Task 5

Dataset overview 👍

The used in this task is performance of students in exams , it contains 8 columns and 1000 rows .

1. **Loading data set** 👍

---

[27]:
```
import numpy as np
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
```

[2]:
```
df=pd.read_csv(r"C:\Users\Ghulam Mustafa\Downloads\archive (2)\exams.csv")
df.head()
```

[2]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score
0	male	group A	high school	standard	completed	67	67	63
1	female	group D	some high school	free/reduced	none	40	59	55
2	male	group E	some college	free/reduced	none	59	60	50
3	male	group B	high school	standard	none	77	78	68
4	male	group E	associate's degree	standard	completed	78	73	68

2. **Checking Null valuese in data.**

[4]:
```
df.isnull().sum()
```

[4]:
```
gender                         0
race/ethnicity                 0
parental level of education    0
lunch                          0
test preparation course        0
math score                     0
reading score                  0
writing score                  0
dtype: int64
```

3.Creating tow new features Performance and overall score on the basis of three score columns .

Creating tow new feature name perfomance and overalll score

[42]:
```
df['overall_score'] = (df['math score'] + df['reading score'] + df['writing score']) / 3
```

[43]:
```
def performance(row):
    avg = (row['math score'] + row['reading score'] + row['writing score']) / 3

    if avg <= 50:
        return 'Average'
    elif avg <= 70:
        return 'Good'
    elif avg <= 86:
        return 'Excellent'
    else:
        return 'Topper'

df['Performance'] = df.apply(performance, axis=1)
```

4.Encoding Performance column .

Encoding Performance colum

[44]:
```
df['encoded_performance'] = df['Performance'].map({'Bade': 0, 'Topper':1,'Excellent':2,'Good': 3})
```

[45]:
```
df.head()
```

[45]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	Performance	encoded_performance	overall_score	Scaled_overall_score
0	male	group A	high school	standard	completed	67	67	63	Good	3.0	65.666667	0.561702
1	female	group D	some high school	free/reduced	none	40	59	55	Good	3.0	51.333333	0.378723
2	male	group E	some college	free/reduced	none	59	60	50	Good	3.0	56.333333	0.442553
3	male	group B	high school	standard	none	77	78	68	Excellent	2.0	74.333333	0.672340
4	male	group E	associate's degree	standard	completed	78	73	68	Excellent	2.0	73.000000	0.655319

5.Scaling overall score column using MInMaxScaler module of sikit learn

Scaling the overall score column using Min Max module of sklearn library

[46]:
```
obj=MinMaxScaler()
df['Scaled_overall_score']=obj.fit_transform(df[['overall_score']])
```

[47]:
```
df.head()
```

[47]:

	gender	race/ethnicity	parental level of education	lunch	test preparation course	math score	reading score	writing score	Performance	encoded_performance	overall_score	Scaled_overall_score
0	male	group A	high school	standard	completed	67	67	63	Good	3.0	65.666667	0.561702
1	female	group D	some high school	free/reduced	none	40	59	55	Good	3.0	51.333333	0.378723
2	male	group E	some college	free/reduced	none	59	60	50	Good	3.0	56.333333	0.442553
3	male	group B	high school	standard	none	77	78	68	Excellent	2.0	74.333333	0.672340
4	male	group E	associate's degree	standard	completed	78	73	68	Excellent	2.0	73.000000	0.655319