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

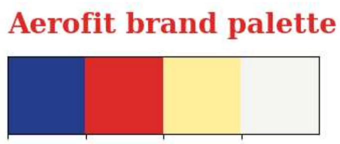
```
df = pd.read_csv("aerofit_treadmill.txt", sep = ",")
```

```
df.head()
```

	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

Next steps: Generate code with df View recommended plots

```
sns.palplot(['#243E8D', '#DB2A27', '#FFEE9A','#f5f5f1'])
font1 = {'family':'serif','color':'#DB2A27','size':20, 'weight': "bold"}
plt.title("Aerofit brand palette ",loc='left',fontfamily='serif',fontsize=20,y=1.2, fontdict = font1)
plt.show()
```



1. Aerofit

Aerofit, a dynamic player in the fitness industry, traces its origins to M/s. Sachdev Sports Co, established in 1928 by Ram Ratan Sachdev. From its modest beginnings in Hyderabad, India, the company evolved into a leading sports equipment supplier across Andhra Pradesh and Telangana. Recognizing the growing need for fitness solutions, M/s. Sachdev Overseas emerged to import quality fitness equipment under the "Aerofit" brand, ensuring affordability and post-sales excellence.

Driven by a dedication to innovation, Nityasach Fitness Pvt Ltd was founded, spearheaded by director Nityesh Sachdev. With the brand "Aerofit" at its core, the company aimed to bridge the gap between international fitness technology and the Indian market. By importing advanced fitness equipment at accessible price points, Aerofit sought to redefine the industry landscape, prioritizing health and vitality while staying true to its legacy of passion and customer focus.

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.**

2. Objective

Create comprehensive customer profiles for each **AeroFit treadmill** product through **descriptive analytics**. Develop two-way contingency tables and analyze **conditional and marginal probabilities** to discern customer characteristics, facilitating improved product recommendations and informed business decisions.

```
df.isnull().sum()
```

```
Product          0
Age              0
Gender           0
Education        0
MaritalStatus    0
Usage            0
Fitness          0
Income           0
```

```
Miles           0
dtype: int64
```

There are no Null values

```
df.duplicated().value_counts()
```

```
False    180
Name: count, dtype: int64
```

There are no duplicates

```
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
```

```
df.head()
```

	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

Next steps: Generate code with df View recommended plots

Lets convert the data type to object which will be helpful to anaylse the profiles based on different dimensions.

```
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 does not have any null values. Seems like data is clean. Our goal is to :-

1.Perform descriptive analytics to **create a customer profile** for each AeroFit treadmill product by developing appropriate tables and charts.
2.For each AeroFit treadmill product, construct two-way contingency tables and compute all conditional and marginal probabilities along with their insights/impact on the business.

Dataset Consists of:

Product: Product Purchased KP281, KP481, or KP781

Age: In years

Gender: Male/Female

Education: In years

MaritalStatus: single or partnered

Usage: 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: average number of miles the customer expects to walk/run each week

The product portfolio is as follows:-

1. The KP281 is an entry-level treadmill that sells for $1,500.
2. The KP481 is for mid-level runners that sell for $1,750.
3. The KP781 treadmill is having advanced features that sell for $2,500.

Let us catagorically divide the product into entry, mid and advance levels.

```
df.loc[df['Product'] == "KP281", 'Product_Category'] = "Entry"
df.loc[df['Product'] == "KP481", 'Product_Category'] = "Mid"
df.loc[df['Product'] == "KP781", 'Product_Category'] = "Advance"
df.head()
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles	Product_Category
0	KP281	18	Male	14	Single	3	4	29562	112	Entry
1	KP281	19	Male	15	Single	2	3	31836	75	Entry
2	KP281	19	Female	14	Partnered	4	3	30699	66	Entry
3	KP281	19	Male	12	Single	3	3	32973	85	Entry
4	KP281	20	Male	13	Partnered	4	2	35247	47	Entry

Next steps: Generate code with df View recommended plots

```
df.describe()
```

	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

```
import copy

df_object = copy.deepcopy(df)

df['Usage'] = df['Usage'].astype('object')
df['Fitness'] = df['Fitness'].astype('object')
df['Income'] = df['Income'].astype('object')
df['Miles'] = df['Miles'].astype('object')
df['Education'] = df['Education'].astype('object')
df['Age'] = df['Age'].astype('object')
```

```
df_object.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 10 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 
 9   Product_Category  180 non-null    object
dtypes: int64(6), object(4)
memory usage: 14.2+ KB
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 10 columns):
 #   Column            Non-Null Count  Dtype 
---  ------            --------------  ----- 
 0   Product           180 non-null    object
 1   Age               180 non-null    object
 2   Gender            180 non-null    object
 3   Education         180 non-null    object
 4   MaritalStatus     180 non-null    object
 5   Usage             180 non-null    object
 6   Fitness           180 non-null    object
 7   Income            180 non-null    object
 8   Miles             180 non-null    object
 9   Product_Category  180 non-null    object
dtypes: object(10)
memory usage: 14.2+ KB
```

```
df.describe(include="object")
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles	Product_Category
count	180	180	180	180	180	180	180	180	180	180
unique	3	32	2	8	2	6	5	62	37	3
top	KP281	25	Male	16	Partnered	3	3	45480	85	Entry
freq	80	25	104	85	107	69	97	14	27	80

The Entry level product is the most selling product, Partnered are most likely purchsing the trademil with Males being dominating purchases.

```
df_object.describe()
```

	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

There are no missing values in the data.

There are 3 unique products in the dataset.

- KP281 is the most frequent product.
- Minimum & Maximum age of the person is 18 & 50, mean is 28.79 and
- 75% of persons have age less than or equal to 33.
- Most of the people are having 16 years of education i.e. 75% of persons are having education <= 16 years.
- Out of 180 data points, 104's gender is Male and rest are the female. So, approximate 57.78 % buyers are Males.
- Out of 180 data points, 107's are Partnered. 59.4 % people are partnered.
- Standard deviation for Income & Miles is very high. These variables might have the outliers in it.

The mean age group that purchased the trademill in past three months have been 28, with the average education being of 15 years, On an average people plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3.3 on an average. Approx 60 % of people who purchased trademill in past three months were partnered. Approximate 58 % people who purchsed the trademill were Male.

Let see whats the relation between buyers who are Male and Partnered. Also let us see outliers in Income and Miles.

1. **Product -** Over the past three months, the KP281 product demonstrated the highest sales performance among the three products, accounting for approximately 44% of total sales.
2. **Gender** - Based on the data of last 3 months, around 58% of the buyers were Male and 42% were female.
3. **Marital Status** - Based on the data of last 3 months, around 60% of the buyers were Married and 40% were single.
4. **Age** - The age range of customers spans from 18 to 50 year, with an average age of 29 years.
5. **Education** - Customer education levels vary between 12 and 21 years, with an average education duration of 16 years.
6. **Usage** - Customers intend to utilize the product anywhere from 2 to 7 times per week, with an average usage frequency of 3 times per week.
7. **Fitness** - On average, customers have rated their fitness at 3 on a 5-point scale, reflecting a moderate level of fitness.
8. **Income** - The annual income of customers falls within the range of USD 30,000 to USD 100,000, with an average income of approximately USD 54,000.
9. **Miles** - Customers' weekly running goals range from 21 to 360 miles, with an average target of 103 miles per week.

```
df_malepartnered = df[(df["Gender"] == "Male" )& (df["MaritalStatus"] == "Partnered")]
df_malesingle = df[(df["Gender"] == "Male" )& (df["MaritalStatus"] == "Single")]
df_femalepartnered = df[(df["Gender"] == "Female" )& (df["MaritalStatus"] == "Partnered")]
df_femalesingle = df[(df["Gender"] == "Female" )& (df["MaritalStatus"] == "Single")]
```

```
df_malepartnered.describe(include="object")
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles	Product_Category
count	61	61	61	61	61	61	61	61	61	61
unique	3	23	1	7	1	6	5	41	26	3
top	KP281	25	Male	16	Partnered	4	3	53439	85	Entry
freq	21	11	61	35	61	22	27	6	6	21

```
df_malepartnered["Income"].median()
```

```
53439.0
```

```
df_femalepartnered.describe(include="object")
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles	Product_Category
count	46	46	46	46	46	46	46	46	46	46
unique	3	21	1	4	1	5	4	26	18	3
top	KP281	25	Female	16	Partnered	3	3	45480	85	Entry
freq	27	7	46	21	46	19	30	4	10	27

```
df_femalepartnered["Income"].median()
```

```
48891.0
```

```
df_malesingle.describe()
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles	Product_Category
count	43	43	43	43	43	43	43	43	43	43
unique	3	20	1	7	1	5	4	29	21	3
top	KP281	22	Male	14	Single	4	3	32973	85	Entry
freq	19	4	43	16	43	16	25	3	8	19

```
df_malesingle["Income"].median()
```

```
48891.0
```

```
df_femalesingle.describe(include="object")
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles	Product_Category
count	30	30	30	30	30	30	30	30	30	30
unique	3	13	1	5	1	4	5	19	18	3
top	KP481	23	Female	16	Single	3	3	46617	75	Mid
freq	14	5	30	14	30	14	15	3	4	14

```
df_femalesingle["Income"].median()
```

```
46617.0
```

Profiling could be based on :-

1. Male Partnered : The mean age group of male partnered purchased the trademill in past three months have been 30, with the average education being of 16 years, On an average male partnered plan to use Trademill 4 days a week. They have a self-rated fitness as 3.4 on an average.The mean income of Male Partnered is around $48000.

2. Female Partnered: The mean age group that purchased the trademill in past three months have been approximately 29, with the average education being of 15 years, On an average partnered females plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3 on an average.The mean income of Female Partnered is around $53000.

3. Male Single : The mean age group that purchased the trademill in past three months have been approximately 27, with the average education being of 15 years, On an average single males plan to use Trademill more than 3.6 days a week. They have a self-rated fitness as 3.3 on an average.The mean income of Male Single is around $48000.

4. Female Single: The mean age group that purchased the trademill in past three months have been approximately 29, with the average education being of 16 years, On an average single females plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3.3 on an average. The median income of Femal Single is around $47000.

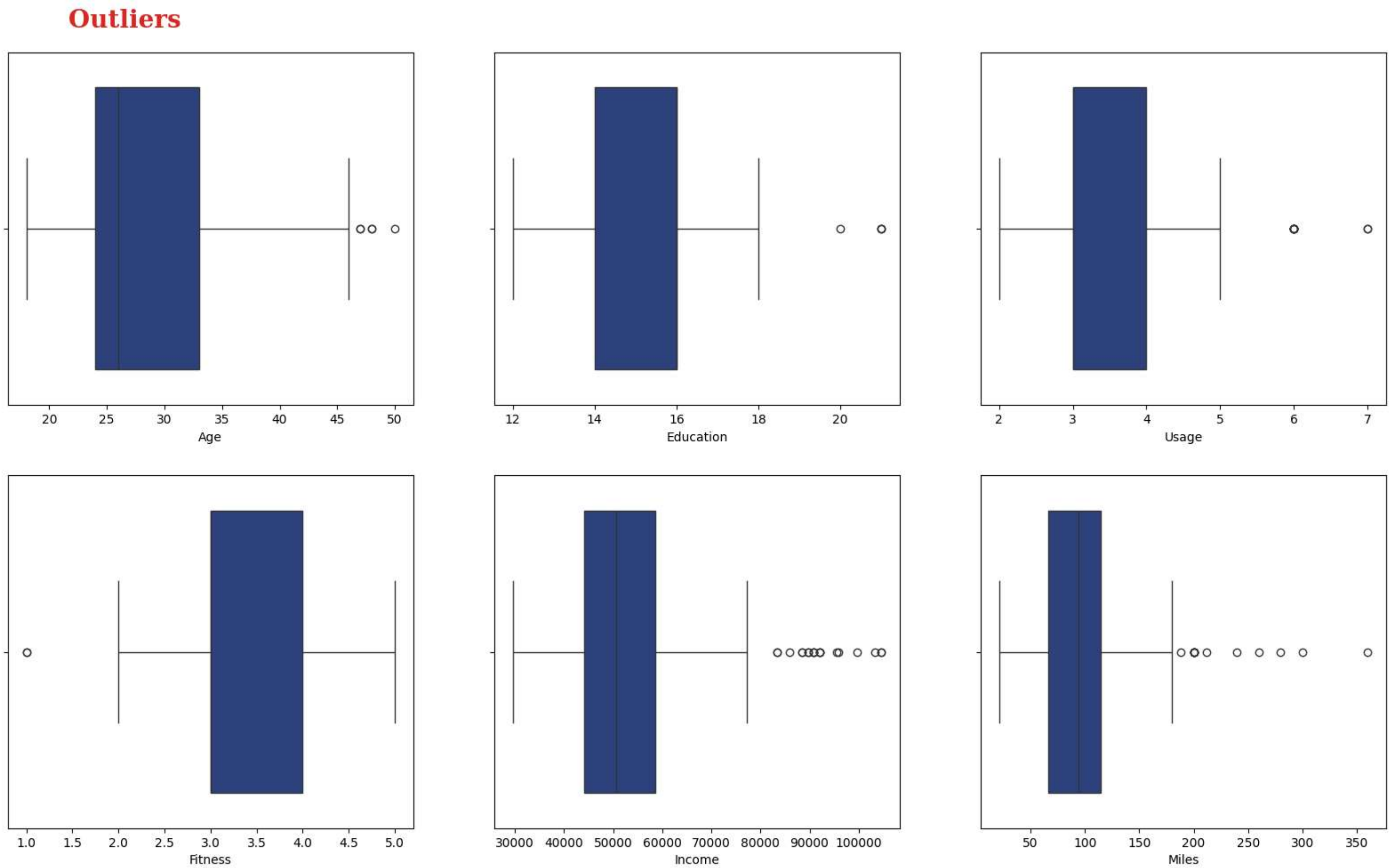
Lets detect outliers in the data.

Detect Outliers (using boxplot, “describe” method by checking the difference between mean and median)

Outliers are datapoints which are exceptional. So, here let us analyse without ouliers and we can analyse the data outliers seprately.

```
# sns.boxplot(data = df,  y = "Income", color = '#243E8D')

# plt.show()
```

```
fig, axis = plt.subplots(nrows=2, ncols=3, figsize=(20, 8))
fig.subplots_adjust(top=1.2)
plt.title("Outliers ",loc='left',y=2.25,x = -2.25, fontdict= font1)
sns.boxplot(data=df, x="Age", color = '#243E8D', orient='h', ax=axis[0,0])
sns.boxplot(data=df, x="Education", color = '#243E8D', orient='h', ax=axis[0,1])
sns.boxplot(data=df, x="Usage", color = '#243E8D', orient='h', ax=axis[0,2])
sns.boxplot(data=df, x="Fitness", color = '#243E8D', orient='h', ax=axis[1,0])
sns.boxplot(data=df, x="Income", color = '#243E8D', orient='h', ax=axis[1,1])
sns.boxplot(data=df, x="Miles", color = '#243E8D', orient='h', ax=axis[1,2])
plt.show()
```



The income and Miles have lot of outliers, hence we can use median values for them in the analysis. Moreover we can analyse the outliers profile seprately.

```
df_Outliers_income = copy.deepcopy(df)
df_woOutliers = copy.deepcopy(df)

Q3 = np.quantile(df["Income"], 0.75)
Q1 = np.quantile(df["Income"], 0.25)
IQR = Q3 - Q1

lower_range = Q1 - 1.5 * IQR
upper_range = Q3 + 1.5 * IQR

df_Outliers_income = df[(df['Income'] < lower_range) | (df['Income'] > upper_range)]
df_woOutliers = df[(df['Income'] > lower_range) & (df['Income'] < upper_range)]
df_woOutliers
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles	Product_Category
0	KP281	18	Male	14	Single	3	4	29562	112	Entry
1	KP281	19	Male	15	Single	2	3	31836	75	Entry
2	KP281	19	Female	14	Partnered	4	3	30699	66	Entry
3	KP281	19	Male	12	Single	3	3	32973	85	Entry
4	KP281	20	Male	13	Partnered	4	2	35247	47	Entry
...	...	...	...	...	...	...	...	...	...	...
156	KP781	25	Male	20	Partnered	4	5	74701	170	Advance
157	KP781	26	Female	21	Single	4	3	69721	100	Advance
158	KP781	26	Male	16	Partnered	5	4	64741	180	Advance
163	KP781	28	Male	18	Partnered	7	5	77191	180	Advance
165	KP781	29	Male	18	Single	5	5	52290	180	Advance

161 rows × 10 columns

Next steps: Generate code with df_woOutliers View recommended plots

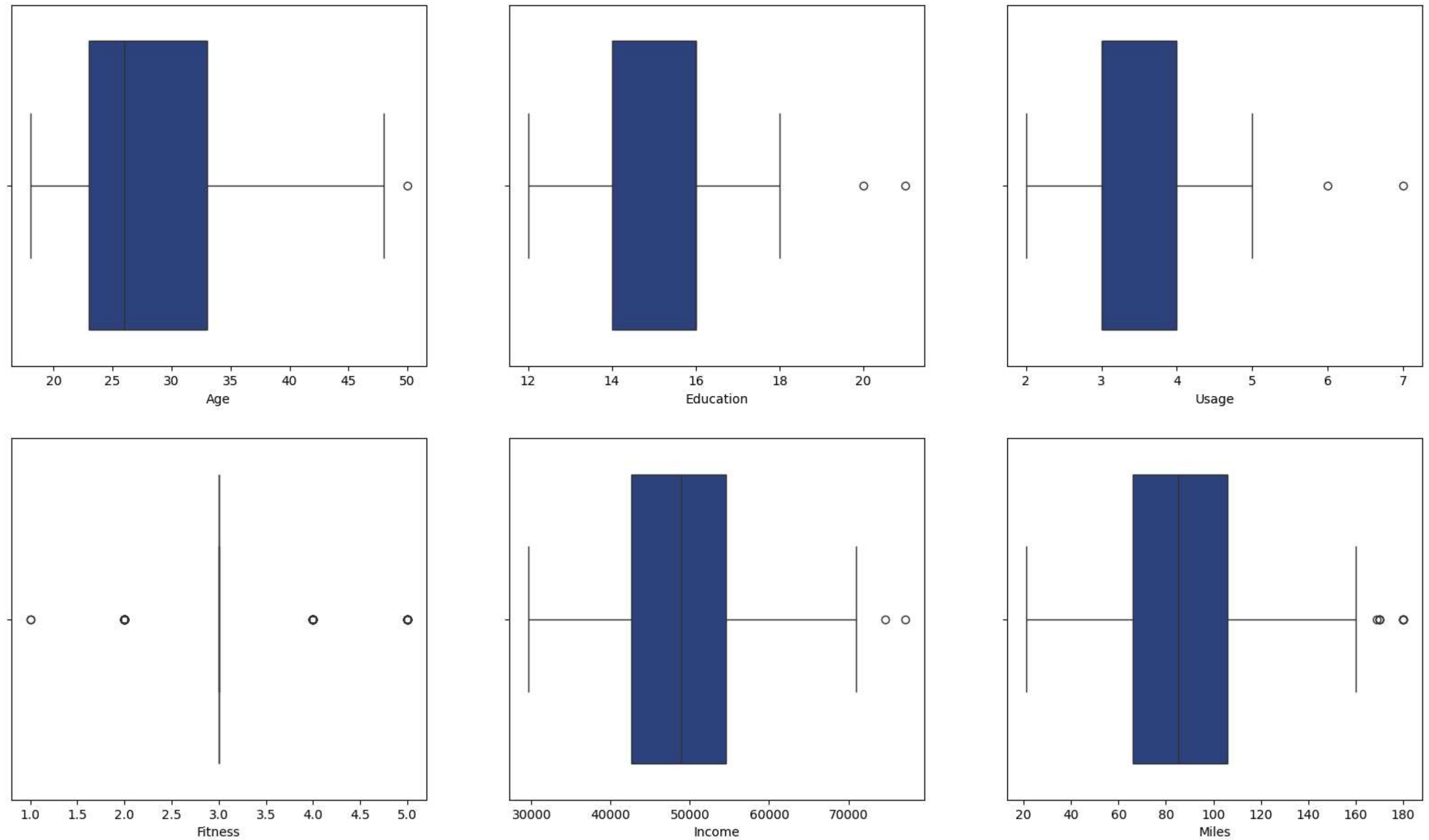
```
df_Outliers_miles = copy.deepcopy(df)


Q3 = np.quantile(df["Miles"], 0.75)
Q1 = np.quantile(df["Miles"], 0.25)
IQR = Q3 - Q1

lower_range = Q1 - 1.5 * IQR
upper_range = Q3 + 1.5 * IQR

df_Outliers_miles = df[(df['Miles'] < lower_range) | (df['Miles'] > upper_range)]
df_woOutliers = df_woOutliers[(df_woOutliers['Miles'] > lower_range) & (df_woOutliers['Miles'] < upper_range)]
```

```
df_woOutliers.shape
```

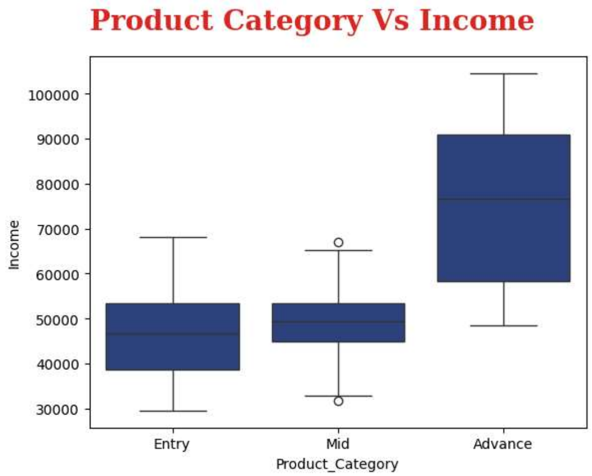
```
(155, 10)
```

```
fig, axis = plt.subplots(nrows=2, ncols=3, figsize=(20, 8))
fig.subplots_adjust(top=1.2)
plt.title(" Without Outliers ",loc='left',y=2.25,x = -2.25, fontdict= font1)
sns.boxplot(data=df_woOutliers, x="Age", color = '#243E8D', orient='h', ax=axis[0,0])
sns.boxplot(data=df_woOutliers, x="Education", color = '#243E8D', orient='h', ax=axis[0,1])
sns.boxplot(data=df_woOutliers, x="Usage", color = '#243E8D', orient='h', ax=axis[0,2])
sns.boxplot(data=df_woOutliers, x="Fitness", color = '#243E8D', orient='h', ax=axis[1,0])
sns.boxplot(data=df_woOutliers, x="Income", color = '#243E8D', orient='h', ax=axis[1,1])
sns.boxplot(data=df_woOutliers, x="Miles", color = '#243E8D', orient='h', ax=axis[1,2])
plt.show()
```

Without Outliers

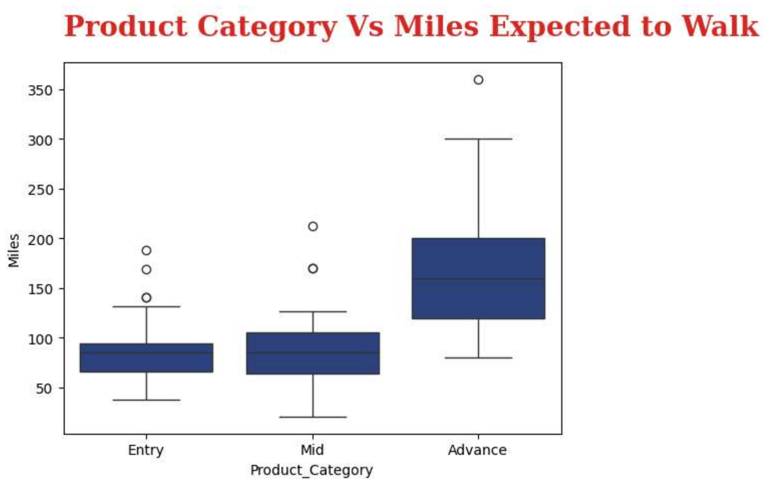



```
sns.boxplot(data = df, x = "Product_Category", y = "Income", color = '#243E8D')
plt.title("Product Category Vs Income ",loc='left',y=1.05, fontdict= font1)
plt.show()
```



The median income category for Entry level is 46000 dollars. With Interquartile range being

```
# Age Education Usage Fitness Income  Miles

sns.boxplot(data = df, x = "Product_Category", y = "Miles", color = '#243E8D')
plt.title("Product Category Vs Miles Expected to Walk ",loc='left',y=1.05, fontdict= font1)
plt.show()
```

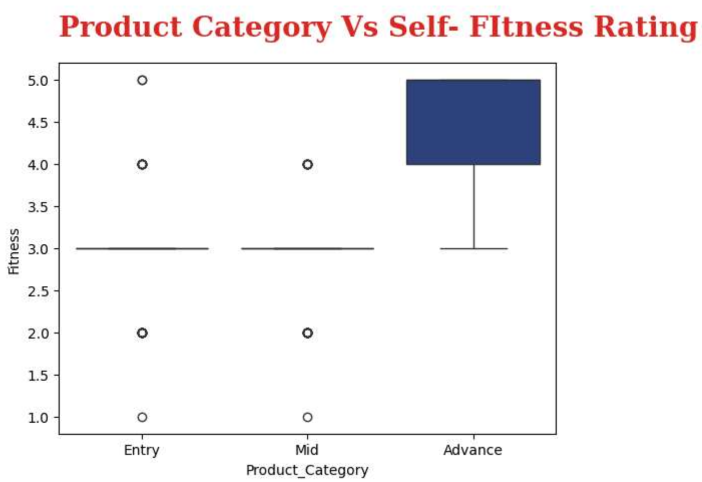


```
p_25_miles = np.percentile(df["Miles"], 25)
p_75_miles = np.percentile(df["Miles"], 75)
iqr_miles = p_75_miles - p_25_miles
iqr_miles
normal_range_miles = (df["Miles"].max() - df["Miles"].min())
iqr_miles, normal_range_miles


# upper limit = Q3 + 1.5 * IQR
upper_miles = p_75_miles + 1.5*(iqr_miles)

# lower limit = Q3 - 1.5 * IQR
lower_miles = p_25_miles - 1.5*(iqr_miles)

iqr_miles, normal_range_miles,upper_miles,lower_miles
```

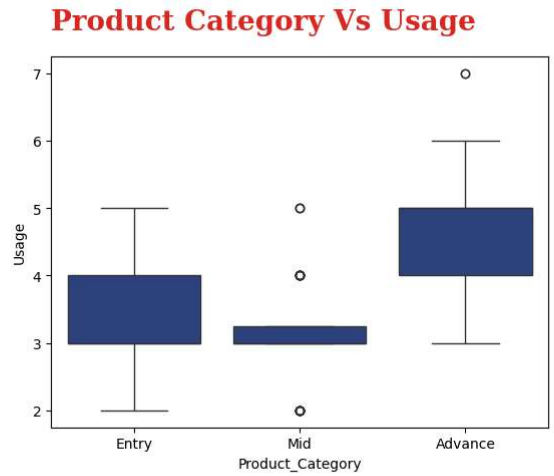
```
(48.75, 339, 187.875, -7.125)
```

Start coding or generate with AI.

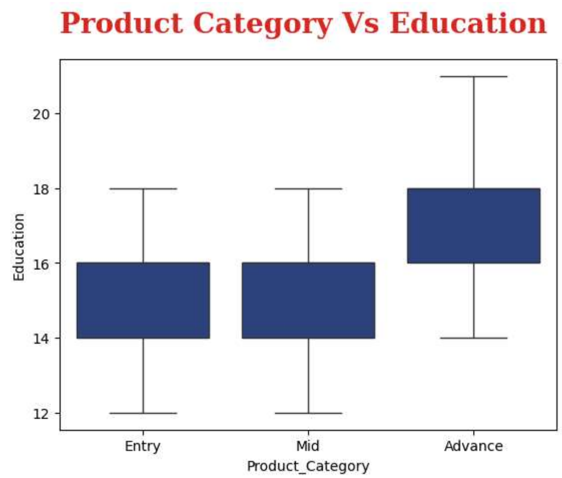
```
# Age   Education   Usage   Fitness Income  Miles

sns.boxplot(data = df, x = "Product_Category", y = "Fitness", color = '#243E8D')
plt.title("Product Category Vs Self- FItness Rating ",loc='left',y=1.05, fontdict= font1)
plt.show()
```



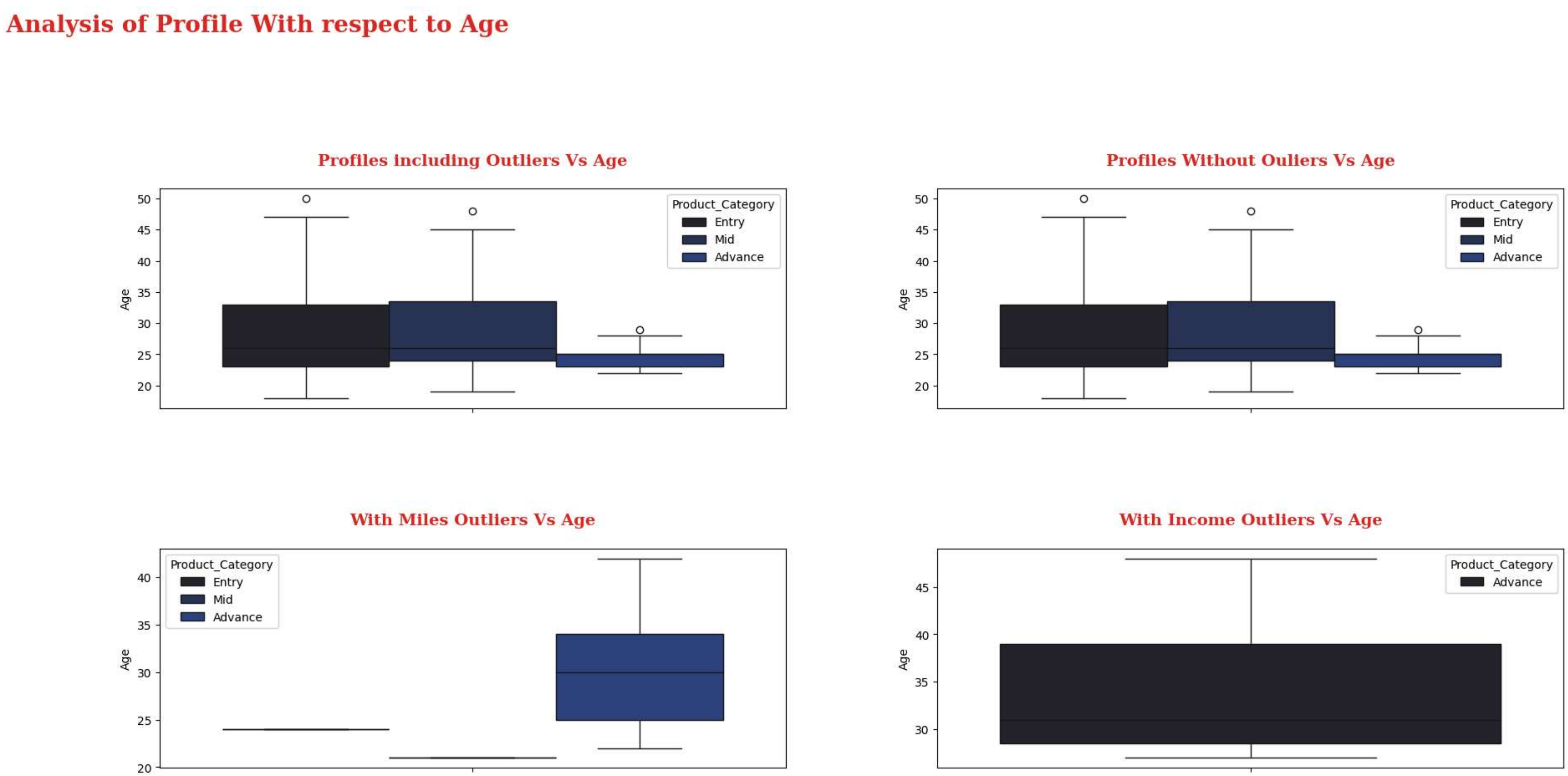

```
# Age   Education   Usage   Fitness Income  Miles

sns.boxplot(data = df, x = "Product_Category", y = "Usage", color = '#243E8D')
plt.title("Product Category Vs Usage ",loc='left',y=1.05, fontdict= font1)
plt.show()
```




```
# Age   Education   Usage   Fitness Income  Miles

sns.boxplot(data = df, x = "Product_Category", y = "Education", color = '#243E8D')
plt.title("Product Category Vs Education ",loc='left',y=1.05, fontdict= font1)
plt.show()
```



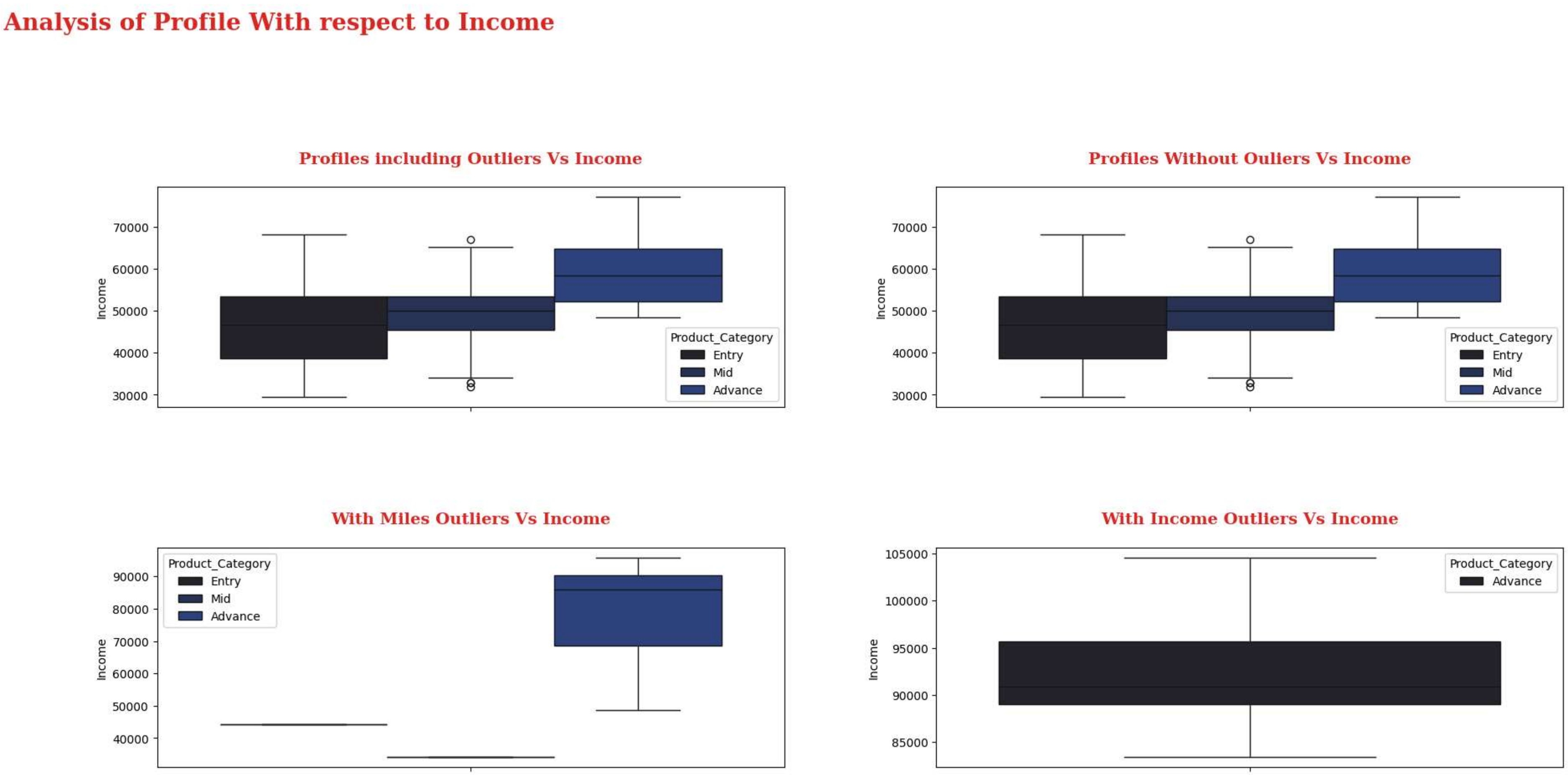
```
# Age   Education   Usage   Fitness Income  Miles

fig, axis = plt.subplots(nrows=2, ncols=2, figsize=(20, 10))
fig.subplots_adjust(top=0.8)
fig.tight_layout(pad=10)
plt.title(" Analysis of Profile With respect to Age ",loc='left',y=3.25,x = -1.5, fontdict= font1)
sns.boxplot(data=df_woOutliers, y="Age", palette='dark:#243E8D',hue="Product_Category", ax=axis[0,0])
sns.boxplot(data=df_woOutliers, y="Age", palette='dark:#243E8D' ,hue="Product_Category", ax=axis[0,1])
sns.boxplot(data=df_Outliers_miles, y="Age", palette='dark:#243E8D', hue="Product_Category", ax=axis[1,0])
sns.boxplot(data=df_Outliers_income, y="Age", palette='dark:#243E8D', hue="Product_Category", ax=axis[1,1])

axis[0,0].set_title("Profiles including Outliers Vs Age", pad=20, fontsize=14, fontdict= font1)
axis[0,1].set_title("Profiles Without Ouliers Vs Age", pad=20, fontsize=14, fontdict= font1)
axis[1,0].set_title("With Miles Outliers Vs Age", pad=20, fontsize=14, fontdict= font1)
axis[1,1].set_title("With Income Outliers Vs Age", pad=20, fontsize=14, fontdict= font1)
plt.show()
```



Advance product is purchased by mostly people above 30s. While mean age of Mid and Entry level product is 25.

```
fig, axis = plt.subplots(nrows=2, ncols=2, figsize=(20, 10))
fig.subplots_adjust(top=0.8)
fig.tight_layout(pad=10)
plt.title(" Analysis of Profile With respect to Income ",loc='left',y=3.25,x = -1.5, fontdict= font1)
sns.boxplot(data=df_woOutliers, y="Income", palette='dark:#243E8D', hue="Product_Category", ax=axis[0,0])
```

```
sns.boxplot(data=df_woOutliers, y="Income", palette='dark:#243E8D' ,hue="Product_Category", ax=axis[0,0])
sns.boxplot(data=df_woOutliers, y="Income", palette='dark:#243E8D' ,hue="Product_Category", ax=axis[0,1])
sns.boxplot(data=df_Outliers_miles, y="Income", palette='dark:#243E8D', hue="Product_Category", ax=axis[1,0])
sns.boxplot(data=df_Outliers_income, y="Income", palette='dark:#243E8D', hue="Product_Category", ax=axis[1,1])

axis[0,0].set_title("Profiles including Outliers Vs Income", pad=20, fontsize=14, fontdict= font1)
axis[0,1].set_title("Profiles Without Ouliers Vs Income", pad=20, fontsize=14, fontdict= font1)
axis[1,0].set_title("With Miles Outliers Vs Income", pad=20, fontsize=14, fontdict= font1)
axis[1,1].set_title("With Income Outliers Vs Income", pad=20, fontsize=14, fontdict= font1)
plt.show()
```




```
df_woOutliers[df_woOutliers["Product_Category"]=="Entry"]["Income"].median(), df_woOutliers[df_woOutliers["Product_Category"]=="Entry"]["Income"].mean()
```

```
(46617.0, 46444.29113924051)
```

```
df_woOutliers[df_woOutliers["Product_Category"]=="Mid"]["Income"].median(), df_woOutliers[df_woOutliers["Product_Category"]=="Mid"]["Income"].mean()
```

```
(50028.0, 49225.57627118644)
```

```
df_woOutliers[df_woOutliers["Product_Category"]=="Advance"]["Income"].median(), df_woOutliers[df_woOutliers["Product_Category"]=="Advance"]["Income"].mean()
```

```
(58516.0, 59913.41176470588)
```

The income group lying around $49000 tends to purchase Entry level. Clearly Male Partnered, Male Single, Female Single lies in this bracket

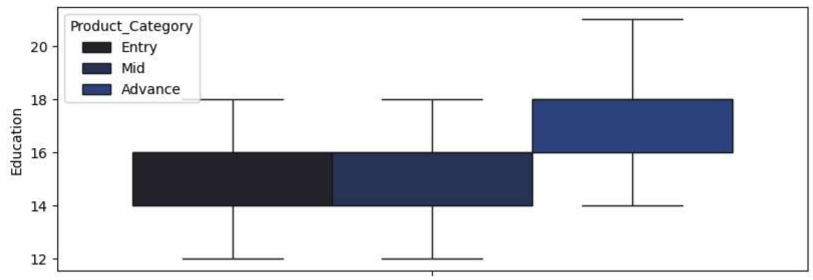
1. Male Partnered : The mean age group of male partnered purchased the trademill in past three months have been 30, with the average education being of 16 years, On an average male partnered plan to use Trademill 4 days a week. They have a self-rated fitness as 3.4 on an average.The mean income of Male Partnered is around $48000. They are more likely to purchase Mid level Tradmill based on Income critera.
2. Female Partnered: The mean age group that purchased the trademill in past three months have been approximately 29, with the average education being of 15 years, On an average partnered females plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3 on an average.The mean income of Female Partnered is around $53000.They are more likely to purchase Mid level or Advance Level Tradmill based on Income critera.
3. Male Single : The mean age group that purchased the trademill in past three months have been approximately 27, with the average education being of 15 years, On an average single males plan to use Trademill more than 3.6 days a week. They have a self-rated fitness as 3.3 on an average.The mean income of Male Single is around $48000.They are more likely to purchase Mid Level Tradmill based on Income critera.
4. Female Single: The mean age group that purchased the trademill in past three months have been approximately 29, with the average education being of 16 years, On an average single females plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3.3 on an average. The median income of Femal Single is around $47000.They are more likely to purchase Entry level or Mid level Tradmill based on Income critera.

Double-click (or enter) to edit

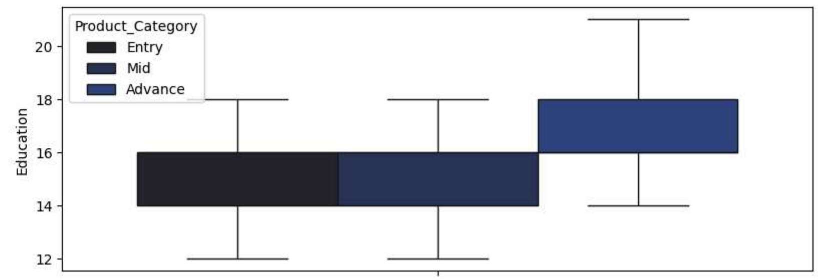
```
fig, axis = plt.subplots(nrows=2, ncols=2, figsize=(20, 10))
fig.subplots_adjust(top=0.8)
fig.tight_layout(pad=10)
plt.title(" Numbers of Years of Education ",loc='left',y=3.25,x = -1.5, fontdict= font1)
sns.boxplot(data=df, y="Education", palette='dark:#243E8D', hue="Product_Category", ax=axis[0,0])
sns.boxplot(data=df_woOutliers, y="Education", palette='dark:#243E8D', hue="Product_Category", ax=axis[0,1])
sns.boxplot(data=df_Outliers_miles, y="Education", palette='dark:#243E8D', hue="Product_Category", ax=axis[1,0])
sns.boxplot(data=df_Outliers_income, y="Education", palette='dark:#243E8D', hue='Product_Category', ax=axis[1,1])

axis[0,0].set_title("Profiles inclusing Outliers Vs Miles", pad=20, fontsize=14, fontdict= font1)
axis[0,1].set_title("Profiles Without Ouliers Vs Miles", pad=20, fontsize=14, fontdict= font1)
axis[1,0].set_title("With Miles Outliers Vs Miles", pad=20, fontsize=14, fontdict= font1)
axis[1,1].set_title("With Income Outliers Vs Miles", pad=20, fontsize=14, fontdict= font1)
plt.show()
```

Numbers of Years of Education

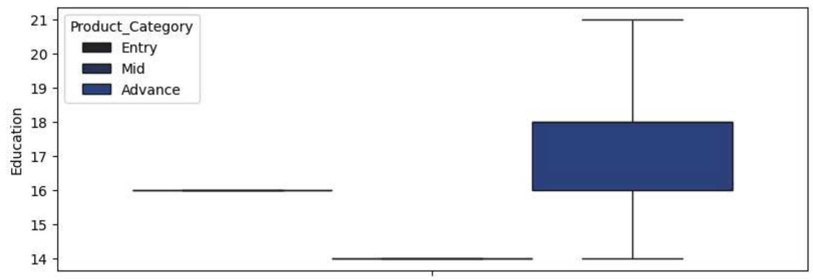
Profiles inclusing Outliers Vs Miles



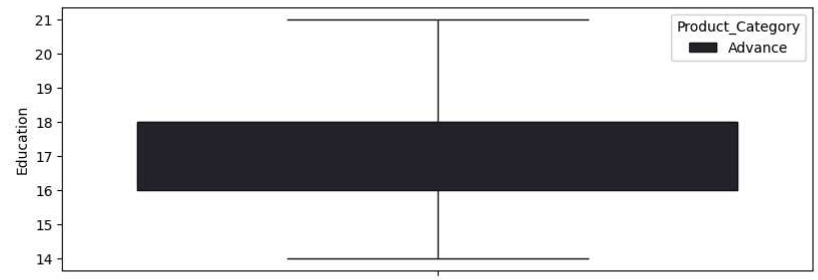
Profiles Without Ouliers Vs Miles



With Miles Outliers Vs Miles



With Income Outliers Vs Miles



The income group lying around $49000 tends to purchase Entry level. Clearly Male Partnered, Male Single, Female Single lies in this bracket

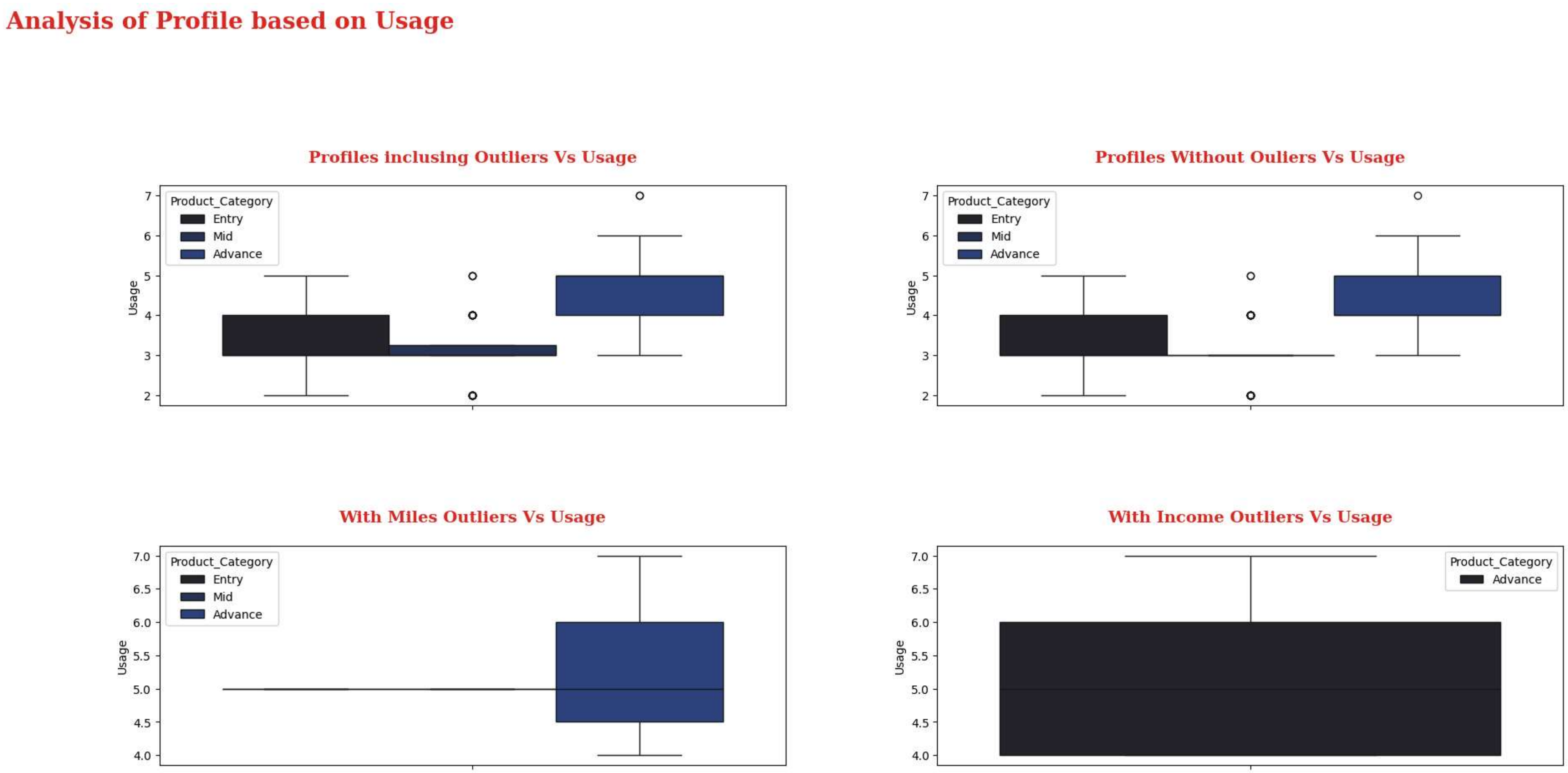
Male Partnered : The mean age group of male partnered purchased the trademill in past three months have been 30, with the average education being of 16 years, On an average male partnered plan to use Trademill 4 days a week. They have a self-rated fitness as 3.4 on an average.The mean income of Male Partnered is around $48000. They are more likely to purchase Mid level Tradmill based on Income critera. They are more likely to purchase Advance level based on Education as Critera.

Female Partnered: The mean age group that purchased the trademill in past three months have been approximately 29, with the average education being of 15 years, On an average partnered females plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3 on an average.The mean income of Female Partnered is around $53000.They are more likely to purchase Mid level or Advance Level Tradmill based on Income critera.They are more likely to purchase Entry or Mild level based on Educational Critera.

Male Single : The mean age group that purchased the trademill in past three months have been approximately 27, with the average education being of 15 years, On an average single males plan to use Trademill more than 3.6 days a week. They have a self-rated fitness as 3.3 on an average.The mean income of Male Single is around $48000.They are more likely to purchase Mid Level Tradmill based on Income critera.They are more likely to purchase Entry or Mild level based on Educational Critera.

Female Single: The mean age group that purchased the trademill in past three months have been approximately 29, with the average education being of 16 years, On an average single females plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3.3 on an average. The median income of Femal Single is around $47000.They are more likely to purchase Entry level or Mid level Tradmill based on Income critera.They are more likely to purchase Advance level based on Education as Critera.

```
fig, axis = plt.subplots(nrows=2, ncols=2, figsize=(20, 10))
fig.subplots_adjust(top=0.8)
fig.tight_layout(pad=10)
plt.title(" Analysis of Profile based on Usage ",loc='left',y=3.25,x = -1.5, fontdict= font1)
sns.boxplot(data=df, y="Usage", color = '#243E8D', palette='dark:#243E8D', hue="Product_Category", ax=axis[0,0])
sns.boxplot(data=df_woOutliers, y="Usage", color = '#243E8D', palette='dark:#243E8D', hue="Product_Category", ax=axis[0,1])
sns.boxplot(data=df_Outliers_miles, y="Usage", color = '#243E8D', palette='dark:#243E8D', hue="Product_Category", ax=axis[1,0])
sns.boxplot(data=df_Outliers_income, y="Usage", color = '#243E8D', palette='dark:#243E8D', hue="Product_Category", ax=axis[1,1])

axis[0,0].set_title("Profiles inclusing Outliers Vs Usage", pad=20, fontsize=14, fontdict= font1)
axis[0,1].set_title("Profiles Without Ouliers Vs Usage", pad=20, fontsize=14, fontdict= font1)
axis[1,0].set_title("With Miles Outliers Vs Usage", pad=20, fontsize=14, fontdict= font1)
axis[1,1].set_title("With Income Outliers Vs Usage", pad=20, fontsize=14, fontdict= font1)
plt.show()
```



The income group lying around $49000 tends to purchase Entry level. Clearly Male Partnered, Male Single, Female Single lies in this bracket

Male Partnered : The mean age group of male partnered purchased the trademill in past three months have been 30, with the average education being of 16 years, On an average male partnered plan to use Trademill 4 days a week. They have a self-rated fitness as 3.4 on an average.The mean income of Male Partnered is around $48000. They are more likely to purchase Mid level Tradmill based on Income critera. They are more likely to purchase Advance level based on Education as Critera.They are more likely to purchase Entry level based on days of Usage in a week.

Female Partnered: The mean age group that purchased the trademill in past three months have been approximately 29, with the average education being of 15 years, On an average partnered females plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3 on an average.The mean income of Female Partnered is around $53000.They are more likely to purchase Mid level or Advance Level Tradmill based on Income critera.They are more likely to purchase Entry or Mild level based on Educational Critera.They are more likely to purchase Entry level based on days of Usage in a week.

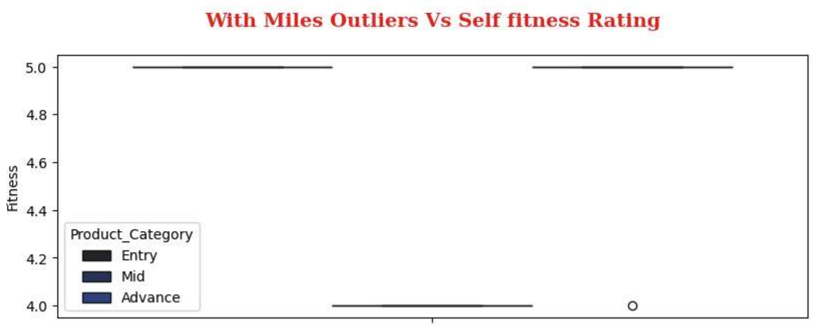
Male Single : The mean age group that purchased the trademill in past three months have been approximately 27, with the average education being of 15 years, On an average single males plan to use Trademill more than 3.6 days a week. They have a self-rated fitness as 3.3 on an average.The mean income of Male Single is around $48000.They are more likely to purchase Mid Level Tradmill based on Income critera.They are more likely to purchase Entry or Mild level based on Educational Critera.They are more likely to purchase Entry level based on days of Usage in a week.

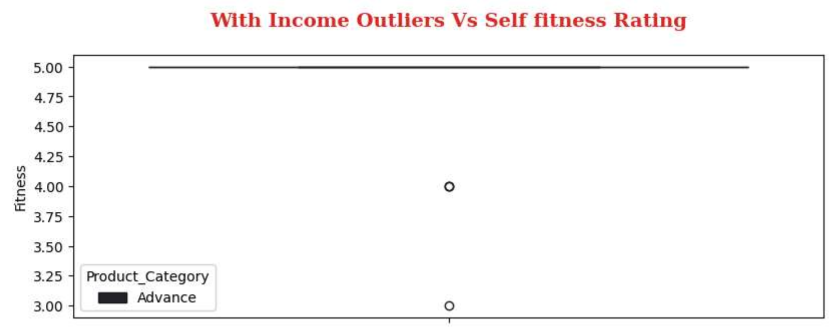
Female Single: The mean age group that purchased the trademill in past three months have been approximately 29, with the average education being of 16 years, On an average single females plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3.3 on an average. The median income of Femal Single is around $47000.They are more likely to purchase Entry level or Mid level Tradmill based on Income critera.They are more likely to purchase Advance level based on Education as Critera.They are more likely to purchase Entry level based on days of Usage in a week.

Analysis of Outliers Customer Profile:- Customers who tend to use for more than 5 days a week are more likely to purchase Advance Trademills.

```
fig, axis = plt.subplots(nrows=2, ncols=2, figsize=(20,10 ))
fig.subplots_adjust(top=0.8)
fig.tight_layout(pad=10)
plt.title(" Self Image of Fitness ",loc='left',y=3.25,x = -1.5, fontdict= font1)
sns.boxplot(data=df, y="Fitness", color = '#243E8D', palette='dark:#243E8D', hue="Product_Category", ax=axis[0,0])
sns.boxplot(data=df_woOutliers, y="Fitness", color = '#243E8D',  palette='dark:#243E8D', hue="Product_Category", ax=axis[0,1])
sns.boxplot(data=df_Outliers_miles, y="Fitness", color = '#243E8D',  palette='dark:#243E8D', hue="Product_Category", ax=axis[1,0])
sns.boxplot(data=df_Outliers_income, y="Fitness", color = '#243E8D',  palette='dark:#243E8D', hue="Product_Category", ax=axis[1,1])



axis[0,0].set_title("Profiles inclusing Outliers Vs Self fitness Rating", pad=20, fontsize=14, fontdict= font1)
axis[0,1].set_title("Profiles Without Ouliers Vs Self fitness Rating", pad=20, fontsize=14, fontdict= font1)
axis[1,0].set_title("With Miles Outliers Vs Self fitness Rating", pad=20, fontsize=14, fontdict= font1)
axis[1,1].set_title("With Income Outliers Vs Self fitness Rating", pad=20, fontsize=14, fontdict= font1)
plt.show()
```

Self Image of Fitness





The income group lying around $49000 tends to purchase Entry level. Clearly Male Partnered, Male Single, Female Single lies in this bracket

Male Partnered : The mean age group of male partnered purchased the trademill in past three months have been 30, with the average education being of 16 years, On an average male partnered plan to use Trademill 4 days a week. They have a self-rated fitness as 3.4 on an average.The mean income of Male Partnered is around $48000. They are more likely to purchase Mid level Tradmill based on Income critera. They are more likely to purchase Advance level based on Education as Critera.They are more likely to purchase Entry level based on days of Usage in a week.They are more likely to purchase Entry level or Mid Level based on self- rated fitness.

Female Partnered: The mean age group that purchased the trademill in past three months have been approximately 29, with the average education being of 15 years, On an average partnered females plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3 on an average.The mean income of Female Partnered is around $53000.They are more likely to purchase Mid level or Advance Level Tradmill based on Income critera.They are more likely to purchase Entry or Mild level based on Educational Critera.They are more likely to purchase Entry level based on days of Usage in a week.They are more likely to purchase Entry level or Mid Level based on self- rated fitness.

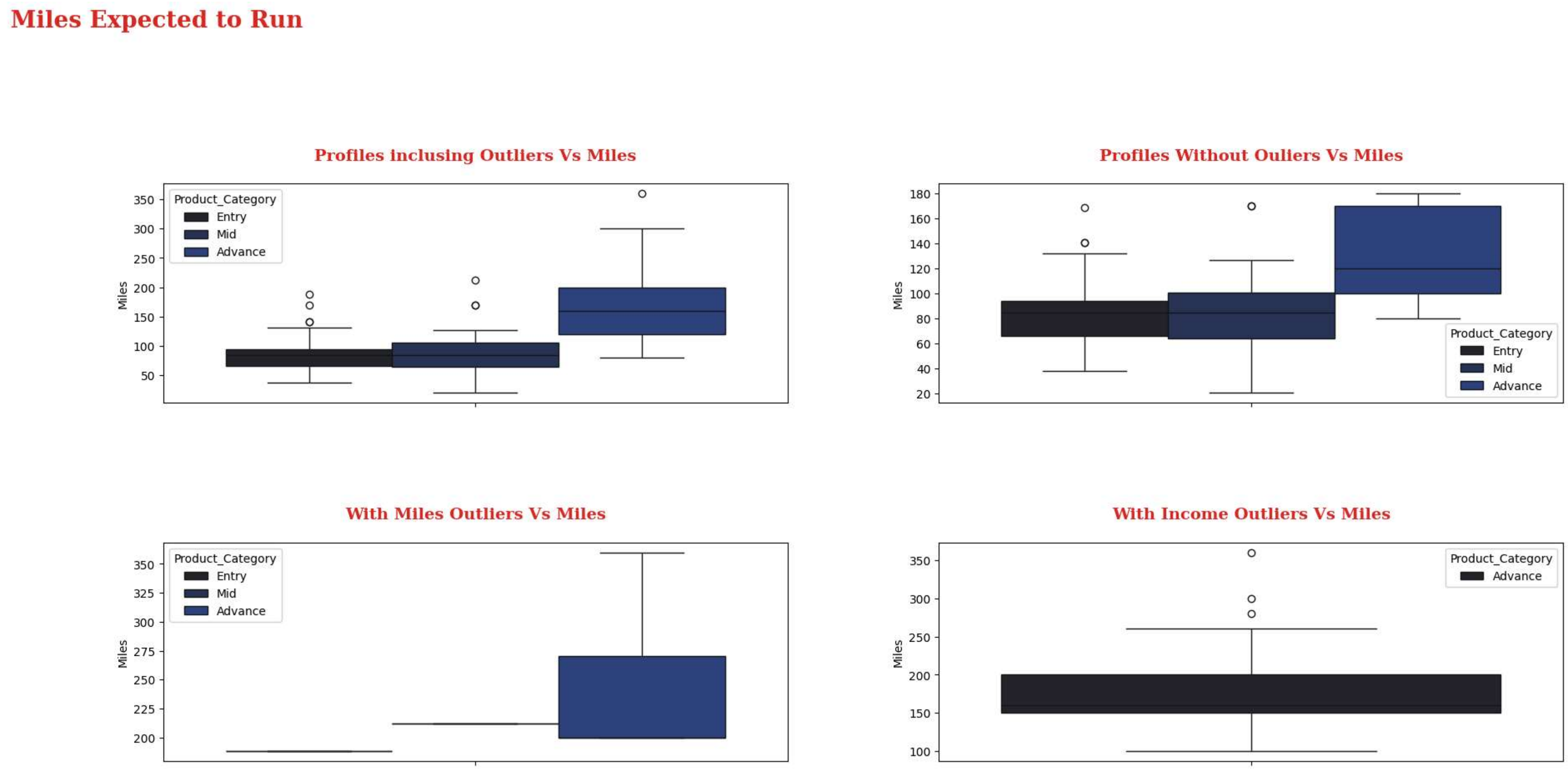
Male Single : The mean age group that purchased the trademill in past three months have been approximately 27, with the average education being of 15 years, On an average single males plan to use Trademill more than 3.6 days a week. They have a self-rated fitness as 3.3 on an average.The mean income of Male Single is around $48000.They are more likely to purchase Mid Level Tradmill based on Income critera.They are more likely to purchase Entry or Mild level based on Educational Critera.They are more likely to purchase Entry level based on days of Usage in a week. They are more likely to purchase Entry level or Mid Level based on self- rated fitness.

Female Single: The mean age group that purchased the trademill in past three months have been approximately 29, with the average education being of 16 years, On an average single females plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3.3 on an average. The median income of Femal Single is around $47000.They are more likely to purchase Entry level or Mid level Tradmill based on Income critera.They are more likely to purchase Advance level based on Education as Critera.They are more likely to purchase Entry level based on days of Usage in a week.They are more likely to purchase Entry level or Mid Level based on self- rated fitness.

Analysis of Outliers Customer Profile:-

1. Customers who tend to use for more than 5 days a week are more likely to purchase Advance Trademills.
2. Customers who tend have high self fitness rating of 4 or more are more likely to purchase Advance Trademills.

```
fig, axis = plt.subplots(nrows=2, ncols=2, figsize=(20, 10))
fig.subplots_adjust(top=0.8)
fig.tight_layout(pad=10)
plt.title(" Miles Expected to Run ",loc='left',y=3.25,x = -1.5, fontdict= font1)
sns.boxplot(data=df, y="Miles", color = '#243E8D', palette='dark:#243E8D', hue="Product_Category", ax=axis[0,0])
sns.boxplot(data=df_woOutliers, y="Miles", color = '#243E8D', palette='dark:#243E8D', hue="Product_Category", ax=axis[0,1])
sns.boxplot(data=df_Outliers_miles, y="Miles", color = '#243E8D', palette='dark:#243E8D', hue="Product_Category", ax=axis[1,0])
sns.boxplot(data=df_Outliers_income, y="Miles", color = '#243E8D',palette='dark:#243E8D', hue="Product_Category", ax=axis[1,1])

axis[0,0].set_title("Profiles inclusing Outliers Vs Miles", pad=20, fontsize=14, fontdict= font1)
axis[0,1].set_title("Profiles Without Ouliers Vs Miles", pad=20, fontsize=14, fontdict= font1)
axis[1,0].set_title("With Miles Outliers Vs Miles", pad=20, fontsize=14, fontdict= font1)
axis[1,1].set_title("With Income Outliers Vs Miles", pad=20, fontsize=14, fontdict= font1)
plt.show()
```



The income group lying around $49000 tends to purchase Entry level. Clearly Male Partnered, Male Single, Female Single lies in this bracket

Male Partnered : The mean age group of male partnered purchased the trademill in past three months have been 30, with the average education being of 16 years, On an average male partnered plan to use Trademill 4 days a week. They have a self-rated fitness as 3.4 on an average.The mean income of Male Partnered is around $48000. They are more likely to purchase Mid level Tradmill based on Income critera. They are more likely to purchase Advance level based on Education as Critera.They are more likely to purchase Entry level based on days of Usage in a week.They are more likely to purchase Entry level or Mid Level based on self- rated fitness.

Female Partnered: The mean age group that purchased the trademill in past three months have been approximately 29, with the average education being of 15 years, On an average partnered females plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3 on an average.The mean income of Female Partnered is around $53000.They are more likely to purchase Mid level or Advance Level Tradmill based on Income critera.They are more likely to purchase Entry or Mild level based on Educational Critera.They are more likely to purchase Entry level based on days of Usage in a week.They are more likely to purchase Entry level or Mid Level based on self- rated fitness.

Male Single : The mean age group that purchased the trademill in past three months have been approximately 27, with the average education being of 15 years, On an average single males plan to use Trademill more than 3.6 days a week. They have a self-rated fitness as 3.3 on an average.The mean income of Male Single is around $48000.They are more likely to purchase Mid Level Tradmill based on Income critera.They are more likely to purchase Entry or Mild level based on Educational Critera.They are more likely to purchase Entry level based on days of Usage in a week. They are more likely to purchase Entry level or Mid Level based on self- rated fitness.

Female Single: The mean age group that purchased the trademill in past three months have been approximately 29, with the average education being of 16 years, On an average single females plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3.3 on an average. The median income of Femal Single is around $47000.They are more likely to purchase Entry level or Mid level Tradmill based on Income critera.They are more likely to purchase Advance level based on Education as Critera.They are more likely to purchase Entry level based on days of Usage in a week.They are more likely to purchase Entry level or Mid Level based on self- rated fitness.

Analysis of Outliers Customer Profile:-

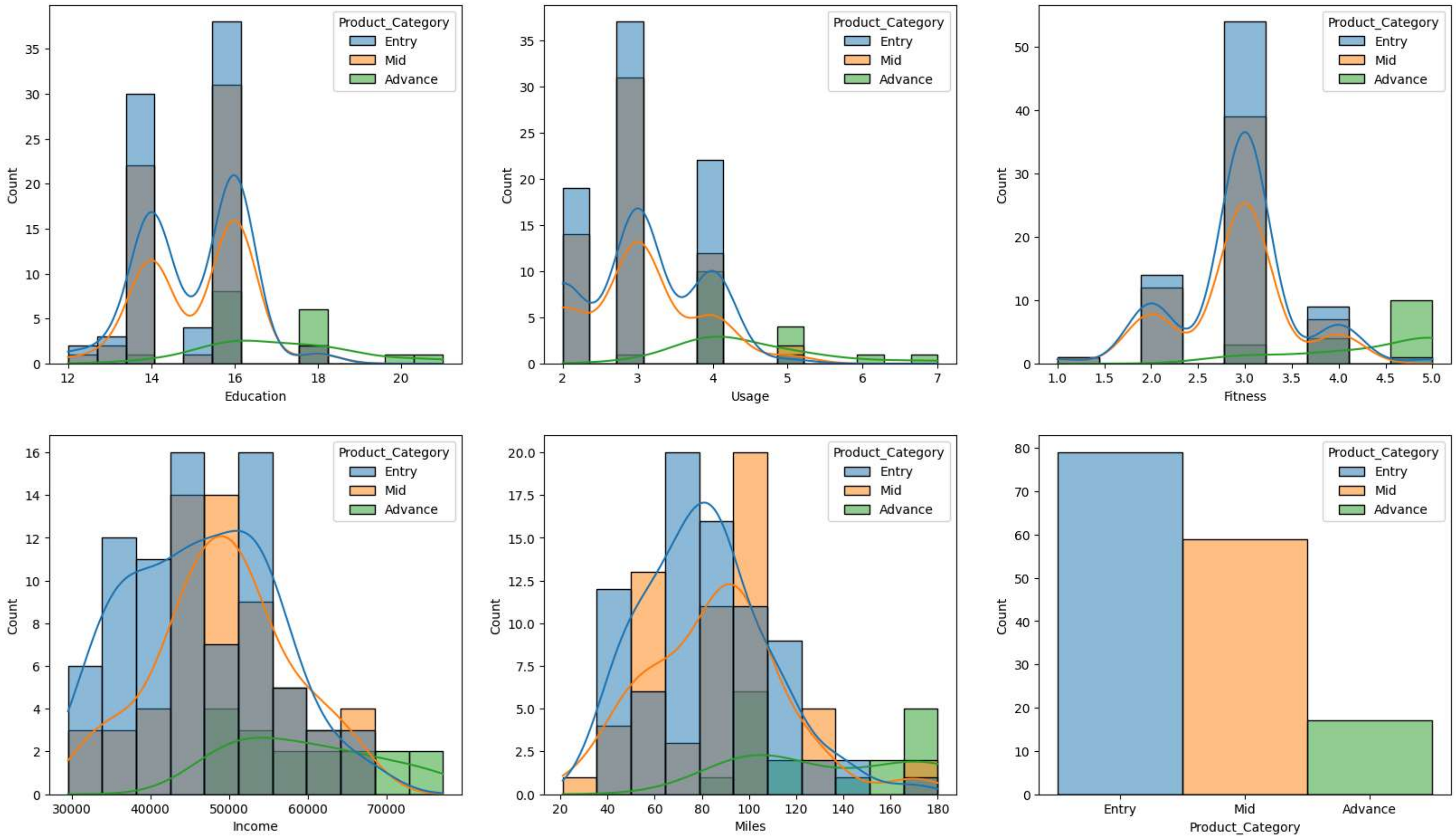
Customers who tend to use for more than 5 days a week are more likely to purchase Advance Trademills.

Customers who tend have high self fitness rating of 4 or more are more likely to purchase Advance Trademills.

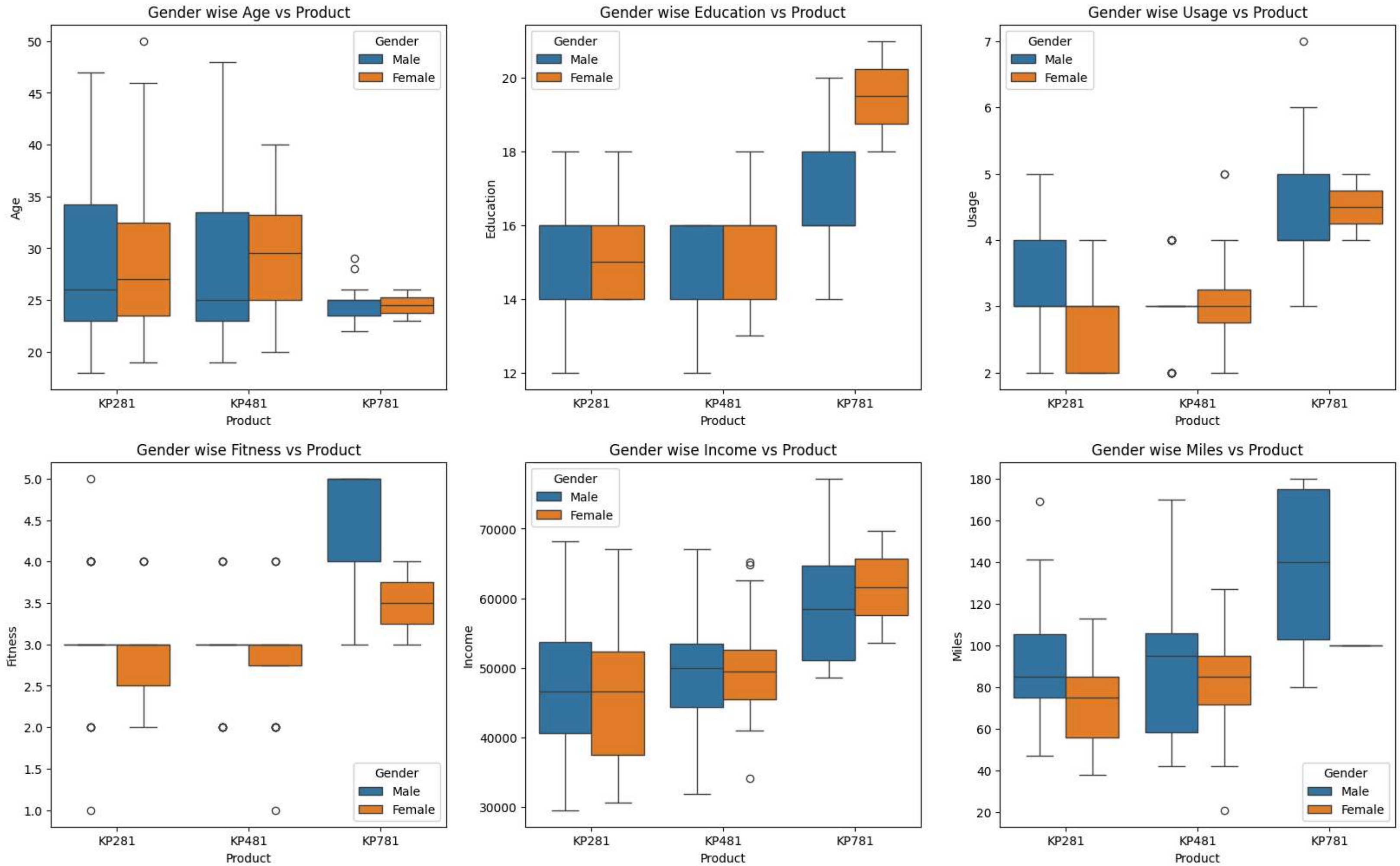
Customers who aim to run more than 150 miles are more likely to puchase Advance Trademills

```
fig, axis = plt.subplots(nrows=2, ncols=3, figsize=(20, 8))
fig.subplots_adjust(top=1.2)
plt.title("Analysis of Profiles Without Outliers ",loc='left',y=2.25,x = -2.25, fontdict= font1)
# sns.histplot(data=df_woOutliers, x="Education", color = '#243E8D', orient='h', ax=axis[0,0])
sns.histplot(data=df_woOutliers, kde=True,x="Education", color = '#243E8D', hue= "Product_Category", ax=axis[0,0])
sns.histplot(data=df_woOutliers, kde=True,x="Usage", color = '#243E8D', hue= "Product_Category",ax=axis[0,1])
sns.histplot(data=df_woOutliers, kde=True,x="Fitness", color = '#243E8D',hue= "Product_Category", ax=axis[0,2])
sns.histplot(data=df_woOutliers, kde=True,x="Income", color = '#243E8D',hue= "Product_Category", ax=axis[1,0])
sns.histplot(data=df_woOutliers, kde=True,x="Miles", color = '#243E8D',hue= "Product_Category", ax=axis[1,1])
sns.histplot(data=df_woOutliers, kde=True,x="Product_Category", color = '#243E8D',hue= "Product_Category", ax=axis[1,2])
plt.show()
```

Analysis of Profiles Without Outliers




```
variables = ["Age", "Education", "Usage", "Fitness", "Income", "Miles"]

fig, axes = plt.subplots(2, 3, figsize=(20, 12))

for i in range(2):
    for j in range(3):
        variable = variables[i * 3 + j]
        sns.boxplot(ax=axes[i, j], data=df_woOutliers, x="Product", y=variable, hue="Gender")
        axes[i, j].set_title(f"Gender wise {variable} vs Product")

plt.show();
```

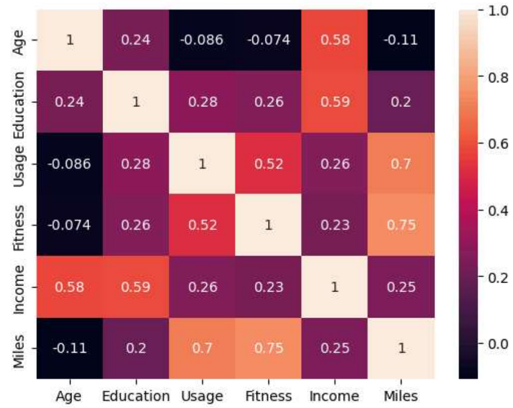



```
sns.pairplot(data=df_woOutliers[['Age','Income','Miles','Education','Fitness','Usage','Product']], hue='Product')
plt.show()
```


Age
Income
Miles
Education
Fitness
Usage
Product
KP281
KP481
KP781

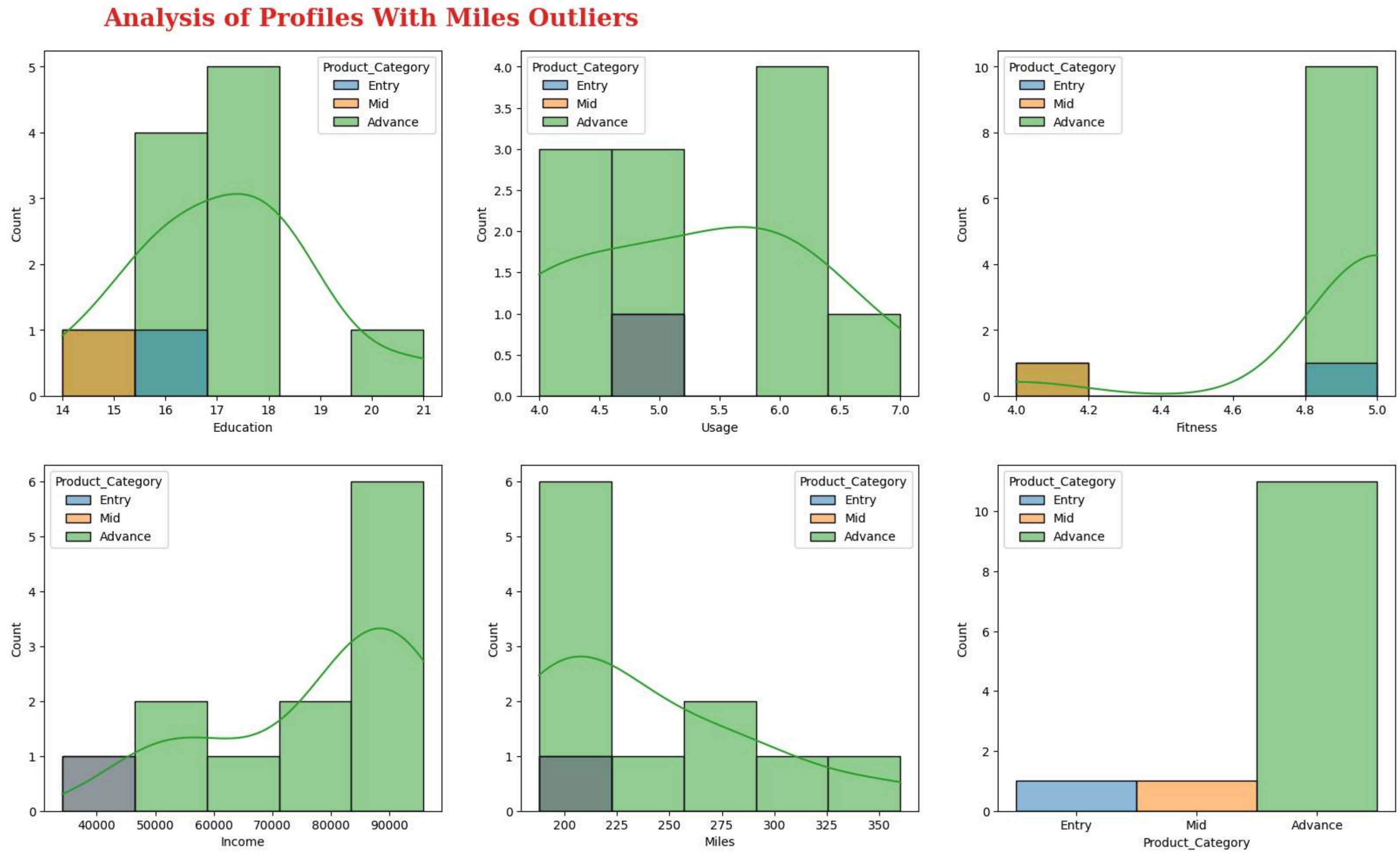
```
sns.heatmap(df_woOutliers[['Age','Education','Usage','Fitness','Income','Miles']].corr(), annot=True)
```

```
<Axes: >
```



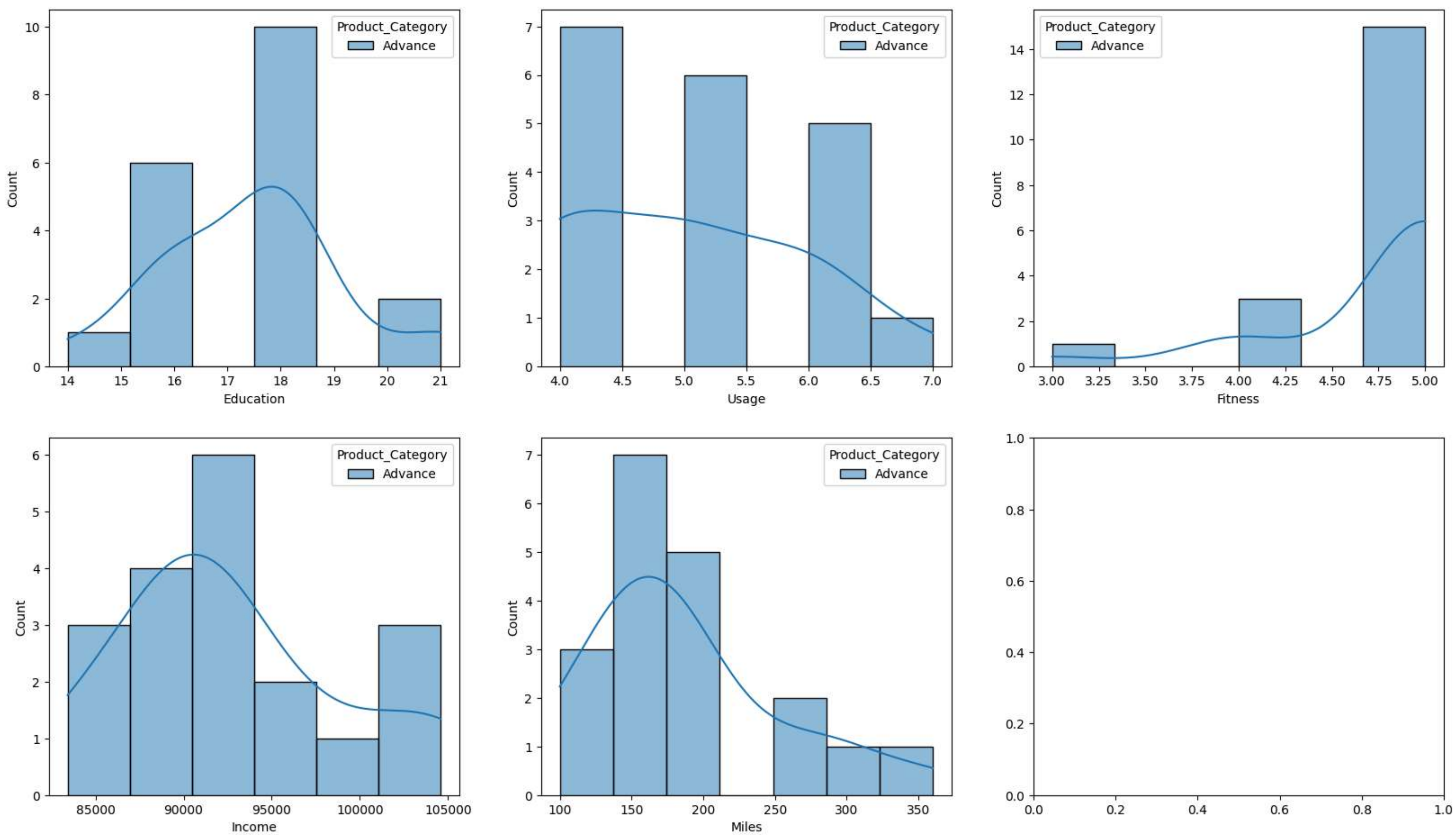
Double-click (or enter) to edit

```
fig, axis = plt.subplots(nrows=2, ncols=3, figsize=(20, 8))
fig.subplots_adjust(top=1.2)
plt.title("Analysis of Profiles With Miles Outliers ",loc='left',y=2.25,x = -2.25, fontdict= font1)
# sns.histplot(data=df_woOutliers, x="Education", color = '#243E8D', orient='h', ax=axis[0,0])
sns.histplot(data=df_Outliers_miles, kde=True,x="Education", color = '#243E8D', hue= "Product_Category", ax=axis[0,0])
sns.histplot(data=df_Outliers_miles, kde=True,x="Usage", color = '#243E8D', hue= "Product_Category", ax=axis[0,1])
sns.histplot(data=df_Outliers_miles, kde=True,x="Fitness", color = '#243E8D', hue= "Product_Category", ax=axis[0,2])
sns.histplot(data=df_Outliers_miles, kde=True,x="Income", color = '#243E8D',  hue= "Product_Category",ax=axis[1,0])
sns.histplot(data=df_Outliers_miles, kde=True,x="Miles", color = '#243E8D',  hue= "Product_Category",ax=axis[1,1])
sns.histplot(data=df_Outliers_miles,x="Product_Category", color = '#243E8D',  hue= "Product_Category",kde=True, ax=axis[1,2])
plt.show()
```




```
fig, axis = plt.subplots(nrows=2, ncols=3, figsize=(20, 8))
fig.subplots_adjust(top=1.2)
plt.title("Analysis of Profiles With Miles Outliers ",loc='left',y=2.25,x = -2.25, fontdict= font1)
# sns.histplot(data=df_woOutliers, x="Education", color = '#243E8D', orient='h', ax=axis[0,0])
sns.histplot(data=df_Outliers_income, kde=True,x="Education", color = '#243E8D', hue= "Product_Category", ax=axis[0,0])
sns.histplot(data=df_Outliers_income, kde=True,x="Usage", color = '#243E8D', hue= "Product_Category", ax=axis[0,1])
sns.histplot(data=df_Outliers_income, kde=True,x="Fitness", color = '#243E8D', hue= "Product_Category", ax=axis[0,2])
sns.histplot(data=df_Outliers_income, kde=True,x="Income", color = '#243E8D',  hue= "Product_Category",ax=axis[1,0])
sns.histplot(data=df_Outliers_income, kde=True,x="Miles", color = '#243E8D',  hue= "Product_Category",ax=axis[1,1])
# sns.histplot(data=df_Outliers_income, x="Product_Category", color = '#243E8D',  hue= "Product_Category",kde=True, ax=axis[1,2])
plt.show()
```

Analysis of Profiles With Miles Outliers



Double-click (or enter) to edit

The income group lying around $49000 tends to purchase Entry level. Clearly Male Partnered, Male Single, Female Single lies in this bracket

Male Partnered : The mean age group of male partnered purchased the trademill in past three months have been 30, with the average education being of 16 years, On an average male partnered plan to use Trademill 4 days a week. They have a self-rated fitness as 3.4 on an average.The mean income of Male Partnered is around $48000. They are more likely to purchase Mid level Tradmill based on Income critera. They are more likely to purchase Advance level based on Education as Critera.They are more likely to purchase Entry level based on days of Usage in a week.They are more likely to purchase Entry level or Mid Level based on self- rated fitness.

Female Partnered: The mean age group that purchased the trademill in past three months have been approximately 29, with the average education being of 15 years, On an average partnered females plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3 on an average.The mean income of Female Partnered is around $53000.They are more likely to purchase Mid level or Advance Level Tradmill based on Income critera.They are more likely to purchase Entry or Mild level based on Educational Critera.They are more likely to purchase Entry level based on days of Usage in a week.They are more likely to purchase Entry level or Mid Level based on self- rated fitness.

Male Single : The mean age group that purchased the trademill in past three months have been approximately 27, with the average education being of 15 years, On an average single males plan to use Trademill more than 3.6 days a week. They have a self-rated fitness as 3.3 on an average.The mean income of Male Single is around $48000.They are more likely to purchase Mid Level Tradmill based on Income critera.They are more likely to purchase Entry or Mild level based on Educational Critera.They are more likely to purchase Entry level based on days of Usage in a week. They are more likely to purchase Entry level or Mid Level based on self- rated fitness.

Female Single: The mean age group that purchased the trademill in past three months have been approximately 29, with the average education being of 16 years, On an average single females plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3.3 on an average. The median income of Femal Single is around $47000.They are more likely to purchase Entry level or Mid level Tradmill based on Income critera.They are more likely to purchase Advance level based on Education as Critera.They are more likely to purchase Entry level based on days of Usage in a week.They are more likely to purchase Entry level or Mid Level based on self- rated fitness.

Analysis of Outliers Customer Profile:-

1. Customers who tend to use for more than** 5 days a week** are more likely to purchase Advance Trademills.
2. Customers who tend have high **self fitness rating of 4 or more* *are more likely to purchase Advance Trademills.
3. Customers who aim to run more than **150 miles** are more likely to puchase Advance Trademills
4. Customers who have **income more than $85000** are more likely to purchase Advance Trademills.

Product vs Education

Customers whose Education is greater than 16, have more chances to purchase the KP781 (Advance) product. While the customers with Education less than 16 have equal chances of purchasing KP281 or KP481.

Product vs Usage

Customers who are planning to use the treadmill greater than 4 times a week, are more likely to purchase the KP781 (Advance) product. While the other customers are likely to purchasing KP281 or KP481.

Product vs Fitness

The more the customer is fit (fitness >= 3), higher the chances of the customer to purchase the KP781 (Advance) product.

Product vs Income

Higher the Income of the customer (Income >= 60000), higher the chances of the customer to purchase the KP781 (Advance) product.

Product vs Miles

If the customer expects to walk/run greater than 120 Miles per week, it is more likely that the customer will buy KP781 (Advance) product.

Marginal Probablity

```
df_woOutliers['Product_Category'].value_counts(normalize=True)
```

```
Product_Category
Entry      0.509677
Mid        0.380645
Advance    0.109677
Name: proportion, dtype: float64
```

If we ignore data of outliers, then 50 percent of people tend to purchase KP281 (Entry Level) product.

```
df_Outliers_miles['Product_Category'].value_counts(normalize=True)
```

```
Product_Category
Advance    0.846154
Entry      0.076923
Mid        0.076923
Name: proportion, dtype: float64
```

```
df_Outliers_income['Product_Category'].value_counts(normalize=True)
```

```
Product_Category
Advance    1.0
Name: proportion, dtype: float64
```

Most of the outliers goes for KP781 (Advance Level) Product. So, if a profile is outlier in terms of Income or Miles to run then customer would likely purchase Advance product.

Marginal Probablity of Usage

```
df["Usage"].value_counts(normalize=True)*100
```

```
Usage
3    38.333333
4    28.888889
2    18.333333
5     9.444444
6     3.888889
7     1.111111
Name: proportion, dtype: float64
```

```
#Creating Age and income Bins
df_new=df
bins=[10,20,30,40,50,100]
labels=["10-20","20-30","30-40","40-50","50-100"]
df_new["Age_bins"]=pd.cut(df_new["Age"],bins=bins,labels=labels)
df_new

bins1=[0,35000,60000,200000]
label1=["Lower_Income","Mid_Income","High_Income"]
df_new["Income_bins"]=pd.cut(df_new["Income"],bins=bins1,labels=label1)
df_new
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles	Product_Category	Age_bins	Income_bins
0	KP281	18	Male	14	Single	3	4	29562	112	Entry	10-20	Lower_Income
1	KP281	19	Male	15	Single	2	3	31836	75	Entry	10-20	Lower_Income
2	KP281	19	Female	14	Partnered	4	3	30699	66	Entry	10-20	Lower_Income
3	KP281	19	Male	12	Single	3	3	32973	85	Entry	10-20	Lower_Income
4	KP281	20	Male	13	Partnered	4	2	35247	47	Entry	10-20	Mid_Income
...	...	...	...	...	...	...	...	...	...	...	...	...
175	KP781	40	Male	21	Single	6	5	83416	200	Advance	30-40	High_Income
176	KP781	42	Male	18	Single	5	4	89641	200	Advance	40-50	High_Income
177	KP781	45	Male	16	Single	5	5	90886	160	Advance	40-50	High_Income
178	KP781	47	Male	18	Partnered	4	5	104581	120	Advance	40-50	High_Income
179	KP781	48	Male	18	Partnered	4	5	95508	180	Advance	40-50	High_Income

180 rows × 12 columns

Next steps: Generate code with df View recommended plots

```
#Marginal Probability education
ME=df_new["Education"].value_counts(normalize=True)*100
ME.reset_index().sort_values(["Education"],ascending=False).rename(columns={"index":"Years_of_Education","Education":"Proportion"})
```

	Proportion	proportion
6	21	1.666667
7	20	0.555556
2	18	12.777778
0	16	47.222222
3	15	2.777778
1	14	30.555556
4	13	2.777778
5	12	1.666667

```
#Marginal Probability income_bins
MI=df_new["Income_bins"].value_counts(normalize=True)*100
MI.reset_index().sort_values(["Income_bins"],ascending=False).rename(columns={"index":"Income_Bins","Income_bins":"Proportion"})
```

	Proportion	proportion
1	High_Income	23.333333
0	Mid_Income	68.888889
2	Lower_Income	7.777778

```
#Marginal Probability age_bins
MA=df_new["Age_bins"].value_counts(normalize=True)*100
MA.reset_index().sort_values(["Age_bins"],ascending=False).rename(columns={"index":"Age_Bins","Age_bins":"Proportion"})
```

	Proportion	proportion
4	50-100	0.000000
2	40-50	6.666667
1	30-40	26.666667
0	20-30	61.111111
3	10-20	5.555556

Conditional Probabilities

1.Probability of each product given gender

```
#Conditional probability - Product vs gender
cpg=pd.crosstab(df_new["Product"],df_new["Gender"],margins=True,margins_name="Total_sales",normalize=True)*100
cpg.round(1)
```

Gender Product	Female	Male	Total_sales
KP281	22.2	22.2	44.4
KP481	16.1	17.2	33.3
KP781	3.9	18.3	22.2
Total_sales	42.2	57.8	100.0

2. Probablity of each given Maritial Status

```
#Conditional probability - Product vs MaritalStatus
Cpm=pd.crosstab(df_new["Product"],df_new["MaritalStatus"],margins=True,margins_name="Total_sales",normalize=True)*100
Cpm.round(1)
```

MaritalStatus Product	Partnered	Single	Total_sales
KP281	26.7	17.8	44.4
KP481	20.0	13.3	33.3
KP781	12.8	9.4	22.2
Total_sales	59.4	40.6	100.0

3. Probablity of Product given usage

```
#Conditional probability - Product vs Usage
Cpu=pd.crosstab(df_new["Product"],df_new["Usage"],margins=True,margins_name="Total_Usage",normalize=True)*100
Cpu.round(1)
```

Usage Product	2	3	4	5	6	7	Total_Usage
KP281	10.6	20.6	12.2	1.1	0.0	0.0	44.4
KP481	7.8	17.2	6.7	1.7	0.0	0.0	33.3
KP781	0.0	0.6	10.0	6.7	3.9	1.1	22.2
Total_Usage	18.3	38.3	28.9	9.4	3.9	1.1	100.0

4. Probablity of Product given Fitness

```
#Conditional probability - Product vs Fitness
Cpf=pd.crosstab(df_new["Product"],df_new["Fitness"],margins=True,margins_name="Proportion",normalize=True)*100
Cpf.round(1)
```

Fitness Product	1	2	3	4	5	Proportion
KP281	0.6	7.8	30.0	5.0	1.1	44.4
KP481	0.6	6.7	21.7	4.4	0.0	33.3
KP781	0.0	0.0	2.2	3.9	16.1	22.2
Proportion	1.1	14.4	53.9	13.3	17.2	100.0

5. Probablity of Product given Education

```
#Conditional probability - Product vs Education
Cpe=pd.crosstab(df_new["Product"],df_new["Education"],margins=True,margins_name="Proportion",normalize=True)*100
Cpe.round(1)
```

Education Product	12	13	14	15	16	18	20	21	Proportion
KP281	1.1	1.7	16.7	2.2	21.7	1.1	0.0	0.0	44.4
KP481	0.6	1.1	12.8	0.6	17.2	1.1	0.0	0.0	33.3
KP781	0.0	0.0	1.1	0.0	8.3	10.6	0.6	1.7	22.2
Proportion	1.7	2.8	30.6	2.8	47.2	12.8	0.6	1.7	100.0

6. Probablity of product given Income

```
#Conditional probability - Product vs Income_bins
Cpi=pd.crosstab(df_new["Product"],df_new["Income_bins"],margins=True,margins_name="Proportion",normalize=True)*100
Cpi.round(1)
```

Income_bins Product	Lower_Income	Mid_Income	High_Income	Proportion
KP281	4.4	36.7	3.3	44.4
KP481	3.3	26.1	3.9	33.3
KP781	0.0	6.1	16.1	22.2
Proportion	7.8	68.9	23.3	100.0

The product portfolio is as follows:-

1. The KP281 is an entry-level treadmill that sells for $1,500.
2. The KP481 is for mid-level runners that sell for $1,750.
3. The KP781 treadmill is having advanced features that sell for $2,500.

Let us catagorically divide the product into entry, mid and advance levels.

Insights :-

- Break up of sales based on product
  KP281(44.4%),KP481(33.3%),KP781(22.2%) with outliers and
  KP281(50.96%),KP481(38.06%),KP781(10.9%) without outliers.
- Breakup of sales based on Gender level - Male(57.8%),Female(42.2%)
- Breakup of sales based on marital Status - Partnered(59.4%),single(40.6%)
- KP281 is the most preffered product in terms of both the gender
- KP281 is the most preffered product by partners
- KP781 is bought by individuals and partnered are of higher income bracket
- Females and partners usage is the most with the product KP781
- Among the partners - male influenced to buy 62 treadmillsoverall,females influenced to buy 46 treadmills overall
- Age between 20-25 in both male and female category tend to buy the most treadmills, irrespective of the product
- In partners, Male tends to influence the most to buy treadmills in KP481,KP781
- Partners with higher usage tend to buy the KP781 the more
- Education,Fitness, Income, Usage, and Miles have big impact on sales of KP781

Customer Profiles and Likely Product Purchased

The income group lying around $49000 tends to purchase Entry level. Clearly Male Partnered, Male Single, Female Single lies in this bracket

Male Partnered : The mean age group of male partnered purchased the trademill in past three months have been 30, with the average education being of 16 years, On an average male partnered plan to use Trademill 4 days a week. They have a self-rated fitness as 3.4 on an average.The mean income of Male Partnered is around $48000. They are more likely to purchase Mid level Tradmill based on Income critera. They are more likely to purchase Advance level based on Education as Critera.They are more likely to purchase Entry level based on days of Usage in a week.They are more likely to purchase Entry level or Mid Level based on self- rated fitness.

Female Partnered: The mean age group that purchased the trademill in past three months have been approximately 29, with the average education being of 15 years, On an average partnered females plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3 on an average.The mean income of Female Partnered is around $53000.They are more likely to purchase Mid level or Advance Level Tradmill based on Income critera.They are more likely to purchase Entry or Mild level based on Educational Critera.They are more likely to purchase Entry level based on days of Usage in a week.They are more likely to purchase Entry level or Mid Level based on self- rated fitness.

Male Single : The mean age group that purchased the trademill in past three months have been approximately 27, with the average education being of 15 years, On an average single males plan to use Trademill more than 3.6 days a week. They have a self-rated fitness as 3.3 on an average.The mean income of Male Single is around $48000.They are more likely to purchase Mid Level Tradmill based on Income critera.They are more likely to purchase Entry or Mild level based on Educational Critera.They are more likely to purchase Entry level based on days of Usage in a week. They are more likely to purchase Entry level or Mid Level based on self- rated fitness.

Female Single: The mean age group that purchased the trademill in past three months have been approximately 29, with the average education being of 16 years, On an average single females plan to use Trademill more than 3 days a week. They have a self-rated fitness as 3.3 on an average. The median income of Femal Single is around $47000.They are more likely to purchase Entry level or Mid level Tradmill based on Income critera.They are more likely to purchase Advance level based on Education as Critera.They are more likely to purchase Entry level based on days of Usage in a week.They are more likely to purchase Entry level or Mid Level based on self- rated fitness.

Analysis of Outliers Customer Profile:-

1. Customers who tend to use for more than** 5 days a week** are more likely to purchase Advance Trademills.
2. Customers who tend have high **self fitness rating of 4 or more* *are more likely to purchase Advance Trademills.

3. Customers who aim to run more than **150 miles** are more likely to puchase Advance Trademills
4. Customers who have **income more than $85000** are more likely to purchase Advance Trademills.

Product Type and likely Customer Profile

KP281 Age : Around 28, but under 35 Income : Less than 50000 Fitness: under 3 Miles : Under 90 Usage: 3-4 Education : less than 16 Marital Status : Both, but targeted more towards Partnered (60% Probability) Gender: Both

KP481 Age: Around 28, but under 35 Income : If Partnered then around 50000 else less than 50000 Education : less than 16 Fitness: under 3 Miles : Around 100 Usage: 3 Marital Status : Both, but targeted more towards Partnered (60% Probability) Gender: Both, but targeted more towards Male (51.7% Probability)

KP781 Age: Under 30 Income : Above 60000 Fitness: Above 3 Education : Above 16 Usage : Above 4 Miles : Above 120 Gender: Male (82.7% Probability) Maritial Status : Both, but targeted more towards Partnered (57% Probability)

Business Insights

1. KP281 & KP481 products preferred by almost similar Characteristics and KP281 is most sold product, we can promote KP481 products more by adding some features from KP781(top level) . This will increase the sales of KP481(mid level) as users will get better deal by paying little extra.
2. Discount on KP781 can be given to female users to increase the sale.
3. Gamification of fitness score, miles covered should be done. This will lead to increase in usage.
4. The marketing strategy should be focused, targeted to partnered, Males, High Income group people. Event on Valentines Day, Marathons on National Importance Days, Marketing by Celebritiy Business Tycoon could be probable marketing strategies.
5. Female Singles could be given discount or extra goodies. This may lead to increase in Sales.
6. KP281 could be made cheaper by reducing some adjunct functions or cost cutting, ensuring quality reamains the same. Therefore catering to the purchasing ability of lower income group.

```
fig, axis = plt.subplots(nrows=2, ncols=3, figsize=(20, 8))
fig.subplots_adjust(top=1.2)
plt.title("Analysis of Profiles With Miles Outliers ",loc='left',y=2.25,x = -2.25, fontdict= font1)
# sns.histplot(data=df_woOutliers, x="Education", color = '#243E8D', orient='h', ax=axis[0,0])
sns.histplot(data=df_Outliers_miles, kde=True,x="Education", color = '#243E8D', hue= "Product_Category", ax=axis[0,0])
sns.histplot(data=df_Outliers_miles, kde=True,x="Usage", color = '#243E8D', hue= "Product_Category", ax=axis[0,1])
sns.histplot(data=df_Outliers_miles, kde=True,x="Fitness", color = '#243E8D', hue= "Product_Category", ax=axis[0,2])
sns.histplot(data=df_Outliers_miles, kde=True,x="Income", color = '#243E8D',  hue= "Product_Category",ax=axis[1,0])
sns.histplot(data=df_Outliers_miles, kde=True,x="Miles", color = '#243E8D',  hue= "Product_Category",ax=axis[1,1])
sns.histplot(data=df_Outliers_miles,x="Product_Category", color = '#243E8D',  hue= "Product_Category",kde=True, ax=axis[1,2])
plt.show()
```