Time: 1 hour

In each of the questions, there are input tables which you can either create directly in R or Python or you can 1st create it in excel or spreadsheet or any other format and then import it in the software.

You have to use either R or Python for all the questions.

1. Given below are 3 input data tables:

Data X

|  |  |
| --- | --- |
| Name | Age |
| RAM | 30 |
| SAM | 28 |
| TAM | 26 |

Data **Y**

|  |  |
| --- | --- |
| Name | Age |
| Ram | 30 |
| Nam | 20 |

Data Z

|  |  |
| --- | --- |
| Name | Salary |
| Ram | 30000 |
| Sam | 20008 |
| Kam | 29006 |

In above data tables, the data-entry guy was not consistent with letters, i.e. he used capital letters and small letters interchangeably.

1. Please append the data X and Y.

Once appended, please remove duplicate entries from the output data.

1. Please integrate the tables in a way to get following output:

|  |  |  |
| --- | --- | --- |
| Name | Age | Salary |
| Ram | 30 | 30000 |
| Sam | 28 | 20008 |

1. Please integrate the tables in a way to get following output:

|  |  |  |
| --- | --- | --- |
| Name | Age | Salary |
| Ram | 30 | 30000 |
| Sam | 28 | 20008 |
| Tam | 26 |  |
| Nam | 20 |  |
| Kam |  | 29006 |

(blanks or . or null means same in above illustration)

1. Please integrate the tables in a way to get following output:

|  |  |  |
| --- | --- | --- |
| Name | Age | Salary |
| Tam | 26 |  |
| Nam | 20 |  |
| Kam |  | 29006 |

1. Given below is the input data table:

|  |  |  |
| --- | --- | --- |
| Name | Salary | Dob |
| Ram Sharma | 55000 | 20aug1990 |
| Tam joshi | 85000 | 19nov1993 |
| Raka Thomas | 44000 | 10aug1980 |
| Kaka joshi | 35000 |  |

Create a variable Age from the variable dob. If some person’s age is finally missing, impute it as 100.

Create a quality variable called salary\_remark on basis of this logic:  If someone’s salary is less than 40000, then call him/her as **Okay**, if it’s more than 40000 but below 50000 then call it **Growing** and remaining call it as **Good enough**.

1. Given below is the input data table:

Data X

|  |  |  |
| --- | --- | --- |
| Customerid | accountid | Balance |
| 100 | 1000 | 5000 |
| 200 | 2000 | 7000 |
| 100 | 3000 | 12900 |
| 250 | 2500 | 11000 |
| 100 | 4000 | 5500 |
| 200 | 3300 | -1000 |
| 200 | 1200 | -0.90 |

Please create the  following output dataset from X.

|  |  |  |
| --- | --- | --- |
| Customerid | Num\_accounts | Tot\_balance |
| 100 | 3 | 23400 |
| 200 | 3 | 5999.1 |
| 250 | 1 | 11000 |

1. Given below are the 3 input data tables:

Data X: list of candidates with their scores. Candidates are from 2 sections A and B

|  |  |  |
| --- | --- | --- |
| **Name** | **Section** | **Score** |
| Ram kishen | A | 68 |
| Vikas Suri | B | 79 |
| Vinay Sharma | B | 80 |
| Vimal jain | A | 79 |

Data Y: another list of candidates with their scores. Candidates are from 2 sections A and B

|  |  |  |
| --- | --- | --- |
| **Name** | **Section** | **Score** |
| Viki taneja | A | 90 |
| Tiki verma | B | 35 |
| amit Kumar | A | 70 |
| RAMAN SINGH | A | 65 |
| RAM SINGH | B | 50 |

Data Z: another list of candidates with their scores. Candidates are from sections A, B and C

|  |  |  |
| --- | --- | --- |
| **Name** | **section** | **Score** |
| Aap Kumar | C | 80 |
| Sushil modi | A | 78 |
| aamir khan | C | 67 |
| Azad Khanna | A | 92 |
| preeti kumari | B | 56 |
| riTa joShi | C | 70 |

Final output dataset required is:

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **section** | **Score** | **CumulativeScore** |
| Raman Singh | A | 65 | 65 |
| Ram Kishen | A | 68 | 133 |
| Amit Kumar | A | 70 | 203 |
| Sushil Modi | A | 78 | 281 |
| Vimal Jain | A | 79 | 360 |
| Azad Khanna | A | 92 | 452 |
| Viki Taneja | A | 90 | 542 |
| Tiki Verma | B | 35 | 35 |
| Ram Singh | B | 50 | 85 |
| Preeti Kumari | B | 56 | 141 |
| Vikas Suri | B | 79 | 220 |
| Vinay Sharma | B | 80 | 300 |
| Aamir Khan | C | 67 | 67 |
| Rita Joshi | C | 70 | 137 |
| Aap Kumar | C | 80 | 217 |

1. Table given below shows the scores obtained by students in maths olympiad

|  |  |  |  |
| --- | --- | --- | --- |
| Class | Section | Student | Marks |
| 8 | A | Ramu | 80 |
| 9 | A | Kamu | 78 |
| 7 | B | Tamu | 77 |
| 8 | A | Namu | 89 |
| 7 | A | LaMu | 90 |
| 7 | A | Vamu | 67 |
| 7 | B | Xamu | 80 |
| 9 | A | Yamu | 88 |
| 8 | B | Ramu | 65 |
| 8 | C | Ramu | 99 |
| 8 | C | Hamu | 39 |
| 8 | B | Famu | 65 |
| 7 | A | Gamu | 45 |
| 7 | A | Namu | 30 |
| 7 | A | Namu | 98 |
| 9 | B | Pamu | 90 |

We need the following output table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Class | Section | Student | Marks | sect\_cum | csect\_cum |
| 7 | A | Gamu | 45 | 45 | 45 |
| 7 | A | LaMu | 90 | 213 | 135 |
| 7 | A | Namu | 30 | 243 | 165 |
| 7 | A | Namu | 98 | 430 | 263 |
| 7 | A | Vamu | 67 | 577 | 330 |
| 7 | B | Tamu | 77 | 297 | 77 |
| 7 | B | Xamu | 80 | 377 | 157 |
| 8 | A | Namu | 89 | 332 | 89 |
| 8 | A | Ramu | 80 | 510 | 169 |
| 8 | B | Famu | 65 | 65 | 65 |
| 8 | B | Ramu | 65 | 220 | 130 |
| 8 | C | Hamu | 39 | 39 | 39 |
| 8 | C | Ramu | 99 | 138 | 138 |
| 9 | A | Kamu | 78 | 123 | 78 |
| 9 | A | Yamu | 88 | 665 | 166 |
| 9 | B | Pamu | 90 | 155 | 90 |