

Understanding Flight Delays: Insights for U.S. Aviation



GROUP

BISHAL RAUNIYAR | DEVANSH TRIVEDI | MANISH THAPA

Flight Delays Represent a Persistent and Costly Challenge for the Aviation Industry

The Impact



Millions of passengers affected annually



Billions in economic costs to airlines and travelers



Cascading effects throughout the entire air transportation network



A critical operational challenge for every carrier

Our Research Objective

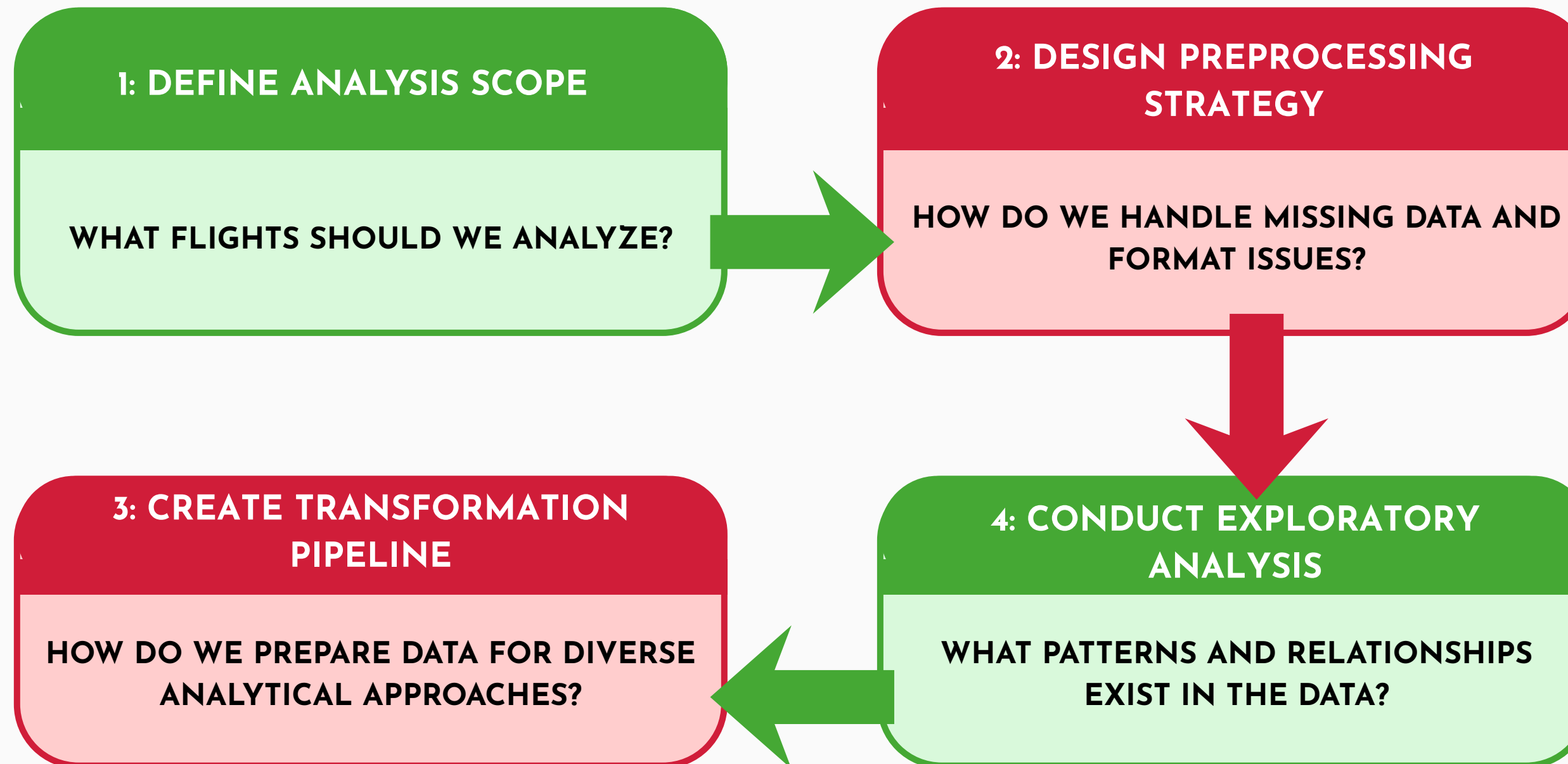
Our goal was to move beyond the simple fact that delays happen and instead identify the specific patterns and root causes. We analyzed comprehensive Department of Transportation data to answer:

- When do delays occur?
- Which airlines and airports are most affected?
- What are the primary, data-backed causes?

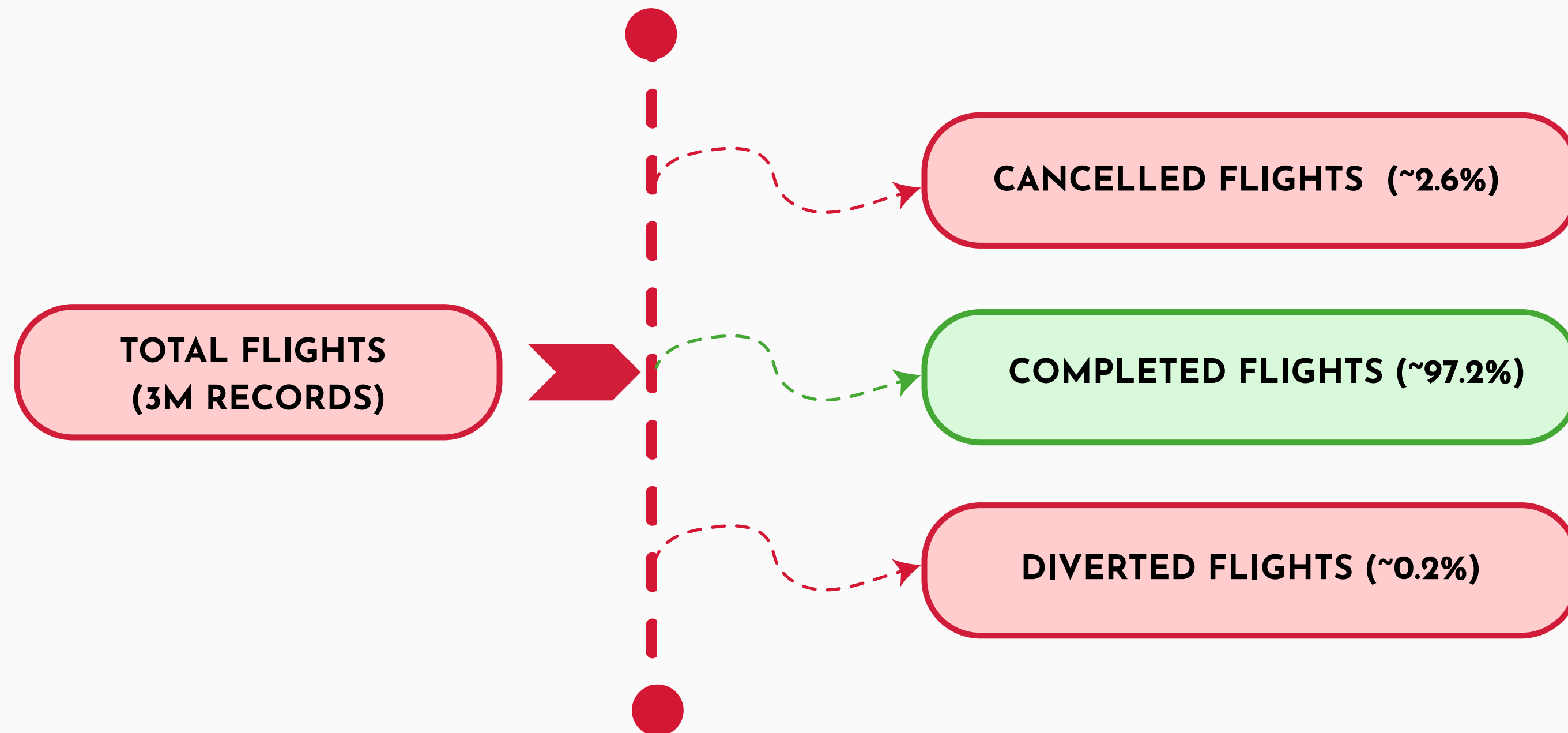
Our Analysis is Built on 3 Million Official DOT Flight Records from 2019-2023

Dataset Characteristics		Key Attributes Available			
Source:	U.S. Department of Transportation, Bureau of Transportation Statistics	Flight Identifiers	Timing Data	Delay Metrics	Operational Status
Time Period:	January 2019 - August 2023	Date, airline, flight number, origin/destination	Scheduled vs. actual departure/arrival times, taxi times	Total delay minutes with breakdowns by specific cause	Completion, Cancellation, and diversion indicators
Scale:	Approx. 3 million flight records				
Coverage:	Domestic US flights with comprehensive operational details				

We Employed a Deliberate Four-Point Analytical Decision Framework



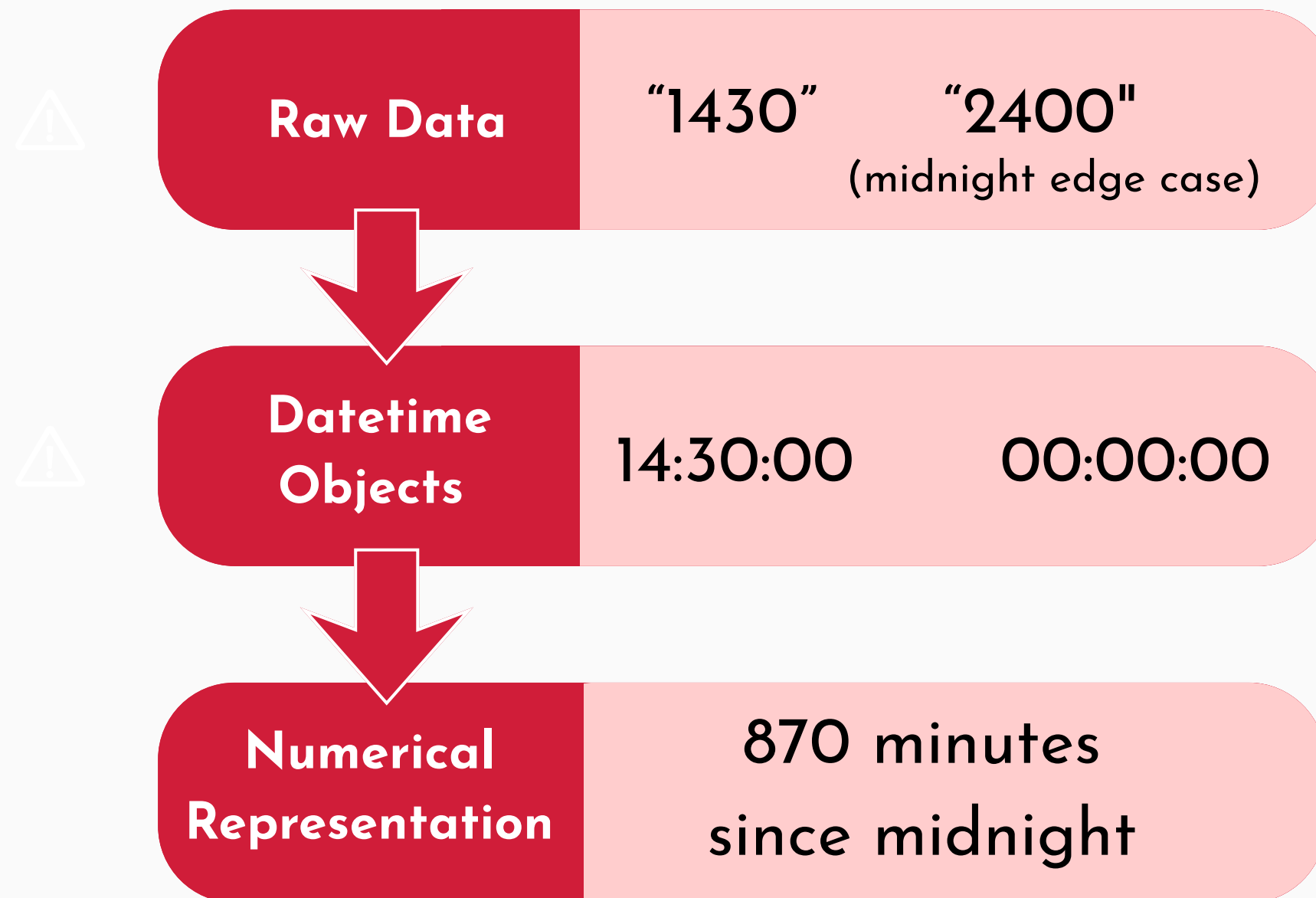
Decision 1 - We Focused the Analysis on Completed



Decision 2 - Missing Data Strategy

Data Category	Our Approach	Justification
CRITICAL IDENTIFIERS (FL_DATE, AIRLINE, ORIGIN, DEST)	DROP INCOMPLETE RECORDS	CANNOT IMPUTE FLIGHT IDENTITY; RECORDS WITHOUT THESE FIELDS ARE UNUSABLE FOR ANALYSIS
DELAY METRICS (ARR_DELAY, DEP_DELAY, DELAY CAUSE COLUMNS)	FILL MISSING WITH ZERO	MISSING DELAY VALUES MOST LIKELY INDICATE ON-TIME PERFORMANCE; CONFIRMED BY CROSS-CHECKING WITH TIMING DATA
TIMING DETAILS (WHEELS_OFF, WHEELS_ON, TAXI TIMES)	DROP RECORDS WITH MISSING VALUES	THESE FIELDS HAD MINIMAL MISSING DATA (<2%) AND ARE NEEDED FOR VALIDATION CALCULATIONS

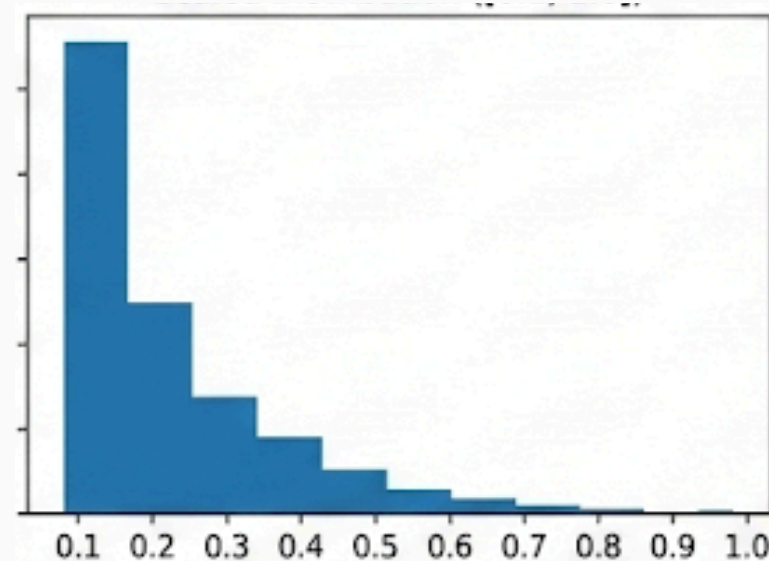
Decision 3 - Time Data Transformation



This multi-step process enables proper time arithmetic, handles critical edge cases, and produces a clean numerical format required for modeling and calculations.

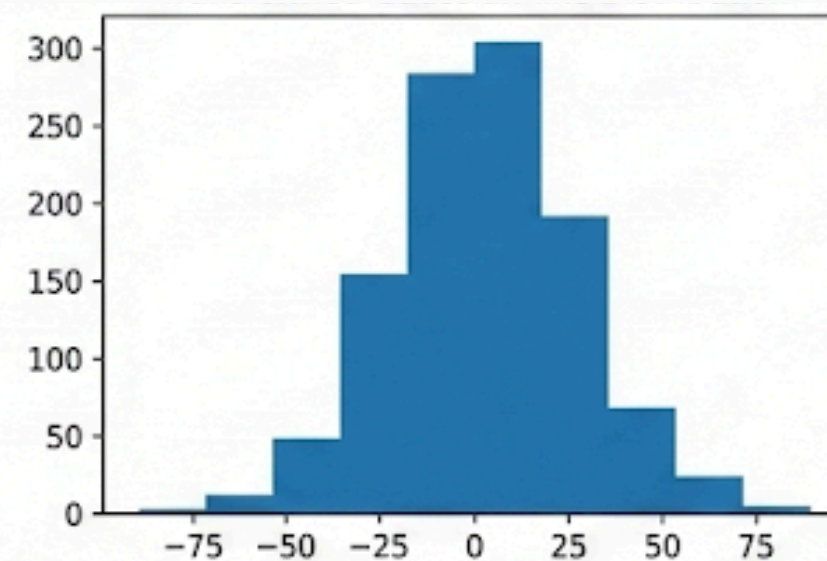
Decision 4 - Multi-Scale Preprocessing

Min Max



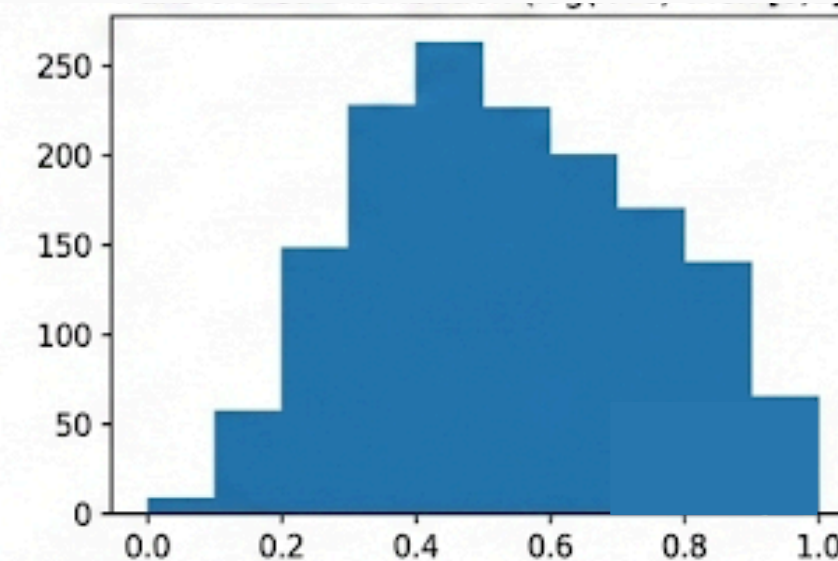
Scale to range
[0.1, 1.0]

Yeo-Johnson Power Transform



Power transformation
for normalization

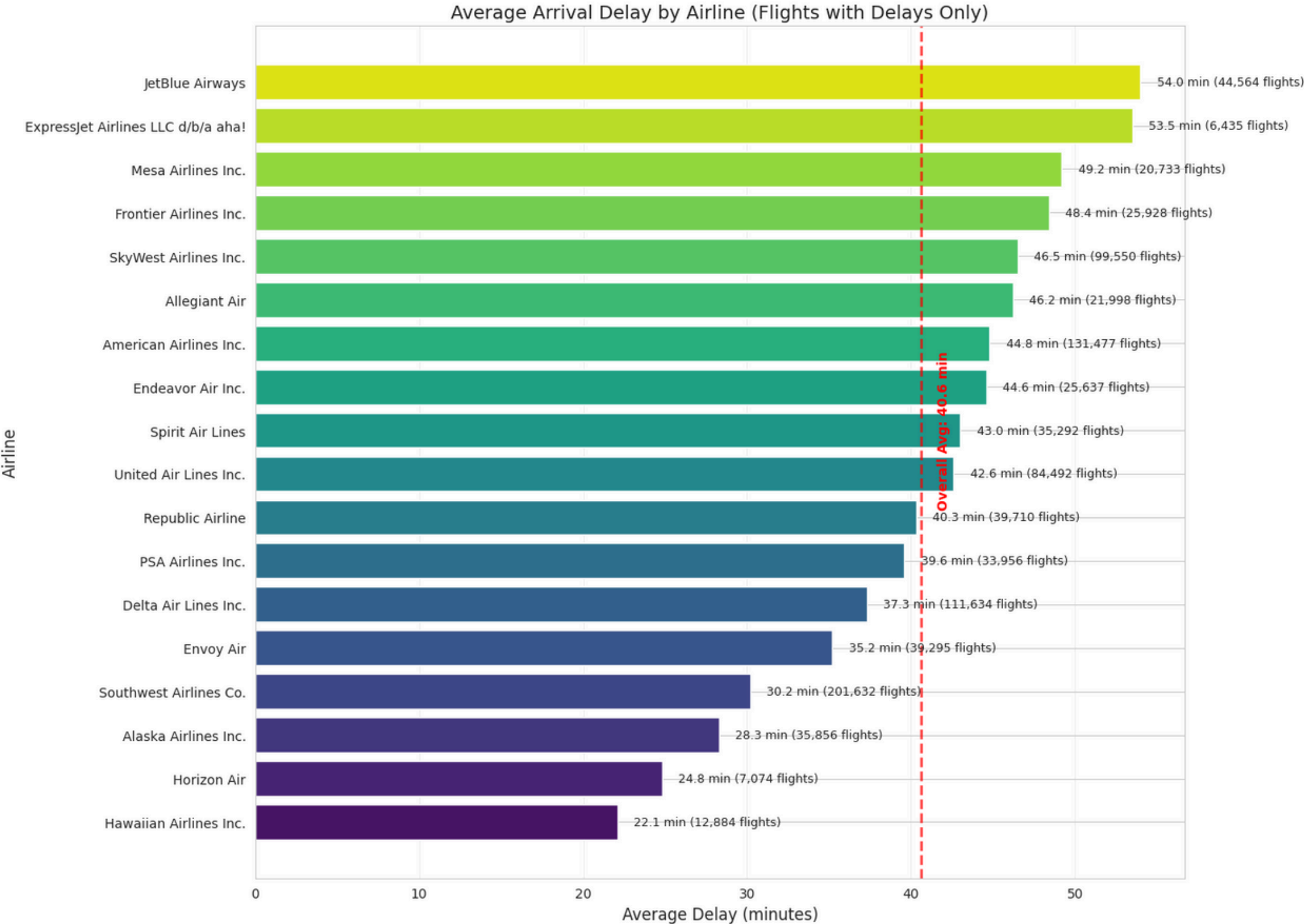
Log Scaling



$\log(x + \epsilon)$ then
normalize to [0,1]

Airline-Wise Average Arrival Delay

2019-2023 sample from DOT flight records



Best On Time Performance

Delta (DL) has the lowest average arrival delay.



Most Delayed Carrier

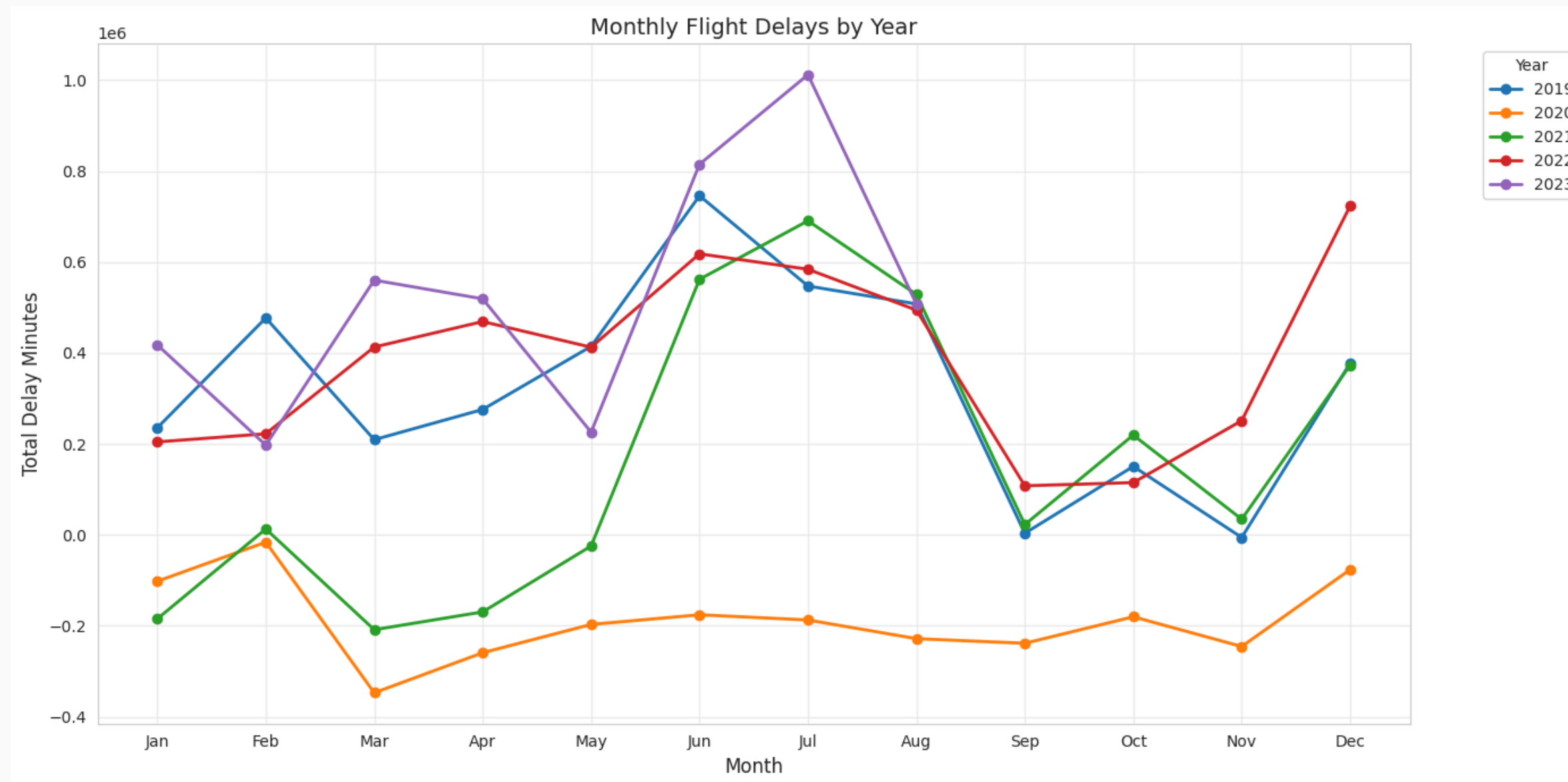
Spirit (NK) shows highest average arrival delay.



Overall Pattern

Most major carriers cluster are within a small delay range, but low cost carrier skew higher.

What We Learned



What We Did

Aggregate delays by month and year, EDA of Temporal trends.



Why We Did it

To understand seasonal patterns and pandemic impact.

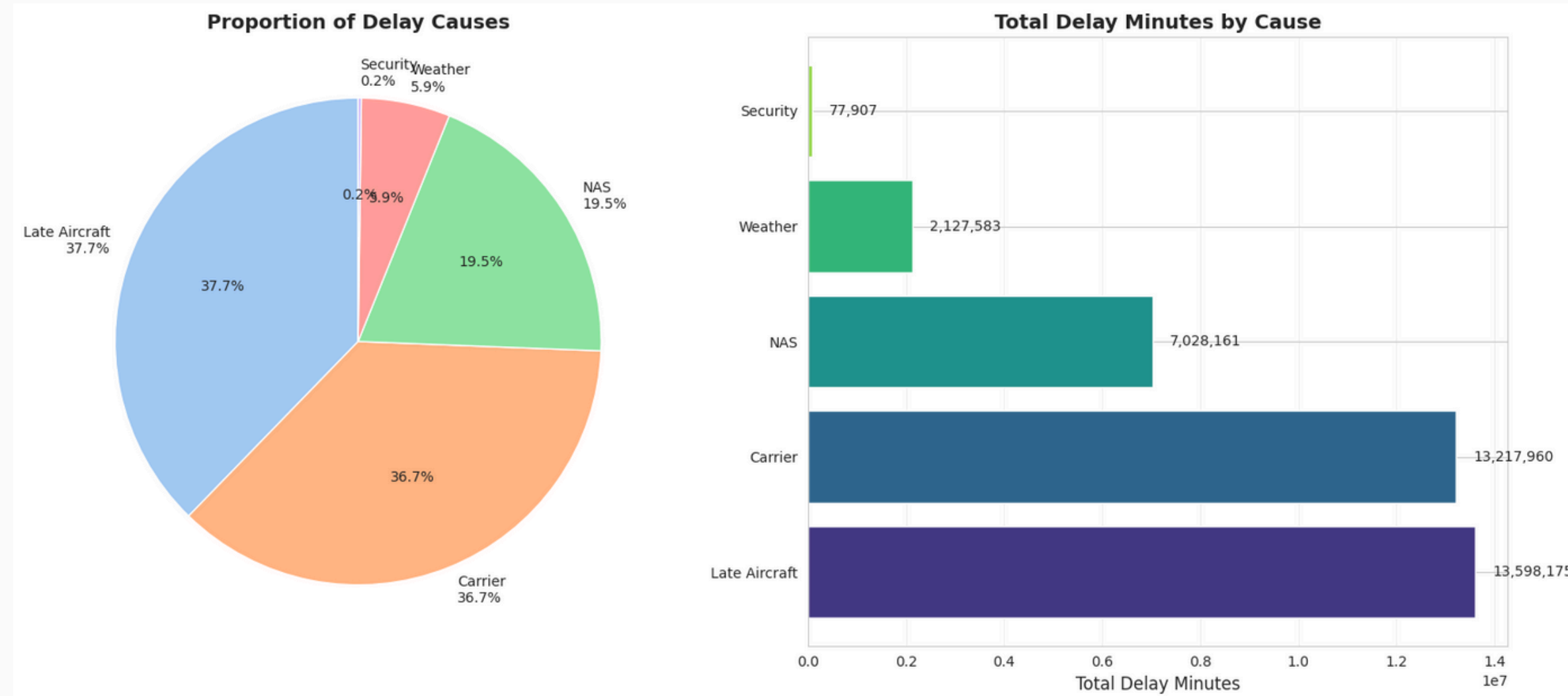


What We Learned

2020 drop, consistent Summer peak, 2022 spike, variability in Winter.

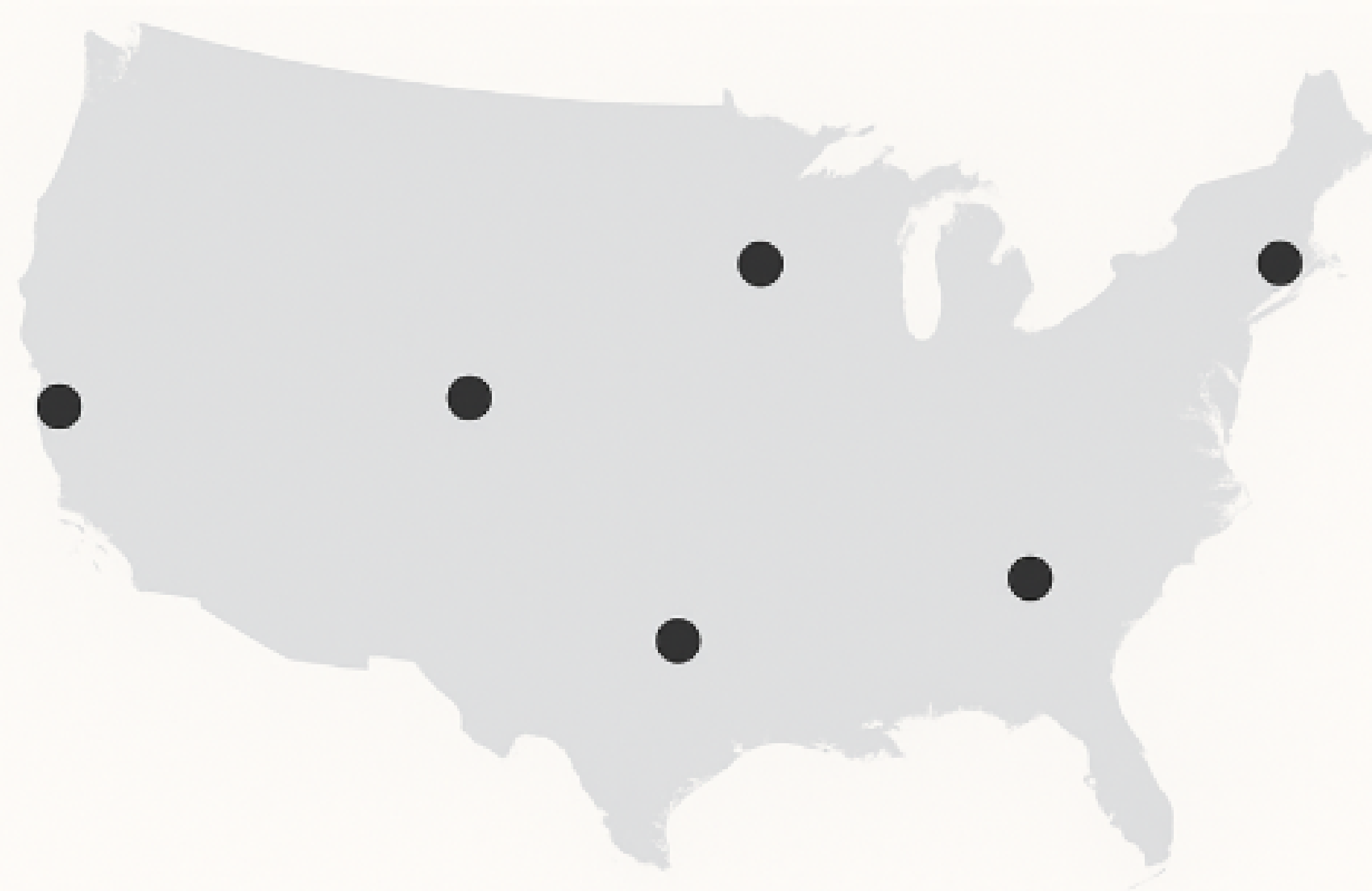
What Drives Flight Delays?

Delay Minutes by Cause 2019 - 2023



We focused on known delay facts to pinpoint largest contributors, Late Aircraft followed by carrier and NAS.

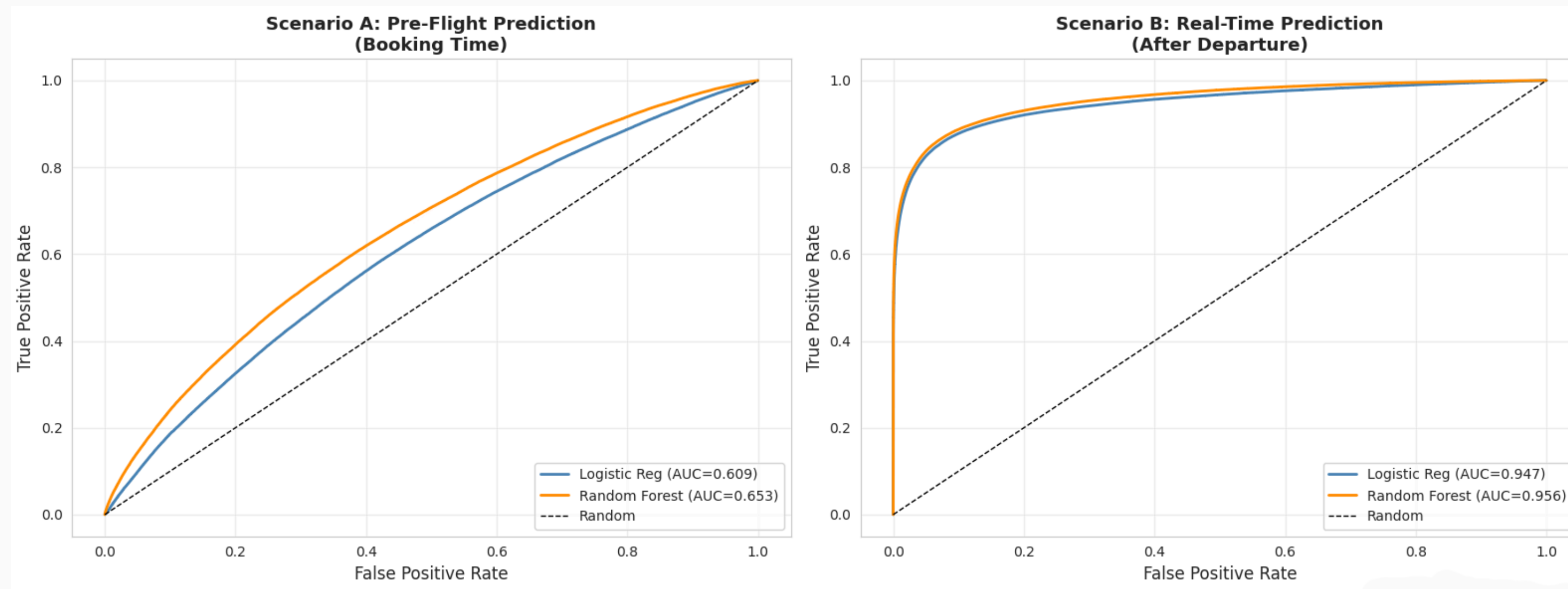
Top Destination Airports



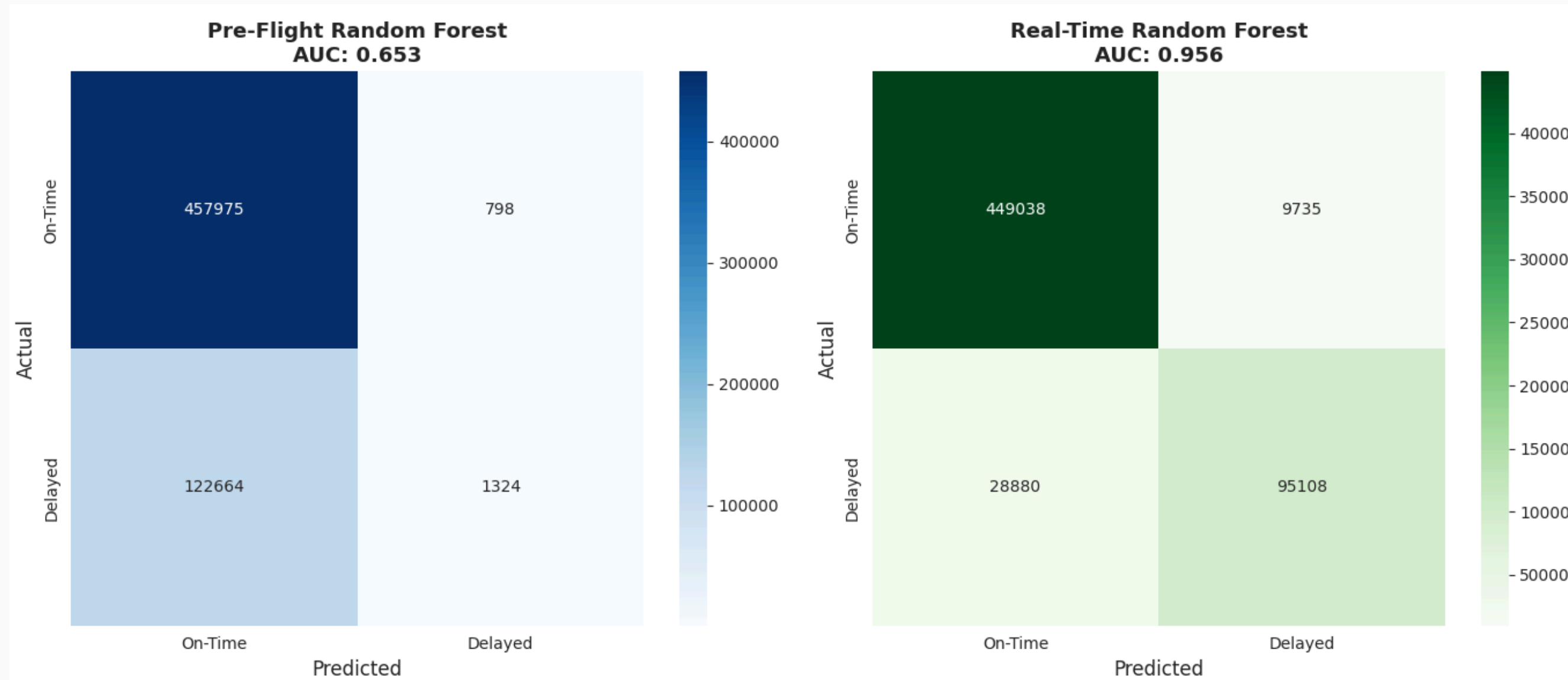
- We've filtered data to focus on most frequently visited airports.
- This is to ensure statistical reliability and meaningful results.
- We noticed the patterns of delays are common at major airport hubs.

Predictive Modeling: Can We Forecast Flight Delays?

ROC-AUC Curve



Confusion Matrix



Limitations

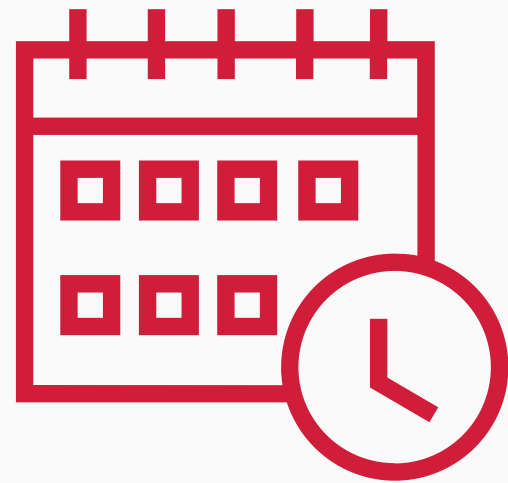
- Analyzed only top 6 airports
- We've excluded the cancelled flights
- Time zones not normalized

Future Work

- Cancellation classification
- Stratified sampling & cancellation
- Time series & analysis

Key Takeaways

Seasonal Patterns



Clear Summer Peaks
each year

Carrier Differences



Strong variation in
airline by mean delay

Operations Matter



Internal processes huge
drivers of delay

THANK YOU