**Data Analysis Project Using Python**

**Objective**:

Analyze a dataset of student exam scores and answer specific questions, presenting findings using Python libraries and techniques for data analysis.

**#Importing Libraries**

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

**Step1:Load the dataset into a DataFrame**

data=pd.read\_csv(r"C:\Users\rudra\OneDrive\Desktop\Main\_Flow\_Tech\student-mat.csv", sep=";")

print("Data loaded successfully!")

print(data.info())

**Step2: Data Exploration**

print(data.head())

print("\nDataset Info:")

**Step3:Data Cleaning**

print("\nMissing Values:")

print(data.isnull().sum())

#Remove Duplicates

data = data.drop\_duplicates()

**Step4:Data Analytics**

average\_score = data['G3'].mean()

print(f"\nAverage Math Score (G3): {average\_score:.2f}")

1) students\_above\_15 = len(data[data['G3'] > 15])

print(f"Number of students scoring above 15: {students\_above\_15}")

2) correlation = data['studytime'].corr(data['G3'])

print(f"Correlation between study time and final grade: {correlation:.2f}")

3)average\_grade\_by\_gender=data.groupby('sex')['G3'].mean()

print("\nAverage Final Grade by Gender:")

print(average\_grade\_by\_gender)

**Step5:Data Visualizations**

plt.figure(figsize=(8, 5))

plt.hist(data['G3'], bins=10, color='skyblue', edgecolor='black')

plt.title("Distribution of Final Grades (G3)")

plt.xlabel("Final Grade")

plt.ylabel("Frequency")

plt.show()

**#Scatter Plot of study time vs final grade**

plt.figure(figsize=(8, 5))

sns.scatterplot(data=data, x='studytime', y='G3', hue='sex')

plt.title("Study Time vs Final Grade")

plt.xlabel("Study Time (hours)")

plt.ylabel("Final Grade")

plt.legend(title="Gender")

plt.show()

**#Bar chart of average scores by gender**

plt.figure(figsize=(8, 5))

average\_grade\_by\_gender.plot(kind='bar', color=['blue', 'pink'])

plt.title("Average Final Grade by Gender")

plt.ylabel("Average Final Grade")

plt.xlabel("Gender")

plt.xticks(rotation=0)

plt.show()

**Conclusion:** The analysis of the dataset of student exam scores provided meaningful insights into the performance trends and factors influencing student outcomes.