

```
In [8]: import pandas as pd
import matplotlib.pyplot as plt # !pip install matplotlib / pip terminal
```

```
In [9]: df = pd.read_csv("Academic-Performance-Dataset.csv")
```

```
In [10]: df
```

	Rollno	Name	Gender	Branch	Attendance	Phy_marks	Che_marks	EM1_marks	PPS_marks	SME_marks	Total Marks	Percentage
0	1	Mohammed	M	Comp	72.0	62.0	98.0	63.0	89.0	36.0	368	73.6
1	2	Reyansh	M	IT	58.0	62.0	83.0	83.0	88.0	34.0	350	70.0
2	3	Aarav	M	IT	57.0	-20.0	100.0	NaN	56.0	36.0	192	38.4
3	4	Atharv	M	IT	60.0	89.0	83.0	70.0	33.0	23.0	298	59.6
4	5	Vivaan	M	Comp	85.0	90.0	NaN	78.0	23.0	56.0	247	49.4
5	6	Advik	M	ENTC	94.0	99.0	84.0	100.0	56.0	99.0	438	87.6
6	7	Ansh	M	ENTC	98.0	88.0	95.0	81.0	78.0	78.0	420	84.0
7	8	Ishaan	M	ENTC	75.0	66.0	51.0	83.0	-99.0	76.0	192	38.4
8	9	Dhruv	M	ENTC	63.0	NaN	NaN	97.0	56.0	55.0	208	41.6
9	10	Siddharth	M	ENTC	96.0	67.0	78.0	95.0	NaN	98.0	338	67.6
10	11	Vihaan	M	ENTC	82.0	54.0	70.0	88.0	55.0	56.0	323	64.6
11	12	NaN	M	IT	75.0	64.0	67.0	71.0	66.0	87.0	355	71.0
12	13	Aarush	M	IT	67.0	56.0	81.0	NaN	90.0	55.0	282	56.4
13	14	Leo	M	IT	98.0	-34.0	70.0	94.0	77.0	66.0	273	54.6
14	15	Maryam	F	IT	64.0	87.0	60.0	90.0	65.0	90.0	392	78.4
15	16	Saanvi	F	Comp	66.0	90.0	95.0	67.0	99.0	77.0	428	85.6
16	17	Zaranev	F	Comp	93.0	54.0	NaN	75.0	90.0	65.0	284	56.8
17	18	Inaya	F	Comp	74.0	67.0	93.0	93.0	87.0	99.0	439	87.8
18	19	Aarya	F	Comp	72.0	88.0	84.0	81.0	80.0	45.0	378	75.6
19	20	NaN	F	Comp	53.0	76.0	81.0	93.0	65.0	23.0	338	67.6

```
In [11]: # 2
df.isnull().sum()
```

```
Out[11]: Rollno      0
Name            2
Gender          0
Branch          0
Attendance      0
Phy_marks       1
Che_marks       3
EM1_marks       2
PPS_marks       1
SME_marks       0
Total Marks     0
Percentage      0
dtype: int64
```

```
In [12]: df.dropna() # it remove the null rows technique 1
```

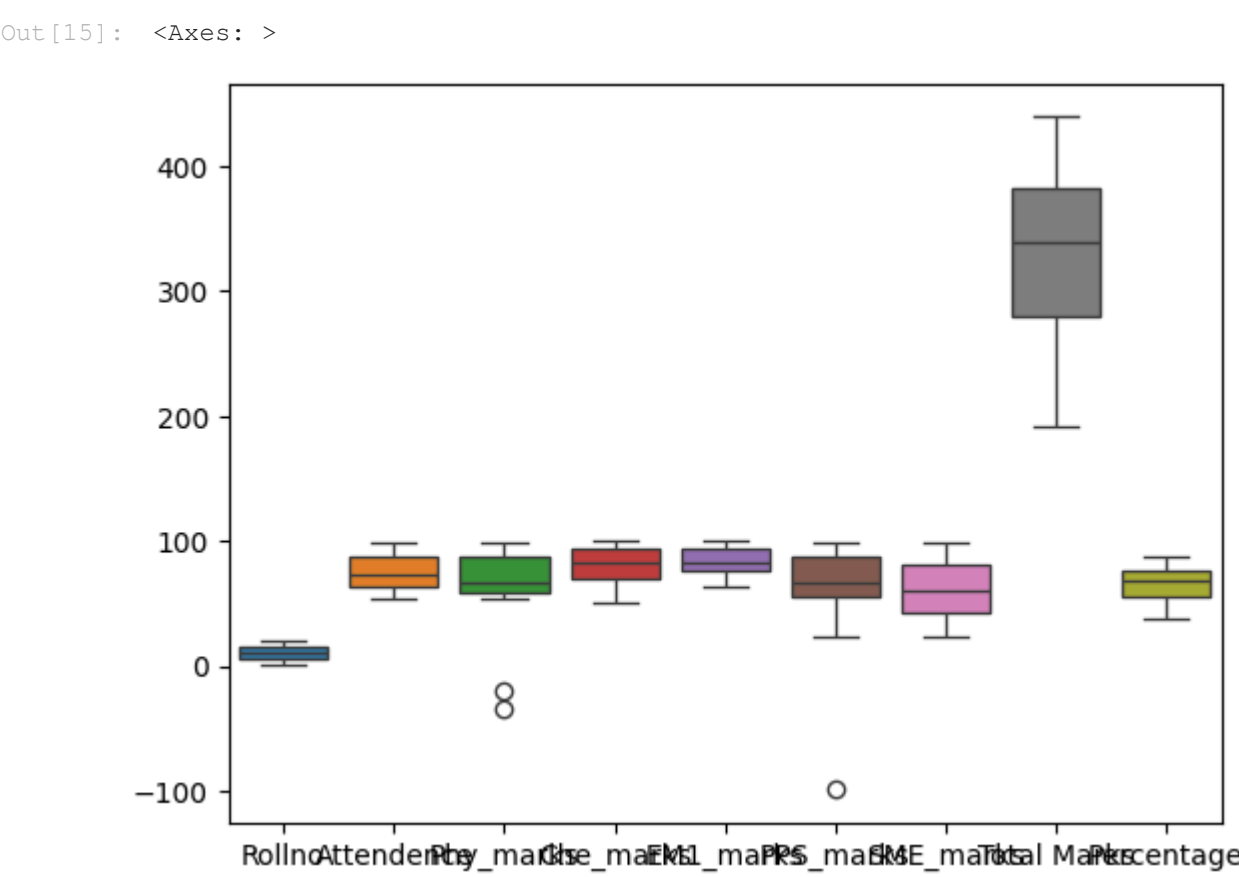
	Rollno	Name	Gender	Branch	Attendance	Phy_marks	Che_marks	EM1_marks	PPS_marks	SME_marks	Total Marks	Percentage
0	1	Mohammed	M	Comp	72.0	62.0	98.0	63.0	89.0	36.0	368	73.6
1	2	Reyansh	M	IT	58.0	62.0	83.0	83.0	88.0	34.0	350	70.0
3	4	Atharv	M	IT	60.0	89.0	83.0	70.0	33.0	23.0	298	59.6
5	6	Advik	M	ENTC	94.0	99.0	84.0	100.0	56.0	99.0	438	87.6
6	7	Ansh	M	ENTC	98.0	88.0	95.0	81.0	78.0	78.0	420	84.0
7	8	Ishaan	M	ENTC	75.0	66.0	51.0	83.0	-99.0	76.0	192	38.4
10	11	Vihaan	M	ENTC	82.0	54.0	70.0	88.0	55.0	56.0	323	64.6
13	14	Leo	M	IT	98.0	-34.0	70.0	94.0	77.0	66.0	273	54.6
14	15	Maryam	F	IT	64.0	87.0	60.0	90.0	65.0	90.0	392	78.4
15	16	Saanvi	F	Comp	66.0	90.0	95.0	67.0	99.0	77.0	428	85.6
17	18	Inaya	F	Comp	74.0	67.0	93.0	93.0	87.0	99.0	439	87.8
18	19	Aarya	F	Comp	72.0	88.0	84.0	81.0	80.0	45.0	378	75.6

```
In [13]: # technique 2
df["Name"] = df["Name"].fillna("Ritesh")
```

```
In [14]: df
```

	Rollno	Name	Gender	Branch	Attendance	Phy_marks	Che_marks	EM1_marks	PPS_marks	SME_marks	Total Marks	Percentage
0	1	Mohammed	M	Comp	72.0	62.0	98.0	63.0	89.0	36.0	368	73.6
1	2	Reyansh	M	IT	58.0	62.0	83.0	83.0	88.0	34.0	350	70.0
2	3	Aarav	M	IT	57.0	-20.0	100.0	NaN	56.0	36.0	192	38.4
3	4	Atharv	M	IT	60.0	89.0	83.0	70.0	33.0	23.0	298	59.6
4	5	Vivaan	M	Comp	85.0	90.0	NaN	78.0	23.0	56.0	247	49.4
5	6	Advik	M	ENTC	94.0	99.0	84.0	100.0	56.0	99.0	438	87.6
6	7	Ansh	M	ENTC	98.0	88.0	95.0	81.0	78.0	78.0	420	84.0
7	8	Ishaan	M	ENTC	75.0	66.0	51.0	83.0	-99.0	76.0	192	38.4
8	9	Dhruv	M	ENTC	63.0	NaN	NaN	97.0	56.0	55.0	208	41.6
9	10	Siddharth	M	ENTC	96.0	67.0	78.0	95.0	NaN	98.0	338	67.6
10	11	Vihaan	M	ENTC	82.0	54.0	70.0	88.0	55.0	56.0	323	64.6
11	12	Ritesh	M	IT	75.0	64.0	67.0	71.0	66.0	87.0	355	71.0
12	13	Aarush	M	IT	67.0	56.0	81.0	NaN	90.0	55.0	282	56.4
13	14	Leo	M	IT	98.0	-34.0	70.0	94.0	77.0	66.0	273	54.6
14	15	Maryam	F	IT	64.0	87.0	60.0	90.0	65.0	90.0	392	78.4
15	16	Saanvi	F	Comp	66.0	90.0	95.0	67.0	99.0	77.0	428	85.6
16	17	Zaranev	F	Comp	93.0	54.0	NaN	75.0	90.0	65.0	284	56.8
17	18	Inaya	F	Comp	74.0	67.0	93.0	93.0	87.0	99.0	439	87.8
18	19	Aarya	F	Comp	72.0	88.0	84.0	81.0	80.0	45.0	378	75.6
19	20	Ritesh	F	Comp	53.0	76.0	81.0	93.0	65.0	23.0	338	67.6

```
In [15]: #scan outlier
import seaborn as sns # !pip install seaborn / pip for terminal
sns.boxplot(df)
```



```
In [16]: # IQR method
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

```
In [17]: dataset = [12,35,40,60,70,80,25,41,20,30,145,55,68,75,56,150]
sorted(dataset)

quantile1, quantile3 = np.percentile(dataset , (25,75))
print(quantile1, quantile3)

33.75 58.199999999999996
```

```
In [18]: iqr_value = quantile3 - quantile1
print(iqr_value)

24.449999999999996
```

```
In [19]: lower_bound_value = quantile1 - (1.5*iqr_value)
upper_bound_value = quantile3 + (1.5*iqr_value)
print(upper_bound_value, lower_bound_value)

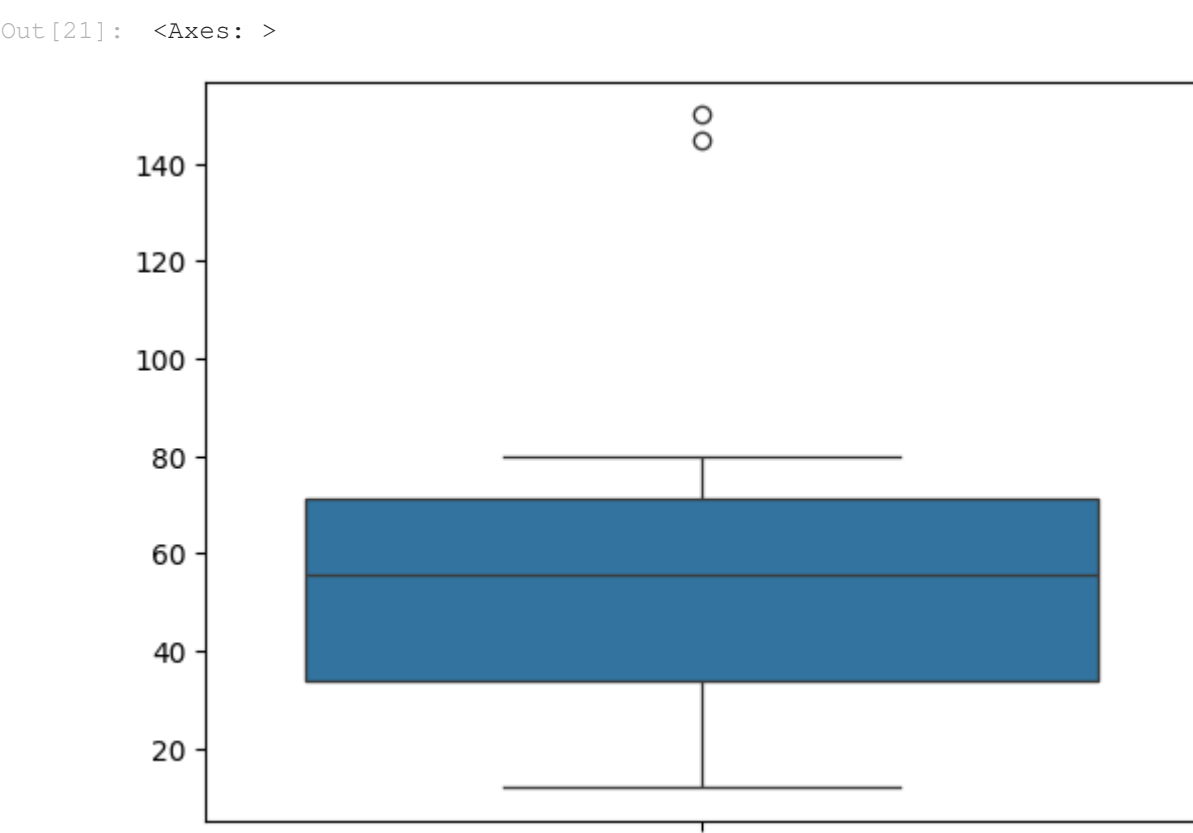
94.875 -2.924999999999997
```

```
In [20]: outlier = []
for x in dataset:
    if (x>upper_bound_value) or (x<lower_bound_value):
        outlier.append(x)

print(outlier)

[145, 150]
```

```
In [21]: import seaborn as sns
sns.boxplot(dataset)
```



```
In [26]: # 3
from sklearn.preprocessing import MinMaxScaler # !pip install scikit-learn / pip for terminal

mms = MinMaxScaler() # - Add parentheses here

df[["Phy_marks", "Che_marks", "EM1_marks", "PPS_marks", "SME_marks"]] = mms.fit_transform(
    df[["Phy_marks", "Che_marks", "EM1_marks", "PPS_marks", "SME_marks"]]
)
```

```
In [27]: df
```

	Rollno	Name	Gender	Branch	Attendance	Phy_marks	Che_marks	EM1_marks	PPS_marks	SME_marks	Total Marks	Percentage
0	1	Mohammed	M	Comp	72.0	0.721805	0.959184	0.000000	0.949495	0.171053	368	73.6
1	2	Reyansh	M	IT	58.0	0.721805	0.653061	0.540541	0.944444	0.144737	350	70.0
2	3	Aarav	M	IT	57.0	0.105263	1.000000	NaN	0.782828	0.171053	192	38.4
3	4	Atharv	M	IT	60.0	0.924812	0.653061	0.189189	0.666667	0.000000	298	59.6
4	5	Vivaan	M	Comp	85.0	0.932331	NaN	0.405405	0.616162	0.434211	247	49.4
5	6	Advik	M	ENTC	94.0	1.000000	0.673469	1.000000	0.782828	1.000000	438	87.6
6	7	Ansh	M	ENTC	98.0	0.917293	0.897959	0.486486	0.893939	0.723684	420	84.0
7	8	Ishaan	M	ENTC	75.0	0.751880	0.000000	0.540541	0.000000	0.697368	192	38.4
8	9	Dhruv	M	ENTC	63.0	NaN	NaN	0.918919	0.782828	0.421053	208	41.6
9	10	Siddharth	M	ENTC	96.0	0.759398	0.551020	0.864865	NaN	0.986842	338	67.6
10	11	Vihaan	M	ENTC	82.0	0.661654	0.387755	0.675676	0.777778	0.434211	323	64.6
11	12	Ritesh	M	IT	75.0	0.736842	0.326531	0.216216	0.833333	0.842105	355	71.0
12	13	Aarush	M	IT	67.0	0.676692	0.612245	NaN	0.954545	0.421053	282	56.4
13	14	Leo	M	IT	98.0	0.000000	0.387755	0.837838	0.888889	0.565789	273	54.6
14	15	Maryam	F	IT	64.0	0.909774	0.183673	0.729730	0.828283	0.881579	392	78.4
15	16	Saanvi	F	Comp	66.0	0.932331	0.897959	0.108108	1.000000	0.710526	428	85.6
16	17	Zaranev	F	Comp	93.0	0.661654	NaN	0.324324	0.954545	0.552632	284	56.8
17	18	Inaya	F	Comp	74.0	0.759398	0.857143	0.810811	0.939394	1.000000	439	87.8
18	19	Aarya	F	Comp	72.0	0.917293	0.673469	0.486486	0.904040	0.289474	378	75.6

19	20	Ritesh	F	Comp	53.0	0.827068	0.612245	0.810811	0.828283	0.000000	338	67.6
----	----	--------	---	------	------	----------	----------	----------	----------	----------	-----	------

In []: