STAT560 Course Project

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. The data are stored in the CardioGoodFitness.csv file. The team identifies the following customer variables to study: product purchased, TM195, TM498, or TM798; gender; age in years; education in years; relationship status, single or partnered; annual household income ($); average number of times the customer plans to use the treadmill each week; average number of miles the customer expects to walk/run each week; and self-rated fitness on an 1-to-5 scale, where 1 is poor shape and 5 is excellent shape. Perform descriptive analytics to create a customer profile for each Cardio Good Fitness treadmill product line.

Objective:

In this project, you are expected to generate a set of insights and recommendations that will help the company in targeting new customers. You will explore insights from the data to improve treadmill recommendation based on the income, marital status or age, etc and boosting out sales with potential buyers of a specific category of the treadmill.

You will write a report summarize your findings with visualizations or outputs from R. You will need to prepare a report that is accessible to a broad audience. It will need to outline your motivation, analysis steps, findings, and conclusions. Do not include your R code in the report. Submit the R code you used in appendix or in a separate file.

Some questions to be answered:

Below is a list of some possible questions that you may try to explore using the data. You do not have to answer all those questions and you should certainly not limit yourself to the list of the questions.

- How many models does store have?
- Which is most sold Model?
- Are Male customers buying treadmill more than female customers?
- What is the average Income, Age, Education of people buying treadmill?
- How many days and miles customer expect to run on treadmill?
- What is the average self-rated fitness of customers buying treadmill?
- Are married customers buying Treadmill more than Single customers?
- Is there any relation between Income and model?
- Is there any relation between Age and model?
- Is there any relation between self-rated fitness and model?
- Is there any relation between education and model?
- Does gender have any effect on model customer buy?
- Does Martial status have any effect model customer buy?
- Are different age groups buying different models?
- Relation between Age, Income and education and model bought?

Dataset Information:

cardiogoodfitness.csv: The csv contains data related to customers who have purchased different model from Cardio Good Fitness.

- 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.

Things to do (with examples):

- Read the data file into R and observe a glimpse of the data.

  e.g. What is the dimensions of the data? How may observations and how many variables?

- Data preprocessing and exploration.

  e.g. Are there missing data? What are the numerical / categorical variables? What is the range of customer education and what is the median? What is the average age of customer who purchases TM195?

- Generate various visualizations of the data and get more insights.

  e.g. Generate univariate plots to check outliers and the shape of the distribution. Generate multivariate graphics such as side by side boxplot of income v.s. models purchased, and discuss findings

- Perform univariate, bivariate and multi-variate tests to draw inferences.

  e.g. Perform appropriate tests to see if gender affects the model purchased. Perform tests to see if there any difference in mean age among customers that purchased different models.

- Summarize your major observations and recommendations.

  e.g. Which model is most popular? Which model attracts people with lower income? we should market it as a budget Treadmill. Which model attracts the professionals with higher income? We should market it as a high-end luxurious Treadmill.