**CHALLANGE 5**

**TASK 1 : Provide the data dictionary for the given data.**

1. **Student Name**: The name of the student.
2. **University**: The university where the student studies.
3. **Package Offered**: The salary or package offered to the student.
4. **Type of Company**: The industry or type of company the student is placed in.
5. **Student Native Country**: The country of origin of the student.
6. **Prior Work Experience in years**: The number of years of work experience the student had before admission.
7. **Domain**: The specific domain or field the student is associated with (e.g., Finance, IT, Pharmaceutical).
8. **Passport Number**: The passport number of the student.
9. **Date of Birth**: The birthdate of the student.
10. **Date of Admission**: The date when the student was admitted to the university.
11. **Date of Placement**: The date when the student was placed or got a job.

**TASK 1.2: Determine how many rows and columns are present in the given data**

1. Student Name
2. University
3. Package Offered
4. Type of Company
5. Student Native Country
6. Prior Work Experience in years
7. Domain
8. Passport Number
9. Date of Birth
10. Dates (Admission and Placement)

**TASK 2: Identify whether there are any missing values in the data. If so, specify which columns have missing values**

1. **Student Name**: No missing values observed as per the provided data.
2. **University**: No missing values observed.
3. **Package Offered**: No missing values observed.
4. **Type of Company**: No missing values observed.
5. **Student Native Country**: No missing values observed.
6. **Prior Work Experience in years**: There's an inconsistency in one row where the value is missing. (Row for Atharv Mahajan under the "Prior Work Experience in years" column)
7. **Domain**: No missing values observed.
8. **Passport Number**: No missing values observed.
9. **Date of Birth**: No missing values observed.
10. **Dates (Admission and Placement)**: No missing values observed.

**TASK 2.2: Determine if there are any duplicate rows in the data. If duplicates exist, report how many**

1. **Sort the Data**: Sort your dataset based on all columns. This step is optional but can help in visually identifying duplicates.
2. **Scan for Duplicates**: Compare each row with the subsequent rows to check for identical entries across all columns.
3. **Count the Duplicates**: As you compare rows, keep track of any duplicates you find.

**TASK 3:** **Find the maximum age of the students in the given dataset**

|  |  |  |
| --- | --- | --- |
| Rudra Verma | 25-02-1990 33 |  |
| Aayansh Sharma | 17-02-1989 34 |  |

**TASK 3.2:** **. Identify the type of company in which the highest number of graduates from the three universities got placed**

Type of Company Harvard Graduates Boston Graduate Washington Graduates

Finance 1 2 4 IT 1 0 2 Pharmaceutical 0 1 0

**TASK 3.3: Display the packages offered by the companies in increasing order.**

| **Company** | **Packages Offered (in increasing order)** |
| --- | --- |
| Harvard Company | $2,30,000 |
| Washington Company | $40,000, $80,000, $89,700 |
| Boston Company | $45,000, $90,000 |

**TASK 3.4: Calculate the mean value for the column "Prior Work Experience in Years."**

the mean value for the column "Prior Work Experience in Years." IS 4

**CHALLENGE 5.2**

**TASK 1:** **Provide the data dictionary for the given data.**

1. **Student Name**: The name of the student.
2. **University**: The university where the student studied.
3. **Package Offered**: The salary package offered to the student.
4. **Type of Company**: The industry or sector of the company where the student was placed.
5. **Student Native Country**: The native country of the student.
6. **Prior Work Experience in Years**: The number of years of work experience the student had before placement.
7. **Domain**: The field or domain in which the student was placed.
8. **Passport Number**: The passport number of the student.
9. **Date of Birth**: The birth date of the student.
10. **Date of Admission**: The date when the student was admitted to the university.
11. **Date of Placement**: The date when the student was placed in a company.

**TASK 1.2:** **How many rows and columns present in the complete data?**

it appears there are 12 columns and 49 rows of information. Each row represents a student's details across various attributes, and there are a total of 49 such entries in this dataset.

**TASK 2 : Check for missing values in the dataset and identify the columns in which they occur**.

Column Name Missing Values

University 0 Package Offered 0 Type of Company 0 Student Native Country 0 Prior Work Experience in years 1 Domain 0 Passport Number 0 Date of Birth 0 Date of Admission 0 Date of Placement 1

**TASK 2.2: 2. Are there any duplicate rows in the dataset, and if yes, how many?**

I don't have direct access to your dataset, so I can't identify duplicate rows or provide a specific count without analysing the data. However, identifying duplicate rows involves comparing each row with others in the dataset to find identical entries

**TASK 3 : Determine the maximum age of the student in the complete data**

| **Student Name** | **Date of Birth** | **Current Date** | **Calculated Age** |
| --- | --- | --- | --- |
| Rudra Verma | 25-02-1990 | 21-12-2023 | 33 |
| Aayansh Sharma | 17-02-1989 | 21-12-2023 | 34 |

**TASK 3.2: Identify the type of company in which the highest number of graduates from the complete data got placed**

| **Type of Company** | **Number of Graduates** |
| --- | --- |
| Finance | 21 |
| IT | 14 |
| Pharmaceutical | 14 |
| Bank | 4 |

**TASK 3.3: Display the salary packages offered by companies in increasing order**

| **Company** | **Salary Package** |
| --- | --- |
| IT | $10,000 - $150,000 |
| Bank | $40,000 - $89,700 |
| Pharmaceutical | $40,000 - $150,000 |
| Finance | $45,000 - $2,30,000 |

**TASK 3.4: Find the exact mean value for the "Prior Work Experience in Years" column**

The mean value for the "Prior Work Experience in Years" column is 4

**TASK 4:** **Write the following prompt for the complete employee data in the chat window of ChatGPT Provide the data dictionary for the given data. (Copy the full data and paste it in ChatGPT)**

!

**ChatGPT**

The message you submitted was too long, please reload the conversation and submit something shorter.

**TASK 4:** **Write your interpretation about ChatGPT for this data.**

!

**ChatGPT**

The message you submitted was too long, please reload the conversation and submit something shorter.