Analysing Factors Affecting Airline Passenger Satisfaction

# Research Topic Proposals

Our research aims to examine certain aspects of the airline flight experience and develop a regression model to predict the likelihood of passenger satisfaction. This analysis could be valuable for airlines looking to improve their services and customer experiences in a competitive travel industry.

Our model will leverage a survey dataset that includes certain passenger characteristics, flight details, and satisfaction ratings for select pre-flight / in-flight experience components. We will conduct exploratory data analysis (EDA) to ensure modelling suitability. Since the dependent variable is binary (satisfaction or neutral/dissatisfaction), we will utilize logistic regression modelling and test certain key assumptions for different types of independent variables; this process includes:

* Encoding binary variables (e.g., gender, travel class)
* Testing for multicollinearity using methods such as correlation matrices and variance inflation factors (VIFs)
* Evaluating relationships with log odds for continuous variables (e.g., age, flight distance), as well as ordinal ratings variables, and applying transformations as needed

# SMART Question

1. To what extent do certain surveyed passenger characteristics and flight experience components impact the likelihood that a passenger will be satisfied – rather than neutral or dissatisfied – with their trip?
2. How can we model the likelihood of passenger satisfaction—using certain surveyed passenger characteristics and flight experience components—in a manner that minimizes predictive bias?
3. To what extent can we predict the likelihood that a flight passenger will be satisfied with their experience using multiple different types of variables (including categorical, ordinal, and continuous inputs)?

# Data Source

The dataset for this research is sourced from Kaggle. The dataset contains information related to airline passenger satisfaction and has a shape of (103,904, 25) where there are 103,904 rows and 25 columns.

# GitHub Repository

[Link to our team's GitHub repository](https://github.com/parv-bhargava/DATS_6101_13_mid_term_project)