# Explanation of My Work

## 1. Objective of the Project

- The goal was to work with an unlabeled dataset and use the K-Means clustering algorithm to create two clusters (binary classification).  
- After clustering, I added these cluster labels (0 and 1) as a new column to the dataset.  
- Finally, I used the updated dataset to train and test a supervised machine learning model to classify the data.

## 2. Steps I Followed

### Step 1: Data Preparation

- First, I handled missing values by filling them with the median for numeric columns.  
- Then, I scaled the numeric features to a range between 0 and 1, so the clustering algorithm would work properly.

### Step 2: K-Means Clustering

- I applied the K-Means algorithm to group the data into two clusters.  
- These clusters were treated as binary class labels (0 and 1).  
- I added these labels as a new column to the dataset.

### Step 3: Train-Test Split

- I divided the dataset into three parts: training, validation, and test sets.  
- The training set was used to train the model, the validation set to fine-tune it, and the test set to evaluate it.

### Step 4: Training a Decision Tree Classifier

- I trained a Decision Tree model using the labeled dataset.  
- This model learned to classify data points into Cluster 0 or Cluster 1 based on their features.

### Step 5: Model Evaluation

- I tested the model on the validation and test sets.  
- The accuracy was very high, around 99% for both the validation and test sets.  
- This means the model is performing very well and can classify the data accurately.  
- I also checked metrics like precision, recall, and F1-score, which were also close to 99%.

## 3. Key Insights

- The Decision Tree model performed well, and the results show it is reliable and not overfitting.  
- By creating clusters with K-Means and training a classifier, I combined unsupervised and supervised learning in this project.