Linear Models Project Proposal

Team members: Cristian Del Toro, Arpon Kundu, Hadiseh Azadehyaei

Dataset

The objective of the analysis is to explore the Flight Price Prediction dataset obtained from Kaggle/“Ease My Trip” website and to conduct various statistical methods and linear models in order to get meaningful information. The Flight Price Prediction dataset has 300153 samples, with 8 categorical and 2 numerical predictors; and price ticket is our target variable.

Categorical Variables:

- airline: The airline operating the flight (e.g., IndiGo, Air India, Jet Airways).
- flight: The flight number assigned by the airline.
- source_city: The city from which the flight departs.
- departure_time: The time of day when the flight departs (e.g., Morning, Afternoon, Evening, Night).
- stops: The number of stops the flight makes between the source and destination cities (e.g., non-stop, 1 stop, 2 stops).
- arrival_time: The time of day when the flight arrives at the destination (e.g., Morning, Afternoon, Evening, Night).
- destination_city: The city where the flight lands.
- class: The travel class of the passenger (e.g., Economy, Business).

Numerical Variables:

- duration: The total flight duration from departure to arrival, measured in hours.
- days_left: The number of days between the booking date and the departure date.

Target Variable:

- price: The cost of the flight ticket (Indian Rupees).

Exercises

- Apply multiple regression
- Use indicator/dummy variables
- Perform variable selection
- Optional: Apply transformations with investigation

Attestation

This dataset has not been analyzed in this specific way by any of the team members before for any previous coursework or work.