Predicting Flight Delays: A Data-Driven Dashboard

Navya Bingi, Andrew Guerrero, Agha Yusuf Khan, Zone Li, Ariana Richardson, Tyra Silaphet

Summary

Why are flight delays so unpredictable? We analyzed **~2 million** U.S. domestic flights (2022–2024), enriched with **hourly weather reports from the world's busiest airport (ATL)**. Our interpretable model forecasts delay risk, showing strong influence from weather and timing factors. We present the results in an interactive dashboard to help travelers and airlines assess flight delay risk in advance.

Why Predict Flight Delays?

Flight delays frustrate travelers and strain airline operations. Most tools only give real-time updates, not early warning or explanation. But many delays follow patterns based on weather, congestion, and timing. Forecasting them helps people plan smarter and improves airport operations.

Datasets

We combined two primary datasets to model and visualize flight delays across the U.S.:

- **Flight Delay & Cancellation Dataset**
  Source: U.S. Department of Transportation (via Kaggle)
  Flights: 1992511
  Airlines: 14 major U.S. carriers
  Time span: January 2022 – December 2024
- **METAR Weather Dataset**
  Source: Hourly METAR observations for ATL airport
  Matched by origin airport + departure hour
  Features: wind speed, visibility, temperature, participation

What Others Have Done

Most flight delay tools focus on live updates — not prediction. Carvalho et al. (2021) reviewed delay forecasting models and found that while machine learning is increasingly used, few projects integrate prediction with user-facing tools. Mamdouh et al. (2023) built a delay classifier but excluded weather data and interpretability. Kim & Park (2024) emphasized weather as a key delay driver, validating our use of METAR features.

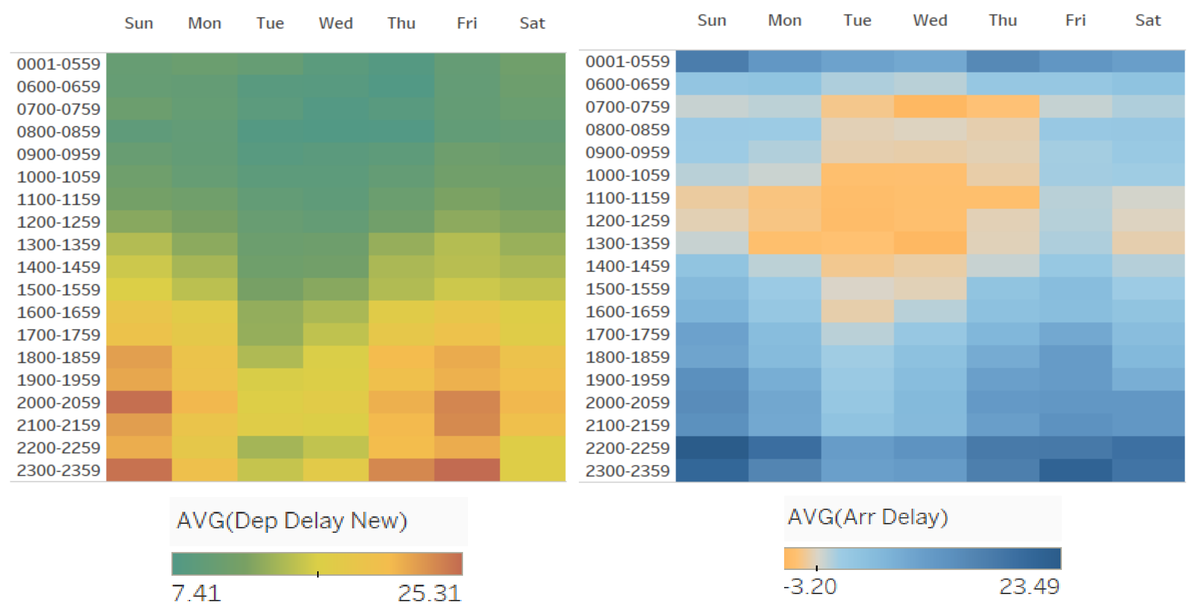
Our project is the first to combine weather-based delay prediction with an interpretable, traveler-facing dashboard focused on the ATL airport.

From Prediction to Visualization

We trained a **random forest model** to predict whether a flight from ATL would be *delayed* — defined as an arrival or departure delay greater than 15 minutes. Features include weather, timing, and flight characteristics. The model outputs feed into an interactive Tableau dashboard, allowing users to explore delay risk by route, time, and cause. Our model achieved an ROC AUC of **0.73** on held-out test data. Despite class imbalance, it detected delays with recall of **0.53** and precision of **0.41**, showing strong predictive value for real-world delay forecasting.

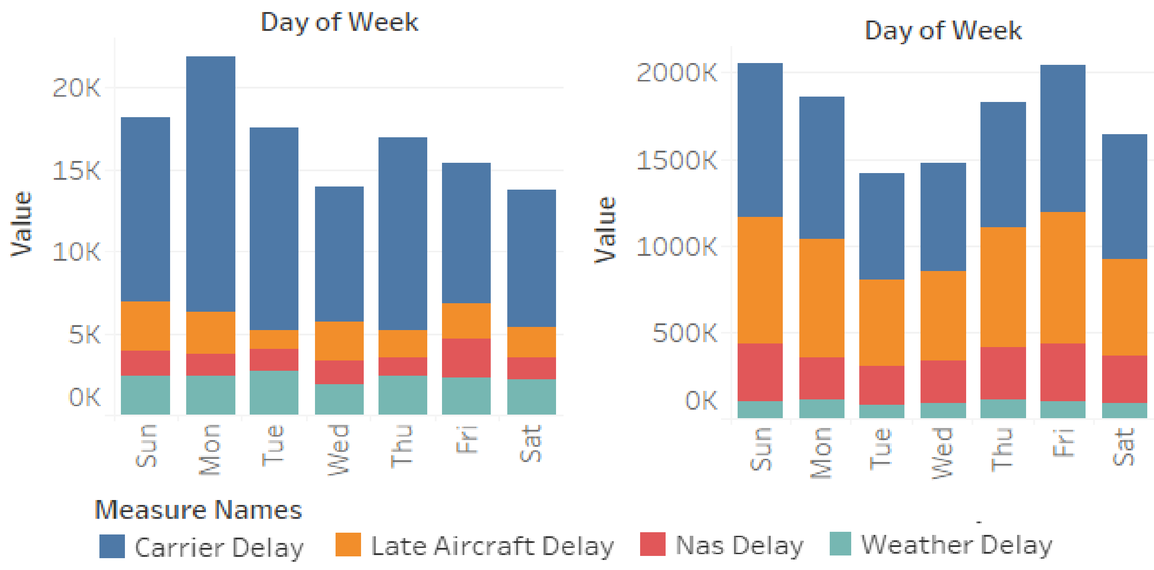
Departure and Arrival Delay Heatmaps by Hour and Day

Departure delays (left) peak during evening hours, especially on Sundays and Fridays. Arrival delays (right) are more dispersed, with some midday congestion midweek. Both show clear time-of-day patterns at ATL.



Breakdown of Delay Causes by Day

Carrier-related issues are the most common cause of both departure (left) and arrival (right) delays at ATL. Weather delays remain relatively consistent, while late aircraft and NAS delays peak on weekends and Mondays.



Evaluation & Feedback

We surveyed 28 users, including laypeople and airline professionals, using Likert-scale questions across four key areas: accessibility, aesthetics, interpretability, and travel usefulness. Overall, there was consensus on the potential usefulness of the product as there is demand for information being presented in this manner, but this perceived value is diminished due to a lack of easily digestible visualizations. Based on this feedback, we iterated on our design to improve visual interpretability and enhance accessibility.