Transportation Project: Predicting and Analyzing Flight Delays and Cancellations

This project investigates the complex factors that lead to **flight delays and cancellations** in the airline industry, using machine learning and data analytics. It aims to uncover patterns related to weather, airline competition, flight routing, and operational efficiency to help improve scheduling, customer experience, and airline performance.

Dataset Creator Question: Which airlines or routes experience the most frequent delays or cancellations?

Datasets:

1. **Flight Delays and Cancellations (1987–2023)**
   A massive dataset of 29 million U.S. domestic flight records with delay and cancellation reasons (weather, security, carrier, etc.), airport locations, and flight status. Supports predictive modeling for flight outcomes.

   https://www.transtats.bts.gov/DL_SelectFields.aspx?gnoyr_VQ=FGJ&QO_fu146_anzr=b0-gvzr

2. **Airline Delay Data with Weather Features**
   A dataset of 1 million records including detailed environmental features (temperature, wind, rain, snow) tied to U.S. domestic flight delays (2004–2017). Enables deeper prediction of weather-related disruptions.
   https://data.mendeley.com/datasets/j3z5bm7496/1

3. **Comprehensive Flight Data from Priceline**
   Includes 2,000+ flight records with stoppage points, ticket pricing, layover times, and airline routes. Useful for linking flight itinerary complexity and cost to delay probabilities.
   https://www.kaggle.com/datasets/joyshil0599/comprehensive-flight-data-from-priceline/data