

Table of Contents

Leveraging Descriptive Analytics for Customer Segmentation: A Comprehensive Analysis of CardioGood Fitness Treadmill Users.....	2
Introduction	2
Methodology.....	3
Data preparation	3
Analysis.....	4
Customer Profiles By Treadmill Product Line:.....	7
Marital and Gender Distribution of CardioGood Fitness Customers by Treadmill Product Line.....	9
Correlation Analysis between Income and the Product Model	11
Conclusion and Recommendations.....	12

Leveraging Descriptive Analytics for Customer Segmentation: A Comprehensive Analysis of CardioGood Fitness Treadmill Users

Introduction

In today's competitive fitness equipment market, understanding customer preferences and tailoring products to cater to their needs is crucial for business to thrive. CardioGood Fitness, a leading manufacturer of high-quality treadmills, aims to enhance its marketing strategies and improve its product offerings by delving deeper into its customer base. This report focuses on leveraging descriptive analytics to segment CardioGood Fitness treadmill users, providing valuable insights into the typical customer profile for each treadmill product line.

With a diverse range of treadmills offered by the company, namely TM195, TM498, and TM798, it is essential to identify unique characteristics of the customers who purchase each product. By studying variables such as product purchased, gender, age, education, relationship status, annual household income, average planned weekly treadmill usage, average expected weekly miles, and self-rated fitness, we aim to create a comprehensive profile for each product line's target.

The objective of the analysis is to uncover patterns and trends in customer preferences, enabling the marketing team at AdRight to devise more effective marketing campaigns and promotional strategies for CardioGood Fitness. Additionally, the findings may help the company in refining its product development approach, allowing them to cater to the specific needs and desires of their customers more effectively.

This report will detail the methodology employed to collect, prepare and analyze the data, followed by the results of the descriptive analytics and a comparative analysis of customer profiles across the treadmill product lines. Finally, we will present the conclusions drawn from our findings and offer recommendations to support AdRight's marketing initiatives and CardioGood Fitness's business growth.

Methodology

The methodology for this report is structured into three main parts: Data collection, data preparation and data analysis.

1. Data Collections:

The primary data source for this analysis is the CardioGoodFitness.csv file, which contains customer data for individuals who purchased a treadmill at CardioGood Fitness retail store during the prior three months. The dataset has 180 rows, each representing a unique customer, and includes the following variables:

- Product: The treadmill product purchased(TM195, TM498, TM798)
- Age: Customers age in years
- Gender: Customers gender(Male or Female)
- Education: Customers education in years
- MaritalStatus: Customers relationship status(Single or Partnered)
- Usage: Average number of times the customer plans to use the treadmill each week
- Fitness: Self-rated fitness on a 1 to 5 scale, where 1 is poor shape and 5 is excellent shape
- Income: Customer annual household income in dollars
- Miles: Average number of miles the customer expects to walk/run each week.

Data preparation

Before conducting the analysis, the data will be prepared using the following steps:

- Import necessary libraries, such as pandas and numpy for data manipulation, and matplotlib and seaborn for visualization.
- Load the CardioGoodFitness.csv file into a pandas DataFrame for easy data manipulation and analysis.

- Check for missing values, duplicates or outliers and handle them accordingly(e.g impute missing values, remove duplicates, or winsorize outliers)
- Ensure data is in the appropriate format for each variable(e.g numerical variables as integers or floats, and categorical variables as strings or categorical data types).

Analysis

After preparing the data, descriptive analytics techniques will be applied to create customer profiles for each treadmill product line. This process will involve:

- Grouping the data by the products purchased(TM195, TM498 and TM798) to analyse customer characteristics separately for each treadmill product line.
- Calculating summary statistics(mean, median, mode, standard deviation, and range) for numerical variables within each product group.
- Analysing the distribution of categorical variables(gender and relationship status) within each product group, using frequency tables or visualizations such as bar charts.
- Conducting a comparative analysis of customer profiles across the treadmill product lines, using visualizations like box plots, bar charts, or histograms to illustrate similarities and differences among customer characteristics.

First we start by importing the necessary libraries:

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

Then we load the data into our DataFrame aliased as “cgf”:

```
#Load data
cgf = pd.read_csv("CardioGoodFitness.csv")
```

Explore the data:

```
#Explore the data
cgf.head()
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles
0	TM195	18	Male	14	Single	3	4	29562	112
1	TM195	19	Male	15	Single	2	3	31836	75
2	TM195	19	Female	14	Partnered	4	3	30699	66
3	TM195	19	Male	12	Single	3	3	32973	85
4	TM195	20	Male	13	Partnered	4	2	35247	47

```
#check data type
```

```
cgf.info()  
cgf.describe()
```

```
<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
```

	Age	Education	Usage	Fitness	Income	Miles
count	180.000000	180.000000	180.000000	180.000000	180.000000	180.000000
mean	28.788889	15.572222	3.455556	3.311111	53719.577778	103.194444
std	6.943498	1.617055	1.084797	0.958869	16506.684226	51.863605
min	18.000000	12.000000	2.000000	1.000000	29562.000000	21.000000
25%	24.000000	14.000000	3.000000	3.000000	44058.750000	66.000000
50%	26.000000	16.000000	3.000000	3.000000	50596.500000	94.000000
75%	33.000000	16.000000	4.000000	4.000000	58668.000000	114.750000
max	50.000000	21.000000	7.000000	5.000000	104581.000000	360.000000

Group the Products by their respective means:

```
#group data by product purchased
```

```
grouped_cgf = cgf.groupby("Product")
```

```
grouped_cgf_means = grouped_cgf.mean()  
grouped_cgf_means
```

	Age	Education	Usage	Fitness	Income	Miles
Product						
TM195	28.55	15.037500	3.087500	2.9625	46418.025	82.787500
TM498	28.90	15.116667	3.066667	2.9000	48973.650	87.933333
TM798	29.10	17.325000	4.775000	4.6250	75441.575	166.900000

The snippet above shows that there are differences in customer characteristics across the product lines. For instance, TM798 customers have a higher education level, plan to use the treadmill more frequently, rate their fitness higher, have a higher income, and expect to walk/run more miles per week compared to customers of TM195 and TM498.

Customer Profiles By Treadmill Product Line:

```
grouped_cgf[['Age', 'Education', 'Usage', 'Fitness', 'Income', 'Miles']].agg([np.mean, np.median, np.std, np.min, np.max])
```

Education						... Income						Miles								
amin	amax	mean	median	std		amin	amax	...	mean	median	std		amin	amax	mean	median	std		amin	amax
1452	18	50	15.037500	16.0	1.216383	12	18	...	46418.025	46617.0	9075.783190	29562	68220	82.787500	85.0	28.874102		38	188	
5248	19	48	15.116667	16.0	1.222552	12	18	...	48973.650	49459.5	8653.989388	31836	67083	87.933333	85.0	33.263135		21	212	
1738	22	48	17.325000	18.0	1.639066	14	21	...	75441.575	76568.5	18505.836720	48556	104581	166.900000	160.0	60.066544		80	360	

The above snippet provides us with additional statistics for each variable within each product group. The output includes the mean, median, standard deviation, minimum and maximum for each variable. From the result, we can infer that:

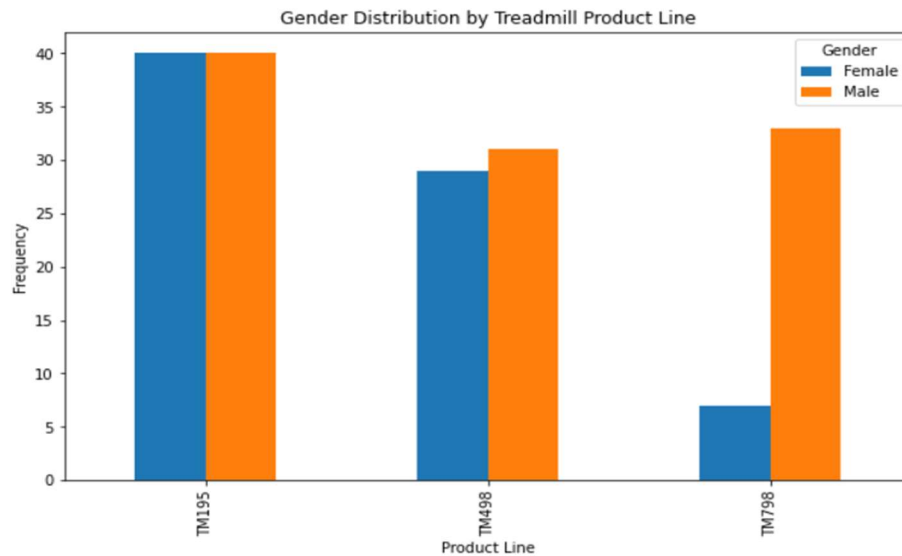
- TM195 and TM498 customers have similar age distributions, with TM798 customers being slightly older on an average. The median age for each product is close to the mean, which suggests a relatively symmetrical age distribution.
- Customers who purchase the TM798 treadmill, like we have said before, have a higher average education level compared to those who purchase the TM195 and TM498. This could indicate that higher educated customers are more likely to invest in a premium treadmill model.
- Average usage and fitness levels of TM798 customers are notably higher than those of TM195 and TM498 customers which suggest that customers who plan to use the treadmill more frequently and rate their fitness higher are more likely to choose the TM798 model.

- The income range for each product line varies significantly, with TM798 customers having the highest average and median income. This could indicate that customers with higher incomes are more likely to purchase a premium treadmill model.
- Lastly, the average and median miles that customers expect to walk/run each week are the highest for the TM798 model, reflecting the higher fitness and usage levels of these customers. The standard deviation for miles is also larger for TM798 customers, indicating a wider range of expected mileage among this group.

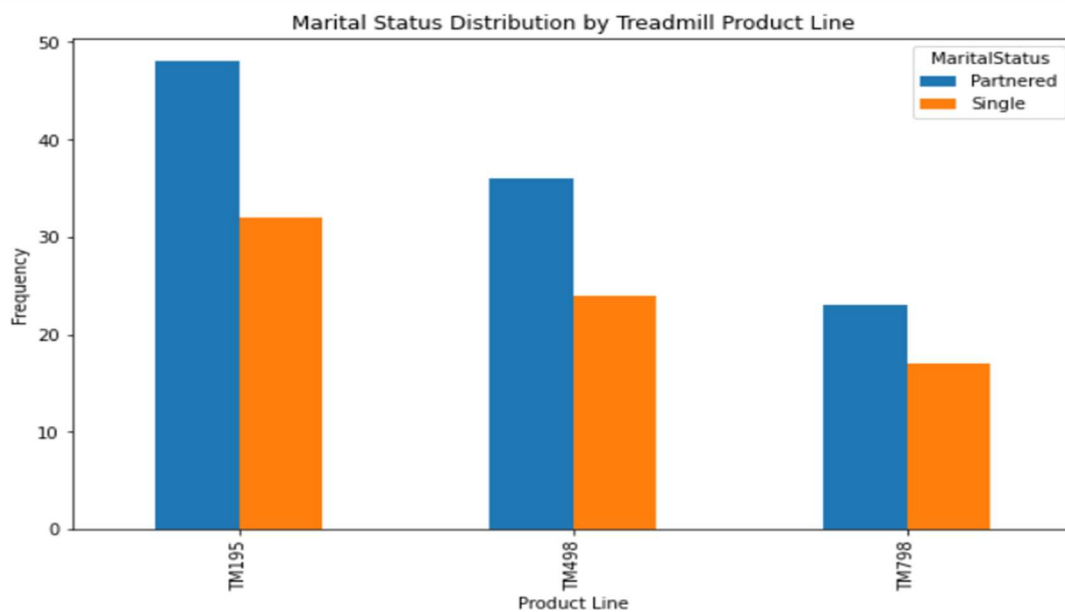
Concisely, the AdRight market research team can use this information to develop targeted marketing strategies for each CardioGood Fitness treadmill product.

Marital and Gender Distribution of CardioGood Fitness Customers by Treadmill Product Line

```
#visualize gender distribution using bar chart
gender.plot(kind='bar', figsize=(10, 6))
plt.title("Gender Distribution by Treadmill Product Line")
plt.xlabel("Product Line")
plt.ylabel("Frequency")
plt.show()
```



```
#visualize maritalstatus distribution using bar chart
maritals.plot(kind='bar', figsize=(10, 6))
plt.title("Marital Status Distribution by Treadmill Product Line")
plt.xlabel("Product Line")
plt.ylabel("Frequency")
plt.show()
```



From the snippet above, the TM195 model has a ratio of approximately 3:2 for partnered to single users. This suggests that it's more favored by those in a relationship or living together. One reason could be that it offers features or pricing that appeal to a dual-income household, or it's simply more appealing to people who live together and potentially share fitness equipment. It might also infer that this model provides a more balanced mix of features that cater to a wider range of fitness levels or workout preferences. Also, in terms of gender distribution, there is an even distribution between male and female users (40 each).

The question is, what makes it so appealing to both genders?

Perhaps its features, pricing and marketing position the model as versatile, all-inclusive and therefore caters to a broad range of fitness levels and workout preferences. This gender mix might indicate that the TM195 model is suitable for marketing campaigns that highlight its broad appeal and versatility.

The TM498 model has more single users than partnered users. It seems to be favored by individuals who are not in partnered relationships. This might imply that this model is more compact, cost-effective, or user-specific model, suitable for individuals living alone. The product could be perceived as practical and straightforward, suitable for users who are not looking to share equipment. Gender-wise, the distribution is slightly skewed. The slight male preference could indicate that this model offers features or pricing that slightly more males find appealing, but the difference is not significant enough to label it a male-oriented product.

The user base of the TM798 model is skewed towards partnered individuals. This could indicate that the model is a premium product, appealing to those with more disposable income or those who take their fitness regimen more seriously. It may have features that are geared towards advanced users or offer more customization, which may explain the appeal to a wider range of users within a household. Like the previous model, the TM798 is skewed toward male users, with 33 males compared to only 6 females. This suggests that the model offers features, pricing or branding that strongly appeals to males. The male dominance might imply that the TM798 is a high-performance model or has features that are more appealing to serious fitness enthusiasts, a category often associated with males due to social norms.

To put it all together, the TM195 seems like a 'one-size-fits-all' model suitable for a variety of users, while the TM498 may be a 'individualistic' model, and the TM798 seems like a 'premium' model. Each of these treadmills could be marketed towards their respective demographic for optimal results. In combination with the marital status data, the gender data can also inform more nuanced marketing strategies. For example, the TM195 could be marketed as a versatile, family-friendly model suitable for everyone. The TM498, with a slight lean towards male and single users, could be positioned as a practical, user friendly model for individual users. The TM798, with its clear preference among partnered males, could be marketed as a premium. High-performance model for fitness enthusiasts.

Correlation Analysis between Income and the Product Model

The correlation analysis reveals important insights into the relationship between customers income levels and their product preferences.

First, as seen in the snippet, Product TM195 shows a negative correlation of -0.396744 with the customers' income. This moderate negative correlation suggests that the TM195 treadmill model is typically preferred by the customers belonging to the lower income bracket. This could be due to the affordability of the model as compared to other models. TM498 also reveals a weak correlation of -0.203871 with income. While this is not as pronounced as the former, it suggests that the model also finds some favor with customers in the lower to middle income range. However, given the weaker correlation, it is likely that the choice for this model is influenced by factors beyond just income levels. Rather, it could be that this model is seen as a 'Middle ground' option offering a balance between affordability and advanced features.

```

cgf_encoded = pd.get_dummies(cgf, columns=['Product'])

corr = cgf_encoded.corrwith(cgf_encoded['Income']).filter(like='Product')

print(corr)

```

Product_TM195	-0.396744
Product_TM498	-0.203871
Product_TM798	0.705368
dtype:	float64

Lastly, the TM798 on the other hand, demonstrates a strong positive correlation with income. This implies that customers who opt for this model generally have higher income levels. This model could be priced higher due to superior features or a more prestigious brand image. The strong correlation suggests that income is a significant factor influencing the choice for this model. Therefore, marketing campaigns for the TM798 model may be more effective if they are aimed at affluent neighbourhoods, high-end fitness centers, or platforms frequented by a wealthier demographic.

Conclusion and Recommendations

This comprehensive descriptive analytics study has gleaned profound insights into CardioGood Fitness's customer base, highlighting the distinctive profiles for each of the treadmill product lines – TM195, TM498 and TM798.

1. The TM195 model, boasting an even gender distribution and being more favoured by partnered customers, portrays itself as a versatile and inclusive option. It caters to a wide array of fitness enthusiasts irrespective of their fitness levels and workout preferences. It appeals to individuals across different demographics too, making it a go-to choice for this seeking a flexible fitness solution. Accordingly, marketing strategies for this model should accentuate its versatility and universal appeal, targeting a broad range of people and emphasizing its capability to cater to a diverse range of fitness goals.
2. The TM498 model leans slightly towards the male demographic and has a higher prevalence among the single users. This pattern suggests its status as

a practical, efficient model, which is particularly suited for individuals living alone or who prefer solo workouts. This trend might be indicative of model's value for money proposition, which amalgamates affordability with different features, making it an attractive option for cost-conscious fitness enthusiasts. As such, marketing campaigns for TM498 should underline its cost-efficiency and suitability for individual users, tailoring the messaging towards single, male users who seek a combination of affordability and functionality.

3. The TM798 model shows a clear inclination towards partnered male users, indicating its image as a premium, high-performance product. This user preference suggests that the model offers advanced features and more comprehensive fitness solution that caters to dedicated fitness enthusiasts, typically associated with higher income brackets. The marketing efforts for TM798 should therefore capitalize on its advanced features, premium build, and benefits for committed fitness enthusiasts. The campaigns should specifically target higher-income, fitness-conscious demographics, underscoring the model's unique selling points and positioning it as a top-tier choice for those who take their fitness regimes seriously.

The Correlation analysis fortifies these insights, evidencing a negative correlation between income and preference for TM195 and TM498 models and a strong positive correlation between income and the preference for the TM798 model. This pattern suggests that the TM195 and TM498 models are preferred by customers in the lower to middle income brackets, the TM798 model is particularly appealing to customers in the higher income bracket. This information can be used to guide pricing strategies, with potential scope for introducing more upscale features in the lower-end models or a budget-friendly option in the premium segment.

On the basis of these insights, we recommend AdRight to adopt a differentiated marketing approach for each product line:

- For the TM195, employ a marketing strategy that emphasizes its universal appeal and versatility. Promotion could highlight its suitability for both genders and all relationship statuses, perhaps positioning it as a great choice for shared usage within households.
- The TM498, marketing campaigns could focus on its practicality and cost-efficiency, appealing to individual users who seek user-friendly, practical model. The slight male bias could also be considered in creating targeted promotions.
- Lastly, the TM798 marketing should underscore its advanced features, high performance, and appeal to fitness people. Targeting strategies should aim at affluent neighbourhoods, high-end fitness centres, or platforms frequented by wealthier demographics.