untitled0

March 20, 2024

#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. 1. Perform descriptive analytics to create a customer profile for each AeroFit treadmill product by developing appropriate tables and charts. 2. 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. Dataset The company collected the data on individuals who purchased a treadmill from the AeroFit stores during the prior three months. The dataset has the following features:

[4]:
```
!gdown https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/001/125/
↪original/aerofit_treadmill.csv?1639992749
```

```
Downloading…
From: https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/001/125/ori
ginal/aerofit_treadmill.csv?1639992749
To: /content/aerofit_treadmill.csv?1639992749
100% 7.28k/7.28k [00:00<00:00, 12.3MB/s]
```

[2]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[3]:
```
df = pd.read_csv('https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/
↪000/001/125/original/aerofit_treadmill.csv?1639992749')
```

[7]:
```
df
```

[7]:

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	\
0	KP281	18	Male	14	Single	3	4	29562	
1	KP281	19	Male	15	Single	2	3	31836	
2	KP281	19	Female	14	Partnered	4	3	30699	
3	KP281	19	Male	12	Single	3	3	32973	

```
4     KP281   20    Male          13     Partnered      4         2   35247
..      …   …     …          …            …    …       …      …
175   KP781   40    Male          21        Single      6         5   83416
176   KP781   42    Male          18        Single      5         4   89641
177   KP781   45    Male          16        Single      5         5   90886
178   KP781   47    Male          18     Partnered      4         5  104581
179   KP781   48    Male          18     Partnered      4         5   95508

     Miles
0      112
1       75
2       66
3       85
4       47
..     …
175    200
176    200
177    160
178    120
179    180

[180 rows x 9 columns]
```

[6]:
```
df.head(10)
```

[6]:

	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

[8]:
```
df.describe(include='object').T
```

[8]:

	count	unique	top	freq
Product	180	3	KP281	80
Gender	180	2	Male	104
MaritalStatus	180	2	Partnered	107

[9]:
```
df.shape
```

[9]:
```
(180, 9)
```

This tells us the data is made of 180 rows and 9 columns

- Columns consist:

1. Products
2. Age
3. Gender
4. Education
5. MaritalStatus
6. Usage
7. Fitness
8. Income
9. Miles

[10]:
```
df.describe().T
```

[10]:

	count	mean	std	min	25%	50%	\
Age	180.0	28.788889	6.943498	18.0	24.00	26.0	
Education	180.0	15.572222	1.617055	12.0	14.00	16.0	
Usage	180.0	3.455556	1.084797	2.0	3.00	3.0	
Fitness	180.0	3.311111	0.958869	1.0	3.00	3.0	
Income	180.0	53719.577778	16506.684226	29562.0	44058.75	50596.5	
Miles	180.0	103.194444	51.863605	21.0	66.00	94.0	

	75%	max
Age	33.00	50.0
Education	16.00	21.0
Usage	4.00	7.0
Fitness	4.00	5.0
Income	58668.00	104581.0
Miles	114.75	360.0

This statistical summary gives us the

1. count: The number of non-null values in each column.
2. mean: The average value of each column.
3. std: The standard deviation, a measure of the dispersion of values around the mean.
4. min: The minimum value in each column.
5. 25%: The first quartile, also known as the lower quartile or 25th percentile.
6. 50%: The second quartile, also known as the median or 50th percentile.
7. 75%: The third quartile, also known as the upper quartile or 75th percentile.
8. max: The maximum value in each column.

[11]:
```
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
```

This gives us

1. Index: It tells about the index type and number of entries.
2. Columns: It lists the column names and the number of non-null values in each column.
3. Data Type: For each column, it displays the data type (e.g., integer, float, object/string).
4. Memory Usage: It provides an estimate of the memory usage of the DataFrame.

[12]:
```
df['Product'].unique()
```

[12]:
```
array(['KP281', 'KP481', 'KP781'], dtype=object)
```

gives us 3 models (their data type is obejct)

[13]:
```
df['Age'].unique()
```

[13]:
```
array([18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34,
       35, 36, 37, 38, 39, 40, 41, 43, 44, 46, 47, 50, 45, 48, 42])
```

[14]:
```
age_count =df['Age'].value_counts(normalize=True)*100
age_count.round(0)
```

[14]:
```
25    14.0
23    10.0
24     7.0
26     7.0
28     5.0
35     4.0
33     4.0
30     4.0
38     4.0
21     4.0
22     4.0
27     4.0
31     3.0
34     3.0
```

```
29     3.0
20     3.0
40     3.0
32     2.0
19     2.0
48     1.0
37     1.0
45     1.0
47     1.0
46     1.0
50     1.0
18     1.0
44     1.0
43     1.0
41     1.0
39     1.0
36     1.0
42     1.0
Name: Age, dtype: float64
```

This gives us the mode of the data that is 25 as the age therefore telling us that most number of buyers are 25 years old.

[15]:
```
age_count[(age_count.index>=20)&(age_count.index<=35)].sum().round(0)
```

[15]:
```
82.0
```

We can clearly see that 82% of all customers fall within the age range of 20 to 35

[16]:
```
df['Gender'].value_counts(normalize=True)*100
```

[16]:
```
Male      57.777778
Female    42.222222
Name: Gender, dtype: float64
```

58% customers are male while the other 42% are females

[17]:
```
df['MaritalStatus'].value_counts(normalize=True)*100
```

[17]:
```
Partnered    59.444444
Single       40.555556
Name: MaritalStatus, dtype: float64
```

60% customers have a marital partner

[18]:
```
df['Income'].describe()
```

[18]:
```
count       180.000000
mean      53719.577778
std       16506.684226
min       29562.000000
25%       44058.750000
50%       50596.500000
75%       58668.000000
max      104581.000000
Name: Income, dtype: float64
```

[19]:
```
count_of_product = df['Product'].value_counts(normalize=True)*100.0
print(count_of_product.round(3))
```

```
KP281    44.444
KP481    33.333
KP781    22.222
Name: Product, dtype: float64
```

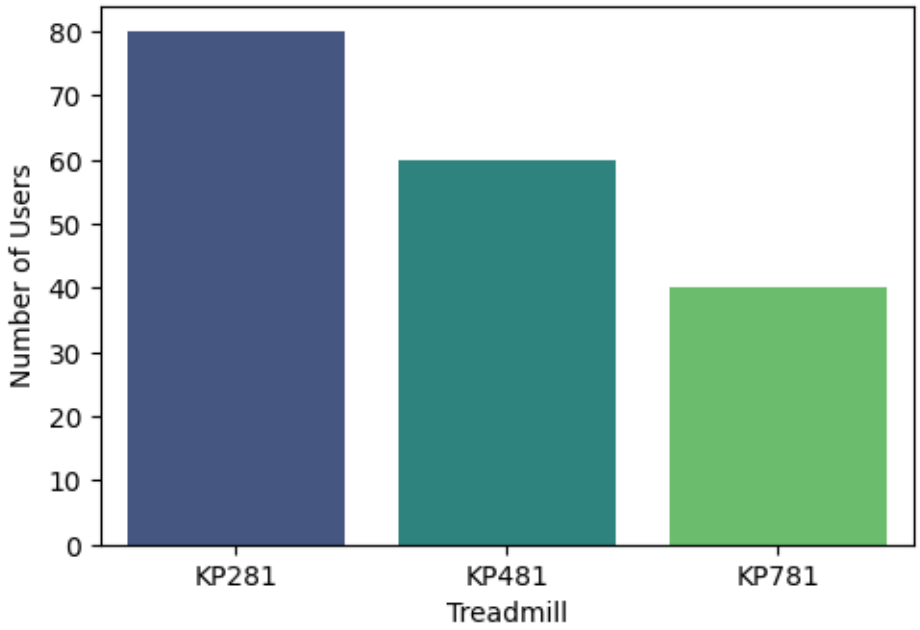
0.1 *We can clearly see that the KP281 is preferred by 44% of all customers while KP481 is preferred by 33% and KP781 22% respectively.*

[60]:
```
plt.figure(figsize=(12, 8))
plt.subplot(2, 2, 1)
sns.countplot(data=df, x=df['Product'], palette='viridis')
plt.xlabel('Treadmill')
plt.ylabel('Number of Users')
plt.show()
```

```
<ipython-input-60-85d5f751df35>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.countplot(data=df, x=df['Product'], palette='viridis')
```



It is Clearly visible that KP281 is the most sold model while KP781 does not perform that well in the market Hence pushing for KP281 even more and positioning KP781 in the right markets

[21]:
```
df['Age Group']= pd.cut(df['Age'],bins=[17,29,39,50],labels=['young','middle␣
↪aged','old'])
df['Age Group'].value_counts()
```

[21]:
```
young          113
middle aged     50
old             17
Name: Age Group, dtype: int64
```

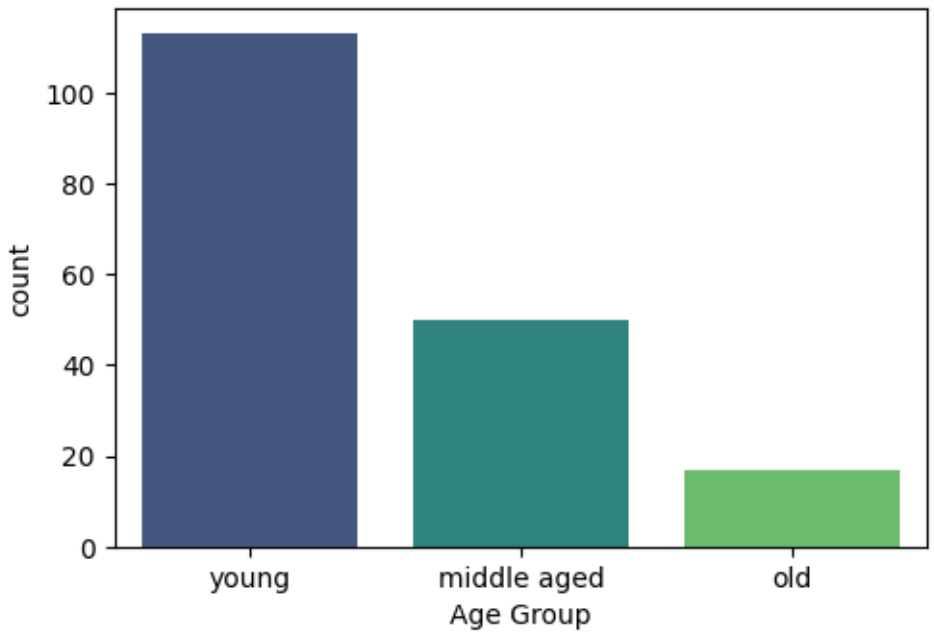
[22]:
```
plt.figure(figsize=(12, 8))
plt.subplot(2, 2, 1)
sns.countplot(data=df, x=df['Age Group'], palette='viridis')
```

```
<ipython-input-22-f3333deceb9c>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.countplot(data=df, x=df['Age Group'], palette='viridis')
```

[22]:
```
<Axes: xlabel='Age Group', ylabel='count'>
```



Majority of users are young hence more gen-z and millenial targeted marketing can benefit while also assisting the older individuals by educating them about the advantages of regualr walking in middle and old age

[23]:
```
df['Income Group']= pd.
 ↪cut(df['Income'],bins=[29000,50000,75000,105000],labels=['low','mid','high'])
df['Income Group'].value_counts()
```

[23]:
```
low     83
mid     76
high    21
Name: Income Group, dtype: int64
```

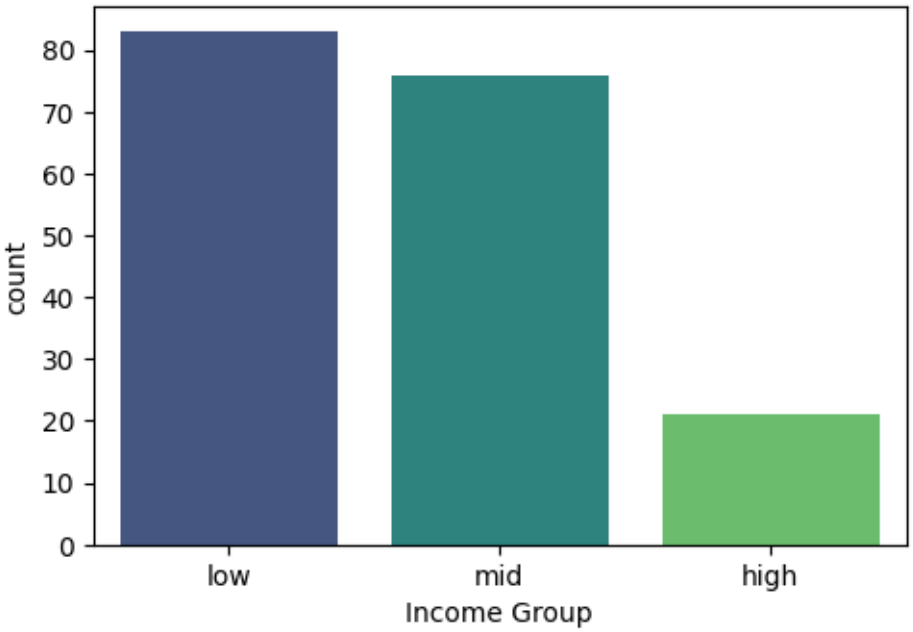
[24]:
```
plt.figure(figsize=(12, 8))
plt.subplot(2, 2, 1)
sns.countplot(data=df, x=df['Income Group'], palette='viridis')
```

```
<ipython-input-24-42acd96ddb1c>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.countplot(data=df, x=df['Income Group'], palette='viridis')
```

[24]:
```
<Axes: xlabel='Income Group', ylabel='count'>
```



We see low and mid income individuals buying the major chunk of Treadmills therefore more focused marketing towards these individuals alongside a premium offering for the high income individuals

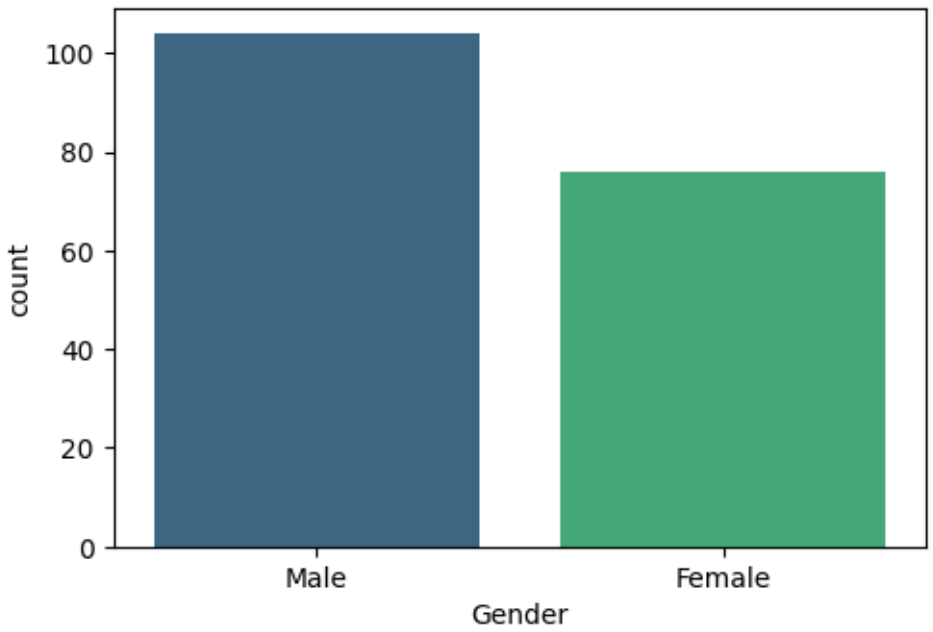
[63]:
```
plt.figure(figsize=(12, 8))
plt.subplot(2, 2, 1)
sns.countplot(data=df, x=df['Gender'], palette='viridis')
```

```
<ipython-input-63-744953f6b716>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.countplot(data=df, x=df['Gender'], palette='viridis')
```

[63]:
```
<Axes: xlabel='Gender', ylabel='count'>
```



Majority are Male customers hence there is scope for attracting more female audience as well with proper marketing and education campaigns Huge market of housewives who don't have much opportunities to exercise regularly

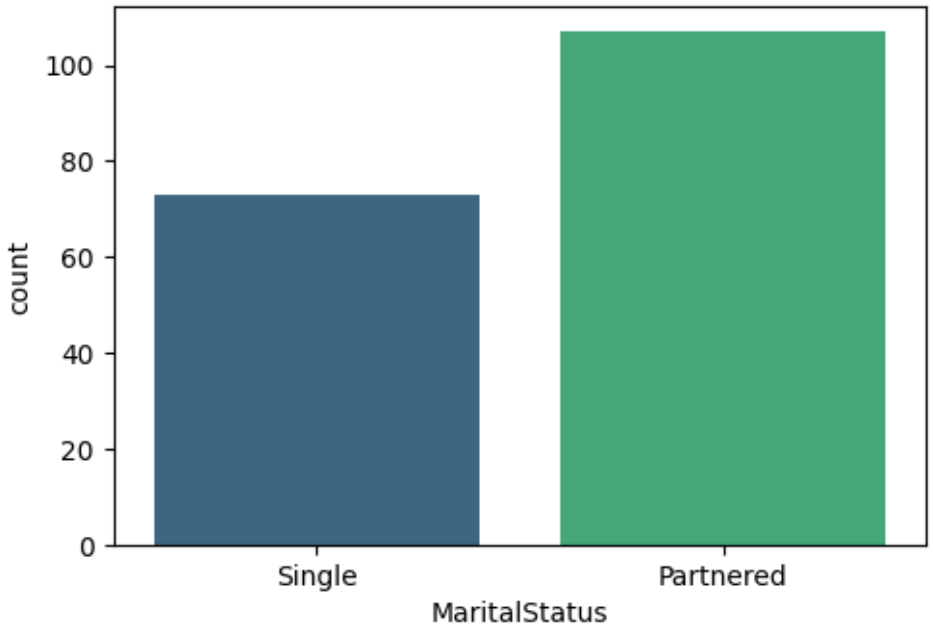
[26]:
```
plt.figure(figsize=(12, 8))
plt.subplot(2, 2, 1)
sns.countplot(data=df, x=df['MaritalStatus'], palette='viridis')
```

```
<ipython-input-26-4e6c40cada1d>:3: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in
v0.14.0. Assign the `x` variable to `hue` and set `legend=False` for the same
effect.

  sns.countplot(data=df, x=df['MaritalStatus'], palette='viridis')
```

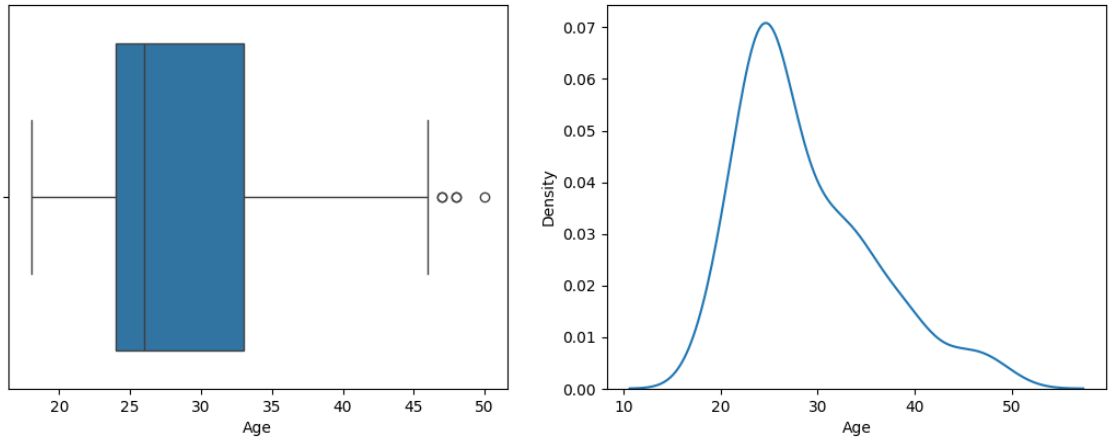
[26]: `<Axes: xlabel='MaritalStatus', ylabel='count'>`



We can clearly notice that Partnered individuals tend to buy more treadmills Hence better targetting for singles can be done with quirky and innovative marketing campaigns

[27]:
```
plt.figure(figsize=(20,10))
plt.subplot(2,3,1)
sns.boxplot(data=df, x='Age')
plt.subplot(2,3,2)
sns.kdeplot(data=df, x='Age')
```

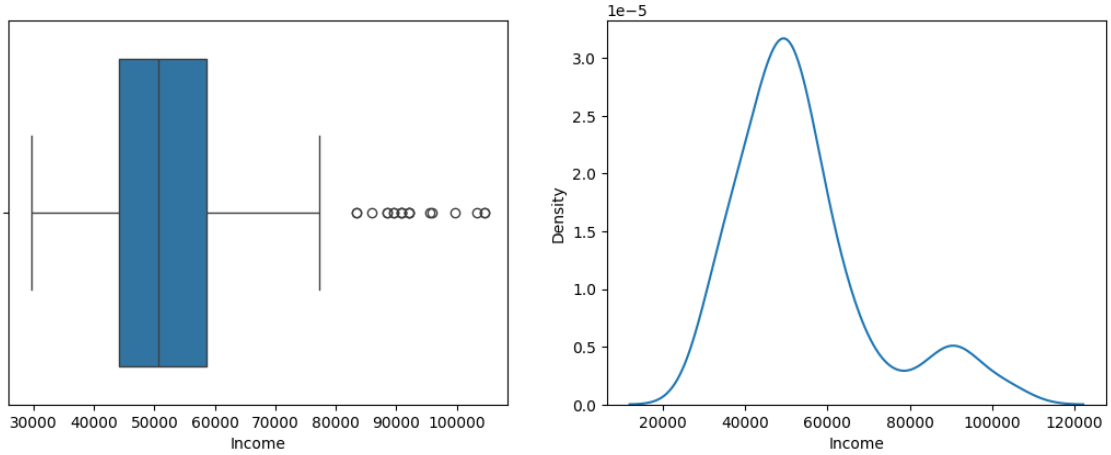
[27]: <Axes: xlabel='Age', ylabel='Density'>



We can clearly see there are 3 outliers in the Age category and the most customers belong to the age group around 25 years old

[28]:
```
plt.figure(figsize=(20,10))
plt.subplot(2,3,1)
sns.boxplot(data=df, x='Income')
plt.subplot(2,3,2)
sns.kdeplot(data=df, x='Income')
```

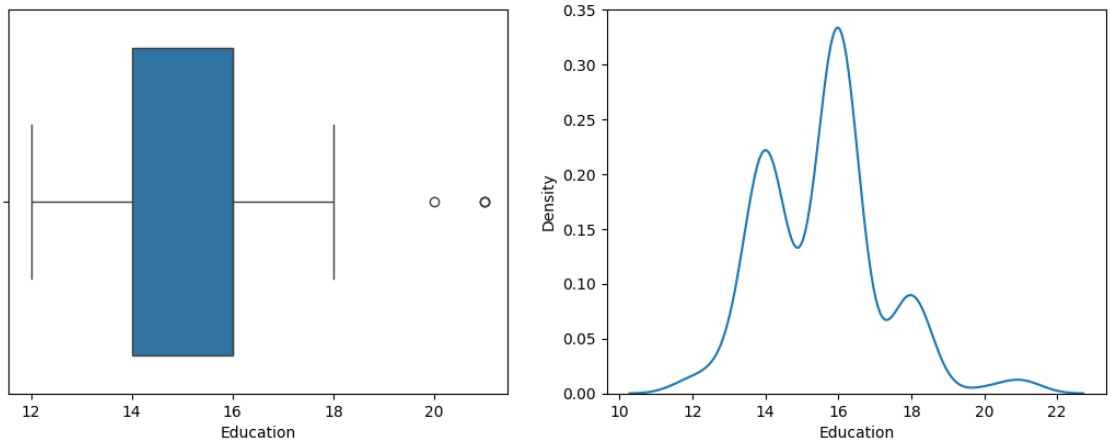
[28]: <Axes: xlabel='Income', ylabel='Density'>



We can clearly see that there are 11 outliers in the Income category and the most number of cutomers earn around 50000

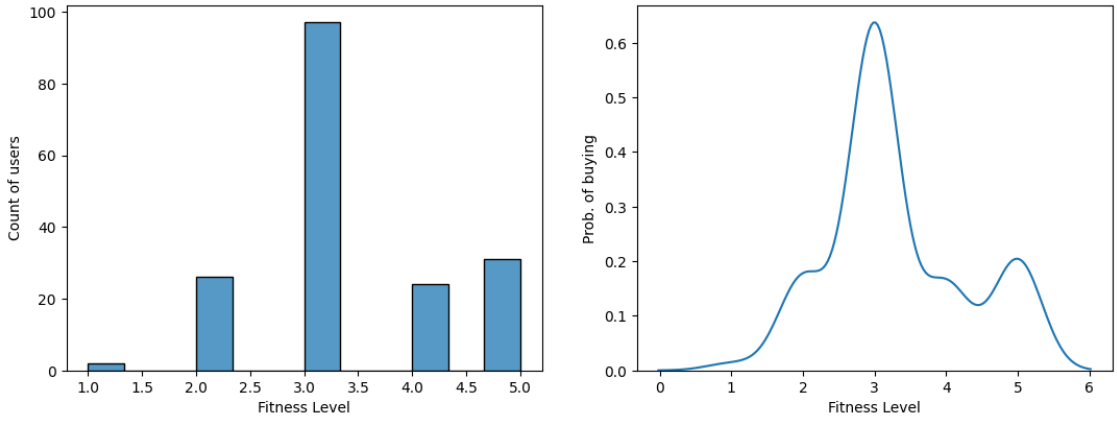
[29]:
```
plt.figure(figsize=(20,10))
plt.subplot(2,3,1)
sns.boxplot(data=df, x='Education')
plt.subplot(2,3,2)
sns.kdeplot(data=df, x='Education')
```

[29]: <Axes: xlabel='Education', ylabel='Density'>



[30]:
```
plt.figure(figsize=(20,10))
plt.subplot(2,3,1)
sns.histplot(data=df,x='Fitness')
plt.xlabel('Fitness Level')
plt.ylabel('Count of users')
plt.subplot(2,3,2)
sns.kdeplot(data=df,x='Fitness')
plt.xlabel('Fitness Level')
plt.ylabel('Prob. of buying')
```

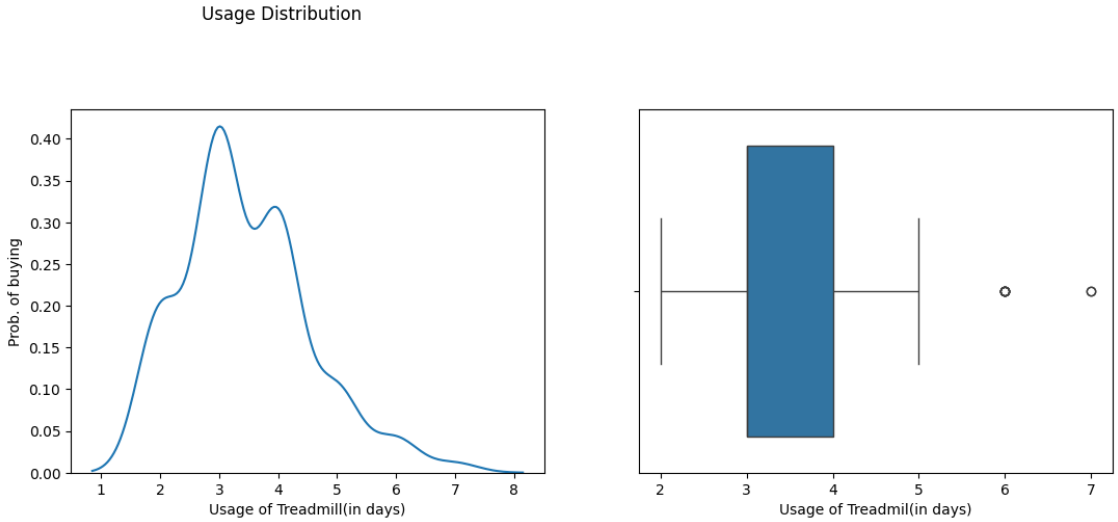
[30]:
```
Text(0, 0.5, 'Prob. of buying')
```



WE can clearly observe people with a fitness level of around 3 or above have a higher tendancy to buy and use the treadmills

[31]:
```
plt.figure(figsize=(20,10))
plt.subplot(2,3,2)
sns.kdeplot(data=df,x='Usage')
plt.xlabel('Usage of Treadmill(in days)')
plt.ylabel('Prob. of buying')
plt.subplot(2,3,3)
sns.boxplot(data=df,x= 'Usage')
plt.xlabel('Usage of Treadmil(in days)')
plt.suptitle('Usage Distribution')
```

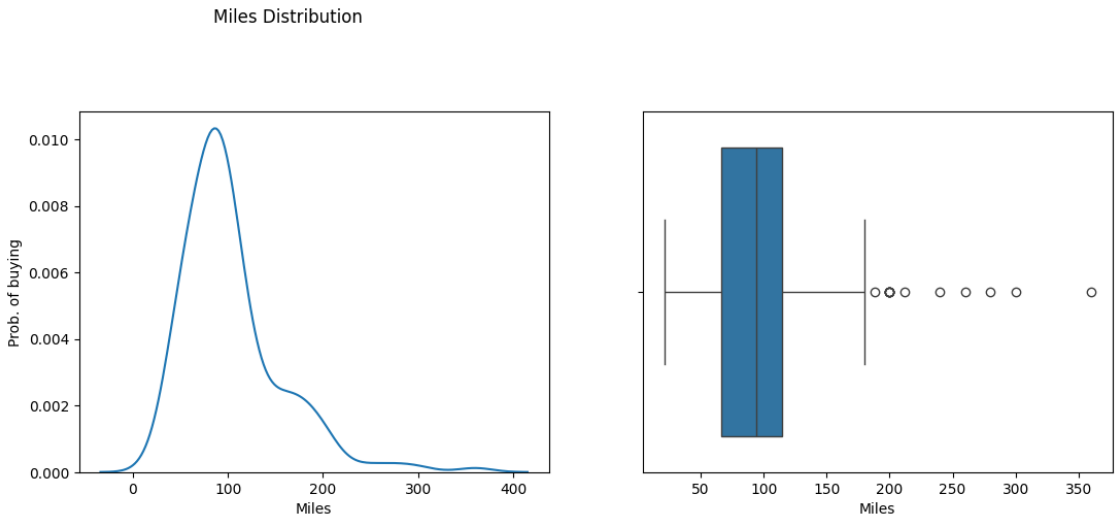
[31]:
```
Text(0.5, 0.98, 'Usage Distribution')
```



The probability of buying a treadmill increases when a person uses it for almost 3-4 days a week

[32]:
```
plt.figure(figsize=(20,10))
plt.subplot(2,3,2)
sns.kdeplot(data=df,x='Miles')
plt.xlabel('Miles')
plt.ylabel('Prob. of buying')
plt.subplot(2,3,3)
sns.boxplot(data=df,x='Miles')
plt.xlabel('Miles')
plt.suptitle('Miles Distribution')
```

[32]:
```
Text(0.5, 0.98, 'Miles Distribution')
```



a person walking around 100 miles is more likely to invest ina. treadmill

[33]:
```
for col in df.select_dtypes(np.number):
  mean = df[col].mean().round(2)
  standard_deviation = df[col].std().round(2)
  median = df[col].median().round(2)
  minimum = df[col].min()
  maximum = df[col].max()
  q1 = np.percentile(df[col],25)
  q3 = np.percentile(df[col],75)
  IQR = q3-q1
  upper_bound = q3+1.5*IQR
  lower_bound = q1-1.5*IQR
  print(f'--- Descriptive Statistics of', col, 'column ---')
  print(f'Mean :', mean)
  print(f'Standard Deviation :', standard_deviation)
  print(f'Median :',median)
  print(f'Minimum :',minimum)
  print(f'Maximum :', maximum)
  print(f'25th Percentile :',q1)
  print(f'75th Percentile :',q3)
  print(f'Inter Quartile Range :', IQR)
  print(f'Upper bound:',upper_bound)
  print(f'Lower bound:',lower_bound)
  print()
```

```
--- Descriptive Statistics of Age column ---
Mean : 28.79
Standard Deviation : 6.94
Median : 26.0
```

```
Minimum : 18
Maximum : 50
25th Percentile : 24.0
75th Percentile : 33.0
Inter Quartile Range : 9.0
Upper bound: 46.5
Lower bound: 10.5

--- Descriptive Statistics of Education column ---
Mean : 15.57
Standard Deviation : 1.62
Median : 16.0
Minimum : 12
Maximum : 21
25th Percentile : 14.0
75th Percentile : 16.0
Inter Quartile Range : 2.0
Upper bound: 19.0
Lower bound: 11.0

--- Descriptive Statistics of Usage column ---
Mean : 3.46
Standard Deviation : 1.08
Median : 3.0
Minimum : 2
Maximum : 7
25th Percentile : 3.0
75th Percentile : 4.0
Inter Quartile Range : 1.0
Upper bound: 5.5
Lower bound: 1.5

--- Descriptive Statistics of Fitness column ---
Mean : 3.31
Standard Deviation : 0.96
Median : 3.0
Minimum : 1
Maximum : 5
25th Percentile : 3.0
75th Percentile : 4.0
Inter Quartile Range : 1.0
Upper bound: 5.5
Lower bound: 1.5

--- Descriptive Statistics of Income column ---
Mean : 53719.58
Standard Deviation : 16506.68
Median : 50596.5
```

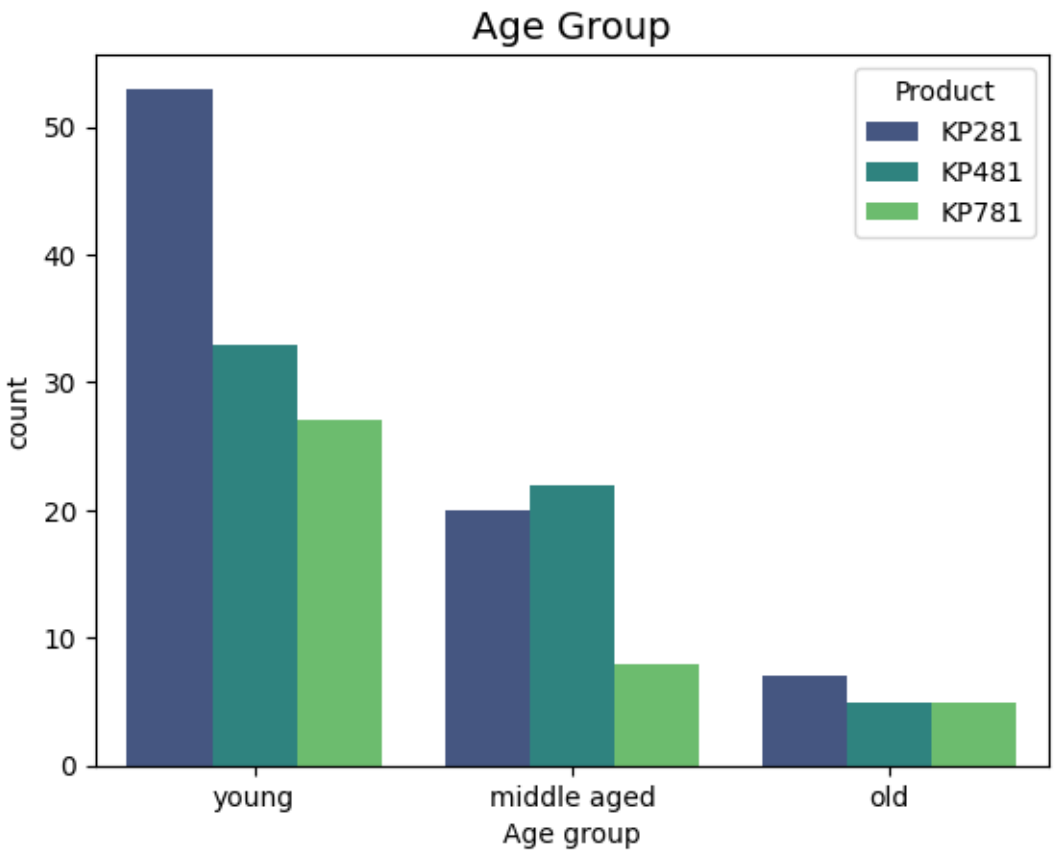
```
Minimum : 29562
Maximum : 104581
25th Percentile : 44058.75
75th Percentile : 58668.0
Inter Quartile Range : 14609.25
Upper bound: 80581.875
Lower bound: 22144.875

--- Descriptive Statistics of Miles column ---
Mean : 103.19
Standard Deviation : 51.86
Median : 94.0
Minimum : 21
Maximum : 360
25th Percentile : 66.0
75th Percentile : 114.75
Inter Quartile Range : 48.75
Upper bound: 187.875
Lower bound: -7.125
```

[34]:
```
sns.countplot(data=df,x='Gender',hue='Product' ,palette='viridis')
plt.suptitle('Gender Distribution', fontsize=18)
plt.xlabel('Gender')
plt.show()
```

Both genders equally prefer KP281 but males tend to buy more KP781s as compared to females

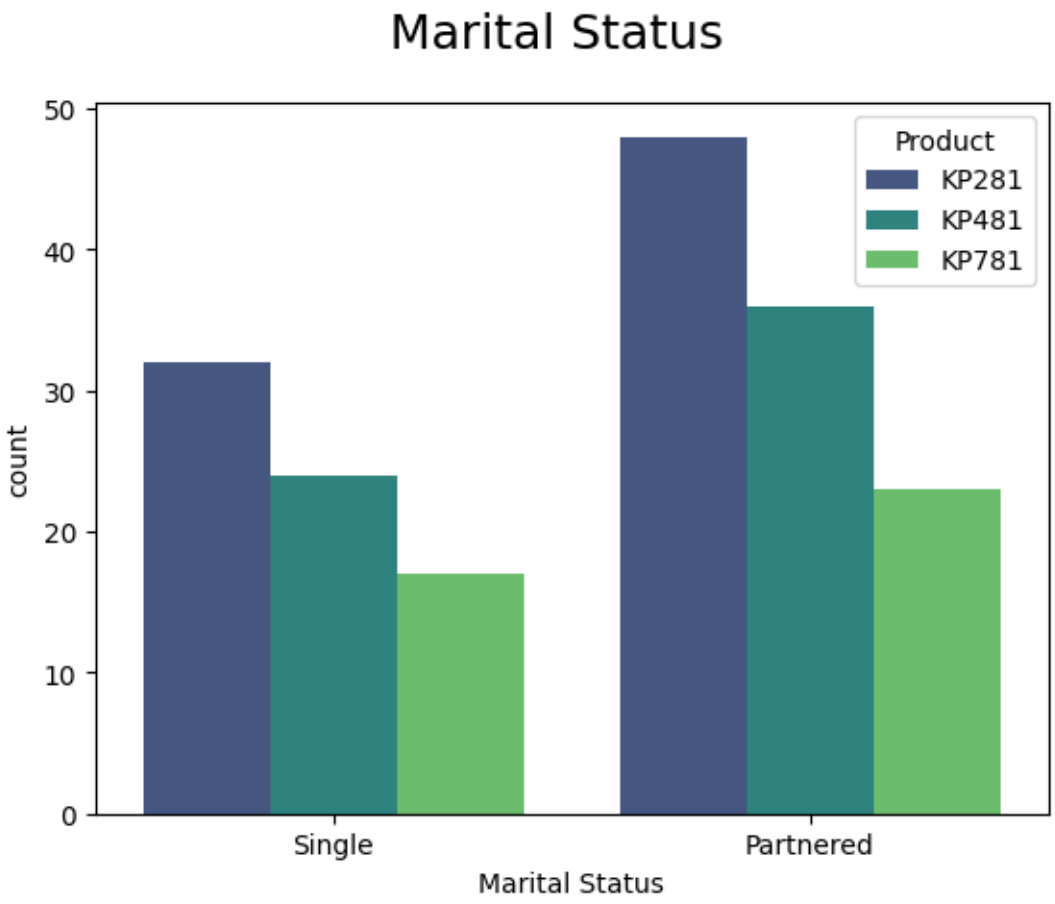
[64]:
```
sns.countplot(data=df,x='Age Group', hue='Product', palette='viridis')
plt.title('Age Group', fontsize=14)
plt.xlabel('Age group')
plt.show()
print()
print ('18-29: Young')
print('30-39: Middle-aged')
print ('40-50: 01d')
```




```
18-29: Young
30-39: Middle-aged
40-50: 01d
```

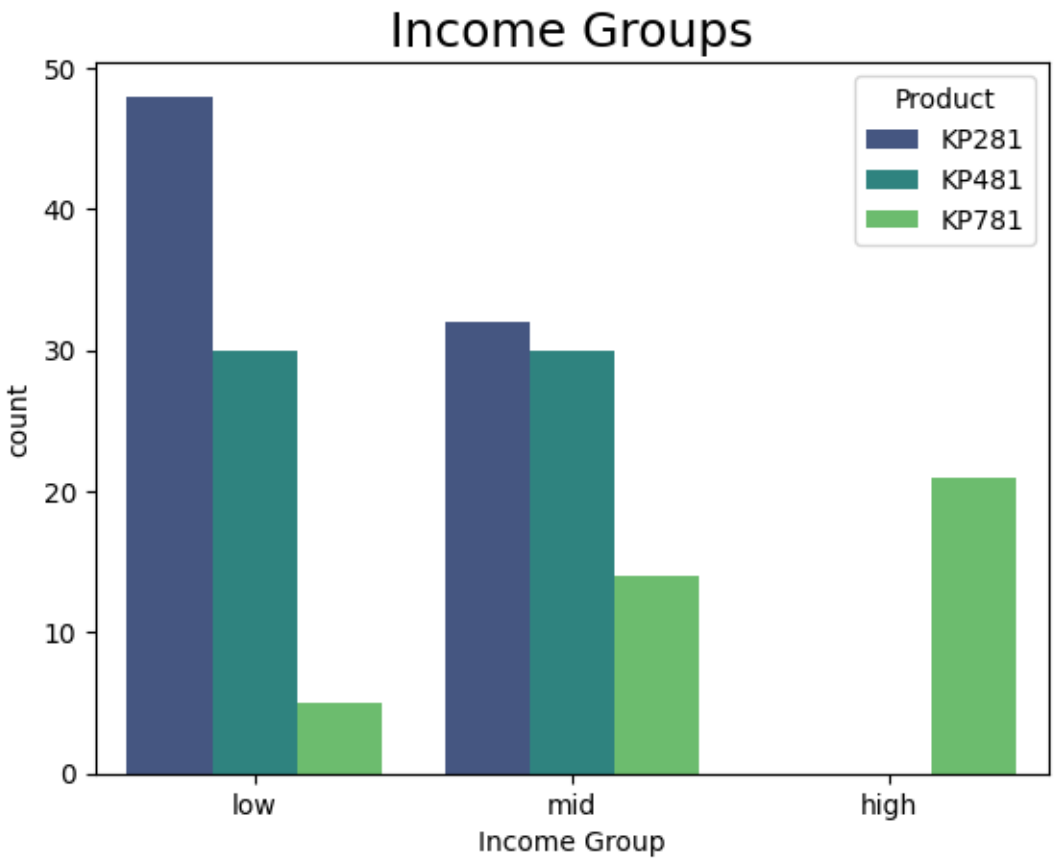
The young fitness and health senstive generation is more inclined towards buying treadmills

[65]:
```
sns.countplot(data=df, x='MaritalStatus', hue='Product', palette='viridis')
plt.suptitle('Marital Status', fontsize=18)
plt.xlabel('Marital Status')
plt.show()
```



partnered people invest more in treadmills than singles

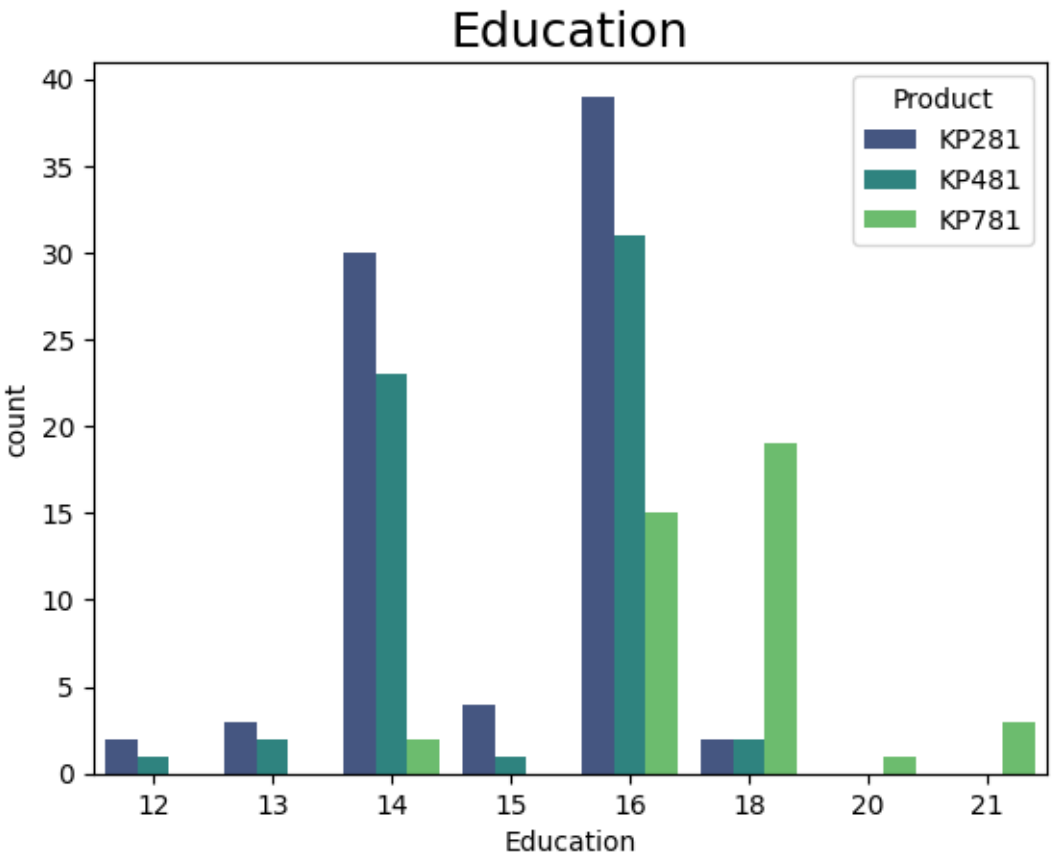
[57]:
```
sns.countplot(data=df, x='Income Group', hue='Product', palette='viridis')
plt.title('Income Groups', fontsize=18)
plt.xlabel('Income Group')
plt.show()
```



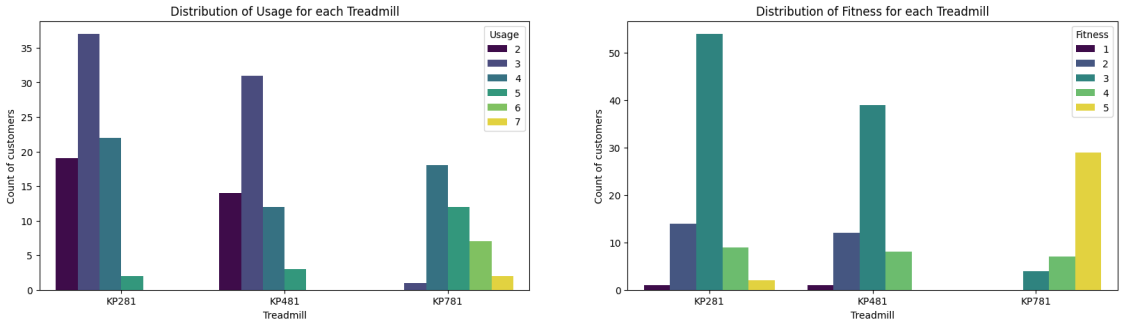
It is clearly visible that low income individuals prefer KP281 in huge numbers while the high incoome individuals only buy KP781

[56]:
```
sns.countplot(data=df, x='Education', hue='Product', palette='viridis')
plt.title('Education', fontsize=18)
```

[56]:
```
Text(0.5, 1.0, 'Education')
```



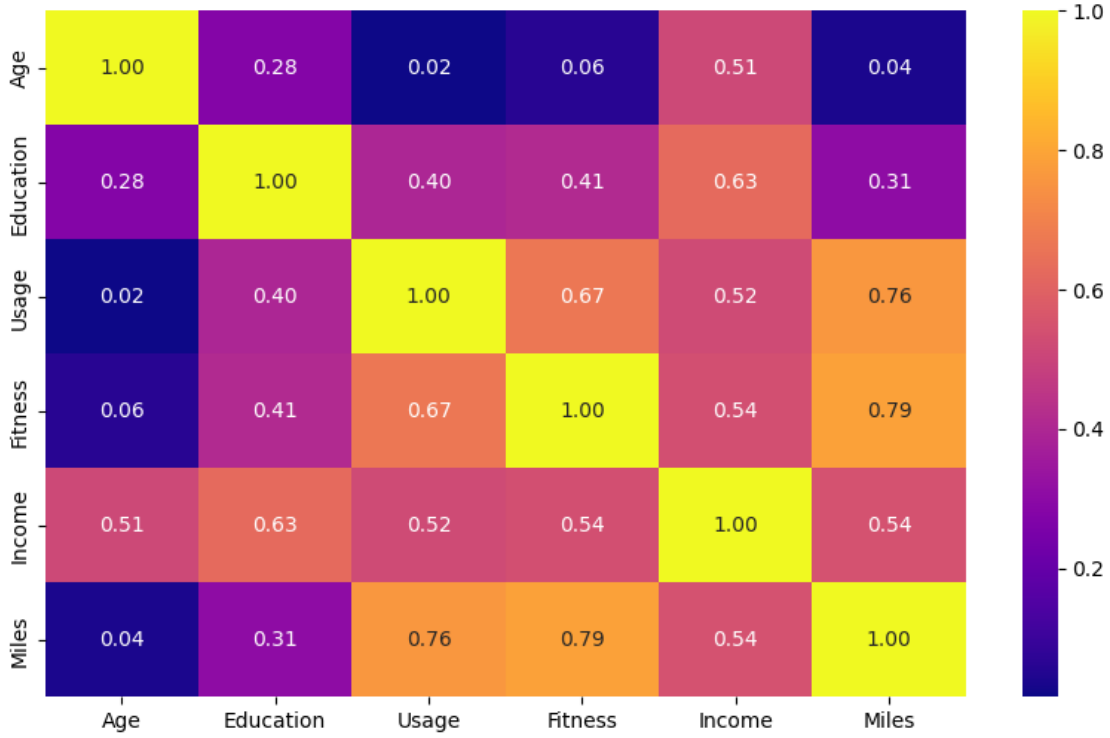
[41]:
```
plt.figure(figsize=(20,5))
plt.subplot(1,2,1)
sns.countplot(data=df, x='Product', hue='Usage' , palette='viridis')
plt.xlabel('Treadmill')
plt.ylabel('Count of customers')
plt.title('Distribution of Usage for each Treadmill')
plt.subplot(1,2,2)
sns.countplot(data=df, x='Product', hue='Fitness', palette='viridis')
plt.xlabel('Treadmill')
plt.ylabel('Count of customers')
plt.title('Distribution of Fitness for each Treadmill')
plt.show()
```



• KP281 and KP481 are popular among users exercising 3 times weekly, while KP781 is the choice for those using treadmills 4-5 times a week.

• KP281 and KP481 are preferred by customers with a fitness level of 3, whereas those at level 5 mainly opt for the advanced KP781.

[38]:
```
numeric_df = df.select_dtypes(include=[np.number])
plt.figure(figsize=(10,6))
sns.heatmap(numeric_df.corr(), annot=True, cmap='plasma', fmt='.2f')
plt.show()
```

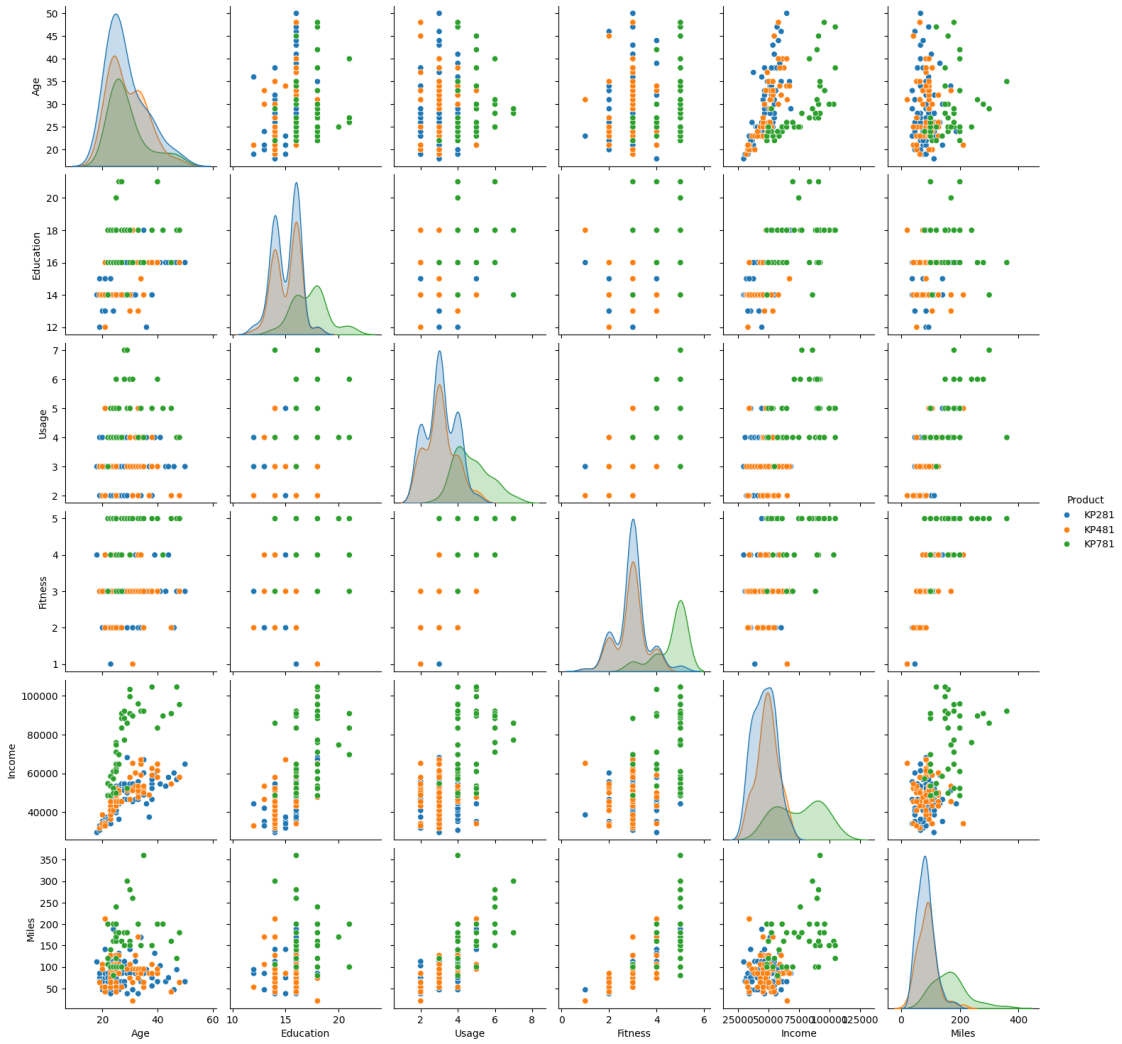


This shows that age is related to education with a unit of 0.28

Usage and fitness relation of 0.67 shows that fitter individuals use the treadmill more

[5]:
```
sns.pairplot(df,hue ='Product')
```

[5]:
```
<seaborn.axisgrid.PairGrid at 0x7bc14ecd92d0>
```



##Probablities and how each aspect affects whether someone will buy a treadmill or not

[45]:
```
pd.crosstab(index =df['Product'],columns = df['Age Group'],margins␣
  ↪=True,normalize = True).round(2)
```

[45]:

Age Group	young	middle aged	old	All
Product				
KP281	0.29	0.11	0.04	0.44
KP481	0.18	0.12	0.03	0.33
KP781	0.15	0.04	0.03	0.22

```
All          0.63          0.28   0.09   1.00
```

[46]:
```
pd.crosstab(index =df['Product'],columns = df['Education'],margins
  ↪=True,normalize = True).round(2)
```

[46]:
```
Education    12    13    14    15    16    18    20    21   All
Product
KP281      0.01  0.02  0.17  0.02  0.22  0.01  0.00  0.00  0.44
KP481      0.01  0.01  0.13  0.01  0.17  0.01  0.00  0.00  0.33
KP781      0.00  0.00  0.01  0.00  0.08  0.11  0.01  0.02  0.22
All        0.02  0.03  0.31  0.03  0.47  0.13  0.01  0.02  1.00
```

[47]:
```
pd.crosstab(index =df['Product'],columns = df['Gender'],margins =True,normalize
  ↪= True).round(2)
```

[47]:
```
Gender   Female  Male   All
Product
KP281      0.22  0.22  0.44
KP481      0.16  0.17  0.33
KP781      0.04  0.18  0.22
All        0.42  0.58  1.00
```

[48]:
```
pd.crosstab(index =df['Product'],columns = df['MaritalStatus'],margins
  ↪=True,normalize = True).round(2)
```

[48]:
```
MaritalStatus  Partnered  Single   All
Product
KP281                0.27    0.18  0.44
KP481                0.20    0.13  0.33
KP781                0.13    0.09  0.22
All                  0.59    0.41  1.00
```

[49]:
```
pd.crosstab(index =df['Product'],columns = df['Usage'],margins =True,normalize
  ↪= True).round(2)
```

[49]:
```
Usage       2     3     4     5     6     7   All
Product
KP281    0.11  0.21  0.12  0.01  0.00  0.00  0.44
KP481    0.08  0.17  0.07  0.02  0.00  0.00  0.33
KP781    0.00  0.01  0.10  0.07  0.04  0.01  0.22
All      0.18  0.38  0.29  0.09  0.04  0.01  1.00
```

[51]:
```
pd.crosstab(index =df['Product'],columns = df['Income Group'],margins
  ↪=True,normalize = True).round(2)
```

[51]:
```
Income Group   low   mid  high   All
Product
```

```
KP281          0.27  0.18  0.00  0.44
KP481          0.17  0.17  0.00  0.33
KP781          0.03  0.08  0.12  0.22
All            0.46  0.42  0.12  1.00
```

[52]:
```
pd.crosstab(index =df['Product'],columns = df['Fitness'],margins␣
↪=True,normalize = True).round(2)
```

[52]:
```
Fitness     1     2     3     4     5   All
Product
KP281    0.01  0.08  0.30  0.05  0.01  0.44
KP481    0.01  0.07  0.22  0.04  0.00  0.33
KP781    0.00  0.00  0.02  0.04  0.16  0.22
All      0.01  0.14  0.54  0.13  0.17  1.00
```

All these probabilities give us the following observations

- The Probability of a treadmill being purchased by a Young Adult(18-25) is 44%.
- The Probability of a treadmill being purchased by a customer with Higher Education(Above 15 Years) is 62%.
- P(KP281|Low) = 0.27 which is the highest of all conditional probabilities
- The Probability of a treadmill being purchased by a Married Customer is 59%.
- The Probability of a treadmill being purchased by a customer with Usage 3 per week is 38%.
- The Probability of a treadmill being purchased by a customer with Average(3) Fitness is 54%.
- The Probability of a treadmill being purchased by a male is 58%.

#My Business Insights and Observations

1. Most potential customers for KP281 are:

- Young (18-29)
- Males and Females both
- Couples (Married)
- Education between 14 to 16 years
- Fitness level>=3
- Usage of 3 times a week
- running 60-100 miles
- Income between 29000 to 50000

2. Most potential customers for KP481 are:

- Young (18-29)
- Males and Females both
- Married or unmarried
- Education between 14 to 16 years
- Fitness level>=3
- Usage of 3 times a week
- running 80-120 miles
- Income between 29000 to 50000

3. Most potential customers for KP781 are:

- Young (18-25)
- Males
- Couples (Married)
- Education between 16 to 18 years
- Fitness level = 5
- Usage of 4-6 times a week
- running 120-200 miles
- Income between 75000 to 105000

###Concluding Observations

1. KP281 is the top choice for 44.44% of users, followed by KP481 with 33.33%, and KP781 appeals to 22.22%.
2. Young, educated, low income, married customers are more likely to get a treadmill
3. KP281 is the cheapest and therefore is the most sold one
4. 33% people buy KP481 for a balance in features and price
5. Fitness freaks and high income individuals go for the best in line KP781
6. Women don't buy KP781 at all probably due to high costs

#My recommendations and Suggestions

Integrated Business Analysis and Growth Strategy for Aerofit:

1. Market Analysis: - **Targeting Gen-Z and Millennials:** Focus marketing efforts on understanding preferences and lifestyle choices of Gen-Z and Millennials, utilizing platforms like Instagram, TikTok, and YouTube for product showcases and testimonials. - **Women-Centric Approach:** Develop marketing strategies highlighting treadmill features tailored to women's needs and preferences, emphasizing comfort, design, and customizable workout programs. - **Competitive Analysis:** Analyze competitors' strategies to identify differentiation opportunities and market gaps.

2. Product Analysis: - **Decoy Effect in Pricing:** Use pricing strategies to position KP481 as the best value option between KP281 and KP781, appealing to value-conscious customers. - **Product Performance Evaluation:** Continuously assess performance and gather customer feedback to drive innovation and improvement, particularly focusing on features that cater to women's needs and older adults' comfort. - **Product Diversification:** Explore opportunities to introduce complementary fitness equipment or accessories targeting specific customer segments, such as low-impact options for older adults.

3. Customer Analysis: - **Segmentation and Personalization:** Segment customers based on demographics and psychographics for targeted marketing and personalized recommendations, with specific emphasis on women and older adults. - **Enhanced Customer Experience:** Implement efficient customer service and post-purchase support to improve satisfaction and retention, with tailored support for women and older adults.

4. Pricing Strategy: - **Value-Based Pricing:** Align pricing with the perceived value of Aerofit products, offering strategic discounts and promotions to stimulate demand, including targeted offers for women and older adults. - **Promotional Strategies:** Offer bundle deals and loyalty programs to incentivize purchases, especially targeting Gen-Z, Millennials, women, and older adults.

5. Marketing and Promotion: - **Targeted Marketing Campaigns:** Develop campaigns resonating with Gen-Z, Millennials, women, and older adults, emphasizing Aerofit's benefits in

achieving fitness goals and addressing specific needs. - **Digital Marketing Channels:** Utilize social media and influencer partnerships for effective outreach and engagement, with content tailored to the interests and preferences of each demographic. - **Brand Building:** Emphasize quality, innovation, and sustainability in branding to appeal to environmentally conscious consumers, including messaging focused on Aerofit's commitment to providing inclusive and age-friendly fitness solutions.

6. Distribution and Sales Channels: - **Expanding Reach:** Forge partnerships with retailers and e-commerce platforms to expand distribution networks, ensuring accessibility to Aerofit products for women and older adults in various regions. - **Streamlined Sales Processes:** Optimize online and offline sales processes for a seamless buying experience, with special attention to ease of use for older adults and first-time buyers.

7. Innovation and Technology: - **Research and Development:** Invest in R&D for product innovation addressing evolving customer needs, including features that enhance comfort, safety, and ease of use for women and older adults. - **Smart Features Integration:** Incorporate IoT connectivity and data analytics to enhance functionality and user experience, with intuitive interfaces suitable for users of all ages and fitness levels.

8. Customer Retention and Loyalty: - **Loyalty Programs:** Implement programs to incentivize repeat purchases and foster brand loyalty, with rewards and benefits tailored to the preferences of women and older adults. - **Feedback Mechanisms:** Regularly gather feedback to demonstrate responsiveness to customer needs, with dedicated channels for women and older adults to provide input on product preferences and experiences.

9. Expansion and International Growth: - **Market Exploration:** Identify opportunities for expansion into new geographical markets, adapting strategies to suit local preferences, including offerings that cater to the specific fitness needs of women and older adults. - **Localization Strategies:** Tailor marketing and product offerings to meet the needs of target markets, with cultural sensitivity and inclusivity as key considerations.

10. Sustainability and CSR:

- **Environmental Initiatives:** Integrate sustainable practices to appeal to environmentally conscious consumers, with transparency in sourcing and manufacturing processes, appealing to women and older adults concerned about the environment.
- **Community Engagement:** Participate in CSR initiatives to enhance brand reputation, especially among educated and affluent youth demographics, with initiatives focused on promoting health and fitness for women and older adults in local communities.