

Data Analyst Internship Project Documentation

Project 1: Iris Dataset Basic Analysis

1.1 Exploratory Data Analysis (EDA) with Python:

The Iris dataset contains 150 samples of flowers, categorized into three species: Setosa, Versicolor, and Virginica. The features include Sepal Length, Sepal Width, Petal Length, and Petal Width.

The EDA was performed using Python libraries such as Pandas, Seaborn, and Matplotlib to analyze the data and visualize its statistical properties.

1.2 Data Visualization with Power BI:

Power BI was used to visualize the insights derived from Python. The Iris dataset was exported to CSV, then imported into Power BI for visualization. Key visuals include scatterplots, bar charts, and heatmaps.

Project 2: Weather Dataset Analysis

2.1 Data Preparation with Python:

The weather dataset contains parameters such as temperature, humidity, and wind speed. Initial data exploration revealed missing values and outliers. Data preprocessing was performed to clean the dataset, handle missing values, and remove outliers.

2.2 Correlation and Regression Analysis:

Correlation matrix was created to identify relationships between weather parameters. Regression was used to predict temperature based on humidity and wind speed.

2.3 Advanced Analysis with Power BI:

Power BI was used to build interactive dashboards with temperature trends, correlation plots, and interactive filters for date and location. This facilitated deeper analysis.

Conclusion:

Python was used for data analysis, EDA, and regression, while Power BI provided interactive visualizations. Together, these tools offered deep insights into the datasets.