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

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

```
df.head(5)
```

	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.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.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

Age: The age of customers ranges from around 18 to 50 years, indicating a diverse age group among customers.

Education: Education years range from 12 to 21 years. This suggests that customers have varying levels of education, which may influence their preferences and choices.

Income: Annual income varies significantly, with a range from approximately 29,562 to 104,561. This wide income range implies that AeroFit caters to customers with different income levels.

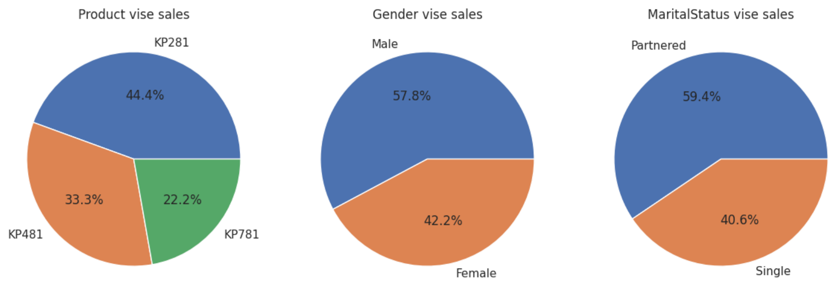
Fitness: Mean of fitness(3.3) indicates that customers having good fitness level.

Miles: Customers expect to walk/run different average miles per week, with the range covering 21 to approximately 360 miles.

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

	Product	Gender	MaritalStatus
count	180	180	180
unique	3	2	2

```
plt.figure(figsize=(12,4))

product_counts = df['Product'].value_counts()
gender_counts = df['Gender'].value_counts()
marital_counts = df['MaritalStatus'].value_counts()

plt.subplot(131)
plt.pie(product_counts, labels=product_counts.index, autopct='%1.1f%%')
plt.title("Product vise sales")

plt.subplot(132)
plt.pie(gender_counts, labels=gender_counts.index, autopct='%1.1f%%')
plt.title("Gender vise sales")

plt.subplot(133)
plt.pie(marital_counts, labels=marital_counts.index, autopct='%1.1f%%')
plt.title("MaritalStatus vise sales")

plt.tight_layout()
plt.show()
```

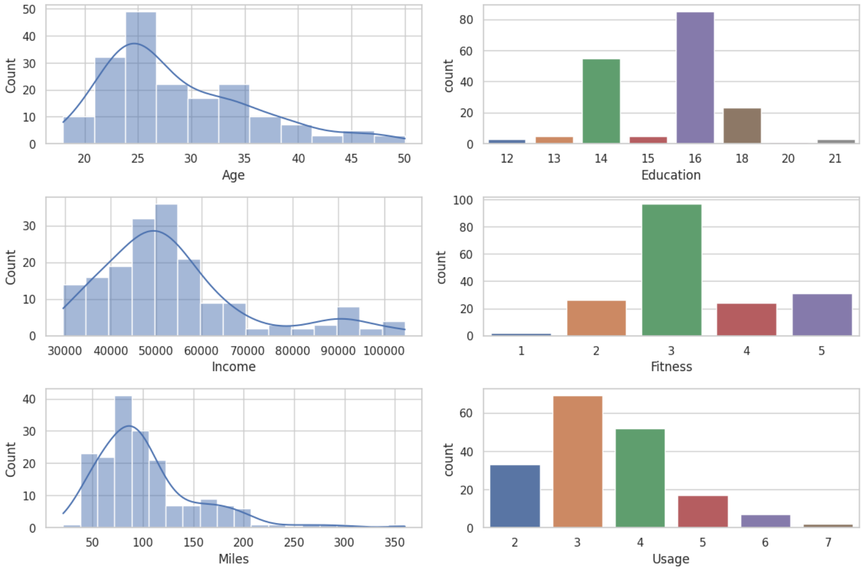


KP281, is the most popular choice among customers, followed by the mid-level KP481 and the advanced KP781.

Nearly 58% of the customers who purchase AeroFit's treadmills are male.

Approximately 59% of customers who buy AeroFit's products are partnered.

```
sns.set(style="whitegrid")

fig, axes = plt.subplots(nrows=3, ncols=2, figsize=(12,8))

sns.histplot(data=df, x='Age', ax=axes[0, 0], kde=True)
sns.histplot(data=df, x='Income', ax=axes[1, 0], kde=True)
sns.histplot(data=df, x='Miles', ax=axes[2, 0], kde=True)

sns.countplot(data=df, x='Usage', ax=axes[2,1])
sns.countplot(data=df, x='Fitness', ax=axes[1,1])
sns.countplot(data=df, x='Education', ax=axes[0,1])

plt.tight_layout()
plt.show()
```



A substantial 78% of AeroFit's customers fall within the age range of 21-35. This age group shows a strong inclination towards purchasing AeroFit's products.

An impressive 75% of AeroFit's customers report an annual income falling within the range of 30000-60000 dollars

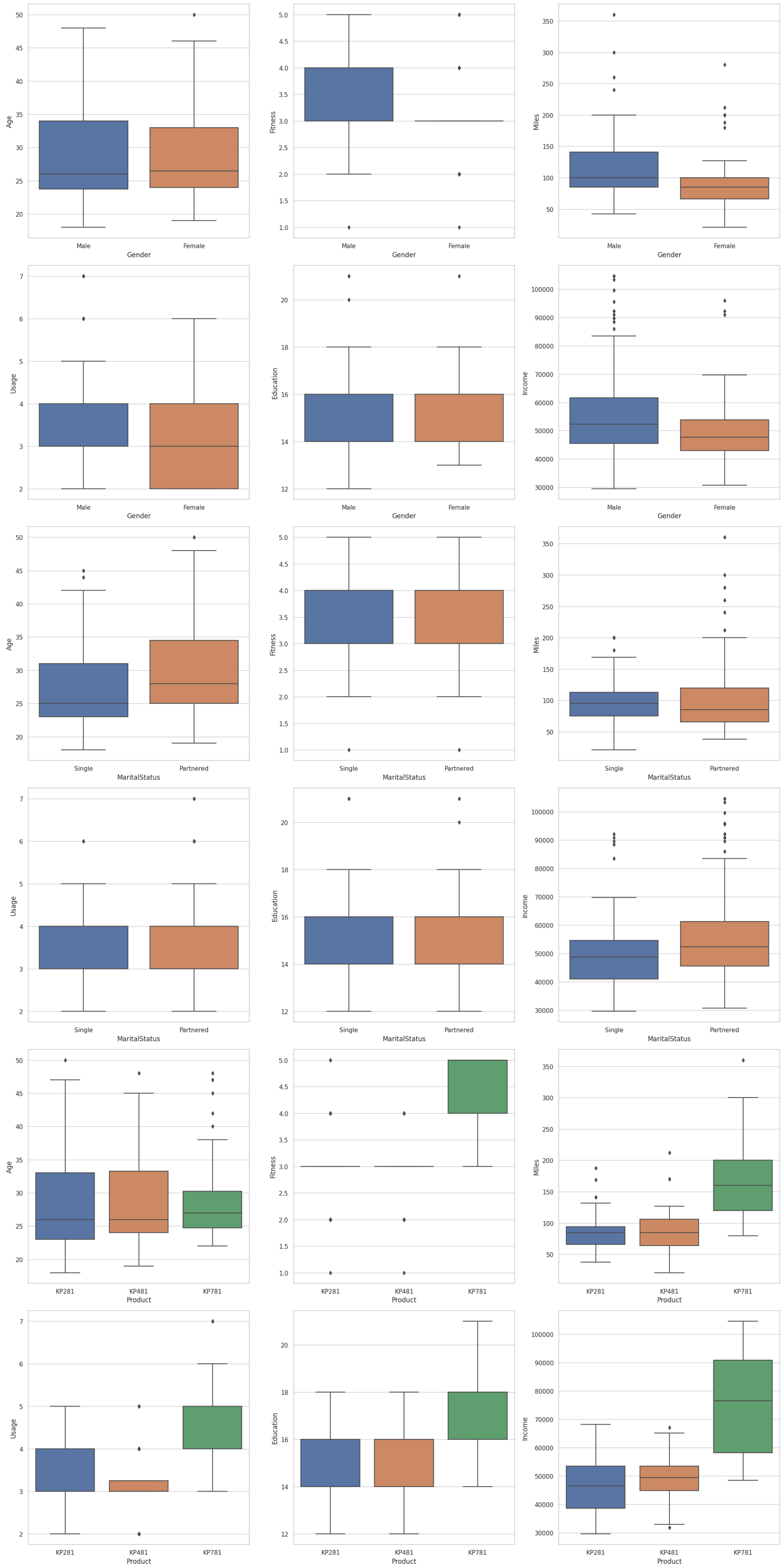
Approximately 75% of AeroFit's customers share a common expectation when it comes to their weekly walking or running goals, with a significant number aiming to cover a range of 40-120 miles.

The majority of customers (approximately 75%) who purchase AeroFit's products have a self-rated fitness level of 3, aligning with a moderate fitness level.

Furthermore, these customers typically hold educational qualifications of 14 or 16 years.

Around 39% of AeroFit's customers intend to use the treadmill three times a week, making this the most common usage frequency. Following closely, 4 times a week and 2 times a week are also popular choices.

```
fig, axes = plt.subplots(nrows=6, ncols=3, figsize=(20, 40))

sns.boxplot(data=df, x='Gender', y='Age', ax=axes[0,0])
sns.boxplot(data=df, x='Gender', y='Fitness', ax=axes[0,1])
sns.boxplot(data=df, x='Gender', y='Miles', ax=axes[0,2])
sns.boxplot(data=df, x='Gender', y='Usage', ax=axes[1,0])
sns.boxplot(data=df, x='Gender', y='Education', ax=axes[1,1])
sns.boxplot(data=df, x='Gender', y='Income', ax=axes[1,2])
sns.boxplot(data=df, x='MaritalStatus', y='Age', ax=axes[2,0])
sns.boxplot(data=df, x='MaritalStatus', y='Fitness', ax=axes[2,1])
sns.boxplot(data=df, x='MaritalStatus', y='Miles', ax=axes[2,2])
sns.boxplot(data=df, x='MaritalStatus', y='Usage', ax=axes[3,0])
sns.boxplot(data=df, x='MaritalStatus', y='Education', ax=axes[3,1])
sns.boxplot(data=df, x='MaritalStatus', y='Income', ax=axes[3,2])
sns.boxplot(data=df, x='Product', y='Age', ax=axes[4,0])
sns.boxplot(data=df, x='Product', y='Fitness', ax=axes[4,1])
sns.boxplot(data=df, x='Product', y='Miles', ax=axes[4,2])
sns.boxplot(data=df, x='Product', y='Usage', ax=axes[5,0])
sns.boxplot(data=df, x='Product', y='Education', ax=axes[5,1])
sns.boxplot(data=df, x='Product', y='Income', ax=axes[5,2])

plt.tight_layout()
plt.show()
```


Age
Fitness
Miles
Usage
Education
Income
Gender
Male
Female
MaritalStatus
Single
Partnered
Product
KP281
KP481
KP781

Age distribution for both genders is fairly similar, with a median age around 25-30. However, there are slightly more older customers (outliers) among males compared to females.

The expected miles to be covered per week are relatively more for male with compared to female.

The income distribution is fairly consistent between genders, with medians indicating similar earning levels. However, there are more high-income outliers among males.

The income and age distribution is fairly consistent between single and partnered customers. Hoever, partnered customers are having slighlty more median with more outliers.

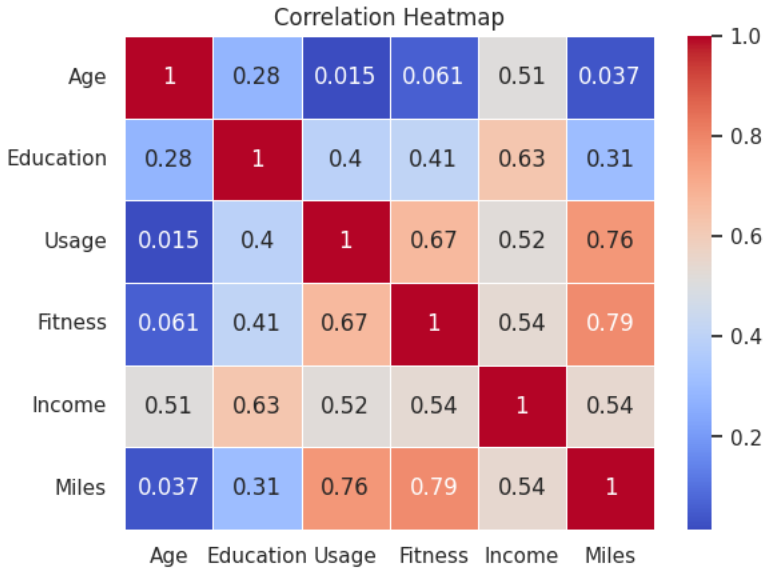
The age distribution varies by product, with KP281 having a younger customer base and KP781 having relatively older customers. KP481 falls in between.

The expected miles to be covered per week vary by product, with KP781 customers having higher mileage expectations, followed by KP481 and KP281.

Income levels vary slightly between products, with KP781 customers having higher median incomes, followed by KP481 and KP281.

```
correlation_matrix = df.corr()
sns.heatmap(correlation_matrix, annot=True, cmap="coolwarm", linewidths=.5)
plt.title("Correlation Heatmap")
plt.show()

sns.pairplot(data=df, vars=['Age', 'Education', 'Income', 'Fitness', 'Miles'], hue='Product')
plt.suptitle("Pairplot of Numerical Variables", y=1.02)
plt.show()
```

```
<ipython-input-27-6adb63b42bf9>:2: FutureWarning: The default value of numeric_only in DataFrame.corr is deprecated. In a future ver
  correlation_matrix = df.corr()
```



There is a positive correlation between income and education, suggesting that individuals with higher levels of education tend to have higher incomes.

Fitness and income also show a positive correlation, suggesting that individuals with higher incomes tend to rate their fitness higher.

There is a moderate positive correlation between income and age, indicating that older customers tend to have higher incomes.

Miles and usage have a positive correlation, suggesting that customers who plan to run or walk more miles per week tend to use the treadmill more frequently.

Miles and fitness have a positive correlation, indicating that customers who expect to cover more miles tend to rate their fitness higher.

```
# Contingency Table 1: Product vs. Gender
contingency_table1 = pd.crosstab(df['Product'], df['Gender'])

# Contingency Table 2: Product vs. Marital Status
contingency_table2 = pd.crosstab(df['Product'], df['MaritalStatus'])

# Contingency Table 3: Product vs. Usage
contingency_table3 = pd.crosstab(df['Product'], pd.cut(df['Usage'], [0, 2, 4, 7]))

# Contingency Table 4: Product vs. Fitness
contingency_table4 = pd.crosstab(df['Product'], df['Fitness'])

# Calculate conditional probabilities
conditional_probabilities1 = contingency_table1 / contingency_table1.sum(axis=0)
conditional_probabilities2 = contingency_table2 / contingency_table2.sum(axis=0)
conditional_probabilities3 = contingency_table3 / contingency_table3.sum(axis=0)
conditional_probabilities4 = contingency_table4 / contingency_table4.sum(axis=0)

# Calculate marginal probabilities
marginal_probabilities1 = contingency_table1 / contingency_table1.sum().sum()
marginal_probabilities2 = contingency_table2 / contingency_table2.sum().sum()
marginal_probabilities3 = contingency_table3 / contingency_table3.sum().sum()
marginal_probabilities4 = contingency_table4 / contingency_table4.sum().sum()

print("Contingency Table 1: Product vs. Gender")
print(contingency_table1)
print("\nConditional Probabilities for Table 1")
print(conditional_probabilities1)
print("\nMarginal Probabilities for Table 1")
print(marginal_probabilities1)
```

```
Contingency Table 1: Product vs. Gender
Gender   Female  Male
Product
KP281        40    40
KP481        29    31
KP781         7    33

Conditional Probabilities for Table 1
Gender     Female      Male
Product
KP281    0.526316  0.384615
KP481    0.381579  0.298077
KP781    0.092105  0.317308

Marginal Probabilities for Table 1
Gender     Female      Male
Product
KP281    0.222222  0.222222
KP481    0.161111  0.172222
KP781    0.038889  0.183333
```

KP281 has a relatively balanced gender distribution, suggesting its broad appeal. KP781 seems to be more popular among male customers, indicating a potential gender-specific appeal.

```
print("\nContingency Table 2: Product vs. Marital Status")
print(contingency_table2)
print("\nConditional Probabilities for Table 2")
print(conditional_probabilities2)
print("\nMarginal Probabilities for Table 2")
print(marginal_probabilities2)
```

```
Contingency Table 2: Product vs. Marital Status
MaritalStatus  Partnered  Single
Product
KP281                 48      32
KP481                 36      24
```

```
KP781                 23      17

Conditional Probabilities for Table 2
MaritalStatus  Partnered    Single
Product
KP281           0.448598  0.438356
KP481           0.336449  0.328767
KP781           0.214953  0.232877

Marginal Probabilities for Table 2
MaritalStatus  Partnered    Single
Product
KP281           0.266667  0.177778
KP481           0.200000  0.133333
KP781           0.127778  0.094444
```

```
print("\nContingency Table 3: Product vs. Usage")
print(contingency_table3)
print("\nConditional Probabilities for Table 3")
print(conditional_probabilities3)
print("\nMarginal Probabilities for Table 3")
print(marginal_probabilities3)
```

```
Contingency Table 3: Product vs. Usage
Usage    (0, 2]  (2, 4]  (4, 7]
Product
KP281        19      59       2
KP481        14      43       3
KP781         0      19      21

Conditional Probabilities for Table 3
Usage      (0, 2]    (2, 4]    (4, 7]
Product
KP281    0.575758  0.487603  0.076923
KP481    0.424242  0.355372  0.115385
KP781    0.000000  0.157025  0.807692

Marginal Probabilities for Table 3
Usage      (0, 2]    (2, 4]    (4, 7]
Product
KP281    0.105556  0.327778  0.011111
KP481    0.077778  0.238889  0.016667
KP781    0.000000  0.105556  0.116667
```

KP281 and KP481 are used most frequently in the "2-4 times a week" usage category. KP781 is used primarily by customers who use the treadmill "4-7 times a week."

```
print("\nContingency Table 4: Product vs. Fitness")
print(contingency_table4)
print("\nConditional Probabilities for Table 4")
print(conditional_probabilities4)
print("\nMarginal Probabilities for Table 4")
print(marginal_probabilities4)
```

```
Contingency Table 4: Product vs. Fitness
Fitness  1   2   3  4   5
Product
KP281    1  14  54  9   2
KP481    1  12  39  8   0
KP781    0   0   4  7  29

Conditional Probabilities for Table 4
Fitness    1         2         3         4         5
Product
KP281    0.5  0.538462  0.556701  0.375000  0.064516
KP481    0.5  0.461538  0.402062  0.333333  0.000000
KP781    0.0  0.000000  0.041237  0.291667  0.935484

Marginal Probabilities for Table 4
Fitness         1         2         3         4         5
Product
KP281    0.005556  0.077778  0.300000  0.050000  0.011111
KP481    0.005556  0.066667  0.216667  0.044444  0.000000
KP781    0.000000  0.000000  0.022222  0.038889  0.161111
```

The majority of customers using all products have a fitness level of 3, indicating that fitness level 3 is the most common among customers.

KP781 stands out with a significant number of customers having a fitness level of 5, suggesting that this product may appeal to more fitness enthusiasts.

Recommendations for AeroFit

- Given that "KP281" is the top-selling product, consider promoting it more prominently.
- It is evident that "KP781" attracts customers with a fitness level of 5, consider promoting it to fitness enthusiasts.
- With almost 59% of customers being partnered, consider marketing strategies that target couples or promote the idea of sharing fitness goals with a partner.
- As nearly 75% of customers fall within the income range of $30, 000 to$60,000, offer financing options, discounts, or special packages that cater to customers within this income bracket.
- Since around 75% of customers expect to cover 40-120 miles per week, provide training plans, tracking apps, or support to help them achieve their mileage goals.
- Since 39% of customers plan to use the treadmill 3 times a week, consider offering special promotions, maintenance plans, or reminders to encourage consistent usage.
- With most customers having a fitness rating of 3, consider creating fitness programs or classes for this specific fitness level, catering to their needs and helping them progress.