**INTRODUCTION**

The wellness sector is a rapidly changing environment, continually adapting to satisfy the varied demands of health-aware clients. In this competitive arena, staying ahead necessitates a profound comprehension of your customer base. To optimize sales and target the right audience, this project analyzes Cardio Good Fitness's treadmill customer data, aiming to develop detailed buyer personas for each model. This translates to understanding not just who purchases their treadmills, but also why they prefer certain models. By examining customer data, Cardio Good Fitness can obtain valuable insights to improve sales, reach the appropriate audience, and ultimately win over potential treadmill buyers' preferences (and finances). This project explores an examination of customer trends in treadmill purchases at Cardio Good Fitness over the last three months. Our objective is to create in-depth customer profiles for each treadmill model (TM195, TM498, TM798) and use these insights to enhance our marketing strategies and recommendations for sales.

**DATA DESCRIPTION**

The variables to be examined include:

* Product purchased (TM195, TM498, or TM798): This specifies which treadmill model the customer has selected. Each model is tailored to meet distinct user preferences, requirements, and budget constraints.
* Age (in years): This specifies the customer's age, which can impact fitness levels, workout routines, and the preferred features or programs in a treadmill.
* Education (in years): This quantifies the number of years of formal education completed by the customer. Education level may be associated with income levels and awareness of health-related matters.
* Marital status (single or partnered): This denotes whether the customer is single or in a relationship. Relationship status can affect fitness routines and choices regarding purchases.
* **Gender:** This denotes whether the customer identifies as male or female. Gender can occasionally impact fitness objectives, equipment preferences, and exercise routines.
* Annual household income ($): This reflects the total income earned by all members of the customer's household per year. Income can affect affordability and the readiness to invest in fitness equipment.
* The average number of times the customer plans to use the treadmill each week: This represents the frequency of treadmill usage per week, indicating the customer's dedication to fitness and their perceived value of owning a treadmill.
* Average number of miles the customer expects to walk/run each week: This quantifies the projected weekly distance the customer plans to cover on the treadmill, reflecting their fitness goals and exercise intensity.
* Self-rated fitness on a 1-to-5 scales (1 = poor shape, 5 = excellent shape) (Fitness): This is a subjective assessment provided by the customer to evaluate their current fitness level. It offers insight into how they perceive their health and fitness, influencing their exercise routines and preferences for equipment.

**EXPLORATORY DATA ANALYSIS**

The dataset consists of 180 rows and 9 columns, presenting a detailed overview of the collected information. Each of the 180 rows signifies a distinct observation or entry, documenting specific instances relevant to the study. The 9 columns in the dataset encompass a range of attributes and details, including demographic data and key metrics pertinent to the analysis. These columns cover aspects such as age, gender, education level, marital status, and other pertinent variables. The arrangement of this dataset enables a thorough examination of the data, supporting both descriptive and inferential analysis to derive valuable insights.

**I conducted a thorough examination of the dataset's structure, revealing no missing data.** Categorical variables, including Marital Status, Gender, and Product, were differentiated from numerical variables such as Age, Education, Usage, Fitness, Income, and Miles. This distinction proved crucial for selecting appropriate analytical approaches. Categorical data necessitated frequency distributions and cross-tabulations to uncover patterns, while numerical data required the calculation of central tendency and dispersion measures. This foundational step was essential for subsequent in-depth exploration and analysis of the dataset.

In the course of exploratory data analysis, I examined the following variables:

A chart of different colored rectangles

Description automatically generated I analyzed the categorical variable of treadmill products (TM195, TM498, and TM798) to understand their frequency distribution. I counted the occurrences of each product i.e (80 (44.44%), 60 (33.33%), and 40 (22.22%) of TM195, TM498, and TM798 were bought respectively) and visualized the distribution using a bar plot as shown in Figure 1. Clearly we can see, there were 3 types of unique models in the store i.e TM195, TM498, and TM798. The most bought treadmill was TM195. The average age of customers purchasing the TM195 model is 28.55 years old.  
 A blue and pink rectangles

Description automatically generated

Figure 1.Histogram of Product Purchase

Figure 2. Histogram of Gender

**In Figure 2, Male customers dominated treadmill purchases, accounting for 104 out of the 180 sales.** This gender disparity highlights a potential area for further investigation to understand factors influencing product preference.

The educational backgrounds of customers displayed notable variability within the dataset. The range of educational attainment extended from 12 to 21 years, reflecting a diverse spectrum of academic qualifications among customers. The median education level was found to be 16 years, indicating that a substantial portion of the customer base had achieved a higher secondary education or some level of tertiary education. These insights into educational demographics offer a clear understanding of the academic qualifications prevalent among the customers.

Based on the summary, the average age of treadmill purchasers is 28.79 years, with an average education level of 15.57 years. Additionally, the average household income of these customers is $53,720, reflecting a moderate to high earning capacity within the customer base. These demographic characteristics provide a foundational understanding of the target market and can be leveraged to refine marketing strategies and product offerings.

A green and blue squares

Description automatically generated As illustrated in Figure 3, partnered individuals represent the majority of treadmill purchasers, with 107 partnered customers compared to 73 single customers. This observation suggests that couples are more inclined to invest in home fitness equipment than single individuals. This insight can be utilized to tailor marketing strategies and product offerings to better address the preferences and needs of partnered consumers.

Figure 3. Histogram of Marital Status

Treadmill customers report a moderate level of perceived fitness, with an average self-rated fitness score of 3.31 on a 1-to-5 scale. On average, customers expect to use the treadmill approximately 3.46 times per week and cover 103.2 miles weekly. These metrics provide valuable insights into customer exercise habits and expectations, which can be instrumental in guiding product development and tailoring marketing strategies.

A screenshot of a graph

Description automatically generatedThere is a strong positive correlation between several key variables: fitness and usage, fitness and income, miles and usage, product TM798 and fitness, product TM798 and usage, and product TM498 and income. These correlations suggest that customers with higher fitness levels tend to use the treadmill more frequently, have higher incomes, and are more likely to purchase higher-end models like the TM798 and TM498. This information is crucial for understanding customer behavior and preferences, aiding in targeted marketing and product positioning strategies.   
There is a strong negative correlation between the TM195 model and several key variables such as income, fitness, and usage. This indicates that customers who purchase the TM195 model tend to have lower incomes, lower fitness levels, and use the treadmill less frequently. Understanding these patterns can help tailor marketing efforts and product offerings to better meet the needs of this customer segment.

Figure 4. Correlation Matrix

A diagram of a graph

Description automatically generated with medium confidence

Figure 5 illustrates the income distribution across different treadmill models. As previously discussed, customers who purchased the TM798 model tend to have significantly higher incomes compared to those who bought the TM195 and TM498 models. Conversely, customers who opted for the TM195 model tend to have lower incomes. This visual representation reinforces the relationship between income level and product choice, clearly distinguishing among the three customer segments.

Figure 5. boxplot of income vs treadmills

A diagram of a graph

Description automatically generated with medium confidence

The provided box plot in Figure 6 visually compares the age distribution of customers who purchased different treadmill models (TM195, TM498, and TM798). The age range of customers purchasing all three models is relatively similar, from approximately 20 to 50 years. The median age for customers purchasing the TM195 model appears slightly lower than that of the other two models, suggesting a younger demographic for TM195 buyers. There are some outliers, particularly in the higher age range, for all three models, indicating that a few older customers are purchasing each type of treadmill. Additionally, the distribution of ages within each model group seems somewhat skewed to the right, with a larger proportion of younger customers. This visual comparison highlights the central tendency, variability, and potential outliers in the age data for each model, allowing for a clear understanding of the demographic differences among customers who prefer each treadmill model. These insights can inform targeted marketing strategies and product development decisions.

Figure 6. Boxplot of Age vs Treadmills

A graph of a number of people

Description automatically generatedFigure 7 presents a histogram of customer age distribution, highlighting that the majority of customers are in the 25-30 age range. There is a noticeable drop in the number of customers as age increases beyond 30, with fewer customers in the 20-25 age range and progressively fewer from 30 to 50 years. This indicates that the primary customer base consists mainly of individuals in their mid-20s to early 30s. Marketing and product development should focus on this demographic, while also exploring strategies to attract older age groups.

Figure 7 Histogram of age of customers

Figure 8 displays a pair plot illustrating the relationships between age, income, and education for customers, differentiated by treadmill model (TM195 in red, TM498 in green, and TM798 in blue). The diagonal panels show the distribution (density plots) of each variable, revealing that TM498 customers generally have higher income and education levels. The upper off-diagonal panels display the correlation coefficients between the variables, indicating significant positive correlations, particularly between income and age (Corr: 0.513) and between income and education (Corr: 0.626). The lower off-diagonal panels provide scatter plots of the variable pairs, demonstrating positive trends and clustering by treadmill model, with TM498 customers forming distinct groups with higher values in income and education. This pair plot helps to visualize the distribution and interrelationships of these variables, highlighting distinct demographic characteristics associated with each treadmill model.A screenshot of a graph

Description automatically generated

Figure 8 pair plot of age, income and education by product

Figure 9 shows a box plot comparing fitness scores across different treadmill models. TM798 exhibits a significantly higher and more variable range of fitness scores compared to TM195 and TM498, which both have lower and nearly identical fitness scores. This suggests that users of TM798 generally rate their fitness much higher than those using the other two models. The greater variability in fitness scores for TM798 also indicates that this model appeals to a broader spectrum of fitness levels, potentially attracting users with higher fitness aspirations.A blue square on a white background

Description automatically generated

Figure 9 boxplot of fitness score vs product

A graph with different colored rectangles

Description automatically generated Figure 10 shows a box plot comparing weekly treadmill usage across different models. TM195 exhibits moderate weekly usage with some variability, TM498 has the lowest and most consistent usage, and TM798 demonstrates the highest and most variable usage. This indicates that users of TM798 tend to use their treadmills more frequently and with greater variation compared to users of TM195 and TM498.

Figure 10 boxplot of weekly usage vs product

**CONCLUSION AND RECOMMENDATIONS**

Based on the analysis, TM195 emerges as the most popular treadmill model, representing 44.44% of sales. This model should be marketed as a reliable and budget-friendly option, highlighting its affordability and essential features to attract a broad range of customers. TM195 is associated with lower income levels, indicating that it is a preferred choice for cost-conscious consumers seeking value for money.

In contrast, TM798 stands out as the model attracting higher-income customers. It also shows significantly higher and more variable fitness scores, suggesting that it appeals to those looking for advanced features and a premium fitness experience. Marketing TM798 as a high-end, luxurious treadmill with superior capabilities will cater to professionals and higher-income individuals willing to invest more in their fitness equipment.

TM498, associated with higher income and education levels compared to TM195, represents a mid-range option. It should be promoted as a treadmill that balances affordability and enhanced features, appealing to customers seeking quality without the highest price tag.

Furthermore, TM798 users exhibit the highest and most variable treadmill usage, indicating strong engagement and varied fitness goals. Emphasizing the versatility and advanced features of TM798 can attract users with diverse fitness needs and higher aspirations. By aligning marketing strategies with these observations, Cardio Good Fitness can effectively target different customer segments, improve product positioning, and enhance overall sales and satisfaction.