Assignment 01

Fundamentals of Data Science

Total Marks: 30 (10 marks per question)

Question 1: The Big Picture of Data Science

Imagine you are explaining Data Science to a friend who thinks it's just about "coding and numbers."

- How would you describe Data Science as an interdisciplinary field? Mention at least three core areas it combines and explain why each is important.
- How is Data Science different from Machine Learning (ML)? Provide a real-world example (e.g., predicting weather vs. analyzing climate trends) to highlight the difference.
- Why are soft skills like storytelling and communication critical for a Data Scientist? Give an example of how poor communication could lead to a failed project.

Question 2: The Data Science Process in Action

You are tasked with building a system to recommend books to users based on their preferences.

- List and briefly explain the key stages of the Data Science process you would follow for this project.
- Why is Exploratory Data Analysis (EDA) important before building the model? Mention two specific tasks you'd perform during EDA (e.g., detecting outliers, checking data types).
- How would you evaluate the final model? Name one metric to assess its performance.

Question 3: Understanding Data Attributes

A dataset contains information about students in a school, including:

- Height (in cm)
- Favorite Subject (Math, Science, Arts)
- Exam Pass/Fail Status (Yes/No)
- Student ID (e.g., S001, S002)

For each attribute above:

- Classify its type (Nominal, Binary, or Other) and justify your answer. Which attribute
- is asymmetric binary? Explain why it's asymmetric with a real-world consequence (e.g., how misclassifying a "Fail" as "Pass" could impact students).
- Why can't we calculate the "average" of Student ID? Relate your answer to the properties of nominal attributes.