Aerofit Case Study

Aerofit is a leading brand in the field of fitness equipment. Aerofit provides a product range including machines such as treadmills, exercise bikes, gym equipment, and fitness accessories to cater to the needs of all categories of people.

The objective of this study is to identify the characteristics of the target audience for each type of treadmill offered by the company, to provide a better recommendation of the treadmills to the new customers.

To access the same, the company collected the data on individuals who purchased a treadmill from the AeroFit stores during the prior three months.

Here is a brief overview of the type of data that is present in the dataset:

There are a total of 180 rows and 9 columns in the dataset with the following data types:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 9 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 
dtypes: int64(6), object(3)
memory usage: 12.8+ KB
```

Also to provide the descriptive statistics for all the columns, we have:

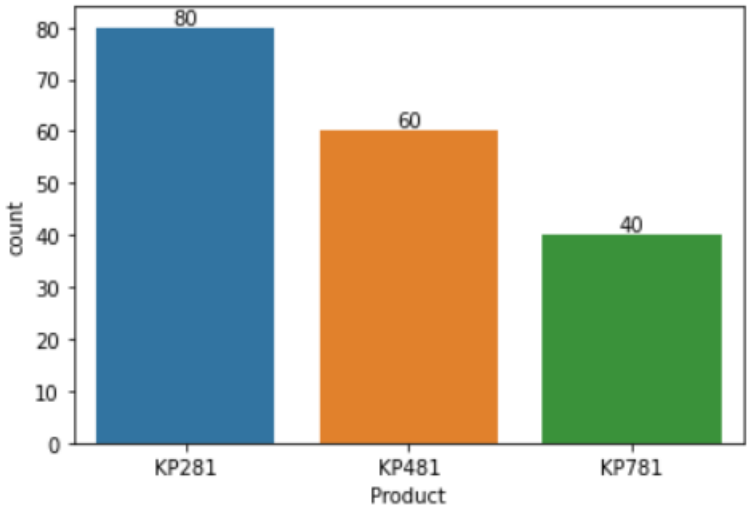
```
df.describe(include = 'all')
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles
count	180	180.000000	180	180.000000	180	180.000000	180.000000	180.000000	180.000000
unique	3	NaN	2	NaN	2	NaN	NaN	NaN	NaN
top	KP281	NaN	Male	NaN	Partnered	NaN	NaN	NaN	NaN
freq	80	NaN	104	NaN	107	NaN	NaN	NaN	NaN
mean	NaN	28.788889	NaN	15.572222	NaN	3.455556	3.311111	53719.577778	103.194444
std	NaN	6.943498	NaN	1.617055	NaN	1.084797	0.958869	16506.684226	51.863605
min	NaN	18.000000	NaN	12.000000	NaN	2.000000	1.000000	29562.000000	21.000000
25%	NaN	24.000000	NaN	14.000000	NaN	3.000000	3.000000	44058.750000	66.000000
50%	NaN	26.000000	NaN	16.000000	NaN	3.000000	3.000000	50596.500000	94.000000
75%	NaN	33.000000	NaN	16.000000	NaN	4.000000	4.000000	58668.000000	114.750000
max	NaN	50.000000	NaN	21.000000	NaN	7.000000	5.000000	104581.000000	360.000000

1. Product – The Product column has information about the product being purchased namely KP281, KP481, or KP781. These are the model numbers of the treadmills, with the following features of the products:

- The KP281 is an entry-level treadmill that sells for $1,500.
- The KP481 is for mid-level runners that sell for $1,750.
- The KP781 treadmill is having advanced features that sell for $2,500.

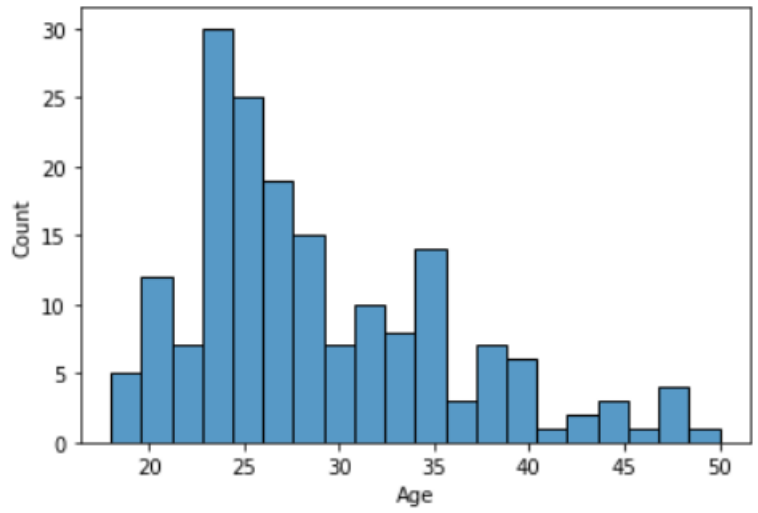
The distribution of the products in the dataset is as follows:



2. Age – Age of the customer in years. For this, we have plotted a histogram to see the distribution of the customer ages.

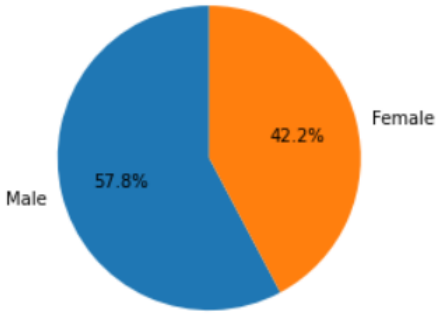
```
sns.histplot(x= 'Age', data = df, bins = 20)
```

```
<AxesSubplot:xlabel='Age', ylabel='Count'>
```



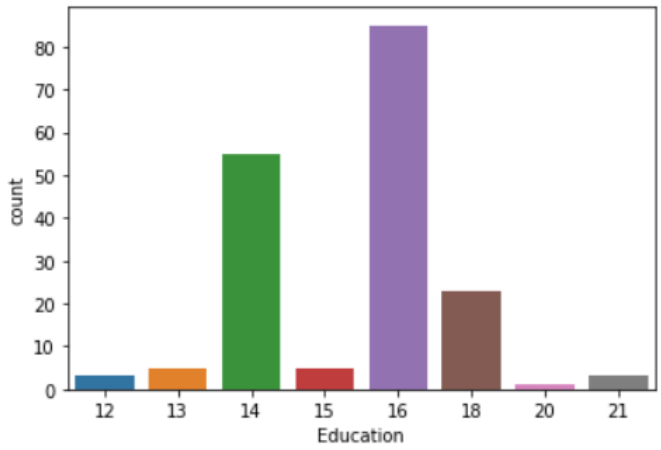
We can observe that the maximum people buying the treadmills are in the typical young bracket 20-30 years.

3. Gender – Male/Female

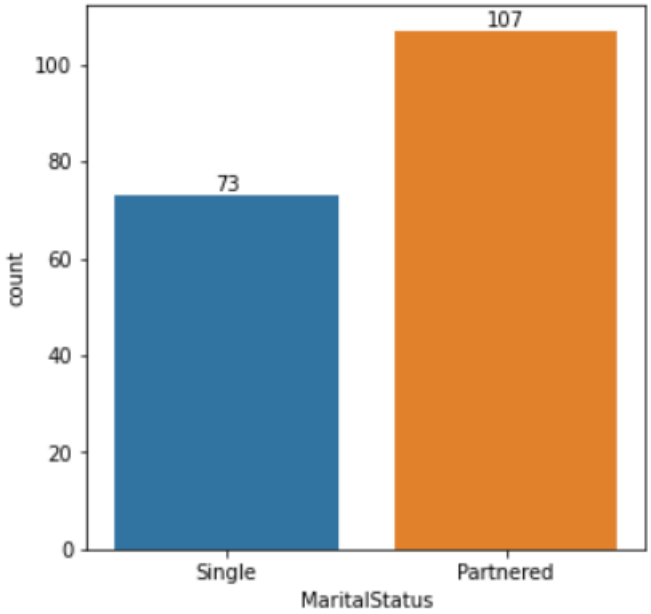


Males are a little more than females in purchasing the treadmills.

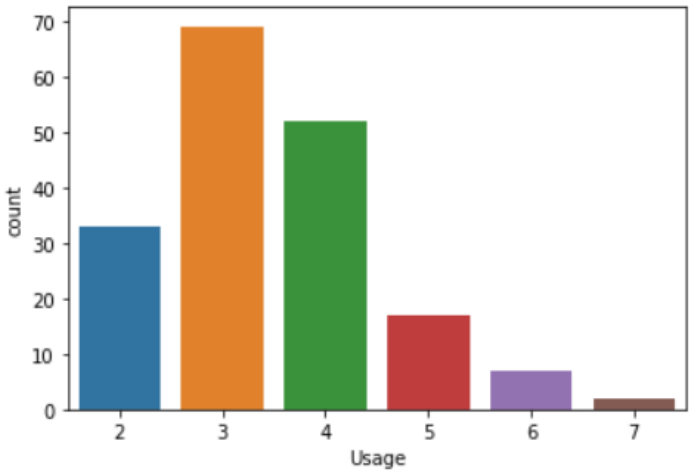
4. Education - It is the number of years of education that one has received. We observe that maximum customers have received 16 or 14 years of education.



5. Marital Status – This column is a categorical variable indicating whether a person is single or partnered. Maximum people are partnered.

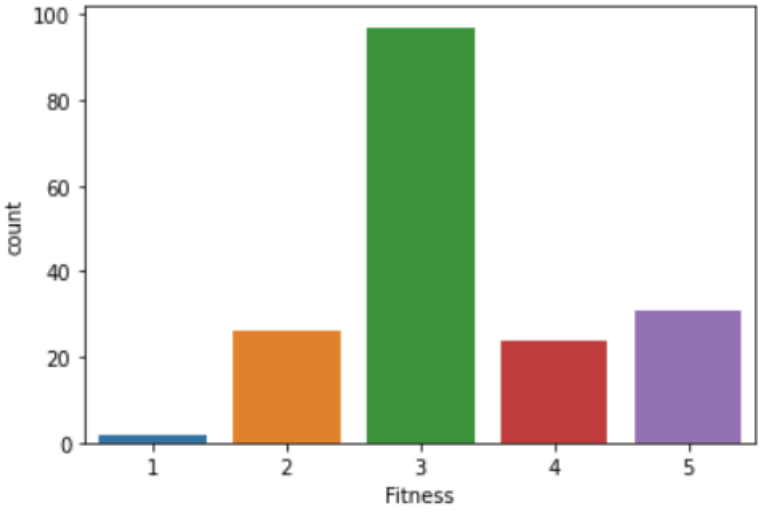


6. Usage - The average number of times the customer plans to use the treadmill each week.



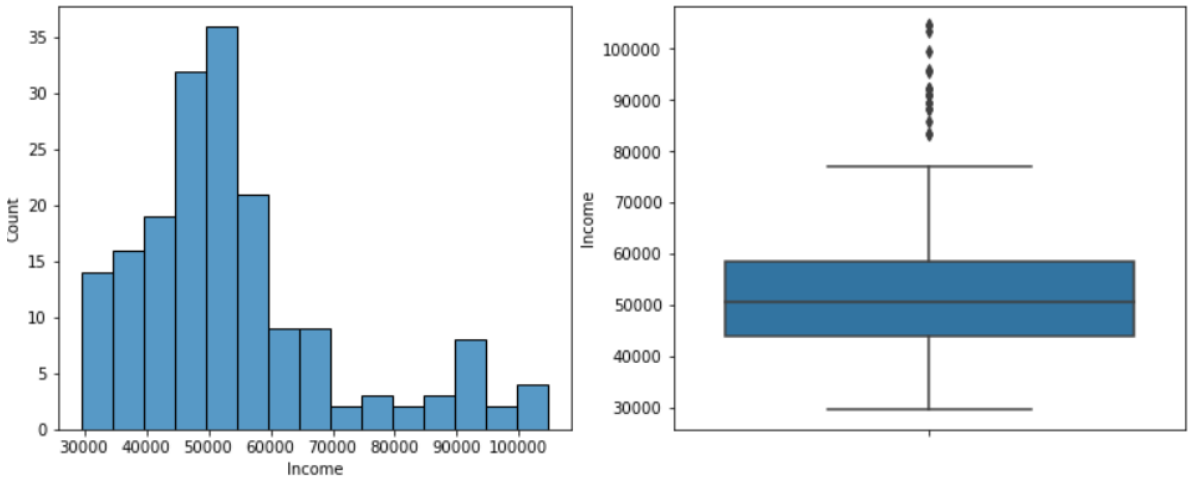
We can observe that maximum customers plan to use the treadmill for 3 to 4 days in a week.

6. Fitness - Self-rated fitness on a 1-to-5 scale, where 1 is the poor shape and 5 is the excellent shape.



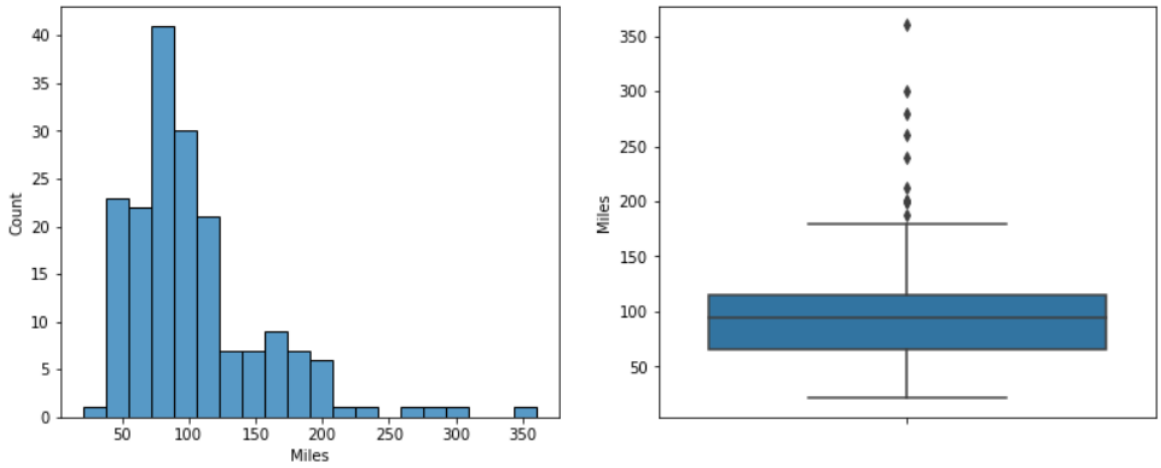
Maximum customers think their fitness to be on an average.

7. Income - Annual income (in $)



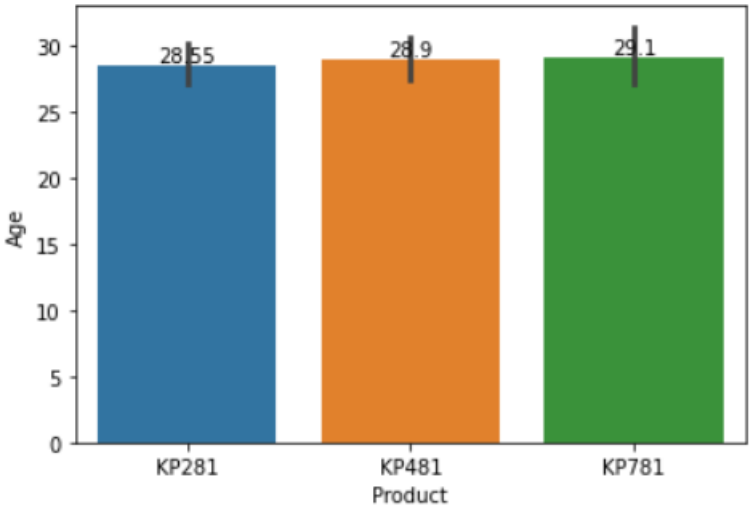
This shows that there are some outliers for the income but from the histogram, it is clear that we can consider those cases as they are not so much out of the bounds and it is completely possible that a few people can earn double the median income. The median of income column lies at $ 50596.5, mean at $ 53720 and the max is $ 104581.

8. Miles - The average number of miles the customer expects to walk/run each week. The median miles that the customer expects to run each week lies at 94 and the max of this column is at 360 miles per week. This may not be considered as an outlier as in reality that may be case with a sportsman or with a fitness freak.



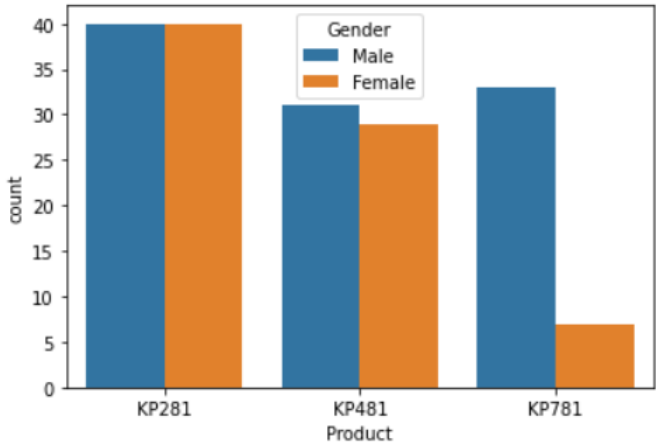
Bivariate Analysis

1. Product and Age: This is done to understand how different age grouped people make their choices regarding the kind of treadmill they buy.

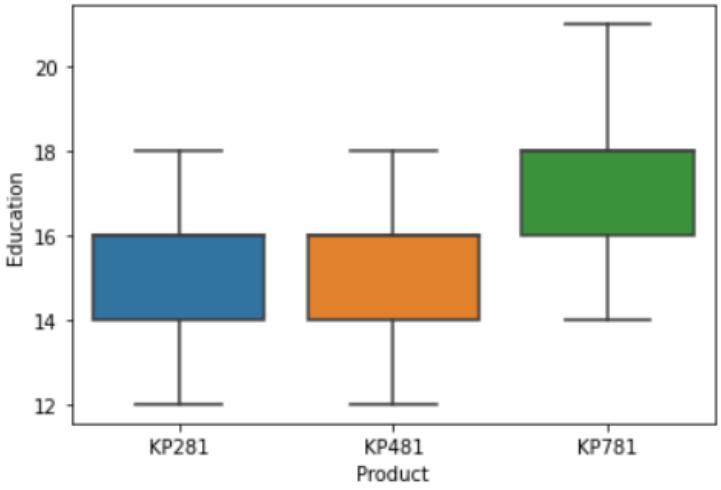


We can observe that there is not much difference between the buying patterns of customers basis their age. In other words, all sorts of treadmills are popular across all age brackets.

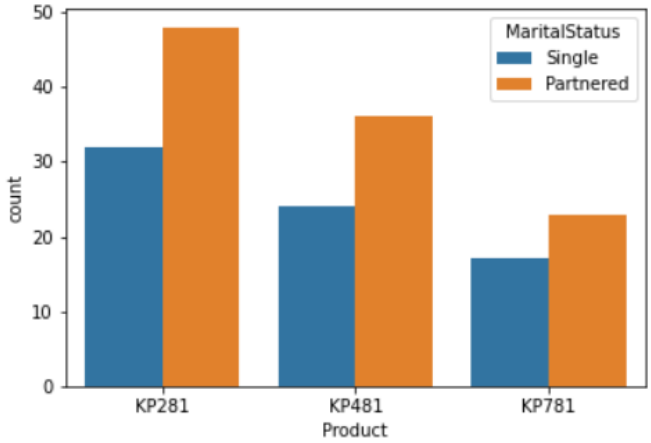
2. Product and Gender – Evident from the following graph, females are far less likely to buy the most premium treadmill (KP781) than males.



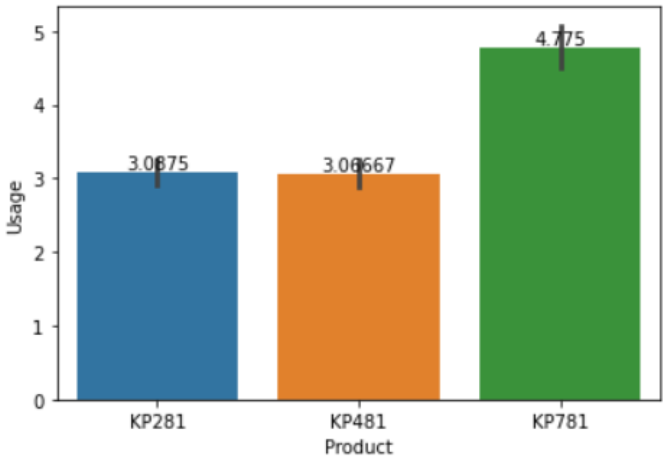
3. Product and Education – KP781 is more popular amongst people with more years of education. That may be because they would be able to appreciate and value its advance features.



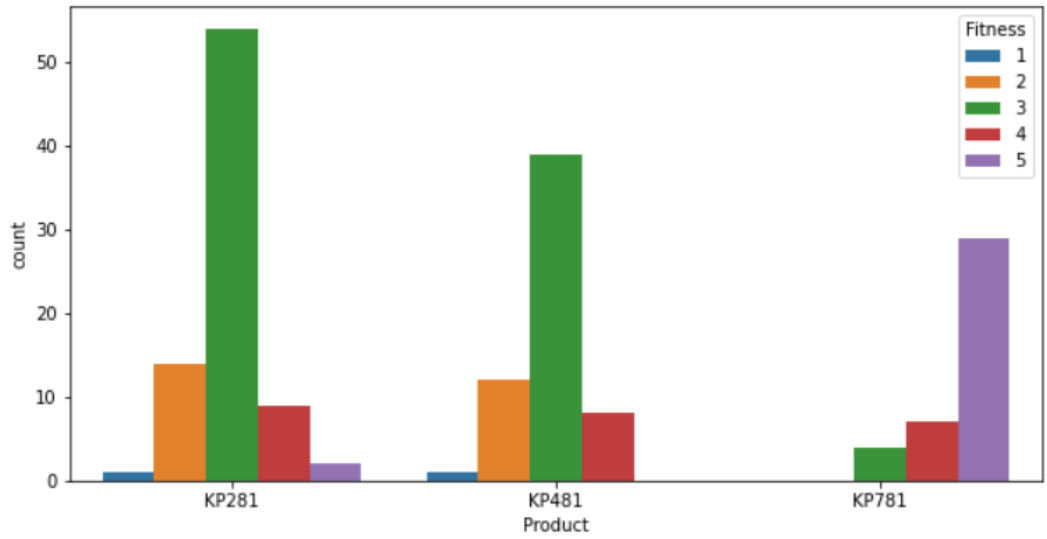
4. Product and Marital Status – There is a clear cut preference of buying a treadmill amongst people who have a partner against those who are single.



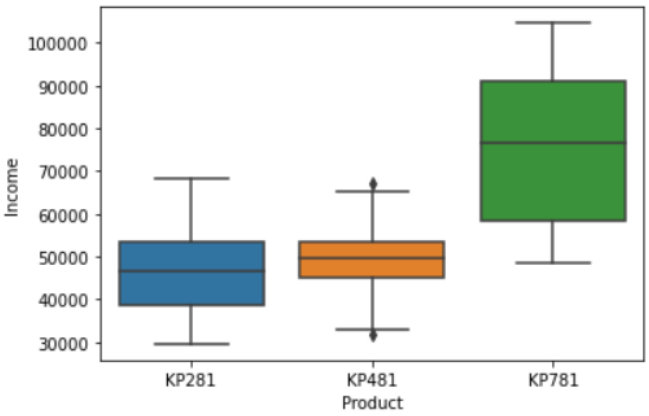
5. Product and Usage - People who have a higher usage or who think are going to use the treadmill extensively seem to prefer the most expensive version of treadmill against the economical versions of the same.



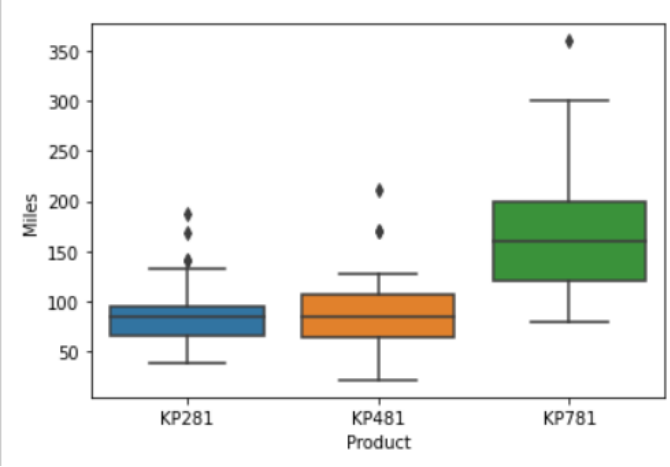
6. Product and Fitness – People who have better fitness levels seem to prefer KP781 against its cheaper counterparts. People with average fitness want to invest in KP281 and KP581 only.



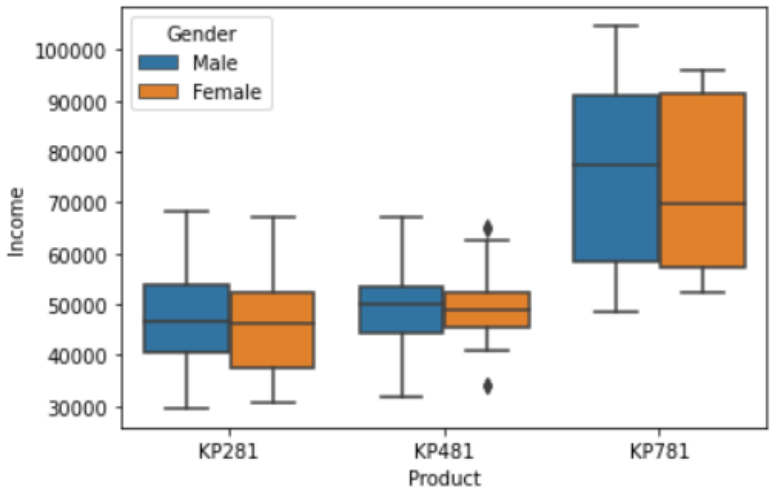
7. Product and Income – people with higher incomes have a bias for the premium product as against those with relatively lower income.



8. Product and Miles – It has a similar diagram as the previous one where customers who run more miles per week tend to prefer a better version of treadmill for themselves.

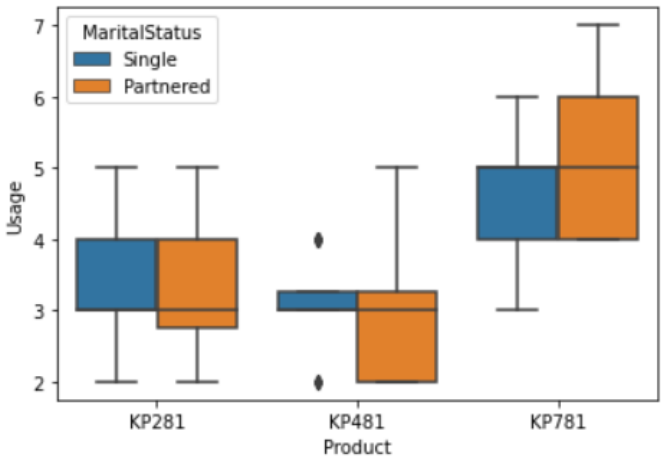


9. Product, gender and Income – Similar preferences for the treadmill are displayed across gender depending on the age group they fall in.

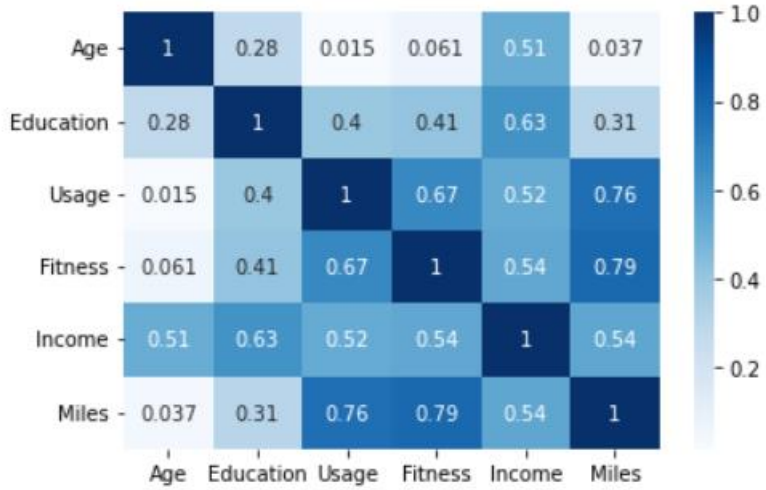


10. Product, Usage and Marital Status – People who have partners are more likely to buy a treadmill depending on their usage. If usage is more, then they will prefer the premium

version over the economical version of the treadmill. For average usage, they may buy the relatively cheaper versions of treadmill.



11. Correlation heatmap – There is high correlation between Usage and Miles and Fitness and Miles, which makes intuitive sense. The more number of miles an individual will plan to run, more will be his usage and correspondingly higher will be his fitness. The interesting thing is that all three of these variables have very less correlation with age. This indicates that irrespective of age, one can choose to be as fit as possible and that age is just a number.



From the above analysis, we can conclude some points:

1. We can see that gender is a big differentiator when it comes to which product category are they buying.

2. Similarly, partnered people are more likely to buy treadmill vs those who are single.

3. Education and income play a differentiating role when it comes to selecting which product category customers are buying.

4. People with higher usage, fitness and miles are more likely to buy an advanced version of treadmill than its economical counterparts.

5. Usage, fitness and miles are closely correlated and together they are uncorrelated with age.

Calculating Marginal, Joint and Conditional probabilities

Step 1 : Binning Data

For calculating the probabilities, we will first bin the numerical data columns. This will not only help in easier analysis but will also solve the problem of outliers as we have observed in the related box plots of income, miles etc. When outliers get covered in a bin, they won't affect our analysis. For this, we have used the pandas cut function and have created bins for the following columns:

1. Age as Age_bins

2. Income as Income_bins

3. Miles as Mile_bins

4. Education as Education_bins

Step 2: For calculating the conditional probabilities we have used pandas crosstab function and calculated conditional probabilities for all the variables with treadmill product one by one. The results and analysis for the same is mentioned below:

1. Gender vs Treadmill Product

```
Table for  Gender vs Treadmill Product
```

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

```
Table for  Gender vs Treadmill Product
```

Product	KP281	KP481	KP781
Gender			
Female	0.526316	0.381579	0.092105
Male	0.384615	0.298077	0.317308
All	0.444444	0.333333	0.222222

Here, we can easily see that if a female were to buy a treadmill, her clear preference will be for the most economical range of the treadmill as P[KP281|Female] = 0.52 (highest value) and also there is a very remote chance that she will buy KP781 as P[KP781|Female] = .092 ~ 9%. For males, there is not much of a difference between the likelihood of buying a particular product category.

2. MaritalStatus vs Treadmill Product

Product	KP281	KP481	KP781	All
MaritalStatus				
Partnered	48	36	23	107
Single	32	24	17	73
All	80	60	40	180

Product	KP281	KP481	KP781
MaritalStatus			
Partnered	0.448598	0.336449	0.214953
Single	0.438356	0.328767	0.232877
All	0.444444	0.333333	0.222222

Only thing noticeable is that partnered people are more likely to buy a treadmill with no clear distinction between any product category.

3. Education Bins vs Treadmill product

Product	KP281	KP481	KP781	All
Education_bins				
Less Edu	39	27	2	68
Avg Edu	41	33	34	108
High Edu	0	0	4	4
All	80	60	40	180

Product	KP281	KP481	KP781
Education_bins			
Less Edu	0.573529	0.397059	0.029412
Avg Edu	0.379630	0.305556	0.314815
High Edu	0.000000	0.000000	1.000000
All	0.444444	0.333333	0.222222

Here, clearly less educated have clear preference for KP281 while the avg educated have an almost equal spread across product category but highly educated buy KP781 only.

4. Age_bins vs Treadmill Product

Product	KP281	KP481	KP781	All
Age_bins				
Till 20	6	4	0	10
20-30yrs	49	31	30	110
30-40yrs	19	23	6	48
40-50yrs	6	2	4	12
All	80	60	40	180

Product	KP281	KP481	KP781
Age_bins			
Till 20	0.600000	0.400000	0.000000
20-30yrs	0.445455	0.281818	0.272727
30-40yrs	0.395833	0.479167	0.125000
40-50yrs	0.500000	0.166667	0.333333
All	0.444444	0.333333	0.222222

Relatively younger till 30 years (modal age group) and relatively older population have a high preference for KP281, whereas those in the middle age of 30-40yrs buy KP481.

5. Income_bins and Treadmill product

Product Income_bins	KP281	KP481	KP781	All
Low	23	9	0	32
Medium	51	44	11	106
High	6	7	10	23
Very High	0	0	19	19
All	80	60	40	180

Product Income_bins	KP281	KP481	KP781
Low	0.718750	0.281250	0.000000
Medium	0.481132	0.415094	0.103774
High	0.260870	0.304348	0.434783
Very High	0.000000	0.000000	1.000000
All	0.444444	0.333333	0.222222

Low income group have a clear cut preference for KP281, medium income group (mode of the population) although have a preference for KP281 but they can buy higher versions of the product if persuaded properly. High and very high income group is already buying KP781.

6. Usage vs Treadmill Product

Product Usage	KP281	KP481	KP781	All
2	19	14	0	33
3	37	31	1	69
4	22	12	18	52
5	2	3	12	17
6	0	0	7	7
7	0	0	2	2
All	80	60	40	180

Product Usage	KP281	KP481	KP781
2	0.575758	0.424242	0.000000
3	0.536232	0.449275	0.014493
4	0.423077	0.230769	0.346154
5	0.117647	0.176471	0.705882
6	0.000000	0.000000	1.000000
7	0.000000	0.000000	1.000000
All	0.444444	0.333333	0.222222

Less usage indicates a preference for KP281 whereas high usage>=5 days per week indicates a clear preference for a high end product and those people are more likely to buy KP781. Customers with average usage show a scope for upselling the product.

7. Mile_bins vs Treadmill Product

Product Mile_bins	KP281	KP481	KP781	All
0-50	12	5	0	17
50-100	50	39	8	97
100-150	16	13	9	38
150-200	2	2	18	22
200+	0	1	5	6
All	80	60	40	180

Product Mile_bins	KP281	KP481	KP781
0-50	0.705882	0.294118	0.000000
50-100	0.515464	0.402062	0.082474
100-150	0.421053	0.342105	0.236842
150-200	0.090909	0.090909	0.818182
200+	0.000000	0.166667	0.833333
All	0.444444	0.333333	0.222222

Similar analysis as that in the case of usage.

8. Fitness vs Treadmill Product

Product Fitness	KP281	KP481	KP781	All
1	1	1	0	2
2	14	12	0	26
3	54	39	4	97
4	9	8	7	24
5	2	0	29	31
All	80	60	40	180

Product Fitness	KP281	KP481	KP781
1	0.500000	0.500000	0.000000
2	0.538462	0.461538	0.000000
3	0.556701	0.402062	0.041237
4	0.375000	0.333333	0.291667
5	0.064516	0.000000	0.935484
All	0.444444	0.333333	0.222222

Maximum people have average fitness and such people can be easily persuaded to buy a better product than the one they prefer. Therein lies the scope for profitability. People with low fitness or usage won't be flexible enough to switch their preference and fit people or high usage people are already buying the premium product.

Multivariate Analysis and Conditional probablities

At this stage, since we are trying to understand which customer has a higher likelihood to buy a particular kind of treadmill few things must be kept in mind. The end goal is to understand the customer characteristics which separate one kind of customer segment from another. This, when achieved will act as a guideline for sales executive to market the right kind of product to the targeted customer.

Age, gender and fitness are the most apparent factors basis which one can easily categorize a kind of customer. Other attributes like information about his MaritalStatus, income, education etc would require a customer to fill some kind of a form, which may or may not be possible at the time of sale or in the first meeting itself. The other more important and relevant matrix that one can easily get information about is the usage and the miles that he is planning to run in a week.

Basis this understanding, we have calculated various conditional probabilities to arrive at customer profiling.

We are just enlisting the customer profiles who are most likely to buy KP281, KP481 and KP781 respectively.

Profiling for KP781:

This product is the most premium and has a straight 1000$ difference from its cheapest counterpart KP281. However, this product will have the highest profit margin and hence will the most popular to sell. Typical characteristics of customers who are buying this product segment are as follows:

1. Males with average and above average usage (3-5) and females with very high usage(5).

2. Clear preference in very high income groups (80000p.a. and above) irrespective of gender and age <40 years for females.

3. Highly educated females with 20 or more than 20 years of education. Males with medium level of education also buy this product.

4. If a person belongs to very high income category, then irrespective of the education he has received, he will clearly buy this product. However, as the education level increases we see that some medium level income group also buy this product.

5. People with more than 16 years of education and planning to run more than 150 miles a week clearly prefer this product over all others.

Profiling for KP481

KP481 is the second most popular product amongst customers. It sells for $1750 and is just $250 more than its cheapest counterpart KP281. Hence, it is one of the best upselling option for customers who are considering to buy KP281.

The company must strive at training its customer/sales executives to sell this product aggressively and also announce some incentives or discounts on the same to improve the total profitability of the company.

1. Popular amongst females and males in the age group of 30-40 yrs.

2. Popular amongst females with avg or above avg usage(3-4), in males, the popularity of this product is seen even if the usage is only at the scale of 2.

3. Popular with individuals with less education but high income group.

4. Popular with single females.

5. People with less education, low to medium income groups with plans to run more than 100 miles per week buy this treadmill.

Profiling for KP281

1. Super favourite amongst women of all ages.

2. Popular amongst people with less education, low budget and low to medium usage.

Recommendations for Aerofit

1. First and foremost, the sales executives at Aerofit must be trained to understand their target customer basis the customer profiling that has been done in this study.

2. Each executive must be able to identify which product is the customer going to buy within the first 5 minutes of his interaction with the prospective customer. Once that is done, then clearly the executive must spend time in either upselling a product or else in quickening the decision making process of the customer (in case he is going to buy KP781).

3. As a general guideline, Aerofit must use this customer profiling to strongly upsell KP481 to all the customers especially females who fall in the medium income group. Instead of KP281 being the most sold product, KP481 should be the most popular product being sold.

4. To people who have more than 16 years of education and who have the capacity to appreciate the advanced features of KP781 and fall in the medium to high income group should be targeted for upselling KP781.

5. Various strategies can be devised to upsell these products like giving couple discounts or rolling out women's day special discounts to promote KP481 at Aerofit level to facilitate the same.