PRACTICE SESSION

Download the student dataset from Moodle

Perform the following activities on Google Collab or Jupyter Notebook

1. Import pandas library

2. Read the csv file to a dataframe student

3. When you read the file you get index numbers automatically, try to read the file without the index numbers and name the dataframe as student_no_index

	sid	sname	marks	age	Enrollment Date	Gender	Country
0	101.0	Alex	50.0	30.0	12/08/2021	Male	Belgium
1	102.0	Reena	50.0	30.0	12/08/2021	Female	NaN
2	103.0	Teena	52.0	40.0	12/08/2021	Female	NaN
3	104.0	Victoria	56.0	38.0	12/08/2021	Female	NaN
4	105.0	Andrew	65.0	37.0	12/08/2021	Male	NaN
5	106.0	William	67.0	36.0	12/08/2021	Male	NaN
6	107.0	Abdul	88.0	30.0	12/08/2021	Male	NaN
7	108.0	Tim	12.0	43.0	01/02/2022	Male	NaN
8	109.0	Veronica	90.0	30.0	12/08/2021	Female	NaN

4. Now, use student dataframe for perform the following activities

5. Check the datatype of the student dataframe

6. Try to change the datatype

7. You get error message. What is the error message?

   Why are you getting the error? How will resolve it?

8. Check the dataframe for missing values

9. For the 'Country' column, fill the missing values with 'United Kingdom'

10. Now covert all names in the 'sname' column to lowercase

11. convert all column names to lowercase

12. convert the name of Enrollment Date to enrollment_date

13. check for outliers in the marks and age column

14. Replace the outliers and missing values in the marks and age column by their mean values

15. Make sure that all entries in the Gender columns have the either male or female values.

16. Check for missing values

17. Remove all rows with missing values

18. Display the original and modified dataframe

19. Check the datatypes and convert the columns with the correct datatypes

19. select all rows where enrollment_date is 12th August 2021

20. select all rows where marks is less than 50

21. select all rows where name starts with 'A'

22. Create another column Status.

23. Perform the subsequent steps using loops and conditional statements

24. If the enrollment_date is beyond 1st November 2021, add value 'Needs Validation' otherwise the status should be 'Validated'

25. Add another column 'Grades'. If marks are greater than 70, set grade as A, if its between 60 and 70, set grade as B, if its between 50 and 60, grade as C, otherwise set Grade as F

26. Use functions, to create code snippet that display the number of students getting grade 'F'

27. Create another function, create appropriate plot to show relationship between marks and age column.

28. Save the notebook

29. Save the dataframe to csv file

30. Upload the notebook to Moodle