

Final Project

Cardio Good Fitness Report

Rolando Ng

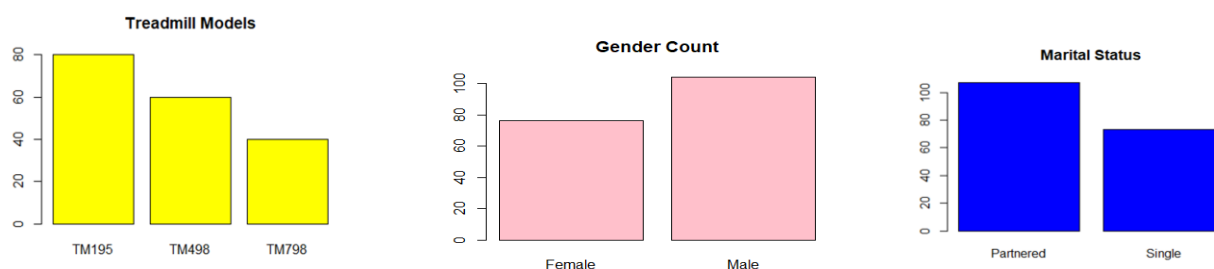
DSCI 30100 Data Analysis with R

Introduction

The Cardio Good Fitness dataset contains information about customers who bought treadmills from a fitness equipment company. Each row represents a single consumer, and the columns show the treadmill model they bought as well as their personal details such as age, gender, income, education, and level of fitness. The dataset helps us understand the kinds of consumers that purchase each model of treadmill and how variables like income, level of fitness, and usage patterns may affect their choices.

The goal of this analysis is to discover the characteristics of consumers who purchase various treadmill models and recognize trends in their demographics, income, exercise routines, and usage. In order to help Cardio Good Fitness make better marketing and product choices.

Categorical Data



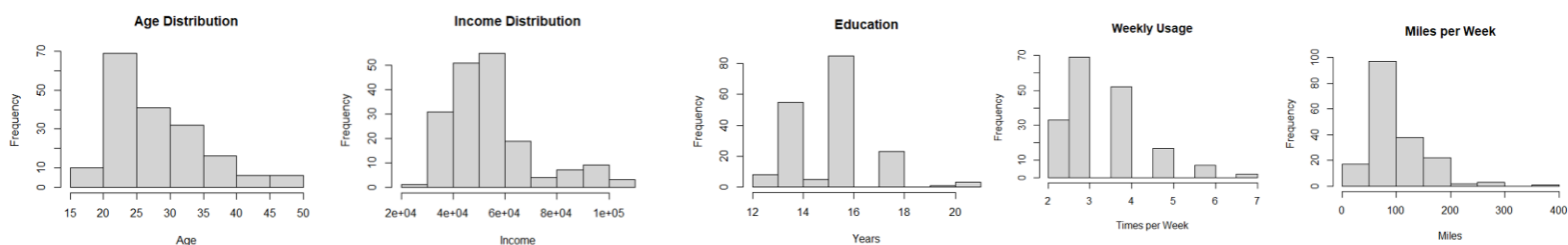
In first, the barplots show that TM195 is the most popular treadmill model, followed by the TM498, while TM798 is the least popular, indicating that customers like more the less expensive or mid-range models options. The gender distribution slightly show that male purchase more

treadmills somewhat more frequently than women, meaning both groups participate but men buy treadmills a bit more often. The marital status plot reveals that the majority of purchases are made by partners, suggesting that households with couples are more likely than single people to buy exercise equipment.

Key points:

- TM195 is the most popular treadmill model.
- More males than females appear in the dataset.
- Partnered customers buy treadmills more often than single customers.

Numerical Data

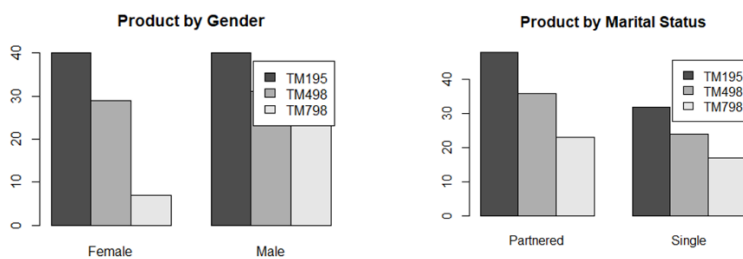


Then the histograms shows a clear patterns in the numerical variables. Age is mostly centered around people in their 20s and 30s, meaning low-aged adults are the main buyers and decrease by the age, meaning that when getting older, buying the less. Income has a wide spread, but most customers fall into the middle-income range. Education is mostly between 12–16 years. Usage and Miles show that customers plan for moderate exercise each week.

Key points:

- 20s people buy the most and decrease by the age.
- Most customers use the treadmill 3–4 times per week.
- Most plan to run around 30–40 miles per week.

Categorical vs Categorical Data

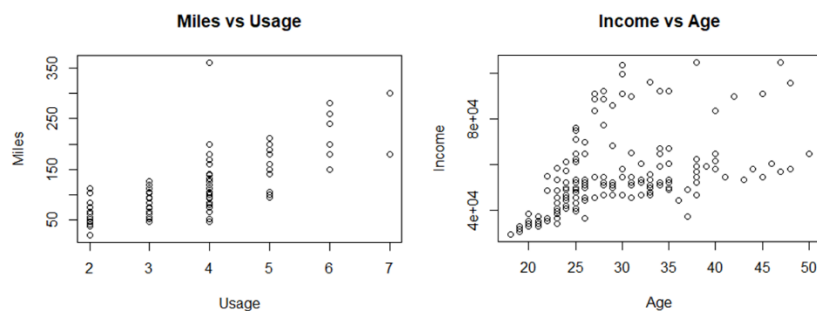


The barplots show clear differences across customer groups. In terms of Product vs. Gender, both genders buy every model; however, men buy somewhat more in total, with TM195 being the most popular for both. For Product vs. Marital Status, married consumers purchase much more treadmills than single consumers, been TM195 the most popular option.

Key points:

- Males buy slightly more treadmills than females.
- Partnered customers purchase more treadmills than single customers.
- TM195 is the most popular model across all groups.

Numerical vs Numerical Data



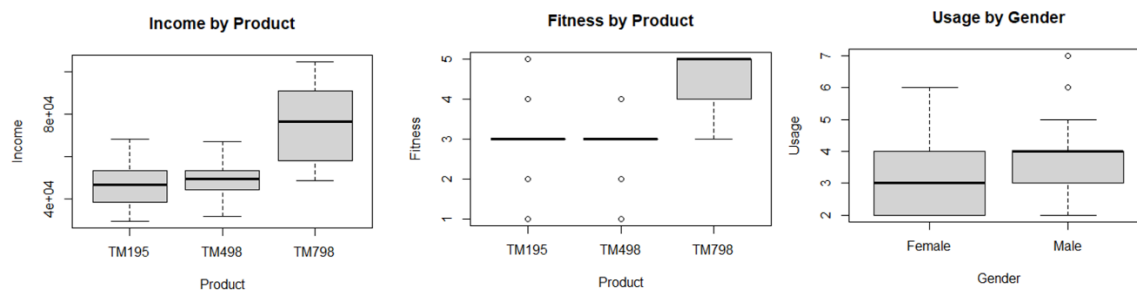
The Miles vs Usage scatterplot shows a clear upward pattern, meaning consumers who use the treadmill more times per week tend to run more miles. This shows a good relationship between weekly usage and weekly mileage.

The Income vs Age scatterplot doesn't show a strong pattern. Income varies widely at all ages, indicating that age does not determine income in this group.

Key points:

- More usage = more miles per week.
- Income doesn't show a clear trend with age.

Categorical vs Numerical Data



The Income x Product boxplot shows that TM798 buyers have higher incomes, while TM195 buyers tend to have lower incomes, indicating that premium model attracts wealthier customers.

The Fitness x Product boxplot shows that TM798 customers also have higher fitness levels, while TM195 buyers have lower level, meaning that fitter customers prefer the higher-end model. The Usage x Gender boxplot shows very similar usage for both males and females, indicating that gender does not affect how often the customers plan to use the treadmill.

Key points:

- TM798 buyers = higher income + higher fitness.
- TM195 buyers = lower income + lower fitness.
- Treadmill usage is similar for both genders.

Simple Linear Model

```
Call:
lm(formula = Usage ~ Age + Income + Education, data = data)

Residuals:
    Min       1Q   Median       3Q      Max
-1.96946 -0.60755 -0.06621  0.63709  2.37361

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.763e+00  7.115e-01   2.478  0.0141 *
Age          -5.259e-02  1.099e-02  -4.786 3.59e-06 ***
Income        4.170e-05  5.689e-06   7.330 8.05e-12 ***
Education     6.206e-02  5.192e-02   1.195  0.2336
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

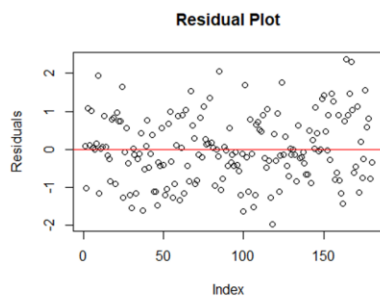
Residual standard error: 0.8744 on 176 degrees of freedom
Multiple R-squared:  0.3611, Adjusted R-squared:  0.3502
F-statistic: 33.16 on 3 and 176 DF, p-value: < 2.2e-16
```

The linear model shows the relationship between customers' intended frequency of treadmill use and their age, income, and level of education. The findings show that while age and education

have very small or negligible influence on weekly consumption, financial status has the biggest favorable impact.

Key points:

- The income is the only meaningful predictor of treadmill usage.
- Age and education have little to no impact.
- The model explains only a small portion of customer behavior.

Residual Plot

The residual plot demonstrates that there is no obvious trend and that the spots are dispersed above and below the zero line. However, the distribution of the points also demonstrates that the model is not very good at predicting consumption, which is consistent with the previously observed low R^2 value.

Key points:

- Model is acceptable.
- There is not a strong pattern.

- There is a wide spread of points.

Conclusion

The Cardio Good Fitness dataset analysis reveals different consumer trends that might inform product and marketing choices. The most popular treadmill model, the TM195, attracts to a wide range of consumers, while the TM798 often attracts buyers who are more financially stable and fit. Treadmills seem to be slightly more popular among male and partnered consumers, and the majority of buyers want to engage in moderate weekly activity. While the basic linear model indicates that weekly consumption is somewhat influenced by income, age and education have little impact and it is challenging to predict overall usage using these factors alone. According to these findings, Cardio Good Fitness may profit by focusing on various treadmill models for certain customers and providing marketing tactics based on household type, income, and degree of fitness.