Question37:

37. Scenario: You are a data scientist working for an educational institution, and you want to

explore the correlation between students' study time and their exam scores. You have collected data

from a group of students, noting their study time in hours and their corresponding scores in an

exam.

Question: Identify any potential correlation between study time and exam scores and explore

various plotting functions to visualize this relationship effectively.

Answer:

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

from scipy.stats import pearsonr

# Load the student study and score data

file\_path =r"C:\Users\jampa\Downloads\student\_study\_scores.csv"

df = pd.read\_csv(file\_path)

def analyze\_study\_score\_correlation(df):

# Calculate correlation coefficient

correlation, p\_value = pearsonr(df['Study\_Time\_Hours'], df['Exam\_Score'])

print(f"Correlation Coefficient: {correlation:.2f}")

print(f"P-Value: {p\_value:.4f}")

# Scatter plot with regression line

plt.figure(figsize=(8,4))

sns.regplot(x='Study\_Time\_Hours', y='Exam\_Score', data=df, ci=None, scatter\_kws={'color': 'blue'}, line\_kws={'color': 'red'})

plt.title('Study Time vs Exam Score')

plt.xlabel('Study Time (Hours)')

plt.ylabel('Exam Score')

plt.grid(True, linestyle='--', alpha=0.6)

plt.show()

# Distribution plot

plt.figure(figsize=(8,4))

sns.histplot(df['Exam\_Score'], kde=True, color='purple')

plt.title('Distribution of Exam Scores')

plt.xlabel('Exam Score')

plt.ylabel('Frequency')

plt.grid(True, linestyle='--', alpha=0.6)

plt.show()

# Run the analysis

analyze\_study\_score\_correlation(df)

Output:



