In [1]:

```
import pandas as pd
df = pd.read_csv("C:/Users/maheshmangaonkar/Desktop/aerofit_treadmill.csv")
```

In [2]:

```
df
```

Out[2]:

	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

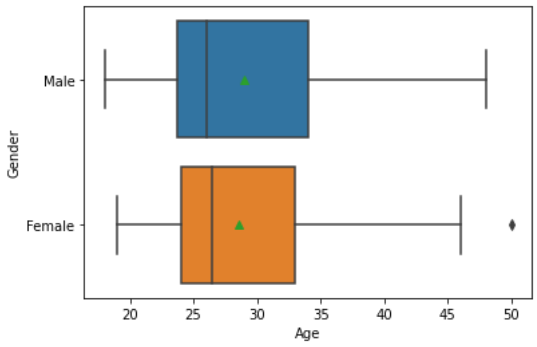
180 rows × 9 columns

In [3]:

```
#structure*
#There are 180* 9 columns in the dataset.It describes the product KP281, KP481,KP781 and the description of people using it
```

In [4]:

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

bp = sns.boxplot(x = 'Age', y = 'Gender',data = df, orient = 'h',showmeans = True,)
```



In [ ]:

```

```

In [5]:

```
# For male there are no values outside the outliers which means that the age group of male lies within the IQR
# Lower age for male is less than that for female and upper age of male is greater than that for females
```

In [6]:

```
Q3 = np.quantile(df['Age'], 0.75)
Q1 = np.quantile(df['Age'], 0.25)
IQR = Q3 - Q1
IQR
```

Out[6]:

```
9.0
```

In [7]:

```
a = np.mean(df['Age'])
```

In [8]:

```
b = np.median(df['Age'])
difference = a - b
difference
```

Out[8]:

```
2.788888888888888
```

In [9]:

```
lower = max(Q1 - 1.5*IQR, 0)
lower
```

Out[9]:

```
10.5
```

In [10]:

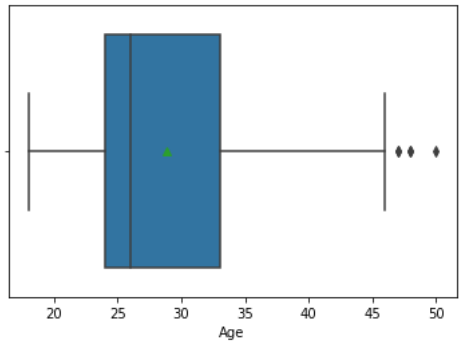
```
upper = Q3 + 1.5 * IQR
upper
```

Out[10]:

```
46.5
```

In [11]:

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

bp = sns.boxplot(x = 'Age', data = df, orient = 'h',showmeans = True,)
```



In [12]:

```
age_outlier = df[(df["Age"] < lower) | (df["Age"]>upper)]
len(age_outlier)
```

Out[12]:

```
5
```

In [13]:

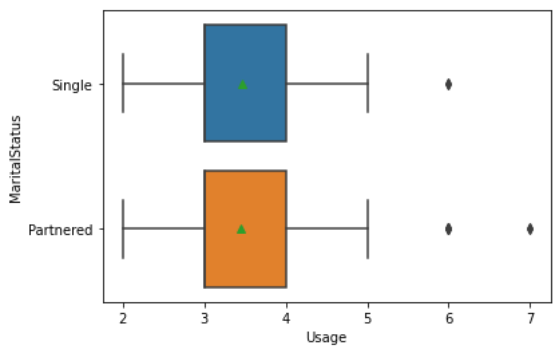
```
#There are 5 people with age greater than upper age in the data set
```

In [14]:

```
sns.boxplot(y = 'MaritalStatus', x = 'Usage',data = df, orient = 'h',showmeans = True)
```

Out[14]:

```
<AxesSubplot:xlabel='Usage', ylabel='MaritalStatus'>
```



In [15]:

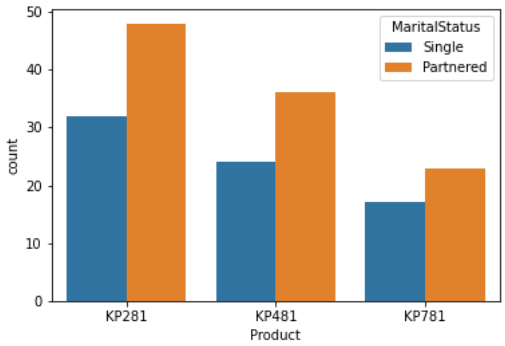
```
#The  mean usage per week is seem to be same for partnered or single persons
```

In [16]:

```
sns.countplot(data=df, x="Product", hue="MaritalStatus")
```

Out[16]:

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



In [17]:

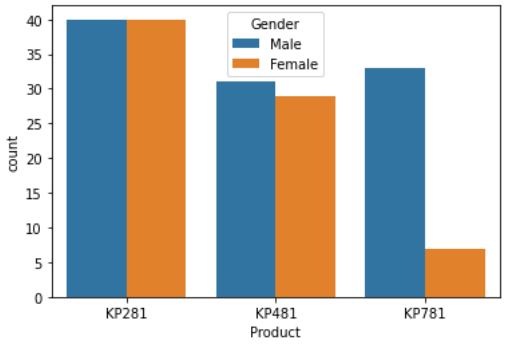
```
#Partenerd people are more likely to purchae the product
```

In [18]:

```
sns.countplot(data=df, x="Product", hue="Gender")
```

Out[18]:

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



In [ ]:

```

```

In [19]:

```
# From the countplot it is clear that are focused on purchasing KP281 which is cheaper than other two products. The trend is
#same for single and partnered persons.
```

In [20]:

```
sns.histplot(data=df, x="Age", hue="Product
```

```
  Input In [20]
    sns.histplot(data=df, x="Age", hue="Product
                                               ^
SyntaxError: EOL while scanning string literal
```

In [ ]:

```
# People in the age group of 21-27 are interested with KP781 and KP481 as they prefer product with advance features
```

In [ ]:

```
pd.crosstab(index = df['Product'],columns =df['Age'],margins = True )
```

In [21]:

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

Out[21]:

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

In [28]:

```
#Conditional Probabilities
#
sns.pairplot(df, hue ='Product')
```

Out[28]:

```
<seaborn.axisgrid.PairGrid at 0x54be12f970>
```

In [ ]:

```
#People with lower Education likely to purchase KP281 and KP481. And people with Education higher levels greater  are likely to purchase K
```

In [ ]:

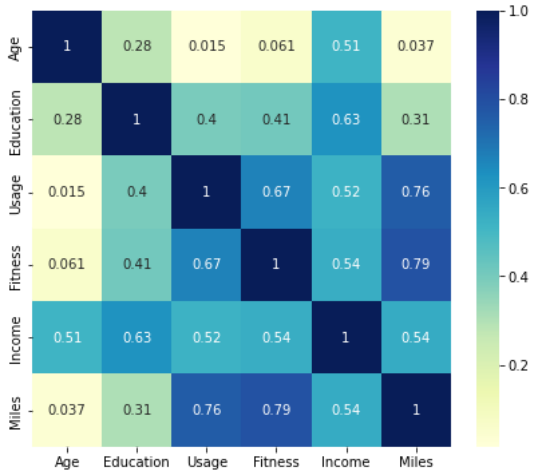
```

```

In [29]:

```
plt.figure(figsize=(7,6))
sns.heatmap(df.corr(), cmap="YlGnBu", annot=True)
```

Out[29]:

<AxesSubplot:>



In [30]:

```
#What is the probability of a male customer buying a KP781 treadmill?
pd.crosstab(df['Product'],[df['Gender']], normalize=True, margins=True, margins_name='Total').round(2)
```

Out[30]:

Gender	Female	Male	Total
Product			
KP281	0.22	0.22	0.44
KP481	0.16	0.17	0.33
KP781	0.04	0.18	0.22
Total	0.42	0.58	1.00

In [32]:

```
#P(P781/Male) = 0.18/0.58 = 0.31034
0.18/0.58
```

Out[32]:

```
0.3103448275862069
```

In [33]:

```
#P(P781/Female) = 0.04/0.42 = 0.095238
0.04/0.42
```

Out[33]:

```
0.09523809523809525
```

In [34]:

```
pd.crosstab(df['Product'],[df['Fitness']], normalize=True, margins=True, margins_name='Total').round(2)
```

Out[34]:

Fitness	1	2	3	4	5	Total
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
Total	0.01	0.14	0.54	0.13	0.17	1.00

In [ ]:

```
# People with Fitness Level 3 or less are likely to purchase KP281 and KP481. And with Fitness Level 5 are likely to purchase KP781.
```

In [35]:

```
pd.crosstab(df['Product'],[df['Income']], normalize=True, margins=True, margins_name='Total').round(2)
```

Out[35]:

Income	29562	30699	31836	32973	34110	35247	36384	37521	38658	39795	...	88396	89641	90886	92131	95508	95866	99601	103336	104581
Product																				
KP281	0.01	0.01	0.01	0.02	0.01	0.03	0.02	0.01	0.02	0.01	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KP481	0.00	0.00	0.01	0.01	0.02	0.00	0.01	0.00	0.01	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
KP781	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01
Total	0.01	0.01	0.01	0.03	0.03	0.03	0.02	0.01	0.03	0.01	...	0.01	0.01	0.02	0.02	0.01	0.01	0.01	0.01	0.01

4 rows × 63 columns

In [ ]:

```
# People with Income less than 60000 are likely to purchase KP281 and KP481. And people with Income greater than 60000 are likely to purch
```