```
import numpy as np
import pandas as pd
```

```
from google.colab import files
uploaded = files.upload()
```

Choose Files StudentsPe…ceTest1.xlsx

- **StudentsPerformanceTest1.xlsx**(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 9301 bytes, last modified: 2/14/2023 - 100% done

Saving StudentsPerformanceTest1.xlsx to StudentsPerformanceTest1.xlsx

```
df = pd.read_excel('StudentsPerformanceTest1.xlsx')
```

```
df
```

	gender	math score	reading score	writing score	Placement Score	placement offer count	Region
0	female	72	72	74.0	78.0	1	Pune
1	female	69	90	88.0	NaN	2	na
2	female	90	95	93.0	74.0	2	Nashik
3	male	47	57	NaN	78.0	1	Na
4	male	na	78	75.0	81.0	3	Pune
5	female	71	Na	78.0	70.0	4	na
6	male	12	44	52.0	12.0	2	Nashik
7	male	NaN	65	67.0	49.0	1	Pune
8	male	5	77	89.0	55.0	0	NaN

```
df.isnull()
```

	gender	math score	reading score	writing score	Placement Score	placement offer count	Region
0	False	False	False	False	False	False	False
1	False	False	False	False	True	False	False
2	False	False	False	False	False	False	False
3	False	False	False	True	False	False	False
4	False	False	False	False	False	False	False
5	False	False	False	False	False	False	False
6	False	False	False	False	False	False	False
7	False	True	False	False	False	False	False
8	False	False	False	False	False	False	True

```
df.isnull().sum()
```

```
gender                  0
math score              1
reading score           0
writing score           1
Placement Score         1
placement offer count   0
Region                  1
dtype: int64
```

```
df.isna().sum()
```

```
gender                  0
math score              1
reading score           0
writing score           1
Placement Score         1
placement offer count   0
Region                  1
dtype: int64
```

```
df.mean()
```

```
<ipython-input-9-c61f0c8f89b5>:1: FutureWarning: Dropping of nuisance columns in DataFrame reductions (with 'numeric_only=None') is
  df.mean()
writing score           77.000000
```

```
Placement Score          62.125000
placement offer count     1.777778
dtype: float64
```

```
df['writing score'].fillna(77.000000,axis=0,inplace=True)
```

```
df['Placement Score'].fillna(62.125000,axis=0,inplace=True)
```

```
df
```

	gender	math score	reading score	writing score	Placement Score	placement offer count	Region
0	female	72	72	74.0	78.000	1	Pune
1	female	69	90	88.0	62.125	2	na
2	female	90	95	93.0	74.000	2	Nashik
3	male	47	57	77.0	78.000	1	Na
4	male	na	78	75.0	81.000	3	Pune
5	female	71	Na	78.0	70.000	4	na
6	male	12	44	52.0	12.000	2	Nashik
7	male	NaN	65	67.0	49.000	1	Pune
8	male	5	77	89.0	55.000	0	NaN

```
df['placement offer count'].fillna(1.777778,axis=0,inplace=True)
```

```
df
```

	gender	math score	reading score	writing score	Placement Score	placement offer count	Region
0	female	72	72	74.0	78.000	1	Pune
1	female	69	90	88.0	62.125	2	na
2	female	90	95	93.0	74.000	2	Nashik
3	male	47	57	77.0	78.000	1	Na
4	male	na	78	75.0	81.000	3	Pune
5	female	71	Na	78.0	70.000	4	na
6	male	12	44	52.0	12.000	2	Nashik
7	male	NaN	65	67.0	49.000	1	Pune
8	male	5	77	89.0	55.000	0	NaN

```
df.drop('Region',axis=1,inplace=True)
```

```
df
```

	gender	math score	reading score	writing score	Placement Score	placement offer count
0	female	72	72	74.0	78.000	1
1	female	69	90	88.0	62.125	2
2	female	90	95	93.0	74.000	2
3	male	47	57	77.0	78.000	1
4	male	na	78	75.0	81.000	3
5	female	71	Na	78.0	70.000	4
6	male	12	44	52.0	12.000	2
7	male	NaN	65	67.0	49.000	1
8	male	5	77	89.0	55.000	0

```
df2 = df.drop('gender',axis=1)
```

```
df2
```

	reading score	writing score	Placement Score	placement offer count
0	72	74.0	78.000	1
1	90	88.0	62.125	2
2	95	93.0	74.000	2
3	57	77.0	78.000	1
4	78	75.0	81.000	3
5	Na	78.0	70.000	4
6	44	52.0	12.000	2
7	65	67.0	49.000	1
8	77	89.0	55.000	0

```
df2.to_csv('raw_data.csv',index=False)
```