Optimizing Air Travel:

Flight Delay Analysis and Prediction

Zawar Ishita Aashish
(23118082)

Introduction:

- **Background:**

Flight delays are a persistent challenge in the aviation industry, causing significant financial losses, passenger dissatisfaction, and operational inefficiencies.

- **Project Motivation:**

With increasing passenger volumes and complex scheduling networks, airlines need intelligent systems to proactively manage delays.

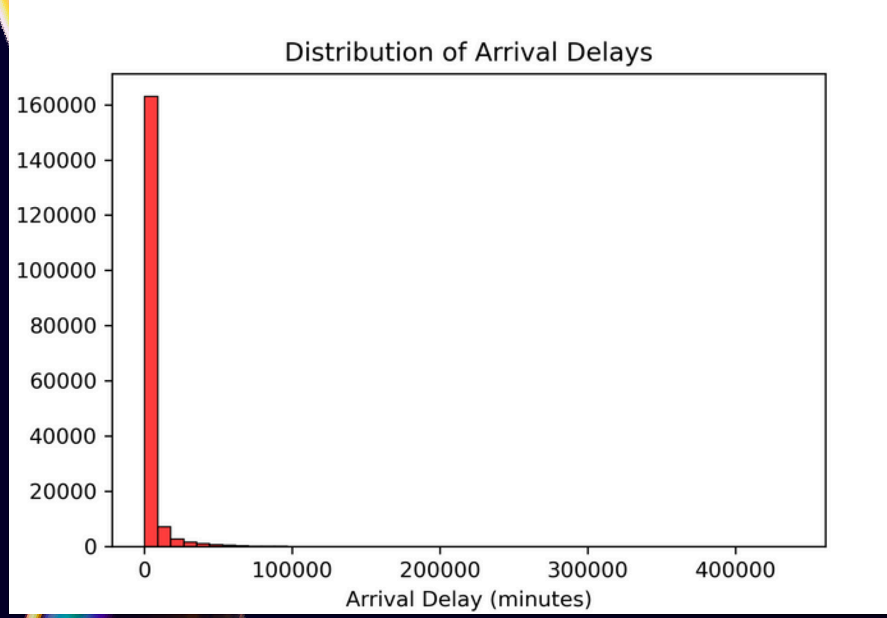
- **Approach:**

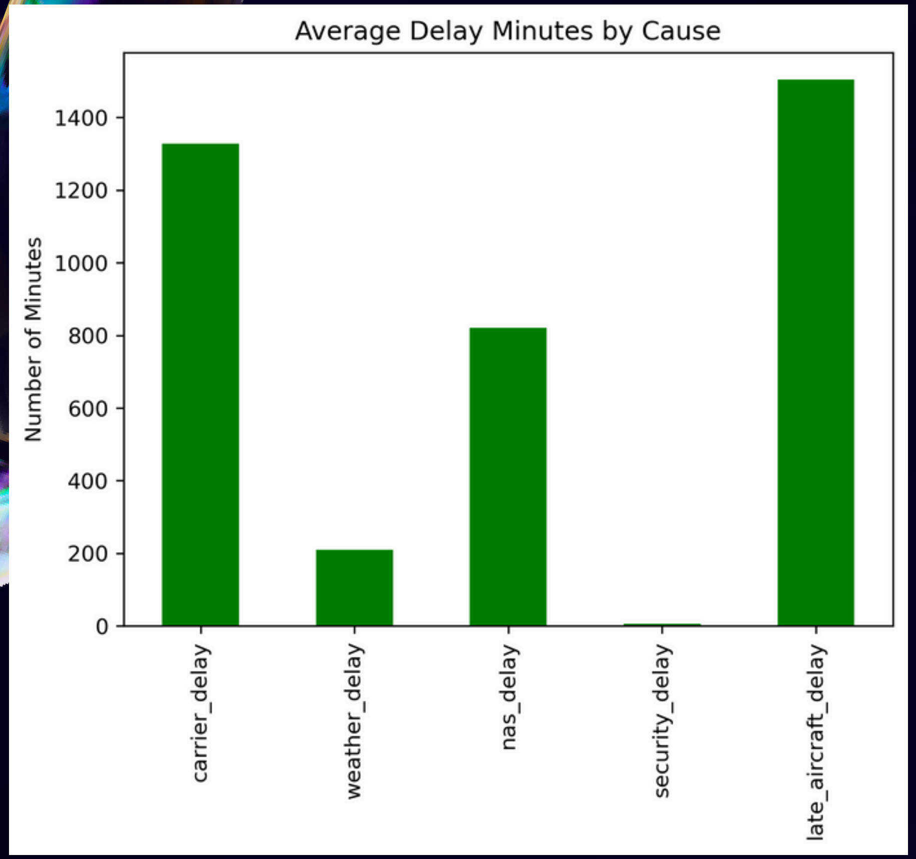
This project applies machine learning and design ML models to predict flight delays and explain their causes, enabling data-driven decision-making for improved airline operations.

- **Core Deliverables:**

Delay prediction models (classification and regression)
Operational insights through OAI and SHAP
Actionable recommendations tailored to controllable delays

Key findings from EDA:





Primary Delay Drivers Identified:

- late_aircraft_delay and carrier_delay are the top contributors to arrival delays.

Delay by Months:

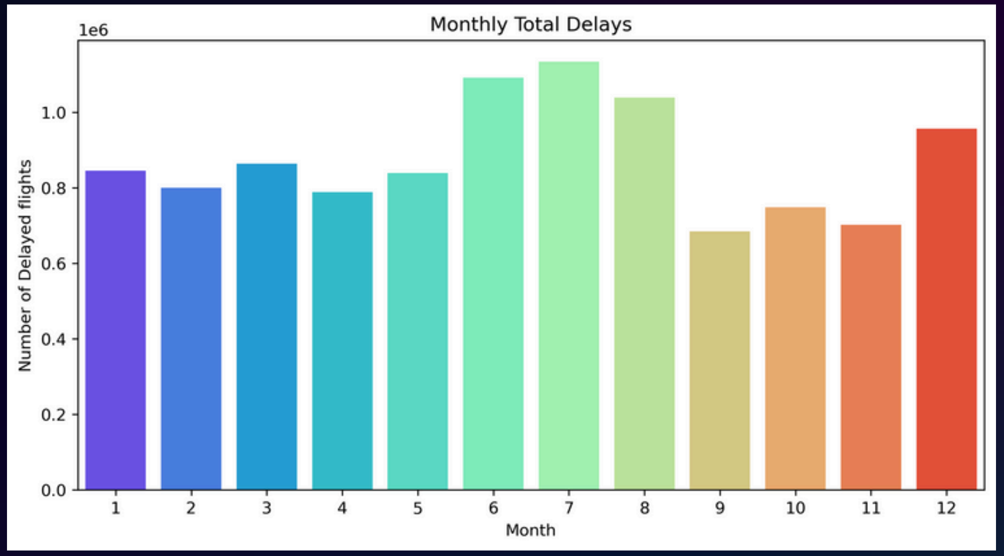
- More delays obeserved in the months of July,August ,June and December.

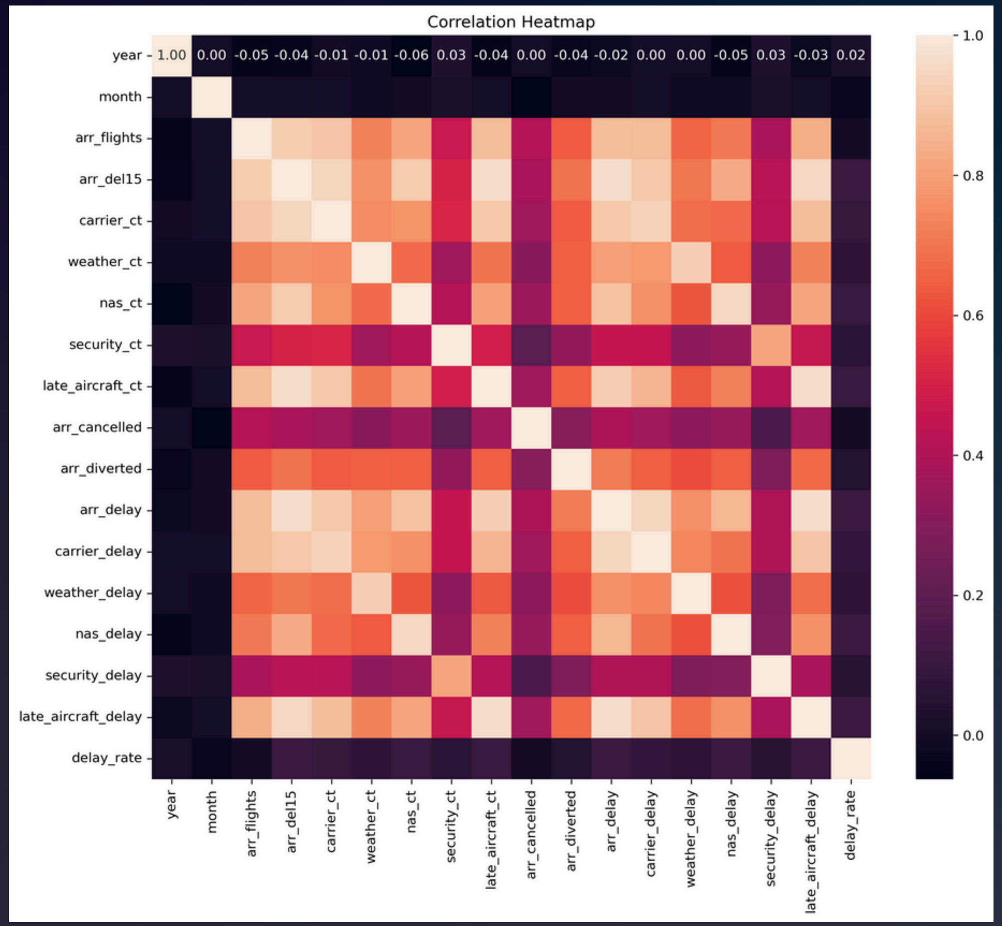
Delay by Airport:

- Certain hub airports(like ORD, ATL, DFW, etc) exhibit consistently higher average delays, likely due to traffic congestion or operational bottlenecks.

Data Quality and Distribution:

- Handling of missing values was done by imputing them by corresponding mean values. Very less percent of total values were missing.





Project Objectives:

Uncover Hidden Patterns:

01.

Conduct a comprehensive Exploratory Data Analysis (EDA) to identify:

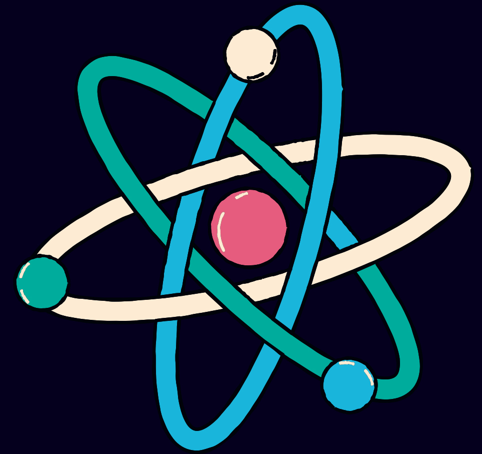
- Recurring delay trends
- Influential factors behind delays
- Significant correlations across time, airport, and carrier dimensions

Develop Predictive Capability:

02.

Build and evaluate machine learning models that can:

- Classify whether a flight will be delayed (Yes/No) -Classification Model
- Predict the expected duration of the delay (in minutes)-Regression Model



Deliver Strategic Recommendations:

03.

Translate model outputs into strategic airline guidance:

- Use OAI and SHAP to separate controllable vs. external delay causes
- Recommend improvements for scheduling, ground handling, and communication

Methodology:

01. 🗂️ Data Collection & Understanding

- Dataset: Airline Delay Cause Dataset
- Features : carrier, airport, year, month, delay types (late aircraft, security, carrier, weather, NAS, etc.)

02. 🛠️ Data Preparation

- Treated missing values and imputed them with mean.
- Created new features: delay likely, OAI score
- Label encoded categorical variables (e.g., airport, carrier,month)

03. 📊 Prediction Model Development

Classification Model:

- Objective: Predict if a flight is likely to be delayed (Yes/No)
- Algorithm: Random Forest Classifier

Regression Model:

- Objective: Estimate delay duration (in minutes)
- Algorithm: Results showed best for Random Forest Regressor when compared to XGBoost and LightGBM models.

04. 🔍 Model Interpretation & Focus Areas

- **OAI (Operational Adjustability Index):** To weight predictions based on controllability of delay causes
- **SHAP:** To explain how each feature affects individual predictions

🧰 Tools & Technologies

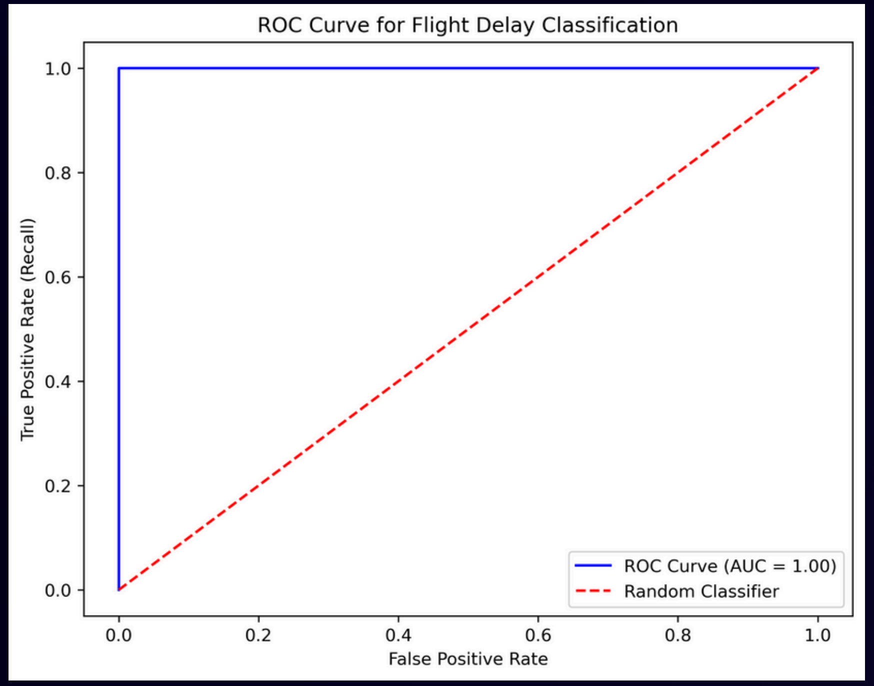
- Python, Pandas, NumPy, Scikit-learn, SHAP, Matplotlib, Seaborn, Jupyter Notebook, Random Forest, XGBoost, LightGM

Model Performance:

Classification Model:

Predict Delay Likelihood: (Yes/No)

- **Accuracy:** 100%
- **F1-Score:** 1.0
- **AUC Score:** 1.0
- **Confusion Matrix:** No false positives or false negatives
- Model demonstrates exceptional ability to distinguish delayed from on-time flights. Ideal for real-time operational alerts and flight risk assessment

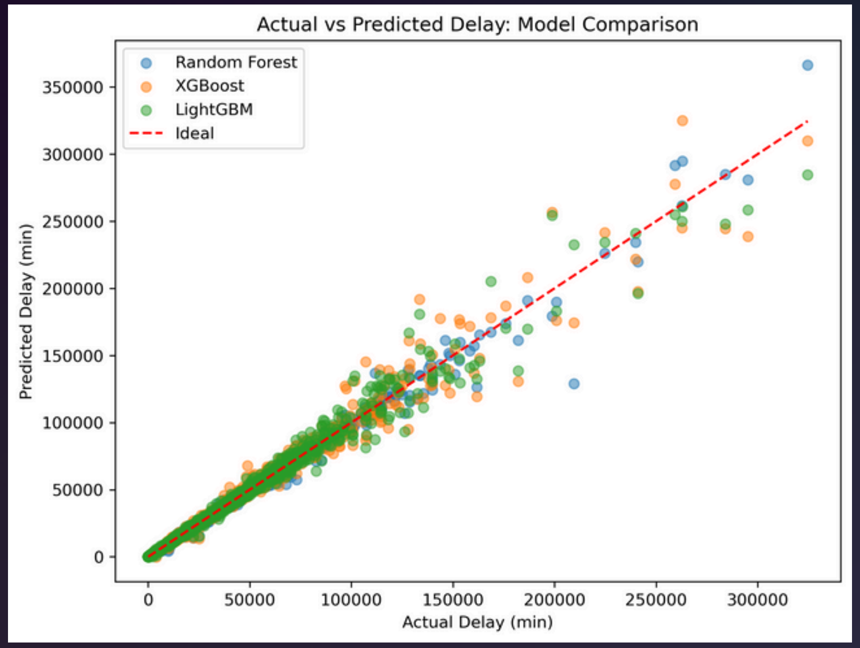


CONFUSION MATRIX	Predicted No	Predicted Yes
Actual No	8526	0
Actual Yes	0	27342

Regression Model:

Delay Duration Prediction (in mins)

Model	MAE	RMSE
Random Forest	83.46	762.34
LightGBM	172.11	1066.54
XGBoost	183.59	1202.46



Random Forest outperformed other models in delay duration estimation, offering more accurate predictions for operational use

Actionable Recommendations:

Flight Delay Optimization

Insights:

Data shows delays spike during peak travel seasons (e.g., July, December) and at congested airports.

Recommendation:

Introduce time buffers during peak hours and reassign flights away from chronically delayed slots or airports

Expected Impact:

Enables smoother operations during peak traffic periods, reduces cumulative delays, and improves adherence to published flight schedules.

Improved Ground Operations

Insights:

Carrier and late aircraft delays were strongly associated with SHAP importance and high OAI scores — meaning these are delays within operational control

Recommendation:

Deploy real-time monitoring tools for gate turnaround and dynamically allocate staff based on risk prediction.

Expected Impact:

Improves on-time departures and enhances aircraft utilization.

Proactive Communication

Insights:

The regression model effectively predicts delay duration. Passengers often experience frustration due to delayed notifications, not just delayed flights.

Recommendation:

Use model predictions to alert passengers of delays over 15 minutes, support timely rebooking, and equip gate agents with delay risk insights to improve communication and coordination.

Expected Impact:

Boosts passenger satisfaction and loyalty while reducing complaints and missed connections.

Continued...

Dynamic Resource Allocation

Insights:

OAI scores revealed that some delays (carrier_delay, late_aircraft_delay) can be actively mitigated by better planning and resource distribution.

Recommendation:

Use SHAP and OAI insights to optimize crew and aircraft allocation, improve gate assignments for delay-prone connections, and support real-time decisions via an operations dashboard.

Expected Impact:

Reduces avoidable delays and ensures smarter deployment of limited operational assets

Predictive Maintenance & Aircraft Readiness

Insights:

Delays due to late_aircraft_delay often stem from aircraft being turned around too quickly or being taken out of rotation unexpectedly.

Recommendation:

Use delay forecasts and SHAP insights to trigger predictive maintenance, avoid risky scheduling, and add buffers for high-delay aircraft.

Expected Impact:

Enhances operational stability and reduces the risk of cascading delays.

Weather-Adapted Planning

Insights:

Though weather is uncontrollable, SHAP values can help isolate seasonal weather effects on specific routes or time windows.

Recommendation:

Combine predicted delay risk with live weather data to adjust slot priorities and add seasonal buffers on weather-sensitive routes.

Expected Impact:

Reduces unexpected delays due to forecastable conditions.

Thank you!!

Zawar Ishita Aashish(23118082)

zawar_ia@mt.iitr.ac.in