

Project Overview & Objectives

- Title: Optimizing Air Travel: Data-Driven Flight Delay Analysis & Prediction1.
- Background: Flight delays disrupt global connectivity, causing inconvenience for passengers and significant operational costs for airlines1.
- Problem Statement: Delays stem from both controllable (carrier, late aircraft) and uncontrollable (weather, NAS) factors, leading to inefficiencies and poor customer experiences1.
- Objectives:
 - Identify hidden delay patterns and influential factors through rigorous EDA1.
 - Build robust models to predict delay occurrence and duration, including explainability and operational focus1.
 - Deliver actionable, data-driven recommendations to minimize delays and improve airline operations1.

Data & Methodology

- Dataset: Historical flight records with fields for year, month, carrier, airport, number of flights, delay causes, cancellations, and diversions1.
- Key Features:
 - Delay metrics: arr_delay, carrier_delay, weather_delay, nas_delay, security_delay, late_aircraft_delay1.
 - Engineered: Delay indicator, total delay causes, season, Operational Adjustability Index (OAI)1.
- Methodology:
 - Data cleaning and imputation for missing values1.
 - Feature engineering to enhance predictive insight and operational relevance1.
 - EDA to reveal trends, bottlenecks, and correlations1.
 - Predictive modeling (classification for delay, regression for duration), explainability with SHAP, and OAI to prioritize controllable delays

EDA - Analytical Insights

- Delay Distribution: Most flights are on time, but a significant minority experience long delays, resulting in a right-skewed distribution¹.
- Temporal Patterns: Delays peak in winter and late evenings; spring and early mornings see fewer delays¹.
- Airport & Carrier Impact:
 - Top 10 airports and carriers identified for highest average delays, revealing persistent hotspots¹.
- Delay Causes:
 - Late aircraft and carrier delays are the largest contributors, followed by NAS and weather; weather is a major cause of cancellations¹.
- Analytical Visuals: Histogram of arrival delays, bar charts for monthly/seasonal delays, pie chart for delay causes, bar chart for top airports/carriers

Root Cause Analysis & Operational Bottlenecks

- Root Cause Breakdown:
- Late aircraft and carrier delays are the most controllable and impactful; weather and NAS are significant but less controllable¹.
- Operational Bottlenecks:
- Delay propagation is evident: delays at major hubs often cascade to subsequent flights¹.
- Cancellation and diversion rates highlight inefficiencies at specific airports¹.
- Correlation Analysis: Strong positive correlation between late aircraft delays and overall arrival delays, confirming systemic operational issues¹.
- Analytical Visuals: Correlation heatmap, bar chart of cancellations/diversions, summary of bottleneck airports
-

Predictive Modeling & Explainability

- Classification Model: Predicts if a flight will be delayed (Yes/No) using Random Forest; achieves high accuracy and ROC AUC, supporting robust operational forecasting¹.
- Regression Model: Predicts expected delay duration in minutes; evaluated using MAE and RMSE, providing actionable time estimates¹.
- Operational Adjustability Index (OAI): Custom metric prioritizing controllable delays, focusing the model on actionable outcomes for airlines¹.
- Explainable ML (SHAP): SHAP values highlight the most influential features for each prediction, supporting transparency and trust¹. OAI-weighted SHAP analysis distinguishes controllable from uncontrollable delay drivers, empowering targeted interventions¹.
- Analytical Visuals: ROC curve, confusion matrix, SHAP summary plot, OAI distribution

Actionable Recommendations & Impact

- For Airlines:
 - Prioritize operational improvements for late aircraft and carrier-related delays1.
 - Optimize schedules to avoid peak delay periods and high-risk airports1.
 - Enhance weather preparedness and contingency planning1.
 - Use predictive models for proactive passenger communication and resource allocation1.
- For Airports:
 - Collaborate with airlines to streamline ground operations and reduce bottlenecks1.
- Business Impact: Data-driven strategies can reduce delays, lower costs, and improve passenger satisfaction1.
- Next Steps: Continuous monitoring, periodic model retraining, and integration of real-time data for ongoing improvement

THANK YOU!

FOR YOUR ATTENTION