Airline Flight Delays Analytics

Uncovering Patterns, Reducing Delays, Improving Reliability

Data-Driven Insights from 6 Million U.S. Commercial Flights

Dataset: Airline Flight Delays (Maven Analytics)

Executive Summary

- **Dataset:** Airline Flight Delays

- **Tools:** SQL (MySQL) + Power BI for data transformation & visualization

- **Deliverables:** 30+ KPI-driven analytical views across 5 dashboards

- **Goal:** Identify inefficiencies and improve flight reliability

The Business Problem

The aviation industry faces billions in annual losses from operational inefficiencies:

- Flight delays impacting customer satisfaction

- Congestion reducing airport & airline efficiency

- Scheduling gaps and route bottlenecks

- Ground-time inefficiencies increasing costs

Central Question

How can airlines and airports reduce delays, cancellations, and congestion

while improving on-time performance?

Project Scope

- Airline performance metrics & rankings

- Root causes of delays and cancellations

- Seasonality and time-based patterns

- Airport & route efficiency analysis

- Operational cost impact assessment

Project Objectives

- Measure airline & airport performance via quantifiable KPIs

- Identify why, when, and where delays occur

- Determine most reliable airlines

- Analyze seasonality for scheduling optimization

- Provide actionable recommendations

Key KPIs Measured (Part 1)

Airline Performance

On-time rates, delay averages, cancellation rates, diversion rates

Delay Analysis

Distribution, causes, cost impact, delay brackets

Key KPIs Measured (Part 2)

Seasonality & Trends

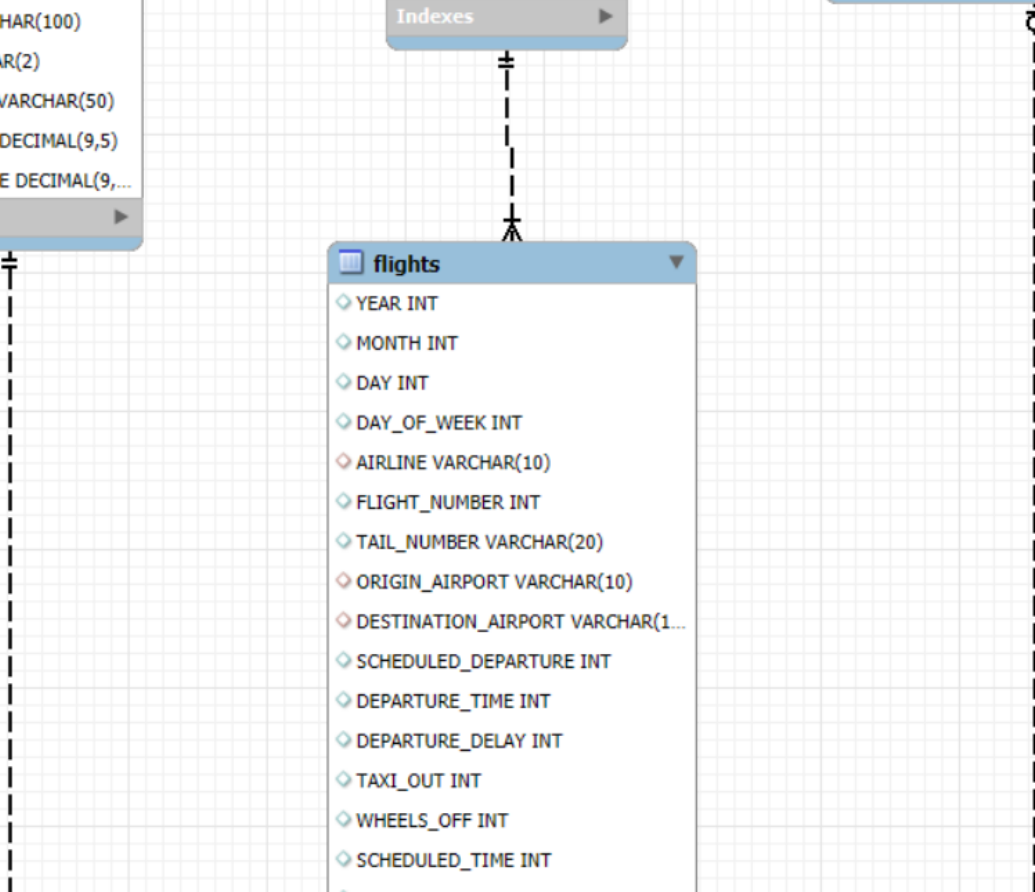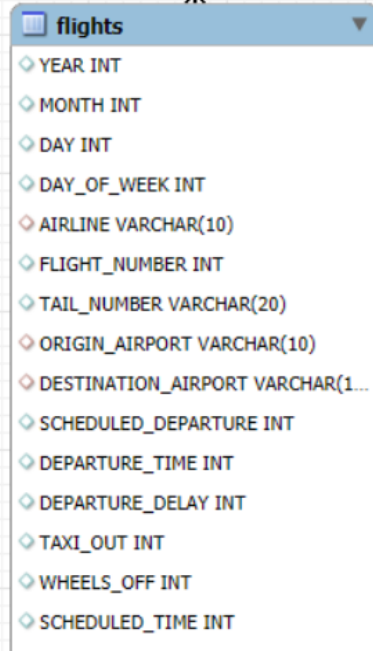Monthly trends, hourly patterns, time-of-day analysis
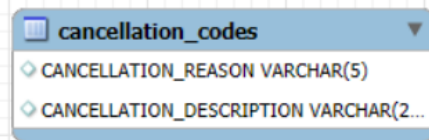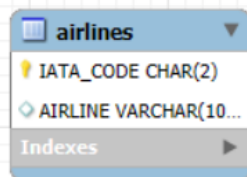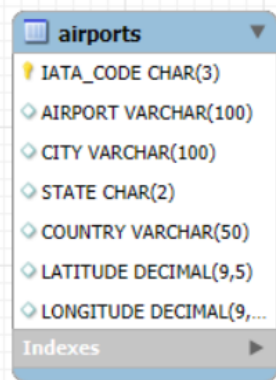
Airport & Routes

Top airports, busiest routes, taxi times, performance

Data Source & Preparation

- **Dataset:** Maven Analytics - U.S. domestic flight records

- **Size:** ~6 million flight records

- **Cleaning:** Removed duplicates, handled NULLs, standardized codes

- **Engineering:** Time-of-day extraction, delay brackets, window functions

Data Modeling Architecture

- **Star Schema:** Optimized for reporting performance

- **Entities:** Airports, Airlines, Cancellation Codes, Flights (fact table)

- **Relationships:**
    - **(One → many) relationships:**
        - Airports → Flights (Origin & Destination)
        - Airlines → Flights
        - Cancellation Codes → Flights

- **Validation:** Cross-checked codes, values, and data integrity

airports
- IATA_CODE CHAR(3)
- AIRPORT VARCHAR(100)
- CITY VARCHAR(100)
- STATE CHAR(2)
- COUNTRY VARCHAR(50)
- LATITUDE DECIMAL(9,5)
- LONGITUDE DECIMAL(9,...
- Indexes

airlines
- IATA_CODE CHAR(2)
- AIRLINE VARCHAR(10...
- Indexes

cancellation_codes
- CANCELLATION_REASON VARCHAR(5)
- CANCELLATION_DESCRIPTION VARCHAR(2...

flights
- YEAR INT
- MONTH INT
- DAY INT
- DAY_OF_WEEK INT
- AIRLINE VARCHAR(10)
- FLIGHT_NUMBER INT
- TAIL_NUMBER VARCHAR(20)
- ORIGIN_AIRPORT VARCHAR(10)
- DESTINATION_AIRPORT VARCHAR(1...
- SCHEDULED_DEPARTURE INT
- DEPARTURE_TIME INT
- DEPARTURE_DELAY INT
- TAXI_OUT INT
- WHEELS_OFF INT
- SCHEDULED_TIME INT

1:1

1:n

1:1

1:n

n:m
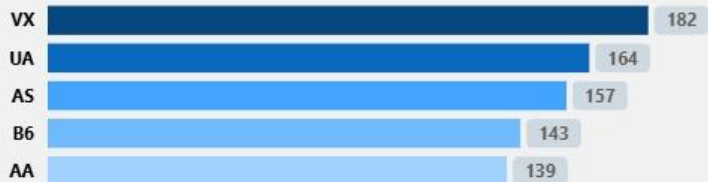
1:n

Five Analytical Dashboards

- **1. Airline Performance:** On-time metrics, delay averages, rankings

- **2. Delay Analysis:** Distribution, causes, flight status breakdown

- **3. Cancellation Analysis:** Reasons, frequencies, impact metrics

- **4. Seasonality & Trends:** Hour/month/period patterns

- **5. Airport & Routes:** Busiest hubs, taxi times, geographic view

Airline Performance Insights

- **Top volumes:** WN, DL, AA handle highest flight loads

- **Best performer:** DL consistently delivers best on-time performance

- **Challenge airlines:** UA & NK show higher departure delays

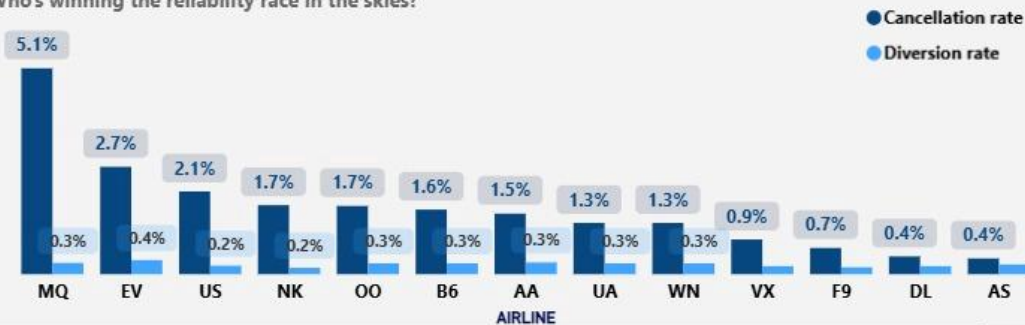- **Efficiency leaders:** VX and AS best for average airtime

Airline Performance

Average flight airtime by Airline

Airline	Airtime
VX	182
UA	164
AS	157
B6	143
AA	139

Tracking Flight Disruptions by Airline

Who's winning the reliability race in the skies?

● Cancellation rate
● Diversion rate

AIRLINE	Cancellation rate	Diversion rate
MQ	5.1%	0.3%
EV	2.7%	0.4%
US	2.1%	0.2%
NK	1.7%	0.2%
OO	1.7%	0.3%
B6	1.6%	0.3%
AA	1.5%	0.3%
UA	1.3%	0.3%
WN	1.3%	0.3%
VX	0.9%	
F9	0.7%	
DL	0.4%	
AS	0.4%	

AIRLINE	Total flights	# of on time arrivals	% of on time arrival	% of total flights (arrives on time)	Average arrival delay	# of on time departures	% of on time departure	% of total flights (departure on time)	Average departure delay
WN	1245812	775045	62.21	13.53%	4.36	679229	0.55	11.86%	10.57
DL	872057	621217	71.24	10.84%	0.19	589672	0.68	10.29%	7.35
AA	715065	462874	64.73	8.08%	3.44	469515	0.66	8.20%	8.86
OO	578393	355958	61.54	6.21%	5.83	407212	0.70	7.11%	7.77
EV	556746	343529	61.70	6.00%	6.56	387243	0.70	6.76%	8.66
UA	509150	322923	63.42	5.64%	5.42	252909	0.50	4.41%	14.38
MQ	279607	176102	62.98	3.07%	6.44	186375	0.67	3.25%	10.01
B6	262772	160774	61.18	2.81%	6.66	160760	0.61	2.81%	11.50
US	194648	118363	60.81	2.07%	3.70	132196	0.68	2.31%	6.11
AS	171852	114899	66.86	2.01%	-0.97	128311	0.75	2.24%	1.78
NK	115375	58488	50.69	1.02%	14.45	63342	0.55	1.11%	15.91
F9	90248	49016	54.31	0.86%	12.48	55389	0.61	0.97%	13.32
HA	76101	45922	60.34	0.80%	2.02	55961	0.74	0.98%	0.48
VX	61369	37189	60.60	0.65%	4.73	38003	0.62	0.66%	9.01
Total	5729195	3642299		63.57%	5.38	3606117		62.94%	8.98

Delay Analysis Insights

46% of flights experience delay — systemic inefficiency across industry

- Evening & Night periods show highest delays

- Late aircraft is largest delay contributor

- Major delays (60+ mins) concentrated in WN & DL

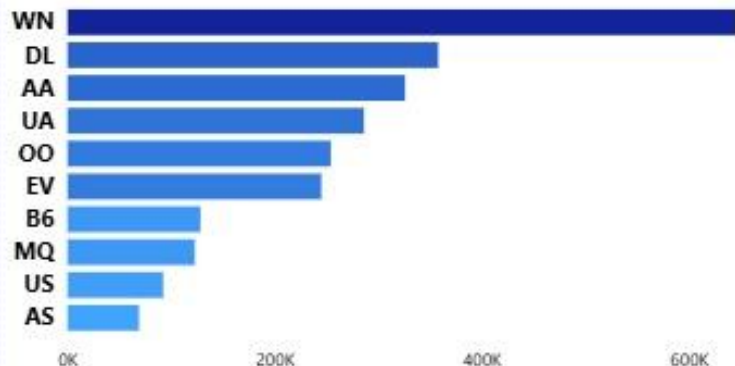- Cascading effect from upstream flight disruptions
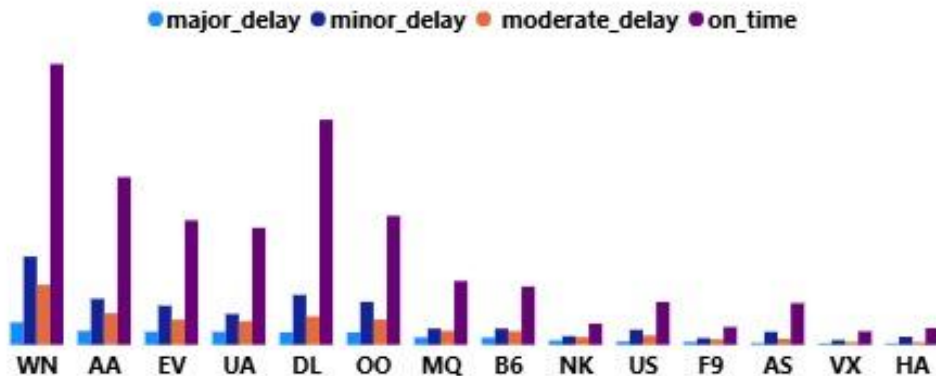
Delay Analysis

TOTAL FLIGHTS
6M
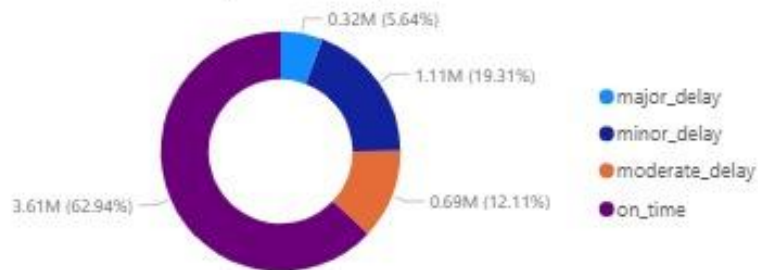
AVG DELAY / FLIGHT
10.77

Delayed flights
0M 2.09M 6M

Top 10 Delayed Airlines

Airline	
WN	(largest, ~620K)
DL	~370K
AA	~340K
UA	~280K
OO	~260K
EV	~250K
B6	~140K
MQ	~130K
US	~90K
AS	~60K

(axis: 0K, 200K, 400K, 600K)

Arrival Delay Distribution

● major_delay ● minor_delay ● moderate_delay ● on_time

Airlines: WN, AA, EV, UA, DL, OO, MQ, B6, NK, US, F9, AS, VX, HA

Departure Delay Distribution

- 0.32M (5.64%) — major_delay
- 1.11M (19.31%) — minor_delay
- 0.69M (12.11%) — moderate_delay
- 3.61M (62.94%) — on_time

● major_delay
● minor_delay
● moderate_delay
● on_time

Delay Causes

(axis: 0%, 50%, 100%)

● Air_system_delay
● Airline_delay
● Aircraft_delay
● Security_delay
● Weather_delay
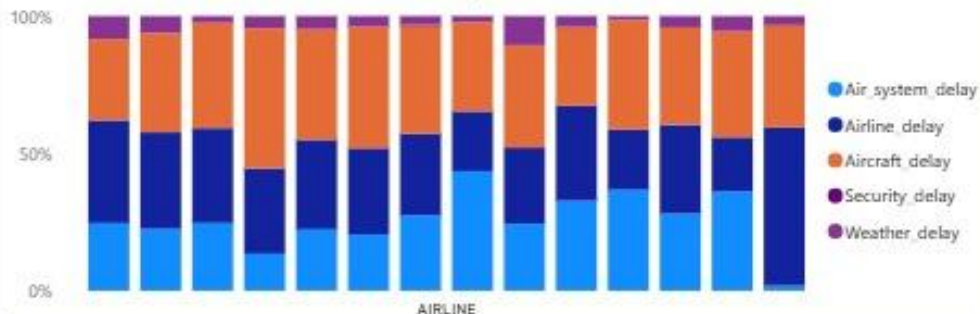
AIRLINE

Cancellation Root Causes

54%

National Air System issues

28%

Weather-related

High count

MQ, EV, WN airlines

Cancellation Analysis

TOTAL FLIGHTS	TOTAL CANCELLATION	# AIRLINES
6M	89.88K	14

Flights VS Cancellations

Sum of cancellations

CANCELLATION_REASON

16K (17.52%)

25K (28.11%)

49K (54.35%)

REASONS
- Weather
- Airline/Carrier
- National Air System
- Security
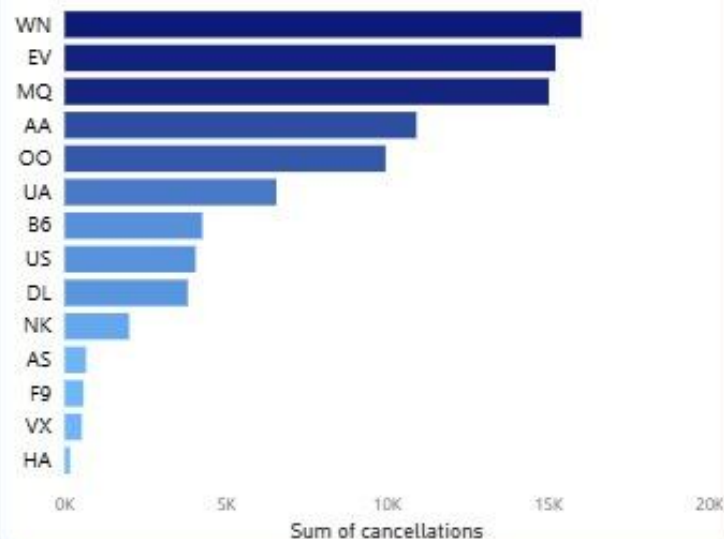
FLIGHT STATUS

63.57%	36.43%	1.54%
On Time	Delayed	Cancelled
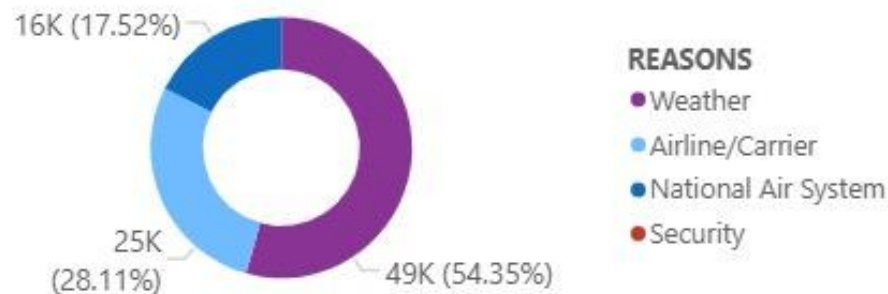
Diversions by Airlines

3.4K

60.0

Sum of diversions

AIRLINE
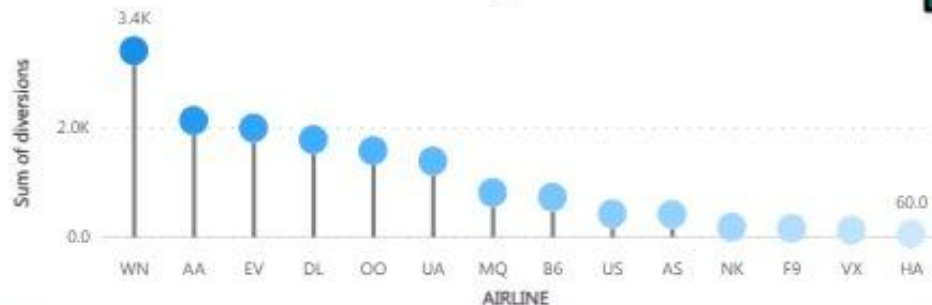
Airport & Route Insights

- **Busiest hubs:** ATL, ORD, DFW handle ~34% of total traffic

- **High taxi times:** JFK, LGA, ORD → major delay contributors

- **Departure delays:** LAX shows congestion & ground inefficiencies

- **Implication:** Hub airports critical for system-wide optimization

Airport & Route Analysis

Top 10 Origin Airports
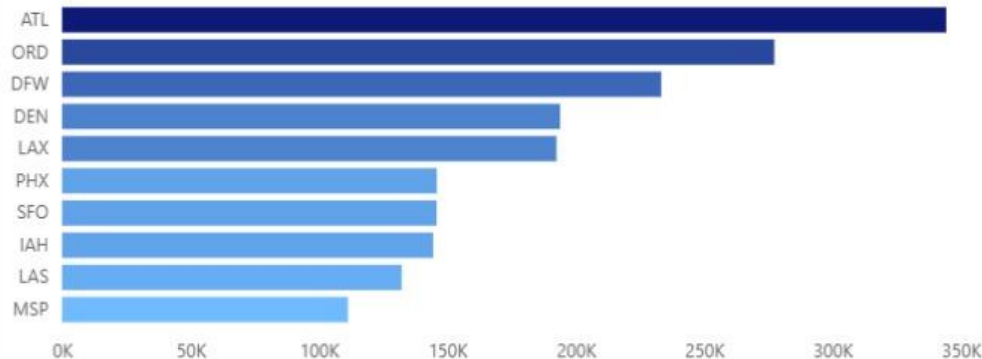
Total Flights

5.73M

Top 10 Airports Flight Number

1.92M

Pct of Top 10 Flights to Total Flights

33.49%

Top 10 Destination Airports

Top 10 Destinations

Microsoft Azure

©2025 EarthStar Geographics Feedback
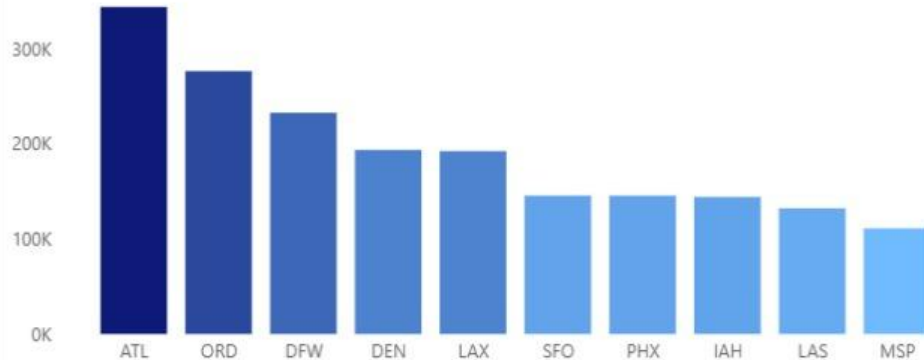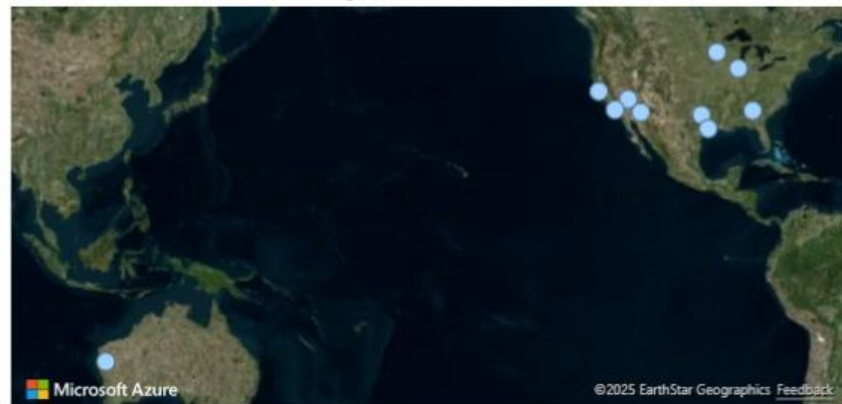
Seasonality & Time Patterns

- **Peak volume:** July with 521K flights

- **Lowest volume:** February with elevated delays despite fewer flights

- **Time-of-day:** Evening (6-8 PM) worst delays; early morning optimal

- **Root cause:** Cumulative operational drift throughout the day

Seasonality & Trends

Number of flights by departure hour

At 4 am, the least departures of 531 flights happen. Meanwhile at 6 am, the highest departures happ...
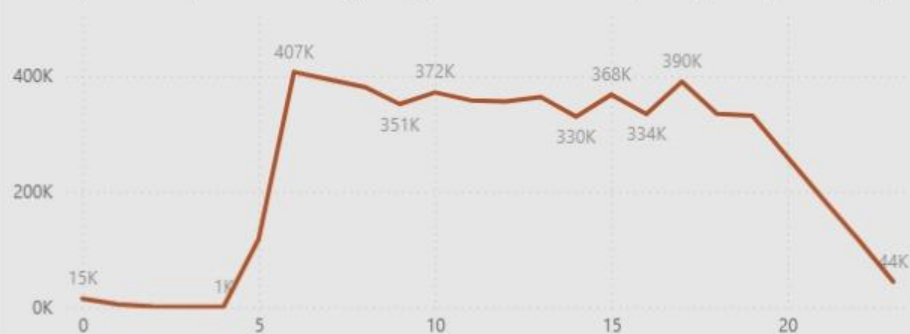
When most delays happen by time of day.

● Average arrival delay ● Average departure delay

Monthly and Daily Flight Trends

Busiest Airports

- **LAX is the busiest airport** and shows the **highest departure delays**, indicating heavy congestion.

- **JFK and LGA have the longest taxi-out times**, reflecting severe takeoff congestion in the New York area.

- **ORD has the highest taxi-in time**, slowing arrivals and increasing turnaround delays.

- Despite high traffic, **LAS and SFO maintain relatively lower delays**, showing stronger operational efficiency.

- Traffic is highly concentrated: **106K+ flights** occur between a small group of major hubs.

- Because of this concentration, **disruptions at LAX, JFK, or ORD can cause widespread delays** across the network.
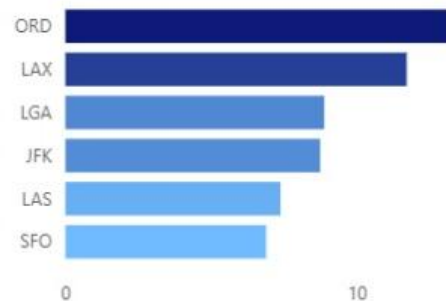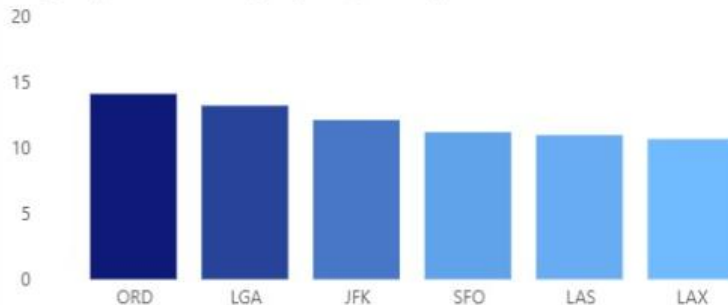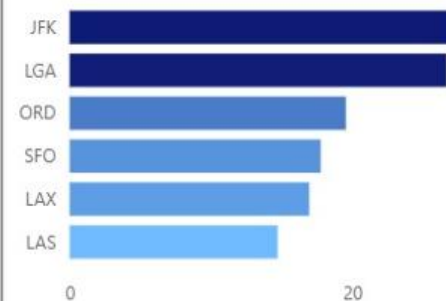
Recommendations (Part 1)

- **Study & replicate:** DL's operational excellence at scale

- **Aircraft turnaround:** Optimize gate scheduling for high-delay airlines

- **Buffer time:** Add capacity during peak periods to reduce cascades

- **Taxi optimization:** Redesign routes at JFK, LGA, ORD

Recommendations (Part 2)

- **Scheduling:** Reduce 6-8 PM departures; increase early morning flights

- **Cancellations:** Enhance air traffic coordination during weather

- **Technology:** Deploy ML predictive delay models

- **Integration:** Real-time weather & ATC alerts in scheduling systems

Future Enhancements

- Machine Learning delay prediction model

- Passenger impact & missed connection analysis

- Cost optimization & savings modeling

- Real-time dashboard automation with live APIs

- Advanced route optimization algorithms

Key Takeaway

Data-driven operational insights enable airlines and airports to systematically reduce delays, improve reliability, and enhance customer experience while optimizing costs.

Questions?

Let's explore the dashboards and dive deeper into the data