1.Import the dataset and do usual data analysis steps like checking the structure & characteristics of the dataset

```
!gdown https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/001/125/original/aerofit_treadmill.csv?16399927
```

```
Downloading...
From: https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/001/125/original/aerofit_treadmill.csv?16399927
To: /content/aerofit_treadmill.csv?16399927
100% 7.28k/7.28k [00:00<00:00, 7.48MB/s]
```

```python
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```python
df = pd.read_csv('aerofit_treadmill.csv?16399927')
df
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles
0	KP281	18	Male	14	Single	3	4	29562	112
1	KP281	19	Male	15	Single	2	3	31836	75
2	KP281	19	Female	14	Partnered	4	3	30699	66
3	KP281	19	Male	12	Single	3	3	32973	85
4	KP281	20	Male	13	Partnered	4	2	35247	47
...	...	...	...	...	...	...	...	...	...
175	KP781	40	Male	21	Single	6	5	83416	200
176	KP781	42	Male	18	Single	5	4	89641	200
177	KP781	45	Male	16	Single	5	5	90886	160
178	KP781	47	Male	18	Partnered	4	5	104581	120
179	KP781	48	Male	18	Partnered	4	5	95508	180

180 rows × 9 columns

```python
df[['Product','Age','Gender','Education','MaritalStatus','Usage','Fitness','Income','Miles']].isnull().sum()
```

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

```python
df.dtypes
```

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

```python
df.nunique()
```

```
Product           3
Age              32
Gender            2
Education         8
```

```
MaritalStatus      2
Usage              6
Fitness            5
Income            62
Miles             37
dtype: int64
```

2.Detect Outliers (using boxplot, "describe" method by checking the difference between mean and median)

```
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

```
sns.boxplot(data=df[["Usage", "Fitness"]])
plt.title('Oultliers for USAGE and FITNESS')
plt.show()
```
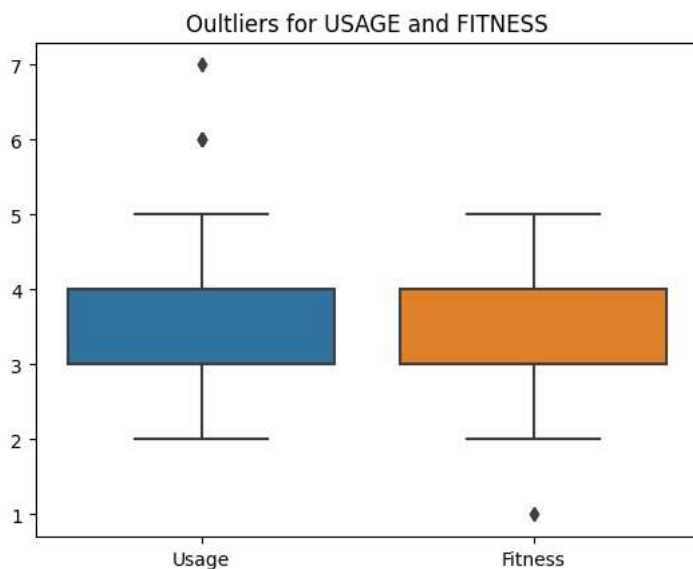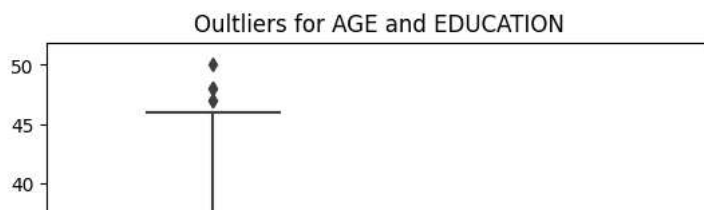


```
sns.boxplot(data=df[["Age", "Education"]])
plt.title('Oultliers for AGE and EDUCATION')

plt.show()
```
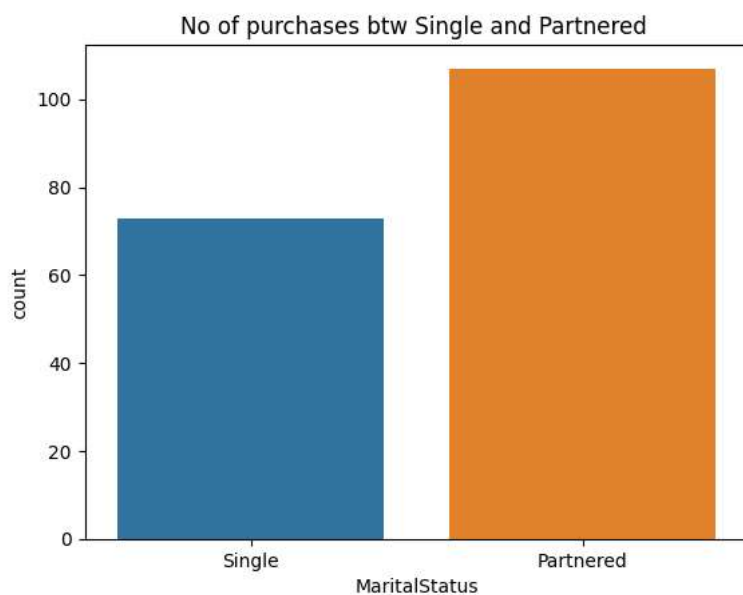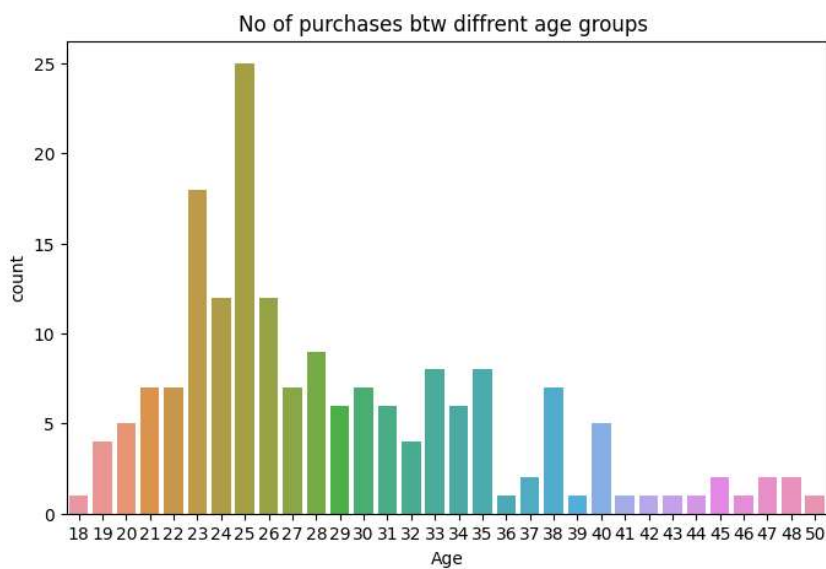
Oultliers for AGE and EDUCATION

3.Check if features like marital status, age have any effect on the product purchased (using countplot, histplots, boxplots etc)

```
sns.countplot(data=df,x='MaritalStatus')
plt.title('No of purchases btw Single and Partnered ')
plt.show()
```
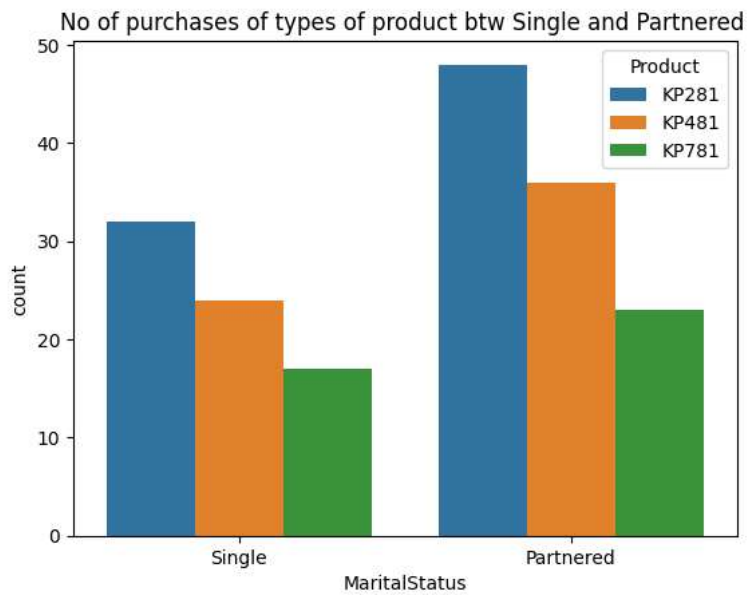


```
plt.figure(figsize=(8,5))
sns.countplot(data=df,x='Age')
plt.title('No of purchases btw diffrent age groups ')

plt.show()
```



```
sns.countplot(data=df,x='MaritalStatus',hue='Product')
plt.title('No of purchases of types of product btw Single and Partnered ')
```

```
plt.show()
```

No of purchases of types of product btw Single and Partnered

4.Representing the marginal probability like - what percent of customers have purchased KP281, KP481, or KP781 in a table

```
len(df.loc[df["Product"]=='KP281'])/len(df)
```

> 0.4444444444444444

```
len(df.loc[df["Product"]=='KP481'])/len(df)
```
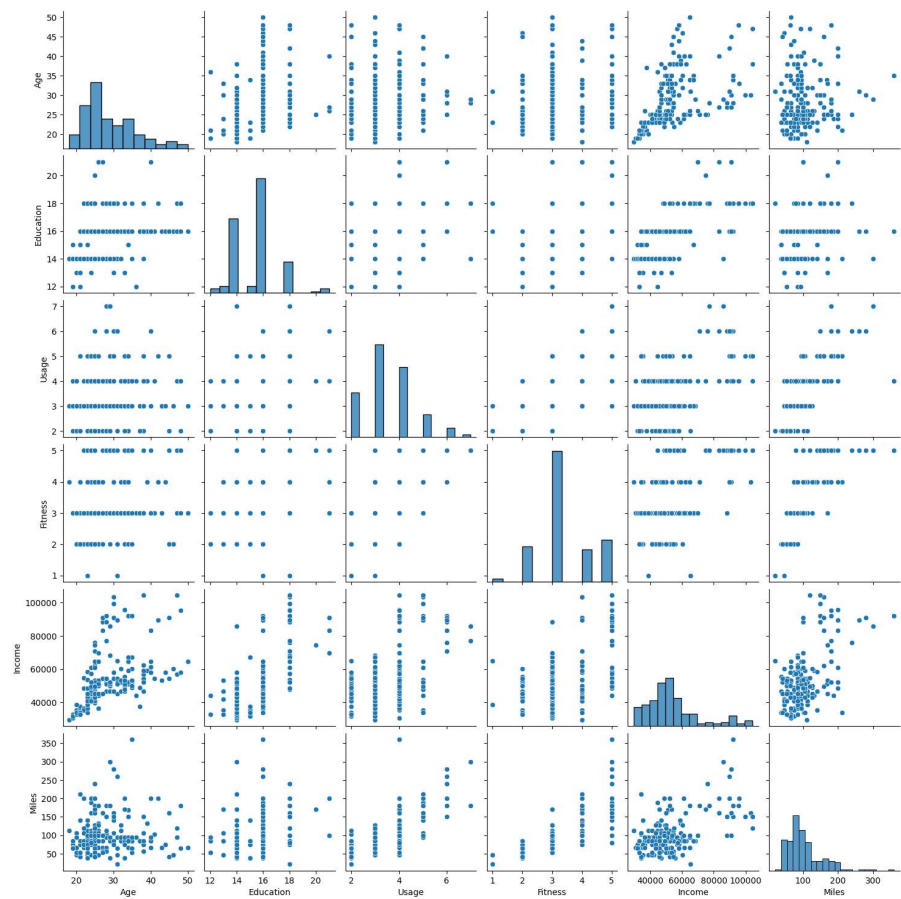
> 0.3333333333333333

```
len(df.loc[df["Product"]=='KP781'])/len(df)
```

> 0.2222222222222222

5.Check correlation among different factors using heat maps or pair plots.

```
sns.pairplot(data=df)
plt.show()
```

6.With all the above steps you can answer questions like: What is the probability of a male customer buying a KP781 treadmill?

```
pd.crosstab(index=df['Product'],columns=df['Gender'],margins=True)
```

Gender Product	Female	Male	All
KP281	40	40	80
KP481	29	31	60
KP781	7	33	40
All	76	104	180

```
33/104
```

    0.3173076923076923

```
7/76
```

    0.09210526315789473

7.Customer Profiling - Categorization of users.
8.Probability- marginal, conditional probability.

```
pd.crosstab(index=df['Product'],columns=df['MaritalStatus'],normalize=True,margins=True)
```

MaritalStatus	Partnered	Single	All
Product			
KP281	0.266667	0.177778	0.444444

```
pd.crosstab(index=df['Product'],columns=df['Fitness'],normalize=True,margins=True)
```

Fitness	1	2	3	4	5	All
Product						
KP281	0.005556	0.077778	0.300000	0.050000	0.011111	0.444444
KP481	0.005556	0.066667	0.216667	0.044444	0.000000	0.333333
KP781	0.000000	0.000000	0.022222	0.038889	0.161111	0.222222
All	0.011111	0.144444	0.538889	0.133333	0.172222	1.000000

```
sns.boxplot(data=df,x='Income',y='Product')
plt.title('Types of product purchased btw different salary catagories ')
plt.show()
```


Insights-

1. There are no null or Nan values in the dataset and there are 3 different types of product KP281,KP481,KP781

2. The difference between mean and median is almost the same for every column in the dataset

3. The graph shows that Partnered people have bought more no of products and people around the age of 23 - 26 have bought more products

4. The most puchased and desired product is KP281

5. The probability of customers who bought KP281,KP481 and KP781 are 0.44, 0.33 and 0.22

6.

- KP281 bought by customers with salary around $4000-5500$
- KP281 bought by customers with salary around $4500-5500$
- KP281 bought by customers with salary around $5800-91000$

Recommendations-

1. By giving more coupons , discounts and payment options on KP481 and KP781 will result in more no of products purchased by the customers

2. We can also give a free trial based period of the products to be used by customers so they would feel more comfortable and confident buying the product

3. Recommending targeted discounts to customers above the age of 35 so those group of people would buy the products.

4. We could alse recommend to give the products for rental or to business or sell the products to gyms to increase diffrent line of revenu stream

● ✕