Google Drive Link: [https://colab.research.google.com/drive/1LRvfAyXtWDImuPxNy5gK8hjnOrVS8IUm?usp=sharing (https://colab.research.google.com/drive/1LRvfAyXtWDImuPxNy5gK8hjnOrVS8IUm?usp=sharing)](https://colab.research.google.com/drive/1LRvfAyXtWDImuPxNy5gK8hjnOrVS8IUm?usp=sharing)

In [43]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [44]:
```
df = pd.read_csv('/content/drive/MyDrive/Dataset/aerofit_treadmill.csv')
df
```

Out[44]:

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

180 rows × 9 columns

In [45]:
```
df.head()
```

Out[45]:

	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

In [46]:
```
df.tail()
```

Out[46]:

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles
175	KP781	40	Male	21	Single	6	5	83416	200
176	KP781	42	Male	18	Single	5	4	89641	200
177	KP781	45	Male	16	Single	5	5	90886	160
178	KP781	47	Male	18	Partnered	4	5	104581	120
179	KP781	48	Male	18	Partnered	4	5	95508	180

In [47]:
```
df.shape
```

Out[47]:
```
(180, 9)
```

In [48]:
```
df.describe(include="all")
```

Out[48]:

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	
count	180	180.000000	180	180.000000	180	180.000000	180.000000	1
unique	3	NaN	2	NaN	2	NaN	NaN	
top	KP281	NaN	Male	NaN	Partnered	NaN	NaN	
freq	80	NaN	104	NaN	107	NaN	NaN	
mean	NaN	28.788889	NaN	15.572222	NaN	3.455556	3.311111	537
std	NaN	6.943498	NaN	1.617055	NaN	1.084797	0.958869	165
min	NaN	18.000000	NaN	12.000000	NaN	2.000000	1.000000	295
25%	NaN	24.000000	NaN	14.000000	NaN	3.000000	3.000000	440
50%	NaN	26.000000	NaN	16.000000	NaN	3.000000	3.000000	505
75%	NaN	33.000000	NaN	16.000000	NaN	4.000000	4.000000	586
max	NaN	50.000000	NaN	21.000000	NaN	7.000000	5.000000	1045

In [49]:
```
# Checking number of nunique values in our dataset
for i in df.columns:
  print(i,":",df[i].nunique())
```

```
Product : 3
Age : 32
Gender : 2
Education : 8
MaritalStatus : 2
Usage : 6
Fitness : 5
Income : 62
Miles : 37
```

In [50]:
```
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
```

Observations:

1. The dataset contains no missing values and encompasses records for three distinct treadmill products: KP281, KP481, and KP781. Among these, KP281 appears to be the most commonly purchased.
2. Age distribution spans from 18 to 50 years, with a mean age of 28.79. Notably, 75% of individuals are aged 33 or younger, indicating a relatively youthful customer base.
3. Education levels primarily cluster around 16 years, with 75% of customers reporting 16 years of education or fewer.
4. Gender distribution shows that out of 180 data points, 104 customers identify as male, while the remaining identify as female, suggesting a slight male predominance.
5. Income and Miles variables exhibit notably high standard deviations, hinting at potential outliers within these features.

1. Univariate Analysis

Understanding the distribution of the data for the quantitative attributes:

Age

Education

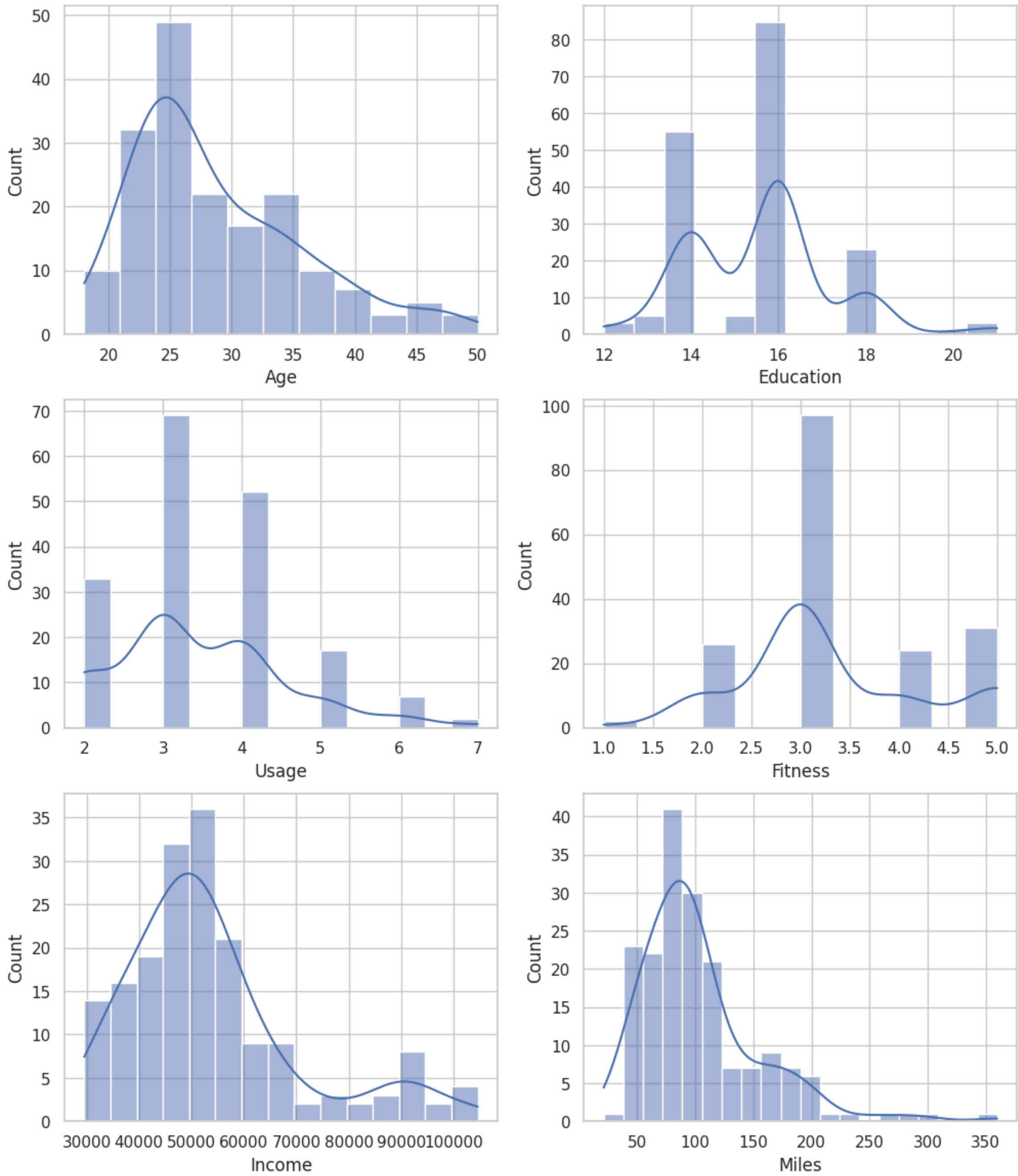
Usage

Fitness

Income

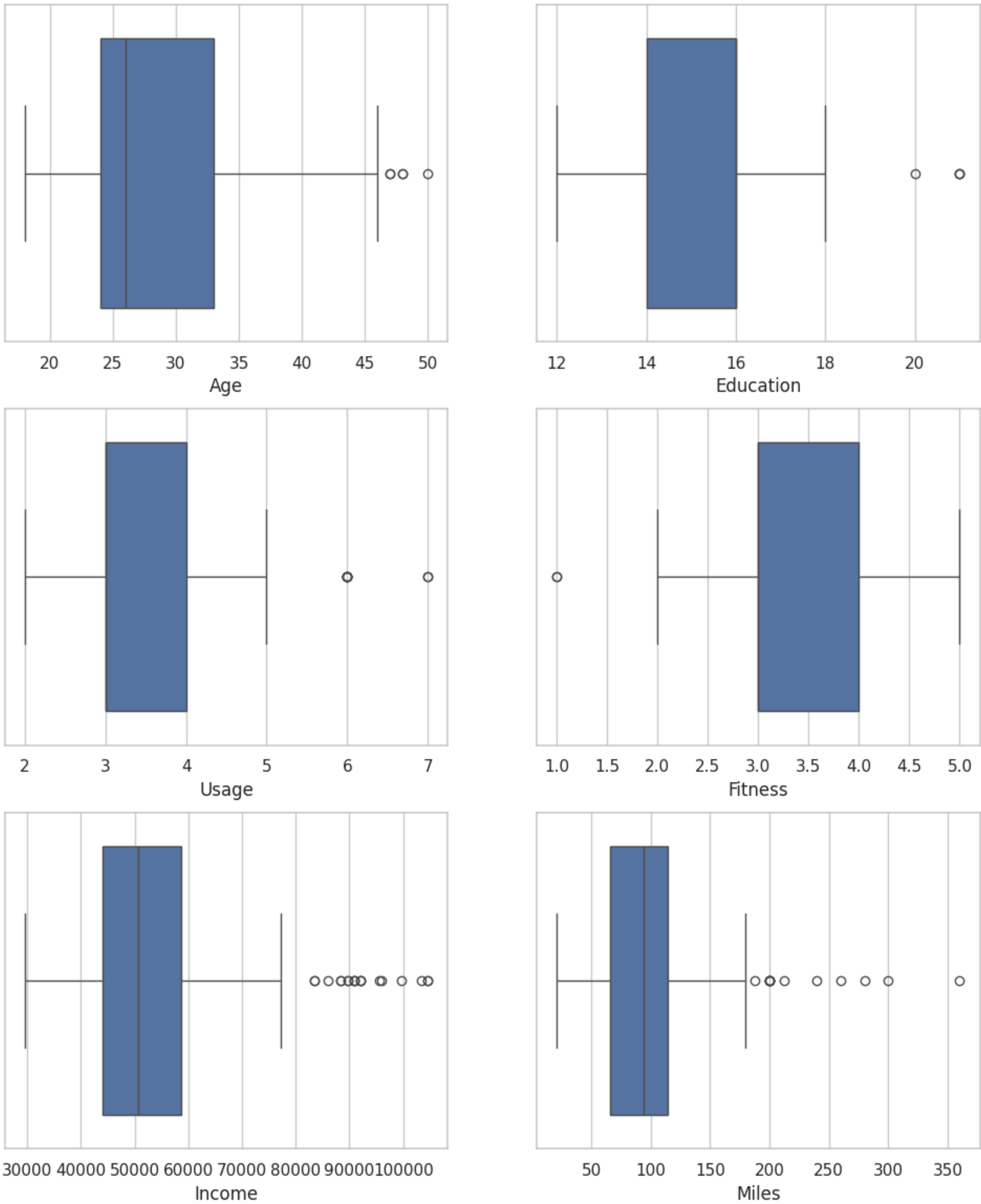
Miles

In [51]:
```
fig, axis = plt.subplots(nrows=3, ncols=2, figsize=(12, 10))
fig.subplots_adjust(top=1.2)

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])
sns.histplot(data=df, x="Income", kde=True, ax=axis[2,0])
sns.histplot(data=df, x="Miles", kde=True, ax=axis[2,1])
plt.show()
```



Outliers detection using BoxPlots

In [52]:
```
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()
```

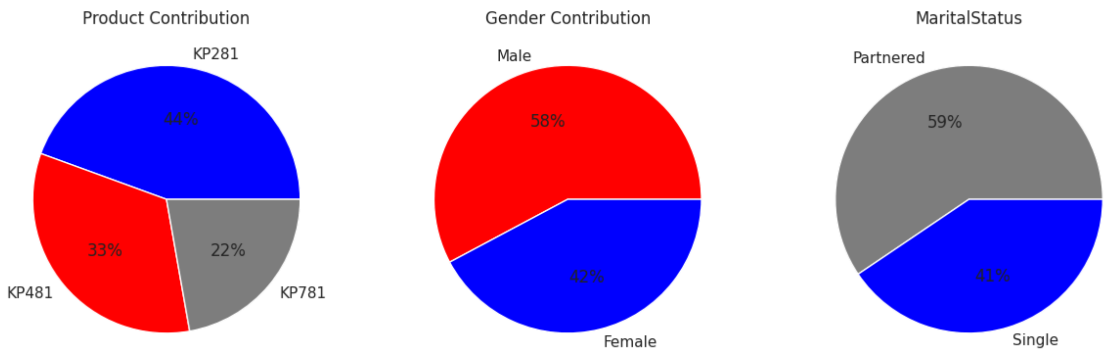


Observation:

From the boxplots, it's evident that:

1. Age, Education, and Usage exhibit minimal outliers, indicating relatively consistent distributions within these variables.
2. Conversely, Income and Miles display a higher prevalence of outliers, suggesting greater variability or dispersion within these features.

In [53]:
```
# Analysis through pie-chart
fig = plt.figure(figsize=(15,5))

f1=plt.subplot(1, 3, 1)
f1.set_title('Product Contribution')
data = df["Product"].value_counts()
labels = ['KP281','KP481','KP781']
colors=['Blue','Red','Grey']
plt.pie(data, labels = labels, colors = colors, autopct='%.0f%%')

f1=plt.subplot(1, 3, 2)
f1.set_title('Gender Contribution')
data = df["Gender"].value_counts()
labels=df['Gender'].value_counts().index
colors=['Red','Blue']
plt.pie(data, labels = labels, colors = colors, autopct='%.0f%%')

f1=plt.subplot(1, 3, 3)
f1.set_title('MaritalStatus')
data = df["MaritalStatus"].value_counts()
labels=df['MaritalStatus'].value_counts().index
colors=['Grey','blue']
plt.pie(data, labels = labels, colors = colors, autopct='%.0f%%')
plt.show()
```



Observations:

1. KP281 emerges as the most commonly purchased product among the available treadmill options.
2. The dataset skews slightly towards male customers, indicating a higher proportion of males compared to females.
3. Partnered individuals outnumber single individuals within the dataset.

In [54]:
```
df1 = df[['Product', 'Gender', 'MaritalStatus']].melt()
df1.groupby(['variable', 'value'])[['value']].count() / len(df)
```

Out[54]:

		value
variable	**value**	
Gender	**Female**	0.422222
	Male	0.577778
MaritalStatus	**Partnered**	0.594444
	Single	0.405556
Product	**KP281**	0.444444
	KP481	0.333333
	KP781	0.222222

Observations:

1. Product:

   44.44% of customers have purchased the KP281 product.

   33.33% of customers have purchased the KP481 product.

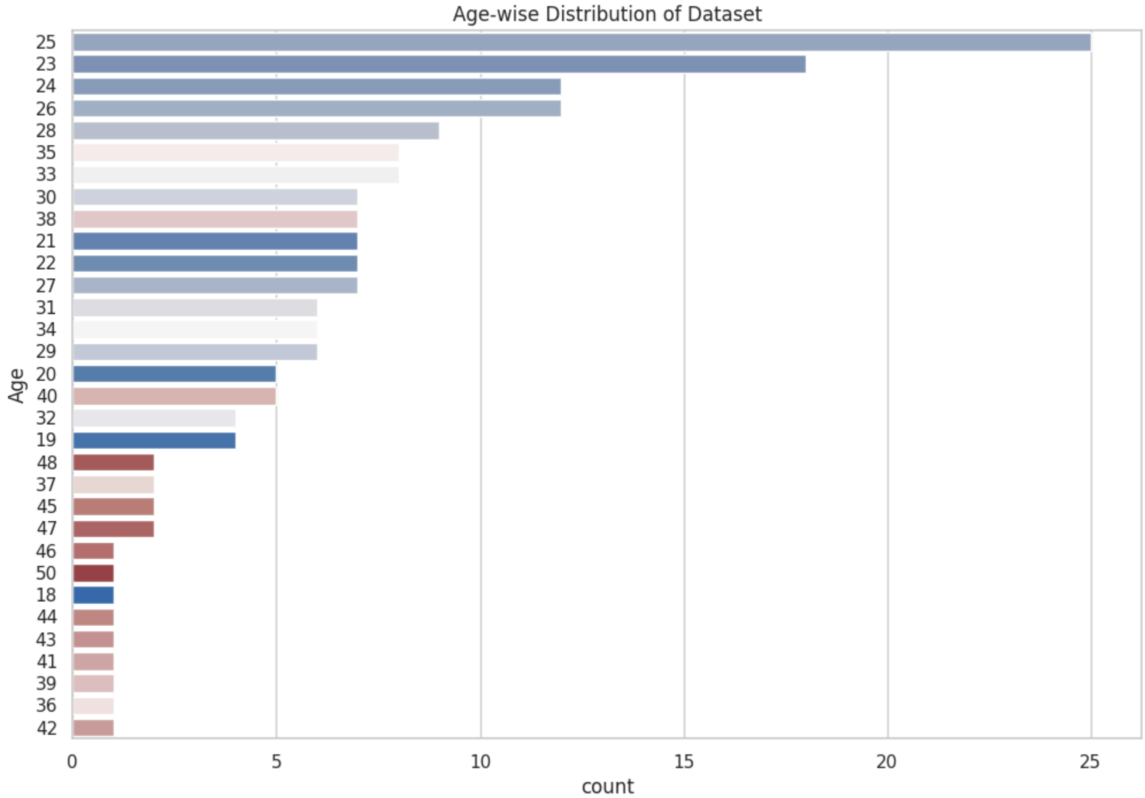
   22.22% of customers have purchased the KP781 product.
2. Gender:

   57.78% of customers are male.

   MaritalStatus:

1. 59.44% of customers are partnered.

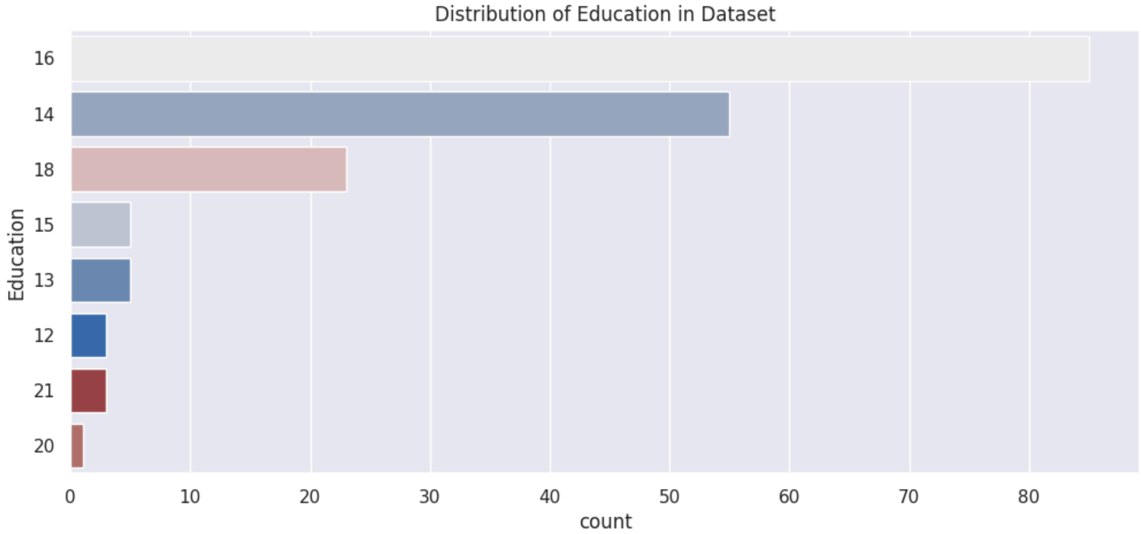
In [55]:
```
# Analysis through countplot
plt.figure(figsize=(12,8))
plt.title('Age-wise Distribution of Dataset', fontsize=12)
sns.set(style="darkgrid")
ax = sns.countplot(y="Age", data=df, palette="vlag", order=df['Age'].value_
counts().index[0:35],hue='Age',legend=False)
plt.show()
```



Observations :

1. Among the age groups, users aged 25 years have the maximum count, indicating a peak in treadmill purchases within this demographic.
2. Additionally, the majority of customers fall within the age range of 18 years and above, highlighting the prevalence of adult customers in the dataset.

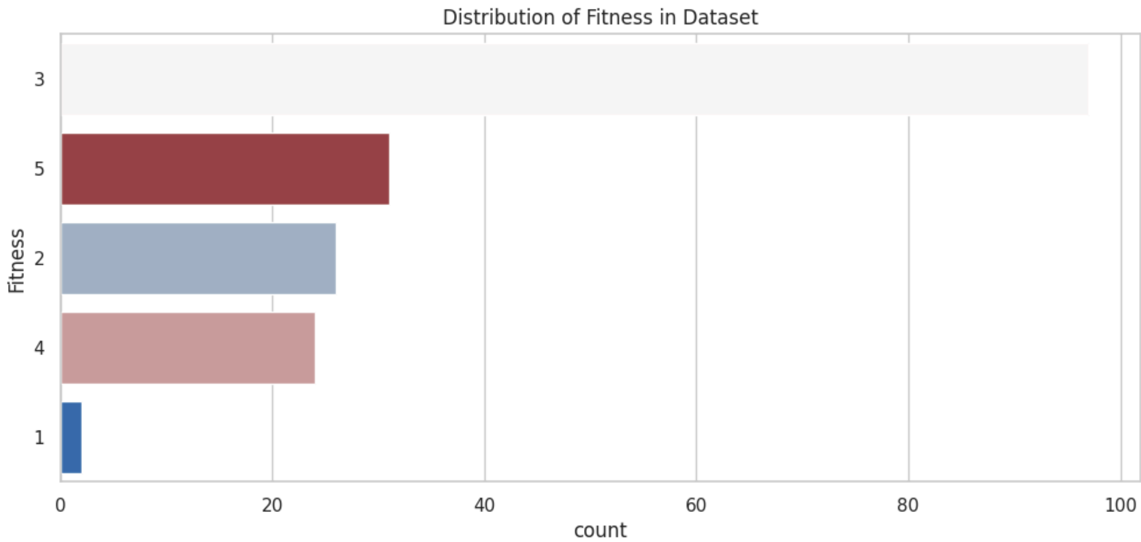
In [56]:
```
# Analysis through countplot
plt.figure(figsize=(12,5))
plt.title('Distribution of Education in Dataset', fontsize=12)
sns.set(style="whitegrid")
ax = sns.countplot(y="Education", data=df, palette="vlag", order=df["Educat
ion"].value_counts().index[0:10],legend=False,hue='Education')
plt.show()
```



Observations :

The majority of customers report their education years falling within the range of 16 to 18, indicating a concentration of educational attainment within this bracket among the customer base.

In [57]:
```
# Analysis through countplot
plt.figure(figsize=(12,5))
plt.title('Distribution of Fitness in Dataset', fontsize=12)
sns.set(style="darkgrid")
ax = sns.countplot(y="Fitness", data=df, palette="vlag", order=df["Fitnes
s"].value_counts().index[0:10],hue='Fitness',legend=False)
plt.show()
```



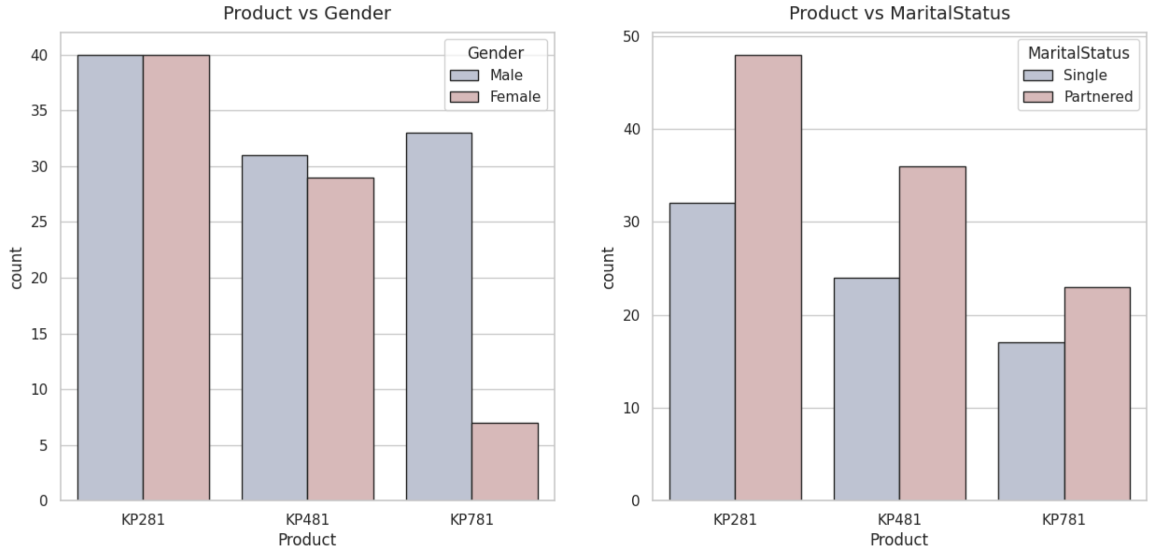
Observations:

Individuals categorized as medium active have predominantly rated themselves with a fitness level of 3 on the scale provided.

2. Bivariate Analysis

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

In [58]:
```
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
='vlag', ax=axs[0])
sns.countplot(data=df, x='Product', hue='MaritalStatus', edgecolor="0.15",
palette='vlag', 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()
```



Observations:

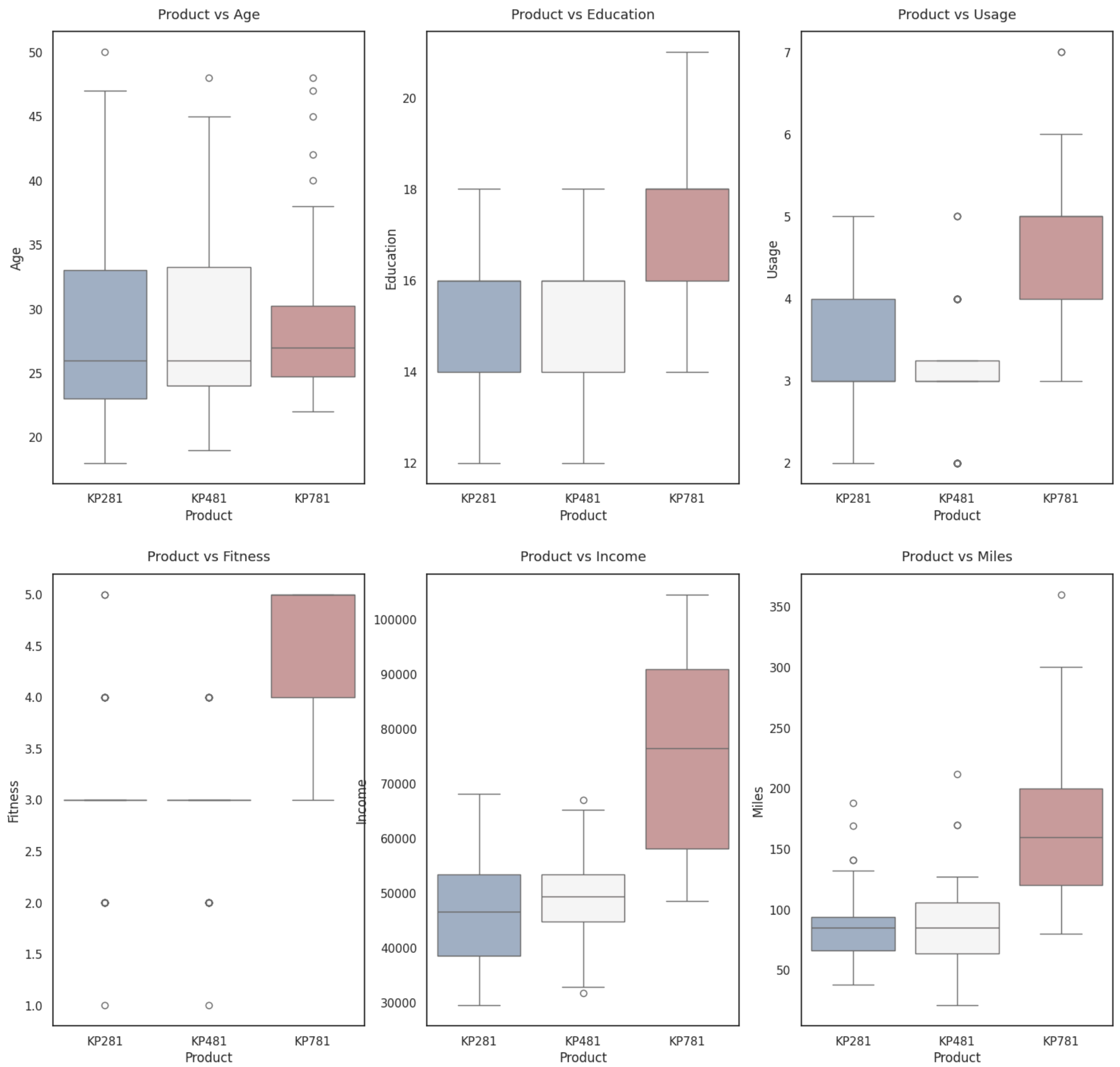
1. Product vs Gender:

   Equal numbers of males and females have purchased the KP281 product indicating a balanced gender distribution for this product. Similarly, the distribution is almost equal for the KP481 product. However, a significant proportion of male customers have purchased the KP781 product compared to females, suggesting a gender preference for this particular treadmill model.
2. Product vs MaritalStatus:

   Partnered customers are more likely to purchase the product, as indicated by the data. This suggests a correlation between marital status and treadmill purchase, with partnered individuals showing a higher propensity to buy.

Checking if following features have any effect on the product purchased: Age Education Usage Fitness Income Miles

In [59]:
```
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], pale
tte='vlag',hue='Product',legend=False)
        axs[i,j].set_title(f"Product vs {attrs[count]}", pad=12, fontsize=1
3)
        count += 1
```



Observations:

1. Product vs Age:

   Customers purchasing products KP281 and KP481 exhibit the same median age value. Customers within the age range of 25-30 are more inclined to purchase the KP781 product.
2. Product vs Education:

   Customers with an education level greater than 16 are more likely to purchase the KP781 product. Conversely, customers with an education level less than or equal to 16 have an equal likelihood of purchasing either the KP281 or KP481 product.
3. Product vs Usage:

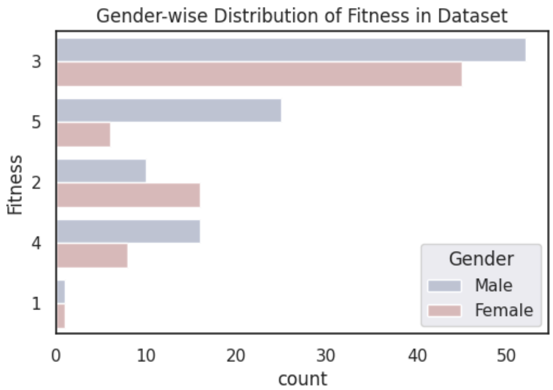
   Customers planning to use the treadmill more than 4 times a week are more inclined to purchase the KP781 product. Other customers are more likely to purchase either the KP281 or KP481 product.
4. Product vs Fitness:

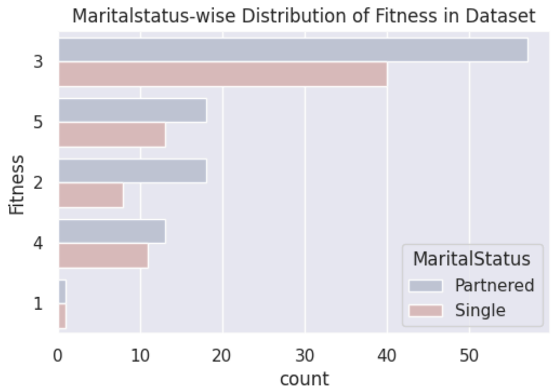
   Customers with a higher fitness level (fitness score >= 3) are more inclined to purchase the KP781 product.
5. Product vs Income:

   Customers with higher incomes (income >= $60,000) are more likely to purchase the KP781 product.
6. Product vs Miles:

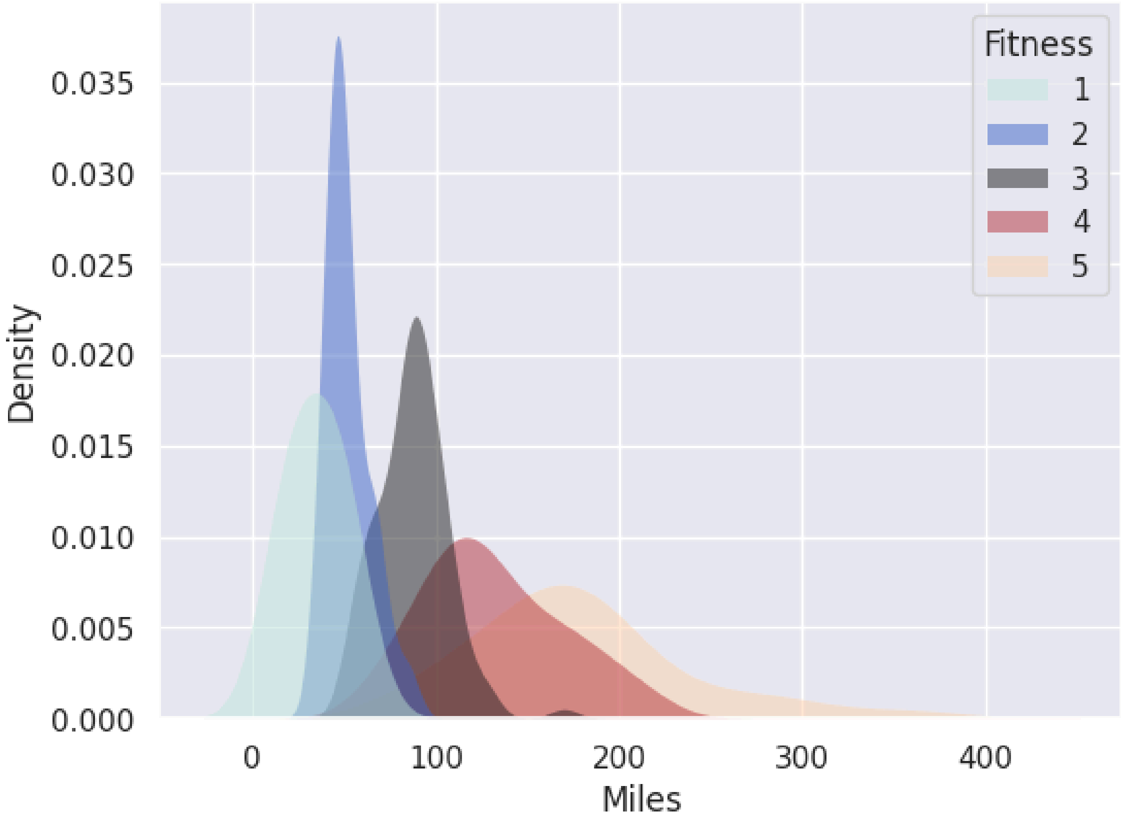
   Customers expecting to walk/run more than 120 miles per week are more likely to purchase the KP781 product.

In [60]:
```
# Bi-Variate Analysis through countplot
plt.figure(figsize=(20,8))
plt.subplot(2, 3, 1)
plt.title('Gender-wise Distribution of Fitness in Dataset', fontsize=12)
sns.set(style="darkgrid")
ax = sns.countplot(y="Fitness", data=df, palette="vlag", order=df["Fitnes
s"].value_counts().index[0:10],hue="Gender")

plt.subplot(2, 3, 2)
plt.title('Maritalstatus-wise Distribution of Fitness in Dataset', fontsize
=12)
sns.set(style="darkgrid")
ax = sns.countplot(y="Fitness", data=df, palette="vlag", order=df["Fitnes
s"].value_counts().index[0:10],hue="MaritalStatus")
plt.show()
```





In [61]:
```
# Bi-Variate Analysis through kde plot
sns.kdeplot( data=df, x="Miles", hue="Fitness",
   fill=True, common_norm=False, palette="icefire",
   alpha=.5, linewidth=0)
plt.show()
```

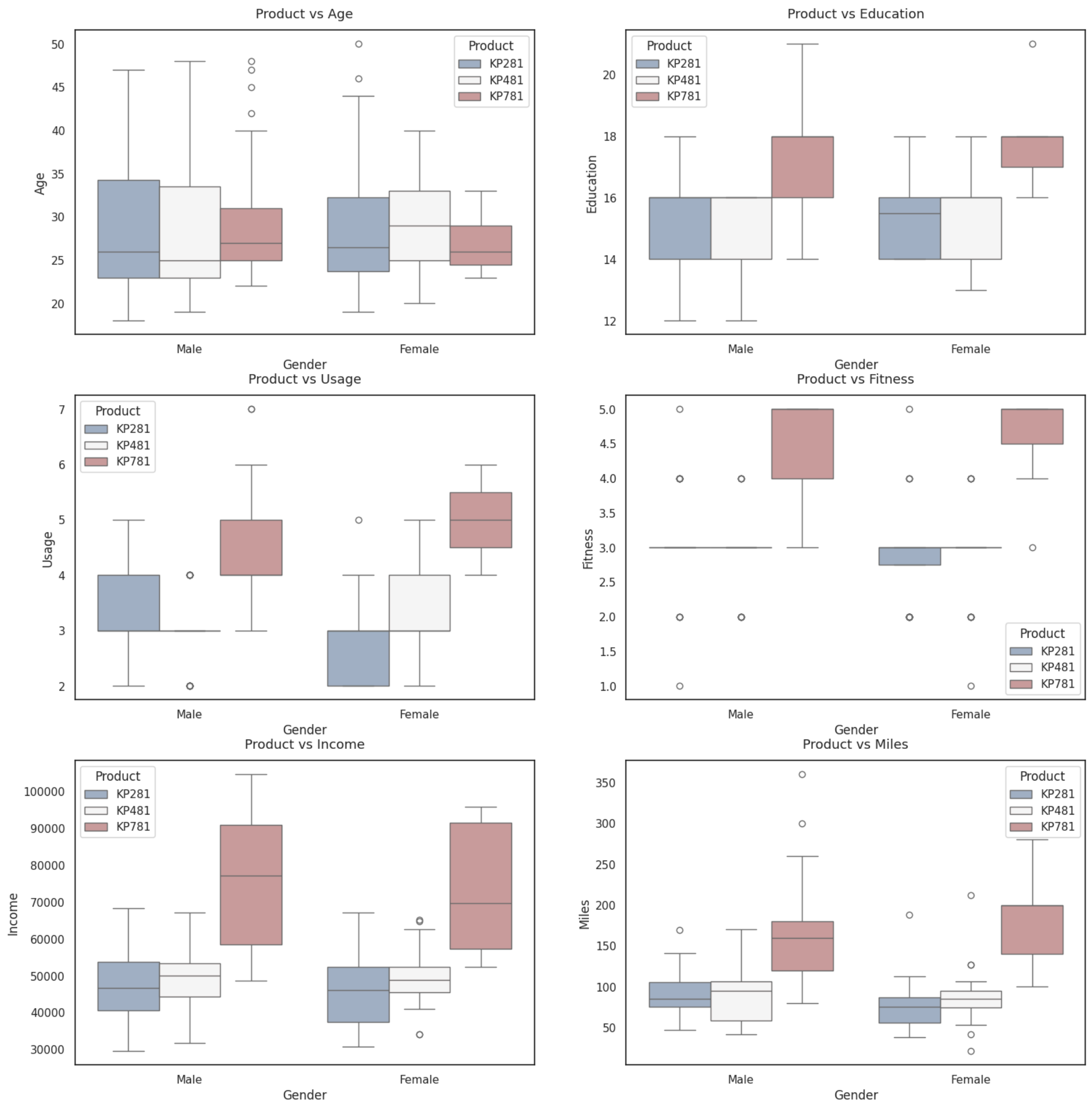


Observations:

There is a direct relationship between fitness level (self-rated body shape) and the average number of miles per week maintained by the user. Users with an excellent body shape rating of 5 tend to maintain the highest average number of miles per week.

3. Multivariate Analysis

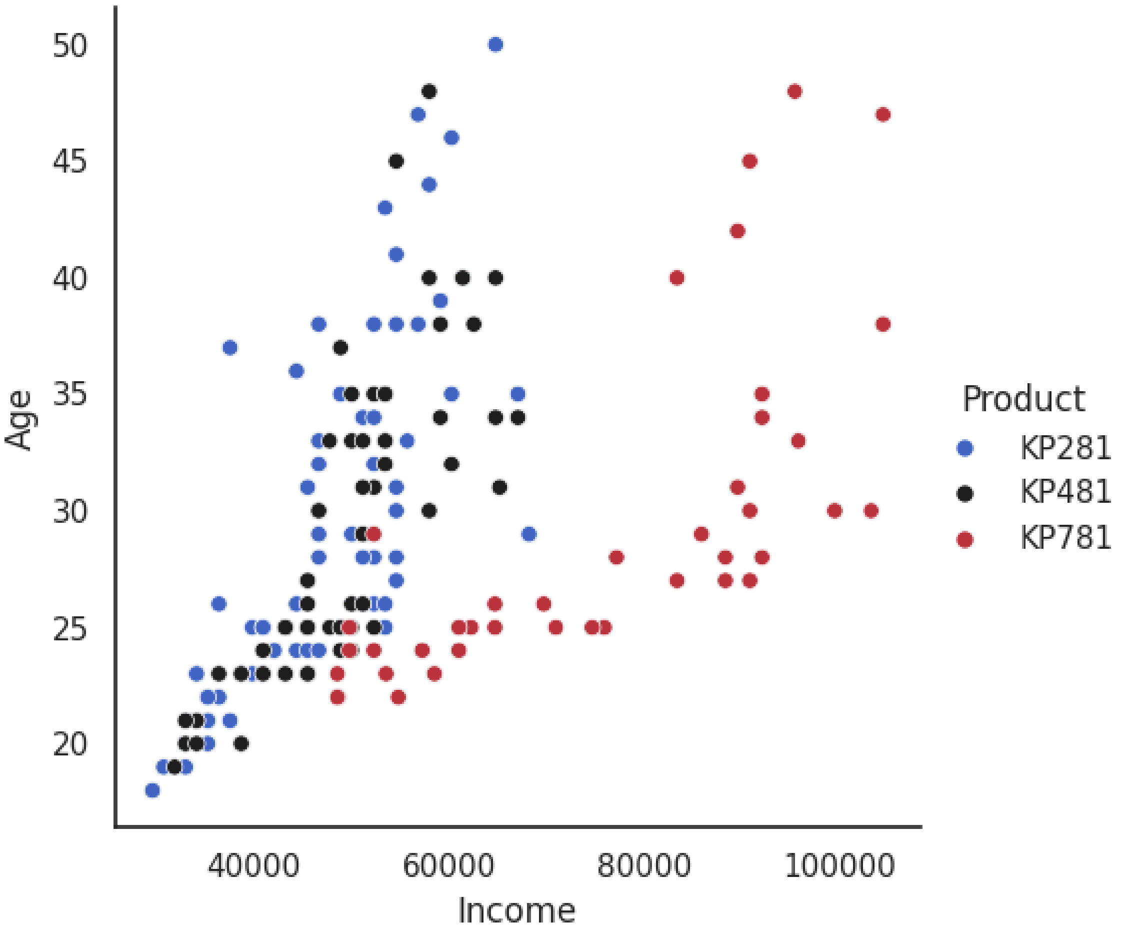
In [62]:
```
attrs = ['Age', 'Education', 'Usage', 'Fitness', 'Income', 'Miles']
sns.set_style("white")
fig, axs = plt.subplots(nrows=3, ncols=2, figsize=(18, 12))
fig.subplots_adjust(top=1.3)
count = 0
for i in range(3):
    for j in range(2):
        sns.boxplot(data=df, x='Gender', y=attrs[count], hue='Product', ax=
axs[i,j], palette='vlag')
        axs[i,j].set_title(f"Product vs {attrs[count]}", pad=12, fontsize=1
3)
        count += 1
```



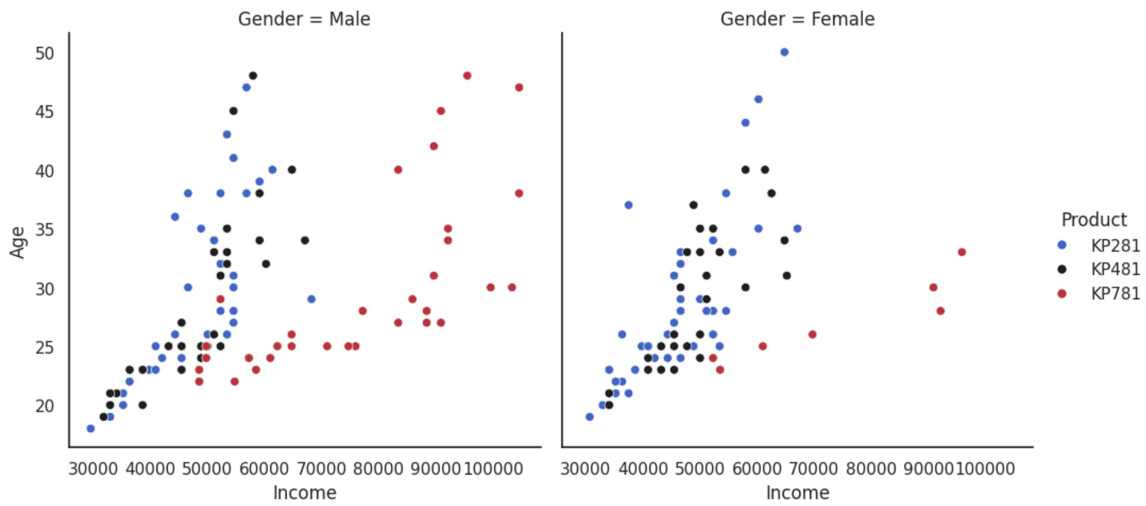
Observations:

Female customers who plan to use the treadmill 3-4 times a week are more inclined to purchase the KP481 product.

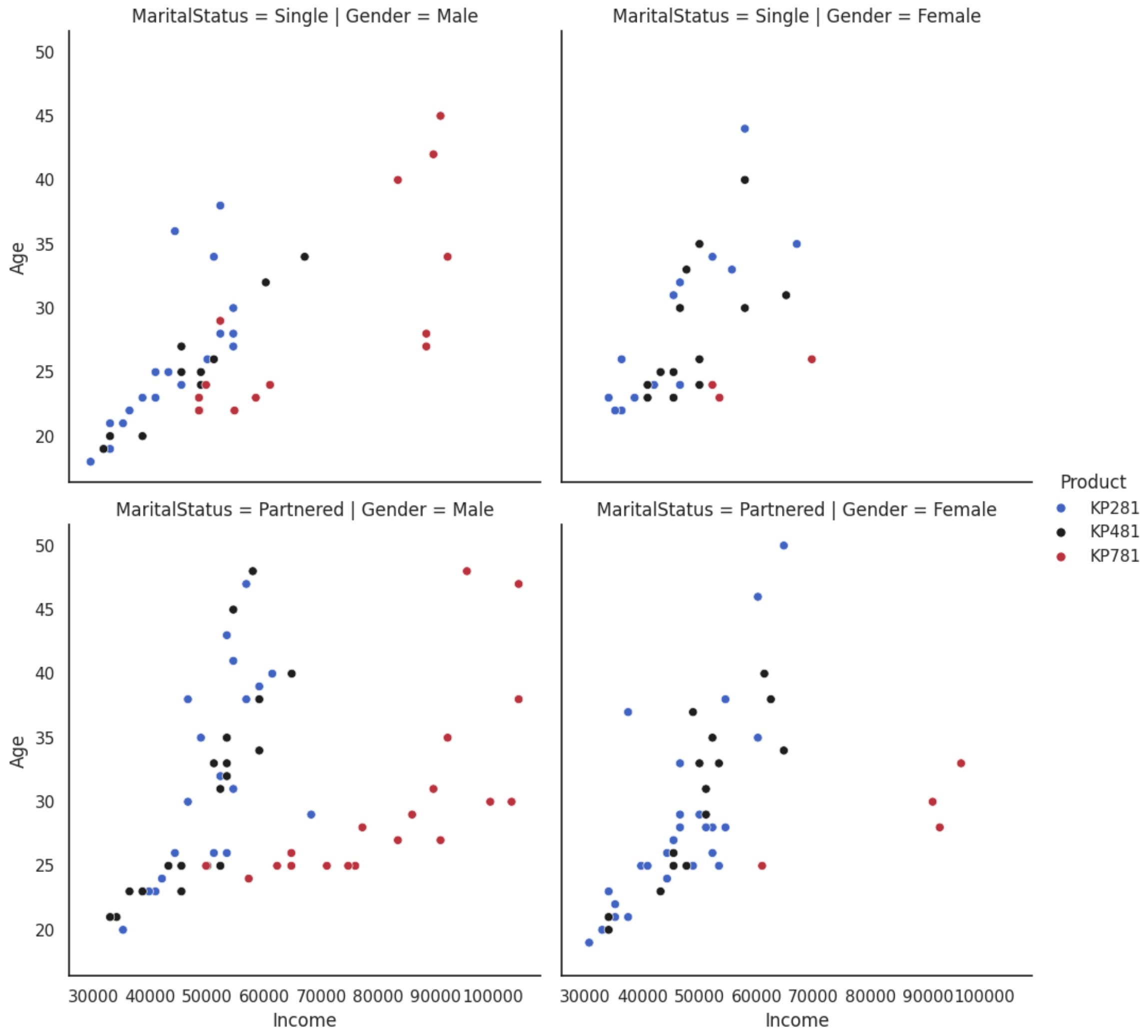
In [63]:
```
# Multi-Variate Analysis through scattered plot
sns.relplot(data=df, x="Income", y="Age", hue="Product",palette='icefire')
plt.show()
```



In [64]:
```
# Multi-Variate Analysis through scattered plot
sns.relplot(data=df, x="Income", y="Age", hue="Product", col="Gender",palet
te='icefire')
plt.show()
```



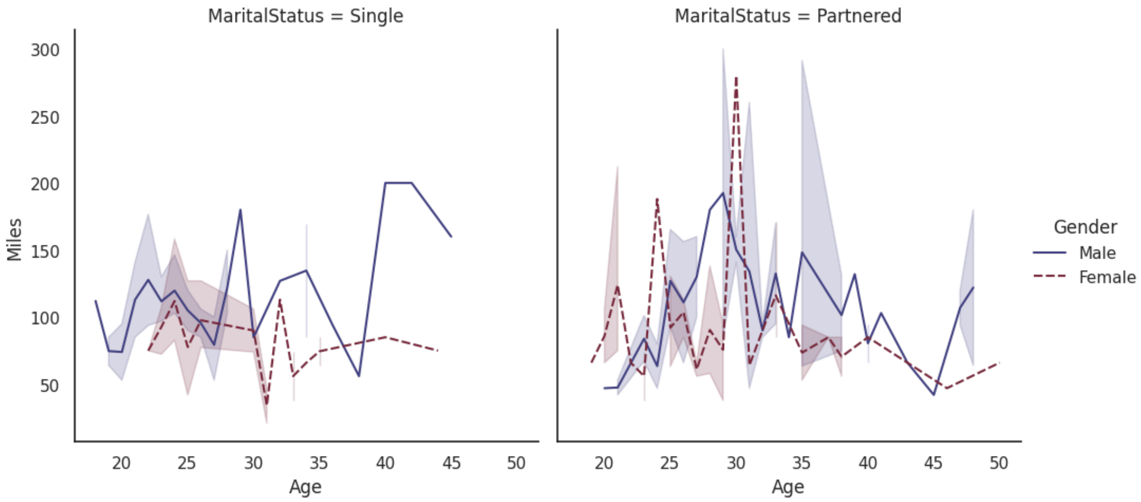
In [65]:
```
# Multi-Variate Analysis through scattered plot
sns.relplot(data=df, x="Income", y="Age", hue="Product", col="Gender",row
="MaritalStatus",palette='icefire')
plt.show()
```



Observations:

1. Only three females in the dataset have incomes above $70,000.
2. There appears to be a decreasing trend in the number of women with higher salaries as age increases.
3. The KP281 and KP481 products are primarily purchased by customers earning between $30,000-$70,000.

In [66]:
```
# Multi-Variate Analysis through line plot
sns.relplot(data=df, x="Age", y="Miles", col="MaritalStatus",
    hue="Gender", style="Gender", kind="line",palette='icefire')
plt.show()
```

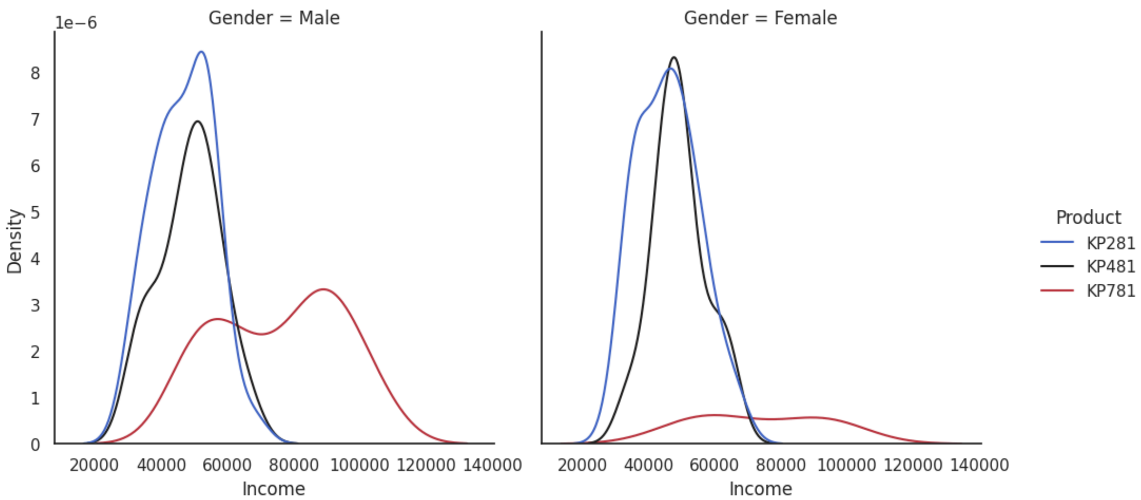


Observations:

Unmarried male customers exhibit the highest mileage, approximately 200 miles per week, around the age of 40 years. Unmarried female customers demonstrate peak mileage at approximately 33 years of age, averaging 110 miles per week.

Married male customers reach their highest mileage, approximately 190 miles per week, around the age of 28 years. Married female customers, on the other hand, show their peak mileage at around 30 years of age, averaging 280 miles per week.

In [67]:
```
# Multi-Variate Analysis through kde plot
sns.displot(data=df, x="Income", hue="Product", col="Gender", kind="kde",pa
lette='icefire')
plt.show()
```

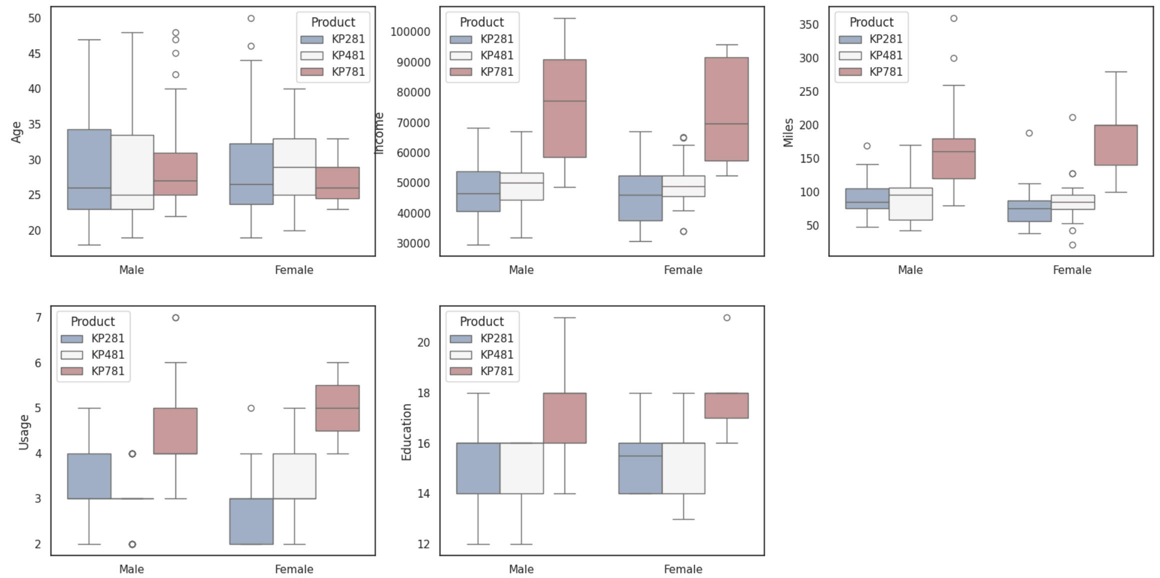


Observations:

Customers with incomes ranging from $30,000to$70,000 tend to purchase the KP281 and KP481 products. The KP781 product is primarily bought by customers earning more than $50,000.

3. Missing Value & Outlier Detection

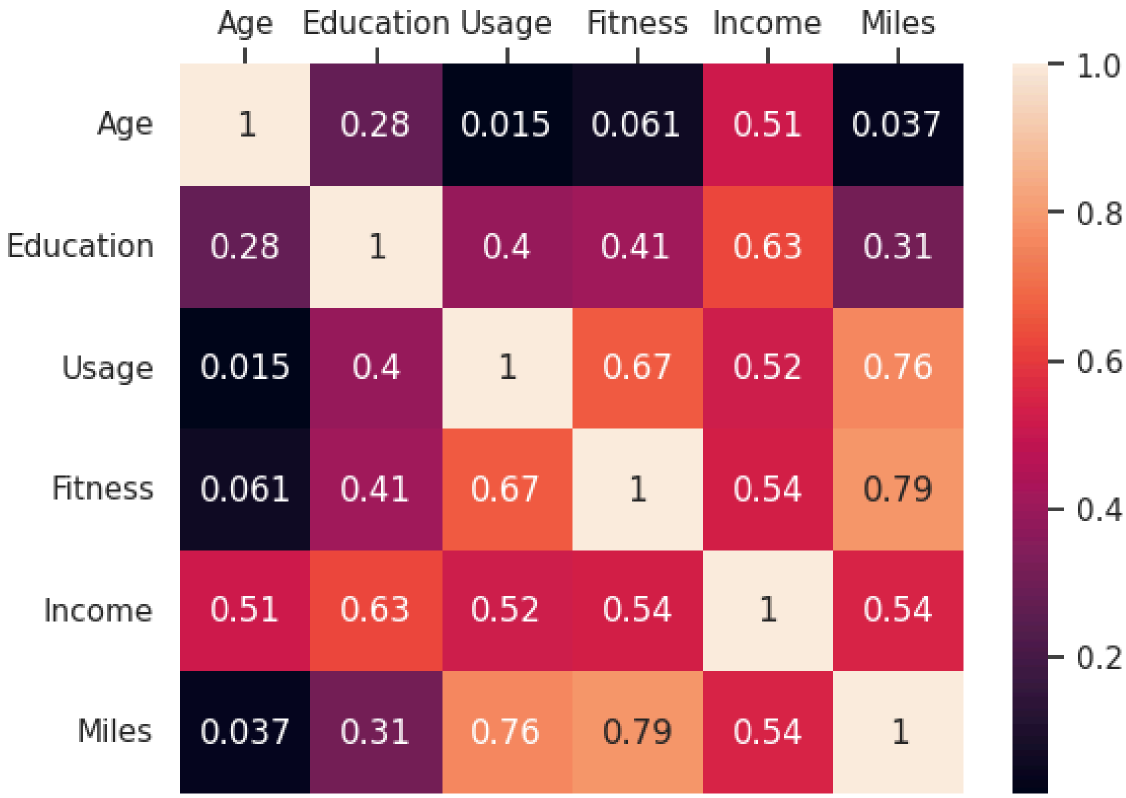
In [68]:
```
fig=plt.figure(figsize=(20,10))
plt.subplot(2,3,1)
sns.boxplot(data=df, x="Gender", y="Age", hue="Product",palette='vlag')
plt.ylabel("Age",fontsize=12)
plt.xlabel("",fontsize=12)

plt.subplot(2,3,2)
sns.boxplot(data=df, x="Gender", y="Income", hue="Product",palette='vlag')
plt.ylabel("Income",fontsize=12)
plt.xlabel("",fontsize=12)

plt.subplot(2,3,3)
sns.boxplot(data=df, x="Gender", y="Miles", hue="Product",palette='vlag')
plt.ylabel("Miles",fontsize=12)
plt.xlabel("",fontsize=12)

plt.subplot(2,3,4)
sns.boxplot(data=df, x="Gender", y="Usage", hue="Product",palette='vlag')
plt.ylabel("Usage",fontsize=12)
plt.xlabel("",fontsize=12)

plt.subplot(2,3,5)
sns.boxplot(data=df, x="Gender", y="Education", hue="Product",palette='vla
g')
plt.ylabel("Education",fontsize=12)
plt.xlabel("",fontsize=12)
plt.show()
```



Observations:

1. Users with higher salaries show a preference for the most expensive treadmill model, KP781.
2. Customers who purchase the KP781 treadmill tend to use the machine regularly and maintain a higher average number of miles per week.
3. Users with more years of education are inclined towards purchasing the most premium treadmill model.
4. Female users demonstrate higher weekly usage of the KP781 and KP481 treadmills, while the KP281 treadmill is primarily used by male users.

In [69]:
```
d_f=sns.heatmap(df[['Age','Education','Usage','Fitness','Income','Miles']].
corr(), annot=True)
d_f.set(xlabel="", ylabel="")
d_f.xaxis.tick_top()
```



Observations:

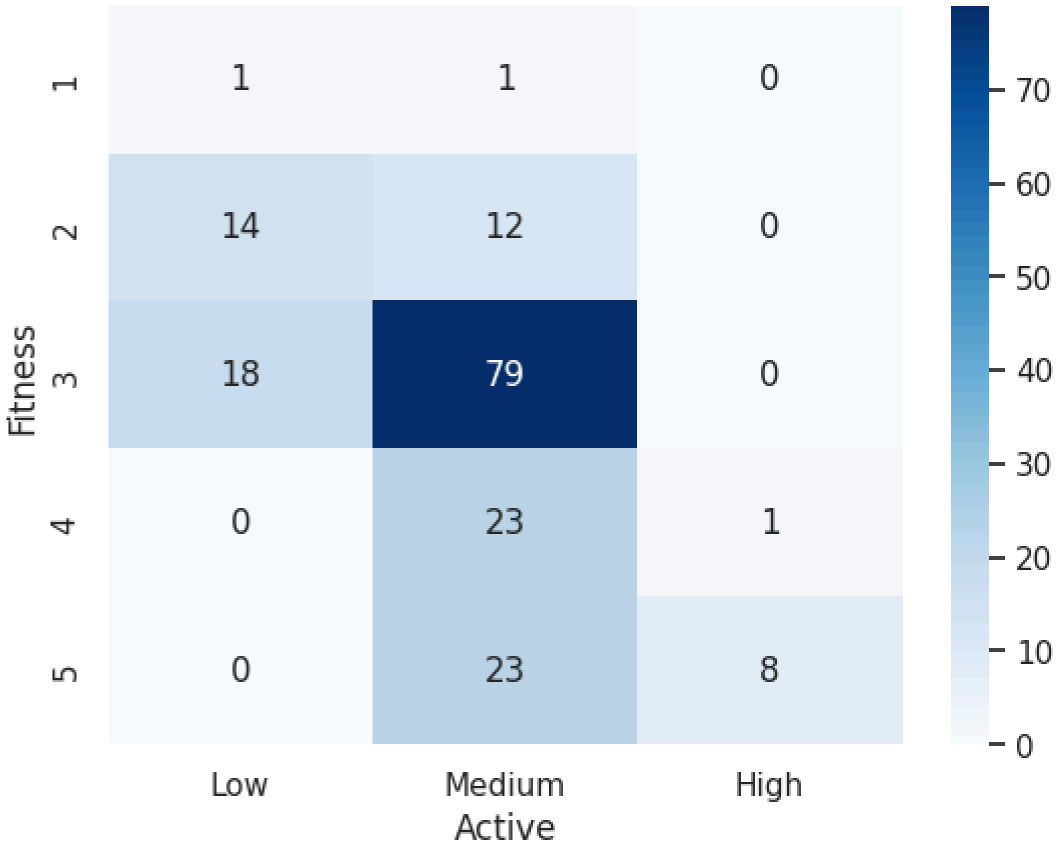
1. The correlation coefficient between fitness score and miles covered per week is 0.79, indicating a strong positive correlation between the two variables. This suggests that individuals with higher fitness scores tend to cover more miles per week on average.
2. The correlation coefficient between fitness score and usage per week is 0.67, indicating a moderately strong positive correlation. This implies that individuals with higher fitness scores tend to use the treadmill more frequently per week on average.
3. The correlation coefficient between miles covered per week and usage per week is 0.76, indicating a strong positive correlation. This suggests that individuals who cover more miles per week also tend to use the treadmill more frequently per week on average..

4. Categorization

In [70]:
```
# categorization of customers as per their usages
bins=[0,2,5,7]
labels=["Low","Medium","High"]
df["Active"]=pd.cut(x=df["Usage"],bins=bins,labels=labels,include_lowest=Fa
lse)
df["Active"].value_counts()
```

Out[70]:
```
Active
Medium    138
Low        33
High        9
Name: count, dtype: int64
```

In [71]:
```
cp_5=pd.crosstab(df["Fitness"],df["Active"])
sns.heatmap(cp_5,cmap="Blues",annot=True)
plt.show()
```



Observations:

Highly active individuals typically rate themselves with a fitness score of 5.

Average active individuals tend to rate themselves with a fitness score of 3.

Least active individuals generally rate themselves with a fitness score of 3 or below 3.

In [72]:
```
# defining category based on the income in three segment.
bins=[25000, 44058.75, 58668, 104581]
labels=["Low","Middle","High"]
df["Income Segment"]=pd.cut(x=df["Income"], bins=bins, labels=labels, inclu
de_lowest=False)
df["Income Segment"].value_counts()
```

Out[72]:
```
Income Segment
Middle    90
Low       45
High      45
Name: count, dtype: int64
```

5. Probability Analysis

5.1 Marginal Probabilities

In [73]:
```
Prod=df["Product"].value_counts(normalize=True).round(2)
print("P(KP281):",Prod["KP281"],'\n'"P(KP481):",Prod["KP481"],'\n'"P(KP78
1):",Prod["KP781"])
```

```
P(KP281): 0.44
P(KP481): 0.33
P(KP781): 0.22
```

In [74]:
```
Gen=df["Gender"].value_counts(normalize=True).round(2)
print("P(Male):",Gen["Male"],'\n'"P(Female):",Gen["Female"])
```

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

In [75]:
```
Mar=df["MaritalStatus"].value_counts(normalize=True).round(2)
print("P(Single):",Mar["Single"],'\n'"P(Partnered):",Mar["Partnered"])
```

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

In [76]:
```
df["Usage"].value_counts(normalize=True).round(2)
```

Out[76]:
```
Usage
3    0.38
4    0.29
2    0.18
5    0.09
6    0.04
7    0.01
Name: proportion, dtype: float64
```

5.2 Conditional Probability

5.2.1 Probability of puchasing treadmill for the given gender

In [77]:
```
cp_1=pd.crosstab(df["Product"],columns=df["Gender"])
cp_1
```

Out[77]:

Gender	Female	Male
Product		
KP281	40	40
KP481	29	31
KP781	7	33

In [78]:
```
kp281_M=cp_1["Male"]["KP281"].sum()/cp_1["Male"].sum()
kp281_F=cp_1["Female"]["KP281"].sum()/cp_1["Female"].sum()
kp481_M=cp_1["Male"]["KP481"].sum()/cp_1["Male"].sum()
kp481_F=cp_1["Female"]["KP481"].sum()/cp_1["Female"].sum()
kp781_M=cp_1["Male"]["KP781"].sum()/cp_1["Male"].sum()
kp781_F=cp_1["Female"]["KP781"].sum()/cp_1["Female"].sum()
print("P[KP281/Male]",":",kp281_M.round(2))
print("P[KP281/Female]",":",kp281_F.round(2))
print()
print("P[KP481/Male]",":",kp481_M.round(2))
print("P[KP481/Female]",":",kp481_F.round(2))
print()
print("P[KP781/Male]",":",kp781_M.round(2))
print("P[KP781/Female]",":",kp781_F.round(2))
```

```
P[KP281/Male] : 0.38
P[KP281/Female] : 0.53

P[KP481/Male] : 0.3
P[KP481/Female] : 0.38

P[KP781/Male] : 0.32
P[KP781/Female] : 0.09
```

5.2.2 Probability of puchasing treadmill for the given maritalstatus

In [79]:
```
cp_2=pd.crosstab(df["Product"],columns=df["MaritalStatus"])
cp_2
```

Out[79]:

MaritalStatus	Partnered	Single
Product		
KP281	48	32
KP481	36	24
KP781	23	17

In [80]:
```
kp281_P=cp_2["Partnered"]["KP281"].sum()/cp_2["Partnered"].sum()
kp281_S=cp_2["Single"]["KP281"].sum()/cp_2["Single"].sum()
kp481_P=cp_2["Partnered"]["KP481"].sum()/cp_2["Partnered"].sum()
kp481_S=cp_2["Single"]["KP481"].sum()/cp_2["Single"].sum()
kp781_P=cp_2["Partnered"]["KP781"].sum()/cp_2["Partnered"].sum()
kp781_S=cp_2["Single"]["KP781"].sum()/cp_2["Single"].sum()
print("P[KP281/Partnered]",":",kp281_P.round(2))
print("P[KP281/Single]",":",kp281_S.round(2))
print()
print("P[KP481/Partnered]",":",kp481_P.round(2))
print("P[KP481/Single]",":",kp481_S.round(2))
print()
print("P[KP781/Partnered]",":",kp781_P.round(2))
print("P[KP781/Single]",":",kp781_S.round(2))
```

```
P[KP281/Partnered] : 0.45
P[KP281/Single] : 0.44

P[KP481/Partnered] : 0.34
P[KP481/Single] : 0.33

P[KP781/Partnered] : 0.21
P[KP781/Single] : 0.23
```

5.2.3 Probability of puchasing treadmill for the given Income Segmen

In [81]:
```
cp_3=pd.crosstab(df["Product"],columns=df["Income Segment"])
cp_3
```

Out[81]:

Income Segment	Low	Middle	High
Product			
KP281	30	43	7
KP481	15	36	9
KP781	0	11	29

In [82]:
```
kp281_H=cp_3["High"]["KP281"].sum()/cp_3["High"].sum()
kp281_M=cp_3["Middle"]["KP281"].sum()/cp_3["Middle"].sum()
kp281_L=cp_3["Low"]["KP281"].sum()/cp_3["Low"].sum()
kp481_H=cp_3["High"]["KP481"].sum()/cp_3["High"].sum()
kp481_M=cp_3["Middle"]["KP481"].sum()/cp_3["Middle"].sum()
kp481_L=cp_3["Low"]["KP481"].sum()/cp_3["Low"].sum()
kp781_H=cp_3["High"]["KP781"].sum()/cp_3["High"].sum()
kp781_M=cp_3["Middle"]["KP781"].sum()/cp_3["Middle"].sum()
kp781_L=cp_3["Low"]["KP781"].sum()/cp_3["Low"].sum()
print("P[KP281/High]",":",kp281_H.round(2))
print("P[KP281/Middle]",":",kp281_M.round(2))
print("P[KP281/Low]",":",kp281_L.round(2))
print()
print("P[KP481/High]",":",kp481_H.round(2))
print("P[KP481/Middle]",":",kp481_M.round(2))
print("P[KP481/Low]",":",kp481_L.round(2))
print()
print("P[KP781/High]",":",kp781_H.round(2))
print("P[KP781/Middle]",":",kp781_M.round(2))
print("P[KP781/Low]",":",kp781_L.round(2))
```

```
P[KP281/High] : 0.16
P[KP281/Middle] : 0.48
P[KP281/Low] : 0.67

P[KP481/High] : 0.2
P[KP481/Middle] : 0.4
P[KP481/Low] : 0.33

P[KP781/High] : 0.64
P[KP781/Middle] : 0.12
P[KP781/Low] : 0.0
```

6. Customer Profiling

KP781:

Targeted towards high-income users.

Predominantly preferred by married individuals.

Users typically rate their fitness level higher than 3.

High average miles covered per week.

Male customers are the primary purchasers.

Users generally have 16 or more years of education.

Regular exercisers are the primary target customers.

KP481:

Positioned in the middle price segment.

Users typically have a usage frequency of 4 or below.

Slightly higher average miles covered per week compared to KP281.

Also popular among married couples.

Users fall under the middle-income category.

KP281:

Represents 44% of the total customer base.

Attracts users with lower income and lower weekly usage frequencies.

Average usage is below 150 miles per week.

Preferred by users with less than 16 years of education.

Fitness rating is typically under 3.

7. Business Insights:

1. KP281 accounts for 44% of total treadmill sales, indicating its popularity among customers.
2. The majority of customers (58%) are male, suggesting a gender imbalance in treadmill purchases.
3. 60% of customers are married, highlighting the importance of family demographics in purchasing decisions.
4. There's an inverse relationship between unit price and sales volume, indicating price sensitivity among customers.
5. KP781 is predominantly preferred by males, with a significantly higher adoption rate compared to females.
6. The probability of low-income users purchasing KP781 is negligible (0.00), while high-income users show a higher likelihood (0.64).
7. Low-income users are more likely to purchase KP281 (0.67) compared to high-income users (0.16), indicating price sensitivity.
8. The probability of unmarried users purchasing KP281 is similar to married users, suggesting broad appeal across marital statuses.
9. However, unmarried users show a slightly higher probability of purchasing KP781 compared to married users.
10. Male users are more inclined to purchase KP781 compared to female users, indicating a gender preference for this premium product.

8. Recommendations:

1. Diversify Product Range: Consider expanding the product range to cater to a wider audience. Introducing more mid-range options could attract customers who are currently underserved by the existing product lineup.
2. Premium Product Promotion: KP781 being the premium product should be promoted in the premium segment, targeting high-income individuals who value advanced features and superior quality.
3. Segment-Wise Targeting: Utilize customer profiles to segment the market effectively. Focus marketing efforts on specific customer segments based on income, marital status, gender, and usage patterns.
4. Promotion of KP481: Since KP281 and KP481 share similar user profiles, prioritize the promotion of KP481 to leverage its higher price point and potentially increase revenue.
5. Highlighting KP781 Benefits: Position KP781 as the ultimate treadmill choice for long-duration workouts and superior exercise experience, emphasizing its advanced features and capabilities.
6. Targeted Advertising: Use customer purchase history data to tailor ads and marketing campaigns to relevant customer segments, maximizing the impact of advertising efforts.
7. Add-Ons for KP481: Offer additional incentives or add-ons with the purchase of KP481 to attract more customers and increase its appeal compared to other models. This could include free accessories, extended warranties, or discounted gym memberships.

9. Conclusion

In conclusion, the analysis of Aerofit's treadmill sales data has provided valuable insights into customer behavior, product preferences, and market trends. The observations and recommendations derived from the analysis can inform strategic decisions to optimize sales, enhance customer satisfaction, and drive business growth.

Key findings include the popularity of the KP281 treadmill among customers, particularly those with lower incomes and usage frequencies. Conversely, the KP781 treadmill appeals to high-income individuals seeking premium features and advanced functionality. The KP481 treadmill occupies a middle ground, offering a balance between price and features.

Recommendations for Aerofit include diversifying the product range to cater to different customer segments, implementing targeted marketing campaigns, optimizing pricing strategies, and enhancing the overall customer experience.

By implementing these recommendations and staying agile in response to market dynamics, Aerofit can position itself for sustained success in the competitive fitness equipment industry.