Society of Business

STUDENTS' TECHNICAL COUNCIL

Open Projects 2025

Analytics

Optimizing Air Travel: A Data-Driven Approach to Flight Delay Analysis and Prediction

Problem Statement

In today's interconnected world, air travel has evolved from a luxury to an indispensable mode of transportation, underpinning global commerce, tourism, and personal connections. However, the pervasive nature of flight delays frequently undermines the efficiency and convenience of this essential service. These delays not only cause significant inconvenience and stress for passengers but also incur substantial operational costs for airlines, ranging from fuel expenses and crew repositioning to missed connections and reputational damage. Understanding the underlying causes of these disruptions and proactively anticipating them is paramount for enhancing operational efficiency, improving customer satisfaction, and fostering a more reliable air travel ecosystem.

This 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:

The primary objectives of this project are to:

- **Uncover Hidden Patterns:** Conduct an in-depth exploratory data analysis (EDA) to identify recurring trends, influential factors, and significant correlations contributing to flight delays.
- **Develop Predictive Capability**: Build a robust analytical model capable of predicting the likelihood or duration of flight delays, providing an early warning system for stakeholders.
- **Generate Actionable Insights**: Formulate data-backed recommendations and strategic guidance for airlines and relevant stakeholders to mitigate delay occurrences and enhance operational resilience.

Deliverables:

This project will deliver a comprehensive set of artifacts designed to provide both analytical depth and practical application:

3.1. Detailed Exploratory Data Analysis (EDA) Report:

A comprehensive document detailing the initial findings from the dataset.
Includes visualizations showcasing delay distributions, common delay causes, temporal patterns (e.g, time of day, day of week, seasonal variations), and the influence of various operational and environmental factors.

Highlights key insights and potential areas for intervention.

Root Cause Analysis:

Detailed breakdown of delay reasons and their impact.

Identification of specific operational bottlenecks or common issues leading to delays.

3.2. Predictive Model:

While air travel is now integral to modern life, it often brings with it the unwelcome challenge of flight delays. Using the Flight Delay dataset build a capable model which predicts-

- Whether a flight is likely to be delayed (Yes/No).
- Estimates the expected delay duration (in minutes).

In addition to conventional performance tracking, the model should incorporate two advanced features:

Operational Adjustability Index (OAI):

A custom evaluation metric designed to prioritize controllable delays such as carrier delay or late aircraft by assigning higher weights to these causes. The model should be trained not just to minimize the total predicted delay, but specifically to minimize the OAI. This ensures the model focuses on delays where operational intervention is feasible, enabling airlines to reduce inefficiencies and costs more effectively.

Explainable ML using SHAP:

To enhance transparency, SHAP (SHapley Additive exPlanations) is employed to interpret the model's predictions. SHAP values indicate the contribution of each feature to a specific delay prediction. By applying OAI-based weights to these values, the project will identify whether delay causes are controllable or external—empowering airlines to take focused and meaningful action.

Model Performance Report:

- Metrics like accuracy, precision, recall, F1-score (for classification models).
- Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) (for regression models).
- Confusion matrix , ROC , AUC curve.

3.3. Actionable Recommendations & Consulting Insights:

Actionable Recommendations for Delay Mitigation based on the analysis, focusing on concrete suggestions for airlines, including schedule adjustments, improved ground operations, proactive communication, and better resource allocation.

Submission Guidelines:

Code File:
The complete project codebase, including all scripts, notebooks, and any necessary data files, should be provided in a well-organized and clearly commented format in .ipynb

Presentation Deck: A comprehensive and visually appealing presentation deck should be submitted with a maximum length of six content pages, excluding the appendix, introduction, and thank-you slides, summarizing the project's objectives, methodology, key findings from EDA, model performance, and actionable recommendations. This deck will serve as the primary communication tool for presenting the project's outcomes.

Timeline:
The project timeline is scheduled from June 6th to June 14th, 2025, with all deliverables to be completed by end of day on June 14th.

Dataset: Link