**Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Roll No: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**1- Differentiate between provenance and verifiable? (3)**

**2- What is the difference between Matplotlib and Seaborn? (3)**

***Dataset: Employee.csv***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **RegNo** | **Name** | **Department** | **Salary** | **Age** | **Bloodgroup** | **Scale** |
| **01** | **Aslam** | **BBA** | **25000** | **50** | **B+** | **17** |
| **02** | **Ali** | **BBA** | **50000** | **42** | **O+** | **18** |
| **03** | **Khan** | **BBA** | **75000** | **35** | **B-** | **N/A** |
| **04** | **Haider** | **CS** | **99000** |  | **A+** | **17** |
| **05** | **Alam** | **CS** | **65000** |  | **AB+** | **18** |
| **07** | **Khan** | **BBA** | **75000** | **35** | **B-** | **N/A** |
| **06** | **John** | **CS** | **45000** | **36** | **B+** | **18** |

**Employeedetails.csv**

|  |  |  |  |
| --- | --- | --- | --- |
| **RegNo** | **Age** | **Bloodgroup** | **Scale** |
| **01** | **50** | **B+** | **17** |
| **02** | **42** | **O+** | **18** |
| **03** | **35** | **B-** | **N/A** |
| **04** | **46** | **A+** | **17** |
| **05** | **45** | **AB+** | **18** |
| **07** | **35** | **B-** | **N/A** |
| **06** | **36** | **B+** | **18** |

**Do the following using PETL / Pandas / Visualization: (2\*7 = 14)**

Import petl

Import pandas

Table1 = fromcsv(“*Employee.csv”)*

Table2 = fromcsv(“Employeedetails.csv*”)*

**1- Salary is string, convert to integer**

Table1 = convert(Table1, “Salary”, int)

**2- Sort both the tables on RegNo**

Table1 = sort(Table1, “RegNo”)

Table2 = sort(Table2, “RegNo”)

**3- Combine Employee and Employeedetails**

Table3 = join(Table1, Table2, key = “RegNo”)

Table.tocsv(“combine.csv”)

**4- Remove the age column**

Data = pd.read\_csv(“combine.csv”)

Data = Data.drop (“Age”, axis=1)

**5- Fill the null value of scale with median**

Data["Scale"].fillna(Data["Scale "].median(), inplace=True)

**6- Make a line chart of salary and scale**

import matplotlib.pyplot as plt

plt.plot(Data[“Salary”], Data[“Scale”])

**7- Make histogram of age**

**(Since I didn’t mention that these questions are linked or not, both the following are correct)**

1- Since Age is deleted so no histogram

**OR**

2- plt.histData[“Age”])