# Flight Delay Prediction at Hartsfield-Jackson Atlanta International Airport

### 1. Problem Description

This project aims to develop a machine learning system for predicting flight delays at Hartsfield-Jackson Atlanta International Airport, one of the busiest airport by passenger traffic in the world and has one of the largest airline hub.

The predictive model will be trained on historical data retrieved through USA bureau of transportation. The real-time predictions will utilize live flight schedules fetched from FlightAPI to provide users with up-to-the-minute delay estimates.

To ensure model performance reflects changing trends, the pipeline will dynamically update using daily real-time flight data.

### 2. Tools

- HuggingFace: real-time inference and GUI server hosting.
- Hopsworks: feature store and model registry.
- Gradio: GUI creation.
- Modal/GitHub Actions: Weekly data fetch and model retraining scripts.

### 3. Data

i. Historical Data:

- Bureau of Transportation Statistics (BTS): Historical flight delay causes will be sourced from the BTS On-Time Delay Causes Database. This dataset provides detailed insights into delay causes across different airports and carriers.

ii. Real-Time Data:

- FlightAPI: Real-time flight schedules, including carrier, origin and destination, departure times, and cancellations, will be integrated through the FlightAPI service.

iii. Weather Data:

- Open-Meteo: Real-time and forecast weather data will be used to factor in weather-related disruptions, such as storms or heavy winds, into the predictions.

### 4. Methodology

Combining flight and weather features to enhance model predictive power. For example, features like the likelihood of weather-related delays based on past trends at specific airports will be included.

Furthermore, the model will undergo hyperparameter optimization to ensure maximum performance.