

Session 1: Google Colab link

https://colab.research.google.com/drive/1YhWH6HPOM-pGsozSMkd5gRGALfZ_CqkL?usp=sharing

Session 2 : Google Colab link

<https://colab.research.google.com/drive/13qZGny5x5g0Do2j2JmrBMmfTETSssaGr?usp=sharing>

Session 3: Google Colab link

<https://colab.research.google.com/drive/11sD46O2apkTLvMXjTz9MxrxHQf5kO8V#scrollTo=SfZky2dlVzIA>

Dataset 01 for reference:

https://github.com/rasbt/pattern_classification/blob/master/data/wine_data.csv

- usecols=[0,1,2]

Dataset 02: Reference

- <https://github.com/sharmaroshan/Heart-UCI-Dataset/blob/master/heart.csv>

Assignment: Data Preprocessing and Feature Scaling

Part A: File Importing and Display

Q1. Upload a CSV file named student_data.csv which contains the following columns:

Name, Age, Gender, Score, Attendance, Remarks

Tip: Create a small dataset manually in Excel or Google Sheets and download as CSV.

Q2. Read the file into a DataFrame using Pandas and display the first 5 rows.

Part B: Handling Missing Data

Q3. Check for missing values in all columns and print the total count.

Q4. Replace missing values:

- Numerical columns (Age, Score, Attendance) → Fill with mean
- Categorical columns (Gender, Remarks) → Fill with mode

Part C: Categorical Data Encoding

Q5. Convert the Gender column into numeric format using Label Encoding.

Q6. Convert the Remarks column using One-Hot Encoding.

Part D: Feature Scaling

- Q7. Apply Min-Max Scaling on Score and Attendance columns and print the scaled result.
- Q8. Apply Standardization (Z-score) on the same columns and print the result.
- Q9. Apply Max Absolute Scaling on these columns and print the result.
- Q10. Apply L2 Normalization row-wise on the original Score and Attendance columns.

Part E: Summary Statistics

- Q11. Calculate and display the mean and standard deviation for all numerical columns.
- Q12. Calculate the correlation matrix for numerical features in the dataset and explain what it means.