Exercise: Analyzing the Iris Dataset

Now that you have a good understanding of exploratory data analysis and its importance, it's time to put your knowledge to a more practical example. We'll be focusing on an iris public dataset from the scikit-learn library.

We've included a checklist to guide you into starting your Sagemaker Studio instance, just in case.

Exercise: Iris Dataset Setup Checklist

Before you begin, here is a checklist to get you ready for the exercise.

Launch the AWS Gateway

Open up Sagemaker Studio

Make sure you have a cloned repository of the course

Navigate to the `Exercise: Iris Dataset` directory and open the starter Jupyter notebook.

Now that you have your Sagemaker Studio instance up and running, use the checklist below to confirm you finished all of the steps for the exercise. You'll open up Lesson 2's Jupyter Notebook that is related to this exercise and complete it by filling in all the `?` locations in the notebook.

Final Exercise: Iris Dataset Exercise Checklist

Make sure you finish these main steps.

Load the iris dataset into a pandas dataframe

Create a table summary of the features and target values

Create a histogram of all the features and target

Create a correlation matrix of the features and target

Create scatter plots of all the features and target

Reminder

After you are finished, remember to shut down all instances running on Sagemaker Studio to prevent any unused resources being billed!