

# **Exploratory Data Analysis Report**

## **Internship Task 3 – AI & ML**

**Name:** Your Name

**Dataset:** Iris Dataset

**Tools Used:** Python, Pandas, Matplotlib, Seaborn

**Date:** 19-01-2026

### **1. Introduction**

Exploratory Data Analysis (EDA) is the process of analyzing datasets using statistics and visualization techniques to understand data patterns, relationships, and anomalies. This report presents EDA performed on the Iris dataset to study feature behavior and relationships.

### **2. Dataset Overview**

- Total records: 150
- Total features: 4 numerical + 1 categorical
- Target variable: Species
- Missing values: None

The dataset is clean and ready for analysis.

### **3. Histogram Analysis**

Histograms were plotted for numerical features to observe their distribution.

#### **Observation:**

- Sepal length and petal length show near normal distribution.
- Petal features show better separation across species.
- No extreme skewness observed.

### **4. Species Distribution (Count Plot)**

A count plot was used to analyze categorical data.

#### **Observation:**

- All three species (Setosa, Versicolor, Virginica) have equal number of samples.
- Dataset is balanced.

## 5. Outlier Detection (Box Plot)

Box plots were used to identify outliers.

### Observation:

- Minor outliers observed in sepal width.
- No extreme anomalies detected.

## 6. Correlation Analysis (Heatmap)

Correlation heatmap was plotted to understand feature relationships.

### Observation:

- Strong positive correlation between petal length and petal width.
- Sepal features show weaker correlation.
- Petal features are more important for prediction.

## 7. Important Features Identified

Based on analysis:

- Petal Length
- Petal Width

These features contribute most towards classification accuracy.

## 8. Conclusion

The Iris dataset is clean, balanced, and well-structured. Strong relationships exist between petal features, making them important predictors for machine learning models. EDA helped understand feature behaviour and improved data understanding.