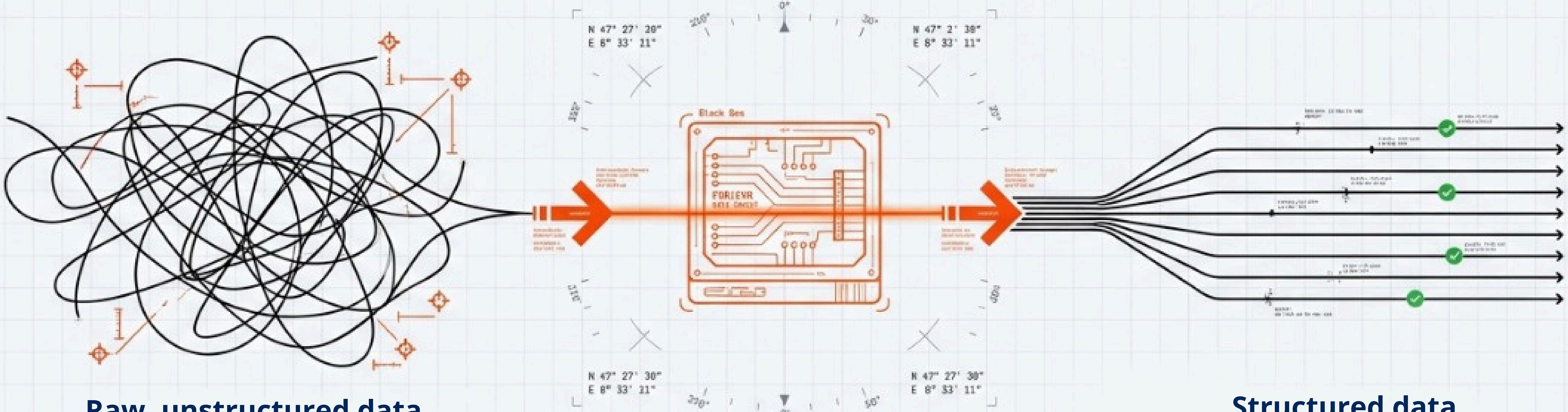
Project: Wings of data

Optimisation of air performance by business intelligence
a data-to-action approach to managing delays

Ayoub El Atri - Ala Harrabi - Nourine Brahem

This project leverages historical flight data from a large data network to transform it to operational planning



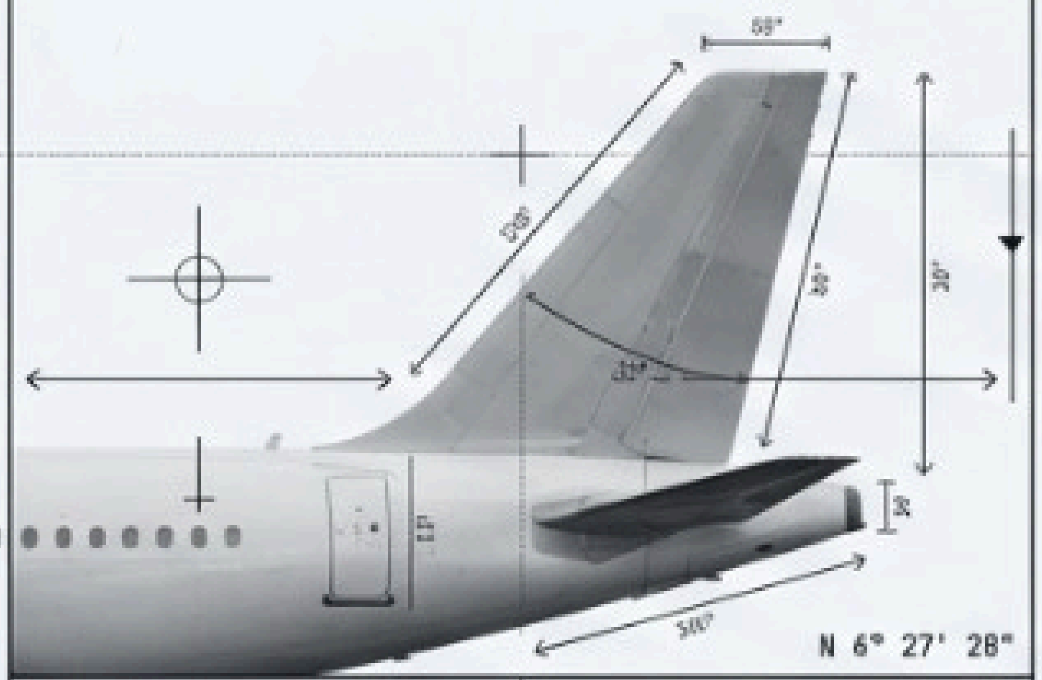
Raw, unstructured data

Structured data

Data scope: simulation of real operational data // multiple compaigns // hundreds of daily flights

Measurement indicators: volume and punctuality

KPI 1 : Total Flights



$$\text{TotalFlights} = |\{\text{Flight ID}\}|$$

KPI 2 : Flights per Airline

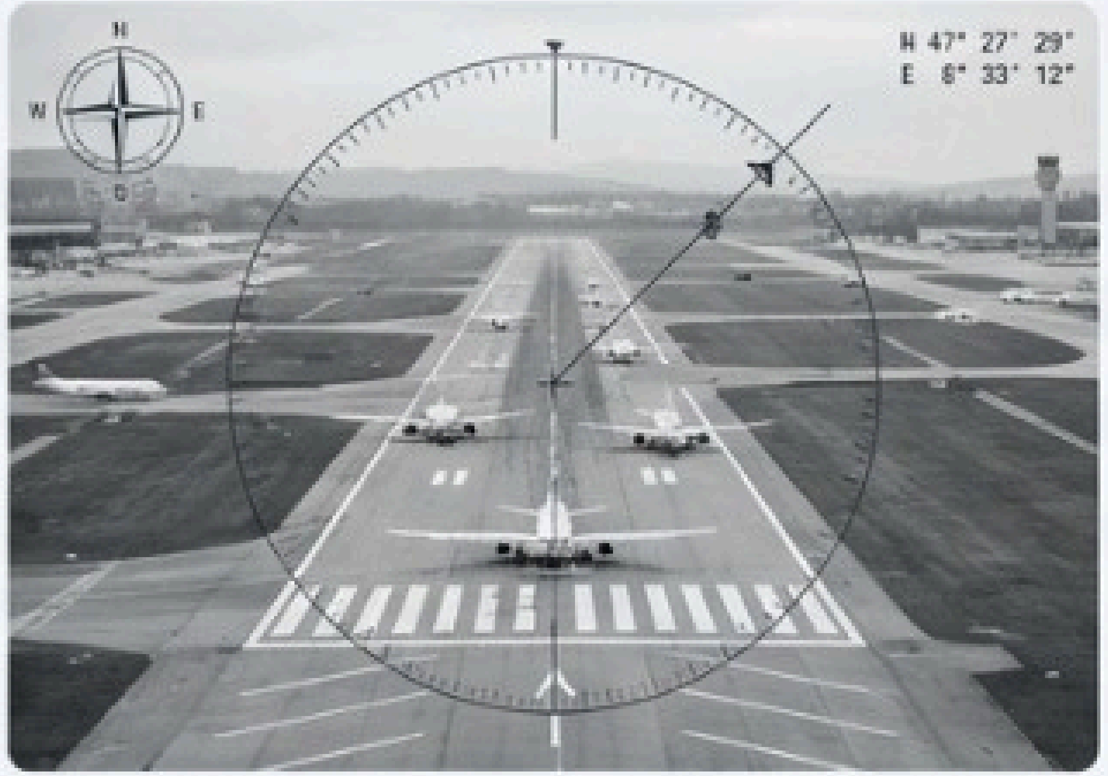
$$\text{FlightsPerAirline}(a) = |\{\text{Flight ID} \mid \text{Airline} = a\}|$$

KPI 3/4 : Average delay

$$\text{AvgDelay}(a) = \frac{1}{n_a}\sum_{a} \text{Delay}_i$$

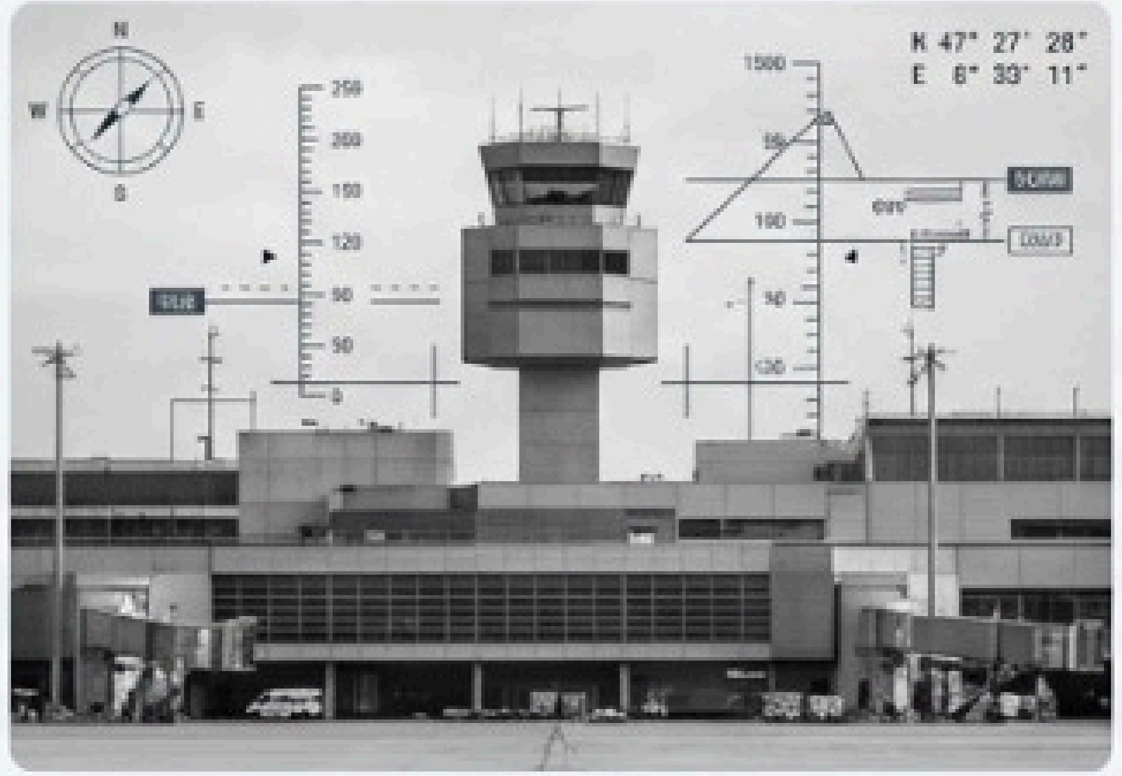
Measurement indicators: intensity and cycles

KPI 6 : Distribution of bands of delay



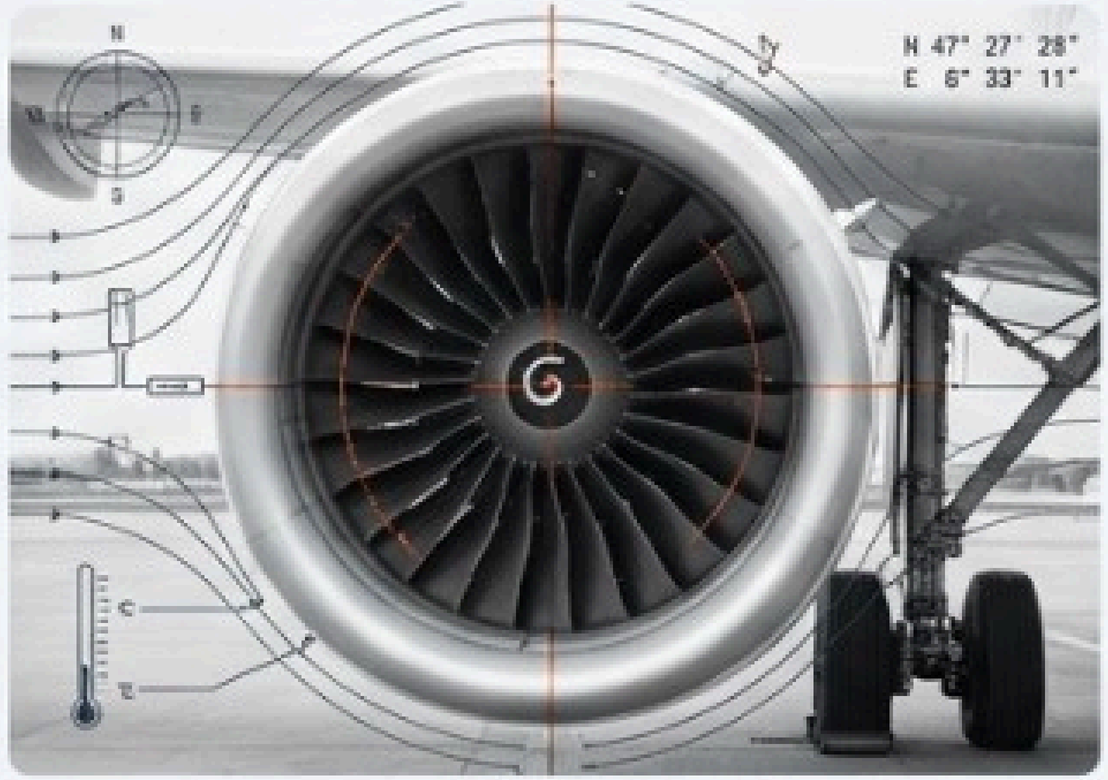
$$\%FlightsInBand(b) = \frac{|\{Flight\ ID \mid DelayBand = b\}|}{TotalFlights} \times 100$$

KPI 7 : Airport delay intensity



$$AvgDelayAtAirport(p) = \frac{1}{n_p} \times \sum DepartureDelay_i$$

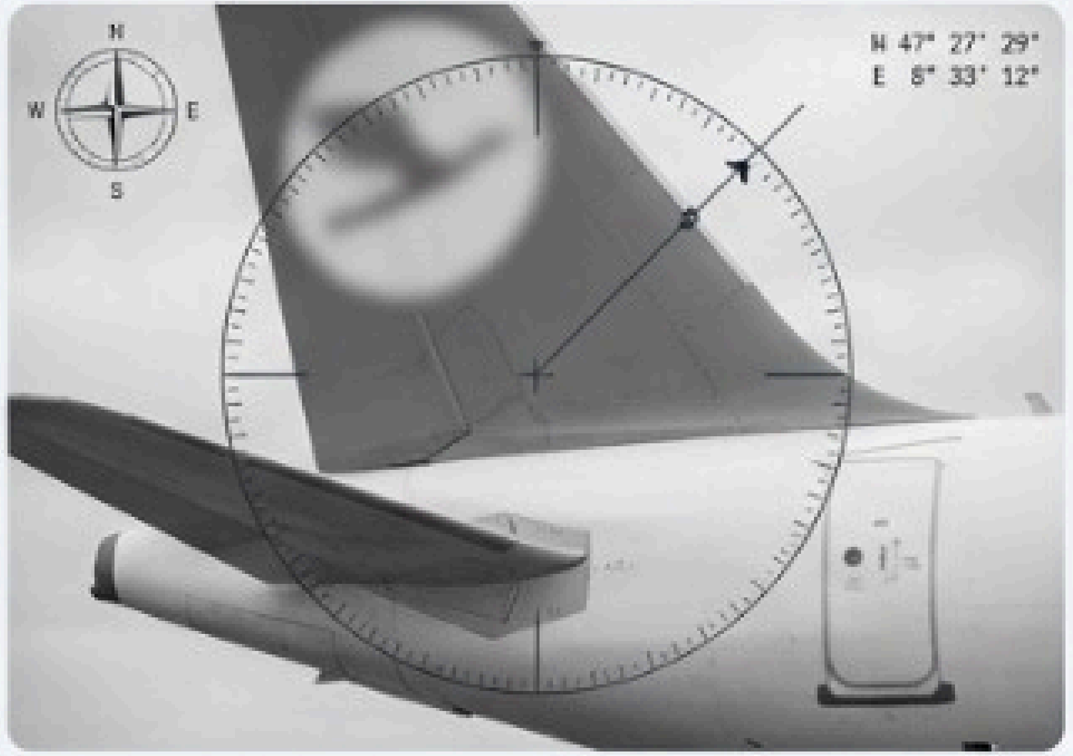
KPI 8 : Rush hour delay index



$$PeakDelayIndex = \frac{AvgDelay\_peak}{AvgDelay\_off\text{-}peak}$$

Measurement indicators: risk and structure

KPI 5 : Delay rate

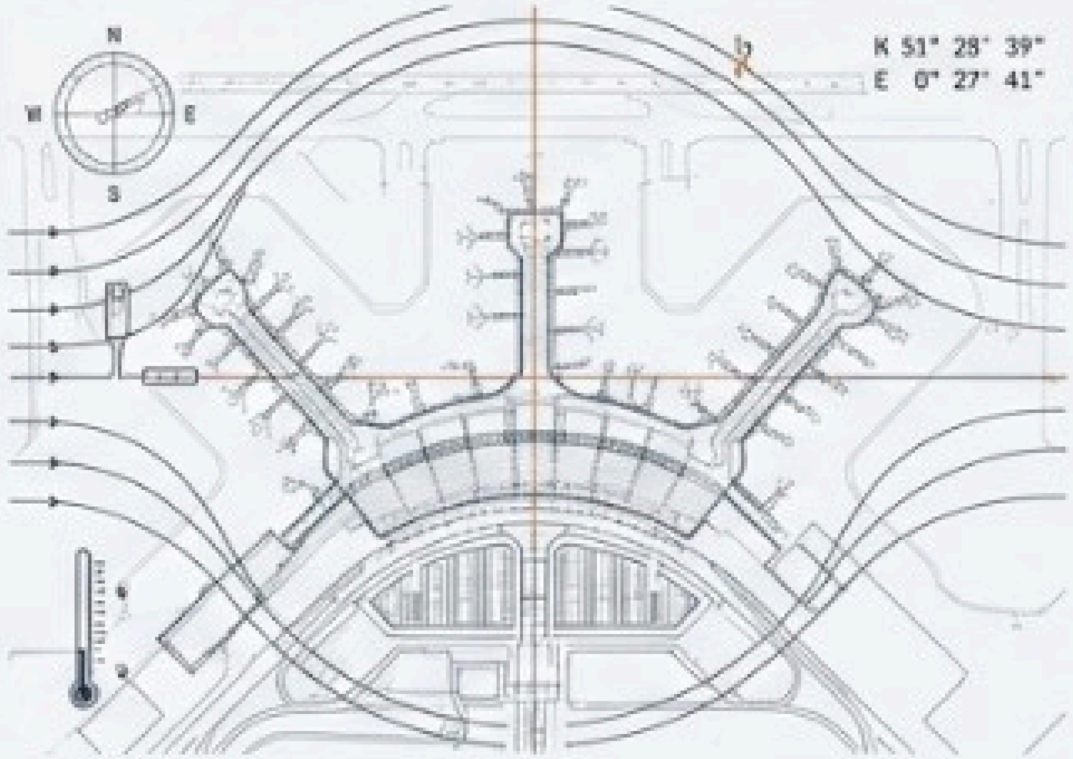


$$\text{DelayRate} = \frac{|\{\text{Flight ID} \mid \text{DepartureDelay} > 0\}|}{\text{TotalFlights}}$$

KPI 9 : Propagation rate

$$\text{PropagationRatio} = \frac{\text{AvgArrivalDelay}}{\text{AvgDepartureDelay}}$$

KPI 10 : Concentration rate



$$\text{CR} = \frac{\text{Flights\_Top5}}{\text{TotalFlights}}$$

From collection to visualization

1. Data collection: aggregation of wings of data dataset

2. ETL process: cleaning, standardizing and transformation

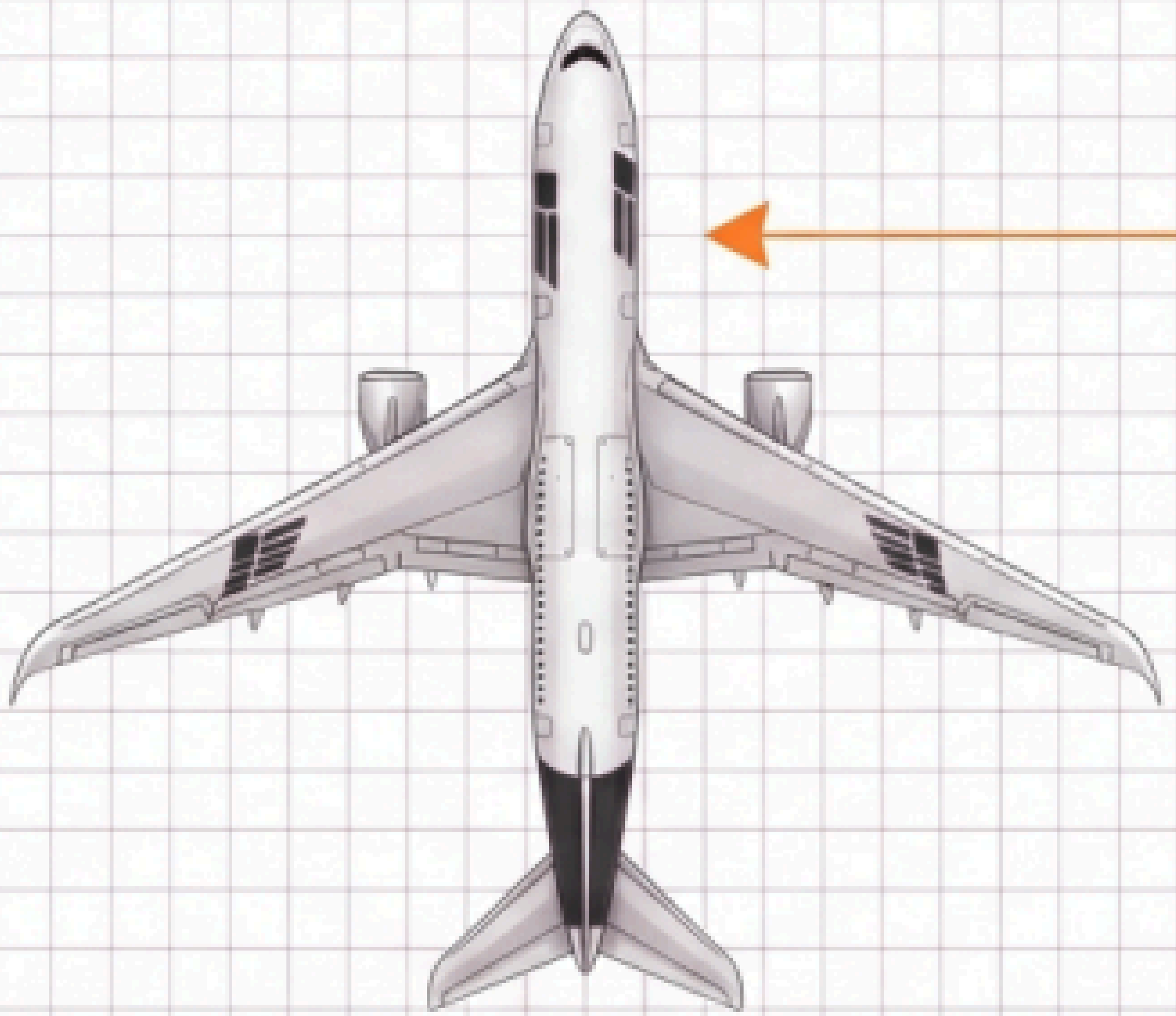
3. Modelling: Star schema architecture for performance

4. Visualization Dashboards power BI and KPIs

ETL PROCESS

TRANSFORM: The imported raw data is cleaned, standardized, validated, and converted to match the target system's structure, often involving complex business rules.

EXTRACT: The data is collected from various sources in its original form.



LOAD: The cleaned, transformed data is loaded into the destination, making it ready for analysis.

ETL EXECUTION: PYTHON IMPLEMENTATION

EXTRACT

```
#loading the data:
print("loading the data ...")
df = pd.read_csv("Flights_raw_data.csv")
print("raw data loaded !")
```

EXTRACT: Extract data about records from CSV using pandas

```
FlightDate,Airline,Origin,Dest,Cancelled,Diverted,CRSDepTime,DepTime,DepDelayMinutes,DepDelay,ArrTime,
2018-01-23,Endeavor Air Inc.,ABY,ATL,False,False,1202,1157.0,0.0,-5.0,1256.0,0.0,38.0,62.0,59.0,145.0,
2018-01-24,Endeavor Air Inc.,ABY,ATL,False,False,1202,1157.0,0.0,-5.0,1258.0,0.0,36.0,62.0,61.0,145.0,
2018-01-25,Endeavor Air Inc.,ABY,ATL,False,False,1202,1153.0,0.0,-9.0,1302.0,0.0,40.0,62.0,69.0,145.0,
2018-01-26,Endeavor Air Inc.,ABY,ATL,False,False,1202,1150.0,0.0,-12.0,1253.0,0.0,35.0,62.0,63.0,145.0
2018-01-27,Endeavor Air Inc.,ABY,ATL,False,False,1400,1355.0,0.0,-5.0,1459.0,0.0,36.0,60.0,64.0,145.0,
2018-01-28,Endeavor Air Inc.,ABY,ATL,False,False,1202,1202.0,0.0,0.0,1326.0,20.0,37.0,62.0,84.0,145.0,
```

TRANSFORM: Cleaned flights dataset with no duplicates, missing values and handeled outliers ready for analysis

LOAD

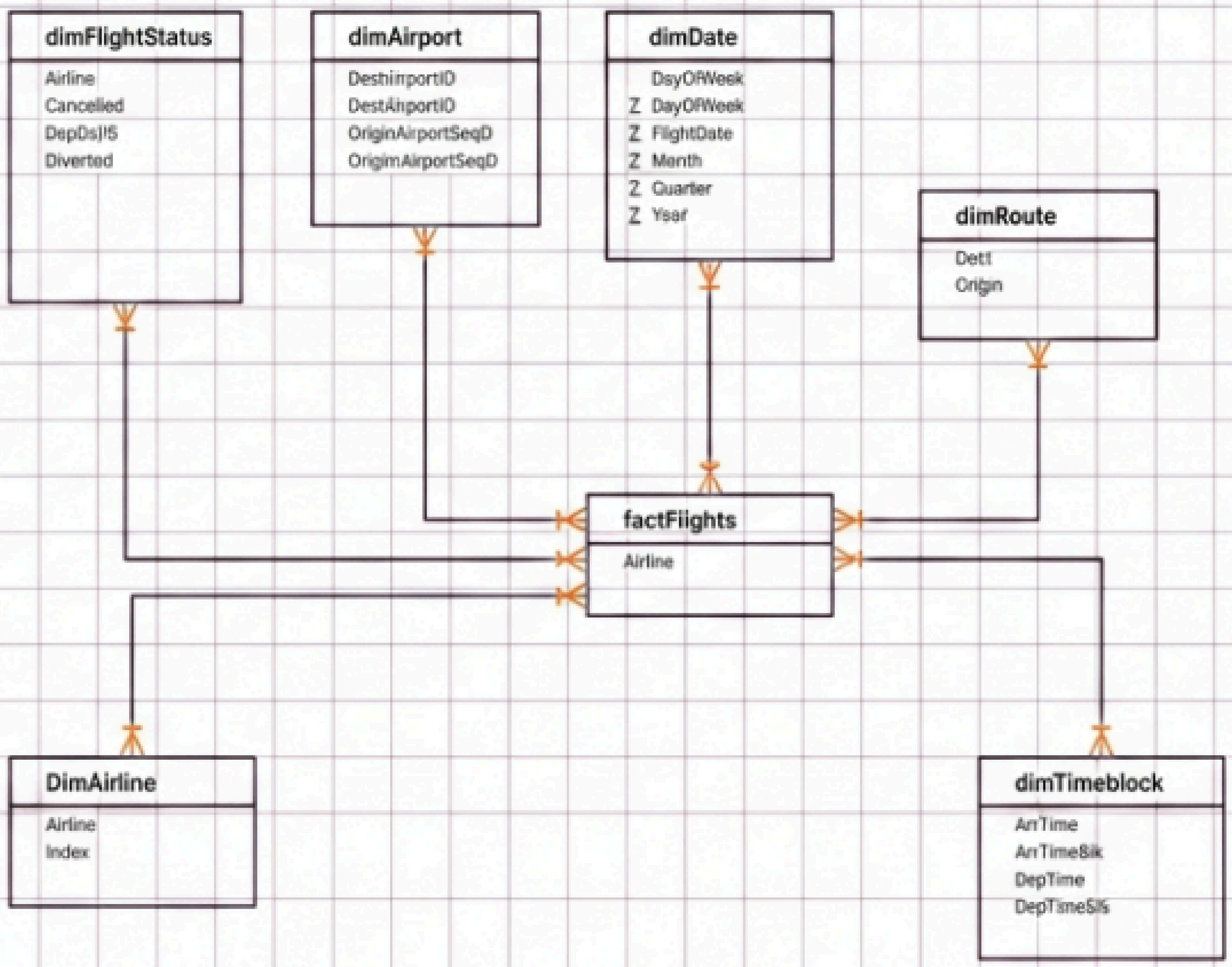
```
#exporting the clean data in csv format
print("exporting the clean data...")
df.to_csv("flights_data_cleaned.csv", index=False)
```

LOAD: loaded cleaned flights dataset as a csv file

DATA STORAGE

-After the ETL process, the cleaned and transformed data was stored in a CSV file.
-CSV storage allows easy data sharing and offline analysis.

DATA MODELING



6M
Total Flights

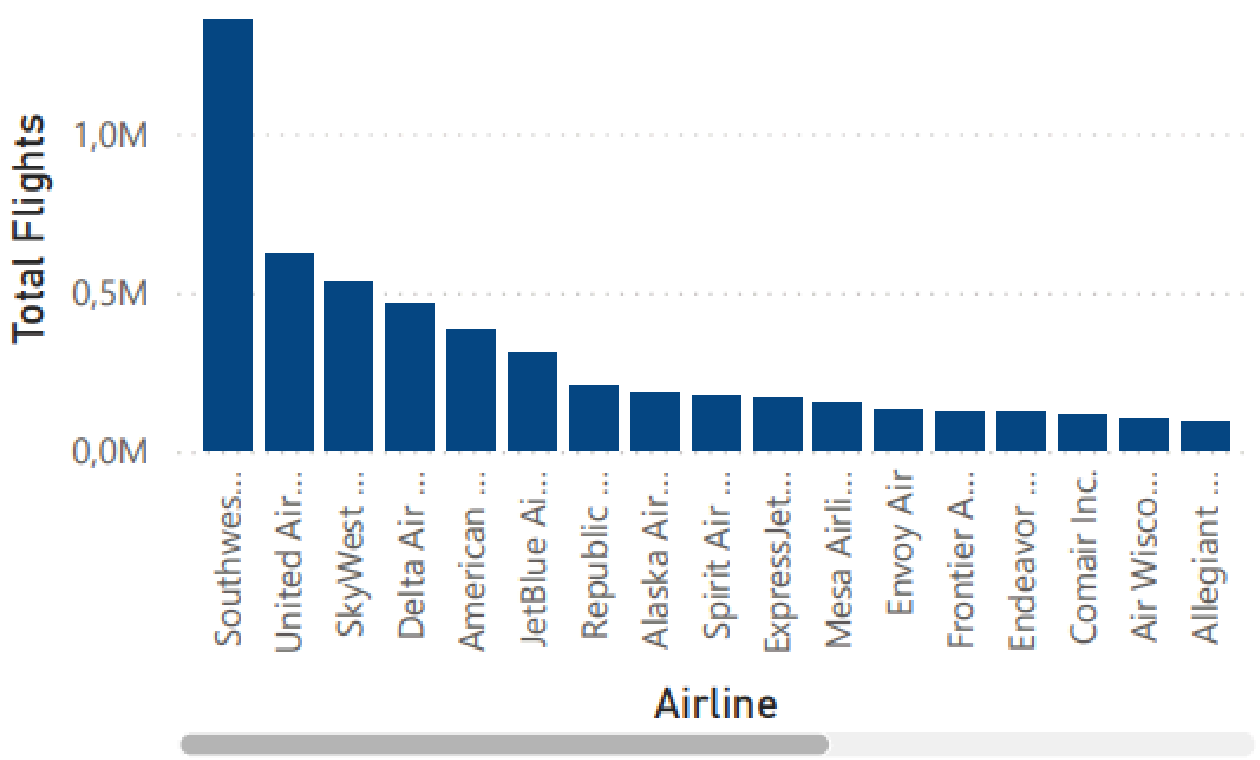
-0,81
avg arrival delay

2,24
avg departure delay

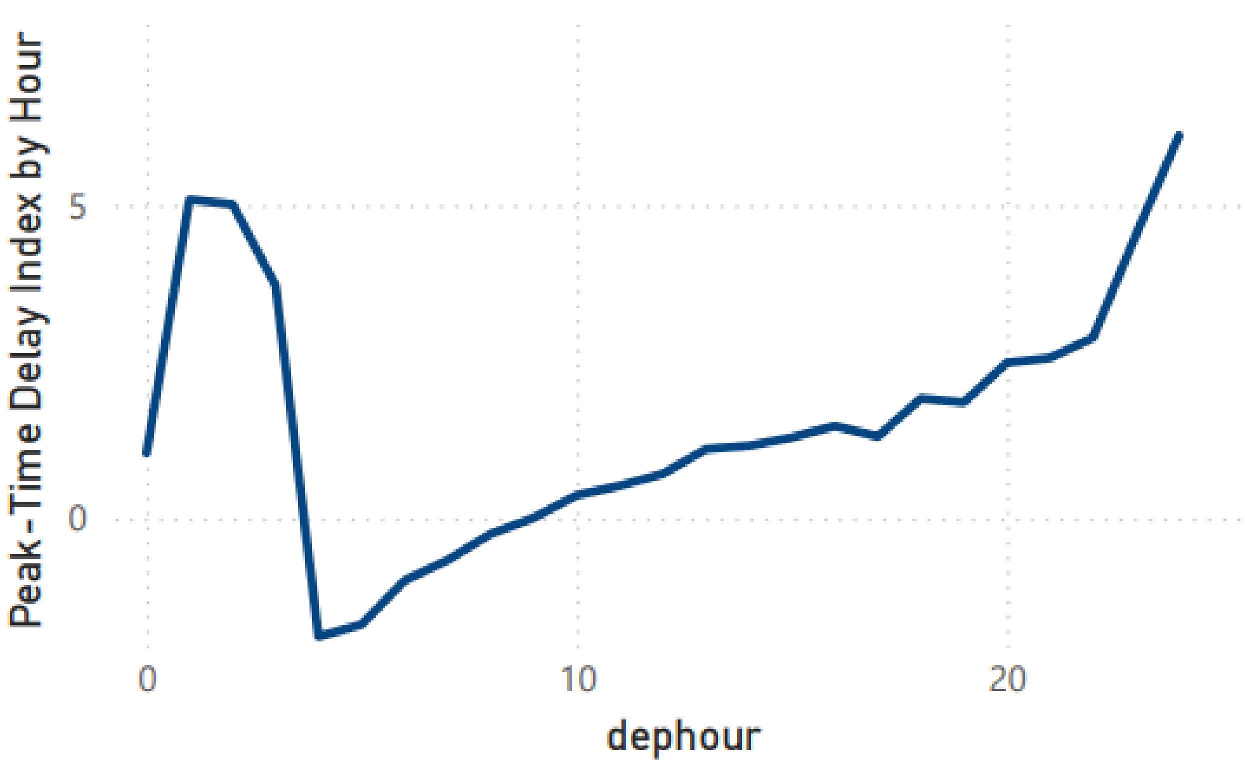
-0,36
delay propagation ratio

0,59
top 5 airlines concentration ratio

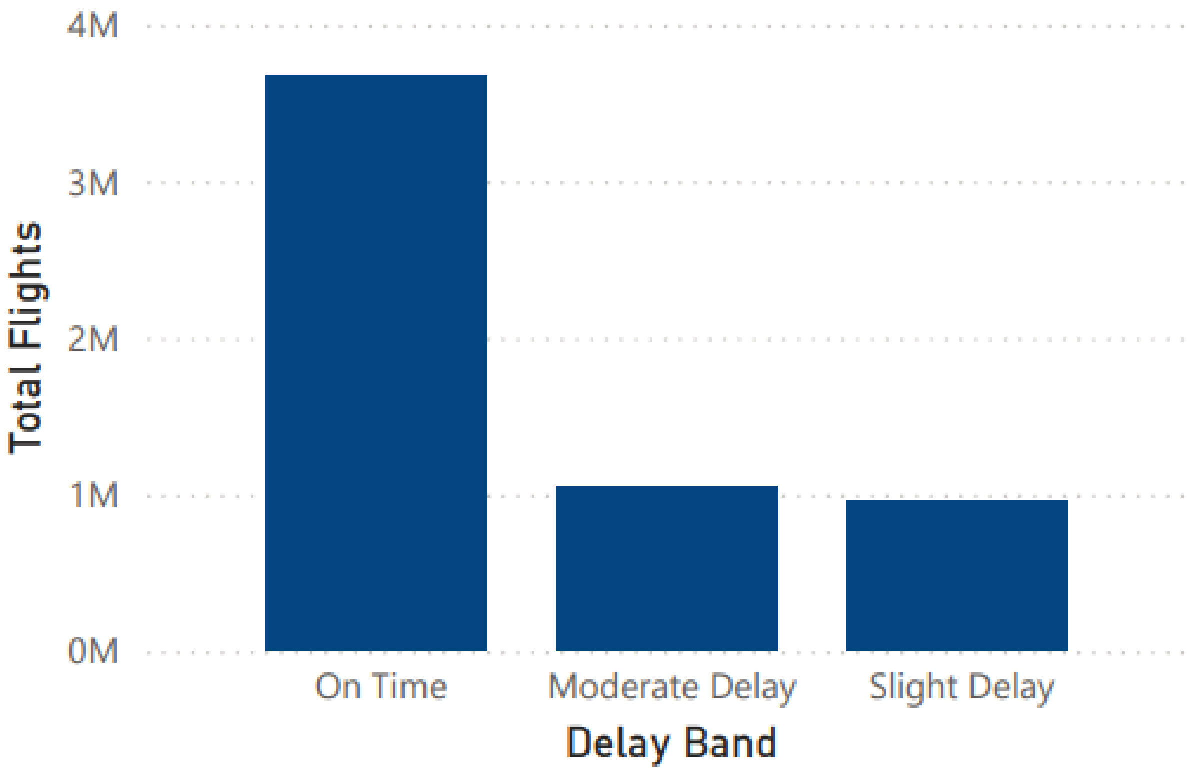
Total Flights by Airline



Peak-Time Delay Index by Hour by dephour



Total Flights by Delay Band



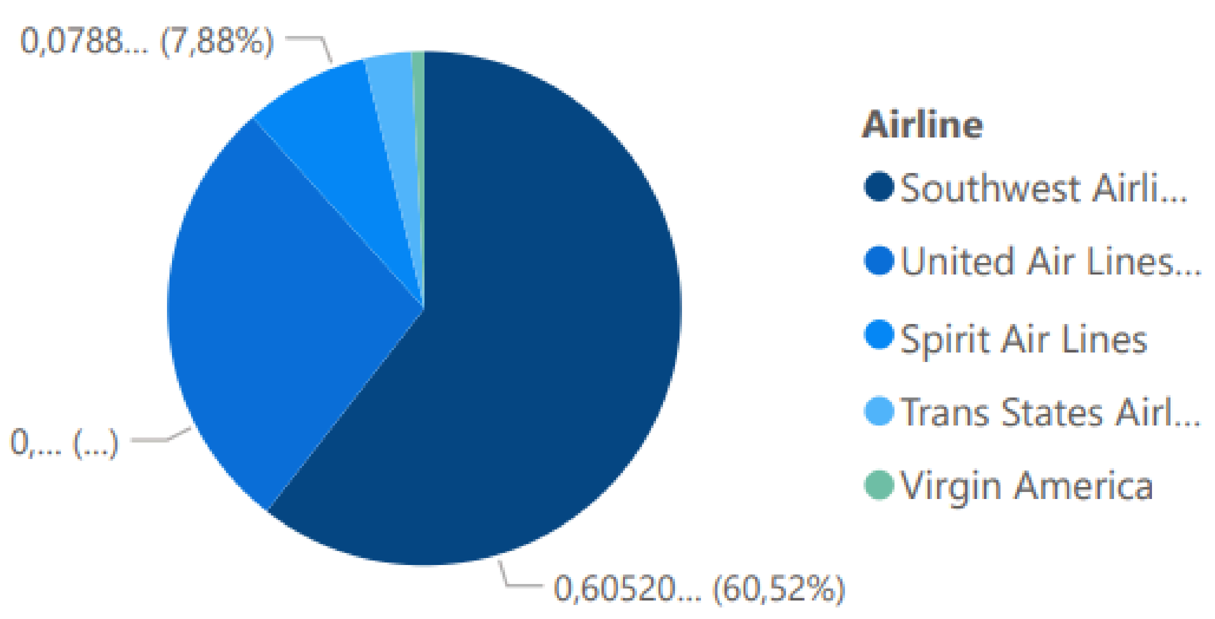
FlightDate
- lundi 1 janvier 2018
- mardi 2 janvier 2018
- mercredi 3 janvier 2018

Dest
- (Blank)
- ABY
- ANC

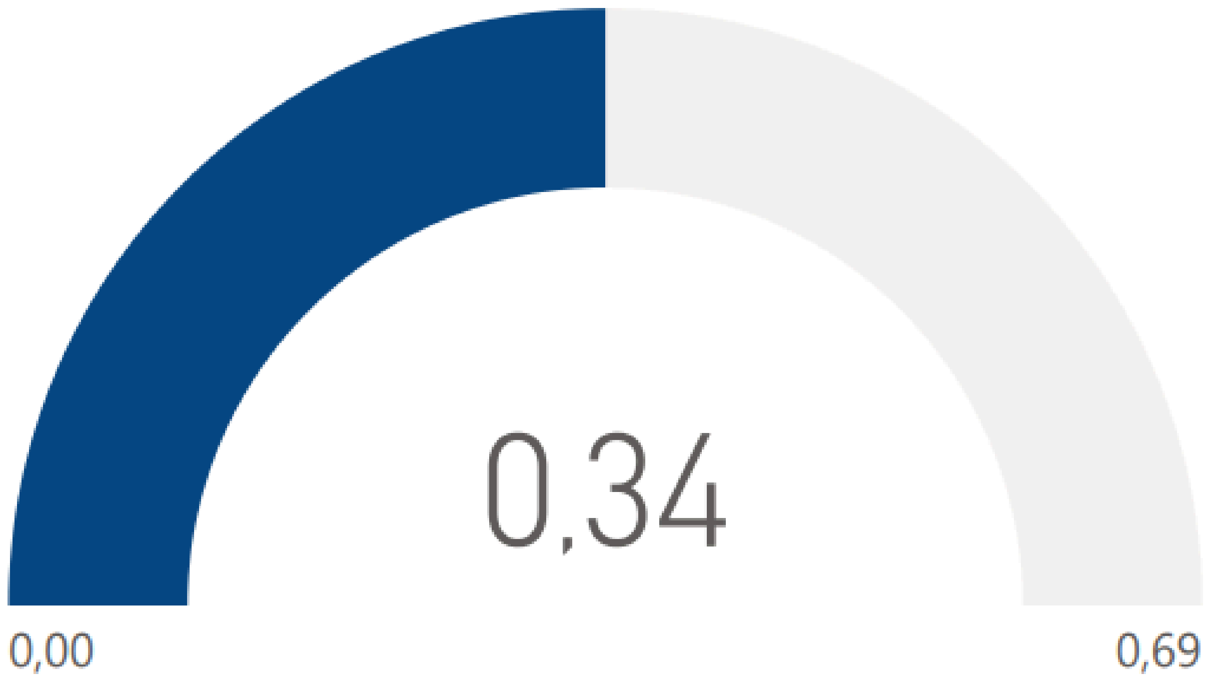
Airline
- Air Wisconsin Airlines...
- Alaska Airlines Inc.
- Allegiant Air
- American Airlines Inc.

Delay Band
- Moderate Delay
- On Time
- Slight Delay

Airline Market Share by Airline



Delay rate



Dimensional Logic and Dictionary

The model enriches raw flight data with dimensions that allow for slicing data across multiple axes, answering the Who, Where, and When of operational performance.

Dimension	Analytical Capability
Airline Dimension	Allows aggregation by Region and Fleet Size. Enables comparative analysis of market share and normalization of metrics between large and small carriers.
Airport Dimension	Connects to both Origin and Destination fields in the fact table. Critical for identifying infrastructure bottlenecks.
Date Dimension	Contains standard hierarchies (Year > Quarter > Month > Day) to track seasonal variations and peak operational periods.
Delay Band Dimension	Categorizes numerical delay minutes into severity groups (e.g., On Time, Slight, Severe) to reveal patterns: are delays frequent but short, or rare but catastrophic?

Establishing Baselines: Volume and Scale Metrics

Total Flights	Flights per Airline	Flights by Time Period
Business Context Measures overall operational scale. Serves as the denominator for rate-based metrics.	**Business Context** Enables comparative analysis of market share and resource utilization.	**Business Context** Identifies peak operational periods and seasonal growth trends.
Technical Logic `DISTINCTCOUNT(Flights[FlightID])`	**Technical Logic** `CALCULATE( [Total Flights], ... )`	**Technical Logic** `Time Intelligence Hierarchies`
Insight Ensures each flight is counted exactly once, handling duplicate records.	**Insight** Uses CALCULATE with the Airline dimension in the filter context.	**Insight** Leverages Date dimension hierarchies for time series visualization.

Delay Analytics: Departure vs. Arrival

Departure Delay

AVERAGE(DepartureDelay)

Minutes delayed at the gate (includes negative for early).

Measures Schedule Adherence. Reflects gate management and ground ops efficiency.

Arrival Delay

AVERAGE(ArrivalDelay)

Minutes delayed at destination.

Measures Passenger Perspective. Typically higher than departure delay due to en-route propagation.

Key Metric: Delay Rate

The proportion of flights with ANY delay (>0 minutes). This differentiates between a system with many small delays vs. one with huge averages caused by outliers.

Advanced Operational Intelligence

Peak-Time Delay Index	Delay Propagation Ratio	Airport Delay Intensity
Concept Compares performance during peak hours (7-9 AM, 5-7 PM) vs. off-peak.	**Concept** Quantifies the "Ripple Effect". How much does an initial delay grow by the time the plane lands?	**Concept** Identifies specific bottleneck airports.
Mathematical formula $\frac{\text{Avg Delay Peak}}{\text{Avg Delay Off–Peak}}$	**Formula** $\text{DIVIDE}\left(\frac{\text{ArrivalDelay}}{\text{DepartureDelay}}\right)$	**Formula** $\rightarrow \text{AVERAGE}$
Insight Values > 1 indicate specific congestion impacts on reliability.	**Insight** High ratios indicate poor recovery during the flight; low ratios suggest the pilot made up time in the air.	**Insight** Used to prioritize infrastructure investments based on delay attribution.

The DAX Framework: Core Calculation Patterns

Conditional Aggregation (The Delay Rate)

```
Delay Rate =
DIVIDE(
   CALCULATE( COUNT(Flights
      [FlightID]), Flights
      [DepartureDelay] > 0),
   [Total Flights]
)
```

Demonstrates **CALCULATE** to modify filter context and **DIVIDE** for safe division (handling divide-by-zero).

Time Intelligence

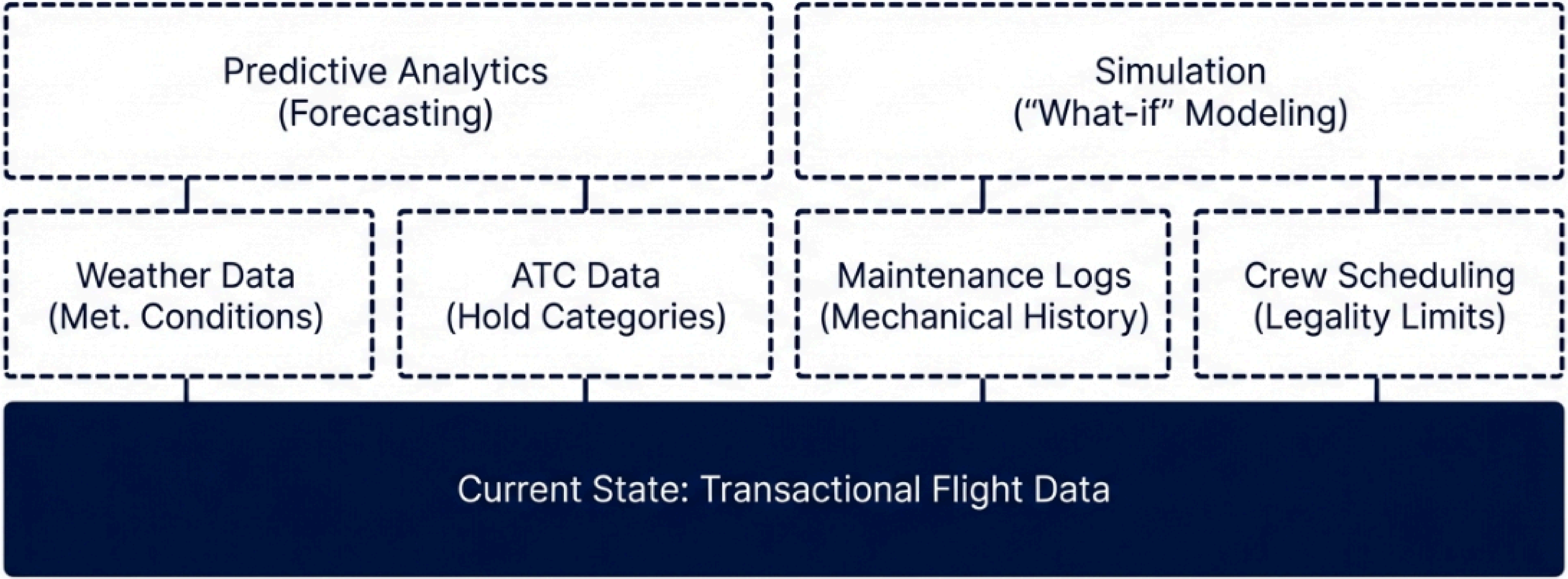
```
Prev Month Flights =
CALCULATE(
    [Total Flights],
    PREVIOUSMONTH(Date[Da
te])
)
```

Enables period-over-period trend analysis using the Date table.

Context Propagation Strategy

The model relies on the distinction between Row Context (evaluating specific airlines in a table iteration) and Filter Context (slicers for Date or Airport) to ensure metrics update dynamically across the report.

Future Extensibility and Integration



Strategic Recommendations: Operational Optimization

Peak Management

inding: High concentration f delays **during** peak hours.

ction :

Redefine planning and increase rotation buffers.

Temporal Planning

Finding: Seasonal delay patterns identified.

Action :

- Adjust staffing and maintenance based on risk history.

Predictive Maintenance

Finding: Inefficiencies relate to equipment availability.

Action :

- Schedule maintenance outside peak periods based on historical data.

Strategic Recommendations: Real-Time Intelligence

Machine Learning Prediction

- Deploy ML models to alert teams of probable disruptions.

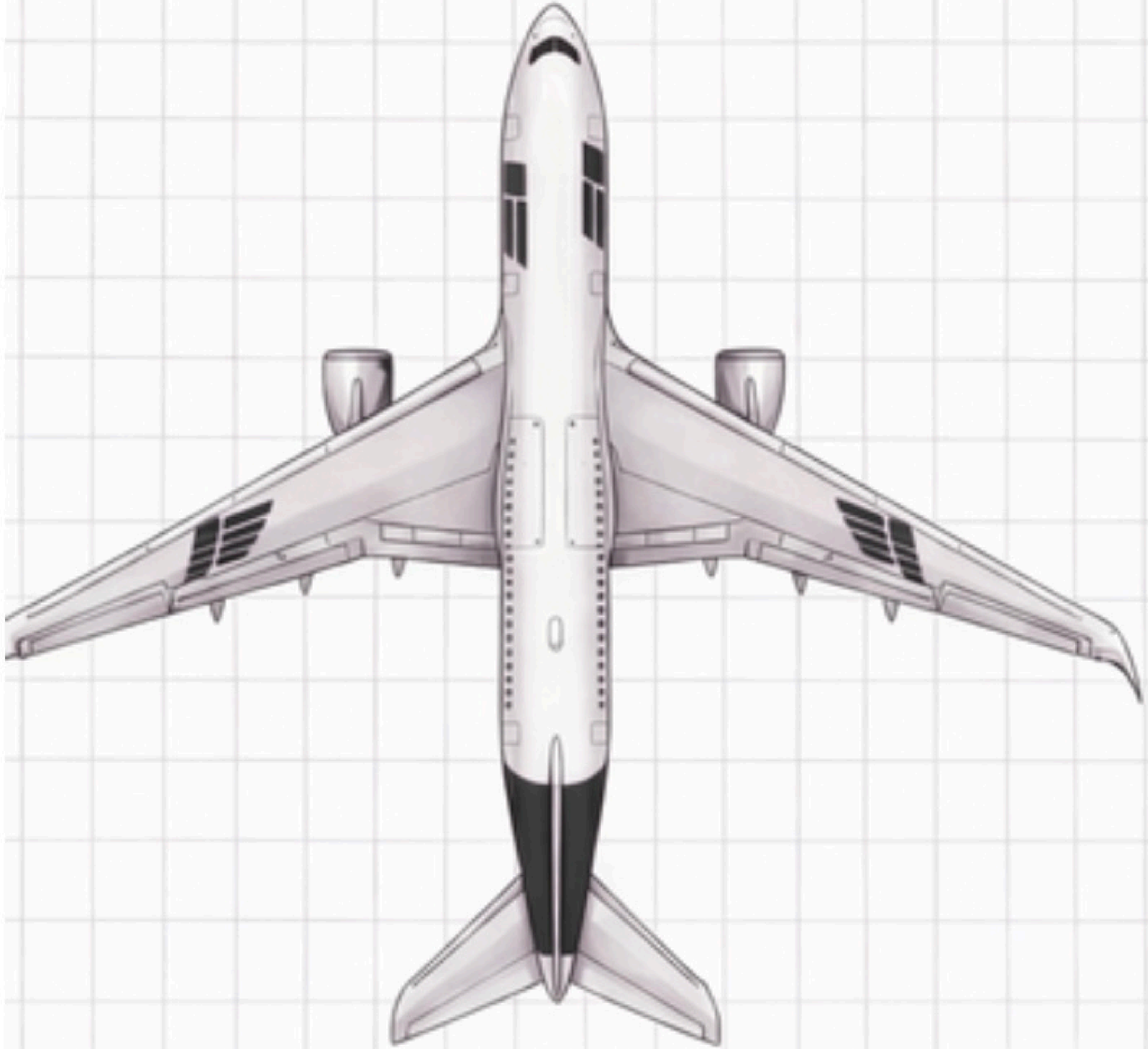
Task-Forces Collaboratives

- Create joint airport-airline teams for real-time resource allocation.

Recovery Mechanisms

- Enhance crew coordination and in-flight recovery strategies to limit delay propagation.

Strategic Recommendations: Network Structure



Hub Decentralization

Finding : Strong reliance and concentration on major airports.
Action : Diversify routing strategies and assess usage of secondary airports.
Impact : Reduced systemic risk.

Company Benchmarking

Finding: Significant variability in performance among airlines.
Action : Implement internal benchmarking programs to replicate best practices.
Impact : Improved overall punctuality.