pg_3

```
import pandas as pd

df=pd.read_excel('/content/dataset_program_3.xlsx')
df.head()
#df.shape
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	origin	car name
0	22.0	4	140.0	72.0	2408	19.0	71	1	chevrolet vega (sw)
1	23.0	8	350.0	NaN	3900	17.4	79	1	cadillac eldorado
2	36.0	4	98.0	70.0	2125	17.3	82	1	mercury lynx l

Checking For Single valued Row

```
a=list(df.columns)
for i in a:
  print(len(df[i].unique())<3)
```

```
False
False
False
False
False
False
False
True
False
```

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

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	car name
0	22.0	4	140.0	72.0	2408	19.0	71	chevrolet vega (sw)
1	23.0	8	350.0	NaN	3900	17.4	79	cadillac eldorado
2	36.0	4	98.0	70.0	2125	17.3	82	mercury lynx l
3	26.8	6	173.0	115.0	2700	12.9	79	oldsmobile omega brougham
4	22.5	6	232.0	90.0	3085	17.6	76	amc hornet

Checking and removing null value

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

```
mpg             0
cylinders       0
displacement    12
horsepower      6
weight          0
acceleration    0
model year      0
car name        0
dtype: int64
```

```
a=df['displacement'].mean()
# df['displacement'].median()
# df['displacement'].mode()
b=df['horsepower'].mean()

values = {"displacement": a, "horsepower":b }#250.734,122

df.fillna(value=values,inplace=True)
df.isnull().sum()
```

```
mpg             0
cylinders       0
displacement    0
horsepower      0
weight          0
acceleration    0
model year      0
car name        0
dtype: int64
```

Removing Duplicate rows

```
data=df.duplicated(keep=False)
data
```

```
0       True
1      False
2      False
3      False
4      False
       ...
293     True
294     True
295    False
296    False
297    False
Length: 298, dtype: bool
```

```
data=df[df.duplicated(keep=False)]
data.sort_values(by='mpg')
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	car name
225	11.0	8	250.734266	150.0	4997	14.0	73	chevrolet impala
212	11.0	8	429.000000	208.0	4633	11.0	72	mercury marquis
208	11.0	8	250.734266	150.0	4997	14.0	73	chevrolet impala
49	11.0	8	429.000000	208.0	4633	11.0	72	mercury marquis
115	12.0	8	429.000000	198.0	4952	11.5	73	mercury marquis brougham
...	...	...	...	...	...	...	...	...
99	25.0	4	97.500000	80.0	2126	17.0	72	dodge colt hardtop
251	26.0	4	91.000000	70.0	1955	20.5	71	plymouth cricket
289	26.0	4	91.000000	70.0	1955	20.5	71	plymouth cricket
194	28.0	4	98.000000	80.0	2164	15.0	72	dodge colt (sw)
130	28.0	4	98.000000	80.0	2164	15.0	72	dodge colt (sw)

98 rows × 8 columns

```
data = pd.DataFrame(df)
display(data.drop_duplicates())
```

	mpg	cylinders	displacement	horsepower	weight	acceleration	model year	car name
0	22.0	4	140.0	72.000000	2408	19.0	71	chevrolet vega (sw)
1	23.0	8	350.0	122.284247	3900	17.4	79	cadillac eldorado
2	36.0	4	98.0	70.000000	2125	17.3	82	mercury lynx l
3	26.8	6	173.0	115.000000	2700	12.9	79	oldsmobile omega brougham
4	22.5	6	232.0	90.000000	3085	17.6	76	amc hornet
...	...	...	...	...	...	...	...	...
291	17.7	6	231.0	165.000000	3445	13.4	78	buick regal sport coupe (turbo)
292	17.5	8	318.0	140.000000	4080	13.7	78	dodge magnum xe
295	14.0	8	455.0	225.000000	3086	10.0	70	buick estate wagon (sw)
296	15.0	6	250.0	72.000000	3432	21.0	75	mercury monarch
297	18.6	6	225.0	110.000000	3620	18.7	78	dodge aspen

249 rows × 8 columns