SocBiz Open Project '25

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

Leveraging Predictive Modeling, SHAP, and OAI for Airline Operational Efficiency

Presented by:

Name - Anuj Singh

Enrollment No.: 23112018

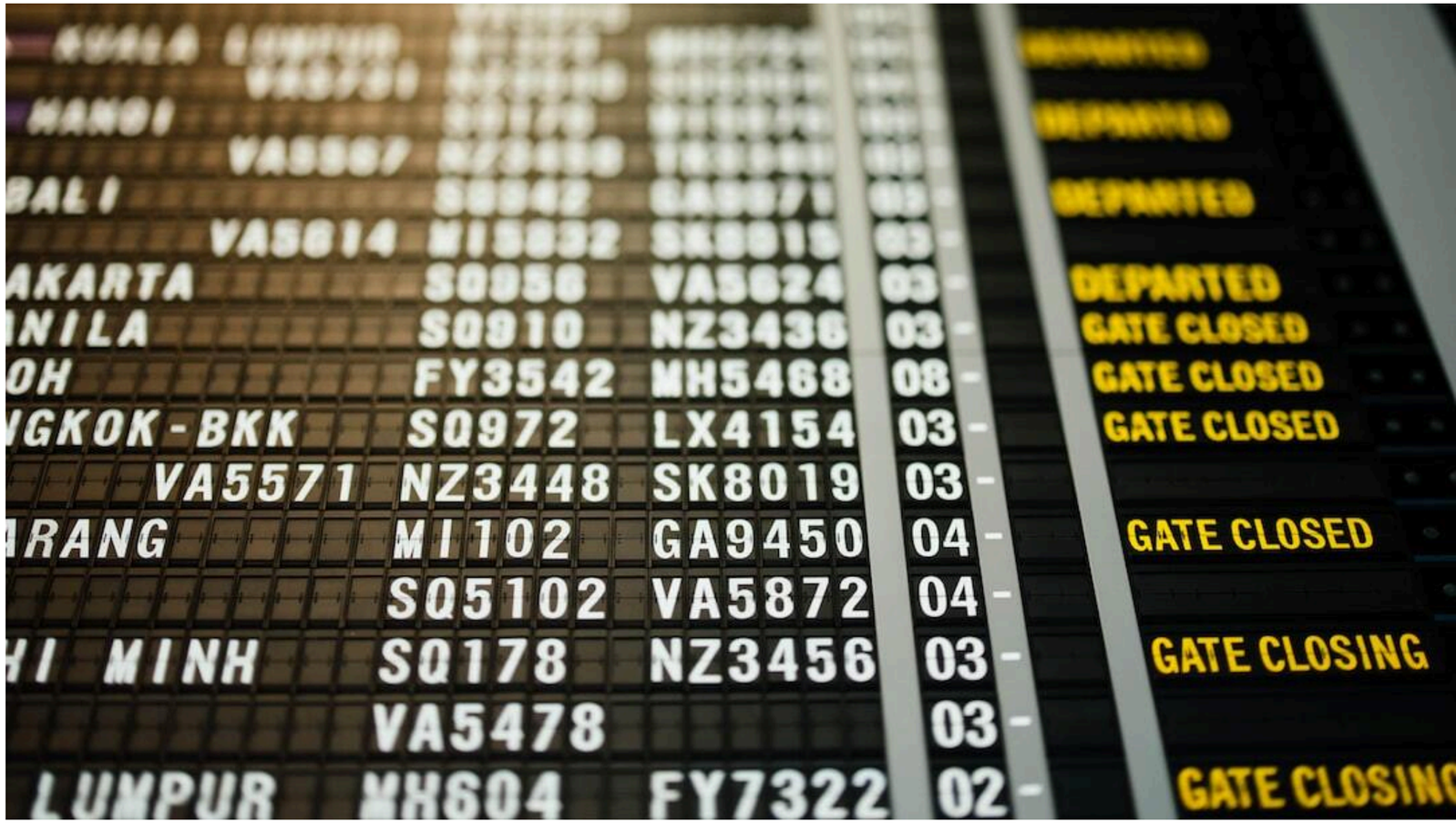
Branch: Chemical Engineering

Problem Context:

- Flight delays cost airlines billions and inconvenience passengers.
- Key causes range from controllable (e.g., carrier delay) to uncontrollable (e.g., weather).

Project Objectives:

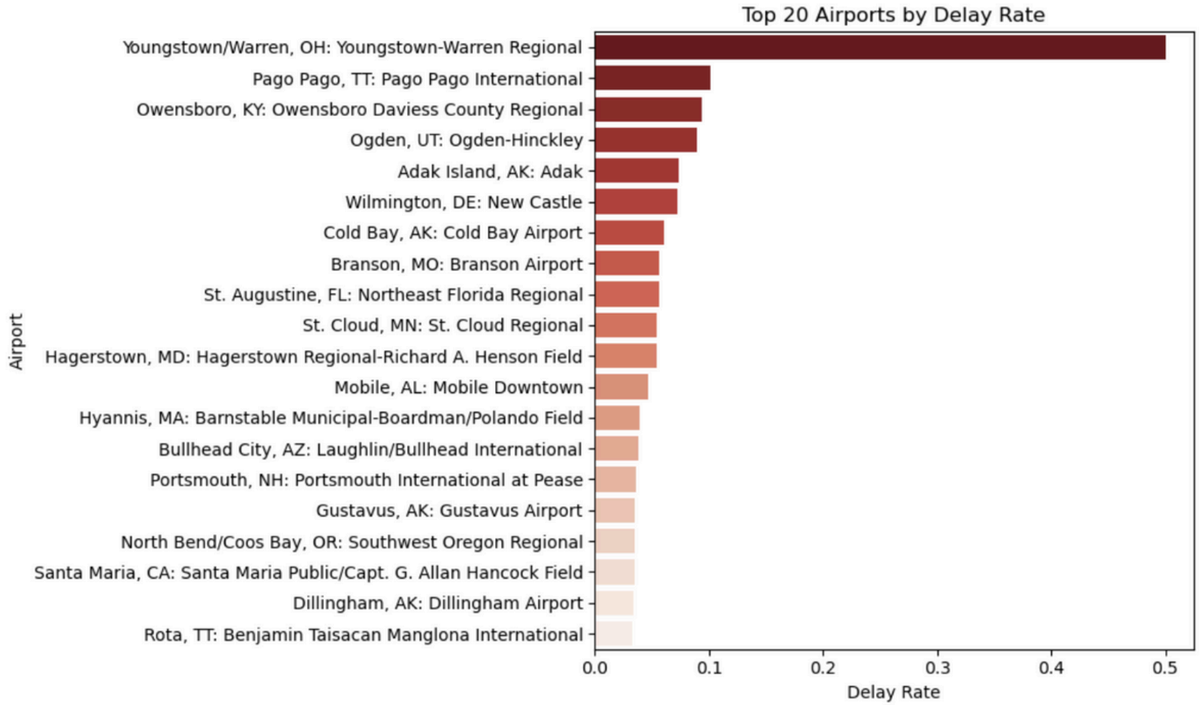
- 📊 Perform EDA to uncover delay trends.
- 🤖 Build two models:
- Classification (delay: Yes/No)
- Regression (delay duration in minutes)
- ⚙️ Minimize OAI: Focus on controllable delays.
- 🔍 Apply SHAP for explainable predictions.

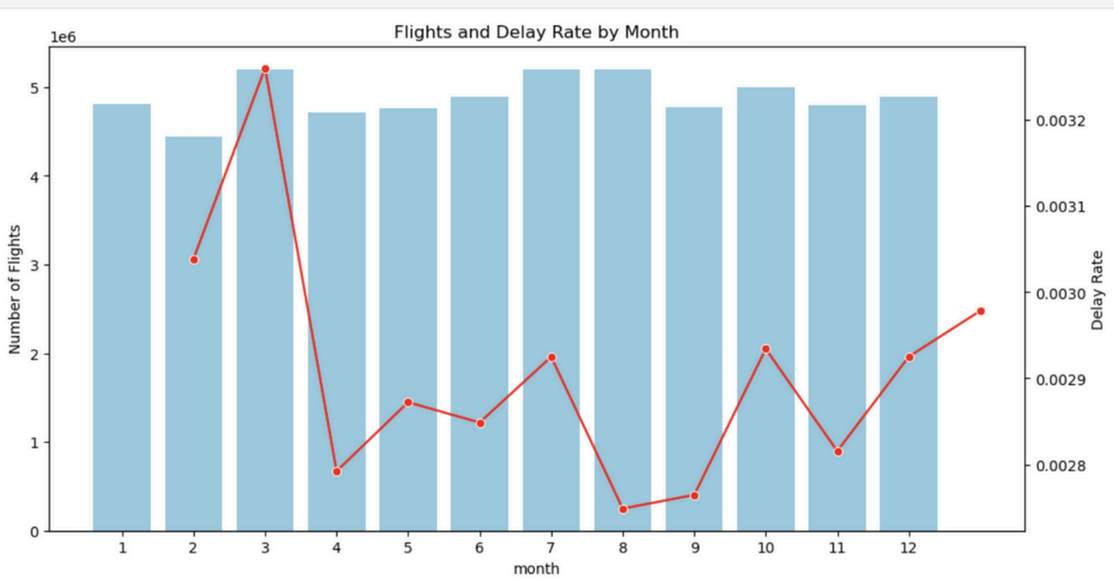


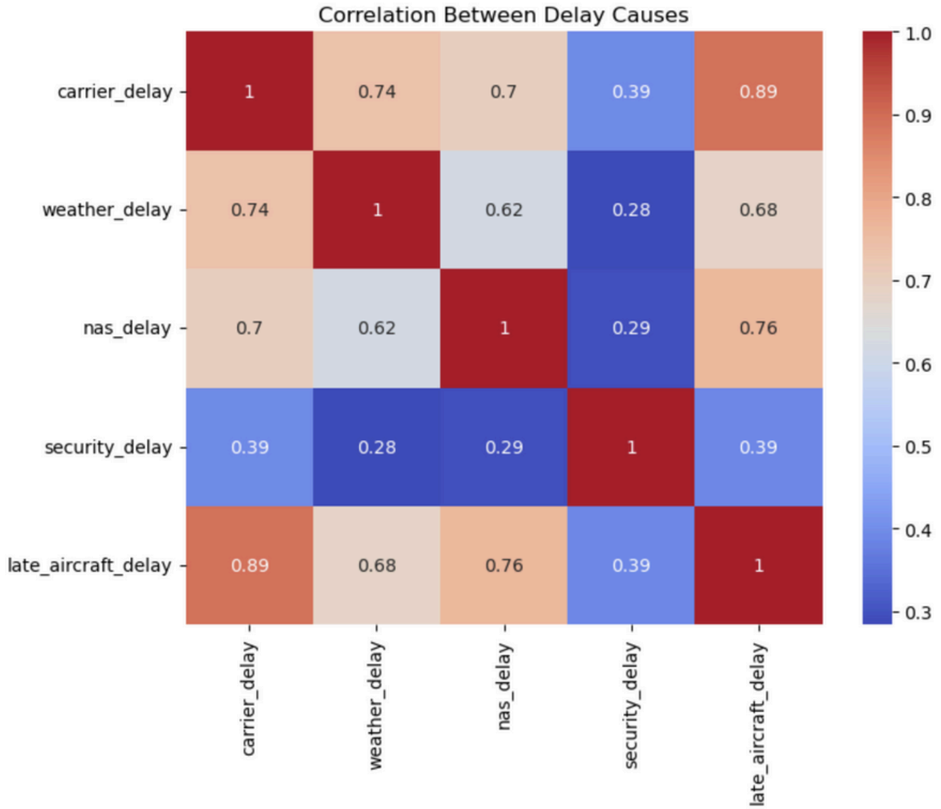
EDA & Insights

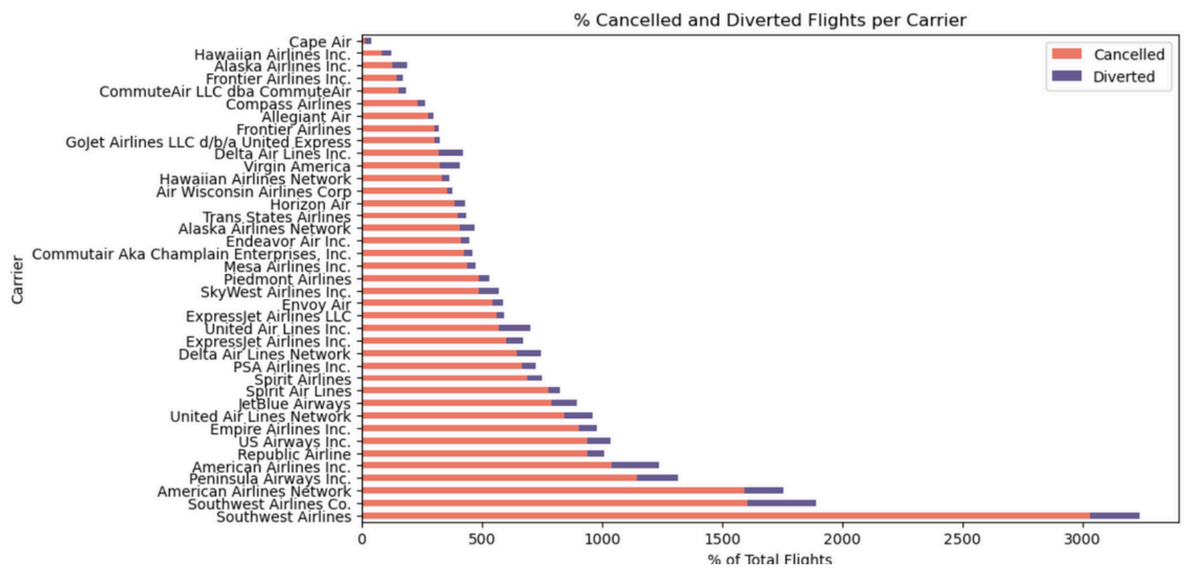
Key Insights from Data:

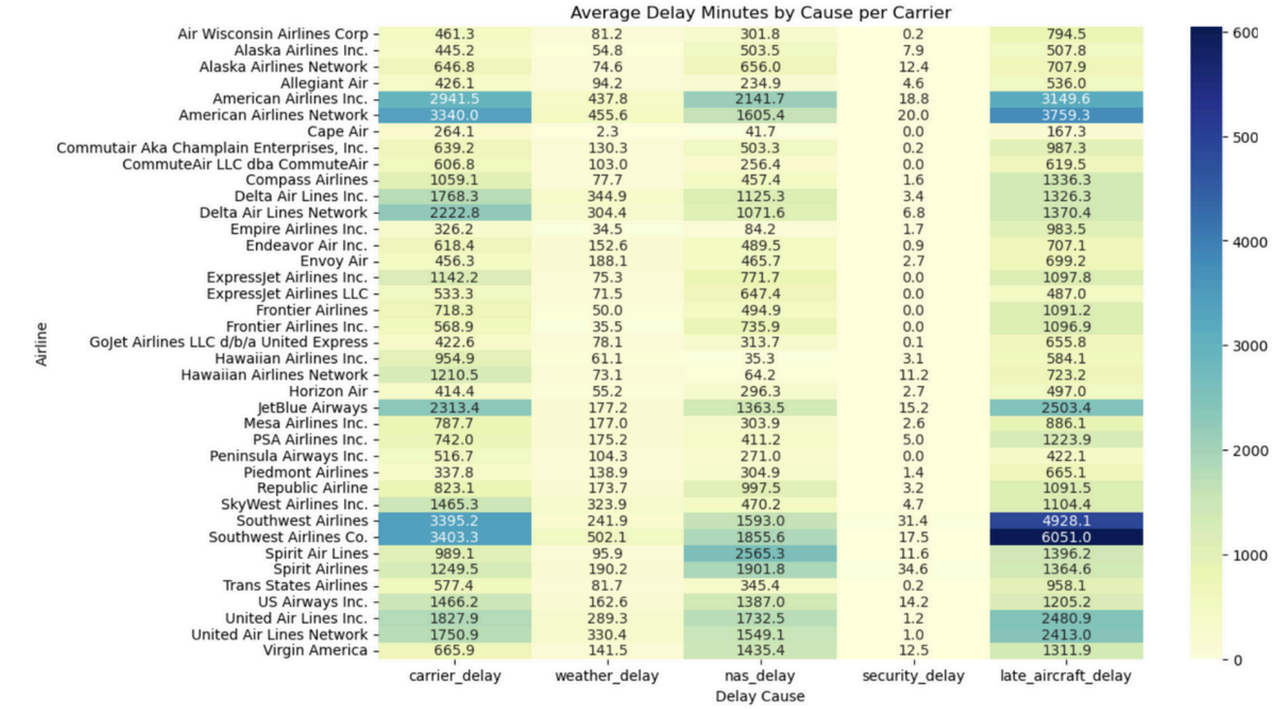
- *96% of flights were delayed.*
- *Late Aircraft & Carrier Delay in Majority*
- *Highest delays: Fridays, weekends, winter months.*
- *Top 20 airports having highest average delays.*
- *Certain carriers show consistent bottlenecks.*

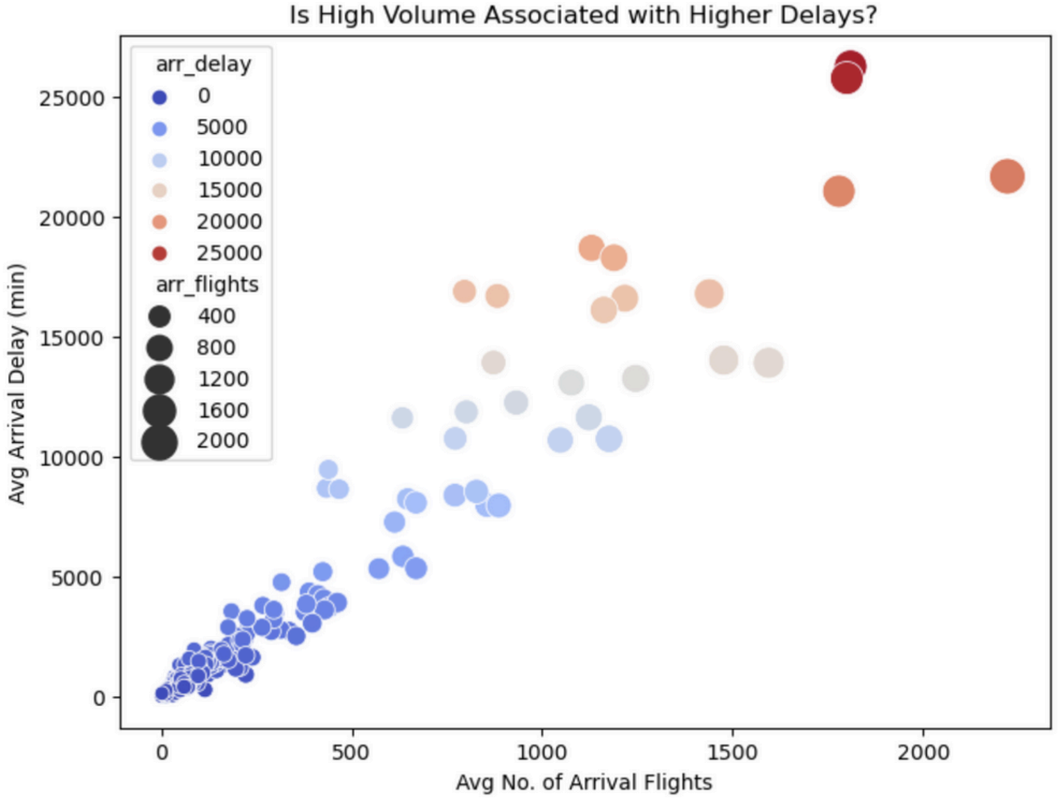


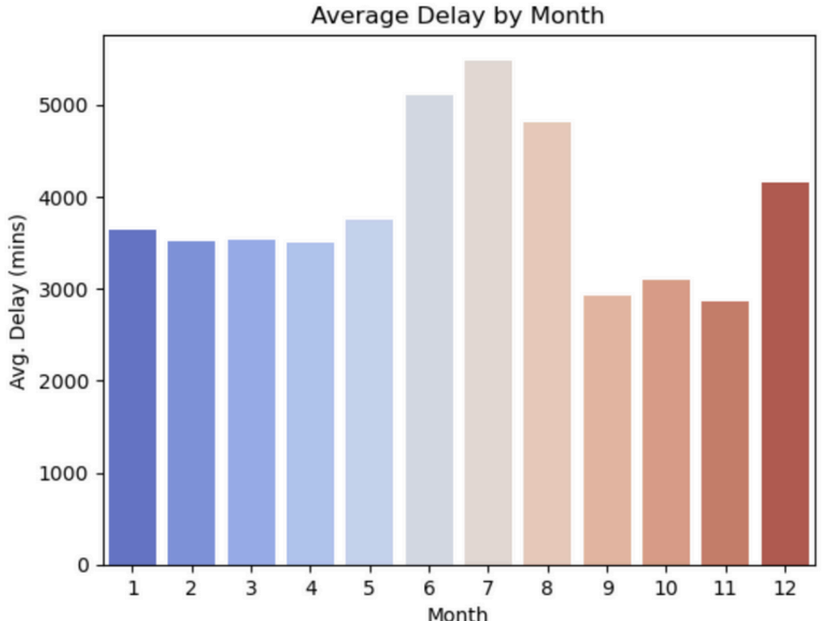





% Cancelled and Diverted Flights per Carrier
Cancelled
Diverted
Carrier
Cape Air
Hawaiian Airlines Inc.
Alaska Airlines Inc.
Frontier Airlines Inc.
CommuteAir LLC dba CommuteAir
Compass Airlines
Allegiant Air
Frontier Airlines
GoJet Airlines LLC d/b/a United Express
Delta Air Lines Inc.
Virgin America
Hawaiian Airlines Network
Air Wisconsin Airlines Corp
Horizon Air
Trans States Airlines
Alaska Airlines Network
Endeavor Air Inc.
Commutair Aka Champlain Enterprises, Inc.
Mesa Airlines Inc.
Piedmont Airlines
SkyWest Airlines Inc.
Envoy Air
ExpressJet Airlines LLC
United Air Lines Inc.
ExpressJet Airlines Inc.
Delta Air Lines Network
PSA Airlines Inc.
Spirit Airlines
Spirit Air Lines
JetBlue Airways
United Air Lines Network
Empire Airlines Inc.
US Airways Inc.
Republic Airline
American Airlines Inc.
Peninsula Airways Inc.
American Airlines Network
Southwest Airlines Co.
Southwest Airlines
0
500
1000
1500
2000
2500
3000
% of Total Flights


Average Delay Minutes by Cause per Carrier
Airline
Air Wisconsin Airlines Corp 461.3 81.2 301.8 0.2 794.5
Alaska Airlines Inc. 445.2 54.8 503.5 7.9 507.8
Alaska Airlines Network 646.8 74.6 656.0 12.4 707.9
Allegiant Air 426.1 94.2 234.9 4.6 536.0
American Airlines Inc. 2941.5 437.8 2141.7 18.8 3149.6
American Airlines Network 3340.0 455.6 1605.4 20.0 3759.3
Cape Air 264.1 2.3 41.7 0.0 167.3
Commutair Aka Champlain Enterprises, Inc. 639.2 130.3 503.3 0.2 987.3
CommuteAir LLC dba CommuteAir 606.8 103.0 256.4 0.0 619.5
Compass Airlines 1059.1 77.7 457.4 1.6 1336.3
Delta Air Lines Inc. 1768.3 344.9 1125.3 3.4 1326.3
Delta Air Lines Network 2222.8 304.4 1071.6 6.8 1370.4
Empire Airlines Inc. 326.2 34.5 84.2 1.7 983.5
Endeavor Air Inc. 618.4 152.6 489.5 0.9 707.1
Envoy Air 456.3 188.1 465.7 2.7 699.2
ExpressJet Airlines Inc. 1142.2 75.3 771.7 0.0 1097.8
ExpressJet Airlines LLC 533.3 71.5 647.4 0.0 487.0
Frontier Airlines 718.3 50.0 494.9 0.0 1091.2
Frontier Airlines Inc. 568.9 35.5 735.9 0.0 1096.9
GoJet Airlines LLC d/b/a United Express 422.6 78.1 313.7 0.1 655.8
Hawaiian Airlines Inc. 954.9 61.1 35.3 3.1 584.1
Hawaiian Airlines Network 1210.5 73.1 64.2 11.2 723.2
Horizon Air 414.4 55.2 296.3 2.7 497.0
JetBlue Airways 2313.4 177.2 1363.5 15.2 2503.4
Mesa Airlines Inc. 787.7 177.0 303.9 2.6 886.1
PSA Airlines Inc. 742.0 175.2 411.2 5.0 1223.9
Peninsula Airways Inc. 516.7 104.3 271.0 0.0 422.1
Piedmont Airlines 337.8 138.9 304.9 1.4 665.1
Republic Airline 823.1 173.7 997.5 3.2 1091.5
SkyWest Airlines Inc. 1465.3 323.9 470.2 4.7 1104.4
Southwest Airlines 3395.2 241.9 1593.0 31.4 4928.1
Southwest Airlines Co. 3403.3 502.1 1855.6 17.5 6051.0
Spirit Air Lines 989.1 95.9 2565.3 11.6 1396.2
Spirit Airlines 1249.5 190.2 1901.8 34.6 1364.6
Trans States Airlines 577.4 81.7 345.4 0.2 958.1
US Airways Inc. 1466.2 162.6 1387.0 14.2 1205.2
United Air Lines Inc. 1827.9 289.3 1732.5 1.2 2480.9
United Air Lines Network 1750.9 330.4 1549.1 1.0 2413.0
Virgin America 665.9 141.5 1435.4 12.5 1311.9
carrier_delay
weather_delay
nas_delay
security_delay
late_aircraft_delay
Delay Cause
600
500
400
300
200
100
0


Is High Volume Associated with Higher Delays?
arr_delay
0
5000
10000
15000
20000
25000
arr_flights
400
800
1200
1600
2000
Avg Arrival Delay (min)
Avg No. of Arrival Flights
0
5000
10000
15000
20000
25000
0
500
1000
1500
2000


Average Delay by Month
Avg. Delay (mins)
Month
0
1000
2000
3000
4000
5000
1
2
3
4
5
6
7
8
9
10
11
12

Classification Model

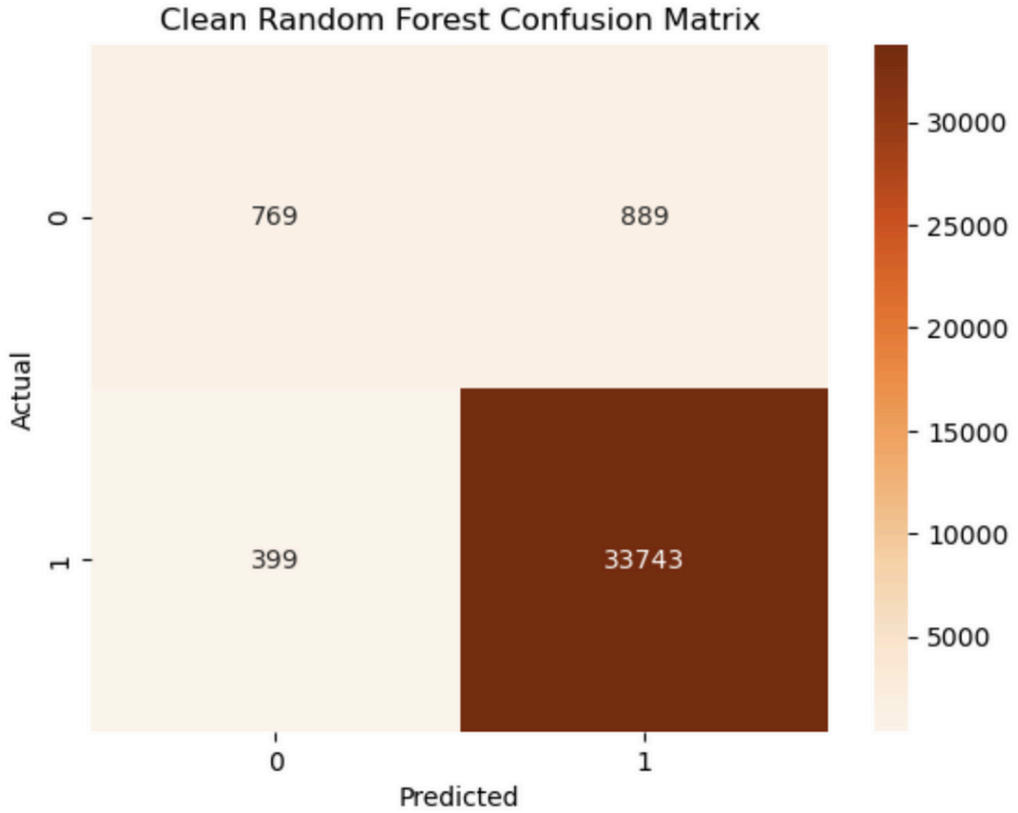
Random Forest Classifier has been used

- Flags high-risk flights reliably
- Enables early warnings for airline ops
- SHAP-supported → Explainable output

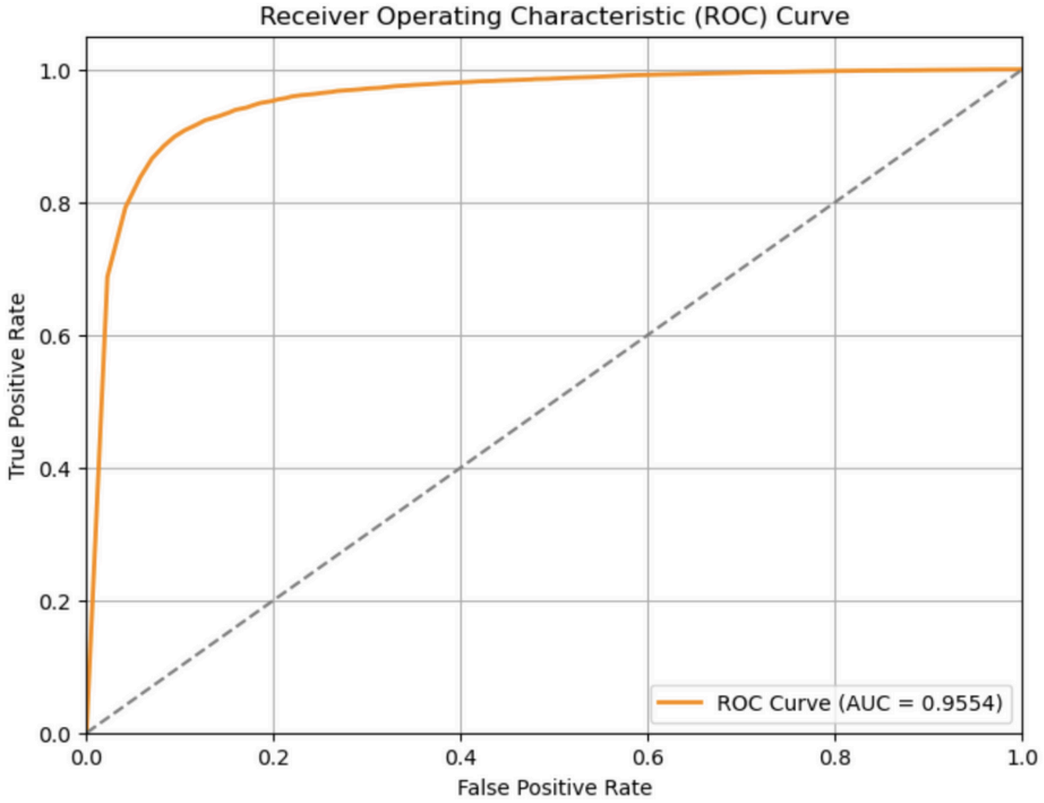
```
Clean Random Forest Classification Report:
              precision    recall  f1-score   support

           0       0.66      0.46      0.54      1658
           1       0.97      0.99      0.98     34142

    accuracy                           0.96     35800
   macro avg       0.82      0.73      0.76     35800
weighted avg       0.96      0.96      0.96     35800

ROC AUC: 0.9554433625999241
```

Precision , Recall , F1 score , Accuracy

Confusion Matrix



ROC Curve



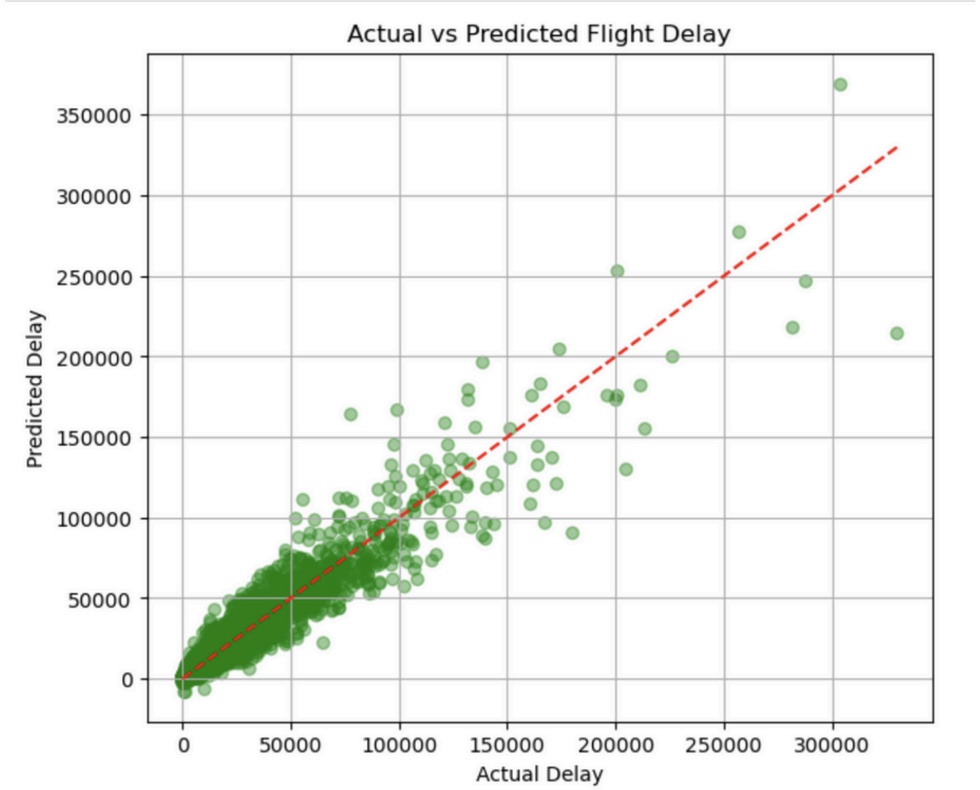
Regression Model

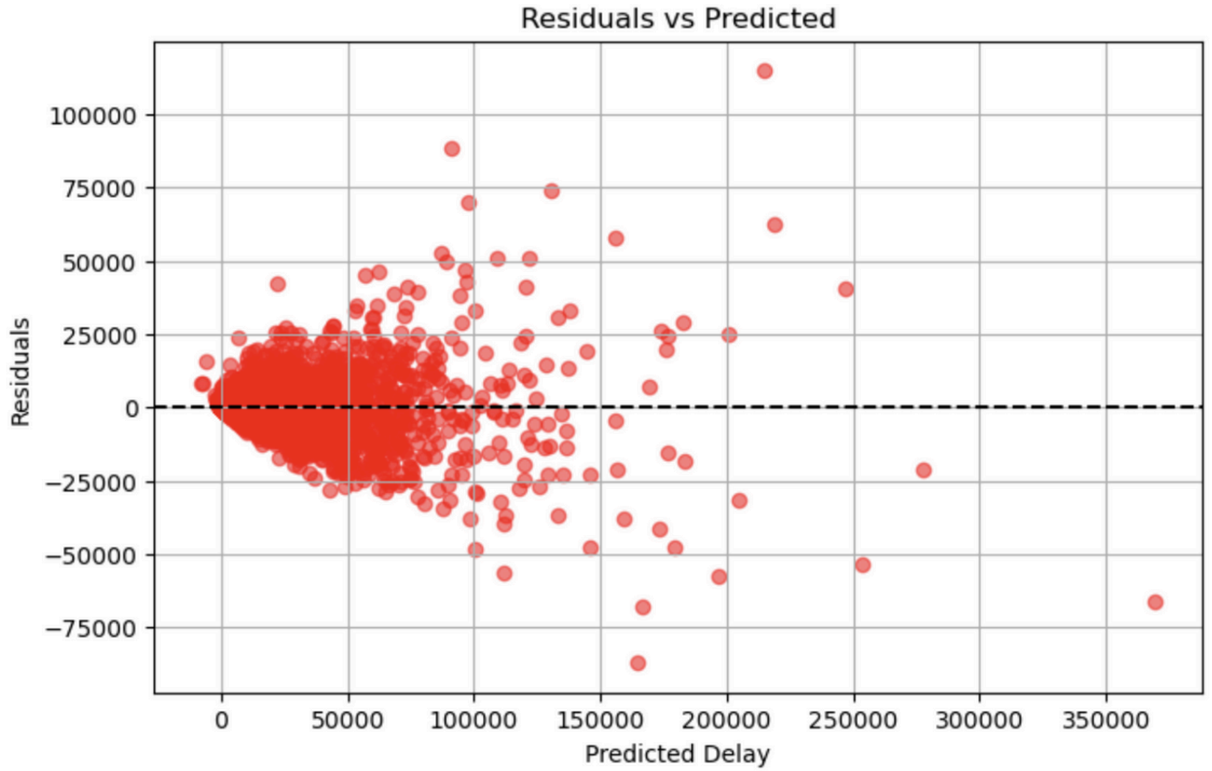
XGBoost regressor is used here

- Accurately estimates delay duration
- Helps plan crew shifts, rebooking, rerouting
- OAI-focused to reduce preventable delay minutes

	Model	MAE	RMSE	R2
3	XGBoost	1025.271	3038.037	0.930
2	LightGBM	1080.644	3170.326	0.923
1	HistGradientBoosting	1077.795	3232.163	0.920
0	Linear Regression	1873.525	4792.888	0.825

RootMeanSqError, R2 score comparison





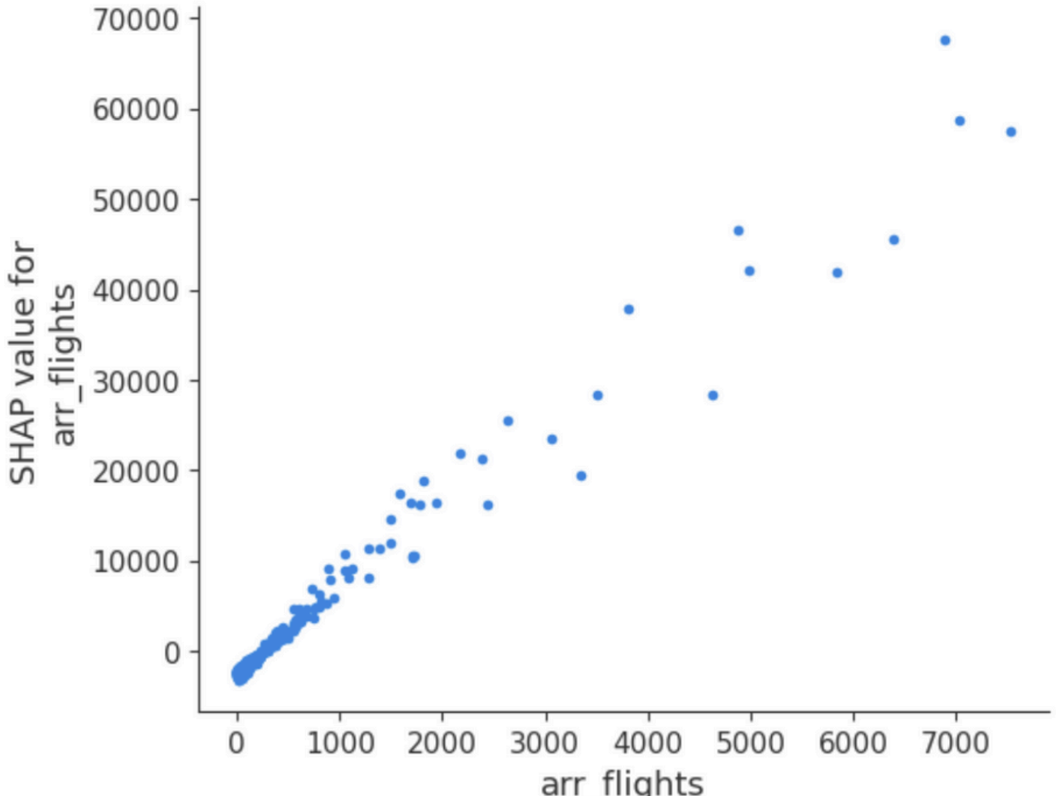
SHAP & OAI Analysis

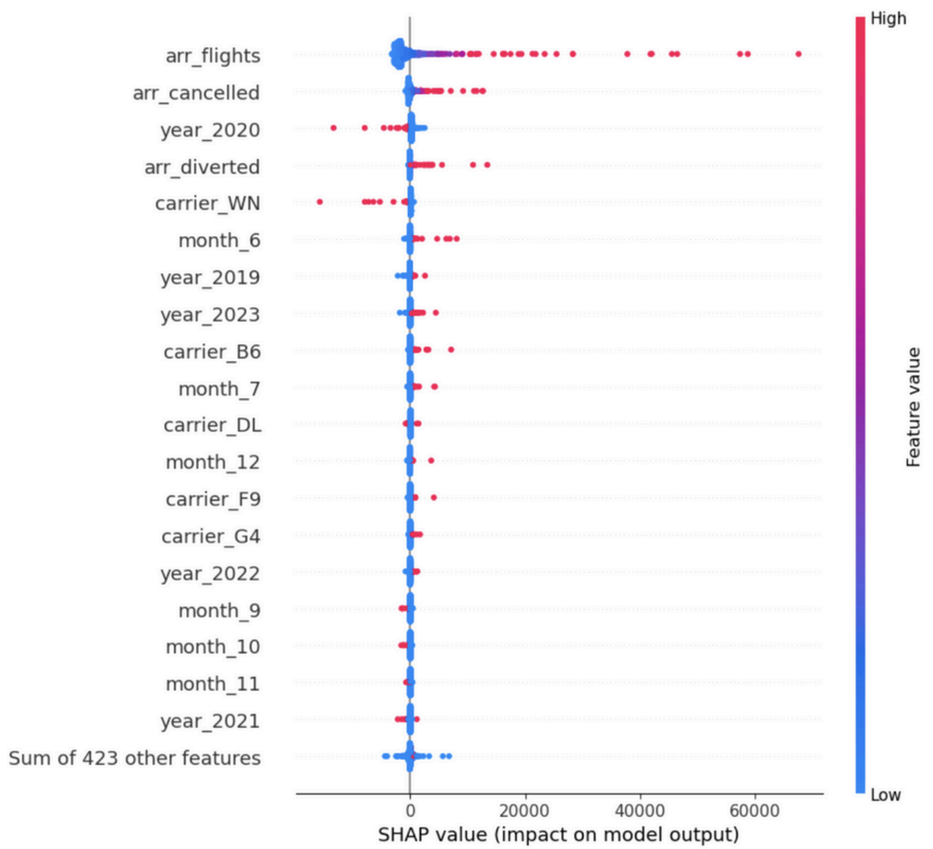
SHAP (Explainability):

- Identifies why a delay is predicted
- Ranks features by impact (e.g., month, carrier, airport)

OAI (Operational Adjustability Index):

- Weights delays by controllability
- Used as evaluation metric for business impact
- High correlation with predicted delay (0.9643)





SHAP + OAI = Strategic decision support

🔍 OAI Regression Correlation (Actual vs Pred): 0.9643

Actionable Business Recommendations

✈️ Operational Scheduling & Flight Design

- Add minimum turnaround buffer for high-delay aircraft and crew reuse cases.
- Adjust departure slots for delay-prone routes, especially during peak risk hours (3–9 PM) and Fridays/winter months.
- Use historical delay trends + ML outputs to recommend dynamic scheduling windows per route.

👨‍✈️ Ground Operations & Resource Allocation

- Prioritize Late Aircraft mitigation through:
  - Faster baggage handling
  - Real-time aircraft maintenance triggers
- Use delay probability heatmaps to assign gate priorities for tight-transfer flights.

📲 Passenger-Facing Enhancements

- Integrate model outputs into real-time check-in systems and mobile apps:
  - Early alert for predicted delays
  - Offer on-the-fly rebooking or voucher options
- Enable personalized alerts based on route and historical behavior

📊 Data-Driven Intelligence & Coordination

- Use OAI-weighted dashboards to highlight:
  - Delay trends by controllability
  - Carrier-specific performance
- Conduct weekly operations huddles to review model predictions and real outcomes.
- Share delay risk scores with:
  - Airport operations control
  - Crew planning
  - Maintenance scheduling teams

💡 Strategic Long-Term Value

- Use SHAP explanations per route/airport to create a Delay Playbook – a living document for seasonal and operational strategy.
- Apply Root Cause Trends + Model Forecasts to negotiate better slot timings with ATC and airport authorities.

“This project moves airlines from reactive delay handling to proactive delay prevention — using data, SHAP transparency, and OAI-driven prioritization.”

Thank You

+91- 9667613369

anuj_s@ch.iitr.ac.in