Project: Expectation Decider

Project Objective:

You are hired as a junior data analyst in an educational research institute. The institute is launching a new "Expectation Decider" model that predicts whether a student will pass a competitive mathematics exam based on historical data. You have been given sample data of 200 students, including their study hours, attendance, participation in group discussions, and previous test performance. Your task is to **analyze the probability patterns** from the data and derive meaningful insights using the topics below.

Scenario:

The dataset contains the following information for each student:

Field	Description
study_hours	Number of hours a student studied per week
attendance	Percentage attendance in lectures
group_discussion	Participation in group discussions (Yes/No)
previous_test_score	Marks out of 100 from the last internal test
final_exam_pass	Whether the student passed the competitive exam (Pass/Fail)

Note: Generate the dataset from the AI tool in such a way that you can solve all the tasks.

Question:

Using the given dataset and **appropriate statistical techniques**, complete the following tasks in **one integrated analysis report**. All work must be shown clearly.

1. **Understanding the Basics**
   - Explain in your own words:
     - What is Probability?
     - Key Probability Terminology.
     - Give at least **three probability event examples** from the dataset.

2. **Types of Events**
   - Identify one **empirical probability** and one **theoretical probability** scenario from the dataset and calculate both.

3. **Random Variable & Probability Distribution**
   - Define a **random variable** for the event "Number of students passing the final exam" out of 3 randomly selected students.
   - Construct its **probability distribution table**.
   - Find the **mean** and **variance** of this random variable.

4. **Venn Diagram in Probability**
   - Draw a **Venn diagram** showing:
     - Students who **study more than 10 hours/week**.
     - Students who **attend more than 80% of classes**.
     - Overlap showing students who satisfy both conditions.

5. **Contingency Table & Probability Calculations**
   - Create a **contingency table** for group_discussion (Yes/No) vs. final_exam_pass (Pass/Fail).
   - From the table, calculate:
     - **Joint probability** of "Participates in group discussion AND Passes exam".
     - **Marginal probability** of "Passes exam".
     - **Conditional probability** of "Passes exam given participation in group discussion".

6. **Understanding Relationships**
   - Interpret **conditional probability** results in plain language ("intuition" behind the formula).
   - Identify if "participating in group discussions" and "passing exam" are **independent, dependent, or mutually exclusive** events. Justify your answer.

7. **Bayes Theorem Application**
   - Suppose historical data shows:
     - 70% of students who pass had high attendance (>80%).
     - 40% of students who fail also had high attendance.
     - 60% of all students had high attendance.
   - Use **Bayes Theorem** to find the probability that a student **passed the exam given they had high attendance**.

Submission Requirements:

- Clearly show all formulas, substitutions, and step-by-step calculations.
- Include **tables, diagrams, and charts** wherever applicable.
- Provide a **final summary** explaining what factors most affect the probability of passing.

Expected Time: 4 Hours