```
In [1]: import pandas as pd
```

```
In [2]: df=pd.read_excel("research_student.xlsx")
```

```
In [3]: df.shape
```

Out[3]: (223, 24)

```
In [6]: df.head(5)
```

Out[6]:

	Branch	Marks[10th]	Marks[12th]	Gender	Board[10th]	Board[12th]	Category	GPA 1	Rank
0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	CIVIL	77.57	64.6	Male	BSEB Patna	BSEB Patna	OBC	6.29	44718.0
2	CSE	86.40	71.8	Male	CBSE	CBSE	GEN	6.47	24222.0
3	CSE	88.14	78.0	Male	ICSE	ICSE	GEN	7.35	24723.0
4	CSE	65.40	59.8	Female	CBSE	CBSE	ST	6.41	232157.0

5 rows × 24 columns

```
In [7]: df.tail(5)
```

Out[7]:

	Branch	Marks[10th]	Marks[12th]	Gender	Board[10th]	Board[12th]	Category	GPA 1	
218	PROD	91.2	80.6	Male	ICSE	CBSE	GEN	74.70	39
219	PROD	79.4	63.2	Male	CENTRAL BOARD OF SECONDARY EDUCATION	CENTRAL BOARD OF SECONDARY EDUCATION	OBC	6.71	114
220	PROD	87.4	83.2	Male	CBSE	CBSE	GEN	7.18	40
221	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
222	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	279

5 rows × 24 columns

In [8]: `df.describe()`

Out[8]:

	Marks[10th]	Marks[12th]	GPA 1	Rank	Normalized Rank	CGPA	Current Back
count	220.000000	220.000000	220.000000	221.000000	220.000000	220.000000	220.000000
mean	84.307455	77.783227	7.534409	42312.122172	14.725877	7.199468	0.077273
std	8.519507	9.044172	4.602384	38503.691510	12.528090	0.700186	0.267633
min	53.700000	56.800000	5.760000	11814.000000	4.219286	5.890000	0.000000
25%	79.000000	71.550000	6.710000	23949.000000	8.544196	6.637500	0.000000
50%	86.550000	79.100000	7.180000	30080.000000	10.694107	7.125000	0.000000
75%	91.000000	85.250000	7.760000	41527.000000	14.718036	7.702500	0.000000
max	96.600000	96.500000	74.700000	279839.000000	99.942500	9.010000	1.000000

In [9]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 223 entries, 0 to 222
Data columns (total 24 columns):
 #   Column                 Non-Null Count   Dtype
---  ------                 --------------   -----
 0   Branch                 220 non-null     object
 1   Marks[10th]            220 non-null     float64
 2   Marks[12th]            220 non-null     float64
 3   Gender                 220 non-null     object
 4   Board[10th]            220 non-null     object
 5   Board[12th]            220 non-null     object
 6   Category               220 non-null     object
 7   GPA 1                  220 non-null     float64
 8   Rank                   221 non-null     float64
 9   Normalized Rank        220 non-null     float64
 10  CGPA                   220 non-null     float64
 11  Current Back           220 non-null     float64
 12  Ever Back              219 non-null     float64
 13  GPA 2                  220 non-null     float64
 14  GPA 3                  220 non-null     float64
 15  GPA 4                  220 non-null     float64
 16  GPA 5                  220 non-null     float64
 17  GPA 6                  220 non-null     float64
 18  Olympiads Qualified    220 non-null     float64
 19  Technical Projects     220 non-null     float64
 20  Tech Quiz              220 non-null     float64
 21  Engg. Coaching         220 non-null     float64
 22  NTSE Scholarships      220 non-null     float64
 23  Miscellany Tech Events 220 non-null     float64
dtypes: float64(19), object(5)
memory usage: 41.9+ KB
```

In [10]: `df.isnull()`

Out[10]:

	Branch	Marks[10th]	Marks[12th]	Gender	Board[10th]	Board[12th]	Category	GPA 1	Rank
0	True	True	True	True	True	True	True	True	True
1	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...
218	False	False	False	False	False	False	False	False	False
219	False	False	False	False	False	False	False	False	False
220	False	False	False	False	False	False	False	False	False
221	True	True	True	True	True	True	True	True	True
222	True	True	True	True	True	True	True	True	False

223 rows × 24 columns

In [11]: `df.isnull().sum()`

Out[11]:
```
Branch                   3
Marks[10th]              3
Marks[12th]              3
Gender                   3
Board[10th]              3
Board[12th]              3
Category                 3
GPA 1                    3
Rank                     2
Normalized Rank          3
CGPA                     3
Current Back             3
Ever Back                4
GPA 2                    3
GPA 3                    3
GPA 4                    3
GPA 5                    3
GPA 6                    3
Olympiads Qualified      3
Technical Projects       3
Tech Quiz                3
Engg. Coaching           3
NTSE Scholarships        3
Miscellany Tech Events   3
dtype: int64
```

In [12]: df=df.drop([0,221,222])

In [13]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 220 entries, 1 to 220
Data columns (total 24 columns):
 #   Column               Non-Null Count   Dtype
---  ------               --------------   -----
 0   Branch               220 non-null     object
 1   Marks[10th]          220 non-null     float64
 2   Marks[12th]          220 non-null     float64
 3   Gender               220 non-null     object
 4   Board[10th]          220 non-null     object
 5   Board[12th]          220 non-null     object
 6   Category             220 non-null     object
 7   GPA 1                220 non-null     float64
 8   Rank                 220 non-null     float64
 9   Normalized Rank      220 non-null     float64
 10  CGPA                 220 non-null     float64
 11  Current Back         220 non-null     float64
 12  Ever Back            219 non-null     float64
 13  GPA 2                220 non-null     float64
 14  GPA 3                220 non-null     float64
 15  GPA 4                220 non-null     float64
 16  GPA 5                220 non-null     float64
 17  GPA 6                220 non-null     float64
 18  Olympiads Qualified  220 non-null     float64
 19  Technical Projects   220 non-null     float64
 20  Tech Quiz            220 non-null     float64
 21  Engg. Coaching       220 non-null     float64
 22  NTSE Scholarships    220 non-null     float64
 23  Miscellany Tech Events  220 non-null  float64
dtypes: float64(19), object(5)
memory usage: 43.0+ KB
```

In [14]: df.shape

Out[14]: (220, 24)

In [15]: `df.dtypes`

Out[15]:
```
Branch                   object
Marks[10th]             float64
Marks[12th]             float64
Gender                   object
Board[10th]              object
Board[12th]              object
Category                 object
GPA 1                   float64
Rank                    float64
Normalized Rank         float64
CGPA                    float64
Current Back            float64
Ever Back               float64
GPA 2                   float64
GPA 3                   float64
GPA 4                   float64
GPA 5                   float64
GPA 6                   float64
Olympiads Qualified     float64
Technical Projects      float64
Tech Quiz               float64
Engg. Coaching          float64
NTSE Scholarships       float64
Miscellany Tech Events  float64
dtype: object
```

In [16]: `df=df.fillna(0)`

```
In [17]: df.isnull().sum()
```

```
Out[17]: Branch                     0
         Marks[10th]                0
         Marks[12th]                0
         Gender                     0
         Board[10th]                0
         Board[12th]                0
         Category                   0
         GPA 1                      0
         Rank                       0
         Normalized Rank            0
         CGPA                       0
         Current Back               0
         Ever Back                  0
         GPA 2                      0
         GPA 3                      0
         GPA 4                      0
         GPA 5                      0
         GPA 6                      0
         Olympiads Qualified        0
         Technical Projects         0
         Tech Quiz                  0
         Engg. Coaching             0
         NTSE Scholarships          0
         Miscellany Tech Events     0
         dtype: int64
```

```
In [18]: df.columns
```

```
Out[18]: Index(['Branch', 'Marks[10th]', 'Marks[12th]', 'Gender', 'Board[10th]',
                'Board[12th]', 'Category', 'GPA 1', 'Rank', 'Normalized Rank', 'CGPA',
                'Current Back', 'Ever Back', 'GPA 2', 'GPA 3', 'GPA 4', 'GPA 5',
                'GPA 6', 'Olympiads Qualified', 'Technical Projects', 'Tech Quiz',
                'Engg. Coaching', 'NTSE Scholarships', 'Miscellany Tech Events'],
               dtype='object')
```

```
In [20]: imp_columns=['Marks[10th]', 'Marks[12th]', 'GPA 1', 'Rank', 'Normalized Rank',
         'Current Back', 'Ever Back', 'GPA 2', 'GPA 3', 'GPA 4', 'GPA 5',
         'GPA 6', 'Olympiads Qualified', 'Technical Projects', 'Tech Quiz',
         'Engg. Coaching', 'NTSE Scholarships', 'Miscellany Tech Events']
```

In [21]: `df[imp_columns]`

Out[21]:

	Marks[10th]	Marks[12th]	GPA 1	Rank	Normalized Rank	CGPA	Current Back	Ever Back	GPA 2	GPA 3
1	77.57	64.6	6.29	44718.0	15.970714	6.02	1.0	4.0	6.12	5.94
2	86.40	71.8	6.47	24222.0	8.650714	6.10	1.0	7.0	6.12	5.88
3	88.14	78.0	7.35	24723.0	8.829643	6.65	1.0	1.0	6.35	6.54
4	65.40	59.8	6.41	232157.0	82.913214	6.09	1.0	11.0	6.00	5.71
5	81.00	74.0	6.80	23252.0	8.304286	6.13	1.0	0.0	6.06	5.88
...	...	...	...	...	...	...	...	...	...	...
216	78.80	66.0	6.35	100000.0	35.714286	6.44	0.0	3.0	6.35	6.06
217	91.00	81.0	7.00	36706.0	13.109286	7.07	0.0	0.0	6.65	6.47
218	91.20	80.6	74.70	39792.0	14.211429	7.36	0.0	2.0	71.80	67.20
219	79.40	63.2	6.71	114306.0	40.823571	6.89	0.0	0.0	6.12	6.41
220	87.40	83.2	7.18	40000.0	14.285714	7.69	0.0	0.0	6.88	7.06

220 rows × 19 columns

In [22]: `from sklearn.preprocessing import StandardScaler`

In [23]: `scaler=StandardScaler()`

In [26]: `df[imp_columns]=scaler.fit_transform(df[imp_columns])`

In [27]: `df[imp_columns]`

Out[27]:

	Marks[10th]	Marks[12th]	GPA 1	Rank	Normalized Rank	CGPA	Current Back	Ever Back
1	-0.792630	-1.460973	-0.271000	0.099590	0.099590	-1.688349	3.455601	0.968301
2	0.246178	-0.663065	-0.231801	-0.486029	-0.486029	-1.573833	3.455601	2.101418
3	0.450881	0.024023	-0.040160	-0.471714	-0.471714	-0.786536	3.455601	-0.164817
4	-2.224374	-1.992912	-0.244867	5.455168	5.455168	-1.588147	3.455601	3.612242
5	-0.389107	-0.419259	-0.159935	-0.513744	-0.513744	-1.530889	3.455601	-0.542523
...	...	...	...	...	...	...	...	...
216	-0.647926	-1.305824	-0.257934	1.679129	1.679129	-1.087140	-0.289385	0.590595
217	0.787347	0.356485	-0.116381	-0.129332	-0.129332	-0.185327	-0.289385	-0.542523
218	0.810876	0.312156	14.626933	-0.041157	-0.041157	0.229793	-0.289385	0.212889
219	-0.577339	-1.616122	-0.179535	2.087885	2.087885	-0.442988	-0.289385	-0.542523
220	0.363824	0.600290	-0.077181	-0.035214	-0.035214	0.702171	-0.289385	-0.542523

220 rows × 19 columns

In [29]:
```python
from sklearn.preprocessing import LabelEncoder
encoding_list = ['Branch','Gender','Board[10th]','Board[12th]','Category']
df[encoding_list] = df[encoding_list].apply(LabelEncoder().fit_transform)
```

In [30]: `df[encoding_list]`

Out[30]:

	Branch	Gender	Board[10th]	Board[12th]	Category
1	0	1	11	11	1
2	1	1	18	19	0
3	1	1	25	29	0
4	1	0	18	19	3
5	1	1	18	19	0
...	...	...	...	...	...
216	6	1	18	19	3
217	6	1	18	19	0
218	6	1	25	19	0
219	6	1	20	22	1
220	6	1	18	19	0

220 rows × 5 columns

In [31]: `df.info()`

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 220 entries, 1 to 220
Data columns (total 24 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   Branch                220 non-null    int32
 1   Marks[10th]           220 non-null    float64
 2   Marks[12th]           220 non-null    float64
 3   Gender                220 non-null    int32
 4   Board[10th]           220 non-null    int32
 5   Board[12th]           220 non-null    int32
 6   Category              220 non-null    int32
 7   GPA 1                 220 non-null    float64
 8   Rank                  220 non-null    float64
 9   Normalized Rank       220 non-null    float64
 10  CGPA                  220 non-null    float64
 11  Current Back          220 non-null    float64
 12  Ever Back             220 non-null    float64
 13  GPA 2                 220 non-null    float64
 14  GPA 3                 220 non-null    float64
 15  GPA 4                 220 non-null    float64
 16  GPA 5                 220 non-null    float64
 17  GPA 6                 220 non-null    float64
 18  Olympiads Qualified   220 non-null    float64
 19  Technical Projects    220 non-null    float64
 20  Tech Quiz             220 non-null    float64
 21  Engg. Coaching        220 non-null    float64
 22  NTSE Scholarships     220 non-null    float64
 23  Miscellany Tech Events 220 non-null   float64
dtypes: float64(19), int32(5)
memory usage: 38.7 KB
```

In [ ]: