Business Case: Aerofit - Descriptive Statistics & Probability (Sabyasachi Banerjee)

Aerofit

About

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.

In [ ]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings("ignore")
```

In [ ]:
```
!wget "https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/001/125/original/aerofit_treadmill.csv?1639992749" -O aerofit.csv
```

```
--2024-07-19 14:02:13--  https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/001/125/original/aerofit_treadmill.csv?1639992749
Resolving d2beiqkhq929f0.cloudfront.net (d2beiqkhq929f0.cloudfront.net)... 3.162.130.14, 3.162.130.111, 3.162.130.189, ...
Connecting to d2beiqkhq929f0.cloudfront.net (d2beiqkhq929f0.cloudfront.net)|3.162.130.14|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 7279 (7.1K) [text/plain]
Saving to: 'aerofit.csv'

aerofit.csv         100%[===================>]   7.11K  --.-KB/s    in 0s

2024-07-19 14:02:13 (1.71 GB/s) - 'aerofit.csv' saved [7279/7279]
```

In [ ]:
```
data = pd.read_csv("aerofit.csv")
```

📃 Dataset Details

- Product Purchased: KP281, KP481, or KP781
- Age: In years
- Gender: Male/Female
- Education: In years
- MaritalStatus: Single or partnered
- Usage: The average number of times the customer plans to use the treadmill each week.

- Income: Annual income (in $)
- Fitness: Self-rated fitness on a 1-to-5 scale, where 1 is the poor shape and 5 is the excellent shape.
- Miles: The average number of miles the customer expects to walk/run each week

📚 Exploratory Data Analysis

In [ ]:
```
data.head()
```

Out[ ]:

	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 [ ]:
```
data.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
```

In [ ]:
```
data.shape
```

Out[ ]: `(180, 9)`

In [ ]:
```
data.duplicated().sum()
```

Out[ ]: `0`

🔍 Insights

- The dataset consists of 9 Columns and no missing values
- There are no duplicate entries in the dataset

📊 Statistical Analysis

- Descriptive Analysis

In [ ]:
```
data.describe()
```

Out[ ]:

	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

🔍 Insights

- Total count of all columns is 180.
- Age: The mean age of customers is 28 years, with the median age at 26.
- Education: The mean education level is 15, with a maximum of 21 and a minimum of 12.
- Usage: The mean weekly usage is 3.4 times, with a maximum of 7 and a minimum of 2.
- Fitness: The average fitness rating is 3.3 on a scale of 1 to 5.
- Miles: Customers walk an average of 103 miles, with the maximum distance being almost 115 miles and the minimum 21 miles.

- Income: Most customers earn around 58K annually, with a maximum of 104K and a minimum of approximately $30K.

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

Out[ ]:

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

🔍 Insights

- Product - Over the past three months, the KP281 product led sales performance, comprising roughly 44% of total sales.
- Gender - In the last three months, approximately 58% of buyers were male, and 42% were female.
- Marital Status - In the past three months, about 60% of buyers were married, while 40% were single.

In [ ]:
```
skip = ["Age", "Income","Miles"]

for i in data.columns:
  print(f" Unique value of -> {i} ")
  print(data[i].unique())
  if i in skip:
    pass
  else:
    print("*"* 50)
    print(f"Value Count of -> {i}")
    print(data[i].value_counts().sort_values( ascending = False))
  print("-"* 50)
```

```
 Unique value of -> Product
['KP281' 'KP481' 'KP781']
**************************************************
Value Count of -> Product
Product
KP281    80
KP481    60
KP781    40
Name: count, dtype: int64
--------------------------------------------------
 Unique value of -> Age
[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]
--------------------------------------------------
 Unique value of -> Gender
['Male' 'Female']
**************************************************
Value Count of -> Gender
Gender
Male      104
Female     76
Name: count, dtype: int64
--------------------------------------------------
 Unique value of -> Education
[14 15 12 13 16 18 20 21]
**************************************************
Value Count of -> Education
Education
16    85
14    55
18    23
15     5
13     5
12     3
21     3
20     1
Name: count, dtype: int64
--------------------------------------------------
 Unique value of -> MaritalStatus
['Single' 'Partnered']
**************************************************
Value Count of -> MaritalStatus
MaritalStatus
Partnered    107
Single        73
Name: count, dtype: int64
--------------------------------------------------
 Unique value of -> Usage
[3 2 4 5 6 7]
**************************************************
Value Count of -> Usage
Usage
3    69
4    52
2    33
5    17
6     7
7     2
Name: count, dtype: int64
--------------------------------------------------
 Unique value of -> Fitness
[4 3 2 1 5]
**************************************************
Value Count of -> Fitness
Fitness
3    97
5    31
2    26
4    24
1     2
Name: count, dtype: int64
--------------------------------------------------
 Unique value of -> Income
[ 29562  31836  30699  32973  35247  37521  36384  38658  40932  34110
  39795  42069  44343  45480  46617  48891  53439  43206  52302  51165
  50028  54576  68220  55713  60261  67083  56850  59124  61398  57987
  64809  47754  65220  62535  48658  54781  48556  58516  53536  61006
  57271  52291  49801  62251  64741  70966  75946  74701  69721  83416
  88396  90886  92131  77191  52290  85906 103336  99601  89641  95866
 104581  95508]
--------------------------------------------------
 Unique value of -> Miles
[112  75  66  85  47 141 103  94 113  38 188  56 132 169  64  53 106  95
 212  42 127  74 170  21 120 200 140 100  80 160 180 240 150 300 280 260
 360]
--------------------------------------------------
```

🔍 Insights

- The three different treadmill products are KP281, KP481, and KP781.
- The most commonly purchased treadmill product type is KP281.
- There are 32 unique ages among customers.

- The customer list includes 104 males and 76 females.
- There are 8 unique education levels (14, 15, 12, 13, 16, 18, 20, 21).
- The highest fitness rating is 3.
- Most customers use the treadmill at least 3 days per week.

The majority of customers are married or partnered.

Creating New Columns

Age Group Column

In [ ]:
```
age_bin = [17,29,45,data["Age"].max()]
age_label = ["Adult","Middle_Age", "Old"]
data["Age_group"] = pd.cut(data["Age"], bins = age_bin , labels = age_label)
```

- Bucketing **Age** column in 3 categorise
- Adult : 18 - 29
- Middle_age : 30 - 45
- Old : 46 and above

Fitness Level Group Column

In [ ]:
```
fitness_bin = [0,2,3,5]
fitness_label = ["low_fit","mid_fit","high_fit"]
data["Fitness_group"] = pd.cut(data["Fitness"], bins = fitness_bin , labels = fitness_label)
```

- Bucketing **Fitness** column in 3 categorise
- low_fit : 1 - 2
- mid_fit : 3
- high_fit : 4 - 5

Income Level Group Column

In [ ]:
```
income_bin = [0,50000,70000,data["Income"].max()]
income_label = ["Moderate_income","High_income","Veryhigh_income"]
data["Income_group"] = pd.cut(data["Income"], bins = income_bin , labels = income_label)
```

- Bucketing **Income** column in 3 categorise
- Moderate_income : <= 50000
- High_income : 50001 - 70000
- Veryhigh_income : Above 70000

Miles walk Level Group Column

In [ ]:
```
miles_bin = [0,50,100,200,data["Miles"].max()]
miles_label = ["Light_activity","Moderate_activity","High_activity","Veryhigh_activity"]
data["Miles_group"] = pd.cut(data["Miles"], bins = miles_bin , labels = miles_label)
```

- Bucketing **Miles** column in 4 categorises
- Light Activity - Upto 50 miles
- Moderate Activity - 51 to 100 miles
- High Activity - 101 to 200 miles
- Veryhigh Activity - Above 200 miles

In [ ]:
```
data.sample(5)
```

Out[ ]:

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles	Age_group	Fitness_group	Income_group	Miles_group
113	KP481	30	Female	14	Single	3	3	57987	74	Middle_Age	mid_fit	High_income	Moderate_activity
144	KP781	23	Female	18	Single	5	4	53536	100	Adult	high_fit	High_income	Moderate_activity
141	KP781	22	Male	16	Single	3	5	54781	120	Adult	high_fit	High_income	High_activity
48	KP281	28	Male	14	Single	4	3	54576	113	Adult	mid_fit	High_income	High_activity
95	KP481	24	Male	14	Single	3	4	48891	106	Adult	high_fit	Moderate_income	High_activity

Product Portfolio:

- The KP281 is an entry-level treadmill that sells for 1500.

- The KP481 is for mid-level runners that sell for 1750.
- The KP781 treadmill is having advanced features that sell for 2500.

In [ ]:
```
data["Product"].value_counts()
```

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

Adding Products price to a new Column

In [ ]:
```
def product_price(x):
  if x == "KP281":
    return 1500
  elif x == "KP481" :
    return 1750
  else :
    return 2500
```

In [ ]:
```
data["Product_price"]  = data["Product"].apply(lambda x: product_price(x))
```

In [ ]:
```
product_sale = data.groupby("Product")["Product_price"].sum().reset_index()
product_sale
```

Out[ ]:

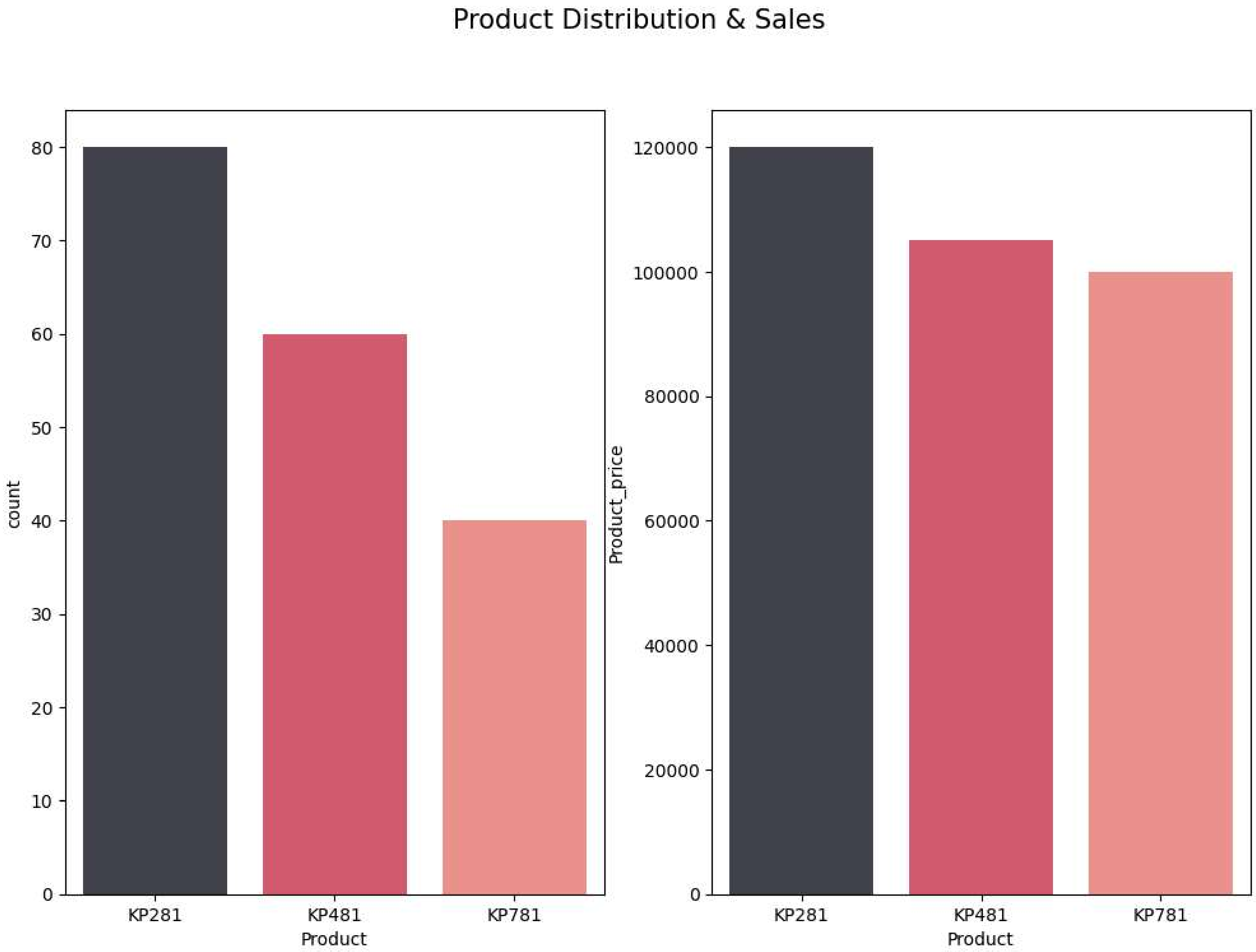
	Product	Product_price
0	KP281	120000
1	KP481	105000
2	KP781	100000

Visual Analysis - Univariate and Bivariate Analysis

Product Distribution

In [ ]:
```
palette = ["#3F3F4E", "#E84A5F", "#FF847C", "#FECEA8", "#99B898"]
sns.set_palette(palette)
```

In [ ]:
```
plt.figure(figsize = (12,8))
plt.subplot(1,2,1)
sns.countplot(x = "Product", data = data  , hue = "Product" , legend = False)
plt.subplot(1,2,2)
sns.barplot(y = "Product_price", x = "Product",data = product_sale , hue = "Product", legend = False)
plt.suptitle("Product Distribution & Sales" , fontsize = (15))
plt.show()
```



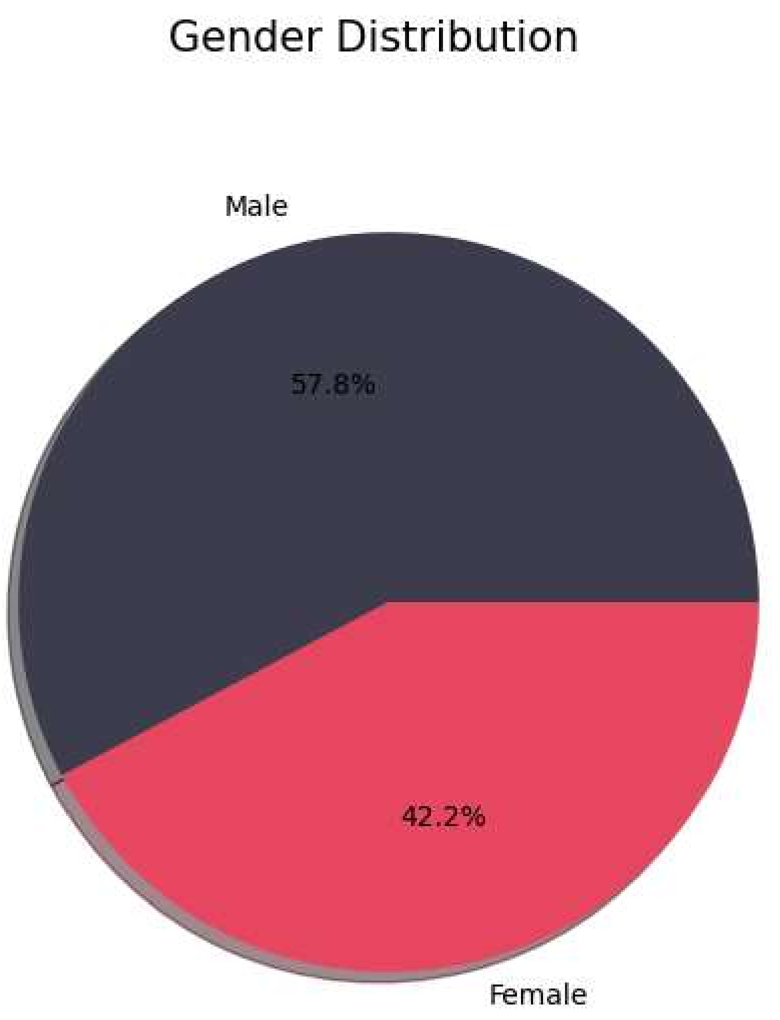
🔍 **Insights**

- The KP281 treadmill, positioned as an entry-level model, leads in unit sales, followed by the mid-level KP481 and the advanced KP781 models.
- Despite different target markets, all three treadmill models contribute equally to overall revenue generation, indicating balanced sales performance across product tiers.
- Among the treadmill models, KP281 emerges as the most frequently purchased, while KP481 follows as the second most popular choice, with KP781 being the least preferred option.

Gender Distribution

In [ ]:
```
plt.figure(figsize = (6,6))

labels = data["Gender"].value_counts().index
values = data["Gender"].value_counts().values
plt.pie(values, labels = labels , autopct = "%1.1f%%", shadow = True , colors = palette[0:2])
plt.suptitle("Gender Distribution" , fontsize = (15))
plt.show()
```

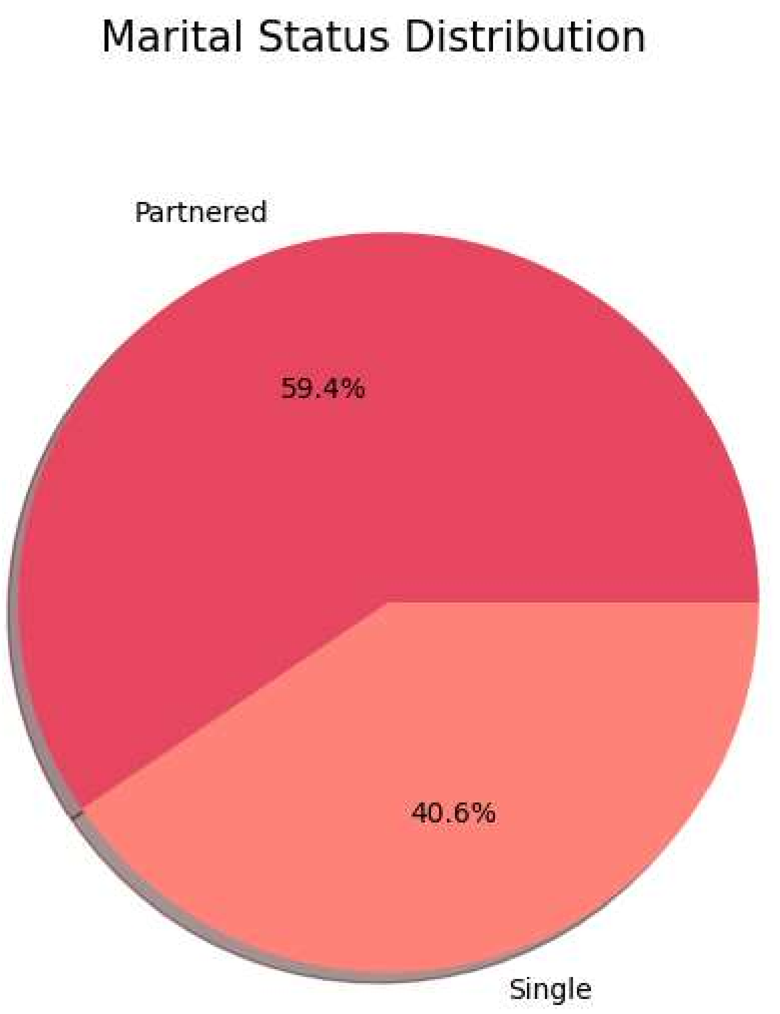


- There is a higher preference among males for purchasing the products compared to females

Marital Status Disribution

In [ ]:
```
plt.figure(figsize = (6,6))

labels = data["MaritalStatus"].value_counts().index
values = data["MaritalStatus"].value_counts().values
plt.pie(values, labels = labels , autopct = "%1.1f%%", shadow = True , colors = palette[1:3])
plt.suptitle("Marital Status Distribution" , fontsize = (15))
plt.show()
```

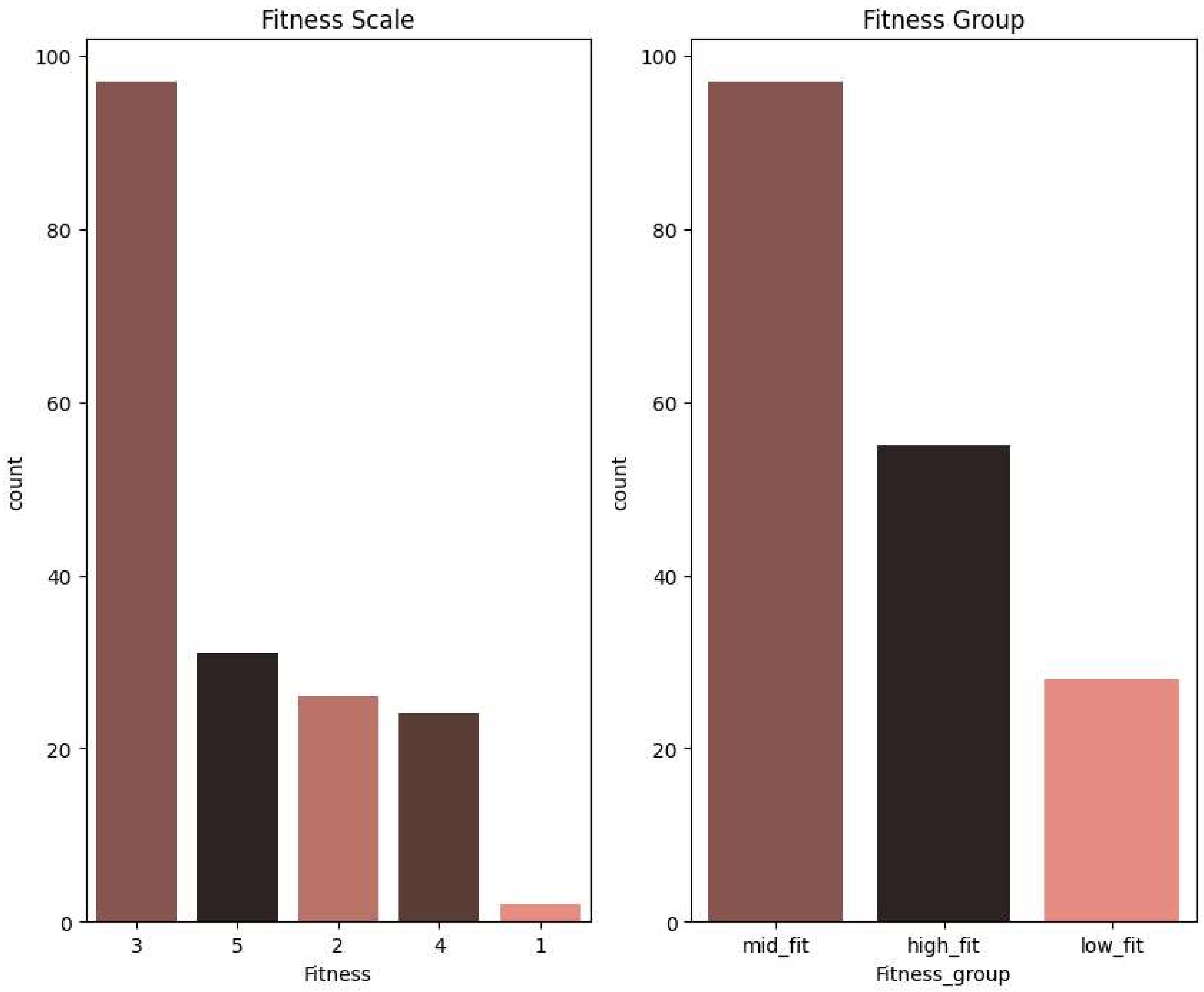


Fitness Distribution

In [ ]:
```
round(data["Fitness"].value_counts(normalize = True)*100,2)
```

Out[ ]:
```
Fitness
3    53.89
5    17.22
2    14.44
4    13.33
1     1.11
Name: proportion, dtype: float64
```

In [ ]:
```
plt.figure(figsize = (10,8))
plt.subplot(1,2,1)
sns.countplot(x = "Fitness", data = data, hue = "Fitness" , order = data["Fitness"].value_counts().index, legend = False , palette = "dark:
plt.title("Fitness Scale")
plt.subplot(1,2,2)
sns.countplot(x = "Fitness_group", data = data , hue = "Fitness_group", order = data["Fitness_group"].value_counts().index, legend = False
plt.title("Fitness Group")
plt.show()
```



🔍 **Insights**

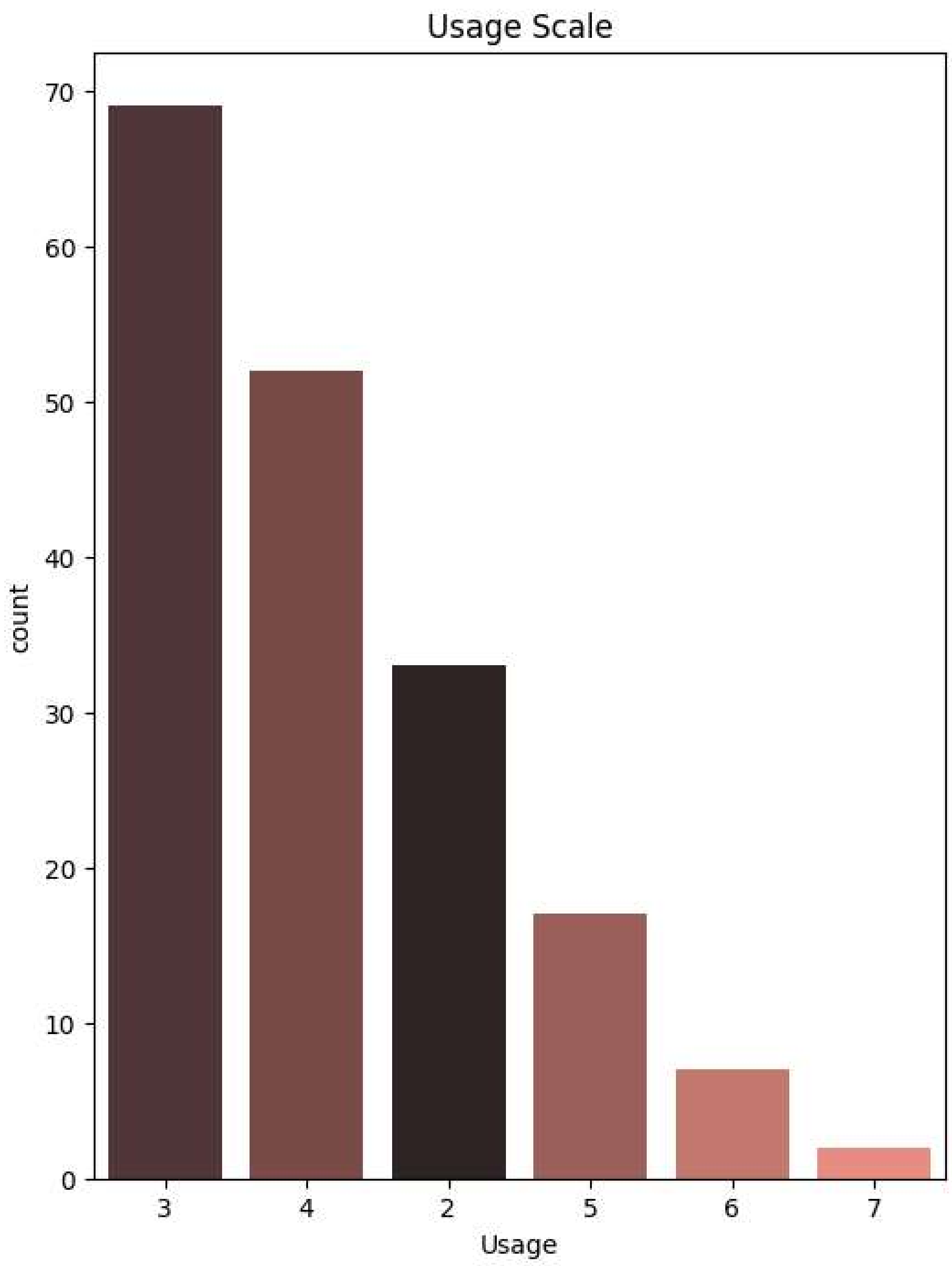
- Distribution Insight Across Fitness Levels:
- The majority of participants (68.33%) fall within the "mid_fit" category (fitness level 3), indicating a significant portion of the surveyed population maintains a moderate level of fitness.
- This reinforces that "mid_fit" (53.89% initially + 14.44% from fitness levels 2) is the predominant fitness group, central to the overall distribution.
- Higher Fitness Levels Concentration:
- Approximately 31.67% of participants are in the "high_fit" category (fitness levels 4 and 5), highlighting a notable segment of the population that maintains a high level of fitness.
- This distribution underscores the presence of a substantial number of individuals committed to maintaining a higher fitness regimen within the surveyed group.

Trademill Usage Distribution

In [ ]:
```
round(data["Usage"].value_counts(normalize = True)*100,2)
```

Out[ ]:
```
Usage
3    38.33
4    28.89
2    18.33
5     9.44
6     3.89
7     1.11
Name: proportion, dtype: float64
```

In [ ]:
```
plt.figure(figsize = (6,8))
sns.countplot(x = "Usage", data = data, hue = "Usage" , legend = False , palette = "dark:salmon",order = data["Usage"].value_counts().index
plt.title("Usage Scale")
plt.show()
```



🔍 **Insights**

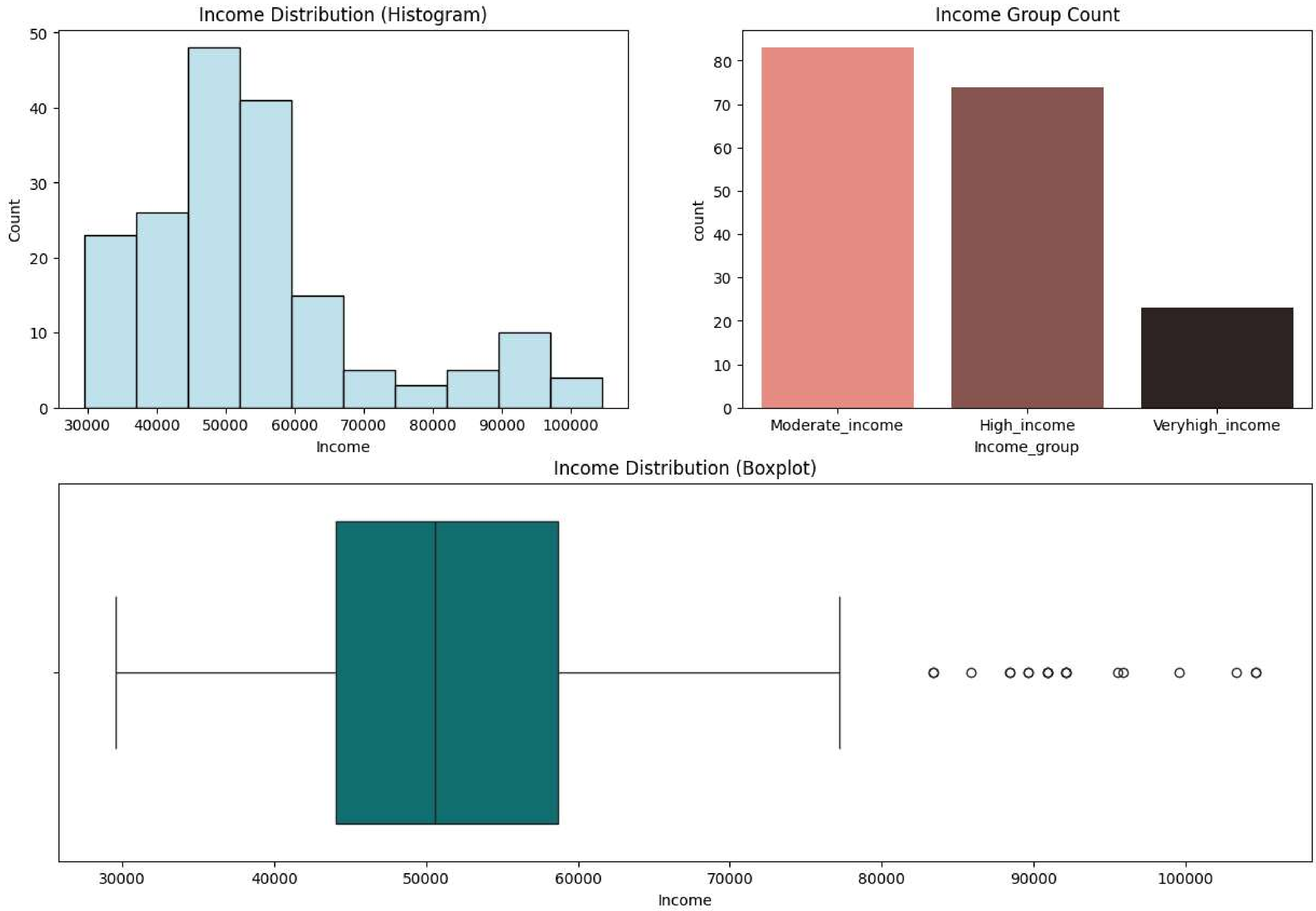
Nearly 85% of customers intend to use the treadmill between 2 to 4 times per week, while only 15% plan to use it 5 times or more weekly.

Income Distribution

In [ ]:
```
round(data["Income_group"].value_counts(normalize = True)*100,2)
```

Out[ ]:
```
Income_group
Moderate_income    46.11
High_income        41.11
Veryhigh_income    12.78
Name: proportion, dtype: float64
```

In [ ]:
```
plt.figure(figsize = (15,10))
plt.subplot(2,2,1)
sns.histplot(x = "Income", bins = 10 , data = data , kde = False, edgecolor = "black", color = "lightblue")
plt.title("Income Distribution (Histogram)")
plt.subplot(2,2,2)
sns.countplot(x = "Income_group", data = data , palette = "dark:salmon_r", hue = "Income_group", legend = False , order = data["Income_grou
plt.title("Income Group Count")
plt.subplot(2,1,2)
sns.boxplot(x = "Income", data = data,color = "teal")
plt.title("Income Distribution (Boxplot)")
plt.show()
```



🔍 Insights

- Income Distribution Insight:
- The majority of customers fall into the "Moderate_income" category, comprising 46.11% of the total. This suggests that a significant portion of the customer base has a moderate income level.
- High and Very High Income Groups:
- Combined, the "High_income" and "Veryhigh_income" groups account for 53.89% of the total, indicating that a considerable majority of customers have above-average income levels.

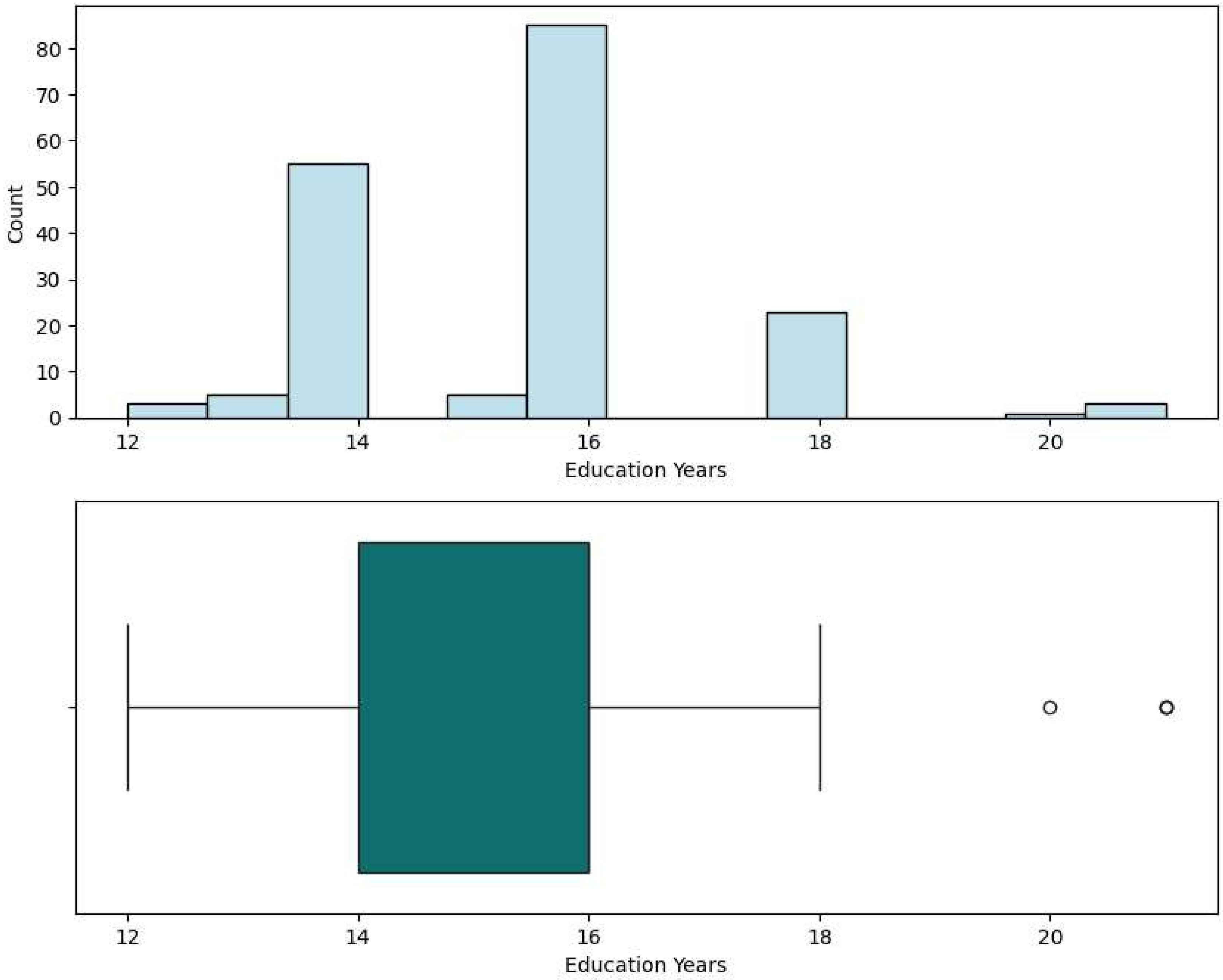
Education Distribution

In [ ]:
```
round(data["Education"].value_counts(normalize = True)*100,2)
```

Out[ ]:
```
Education
16    47.22
14    30.56
18    12.78
15     2.78
13     2.78
12     1.67
21     1.67
20     0.56
Name: proportion, dtype: float64
```

In [ ]:
```
plt.figure(figsize= (10,8))
plt.subplot(2,1,1)
sns.histplot(x = "Education", data = data , color = "lightblue", edgecolor = "black", kde = False)
plt.xlabel("Education Years")
plt.subplot(2,1,2)
sns.boxplot(x = "Education", data = data , color = "teal")
plt.suptitle("Education Level Distribution")
plt.xlabel("Education Years")
plt.show()
```

Education Level Distribution



🔍 **Insights**

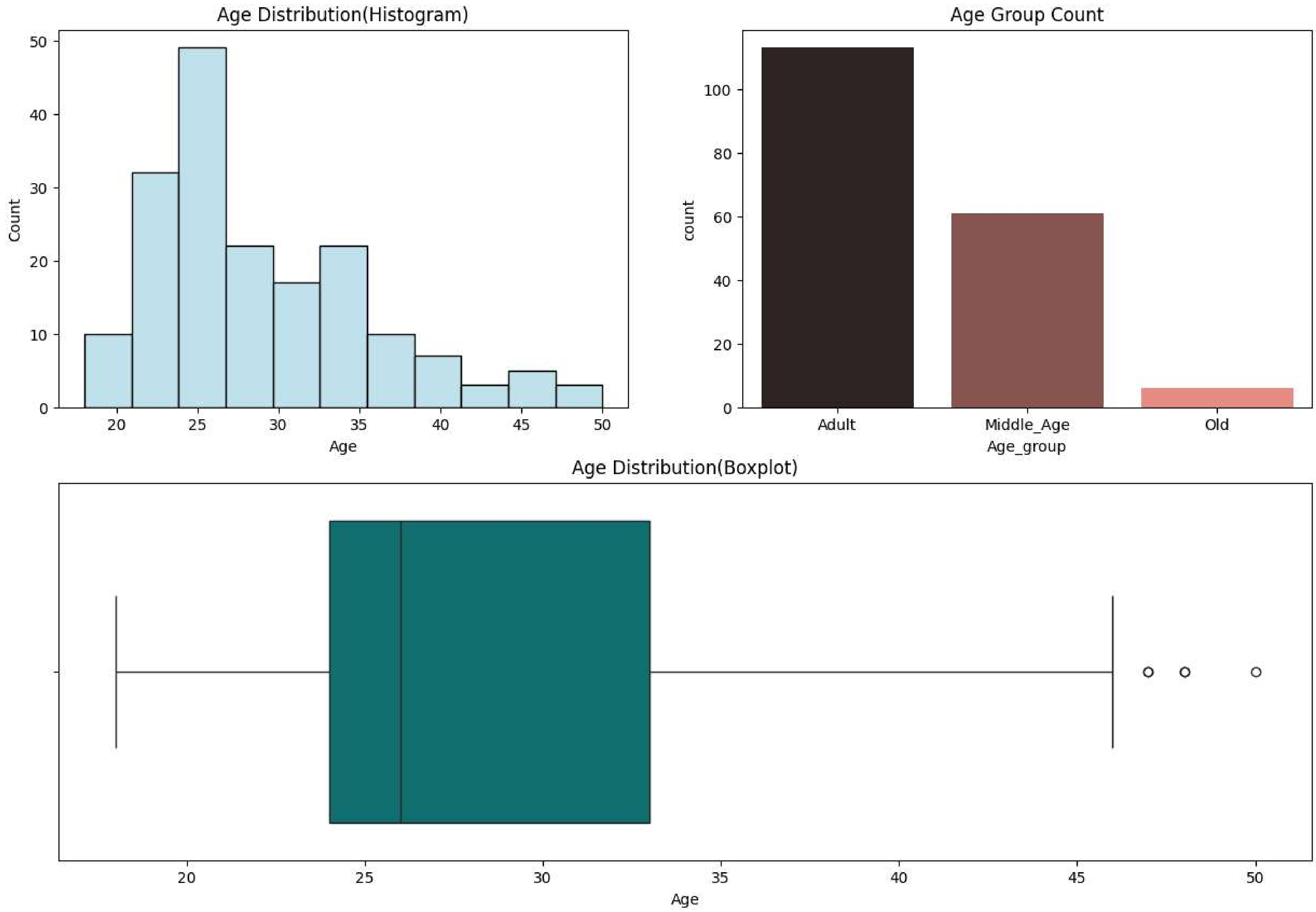
- A significant majority of customers, approximately 91.67%, have education levels of 14 years or higher. This high proportion suggests a strong correlation between higher education and the purchase of aerofit items, possibly driven by increased health awareness and disposable income.

Age Distribution

In [ ]:
```
round(data["Age_group"].value_counts(normalize = True)*100,2)
```

Out[ ]:
```
Age_group
Adult         62.78
Middle_Age    33.89
Old            3.33
Name: proportion, dtype: float64
```

In [ ]:
```
plt.figure(figsize =(15,10))
plt.subplot(2,2,1)
sns.histplot(x = "Age", data = data , edgecolor = "black", kde = False , color = "lightblue")
plt.title("Age Distribution(Histogram)")
plt.subplot(2,2,2)
sns.countplot(x = "Age_group",data = data , palette = "dark:salmon", hue = "Age_group", legend = False , order = data["Age_group"].value_co
plt.title("Age Group Count")
plt.subplot(2,1,2)
sns.boxplot(x = "Age", data = data , color = "teal")
plt.title("Age Distribution(Boxplot)")
plt.show()
```



🔍 Insights:

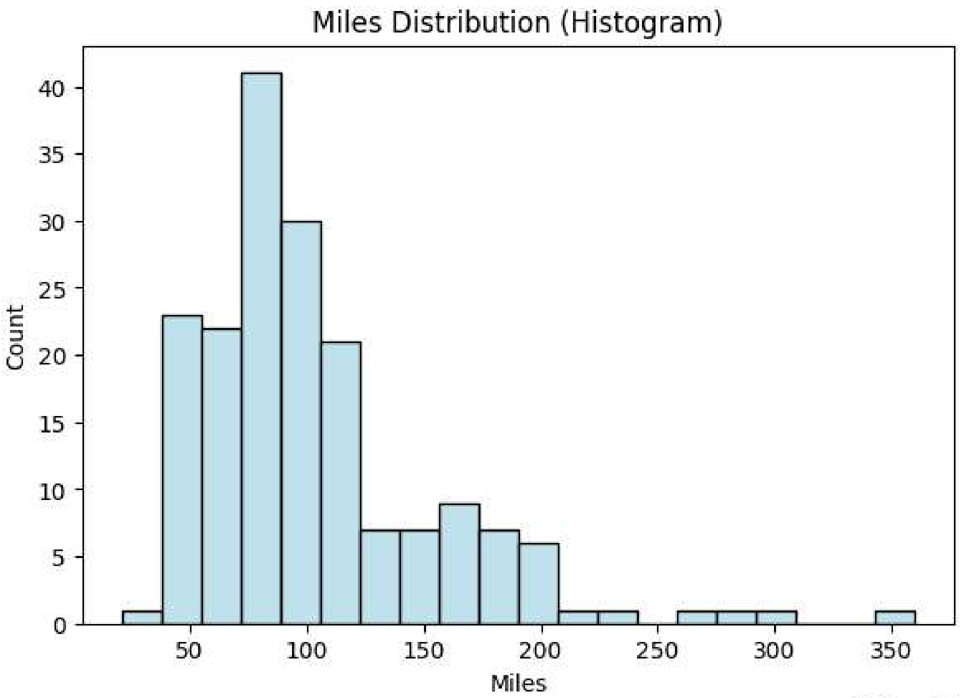
- The majority of customers (96.67%) fall into the "Adult" and "Middle_age" categories, indicating a strong preference among adults and middle-aged individuals for the company's products.
- Specifically, 62.78% of customers are categorized as "Adult" (ages 18-29), and 33.89% fall into the "Middle_age" group (ages 30-45), highlighting these demographics as key target markets.

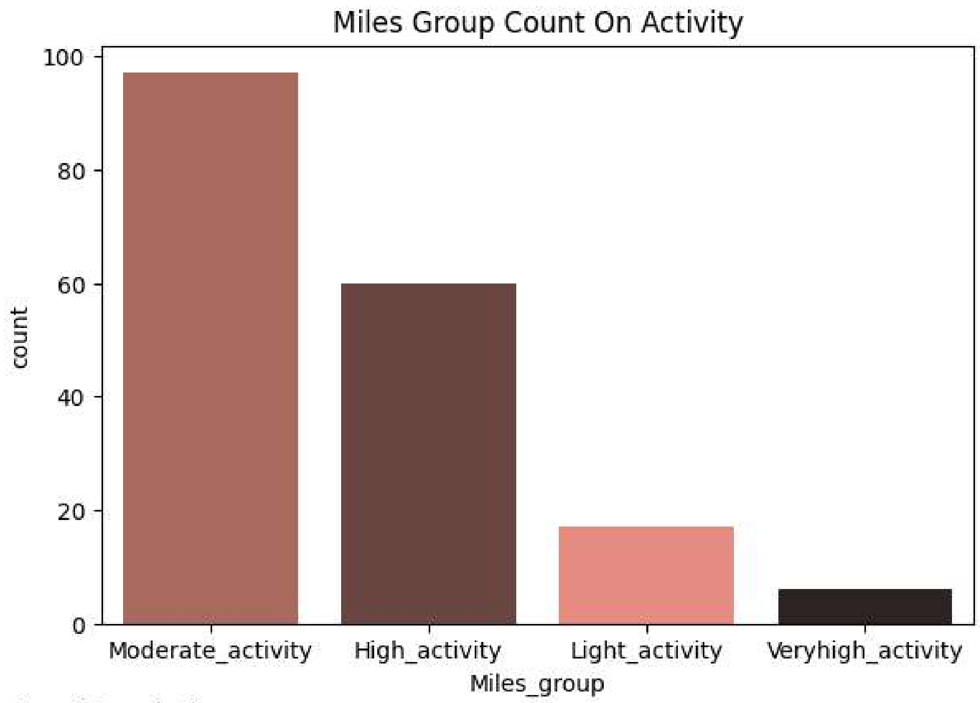
Weekly running MILES Distribution

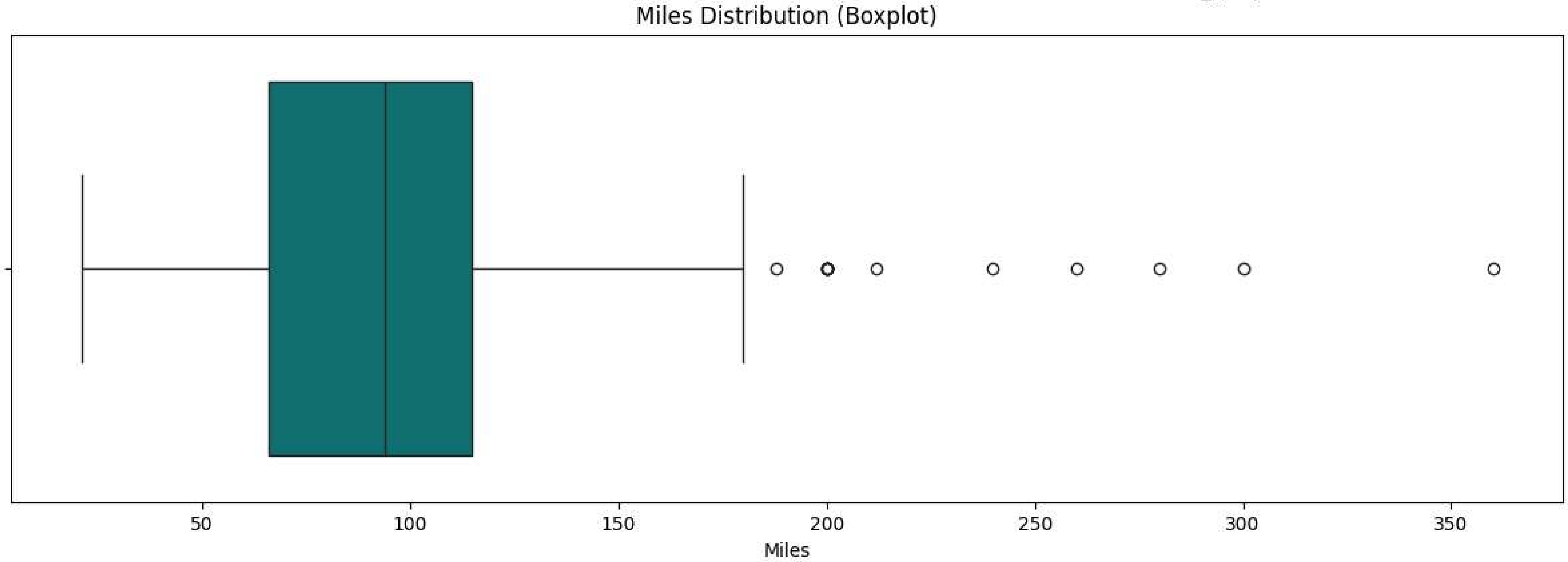
In [ ]:
```
round(data["Miles_group"].value_counts(normalize = True)*100,2)
```

Out[ ]:
```
Miles_group
Moderate_activity    53.89
High_activity        33.33
Light_activity        9.44
Veryhigh_activity     3.33
Name: proportion, dtype: float64
```

In [ ]:
```
plt.figure(figsize = (15,10))
plt.subplot(2,2,1)
sns.histplot(x = "Miles" , data = data , edgecolor = "black", color = "lightblue", kde = False)
plt.title("Miles Distribution (Histogram)")
plt.subplot(2,2,2)
sns.countplot(x = "Miles_group", data = data , hue = "Miles_group", legend = False , palette  = "dark:salmon_r", order = data["Miles_group"
plt.title("Miles Group Count On Activity")
plt.subplot(2,1,2)
sns.boxplot(x = "Miles", data = data,color = "teal")
plt.title("Miles Distribution (Boxplot)")
plt.show()
```





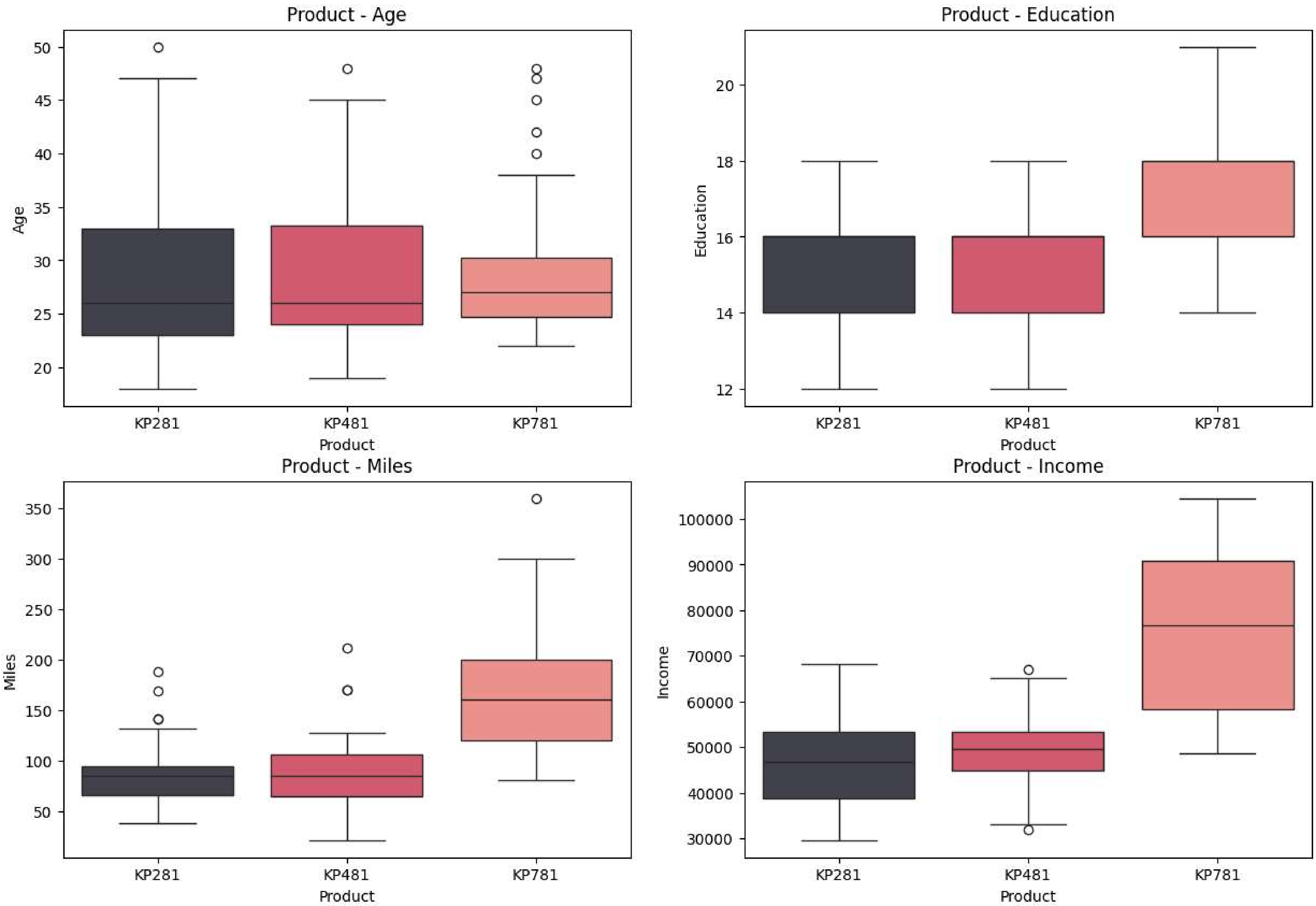


🔍 **Insights**

- The majority of customers (~ 88%) are inclined towards moderate to high levels of treadmill activity, indicating a strong engagement in physical fitness routines that involve running significant distances per week.

Analysis on Product Preference Across Customer Profile

In [ ]:
```
num_var = ["Age","Education","Miles","Income"]
plt.figure(figsize= (15,10))
for i , j in enumerate(num_var):
  i = i+1
  plt.subplot(2,2,i)
  sns.boxplot(x = "Product", y = j , data = data,palette = palette)
  plt.title(f"Product - {j}")
plt.show()
```



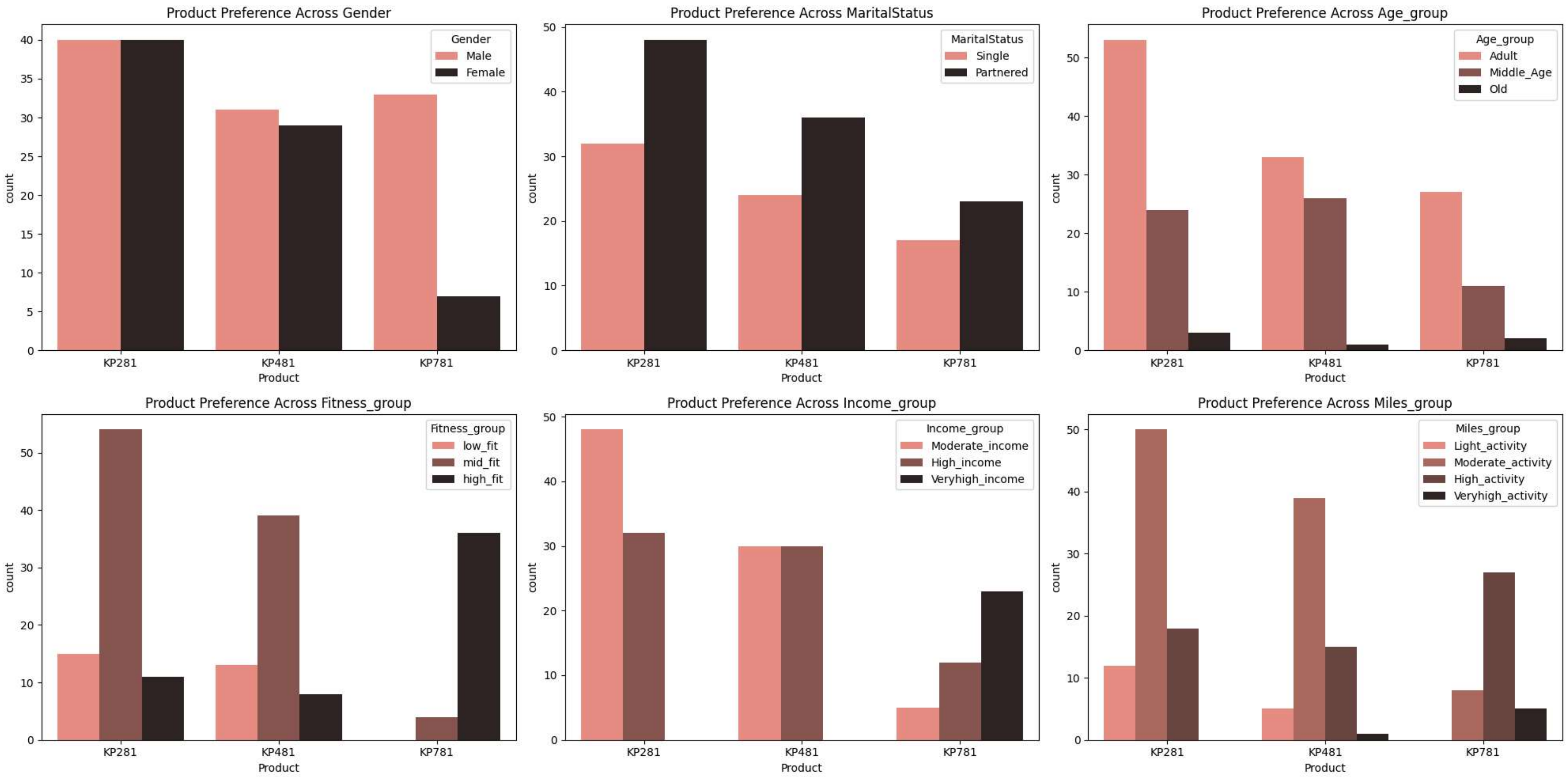
🔍 Insights

- The analysis suggests a clear preference for the KP781 treadmill model among customers with higher education and income levels, who also engage in running activities exceeding 150 miles per week.

In [ ]:
```python
cat_var = ["Gender","MaritalStatus","Age_group","Fitness_group","Income_group","Miles_group"]

plt.figure(figsize= (20,10))

for i,j in enumerate(cat_var):
  i = i+1
  plt.subplot(2,3,i)
  sns.countplot(x = "Product", hue = j, data = data, palette = "dark:salmon_r")
  plt.title(f"Product Preference Across {j}")
plt.tight_layout()
plt.show()
```



🔍 Insights

- Product Preference Across Gender:
- KP781 is significantly more popular among males compared to females.
- Product Preference Across Marital Status:
- Partnered individuals prefer KP281 and KP781 more than singles.
- Product Preference Across Age Group:
- Adults predominantly purchase KP281.
- Product Preference Across Fitness Group:
- Mid fit individuals overwhelmingly prefer KP281 and KP481.
- Product Preference Across Income Group:
- KP281 is most popular among moderate income individuals.
- KP781 is preferred more by customers with Very high income individuals.
- Product Preference Across Miles Group:
- Light activity group predominantly prefers KP281.

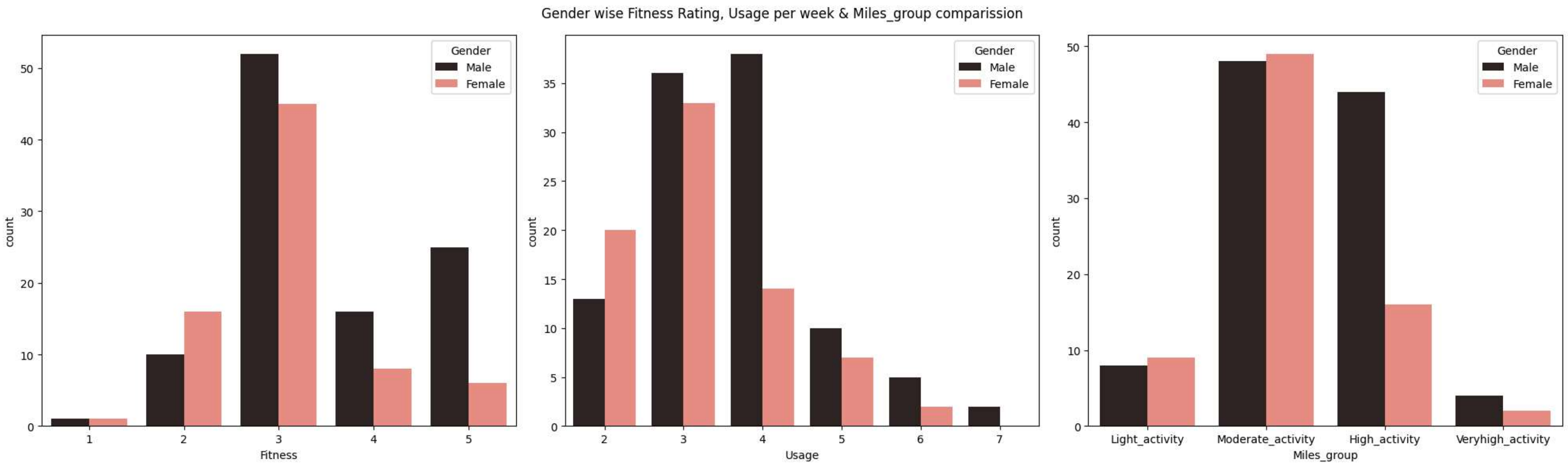
In [ ]:
```
data.head(3)
```

Out[ ]:

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles	Age_group	Fitness_group	Income_group	Miles_group	Produ
0	KP281	18	Male	14	Single	3	4	29562	112	Adult	high_fit	Moderate_income	High_activity	
1	KP281	19	Male	15	Single	2	3	31836	75	Adult	mid_fit	Moderate_income	Moderate_activity	
2	KP281	19	Female	14	Partnered	4	3	30699	66	Adult	mid_fit	Moderate_income	Moderate_activity	

Usage,Fitness,Miles/Week Comapred with Gender

In [ ]:
```
plt.figure(figsize = (20,6))
plt.subplot(1,3,1)
sns.countplot(x = "Fitness",data = data , hue = "Gender", palette = "dark:salmon")
plt.subplot(1,3,2)
sns.countplot(x = "Usage" , data = data , hue = "Gender" , palette = "dark:salmon")
plt.suptitle("Gender wise Fitness Rating, Usage per week & Miles_group comparission ")
plt.subplot(1,3,3)
sns.countplot(x = "Miles_group" , data = data , hue = "Gender" , palette = "dark:salmon")
plt.tight_layout()
plt.show()
```



Correlation

Heatmap

In [ ]:
```
data_corr = data.corr(numeric_only = True)
data_corr
```

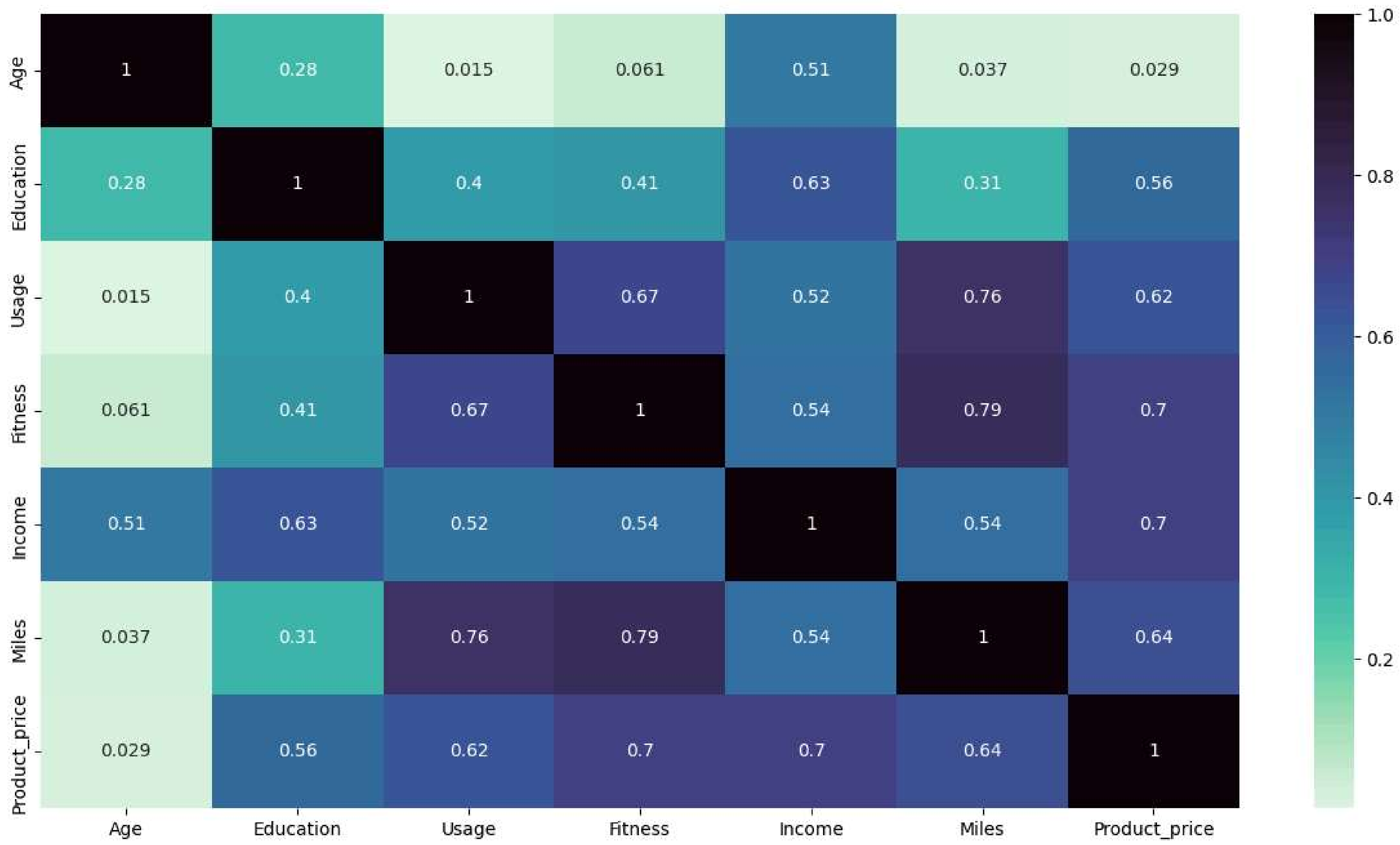
Out[ ]:

	Age	Education	Usage	Fitness	Income	Miles	Product_price
Age	1.000000	0.280496	0.015064	0.061105	0.513414	0.036618	0.029263
Education	0.280496	1.000000	0.395155	0.410581	0.625827	0.307284	0.563463
Usage	0.015064	0.395155	1.000000	0.668606	0.519537	0.759130	0.623124
Fitness	0.061105	0.410581	0.668606	1.000000	0.535005	0.785702	0.696616
Income	0.513414	0.625827	0.519537	0.535005	1.000000	0.543473	0.695847
Miles	0.036618	0.307284	0.759130	0.785702	0.543473	1.000000	0.643923
Product_price	0.029263	0.563463	0.623124	0.696616	0.695847	0.643923	1.000000

In [ ]:

```
plt.figure(figsize = (15,8))
sns.heatmap(data_corr, annot = True , cmap = "mako_r")
```

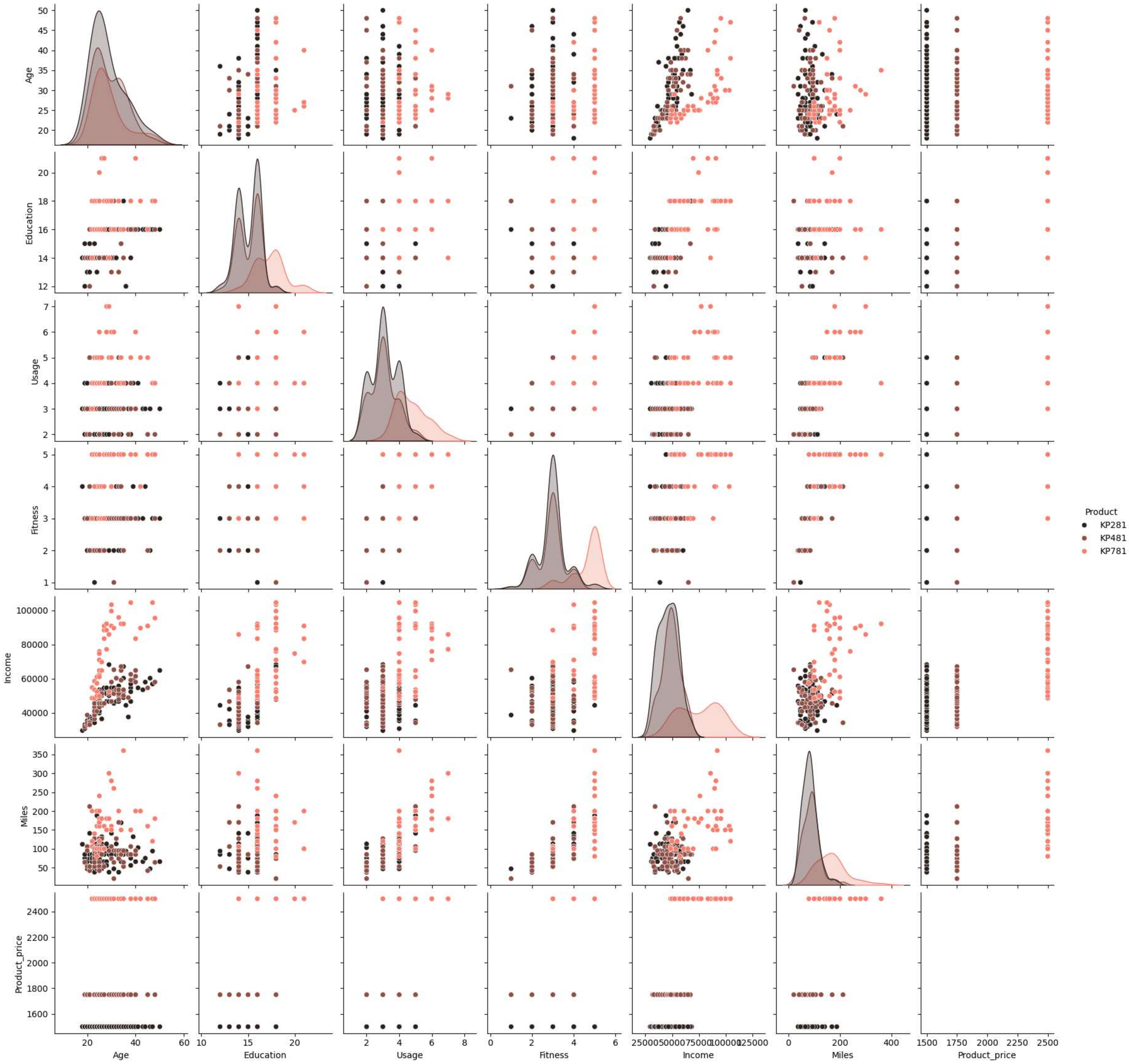
Out[ ]: <Axes: >



In [ ]:

```
sns.pairplot(data,hue= "Product", palette = "dark:salmon")
```

Out[ ]: <seaborn.axisgrid.PairGrid at 0x7caded4db4f0>



🔍 **Insights**

- Age and Income: A positive correlation is evident between age and income, as shown in both the pair plot and heatmap.
- Education and Income: Education level is strongly correlated with income.
- Education and Fitness/Usage: Education also significantly correlates with fitness rating and treadmill usage.
- Usage, Fitness, and Miles: Increased treadmill usage leads to higher fitness levels and greater mileage.

Analysis based on Probability

Probability of product purchase based on gender

In [ ]:
```
round(pd.crosstab(index = data["Gender"], columns = data["Product"], margins = True, normalize = True),2)
```

Out[ ]:

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

🔍 **Insights**

- Male Customers: The probability of a male purchasing a treadmill is 58%. The conditional probabilities for each treadmill model are: KP281 (22%), KP481 (17%), and KP781 (18%).

- Female Customers: The probability of a female purchasing a treadmill is 42%. The conditional probabilities for each treadmill model are: KP281 (22%), KP481 (16%), and KP781 (4%).

In [ ]:
```
data.head(3)
```

Out[ ]:

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles	Age_group	Fitness_group	Income_group	Miles_group	Produ
0	KP281	18	Male	14	Single	3	4	29562	112	Adult	high_fit	Moderate_income	High_activity	
1	KP281	19	Male	15	Single	2	3	31836	75	Adult	mid_fit	Moderate_income	Moderate_activity	
2	KP281	19	Female	14	Partnered	4	3	30699	66	Adult	mid_fit	Moderate_income	Moderate_activity	

In [ ]:
```
col_p = ["Gender" , "Education" , "MaritalStatus","Usage","Fitness","Age_group","Income_group","Miles_group"]
for i in  col_p:
  print(round(pd.crosstab(index = data[i], columns = data["Product"], margins = True, normalize = True),2))
  print("-"*100)
```

```
Product  KP281  KP481  KP781   All
Gender
Female    0.22   0.16   0.04  0.42
Male      0.22   0.17   0.18  0.58
All       0.44   0.33   0.22  1.00
----------------------------------------------------------------------------------------------------
Product    KP281  KP481  KP781   All
Education
12          0.01   0.01   0.00  0.02
13          0.02   0.01   0.00  0.03
14          0.17   0.13   0.01  0.31
15          0.02   0.01   0.00  0.03
16          0.22   0.17   0.08  0.47
18          0.01   0.01   0.11  0.13
20          0.00   0.00   0.01  0.01
21          0.00   0.00   0.02  0.02
All         0.44   0.33   0.22  1.00
----------------------------------------------------------------------------------------------------
Product        KP281  KP481  KP781   All
MaritalStatus
Partnered       0.27   0.20   0.13  0.59
Single          0.18   0.13   0.09  0.41
All             0.44   0.33   0.22  1.00
----------------------------------------------------------------------------------------------------
Product  KP281  KP481  KP781   All
Usage
2         0.11   0.08   0.00  0.18
3         0.21   0.17   0.01  0.38
4         0.12   0.07   0.10  0.29
5         0.01   0.02   0.07  0.09
6         0.00   0.00   0.04  0.04
7         0.00   0.00   0.01  0.01
All       0.44   0.33   0.22  1.00
----------------------------------------------------------------------------------------------------
Product  KP281  KP481  KP781   All
Fitness
1         0.01   0.01   0.00  0.01
2         0.08   0.07   0.00  0.14
3         0.30   0.22   0.02  0.54
4         0.05   0.04   0.04  0.13
5         0.01   0.00   0.16  0.17
All       0.44   0.33   0.22  1.00
----------------------------------------------------------------------------------------------------
Product     KP281  KP481  KP781   All
Age_group
Adult        0.29   0.18   0.15  0.63
Middle_Age   0.13   0.14   0.06  0.34
Old          0.02   0.01   0.01  0.03
All          0.44   0.33   0.22  1.00
----------------------------------------------------------------------------------------------------
Product          KP281  KP481  KP781   All
Income_group
Moderate_income   0.27   0.17   0.03  0.46
High_income       0.18   0.17   0.07  0.41
Veryhigh_income   0.00   0.00   0.13  0.13
All               0.44   0.33   0.22  1.00
----------------------------------------------------------------------------------------------------
Product            KP281  KP481  KP781   All
Miles_group
Light_activity      0.07   0.03   0.00  0.09
Moderate_activity   0.28   0.22   0.04  0.54
High_activity       0.10   0.08   0.15  0.33
Veryhigh_activity   0.00   0.01   0.03  0.03
All                 0.44   0.33   0.22  1.00
----------------------------------------------------------------------------------------------------
```

🔍 Insights

- Gender
  - Females:

1. Probability of purchase: 42%.
    2. Preferred models: KP281 (22%), KP481 (16%), KP781 (4%).
  - Males:
    1. Probability of purchase: 58%.
    2. Preferred models: KP281 (22%), KP481 (17%), KP781 (18%).
- Education
  - Highest Education Level (16 years):
    1. Probability of purchase: 47%.
    2. Preferred models: KP281 (22%), KP481 (17%), KP781 (8%).
  - Moderate Education Level (14 years):
    2. Probability of purchase: 31%.
    3. Preferred models: KP281 (17%), KP481 (13%), KP781 (1%).
- Marital Status
  - Partnered:
    1. Probability of purchase: 59%.
    2. Preferred models: KP281 (27%), KP481 (20%), KP781 (13%).
  - Single:
    1. Probability of purchase: 41%.
    2. Preferred models: KP281 (18%), KP481 (13%), KP781 (9%).
- Usage
  - High Usage (3 times per week):
    1. Probability of purchase: 38%.
    2. Preferred models: KP281 (21%), KP481 (17%), KP781 (1%).
  - Moderate Usage (4 times per week):
    1. Probability of purchase: 29%.
    2. Preferred models: KP281 (12%), KP481 (7%), KP781 (10%).
- Fitness
  - Moderate Fitness Level (3):
    1. Probability of purchase: 54%.
    2. Preferred models: KP281 (30%), KP481 (22%), KP781 (2%).
  - High Fitness Level (5):
    1. Probability of purchase: 17%.
    2. Preferred models: KP281 (1%), KP481 (0%), KP781 (16%).
- Age Group
  - Adults:
    1. Probability of purchase: 63%
    2. Preferred models: KP281 (29%), KP481 (18%), KP781 (15%).
  - Middle-Aged:
    1. Probability of purchase: 34%.
    2. Preferred models: KP281 (13%), KP481 (14%), KP781 (6%).
- Income Group
  - Moderate Income:
    1. Probability of purchase: 46%.
    2. Preferred models: KP281 (27%), KP481 (17%), KP781 (3%).
  - High Income:
    1. Probability of purchase: 41%.
    2. Preferred models: KP281 (18%), KP481 (17%), KP781 (7%).

📊 Customer Profiling

- Based on the above analysis:
  - Probability of purchase of KP281: 44%
  - Probability of purchase of KP481: 33%
  - Probability of purchase of KP781: 22%
- Customer Profile for KP281 Treadmill:
  - Gender: Both genders equally likely

    - Age: Mainly adults
    - Education: 14 to 16 years
    - Marital Status: Partnered
    - Income: Moderate income
    - Weekly Usage: 3 to 4 times
    - Fitness Level: 2 to 3
    - Weekly Running Mileage: Moderate activity
- Customer Profile for KP481 Treadmill:
    - Gender: Both genders equally likely
    - Age: Mainly adults and middle-aged
    - Education: 14 to 16 years
    - Marital Status: Partnered
    - Income: Moderate to high income
    - Weekly Usage: 3 to 4 times
    - Fitness Level: 2 to 3
    - Weekly Running Mileage: Moderate activity
- Customer Profile for KP781 Treadmill:
    - Gender: Mainly males
    - Age: Adults
    - Education: 16 years and above
    - Marital Status: Partnered
    - Income: High to very high income
    - Weekly Usage: 4 to 6 times
    - Fitness Level: 4 to 5
    - Weekly Running Mileage: High to very high activity

Recommendations

- Get More Women Into Exercise
    - Action: Start a focused ad campaign to get women excited about fitness.
    - Details: Show how working out helps and use popular female fitness stars. Give special deals, free tryouts, and classes just for women to get them involved.
- Push KP281 & KP481 as Cheap Choices
    - Action: Sell KP281 and KP481 treadmills as low-cost options for people who make $39,000 to$53,000 a year.
    - Details: Point out how much you get for your money with these models. Use messages about saving cash and offer good payment plans to help more people buy these treadmills.
- Sell KP781 to Pros and Athletes
    - Action: Present KP781 as a top-tier treadmill with cutting-edge features for dedicated fitness buffs and athletes.
    - Details: Team up with fitness influencers and world-class athletes to back the KP781. Showcase its state-of-the-art features and top-notch performance in promotional content. Offer thorough user guides and assistance to help customers make the most of the treadmill.
- Broaden Market to Include Older Age Groups
    - Action: Study and create plans to sell treadmills to people over 50.
    - Details: Evaluate the health advantages and possible risks of treadmill use for seniors.

In [ ]: