Instructions:

Please read the instructions very carefully.

1. This is an **INDIVIDUAL WORK** unless specified otherwise. Students are not allowed to share their answers (actual coding) but are allowed to discuss with each other to solve the problems.

2. The tasks will be checked during the practical session itself therefore students will be assessed based on participation and their answers.

3. To start:

    a. Create a new folder called **StudentID-P1**. Please change student id to your own personal ID and change the 1 to the correct practical number.

    b. Create one .ipynb / .py file for each question with the naming convention question1.ipynb, question2.ipynb and so on.

    c. The following information must be included in each file:
        i. Student Name
        ii. Student ID
        iii. Module Code and Title

4. Please note that **ALL FILE AND NAMING CONVENTIONS** must be followed.

5. The **GREEN** coloured font in the sample output represents an input from the user.

6. The **BLUE** coloured font in the sample output represents a dynamic output.

7. Please note that the coloured fonts will vary on the values specified. In other words, they are just **SAMPLES** only.

8. All tasks must be completed within the session. Students are given enough time to complete the tasks listed.

9. Once completed, please create a zip file with the same name as your folder, and upload it to LMS before the end of the session. A submission link will be provided.

10. Students are encouraged to upload their work to their own GitHub account. Students are also encouraged to include the GitHub link in the submission.

11. Students are encouraged to ask questions during the practical if they encountered a problem.

Question 1:

Handling missing data on Class Grades Dataset:

a. Import class grades' csv and store data in DataFrames and provide a summary of the dataset

Sample Output:

Summary
Row count = 99
No.of Attributes = 6
Mean value of Final attributes = 70

Note: More information can be added to the summary such as highest and lowest value of each attribute.

Output all the rows:

Prefix	Assignment	Tutorial	Midterm	TakeHome	Final
5	57.14	34.09	64.38	51.48	52.5
8	95.05	105.49	67.5	99.07	68.33
…	…	…	…		…

Note: The title of attributes should be the index labels.

b. Identify missing values in the dataset

Sample Output:

Missing values in TakeHome = 2 missing values
Missing values in Final= 1 missing value
Total missing values = 3

Prefix	Assignment	Tutorial	Midterm	TakeHome	Final
7	72.85	86.85	60	na	56.11
8	63.4	86.21	63.12	72.78	na
6	90.74	89.64	61.25	90	na
…	…	…	…		…

c. Replace missing values with constant values

Sample Output:

Constant value use is: 60

Prefix	Assignment	Tutorial	Midterm	TakeHome	Final
7	72.85	86.85	60	60	56.11
8	63.4	86.21	63.12	72.78	60
6	90.74	89.64	61.25	90	60
…	…	…	…		…

Note: Constant value for each attribute can be different.

d. Replace missing values with mean, median and mode

Sample Output:

Mean value for TakeHome Attribute: 65
Mean value for Final Attribute: 60

Prefix	Assignment	Tutorial	Midterm	TakeHome	Final
7	72.85	86.85	60	65	56.11
8	63.4	86.21	63.12	72.78	60
6	90.74	89.64	61.25	90	60
…	…	…	…		…

Median value for TakeHome Attribute: 70
Median value for Final Attribute: 66

Prefix	Assignment	Tutorial	Midterm	TakeHome	Final
7	72.85	86.85	60	70	56.11
8	63.4	86.21	63.12	72.78	66
6	90.74	89.64	61.25	90	66
…	…	…	…		…

Mode value for TakeHome Attribute: 70
Mode value for Final Attribute: 70

Prefix	Assignment	Tutorial	Midterm	TakeHome	Final
7	72.85	86.85	60	70	56.11
8	63.4	86.21	63.12	72.78	70
6	90.74	89.64	61.25	90	70
…	…	…	…		…

Question 2:

Applying discretization and binning to Class Grades Dataset.

a. Create bins with the values of 20, 40, 60, 80, and 100. Only apply these bins on Final attribute.

Sample Output:

Prefix	Assignment	Tutorial	Midterm	TakeHome	Final	FinalR
5	57.14	34.09	64.38	51.48	52.5	60
8	95.05	105.49	67.5	99.07	68.33	60
8	95.9	99.99	95.62	105.56	102.22	100
…	…	…	…		…	

Note: Add the bins to dataframe as new attribute

b. Add labels for each bin. The labels are as follows:

Bins	**Labels**
20	E
40	D
60	C
80	B
100	A

Sample Output:

Prefix	Assignment	Tutorial	Midterm	TakeHome	Old_Final	Final	Labels
5	57.14	34.09	64.38	51.48	52.5	60	C
8	95.05	105.49	67.5	99.07	68.33	60	C
8	95.9	99.99	95.62	105.56	102.22	100	A
…	…	…	…		…		

Note: Add the labels to dataframe as new attribute.

Question 3 (Question 2 must be completed before proceeding):

Applying data transformation to Class Grades Dataset (Use data modified in Question 2).

a. Transform labels into new labels. The following table shows the new labels.

Labels	New_Labels
E	Fail
D	Pass
C	Merit
B	Merit
A	Distinction

Sample Output:

Prefix	Assignment	Tutorial	Midterm	TakeHome	Final	Labels	NewLabels
5	57.14	34.09	64.38	51.48	60	C	Merit
8	95.05	105.49	67.5	99.07	60	C	Merit
8	95.9	99.99	95.62	105.56	100	A	Distinction
…	…	…	…				

Note: Add the labels to dataframe as new attribute.

b. Transform Assignment values into the range of 0 and 1.

Sample Output:

Prefix	Assignment
5	0.57
8	0.95
8	0.95
…	…

Hint: Divide assignment values with 100. Apply transform using map(lambda x: …).

- END OF FILE -