In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Above is the Libraries to fetch the Analytical and Descriptive tool to work on the Analysis

The Below is the way tp get the Data set from the Pandas library to evaluate and get in touch with the deatils of the Aerofit Data

In [2]:
```
df = pd.read_csv('aerofit.csv')
```

To get familier with the data set the head funtion will return the top 5 values of the data to Analyse the Character of Data

Analysis of Data and its Structure

In [9]:
```
df.head()
```

Out[9]:

	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

To get familier with the data set the Tail funtion which is similar to head, will return the Bottom 5 values of the data to Analyse the Character of Data

In [8]:
```
df.tail()
```

Out[8]:

	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

The info function in the library function will say the structure if the data, where the columns are describe into which category
of the data mainly comes from, well like the Data type of the each column name

In [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
```

The data set having no null values, which shows the data is purely clear and easy to navigate with the all present values.

In [5]:
```
df.isna().sum()
```

Out[5]:
```
Product          0
Age              0
Gender           0
Education        0
MaritalStatus    0
Usage            0
Fitness          0
Income           0
Miles            0
dtype: int64
```

The describe function will sets the data into the statistics details of the numrical structure of the data, where the values
from the data set will generally speaks about the count of data. the mean value and the nth percentile value for of the generated columns

In [14]:
```
df.describe()
```

Out[14]:

	Age	Education	Usage	Fitness	Income	Miles
count	180.000000	180.000000	180.000000	180.000000	180.000000	180.000000
mean	28.788889	15.572222	3.455556	3.311111	53719.577778	103.194444
std	6.943498	1.617055	1.084797	0.958869	16506.684226	51.863605
min	18.000000	12.000000	2.000000	1.000000	29562.000000	21.000000
25%	24.000000	14.000000	3.000000	3.000000	44058.750000	66.000000
50%	26.000000	16.000000	3.000000	3.000000	50596.500000	94.000000
75%	33.000000	16.000000	4.000000	4.000000	58668.000000	114.750000
max	50.000000	21.000000	7.000000	5.000000	104581.000000	360.000000

In [20]:
```
df.describe(include=object)
```

Out[20]:

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

Here we have the above details from the data where the data column having the data type as object which also termed as str format
we, know the numerical value having the stats details, but when it comes to str or the object type, the function can also have the
performance to calculate the neccesary deatils for the object data type

In [18]:
```
df.shape
```

Out[18]: (180, 9)

The shape will return the dataa count of the data where the rows and columns, specifies the number of rows and columns
here the data have the 180 rows and 9 columns

In [19]:
```
df.columns
```

Out[19]:
```
Index(['Product', 'Age', 'Gender', 'Education', 'MaritalStatus', 'Usage',
       'Fitness', 'Income', 'Miles'],
      dtype='object')
```

Here is the Columns name startisfies the particular name of the column in data

In [21]:
```
df.size
```

Out[21]: 1620

```
There are the total in the data set 1620 incuding the rows and the columns
```

In [24]:

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

Out[24]:

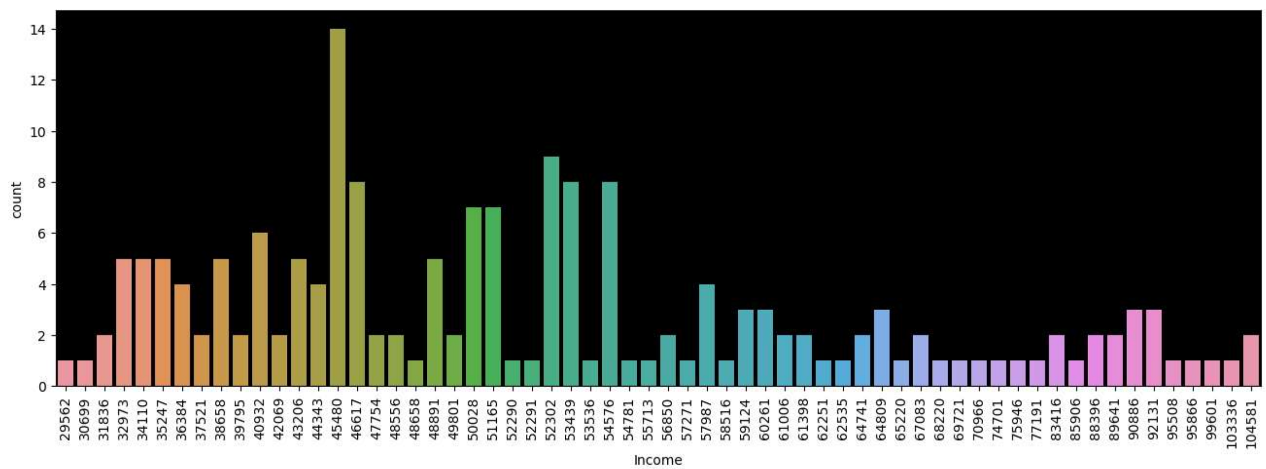
```
Product          False
Age              False
Gender           False
Education        False
MaritalStatus    False
Usage            False
Fitness          False
Income           False
Miles            False
dtype: bool
```

```
In the data set there were no single rows are there which have an any single NAN or null values in the data set
```

```
1. There are 3 unique products in the dataset.
2. KP281 will be the top major product to be sold
3.The average age of the population is 28.79; the minimum and maximum ages are 18 and 50 respectively,
and 75% of people are under the age of 33.
4. Mean income is approx in healthy line where the 53K incomes as satisfies the result for the aerofit to engaged customers
5.Majority people are having 16 years of education i.e., 75% of persons are having
education <= 16 years.
6. Out of 180 data points, 104's gender is Male and rest are the female.
7. Standard deviation for Income & Miles is very high. These variables might have the
outliers in it.
```

In [128]:

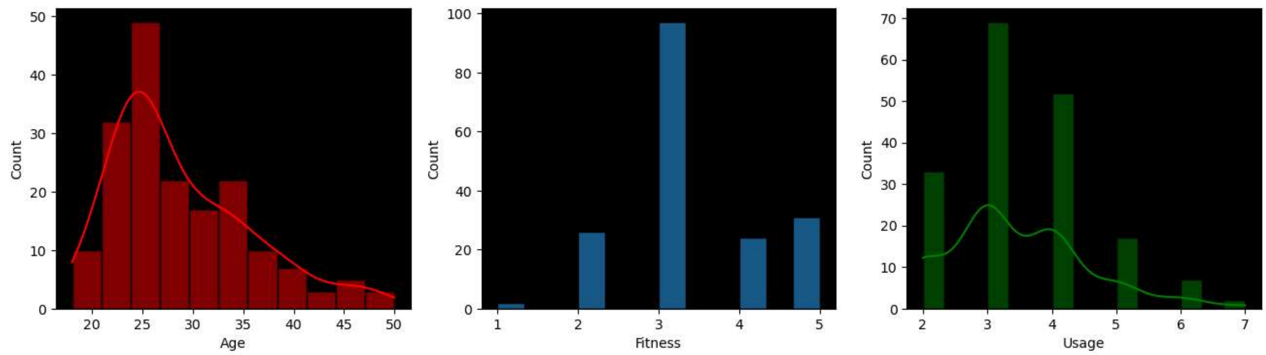
```
plt.figure(figsize=(16,5), dpi=100)
plt.xticks(rotation = 90)
ax=plt.gca()
ax.set_facecolor('black')
sns.countplot(data=df, x='Income')
plt.show()
```




```
The Above graph the income distribution among the people having the gathered tendency to the inrichment of the data incomes
categories of the data
Insights: The graph depicts the income distribution among individuals with a proclivity for accumulating wealth. It
highlights the various income brackets within the dataset. Business insights from this graph could include identifying
segments with high income potential, devising targeted marketing strategies for affluent groups, and tailoring products or
services to cater to their preferences, thereby optimizing revenue generation.
```

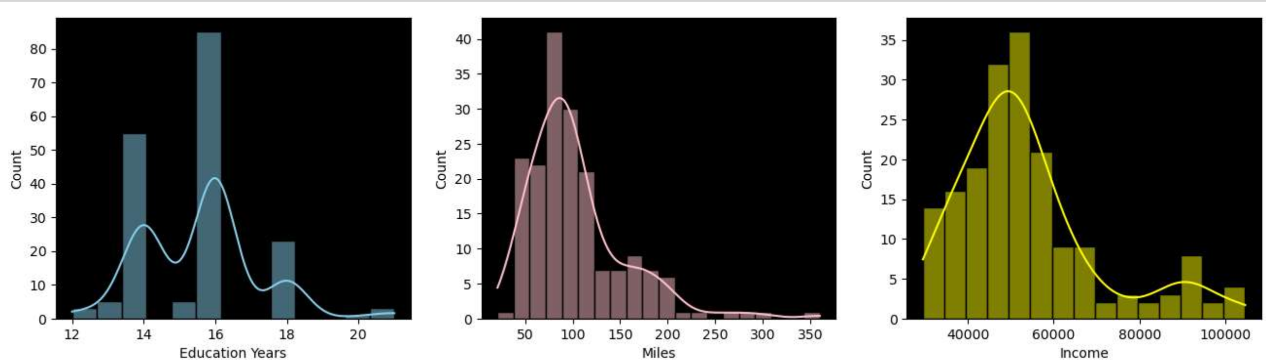
Uni-Variate Analyis

In [153]:
```
plt.figure(figsize=(16,4),dpi=100)
plt.subplot(1,3,1)
ax=plt.gca()
ax.set_facecolor('black')
sns.histplot(data=df, x='Age', kde=True,color='red')
plt.subplot(1,3,2)
ax=plt.gca()
ax.set_facecolor('black')
sns.histplot(data=df, x='Fitness')
plt.subplot(1,3,3)
ax=plt.gca()
ax.set_facecolor('black')
sns.histplot(data=df, x='Usage', kde=True,color='green')
plt.xticks()
plt.show()
```




```
The univariate Analysis for the Particular characterstics behaviour followed by the count with analysis in respect to the
number of the people.
Insights: The maximum ranges of the people who come to the aerofit is from 22 to 27, the aerfoit must and should comsider
the
people from thr age ranges between 22 to 27 with their target and descent wise
The majority of the xustomer who has done the self fitness scale rating is the 3.
Maximum number of the people come to the range in week at the day 3, the aero fit must share their insights wiht the some
beautiful ways to get the attracted customers in the while as well
```

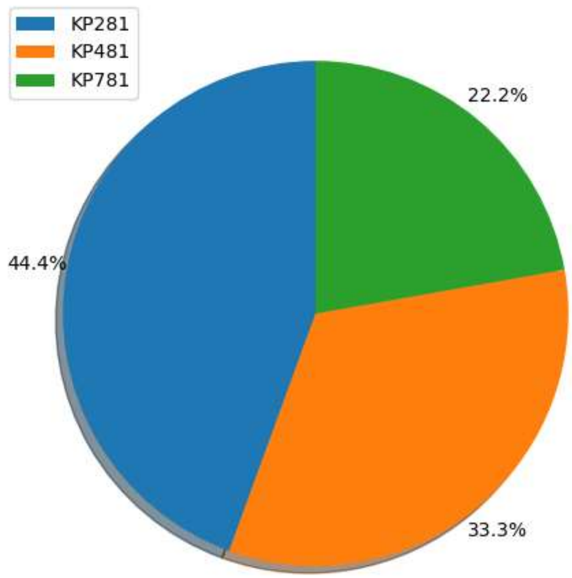
In [133]:
```
plt.figure(figsize=(16,4),dpi=100)
plt.subplot(1,3,1)
ax=plt.gca()
ax.set_facecolor('black')
sns.histplot(data=df, x='Education', kde=True, color='skyblue')
plt.xlabel("Education Years")
plt.subplot(1,3,2)
ax=plt.gca()
ax.set_facecolor('black')
sns.histplot(data=df, x='Miles', kde=True,color="pink")
plt.subplot(1,3,3)
ax=plt.gca()
ax.set_facecolor('black')
sns.histplot(data=df, x='Income',kde=True,color="yellow")
plt.show()
```




```
The univariate Analysis for the Particular characterstics behaviour followed by the count with analysis in respect to the
number of the people.
Insight: From the Education wise here we can observe the the age in between the max to 16 will fall into the major concern
```

with the fitness
While in also the trademill covers the particular 80 to 100 miles for the users used so far at peak level, also the income concern with the 50k to 60k incomes ranges for the whole scenario

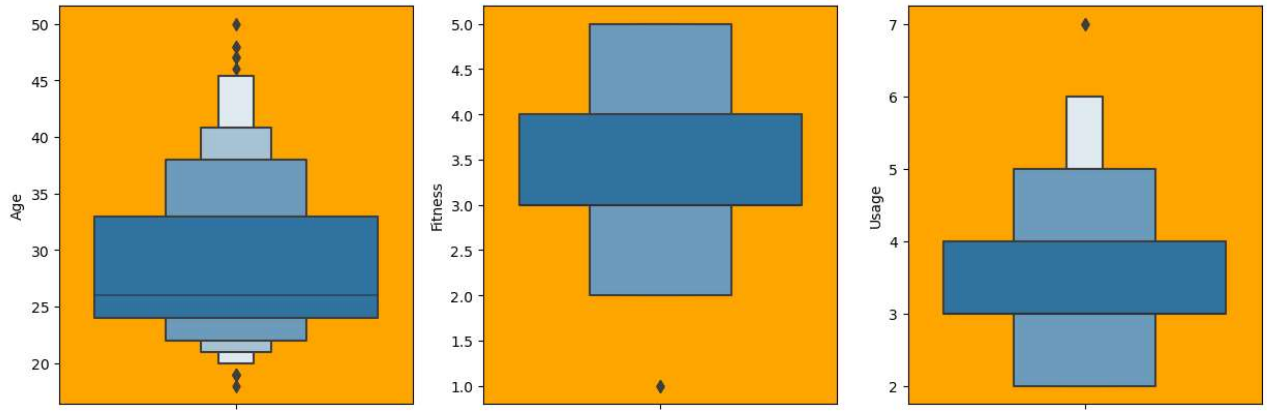
In [147]:
```
for_pie = df['Product'].value_counts()
for_pie
plt.figure(figsize=(15,6))
mylabels = ["KP281","KP481","KP781"]
plt.pie(for_pie.values,autopct='%1.1f%%', startangle=90, shadow=True, pctdistance=1.12)
plt.legend(labels=for_pie.index, loc='upper left')
plt.show()
```



Insights: The table illustrates product sales distribution, showcasing KP281 as the highest seller at 80 units, followed by KP481 at 60 units, and KP781 at 40 units. These figures can be visually represented through a pie chart, highlighting the dominance of KP281, a significant share for KP481, and a smaller segment for KP781 in the overall sales composition.

Detect Outliers for Structure

In [3]:
```
plt.figure(figsize=(15,5))
plt.subplot(1,3,1)
ax=plt.gca()
ax.set_facecolor('orange')
sns.boxenplot(data=df, y="Age")
plt.subplot(1,3,2)
ax=plt.gca()
ax.set_facecolor('orange')
sns.boxenplot(data=df,y='Fitness')
plt.subplot(1,3,3)
ax=plt.gca()
ax.set_facecolor('orange')
sns.boxenplot(data=df,y='Usage')
plt.show()
```



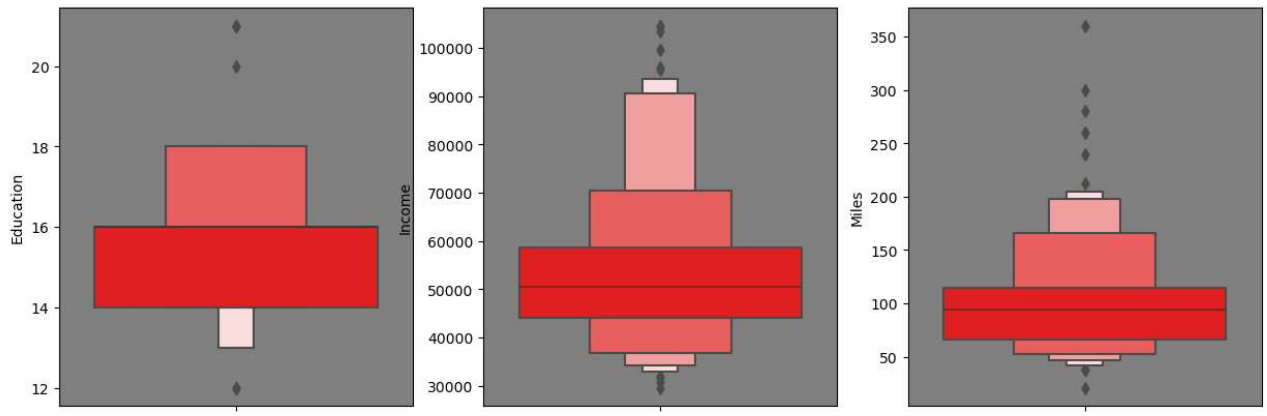

```
Insights: A veryfew Age people lies in the outliers, consisting the more than the age 45 upto the age 50 considered in the range of the outliers.
while the Fitness rating scale only 1 has been detected who has given the rating 1 to the particular person.
while the usage having very less who takes in part of the week section.
```

In [130]:
```
df[['Age','Fitness','Usage']].describe()
```

Out[130]:

	Age	Fitness	Usage
count	180.000000	180.000000	180.000000
mean	28.788889	3.311111	3.455556
std	6.943498	0.958869	1.084797
min	18.000000	1.000000	2.000000
25%	24.000000	3.000000	3.000000
50%	26.000000	3.000000	3.000000
75%	33.000000	4.000000	4.000000
max	50.000000	5.000000	7.000000

In [123]:
```
plt.figure(figsize=(15,5))
plt.subplot(1,3,1)
ax=plt.gca()
ax.set_facecolor('gray')
sns.boxenplot(data=df, y="Education",color='red')
plt.subplot(1,3,2)
ax=plt.gca()
ax.set_facecolor('gray')
sns.boxenplot(data=df,y='Income',color='red')
plt.subplot(1,3,3)
ax=plt.gca()
ax.set_facecolor('gray')
sns.boxenplot(data=df,y='Miles',color='red')
plt.show()
```




```
Insight: The less people also considerd for the education in years are more than 20, tha ranges from the years as outliers where as from the income level who has more than 90k of the income will be considered as large which has the income more than the highest range, and also the from the category miles some of the people are there who are run more than 200 miles.
```

In [131]:
```
df[['Education','Income','Miles']].describe()
```

Out[131]:

	Education	Income	Miles
count	180.000000	180.000000	180.000000
mean	15.572222	53719.577778	103.194444
std	1.617055	16506.684226	51.863605
min	12.000000	29562.000000	21.000000
25%	14.000000	44058.750000	66.000000
50%	16.000000	50596.500000	94.000000
75%	16.000000	58668.000000	114.750000
max	21.000000	104581.000000	360.000000

Featuring the Marital Status and Age with Product

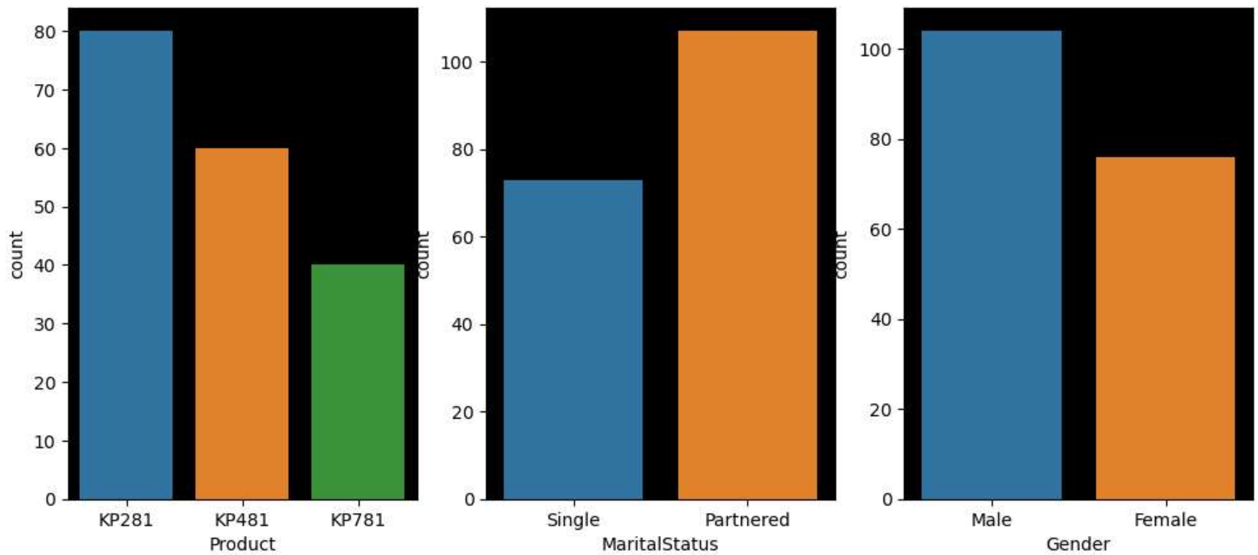
In [136]:
```
df['Product'].value_counts()
```

Out[136]:
```
KP281    80
KP481    60
KP781    40
Name: Product, dtype: int64
```

```
Insights: The unique products having the number of product by the Customers
The items listed with codes KP281 (80), KP481 (60), and KP781 (40) exhibit a descending gradation of values. This suggests potential prioritization or significance, with KP281 holding the highest importance. However, the context behind these values is needed for a comprehensive understanding of their implications and decision-making basis.
```

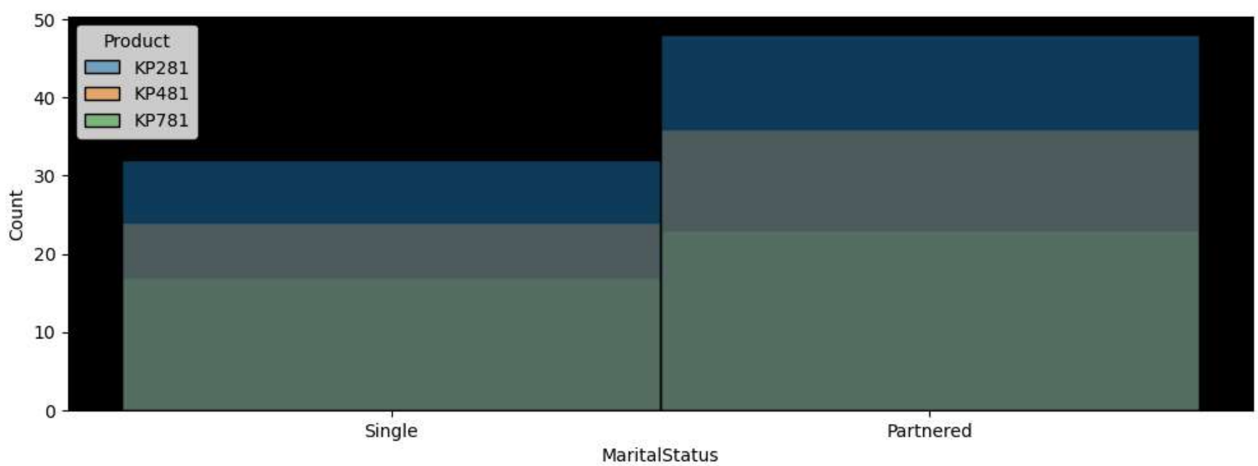
In [121]:
```
plt.figure(figsize=(12,5))
plt.subplot(1,3,1)
ax=plt.gca()
ax.set_facecolor('black')
sns.countplot(data=df, x='Product')
plt.subplot(1,3,2)
ax=plt.gca()
ax.set_facecolor('black')
sns.countplot(data=df, x='MaritalStatus')
plt.subplot(1,3,3)
ax=plt.gca()
ax.set_facecolor('black')
sns.countplot(data=df, x='Gender')

plt.show()
```




```
Insights: The list of the comparisons between the Product, Marital Status and Gender, where it can consider as KP281 has
the
highest selling category of mills which is the Entry level of the trademill and the maximum number os customers has choosen
that insights the regular trademill has the highest selling, although the Partnered people has the majority.
The male people considered as the highest taker customer for general purchase category
```

In [354]:
```
plt.figure(figsize=(12,4))
sns.histplot(data=df, x='MaritalStatus',hue='Product')
ax=plt.gca()
ax.set_facecolor('black')
plt.show()
```

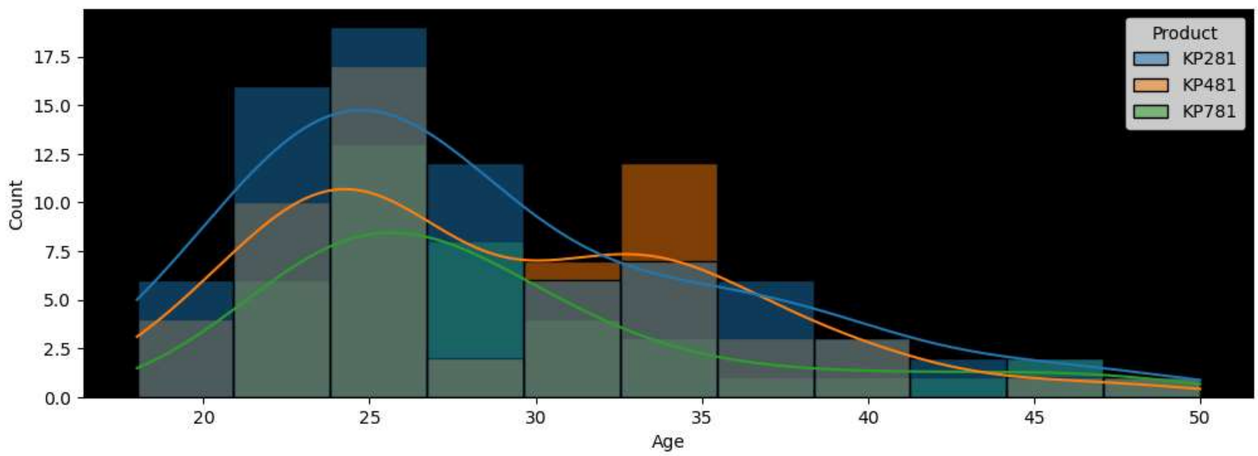



```
Insights: Product has been categorised between the Marital status, where the tendency shows the Partnered people use most
expensive trademill as compare to the singles, but the most selling product is KP781 for all the category and after that
the product called KP281 has the 2nd highest selling product all over in terms of Marital Status.
```

Business view: The product segmentation based on marital status reveals that partnered individuals prefer higher-end treadmills, while singles opt for less expensive options. Despite this trend, the top-selling product across all categories is KP781, indicating its universal appeal. Notably, KP281 holds the second-highest sales position, emphasizing its popularity across marital statuses. This insight guides the organization to tailor marketing strategies to both segments' preferences, potentially enhancing sales and customer satisfaction.

In [169]:

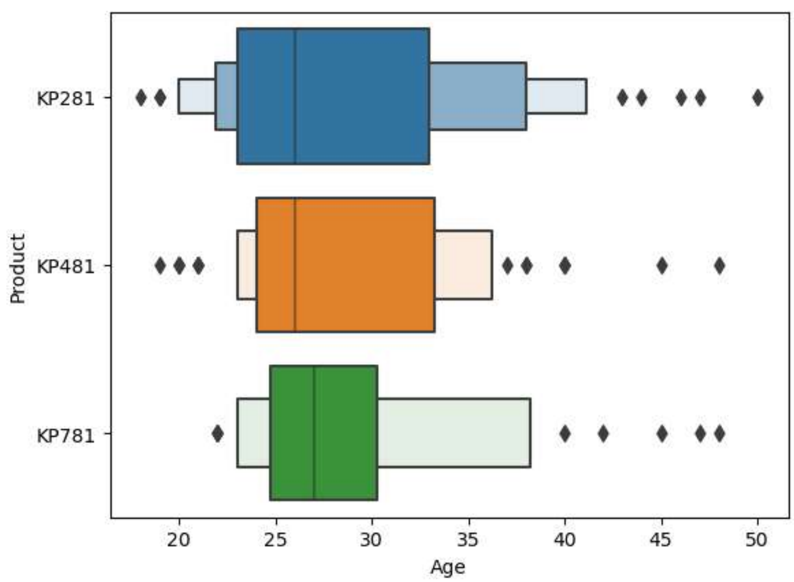
```
plt.figure(figsize=(12,4))
sns.histplot(data=df, x='Age',hue='Product',kde=True)
ax=plt.gca()
ax.set_facecolor('black')
plt.show()
```



Insight: In the compare to the age distsribution for the product wise the tendency of the product called KP281 which is the start range of the product has been sold more at the group of the people ramges from the 20 to 30 a wide range as compare to
the other category of the product. The tendency graph for the product has much have growth at the age limit 25 as least on the 50.
Cusotmers are more focused on the age 25 for buying trademill with the product KP781 which is the highest range category.

In [173]:

```
sns.boxenplot(data=df, x='Age', y='Product')
plt.show()
```



Insight: The largest sales of the product is KP281 all over wise, and also the people from the age more than 40 upto to the age 50, similary the least product has been sale for product is KP781 and max of the people are also comes into the outlier range.

```
Business view: Product KP281 dominates sales across all categories, indicating its strong market presence. Its popularity
is notably high among individuals aged 40 to 50. Conversely, KP781 has the lowest sales, with a majority falling into the
outlier range. This data-driven insight directs the organization to focus on optimizing KP781's appeal and potentially
exploring targeted campaigns to engage outliers, fostering balanced product distribution and increased revenue
opportunities.
```

In [215]:

```
df1=df[['Product','Gender','MaritalStatus']].melt()
df1.groupby(['variable', 'value'])[['value']].count()/len(df)*100
```

Out[215]:

variable	value	value
Gender	Female	42.222222
	Male	57.777778
MaritalStatus	Partnered	59.444444
	Single	40.555556
Product	KP281	44.444444
	KP481	33.333333
	KP781	22.222222

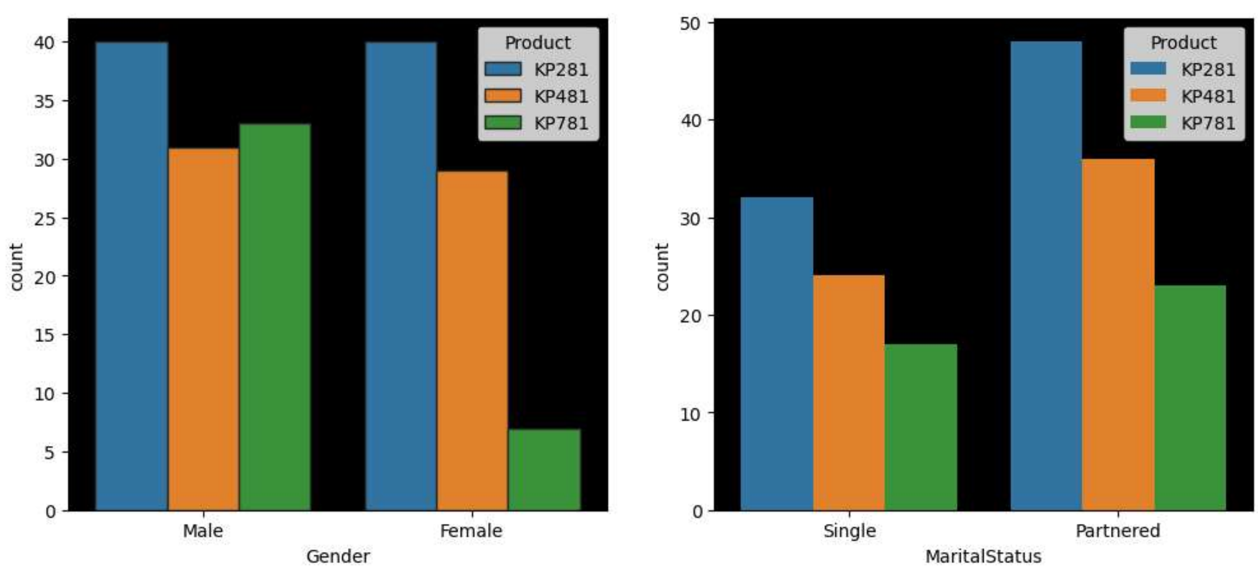
```
Insight: From Product: KP281 has 44% of customer taken product from all over
        From product KP481 has 33% of customers taken product from all over
        From Product KP781 has 22& of customers taken product from all over

Although, Female Purchases more compare to men with the difference of 15%
where as the form the MaritalStaus with the Engaged person considered more than Single.
```

Bi Variate Analysis

In [229]:

```
plt.figure(figsize=(12,5))
plt.subplot(1,2,1)
sns.countplot(data=df, x="Gender",hue='Product',edgecolor='0.15')
ax=plt.gca()
ax.set_facecolor('black')
plt.subplot(1,2,2)
sns.countplot(data=df,x='MaritalStatus', hue="Product")
ax=plt.gca()
ax.set_facecolor('black')
plt.show()
```



In [232]:
```
df.head()
```

Out[232]:

	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

Marginal Probability of Product Purchased

In [238]:
```
pd.crosstab(df.Product, len(df)).apply(lambda x: x/len(df)*100)
```

Out[238]:

col_0	180
Product	
KP281	44.444444
KP481	33.333333
KP781	22.222222

```
Here we have the Marginal Probability for the Product to be get Analysed, Where the newly Launched Product has the KP281
has the Maximum sales where as the Major with the hight range states which is KP781 has the least sales by the customers
```

In [242]:
```
df['Product'].value_counts(normalize=True)*100
```

Out[242]:
```
KP281    44.444444
KP481    33.333333
KP781    22.222222
Name: Product, dtype: float64
```

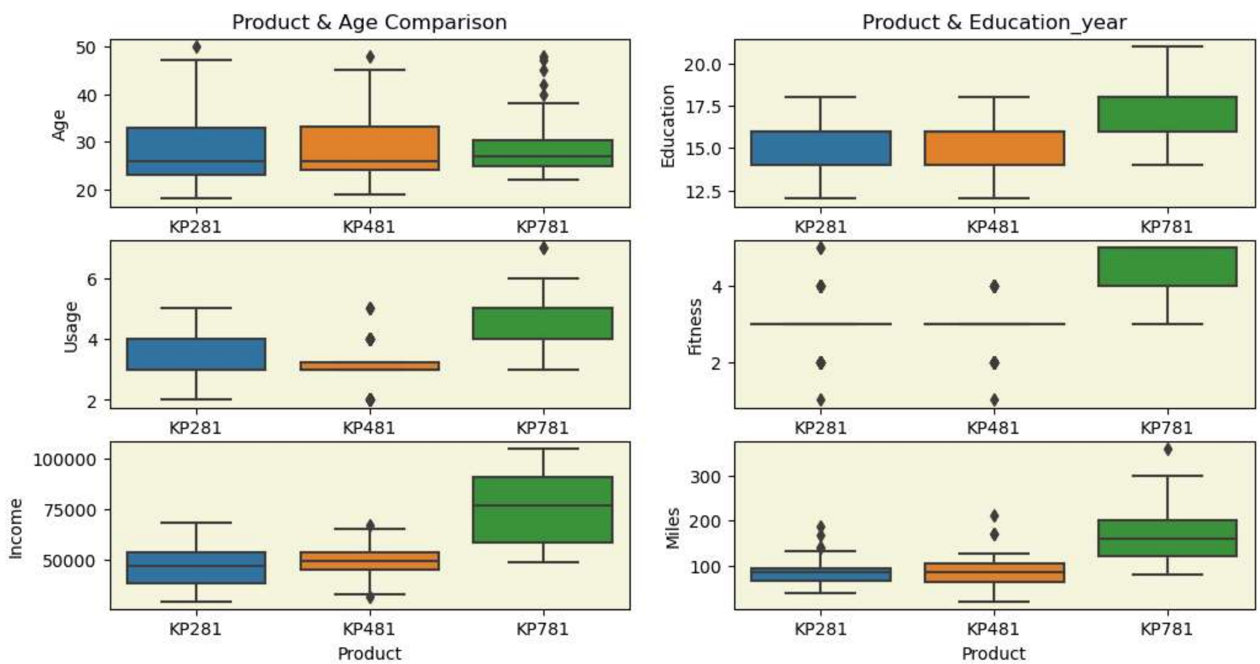
```
The Condition Probability merges with the same asoects of the morginal scene, the another way to find such a behvaiour.
The above stats describes same results for Product
```

Multi-Variate Corelation Among the different Factors

Product Analysis in terms of the all Factors Associated.

In [274]:
```
plt.figure(figsize=(12,6))
plt.subplot(3,2,1)
sns.boxplot(data=df, x='Product', y='Age')
ax=plt.gca()
ax.set_facecolor('beige')
plt.title('Product & Age Comparison')
plt.subplot(3,2,2)
ax=plt.gca()
ax.set_facecolor('beige')
sns.boxplot(data=df, x='Product', y='Education')
plt.title('Product & Education_year')
plt.subplot(3,2,3)
ax=plt.gca()
ax.set_facecolor('beige')
sns.boxplot(data=df, x='Product', y='Usage')
plt.subplot(3,2,4)
ax=plt.gca()
ax.set_facecolor('beige')
sns.boxplot(data=df, x='Product', y='Fitness')
plt.subplot(3,2,5)
ax=plt.gca()
ax.set_facecolor('beige')
sns.boxplot(data=df, x='Product', y='Income')
plt.subplot(3,2,6)
ax=plt.gca()
ax.set_facecolor('beige')
sns.boxplot(data=df, x='Product', y='Miles')

plt.show()
```




```
The analysis shown above is for the product with all segments to take into account and discover significance insight about
the client and their role and thinking for acquiring the units in accordance with the category for singles as well as
partners, with respect to age as well.
Insights: The Age consoderation for the customers mainly lies in the group of 23 to 30 with the less income ranges from the
below 50k with the single instances for the basic product KP281, also the higher educated people with the more than 15
years spots their research for the higher variety end product that is KP781.
With the major scene we can see the rating in the fitness has been observed more from the product KP781 as compare to the
other product. al thought he product code KP781 has also comes on the major role in the usage.
The miles has also been used more than the other product as the KP781 has more usage and runs on particular Product.
With the all varient the more Income customers tends to purchase the best product from the aerofit that is KP781
```

Conditional & Marginal Probability Analysis

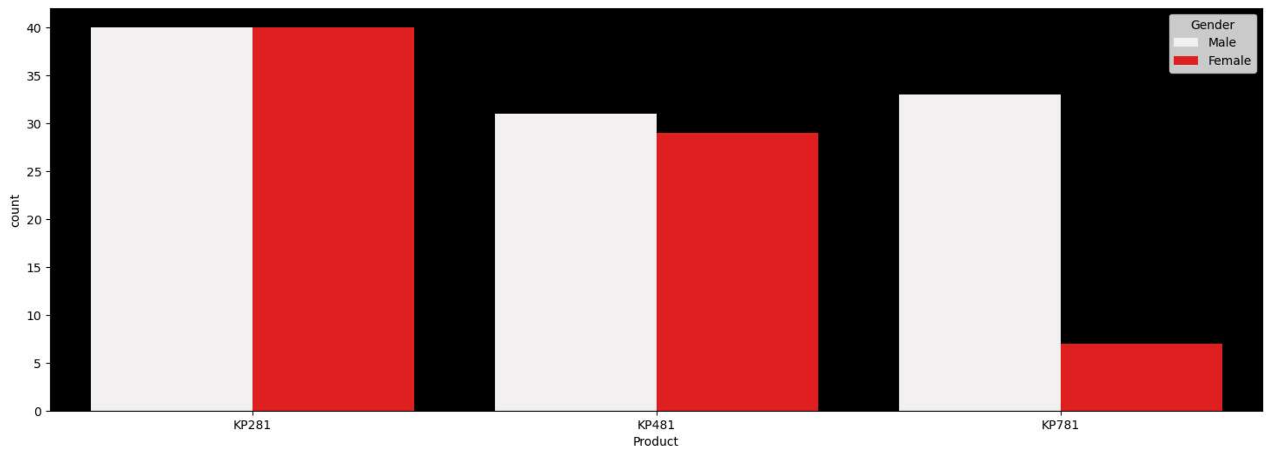
In [3]:
```
custom = pd.crosstab(df['Gender'],df['Product'])
custom
```

Out[3]:

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

```
Insights: The Product has been categorised with the gender wise followed by the product category.
    Here, The male perspective has more indulge in fitness as it tends to be more involvemnmt of male as compare to the
    female ones.
    The Intitial product which is been launched new to the market has been considered with inital silimar distribution
among
    the Male and Female
    Aerofit needs to be more precise and make indulge to the Male customers as compare to female customers.
Let's have Visual representation of the same.
```

In [348]:
```
plt.figure(figsize=(18,6))
sns.countplot(data=df, x='Product', hue='Gender',color="red")
ax=plt.gca()
ax.set_facecolor('black')
plt.show()
```




```
Similar Analysis for the Male and Female customers for the Product KP281 has comes to be equal distribution.
The Enrichess towards purchasing the Trademill consider to be slightly more than Female although when it compares to the
most advanced trademill, male customers move to higher end rather than to choosing the intital with the less feature.
```

In [355]:
```
pd.crosstab(df.Gender, len(df['Product'])).apply(lambda x: x/len(df['Product'])*100).round(2)
```

Out[355]:

col_0	180
Gender	
Female	42.22
Male	57.78

```
Insights: The distribution among the products and between the customer with the sub distribution between the Males and
    Female, where majority with the males having the approx 58% of males cusotmers falls and 43% approx of females customer
    falls.
```

In [337]:
```
pd.crosstab(df.Gender,df.Product).apply(lambda x: x/len(df['Product'])*100).round(2)
```

Out[337]:

Product	KP281	KP481	KP781
Gender			
Female	22.22	16.11	3.89
Male	22.22	17.22	18.33

```
Insights: For the Every product we can finalise the term between which of the product has been diversify with the male and
    female where we can see the and also the major poplulation lands for the product KP781 which male customers and ver
less
```

with the female customer around 4% who choose for the most advanced KP781 product.

In [327]:
```python
pd.crosstab(df.MaritalStatus, df.Product).apply(lambda x: x/len(df['Product'])*100).round(2)
```

Out[327]:

Product MaritalStatus	KP281	KP481	KP781
Partnered	26.67	20.00	12.78
Single	17.78	13.33	9.44

Insights: The table displays product preferences based on marital status. Among partnered individuals, KP281 is the most popular (26.67%), followed by KP481 (20.00%), and KP781 (12.78%). Conversely, for singles, KP281 maintains its lead (17.78%), trailed by KP481 (13.33%), and KP781 (9.44%). This data indicates that KP281 has a strong overall appeal, with partnered individuals showing higher preference across all products compared to singles.

In [333]:
```python
pd.crosstab(df.Age, df.Product).apply(lambda x: x/len(df['Product'])*100).round(2)[4:18]
```

Out[333]:

Product Age	KP281	KP481	KP781
22	2.22	0.00	1.67
23	4.44	3.89	1.67
24	2.78	1.67	2.22
25	3.89	6.11	3.89
26	3.89	1.67	1.11
27	1.67	0.56	1.67
28	3.33	0.00	1.67
29	1.67	0.56	1.11
30	1.11	1.11	1.67
31	1.11	1.67	0.56
32	1.11	1.11	0.00
33	1.11	2.78	0.56
34	1.11	1.67	0.56
35	1.67	2.22	0.56

Insights: The data illustrates product preferences across different ages. KP281 is most favored among the 23-26 age range, while KP481 gains popularity from 25-27 and 34-35. KP781 maintains consistent but moderate appeal across various age groups. Understanding these trends can inform targeted marketing strategies, optimizing product promotion for specific age segments and potentially boosting overall sales.

Marinal Condition on Basis of two categories

In [194]:
```python
length_gender = len(df['Gender'])
length_gender
df.groupby(by='Gender')['MaritalStatus'].value_counts()/length_gender*100
```

Out[194]:
```
Gender  MaritalStatus
Female  Partnered        25.555556
        Single           16.666667
Male    Partnered        33.888889
        Single           23.888889
Name: MaritalStatus, dtype: float64
```

Insights: "Analysis of relationship statuses among genders reveals that 33.89% of males are partnered, while 25.56% of females are in relationships. Single females account for 16.67% and single males for 23.89%. These insights offer companies opportunities to tailor marketing strategies, create gender-specific product offerings, and develop relationship-oriented services for optimal customer engagement."

Product Income Analysis with more than a mean value of Income || Conditional Probability

In [205]:
```
avg_income = df[df['Income']>df['Income'].mean()]
avg_income[['Product','Income']][:10]
```

Out[205]:

	Product	Income
42	KP281	54576
47	KP281	54576
48	KP281	54576
50	KP281	68220
54	KP281	54576
55	KP281	54576
59	KP281	55713
64	KP281	60261
65	KP281	67083
69	KP281	54576

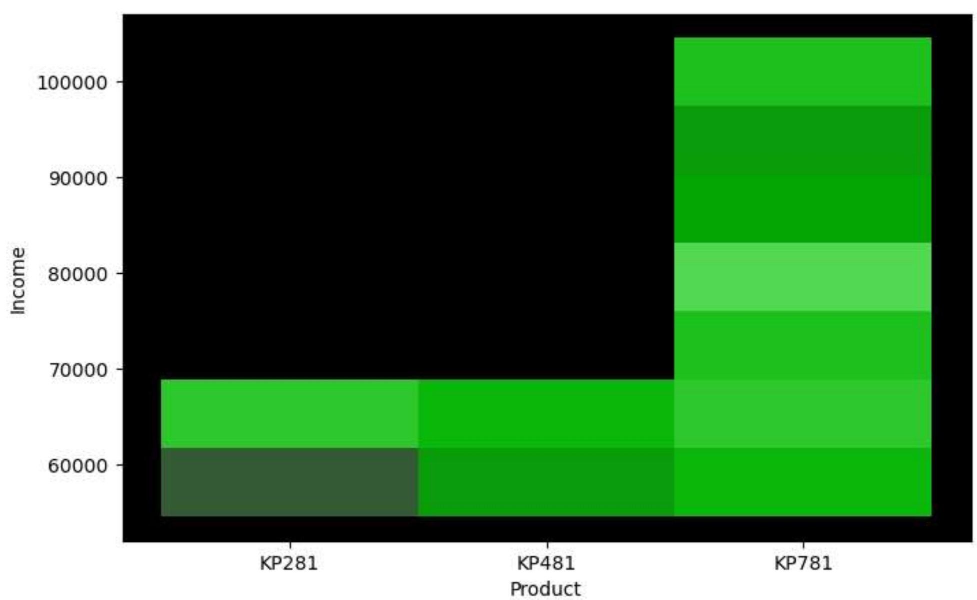
Insights: The product "KP281" appears to be consistently generating income, with multiple transactions recorded. The income values vary, with the highest recorded at 68,220. This suggests a stable demand for this product. Further analysis could reveal trends in customer preferences and potentially guide marketing strategies to capitalize on the product's popularity.

In [206]:
```
avg_income['Product'].value_counts(normalize=True)*100
```

Out[206]:
```
KP781    50.793651
KP281    28.571429
KP481    20.634921
Name: Product, dtype: float64
```

Among the listed products, "KP781" has the highest contribution to total income at 50.79%, indicating its significant revenue generation. "KP281" follows with 28.57% of the income, while "KP481" contributes 20.63%. This breakdown helps allocate resources effectively, potentially focusing more on "KP781" and "KP281" due to their larger shares in the overall income.

In [207]:
```
plt.figure(figsize=(8,5))
sns.histplot(data=avg_income, x='Product', y='Income', color="green")
ax=plt.gca()
ax.set_facecolor('black')
plt.savefig('pro_inc.png')
```



The Same insights has been drawn with the help of the graph to follow the visual tendency among the product and the income level from the customers who have more than the income of 50k, while the hard solid color signifies the detail collab with less tendency where the brightest color tends ton be give more high and advanced product to be purchased.

These allocation of the product has been classified with the customers who has more than 50k of its income all over the data income potentially where their vision to buy the product KP781 more precisely.

Product Income Analysis with less than a mean value of Income || Conditional Probability

In [7]:
```
below_avg_income = df[df['Income']<df['Income'].mean()]
below_avg_income[['Product','Income']][:10]
```

Out[7]:

	Product	Income
0	KP281	29562
1	KP281	31836
2	KP281	30699
3	KP281	32973
4	KP281	35247
5	KP281	32973
6	KP281	35247
7	KP281	32973
8	KP281	35247
9	KP281	37521

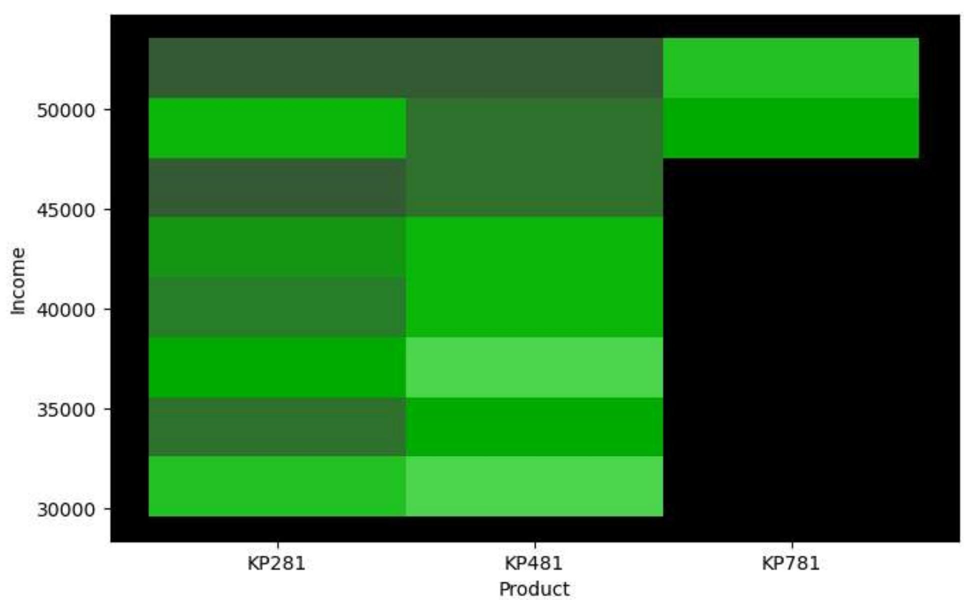
Insights: Repeatedly, product KP281 demonstrates strong income performance, consistently surpassing the 30,000 mark. This indicates a steady demand and likely customer loyalty. The alternating pattern of incremental increases might suggest a potential sales strategy, where periodic price adjustments could maximize profits. Further analysis into the factors influencing this consistent growth could offer valuable insights for sustaining and enhancing overall business profitability.

In [13]:
```
below_avg_income['Product'].value_counts(normalize=True)*100
```

Out[13]:
```
KP281    52.991453
KP481    40.170940
KP781     6.837607
Name: Product, dtype: float64
```

Insights: The table highlights varying sales distribution among products: KP281, KP481, and KP781. KP281 leads with 52.99%, indicating its popularity and robust demand. KP481 follows with 40.17%, suggesting a substantial market presence. Meanwhile, KP781 lags at 6.84%, indicating potential for improvement. A strategic focus on KP781's marketing and sales efforts could enhance its share, while maintaining KP281 and KP481's strengths would be prudent for overall business growth.

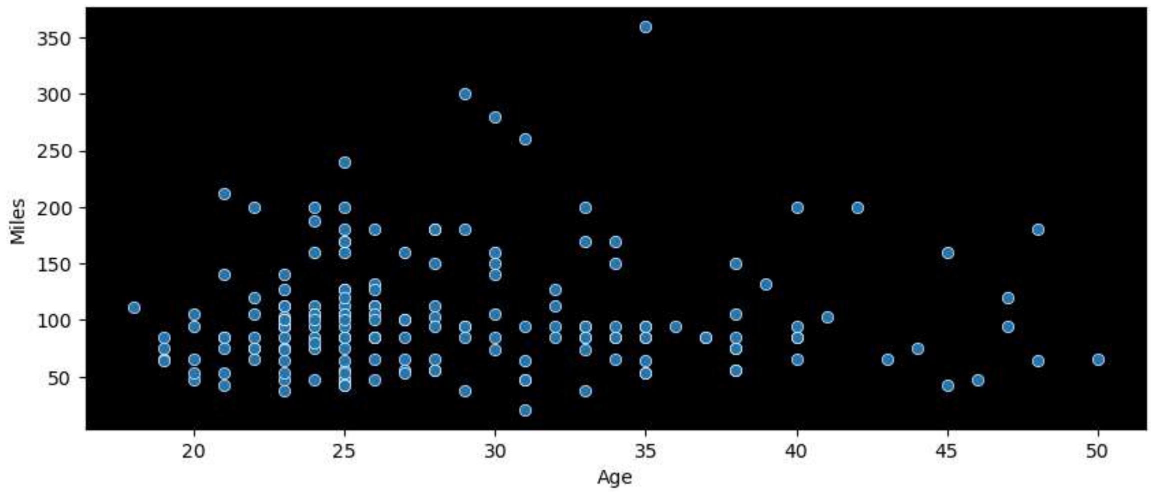
In [15]:
```
plt.figure(figsize=(8,5))
sns.histplot(data=below_avg_income, x='Product', y='Income', color="green")
ax=plt.gca()
ax.set_facecolor('black')
plt.show()
```



```
Insights: Utilizing a graphical representation, the analysis mirrors the previously drawn insights. The graph visually captures sales tendencies among products while correlating income levels below 50k of income. Bold, solid colors signify intricate correlations, with brighter shades indicating a higher affinity for advanced products. This classification aligns with customers' preferences for purchasing the KP281 product with heightened precision, presenting an opportunity for strategic targeting and tailored marketing strategies to further enhance its appeal among this affluent demographic. Signinfies the below mean average value people whos having less income towards all tends to but the KP281 the base product for their own.
```

In [68]:

```
plt.figure(figsize=(10,4))
sns.scatterplot(data=df, x='Age',y='Miles')
ax=plt.gca()
ax.set_facecolor('black')
plt.show()
```




```
Insights: The provided insights highlight a significant business opportunity within the age range of 22 to 27, where individuals exhibit heightened motivation for treadmill usage, resulting in increased weekly activity. Leveraging this trend can drive targeted marketing, product enhancements, and tailored experiences to engage and retain this demographic. Focusing resources on refining product features, personalized training plans, and social integration could amplify user satisfaction and brand loyalty, ultimately yielding higher sales and market share in the fitness industry.
```

Corelation Between All Segments

In [385]:

```
corelation = df.corr()
corelation
```

```
C:\Users\HP\AppData\Local\Temp\ipykernel_7224\1302924411.py:1: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value of numeric_only to silence this warning.
  corelation = df.corr()
```

Out[385]:

	Age	Education	Usage	Fitness	Income	Miles
Age	1.000000	0.280496	0.015064	0.061105	0.513414	0.036618
Education	0.280496	1.000000	0.395155	0.410581	0.625827	0.307284
Usage	0.015064	0.395155	1.000000	0.668606	0.519537	0.759130
Fitness	0.061105	0.410581	0.668606	1.000000	0.535005	0.785702
Income	0.513414	0.625827	0.519537	0.535005	1.000000	0.543473
Miles	0.036618	0.307284	0.759130	0.785702	0.543473	1.000000

Important objects needs to corelate

In [19]:
```
corelation = df.corr()
corelation_stats = corelation[['Income','Usage','Miles']]
corelation_stats
```

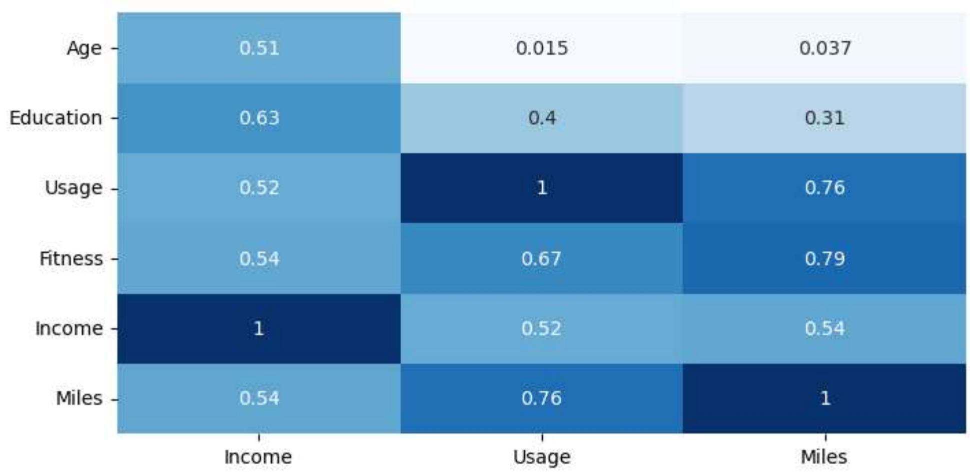
```
C:\Users\HP\AppData\Local\Temp\ipykernel_15344\1166044688.py:1: FutureWarning: The default value of numeric_only in Data
Frame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value
of numeric_only to silence this warning.
  corelation = df.corr()
```

Out[19]:

	Income	Usage	Miles
Age	0.513414	0.015064	0.036618
Education	0.625827	0.395155	0.307284
Usage	0.519537	1.000000	0.759130
Fitness	0.535005	0.668606	0.785702
Income	1.000000	0.519537	0.543473
Miles	0.543473	0.759130	1.000000

```
Insights: The correlation data suggests interesting insights for the product. Income and Education exhibit a moderate
positive correlation, implying that higher income individuals with better education may be more inclined to purchase the
product. Additionally, stronger positive correlations between Usage, Fitness, and Miles indicate that individuals who use
the product more frequently and engage in fitness activities tend to cover more miles. These findings can guide marketing
efforts towards targeting higher-income, educated segments and emphasizing product benefits for active lifestyles.
```

In [20]:
```
plt.figure(figsize=(8,4))
sns.heatmap(data=corelation_stats,cmap="Blues",annot=True,center=None,cbar=False)
plt.show()
```



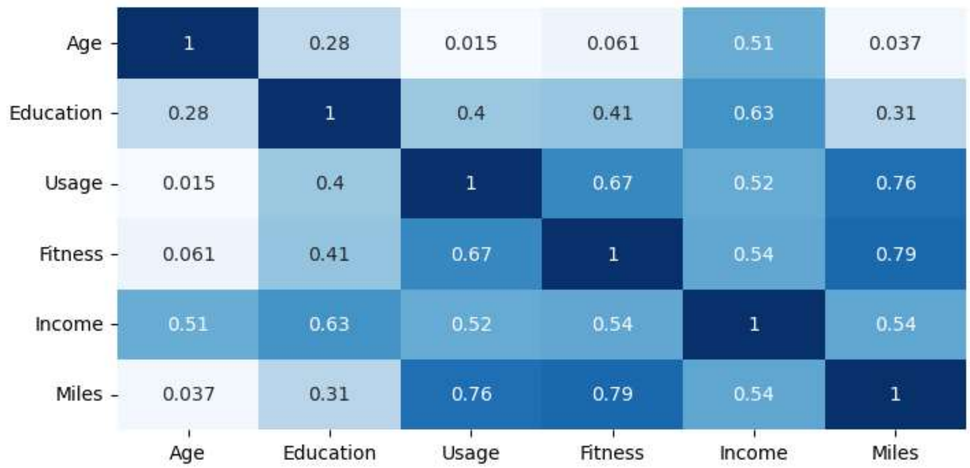

```
Insights: The correlation analysis offers intriguing insights for product strategy. Notably, there is a moderate positive
correlation between Income and Education, implying that individuals with higher income and better education levels show a
stronger inclination to purchase the product. Moreover, the robust positive correlations observed between Usage, Fitness,
and Miles indicate that frequent users who prioritize fitness tend to achieve greater distances. These revelations provide
valuable guidance for marketing endeavors, suggesting a focus on affluent, educated segments and highlighting the product's
advantages for active lifestyles.

Business view: The correlation analysis reveals key insights for product strategy. Higher income and education correlate
with increased product interest. Strong positive correlations between Usage, Fitness, and Miles emphasize active users'
engagement. Marketing should target educated, affluent segments and highlight benefits for active lifestyles, aligning with
company goals.
```

In [442]:

```
plt.figure(figsize=(8,4))
sns.heatmap(df.corr(),cmap="Blues",annot=True,center=None,cbar=False)
plt.show()
```

```
C:\Users\HP\AppData\Local\Temp\ipykernel_7224\3055780787.py:2: FutureWarning: The default value of numeric_only in DataF
rame.corr is deprecated. In a future version, it will default to False. Select only valid columns or specify the value o
f numeric_only to silence this warning.
  sns.heatmap(df.corr(),cmap="Blues",annot=True,center=None,cbar=False)
```



Insights: The above graph speaks the same coorelation as we talked about the previous descriptive corelation, where the Income, Education and with repsect to the usgae and the fitness will empower the more strong relationship between them which has inhabit this findings with the delivery of product into the main stream of the sellings

Business view: The graph reaffirms the earlier correlation. Income and education, coupled with usage and fitness, fortify the relationship. These factors underscore a robust connection, enhancing the product's integration into mainstream sales. This alignment highlights the importance of considering income, education, and consistent usage to drive stronger customer engagement and overall sales success.

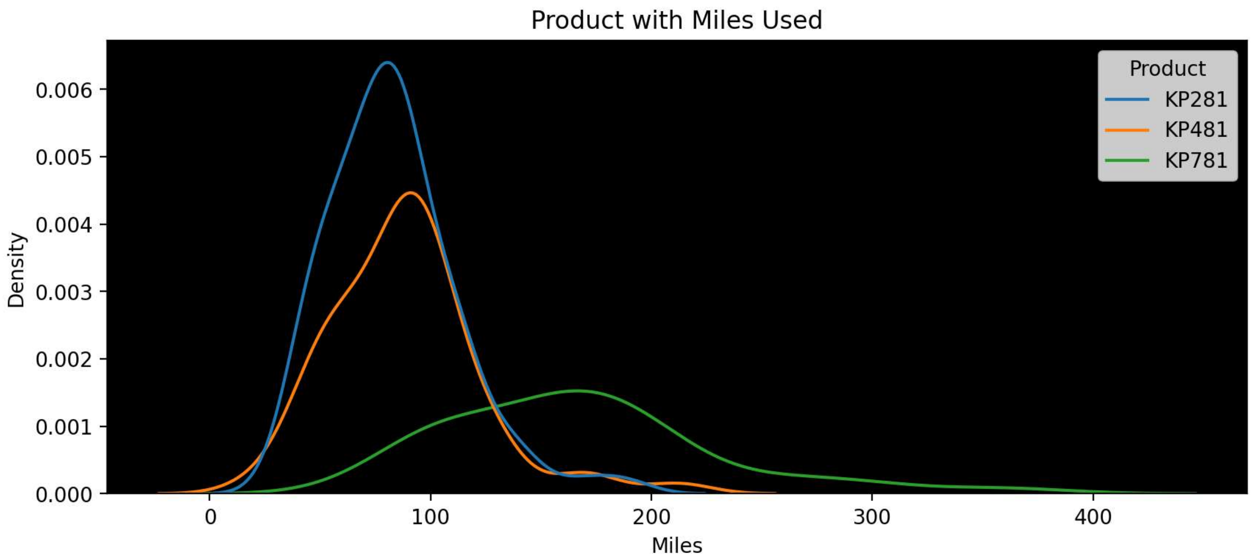
In [69]:

```
df.head()
```

Out[69]:

	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 [229]:
```
plt.figure(figsize=(10,4),dpi=200)
sns.kdeplot(data=df,x='Miles', hue='Product')
plt.title("Product with Miles Used")
ax=plt.gca()
ax.set_facecolor('black')
plt.show()
```



Insights: The acknowledge of the product is been used by the customers when the product over ruled with the customer with the maximum number of miles has been covereed with eaxch product which satisfies the needs of the customers in respect of the miles used product with the highest frequwncy level with the KP281 has the max number of the needs where the advanced features product which is KP781 has been least used in respect of the cusotmers purchased. Also the maximum miles has been used with the product KP781 whcih covers atmost more than 400 miles.
Likely product which is to be considered as KP781 among the all the product.

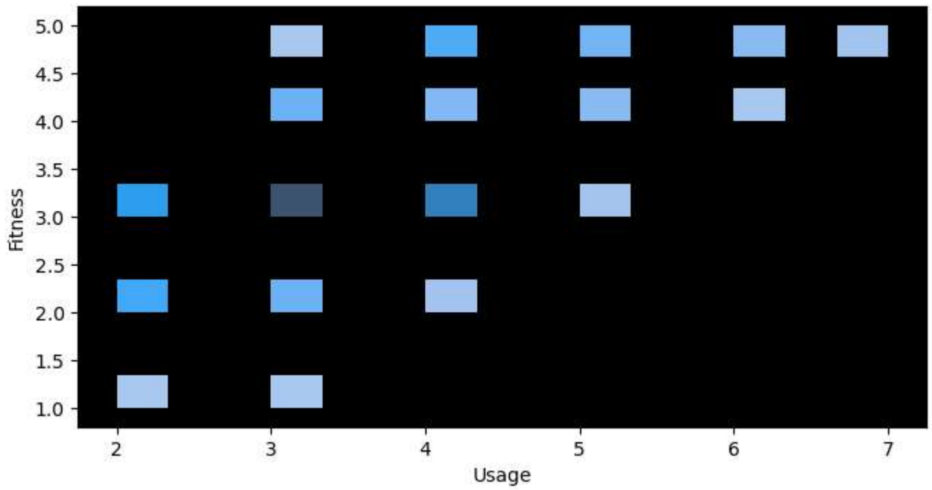
In [94]:
```
df.groupby(by='Usage')['Fitness'].value_counts(normalize=True)
```

Out[94]:
```
Usage  Fitness
2      3          0.545455
       2          0.424242
       1          0.030303
3      3          0.681159
       2          0.144928
       4          0.144928
       1          0.014493
       5          0.014493
4      3          0.576923
       5          0.250000
       4          0.134615
       2          0.038462
5      5          0.529412
       4          0.352941
       3          0.117647
6      5          0.857143
       4          0.142857
7      5          1.000000
Name: Fitness, dtype: float64
```

Insights: The table presents usage-fitness associations for different fitness levels. Fitness level 5 which is a rating scale is been given shows the highest adherence, indicating a strong commitment to exercise. Lower fitness levels demonstrate lesser dedication. Notably, fitness level 7 boasts a perfect adherence rate, suggesting a dedicated and consistent workout routine. The table underscores the positive correlation between higher fitness levels and consistent usage, emphasizing the importance of maintaining a structured fitness regimen for optimal results.

Business view: The table outlines usage-fitness links across levels. With level 5 displaying the strongest adherence, dedication to exercise is evident. Lower fitness tiers show less commitment. Remarkably, level 7 attains perfect adherence, reflecting unwavering routine. The data underscores higher fitness aligning with steady usage, highlighting the need for structured fitness for prime outcomes. A focus on consistent routines emerges as pivotal from a corporate standpoint.

In [91]:
```
plt.figure(figsize=(8,4))
sns.histplot(data=df, x='Usage', y='Fitness')
ax = plt.gca()
ax.set_facecolor('black')
plt.show()
```



Insights: The graph illustrates the relationship between usage and fitness levels. Fitness level 5 demonstrates the highest adherence across usage levels, indicating a strong commitment to exercise. Notably, fitness level 7 exhibits a perfect adherence rate. However, fitness level 1 shows consistently low usage, suggesting potential room for improvement in engagement strategies. Overall, the graph emphasizes the positive correlation between higher fitness levels and increased usage, highlighting the significance of fostering consistent workout habits for optimal business outcomes.

Business View: The graph underscores a direct connection between usage and fitness levels. Notably, top-tier fitness (level 5) reflects remarkable adherence, while level 7 boasts perfect adherence. On the flip side, fitness level 1 displays subpar usage, indicating room for better engagement approaches. The data underscores a crucial tie between higher fitness levels and usage, stressing the need for sustained workout habits to drive positive business outcomes.

In [386]:
```
sns.pairplot(data=df, hue='Product')
plt.show()
```


```
Insights: These number of pairplots describes the same behaviour as we used in the corelation above! where the more systematic
and natural behaviour between every objectives where the customer needs and follows with the product Analysis. The KDE plots describes the main stream for the strong hold when the category falls 1 to 1 which comes to be the equal distribution with them.

Business view: In essence, the pairplots corroborate our earlier findings and lend a holistic understanding of the coherent dynamics between customer requirements and product analysis. This knowledge can guide strategic decision-making, enabling us to better tailor our offerings to meet customer demands and optimize product strategies.
```