

Metadata Documentation

New DataFrames Created:

Data Frame Name	Description
df_flights_active	Cleaned version of original flights dataset.
df_tickets	Cleaned version of original tickets dataset.
df_airports	Cleaned version of original airports dataset.
valid_airports	A set of valid US airports IATA codes.
direction_counts	Direction wise flight counts per route.
round_trip_counts	Contains the actual number of round trips per route.
route_flight_agg	Route level aggregated metrics.
ticket_fare_agg	Aggregated ticket level metrics per route.
df_routes	Final combined route level dataset with cost, revenue, and profit metrics.

Metadata For New Columns Created

For DataFrame: `df_flights_active`

Field Name	Description
ROUTE_ID	Unique identifier for a flight route.
ORIGIN_SIZE	Size category (e.g., medium/large) of the origin airport.
DEST_SIZE	Size category (e.g., medium/large) of the destination airport.
PASSENGERS	Estimated number of passengers per flight.
BAGGAGE_REVENUE	Revenue from checked baggage per flight.
FUEL_MAINT_COST	Estimated fuel and maintenance cost per flight.
DEPREC_INS_COST	Depreciation and insurance cost per flight.
AIRPORT_FEE	Fee based on destination airport size.
DEP_DELAY_COST	Cost due to departure delays (post 15 mins grace).
ARR_DELAY_COST	Cost due to arrival delays (post 15 mins grace).
TOTAL_DELAY_COST	Combined cost from both arrival and departure delays.
DIRECTION_KEY	Directional route ID in the form ORIGIN–DESTINATION

For DataFrame: `df_tickets`

Field Name	Description
ROUTE_ID	Unique identifier for a flight route.

For DataFrame: route_flight_agg

This dataframe was created to show the difference between the two approaches to calculate the total roundtrips per direction

Field Name	Description
NUM_FLIGHTS	Total number of flights for the route.
ROUND_TRIPS	Minimum count of round trips calculated based on direction symmetry.
AVG_PAX_PER_FLIGHT	Average number of passengers per flight for each route.
OLD_ROUND_TRIPS	Approximate round trips computed as half of total flights.

For DataFrame: ticket_fare_agg

Field Name	Description
AVG_FARE	Weighted average fare per passenger for the route.
TOTAL_SAMPLE_PASSENGERS	Total passengers represented in the ticket sample for the route.

For DataFrame: df_routes

Field Name	Description
TICKET_REVENUE	Total estimated revenue from ticket sales per route.
TOTAL_REVENUE	Total revenue including ticket and baggage per route.
PROFIT	Profit = Total Revenue - Total Cost.
PROFIT_PER_ROUND_TRIP	Average profit per round trip on that route.