

```
import pandas as pd
import numpy as np
import matplotlib as plt
import seaborn as sns
```

```
df=pd.read_csv(r"C:\Users\ASUS\Documents\pythonStack\DS_PR\AcademicPerformanc
e_New.csv")
```

```
df.head()
```

	Roll No	NationalITY	WT	DSBDA	AI	Average	Grade	PG
0	1.0	India	72.0	72.0	74.0	72.666667	C	Yes
1	2.0	Japan	69.0	90.0	NaN	53.000000	B	No
2	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	4.0	China	47.0	NaN	44.0	30.333333	D	Yes
4	5.0	India	76.0	78.0	NaN	51.333333	C	No

```
df.describe(include='all')
```

	Roll No	NationalITY	WT	DSBDA	AI	Average	\
count	98.000000	98	91.000000	95.000000	89.000000	98.000000	
unique	NaN	5	NaN	NaN	NaN	NaN	
top	NaN	USA	NaN	NaN	NaN	NaN	
freq	NaN	30	NaN	NaN	NaN	NaN	
mean	50.479592	NaN	61.164835	64.652632	62.235955	58.663265	
std	28.469274	NaN	14.995009	14.893072	16.216590	16.518482	
min	1.000000	NaN	18.000000	26.000000	20.000000	26.000000	
25%	26.250000	NaN	51.000000	54.500000	52.000000	45.250000	
50%	50.500000	NaN	63.000000	66.000000	63.000000	59.333333	
75%	74.750000	NaN	71.000000	74.000000	74.000000	71.916667	
max	99.000000	NaN	97.000000	95.000000	92.000000	91.666667	

	Grade	PG
count	98	98
unique	4	2
top	D	Yes
freq	52	49
mean	NaN	NaN
std	NaN	NaN
min	NaN	NaN
25%	NaN	NaN
50%	NaN	NaN
75%	NaN	NaN
max	NaN	NaN

```
df['WT'].isnull().sum()
```

```
10
```

```
df['WT']=df['WT'].fillna(df['WT'].mean())
```

```
df['AI'].isnull().sum()
```

```

10
df['DSBDA'].isnull().sum()

4

for i in ['DSBDA', 'AI']:
    df[i]=df[i].fillna(df[i].mean())

df['DSBDA'].isnull().sum() ,df['AI'].isnull().sum()

(0, 0)

for i in ['NationalITY', 'Grade']:
    df[i]=df[i].fillna(df[i].mode()[0])

df['NationalITY'].isnull().sum(),df['Grade'].isnull().sum()

(0, 0)

from scipy.stats import zscore

num_cols=['WT', 'AI', 'DSBDA', 'Average']

z_score=df[num_cols].apply(zscore)

z_score

      WT      AI      DSBDA  Average
0  7.578533e-01  0.769438  0.506292     NaN
1  5.480217e-01  0.000000  1.746635     NaN
2  4.969810e-16  0.000000  0.000000     NaN
3 -9.907433e-01 -1.192739  0.000000     NaN
4  1.037629e+00  0.000000  0.919740     NaN
..      ...      ...      ...      ...
94  1.247460e+00  0.000000  0.000000     NaN
95  1.177516e+00  1.292685  1.126464     NaN
96  2.682462e-01 -0.015433  0.092844     NaN
97  1.283585e-01  0.507814  0.506292     NaN
98 -2.213608e-01  0.000000  0.161752     NaN

[99 rows x 4 columns]

((z_score)<3).sum()

WT      99
AI      99
DSBDA   99
Average  0
dtype: int64

df['Performance'] = pd.cut(df['Average']
bins=[0, 50, 70, 100], labels=['Low', 'Medium', 'High'])

```

```
Performance=df['Performance'].value_counts()
```

```
Performance.plot(kind='bar',color=['green','blue','orange'])
```

```
<Axes: xlabel='Performance'>
```

