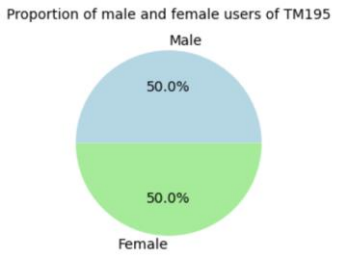
Cardio-Good Fitness – Problem Definition

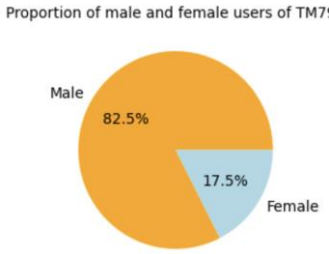
The end goal of this project is to build a product portfolio of various Treadmills offered by Cardio-Good fitness Inc. TM195, TM498, TM798 respectively. We are provided with various features of the customer data including their age, fitness level, expected number of miles they plan to achieve etc. As a first step, it is essential to explore the various dimensions of the data at hand to gain insights into how the customer preferences to purchase a product vary based on the individual characteristics.

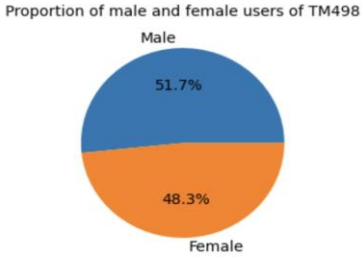
Exploratory Data Analysis

The data has been thoroughly enquired to handle missing values and remove any redundancies present. The following bar plots, and the pie charts have been drawn to visualize as to how each parameter varies with the product range.

The products TM195 and TM498 accounts for an equal share of male and female consumers. There is a significant difference in proportions with respect to the product TM798.







To better analyze Income, Education, Miles and Usage characteristics, these features have been divided into slabs and then compared against each of the products respectively. The following bar plots depict the patterns as to how the features influence product purchases.

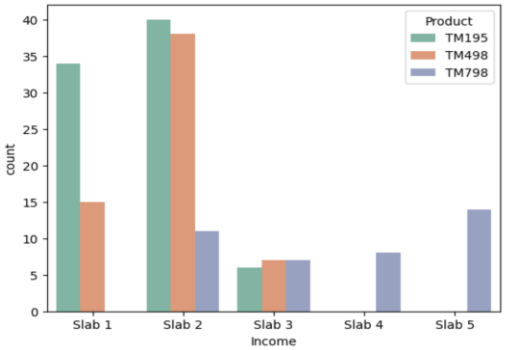


Fig: Income Vs Product type

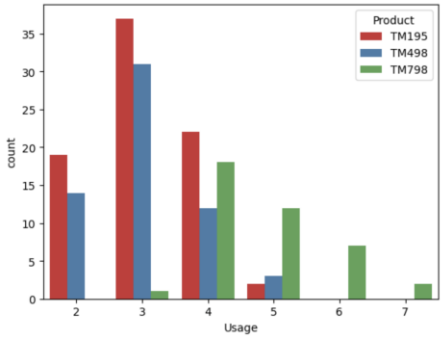


Fig: Usage Vs Product type

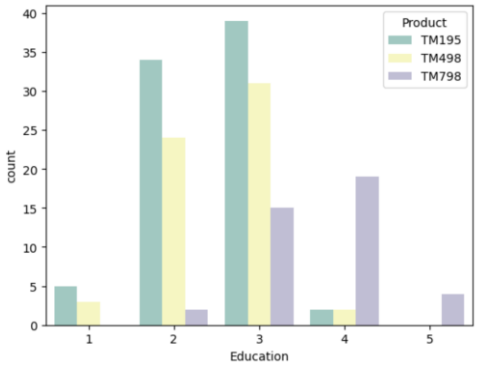


Fig: Education Level Vs Product type

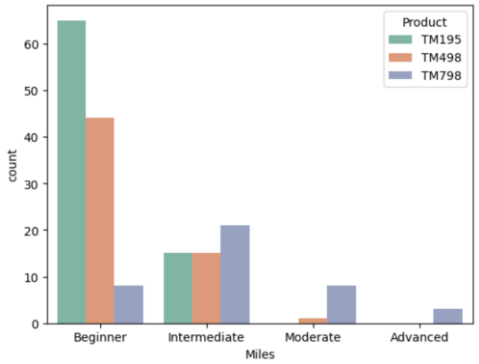


Fig: Miles Vs Product type

To summarize,

- People with low to mid-level income prefer to purchase TM195 and TM498 products.
- As the frequency of usage increases, customers tend to opt for the product TM798. It could be because of its build quality and high-duty performance.
- Education seems to play a significant role in choosing the products. Customers with entry to mid-level education seems to choose TM195 and TM498 products rather than TM798.
- People belonging to “Beginner” and “Intermediate” categories of fitness regime are more inclined towards the treadmill TM195.

Methodology used to construct confidence intervals

Step 1: As the respective data for the products is limited, it is essential to pull out samples from the existing data using the BOOTSTRAPPING technique. It involves sampling the data with replacement. The sample size in consideration is 50. It is because the central limit theorem holds true when the sample size is at least 30.

Step 2: Once the samples are drawn from the data, those are iterated over to obtain the respective sample parameters. It is expected that the distribution of sample means follows Normal distribution.

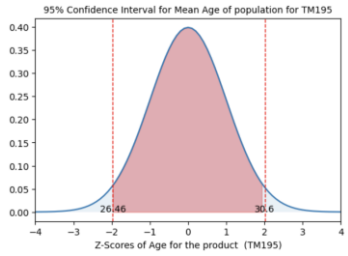
Step 3: Since the sample size is > 30, the Z-distribution can be employed to lower and upper confidence bounds.

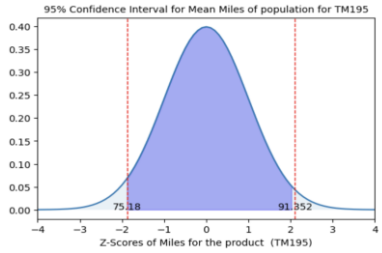
Step 4: Calculate the cumulative frequencies of the bounds.

Step 5: Plot the resulting limits and the interval.

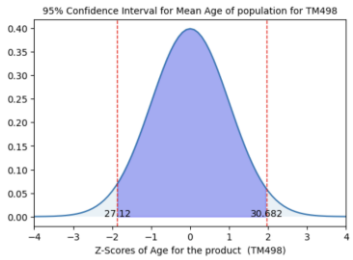
Confidence Intervals: Since Age and Miles seems to have a huge impact on the product type, we are building the respective confidence intervals for their population means.

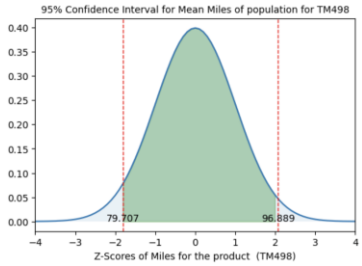
Confidence Intervals for the Product TM195:



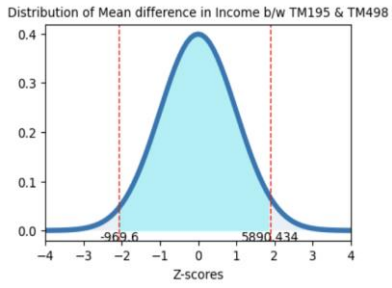


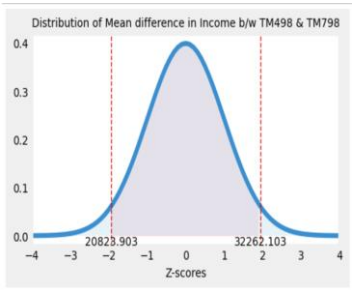
Confidence Intervals for the Product TM498:

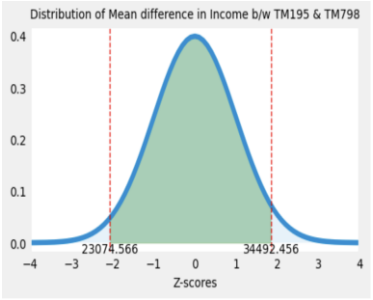




Confidence Intervals for the Mean differences of products







Note: The shaded area represents the confidence interval. And the dotted lines indicate the respective lower and upper bounds. The corresponding limits in the original domain are marked respectively.

How to interpret the confidence intervals?

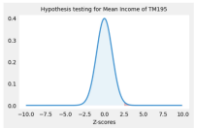
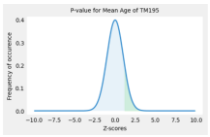
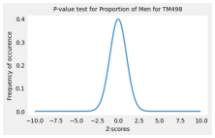
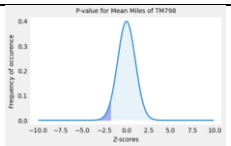
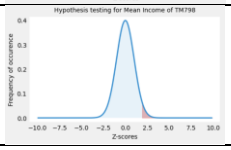
Let us consider the first snippet from above. It speaks of the 95% confidence interval for Mean age of the TM195 product population. The limits are **(26.46, 30.60)** This indicates that 95% of the times the mean age of TM195 population falls in between the above interval.

Summary of Confidence Intervals:

Product Type	Confidence Interval for Age	Confidence Interval for Miles
TM195	(26.46, 30.6)	(75.18, 91.35)
TM498	(27.32, 30.68)	(79.71, 96.89)
TM798	(27.22, 31.14)	(152.43, 184.93)

Hypothesis Testing:

Hypothesis testing is a technique which is used to validate our presumptions about the population based on a performance metric. We are going to use the P-value with a threshold of 0.05 to decide whether to accept or reject the Null Hypothesis. The following hypothesis tests have been performed and the outcomes are tabulated below.

Hypothesis	Null & Alter hypo.	P-value and Decision	Shaded Area
Mean income of population purchasing TM195 exceeds 28,000.	H_0: μ_{income} < 28,000 H_1: μ_{income} >= 28,000	P-value = 0.569 % Null hypothesis is rejected. Mean Income exceeds 25,000.	
Mean age of population purchasing TM195 exceeds 20.	H_0: μ_{age} < 20 H_1: μ_{age} >= 20	P-value = 12.275 % We do not reject the null hypothesis. Null hypothesis is still plausible.	
Proportion of male users for TM498 exceeds 0.3	H_0: p_{male} <= 0.3 H_1: p_{male} > 0.3	P-value = 0.0 % Null hypothesis is rejected. P_{male} exceeds 0.3	
Average miles for TM498 lies within 250.	H_0: μ_{miles} >= 250 H_1: μ_{miles} < 250	P-value = 4.197% Null hypothesis is rejected. Average miles < 250.	
Mean income for TM798 exceeds 60,000.	H_0: μ_{income} >= 60000 H_1: μ_{income} < 60000	P-value = 2.864 % Null hypothesis is rejected. Mean income exceeds 60000.	

Summary table: Below table summarizes the product portfolio for TM195, TM498, and TM798 respectively.

Product	Age Range	Income level	Avg Miles	Education (years)	Usage (per week)
TM195	26 - 30	< 30000	75 - 91	14 - 16	2 - 4
TM498	28 - 31	30000 - 50000	79 - 96	14 - 16	2 - 4
TM798	27 - 32	> 50000	152 - 185	16 - 18	4 - 7