Cardio Fitness Project

By Minh Ngo
June 2021

Table of contents

greatlearning
Power Ahead

Project Background

Background & Objective

- **Background**
  - Cardio Fitness is a retail store where it sells various cardio fitness equipment
  - The retail store has 3 main products with 3 codes :
    - TM195 (A)
    - TM498 (B)
    - TM798 (C)

    For simplicity, in this report, an alias of A,B,C would be used for each product code interchangeably.
  - The store has a customer database that include details about customer demographics

- **Objective**
  - Cardio Fitness wants to understand consumers profile to focus on the most promising segment and target new customers
  - The objective of this project is to explore the customers profile of Cardio Fitness, understand how customers profile vary across 3 main products and gather insights
  - From the insights we would make recommendation for the management team to target new customers for future marketing campaign

Background

- **Key questions**
  - What are main customers segments of the store? What is a typical customer profile like?
  - How are customers profile of each product different?
  - What type of products that high-income customers tend to buy?
  - What customers to target to sell specific products?

- **Approach**
  - EDA is used to explore insights from the data set
  - Univariate analysis were conducted for each variable and then multivariate analysis was used to explore relationship among each pair of variables
  - The customers profile for the store and of each products were assessed
  - Recommendations are derived based on insights

greatlearning
Power Ahead

Data Analysis - EDA

Data information

- The dataset contains product code and demographic information of customers who made the purchase
- There is no sales data associated with each product, hence sale analysis could not be conducted.

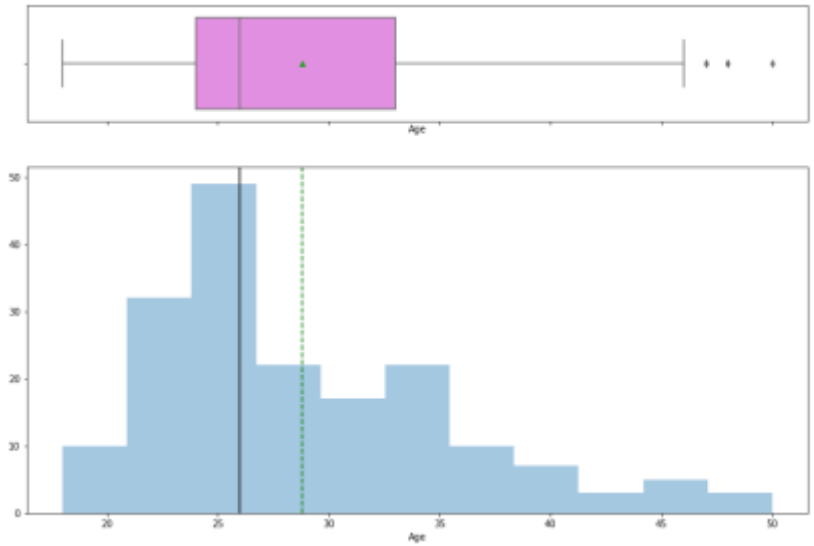
#	Variable	Description
1	Product	the model no. of the treadmill
2	Age	Customers' age
3	Gender	Gender
4	Education	Number of years of education
5	Marital status	Marital status
6	Usage	Average time the customer **wants** to use treadmill every week
7	Fitness	Self-rated fitness score (5 being very fit)
8	Income	Of the customers
9	Miles	Expected to run

Observation	Variable
180	9

Note:
- There is no missing value from the dataset
- There are 3 categorical variables (Product, Gender and Marital Status). The rest are numerical

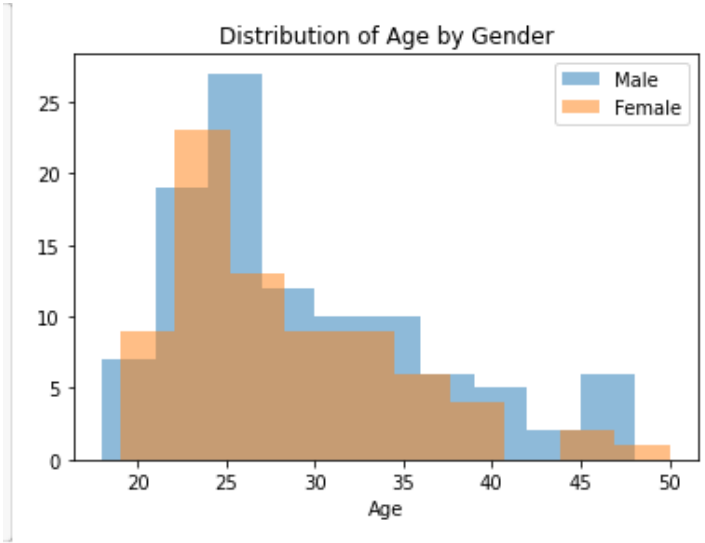
Univariate analysis (1/5)

Age



- The median is ~26 and the mean is ~28. The distribution of age is positively skewed, which means their customer group is of younger age .
- There are some outliers : customers from 40-50 age group

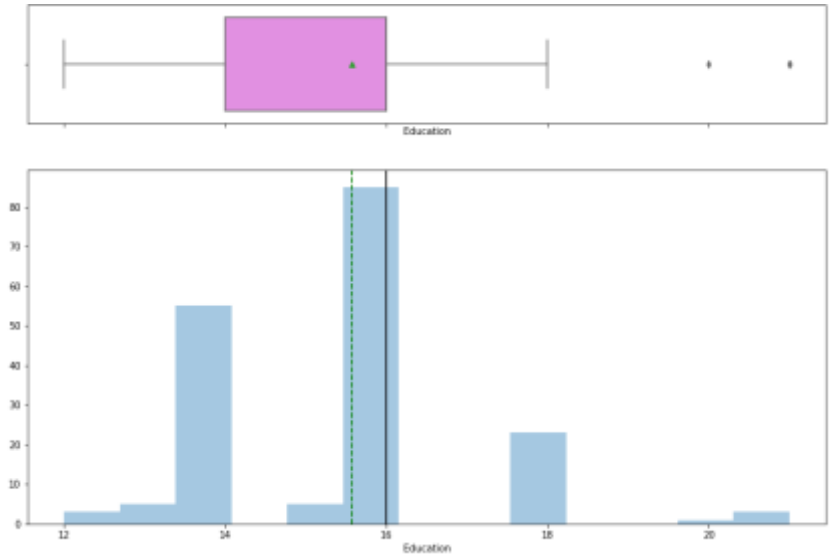
Age distribution by Gender



- The distribution of male vs female customers are slightly different
- There are more male customers of almost every age than female
- From 41-43 there are no female customers, only male

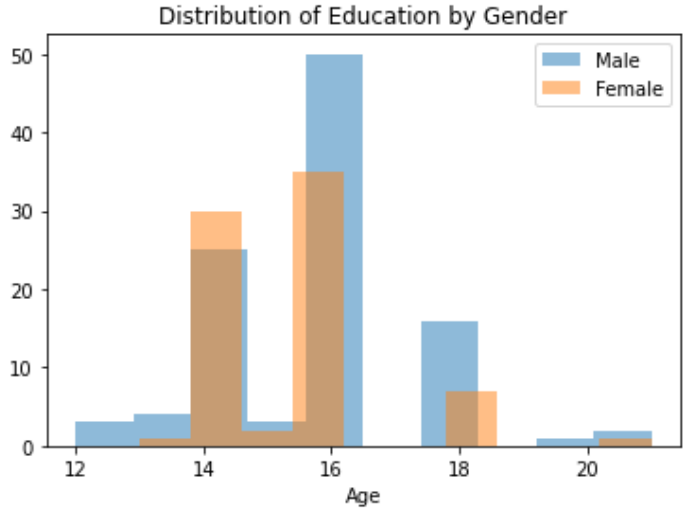
Univariate analysis (2/5)

Education



- The median of education year is ~16 and the mean is ~15.5. The distribution of age is positively skewed, which means their customer group is of younger age .
- Most of the customers have education (number of year) within 14 or 16 years. The 3rd large proportion group is 18 years of education.

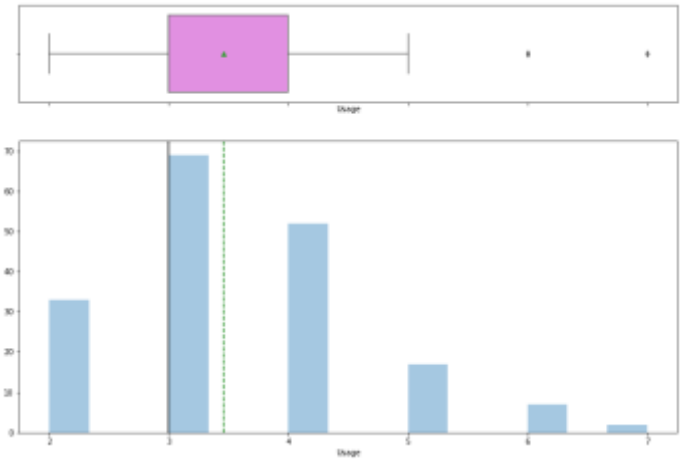
Education by Gender



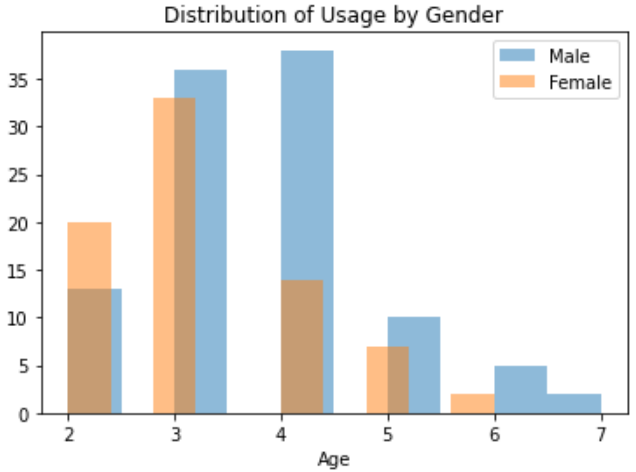
- In the group of 16 and 18 years of education, male customers outnumber female
- But in the group of 14 years of education, female outnumber male customers.
- Overall, male customers have higher years of education

Univariate analysis (3/5)

Usage



Usage by Gender

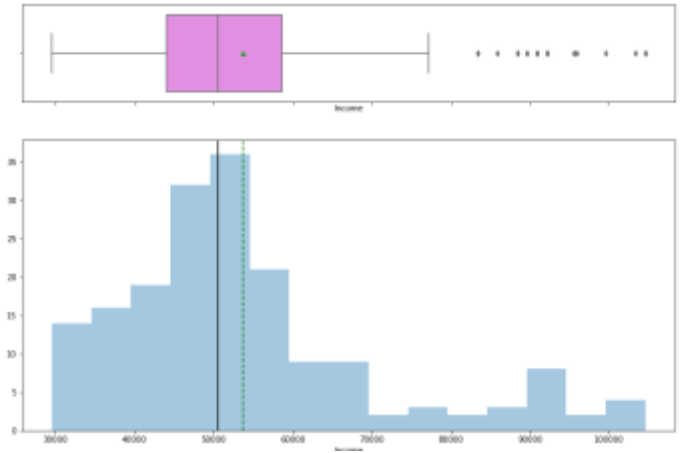


- The median of usage is 3 and the mean is 3.4. The distribution of age is positively skewed,
- Most customers use the fitness equipment around 3 or 4 times per week (every other days_, few use 6-7 times per week

- Men tend to exercise more time per week than women, in other words, they are more active.
- Most men exercise 4 times per week, most women only exercise 3 times per week

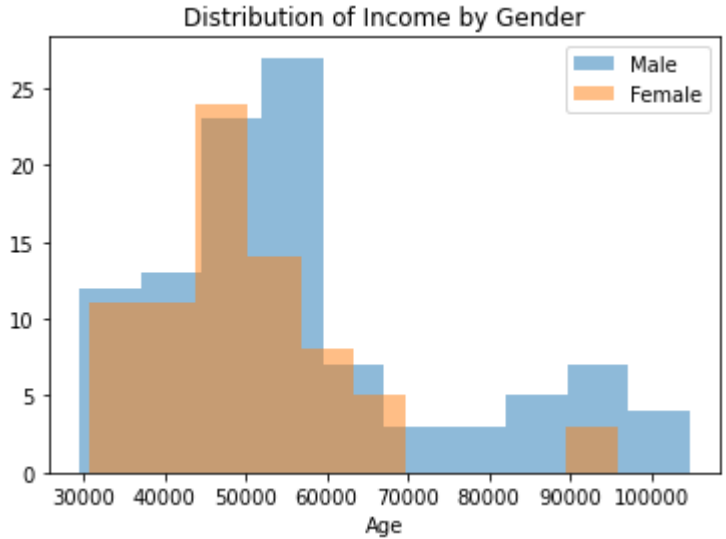
Univariate analysis (4/5)

Income



- The median of income is 50 and the mean is 53. The distribution of age is positively skewed, Most customers have income **within 30-70k range**
- There are some outliers who have income from 83-110k

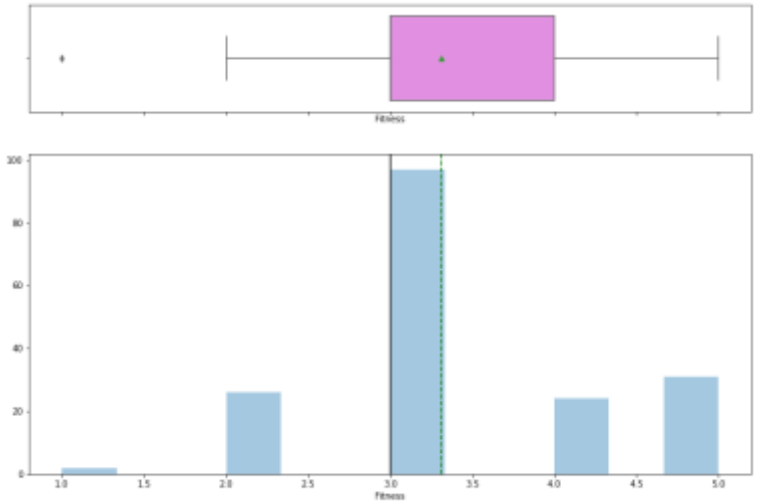
Income by Gender



- There is a big difference in income distribution between man and women
- Most female customers' income are within 30-70k/year. None makes between 70-90k. There are ~3 female customers making 90-95k/year only
- On contrary, male customers have higher income than female at every income group.
- Generally male customers make more per year than female.

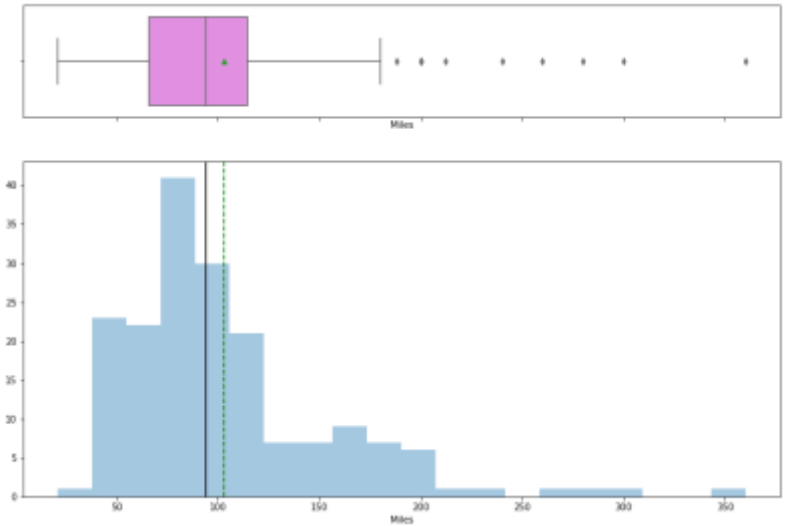
Univariate analysis (5/5)

Fitness



- Most customers assess their fitness level of 3.3 out of 5, which is slightly better than average.
- The next biggest group is the type of client who assess themselves at very fit – from 4.7 to 5.

Miles



- Most customers expect to run 70-120 miles per week with the mean of 101 miles/week

greatlearning
Power Ahead

Multivariate analysis – Pair plot

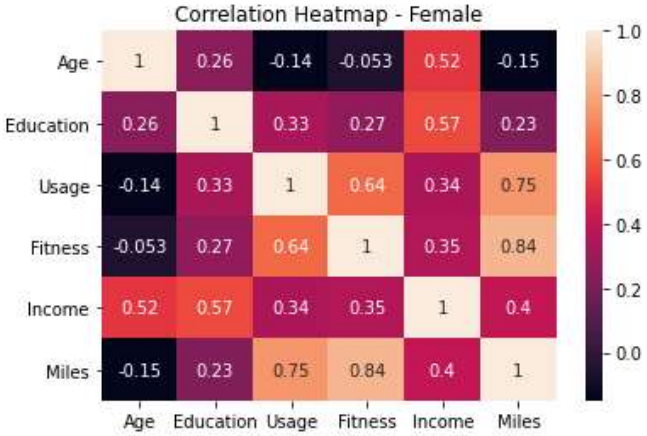
Pair with high correlation

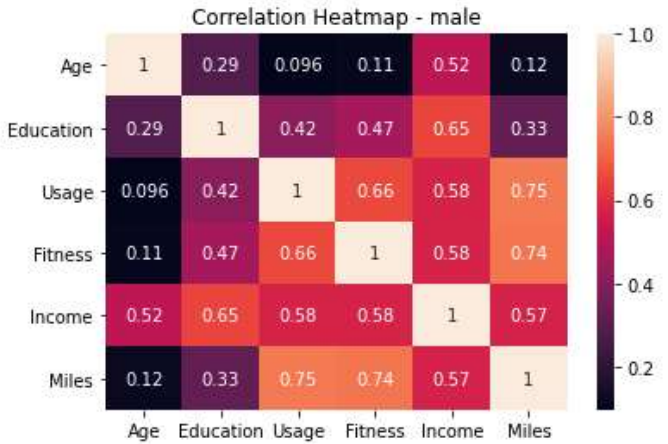
greatlearning
Power Ahead

Multivariate analysis – correlation matrix

- Education and Income have a quite high correlation of 0.6, this mean customers with higher education tend to have higher income
- Usage and miles also have a high correlation: 0.76 and this makes sense because the more you use the product, the higher the miles you can run
- Fitness and miles have the highest correlation: 0.79, this could possibly mean the more you run/use the equipment, the more fit you will get. However, this only show the correlation, it doesn't confirm the causation relationship between fitness and usage. It could also mean people being fit tend to be more health conscious and exercise more often, which in turn make them even more fit.
- Fitness and usage also have a high correlation: 0.66
- Fitness and Age have little correlation: which means age doesn't link with the fitness level of each customers

Multivariate analysis – correlation matrix by Gender



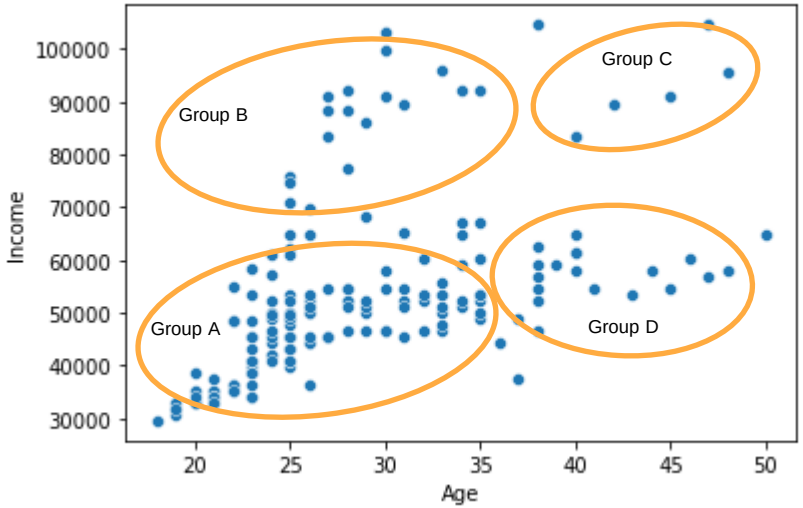


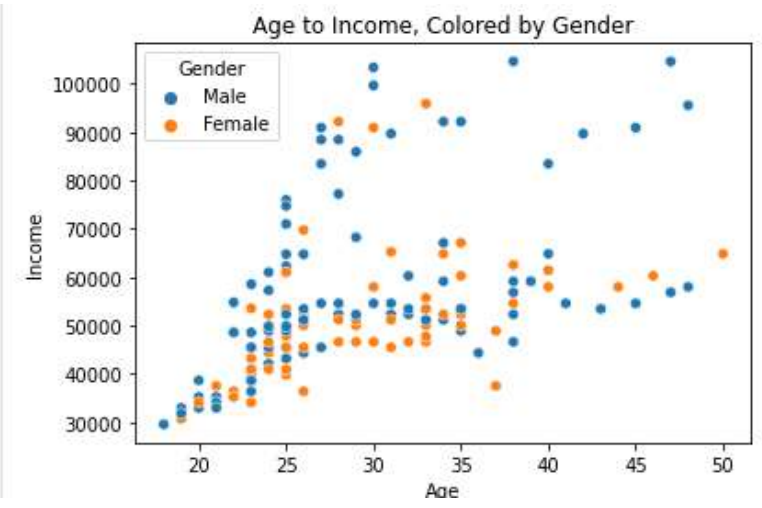
- Overall the correlation heatmap for female and male customers are not significantly different, except for:
- Income vs Fitness: correlation of female is 0.35, quite lower than that of male- which is 0.58
- Miles vs Fitness: correlation of female is 0.84, quite higher than that of male- which is 0.74

Customer profiling

Customer profile: segmentation by Age & Income

CardioFitness may want to target to sell more expensive product for Group B (young + high income)



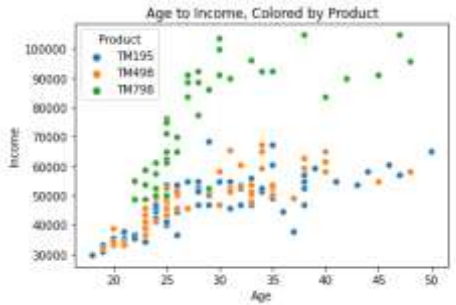


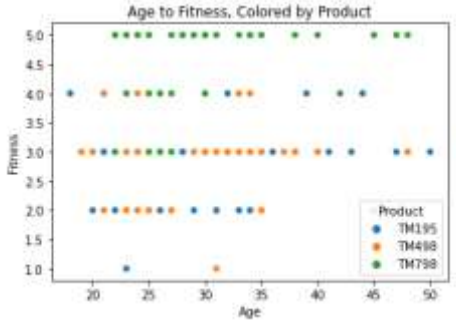
- We can roughly classify customers into 4 main groups based on age and income as above.
- Group A is the core customers group.
- Group C has highest income but has smallest size and may not worth targeting.
- Group B is of same age group like group A yet have very high income, this is the group that CardioFitness may want to grow and sell higher priced items.

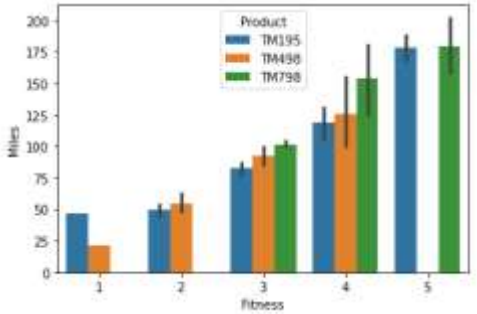
- Similar to our previous findings, male customers have higher income than female customers

Customer analysis by Product

Overall Product 798 seems to be more “hardcore”, more expensive than the other 2 products







- Product TM798 are used by higher income group
- Product TM195 are used by almost every age group and with lower income (<70k)
- Product TM498 are used by younger group (20-35) and with lower income (<70k)

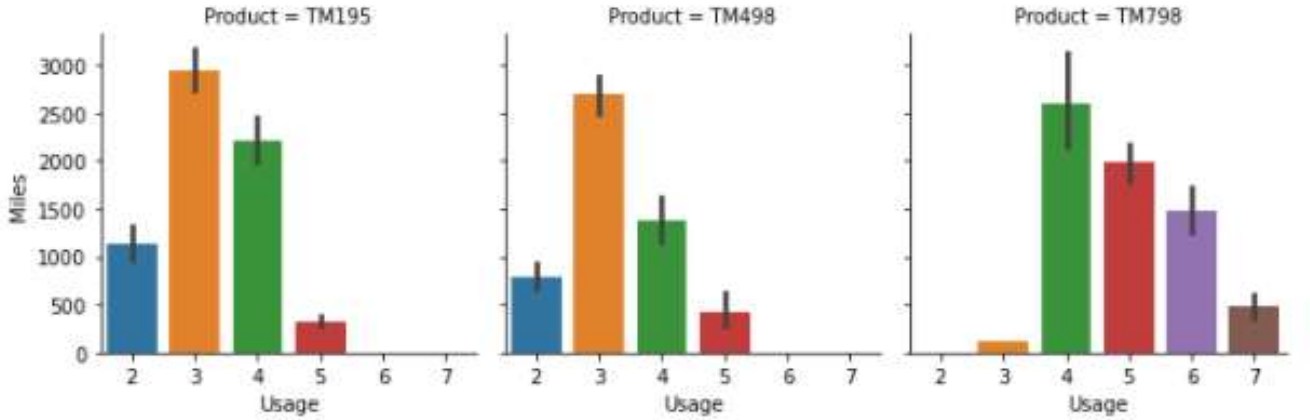
- Product TM798 are used by customers who rate themselves as “most fitness” (score 5)
- There is no clear pattern as to fitness level for the other two products

- Product TM798 are associated with higher miles than other 2 products , meaning people who buy TM798 tend to be more serious with exercise and will run more miles per week

Customer analysis by Product

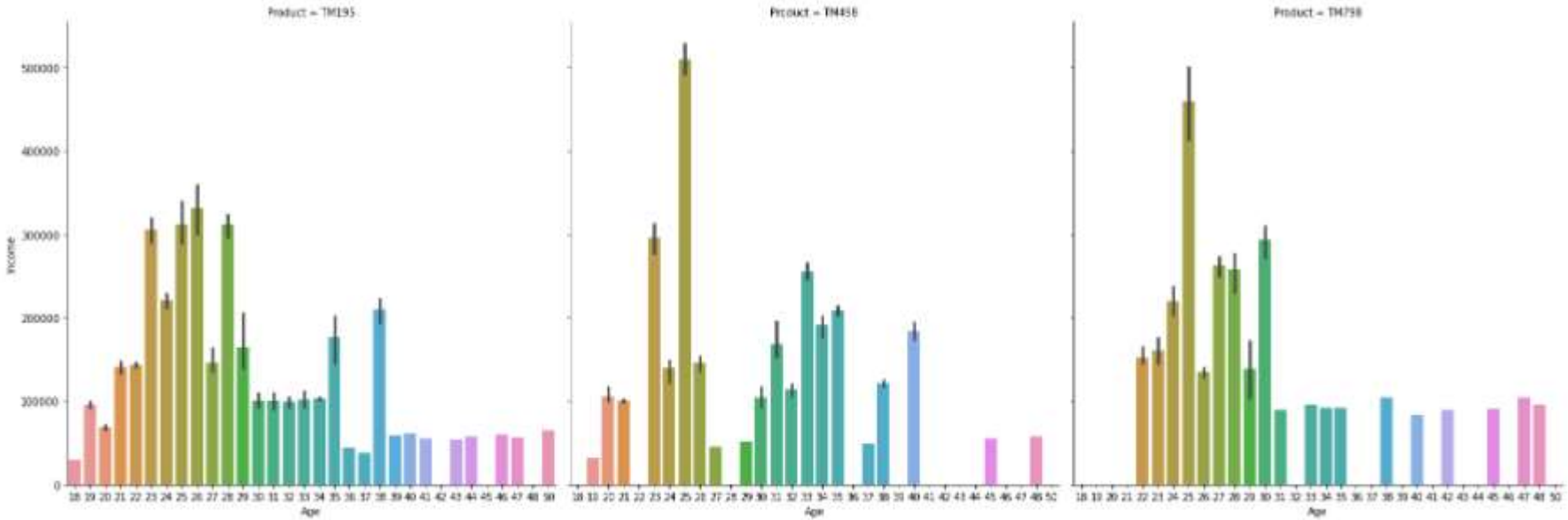
Product TM798 have higher usage per week than other products and could be considered a premium product

Standard 1 | Standard 2 | Premium



- Product TM195: is bought/used by almost every age group
- Product TM498: has the highest sale for people of 25 years old. The sales are not distributed as even as TM195, some age group don't even buy it at all. This product may be age sensitive
- Product TM798: no-one younger than 22 would buy this product, the sales concentrates heavily from 22 to 31 age group. This product could be age-sensitive too

Customer profile – income & age: by Product

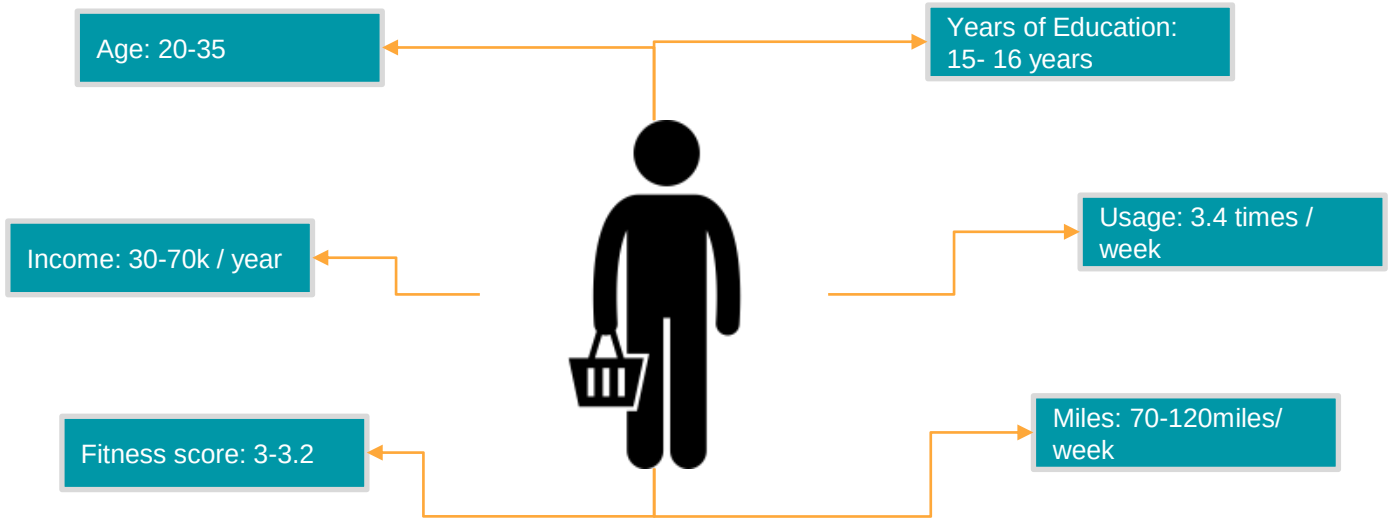


- Product TM195: is bought/used by almost every age group
- Product TM498: has the highest sale for people of 25 years old. The sales are not distributed as even as TM195, some age group don't even buy it at all. This product may be age sensitive
- Product TM798: no-one younger than 22 would buy this product, the sales concentrates heavily from 22 to 31 age group. This product could be age-sensitive too

Key insights

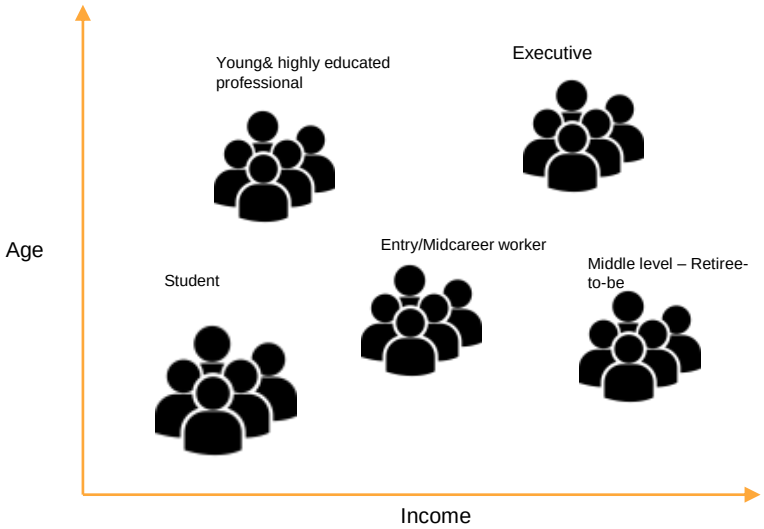
Key findings:
Overall Customer profile

The below summary depicts the typical customer group of Cardio Fitness (not including the outliers)



Key findings & observations

- Years of education are highly correlated with income.
- Male customers have higher income than female customers
- Product TM798 could be considered a premium product with higher price and better fit with higher income customers. People who bought this product tend to be more serious with exercise, they use it more often and run more miles every week.
- Product TM195 is bought by almost every age group, this can be considered a mass product
- Product TM 498 could be age sensitive (some age group don't even buy it). The company needs to explore the reason behind possibly via survey on customer perception and behavior on this product.

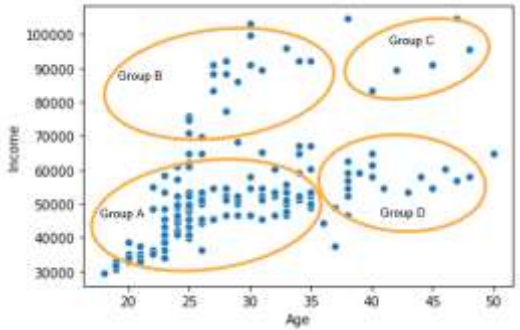


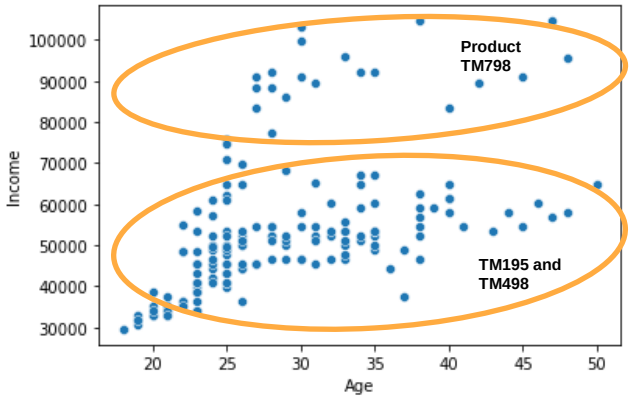
Recommendations

greatlearning
Power Ahead

Recommendations

- Overall, CardioFitness should target its new marketing campaign for customers who have the similar profile as in page 22. Those audiences are likely to be converted into Cardio Customers.
- Sales associate should also use **years of education as a token for income** to sell more pricey product
- Sales associate also could use level of fitness and usage to gauge the possibility to sell products. People who exercise more or rate themselves as very fit tend to buy TM798
- Group B is the most potential group to expand (higher income + 2nd biggest in size) beside group A. The company may want to **understand this group better** (their purchasing criteria, preference, pain point etc.) via survey / in person interview
- In terms of product
    - TM798: sell to group B (young + higher income) and target higher income male customers (who have deeper pocket and have more usage per week)
    - TM195 and TM498 (mass / lower end products): could be targeted to female customers or any customers who have lower income and do not exercise much (=<3 times/week)





greatlearning
Power Ahead

Happy Learning !