Executive Summary

Module 3

Yifan Chen | Jiapeng Wang | Zhixing Liu

The scope of our analysis

In this project, our team plans to use all flight data from May to August 2024, including all airlines and all airports.

Statistical models

This project builds a total of three models, which are used to predict flight cancellations, departure delays, and arrival delays, respectively. Before building the models, we extracted weather data for the departure and arrival airports by calling an API. We considered four weather-related features: snowfall, rainfall, temperature, and wind speed. However, since weather forecasts are typically only available up to one week in advance, if a user wants to predict a flight more than a week ahead, the model will automatically exclude weather features from the input. All of this is handled automatically in the backend of our Shiny application.

Main drivers of flight cancellation

Based on our analysis, snowfall at both the departure and arrival airports is the most significant factor contributing to flight cancellations. Additionally, flights departing from small airports and those scheduled on Sundays are also more likely to be canceled. Flights departing from medium-sized airports also have a certain probability of being canceled, but this likelihood is lower compared to the factors mentioned earlier.

Main drivers of flight on-time (or early) arrivals

Surprisingly, weather is not the most significant factor affecting flight delays. The airline, scheduled departure time, and departure airport have the greatest impact on both departure and arrival delays. Additionally, for departure delays, departing from a medium-sized airport is a relatively important factor. For arrival delays, precipitation at the arrival airport is a relatively significant contributing factor.

Recommendations

- To avoid flight cancellations, it's best to avoid flying on snowy days and to choose larger airports for travel whenever possible.
- To avoid long flight delays, it's advisable to choose flights operated by major airlines such as AA, DL, and UA, and to depart from large airports.
- When booking connecting flights, it's important to avoid short layovers. Make sure to allow sufficient time between flights to account for potential delays on the initial leg of the journey.

Limitations

- Only flight data from 2024 was used, and the limited amount of data may lead to imprecise analysis results.
- All airlines and airports were considered, without focusing on a few representative ones. This may have led to overlooking specific characteristics during the modeling process.
- The model used is relatively simple. It would be beneficial to consider using a variety of different models for prediction and compare their performance.