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

Defining Problem Statement

Airofit is a company that sells three types of treadmills—KP281, KP481, and KP781—each catering to different levels of fitness users (entry, mid, and advanced). The company has customer demographic and usage data and wants to understand the factors influencing a customer's decision to purchase a particular treadmill. This understanding will help in refining marketing strategies, targeting the right customer segments, and optimizing product offerings.

```
df=pd.read_csv('/content/aerofit_treadmill.csv')
```

```
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

```
df.shape
```

```
(180, 9)
```

```
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

```
df.dtypes
```

	0
Product	object
Age	int64
Gender	object
Education	int64
MaritalStatus	object
Usage	int64
Fitness	int64
Income	int64
Miles	int64

dtype: object

```
df.columns
```

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

```
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
```

```
# Check if there is any missing value
df.isnull().sum()
```

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

dtype: int64

so there is no missing value

Converting categorical data to categories

we found Gender, Marital status, Product as categorical data

```
df['Product']=df['Product'].astype('category')
df['MaritalStatus']=df['MaritalStatus'].astype('category')
df['Gender']=df['Gender'].astype('category')
```

```
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    category
 1   Age            180 non-null    int64
 2   Gender         180 non-null    category
 3   Education      180 non-null    int64
 4   MaritalStatus  180 non-null    category
 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: category(3), int64(6)
memory usage: 9.5 KB
```

*Non grafical analysis *

```
df.value_counts()
```

Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles	count
KP281	18	Male	14	Single	3	4	29562	112	1
	19	Female	14	Partnered	4	3	30699	66	1
		Male	12	Single	3	3	32973	85	1
			15	Single	2	3	31836	75	1
	20	Female	14	Partnered	3	3	32973	66	1
...	...	...	...	...	...	...	...	...	...
KP781	40	Male	21	Single	6	5	83416	200	1
	42	Male	18	Single	5	4	89641	200	1
	45	Male	16	Single	5	5	90886	160	1
	47	Male	18	Partnered	4	5	104581	120	1
	48	Male	18	Partnered	4	5	95508	180	1

180 rows × 1 columns

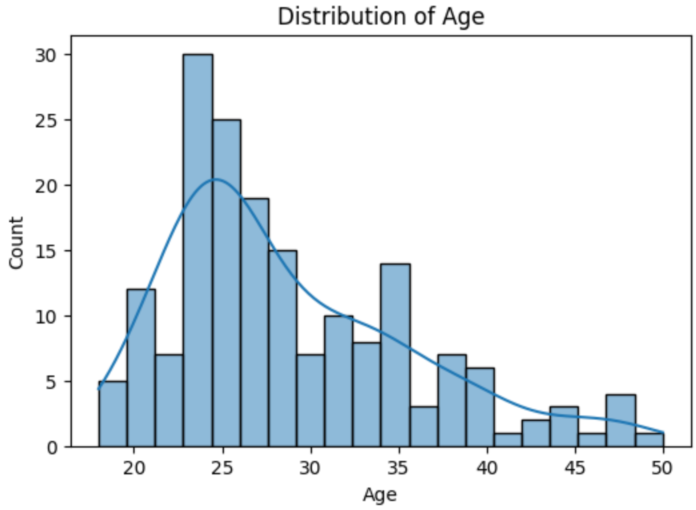
dtype: int64

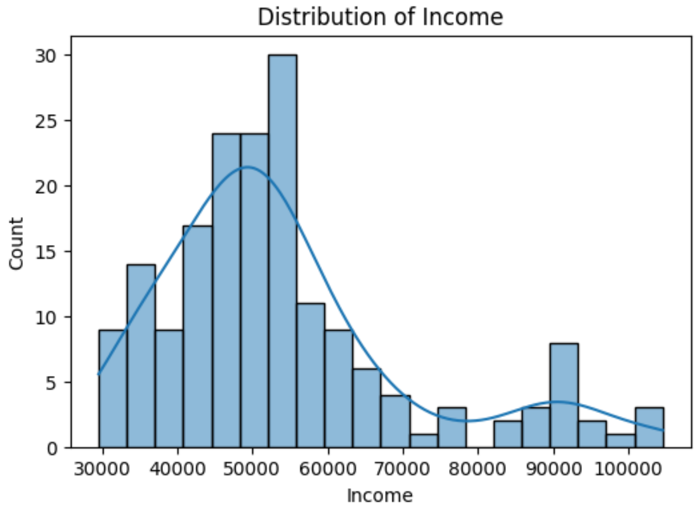
```
df.nunique()
```

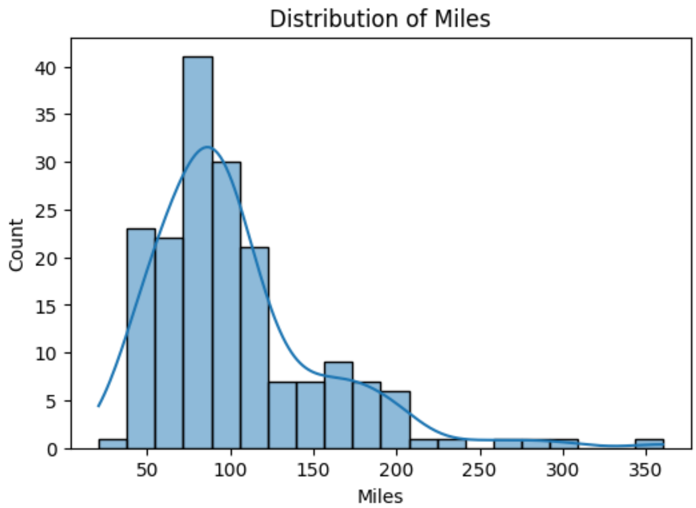
	0
Product	3
Age	32
Gender	2
Education	8
MaritalStatus	2
Usage	6
Fitness	5
Income	62
Miles	37

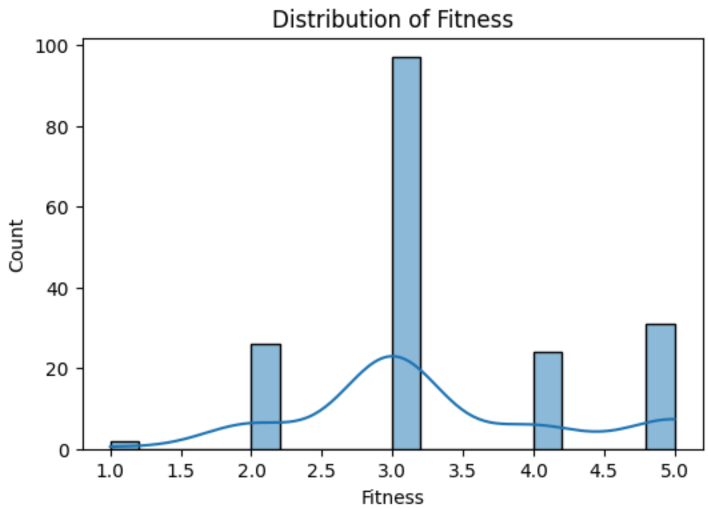
dtype: int64

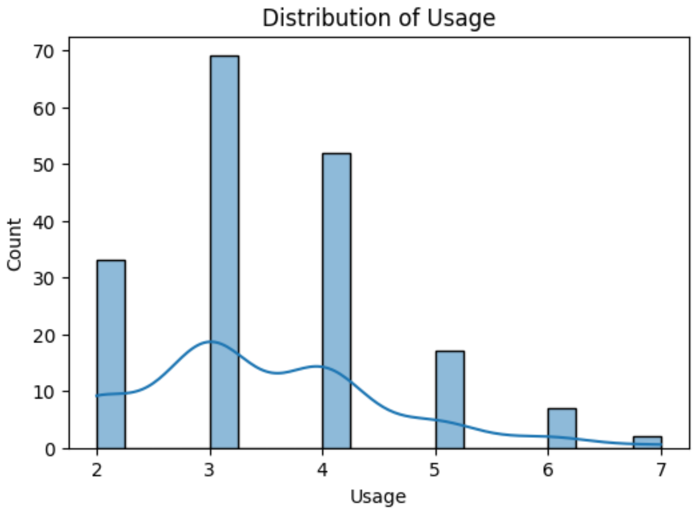
```
cols=['Age','Income','Miles','Fitness','Usage']
for col in cols:
  plt.figure(figsize=(6,4))
  sns.histplot(df[col],kde=True,bins=20)
  plt.title(f'Distribution of {col}')
  plt.show()
```


Distribution of Age
Count
Age
0
5
10
15
20
25
30
20
25
30
35
40
45
50


Distribution of Income
Count
Income
0
5
10
15
20
25
30
30000
40000
50000
60000
70000
80000
90000
100000


Distribution of Miles
Count
Miles
0
5
10
15
20
25
30
35
40
50
100
150
200
250
300
350


Distribution of Fitness
Count
Fitness
0
20
40
60
80
100
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0



Here by the above histogram and kde plot we can see the trends of various Value counts.And by observing this we can conlude the following.

1] The most customers fall within a younger to middle-aged group.we can see there is a spike at 20-30 age in kde plot indicating this age group buys more number of equipments.

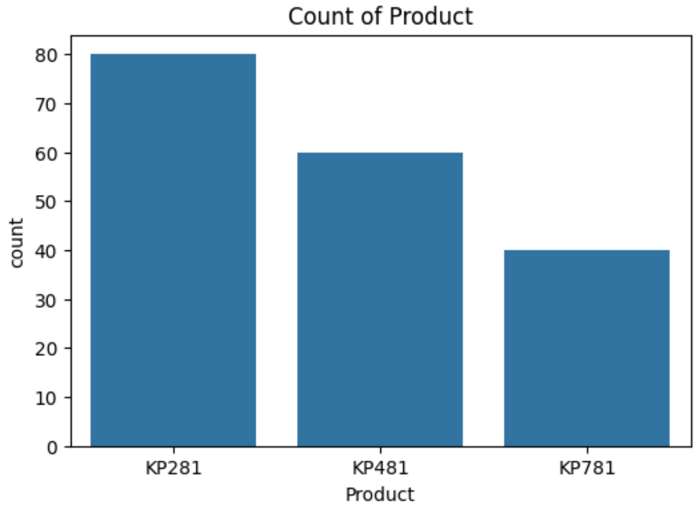
2] It seems that treademills are mostly bought by the medium income groups.From 40000-60000.And higher income groups don't prefer to buy the treadmills.The treadmills may be appealing to middle-income group

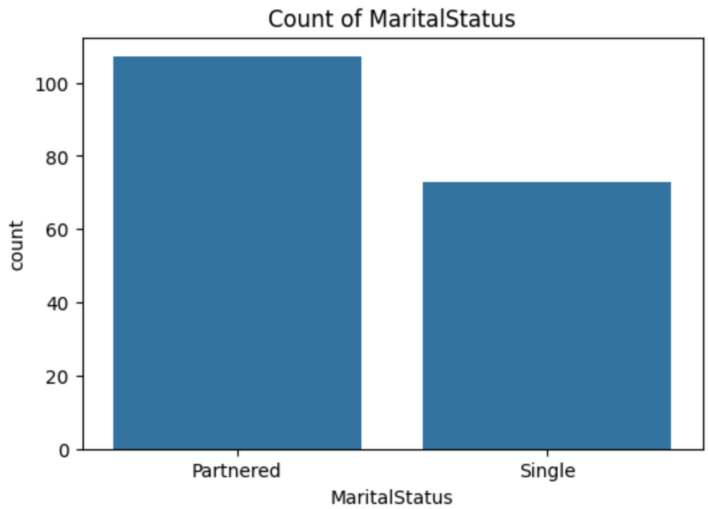
3] The avg number of miles costumer expect to walk per week is between 50-100.

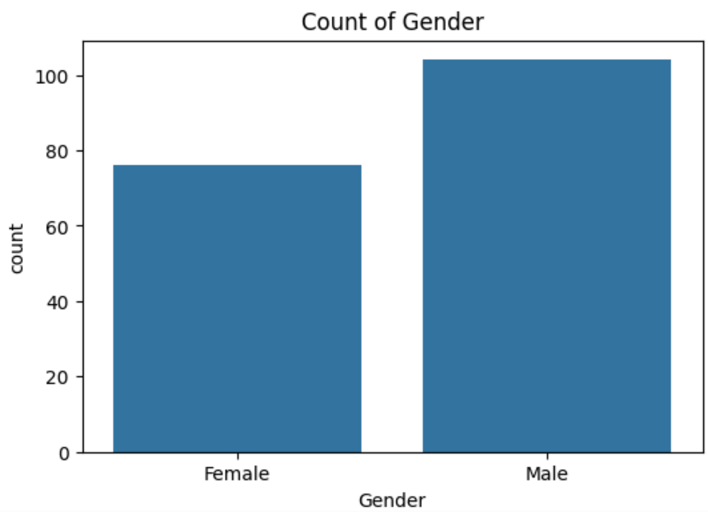
4] Most of the costumers rated themselves 3 from 1-5

5] Costumers expect on an avg 3 times usage of treadmills per week

```
#Countplot for categorical data
cat_cols = ['Product', 'MaritalStatus', 'Gender']
for col in cat_cols:
 plt.figure(figsize=(6, 4))
 sns.countplot(data=df, x=col)
 plt.title(f'Count of {col}')
 plt.show()
```



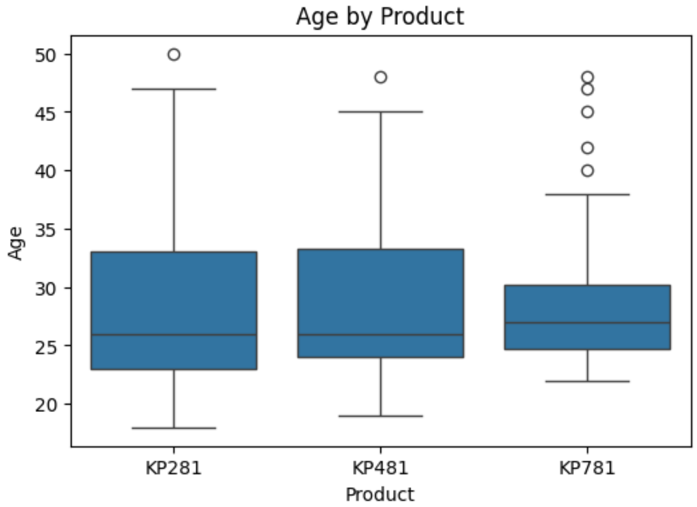


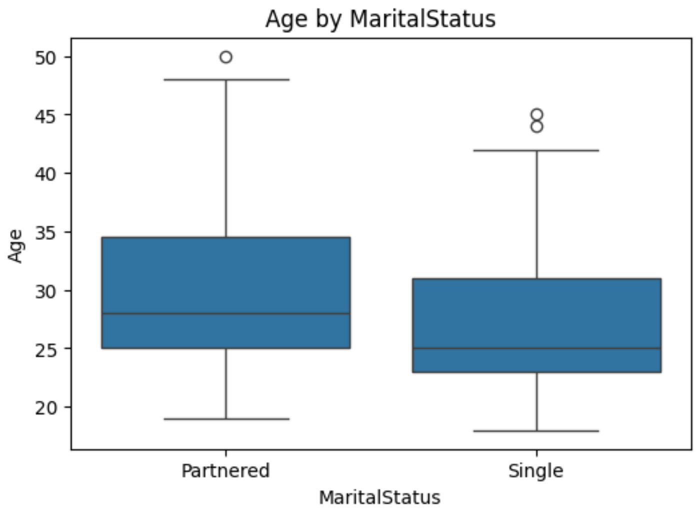


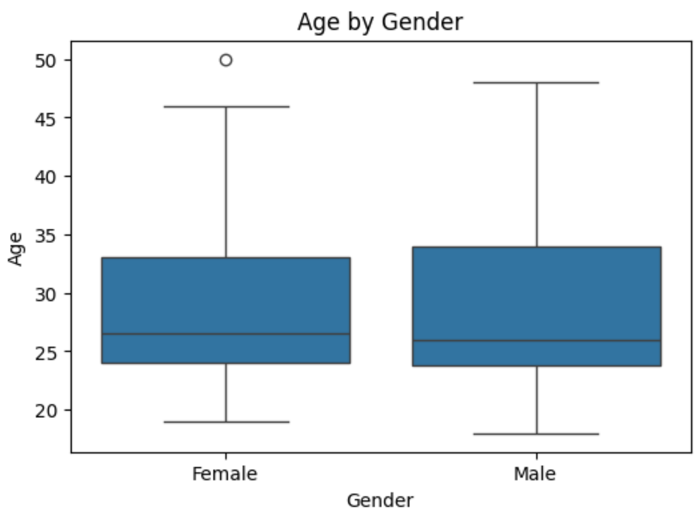
By the above countplot we can conclude that

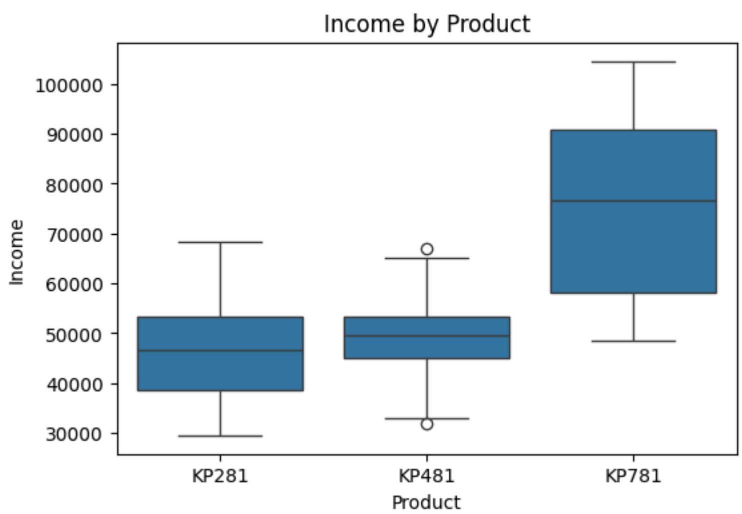
1) KP281 is the most bought treadmill type follwed by KP481 and then KP781.
2) The most treadmills are bought by the married people than the singles.
3) Males buy more euipments than females

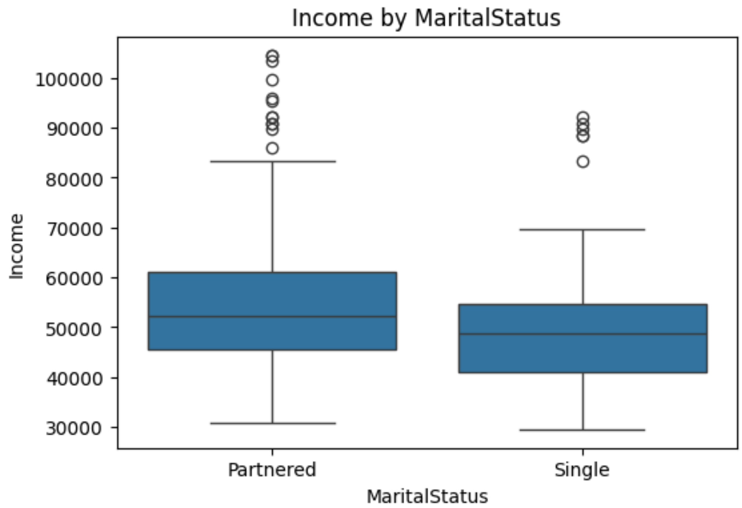
```
#Bivariate analysis for continuos vs categorical
for col in cols:
 for cat in cat_cols:
   plt.figure(figsize=(6, 4))
   sns.boxplot(data=df, x=cat, y=col)
   plt.title(f'{col} by {cat}')
   plt.show()
```

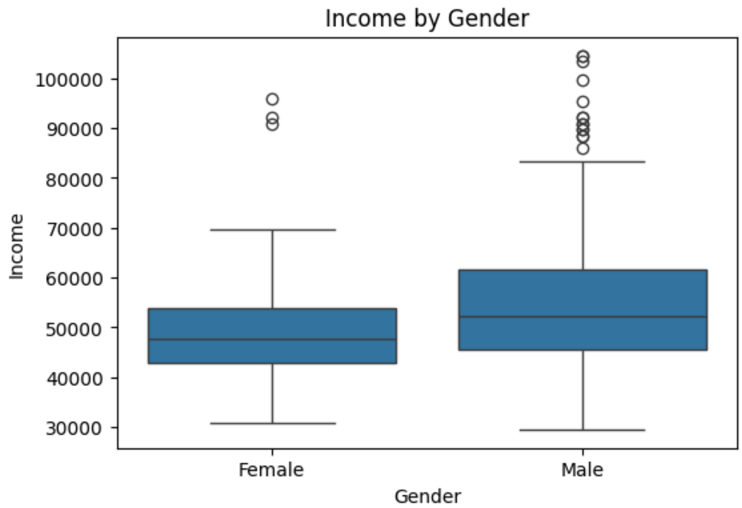

Age by Product
Age
50
45
40
35
30
25
20
KP281
KP481
KP781
Product

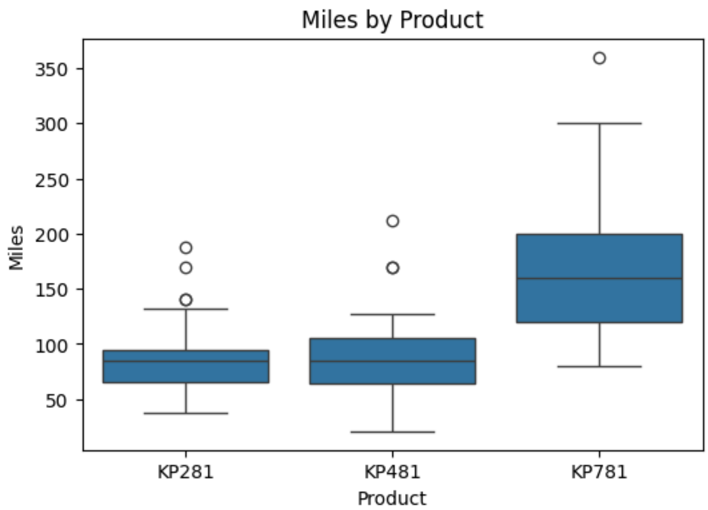

Age by MaritalStatus
Age
50
45
40
35
30
25
20
Partnered
Single
MaritalStatus

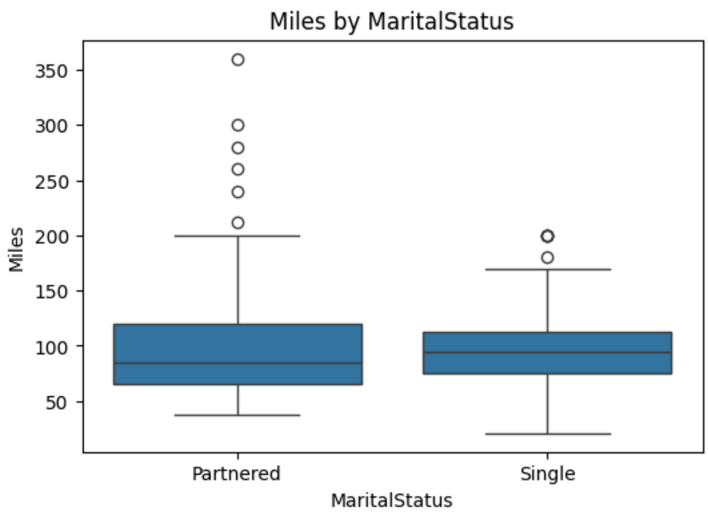

Age by Gender
Age
50
45
40
35
30
25
20
Female
Male
Gender

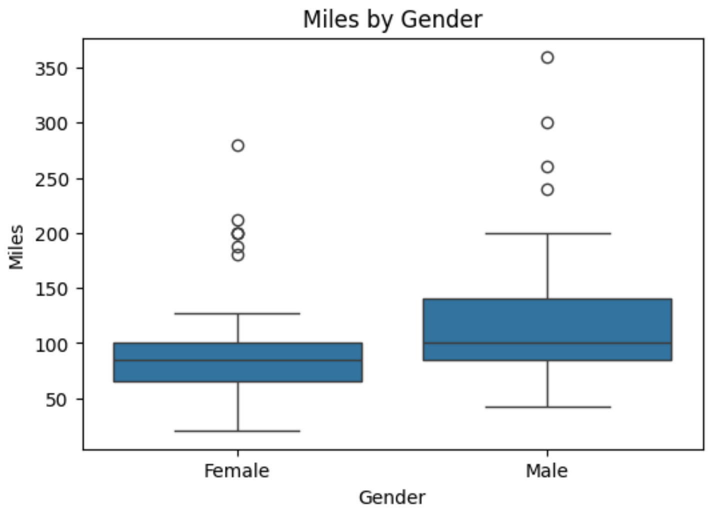

Income by Product
Income
100000
90000
80000
70000
60000
50000
40000
30000
KP281
KP481
KP781
Product

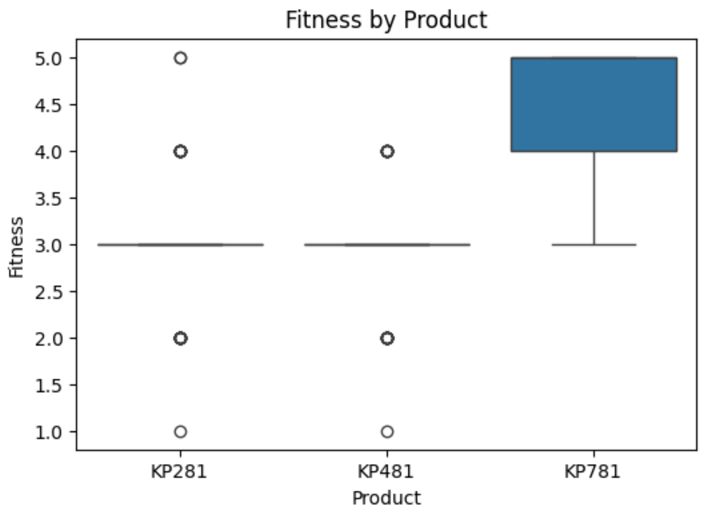

Income by MaritalStatus
Income
MaritalStatus
Partnered
Single
100000
90000
80000
70000
60000
50000
40000
30000

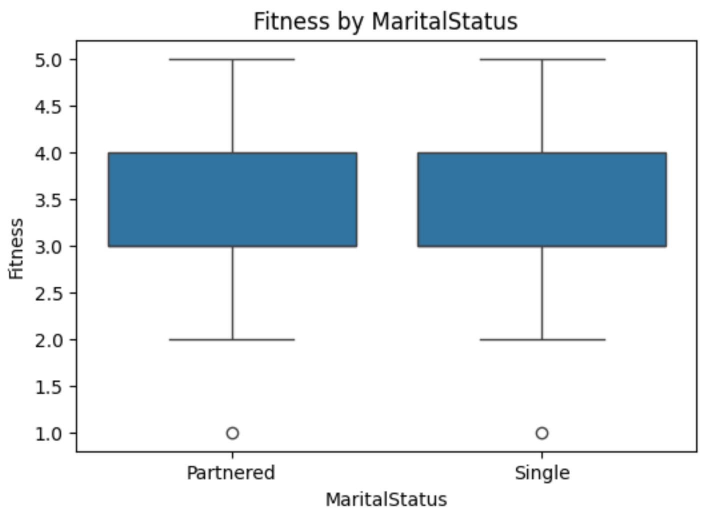

Income by Gender
Income
Gender
Female
Male
100000
90000
80000
70000
60000
50000
40000
30000

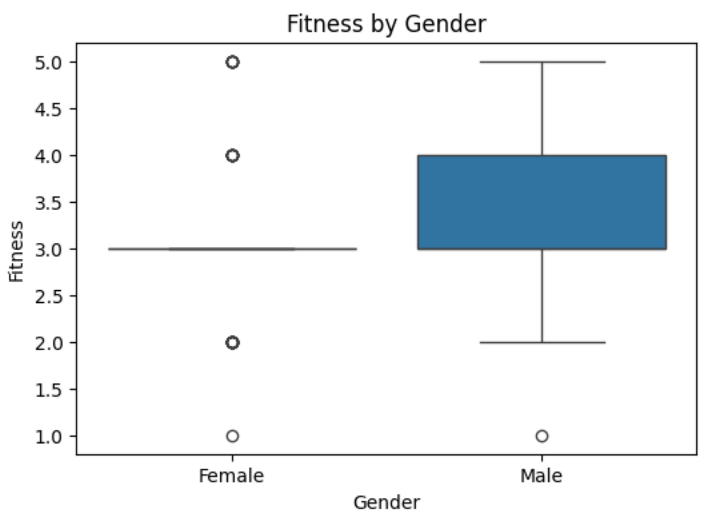

Miles by Product
Miles
Product
KP281
KP481
KP781
350
300
250
200
150
100
50

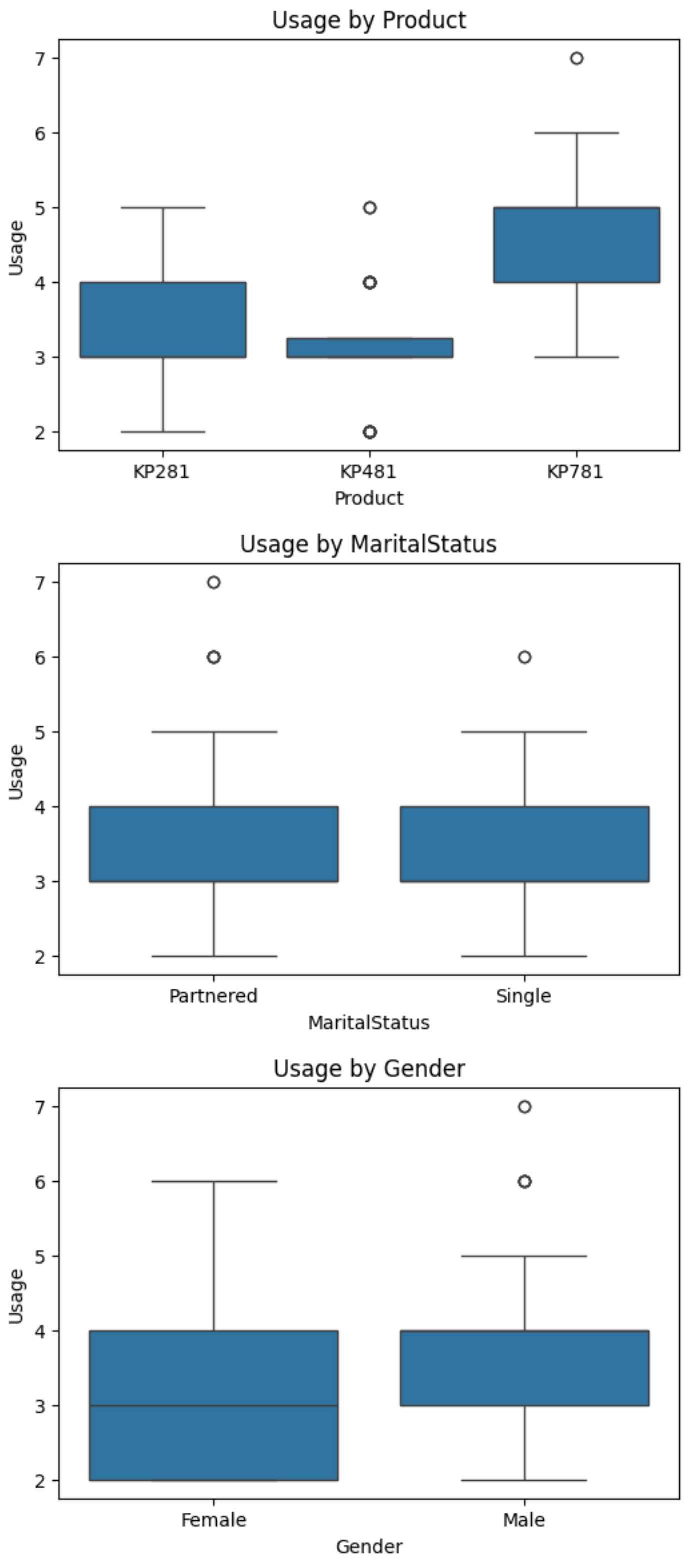

Miles by MaritalStatus
Miles
MaritalStatus
Partnered
Single
350
300
250
200
150
100
50


Miles by Gender
Miles
Female
Male
Gender

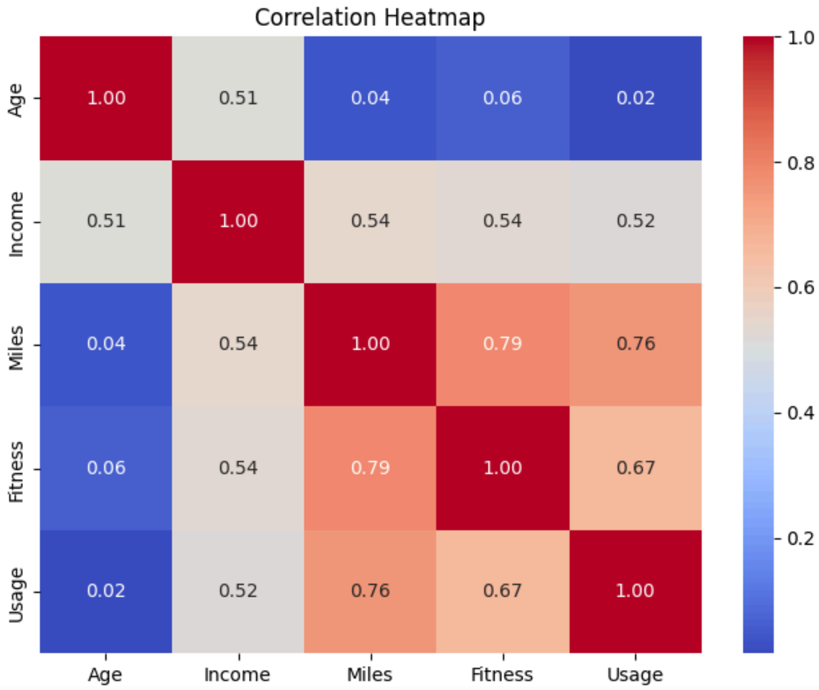

Fitness by Product
Fitness
KP281
KP481
KP781
Product


Fitness by MaritalStatus
Fitness
Partnered
Single
MaritalStatus


Fitness by Gender
Fitness
Female
Male
Gender


Usage by Product
Usage
Product
KP281
KP481
KP781
Usage by MaritalStatus
MaritalStatus
Partnered
Single
Usage by Gender
Gender
Female
Male

```
#Correlation analysis
plt.figure(figsize=(8, 6))
corr = df[cols].corr()
sns.heatmap(corr, annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Correlation Heatmap")
plt.show()
```

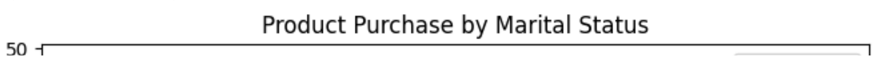



```
#Detecting missing values and outliers
for col in df.select_dtypes(include='number').columns:
 mean_val = df[col].mean()
 median_val = df[col].median()
 diff = abs(mean_val- median_val)
 print(f"\nColumn: {col}")
 print(f"Mean: {mean_val:.2f}, Median: {median_val:.2f}, Difference: {diff:2f}")

 if diff>0.5:
  print('Outlier is detected')
```

```
Column: Age
Mean: 28.79, Median: 26.00, Difference: 2.788889
Outlier is detected

Column: Education
Mean: 15.57, Median: 16.00, Difference: 0.427778

Column: Usage
Mean: 3.46, Median: 3.00, Difference: 0.455556

Column: Fitness
Mean: 3.31, Median: 3.00, Difference: 0.311111

Column: Income
Mean: 53719.58, Median: 50596.50, Difference: 3123.077778
Outlier is detected

Column: Miles
Mean: 103.19, Median: 94.00, Difference: 9.194444
Outlier is detected
```

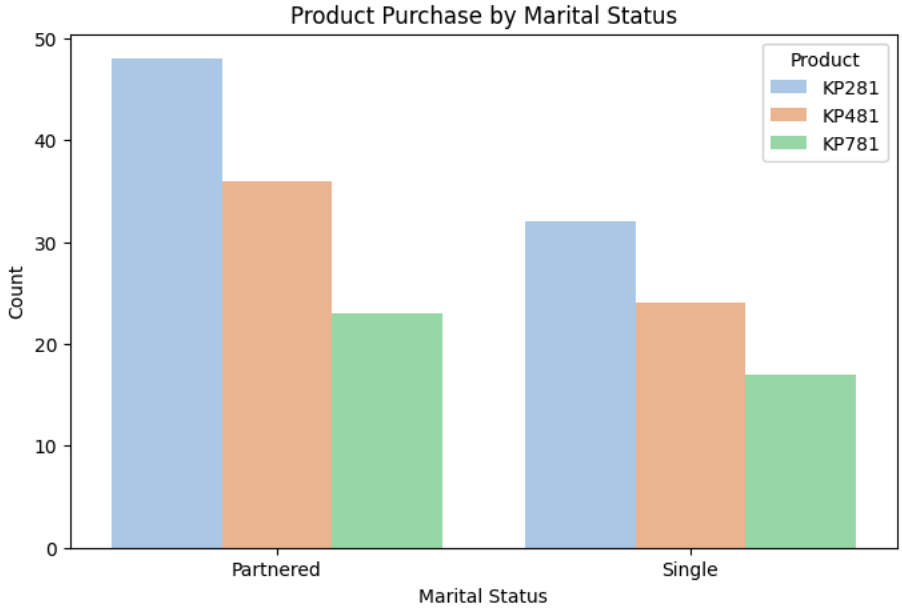
```
plt.figure(figsize=(8, 5))
sns.countplot(data=df, x='MaritalStatus', hue='Product', palette='pastel')
plt.title('Product Purchase by Marital Status')
plt.xlabel('Marital Status')
plt.ylabel('Count')
```

```
Text(0, 0.5, 'Count')
```



```
plt.figure(figsize=(8, 5))
sns.countplot(data=df, x='MaritalStatus', hue='Product', palette='pastel')
plt.title('Product Purchase by Marital Status')
plt.xlabel('Marital Status')
plt.ylabel('Count')
```

```
Text(0, 0.5, 'Count')
```




```
product_counts = pd.crosstab(index=df['Product'], columns='count')
product_percentage = product_counts / product_counts.sum() * 100
product_percentage.columns = ['Percentage']
product_percentage = product_percentage.round(2)
print(product_percentage)
```

```
         Percentage
Product
KP281         44.44
KP481         33.33
KP781         22.22
```

Business insights

1] Most customers fall into the young to middle-aged category. This suggests marketing campaigns should focus on working professionals, millennials, and Gen X customers who are health-conscious. 2]The income distribution is skewed toward mid-income levels, suggesting that the KP281 and KP481 models are more accessible and likely more popular.Customers with higher income may be more inclined to purchase the premium KP781 model.
3]Most customers plan to use the treadmill 3–5 times per week, covering a moderate number of miles

Range of attributes

1] Most users likely fall in the 25 to 45 age group — prime working-age adults interested in health and wellness. 2] Most people fall in between 60k to 100k of income 3]Miles per week range from 20 to 150. 4]Most users probably use the treadmill 3–5 times/week, indicating sustained engagement. 5]Most costumers rated themselves at 3

From the Heatmap given we can conclude,that the higher income buyers are leaning towards the premium models.And we can see the correlation.

Recommendations

Airofit should target mid-aged, fitness-conscious individuals by tailoring marketing based on their income and fitness level. The KP281 can be promoted as an affordable option for beginners, while KP781 should focus on high-income, advanced users. Personalized ads and recommendations can help boost conversion by aligning products with user profiles. Offering loyalty programs and trade-in deals can encourage customers to upgrade over time. Finally, adding smart features or fitness plans to premium models will increase customer engagement and brand value.