Airtime Analytics: Flight Delay Analysis and Prediction

Xinran Shen, Carina Zhang, Xingchen Li, Zeqi Li, Xinzhi Zhang

Project Description

This project aims to analyze and predict flight delays using historical data on flight performance and weather conditions. By integrating these datasets, we seek to uncover patterns and build a predictive model to estimate future delays, enhancing travel planning and management.

Dataset Overview:

https://www.kaggle.com/datasets/patrickzel/flight-delay-and-cancellation-dataset-2019-2023/data?select=2023.csv

The primary dataset used is the "Flight Delay and Cancellation Dataset (2019-2023)" provided by the US Department of Transportation. This dataset contains approximately 30 million rows. We will only be focusing on the 2023 portion. Some key variables include flight routes (origin, destination), time ranges for events (minutes, local time), airline codes, delay and cancellation reasons/attributions.

We will also be integrating this dataset with the following climate data from NOAA, taking local weather information of each flight into consideration when constructing the predictive model.

https://www.ncei.noaa.gov/cdo-web/datasets

Core Research Questions:

- We will be researching questions such as:
- Which airline has the most/least average delay time?
- Which city experiences the most/least average delay time?
- Which airport is most/least affected by delays?
- What time of day is most susceptible to significant delays?
- How to predict future delays given flight information and weather conditions?

Analysis Procedures:

1. Data cleaning and preprocessing
2. Descriptive statistical analysis and visualization
3. Integration with weather data and predictive model development
4. Model evaluation and refinement
5. Final analysis and project reporting

Expected Outcomes:

By the project's conclusion, we aim to provide actionable insights into flight delays and develop a reliable predictive model. This will serve as a tool for airlines and passengers to anticipate and manage delays more effectively.