P2-AIR: Methodology & Quality Assurance

1. Data Sources

IMPORTANT: Flight data is **SYNTHETIC** (generated in BTS TranStats format) for portfolio demonstration purposes. Weather data is **REAL** from NOAA CDO API.

Source	Type	Data Status
BTS TranStats Format	Flight Performance	**SYNTHETIC** (demo)
NOAA CDO API v2	Weather	**REAL** (live API)

Weather Data: GHCND (Global Historical Climatology Network Daily) dataset with PRCP, TMAX, TMIN, AWND, fog, thunder indicators.

2. Data Model

Table	Description	Key Columns
flight_fact	Flight records	flight_date, carrier, origin, dest, delays
weather_fact	Daily weather by airport	airport, date, precip, temp, wind
airport_dim	Airport reference	code, name, city, state, lat/lon
carrier_dim	Carrier reference	carrier_code, carrier_name
flight_weather_fact	Joined fact	All flight + origin/dest weather

3. Weather Attribution Logic

```
Severe Weather Flag:
severe_weather = (precipitation > 25mm) OR (wind > 25mph) OR (fog=1) OR (thunder=1)

Weather Impact:
delta = avg_delay(severe) - avg_delay(normal)
```

⚠ **Attribution Note:** Weather correlation does not imply causation. Delays may result from cascading effects, crew availability, or other operational factors coinciding with weather events.

4. Quality Gates

Gate	Threshold	Logic
Timezone	≥95%	Valid date/time parsing
Join Coverage	≥80%	Flights matched with weather data
Outliers	≥95%	No impossible delays (>24h or <-60min)
Delay Codes	≥75%	Delayed flights have reason codes

Completeness	≥95%	Required fields populated

5. KPI Definitions

Metric	Formula
On-Time Rate	% flights with arr_delay ≤ 15 min
Avg Delay	MEAN(dep_delay) for non-cancelled flights
P90 Delay	90th percentile of dep_delay
Cancel Rate	% flights with cancelled=1

6. Run Metrics

Metric	Value
Duration	36.7 seconds
Flight Records	99,975
Weather Records	1,536
Joined Records	99,975
NOAA API Calls	10
Quality Score	97.5%