EXPLORATORY DATA ANALYSIS (EDA) AND DATA VISUALIZATION ON THE IRIS DATASET**

PROJECT OVERVIEW

Objective: The objective of this project was to perform exploratory data analysis (EDA) on the Iris dataset to uncover initial patterns, spot anomalies, and test hypotheses. Additionally, the project aimed to create visualizations to summarize the main characteristics of the dataset.

Problem Statement: The Iris dataset is a classic benchmark dataset in machine learning, consisting of measurements of sepals and petals of three different species of Iris flowers. The goal was to gain insights into the dataset's features and relationships between them.

Tools & Technologies:

- Python: Primary programming language for data analysis and visualization.

- Pandas & NumPy: Used for data manipulation and analysis.

- Matplotlib & Seaborn: Utilized for data visualization and exploratory data analysis.

Key Steps:

1. **Data Loading**: Loaded the Iris dataset into a Pandas DataFrame using the Seaborn library.

2. **Data Overview**: Obtained a basic understanding of the dataset, including its shape, data types, and summary statistics.

3. **Data Cleaning**: Checked for missing values and anomalies in the data and performed data cleaning if necessary.

4. **Univariate Analysis**: Analyzed the distribution of individual variables using histograms and box plots.

5. **Bivariate Analysis**: Explored relationships between pairs of variables using scatter plots and pair plots.

6. **Multivariate Analysis**: Investigated relationships between multiple variables using heatmaps and other multivariate plots.

7. **Data Visualization**: Created various plots, including violin plots, to visualize the data and summarize findings.

Key Findings:

- The Iris dataset contains 150 samples, with 50 samples for each of the three species.

- Sepal and petal measurements vary significantly among the three species, indicating clear differences.

- Visualizations revealed distinct clusters for each species, suggesting clear separability based on certain features.

GitHub Repository:

Explore the complete project and code here:

https://github.com/sravanthi224/Quanta-1.git