Analysing Factors Affecting Airline Passenger Satisfaction

# Research Topic Proposals

In this research, 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.

Our dataset contains variables such as demographic information, flight details, and satisfaction ratings. To prepare the data for modelling, we will:

* Test for multicollinearity using variance inflation factors (VIFs) and address such issues.
* Handle binary variables (e.g., gender, travel class) by encoding them as 0/1.
* Evaluate continuous variables (e.g., age, flight distance) for their relationship with log odds using scatterplots or other relevant techniques and apply transformations as needed.
* Evaluate satisfaction rating variables; we may calculate average ratings to avoid multicollinearity and test the parallel regression assumption.

Overall, our research aims to uncover insights into the factors affecting airline passenger satisfaction and develop predictive models to enhance the understanding of passenger preferences. This analysis will be valuable for airlines looking to improve their services and customer experiences in a competitive travel industry.

# SMART Question

- To what extent can be predict airline passenger satisfaction with certain factors?

# 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 for this research. [GitHub](https://github.com/parv-bhargava/DATS_6101_13_mid_term_project)