**Report**

1. Data Loading and Exploration

The goal of this project is to predict whether the customers of the airline is satisfied or dissatisfied. As the target variable is categorical, we will be using models that works with classification.

2. Data Preprocessing

Data Cleaning

• autoclean() and klib.datacleaning() were not used

• Removed unnecessary columns which are non-predictive such as id, unnamed: 0

• Handled missing values by filling them as appropriate (mode, median).

• encoding was done as there are multiple categorical columns which prevents our machine learning model to predict

• scaling was done only on features assigning them to x variable with target value included to increase the performance

3. Data Visualisation is also shown with barplot based on value counts of each categorical columns.

According to our visualized data, we can draw several conclusions regarding our dataset. Here are some of them:

* The number of men and women in this sample is approximately the same
* The vast majority of the airline's customers are loyal customers
* most travels’ purpose is not business rather than personal reasons
* About half of the passengers were in business class

according to our countplot from seaborn library:

* most of the passengers who flew in economy plus or economy class were dissatisfied with the flight, and those who were lucky enough to fly in business class were satisfied

4. Model Training

Splitted the dataset into training and test sets using an 80-20 split

Model Selection

One tree and one ensemble class machine learning model was considered since the target output is categorical:  
• Random Forest Classifier and Decision Tree Classifier

6. Evaluation

The following metrics were to evaluate the model

• Accuracy score

• Classification report