```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
aerofit=pd.read_csv('/content/drive/MyDrive/Dataset/Aerofit_treadmill.csv')
aerofit.shape
```

```
(180, 9)
```

```
aerofit.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 12 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Product        180 non-null    object
 1   Age            180 non-null    int64
 2   Gender         180 non-null    object
 3   Education      180 non-null    int64
 4   MaritalStatus  180 non-null    object
 5   Usage          180 non-null    int64
 6   Fitness        180 non-null    int64
 7   Income         180 non-null    int64
 8   Miles          180 non-null    int64
 9   miles_cat      180 non-null    category
 10  edu_cat        180 non-null    category
 11  income_cat     179 non-null    category
dtypes: category(3), int64(6), object(3)
memory usage: 13.8+ KB
```

```
aerofit.head(10)
```

| | 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 |
| **5** | KP281 | 20 | Female | 14 | Partnered | 3 | 3 | 32973 | 66 |
| **6** | KP281 | 21 | Female | 14 | Partnered | 3 | 3 | 35247 | 75 |
| **7** | KP281 | 21 | Male | 13 | Single | 3 | 3 | 32973 | 85 |
| **8** | KP281 | 21 | Male | 15 | Single | 5 | 4 | 35247 | 141 |
| **9** | KP281 | 21 | Female | 15 | Partnered | 2 | 3 | 37521 | 85 |

```
#checking for null values
aerofit.isna().sum().sum()
```

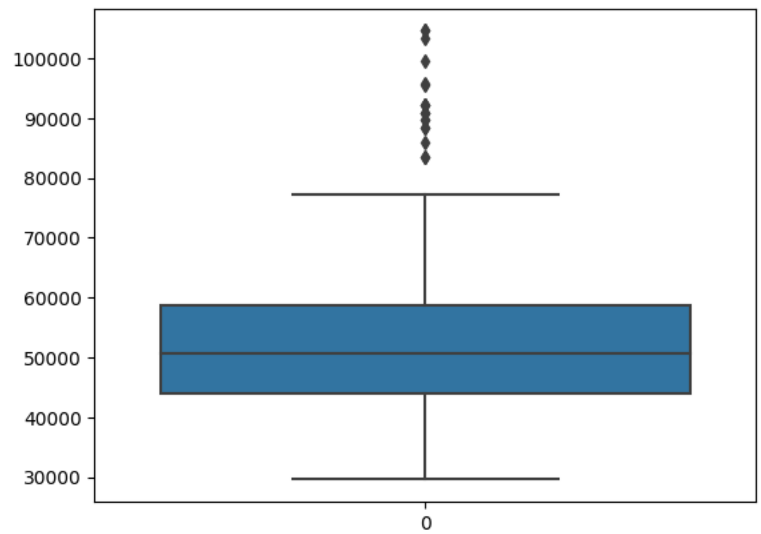
```
0
```

```
#statistical summary
aerofit.describe()
```

```
              Age   Education      Usage    Fitness       Income      Miles
```

```
#Detecting outliers in terms of income and Miles covered
sns.boxplot(aerofit['Income'])
```

```
<Axes: >
```

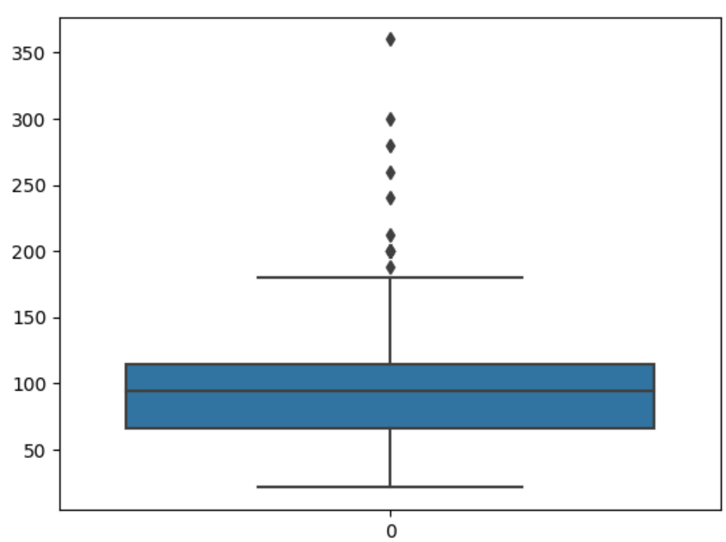


```
#Calculating the outliers
q1=aerofit['Income'].quantile(0.25)
q3=aerofit['Income'].quantile(0.75)
IQR=q3-q1
upper=q3+(1.5*IQR)
lower=q1-(1.5*IQR)
print(upper)
print(lower)
#values above 80581.875 and lower than 22144.875 is detected as Outliers
```

```
80581.875
22144.875
```

```
sns.boxplot(aerofit['Miles'])
```

```
<Axes: >
```

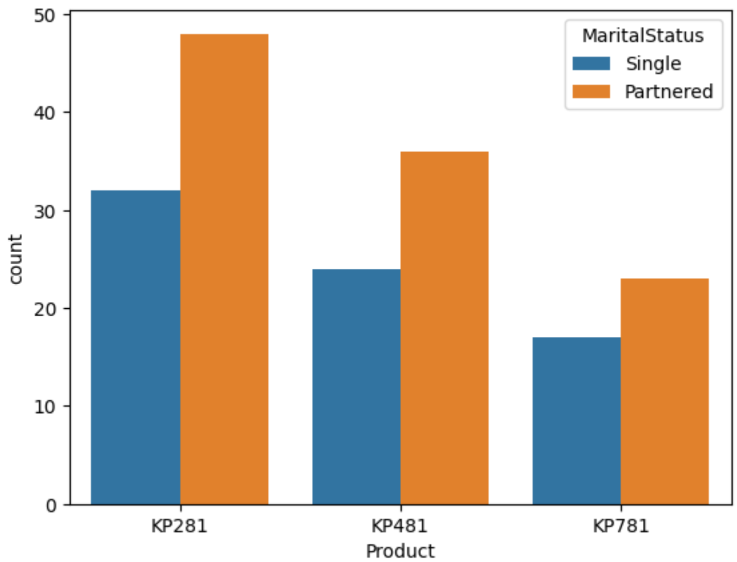


```
#Calculating the outliers
q1=aerofit['Miles'].quantile(0.25)
q3=aerofit['Miles'].quantile(0.75)
IQR=q3-q1
upper=q3+(1.5*IQR)
lower=q1-(1.5*IQR)
```

```
print(upper)
print(lower)
#values above 187.875 and lower than -7.125 is detected as Outliers
```

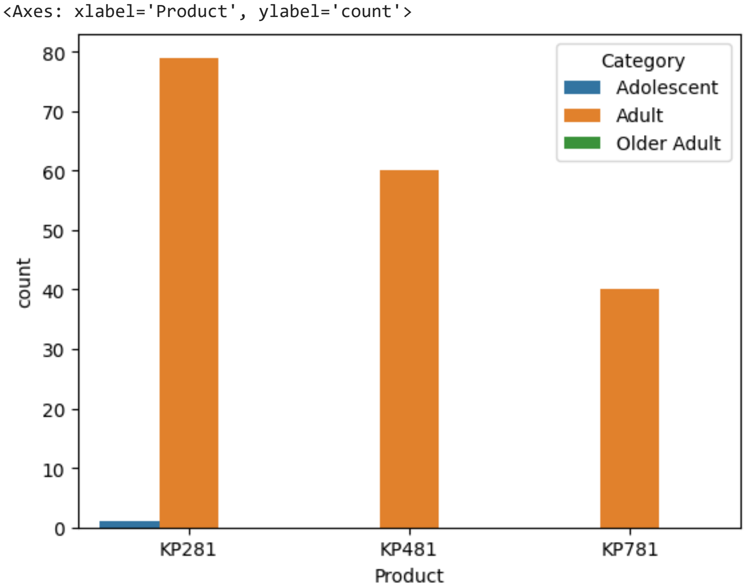
```
187.875
-7.125
```

```
#Checking if marital status affects on the product purchase

ax = sns.countplot(x='Product', data=aerofit,hue='MaritalStatus')
#we can see that most of product of various categories were ordered by married people
```




```
#Relation between Age and no of products purchased
age_labels=[10,18,50,60]
age_cat=['Adolescent ','Adult','Older Adult']
aerofit['Category']=pd.cut(aerofit['Age'],bins=age_labels,labels=age_cat)
aerofit.head(10)
sns.countplot(x='Product', data=aerofit,hue='Category')

#We can see that people above 18 years of age ordered the Products most of the time
```

```
<Axes: xlabel='Product', ylabel='count'>
```



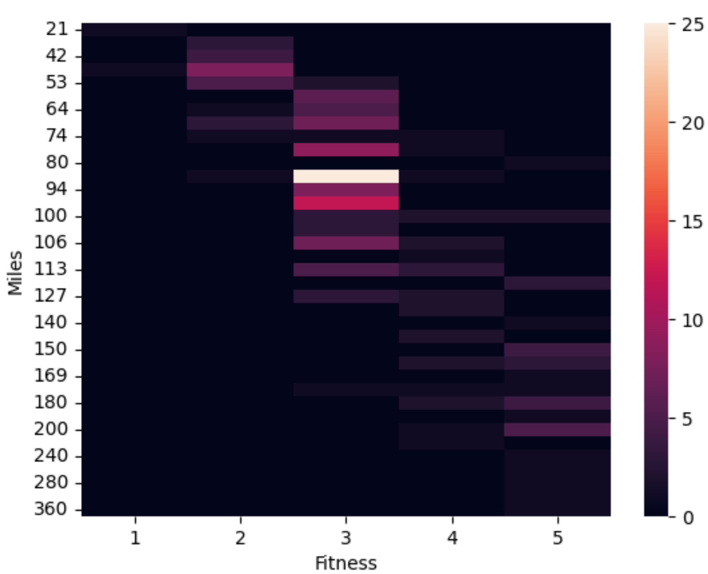
```
#Calculating Marginal Probabilties of Various Products available
aerofit.head(10)
freq=pd.crosstab(aerofit.Product,aerofit.Gender,margins=True)
freq
```

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

```
a=[]
#Probabilty of customers buying KP281:
p1=80/180
print(round(p1,2))
#Probabilty of customers buying KP481:
p2=60/180
print(round(p2,2))
#Probabilty of customers buying KP781:
p3=40/180
print(round(p3,2))


#We can see that as the price of the product increases the purchase rate or demand decreases..KP281 being the cheapest is sold the most.
```

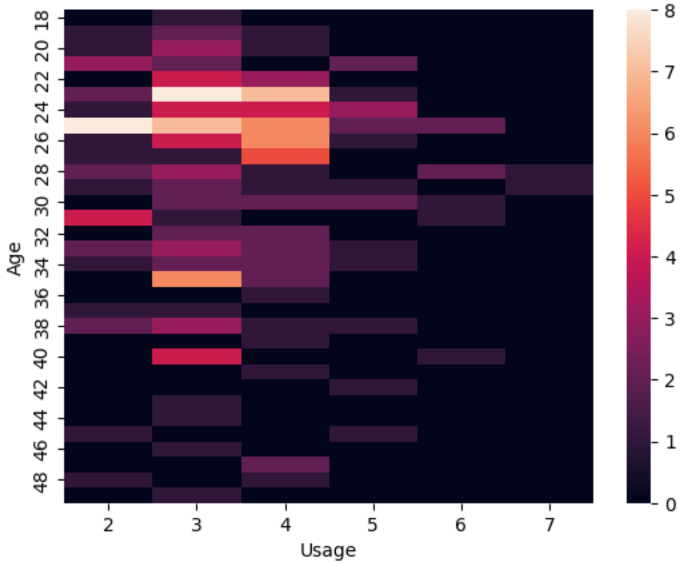
```
0.44
0.33
0.22
```

```
#Checking correlation among different factors:

z=pd.crosstab(aerofit.Miles,aerofit.Fitness)
z.reset_index()
sns.heatmap(z)

#We can see that person rated above 3 are more active and covered more miles.
```




```
u=pd.crosstab(aerofit.Age,aerofit.Usage)
u.reset_index()
sns.heatmap(u)
#From the above analysis from the heatmap...average usage of the trademill  were used by  Adults between  23 to 27 years of age and the avera
```

```
<Axes: xlabel='Usage', ylabel='Age'>
```




```
#What is the probability of a male customer buying a KP781 treadmill?
freq
```

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

```
x1=33/180
x1
```

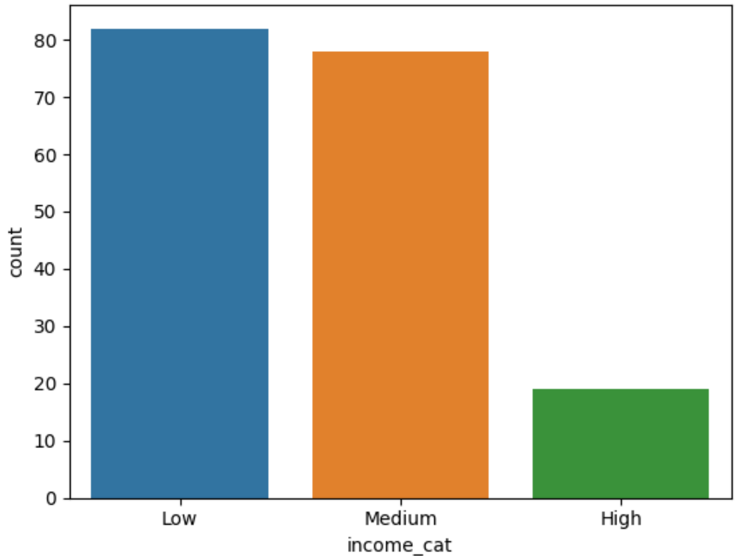
```
0.18333333333333332
```

```
#Analyzing Customer Profiling from the dataset:Understanding Male or Female Customers has frequently brought the product most of the times--
sns.countplot(x='Product',data=aerofit,hue='Gender')
#From the below graph we can see that Male Customers have more purchase rate in the category of KP481,KP781
```

```
<Axes: xlabel='Product', ylabel='count'>
```

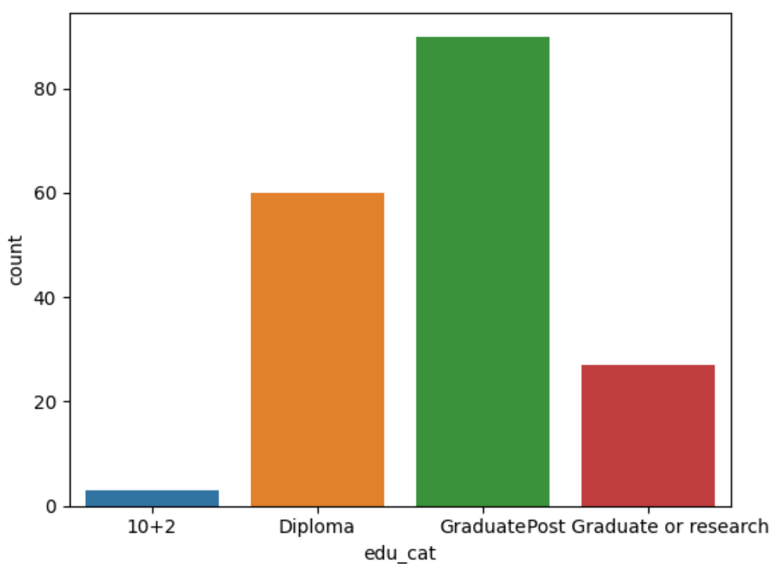
```
#Analysing Customers based on Income:

income_labels=[30000,50000,80000,105000]
income_cat=['Low','Medium','High']
aerofit['income_cat']=pd.cut(aerofit['Income'],bins=income_labels,labels=income_cat)
sns.countplot(x='income_cat',data=aerofit)

#We can see that a person annual income between 30k to 50k purchased most of the trademills..so a special discount should be made available f
```

```
<Axes: xlabel='income_cat', ylabel='count'>
```




```
#Distribution of Customers Based on Education
aerofit['Education'].value_counts()
edu_labels=[10,12,14,16,21]
edu_cat=['10+2','Diploma','Graduate','Post Graduate or research']
aerofit['edu_cat']=pd.cut(aerofit['Education'],bins=edu_labels,labels=edu_cat)
sns.countplot(x='edu_cat',data=aerofit)

#Most of the customers are graduated i.e completed 16 years of full time education
```

```
<Axes: xlabel='edu_cat', ylabel='count'>
```




```
#Given that a person has a high income category..what is the probabilty of buying kp781?
income_labels=[30000,50000,105000]
income_cat=['Low','High']
aerofit['income_cat']=pd.cut(aerofit['Income'],bins=income_labels,labels=income_cat)
pd.crosstab(aerofit.Product,aerofit.income_cat,margins=True)
```

| income_cat | Low | High | All |
|---|---|---|---|
| **Product** | | | |
| **KP281** | 47 | 32 | 79 |
| **KP481** | 30 | 30 | 60 |
| **KP781** | 5 | 35 | 40 |
| **All** | 82 | 97 | 179 |

```
x=35/97
x
```

```
0.36082474226804123
```

```
#What is the probabilty of low income level customers buying KP281?
y=47/179
y
```

```
0.26256983240223464
```

```
edu_labels=[10,12,14,16,21]
edu_cat=['10+2','Diploma','Graduate','Post Graduate or research']
aerofit['edu_cat']=pd.cut(aerofit['Education'],bins=edu_labels,labels=edu_cat)
pd.crosstab([aerofit.Product,aerofit.edu_cat],aerofit.income_cat,margins=True)
```

| | income_cat | Low | High | All |
|---|---|---|---|---|
| **Product** | **edu_cat** | | | |
| **KP281** | **10+2** | 2 | 0 | 2 |
| | **Diploma** | 23 | 9 | 32 |
| | **Graduate** | 22 | 21 | 43 |
| | **Post Graduate or research** | 0 | 2 | 2 |
| **KP481** | **10+2** | 1 | 0 | 1 |
| | **Diploma** | 21 | 4 | 25 |
| | **Graduate** | 7 | 25 | 32 |
| | **Post Graduate or research** | 1 | 1 | 2 |
| **KP781** | **Diploma** | 1 | 1 | 2 |
| | **Graduate** | 3 | 12 | 15 |
| | **Post Graduate or research** | 1 | 22 | 23 |
| **All** | | 82 | 97 | 179 |

```
#What is the probabilty that a customer has a graduate level of education and is buying KP481 and his income category is high?
#Probabilty of high income category buying kp481 and is graduate:
x=25/97
#Probabilty of people buying kp481:
y=60/179
#Probabilty of customer has a graduate level and is buying KP481 and income is high:
x/y
```

```
0.7689003436426116
```

```
#Checking if the person is active(low,moderate,high) according to miles covered:
miles_labels=[20,120,240,360]
miles_cat=['less active','moderately active','highly active']
aerofit['miles_cat']=pd.cut(aerofit['Miles'],bins=miles_labels,labels=miles_cat)
pd.crosstab([aerofit.Product,aerofit.miles_cat],aerofit.income_cat,margins=True)
```

| Product | miles_cat (income_cat) | Low | High | All |
|---|---|---|---|---|
| KP281 | less active | 43 | 30 | 73 |
| | moderately active | 4 | 2 | 6 |
| KP481 | less active | 25 | 27 | 52 |
| | moderately active | 5 | 3 | 8 |
| KP781 | less active | 3 | 9 | 12 |
| | moderately active | 2 | 22 | 24 |
| | highly active | 0 | 4 | 4 |
| All | | 82 | 97 | 179 |

```
#What is the probabilty that a customer has a low activity level and is buying KP281 and his income category is high?

#Probabilty of high income category buying kp281 and is activity level is low:
x=30/97
#Probabilty of people buying kp281:
y=79/179
#Probabilty of customer has a graduate level and is buying KP481 and income is high:
x/y
```

```
0.700769933446431
```

```
#Product Purchased according to Fitness level:
pd.crosstab([aerofit.Product,aerofit.Fitness],aerofit.income_cat,margins=True)
```

| Product | Fitness (income_cat) | Low | High | All |
|---|---|---|---|---|
| KP281 | 1 | 1 | 0 | 1 |
| | 2 | 8 | 6 | 14 |
| | 3 | 31 | 23 | 54 |
| | 4 | 6 | 2 | 8 |
| | 5 | 1 | 1 | 2 |
| KP481 | 1 | 0 | 1 | 1 |
| | 2 | 7 | 5 | 12 |
| | 3 | 18 | 21 | 39 |
| | 4 | 5 | 3 | 8 |
| KP781 | 3 | 1 | 3 | 4 |
| | 4 | 0 | 7 | 7 |
| | 5 | 4 | 25 | 29 |
| All | | 82 | 97 | 179 |

```
#Probabilty of a person having high income category buying kp781 and is rated 5 in fitness level:
#Probabilty of high income category buying kp781 and is rated 5 in fitness:
x=25/179
#Probabilty of people buying kp781:
y=40/179
#Probabilty of customer having 5 rating in fitness and is buying KP781 and income is high:
x/y
```

```
0.625
```

```
#We want to find out the probabilty of a female customer buying KP281 who is less active and have high income level with Graduate Degree?
pd.crosstab([aerofit.Product,aerofit.miles_cat,aerofit.income_cat,aerofit.edu_cat],aerofit.Gender,margins=True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 12 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Product        180 non-null    object
 1   Age            180 non-null    int64
 2   Gender         180 non-null    object
 3   Education      180 non-null    int64
 4   MaritalStatus  180 non-null    object
 5   Usage          180 non-null    int64
 6   Fitness        180 non-null    int64
 7   Income         180 non-null    int64
 8   Miles          180 non-null    int64
 9   miles_cat      180 non-null    category
 10  edu_cat        180 non-null    category
 11  income_cat     179 non-null    category
dtypes: category(3), int64(6), object(3)
memory usage: 13.8+ KB
```

| | | | Gender | Female | Male | All |
|---|---|---|---|---|---|---|
| Product | miles_cat | income_cat | edu_cat | | | |
| **KP281** | **less active** | **Low** | **10+2** | 0 | 2 | 2 |
| | | | **Diploma** | 15 | 7 | 22 |
| | | | **Graduate** | 10 | 9 | 19 |
| | | **High** | **Diploma** | 3 | 6 | 9 |
| | | | **Graduate** | 10 | 9 | 19 |
| | | | **Post Graduate or research** | 1 | 1 | 2 |
| | **moderately active** | **Low** | **Diploma** | 0 | 1 | 1 |
| | | | **Graduate** | 1 | 2 | 3 |
| | | **High** | **Graduate** | 0 | 2 | 2 |
| **KP481** | **less active** | **Low** | **10+2** | 0 | 1 | 1 |
| | | | **Diploma** | 8 | 9 | 17 |
| | | | **Graduate** | 4 | 2 | 6 |
| | | | **Post Graduate or research** | 1 | 0 | 1 |
| | | **High** | **Diploma** | 3 | 0 | 3 |
| | | | **Graduate** | 9 | 14 | 23 |
| | | | **Post Graduate or research** | 1 | 0 | 1 |
| | **moderately active** | **Low** | **Diploma** | 2 | 2 | 4 |
| | | | **Graduate** | 0 | 1 | 1 |
| | | **High** | **Diploma** | 0 | 1 | 1 |
| | | | **Graduate** | 1 | 1 | 2 |
| **KP781** | **less active** | **Low** | **Diploma** | 0 | 1 | 1 |
| | | | **Graduate** | 0 | 2 | 2 |
| | | **High** | **Graduate** | 0 | 2 | 2 |
| | | | **Post Graduate or research** | 2 | 5 | 7 |
| | **moderately active** | **Low** | **Graduate** | 0 | 1 | 1 |
| | | | **Post Graduate or research** | 0 | 1 | 1 |
| | | **High** | **Graduate** | 1 | 6 | 7 |

```
x=10/76
y=79/179
x/y
```

```
0.2981345769487009
```

```
#We want to find out the probabilty of a female customer buying KP481 who is less active and have high income level with Graduate Degree?
x=9/76
y=60/179
x/y
```

```
0.35328947368421054
```

```
#We want to find out the probabilty of a female customer buying KP781 who is less active and have high income level with Graduate Degree?
x=0
y=40/179
x/y
```

```
0.0
```

```
#Performing Descriptive Statistics of the dataset
aerofit['Age'].describe()
#The mean Age of the people is 18 with a Std dev of 6.94 and max Age is 50
```

```
count    180.000000
mean      28.788889
std        6.943498
min       18.000000
25%       24.000000
50%       26.000000
75%       33.000000
max       50.000000
Name: Age, dtype: float64
```

```
x=aerofit['Gender'].value_counts()
t=x[0]+x[1]
m=x[0]/t
f=x[1]/t
print(m,f)
#The data set is almost evenly distributed where percentage of male is 57% and that of female is 42%
```

```
0.5777777777777777 0.4222222222222222
```

```
#What is the average miles covered by a male customer with that of female customer:
aerofit.groupby('Gender')['Miles'].mean()
```

```
Gender
Female     90.013158
Male      112.826923
Name: Miles, dtype: float64
```

```
#As per our analysis we can suggest KP481 to a female customer who is highly educated and has high level of income and whose activity level i
```

```
#In this way we can recommend various other products to our customers based on different criterias(Age,Marital Status,Miles and Fitness level
```

```
#Insights and Recommendations:
'''
1.We can see that most of product of various categories were ordered by married people and people above 18 years of age ordered the Products 
2.The probabilty of buying KP281 is more so additional features should be configured in the system to get customer more views.
3.We can see that a person annual income between 30k to 50k purchased most of the trademills..so a special discount should be made available 
4.Most of the customers are graduated i.e completed 16 years of full time education
5.We can see that highly active individuals and high level of income are purchasing KP781.
6.After performing analysis based on recommending a product we can say that KP481 is a better choice for a female customer who is highly educ
7.From the above analysis from the heatmap...average usage of the trademill  were used by  Adults between  23 to 27 years of age and the aver
'''
```

```
'\n1.We can see that most of product of various categories were ordered by married people and people a
bove 18 years of age ordered the Products most of the time,So a Special Offer or Loyalty program shoul
d be made available to them to get customer attention and increase more purchase Rate.\n2.The probabil
ty of buying KP281 is more so additional features should be configured in the system to get customer m
ore views. \n3.We can see that a person annual income between 30k to 50k purchased most of the trademi
lls..so a special discount should be made available for them.\n4.Most of the customers are graduated 
i.e completed 16 years of full time education\n5.We can see that highly active individuals and high le
```