Data Preprocessing

Aim of the Experiment. The main aim of this experiment is to preprocess the given dataset. The database is created and is available in the file sample.csv.

Start coding or generate with AI.

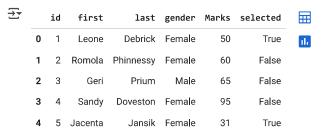
Load the data from "sample.csv", apply Label Encoding, a scaling technique, and Binarization to the data, and display the results after each transformation.

Data loading

Load the data from "sample.csv" into a dataframe.

```
import pandas as pd

df = pd.read_csv('preprocessing.csv')
df.head()
```



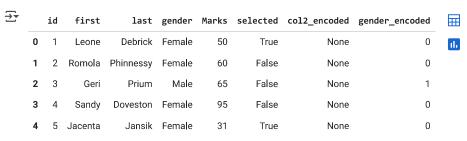
Next steps: Generate code with df View recommended plots New interactive sheet

from sklearn.preprocessing import LabelEncoder

df['gender_encoded'] = label_encoder.fit_transform(df['gender'])

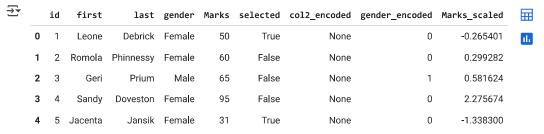
display(df.head())

label_encoder = LabelEncoder()



from sklearn.preprocessing import StandardScaler

```
scaler = StandardScaler()
df['Marks_scaled'] = scaler.fit_transform(df[['Marks']])
display(df.head())
```



from sklearn.preprocessing import Binarizer

binarizer = Binarizer(threshold=50)

df['Marks_binarized'] = binarizer.fit_transform(df[['Marks']])

display(df.head())

| → | | id | first | last | gender | Marks | selected | col2_encoded | gender_encoded | Marks_scaled | Marks_binarized | |
|----------|---|----|---------|-----------|--------|-------|----------|--------------|----------------|--------------|-----------------|-----|
| | 0 | 1 | Leone | Debrick | Female | 50 | True | None | 0 | -0.265401 | 0 | ılı |
| | 1 | 2 | Romola | Phinnessy | Female | 60 | False | None | 0 | 0.299282 | 1 | |
| | 2 | 3 | Geri | Prium | Male | 65 | False | None | 1 | 0.581624 | 1 | |
| | 3 | 4 | Sandy | Doveston | Female | 95 | False | None | 0 | 2.275674 | 1 | |
| | 4 | 5 | Jacenta | Jansik | Female | 31 | True | None | 0 | -1.338300 | 0 | |