

## Business Case: Aerofit - Descriptive Statistics & Probability

### Goal:

The primary goal of this analysis is to identify the characteristics of the target audience for each type of treadmill offered by Aerofit. By examining customer profiles and preferences, we aim to provide actionable recommendations that will help Aerofit better target its marketing strategies and improve product recommendations for new customers. Specifically, the analysis seeks to:

**Profile Customers:** Develop detailed profiles for each treadmill type (KP281, KP481, and KP781) based on customer demographics, fitness level, and usage patterns.

**Identify Key Influences:** Determine the factors that influence the purchase of each treadmill, such as age, gender, income, fitness level, and marital status.

**Calculate Probabilities:** Use statistical methods to compute marginal, conditional, and joint probabilities that can inform marketing and sales strategies.

**Provide Insights and Recommendations:** Derive meaningful business insights from the data and suggest targeted actions to enhance customer engagement and sales.

### Approach:

To achieve the above goal, the following approach is taken:

#### **Data Understanding and Preprocessing:**

- Import and inspect the dataset to understand its structure, data types, and completeness.
- Convert categorical variables as needed and clean the data to ensure accuracy for further analysis.

#### **Descriptive Analysis and Visualization:**

- Perform univariate and bivariate analyses using various plots (boxplots, histograms, count plots) to understand the distribution of key variables and detect any outliers.
- Use cross-tabulations to examine relationships between variables like Gender vs. Product, Age vs. Product, and Usage vs. Product.

#### **Probability Analysis:**

- Calculate marginal, conditional, and joint probabilities to understand the likelihood of different customer segments purchasing each treadmill.
- Perform customer profiling to categorize users based on demographics, fitness level, and treadmill usage patterns.

#### **Correlation and Relationship Analysis:**

Use heatmaps and pair plots to identify correlations between continuous variables, helping to uncover relationships that affect customer preferences.

#### **Insights and Recommendations:**

Based on the data analysis, provide actionable insights and recommendations to Aerofit on how to better target specific customer segments and enhance the product offering.

## 1. Initial Data Exploration

### 1.1 Importing Libraries and imports the data from a CSV file. Printing 1st 5 rows.

#### Code:

```
# Importing necessary libraries
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt

df = pd.read_csv('aerofit.csv')

df.head()
```

#### Output:

	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

### 1.2 Data inspection.

#### Code:

```
df.info()
```

#### Output:

```
<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
```

#### Insights

##### **Data Structure:**

**Dimensions:** The Data Frame has 180 rows (entries) and 9 columns.

**Data Types:** Most columns are of integer data type (int64), indicating numerical values. Three columns (Product, Gender, and Marital Status) are of object data type, likely representing categorical or textual data.

**No Missing Data:** All columns have 180 non-null values, suggesting there are no missing data points in the dataset.

**Compact Dataset:** The Data Frame occupies 12.8 KB of memory, indicating a relatively small dataset.

### Summary:

The `df.info()` output reveals a well-structured dataset with no missing values. The majority of variables are numerical, which is common for datasets related to product usage or customer demographics. The object-type columns likely contain categorical information, such as product names, genders, or marital statuses.

## 2. Non Graphical Analysis

### 2.1 Unique values and counts for categorical attributes.

#### Code:

```
print(f'Dataset shape: {df.shape} ")

# Statistical summary of numerical features
df.describe()

# Checking for unique values and counts in categorical columns
print("Unique values and counts:")
print(df['Product'].value_counts())
print(df['Gender'].value_counts())
print(df['MaritalStatus'].value_counts())
```

#### Output:

```
Dataset shape: (180, 9)
Unique values and counts:
Product
KP281    80
KP481    60
KP781    40
Name: count, dtype: int64
Gender
Male      104
Female     76
Name: count, dtype: int64
MaritalStatus
Partnered  107
Single      73
Name: count, dtype: int64
```

#### Insights:

##### Dataset Shape:

**180 Rows, 9 Columns:** The dataset contains 180 data points (likely individuals) and 9 variables (features) describing each data point.

##### Unique Values and Counts:

###### Product:

**Three Product Variants:** There are three unique product types: KP281, KP481, and KP781.

**Distribution:** KP281 is the most popular with 80 units sold, followed by KP481 with 60 units, and KP781 with 40 units.

###### Gender:

**Female Majority:** The dataset contains more female respondents (104) than male respondents (76).

###### Marital Status:

**Single Majority:** A slightly higher number of respondents are single (107) compared to those who are partnered (73).

##### Overall Observations:

**Product Diversity:** The dataset includes data on three different product variants, allowing for analysis of product preferences or performance differences.

**Gender Balance:** While there's a slight gender imbalance, the dataset contains a reasonable representation of both genders.

**Marital Status Diversity:** The data includes information on both single and partnered individuals, enabling analysis of how marital status might influence product usage or preferences.

### 3. Outliers Diction

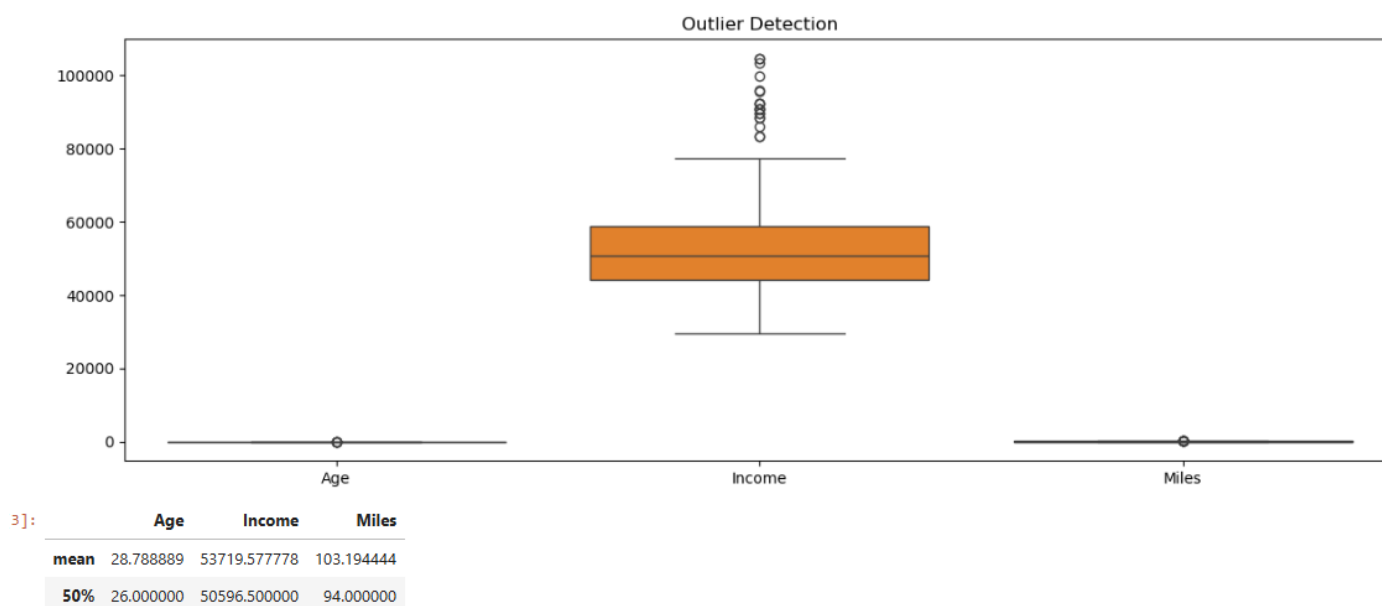
#### 3.1 Boxplot for detecting outliers in Age, Income, and Miles.

##### Code:

```
plt.figure(figsize=(15, 5))
sns.boxplot(data=df[['Age', 'Income', 'Miles']])
plt.title('Outlier Detection')
plt.show()
```

```
# Check for skewness using mean and median
df[['Age', 'Income', 'Miles']].describe().loc[['mean', '50%']]
```

##### Output:



##### Insights:

##### Outlier Detection:

- Age:** The box plot indicates a few potential outliers on the higher end of the age distribution, suggesting individuals who are significantly older than the majority of the sample.
- Income:** There are outliers on both the higher and lower ends of the income distribution, suggesting a few individuals with exceptionally high or low incomes compared to the general population.
- Miles:** The miles distribution also shows some outliers on the higher end, indicating individuals who have logged significantly more miles than the average.

##### Data Distribution:

- Skewness:** The distribution of income appears slightly skewed to the right, indicating a few individuals with very high incomes.
- Central Tendency:** The median values for age and miles are relatively close to the mean, suggesting a relatively symmetric distribution for these variables.

Based on the comparison between the mean and median values, it appears that there are **more outliers on the higher end** of the distribution for **income** and, to a lesser extent, **miles**. This aligns with the observations from the box plot.

### Summary:

The box plot effectively highlights potential outliers in the dataset, particularly for the income and miles variables. These outliers may warrant further investigation to understand their impact on the analysis and potentially exclude them if necessary. Additionally, the visualization provides insights into the overall distribution of the variables, revealing skewness in the income distribution and a relatively symmetric distribution for age and miles.

## 4. Visual Analysis (Univariate & Bivariate)

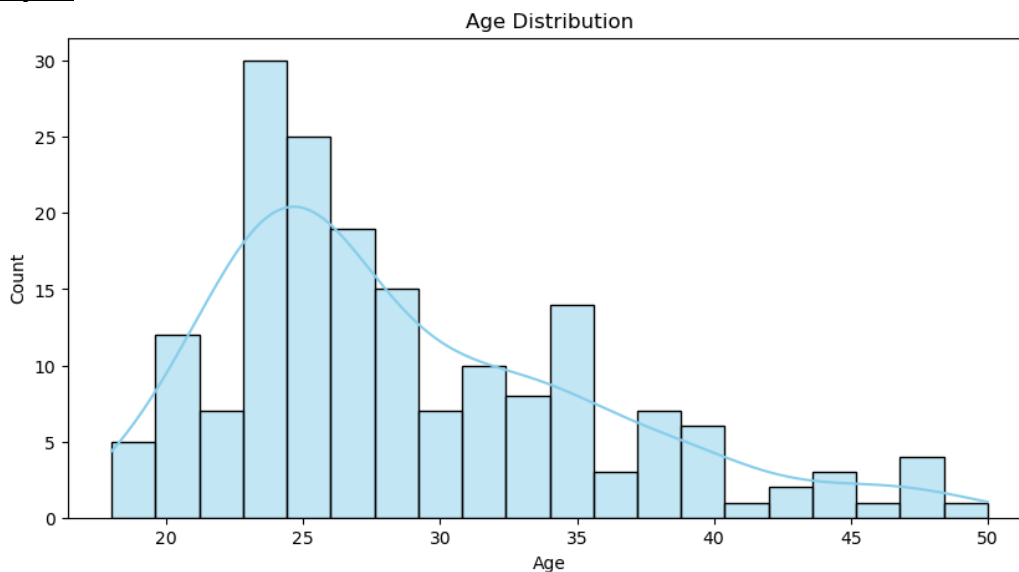
### <Univariate Analysis >

#### 4.1 Histogram for Age distribution:

##### Code:

```
plt.figure(figsize=(10, 5))
sns.histplot(df['Age'], kde=True, bins=20, color='skyblue')
plt.title('Age Distribution')
plt.xlabel('Age')
plt.show()
```

##### Output:



##### Insights:

##### Distribution Shape:

- **Right-Skewed:** The distribution is slightly right-skewed, indicating a few individuals who are significantly older than the majority of the sample. This is evident from the longer tail on the right side of the histogram.
- **Central Tendency:** The peak of the histogram appears to be around the age of 25-27, suggesting that this age group is the most frequent in the dataset.
- **Mode:** The distribution appears to be **multimodal** with two distinct peaks, suggesting the presence of two or more common age groups.

**Summary:** The age distribution histogram reveals a slightly right-skewed distribution with a central tendency around 25-27. The multimodal nature of the distribution suggests the presence of multiple common age groups. The right-skewness indicates the potential for outliers on the higher end, which should be further investigated using statistical methods.

## Recommendations:

### **Target Audience:**

**Focus on Younger Demographics:** The age distribution suggests that the majority of your users are likely between 25 and 30 years old. Tailor your marketing and product offerings to appeal to this demographic.

### **Product Features:**

**Youth-Oriented Features:** Consider incorporating features that resonate with younger users, such as social integration, gamification, or personalized recommendations.

**Customization Options:** Offer customization options to allow users to personalize their experience and make the product feel more tailored to their individual needs.

### **Pricing and Promotions:**

**Student Discounts:** If applicable, offer student discounts or promotions to attract a larger segment of the younger demographic.

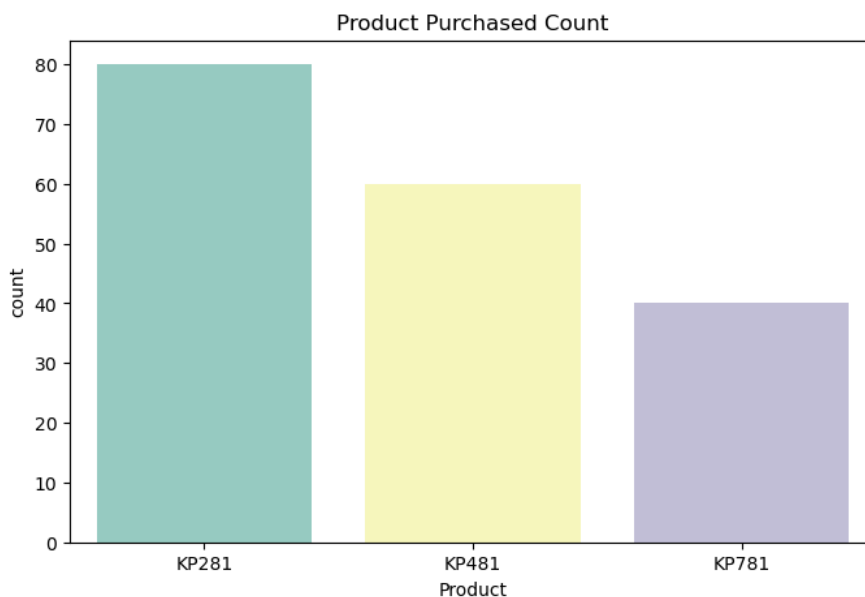
**Subscription Models:** Explore subscription-based models with flexible pricing options to cater to different budgets and usage patterns.

## 4.2 Count plot for Product Purchased

### Code:

```
plt.figure(figsize=(8, 5))  
sns.countplot(x='Product', hue='Product', data=df, palette='Set3', legend=False)  
plt.title('Product Purchased Count')  
plt.show()
```

### Output:



### Insights:

#### **Product Popularity:**

- **KP281:** This product is the most popular among customers, with 80 units sold.
- **KP481:** The second most popular product is KP481, with 60 units sold.
- **KP781:** The least popular product is KP781, with 40 units sold.

#### **Sales Distribution:**

- **Uneven Distribution:** The sales distribution is not uniform, with a significant difference in popularity between the top-selling product (KP281) and the least popular product (KP781).

## Recommendations:

### Marketing and Promotion:

- **Focus on KP281:** Continue to promote KP281 as the flagship product, highlighting its unique features and benefits.
- **Boost KP481 Sales:** Increase marketing efforts for KP481 to boost its sales and narrow the gap with KP281.
- **Re-evaluate KP781:** Consider reviewing KP781's features, pricing, or target market to understand why it is less popular. You might need to make adjustments to improve its appeal.

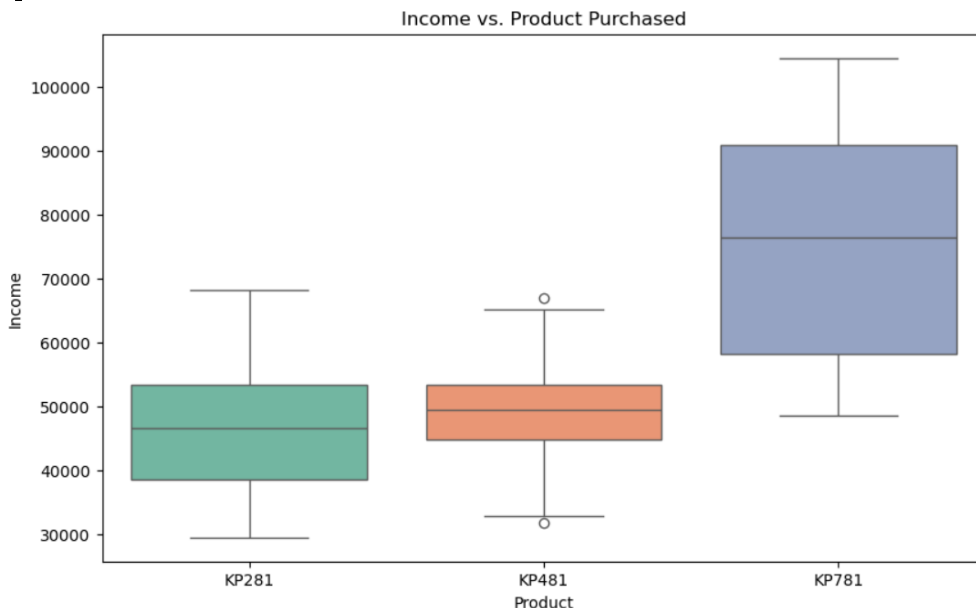
## <Bivariate Analysis>

### 4.3 Boxplot for Product Purchased vs Income.

#### Code:

```
plt.figure(figsize=(10, 6))
sns.boxplot(x='Product', y='Income', hue='Product', data=df, palette='Set2', dodge=False)
plt.title('Income vs. Product Purchased')
plt.show()
```

#### Output:



#### Insights:

##### Income Distribution:

- **KP281:** Customers who purchase KP281 tend to have a lower median income compared to the other two products.
- **KP481:** The income distribution for KP481 purchasers is slightly higher than that of KP281 but lower than KP781.
- **KP781:** Customers purchasing KP781 generally have the highest income levels among the three products.

##### Outliers:

- **KP481:** There is a potential outlier on the lower end of the income distribution for KP481, indicating a customer with a significantly lower income compared to the rest of the KP481 purchasers.

## Recommendations:

### Product Positioning and Pricing:

- **Target Different Segments:** Consider positioning KP281 as a more affordable option, while KP781 can be positioned as a premium product.
- **Adjust Pricing:** Adjust the pricing of each product to align with the income levels of the target customer segments.

### Customer Segmentation:

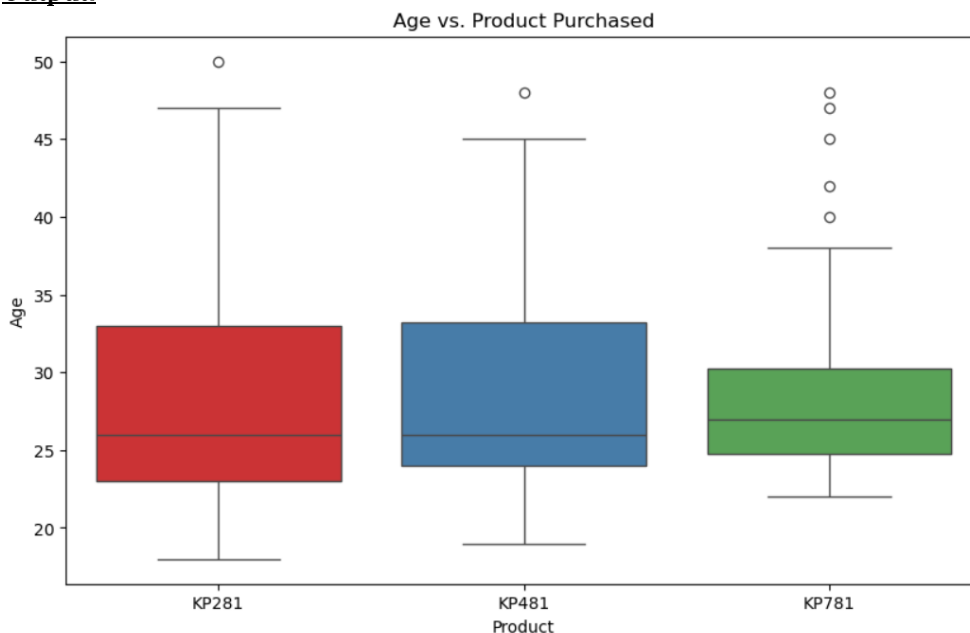
- **Identify Segments:** Use customer segmentation techniques to identify specific customer segments within each product category based on income and other relevant factors.
- **Tailored Offerings:** Develop tailored product offerings or promotions to address the needs and preferences of each segment.

## 4.4 Boxplot for Product Purchased vs Age

### Code:

```
plt.figure(figsize=(10, 6))
sns.boxplot(x='Product', y='Age', hue='Product', data=df, palette='Set1', dodge=False)
plt.title('Age vs. Product Purchased')
plt.show()
```

### Output:



### Insights:

#### Age Distribution:

- **KP281:** Purchasers of KP281 tend to be slightly younger than those of the other two products.
- **KP481:** The age distribution for KP481 purchasers is similar to that of KP281, with a slight overlap.
- **KP781:** Customers purchasing KP781 are generally older than those purchasing the other two products.

#### Outliers:

- **KP281 and KP481:** There are a few outliers on the higher end of the age distribution for KP281 and KP481, indicating some older individuals who purchased these products.



## Recommendations:

### Product Positioning and Targeting:

- **Youth-Oriented:** Position KP281 and KP481 as more youth-oriented products, targeting younger demographics.
- **Mature Audience:** Target KP781 towards a more mature audience.

### Marketing Strategies:

- **Highlight Youth Appeal:** Emphasize features and benefits that resonate with younger audiences for KP281 and KP481.
- **Focus on Experience and Features:** Highlight the unique features and benefits of KP781 that appeal to older customers, such as comfort, durability, or advanced technology.

### Product Development:

- **Age-Specific Features:** Consider incorporating features that cater to the specific needs and preferences of different age groups. For example, you might offer more advanced features for older customers or focus on portability and convenience for younger users.

### Customer Segmentation:

- **Identify Segments:** Use customer segmentation techniques to identify specific age groups within each product category.
- **Tailored Offerings:** Develop tailored product offerings or promotions to address the needs and preferences of each segment.

## 4.5 Count plot for Gender vs. Product Purchased

### Code:

```
plt.figure(figsize=(8, 5))
sns.countplot(x='Product', hue='Gender', data=df, palette='Set2')
plt.title('Product Purchased by Gender')
plt.show()
```

### Output:



## Insights:

## Gender Preferences:

- **KP281:** There is a slightly higher preference for KP281 among males compared to females.
- **KP481:** The distribution of KP481 purchases is relatively balanced between males and females.
- **KP781:** There is a significant preference for KP781 among males compared to females.

## Recommendations:

### Gender-Specific Marketing:

- **Tailored Messaging:** Develop gender-specific marketing messages that resonate with the preferences of each gender.
- **Highlight Unique Features:** Emphasize features that are particularly appealing to males for KP281 and KP781, and highlight features that resonate with females for KP481.

### Product Development:

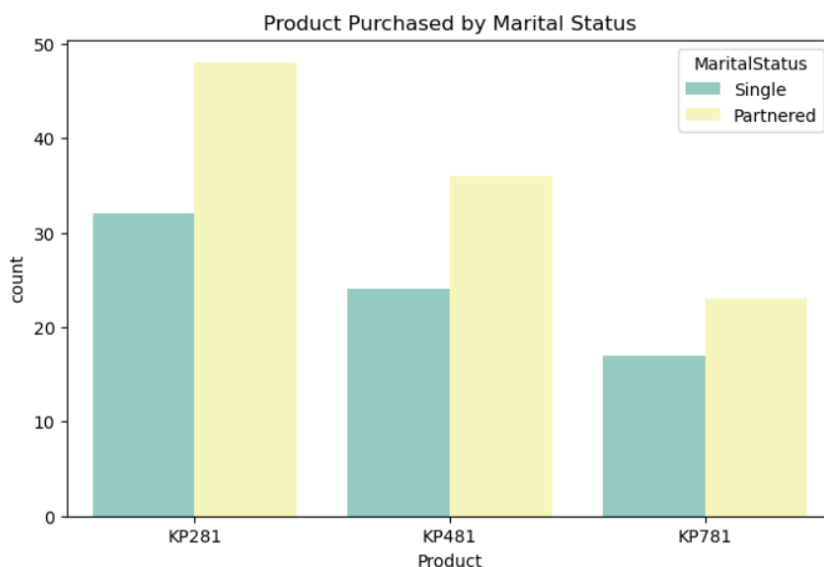
- **Consider Gender-Specific Needs:** Explore opportunities to incorporate features or design elements that cater to the specific needs or preferences of each gender.

## 4.6 Count plot for Marital Status vs. Product Purchased.

### Code:

```
plt.figure(figsize=(8, 5))
sns.countplot(x='Product', hue='MaritalStatus', data=df, palette='Set3')
plt.title('Product Purchased by Marital Status')
plt.show()
```

### Output:



### Insights:

#### Marital Status Preferences:

- **KP281:** There is a slightly higher preference for KP281 among partnered individuals compared to singles.
- **KP481:** The distribution of KP481 purchases is relatively balanced between single and partnered individuals.
- **KP781:** There is a significant preference for KP781 among partnered individuals compared to singles.

## Recommendations:

### Product Positioning and Marketing:

- **Target Specific Segments:** Consider positioning KP281 and KP781 as products that appeal to partnered individuals, while KP481 might be suitable for both single and partnered customers.
- **Tailored Messaging:** Develop marketing messages that resonate with each segment. For example, you could emphasize features that are relevant to couples for KP281 and KP781, while focusing on individual needs and preferences for KP481.

### Product Development:

- **Consider Couple-Specific Features:** Explore opportunities to incorporate features that cater to the needs of couples, such as dual-user modes or personalized settings.
- **Individual-Focused Features:** For KP481, consider features that are appealing to individuals, such as portability or customization options.

## 5. Correlation Analysis

### 5.1 Heat map to visualize correlations among numerical variables like Age, Usage, Fitness, Income, and Miles. Code:

```
from sklearn.preprocessing import LabelEncoder

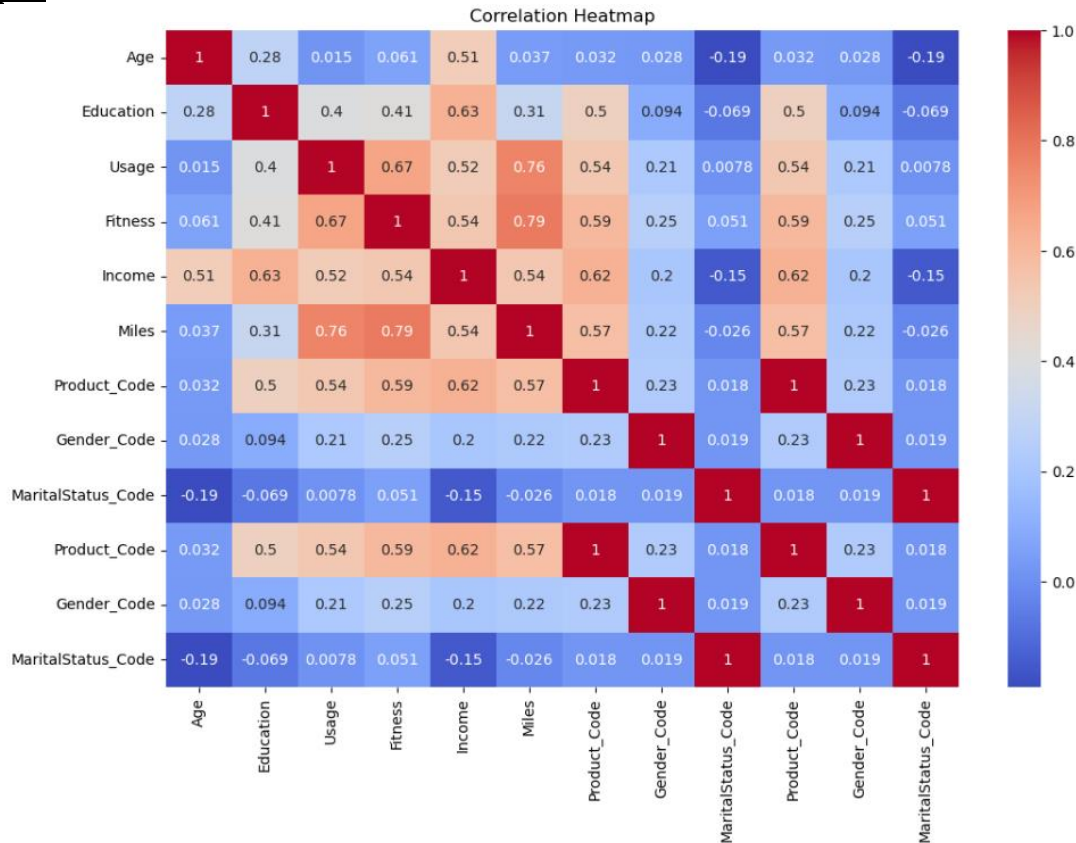
# Convert categorical variables to numeric codes
df['Product_Code'] = df['Product'].astype('category').cat.codes
df['Gender_Code'] = df['Gender'].astype('category').cat.codes
df['MaritalStatus_Code'] = df['MaritalStatus'].astype('category').cat.codes

# Drop the original categorical columns to avoid errors during correlation calculation
numeric_df = df.drop(['Product', 'Gender', 'MaritalStatus'], axis=1)

# Add the newly created numerical columns
numeric_df = pd.concat([numeric_df, df[['Product_Code', 'Gender_Code', 'MaritalStatus_Code']]], axis=1)

# Compute the correlation matrix
plt.figure(figsize=(12, 8))
sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()
```

## Output:



## Insights:

### Key Correlations

- **Age and Income:** There is a moderate positive correlation (0.51) between age and income, suggesting that older customers tend to have higher income levels. This could imply that marketing efforts for high-end products could focus on older demographics.
- **Usage and Miles:** A strong positive correlation (0.79) exists between usage and miles, indicating that customers who use the treadmill more frequently tend to log more miles. This insight can guide user engagement strategies to improve customer satisfaction and retention.
- **Education and Income:** A moderate correlation (0.63) between education and income suggests that higher educational levels are associated with higher income. Tailored marketing messages that resonate with educated individuals may be effective.
- **Fitness and Usage:** The correlation (0.67) indicates that customers who are more fitness-oriented tend to use the treadmill more. This can be leveraged for targeted promotions to fitness enthusiasts.

## Recommendations:

- **Targeted Marketing:** Focus marketing campaigns on older customers with higher income, as they show a higher likelihood of purchasing fitness equipment. Highlight benefits such as improved health and convenience for those in this demographic.
- **Engagement Strategies:** Implement programs that encourage higher usage among existing customers. This could include challenges, community events, or loyalty rewards based on the miles logged.
- **Educational Outreach:** Since education correlates with income, providing information on the long-term health benefits of fitness equipment may appeal more to educated customers. Consider partnerships with fitness trainers or institutions for workshops.
- **Product Development:** Given the strong correlation between fitness and usage, consider developing products or services specifically designed for fitness enthusiasts, such as advanced tracking features or community engagement platforms.

- **Customer Segmentation:** Use the correlation insights to create customer segments. For example, a segment of higher-income, older individuals could be targeted with premium product offerings, while fitness-focused customers might respond better to discounts on accessories or memberships.

**Conclusion:** The correlation analysis provides valuable insights into customer behaviors and preferences. By aligning marketing, product development, and customer engagement strategies with these insights, AeroFit can enhance its customer relationships and increase sales. Implementing these recommendations can lead to actionable strategies that are data-driven and customer-centric.

## 6. Probability Analyses.

### 6.1 Creating a contingency table for Gender vs. Product Purchased.

#### Code:

```
contingency_table = pd.crosstab(df['Gender'], df['Product'])
print("\nContingency Table (Gender vs. Product Purchased):\n", contingency_table)

# Marginal probability of each treadmill model
marginal_prob = contingency_table.sum(axis=0) / contingency_table.values.sum()
print("\nMarginal Probabilities:\n", marginal_prob)

# Conditional probability: Probability of a male buying KP781
prob_male_kp781 = contingency_table.loc['Male', 'KP781'] / contingency_table['KP781'].sum()
print(f"\nProbability of a male buying KP781: {prob_male_kp781:.2f}")
```

#### Output:

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

Marginal Probabilities:
Product
KP281    0.444444
KP481    0.333333
KP781    0.222222
dtype: float64

Probability of a male buying KP781: 0.82
```

#### Insights:

- **KP281:** There is no significant difference in preference for KP281 between males and females.
- **KP481:** There is a slight preference for KP481 among females compared to males.
- **KP781:** There is a significant preference for KP781 among males compared to females.
- **High Preference:** Males have a strong preference for KP781 compared to the other products.
- **Targeted Marketing:** This information can be used to tailor marketing efforts specifically towards male customers, highlighting the features and benefits of KP781 that appeal to them.
- **Product Positioning:** KP781 can be positioned as a product that is particularly appealing to male customers.

## 6.1 Marginal Probability Calculation Using pandas.crosstab.

### Code:

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

### Output:

```
col_0      Count
Product
KP281    44.444444
KP481    33.333333
KP781    22.222222
```

---

### Insights:

#### Product Distribution:

- **KP281:** 44.44% of purchases are for KP281.
- **KP481:** 33.33% of purchases are for KP481.
- **KP781:** 22.22% of purchases are for KP781.

#### Market Share:

- **KP281:** KP281 has the highest market share among the three products, accounting for almost half of all purchases.
- **KP481:** KP481 holds a significant market share, with approximately one-third of all purchases.
- **KP781:** KP781 has the lowest market share, representing around one-fifth of all purchases.

### Recommendations

- **Focus on KP281:** Given its dominant market share, continue to promote KP281 as the flagship product.
- **Boost KP481 Sales:** Increase marketing efforts for KP481 to enhance its market share and narrow the gap with KP281.
- **Re-evaluate KP781:** Consider reviewing KP781's features, pricing, or target market to understand why it has a lower market share. You might need to make adjustments to improve its appeal.

## 6.2 Conditional Probability Calculation for male customer buying KP781.

### Code:

```
# Conditional probability: Probability of a male customer buying KP781
prob_male_kp781 = len(df[(df['Gender'] == 'Male') & (df['Product'] == 'KP781')]) / len(df[df['Gender'] == 'Male'])
print(f'Probability of a male customer buying KP781: {prob_male_kp781:.2f}')
```

### Output:

```
Probability of a male customer buying KP781: 0.32
```

### Insights:

The probability of a male customer buying KP781 is 0.32. This means that there's a 32% chance that a randomly selected male customer will purchase KP781.

## Implications:

- **Moderate Preference:** While there is a preference for KP781 among male customers, it's not as strong as the preference for KP781 among all customers (which was 0.82).
- **Other Factors:** Other factors besides gender might influence a male customer's decision to purchase KP781, such as age, income, or product features.

## Recommendations:

- **Segment Male Customers:** Consider segmenting male customers based on other factors (e.g., age, income) to identify specific sub-segments that might have a higher preference for KP781.
- **Highlight Unique Features:** Emphasize the unique features of KP781 that appeal to male customers, especially those that differentiate it from other products in the same category.
- **Targeted Promotions:** Offer promotions or discounts specifically targeted towards male customers who might be more likely to purchase KP781.
- **Product Development:** Explore opportunities to incorporate additional features or design elements that appeal to male customers and further enhance KP781's appeal.

**6.3 Purpose:** To understand product preferences among different genders, which helps target marketing strategies accordingly

**Analysis:** Calculate the probability of purchasing each product given the gender (e.g., the probability that a customer is female given they purchased KP481).

## Code:

```
# Probability of purchasing KP281 given the customer is Female
prob_kp281_given_female = 40 / (40 + 29 + 7)
print(f"Probability of purchasing KP281 given Female: {prob_kp281_given_female:.2f}")

# Probability of purchasing KP481 given the customer is Male
prob_kp481_given_male = 31 / (40 + 31 + 33)
print(f"Probability of purchasing KP481 given Male: {prob_kp481_given_male:.2f}")
```

## Output:

```
Probability of purchasing KP281 given Female: 0.53
Probability of purchasing KP481 given Male: 0.30
```

## Insights:

- **Probability of purchasing KP281 given Female: 0.53:** This indicates that there's a 53% chance that a female customer will purchase KP281.
- **Probability of purchasing KP481 given Male: 0.30:** This suggests that there's a 30% chance that a male customer will purchase KP481.
- **Female Preference for KP281:** Females have a higher probability of purchasing KP281 compared to males.
- **Male Preference for KP481:** While the probability is lower, males show a higher preference for KP481 compared to females.

## Recommendations:

### Marketing Strategies:

- **Target KP281 to Females:** Focus marketing efforts for KP281 towards female customers, highlighting features that resonate with their preferences.
- **Target KP481 to Males:** Emphasize features of KP481 that appeal to male customers, tailoring marketing messages accordingly.



#### 6.4 Joint Probability of Gender and Product:

**Purpose:** Joint probabilities reveal the overall likelihood of specific combinations occurring, like being male and purchasing KP781.

**Analysis:** Calculate the joint probability of gender and product purchases to determine which combinations are most common.

##### Code:

```
# Joint probability of being Female and purchasing KP281
prob_female_and_kp281 = 40 / 180
print(f'Joint Probability of Female and KP281: {prob_female_and_kp281:.2f}')
```

```
# Joint probability of being Male and purchasing KP781
prob_male_and_kp781 = 33 / 180
print(f'Joint Probability of Male and KP781: {prob_male_and_kp781:.2f}')
```

##### Output:

```
Joint Probability of Female and KP281: 0.22
Joint Probability of Male and KP781: 0.18
```

##### Insights:

- **Joint Probability of Female and KP281: 0.22:** This indicates that 22% of the customers are female and purchased KP281.
- **Joint Probability of Male and KP781: 0.18:** This suggests that 18% of the customers are male and purchased KP781.
- **KP281 is Popular Among Females:** The higher joint probability of Female and KP281 suggests that KP281 is more popular among female customers.
- **KP781 is Popular Among Males:** The higher joint probability of Male and KP781 indicates that KP781 is more popular among male customers.

##### Recommendations:

- **Target Marketing:** Tailor marketing strategies for each product to the respective gender that shows a higher preference. For example, focus on promoting KP281 to female customers and KP781 to male customers.
- **Product Development:** Consider incorporating features or design elements that cater to the specific preferences of the gender that shows a higher affinity for each product.
- **Customer Segmentation:** Use customer segmentation techniques to identify specific segments within each gender based on product preferences and other relevant factors.
- **Tailored Offerings:** Develop tailored product offerings or promotions to address the needs and preferences of each segment.

#### 6.5 Probability of Product Purchase by Age Group:

**Purpose:** Helps understand how age influences product choice, allowing targeted promotions for specific age groups.

**Analysis:** Segment customers into age groups (e.g., <30, 30-50, >50) and calculate the probability of each product being purchased within these segments.

##### Code:

```
# Create age groups
df['Age_Group'] = pd.cut(df['Age'], bins=[0, 30, 50, 100], labels=['<30', '30-50', '>50'])
```

```
# Probability of purchasing KP281 for age group 30-50
prob_kp281_age_30_50 = len(df[(df['Age_Group'] == '30-50') & (df['Product'] == 'KP281')]) / len(df[df['Age_Group'] == '30-50'])
print(f'Probability of purchasing KP281 for age group 30-50: {prob_kp281_age_30_50:.2f}')
```



### Output:

```
Probability of purchasing KP281 for age group 30-50: 0.42
```

### Insights:

The probability of purchasing KP281 for the age group 30-50 is 0.42. This means that there's a 42% chance that a randomly selected customer aged 30-50 will purchase KP281.

- **Moderate Preference:** While there is a preference for KP281 among this age group, it's not overwhelmingly high.
- **Other Factors:** Other factors besides age might influence a customer's decision to purchase KP281, such as income, gender, or product features.

### Recommendations:

- **Targeted Marketing:** Focus marketing efforts for KP281 towards the 30-50 age group, highlighting features that resonate with their preferences.
- **Product Development:** Consider incorporating features that appeal to this age group, such as those related to lifestyle, health, or convenience.
- **Customer Segmentation:** Segment the 30-50 age group further based on other factors (e.g., income, gender) to identify specific sub-segments that might have a higher preference for KP281.
- **Pricing Strategy:** Evaluate the pricing of KP281 to ensure it aligns with the preferences and budget of the 30-50 age group.

## **6.6 Probability of High Fitness Level Given Product Purchased:**

**Purpose:** Determines which product is more likely to be chosen by customers with high fitness levels, aiding in product positioning.

**Analysis:** Calculate the probability of customers having high fitness (rating 4 or 5) for each treadmill type.

### Code:

```
# Probability of having high fitness given KP781 purchase
prob_high_fitness_given_kp781 = len(df[(df['Fitness'] >= 4) & (df['Product'] == 'KP781')]) / len(df[df['Product'] == 'KP781'])
print(f"Probability of high fitness given KP781 purchase: {prob_high_fitness_given_kp781:.2f}")
```

### Output:

```
Probability of high fitness given KP781 purchase: 0.90
```

### Insights:

The probability of high fitness given a KP781 purchase is **0.90**. This means that there's a 90% chance that a customer who purchased KP781 has a high fitness level (rated 4 or 5).

- **Strong Association:** There's a strong association between purchasing KP781 and having a high fitness level.
- **Product Positioning:** KP781 can be positioned as a treadmill that appeals to customers with high fitness goals.

### Recommendations:

- **Target High-Fitness Customers:** Focus marketing efforts for KP781 towards individuals with high fitness levels.
- **Highlight Fitness Features:** Emphasize the features of KP781 that are particularly appealing to fitness enthusiasts, such as advanced workout programs, durability, or performance metrics.
- **Product Development:** Consider incorporating additional features that cater to the needs of high-fitness customers, such as advanced resistance levels or heart rate monitoring.
- **Partnerships:** Collaborate with fitness influencers or organizations to promote KP781 to a target audience of fitness enthusiasts.

## 6.7 Probability of High Income Given Specific Product Purchase:

**Purpose:** Identifies the income levels of customers likely to purchase each treadmill type, aiding in defining target segments by income.

**Analysis:** Calculate the probability of having a high income (e.g., >\$60,000) for each product.

### Code:

```
# Probability of high income (> $60,000) given KP781 purchase
prob_high_income_given_kp781 = len(df[(df['Income'] > 60000) & (df['Product'] == 'KP781')]) / len(df[df['Product'] == 'KP781'])
print(f"Probability of high income given KP781 purchase: {prob_high_income_given_kp781:.2f}")
```

### Output:

```
Probability of high income given KP781 purchase: 0.72
```

### Insights:

The probability of high income given a KP781 purchase is 0.72. This means that there's a 72% chance that a customer who purchased KP781 has a high income (defined as >\$60,000 in this case).

- **Strong Association:** There's a strong association between purchasing KP781 and having a high income.
- **Product Positioning:** KP781 can be positioned as a premium product that appeals to customers with higher incomes.

### Recommendations:

- **Target High-Income Customers:** Focus marketing efforts for KP781 towards individuals with high incomes.
- **Highlight Premium Features:** Emphasize the premium features and benefits of KP781 that are attractive to high-income customers.
- **Pricing Strategy:** Ensure that the pricing of KP781 aligns with the expectations of high-income customers, justifying its premium positioning.
- **Product Development:** Consider incorporating additional features or design elements that appeal to high-income customers, such as luxury materials or advanced technology.
- **Partnerships:** Collaborate with brands or retailers that cater to high-income customers to expand your reach and enhance brand perception.

## 6.8 Conditional Probability of Purchasing Based on Usage:

**Purpose:** Helps identify which treadmill is favored by customers who plan higher usage, informing recommendations based on exercise frequency.

**Analysis:** Calculate the probability of purchasing each product given high usage (e.g., using the treadmill more than 3 times a week).

### Code:

```
# Probability of purchasing KP481 given usage > 3 times a week
prob_kp481_high_usage = len(df[(df['Usage'] > 3) & (df['Product'] == 'KP481')]) / len(df[df['Usage'] > 3])
print(f"Probability of purchasing KP481 given high usage: {prob_kp481_high_usage:.2f}")
```

### Output:

```
Probability of purchasing KP481 given high usage: 0.19
```

## Insights:

### Interpretation:

The probability of purchasing KP481 given high usage is 0.19. This means that there's a 19% chance that a customer who uses the treadmill more than 3 times a week will purchase KP481.

- **Moderate Association:** There is a moderate association between high usage and purchasing KP481. While it's not the strongest association, it still suggests that KP481 might be a suitable choice for customers who plan to use the treadmill frequently.

### Recommendations:

- **Target High-Usage Customers:** Focus marketing efforts for KP481 towards customers who plan to use the treadmill frequently. Highlight features that are relevant to high-usage customers, such as durability, advanced settings, or maintenance-free operation.
- **Product Development:** Consider incorporating features that cater to the needs of high-usage customers, such as enhanced durability, extended warranties, or personalized workout programs.
- **Pricing Strategy:** Evaluate the pricing of KP481 to ensure it aligns with the expectations of customers who plan to use the treadmill frequently.
- **Customer Segmentation:** Segment customers based on their planned usage frequency and other relevant factors to identify specific target segments.

## 7. Customer Profiling.

### 7.1 Customer Profile for KP281:

#### Code:

```
# Profiling customers based on product purchased
profile_kp281 = df[df['Product'] == 'KP281'].describe()
profile_kp481 = df[df['Product'] == 'KP481'].describe()
profile_kp781 = df[df['Product'] == 'KP781'].describe()

print("\nCustomer Profile for KP281:\n", profile_kp281)
print("\nCustomer Profile for KP481:\n", profile_kp481)
print("\nCustomer Profile for KP781:\n", profile_kp781)
```

#### Output:

#### Insights:

##### Demographic Characteristics:

- **Age:** The average age of KP281 customers is 28.55, with a standard deviation of 7.22. This indicates a relatively young customer base.
- **Education:** The average education level is 15.04, suggesting a moderately educated customer group.
- **Income:** The average income is 46,418.03, which is a relatively moderate income level.

##### Usage and Fitness:

- **Usage:** The average usage is 3.09 times per week, indicating moderate usage.
- **Fitness:** The average fitness rating is 2.96 out of 5, suggesting a moderate level of fitness among customers.

## Other Observations:

- **Distribution:** The data shows a relatively even distribution of ages, with a slight skew towards the younger side.
- **Outliers:** There might be a few outliers in the income and miles data, as indicated by the standard deviation and the range between minimum and maximum values.

## 7.2 Customer Profile for KP481:

### Customer Profile for KP481:

	Age	Education	Usage	Fitness	Income	Miles
count	60.000000	60.000000	60.000000	60.000000	60.000000	60.000000
mean	28.900000	15.116667	3.066667	2.900000	48973.650000	87.933333
std	6.645248	1.222552	0.799717	0.62977	8653.989388	33.263135
min	19.000000	12.000000	2.000000	1.000000	31836.000000	21.000000
25%	24.000000	14.000000	3.000000	3.000000	44911.500000	64.000000
50%	26.000000	16.000000	3.000000	3.000000	49459.500000	85.000000
75%	33.250000	16.000000	3.250000	3.000000	53439.000000	106.000000
max	48.000000	18.000000	5.000000	4.000000	67083.000000	212.000000

	Product_Code	Gender_Code	MaritalStatus_Code
count	60.0	60.000000	60.000000
mean	1.0	0.516667	0.400000
std	0.0	0.503939	0.494032
min	1.0	0.000000	0.000000
25%	1.0	0.000000	0.000000
50%	1.0	1.000000	0.000000
75%	1.0	1.000000	1.000000
max	1.0	1.000000	1.000000

### 1. Age Distribution:

- **Mean Age:** The average age of KP481 customers is 28.9 years, indicating that the product appeals primarily to younger individuals.
- **Age Range:** The minimum age is 19, and the maximum is 48, suggesting that the product is being purchased by both younger and middle-aged customers.
- **Recommendation:** KP481 seems to appeal more to a younger audience. Marketing efforts could be directed toward this age group, possibly highlighting features like convenience for busy lifestyles, digital connectivity, or price affordability.

### 2. Education:

- **Mean Education:** The average years of education for KP481 buyers is about 15.1 years, which generally corresponds to individuals with at least some college education or a college degree.
- **Recommendation:** Since this demographic appears well-educated, emphasizing advanced features of the KP481, such as technology integration or fitness monitoring, could resonate well with this group.

### 3. Usage:

- **Mean Usage:** On average, customers use the treadmill 3 times per week, which indicates moderate usage.
- **Recommendation:** Engaging customers through post-purchase communication, like personalized workout plans or usage reminders, could increase treadmill utilization. Offering loyalty programs based on usage might also boost engagement.

## Overall Recommendations:

- **Target Audience:** The KP481 treadmill is most popular among young, moderately educated, fitness-oriented customers with moderate income levels. Focus on this demographic when designing marketing campaigns.
- **Product Positioning:** Position the KP481 as a treadmill suitable for intermediate users who seek affordable, reliable fitness equipment. Highlight features that support a healthy lifestyle without being too technical or advanced.
- **Engagement Strategies:** Post-purchase engagement can include reminders for usage, targeted email campaigns on fitness tips, and the introduction of community or group fitness challenges to encourage more frequent use.

## 7.3 Customer Profile for KP781:

Customer Profile for KP781:

	Age	Education	Usage	Fitness	Income	Miles
count	40.000000	40.000000	40.000000	40.000000	40.000000	40.000000
mean	29.100000	17.325000	4.775000	4.625000	75441.57500	166.900000
std	6.971738	1.639066	0.946993	0.667467	18505.83672	60.066544
min	22.000000	14.000000	3.000000	3.000000	48556.00000	80.000000
25%	24.750000	16.000000	4.000000	4.000000	58204.75000	120.000000
50%	27.000000	18.000000	5.000000	5.000000	76568.50000	160.000000
75%	30.250000	18.000000	5.000000	5.000000	90886.00000	200.000000
max	48.000000	21.000000	7.000000	5.000000	104581.00000	360.000000

	Product_Code	Gender_Code	MaritalStatus_Code
count	40.0	40.000000	40.000000
mean	2.0	0.825000	0.425000
std	0.0	0.384808	0.500641
min	2.0	0.000000	0.000000
25%	2.0	1.000000	0.000000
50%	2.0	1.000000	0.000000
75%	2.0	1.000000	1.000000
max	2.0	1.000000	1.000000

## Insights:

1. **Age:** The average age is 29.1 years, indicating the target audience consists mainly of young to middle-aged adults, active and health-conscious.
2. **Education:** Customers have an average of 17.3 years of education, meaning most hold higher education degrees, making them well-informed consumers likely to value quality and advanced features.
3. **Usage:** The treadmill is used frequently (around 4.78 times per week), suggesting it is popular among dedicated fitness users who value consistent, high-intensity workouts.
4. **Fitness:** High fitness levels, with an average score of 4.63 out of 5, imply that this model is favored by fitness enthusiasts aiming for advanced training.
5. **Income:** Customers are affluent, with an average income of ₹75,442. This reflects a high purchasing power, positioning KP781 as a premium product for customers who are willing to invest in superior fitness equipment.
6. **Miles Run:** Customers log a significant average of 166.9 miles, indicating the treadmill is used for long-distance training, making it appealing to serious runners.
7. **Gender:** The profile is predominantly male (82.5%), suggesting that marketing campaigns should cater more to male customers.
8. **Marital Status:** 42.5% are married, which could suggest a potential market for dual-income households looking to invest in high-end home gym equipment.

**Recommendations:** KP781 should be promoted as a high-performance treadmill for well-educated, affluent, and fitness-driven customers. Focus on advertising its durability, advanced features, and suitability for intense and long-distance training. Marketing should emphasize premium quality, targeting male professionals and fitness enthusiasts.

### **Final Conclusion:**

The analysis of AeroFit treadmill data provides valuable insights into customer preferences, demographic trends, and product performance across various segments. It reveals that older and higher-income customers prefer the premium KP781 model, while younger, more cost-conscious individuals gravitate towards KP281 and KP481. KP781 is especially favored by male customers and partnered individuals, indicating its appeal to a more affluent, older demographic, while KP481 serves as a balanced choice for both genders and moderate-income groups. KP281 stands out as the most popular product, catering to younger customers seeking affordability. Additionally, frequent treadmill users and fitness-conscious individuals are more likely to choose KP781, highlighting its appeal to serious fitness enthusiasts.

### **Final Recommendation:**

Based on these insights, AeroFit should focus on tailored marketing campaigns. KP781 can be positioned as a high-end, advanced treadmill for affluent, older males, emphasizing its durability and advanced features for intensive workouts. KP481 should be marketed as a balanced, gender-neutral option for moderate-income, health-conscious buyers, while KP281 should target younger, budget-conscious consumers by promoting its affordability and value for money. Additionally, AeroFit can consider gender-specific features and advanced safety options for KP781 to cater to its older customer base. To expand the customer base, offering financing options for KP281 and KP481 could attract more buyers. Lastly, targeted promotions for fitness enthusiasts who use the treadmill frequently could enhance customer engagement and satisfaction, leading to increased brand loyalty and sales across all product lines.

## **Jupyter Notebook Analysis**

For a detailed view of the full analysis, including code, visualizations, and insights, please refer to the complete Jupyter notebook available in the PDF format. The notebook documents each step of the analysis process, from data exploration to the final recommendations.

You can access the Jupyter notebook PDF through the **following link**:

[Airofit TredAnalysis](#)