



AEROFIT DATA ANALYSIS

BUSINESS REPORT

ZOHAIB SALEEM

DS 8 RED

EXECUTIVE SUMMARY

The market research team at AeroFit wants to identify the characteristics of the target audience for each type of treadmill offered by the company, to provide a better recommendation of the treadmills to new customers. Consumer data provided by AeroFit consists of their three main types of products priced strategically for different target markets. To effectively analyse the pricing strategy data has been provided with information such as customer age, gender, education, usage, fitness level and income etc. This report analyzes consumer data provided by AeroFit to identify the target audience for each of its three treadmill types. By employing exploratory data analysis techniques, including descriptive statistics, data visualization, and correlation analysis, key insights were uncovered. Outlier detection and conditional probability analysis were utilized to further refine the segmentation. The findings reveal distinct target audience segments for each treadmill type, enabling AeroFit to optimize product recommendations and marketing strategies





DATA HIGHLIGHTS:

The company collected data on individuals who purchased a treadmill from the AeroFit stores during the prior three months. The dataset in `aerofit_treadmill_data.csv` has the following features:

- Product - product purchased: KP281, KP481, or KP781
- Age - in years
- Gender - male/female
- Education - in years
- Marital Status - single or partnered
- Usage - the average number of times the customer plans to use the treadmill each week
- Fitness - self-rated fitness on a 1-5 scale, where 1 is the poor shape and 5 is the excellent shape
- Income - annual income in US dollars
- Miles - the average number of miles the customer expects to walk/run each week

Data was in csv format without any empty rows or columns.

Data has 180 rows and 9 columns.

Data classification is as follows:

#	Column	Non-Null Count	Data Type
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	Marital Status	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



DATA ANALYSIS TOOL: Python using Google colab notebook.

COLAB LINK: https://colab.research.google.com/drive/1VkDaT0-ePBuR7qmIRETyKry_vT54WC7o?usp=sharing

INITIAL SETUP

Installing Dependencies:

- The notebook installs jupyter_dash, for interactive dashboards.

Importing Libraries:

- Common libraries like pandas, numpy, matplotlib.pyplot for data manipulation and visualization.
- dash, plotly.express for creating interactive visualizations.

DATA LOADING

- The dataset aerofit_treadmill_data.csv is loaded into a Pandas DataFrame.

PRELIMINARY EXPLORATION

Preview of the Dataset:

- The first 10 rows of the dataset are displayed using df.head(10).

Data Dimensions:

- The dataset's size is checked, revealing 180 rows and 9 columns.

2. NON-GRAPHICAL ANALYSIS:

Non graphical data analysis produced following results:

- **Age:** The average age of customers is approximately **28.8** years, with a standard deviation of 6.9 years. The youngest customer is 18, the oldest is 50.
- **Education:** Customers have an average of **15.6** years of education, with a standard deviation of 1.6 years. The minimum education level is 12 years, and the maximum is 21.
- **Usage:** On average, customers plan to use the treadmill about **3.5** times per week, with a standard deviation of 1.1 times. The least frequent usage is 2 times/week, and the most is 7 times/week.
- **Fitness:** The average self-rated fitness level is **3.3**, with a standard deviation of 0.96. Fitness is rated on a scale of 1 to 5, with 1 being poor and 5 being excellent.
- **Income:** The average annual income of customers is around **\$53,720**, with a standard deviation of \$53,720, with a standard deviation of \$16,507. The minimum income is \$29,562, and the maximum is \$29,562, and the maximum is \$104,581.



- **Miles:** Customers expect to walk/run an average of **103.2 miles per week**, with a standard deviation of 51.9 miles. The minimum expected mileage is 21 miles/week, and the maximum is 360 miles/week.

Value_Counts:

- **Products:** The dataset contains sales data for three treadmill products: KP281, KP481, and KP781. The KP281 is the most popular product, with **80 units** sold. The KP481 was sold 60 times, and the KP781 is the least popular, with 40 units sold.
- **Gender:** Out of the total customers in the dataset, **104 are male and 76 are female**. This indicates that a larger proportion of customers purchasing treadmills are male
- **Age:** The age of customers purchasing treadmills ranges from **18 to 50**. The most frequent age group is **25**, followed by 23 and 24. A significant portion of customers falls within the age range of **20 to 38**.
- **Marital Status:** Out of the total customers, **107 are partnered, and 73 are single**. This indicates that a larger proportion of customers who purchase treadmills are partnered.

Unique()

Age:

The 'Age' variable in the dataset contains **32 unique** values, ranging from **18 to 50**. This indicates the diversity of age groups among customers who purchase treadmills.

- **Unique Values:** It states that there are **32** different ages recorded in the dataset.
- **Range:** It mentions that the age of customers spans from **18 to 50 years** old.
- **Diversity:** It highlights the presence of various age groups among customers.

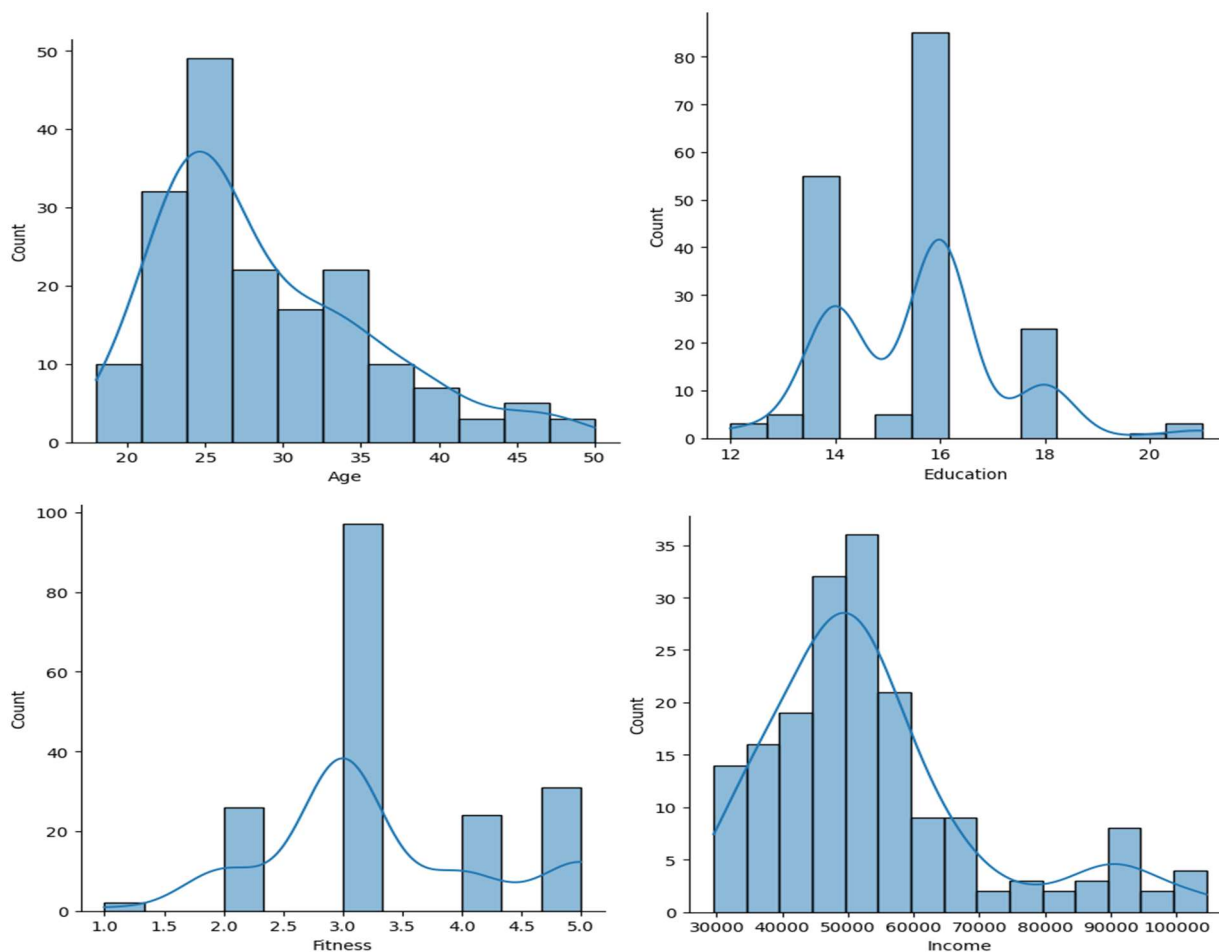
Income:

- **Unique Incomes:** There are **59** distinct income levels recorded for customers in the dataset.
- **Income Range:** The income of customers spans from **\$29,562 to \$104,581**.
- **Income Diversity:** The data suggests a considerable variation in income among treadmill buyers.



3. GRAPHICAL ANALYSIS:

A. DISTPLOTS:



DATA INSIGHTS:

Miles

- **Distribution:** The histogram shows a right-skewed distribution with a peak around 75-100 miles.
- **Insights:** A significant number of observations contribute to this range, indicating that most data points lie within relatively shorter distances. Few outliers exist beyond 200 miles, suggesting less frequent long-distance occurrences.

2. Fitness

- **Distribution:** This graph displays a bimodal distribution with peaks at 2.5 and 3.5.
- **Insights:** A clear clustering around moderate fitness levels indicates varying fitness among the subjects. The population is less represented at lower fitness levels (1-2) and higher levels (4-5), suggesting that most individuals score in the middle range.



3. Income

- **Distribution:** The income data shows a left-skewed distribution with a prominent peak at around \$50,000.
- **Insights:** The majority of individuals fall in the lower to middle-income brackets. As income increases beyond \$70,000, counts decrease significantly, indicating fewer high-income individuals.

4. Education

- **Distribution:** This histogram is bimodal, with spikes at around 12 years (high school) and 16 years (bachelor's degree).
- **Insights:** The data suggests that most individuals either have a high school education or a bachelor's degree, with fewer achieving educational levels beyond this range, reflecting common educational attainment patterns.

5. Age

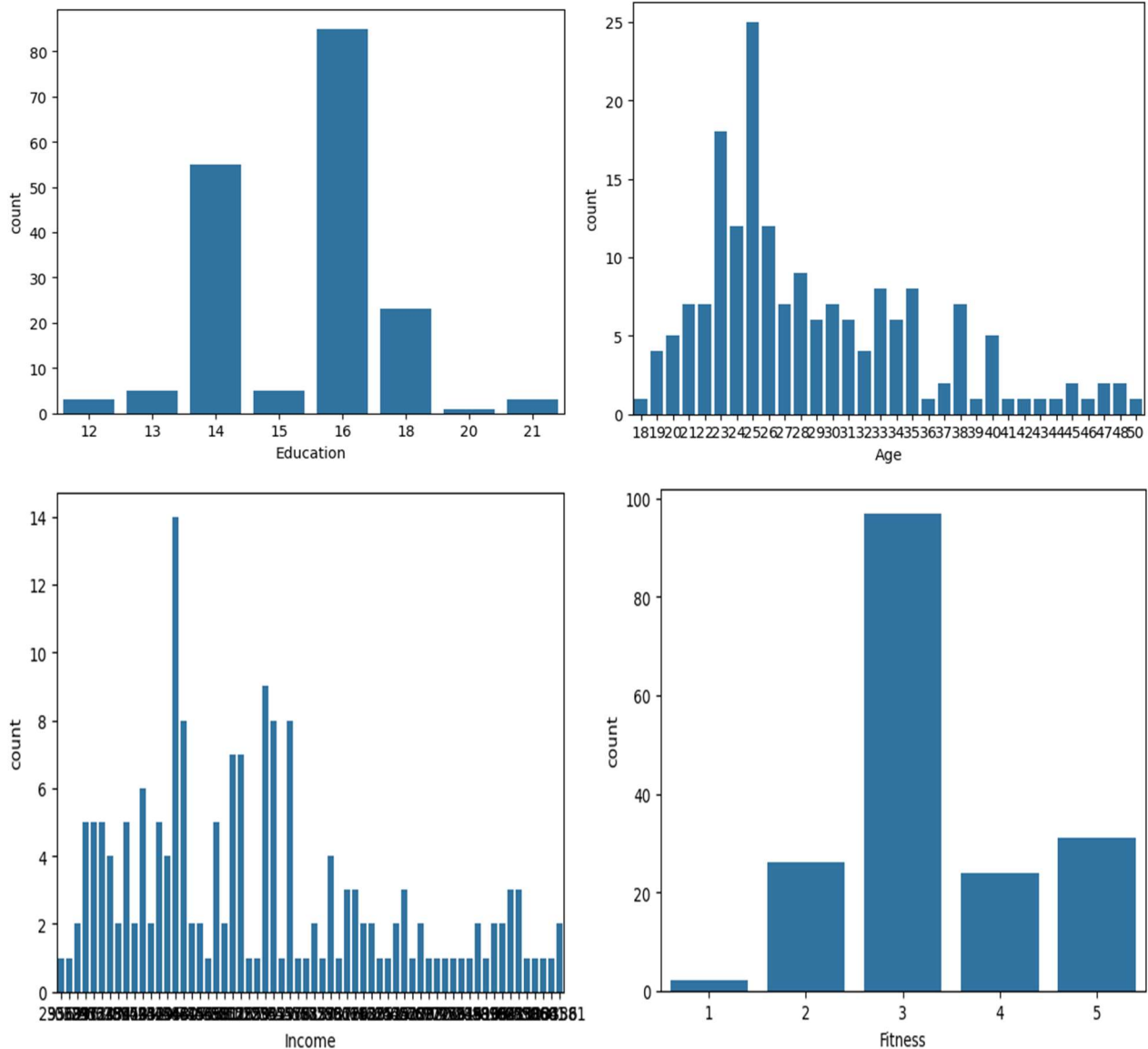
- **Distribution:** The age distribution is slightly skewed to the left, with a peak around 25 years.
- **Insights:** This indicates a younger demographic, with fewer individuals represented in the older age brackets. The data suggests that the population is predominantly youthful, possibly indicating trends related to employment or schooling.

OVERALL INSIGHTS:

- There is a common pattern of clustering in middle ranges across the variables (distance, fitness, income, education).
- Most individuals show moderate levels in fitness, education, and income, while age skews younger, possibly reflecting a younger sample studied.



B. COUNT PLOTS:



1. Fitness Distribution

- **Overview:** The histogram shows fitness levels ranging from 1 to 5.
- **Insights:**
 - The most common fitness rating is **3**, with a significant number of responses (around 100).
 - Ratings **1** and **2** have notably fewer counts, indicating that very low fitness is less common.
 - Higher fitness ratings (4 and 5) show a steady decline in counts, suggesting that fewer individuals report excellent fitness levels.



2. Income Distribution

- **Overview:** This histogram illustrates income levels with a wide range of values.
- **Insights:**
 - There is a **right-skewed distribution**, indicating more individuals earn lower incomes.
 - A distinct peak suggests a popular income range that draws more responses.
 - Outliers in higher income ranges are present but represent a much smaller portion of the dataset.

3. Education Distribution

- **Overview:** This graph depicts education levels, likely measured in years completed.
- **Insights:**
 - The majority of individuals have **16 years of education** (often indicating a bachelor's degree), with the highest count of around 90.
 - There are also significant counts for 12-15 years of education, but counts decrease sharply after reaching around 16 years.
 - Higher education levels (above 16) are less common, reflecting typical educational attainment trends.

4. Age Distribution

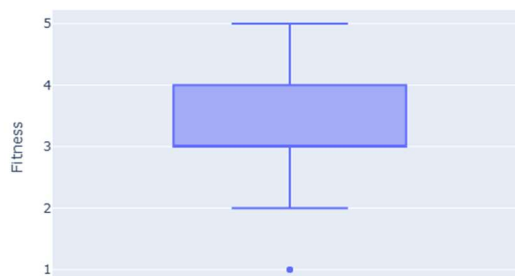
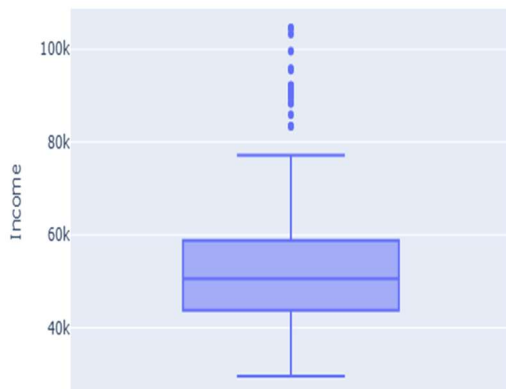
- **Overview:** The histogram shows age data.
- **Insights:**
 - The age distribution appears to have a **bimodal tendency**, with two peaks around certain age ranges, especially between **20-30** and **40-50**.
 - A notable grouping indicates a younger population, particularly with significant counts in the 22-32 age range, which may suggest a sample largely consisting of younger adults or students.
 - Older age groups see a decline in representation, reflecting a potential gap in this age range.

OVERALL INSIGHTS

- **Demographics & Trends:** The data indicates a predominantly young audience with moderate fitness levels, commonly achieving up to bachelor's education, and lower income statuses. These patterns may suggest further examination into lifestyle, economic factors, or educational systems impacting these demographics.
- **Potential Relationships:** There may be interesting correlations among these variables, such as how educational attainment affects income and fitness levels among different age groups.



C. BOX PLOTS:



1. Miles

- **Key Insights:**

- **Median:** The median mileage is around **100 miles**.
- **Interquartile Range (IQR):** The IQR spans from approximately **75 to 125 miles**, indicating that the central 50% of the data is concentrated in this range.
- **Outliers:** There are several outliers exceeding **200 miles**, suggesting a few instances of significantly longer distances compared to the majority.

2. Fitness

- **Key Insights:**



- **Median:** The median fitness score is **3.5**, indicating a generally moderate fitness level among the population.
- **IQR:** The scores range between **3.0 and 4.0**, with most individuals scoring within this range.
- **Outliers:** There are a few lower outliers (scores of **1.0 and 2.0**), suggesting some individuals may have poor fitness levels, but the bulk have scores on the higher end.

3. Income

- **Key Insights:**

- **Median:** Median income is approximately **\$50,000**.
- **IQR:** The IQR spans from around **\$40,000 to \$60,000**, indicating that most incomes are clustered within this range.
- **Outliers:** A number of high-income outliers (over **\$90,000**) indicate a small group earning significantly more than the majority.

4. Education

- **Key Insights:**

- **Median:** The median years of education is about **16 years**, typically corresponding to a bachelor's degree.
- **IQR:** The IQR ranges from **14 to 18 years**, meaning that most individuals have between two years of college and a bachelor's degree.
- **Outliers:** Few outliers are present at higher levels of education, indicating a handful of individuals with significantly more education (beyond 20 years).

5. Age

- **Key Insights:**

- **Median:** The median age is around **30 years**.
- **IQR:** The IQR extends from approximately **25 to 35 years**, indicating that the population is relatively young, with most individuals falling within this age range.
- **Outliers:** A few outliers exist in the higher age range (over **40 years**), but they represent a small segment of the overall data.

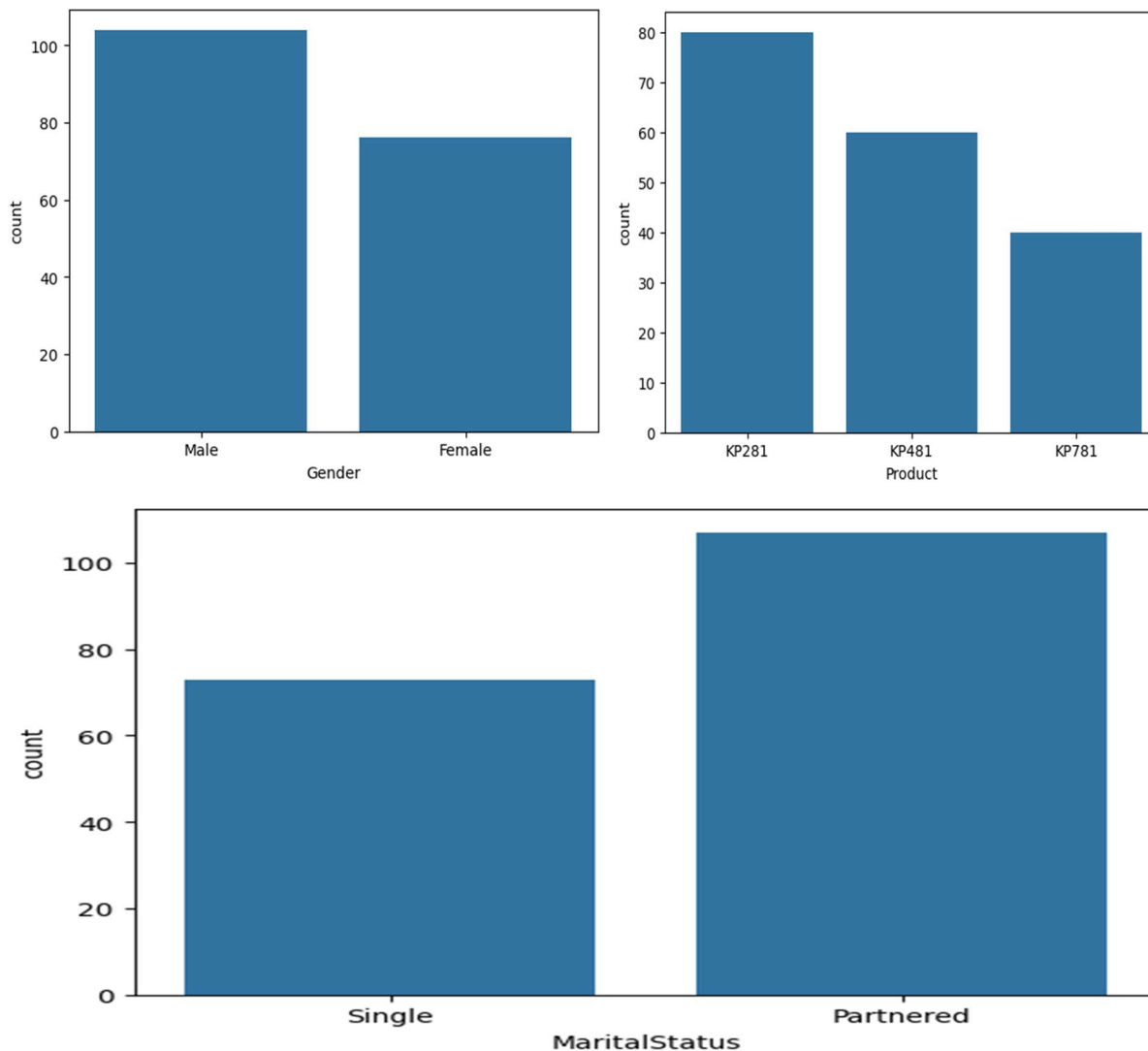
OVERALL INSIGHTS

- **Demographic Profile:** The data indicates a population that is generally young, with moderate fitness levels, typical educational attainment of around a bachelor's degree, and a median income around \$50,000.
- **Variability:** While the majority lies within established ranges for these factors, the presence of outliers suggests diversity in income and education levels, hinting at socioeconomic variability within this demographic.



- **Potential Areas for Further Study:** Investigation into factors affecting fitness levels, income disparities, and the impact of education on both income and fitness could provide valuable insights.

4. CATEGORICAL FEATURES:



1. Marital Status Distribution

- **Overview:** This bar graph displays counts for two categories: Single and Partnered.
- **Insights:**
 - The **Partnered** category has a higher count (over **100** responses) compared to the **Single** category (approximately **70-80**).
 - This suggests a predominantly partnered demographic among respondents.



- **Implications:** Marketing strategies or programs tailored towards partners may be more beneficial based on this demographic split.

2. Gender Distribution

- **Overview:** This graph shows the counts of respondents based on gender, with categories for Male and Female.
- **Insights:**
 - The **Male** category has a significantly higher count (over **100**), while the **Female** category is lower (around **80**).
 - This indicates a male-dominated sample population, which might inform targeted outreach or messaging strategies in future campaigns.
 - **Considerations:** Programs or products may need to consider strategies to engage a more balanced gender representation.

3. Product Preferences

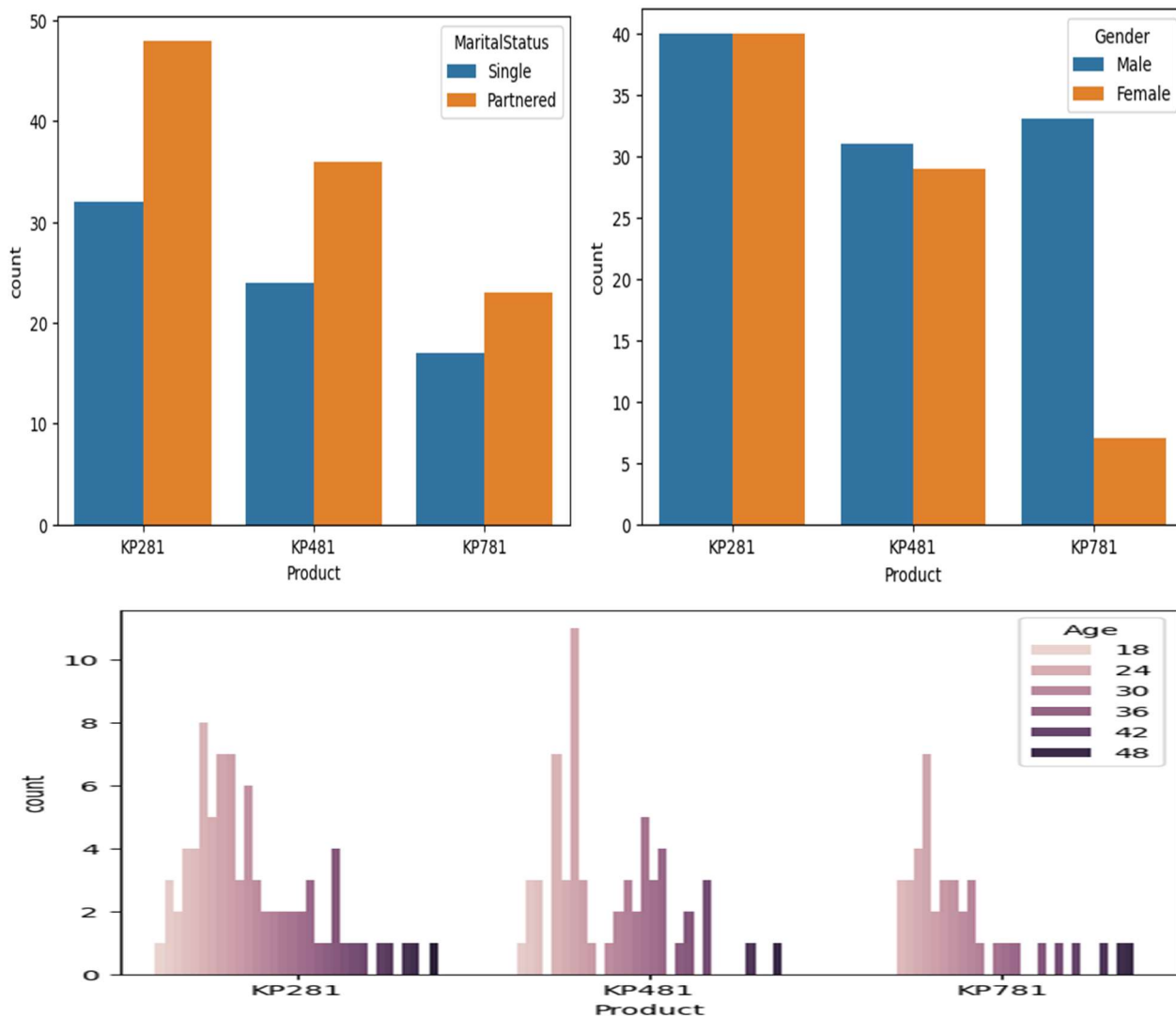
- **Overview:** This bar graph depicts the number of respondents associated with three products: KP281, KP481, and KP781.
- **Insights:**
 - **KP281** is the most popular product, with the highest count (approximately **80-90**).
 - **KP481** follows with a moderate count (around **50-60**), while **KP781** shows the least interest (approximately **30-40**).
 - **Implications:** The preference for KP281 suggests it may have stronger features or marketability, making it a focal point for promotions and inventory management.
 - This insight could lead to further investigation into what drives the popularity of KP281 compared to the others.

OVERALL INSIGHTS:

- The demographic data indicates a skew towards partnered males who show a preference for the product KP281.
- These patterns point to potential areas for focused marketing efforts, as engaging partnered individuals (especially males) may yield more significant returns if targeting the KP281 product.



5. BIVARIATE ANALYSIS:



1. Age Distribution by Product

- **Graph Description:** This histogram shows the age distribution of respondents who prefer different products (KP281, KP481, KP781) using a color gradient to represent ages.
- **Key Insights:**



- **Product KP281** has the highest counts across various age groups, particularly notable at ages **18** and **24**, indicating it is favored by younger consumers.
- **KP481** exhibits moderate popularity but shows a less pronounced peak compared to KP281, indicating diverse preferences among age groups.
- **KP781** has the lowest counts, suggesting it may be less appealing or targeted to a specific demographic.

2. Marital Status Distribution by Product

- **Graph Description:** A bar graph showing the counts of single and partnered individuals preferring each product.
- **Key Insights:**
 - **KP281** stands out with the highest number of partnered respondents, reflecting strong appeal to individuals in relationships.
 - **KP481** follows, but the difference is less pronounced than in KP281, indicating it also has a good mix of single and partnered individuals.
 - **KP781** shows the least interest, particularly among single individuals, which may suggest it doesn't cater effectively to this demographic.

3. Gender Distribution by Product

- **Graph Description:** A bar graph depicting counts of male and female respondents for each product.
- **Key Insights:**
 - Both **KP281** and **KP481** are favored more by males, while **KP781** shows a significant skew towards females, although overall counts are lower.
 - **KP281** has a balanced but slightly higher engagement from males, indicating it might be marketed to appeal to a broader audience.
 - **KP481** has a similar pattern but with slightly less interest overall compared to KP281.

OVERALL INSIGHTS

- KP281 is favored by younger individuals and partnered consumers. It may be beneficial to design marketing strategies that target younger demographics, focusing on relationship-oriented messaging.
- KP481 appears to appeal to a diverse audience but with lower engagement than KP281. Enhancing its visibility could help it capture more market share.



PAIR PLOT:





Pair plot visualizes relationships among several variables: **Age**, **Education**, **Usage**, **Fitness**, **Income**, and **Miles**, categorized by three different products (KP281, KP481, KP781). Here's a detailed description and insights based on the visualizations:

Overview of the Pair Plot

- **Axes and Diagonals:** Each cell represents a scatter plot or density estimation between two variables, while the diagonal exhibits the distribution (density) of each individual variable.
- **Color-Coding:** The data points are color-coded to represent three products:
 - **Blue:** KP281
 - **Orange:** KP481
 - **Green:** KP781

KEY OBSERVATIONS

1. Age Distribution

- **Age & Income:** There is a positive correlation between age and income, with older respondents generally reporting higher incomes. Blue dots (KP281) are more densely populated in higher income brackets.
- **Age & Fitness:** The fitness scores seem to vary across ages, with younger individuals (especially in KP281) appearing more frequently at higher fitness levels.

2. Education Trends

- **Education & Income:** Higher education levels correlate positively with income. KP281 shows a wider distribution in education, indicating greater educational diversity among users.
- **Education & Usage:** Education appears to affect usage patterns, particularly for KP281, with users having higher education utilizing the product more frequently.

3. Usage Insights

- **Usage & Fitness:** Users of KP281 tend to report higher fitness levels, suggesting a potential link between product usage and personal health.
- **Usage Distribution:** Product KP781 shows lower usage across all education levels, indicating either less engagement or a less compelling product for this demographic.

4. Fitness Levels

- **Fitness & Income:** There's a visible correlation where respondents with higher incomes (particularly KP281 users) also report higher fitness levels.
- **Fitness Distribution:** The distribution suggests a concentration of higher fitness levels for KP281, while KP481 and KP781 show lower overall fitness scores.

5. Miles

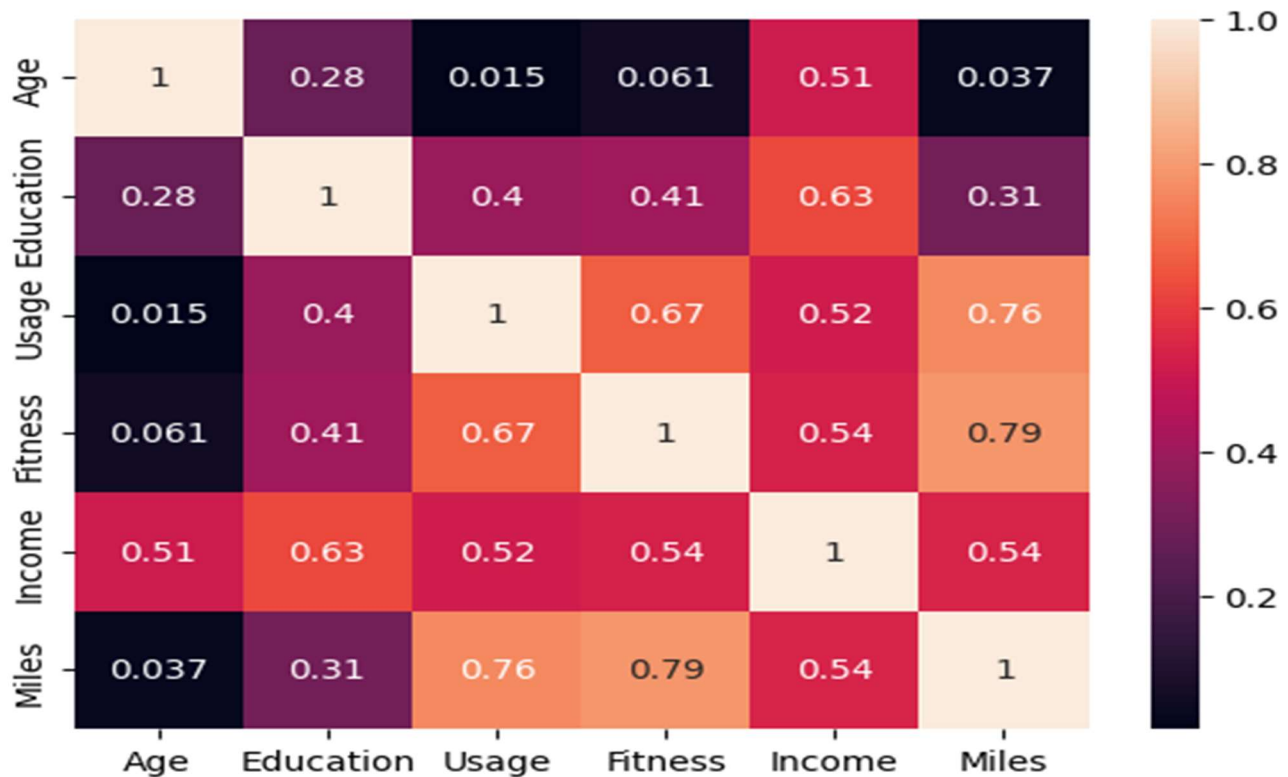
- **Miles & Income:** A moderate correlation is visible; as income increases, so do the miles reported. KP281 users tend to report higher mileage on average.
- **Miles Distribution:** Distribution indicates that most respondents report lower mileage, although there are some higher values associated particularly with KP281.



OVERALL INSIGHTS

- **Product Performance:** KP281 stands out with a strong correlation to several positive indicators (overall fitness, education, and income), suggesting it may be a premium product with a well-targeted audience.
- **Demographic Trends:** There is a trend showing older, more educated individuals tend to prefer KP281, suggesting effective marketing could be focused on health and fitness benefits for this demographic.
- **Product KP781** may require further investigation to understand its lower engagement levels and how to improve its appeal.



HEAT MAP:

Correlation heatmap, which visually represents the relationships among multiple variables: **Age**, **Education**, **Usage**, **Fitness**, **Income**, and **Miles**. Each cell in the matrix displays the correlation coefficient between the variables it intersects, ranging from -1 to 1, where:

- **1** indicates a perfect positive correlation,
- **0** indicates no correlation
- **-1** indicates a perfect negative correlation.

Key Observations from the Heatmap**1. Strongest Correlations:**

- **Fitness and Usage (0.67)**: Indicates a strong positive correlation between fitness levels and product usage. This suggests that those who use the product more frequently tend to have higher fitness scores.
- **Income and Fitness (0.79)**: The highest correlation in the matrix. It implies that individuals with higher income tend to report better fitness levels, possibly due to access to resources such as gyms or health programs.

2. Moderate Correlations:

- **Education and Income (0.63)**: A moderate positive correlation, indicating that as education levels increase, income tends to rise as well.



- **Miles and Usage (0.76):** Suggests a strong relationship where individuals who report using the product more tend to cover more miles, indicating active engagement with the product or related activities.

3. Weak Correlations:

- **Age and Usage (0.015):** An almost negligible correlation, suggesting that age does not significantly impact how users engage with the product.
- **Miles and Age (0.037):** Also very weak, indicating little to no relationship between a respondent's age and the number of miles they report.

4. Other Noteworthy Relationships:

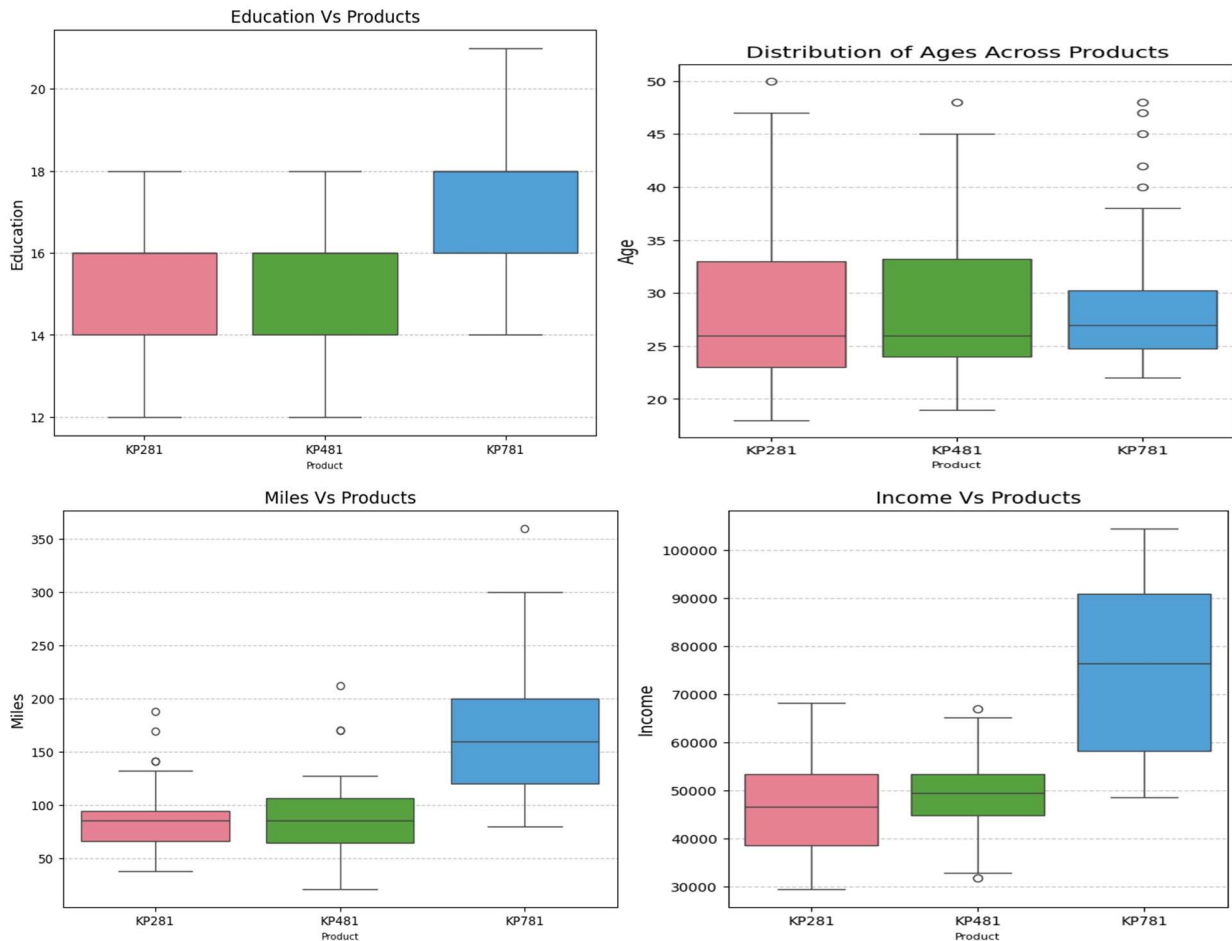
- **Education and Fitness (0.41):** Shows a moderate positive correlation, suggesting that those with higher education levels also report better fitness.
- **Income and Miles (0.54):** Indicates a moderate correlation, suggesting that higher-income individuals tend to report covering more miles.

OVERALL INSIGHTS

- **Health and Economic Factors:** The strong correlation between fitness and income underscores the potential impact of financial resources on health outcomes and fitness levels.
- **Engagement Patterns:** The relationship between usage and miles suggests that active users of the product are more likely to engage in physical activities, which may be a focal point for marketing efforts.
- **Demographic Targeting:** The weak correlation of age with other variables may indicate that marketing strategies should be more focused on lifestyle and health factors rather than age alone.



6. IQR RANGES:



Box plots depict various factors (Fitness, Education, Miles, Income, and Age) in relation to three products: **KP281**, **KP481**, and **KP781**. Each plot summarizes the distribution and statistical properties of these factors for the respective products.

1. Fitness vs. Products

- **KP281:** The fitness scores are primarily around **3.0 to 5.0**, with a median around **4.0**, indicating a high level of fitness among users of this product.
- **KP481:** Scores are very low, with most responses clustered at or below **3.0**, showing that users do not report high fitness levels.
- **KP781:** The median is again higher, close to **4.0**, but with fewer data points. The wide spread suggests variability in fitness scores.
- **Insights:** Users of KP281 and KP781 tend to have better fitness levels than those using KP481.

2. Education vs. Products

- **KP281:** Displays a median education level around **16 years**, reflecting a strong educational background among users.



- **KP481:** The median is also around **16 years**, indicating similar educational attainment to KP281, but with a slightly narrower interquartile range.
- **KP781:** Has a higher median (around **18 years**) with a broader range, suggesting a mix of education levels, including some who have completed advanced degrees.
- **Insights:** KP781 users are generally more educated than those of the other products, indicating potential appeal to a more academically inclined market.

3. Miles vs. Products

- **KP281:** The box shows a median around **100 miles** and IQR (interquartile range) suggesting a moderate mileage among its users.
- **KP481:** Displays a lower median mileage (approximately **75 miles**) and narrower range, indicating less active or engaged users.
- **KP781:** The highest median (around **150 miles**) and broader range indicate higher levels of activity among these users, with some very active individuals contributing to the outliers.
- **Insights:** KP781's users appear the most engaged in physical activity, potentially influencing their product preference.

4. Income vs. Products

- **KP281:** Users have a median income of about **\$50,000** with a range that suggests some lower earners as well.
- **KP481:** The median is around **\$40,000**, indicating lower average earnings among users.
- **KP781:** Users report the highest median income (approximately **\$80,000**), reflecting a wealthier customer base.
- **Insights:** KP781 attracts higher-income individuals, suggesting it could be marketed as a premium product.

5. Age Distribution Across Products

- **KP281:** The median age is around **28 years**, with a relatively narrow range.
- **KP481:** Users are slightly older on average (median around **30 years**), with a more varied age distribution.
- **KP781:** Shows a wider age range, with some older outliers (up to **50 years**) suggesting broader appeal across ages.



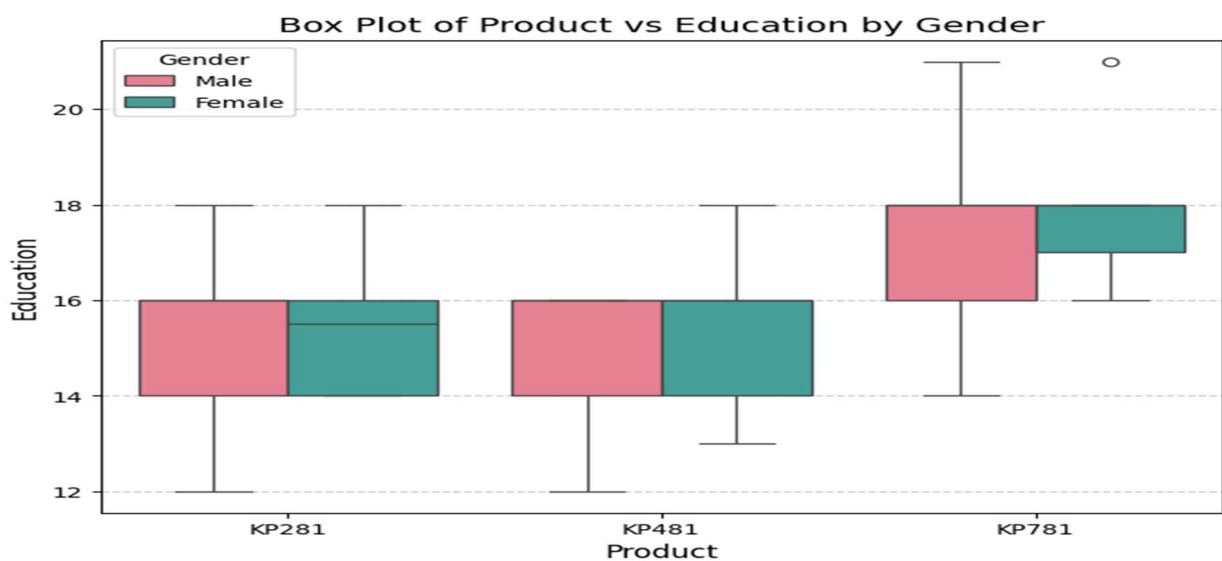
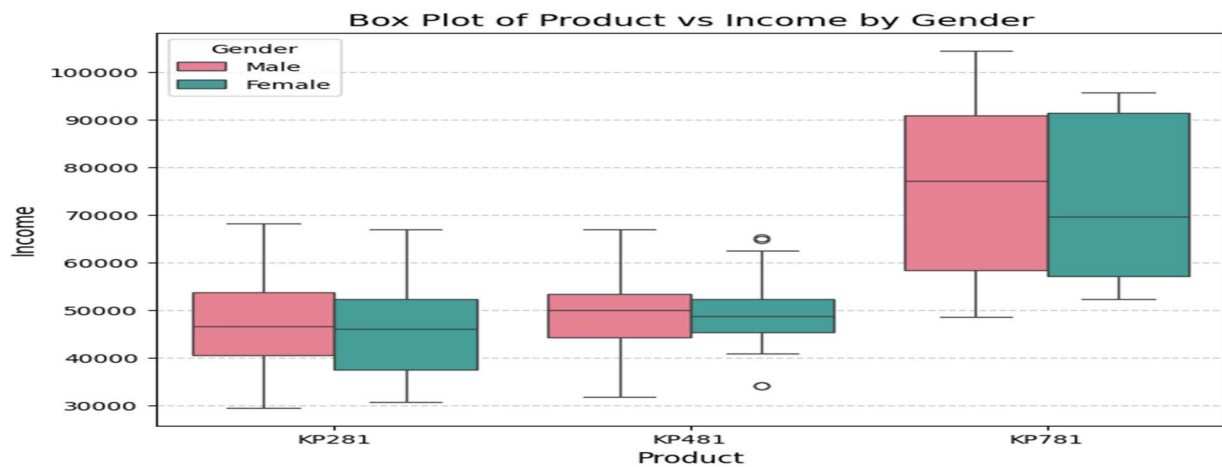
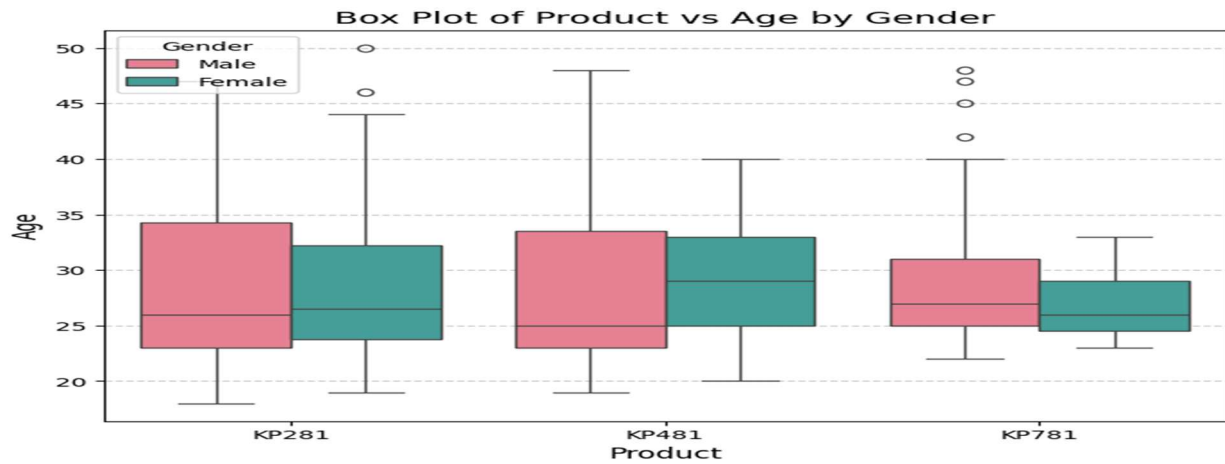
- **Insights:** KP781 appeals to both younger and older demographics, which could be leveraged for marketing strategies.

OVERALL INSIGHTS

- **Product KP281** appeals mainly to fitness-oriented, moderately educated, and younger users.
- **KP481** seems to attract less active individuals with lower incomes and education.
- **KP781** stands out with high-income, highly educated, and active users, representing a potential market segment for premium offerings.



MULTIVARIATE BOX PLOTS



1. Product vs. Education by Gender

- **Education Distribution:**

- **KP281:** Both males and females have similar median education levels around **16 years**. Males show a slightly wider spread, with some outliers at higher levels of education.
- **KP481:** The median education level is slightly lower, about **15.5 years**, for both genders, indicating it appeals to individuals with moderate education.
- **KP781:** Males show a higher median (around 18 years) compared to females (around 16 years), suggesting that KP781 appeals more to **well-educated individuals**, especially among males.

2. Product vs. Fitness by Gender

- **Fitness Distribution:**

- **KP281:** Both genders report higher fitness levels, with **males** averaging around **4.0** and **females** around **3.5**.
- **KP481:** Fitness levels are low for both genders, with **males** around **2.5** and **females** slightly below that (**around 2.0**), indicating KP481 appeals less to fitness-oriented consumers.
- **KP781:** **Males** average a higher fitness score (**around 4.5**) than **females** (**around 3.5**), reflecting that this product attracts more fit individuals across both genders.

3. Product vs. Income by Gender

- **Income Distribution:**

- **KP281:** Males earn slightly more than females, with medians around **\$50,000** for **females** and **\$55,000** for **males**.
- **KP481:** Both genders have lower income levels, with **males** at around **\$40,000** and **females** at **\$38,000**, indicating this product targets a lower-income demographic.
- **KP781:** This product shows the highest earnings, with **males** earning around **\$80,000** and **females** around **\$70,000**, appealing to a more affluent audience.

4. Product vs. Age by Gender

Age Distribution:

KP281: The median age is around **30 years** for **both genders**, with males showing a slightly broader age range than females.

KP481: Displays a similar age range for both genders, with medians around **31 years**. There are some outliers indicating older users.

KP781: **Males** have a lower median age (**around 28 years**) compared to **females** (**around 32 years**), suggesting it attracts both younger males and older females.



OVERALL INSIGHTS

- **Education and Income:** KP281 and KP781 attract more educated and higher-income individuals, particularly among males, which may influence targeted marketing efforts.
- **Fitness Appeal:** KP281 and KP781 appeal to fitness-oriented consumers, while KP481 lacks engagement from fitness-focused individuals, indicating the need for repositioning.
- **Age Factors:** While ages are relatively consistent across products, KP781 shows potential for engaging both younger and older demographics effectively.



7. CONDITIONAL PROBABILITIES

Product Distribution

The data shows the distribution of three different products:

- **KP281:** This is the most popular product, with 80 out of 180 total purchases (approximately 44.4%).
- **KP481:** This is the second most popular product, with 60 purchases (approximately 33.3%).
- **KP781:** This is the least popular product, with 40 purchases (approximately 22.2%).

Product -Gender percentages & Frequencies:

PRODUCT	GENDER		
	MALE	FEMALE	ALL
KP281	40	40	80
KP481	29	31	60
KP781	7	33	40
ALL	76	104	180

The table displays the number of purchases for each product (KP281, KP481, KP781) broken down by gender (Female, Male). It also includes totals for each product and gender, as well as an overall total.

Key Observations

1. **KP281:** This product has an equal split between male and female customers, with 40 purchases each.
2. **KP481:** There's a relatively even distribution for this product as well, with slightly more males (31) than females (29) purchasing it.
3. **KP781:** This product shows a significant difference in gender preference, with males (33) making up the majority of purchasers compared to females (7).
4. **Overall:** Out of the total 180 purchases, 104 were made by males and 76 by females, indicating a higher overall preference among males for these products.



Frequency Table

Based on the frequency table and given conditions, following are inferences:

1. **KP281**: This product appeals equally to both genders.
2. **KP481**: It has a slightly higher preference among males, but it's still relatively gender neutral.
3. **KP781**: This product is more popular among males.
4. Percentage of Male customers purchasing a treadmill: **57.78%**
5. Percentage of Female customers purchasing KP781 treadmill: **3.89%**
6. Probability of a customer being Female given that Product is KP281: **0.5**

Male Treadmill Purchases:

A majority (57.78%) of treadmill purchases are made by male customers. This suggests that males are a significant target market for treadmills in this dataset.

Female KP781 Purchases:

A small percentage (3.89%) of treadmill purchases are made by female customers specifically for the KP781 model. This indicates that the KP781 model is not as popular among female customers.

Female KP281 Purchases:

There's a 50% probability that a customer purchasing the KP281 model is female. This means the KP281 model is equally popular among male and female customers.

OVERALL INSIGHTS:

- **Marketing:** Males are the primary purchasers of treadmills; marketing efforts could be tailored to appeal to their preferences.
- **Product Development:** The KP781 model seems less popular among female customers. Marketing strategies to be devised for attracting female customers
- **Targeted Promotions:** Since the KP281 model is equally popular among both genders, promotions for this model could be targeted to both male and female customers.



AGE VS PRODUCT

- **KP281:** 51 customers (28.33% of total customers) in the 20-30 age range purchased this model.
- **KP481:** 34 customers (18.89% of total customers) in the 20-30 age range purchased this model.
- **KP781:** 30 customers (16.67% of total customers) in the 20-30 age range purchased this model.

Insights

1. **Target Demographic:** A significant portion of customers for all three treadmill models fall within the 20-30 age range.
2. **KP281 Popularity:** The KP281 model is particularly popular among customers in their 20s and 30s, with the highest percentage of purchases within this age group.

INCOME VS PRODUCT

- Percentage of low-income customers buying a treadmill: **46.11%**
- Percentage of high-income customers buying a treadmill given the product is KP781: **47.5%**

Insights

1. **Low-Income Customers:** A significant portion (46.11%) of treadmill purchases are made by customers in the low-income category. This suggests that treadmills are accessible to a wide range of income levels.
2. **High-Income within KP781:** Nearly half (47.5%) of the customers who purchased the KP781 model are high-income earners. This reinforces the idea that the KP781 is a popular choice among high-income customers.

FITNESS VS PRODUCT

- Percentage of customers that have fitness level 5: **17.22%**
- Percentage of customers with fitness level 5 buying KP781 treadmill: **72.5%**

Insights

1. **Fitness Level 5 Customers:** A relatively small percentage (17.22%) of your customers have a fitness level of 5.
2. **Fitness Level 5 within KP781:** A significant percentage (72.5%) of customers who purchased the KP781 treadmill have a fitness level of 5.



PRODUCT VS MARITAL STATUS:

Percentage of customers who are partnered and using treadmills: **59.44%**

Insights

This percentage tells us that a majority (nearly 60%) of treadmill customers are partnered.

This suggests that treadmills are a popular product among couples.



SUMMARY

AeroFit team analyzed consumer data to optimize the marketing and recommendations of their three treadmill products: KP281, KP481, and KP781. The study explored customer demographics, product preferences, and usage patterns using statistical and visualization techniques.

KEY FINDINGS

1. Demographics:

- The average customer is young (**28.8 years**), moderately educated (**15.6 years**), and earns about **\$53,720 annually**.
- A larger portion of the audience is male (**57.78%**) and partnered (**59.44%**).

2. Product Insights:

- KP281: The most popular product, purchased by **44.4%** of customers, appeals equally to both genders and younger age groups (**20–30 years**).
- KP481: A mid-range product bought by **33.3%** of customers, generally appeals to moderately educated individuals with lower incomes and fitness levels.
- KP781: The least popular product (**22.2%**) attracts high-income, highly educated, and highly fit individuals, indicating its premium positioning.

3. Usage and Fitness:

- The average treadmill user plans to exercise **3.5** times per week and walk/run **103.2** miles weekly.
- Fitness and income are **positively correlated**, with higher earners often reporting better fitness levels and more treadmill engagement.

4. Income and Education:

- **Higher income** is strongly associated with **higher education** and **fitness levels**.
- KP781 is the preferred choice among **high-income** earners.

5. Behavioral Trends:

- **Younger** individuals **favor KP281** due to its versatility and appeal to moderate fitness users.
- **KP781** is preferred by fitness enthusiasts with **higher incomes**.

RECOMMENDATIONS:

- Focus on KP281 as a versatile product for a broad demographic.
- Enhance the visibility of KP481 by addressing its lower engagement among fitness-focused individuals.
- Promote KP781 as a premium product for affluent and fitness-oriented audiences, potentially with targeted campaigns for high-income groups.

