Data Science Project

- **Step 1:** Import Necessary Libraries
- **Explanation: Importing essential libraries.**
  - pandas for data manipulation and analysis.
  - numpy for numerical operations.
  - matplotlib.pyplot and seaborn for data visualization.
  - train_test_split for splitting the data into training and testing sets.
  - RandomForestClassifier for building a random forest classification model.
  - classification_report, accuracy_score, and confusion_matrix for evaluating the model.
  - StandardScaler for standardizing feature values

- **Step 2:** Load the Breast Cancer Dataset
- **Explanation:** Loading the Breast Cancer Wisconsin (Diagnostic) dataset from sci-kit-learn.

- **Step 3:** Create a DataFrame
- **Explanation:** Creating a Pandas DataFrame to organize the dataset. The features are stored in columns, and the target variable ('target') is added.

- **Step 4:** Exploratory Data Analysis (EDA)
- **Explanation:** Conducting exploratory data analysis to understand the data.

Visualizing the distribution of target classes using count plot.

Creating a correlation heatmap (sns. heatmap) to identify relationships between features.

- **Step 5:** Data Preprocessing
- **Explanation:** Separating features (X) and the target variable (y) to prepare for model training.

- **Step 6:** Split the Data
- Explanation: Splitting the dataset into training and testing sets using the train_test_split function. A common practice is to use 80% of the data for training and 20% for testing.

- **Step 7:** Standardize the Features
- **Explanation**: Standardizing the features to ensure they are on a similar scale. This is important for many machine learning algorithms.

- **Step 8:** Build and Train the Model
- **Explanation:** Building and training a RandomForestClassifier using the training data.

- **Step 9**: Model Evaluation
- **Explanation:** Evaluating the model's performance on the testing set.

accuracy_score: Calculates the accuracy of the model. Confusion_matrix: Displays the number of true positives, true negatives, false positives, and false negatives.

classification_report: Provides precision, recall, F1-score, and support for each class.