# Logistic Regression – Research and Implementation

Use the dataset at: https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction

### **Research Questions**

1. What are the key features that influence airline passenger satisfaction, and how can logistic regression capture these relationships in a multiclass classification context?
2. How does the performance of logistic regression vary when different preprocessing techniques (e.g., imputation, feature scaling) are applied to the dataset?
3. What is the impact of feature engineering, such as combining or transforming variables, on the accuracy of logistic regression for predicting passenger satisfaction levels?
4. How does logistic regression compare with other classification algorithms (e.g., Random Forest, Support Vector Machines) in predicting passenger satisfaction?

### **Practical Questions**

1. How can missing data in the Airline Passenger Satisfaction dataset be handled to minimize its impact on model performance?
2. What preprocessing steps are necessary to prepare categorical features (e.g., Gender, Class) and numerical features (e.g., Flight Distance, Age) for logistic regression?
3. How can logistic regression's hyperparameters, such as regularization strength, be optimized for multiclass classification in this dataset?
4. How does class imbalance in the satisfaction categories affect the performance of logistic regression, and what methods can mitigate this imbalance?
5. Can interaction terms or derived features (e.g., Total Satisfaction Score) improve the predictive power of the logistic regression model?
6. What insights can be derived from the coefficients of the logistic regression model regarding the factors most strongly associated with passenger satisfaction?