

S3 MCA (2024-26 Batch)

M24CA1L306 Data Science Lab

Lab Exercise #02

Date of Submission: 11 Sept 2025

- Download the dataset **iris.csv**
- Using Python, visualize the various characteristics of the dataset.
- Use Google Colab for implementation.
- Define your own conditions to learn the characteristics
- After each visualization write your findings
- Use the following types of charts for visualization:
 1. Histogram
 2. Quartile (Box) plot
 3. Distribution chart
 4. Scatterplot
 5. Scatter multiple
 6. Scatter matrix
 7. Bubble chart
 8. Density chart
 9. Parallel chart
 10. Deviation chart
 11. Andrews curves
- Sample EDA Questions to discover insights about **iris.csv** dataset:
 1. Which species has the smallest flowers?
 - Use boxplots or histograms of petal length and petal width across species.
 - **Findings expected:** Setosa has the smallest petals compared to Versicolor and Virginica.
 2. Which feature best separates the three species?
 - Use pairplots or scatter plots of features colored by species.
 - **Findings expected:** Petal length and petal width clearly separate species.
 3. Are sepal dimensions good indicators of species?
 - Plot scatter plots of sepal length vs. sepal width with species as hue.
 - **Findings expected:** Sepal dimensions overlap more, making them weaker predictors than petals.
 4. Which two features are most correlated?
 - Draw a heatmap of correlations.
 - **Findings expected:** Petal length and petal width show the strongest positive correlation.
 5. How do the sizes of flowers vary across species?
 - Create violin plots of sepal and petal features.
 - **Findings expected:** Virginica tends to have the largest flowers, followed by Versicolor, while Setosa is the smallest.