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

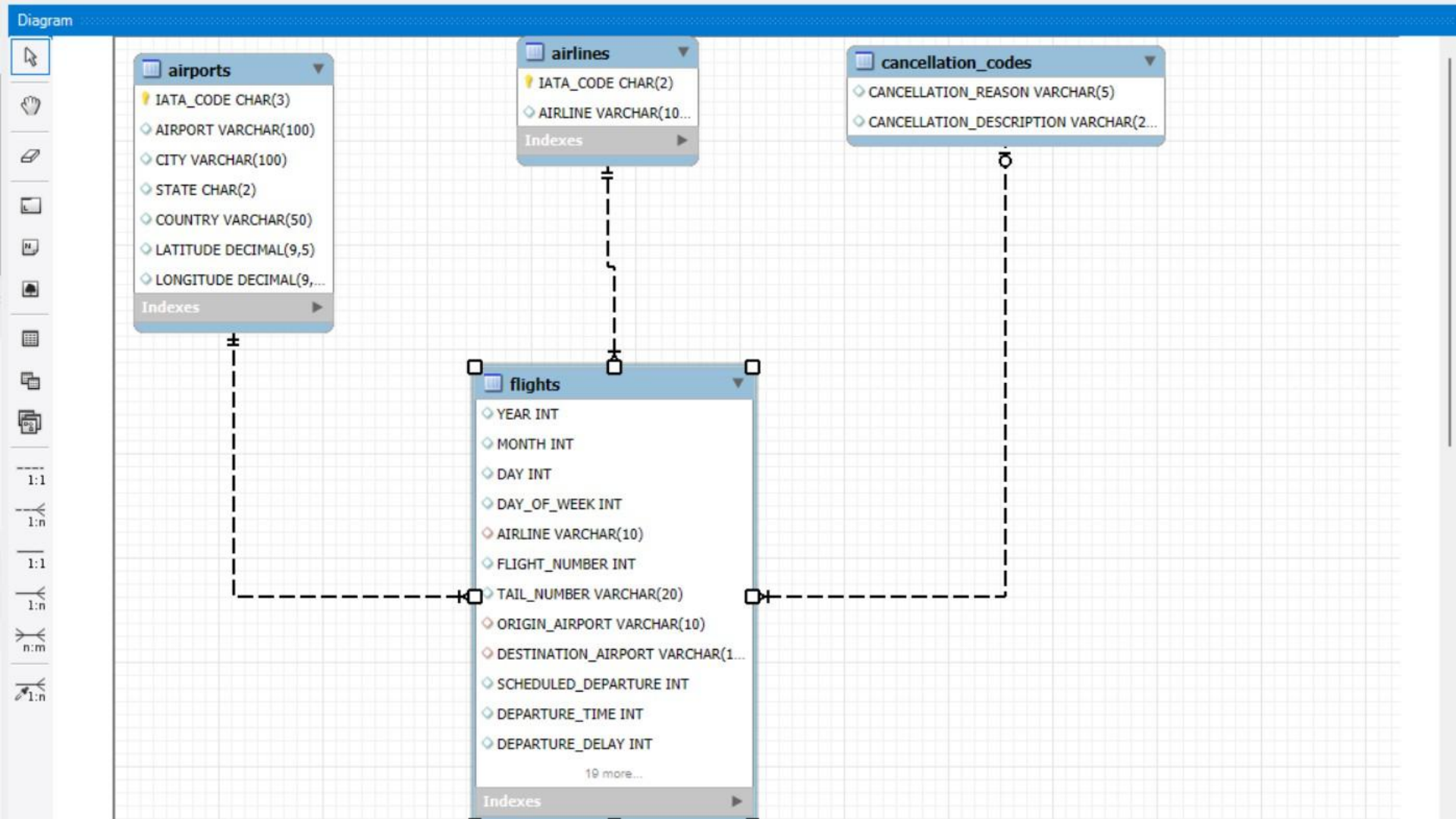


Diagram
airports
IATA_CODE CHAR(3)
AIRPORT VARCHAR(100)
CITY VARCHAR(100)
STATE CHAR(2)
COUNTRY VARCHAR(50)
LATITUDE DECIMAL(9,5)
LONGITUDE DECIMAL(9,...
Indexes
airlines
IATA_CODE CHAR(2)
AIRLINE VARCHAR(10...
Indexes
cancellation_codes
CANCELLATION_REASON VARCHAR(5)
CANCELLATION_DESCRIPTION VARCHAR(2...
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
19 more...
Indexes
1:1
1:n
1:1
1:n
n:m
1:n

7 Analytical Dashboards

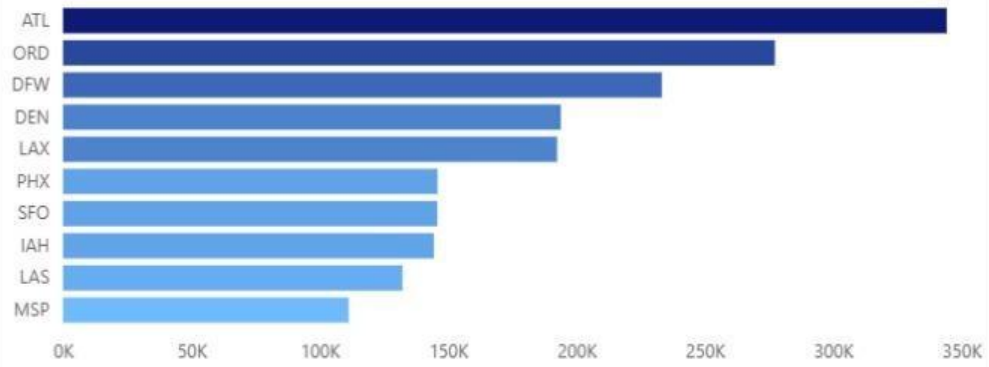
- **1. Airline Performance:** On-time metrics, delay averages, rankings
- **2. Delay Analysis:** Distribution, causes, flight status breakdown
- **3. Cancellation Analysis:** Reasons, frequencies, impact metrics
- **4. Seasonality & Trends:** Hour/month/period patterns
- **5. Airport & Routes:** Busiest hubs, taxi times, geographic view
- 6- Busiest Airports and their effects on Delay
- 7- Aircraft utilization & cost impact

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

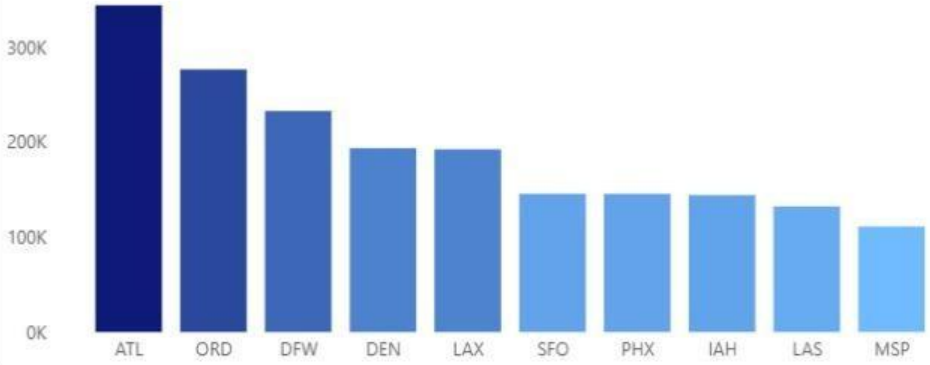
Top 10 Airports Flight Number

1.92M

Pct of Top 10 Flights to Total Flights

33.49%

Top 10 Destination Airports



Top 10 Destinations

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.

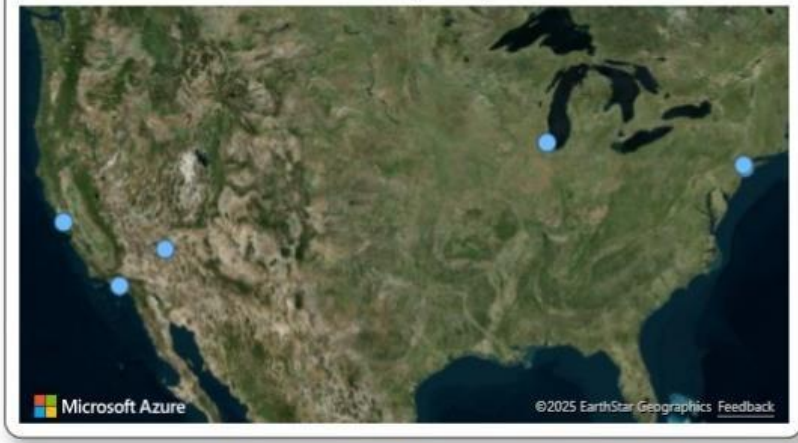
Busiest Airports and their effect on Delay

Total Flights between Busiest Airports

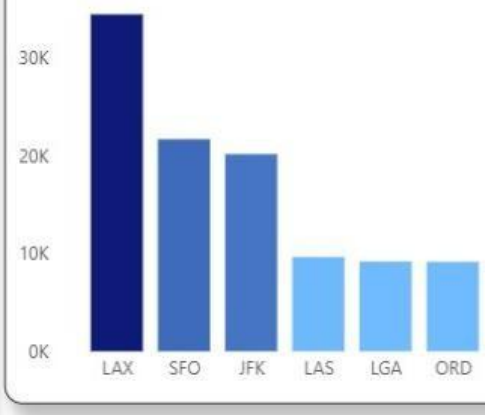
104.5K

Top 6 Destinations Airports

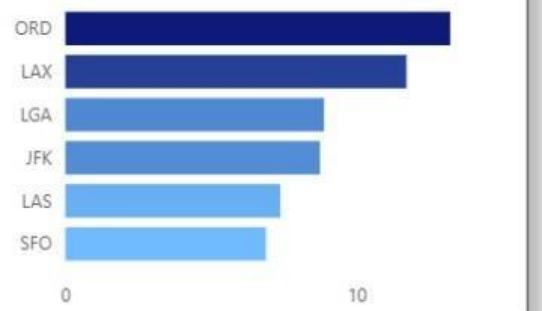
Busiest Origin Airport



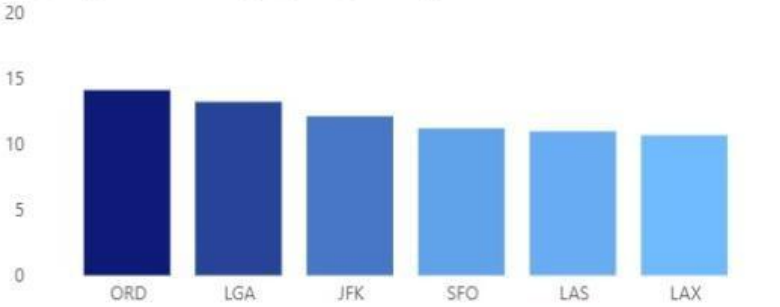
Top 6 Origins Airports



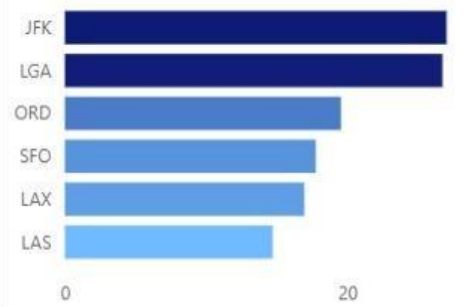
Avg Taxi in of Busiest Destination Airport



Avg departure delay by Origin Airports



Avg Taxi out of Busiest Origin Airport



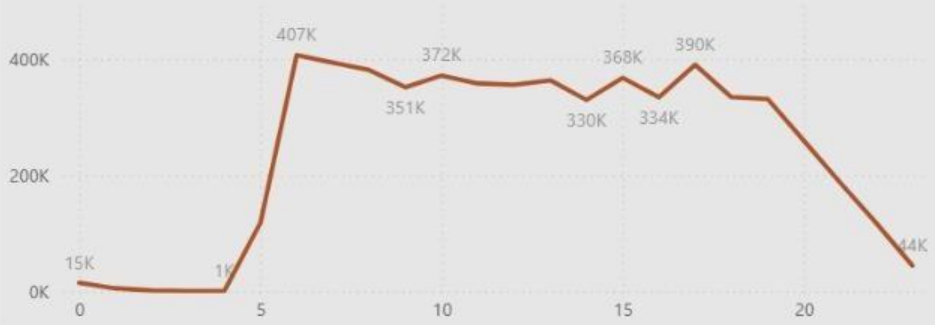
Seasonality & Time Patterns

- **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 happen

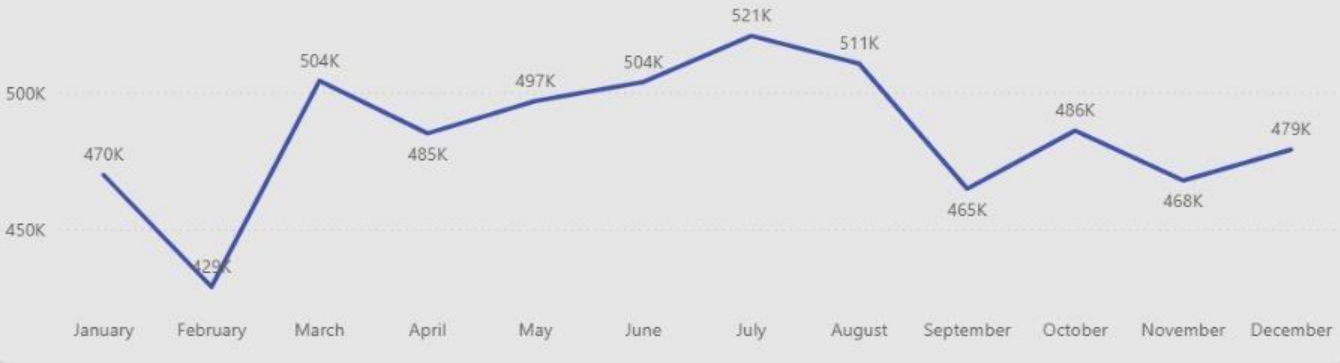


When most delays happen by time of day.

Average number of flights per day

2.28K

Monthly Flight Trends



Delay Analysis Insights

35.86% 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

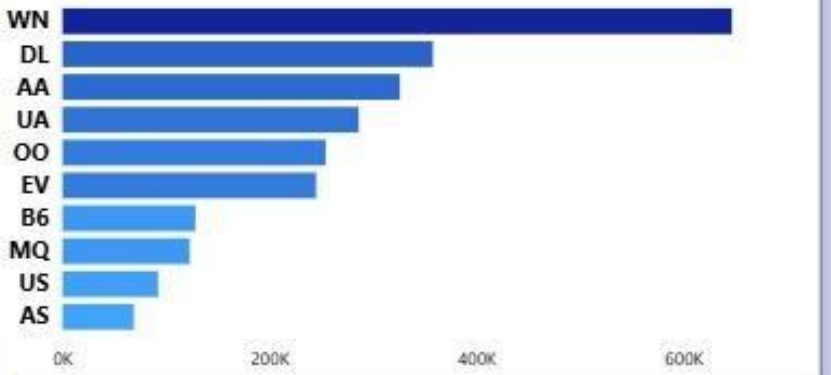
Delay Analysis

TOTAL FLIGHTS
6M

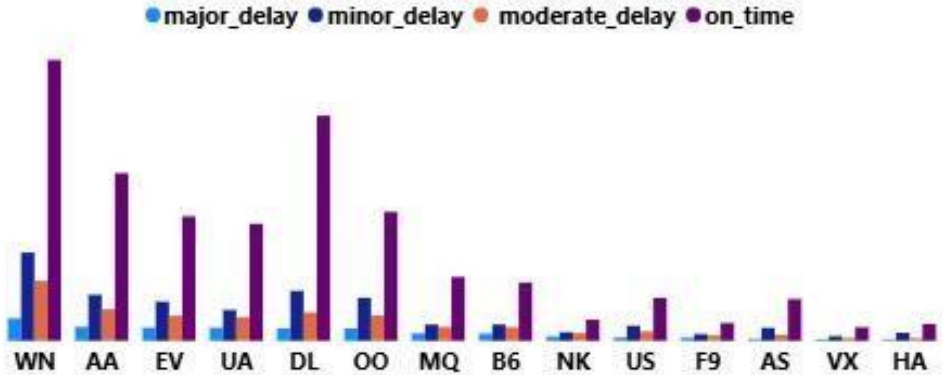
AVG DELAY / FLIGHT
10.77

Delayed flights

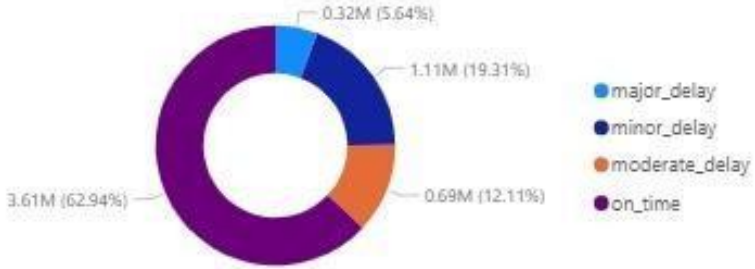
Top 10 Delayed Airlines



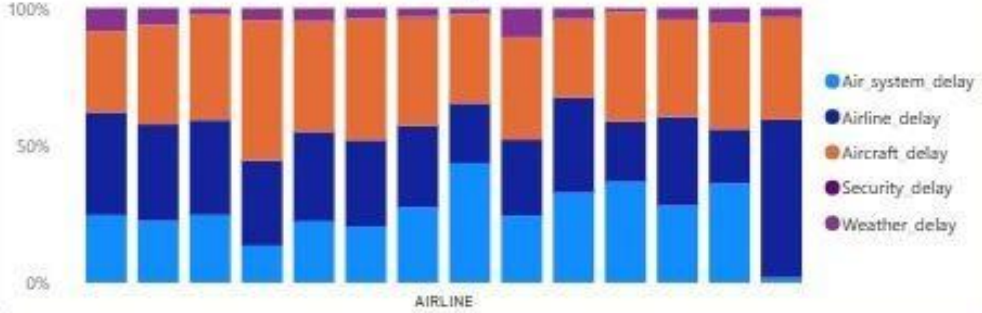
Arrival Delay Distribution



Departure Delay Distribution



Delay Causes



Cancellation Root Insights

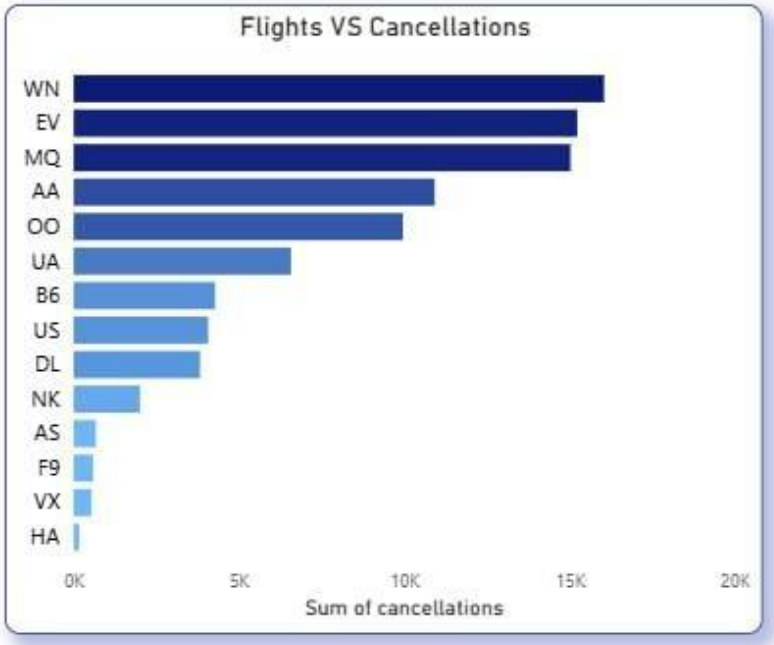
Airline/Carrier issues are the leading cause of cancellations, accounting for **over half (54%)** of all cancelled flights.

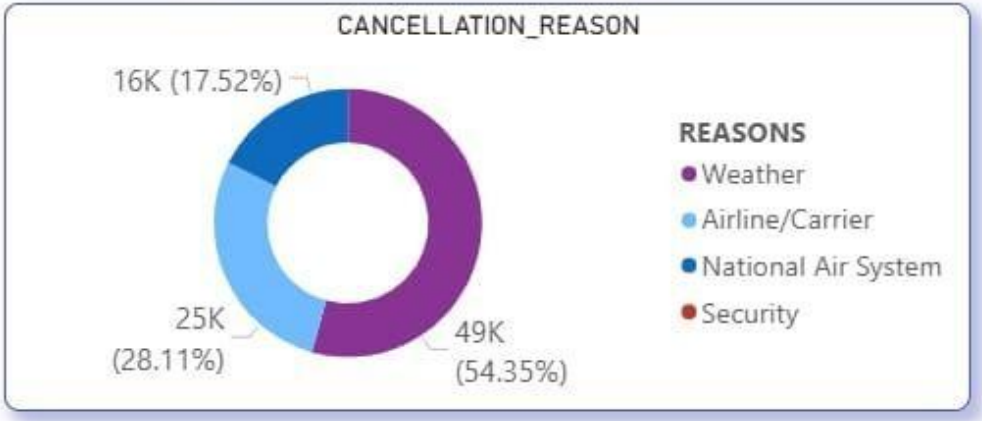
WN, EV, and MQ have the **highest cancellation counts**, significantly more than other airlines.

Despite cancellations, **most flights (62%) operate on time**, while only **1.54%** are cancelled and **0.26%** diverted.

Cancellation Analysis

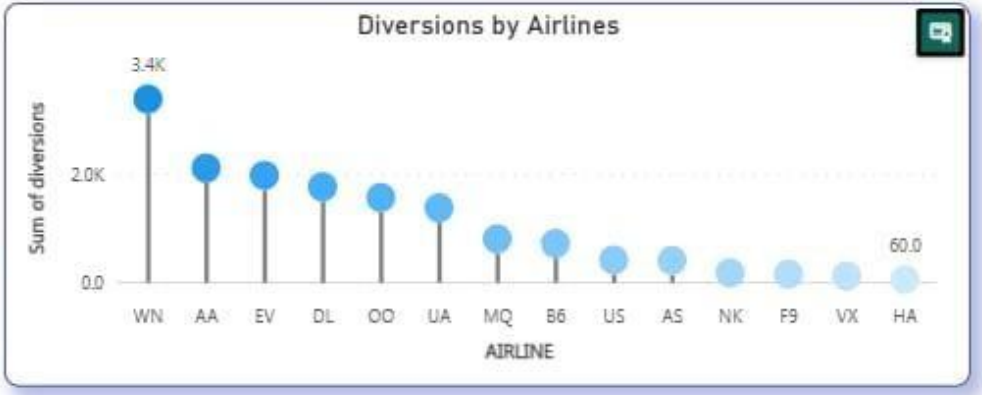
TOTAL FLIGHTS	TOTAL CANCELLATION	# AIRLINES
6M	89.88K	14





FLIGHT STATUS

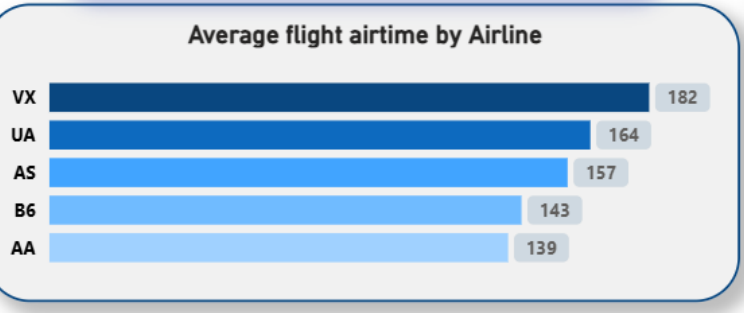
62.33%	35.86%	1.54%	0.26%
On Time	Delayed	Cancelled	Diverted

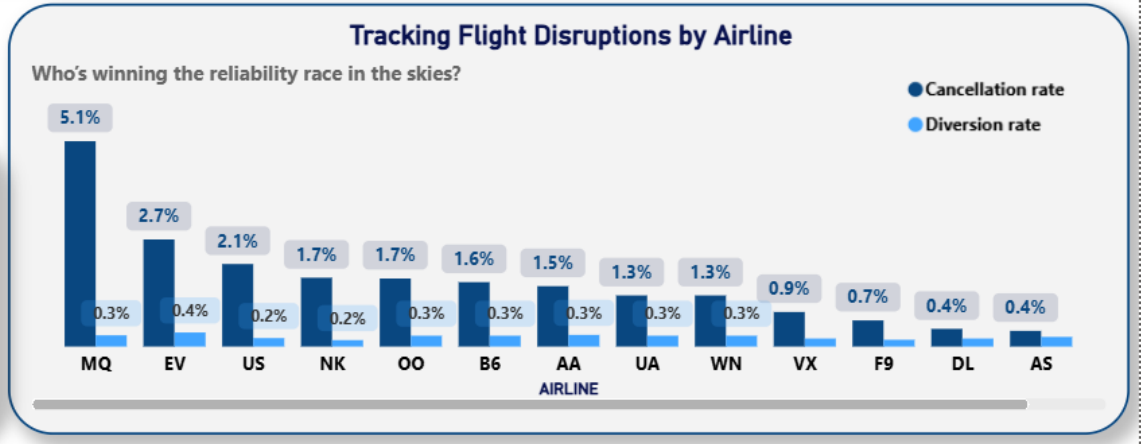


Airline Performance Insights

- **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





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	54.52%	11.86%	10.57
DL	872057	621217	71.24	10.84%	0.19	589672	67.62%	10.29%	7.35
AA	715065	462874	64.73	8.08%	3.44	469515	65.66%	8.20%	8.86
OO	578393	355958	61.54	6.21%	5.83	407212	70.40%	7.11%	7.77
EV	556746	343529	61.70	6.00%	6.56	387243	69.55%	6.76%	8.66
UA	509150	322923	63.42	5.64%	5.42	252909	49.67%	4.41%	14.38
MQ	279607	176102	62.98	3.07%	6.44	186375	66.66%	3.25%	10.01
B6	262772	160774	61.18	2.81%	6.66	160760	61.18%	2.81%	11.50
US	194648	118363	60.81	2.07%	3.70	132196	67.92%	2.31%	6.11
AS	171852	114899	66.86	2.01%	-0.97	128311	74.66%	2.24%	1.78
NK	115375	58488	50.69	1.02%	14.45	63342	54.90%	1.11%	15.91
F9	90248	49016	54.31	0.86%	12.48	55389	61.37%	0.97%	13.32
HA	76101	45922	60.34	0.80%	2.02	55961	73.54%	0.98%	0.48
VX	61369	37189	60.60	0.65%	4.73	38003	61.93%	0.66%	9.01
Total	5729195	3642299			5.38	3606117			8.98

Aircraft utilization & cost impact

Key Insights

Network Scale & Cost Impact

- 5.7M flights operated, yet $69M lost to delays, proving small delays multiply into massive costs at scale.

The Utilization Trade-off

- Aircraft dwell time is only 22 minutes, efficient but risky. Any disruption (late arrival, slow boarding, maintenance) cascades into cascading delays across the network.

Airline Performance

- WN leads in volume (1.25M flights, $13.8M delay cost), scale amplifies impact.
- AA and DL each face $8M costs despite strong performance elsewhere, showing delays are volume × minutes, not just operations.
- EV and UA lose $7M+ despite fewer flights, revealing structural inefficiencies in their hub-and-spoke networks.

Delay Distribution

- Average deviation is -5.88 minutes (early), yet high delay costs show a few severely delayed flights drive the financial impact.

Bottom Line

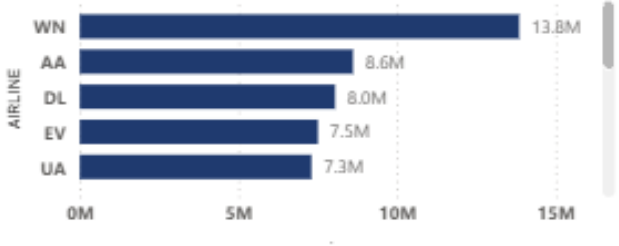
- The network prioritizes utilization over resilience. Adding just 3-5 minutes to dwell time on critical flights could significantly reduce downstream delays and save millions.

Aircraft Utilization & Cost Impact Dashboard

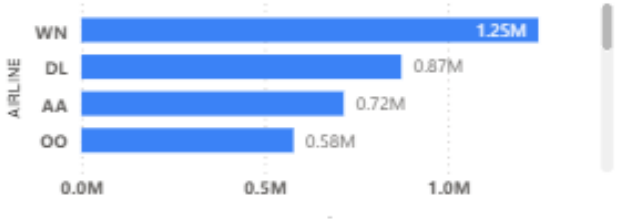
Filter

Total Number of Flights	Average Aircraft Dwell Time	Average Flight Distance	Average Time Deviation	Total Estimated Delay Cost
5.8M	22.10	889.34	-5.88	$69M

Estimated Cost by Airline



Total Flights by Airline



Airline Performance Summary

Airline Code	Total Flights	Avg. Flight Distance (mi)	Avg. Time Deviation (min)	Sum of Total Estimated Delay Cost (USD)
WN	1245812	740.71	-7.92	13,849,257.00
DL	872057	853.62	-8.06	8,046,301.00
AA	715065	1,041.34	-8.26	8,611,910.00
OO	578393	496.77	-3.77	7,215,186.00
EV	556746	462.25	-5.11	7,504,821.00
UA	509150	1,271.55	-11.85	7,301,833.00
MQ	279607	422.32	-8.45	4,089,400.00
B6	262772	1,062.18	-7.87	3,889,470.00
US	194648	911.50	-5.45	2,091,729.00
AS	171852	1,197.42	-3.65	1,284,997.00
NK	115375	985.27	-4.28	2,313,017.00
F9	90248	967.21	-2.11	1,698,364.00
Total	5729195	12,450.72	-82.30	69,103,367.00

Recommendations (Part 1)

Scheduling and buffer optimization

• Review schedule buffers by airline and by key route:
slightly increasing scheduled block time on chronically late routes while reducing unnecessary buffer on consistently early routes can keep the overall average close to zero while reducing extreme delays.

• Re-evaluate connection times at major hubs for the top airlines; adding a few minutes to critical bank waves can significantly cut missed-connection chains and downstream delays.

Turnaround and ground-time management

•For airlines with high delay cost and very short dwell times, pilot targeted turnaround process reviews (boarding, cleaning, refuelling, catering) to identify bottlenecks and standardize best practices.

•Introduce minimum protective dwell time thresholds for specific aircraft types or congested airports to reduce the risk that a single late arrival automatically creates a late departure.

Recommendations (Part 2)

Operational and maintenance focus

- Prioritize preventive maintenance windows for fleets with the highest delay impact per flight, even if this slightly reduces raw utilization, because avoided technical delays often save more money than the extra flight would have generated.
- Monitor delay cost per airline and per aircraft tail as a recurring KPI; use it to select where to invest in reliability improvements, crew training, and better ground coordination.

Cost and performance management

- Build "cost of delay per minute" benchmarks per airline and per route class (short/medium/long haul) to understand where each additional minute is most expensive.
- Use the dashboard in regular performance reviews with airline partners or internal operations teams, setting explicit targets (e.g., reduce total delay cost by 10–15% within 12 months while maintaining current utilization levels).

These actions together aim to keep utilization high, but reduce the volatility and cost of delays, especially for the biggest carriers on the network.

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