Business Case: [Aerofit](#) - Descriptive Statistics & Probability

Table of Contents

---

[1] Buisness_Case_Aerofit.ipynb

Data Exploration and Preprocessing

1.a) Check Data Structure

Data Structure Overview

The dataset consists of 180 rows and 9 columns, containing information on individuals who purchased treadmills from Aerofit. Here's a breakdown of the structure:

- Total Entries: 180
- Total Columns: 9

Each column represents a specific feature related to customer demographics, treadmill usage patterns, and preferences.

Data Columns

1. **Product (Object)**: Indicates the model of treadmill purchased (KP281, KP481, KP781).
2. **Age (int64)**: Age of the customer in years.
3. **Gender (Object)**: Gender of the customer (Male/Female).
4. **Education (int64)**: Number of years of education completed by the customer.
5. **MaritalStatus (Object)**: Customer's marital status (Single/Partnered).
6. **Usage (int64)**: Average number of times the customer plans to use the treadmill per week.
7. **Fitness (int64)**: Self-rated fitness on a scale of 1 (poor) to 5 (excellent).
8. **Income (int64)**: Annual income of the customer in dollars.
9. **Miles (int64)**: Average number of miles the customer expects to run or walk per week.

1.b) Handling Missing Data

The dataset has no missing values across all columns. Each feature is fully populated, ensuring a complete dataset for analysis.This clean dataset allows for robust analysis without the need for any imputation or handling of missing data.

1.c) Outlier Detection

In this set of boxplots, we visualize the distribution and presence of outliers for four continuous variables: Age, Income, Usage, and Miles.

1. **Age**:
   - The distribution is fairly uniform, with minimal outliers, suggesting that most customers fall within a similar age range.
2. **Income**:
   - This boxplot shows significant outliers, with a wider spread in income levels. This indicates that a small subset of customers has much higher income levels compared to the majority.
3. **Usage**:
   - The usage data appears relatively consistent, with few outliers. Most customers tend to use the treadmill a similar number of times per week.
4. **Miles**:
   - Multiple outliers are present in the miles data. These outliers represent customers who expect to run considerably more miles than the average user.
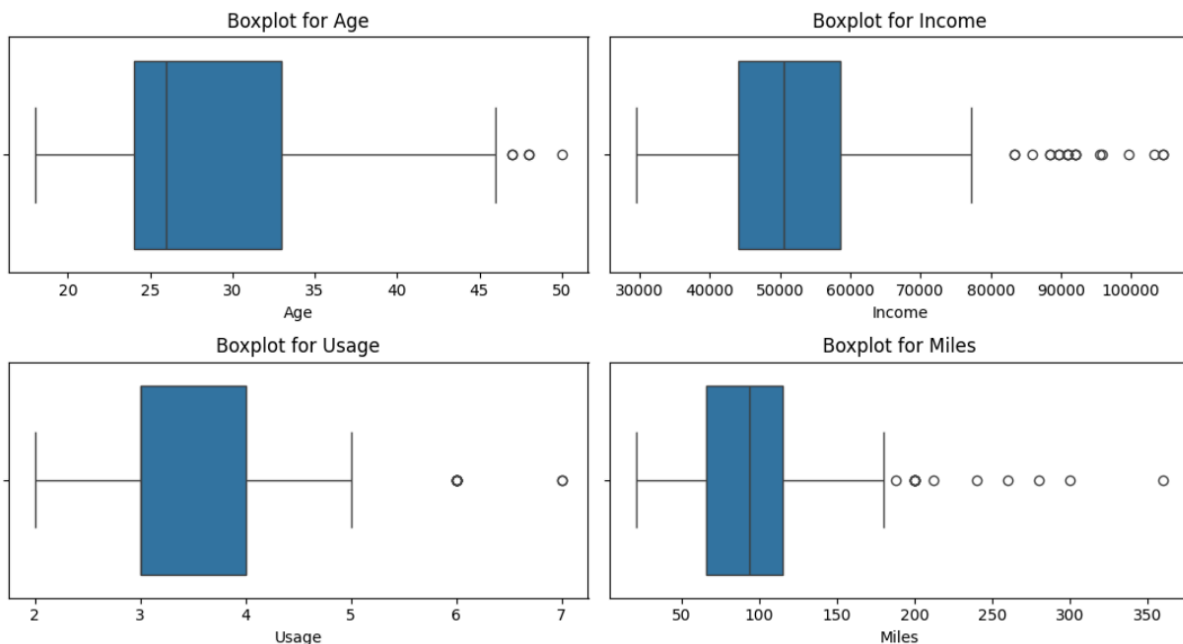
Descriptive Analytics and Customer Profiling

2.a) Summary Statistics

The table highlights the **mean**, **median**, and **mode** values for continuous variables (Age, Usage, Income, Fitness, Miles, Education) across three treadmill products (KP281, KP481, and KP781). Here are the insights:

```
                          Mean     Median      Mode
Product
KP281   Age            28.550000     26.0      23.0
        Usage           3.087500      3.0       3.0
        Income      46418.025000  46617.0   46617.0
        Fitness         2.962500      3.0       3.0
        Miles          82.787500     85.0      85.0
        Education      15.037500     16.0      16.0
KP481   Age            28.900000     26.0      25.0
        Usage           3.066667      3.0       3.0
        Income      48973.650000  49459.5   45480.0
        Fitness         2.900000      3.0       3.0
        Miles          87.933333     85.0      95.0
        Education      15.116667     16.0      16.0
KP781   Age            29.100000     27.0      25.0
        Usage           4.775000      5.0       4.0
        Income      75441.575000  76568.5   90886.0
        Fitness         4.625000      5.0       5.0
        Miles         166.900000    160.0     100.0
        Education      17.325000     18.0      18.0
```

1. **KP281 Customers**:
   - **Age**: Average is around **28.5 years**, mostly young to middle-aged.
   - **Usage**: They use the treadmill **about 3 times a week**.
   - **Income**: Average income is around **$46,418**, relatively lower compared to other products.
   - **Fitness**: Fitness rating is around **3**, indicating moderate fitness levels.
   - **Miles**: Customers expect to run about **83 miles per week**.
   - **Education**: Average education level is around **15 years**.
2. **KP481 Customers**:
   - **Age**: Similar to KP281 with an average age of **28.9 years**.
   - **Usage**: They also use the treadmill around **3 times a week**.
   - **Income**: Average income is **$48,974**, slightly higher than KP281.
   - **Fitness**: Fitness rating is also **3**, showing moderate levels.
   - **Miles**: Expected to run about **88 miles per week**.
   - **Education**: Similar education level as KP281.
3. **KP781 Customers**:
   - **Age**: Slightly older with an average age of **29.1 years**.
   - **Usage**: Higher treadmill usage of around **4.8 times per week**.
   - **Income**: Significantly higher income, with an average of **$75,442**.
   - **Fitness**: Higher fitness level, with an average rating of **4.6**.
   - **Miles**: Expect to run significantly more, about **167 miles per week**.
   - **Education**: Higher education level with an average of **17.3 years**.

2.b) Visualizations

Age Distribution

The mean age for KP281 customers is **28.55** years, with a median of **26** and a mode of **23**, indicating a younger customer base. KP481 customers have a similar average age of **28.90**, with a median of **26** and mode of **25**, suggesting a slightly older group compared to KP281. KP781 customers, on the other hand, have a higher mean age of **29.10**, with a median of **27** and mode of **25**, indicating a trend towards slightly older, more fitness-conscious buyers. Overall, the age range is quite wide (32 years), showing that customers from various age groups purchase treadmills.
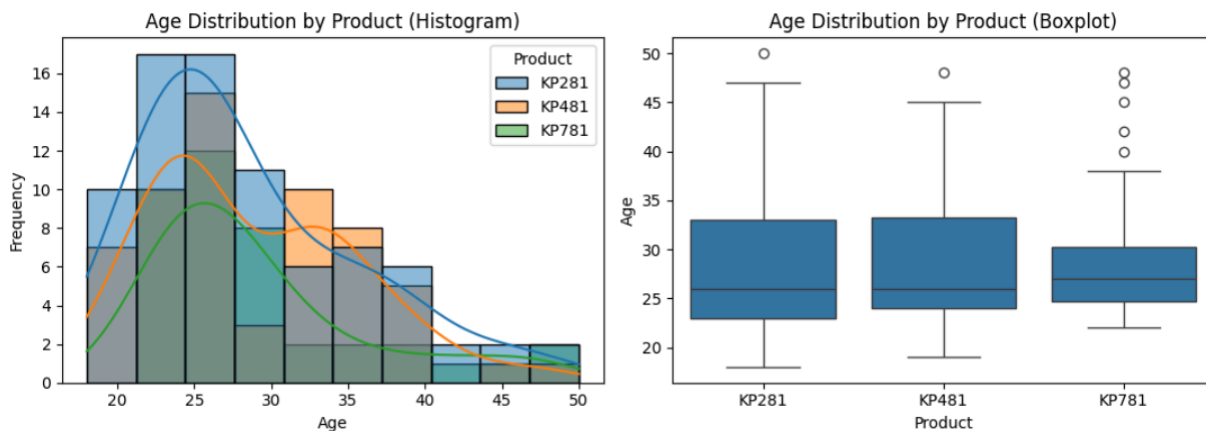

Income Distribution

Use histograms or KDE plots to see income variations across products.KP281 customers have a mean income of $46,418, while KP481 customers have a slightly higher mean of $48,973. KP781 buyers stand out with a significantly higher mean income of $75,441. The wide overall income range of $75,019 shows that customers from diverse income brackets purchase treadmills
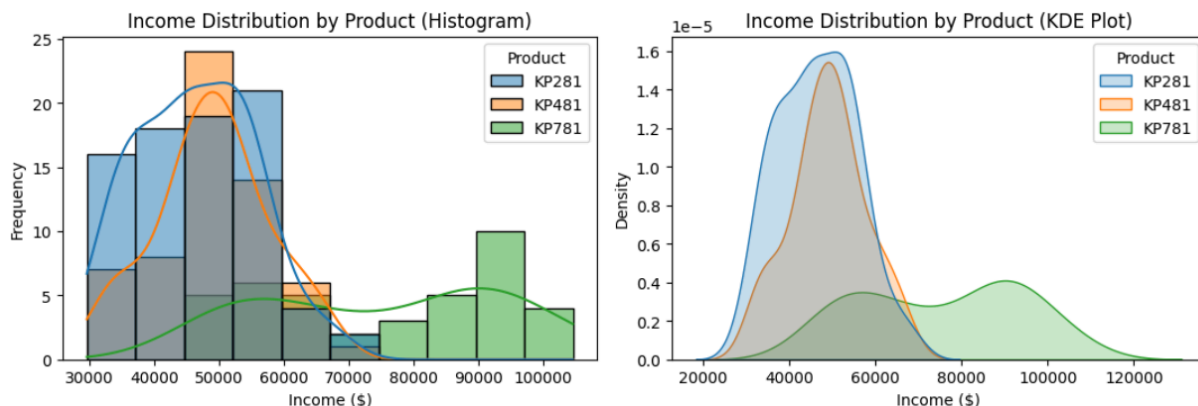
Fitness & Usage

Plot count plots or bar charts to compare fitness levels and usage patterns across the three products.For KP281 and KP481, the most common fitness level among customers is **3**, indicating moderate fitness levels. However, KP781 buyers predominantly have a fitness level of **5**, suggesting they are fitness enthusiasts. This highlights the KP781 treadmill's appeal to customers with higher fitness levels.
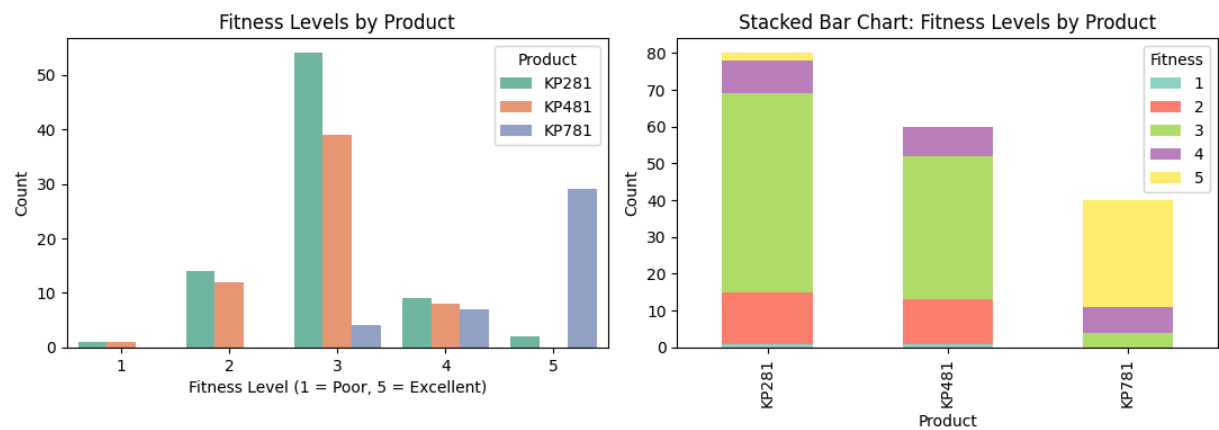

Miles Expected

 Visualize how many miles customers expect to run across different treadmills.For the KP281 treadmill, customers expect to run an average of **82.79 miles**, with a median of **85** miles, indicating a fairly consistent expectation among users. The KP481 has a slightly higher average of **87.93 miles**, with a mode of **95 miles**, suggesting a greater willingness to use the treadmill. In contrast, KP781 users expect to run an average of **166.90 miles**, with a median of **160 miles** and a mode of **100 miles**, reflecting the expectations of more serious fitness enthusiasts. Overall, this reveals significant variation in expected mileage across customers, particularly highlighting the high expectations of KP781 users.
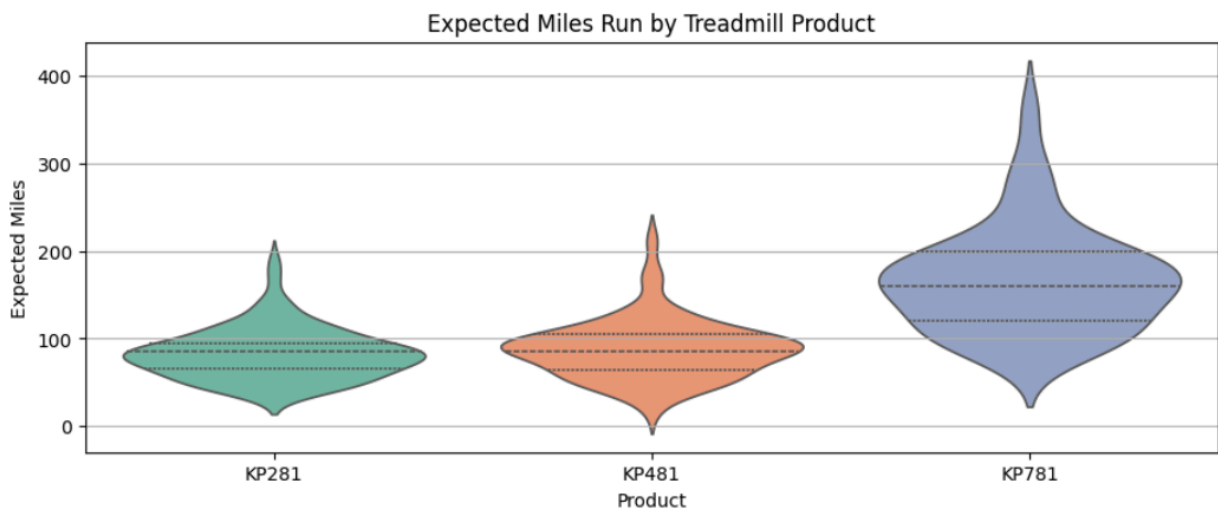
2.c) Gender and Marital Status Impact

In terms of gender distribution, the KP281 treadmill attracts an equal number of male and female customers, each totaling 40. The KP481 shows a slight male predominance, with 31 males and 29 females, while the KP781 has a significant male skew, with 33 males compared to only 7 females. This indicates that KP781 may be more popular among male customers, suggesting potential areas for targeting female customers in marketing efforts.
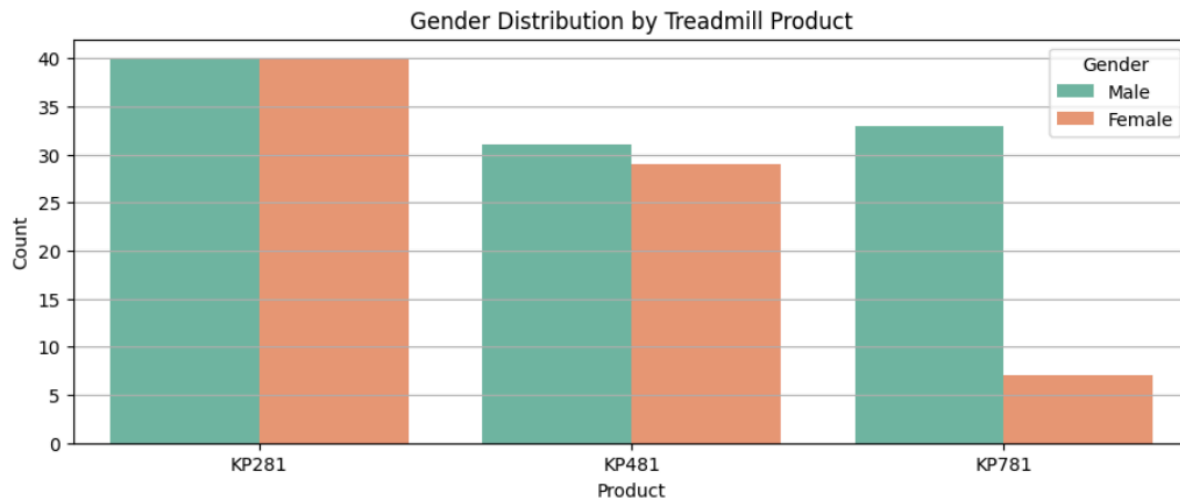
Regarding marital status distribution, KP281 has a majority of 48 partnered customers compared to 32 singles, indicating a strong interest among individuals in relationships. For KP481, there are 24 singles and 36 partners, again showing a preference for partnered individuals. In contrast, KP781 has the lowest counts for both categories, with 17 singles and 23 partnered. This pattern suggests that married or partnered customers may lean toward the KP281 and KP481 models, while KP781 may appeal to a more niche audience.
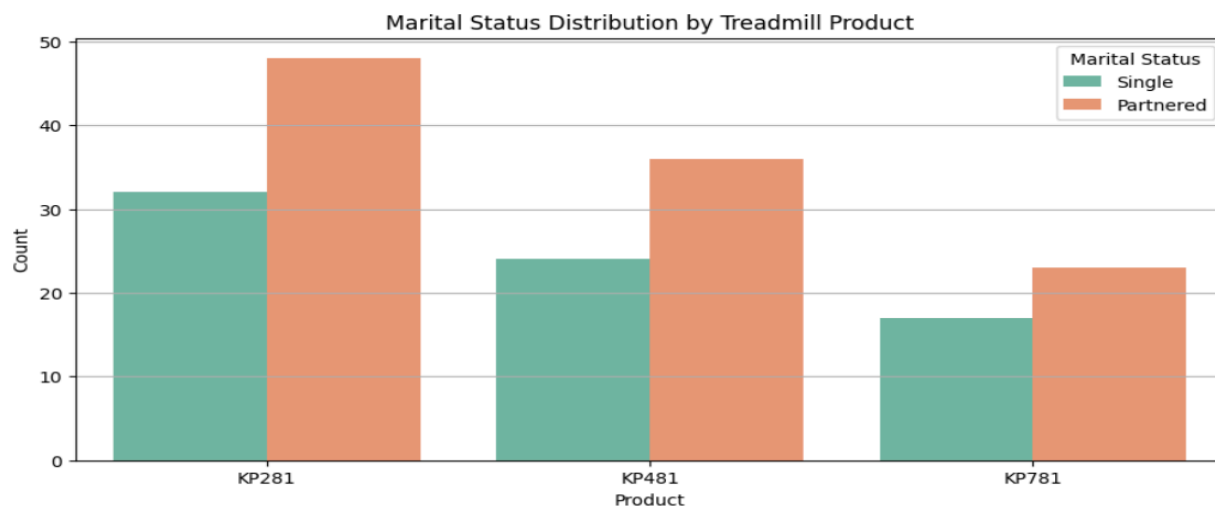
Two-Way Contingency Tables and Probabilities

3.a) Create Contingency Tables

Use pd.crosstab() to create tables showing relationships between categorical variables (e.g., Product Purchased vs. Gender, Marital Status, Fitness).

```
--- Contingency Table: Product Purchased vs. Gender ---
Gender    Female  Male  Total
Product
KP281        40     40     80
KP481        29     31     60
KP781         7     33     40
Total        76    104    180

--- Contingency Table: Product Purchased vs. Marital Status --
MaritalStatus  Partnered  Single  Total
Product
KP281               48      32      80
KP481               36      24      60
KP781               23      17      40
Total              107      73     180

--- Contingency Table: Product Purchased vs. Fitness ---
Fitness   1    2    3   4   5  Total
Product
KP281     1   14   54   9   2     80
KP481     1   12   39   8   0     60
KP781     0    0    4   7  29     40
Total     2   26   97  24  31    180
```

3.b) Marginal Probabilities

The **marginal probabilities** of purchasing each treadmill model indicate the following customer preferences: **KP281** has the highest probability of **44.44%**, suggesting it is the most popular choice among customers. **KP481** follows with a probability of **33.33%**, indicating a moderate level of interest. Lastly, **KP781** has the lowest marginal probability at **22.22%**, suggesting it appeals to a smaller segment of the market. These probabilities highlight the varying levels of demand for each treadmill model, which can inform targeted marketing strategies.

3.c) Conditional Probabilities

Probability of a Male Buying Each Treadmill Model:

- KP281 has the highest probability (38.46%), indicating a strong preference among male customers.
- KP781 follows closely at 31.73%, while KP481 is slightly lower at 29.81%.

Probability of Customers in the Highest Fitness Category Purchasing Each Model:

- A significant 93.55% of customers in the highest fitness category prefer KP781, indicating it is favored by fitness enthusiasts.
- In contrast, only 6.45% of this group purchases KP281, suggesting a lower fitness interest among those customers.

Probability of a Single Customer Buying Each Treadmill Model:

- KP281 leads with a probability of 43.84%, appealing more to single customers.
- KP481 is next at 32.88%, while KP781 has the lowest probability at 23.29%.

Probability of a Partnered Customer Buying Each Treadmill Model:

- KP281 also tops this category with 44.86%, followed by KP481 at 33.64%, and KP781 at 21.50%.

Probability of Customers Aged 30 or Below Buying Each Treadmill Model:

- KP281 is the most popular choice (45.83%) among younger customers, while KP481 and KP781 show lower probabilities at 29.17% and 25.00%, respectively.

Probability of Customers with Income Above $100,000 Buying Each Treadmill Model:

- KP781 stands out with a 100% probability, indicating that all high-income customers in the dataset prefer this model.

Probability of Customers Using the Treadmill More Than 3 Times a Week:

- Half of the customers who use the treadmill more frequently choose KP781 (50%), while KP281 and KP481 show lower usage probabilities at 30.77% and 19.23%, respectively.

Probability of Customers Expecting to Run More Than 10 Miles Per Week:

- KP281 again shows a higher preference at 44.44%, followed by KP481 (33.33%) and KP781 (22.22%).

Probability of Customers with a Fitness Rating of 4 or Higher:

- KP781 leads with 65.45%, indicating that it is preferred by customers with higher fitness ratings, while KP281 and KP481 show significantly lower probabilities at 20.00% and 14.55%**.**

Probability of Customers Aged Between 31 and 50:

- KP281 and KP481 both have an equal probability of 41.67%, while KP781 is significantly lower at 16.67%, indicating less interest in this demographic group.

3.d) Business Insights

Male Purchases: Approximately 38.5% of males purchased the KP281 treadmill, while 31.7% opted for the KP781, and 29.8% chose the KP481.
Insight: The KP281 is more popular among male customers, suggesting that entry-level models may attract a significant male demographic, which could inform targeted marketing efforts.

Highest Fitness Category: A staggering 93.5% of customers in the highest fitness category purchased the KP781, while only 6.5% opted for the KP281.
Insight: This indicates that highly fit customers are drawn to premium products, likely due to the advanced features. Aerofit could consider marketing KP781 to fitness enthusiasts, highlighting its superior capabilities.

Single Customer Purchases: Among single customers, 43.8% purchased the KP281, 32.9% purchased the KP481, and only 23.3% chose the KP781.
Insight: The trend indicates that single customers are more inclined towards entry-level options. Aerofit might consider promotional strategies for the KP281, emphasizing affordability for younger singles.

Partnered Customer Purchases: Partnered customers show a similar trend, with 44.9% purchasing the KP281, 33.6% the KP481, and 21.5% the KP781.
Insight: Since partnered customers tend to gravitate towards lower to mid-range options, Aerofit could tailor marketing campaigns to highlight family-oriented features in these models.

Age Distribution (30 or Below): For customers aged 30 or below, 45.8% purchased the KP281, 29.2% opted for the KP481, and 25.0% chose the KP781.
Insight: The entry-level KP281 remains the most appealing option for younger buyers, indicating that Aerofit can enhance engagement with this demographic through targeted marketing efforts.

High-Income Buyers: All customers with an income above $100,000 purchased the KP781.
Insight: This suggests that high-income customers prioritize premium features and

performance. Aerofit should leverage this insight to target affluent customers, showcasing the KP781's advanced specifications.

High Usage Frequency: Among customers using the treadmill more than three times a week, 50% chose the KP781, 30.8% selected the KP281, and 19.2% picked the KP481.
Insight: Regular users gravitate towards the KP781, indicating that Aerofit should market this model's durability and high-performance features to fitness enthusiasts.

Expected Running Miles: Customers expecting to run more than ten miles per week purchased the KP281 (44.4%), KP481 (33.3%), and KP781 (22.2%).
Insight: While the KP281 is popular, it might also indicate a need to promote the KP781's features for high-mileage users, possibly enhancing its perceived value for those seeking performance.

Fitness Rating of 4 or Higher: 65.5% of customers with a fitness rating of 4 or higher purchased the KP781, while 20.0% chose the KP281, and 14.5% the KP481.
Insight: The KP781's appeal to highly fit individuals suggests that Aerofit can position this model as a must-have for serious fitness enthusiasts, focusing on its advanced features.

Age Group (31 to 50): Among customers aged between 31 and 50, 41.7% purchased the KP281, and 41.7% also opted for the KP481, while only 16.7% chose the KP781.
Insight: This demographic shows a preference for entry to mid-level treadmills, indicating that Aerofit can refine its marketing strategy for these age groups to enhance awareness and interest in the KP481.

Correlation and Heatmap Analysis

Use data.corr() to create a correlation matrix for continuous variables (Age, Income, Usage, Fitness, Miles).
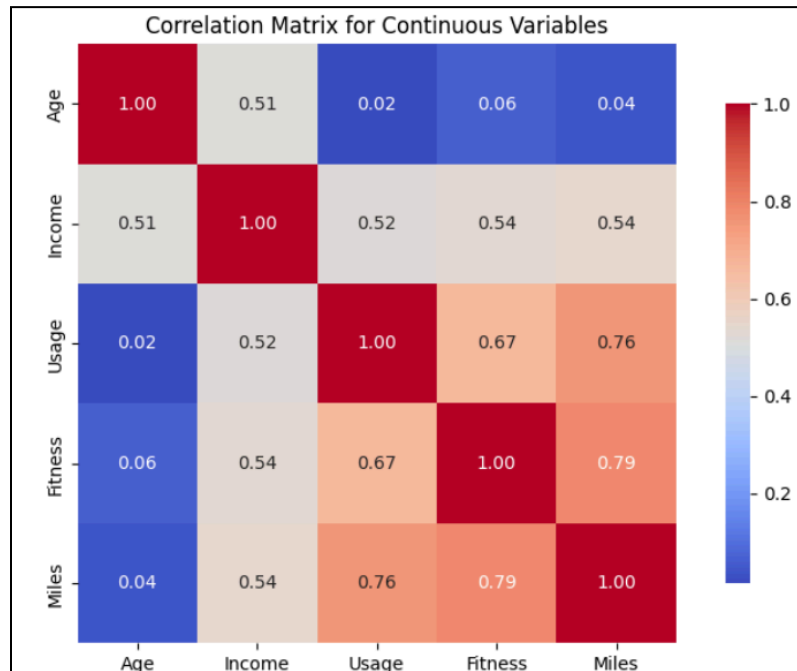
Based on the provided correlation matrix, here are some insights regarding the relationships between the continuous variables (Age, Income, Usage, Fitness, and Miles) for Aerofit treadmills:

Insights from the Correlation Matrix Age and Income (0.51):

There is a moderate positive correlation between Age and Income. This suggests that as customers age, their income tends to increase. Business Insight: Marketing strategies could focus on targeting middle-aged customers, who may have higher incomes and more purchasing power for premium treadmill models.

Income and Usage (0.52):

The moderate positive correlation between Income and Usage indicates that higher-income customers are likely to use the treadmill more frequently. Business Insight: Aerofit may consider offering promotions or loyalty programs aimed at high-income customers to encourage even more usage, potentially increasing their brand loyalty.

Usage and Fitness (0.67):

There is a strong positive correlation between Usage and Fitness. This implies that customers who use the treadmill more frequently also tend to rate their fitness levels higher. Business Insight: Aerofit can leverage this relationship in marketing campaigns by promoting the health benefits of regular treadmill use, positioning their products as tools for enhancing fitness.

Usage and Miles (0.76):

The strong positive correlation between Usage and Miles indicates that customers who use their treadmills more often also expect to run or walk more miles. Business Insight: This insight could inform product development, as Aerofit may want to design and market models that cater to users who are likely to engage in higher mileage running or walking.

Fitness and Miles (0.79):

There is a strong positive correlation between Fitness and Miles, suggesting that customers who rate themselves higher in fitness also expect to cover more miles on their treadmills. Business Insight: Aerofit could target fitness enthusiasts with marketing messages that emphasize the performance capabilities of their treadmills, particularly for those looking to log significant mileage.

Income and Fitness (0.54):

The positive correlation between Income and Fitness indicates that higher-income individuals tend to rate their fitness levels higher. Business Insight: Aerofit may consider positioning higher-end products towards affluent customers, promoting features that enhance fitness, such as advanced tracking technologies or customizable workout plans.

Low Correlation with Age:

Age shows low correlations with Usage, Fitness, and Miles, indicating that age might not be a significant factor influencing how customers use the treadmill or their self-rated fitness levels. Business Insight: Marketing efforts should focus less on age demographics and more on lifestyle and income characteristics to better segment and target customers. Conclusion The correlation matrix reveals valuable insights into the relationships among customer characteristics and their interactions with Aerofit treadmills. These insights can guide marketing strategies, product development, and customer engagement initiatives. By understanding these correlations, Aerofit can better meet the needs of its diverse customer base and enhance customer satisfaction.

Customer Segmentation and Profiling

5.a) Customer Segments Segment Descriptions Based on Descriptive Analysis:

KP281 Customers:

Demographics: Younger customers, typically in their late 20s to early 30s.

Income: Lower-income bracket, likely earning less than $70,000 annually.

Fitness Level: Less fitness-conscious; may be first-time treadmill buyers.

Usage: Moderate usage, averaging 2-3 times per week.

Insights: This segment may be attracted to budget-friendly options and promotional deals.

KP481 Customers:

Demographics: Mid-income customers, generally aged between 30 to 45.

Income: Income range typically between $70,000$ and $100,000$ annually.

Fitness Level: Moderate fitness level; balanced usage pattern.

Usage: Regular users, averaging 3-4 times per week.

Insights: This group is likely looking for value in terms of quality and features, making them responsive to marketing focused on versatility.

KP781 Customers:

Demographics: Primarily high-income individuals aged 35 to 55.

Income: Income above $100,000 annually, often professionals or fitness enthusiasts.
Fitness Level: Highly fitness-conscious; engaged in running and other cardio exercises.
Usage: High usage, typically 5 or more times per week, expecting to run over 10 miles weekly.

Insights: This segment appreciates advanced features, performance tracking, and premium quality, making them less sensitive to price.

5.b) Actionable Profiles :

- **KP281 buyers** are typically younger individuals, predominantly in their late 20s to early 30s, with lower income levels. They may be first-time treadmill owners looking for affordable options. Their usage is moderate, averaging 2-3 times a week, indicating a casual approach to fitness. Marketing Strategy: Promotions should focus on affordability, ease of use, and introductory offers. Consider targeting first-time buyers through social media ads highlighting budget-friendly benefits.

- **KP481 customers** are mid-income individuals aged 30 to 45 who value balance in their fitness routines. They are moderately fitness-conscious and use their treadmills regularly, averaging 3-4 times per week. This segment appreciates the balance between price and quality. Marketing Strategy: Emphasize the treadmill's features, versatility, and value-for-money in marketing campaigns. Highlight customer testimonials to showcase satisfaction and benefits of regular use.

- **KP781 buyers** are primarily high-income males aged 35 to 55 who are serious fitness enthusiasts. They use their treadmills frequently (5+ times per week) and expect to run more than 10 miles weekly. This segment prioritizes performance and advanced features. Marketing Strategy: Focus on premium features, performance analytics, and personalized training programs. High-end marketing channels like fitness magazines and specialized online platforms should be utilized to reach this audience effectively.

- **Conclusion :** By creating distinct customer segments and actionable profiles, Aerofit can tailor its marketing strategies and product offerings to better meet the needs and preferences of each group. This targeted approach is likely to enhance customer satisfaction and drive sales growth across the treadmill product line.

Insights and Recommendations

Male Purchases:

- Insight: The KP281 is the most popular among male customers, suggesting a strong market for entry-level models.
- Recommendation: Aerofit should develop targeted marketing campaigns aimed at men, emphasizing the affordability and value of the KP281.

Fitness Enthusiasts:

- Insight: The KP781 attracts a significant majority (93.5%) of highly fit customers.
- Recommendation: Focus marketing efforts on fitness enthusiasts by showcasing the advanced features of the KP781, positioning it as the ideal choice for serious athletes.

Single Customer Trends:

- Insight: Single customers lean towards the KP281 as their preferred choice.
- Recommendation: Implement promotional strategies for the KP281 that highlight its value and affordability, particularly targeting younger, single demographics.

Partnered Customer Preferences:

- Insight: Partnered customers show a preference for lower to mid-range options like the KP281 and KP481.
- Recommendation: Marketing should emphasize family-oriented features and benefits of the KP281 and KP481 to attract partnered customers.

Youth Engagement:

- Insight: The KP281 is favored by customers aged 30 or below.
- Recommendation: Enhance engagement with younger buyers through social media campaigns and influencer partnerships that highlight the KP281.

High-Income Targeting:

- Insight: High-income customers exclusively purchased the KP781.
- Recommendation: Target affluent consumers with marketing strategies that emphasize the premium quality and advanced specifications of the KP781.

Promoting Usage Features:

- Insight: Regular users prefer the KP781.

- Recommendation: Highlight the durability and performance of the KP781 in marketing materials to attract serious fitness enthusiasts.

High-Mileage Marketing:

- Insight: Customers expecting to run more than ten miles per week often choose the KP281.
- Recommendation: Promote the KP781's high-performance features to high-mileage users to enhance its perceived value.

Focus on Fitness Ratings:

- Insight: The KP781 appeals strongly to customers with a fitness rating of 4 or higher.
- Recommendation: Position the KP781 as a must-have for fitness enthusiasts, emphasizing its superior capabilities in marketing efforts.

Age Group Strategy:

- Insight: Customers aged 31 to 50 prefer the KP281 and KP481.
- Recommendation: Tailor marketing strategies for these age groups to raise awareness and interest in the KP481 by emphasizing its features suitable for this demographic.