CARDIO GOOD FITNESS

EXPLORATORY DATA ANALYSIS

by **JAKE EIDE**

BACKGROUND & OBJECTIVE

Cardio Good Fitness is a retail store that sells treadmills. The objective of this study is to perform preliminary data analysis on the dataset, extract basic observations, and discover lines of questioning that could be relevant for the business. The differences between the customers of each treadmill product will be identified, as well as the relationships between different attributes of customers. This analysis has led to a set of insights and recommendations that will help the company target new customers.

DATA INFORMATION

The data contains information about 180 Cardio Good Fitness customers.

VARIABLE	DESCRIPTION	DATA TYPE	RANGE
Age	of the customer – in number of years	Numerical	18 - 50 years
Education	of the customer – in number of years	Numerical	12 - 21 years
Income	of the customer	Numerical	29562 - 104581
Gender	of the customer	Categorical	Male, Female
Marital Status	of the customer	Categorical	Partnered, Single
Product	the model number of the treadmill	Categorical	TM195, TM498, TM798
Usage	avg. number of times the customer wants to use the treadmill	Numerical	2 - 7 times per week
Fitness	self rated fitness score of the customer (5 - very fit, 1 - very unfit)	Numerical	1 - 5
Miles	expected to run	Numerical	21 - 360 miles

EXPLORATORY DATA ANALYSIS:
CORRELATION MATRIX

Observations:

• *Higher correlation (above .50):*
Fitness / Miles
Usage / Miles
Usage / Fitness
Income / Education
Income / Fitness
Income / Usage
Income / Age

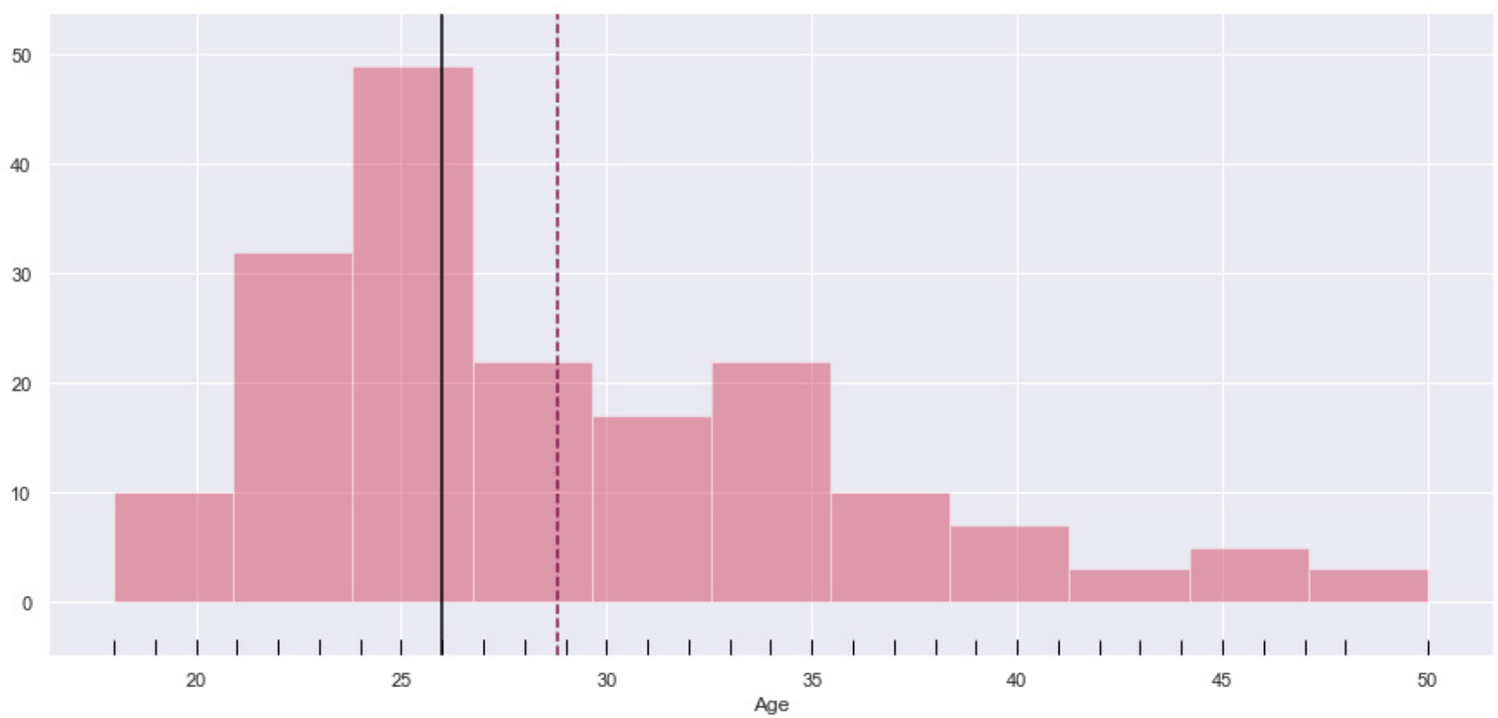
• *Moderate correlation (between 0.30 and 0.49):*
Education / Fitness
Education / Usage
Education / Miles

• *Lower correlation (between 0.29 and 0.10):*
Education / Age

• *Lowest correlation (between 0.09 and 0.00):*
Age / Fitness
Age / Miles
Age / Usage

EXPLORATORY DATA ANALYSIS: **AGE**

From this data set, we can draw some generalizations about the attributes of a typical Cardio Good Fitness customer. Looking at the histogram and box plots to the right, we see the ages of CGF customers have a spread of 32 years – ranging from the youngest customers at 18 years up to the oldest at 50 years.



Observations:

• The distribution of age is right skewed, with the majority of the customers on the younger side

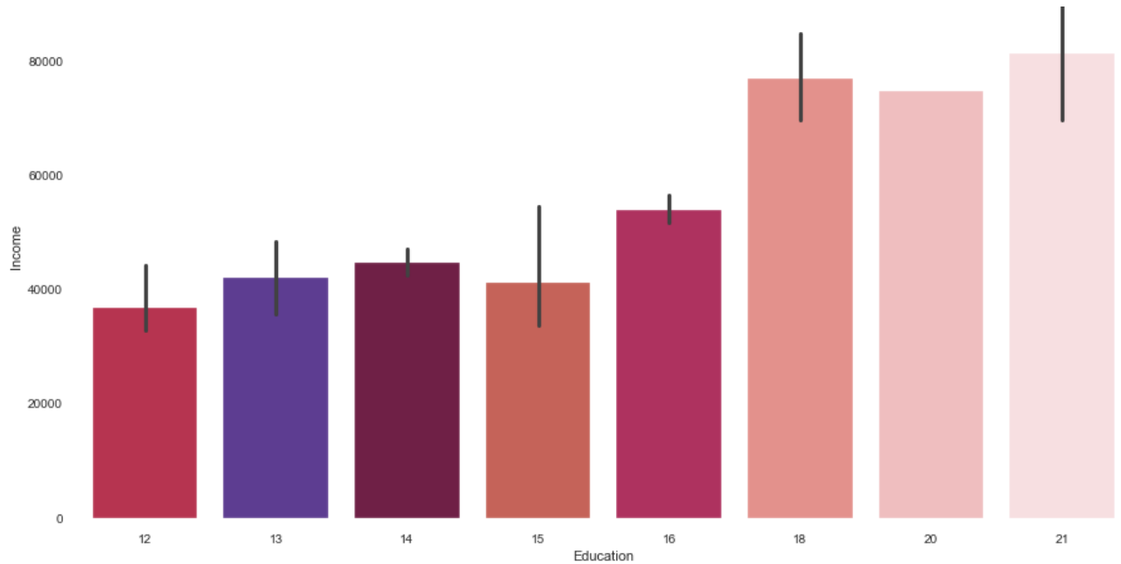
Observations on Central Tendency:

• Mean Age is 28.8 years old
• Median Age is 26 years old
• Mode is 25 years old
• 63% of the customers are between ages 23 and 33

Older and Younger Customers:

• Customers 22 and younger account for 13%
• Customers 34 and older account for 23%

EXPLORATORY DATA ANALYSIS:
INCOME & EDUCATION



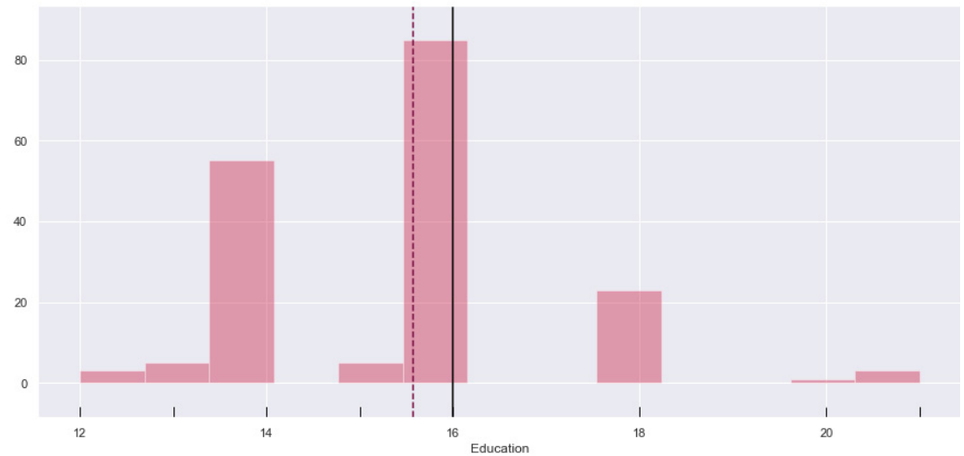
In the correlation matrix, we saw that income and education have a high correlation. The chart above shows those customers with more years of education have higher incomes.

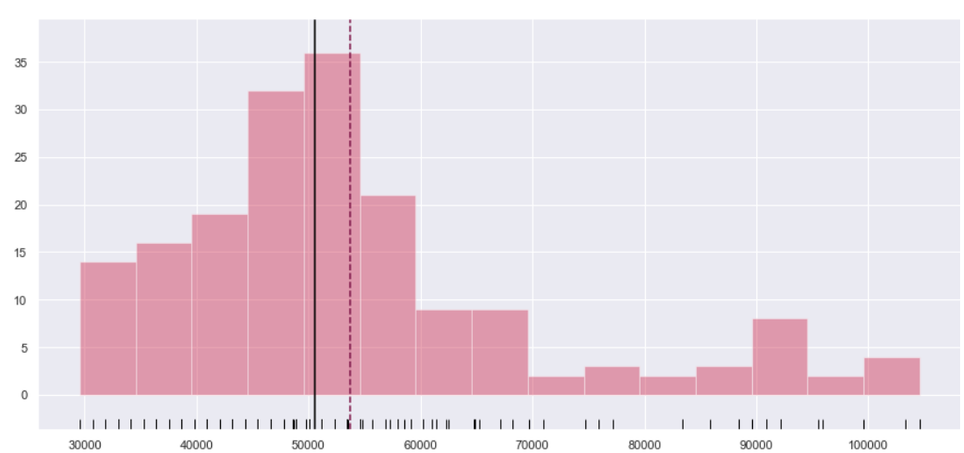
Observations:

- The distributions for educationand income are both right skewed
- The wealthiest 25% of customers have incomes starting at 58,668 and up to 104,581

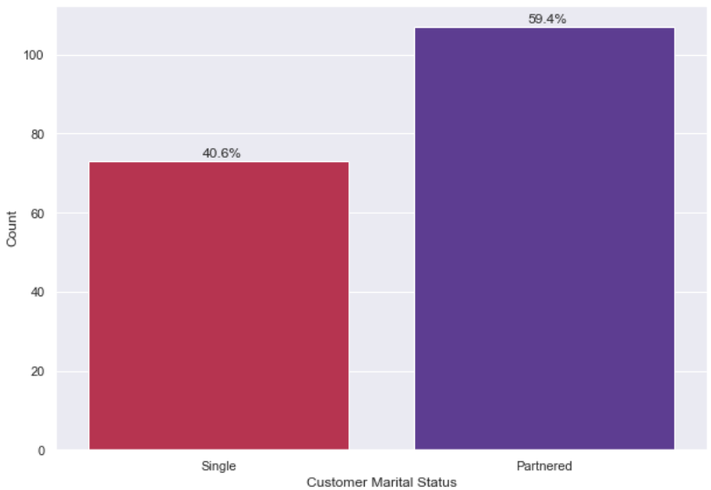
Observations on Central Tendency:

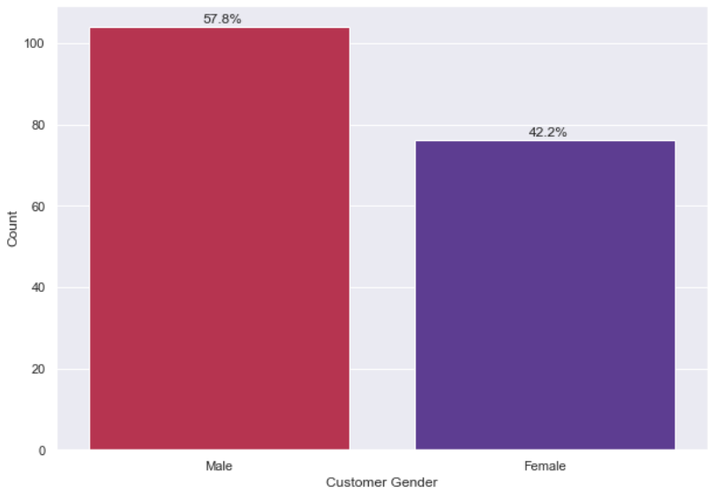
- Mean Education is 15.6 years / Mean Income is 53,720
- Median Education is 16 years / Median Income is 50,597
- From this information, we can infer that the average customer likely has about 16 years of education and an income between 50K and 55K

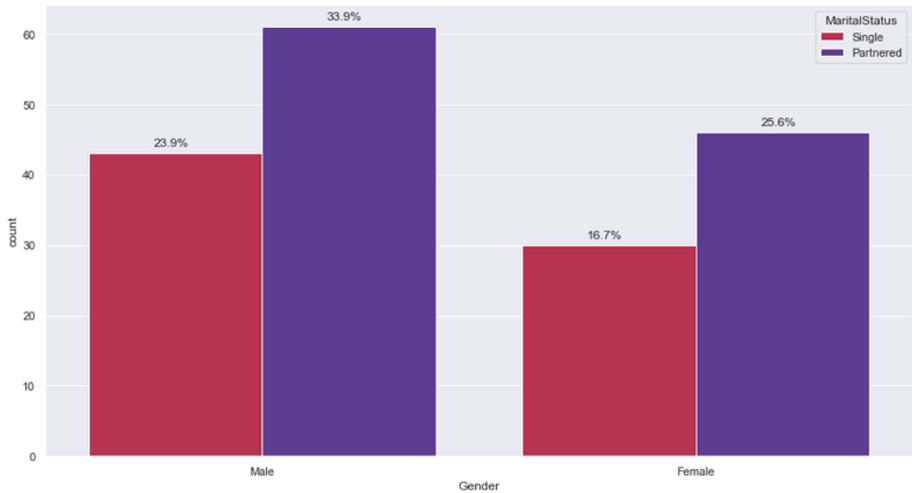




EXPLORATORY DATA ANALYSIS:
MARITAL STATUS & GENDER





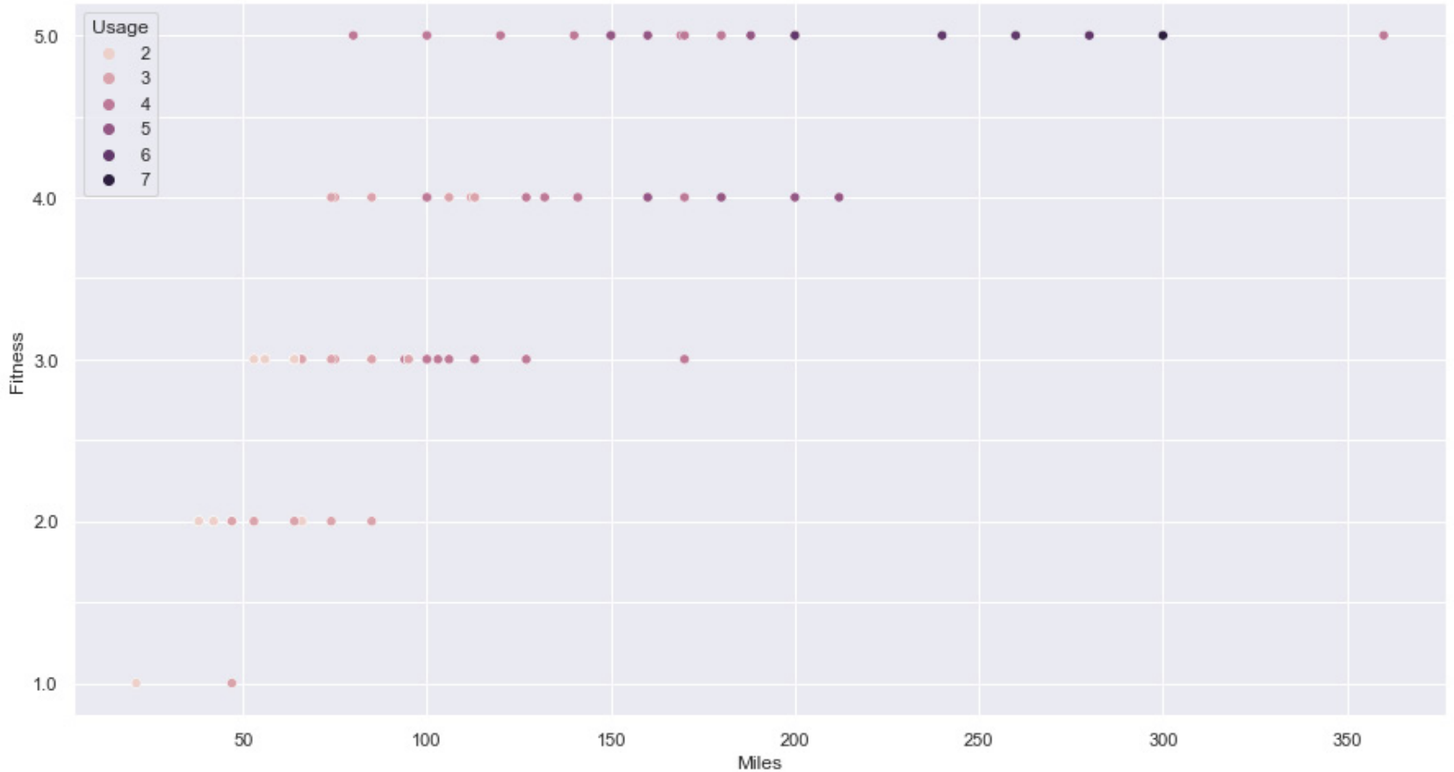


Segmentation by Marital Status
- 59% of customers are partnered
- 41% of customers are single

Segmentation by Gender
- 58% of customers are male
- 42% of customers are female

Male and female customers are married in similar proportions – neither gender tends to be married more than the other.

Segmentation by Marital Status & Gender
- 34% Partnered Males
- 26% Partnered Females
- 24% Single Males
- 17% Single Females

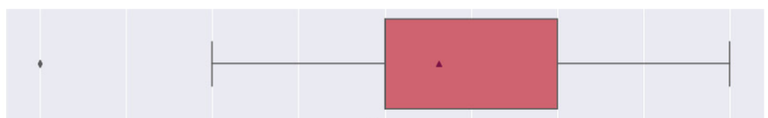


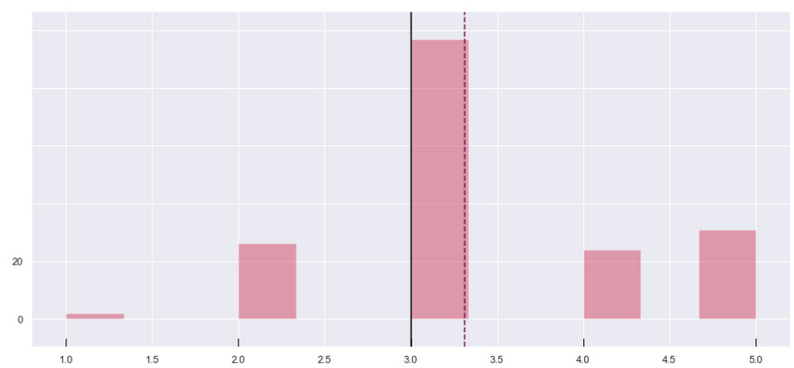
EXPLORATORY DATA ANALYSIS:

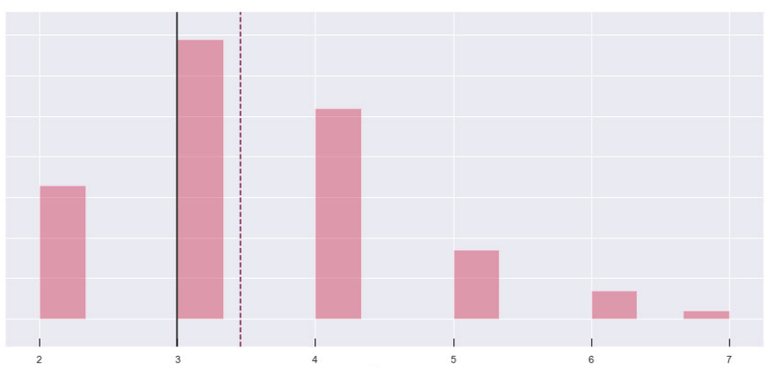
FITNESS, USAGE, & MILES

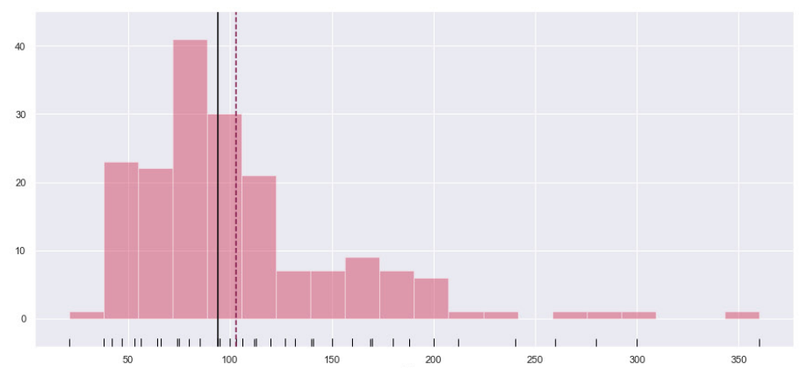
The variables for fitness score, usage, and miles have very high correlation levels.

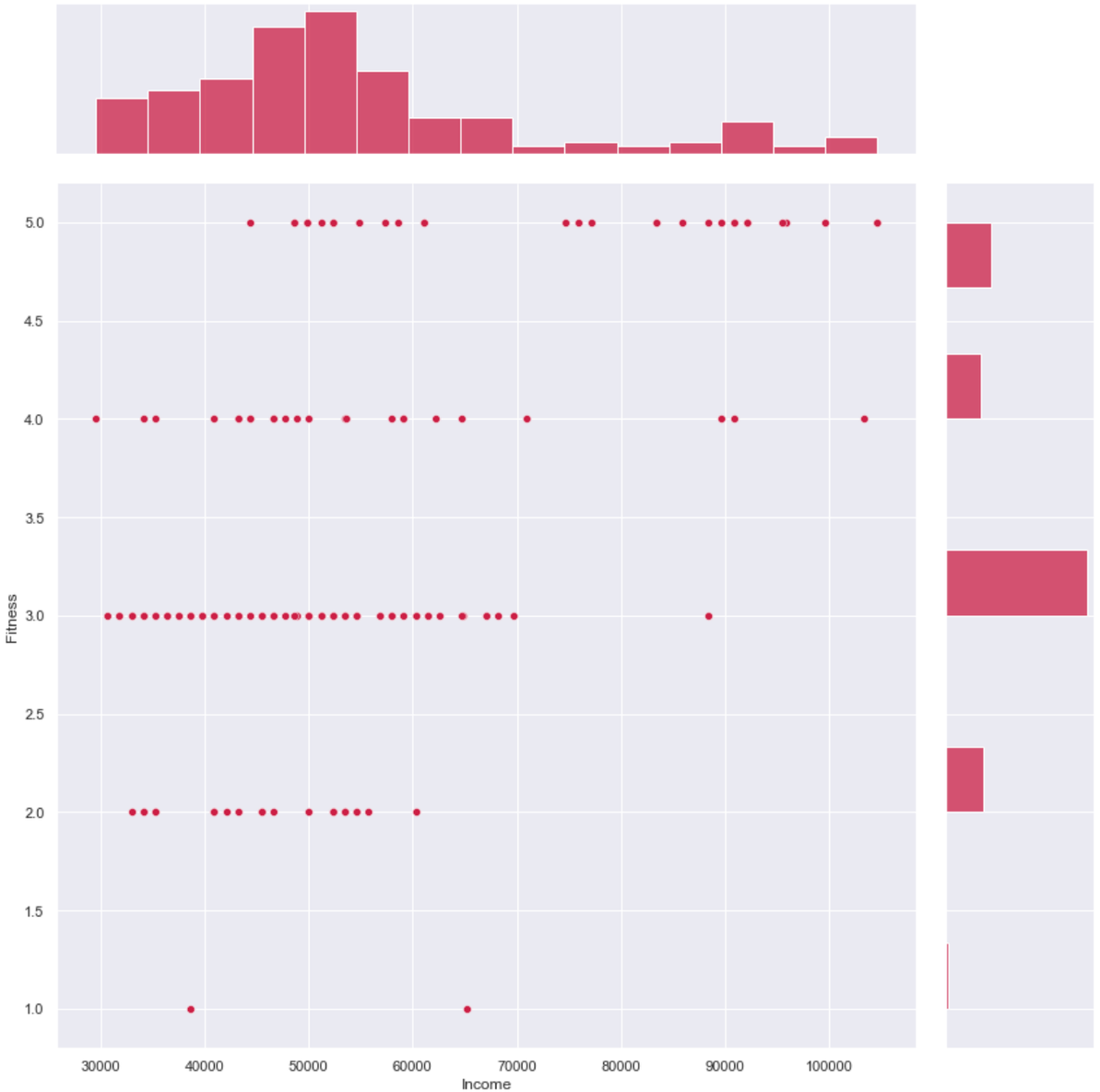
- 16%: fitness levels 1 and 2
    - customers expect to run less than 90 miles
    - expected usage is 2 or 3 times per week
- 54%: fitness level 3
    - most expect to run between 50 and 130 miles
    - expected usage is 2 to 4 times per week
- 30%: fitness levels 4 and 5
    - most expect to run between 75 and 300 miles
    - expected usage is 3 to 7 times per week











EXPLORATORY DATA ANALYSIS:
FITNESS & INCOME

Fitness rating and income have high correlation levels.

Observations:

- Fitness levels 1 and 2
  - almost all incomes in this group are between 30k and 60k

- Fitness levels 3 and 4
  - almost all have incomes between 30k and 70k
  - 4 outliers in this group have higher incomes

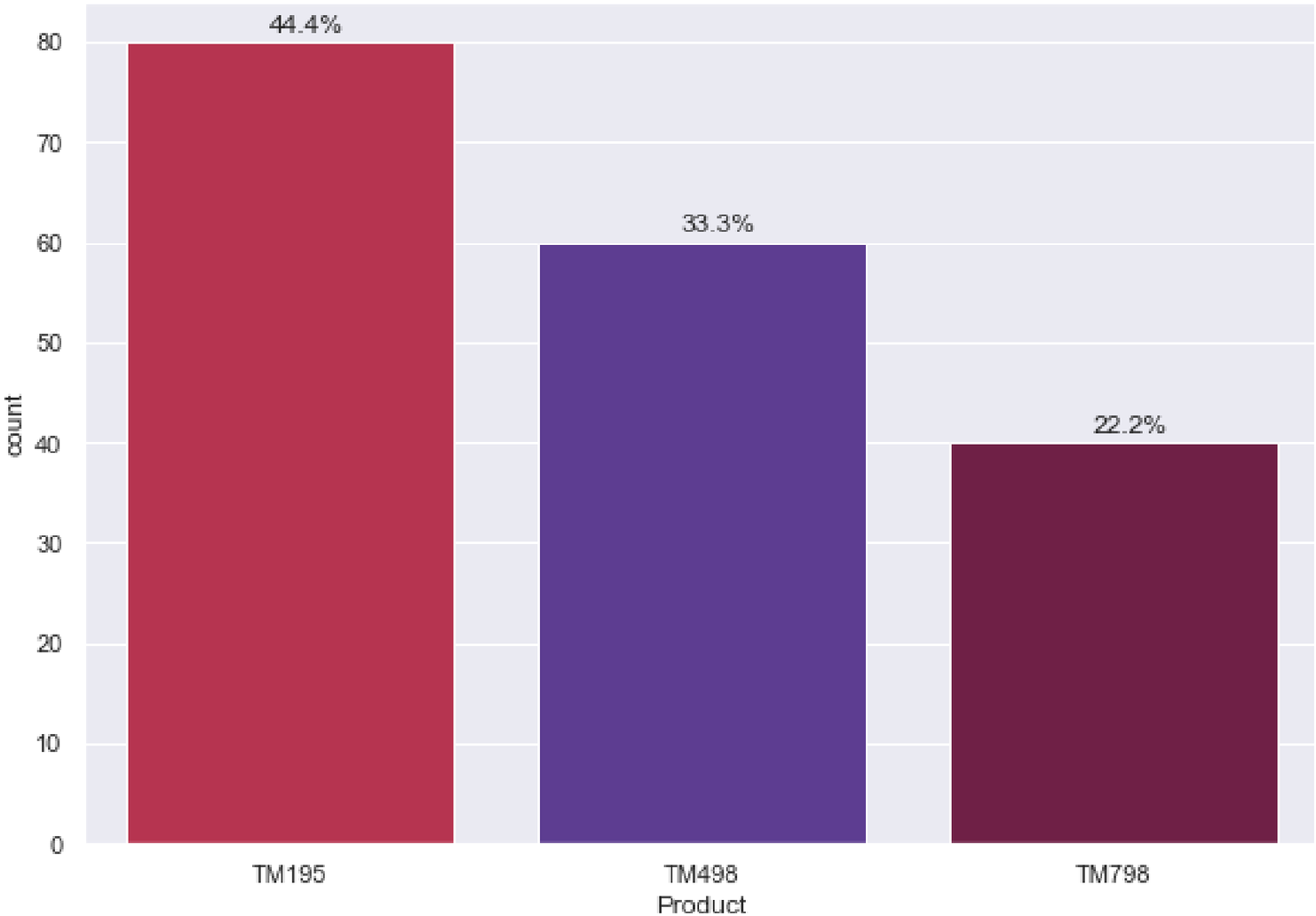
- Fitness level 5
  - the lowest income in this group is about 43k
  - incomes in this group are much higher than the other four levels
  - this group includes the vast majority of incomes above 75k

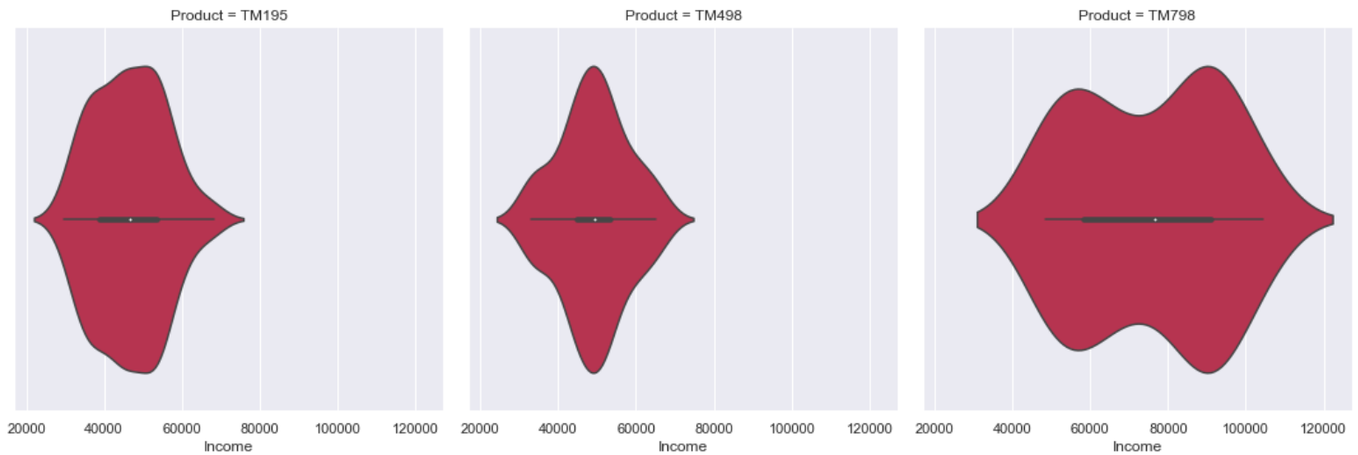
EXPLORATORY DATA ANALYSIS: PRODUCT

The following pages will explore and identify the differences between the customers of each treadmill product.

From the graph on the left we can see that the three products account for differing percentages of treadmill sales.

- ***TM195***: the best-selling product line
- ***TM498***: the mid-selling product
- ***TM798***: the lowest-selling product



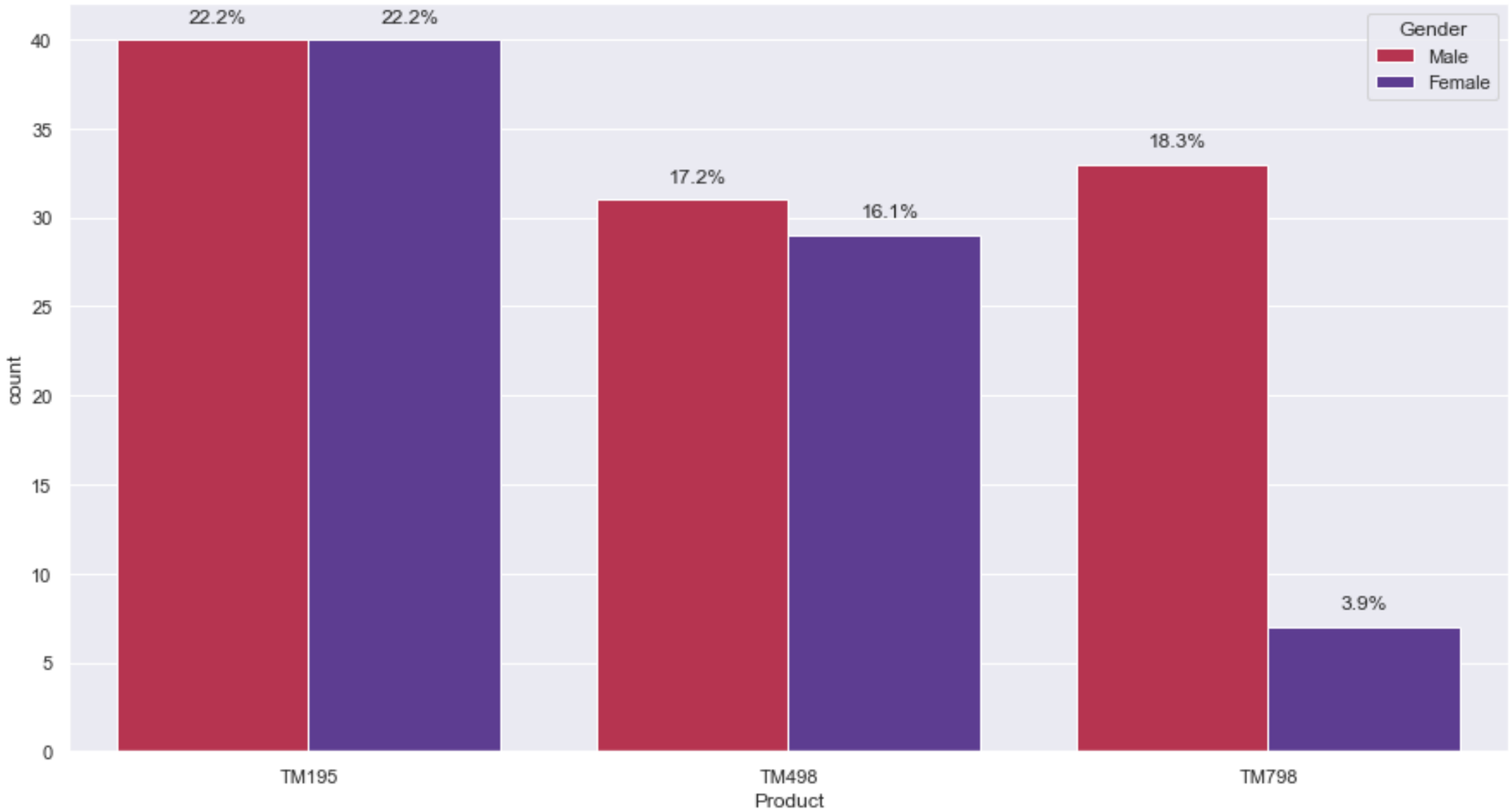


EXPLORATORY DATA ANALYSIS:
PRODUCT & INCOME

From these graphs we can see that the TM798 is the preferred model for those customers with higher income and higher education levels. We do not have data on the price points or feature sets of the different models, but it is quite possible that the TM798 is more expensive and has a larger feature set than the other treadmills

There is not a large income or education differentiator between TM195 and TM498. The TM195 does have a larger cluster of customers below 45000, but the TM498 also has a number of customers in this area as well.

EXPLORATORY DATA ANALYSIS:
PRODUCT & GENDER



- ***TM195***: equal sales to each gender
- ***TM498***: slightly male skewed
- ***TM798***: heavily male skewed

EXPLORATORY DATA ANALYSIS:
PRODUCT, FITNESS, MILES, & USAGE

TM195

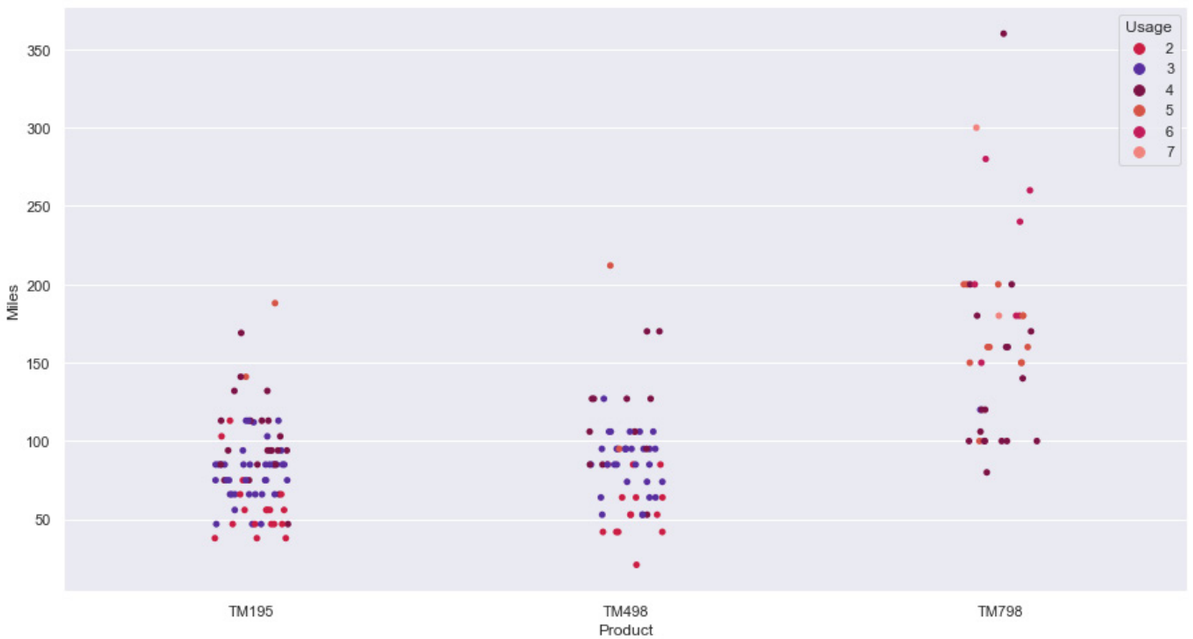
• high appeal to customers with fitness level 3

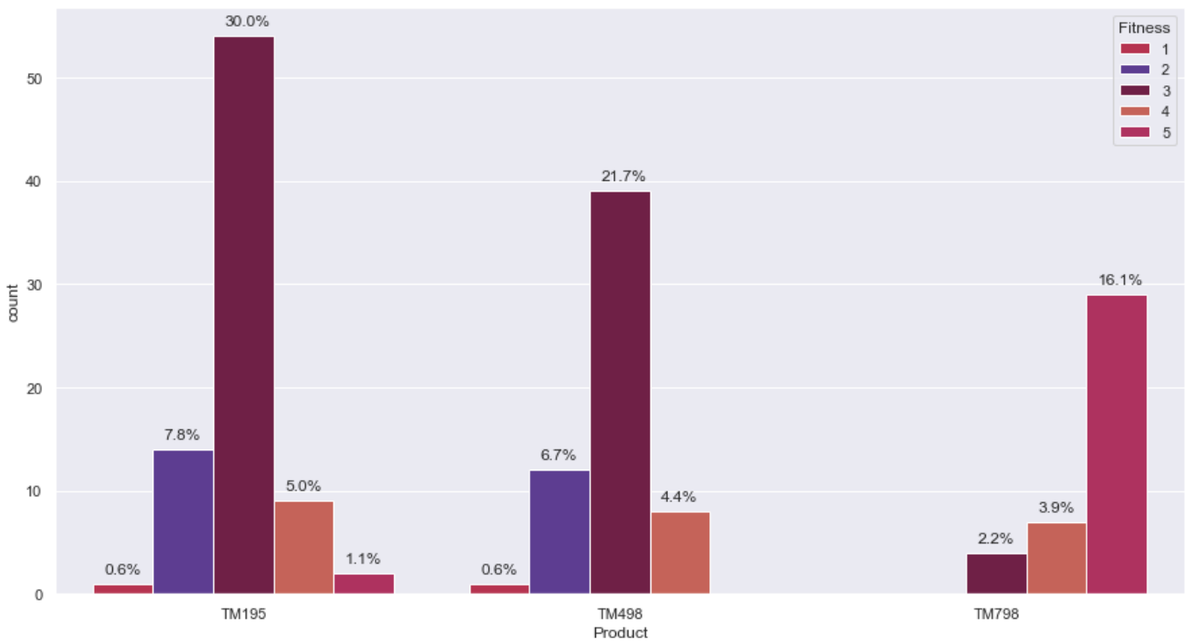
TM498

• very similar to TM195 customers, but not as widely used

TM798

• appeals to higher fitness levels than the other two models, especially those customers with a fitness level of 5 and those planning on running more miles





CONCLUSION

The analysis has brought forth the following conclusions:

1. Cardio Good Fitness products are most popular to customers between the ages of 23 and 33 years

2. The average CGF customer likely has about 16 years of education and an income between 50K and 55K

3. The average CGF customer is more likely to be male and more likely to be married.

4. The largest percentage of CGF customers (54%) give themselves a fitness rating of 3 on the 5-point scale, equating to an average fitness level. These same customers plan on using the treadmill 2 to 4 times per week, and expect to run between 50 and 130 miles,

5. Customers with a fitness level 5 (very fit) have a significantly higher income on average than the other fitness levels.

6. TM195 is the best selling product line, accounting for 44% of GCF treadmill sales. The TM195 customer is equally likely to be male or female, has an income between 30000 and 70000, and is the favorite of those with a fitness level of 3. TM195 customers do not expect to run more than 200 miles.

7. TM498 accounts for 33% of GCF treadmill sales. The TM498 customer is slightly more likely to be male than female, and has an income between 30000 and 70000. TM498 customers generally do not expect to run more than 200 miles.

8. While TM195 and TM498 customers are very similar to each other, TM798 customers are quite different. TM798 accounts for 22% of sales. TM798 customers skew very largely male, have expected incomes between 50000 and 105000, and have higher levels of education. The TM798 especially appeals to customers who are more physically fit: those who rate themselves as a fitness level 5, plan on using the treadmill every day, and/or plan on running more miles compared to other customers.

RECOMMENDATIONS

The following recommendations were created based on the analysis:

1. CGF treadmills are popular with male users, especially the TM798 product. It is worth investigating if CGF could market to a female demographic to grow the business.

2., The TM195 and TM498 customers largely overlap, and are not differentiated. The TM195 is the more popular of the two models, with a large appeal to average fitness levels. Perhaps in the next iteration of product development, the TM498 could be redesigned to appeal to a different segment of the market. Maybe the TM498 could be better positioned as a product between the TM195 and TM798 – one that would appeal to those of a fitness level 4.

3. It is worth bringing attention to the fact that there were no customers over the age of 50. It would be worth looking into why CGF does not appeal to this demographic.

4. The TM195 is the core product in the treadmill lineup. This should be noted so as to keep up with in-store supply for the product, but also so that the company is aware that the TM195 is likely the product customers think of when they imagine the Cardio Good Fitness brand.

5. To build a more detailed picture of the treadmill lineup, we would need to procure more data on the treadmills themselves. This would include price points and feature sets.

6. To create a more robust picture of how Cardio Good Fitness customers compare to rival companies, we would need to procure similar data sets about the treadmills of other brands. This could help illuminate CGF's strengths and weaknesses.

THANK YOU

EXPLORATORY DATA ANALYSIS

by **JAKE EIDE**