Part 3: Documentation

Data Science Task (Part 1):
For the prediction task, I chose to implement a Decision Tree Classifier. My choice was influenced by the model's simplicity and interpretability, making it an ideal match for this foundational dataset. After preprocessing the data and dividing it into training (70%) and testing (30%) sets, I trained the model and used it to predict the species of the Iris flowers in the test set.The performance of my model was evaluated using several key metrics:
Accuracy: The model achieved an accuracy of 1, indicating a commendable level of overall correctness.
Precision: The model achieved the precision score as 1, which reflects the model's proficiency in minimising false positives.
Recall: The model achieved the recall score as 1, indicating the model's capability to identify all pertinent instances.
Classification Report: The classification report provided a granular breakdown of precision, recall, and F1-score for each class, highlighting the model's balanced performance across different Iris species.
These metrics collectively underscore that the Decision Tree Classifier was a good choice, fitting the intricacies of this classification task.

Simple Exploratory Data Analysis (EDA) (Part 2):
My EDA offered invaluable insights into the Iris dataset:
Histograms: I observed that the histograms for each feature illustrated the distribution of measurements. Notably, the distributions of petal length and petal width were bimodal, hinting at distinct species groups.
Pairplot: The pairplot from Seaborn vividly delineated the relationships between feature pairs, distinctly revealing clusters associated with the different species. This visualisation underpinned the correlation between the features and Iris species, validating their utilisation in the classification task.
Box Plots: I crafted box plots for each feature, grouped by species, to visually articulate the spread and variance of the data. These plots were also instrumental in spotting outliers and grasping the distribution nuances of each feature across species.

These visual insights were paramount in comprehending the dataset's characteristics, steering the feature selection for the classification endeavour.

Challenges & Learnings:
Throughout the project, a prominent challenge I encountered was ensuring the model's adaptability across different species. This hurdle was surmounted by meticulously analysing the EDA outcomes and selecting a model adept at navigating the dataset's subtleties. The project underlined the significance of a comprehensive EDA preceding model selection and training, spotlighting how adept data visualisation can unveil patterns that guide more informed decisions regarding model choice and feature selection.