

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

```
In [3]: data = pd.read_csv("C:/Users/shara/OneDrive/Desktop/ML LAB/movies.csv")
```

```
In [9]: data.isnull().sum()
```

```
Out[9]: MOVIES      0
YEAR        644
GENRE       80
RATING     1820
ONE-LINE     0
STARS       0
VOTES     1820
RunTime    2958
Gross      9539
dtype: int64
```

```
In [10]: data_1 = data.dropna(axis=0,how="all")
len(data_1)
# if all values are null
```

```
Out[10]: 9999
```

```
In [11]: data_1 = data.dropna(axis=0,how="any")
len(data_1)
# when any value is null
```

```
Out[11]: 460
```

```
In [13]: data_1.isnull().sum()
```

```
Out[13]: MOVIES      0
YEAR        0
GENRE       0
RATING     0
ONE-LINE    0
STARS      0
VOTES     0
RunTime    0
Gross      0
dtype: int64
```

```
In [14]: data_1 = data.dropna(axis=0,how="any",subset=["GENRE"])
data_1.isnull().sum()
```

```
Out[14]: MOVIES      0
YEAR        603
GENRE       0
RATING     1751
ONE-LINE     0
STARS       0
VOTES     1751
RunTime    2887
Gross      9459
dtype: int64
```

```
In [18]: data_1 = data.drop(["Gross"],axis=1)
data_1.isnull().sum()
```

```
Out[18]: MOVIES      0
YEAR        644
GENRE       80
RATING     1820
ONE-LINE     0
STARS       0
VOTES     1820
RunTime    2958
dtype: int64
```

```
In [19]: data_1["VOTES"] = data_1["VOTES"].fillna("0")
data_1.isnull().sum()
```

```
Out[19]: MOVIES      0
YEAR      644
GENRE     80
RATING    1820
ONE-LINE   0
STARS     0
VOTES     0
RunTime   2958
dtype: int64
```

```
In [22]: meanRT = data_1["RATING"].mean()
meanRT = round(meanRT,1)
print(meanRT)
```

```
6.9
```

```
In [24]: data_1["RATING"] = data_1["RATING"].fillna(meanRT)
data_1.isnull().sum()
```

```
Out[24]: MOVIES      0
YEAR      644
GENRE     80
RATING     0
ONE-LINE   0
STARS     0
VOTES     0
RunTime   2958
dtype: int64
```

```
In [25]: meanRunT = data_1["RunTime"].mean()
meanRunT = round(meanRunT,1)
print(meanRunT)
data_1["RunTime"] = data_1["RunTime"].fillna(meanRunT)
data_1.isnull().sum()
```

```
68.7
```

```
Out[25]: MOVIES      0
YEAR      644
GENRE     80
RATING     0
ONE-LINE   0
STARS     0
VOTES     0
RunTime    0
dtype: int64
```

```
In [26]: data_1 = data_1.dropna(axis=0,how="any",subset=["GENRE"])
data_1.isnull().sum()
```

```
Out[26]: MOVIES      0
YEAR      603
GENRE      0
RATING     0
ONE-LINE   0
STARS     0
VOTES     0
RunTime    0
dtype: int64
```

```
In [27]: data_1["YEAR"] = data_1["YEAR"].fillna("Before 2k")
data_1.isnull().sum()
```

```
Out[27]: MOVIES      0
YEAR      0
GENRE      0
RATING     0
ONE-LINE   0
STARS     0
VOTES     0
RunTime    0
dtype: int64
```