```
#Daksh Gupta
#GH Raisoni Institute of Engineering and Technology
#2nd year
#CSE branch
#MINI PROJECT(EXPLORATORY DATA ANALYSIS)(5+ conclusions)
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
df=pd.read_csv('/content/CardioGoodFitness.csv')
df
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income
0	TM195	18	Male	14	Single	3	4	29562
1	TM195	19	Male	15	Single	2	3	31836
2	TM195	19	Female	14	Partnered	4	3	30699
3	TM195	19	Male	12	Single	3	3	32973
4	TM195	20	Male	13	Partnered	4	2	35247
...	...	...	...	...	...	...	...	...
175	TM798	40	Male	21	Single	6	5	83416
176	TM798	42	Male	18	Single	5	4	89641
177	TM798	45	Male	16	Single	5	5	90886
178	TM798	47	Male	18	Partnered	4	5	104581
179	TM798	48	Male	18	Partnered	4	5	95508

180 rows × 9 columns

```
df.info
```

```
<bound method DataFrame.info of     Product  Age  Gender  Education MaritalStatus  Usage  Fitness  Income  \
0     TM195   18    Male         14        Single      3        4   29562
1     TM195   19    Male         15        Single      2        3   31836
2     TM195   19  Female         14     Partnered      4        3   30699
3     TM195   19    Male         12        Single      3        3   32973
4     TM195   20    Male         13     Partnered      4        2   35247
..      ...  ...     ...        ...           ...    ...      ...     ...
175   TM798   40    Male         21        Single      6        5   83416
176   TM798   42    Male         18        Single      5        4   89641
177   TM798   45    Male         16        Single      5        5   90886
178   TM798   47    Male         18     Partnered      4        5  104581
179   TM798   48    Male         18     Partnered      4        5   95508

     Miles
0      112
1       75
2       66
3       85
4       47
..     ...
175    200
176    200
177    160
178    120
179    180

[180 rows x 9 columns]>
```

```
df.Product.nunique()
```

```
3
```

```
df.Product.unique()
```

```
array(['TM195', 'TM498', 'TM798'], dtype=object)
```

```
df.Age.unique()
```

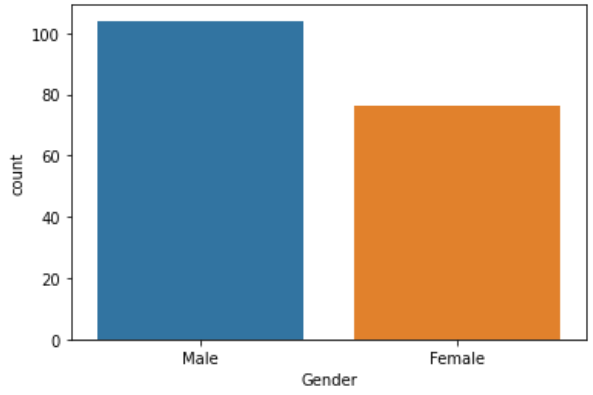
```
array([18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34,
       35, 36, 37, 38, 39, 40, 41, 43, 44, 46, 47, 50, 45, 48, 42])
```

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

```
Product          0
Age              0
Gender           0
Education        0
MaritalStatus    0
Usage            0
Fitness          0
Income           0
Miles            0
dtype: int64
```

```
import seaborn as sns
sns.countplot(df['Gender'])
```

```
/usr/local/lib/python3.8/dist-packages/seaborn/_decorators.py:36: FutureWarning: P
  warnings.warn(
<matplotlib.axes._subplots.AxesSubplot at 0x7f4a9d002a30>
```




```
df.groupby('MaritalStatus').size()
```

```
MaritalStatus
Partnered    107
Single        73
dtype: int64
```

```
df.groupby('Gender').size()
```

```
Gender
Female     76
Male      104
dtype: int64
```

```
df.groupby(['Product','Gender']).size()
```

```
Product  Gender
TM195    Female    40
```

```
          Male      40
TM498     Female    29
          Male      31
TM798     Female     7
          Male      33
dtype: int64
```

```
np.min(df.Education)  #minimum years of education
```

```
12
```

```
np.max(df.Education)   #maximum years of education
```

```
21
```

```
young=np.sum((df.Age>=0)&(df.Age<=20))   #number of people from age 0 to 20
young
```

```
10
```

```
adult=np.sum((df.Age>20)&(df.Age<=40))   #number of people from age 21 to 40
adult
```

```
158
```

```
midaged=np.sum((df.Age>40)&(df.Age<=60))  #number of people from age 41 tp 60
midaged
```

```
12
```

```
old=np.sum((df.Age>60)) #number of people above 60 years
old
```

```
0
```