Aerofit CaseStudy

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.

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

In [ ]:
```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

In [ ]:
```
# loading the dataset
df = pd.read_csv('/content/drive/MyDrive/aerofit_treadmill.txt')
df
```

Out[ ]:

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

180 rows × 9 columns

Defining Problem Statement and Analysing basic metrics

We 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.

- Perform descriptive analytics to create a customer profile for each AeroFit treadmill product by developing appropriate tables and charts.
- 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.

Analysis of basic metrics:

Checking the shape of the dataset

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

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

Checking the structure and characteristics of the dataset

In [ ]:
```
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
```

Checking the missing and duplicate values in the dataset

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

Out[ ]:

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

dtype: int64

There are no missing values in the dataset

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

Out[ ]:
```
np.int64(0)
```

In [ ]:
```

```

There are no duplicated values in the dataset

Statistical summary

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

Out[ ]:

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

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

Out[ ]:

	Age	Education	Usage	Fitness	Income	Mile
count	180.000000	180.000000	180.000000	180.000000	180.000000	180.00000
mean	28.788889	15.572222	3.455556	3.311111	53719.577778	103.19444
std	6.943498	1.617055	1.084797	0.958869	16506.684226	51.86360
min	18.000000	12.000000	2.000000	1.000000	29562.000000	21.00000
25%	24.000000	14.000000	3.000000	3.000000	44058.750000	66.00000
50%	26.000000	16.000000	3.000000	3.000000	50596.500000	94.00000
75%	33.000000	16.000000	4.000000	4.000000	58668.000000	114.75000
max	50.000000	21.000000	7.000000	5.000000	104581.000000	360.00000

Insights:

- There are 3 unique types of products in which KP281 have the highest sale
- Male purchase the product more frequently than female
- Most of the customers are married

Non-Graphical Analysis: Value counts and unique attributes

In [ ]:
```
print("Total count of Product:")
print(df["Product"].value_counts())

print("\nUnique attributes of Product:")
print(df["Product"].unique())
```

```
Total count of Product:
Product
KP281    80
KP481    60
KP781    40
Name: count, dtype: int64

Unique attributes of Product:
['KP281' 'KP481' 'KP781']
```

In [ ]:
```
print("Total count of Age:")
print(df["Age"].value_counts())

print("\nUnique attributes of Age:")
print(df["Age"].unique())
```

```
Total count of Age:
Age
25    25
23    18
24    12
26    12
28     9
33     8
35     8
22     7
30     7
27     7
38     7
21     7
31     6
34     6
29     6
20     5
40     5
19     4
32     4
37     2
45     2
48     2
47     2
18     1
41     1
39     1
36     1
43     1
46     1
44     1
50     1
42     1
Name: count, dtype: int64

Unique attributes 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]
```

In [ ]:
```
print("Total count of Gender:")
gender_count=df["Gender"].value_counts()
print(gender_count)

print("\nUnique attributes of Gender:")
gender_unique=df["Gender"].unique()
print(gender_unique)
```

```
Total count of Gender:
Gender
Male      104
Female     76
Name: count, dtype: int64

Unique attributes of Gender:
['Male' 'Female']
```

In [ ]:
```
print("Total count of Education:")
education=df["Education"].value_counts()
print(education)

print("\nUnique attributes of Education:")
Education=df["Education"].unique()
print(Education)
```

```
Total 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 attributes of Education:
[14 15 12 13 16 18 20 21]
```

In [ ]:
```
print("Total count of MaritalStatus:")
marital_count=df["MaritalStatus"].value_counts()
print(marital_count)

print("\nUnique attributes of MaritalStatus:")
martial_unique=df["MaritalStatus"].unique()
print(martial_unique)
```

```
Total count of MaritalStatus:
MaritalStatus
Partnered    107
Single        73
Name: count, dtype: int64

Unique attributes of MaritalStatus:
['Single' 'Partnered']
```

Insights:

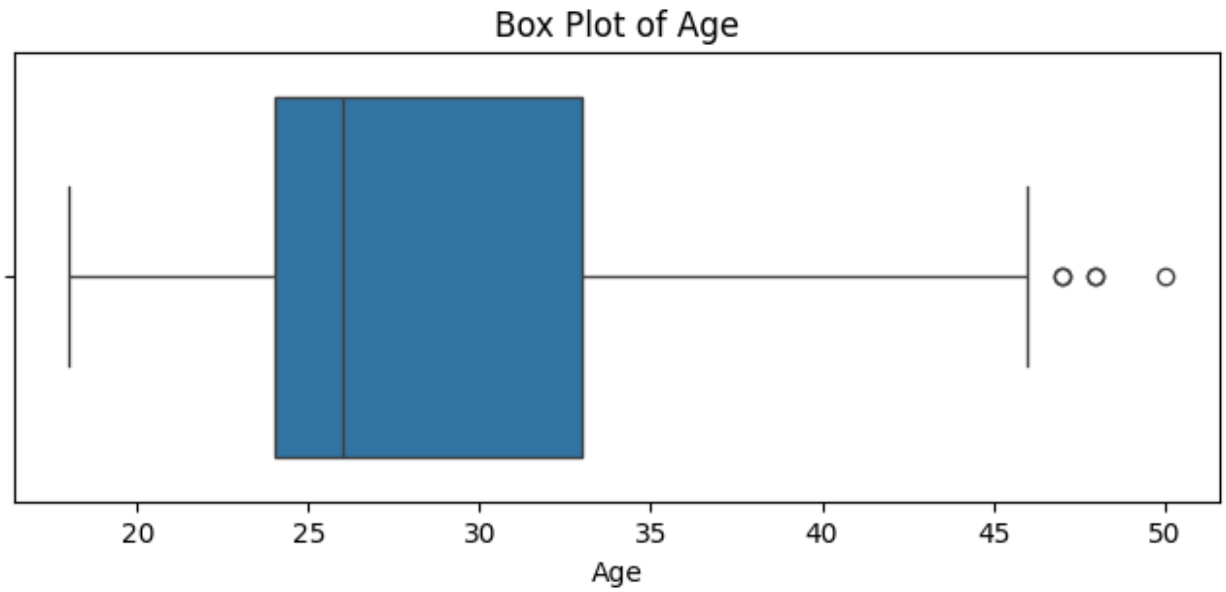
- KP281, KP481, KP781 are the 3 different products
- Most commonly purchased treadmill product type is KP281
- There are 32 unique ages

- 104 Males and 76 Females are in the customers list
- 8 unique set of Educations (14, 15, 12, 13, 16, 18, 20, 21)
- Highest rated Fitness rating is 3
- Most customers usage treadmill atleast 3 days per week
- Majority of the customers who have purchased are Married/Partnered

Missing Value & Outlier Detection

Checking for Outliers

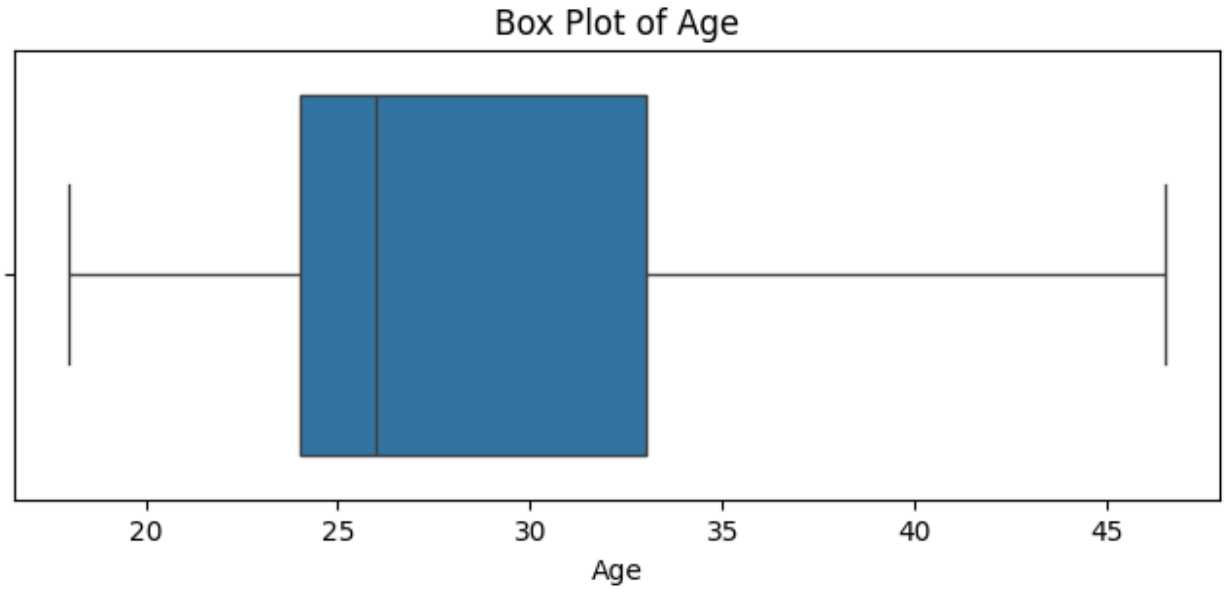
In [ ]:
```
plt.figure(figsize=(8, 3))
sns.boxplot(x="Age", data = df)
plt.xlabel('Age')
plt.title('Box Plot of Age')
plt.show()
```



In [ ]:
```
q1 = df["Age"].quantile(0.25)
q3 = df["Age"].quantile(0.75)
iqr = q3-q1
upperlimit = q3+(1.5*iqr)
lowerlimit = q1-(1.5*iqr)
#create function for condition
def outlier_limit (value):
  if value > upperlimit:
    return upperlimit
  if value < lowerlimit:
    return lowerlimit
  else:
    return value
# Apply def
```

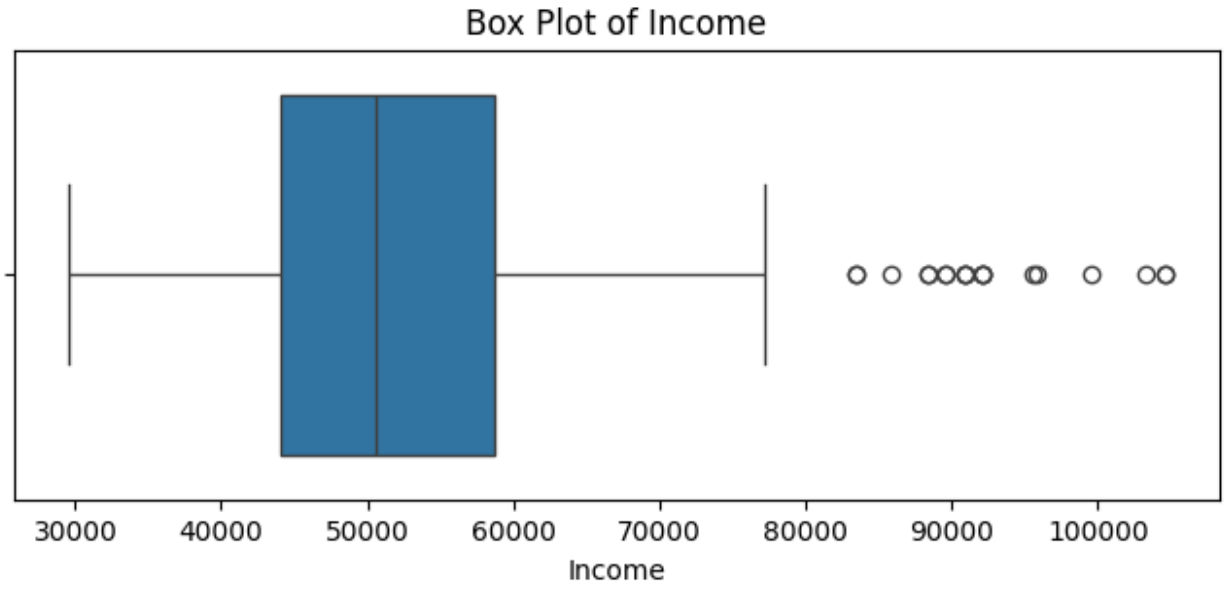
```
df["Age"] = df["Age"].apply(outlier_limit)
```

In [ ]:
```
plt.figure(figsize=(8, 3))
sns.boxplot(x="Age", data = df)
plt.xlabel('Age')
plt.title('Box Plot of Age')
plt.show()
```



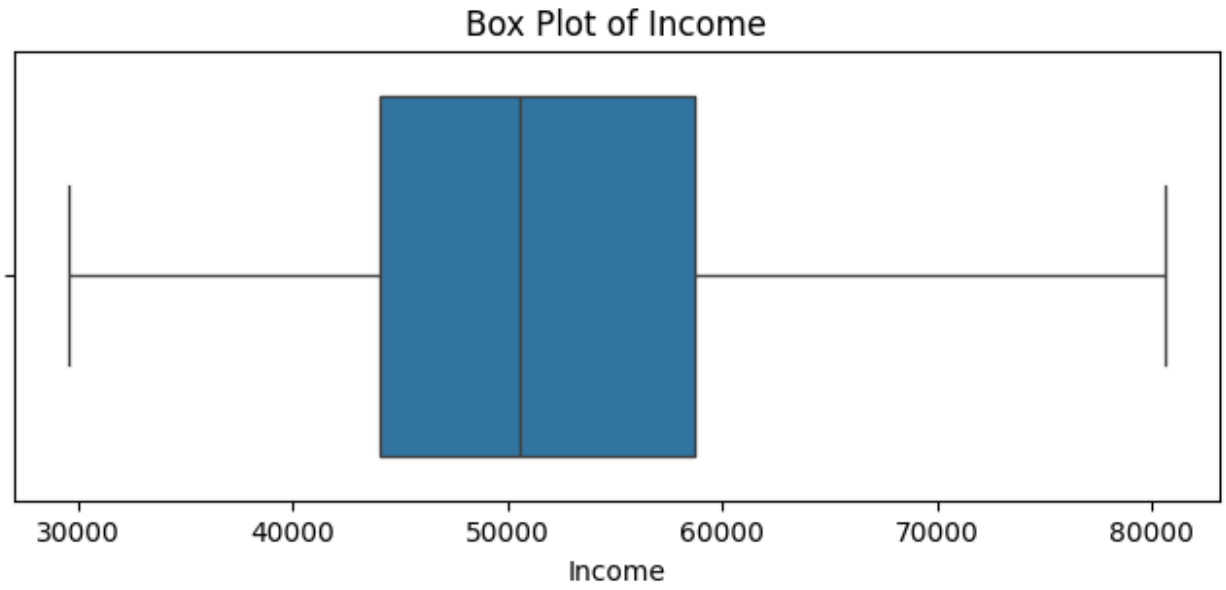
Boxplot of Income to find outlier

In [ ]:
```
plt.figure(figsize=(8, 3))
sns.boxplot(x="Income", data = df)
plt.xlabel('Income')
plt.title('Box Plot of Income')
plt.show()
```



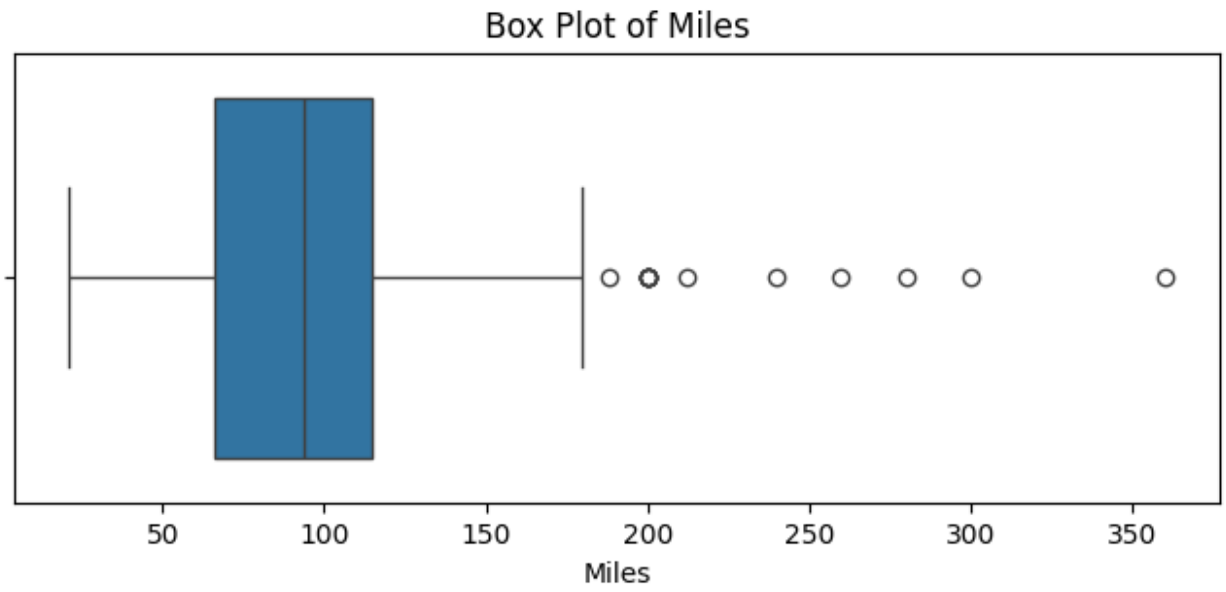
In [ ]:
```
q1 = df["Income"].quantile(0.25)
q3 = df["Income"].quantile(0.75)
iqr = q3-q1
upperlimit = q3+(1.5*iqr)
lowerlimit = q1-(1.5*iqr)
#create function for condition
def outlier_limit(value):
  if value > upperlimit:
    return upperlimit
  if value < lowerlimit:
    return lowerlimit
  else:
    return value
# Apply def
df["Income"] = df["Income"].apply(outlier_limit)
```

In [ ]:
```
plt.figure(figsize=(8, 3))
sns.boxplot(x="Income", data = df)
plt.xlabel('Income')
plt.title('Box Plot of Income')
plt.show()
```



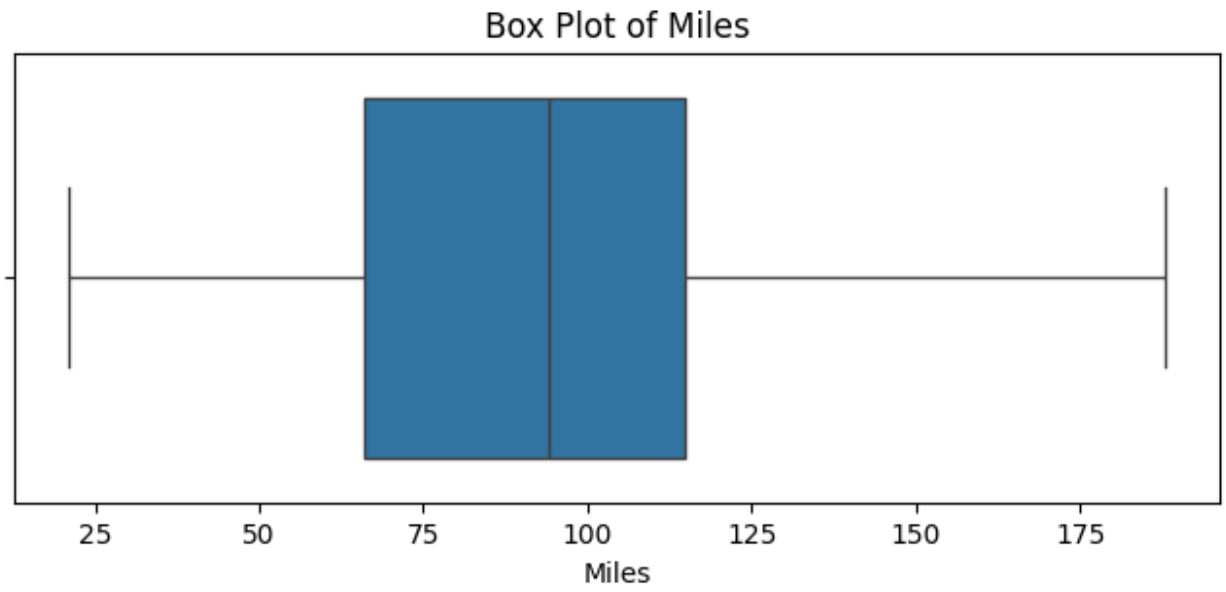
Boxplot of Miles to find Outliers

In [ ]:
```
plt.figure(figsize=(8, 3))
sns.boxplot(x= "Miles", data = df)
plt.xlabel('Miles')
plt.title('Box Plot of Miles')
plt.show()
```



In [ ]:
```
q1 = df["Miles"].quantile(0.25)
q3 = df["Miles"].quantile(0.75)
iqr = q3-q1
upperlimit = q3+(1.5*iqr)
lowerlimit = q1-(1.5*iqr)
#create function for condition
def outlier_limit(value):
  if value > upperlimit:
    return upperlimit
  if value < lowerlimit:
    return lowerlimit
  else:
    return value
# Apply def
df["Miles"] = df["Miles"].apply(outlier_limit)
```

In [ ]:
```
plt.figure(figsize=(8, 3))
sns.boxplot(x= "Miles", data = df)
plt.xlabel('Miles')
plt.title('Box Plot of Miles')
plt.show()
```



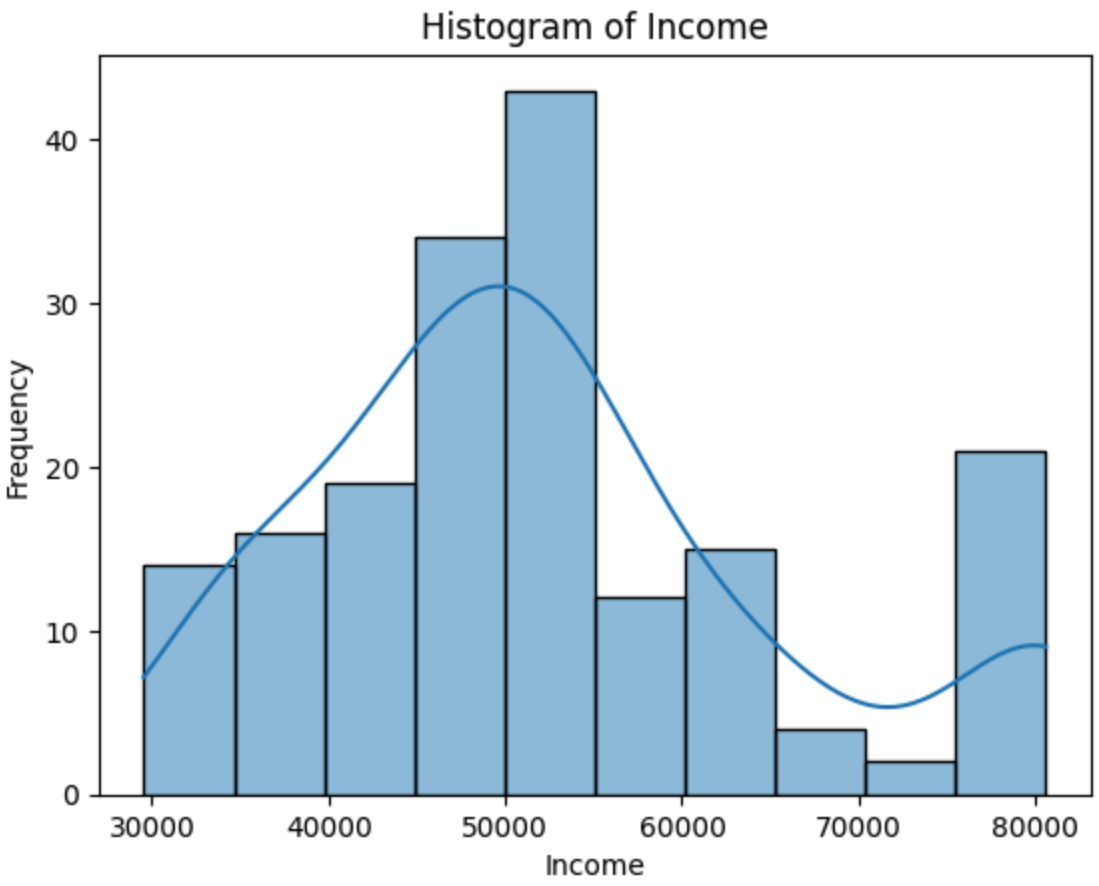
- Outliers of miles are checked and Treated
- Box plot for all the Variables that can have Outliers are checked and treated

Visual Analysis - Univariate & Bivariate

Income Distribution

In [ ]:
```
sns.histplot(df["Income"], kde=True)
plt.xlabel('Income')
plt.ylabel('Frequency')
plt.title('Histogram of Income')
```

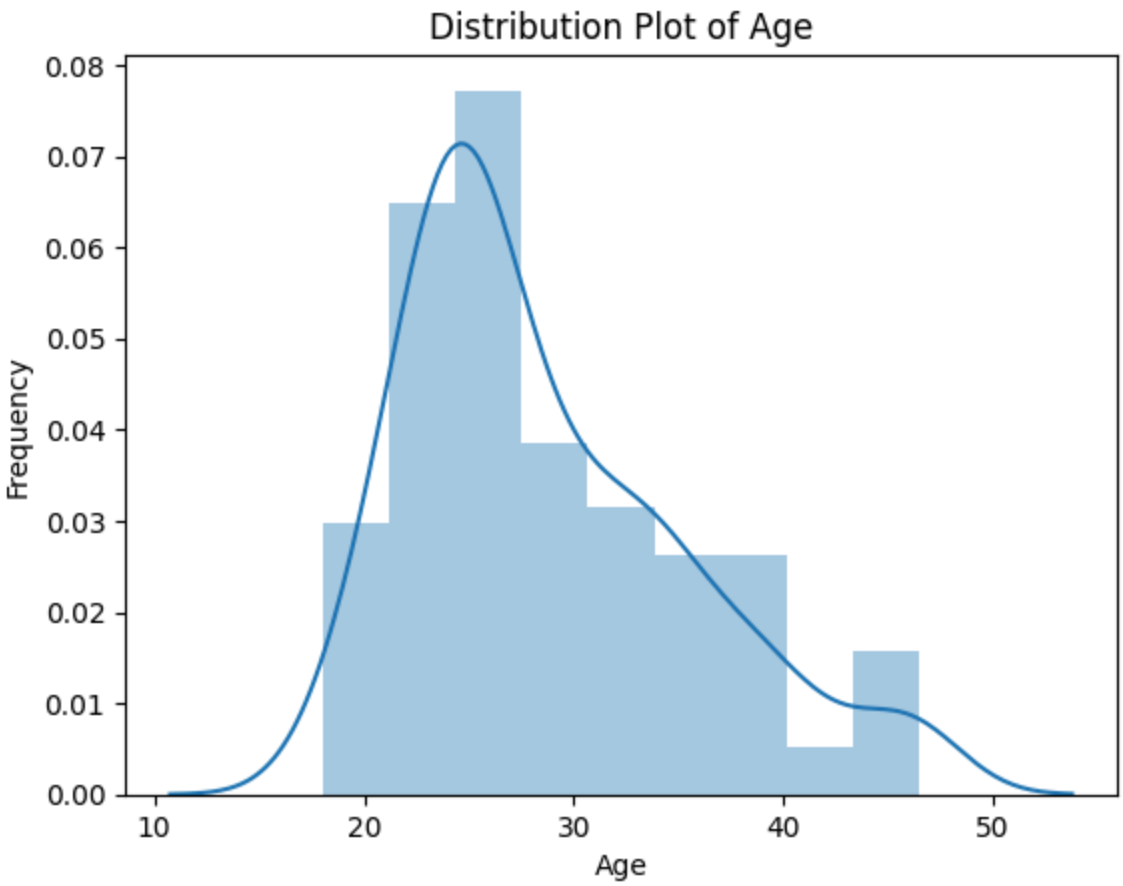
Out[ ]:
```
Text(0.5, 1.0, 'Histogram of Income')
```



- Most of customers who have purchased the product have a average income between 40K to 60K
- Average Income density is over 3.0

Age Distribution

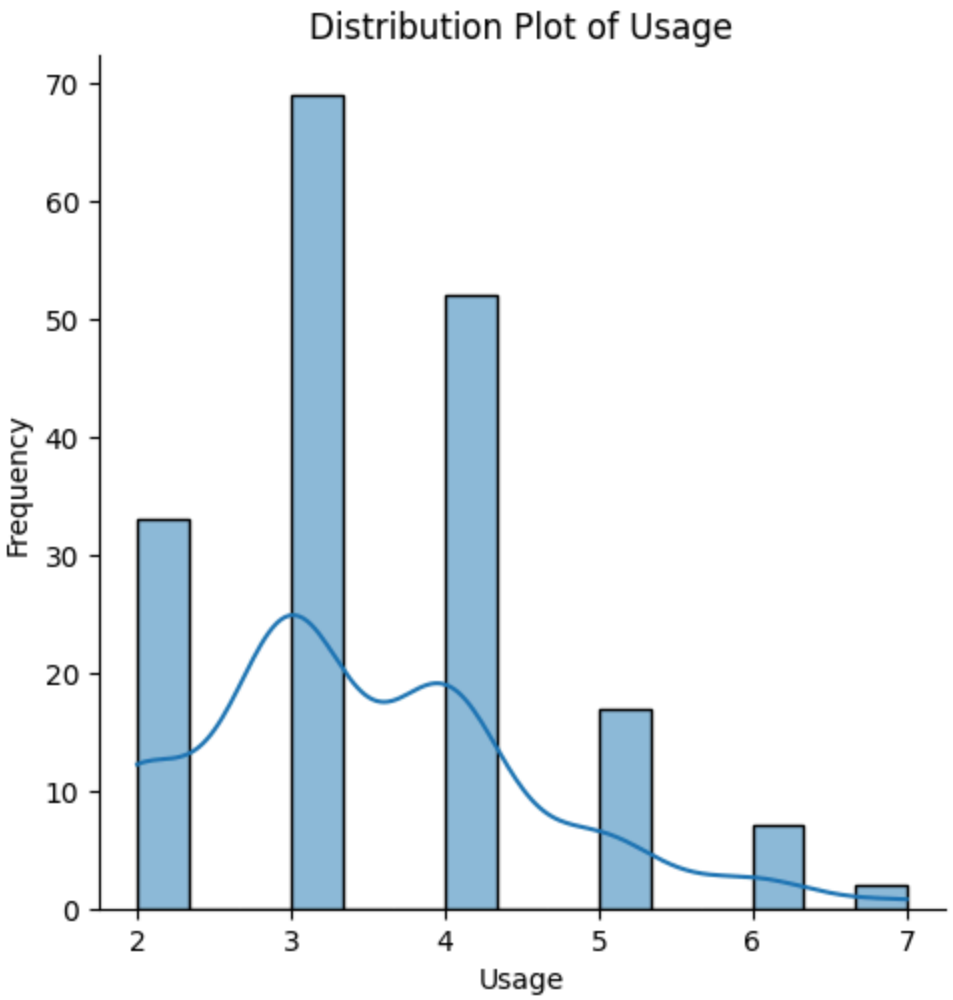
In [ ]:
```
sns.distplot(df["Age"])
plt.xlabel('Age')
plt.ylabel('Frequency')
plt.title('Distribution Plot of Age')
plt.show()
```



- Person in age range 22-28 have the highest tendency to purchase the Product

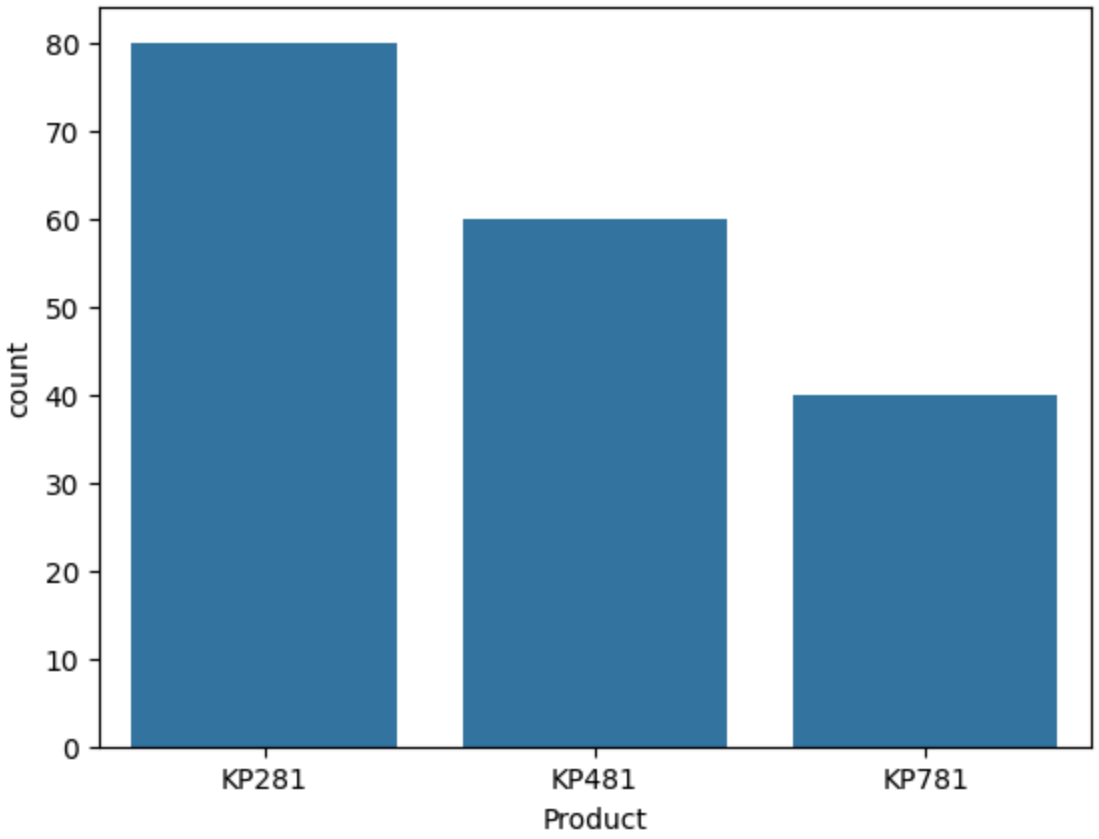
Usage Distribution

In [ ]:
```
sns.displot(df["Usage"], kde=True);
plt.xlabel('Usage')
plt.ylabel('Frequency')
plt.title('Distribution Plot of Usage')
plt.show()
```



- 3 days per week is the most common usage among the customers
- 4 days and 2 days per week is the second and third highest usage among the customers
- Very few customers use product 7 days per week

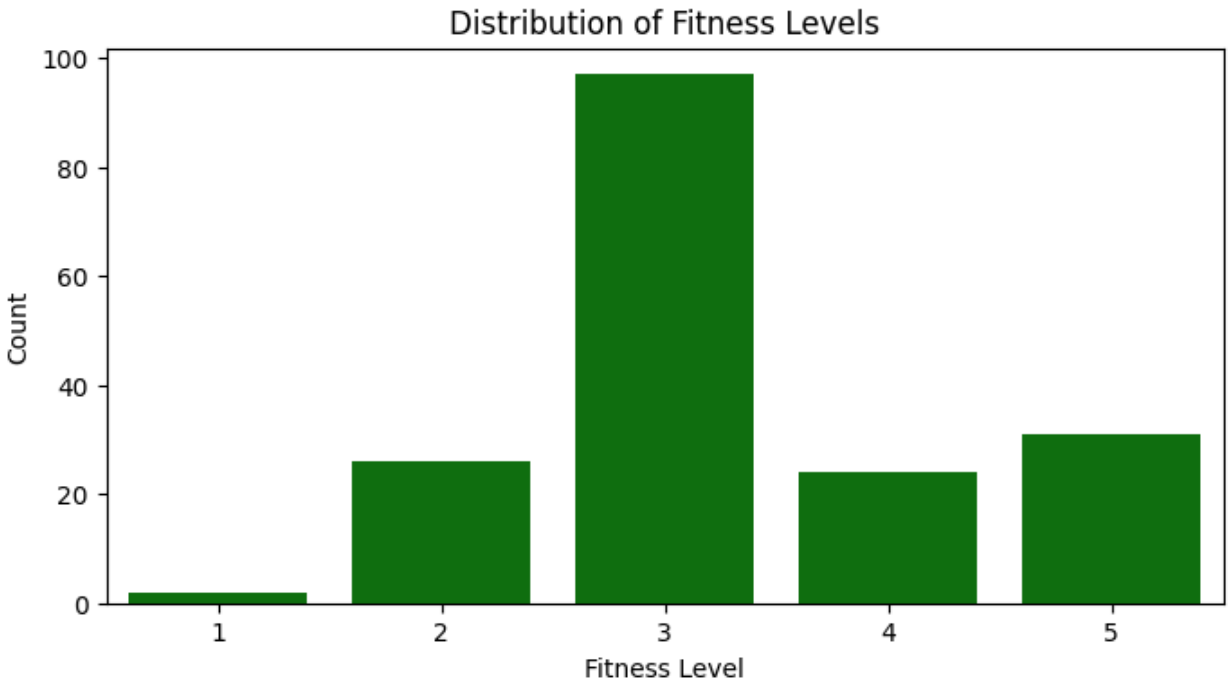
In [ ]:
```
# Product Analysis - count plot
sns.countplot(data=df,x='Product')
plt.show()
```



- KP281 is the most commonly purchase product type
- KP481 is the second most top product type purchased
- KP781 is the least purchased product type

Fitness Analysis-countplot

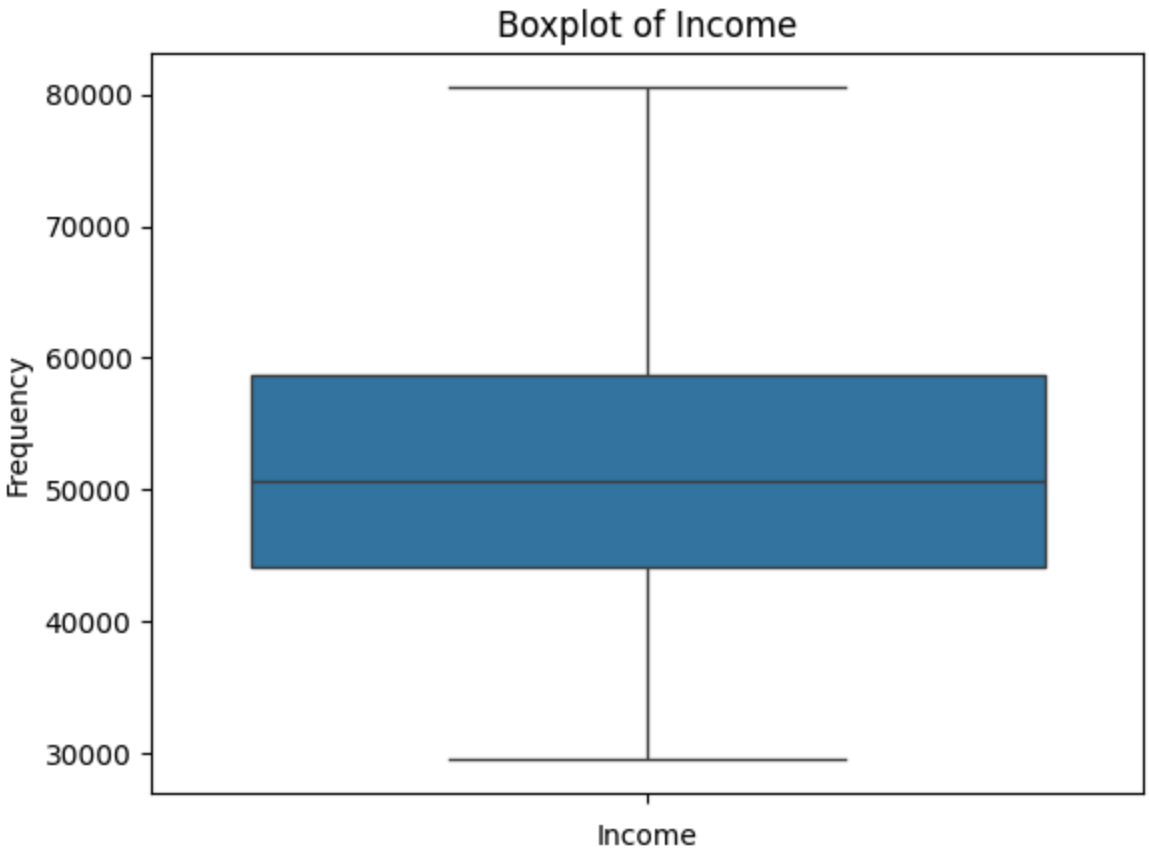
In [ ]:
```
plt.figure(figsize=(8, 4))

# Countplot with color palette
sns.countplot(data=df, x="Fitness", color="green");

# Title and labels
plt.title('Distribution of Fitness Levels')
plt.xlabel('Fitness Level')
plt.ylabel('Count')

# Show plot
plt.show()
```



- More than 90 customers have rated their physical fitness rating as Average
- Excellent shape is the second highest rating provided by the customers

Income Analysis - Box plot

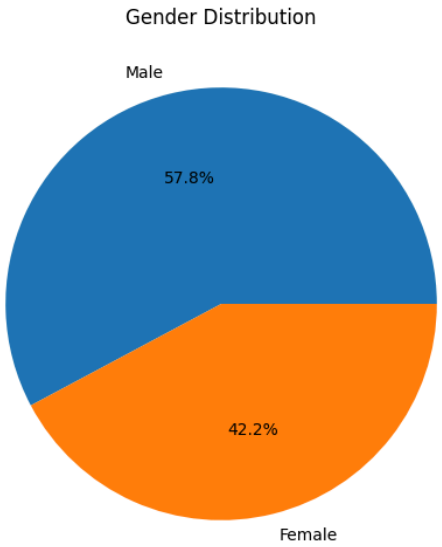
In [ ]:
```
sns.boxplot(data=df, y='Income');
plt.title('Boxplot of Income')
plt.xlabel('Income')
plt.ylabel('Frequency')
plt.show()
```

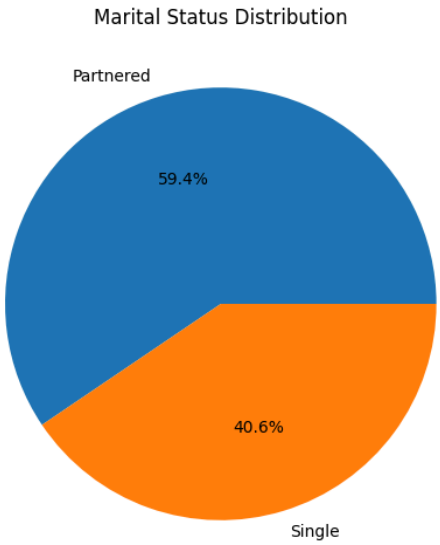


Most customers earn from 45K to around 60K per annum

Gender and Marital Status Disribution

In [ ]:
```
plt.figure(figsize=(14, 6))

# Create the first subplot for gender distribution
plt.subplot(1, 2, 1)
plt.pie(gender_count.values, labels=gender_count.index, autopct='%1.1f%%')
plt.title('Gender Distribution')

# Create the second subplot for marital status distribution
plt.subplot(1, 2, 2)
plt.pie(marital_count.values, labels=marital_count.index, autopct='%1.1f%%')
plt.title('Marital Status Distribution')


plt.show()
```



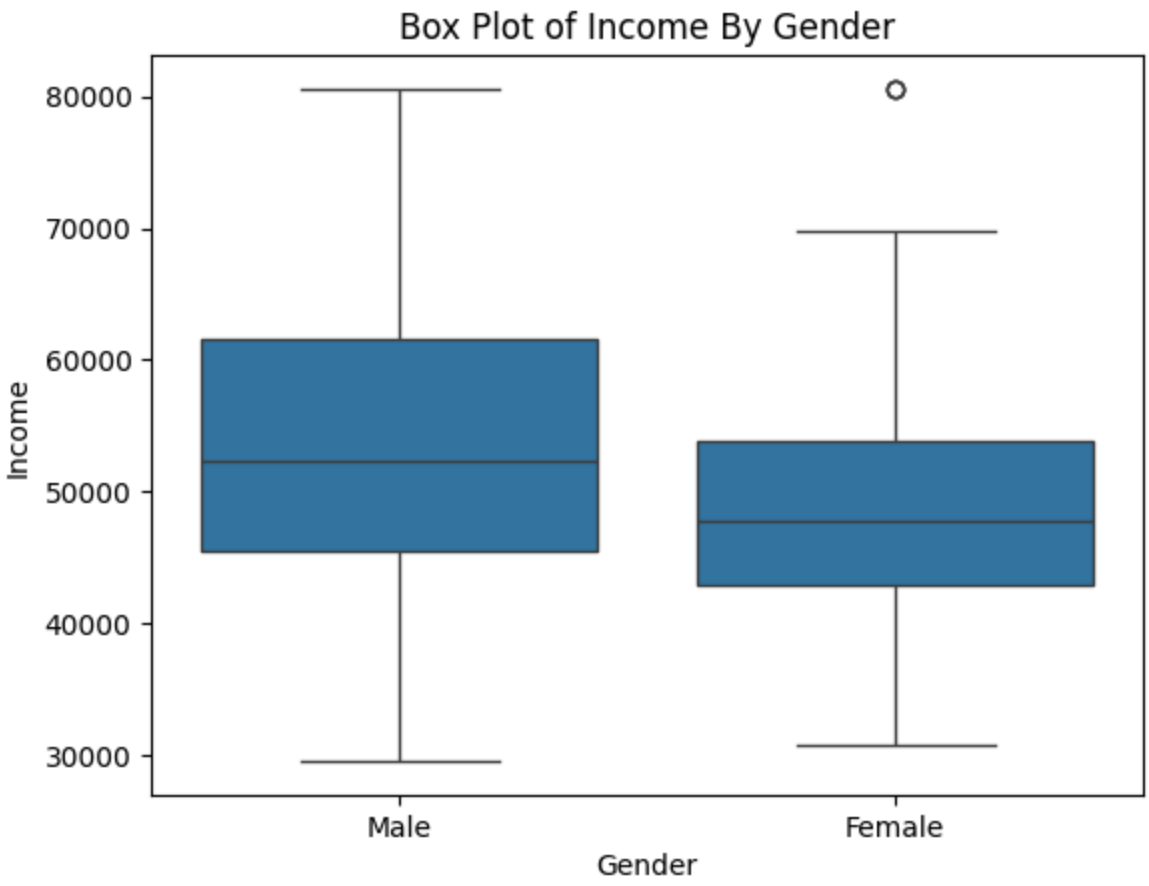


- Approximately 57.8% of the whole records are of male consumers, rest are women.
- Approximately 59.4% of the whole records are of partnered consumers, rest are singles.

Bivariate Analysis

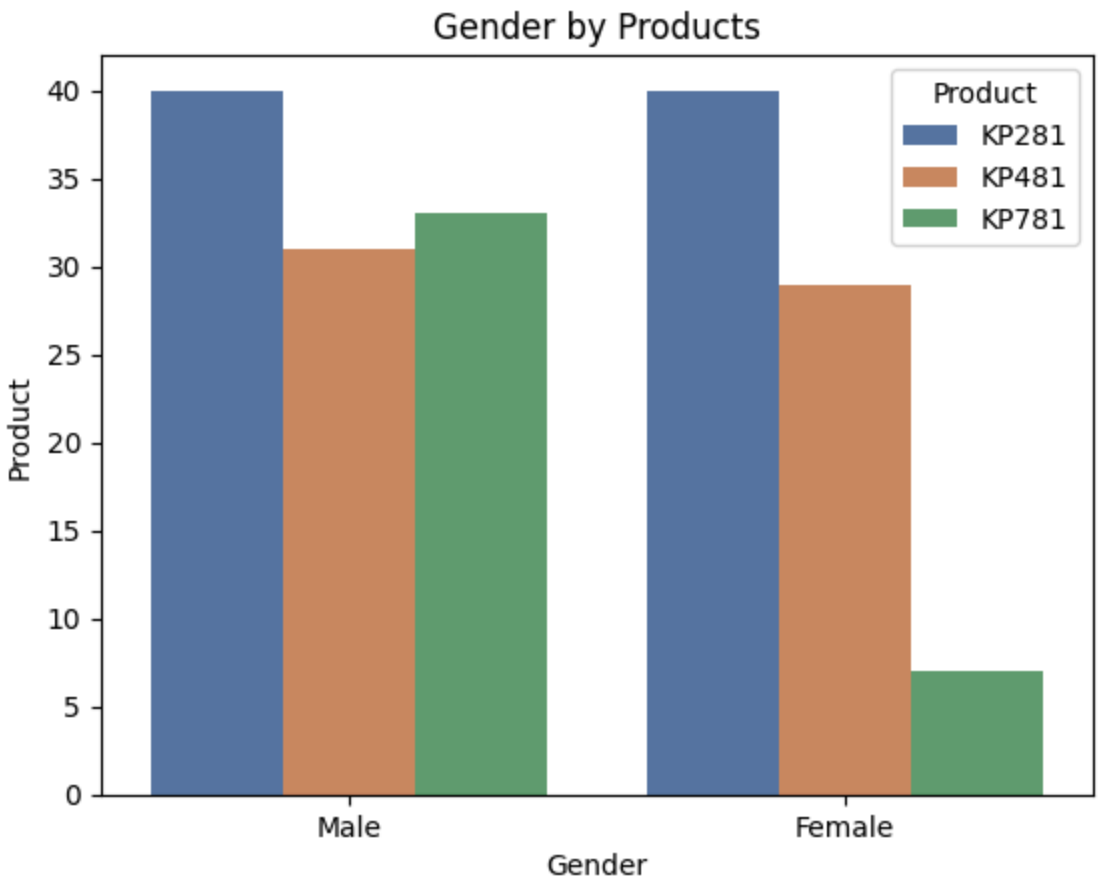
Bivariate box plot of gender vs Income

In [ ]:
```
sns.boxplot(x='Gender', y='Income', data=df)
plt.xlabel('Gender')
plt.ylabel('Income')
plt.title('Box Plot of Income By Gender')
plt.show()
```



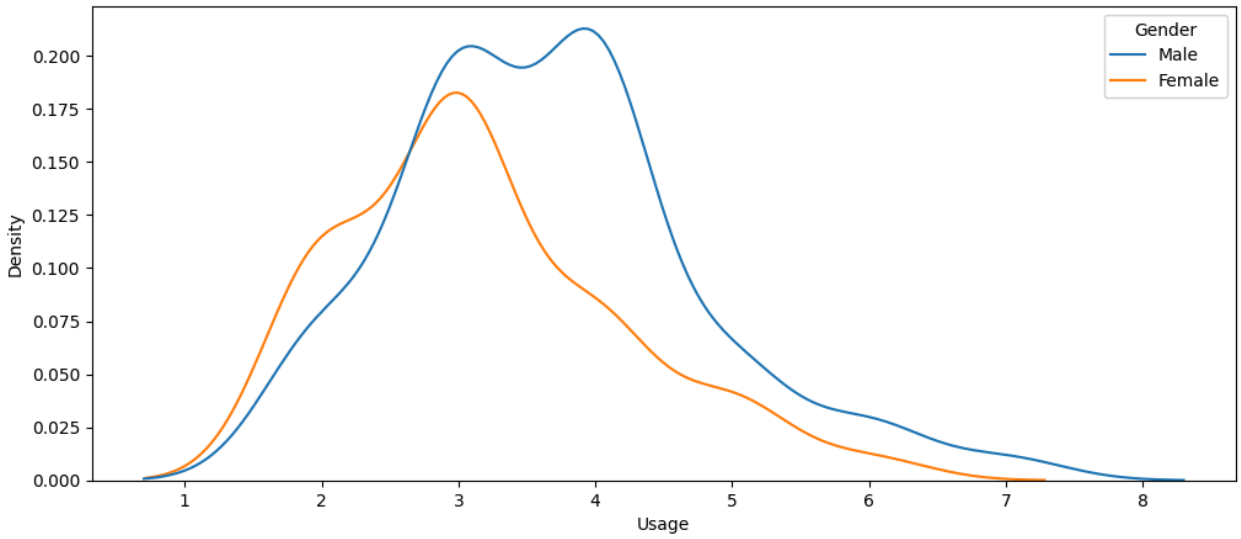
- 75 percentile Income of Male customer is approx 62 K and mean Income is approx 52K
- 75 percentile Income of Female customer is approx 53K and mean Income is approx 46K

In [ ]:
```
sns.countplot(x='Gender', hue='Product', data=df, palette='deep')
plt.xlabel('Gender')
plt.ylabel('Product')
plt.title('Gender by Products')
plt.show()
```



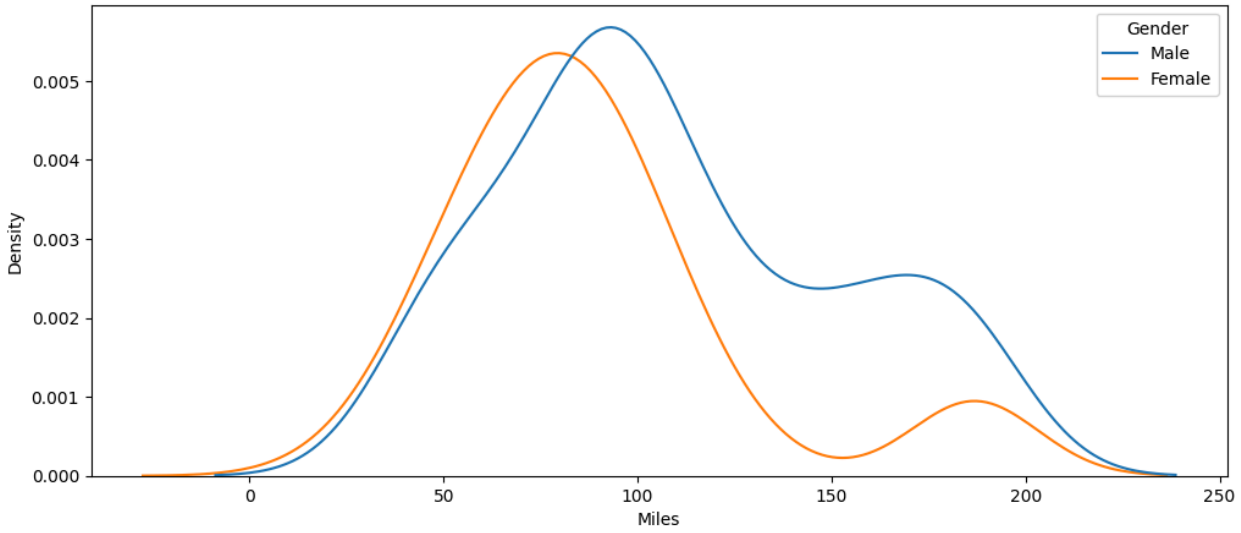
- KP281 Product is the equally preferred by both male and female genders
- KP781 Product is mostly preferred among the Male customers
- Overall Male customers are the highest product purchasers

In [ ]:
```
# Product purchased Customers Income and their Gender
plt.figure(figsize=(12,5))
sns.kdeplot(data=df,x='Usage',hue='Gender')
plt.show()
```



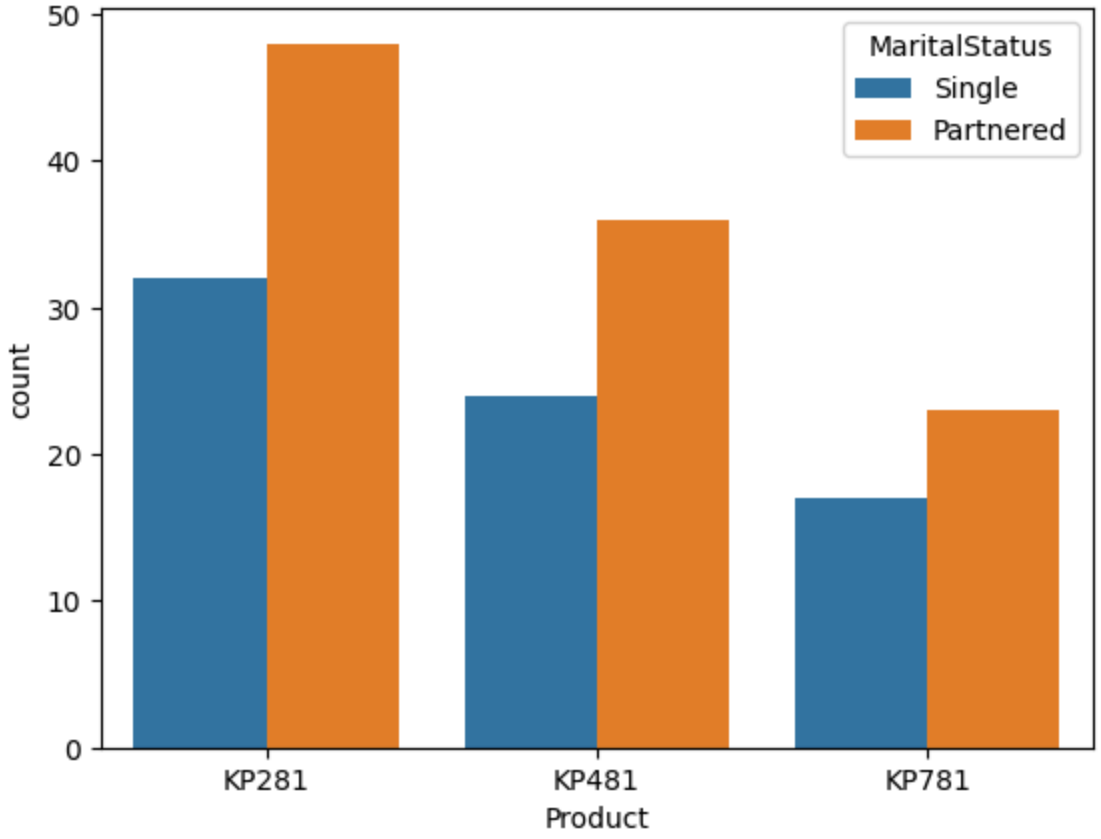
- Male customers usage is significantly higher the female customer
- Female customer's lack consistency after the 3 days per week

In [ ]:
```
# Distance covered by each Gender among the customers
plt.figure(figsize=(12,5))
sns.kdeplot(data=df,x='Miles',hue='Gender')
plt.show()
```



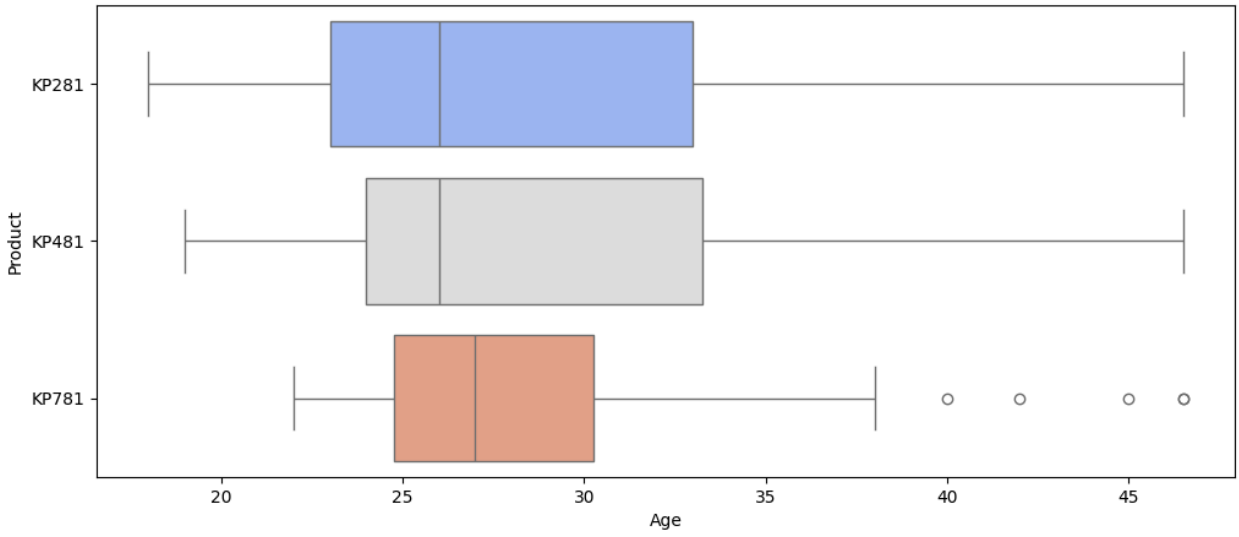
- Male customers have a consistent distance coverage than female customers
- Female customers have max distance covered as just over 300 miles

In [ ]:
```
# Product purchased among Married/Partnered and Single
sns.countplot(data=df,x='Product',hue='MaritalStatus')
plt.show()
```



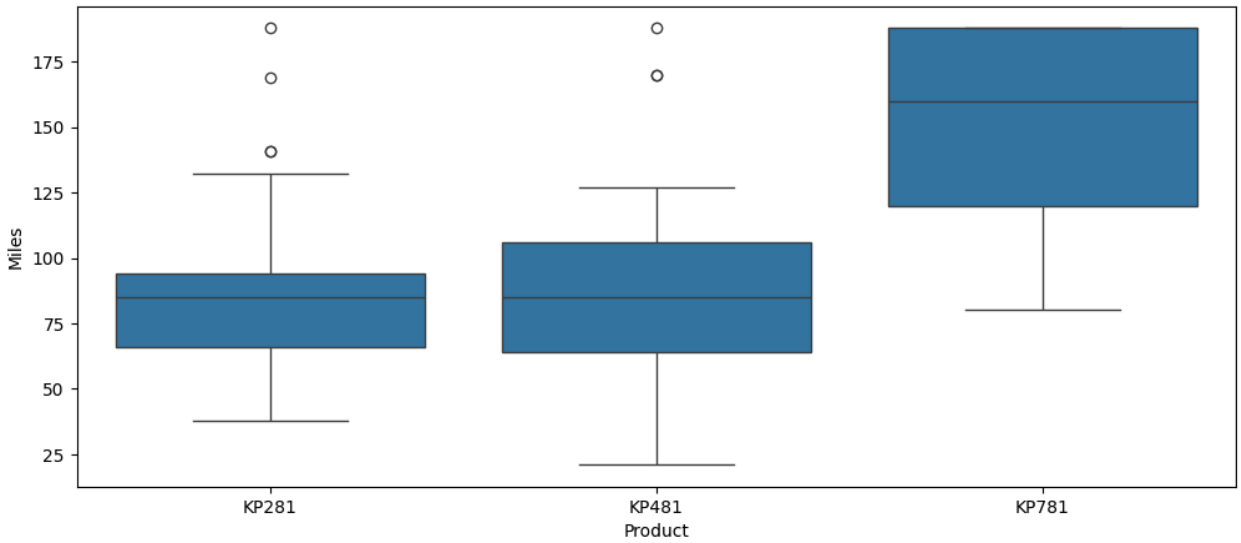
- KP281 is the most preferred product among customers
- KP481 is the second most preferred product among the customers
- Between Singles and Partnered, Partnered customers are the major product purchasers

In [ ]:
```
plt.figure(figsize=(12,5))
sns.boxplot(x='Age',y='Product',data=df,palette="coolwarm")
plt.show()
```



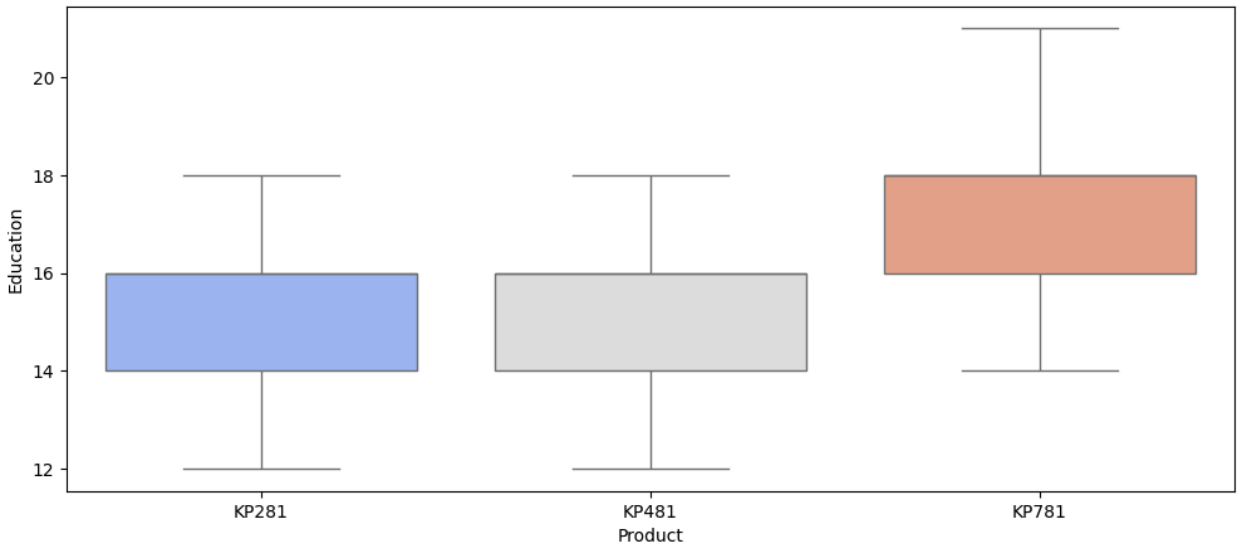
- Roughly few customers with age above 40 use product KP781
- Most of the customers are comfortable with KP281 product type
- KP481 is the second highest popular product among the younger side of the customer

In [ ]:
```
# Miles with each product
plt.figure(figsize=(12,5))
sns.boxplot(x='Product',y='Miles',data=df)
plt.show()
```



- Customers with product KP781, has been able to cover more miles than other two product types
- KP481 product is the second most highest miles covering product among the customers
- KP281 product customer had covered less distance compared with other two product types

In [ ]:
```
# Education of customers with each product purchased
plt.figure(figsize=(12,5))
sns.boxplot(x='Product',y='Education',data=df,palette="coolwarm")
plt.show()
```

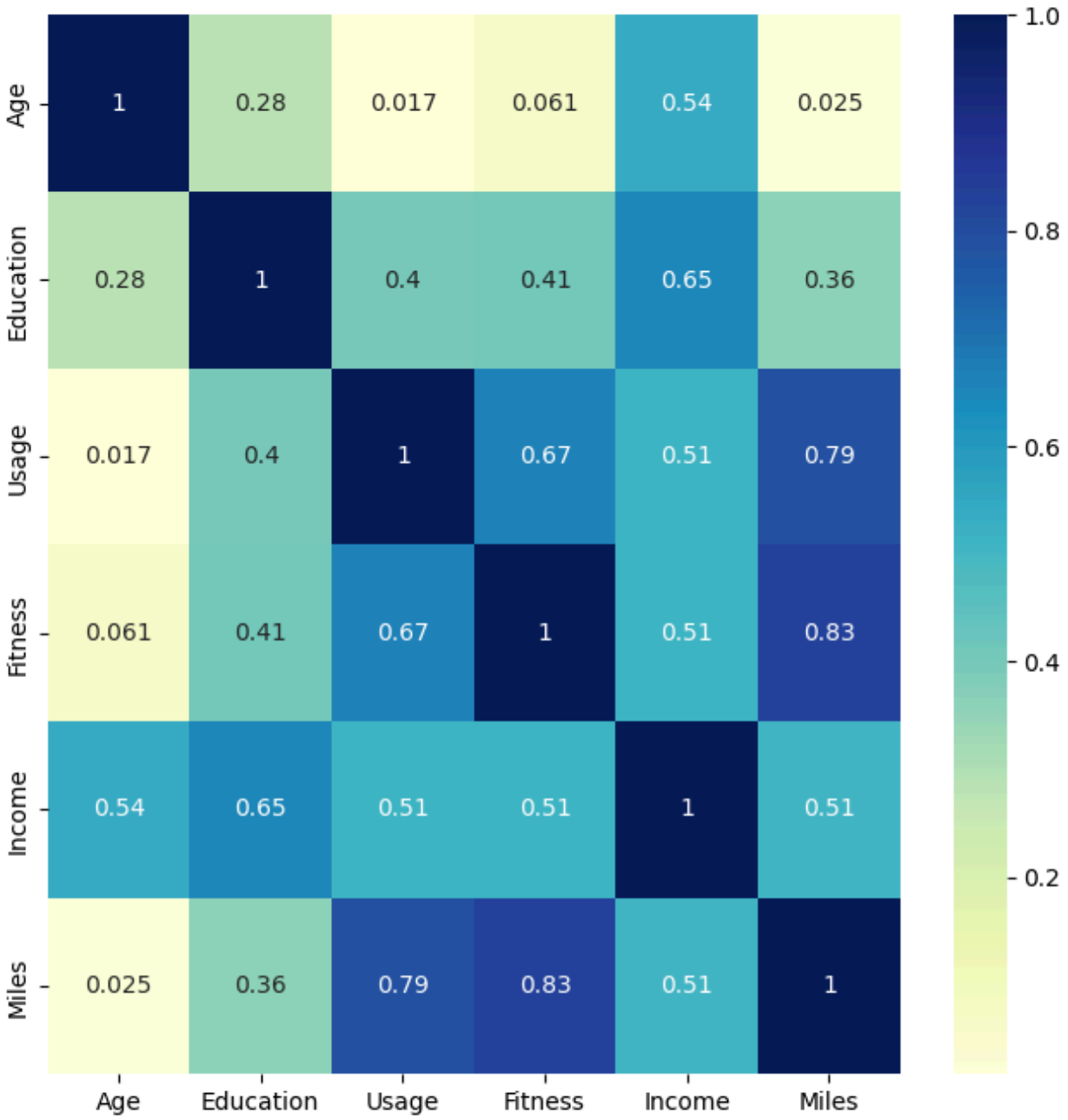


- Customers with Higher education of 16 to 18 have preferred mostly product type KP781
- Customers with education between 14 to 16 prefer KP281 and KP481 equally

Correlation between Variables

Heatmap

In [ ]:
```
plt.figure(figsize = (8, 8))
numerical_df = df.select_dtypes(include=np.number)
sns.heatmap(numerical_df.corr(), annot=True,  cmap="YlGnBu")
plt.show()
```



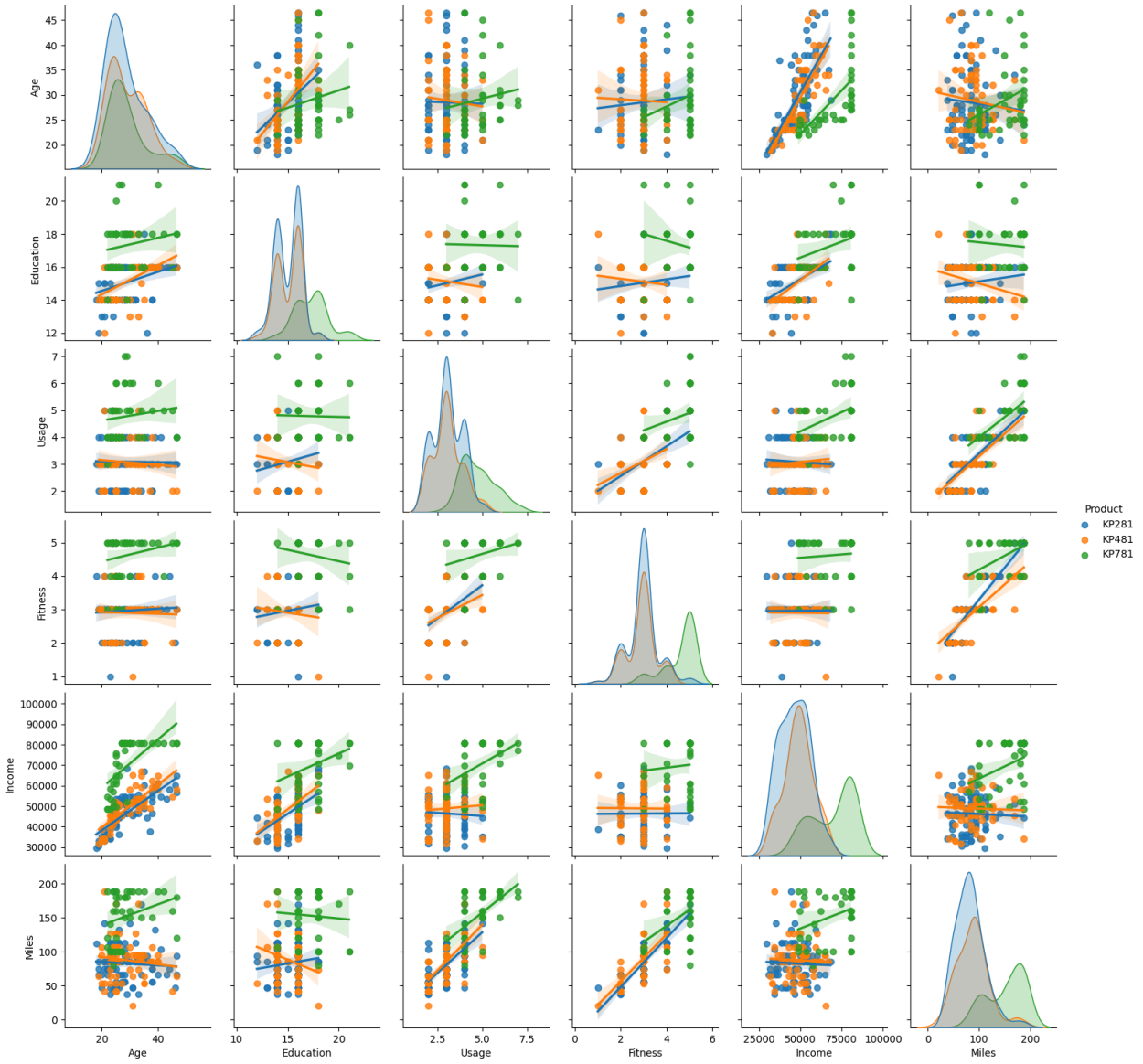
Insights:

- The strong correlation between Miles and Fitness indeed reflects a logical relationship: the fitter a person is, the more they tend to use the treadmill
- Similarly, high correlation between Fitness and Usage suggests that as fitness levels increase, customers are more likely to utilize fitness equipment, driving sales.
- Education and income seems to be highly correlated Higher education

often leads to better job opportunities, thereby increasing income

Pairplot

In [ ]:
```
# Product Analysis - Pair Plot
sns.pairplot(df,hue='Product',kind='reg')
plt.show()
```



From the pair plot we can see Age and Income are positively correlated and heatmap also suggests a strong correlation betwwen them

Probability of product purchase with respect to gender

Customer Profiling for Each Product

Customer profiling based on the 3 product categories provided

KP281

- Easily affordable entry level product, which is also the maximum selling product.
- KP281 is the most popular product among the entry level customers.
- This product is easily afforded by both Male and Female customers.
- Average distance covered in this model is around 70 to 90 miles.
- Product is used 3 to 4 times a week.
- Most of the customer who have purchased the product have rated Average shape as the fitness rating.
- Younger to Elder beginner level customers prefer this product.
- Single female & Partnered male customers bought this product more than single male customers.
- Income range between 39K to 53K have preferred this product.

KP481

- This is an Intermediate level Product.
- KP481 is the second most popular product among the customers.
- Fitness Level of this product users varies from Bad to Average Shape depending on their usage.
- Customers Prefer this product mostly to cover more miles than fitness.
- Average distance covered in this product is from 70 to 130 miles per week.
- More Female customers prefer this product than males.
- Probability of Female customer buying KP481 is significantly higher than male.

- KP481 product is specifically recommended for Female customers who are intermediate user.
- Three different age groups prefer this product - Teen, Adult and middle aged.
- Average Income of the customer who buys KP481 is 49K.
- Average Usage of this product is 3 days per week.
- More Partnered customers prefer this product.
- There are slightly more male buyers of the KP481.
- The distance travelled on the KP481 treadmill is roughly between 75 - 100 Miles. It is also the 2nd most distance travelled model.
- The buyers of KP481 in Single & Partnered, Male & Female are same.
- The age range of KP481 treadmill customers is roughly between 24-34 years.

KP781

- Due to the High Price & being the advanced type, customer prefers less of this product.
- Customers use this product mainly to cover more distance.
- Customers who use this product have rated excelled shape as fitness rating.
- Customer walk/run average 120 to 200 or more miles per week on his product.
- Customers use 4 to 5 times a week at least.
- Female Customers who are running average 180 miles (extensive exercise) , are using product KP781, which is higher than Male average using same product.
- Probability of Male customer buying Product KP781(31.73%) is way more than female(9.21%).
- Probability of a single person buying KP781 is higher than Married customers. So , KP781 is also recommended for people who are single

and exercises more.

- Middle aged to higher age customers tend to use this model to cover more distance.
- Average Income of KP781 buyers are over 75K per annum
- Partnered Female bought KP781 treadmill compared to Partnered Male.
- Customers who have more experience with previous aerofit products tend to buy this product
- This product is preferred by the customer where the correlation between Education and Income is High.

Business Insights based on Non-Graphical and Visual Analysis

Comments on the range of attributes:

- Product - 3 unique products are there. KP281, KP481, KP781.
- Age - Range of age is 18-50
- Gender - 104 are Male and 76 are Female
- MaritalStatus - 107 are Partnered and 73 are Single
- Usage - Mode of usage is 3 times a week
- Fitness- Mode of fitness is 3
- Income- range of Income is 20K-110K

Comments on the distribution of the variables and relationship between them:

Product Analysis

Products: KP281, KP481, and KP781 Unique Count: 80 for KP281, 60 for KP481, and 40 for KP781

Demographic Insights - Age Age Range: 18-50 years 75th Percentile Age: 33 years

Demographic Insights - Gender Male: 104 respondents (57%) Female: 76 respondents (42%)

Demographic Insights - Marital Status Partnered: 107 respondents (60%) Single: 73 respondents (40%)

Usage Patterns 38% of customers plan to use the Treadmill 3 times a week 28% of customers plan to use the Treadmill 4 times a week

Fitness Self-assessment 53% of customers rate themselves as having a fitness level of 3 on a scale of 1 to 5. Income Distribution Income Range: 110,000 75th Percentile Income: Less than $59,000

Recommendation

Female who prefer exercising equipments are very low here. Hence, we should run a marketing campaign on to encourage women to exercise more

KP281 & KP481 treadmills are preferred by the customers whose annual income lies in the range of 39K - 53K Dollars. These models should promoted as budget treadmills.

As KP781 provides more features and functionalities, the treadmill should be marketed for professionals and athletes.

KP781 product should be promotted using influencers and other international atheletes.

Research required for expanding market beyond 50 years of age considering health pros and cons.

Provide customer support and recommend users to upgrade from lower versions to next level versions after consistent usages.

KP781 can be recommended for Female customers who exercises extensively along with easy usage guidance since this type is advanced.

Target the Age group above 40 years to recommend Product KP781.