link for the collab notebook: https://drive.google.com/file/d/1S5MFQvdceClYv4uvFCY7u95u-tEe82IJ/view?usp=sharing

ANALYSING BASIC METRICS:

The dataset given to us have 180 rows and 9 columns which represent the profile of different customers.

Out of the 9 columns ,'Product', 'Gender' and 'MaritalStatus' columns have datatype OBJECT and others have 'integer' datatype.

People with wide range of age groups are there from 8 to 50 years. and a wide variety of different information given in table below.

```
df.describe()
```

	Age	Education	Usage	Fitness	Income	Miles
count	180.000000	180.000000	180.000000	180.000000	180.000000	180.000000
mean	28.788889	15.572222	3.455556	3.311111	53719.577778	103.194444
std	6.943498	1.617055	1.084797	0.958869	16506.684226	51.863605
min	18.000000	12.000000	2.000000	1.000000	29562.000000	21.000000
25%	24.000000	14.000000	3.000000	3.000000	44058.750000	66.000000
50%	26.000000	16.000000	3.000000	3.000000	50596.500000	94.000000
75%	33.000000	16.000000	4.000000	4.000000	58668.000000	114.750000
max	50.000000	21.000000	7.000000	5.000000	104581.000000	360.000000

NON-GRAPHICAL ANALYSIS:

There are a total of 180 customers.Out of which 104 are Males and 76 are Females.

and 107 are Married and 73 are Single.

```
df.Gender.value_counts()
```

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

```
df.MaritalStatus.value_counts()
```

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

Number of people comes into a different fitness rating:

```
df.Fitness.value_counts()
```

```
3    97
5    31
2    26
4    24
1     2
```

No. of People who uses different Treadmill:

```
df.Product.value_counts()
```

```
KP281    80
KP481    60
KP781    40
```

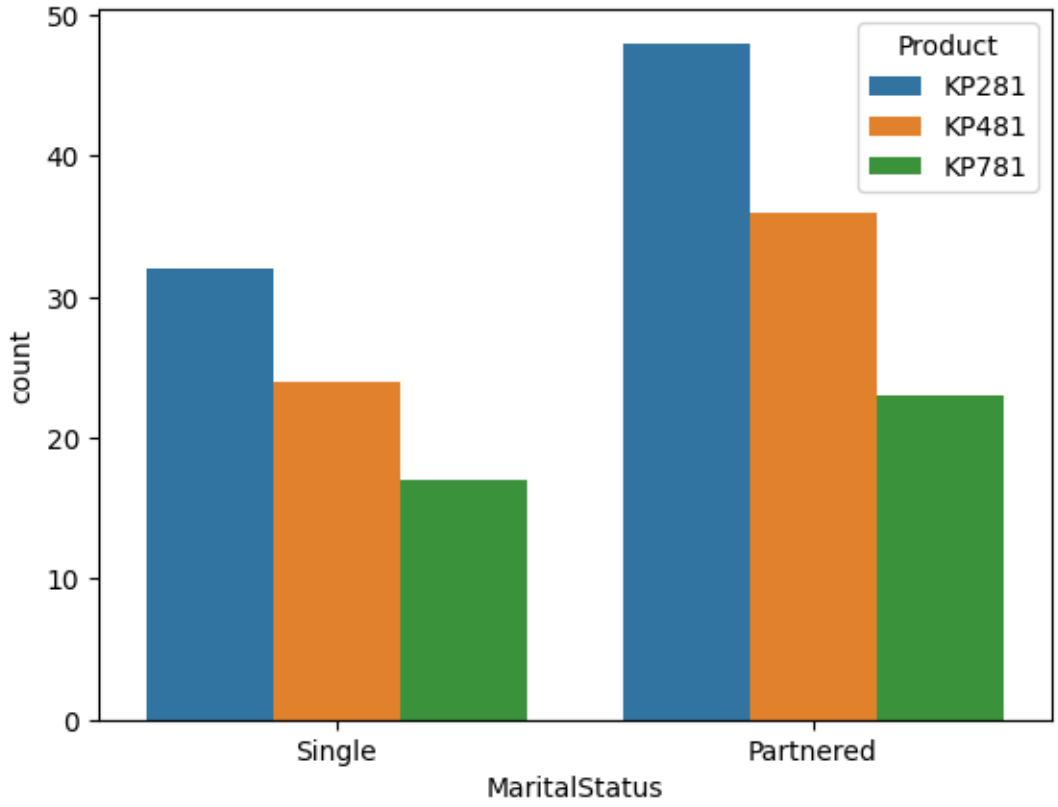
Usage hours of people:

```
df.Usage.value_counts()
```

```
3    69
4    52
2    33
5    17
6     7
7     2
Name: Usage, dtype: int64
```

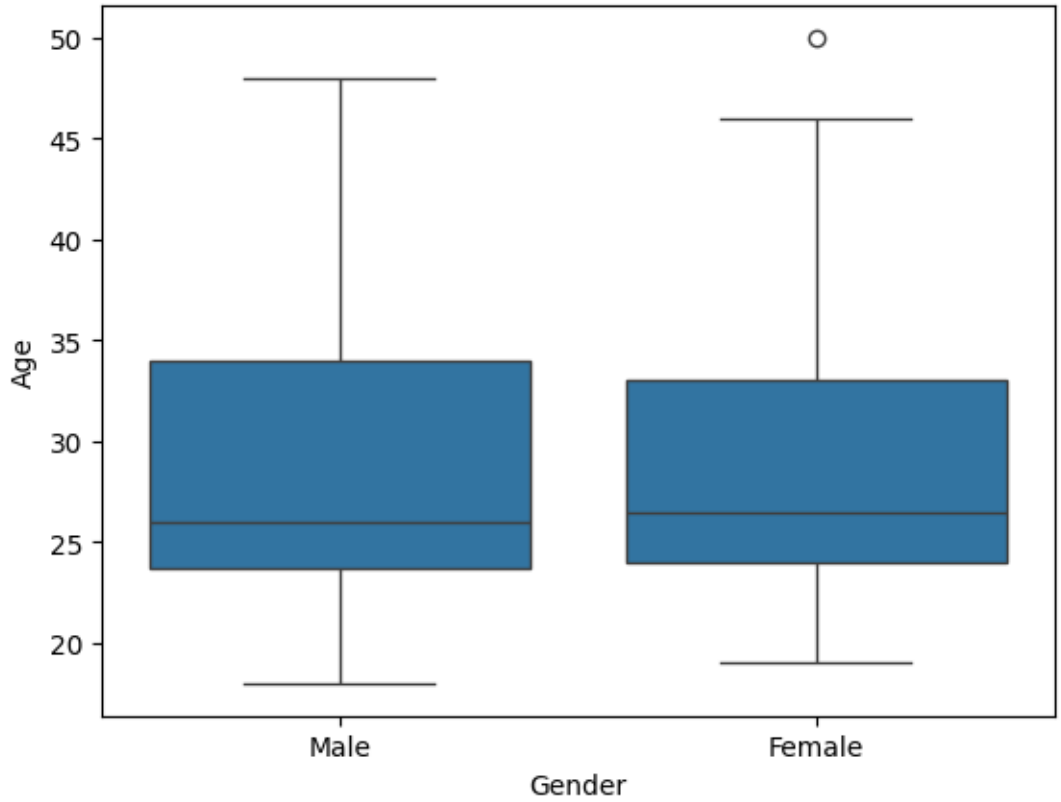
VISUAL ANALYSIS:

#Counts of people who uses different products on the basis of the Marital Status:



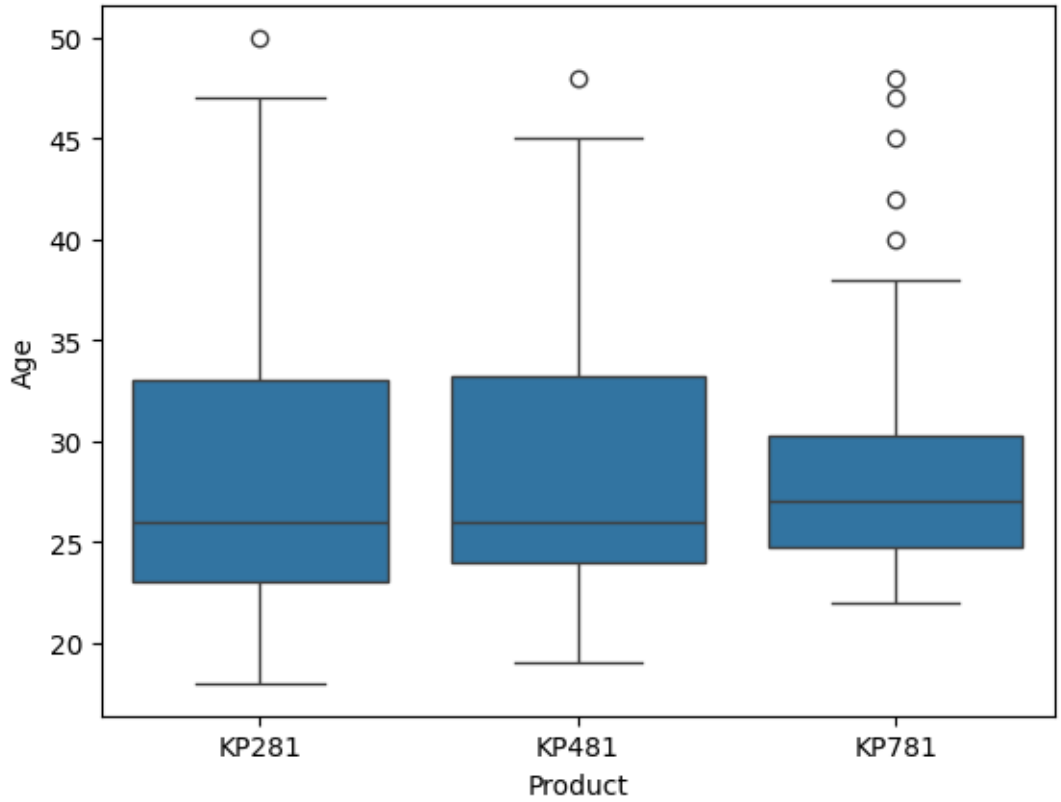
INSIGHTS: We can see that more Married people are buying Treadmills and the most famous treadmill among the 3 is 'KP281' followed by 'KP481' and 'KP781'. as the KP281 is more budget friendly.

#Box plot of Age vs Gender:



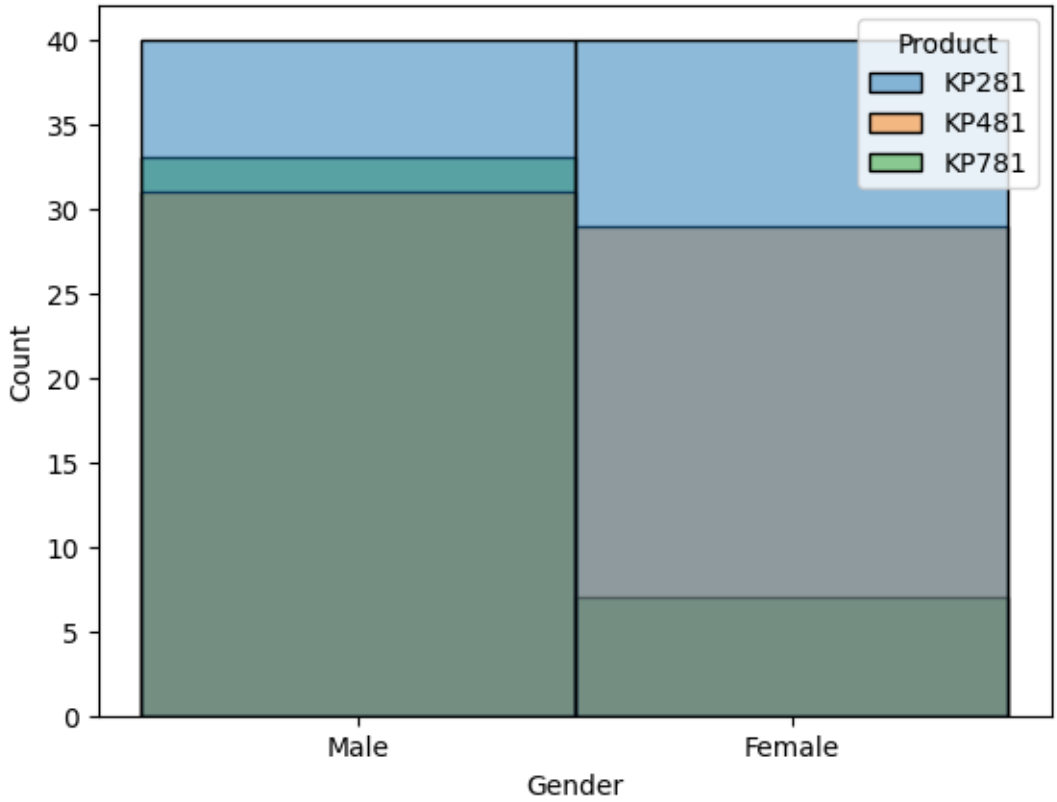
INSIGHTS: here we can see that 50% people are from age group 18-26 years. means that people from 18-26 age are buying more treadmill and people from other age group are are buying less in both the genders 'Male' and 'Female'.

Boxplot of Age vs Product:



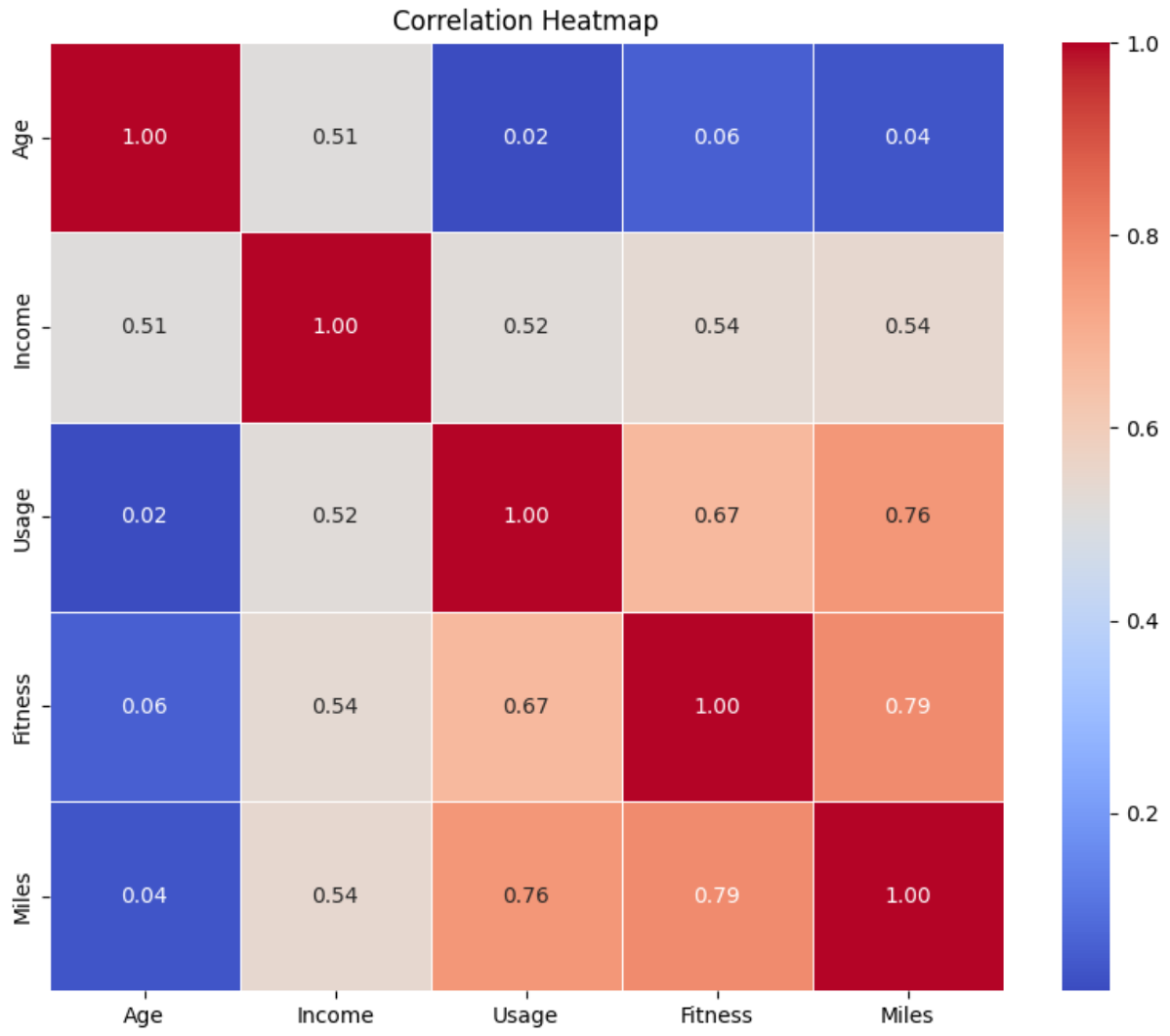
INSIGHTS: we can see that people buying 'KP281' have a wide variety of age group because it is more affordable but age group for people buying 'KP781' are limited.

#Graph of count of gender for different products:



INSIGHTS: Males are buying 'KP481' treadmill less and 'KP281' very high. while in female 'KP481' is the most popular.

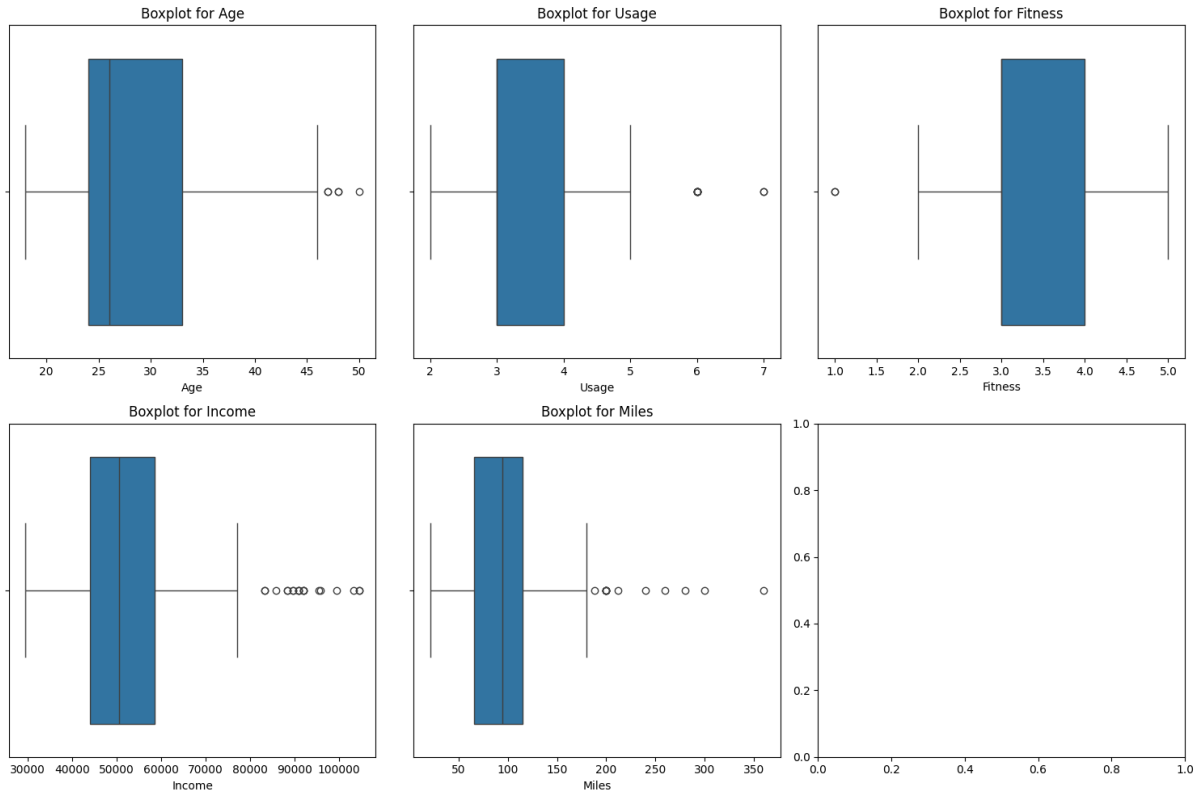
Heatmap of variables:



MISSING VALUES AND OUTLIERS:

There are no missing values in the given dataset.

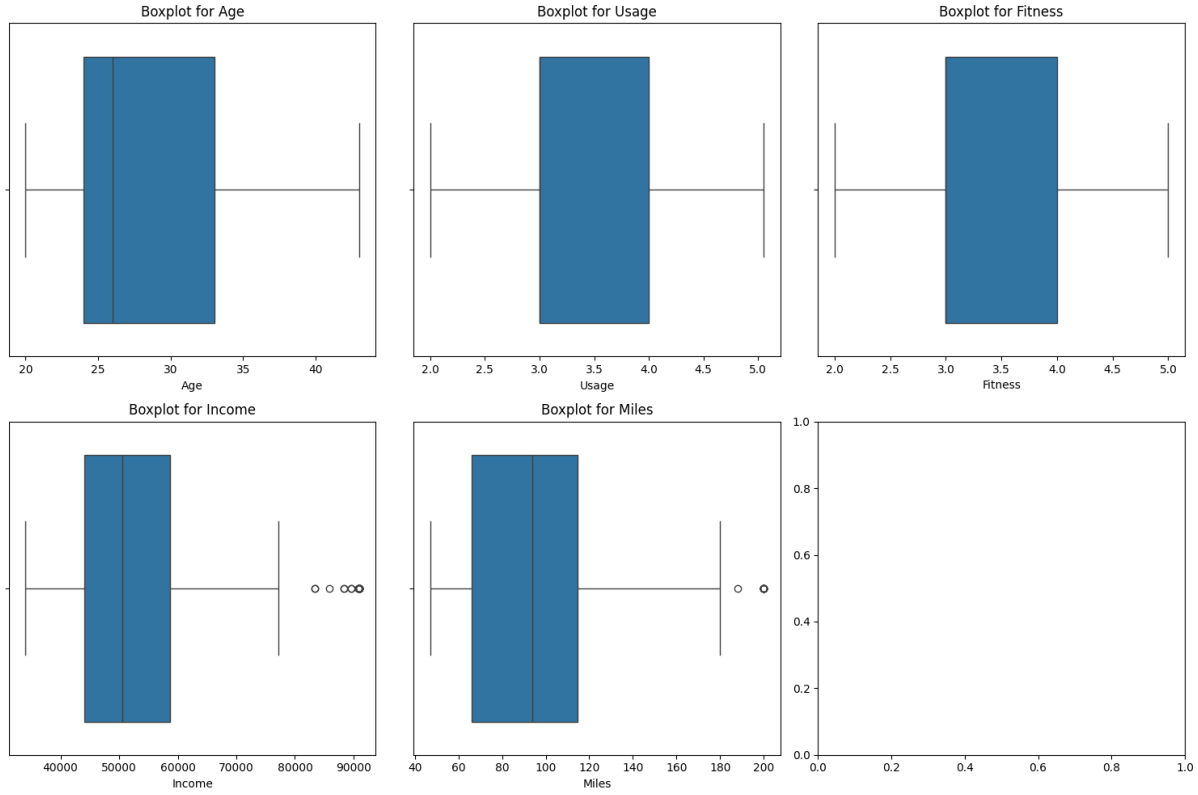
We can detect the outliers using boxplot for different variables:



There are many outliers in each column.
to remove the outliers we will clip the data for below 5% and above 10%.

```
var=['Age','Education','Usage','Fitness','Income','Miles']
for i in var:
  lower=np.percentile(df[i],5)
  upper=np.percentile(df[i],95)
  df[i]=np.clip(df[i],lower,upper)
```

after removing the outliers:



RECCOMENDATIONS:

1. Aerofit should target more on female Customers.
2. Aerofit should make advertisement targeting unmarried customers.
3. Sale of KP781 is very less so Aerofit should advertise their premium product by making less margin on that product.
4. KP481 has very low sale among males. so Aerofit should start making changes in that product so that it will attract more male customers.
5. By launching more Products Aerofit can make a very good Customer base.