

Data Science

Exploratory Data Analysis

Learning Objectives: Using Jupyter notebook to work with a dataset

Due date: 9/11/2025 12:15PM

Open the following Kaggle web-site:

<https://www.kaggle.com/code/rakesh6t184/seaborn-plot-to-visualize-iris-data/notebook>

Create your GitHub homework space, by clicking on this link:

<https://classroom.github.com/a/vB46goBS>

We will be using the code from that dataset to complete the following 14 steps.

Clone the homework space, first you have to accept this link to create the repo, then clone in your computer. You can use any IDE, e.g. PyCharm, VSCode, etc. is up to you. Your repo URL looks like this (replace githubusername with yours)

<https://github.com/computingcelts/f25-ds-hw-githubusername>

Answer the questions for each point in a Markdown cell in your Jupyter Notebook. Include your code in Code cells, one for each task. You can talk to other classmates if you get stuck with anything.

Before starting, make sure you have installed the following python packages (pandas, numpy, seaborn).

Download the following dataset into your Jupyter Notebook.

<https://www.kaggle.com/code/rakesh6184/seaborn-plot-to-visualize-iris-data/input>

1. Importing pandas and Seaborn module.
How do you import packages in python?
2. Importing Iris data set.
How do you know your dataset was downloaded and where was it saved.
When opening the csv file, make sure to use a relative path, don't use a full path, for this you have to become familiar how to reference a file in your file system without using a full path.
3. Displaying data.
How many features and observations are there in your dataset?

4. Checking if there are any missing values.
Why is it important to verify if there are missing values in the dataset?
5. Joint plot.
What is the purpose of the joint plot, what does it show?
6. FacetGrid Plot.
What is the purpose of the FacetGrid plot, what does it show?
7. Boxplot.
What is the purpose of the Boxplot, what does it show?
8. Strip plot.
What is the purpose of the Strip plot, what does it show?
9. Combining Box and Strip Plots.
Why would be a good idea to combine plots?
10. Violin Plot.
What does the violin plot show?
11. Pair Plot.
What is the purpose of the Pair Plot and why would you use it?
12. Plotting heat map.
This heatmap shows correlation coefficients for this dataset? What does it show?
13. Distribution plot.
What is the purpose of a distribution plot? Why would you use it?
14. Swarm plot.
What is the purpose of the Swarm plot.

Once completed, commit and push your code on GitHub. Verify on your browser that the submission went through.