues that airlines can directly control, representing the highest ROI opportunity for improvement initiatives.

## •

## **Model Performance**









## **Controllable Factors - Direct Interventions**



### Optimize Aircraft Turnaround Efficiency

Impact: Late aircraft delays account for 38.9% of total delay minutes

- Implement real-time ground asset tracking
- Streamline baggage handling and refueling processes
- Build operational buffers for historically problematic routes



### **Address Internal Carrier Operations**

Impact: Carrier delays represent 34.3% of total delay minutes

- Enhanced crew management and rostering algorithms
- Shift to predictive maintenance using sensor data
- Detailed root cause analysis system for carrier incidents

### **Refine Disruption Management**

Impact: Cancelled/diverted arrivals significantly increase delay probability

- Comprehensive scenario-based contingency plans
- Automated passenger re-accommodation systems
- Multi-channel transparent communication strategies



### **©** ROI Focus

These controllable factors represent 73.2% of total delays - the highest impact area for operational investments and process improvements.





### **Airport Congestion Management**

Key Finding: Arrival flights volume is the most dominant feature in delay prediction



- Dynamic scheduling to avoid peak congestion windows
- Enhanced ground resource allocation during high-volume periods
- Advocacy for airport infrastructure improvements



### Weather & Seasonal Preparedness

Key Finding: Summer months show consistent peaks across all delay types



### **Adaptive Strategies**

- Advanced meteorological integration for early decision-making
- Seasonal operational readiness protocols
- Flexible routing and diversion strategies

### Mational Air System (NAS) Adaptation

**Impact:** NAS delays contribute 21.2% of total delay minutes



### **Collaborative Approach**

- Maximize internal efficiency to reduce system burden
- Real-time ATC communication channels
- Support for air traffic management system modernization



### Strategic Insight

While external factors are beyond direct control, proactive adaptation and collaboration can significantly minimize their disruptive impact on operations.



## **Predictive Model Implementation**



## **Early Warning Dashboard**

### **Real-time Implementation:**

- Classification: 95% confidence alerts for high-risk flights
- **Regression:** Precise delay duration estimates (±3 min accuracy)
- **SHAP Integration:** Explainable predictions for targeted interventions

### **Operational Benefits**

Proactive passenger communication, dynamic resource allocation, and optimized crew scheduling

### Continuous Improvement Pipeline

### **Model Evolution:**

- CI/CD pipeline for regular model retraining
- Integration of additional real-time data sources
- Performance monitoring and drift detection

### **Future Enhancements**

Incorporate wind speeds, runway closures, staffing levels, and aircraft tail-specific data



### **6** Implementation Roadmap

Phase 1: Deploy

Early warning system with current models

**Phase 2: Enhance** 

Integrate real-time data streams and SHAP explanations

**Phase 3: Scale** 

Industry-wide collaboration and advanced analytics



### **6** Expected Impact

Based on controllable delay analysis (73.2%), airlines implementing these strategies could achieve 20-30% reduction in delay-related costs while significantly improving passenger satisfaction.

# Thank You

Questions & Discussion

### **Key Takeaways**

- ✓ 73.2% of delays are controllable by airlines
- Predictive models achieve 96% accuracy with 3-minute precision
  - SHAP analysis enables targeted, explainable interventions
  - ✓ Data-driven approach can reduce delay costs by 20-30%