

## #About Aerofit

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.

## #Business Problem

The market research team at AeroFit wants 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. The team decides to investigate whether there are differences across the product with respect to customer characteristics.

Perform descriptive analytics to create a customer profile for each AeroFit treadmill product by developing appropriate tables and charts. For each AeroFit treadmill product, construct two-way contingency tables and compute all conditional and marginal probabilities along with their insights/impact on the business.

## #Product Portfolio:

- 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.

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

!wget
https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/001/125
/original/aerofit_treadmill.csv?1639992749

--2023-10-31 13:47:42--
https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/001/125
/original/aerofit_treadmill.csv?1639992749
Resolving d2beiqkhq929f0.cloudfront.net
(d2beiqkhq929f0.cloudfront.net)... 18.65.227.27, 18.65.227.25,
18.65.227.110, ...
Connecting to d2beiqkhq929f0.cloudfront.net
(d2beiqkhq929f0.cloudfront.net)|18.65.227.27|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 7279 (7.1K) [text/plain]
Saving to: 'aerofit_treadmill.csv?1639992749'

aerofit_treadmill.c 100%[=====>] 7.11K --.-KB/s in
0s
```

2023-10-31 13:47:44 (1.67 GB/s) - 'aerofit\_treadmill.csv?1639992749'  
saved [7279/7279]

```
df = pd.read_csv('aerofit_treadmill.csv?1639992749')
df.head()
```

	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					

```
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
```

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

	Product	Age	Gender	Education	MaritalStatus
Usage \					
count	180	180.000000	180	180.000000	180
180.000000					
unique	3	NaN	2	NaN	2
NaN					
top	KP281	NaN	Male	NaN	Partnered
NaN					
freq	80	NaN	104	NaN	107

NaN					
mean	NaN	28.788889	NaN	15.572222	NaN
3.455556					
std	NaN	6.943498	NaN	1.617055	NaN
1.084797					
min	NaN	18.000000	NaN	12.000000	NaN
2.000000					
25%	NaN	24.000000	NaN	14.000000	NaN
3.000000					
50%	NaN	26.000000	NaN	16.000000	NaN
3.000000					
75%	NaN	33.000000	NaN	16.000000	NaN
4.000000					
max	NaN	50.000000	NaN	21.000000	NaN
7.000000					

	Fitness	Income	Miles
count	180.000000	180.000000	180.000000
unique	NaN	NaN	NaN
top	NaN	NaN	NaN
freq	NaN	NaN	NaN
mean	3.311111	53719.577778	103.194444
std	0.958869	16506.684226	51.863605
min	1.000000	29562.000000	21.000000
25%	3.000000	44058.750000	66.000000
50%	3.000000	50596.500000	94.000000
75%	4.000000	58668.000000	114.750000
max	5.000000	104581.000000	360.000000

```
print(df.isnull().any())
```

Product	False
Age	False
Gender	False
Education	False
MaritalStatus	False
Usage	False
Fitness	False
Income	False
Miles	False

dtype: bool

## Insights

- The dataset have 180 rows and 9 columns.
- There is no missing value in the dataset.
- There is 3 unique products in the dataset.
- There is 2 (Male & Female) genders in the datset and MaritalStatus is either Single or Partered.

- KP281 is most frequent product.
- 18 - 50 are the age group of users. 28.79 are the average age of users.
- Males are the top customers. 104 out of 180 customers are male.
- Most of the people have an average education of 16 years.
- The products are used by customers in a week for at least 2 days and at most 7 days, and the mean value of usage is 3.46..
- Self-rated fitness on a 1-to-5 scale, the average value is 3.31.
- Minimum annual income of customers is 29562 USD and maximum is 104581 USD and average income is 53719.58 USD.
- Average usage of tread mills per week is 103.19 miles and minimum usage is 21 miles and maximum is 360 miles.

```
df['Product'].unique()
array(['KP281', 'KP481', 'KP781'], dtype=object)
```

## Insights

- The products in the dataset are KP281, KP481, KP781

## Product portfolio

- 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.

## Univariate Analysis

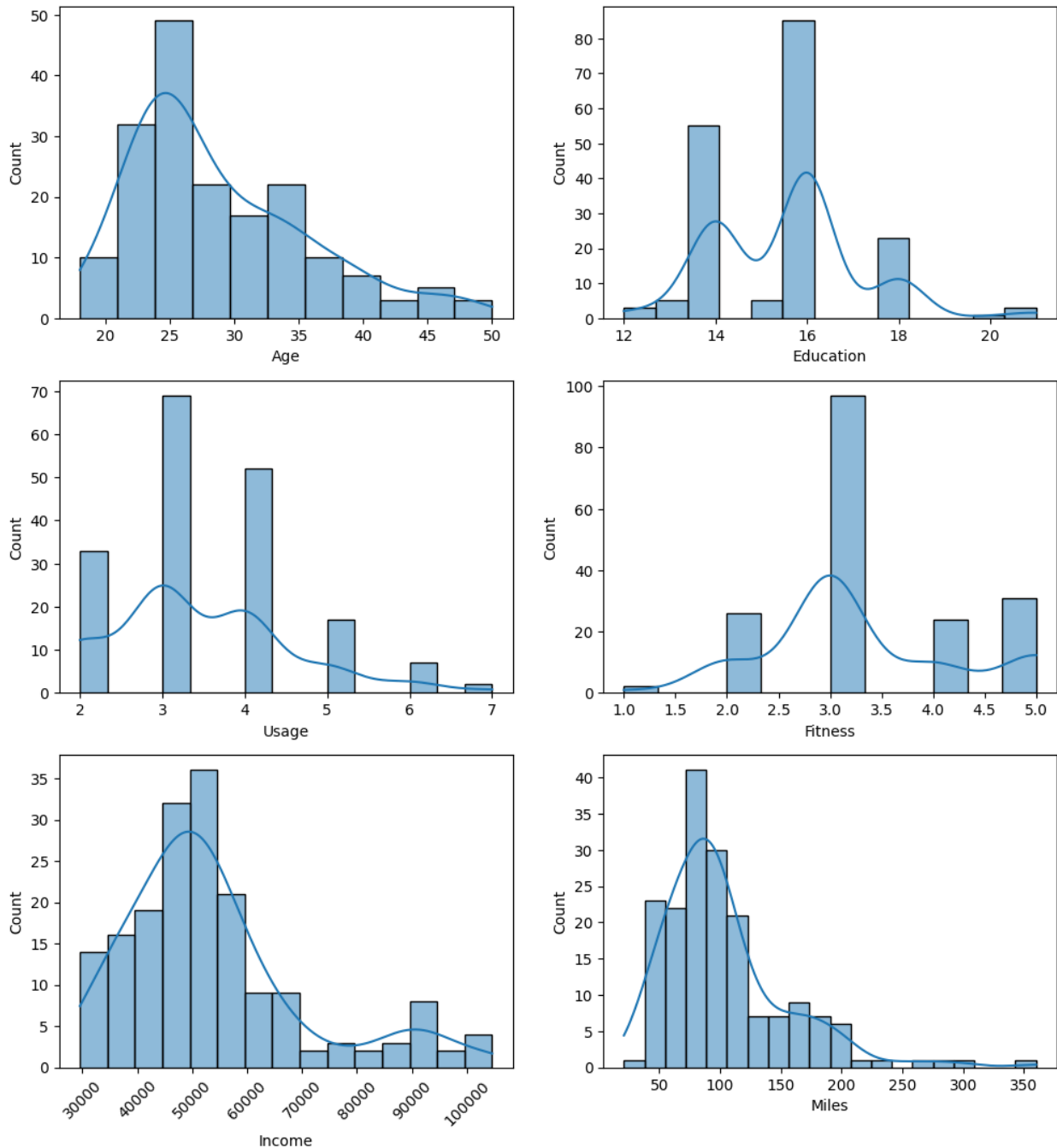
## Understanding the distribution of the data for the quantitative attributes:

1. Age
2. Education
3. Usage
4. Fitness
5. Income
6. Miles

```
fig, axis = plt.subplots(nrows=3, ncols=2, figsize=(12, 10))
fig.subplots_adjust(top=1.1)

sns.histplot(data=df, x="Age", kde=True, ax=axis[0,0])
sns.histplot(data=df, x="Education", kde=True, ax=axis[0,1])
sns.histplot(data=df, x="Usage", kde=True, ax=axis[1,0])
sns.histplot(data=df, x="Fitness", kde=True, ax=axis[1,1])
income_axs=sns.histplot(data=df, x="Income", kde=True, ax=axis[2,0])
income_axs.set_xticklabels(income_axs.get_xticklabels(), rotation=45)
sns.histplot(data=df, x="Miles", kde=True, ax=axis[2,1])
plt.show()
```

```
<ipython-input-16-153941d1bf94>:9: UserWarning: FixedFormatter should
only be used together with FixedLocator
    income_axs.set_xticklabels(income_axs.get_xticklabels(),
rotation=45)
```

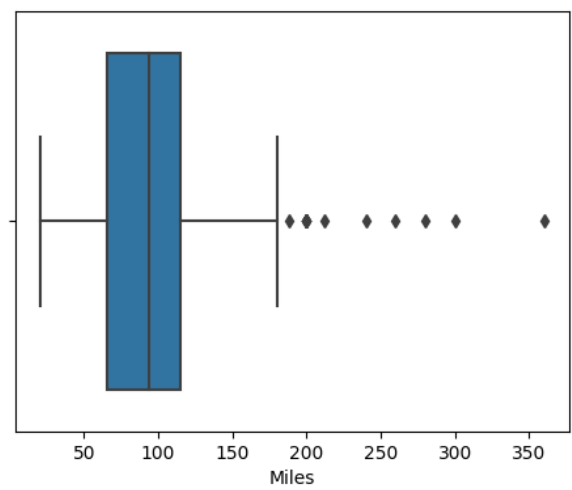
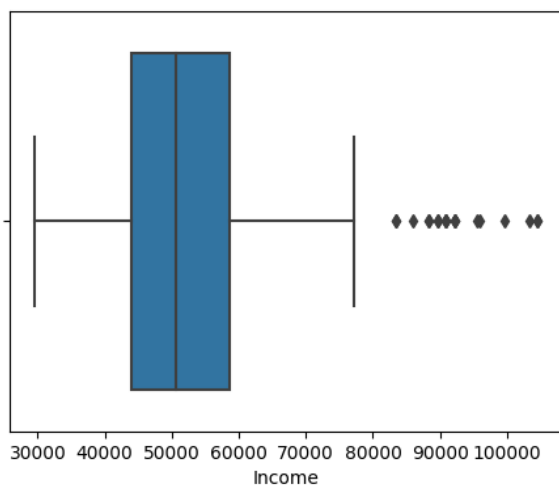
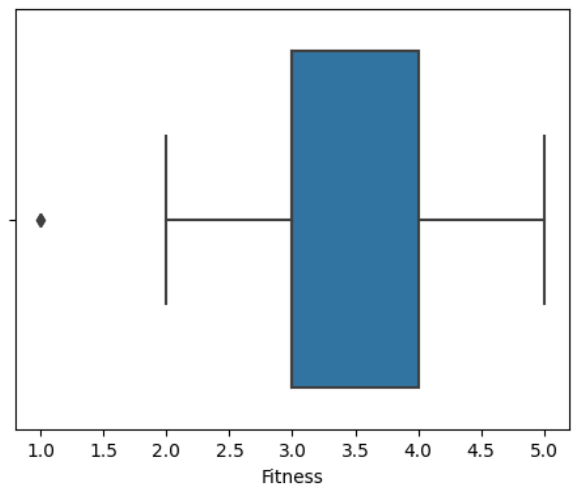
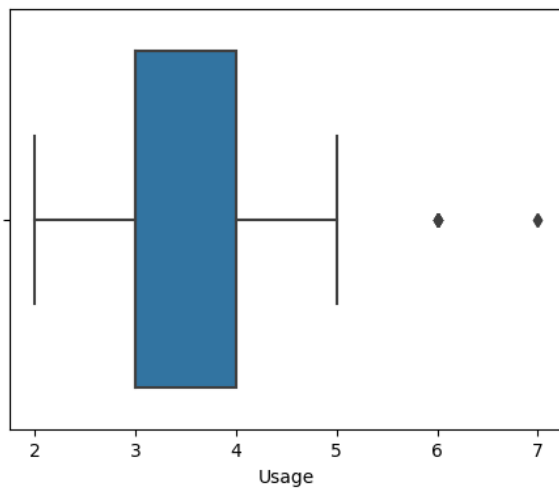
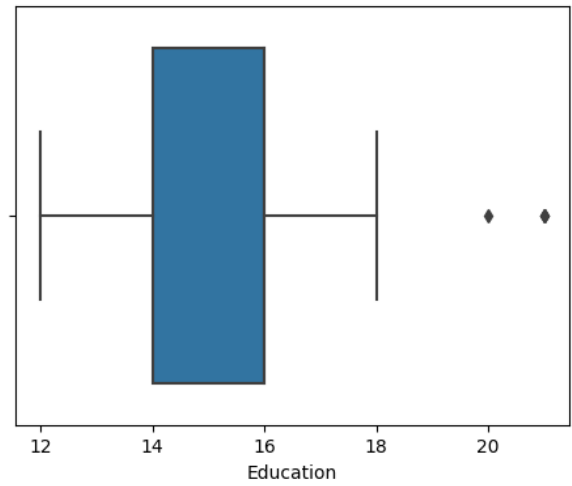
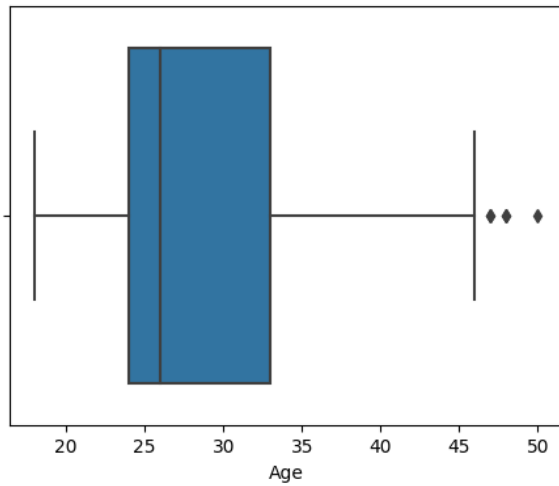


##Finding Outliers using box plot

```
fig, axis = plt.subplots(nrows=3, ncols=2, figsize=(12, 10))
fig.subplots_adjust(top=1.2)

sns.boxplot(data=df, x="Age", orient='h', ax=axis[0,0])
sns.boxplot(data=df, x="Education", orient='h', ax=axis[0,1])
```

```
sns.boxplot(data=df, x="Usage", orient='h', ax=axis[1,0])
sns.boxplot(data=df, x="Fitness", orient='h', ax=axis[1,1])
sns.boxplot(data=df, x="Income", orient='h', ax=axis[2,0])
sns.boxplot(data=df, x="Miles", orient='h', ax=axis[2,1])
plt.show()
```



## Insights

- Age, Education, Usage & Fitness have few outliers.
- Income and Miles have more outliers.

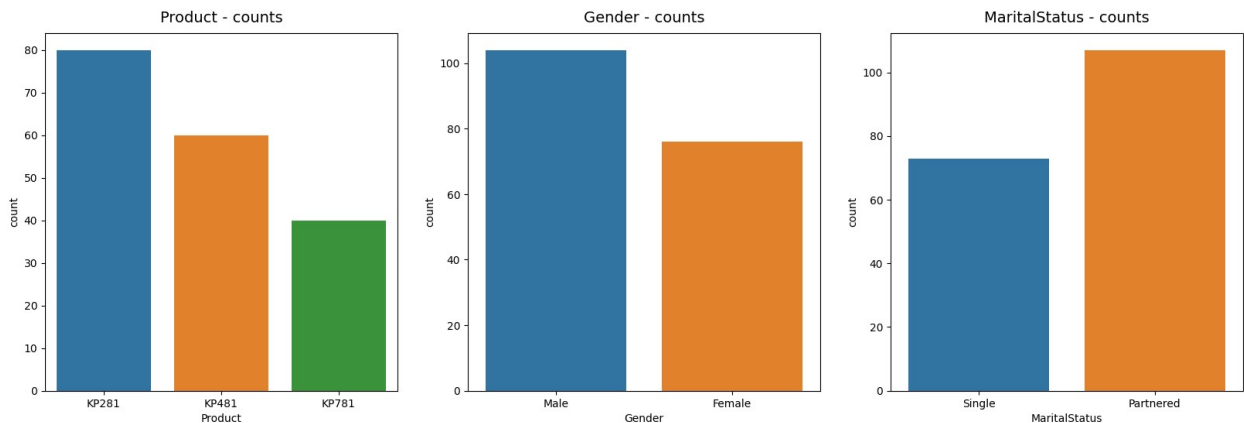


## Understanding the distribution of the data for the qualitative attributes:

1. Product
2. Gender
3. MaritalStatus

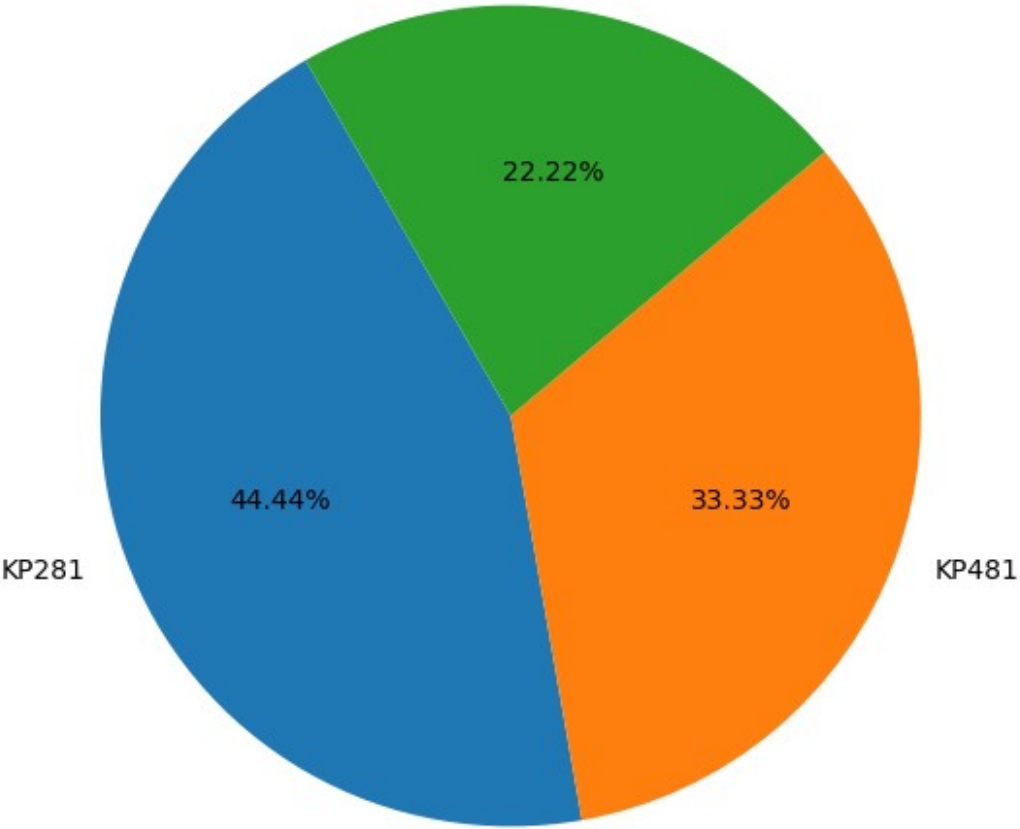
```
fig, axs = plt.subplots(nrows=1, ncols=3, figsize=(20, 6))
sns.countplot(data=df, x='Product', ax=axs[0])
sns.countplot(data=df, x='Gender', ax=axs[1])
sns.countplot(data=df, x='MaritalStatus', ax=axs[2])

axs[0].set_title("Product - counts", pad=10, fontsize=14)
axs[1].set_title("Gender - counts", pad=10, fontsize=14)
axs[2].set_title("MaritalStatus - counts", pad=10, fontsize=14)
plt.show()
```

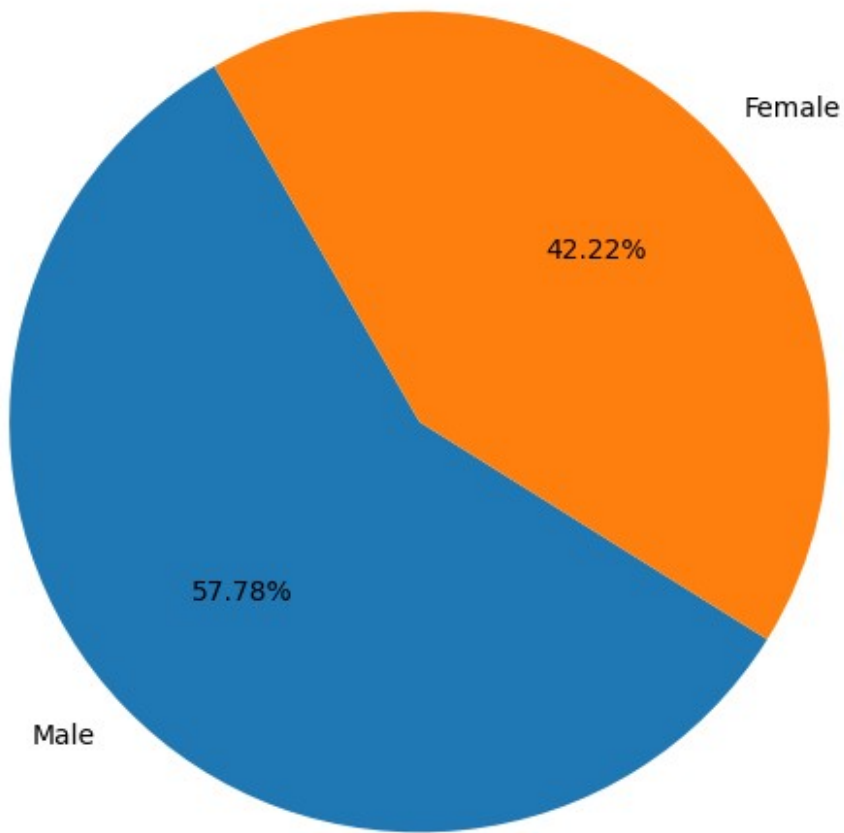


```
tables = ['Product', 'Gender', 'MaritalStatus']
for table in tables:
    table_counts = df[table].value_counts()
    plt.figure(figsize=(6, 6))
    plt.pie(table_counts, labels=table_counts.index, autopct='%1.2f%%',
            startangle=120)
    plt.axis('equal')
    plt.title(f'Distribution of {table}')
    plt.show()
```

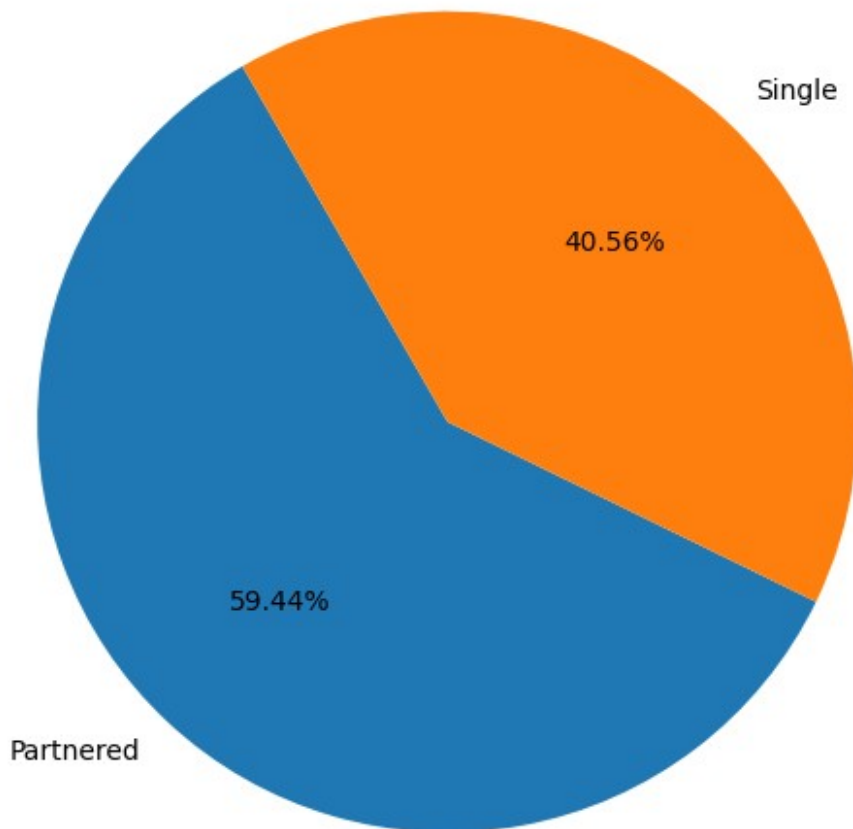
Distribution of Product  
KP781



Distribution of Gender



Distribution of MaritalStatus



## Observation

### ## Products

- 44.44% of the customers have purchased KP2821 treadmill.
- 33.33% of the customers have purchased KP481 treadmill.
- 22.22% of the customers have purchased KP781 treadmill.

## Gender

- 57.78% of the customers are Male.
- 42.22 of the customers are Female.

## MaritalStatus

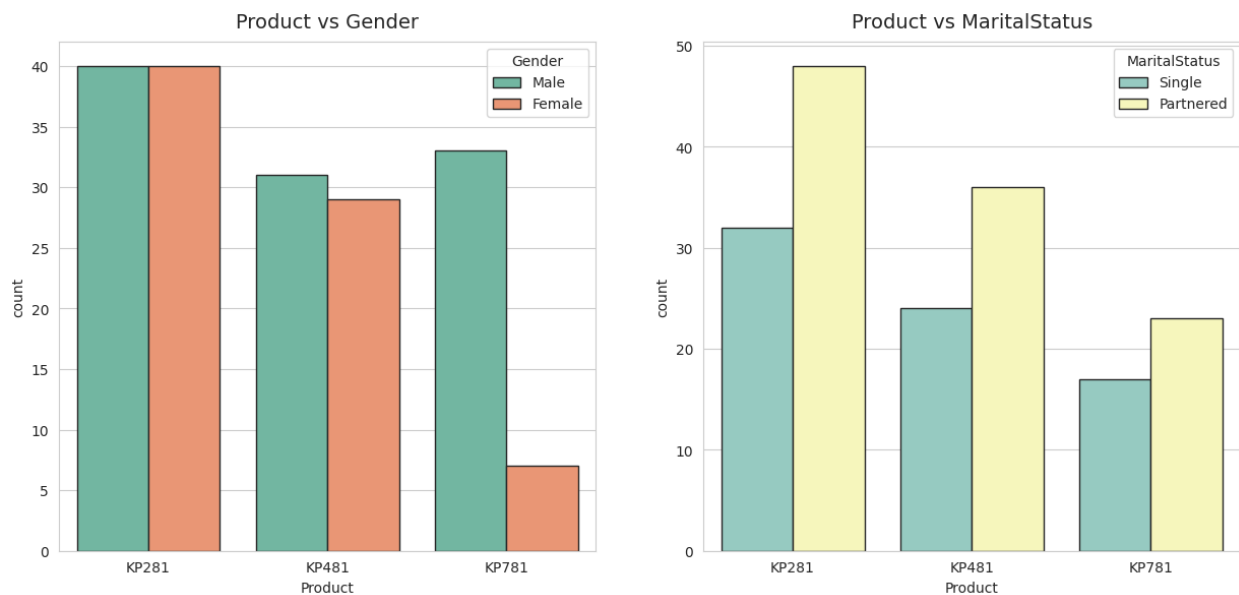
- 59.44% of the customers are Partnered.
- 40.56% of the customers are Single.

## Insights

- KP281 is the most frequent product.
- There are more Males in the data than Females.
- More Partnered persons are there in the data.

#Bivariate Analysis Checking if features - Gender or MaritalStatus have any effect on the product purchased.

```
sns.set_style(style='whitegrid')
fig, axs = plt.subplots(nrows=1, ncols=2, figsize=(15, 6.5))
sns.countplot(data=df, x='Product', hue='Gender', edgecolor="0.15",
palette='Set2', ax=axs[0])
sns.countplot(data=df, x='Product', hue='MaritalStatus',
edgecolor="0.15", palette='Set3', ax=axs[1])
axs[0].set_title("Product vs Gender", pad=10, fontsize=14)
axs[1].set_title("Product vs MaritalStatus", pad=10, fontsize=14)
plt.show()
```



## Insights

### Product vs Gender

- The KP281 treadmill was bought in equal numbers by both males and females.
- For the KP481 treadmill, the number of males who purchased it was slightly greater than the number of females.
- Most customers who bought the KP781 treadmill are males.

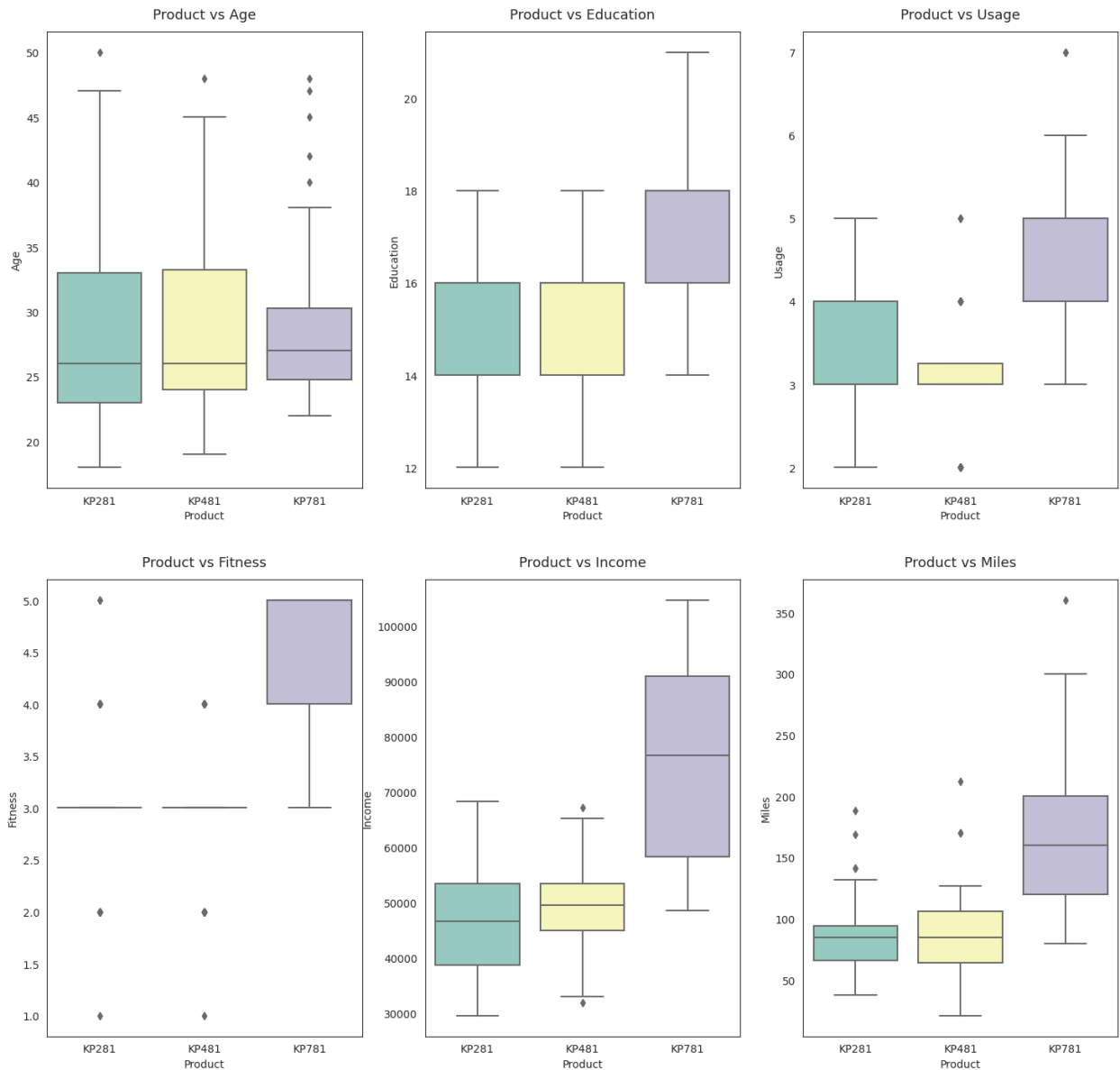
## Product vs MaritalStatus

- Customer who is Partnered, is more likely to purchase the product.

Finding if following features have any effect on the product purchased:

1. Age
2. Education
3. Usage
4. Fitness
5. Income
6. Miles

```
attrs = ['Age', 'Education', 'Usage', 'Fitness', 'Income', 'Miles']
sns.set_style("white")
fig, axs = plt.subplots(nrows=2, ncols=3, figsize=(18, 12))
fig.subplots_adjust(top=1.2)
count = 0
for i in range(2):
    for j in range(3):
        sns.boxplot(data=df, x='Product', y=attrs[count], ax=axs[i,j],
palette='Set3')
        axs[i,j].set_title(f"Product vs {attrs[count]}", pad=12,
        fontsize=13)
        count += 1
```



# Insights

## ###Product vs Age

- Customers purchasing treadmill KP281 & KP481 are having same Age median value.
- Customers whose age lies between 25-30, are more likely to buy KP781 treadmill.

## Product vs Education

- Customers whose Education is greater than 16, have more chances to purchase the KP781 treadmill.

- While the customers with Education less than 16 have equal chances of purchasing KP281 or KP481.

### Product vs Usage

- Customers who are planning to use the treadmill greater than 4 times a week, are more likely to purchase the KP781 treadmill.
- While the other customers are likely to purchasing KP281 or KP481.

### Product vs Fitness

- The more the customer is fit (fitness  $\geq 4$ ), higher the chances of the customer to purchase the KP781 treadmill.

### Product vs Income

- Higher the Income of the customer (Income  $\geq 60000$ ), higher the chances of the customer to purchase the KP781 treadmill.

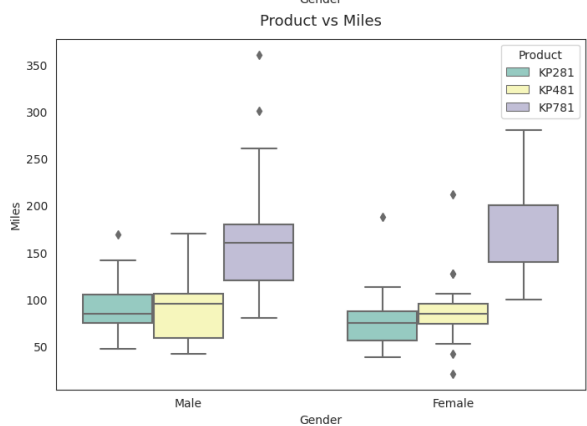
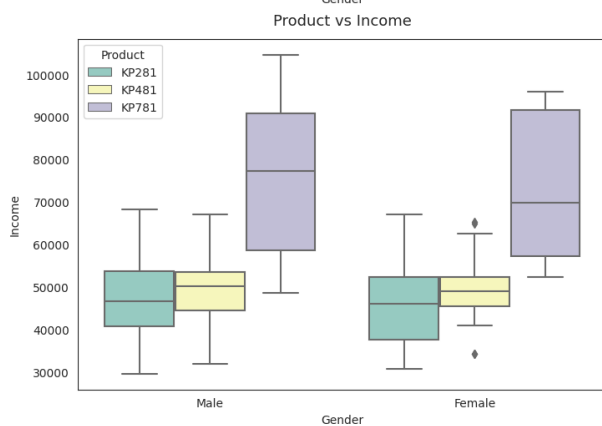
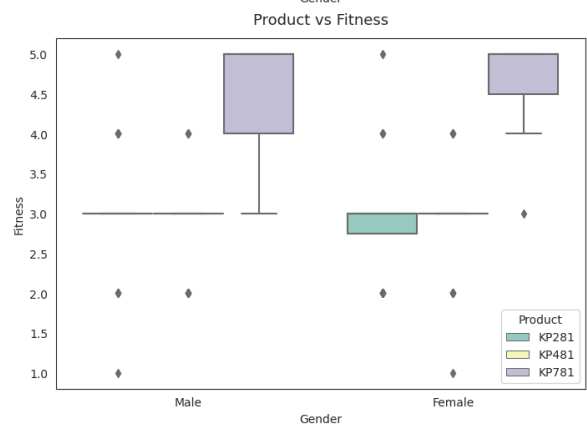
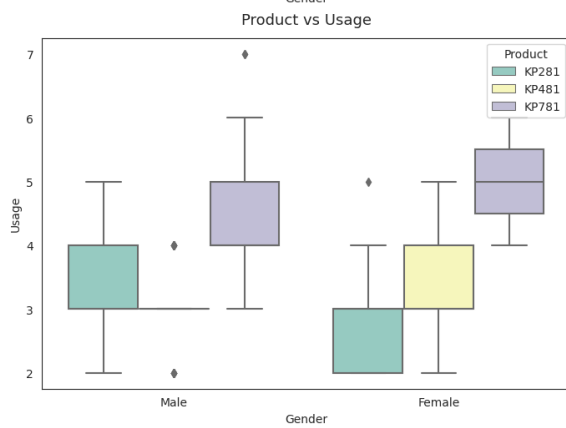
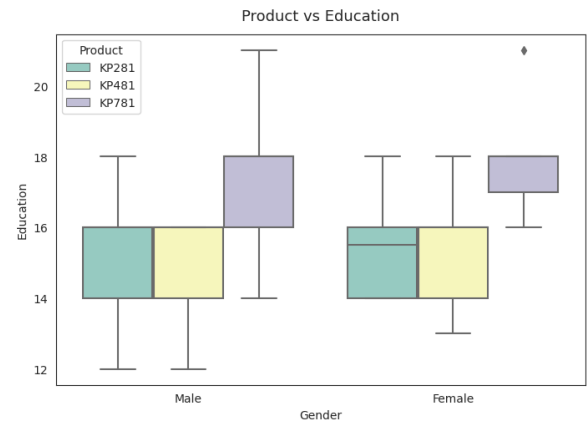
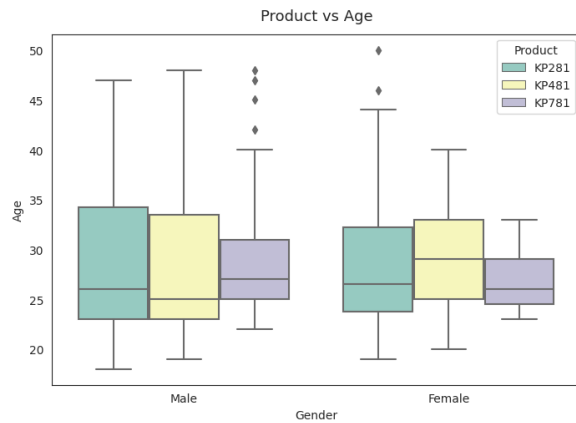
### Product vs Miles

- If the customer expects to walk/run greater than 120 Miles per week, it is more likely that the customer will buy KP781 treadmill.

## Multivariate Analysis

```
attrs = ['Age', 'Education', 'Usage', 'Fitness', 'Income', 'Miles']
sns.set_style("white")
fig, axs = plt.subplots(nrows=3, ncols=2, figsize=(18, 12))
fig.subplots_adjust(top=1.3)
count = 0
for i in range(3):
    for j in range(2):
        sns.boxplot(data=df, x='Gender', y=attrs[count],
hue='Product', ax=axs[i,j], palette='Set3')
        axs[i,j].set_title(f"Product vs {attrs[count]}", pad=12,
        fontsize=13)
        count += 1
```





## Insights

- Females planning to use treadmill 3-4 times a week, are more likely to buy KP481 product.

# Computing Marginal & Conditional Probabilities

- Marginal Probability

```
df['Product'].value_counts(normalize=True)
```

```
KP281    0.444444
KP481    0.333333
KP781    0.222222
Name: Product, dtype: float64
```

## Insights

- Probability of buying treadmill KP281 IS 44.44%
- Probability of buying treadmill KP481 IS 33.33%
- Probability of buying treadmill KP781 IS 22.22%

- Conditional probability

```
def p_prod_given_gender(gender, print_marginal=False):
    if gender != "Female" and gender != "Male":
        return "Invalid gender value."

    df1 = pd.crosstab(index=df['Gender'], columns=[df['Product']])
    p_781 = df1['KP781'][gender] / df1.loc[gender].sum()
    p_481 = df1['KP481'][gender] / df1.loc[gender].sum()
    p_281 = df1['KP281'][gender] / df1.loc[gender].sum()

    if print_marginal:
        print(f"P(Male): {df1.loc['Male'].sum()/len(df):.2f}")
        print(f"P(Female): {df1.loc['Female'].sum()/len(df):.2f}\n")

    print(f"P(KP781/{gender}): {p_781:.2f}")
    print(f"P(KP481/{gender}): {p_481:.2f}")
    print(f"P(KP281/{gender}): {p_281:.2f}\n")

p_prod_given_gender('Male', True)
p_prod_given_gender('Female')

P(Male): 0.58
P(Female): 0.42

P(KP781/Male): 0.32
P(KP481/Male): 0.30
P(KP281/Male): 0.38

P(KP781/Female): 0.09
P(KP481/Female): 0.38
```

P(KP281/Female): 0.53

## Insights

- Probability of male buying any treadmill is 58 % and female is 42%.

### Probability of buying treadmills by gender

- Probability of buying treadmill KP781 by a male is 32% and female is 9%.
- Probability of buying treadmill KP481 by a male is 30% and female is 38%.
- Probability of buying treadmill KP281 by a male is 38% and female is 53%.

## Probability of each product given MaritalStatus

```
def p_prod_given_mstatus(status, print_marginal=False):
    if status != "Single" and status != "Partnered":
        return "Invalid marital status value."

    df1 = pd.crosstab(index=df['MaritalStatus'],
                      columns=df['Product'])
    p_781 = df1['KP781'][status] / df1.loc[status].sum()
    p_481 = df1['KP481'][status] / df1.loc[status].sum()
    p_281 = df1['KP281'][status] / df1.loc[status].sum()

    if print_marginal:
        print(f"P(Single): {df1.loc['Single'].sum()/len(df):.2f}")
        print(f"P(Partnered): {df1.loc['Partnered'].sum()/len(df):.2f}\n")

        print(f"P(KP781/{status}): {p_781:.2f}")
        print(f"P(KP481/{status}): {p_481:.2f}")
        print(f"P(KP281/{status}): {p_281:.2f}\n")

p_prod_given_mstatus('Single', True)
p_prod_given_mstatus('Partnered')
```

P(Single): 0.41  
P(Partnered): 0.59

P(KP781/Single): 0.23  
P(KP481/Single): 0.33

$P(\text{KP281/Single}) : 0.44$

$P(\text{KP781/Partnered}) : 0.21$

$P(\text{KP481/Partnered}) : 0.34$

$P(\text{KP281/Partnered}) : 0.45$

## Insights

- Probability of single buying any treadmill is 41 % and partnered is 42%.

### Probability of buying treadmills by single and partnered

- Probability of buying treadmill KP781 by a single is 23% and partnered is 21%.
- Probability of buying treadmill KP481 by a single is 33% and partnered is 34%.
- Probability of buying treadmill KP281 by a single is 44% and partnered is 45%.

## Recommendations

- Since KP281 treadmill is the most popular, consider promoting it further.
- Partnered individuals are more likely to purchase treadmills. Tailor marketing strategies to this group.
- Customers purchasing KP281 and KP481 treadmills have similar age preferences. Ensure your marketing appeals to this age group.
- For the KP781 treadmill, target customers aged 25-30. Focus on customers with education levels greater than 16 for the KP781 treadmill.
- For customers with education levels less than 16, market both KP281 and KP481 treadmills.
- Customers planning to use the treadmill more than 4 times a week are more likely to buy the KP781 treadmill.
- Target customers with fitness levels of 4 or higher for the KP781 treadmill.
- For higher-income customers (income  $\geq 60000$ ), emphasize the KP781 treadmill.
- Customers expecting to walk/run more than 120 miles per week are more likely to buy the KP781 treadmill.

- Females planning to use treadmills 3-4 times a week are more likely to buy the KP481 treadmill.
- Since males have a higher probability of buying any treadmill, continue to cater to this demographic.
- Focus on marketing the KP781 treadmill to males, as they have a higher likelihood of purchasing it.
- For females, promote the KP481 treadmill, as they have a higher probability of buying it.
- For both single and partnered customers, the probability of buying treadmills is close. Ensure marketing efforts are balanced for both groups.

## Recommended Target Customer for different treadmills

### 1. KP281

For this product the customeras who belong to any of these category is the best option :

- Both Males and Female
- Income between 40000 USD - 55000 USD

### 1. KP481

For this product the customeras who belong to any of these category is the best option :

- Males and Females
- Females planning to use treadmills 3-4 times a week
- Income 45000 USD - 55000 USD

### 1. KP781

For this product the customeras who belong to any of these category is the best option :

- Male
- Age 25 - 30
- Education 16 or more
- Income  $\geq$  60000 USD
- Targetting fitness 4 - 5
- Usage/Planned usage 4 times or more a week
- Usage/Planned usage 120 miles or more per week.