Goal:
The goal of the problem is to **predict whether a passenger was satisfied or not** considering his/her overall experience of travelling on the Shinkansen Bullet Train.

Dataset:

The problem consists of 2 separate datasets: **Travel data** & **Survey data. Travel data** has information related to passengers and attributes related to the Shinkansen train, in which they travelled. The survey data is aggregated data of surveys indicating the post-service experience. You are expected to treat both these datasets as raw data and perform any necessary data cleaning/validation steps as required.

The data has been split into two groups and provided in the Dataset folder. The folder contains both train and test data separately.

- Train_Data
- Test_Data

Target Variable: Overall_Experience (1 represents 'satisfied', and 0 represents 'not satisfied')

The **training set** can be used to build your machine-learning model. The training set has labels for the target column - **Overall_Experience**.

The **testing set** should be used to see how well your model performs on unseen data. For the test set, it is expected to predict the '**Overall_Experience**' level for each participant.

Data Dictionary:

All the data is self-explanatory. The survey levels are explained in the Data Dictionary file.

Submission File Format: You will need to submit a CSV file with exactly 35,602 entries plus a header row. The file should have exactly two columns

- ID
- **Overall_Experience** (contains 0 & 1 values, 1 represents 'Satisfied', and 0 represents 'Not Satisfied')

Evaluation Criteria:

Accuracy Score: The evaluation metric is simply the percentage of predictions made by the model that turned out to be correct. This is also called the **accuracy** of the model. It will be calculated as the total number of correct predictions (True Positives + True Negatives) divided by the total number of observations in the dataset.

In other words, the best possible accuracy is 100% (or 1), and the worst possible accuracy 0%.