# Project Plan Flight Reliability & Performance Dashboard

**🔹 Phase 1: Data Engineering Plan**

**1. Data Collection**

* **API Source**: Your flight data API.
* **Scope**: Filter for flights where airline.name or airline.iata matches the 5 Pakistani airlines.
* **Frequency**: Daily ingestion (scheduled with Airflow or Lambda).

**2. Raw Data Storage**

* **Amazon S3**:
  + Bucket: flight-data-project
  + Folder structure:
  + s3://flight-data-project/raw/YYYY/MM/DD/
  + Store raw JSON responses for reproducibility.

**3. Data Warehouse Design (Snowflake)**

**1. dim\_airline**

From "airline" object.

* **airline\_id** (PK, surrogate key)
* name
* iata\_code
* icao\_code

**2. dim\_airport**

From "departure" and "arrival" objects.

* **airport\_id** (PK, surrogate key)
* airport\_name
* iata\_code
* icao\_code
* timezone
* terminal
* gate

*(Note: “baggage” is only in arrival, you can store it too if useful.)*

**3. dim\_flight**

From "flight" object.

* **flight\_id** (PK, surrogate key)
* flight\_number
* iata\_code
* icao\_code
* codeshared\_flag (Y/N, derived from codeshared)

**4. fact\_flight\_performance**

Main table that ties it all together.

* **fact\_id** (PK)
* flight\_date
* flight\_status (scheduled, cancelled, active, landed, etc.)

🔹 **Departure fields**

* departure\_airport\_id (FK → dim\_airport)
* scheduled\_departure\_time
* estimated\_departure\_time
* actual\_departure\_time
* departure\_delay\_minutes

🔹 **Arrival fields**

* arrival\_airport\_id (FK → dim\_airport)
* scheduled\_arrival\_time
* estimated\_arrival\_time
* actual\_arrival\_time
* arrival\_delay\_minutes
* baggage (if available)

🔹 **Airline link**

* airline\_id (FK → dim\_airline)

🔹 **Flight link**

* flight\_id (FK → dim\_flight)

**🔹 Phase 2: Data Analysis Plan**

The dashboard’s goal is **Flight Reliability & Performance**. Focus on **delays, cancellations, and route efficiency**.

**1. KPIs (Key Metrics)**

* **On-Time Performance (OTP)** = % of flights with <15 min delay.
* **Average Departure Delay** (per airline, per route).
* **Average Arrival Delay**.
* **Cancellation Rate**.
* **Top Routes by Delay**.
* **Airline Comparison** (how the 5 airlines stack up).
* **Daily / Weekly / Monthly trends**.

**2. Example Visuals**

* **Airline Reliability Scorecard** (PIA vs Airblue vs SereneAir, etc.).
* **Heatmap of Delays by Airport** (departure vs arrival).
* **Trend Line**: delays across months.
* **Bar Chart**: top 10 delayed routes.
* **Pie Chart**: cancellation reasons (if available).

**3. Example Analysis Questions**

* Which airline is **most reliable** in terms of delays < 15 minutes?
* Do **international routes** (e.g., DXB → PK) perform worse than domestic?
* Which **Pakistani airport** has the highest average delay?
* Are delays **seasonal** (e.g., more in summer/winter)?

**🔹 Suggested Workflow**

**Data Engineering Flow**  
API → Lambda/Airflow → S3 (Raw JSON) → Snowpipe (load into Snowflake) → Transform (dbt/Snowflake SQL) → Dim/Fact Tables

**Data Analysis Flow**  
Snowflake → Power BI / Tableau / Looker → Flight Reliability Dashboard