INTERN PROJECT PHASE - 1

Data Analyst Projects: Exploring Data Insights

Project 1: Iris Dataset Basic Analysis:

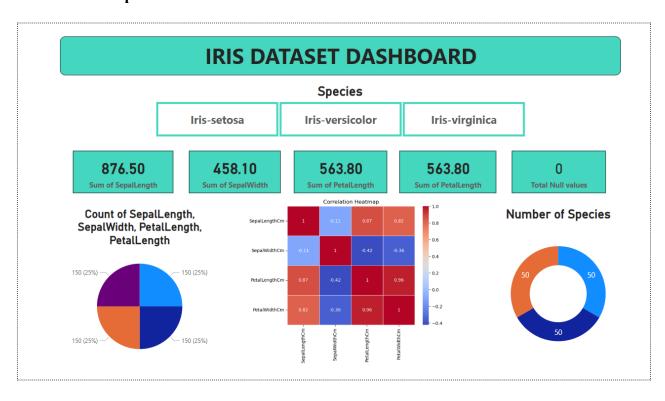
I. Python code explanation:

- First, I import the necessary packages for my Exploratory Data Analysis (EDA).
- Next, I identify which columns are not necessary in this dataset. I remove the 'Id' column because it is not used in the EDA. Then, I check the datatype of each column using the 'info()' method.
- After that, I want to know the null values in my dataset, so I use the 'isna().sum()' method. This function helps to determine the sum of null values in each column. Then, I drop the null values in my dataset.
- I use the 'describe()' method to obtain statistics such as count, mean, standard deviation, minimum, and maximum values for each column.
- Then, I use the 'corr()' method to compute the pairwise correlation of columns. I display the pairplot, histplot, and correlation plot to visualize the relationships between multiple variables, distribution, and the correlation between variables in the dataset.

II. PowerBI explanation:

- In Power BI, I first insert the slicer widget. When I click on a specific species in the slicer, it provides an interactive output for that species.
- Next, I insert five cards to display the sum of Sepal Length, Sepal Width, Petal Length, Petal Width, and the total count of null values in the dataset.
- Then, I use two pie charts to display the count of each Sepal Length, Sepal Width, Petal Length, Petal Width, and to visualize the distribution of species in the dataset.
- Additionally, I import a correlation heatmap to visualize the relationships between the variables and their correlation coefficients.

III. PowerBI output screenshot:



Project 2: Weather Analysis:

I. Python code explanation:

- In Weather Analysis, first, I preprocess the data to determine the datatype, null values, statistics values, etc.
- Then, I visualize the box plot to identify outliers in the category-related columns and to visualize the correlation.
- After that, I use Regression for weather predictions. In the weather predictions, I select some important columns as the predict var, and I select the 'RainTomorrow' column as the target var.
- Before the Regression, I encode the 'RainTomorrow' column using Label Encoder for model fitting, prediction, and printing the MSE, MAE, and R2 score to evaluate the performance of the Regression model.

II. PowerBI explanation:

- In PowerBI, I insert three slicers: Rain Today, Rain Tomorrow, and Wind Gust Direction to obtain interactive output in the table and all cards.
- Then, I insert the table widget. When I click on any slicer, it displays the specific slicer-related rows. Additionally, when I click on any row in the table, it displays the output in the five cards.
- After that, I use the five cards to obtain the output when I click on the slicer or table row. I display the interactive output for Wind Gust Direction, Rain Today, Rain Tomorrow, Wind Direction 9am, and Wind Direction 3pm in the five cards.
- These insights help to determine the timing of rainfall for today and tomorrow. Additionally, they enable advanced visualization techniques.

III. PowerBI output screenshot:

				1	Weathe	er Data	set Da	ashbo	ard				
	F	RainToday			∨ Ra	ainTomorrow		~	WindGustDir			~	
	All				✓ All			~	All			~	
inTemp	MaxTemp	Rainfall	Evaporation	Sunshine	WindGustSpeed	Humidity3pm	Humidity9am	Pressure3pm	Pressure9am	Cloud3pm	Cloud9am	Temp3pm	Temp
12.30	30.20	0.00	6.00	12.60	78	13	43	1,009.20	1,009.60	1	1	28.60	2
14.30	35.00	0.00	7.60	10.50	41	14	43	1,008.70	1,012.60	1	0	33.60	2
9.00	30.70	0.00	7.60	12.10	76	15	38	1,010.80	1,016.10	3	1	30.00	20
7.50	23.30	0.00	6.80	10.90	65	15	46	1,010.90	1,012.50	0	1	21.30	18
13.80	35.20	0.00	6.40	11.20	48	15	47	1,019.30	1,022.50	2	5	34.10	2
12.30	33.80	0.00	7.20	11.30	22	15	57	1,019.10	1,022.60	1	0	32.70	17
10.10	31.20	0.00	8.80	13.10	41	16	45	1,013.70	1,018.20	1	0	30.40	18
1.40	20.60	0.00	5.40	11.10	46	16	56	1,015.40	1,020.80	1	5	19.50	11
3.30	25.50		5.20	10.80	43		57	1,014.60	1,018.80	3	0		
11.30	32.30	0.00	9.40	11.40	28	17	44	1,020.70	1,024.30	2	5	30.50	18
F			No		No			Е		Е			