



DELAY REPORT

WOJCIECH HRYCENKO 473173, CEZARY KUŽMOWICZ 473169, ALEKSANDRA SZPAKOWSKA 473205

WHY WE STARTED THIS PROJECT

- Flight delays and cancellations are one of the most painful parts of air travel – for passengers and for airlines.

OUR CHALLENGE:

*“Can we turn raw historical flight data into a **clear picture of where, when and why delays happen – and who is most at risk?**”*



We wanted an end-to-end analytics pipeline:

raw data



databricks



dashboard

FINDING THE RIGHT DATA & BRINGING IT TO LIFE

We chose the public “Flight Delay and Cancellation Dataset 2019–2023” from **kaggle** – **millions of flights across several years**, with information on:

- airlines, airports, routes
- scheduled vs actual times
- delays, cancellations and diversions

Instead of manual CSV uploads, we built a **small API ingestion layer** that:

- pulls the dataset automatically,
- streams it to Databricks,
- keeps the data ready for SQL analysis and future refreshes.



TURNING RAW FLIGHTS INTO MEANINGFUL METRICS

In **Databricks**, we cleaned and modeled the data with SQL:

- removed corrupted rows, fixed date/time formats, handled missing values.
- derived key flags and metrics:
 - `is_delayed`,
 - delay in minutes,
 - cancellation indicator,
- “**problem flight**” (delayed OR cancelled).

Then we aggregated those into **business-friendly** views:

- **per route** (origin-destination pair),
- **per airline**,
- **per origin airport**.

For each group we calculated:

- Delay Rate, Cancellation Rate, Problem Rate,
- and added a **Risk Category** (Low / Medium / High) based on thresholds.

BRINGING THE STORY TO THE BUSINESS - POWER BI REPORTS

Page 1 - “*Which airports are trouble spots?*”

Zooms in on origin airports and geography:

- bar charts of problem rate by airport,
- a table listing key details and risk for each origin,
- and a map covered with airport bubbles, where color/intensity reflects how problematic each location is.

Helps answer:

“Are there **specific airports or regions** where delays **systematically cluster**? ”



BRINGING THE STORY TO THE BUSINESS - POWER BI REPORTS

Page 2 – “*Which airlines keep their promises?*”

Focuses on airline performance:

- average delay in minutes,
- cancellation rate,
- delayed arrival rate,
- and total flights per airline.

Ranking charts quickly show who is reliable and who struggles, letting managers compare **punctuality vs volume**.



Answers questions like:

“*If we care about **on-time performance**, which carriers are our **best and worst partners**?*”



BRINGING THE STORY TO THE BUSINESS - POWER BI REPORTS

Page 3 – “*Where are the risky routes?*”

Shows the world through the lens of routes:

- big KPI tiles with average distance and overall delay rate,
- bar charts highlighting which cities cause the most delays,
- a detailed route table with risk category and delay %,
- and a world map so you can see the network and spot hot-spots.

Typical question it answers:

“*Which routes combine long distance with **high delay risk** and might need **schedule changes**?*”



HOW PEOPLE USE IT

Operations and Network Planning can:

- quickly locate high-risk routes, airlines and airports,
- simulate “what-if” scenarios using slicers and filters,
- prioritize where to negotiate slots, add buffer time or change partners.

Instead of digging through CSVs, stakeholders now open Power BI and in seconds see:

- “top 10 problematic routes today”,
- “airlines with the highest cancellation rate this quarter”,
- “airports where we consistently arrive late.”

MAIN INSIGHTS

AIRPORTS

- from major airports, Denver Airport (DEN) is the most problematic
- New York area has the biggest delays

AIRLINES

- low-costs (JetBlue, Frontier, ExpressJet) with highest cancellation and delay rate
- the bigger the airline, the fewer the problems

ROUTES

- short and long haul flights with lower delays than medium hauls

OUR TECHNICAL STACK IN ONE VIEW



SINGLE SOLUTION THAT LETS OPERATIONS AND NETWORK PLANNING:

Quickly identify worst routes, airlines and airports.

Monitor delay and cancellation trends over time.

Support data-driven decisions on schedule optimization and service quality.

OUR KEY VALUES





**THANK
YOU**