1. **Define Data Science and Artificial Intelligence. How are they different?**

* Data Science is an interdisciplinary subject that involves collecting, cleaning, and analyzing data to get meaningful insights.
* Artificial Intelligence is the branch of computer science that aims to develop a system or machine that could do the tasks that human does.
* Data Science focuses on extracting knowledge and insights from data to generate reports, visualizations, and predictions. On the other hand, Artificial Intelligence (AI) focuses on building systems that can mimic human behavior, such as making decisions. AI uses predictive models and automated decision-making based on trained datasets. The insights and patterns discovered by data scientists are often used by AI developers to create and train intelligent models.

1. **What is the intersection between AI and Data Science?**

* The intersection between the AI and Data Science is mentioned below:

1. Machine Learning: Data Science use it to build a predictive models whereas AI use it to give machines ability to learn from data.
2. Model Training and Evaluation: Data scientists collect, clean and analyze data to build model that AI engineers deploy in their projects.
3. Big data and analytics: Both data scientists and AI engineers requires tools like python, TensorFlow, dashboards and so on for handling the large datasets.
4. **List three real-world use cases where both are applied together.**
5. **List and explain three types of file formats commonly used to store datasets.**
6. **Define:**

**Population**

**Sample**

**Variable**

1. **Explain four levels of measurement (nominal, ordinal, interval, ratio) with examples.**
2. **Define and differentiate between:**

**Mean**

**Median**

**Mode**

1. **Describe any two probability distributions (e.g., Gaussian, Binomial, Poisson) with real-world use cases.**
2. **Define and distinguish:**

**Type I Error (False Positive)**

**Type II Error (False Negative)**

1. **What is a confusion matrix? Explain with an example.**
2. **Explain the concepts of:**

**Null Hypothesis**

**Alternative Hypothesis**

1. **What is the purpose of a chi-square test? When is it used?**
2. **Describe a real-world example where hypothesis testing would be necessary in business or healthcare.**
3. **What is the difference between Numpy arrays and Pandas DataFrames?**
4. **What is a missing value and how can you handle it in Pandas?**
5. **Use this sample dataset and apply the following (Practical):**

data = {

'Name': ['Alice', 'Bob', 'Charlie', 'David', 'Eva'],

'Age': [25, 30, np.nan, 22, 28],

'Salary': [50000, 54000, 58000, 52000, np.nan],

'Department': ['HR', 'Finance', 'HR', 'IT', 'Finance']

}

1. Fill missing values with:

Mean for Age

Median for Salary

Filter employees with Salary > 52000

Group by Department and show average Salary

Add a new column Tax = Salary × 0.1 (10% tax)

1. Find the employee with the maximum salary
2. What is SQL? Why is it important in data science?
3. Define and differentiate between:

INNER JOIN vs. LEFT JOIN

1. Create a table (Practical):

CREATE TABLE employees (

id INTEGER PRIMARY KEY,

name TEXT,

department TEXT,

salary INTEGER,

experience INTEGER

);

Insert at least 5 rows, then write SQL queries to:

Select all employees in the Finance department

Get employees with experience > 3 years

Calculate average salary by department

Find the employee(s) with salary greater than the overall average

Join with a new table departments(id, name, location) to fetch department details

Apply following operations in pandas DataFrame (Practical):

# Students Table

students = pd.DataFrame({

'student\_id': [1, 2, 3, 4, 5],

'name': ['Alice', 'Bob', 'Charlie', 'David', 'Eve'],

'class': ['10A', '10B', '10A', '10C', '10B']

})

# Scores Table

scores = pd.DataFrame({

'student\_id': [3, 4, 5, 6],

'math\_score': [88, 92, 75, 85],

'science\_score': [90, 85, 80, 70]

})

Apply inner join, left join, right join, outer join

Using the left join fill missing values in math\_score and science\_score with 0 and create a new column total\_score(math\_score+science\_score)