TECHX 3.0 ASSIGNMENT
MACHINE LEARNING

ASSIGNMENT OVERVIEW:

Perform exploratory data analysis (EDA) on a dataset about students' performance. The goal is to understand the structure of the dataset and generate insights using simple data analysis techniques.

DATASET

Use the "Students Performance" dataset, which can be downloaded from here

Tasks:
1. Understand the Dataset:
○ Load the dataset using pandas.
○ Display the first 5 rows of the dataset.
○ Show the structure and summary of the dataset (number of rows, columns, data types).
2. Descriptive Statistics:
○ Calculate the mean, median, and standard deviation for the scores in math, reading, and writing.
○ Find the minimum and maximum scores for each subject.
3. Data Visualizations:
○ Create a histogram for the distribution of scores in math, reading, and writing.
○ Plot a bar chart showing the average score in each subject, categorized by gender.

○ Use a scatter plot to show the relationship between math scores and reading scores.
4. Categorical Data Analysis:
○ Analyze the effect of parental education level on the average scores.
 Does it impact students' performance?
○ Compare the performance of students who completed the test preparation course versus those who did not.
5. Insights:
○ Based on the visualizations and statistics, what insights can you draw about students' performance in relation to their gender, parental education, and test preparation?

SUBMISSION GUIDELINES:

● Submit a Jupyter notebook with your analysis, visualizations, and insights.
● Make sure to explain each step and include comments in your code.

REFERENCES:

● Dataset: Students Performance in Exams - Kaggle
● Python libraries:
   ○ Pandas Documentation
   ○ Matplotlib Documentation
   ○ Seaborn Documentation

This assignment should be accessible and engaging for students who are new to data analysis while providing them with hands-on experience.