PROJECT DOCUMENTATION

IRIS DATASET:

For this project, I conducted Exploratory Data Analysis (EDA) on the well-known Iris dataset. The Iris dataset is a classic dataset in the field of machine learning and data analysis, containing information about three species of iris flowers: Setosa, Versicolor, and Virginica. It comprises four features: sepal length, sepal width, petal length, and petal width.

Methodology:

1. Exploratory Data Analysis (EDA) with Python:

- I used the Python programming language to perform EDA on the Iris dataset.
- I leveraged the Pandas library for data manipulation and initial data exploration.
- Matplotlib and Seaborn libraries were utilized for data visualization.

2. Data Visualization with Power BI or Tableau:

- I opted for Tableau to create visualizations representing the patterns observed during EDA.
- Various chart types, including scatter plots, bar charts, and box plots, were employed to visualize correlations, patterns, and trends within the Iris dataset.

Exploratory Data Analysis Findings:

Data Summary:

- The Iris dataset comprises 150 samples, evenly distributed among the three species.
- The features exhibit variations in their distributions, with petal length and petal width showing the most distinguishable patterns among the species.

Correlation Analysis:

- I observed strong positive correlations between petal length and petal width, as well as between sepal length and petal length.
- Sepal width showed weaker correlations with other features compared to the other features.

Species Differentiation:

- Through visualization, it became evident that the three species exhibit distinct clusters in the feature space, particularly in the petal dimensions.
- Setosa species can be easily distinguished from the other two species due to its smaller petal dimensions.

WEATHER DATASET:

1. Data Preparation with Python:

In this phase, I utilized Python programming language to clean and preprocess the weather dataset. My primary objectives were to handle missing values, outliers, and inconsistencies in the data to ensure its quality for further analysis.

Data Cleaning Process:

- I identified and handled missing values using various techniques such as imputation or deletion based on the extent of missingness and the nature of the data.
- Outlier Detection and Treatment: I employed statistical methods to identify outliers and applied appropriate techniques like removal based on domain knowledge.
- Consistency Checks: I ensured consistency across different variables by checking for data integrity issues such as data type inconsistencies or illogical values.
- Data Transformation: I applied necessary transformations such as scaling, normalization, or encoding categorical variables to prepare the data for analysis.

2. Advanced Analysis with Power BI:

In this phase, I leveraged Power BI for advanced data analysis on the weather dataset. My aim was to create interactive dashboards and visualizations to uncover trends and patterns within the data.

Key Steps in Advanced Analysis:

- Data Integration: I integrated the cleaned weather dataset into Power BI for analysis.
- Dashboard Creation: I developed interactive dashboards comprising various visualizations such as line charts, scatter plots, histograms, and heat maps to explore different aspects of the weather data.
- Drill-Down Functionality: I implemented drill-down functionality within the dashboards to allow users to delve deeper into specific regions, time periods, or weather parameters.
- Trend Identification: I utilized trend analysis techniques to identify long-term patterns and fluctuations in weather parameters.
- Anomaly Detection: I incorporated anomaly detection algorithms to identify unusual patterns or events within the data.
- Correlation Analysis: I performed correlation analysis to identify relationships between different weather parameters. Correlation analysis helped me understand how variables are related to each other and the strength and direction of these relationships.

Key Findings from Correlation Analysis:

- I identified significant correlations between various weather parameters such as temperature, humidity, precipitation, wind speed, etc.
- I determined the degree of correlation using correlation coefficients and visualized the relationships using scatter plots.

Through the data preparation and advanced analysis phases, I gained valuable insights into the weather dataset, uncovering patterns, trends, and relationships between different weather parameters. The correlation analysis provided further understanding of the interdependencies among weather variables, enabling me to make informed predictions and recommendations based on the data.