Airline Delay Analysis and Performance Benchmarking

Summary

Flight delays have significant consequences for travelers, airlines, and airport operations. This project aims to analyze large-scale airline on-time performance data to identify delay hotspots and trends and benchmark airline-specific performance using big data technologies such as Apache Spark and Hadoop. By integrating external dataset, the analysis will provide a deeper understanding of the factors influencing delays due to weather conditions. The findings will be visualized using interactive dashboards to assist airlines and passengers in making informed travel decisions.

Problem Statements

1. **Identifying Delay Hotspots and Trends**
    - Analyze historical flight data to determine which airports and routes experience the highest delays.
    - Examine the impact of external factors, such as weather conditions on flight disruptions.
2. **Airline-Specific Performance Comparison**
    - Benchmark airline on-time performance across different carriers and regions.
    - Identify airlines with the highest and lowest delay rates.

Data Sources

- **On-Time Reporting Carrier Performance Dataset** – U.S. Department of Transportation (https://www.transtats.bts.gov/DL_SelectFields.aspx?gnoyr_VQ=FGJ&QO_fu146_anzr=b0-gvzr)
- **Weather Data** – https://open-meteo.com/

Team Members: (Group 4)

Aishanee Sinha

Debika Choudhary

Jayam Shah

Ketki Maddiwar

Nandhakumar Apparsamy

Sheethal Neelarapu