Descriptive Analysis of Cardio Good Fitness Treadmill Product Lines

Sushrit Kafle

2024-09-23

Introduction

Cardio Good Fitness, a leading retailer in the fitness equipment industry, aims to better understand its customer base to improve its product marketing strategy and increase sales.The market research team is assigned the task to identify the profile of the typical customer for each treadmill product offered by Cardio Good Fitness. The market research team decides to investigate whether there are differences across the product lines with respect to customer characteristics. The team decides to collect data on individuals who purchased a treadmill at a Cardio Good Fitness retail store during the prior three months.

Dataset Description

The team identifies the following customer variables to study:

- **Product** - the model no. of the treadmill
- **Age** - in no of years, of the customer
- **Gender** - of the customer
- **Education** - in no. of years, of the customer
- **Marital Status** - of the customer
- **Usage** - Avg. # times the customer wants to use the treadmill every week
- **Fitness** - Self-rated fitness score of the customer (5 - very fit, 1 - very unfit)
- **Income** - of the customer
- **Miles** - expected to run.

The data are stored in the CardioGoodFitness.csv file.

Methodology

The following steps were undertaken to analyze the data and draw insights:

- **Data Preprocessing**: Identification of missing value, classification of variables into categorical and numerical for analysis.
- **Data Summary**: A preliminary analysis to understand the central tendencies and spread of numerical variables and frequency counts for categorical variables.
- **Univariate, Bivariate, Multivariate Analysis and Tests**: Performed appropriate tests to see if gender affects the model purchased. Performed tests to see if there any difference in mean age among customers that purchased different models.

- **Visualization**: Histograms and density plots were created to assess the distribution of numerical variables, bar charts and boxplots to compare categorical and numerical variables and multivariate plots to explore the combined effect of several customer attributes on treadmill purchases.

Results and Analysis

In this section, the findings from Cardio Good Fitness dataset are discussed. We ask and answer important questions about the dataset to further understand the data.

About The Dataset:

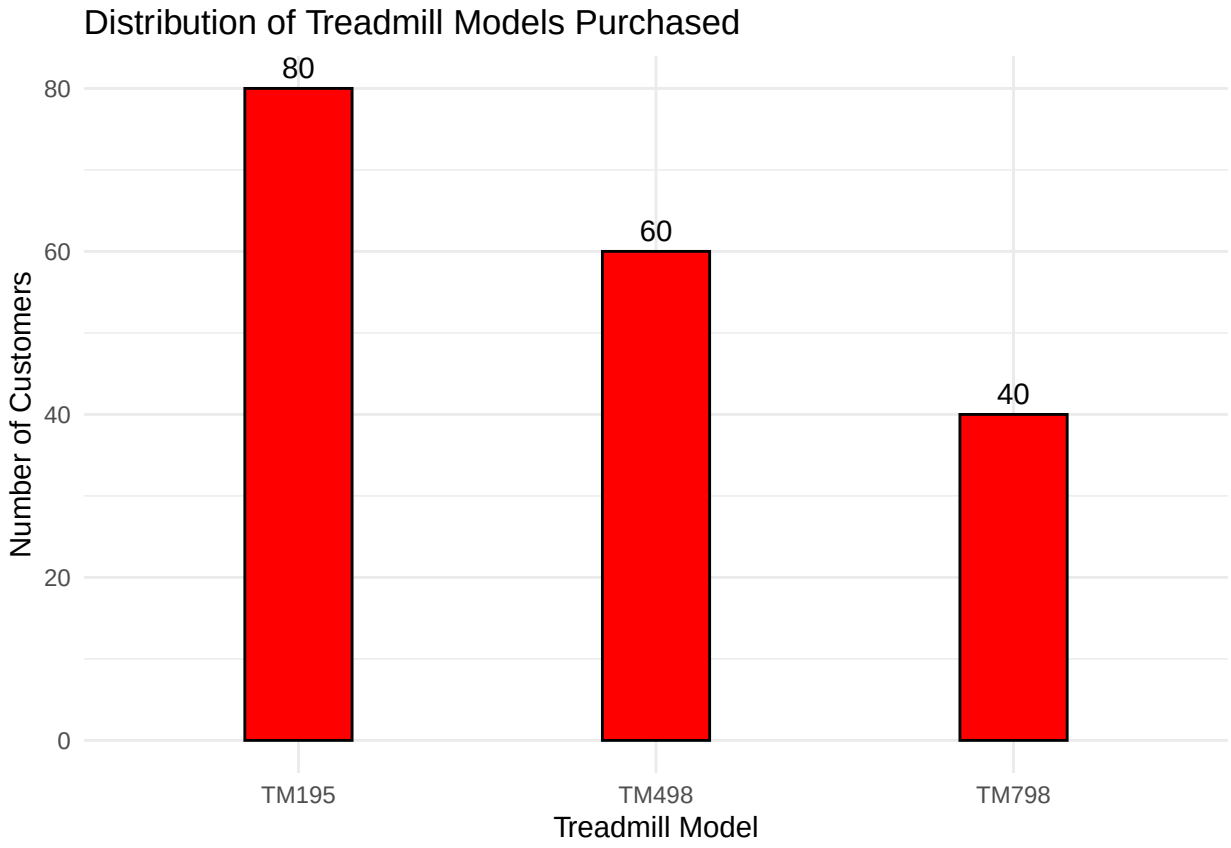
- The dataset has the shape of (180,9),i.e. 180 observations and 9 variables.
- The types of variables include:
  - **Numerical Variable**: Age(Discrete), Education(Discrete), Usage(Discrete), Fitness(Ordinal), Income(Continuous), Miles(Discrete).
  - **Categorical Variable**: Product, Gender, Marital Status.
- There are no missing data in the table.

Table 1: First 5 Rows of CardioGoodFitness Data

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

Univariate Analysis

1. Product



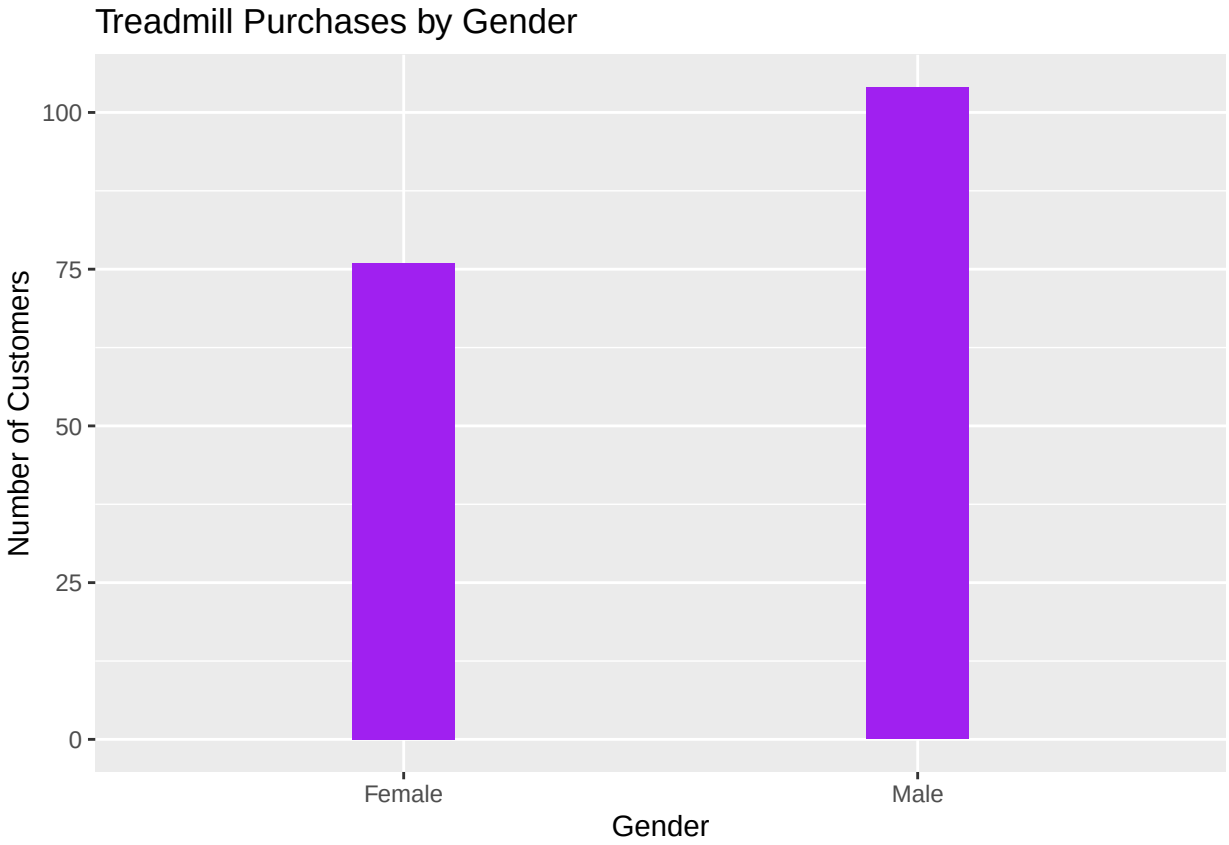
- How many models does store have?

->The store has 3 models: “TM195” “TM498” and “TM798”.

- Which is most sold Model?

-> TM195 is the most sold model.

2. Gender



- Are Male customers buying treadmill more than female customers?

-> Yes, male customers are buying more treadmills.

3. Age

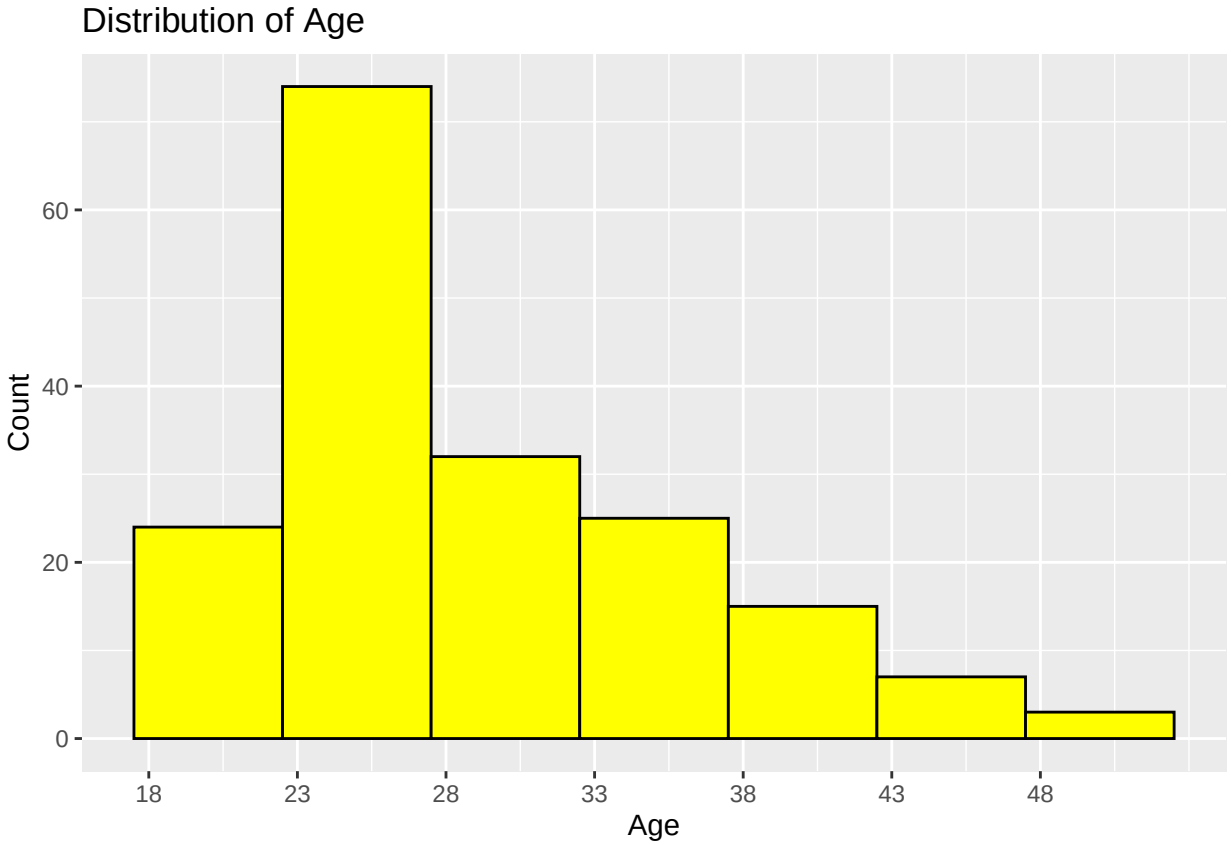
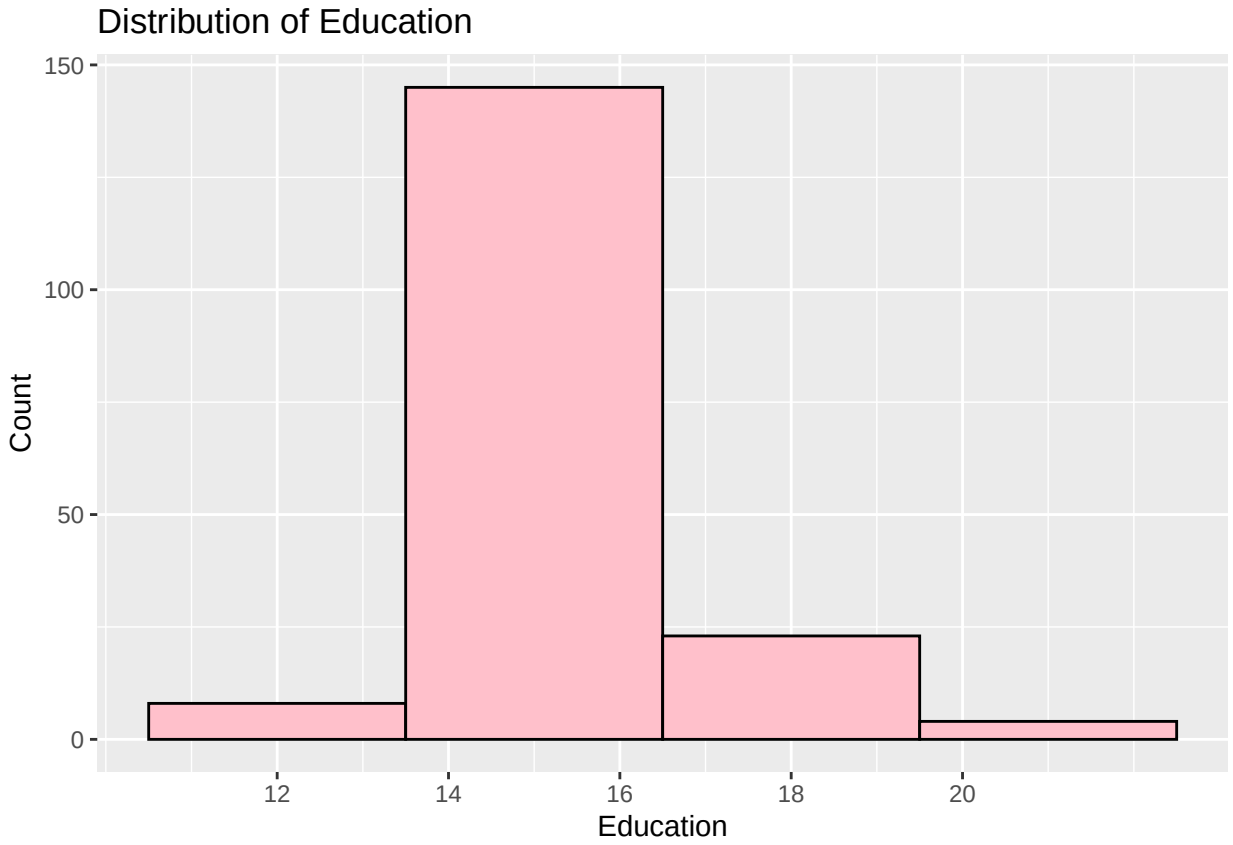


Table 2: Age Summary

Statistic	Value
Min.	18.00000
1st Qu.	24.00000
Median	26.00000
Mean	28.78889
3rd Qu.	33.00000
Max.	50.00000

The age range is (18-50). From the graph and table, we observe that people of age 23-28 are the most popular buyers. After age 28, the sales decrease with the decrease in age.

Education

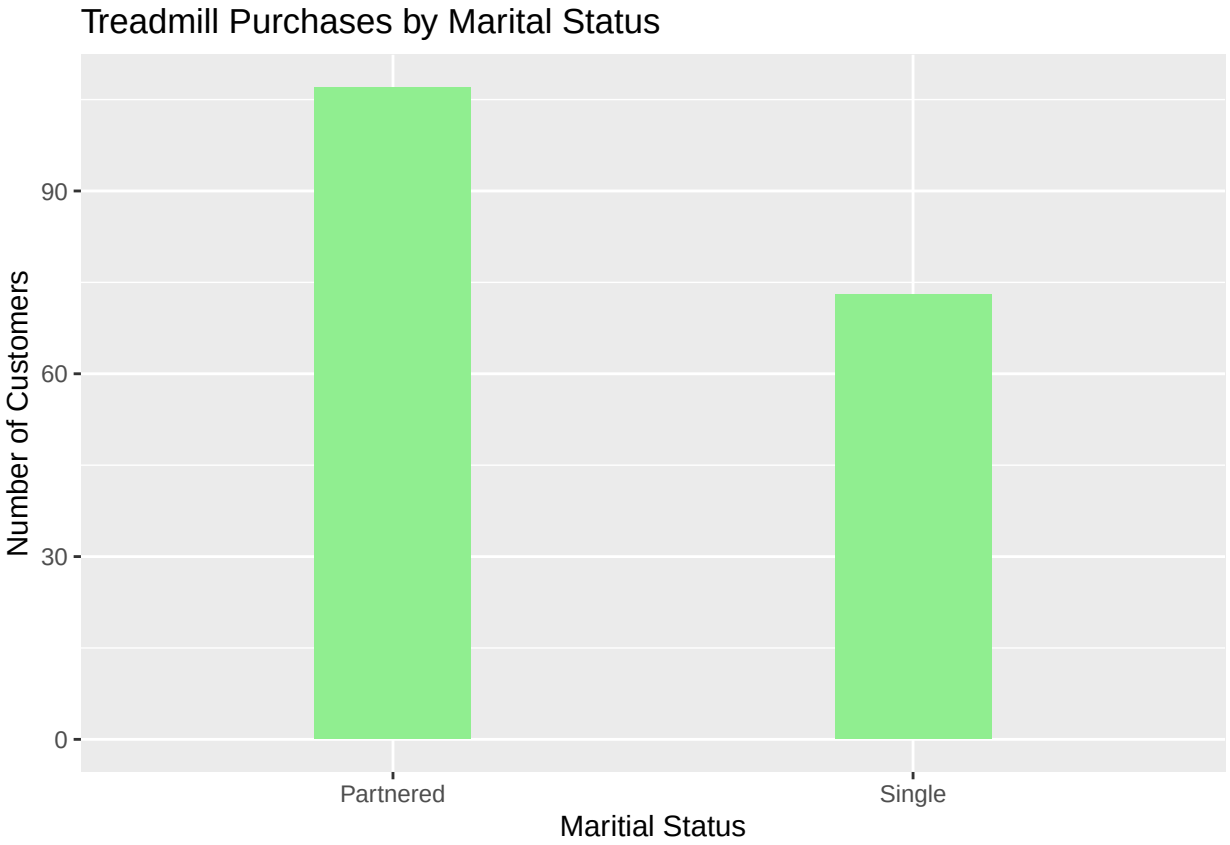



```
education_summary <- summary(data$Education)
education_summary_df <- data.frame(Statistic = names(education_summary), Value = as.numeric(education_su
kable(education_summary_df, caption = "Education Summary")
```

Table 3: Education Summary

Statistic	Value
Min.	12.00000
1st Qu.	14.00000
Median	16.00000
Mean	15.57222
3rd Qu.	16.00000
Max.	21.00000

Martial Status



- Are married customers buying Treadmill more than Single customers?

-> Yes, married people are buying more Treadmills than Single customers.

Usage

Fitness

Income

Miles

- What is the average Income, Age, Education of people buying treadmill?

-> The averages can be observed as below:

Table 4: Averages of Treadmill Buyers

Avg_Income	Avg_Age	Avg_Education
53719.58	28.78889	15.57222

- How many days and miles customer expect to run on treadmill?

-> According to the data,

Table 5: Average Days and Miles Expected on Treadmill

Avg_Usage	Avg_Miles
3.455556	103.1944

An average usage would be 3 times a week with 103.19 miles.

- What is the average self-rated fitness of customers buying treadmill?

-> The average self-rate fitness is given as:

Table 6: Average self-rated fitness of customers buying treadmill

Avg_Rating
3.311111

- Is there any relation between Income and model?

-> Forming hypothesis:

Null: There is no relation between Income and Model of Treadmill.

Alternate: There is relation between Income and Model of Treadmill.

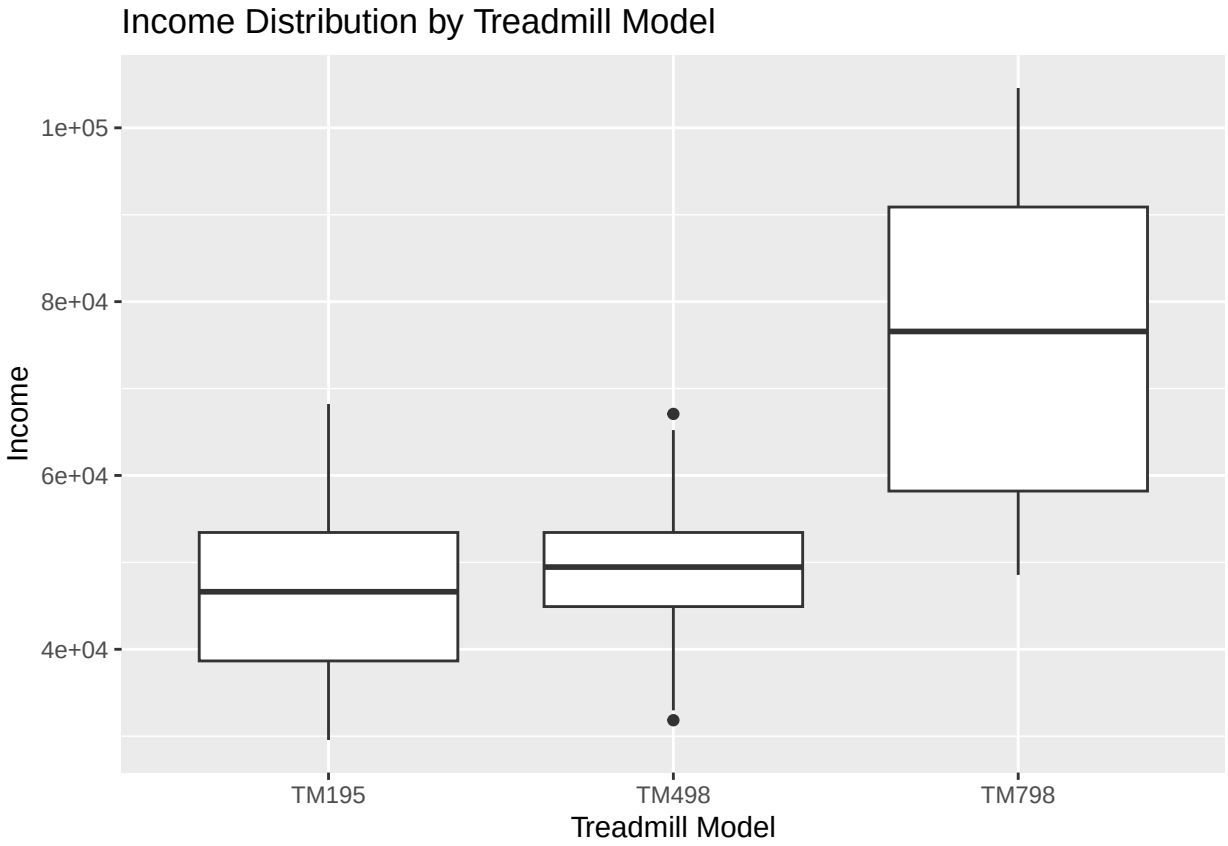
Performing ANOVA test:

Table 7: ANOVA Table: Income ~ Product

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Product	2	24490250199	12245125099	89.25904	0
Residuals	177	24281991523	137186393	NA	NA

As, F-value is very high and p-value < 0.05(typical significant value); we reject the null hypothesis. Hence, there is relation between income and model of treadmill.

We can observe the relation below:



TM798 is more popular among higher income while TM195 and TM498 are popular among low income.

- Is there any relation between Age and model?

-> Forming hypothesis:

Null: No relation between Age and model

Alternate: Relation between Age and model

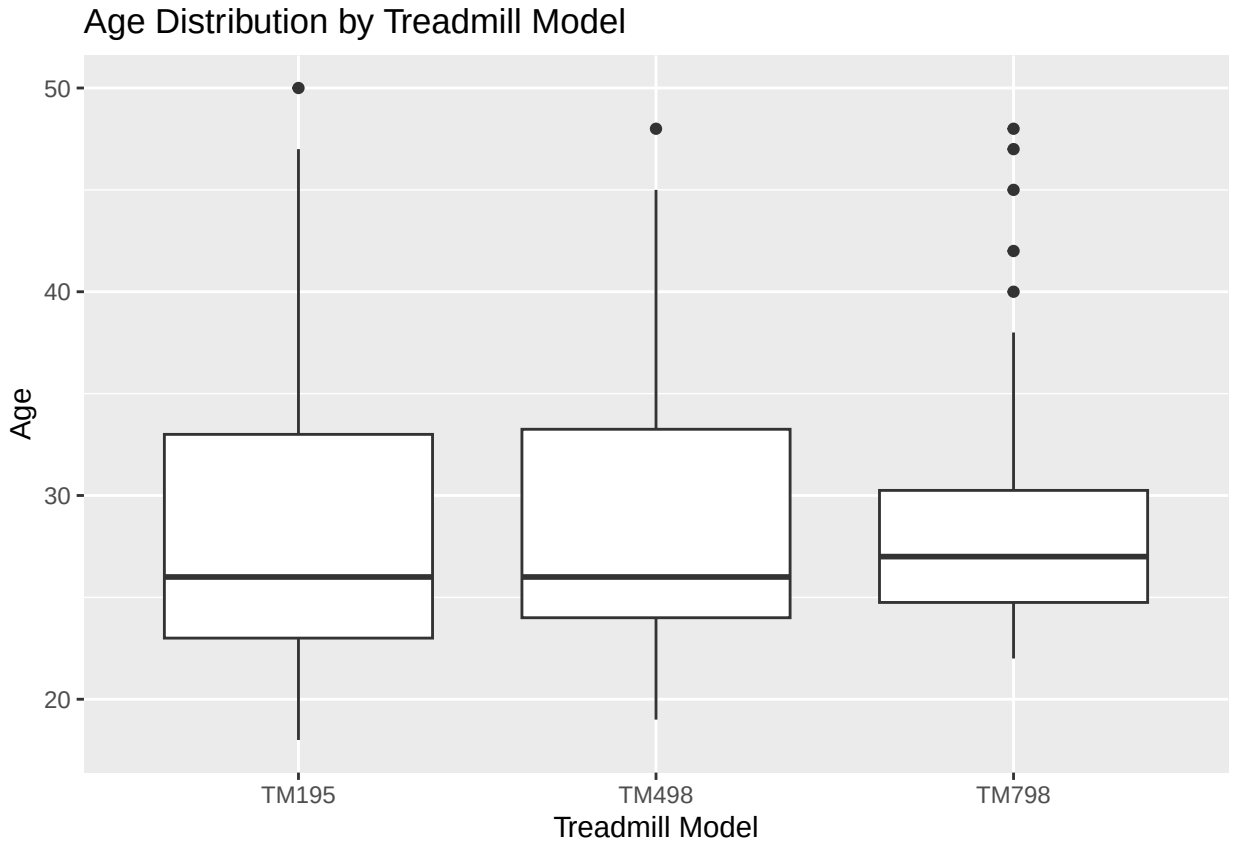
Performing ANOVA test:

Table 8: ANOVA Table: Age vs Product

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Product	2	9.177778	4.588889	0.0942179	0.9101301
Residuals	177	8620.800000	48.705085	NA	NA

As p-value > 0.05, we do not have enough evidence to reject the null hypothesis.

We can the relation with box plot below:



It can be observed that the medians for the different treadmill models are similar.

- Is there any relation between self-rated fitness and model?

-> Forming hypothesis; Null: No relation between self-rated fitness and model. Alternate: There is relation between self-rated fitness and model

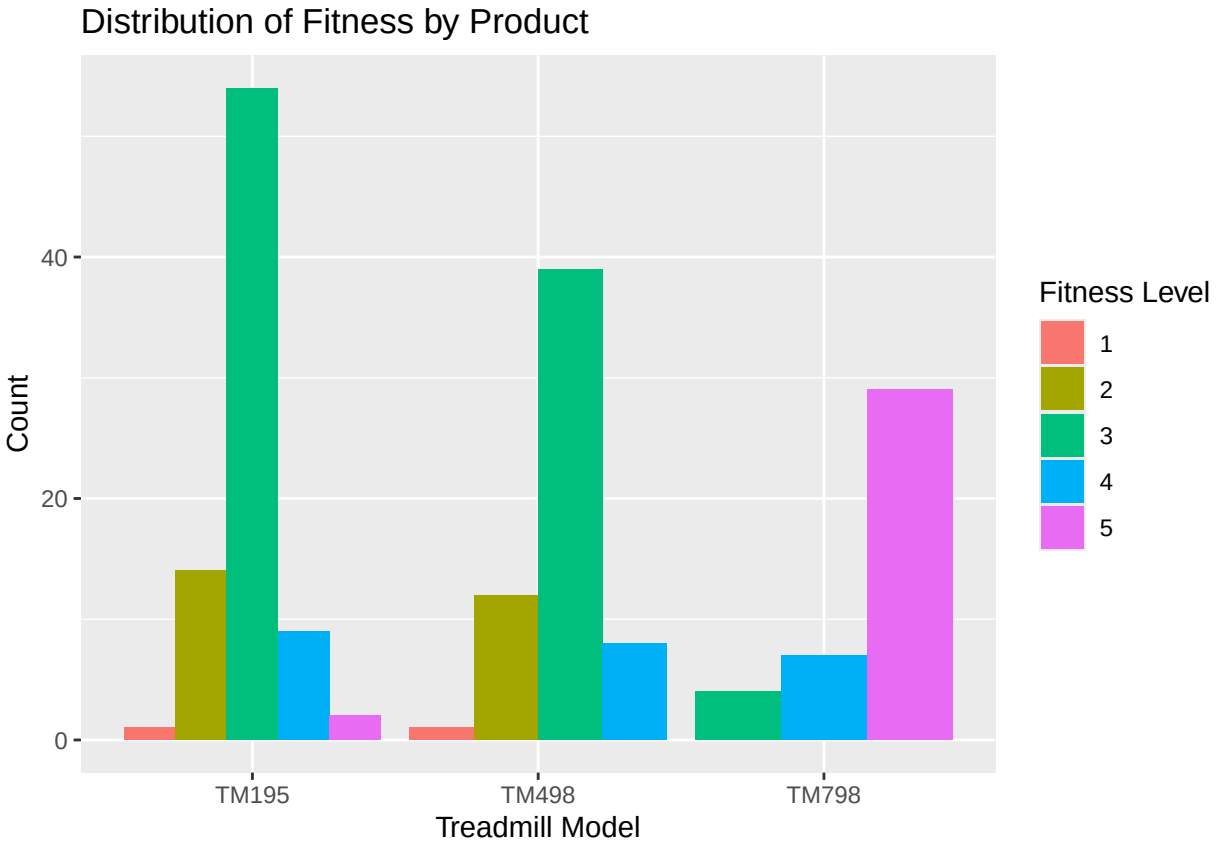
Performing Chi-Square Test for categorical variable:

Table 9: Chi-Square Test Results: Fitness vs Product

	Statistic	DF	P_value
X-squared	118.7768	8	0

For p-value close to 0, reject null hypothesis at 0.05 level. We have sufficient evidence to conclude self-rated fitness and model have a relation.

Observing with plot:



For the model TM195 and TM498, people with fitness level 3 are more attracted. However, the model TM798 is bought by more fitness enthusiasts with fitness level 5.

- Is there any relation between education and model?

-> Forming Hypothesis:

Null: There is no relation between education and model

Alternate: There is relation between education and model.

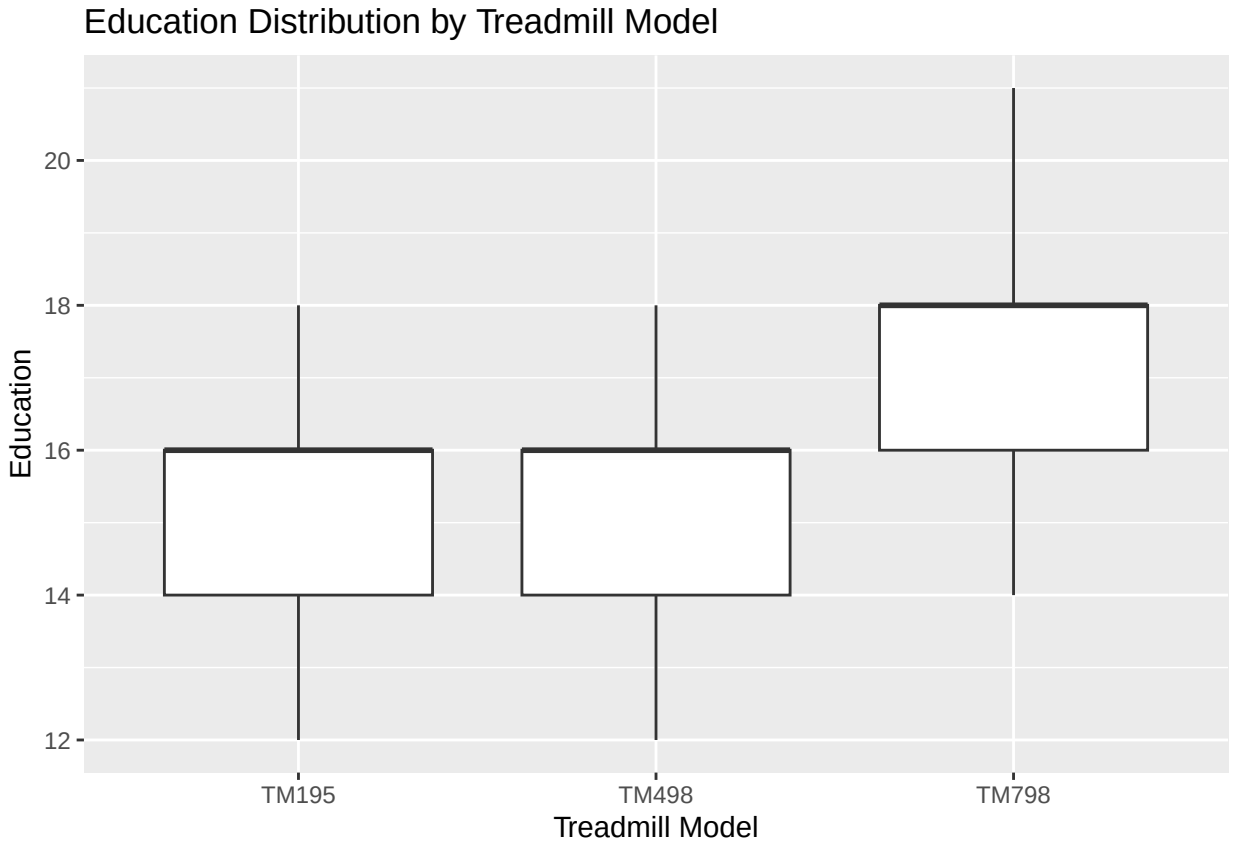
Performing Test:

Table 10: ANOVA Table: Education vs Product

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Product	2	158.2153	79.107639	45.19038	0
Residuals	177	309.8458	1.750541	NA	NA

As the p-value is close to 0, we reject the null hypothesis and conclude that we have enough evidence to support the alternative hypothesis, which says that there is relation between the education level and model of the treadmill.

We can further explore the relation with a plot:



It can be observed that people with more years of eduaction choose TM798 model.

- Does gender have any effect on model customer buy?

->Forming Hypothesis:

Null Hypothesis: Gender doesn't have any effect on model customer buy.

Alternate Hypothesis: Gender has effect on the model customer buy.

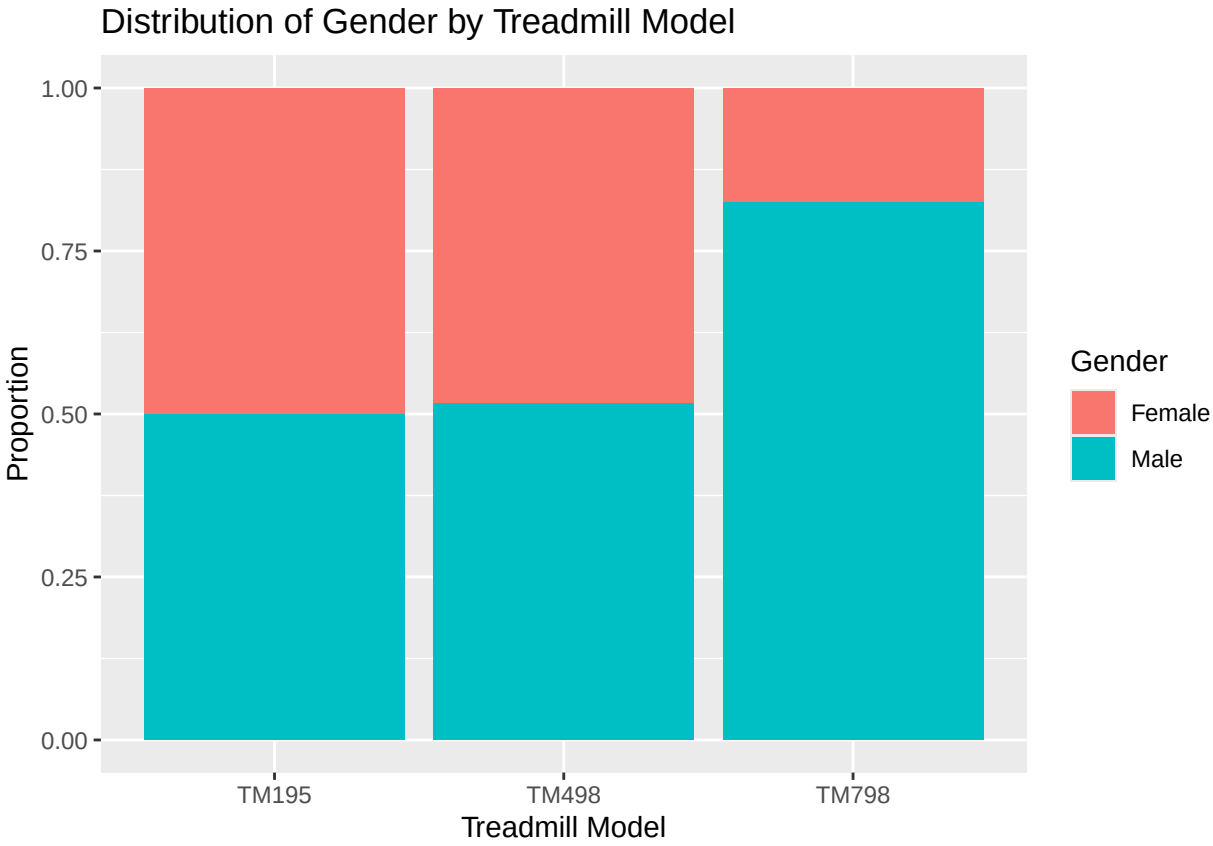
Performing Test:

Table 11: Chi-Square Test: Gender vs Treadmill Model

	Statistic	DF	P_value
X-squared	12.92384	2	0.0015618

As the p-value is close to 0, we reject the null hypothesis and conclude that we have enough evidence to support the alternative hypothesis, which says that there is relation between the gender and model of the treadmill.

Visualizing the relation:



There is equal proportion for male and female for TM195 and TM498 but male are more attracted to the TM798 model than females.

- Does Martial status have any effect model customer buy?

-> Forming hypothesis:

Null Hypothesis: Martial status has no effect on model

Alternate Hypothesis: Martial status has effect on the model customer choose.

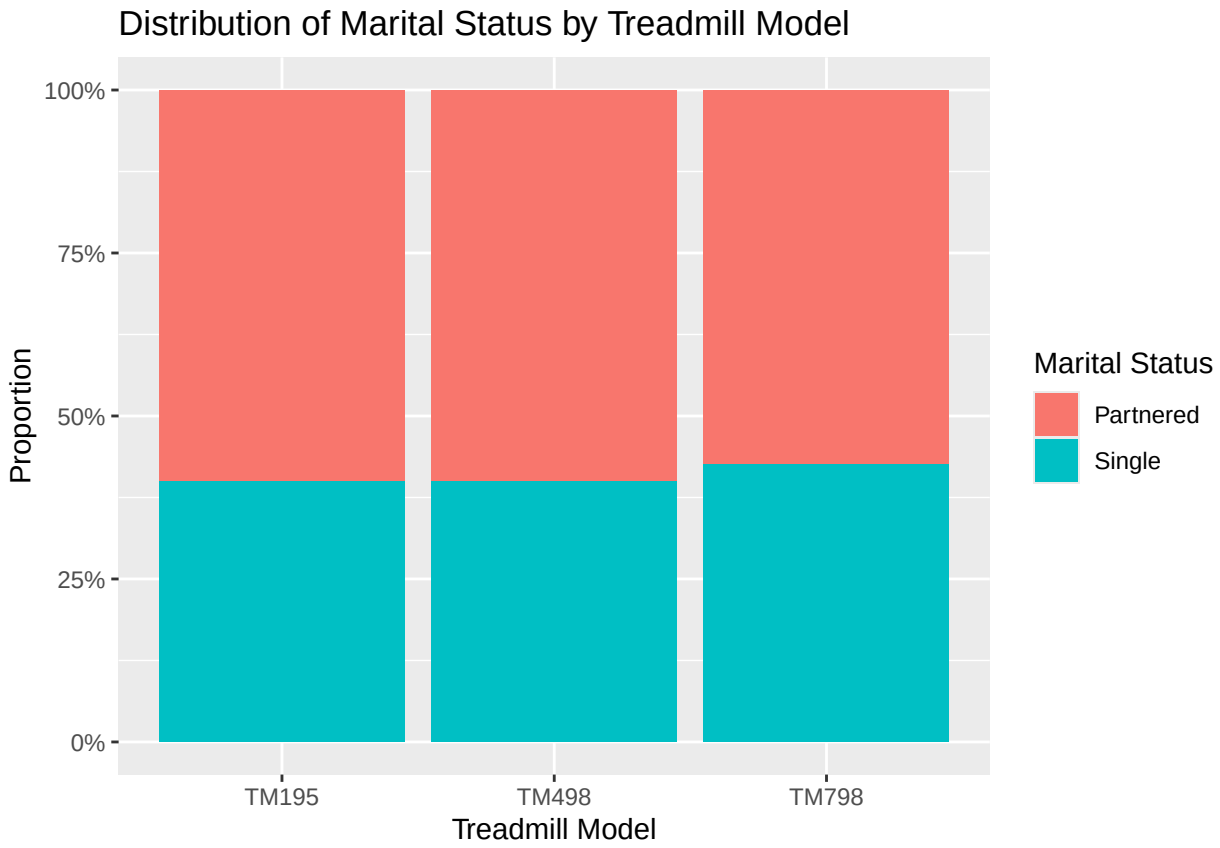
Performing Test:

Table 12: Chi-Square Test: Martial Status vs Treadmill Model

	Statistic	DF	P_value
X-squared	0.0806555	2	0.9604746

As, p-value (0.9604746) > 0.05, we fail to reject the null hypothesis, and conclude that there is not enough evidence to suggest that the martial status has effect on the model customer choose.

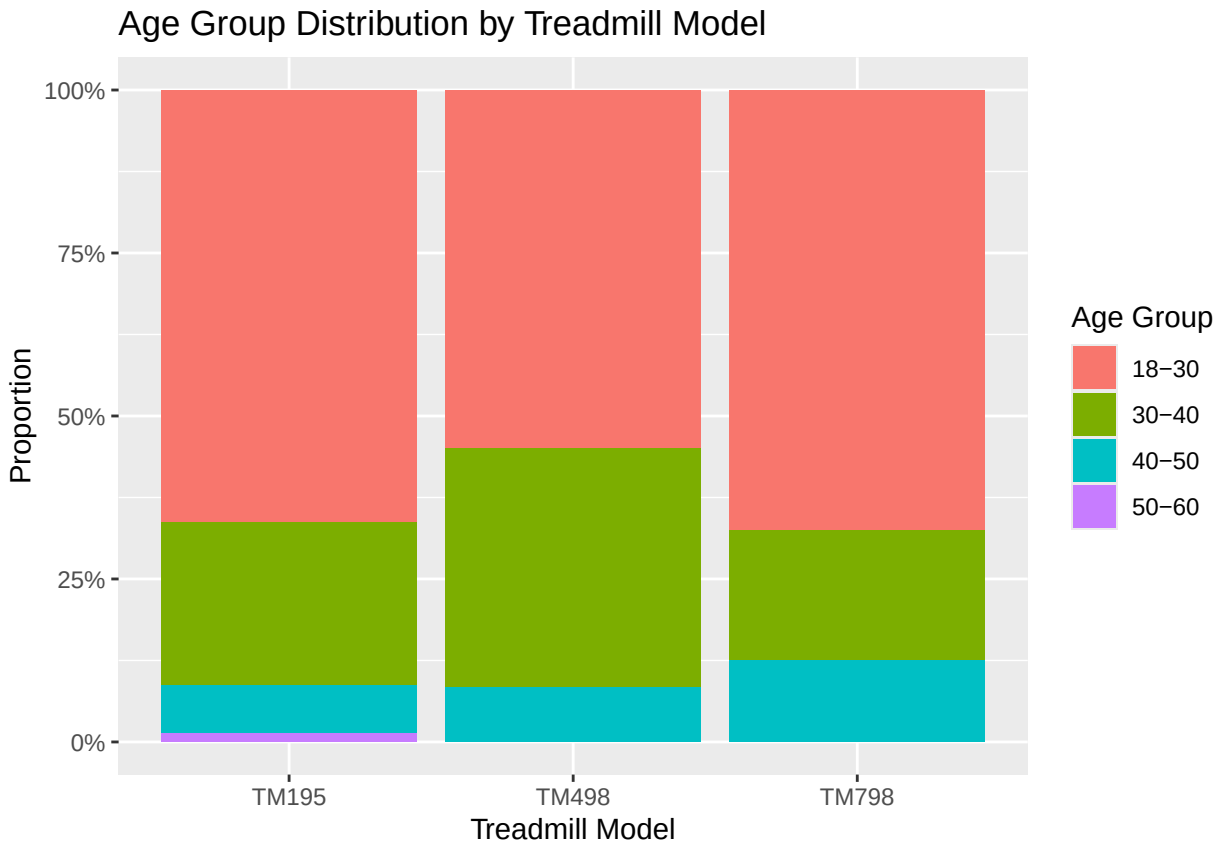
Visualizing:



As per the figure, equal proportion of martial status is found in all the models.

- Are different age groups buying different models?

-> Visualizing:



As per the figure above, people of age 18-30 are the majority buyers of all the models. While people of age 30-40 are slightly more inclined to buy model TM498.

- Relation between Age, Income and education and model bought?

->Visualizing:

From the above figure, it can be seen that people with higher income and in their middle ages tend to choose TM799 while the other two models have broader distribution over the income and ages.

It can also be noticed that people of ages below 30 tend to buy TM798 model even with slightly below mean income.