**DEN VIRTUAL INTERNSHIP**

**TASK NO # 3**

**Implement a data analysis project using pandas and matplotlib to explore and visualize a dataset of your choice.**

**So lets start**

**I will use jupyter notebook for that :**

**# Step 1: Import necessary libraries**

* import pandas as pd
* import matplotlib.pyplot as plt
* import seaborn as sns
* from sklearn.datasets import load\_iris

**# Step 2: Load the dataset**

# The dataset is available in sklearn, but you can load from a CSV if preferred

iris = load\_iris()

iris\_data = pd.DataFrame(data=iris.data, columns=iris.feature\_names)

iris\_data['species'] = iris.target

**# Step 3: Data exploration**

# Viewing the first few rows

print("First five rows of the dataset:")

print(iris\_data.head())

**# Checking for missing values**

print("\nMissing values in the dataset:")

print(iris\_data.isnull().sum())

# Summary statistics

print("\nSummary statistics:")

print(iris\_data.describe())

**# Step 4: Data Visualization**

**# 4.1: Histogram of features**

iris\_data.hist(bins=20, figsize=(10, 8))

plt.suptitle('Histograms of Iris Features')

plt.show()

**# 4.2: Scatter plot matrix (pairplot)**

sns.pairplot(iris\_data, hue='species', markers=["o", "s", "D"], height=2.5)

plt.suptitle('Pairplot of Iris Features by Species')

plt.show()

**# 4.3: Boxplot of features**

plt.figure(figsize=(10, 6))

sns.boxplot(data=iris\_data.drop('species', axis=1), orient="h", palette="Set2")

plt.title('Boxplot of Iris Features')

plt.show()

**# 4.4: Violin plot of feature distribution by species**

plt.figure(figsize=(12, 6))

sns.violinplot(x="species", y="sepal length (cm)", data=iris\_data)

plt.title('Violin Plot of Sepal Length by Species')

plt.show()

now it is upon us how we can understand the data :