



# Project - Cardio Good Fitness

---

Model Report

## Table of Contents

|       |   |   |
|-------|---|---|
| 1     | Project Objective.....                                  | 3 |
| 2     | Assumptions.....  | 3 |
| 3     | Exploratory Data Analysis – Step by step approach ..... | 3 |
| 3.1   | Environment Set up and Data Import.....                 | 3 |
| 3.1.1 | Install necessary Packages and Invoke Libraries.....    | 3 |
| 3.1.2 | Set up working Directory .....                          | 3 |
| 3.1.3 | Import and Read the Dataset.....                        | 4 |
| 3.2   | Variable Identification.....                            | 4 |
| 3.2.1 | Variable Identification – Inferences.....               | 4 |
| 3.3   | Univariate Analysis.....                                | 4 |
| 3.4   | Bi-Variate Analysis.....                                | 5 |
| 3.5   | Missing Value Identification.....                       | 5 |
| 3.6   | Outlier Identification.....                             | 5 |
| 3.7   | Variable Transformation / Feature Creation .....        | 5 |
| 4     | Conclusion.....   | 5 |
| 5     | Appendix A – Source Code.....                           | 5 |

## 1 Project Objective

The objective of the report is to explore the cardio data set ("CardioGoodFitness") in R and generate insights about the data set. This exploration report will consist of the following:

- Importing the dataset in R
- Understanding the structure of dataset
- Graphical exploration
- Descriptive statistics
- Insights from the dataset

## 2 Assumptions

*<Think from practical Project Execution perspective. Add all your assumptions here.>*

## 3 Exploratory Data Analysis – Step by step approach

A Typical Data exploration activity consists of the following steps:

1. Environment Set up and Data Import
2. Variable Identification
3. Univariate Analysis
4. Bi-Variate Analysis
5. Missing Value Treatment (Not in scope for our project)
6. Outlier Treatment (Not in scope for our project)
7. Variable Transformation / Feature Creation
8. Feature Exploration

We shall follow these steps in exploring the provided dataset.

Although Steps 5 and 6 are not in scope for this project, a brief about these steps (and other steps as well) is given, as these are important steps for Data Exploration journey.

### 3.1 Environment Set up and Data Import

#### 3.1.1 Install necessary Packages and Invoke Libraries

Use this section to install necessary packages and invoke associated libraries. Having all the packages at the same places increases code readability.

#### 3.1.2 Set up working Directory

Setting a working directory on starting of the R session makes importing and exporting data files and code files easier. Basically, working directory is the location/ folder on the PC where you have the data, codes etc. related to the project.

Please refer Appendix A for Source Code.

### 3.1.3 Import and Read the Dataset

The given dataset is in .csv format. Hence, the command 'read.csv' is used for importing the file.

Please refer Appendix A for Source Code.

## 3.2 Variable Identification

<Specify which R functions you are using for what purpose in brief. >

### 3.2.1 Variable Identification – Inferences

<Provide your insides from the output of various R functions used to explore the data such as dim, names, str, head, tail, etc.>

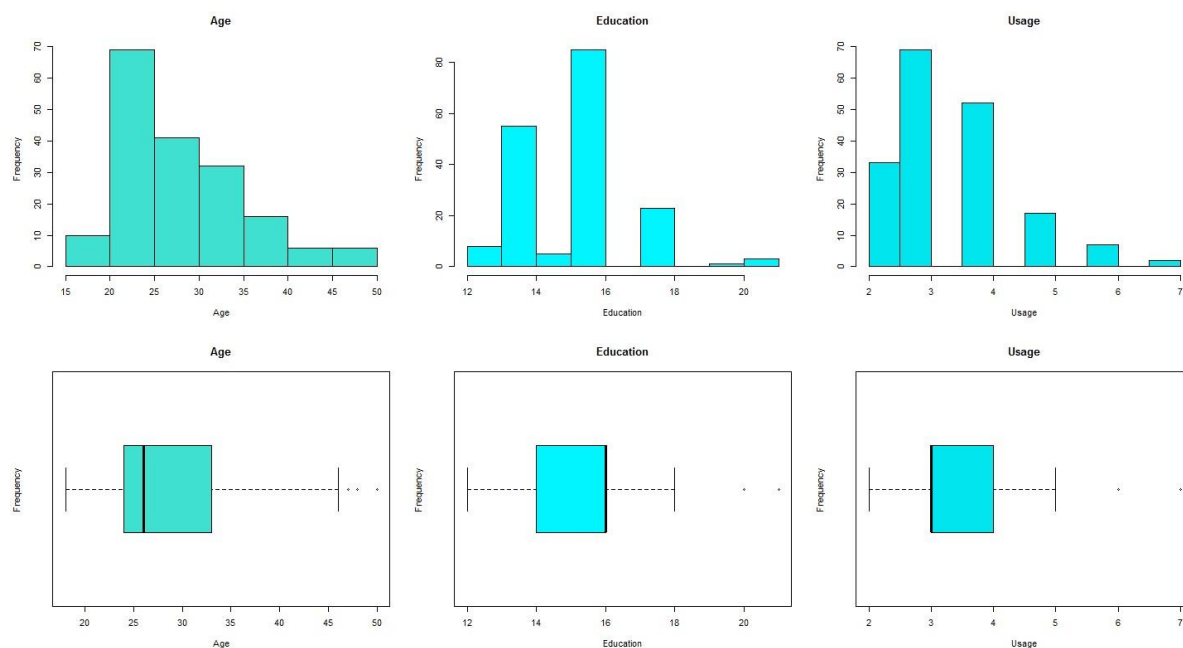
## 3.3 Univariate Analysis

<Explore individual variables one by one. >

<Present your findings in tabular format, make use of appropriate Graphs>

<A Sample graph image is provided below>

<Summarize key observations about each variable>



### 3.4 Bi-Variate Analysis

*<To explore relationship between two variables >*

*<Interpret the findings>*

### 3.5 Missing Value Identification

*<See if any missing values / Outliers?>*

### 3.6 Outlier Identification

### 3.7 Variable Transformation / Feature Creation

*<Do you see a need of transforming a variable / creating new variables for better understanding of the data, or presenting the results to the customer?>*

*<Act accordingly>*

## 4 Conclusion

*<Very important section of the report, often ignored, or just produced technical details.*

*Always keep in mind that the Sr. Management / Customer is going to read this section, so it has to be in business language.*

*You may provide the conclusion for each attribute in brief.>*

## 5 Appendix A – Source Code

*<This is another section often ignored. Although not mandatory, make a practice of providing source code for every assignment you are submitting. R Mark down language is one of the powerful features available in RStudio. If not for this assignment, try to learn it for documentation purpose. It can be self-learnt.*

*Here is a sample piece of code produced in RStudio using R Markdown>*

```
#=====
#
#   Exploratory Data Analysis - CardioFitness
#
#=====
# Environment Set up and Data Import
# Setup Working Directory
setwd("D:/M1 Project")
getwd()

#
# Read Input File
cgf_data=read.csv("CardioGoodFitness.csv")
attach(cgf_data)
```

```
#
# Find out Total Number of Rows and Columns
dim(cgf_data)

## [1] 180 9

# Find out Names of the Columns (Features)
names(cgf_data)

## [1] "Product"      "Age"          "Gender"       "Education"
## [5] "MaritalStatus" "Usage"        "Fitness"      "Income"
## [9] "Miles"

# Find out Class of each Feature, along with internal structure
str(cgf_data)

## 'data.frame': 180 obs. of 9 variables:
## $ Product      : Factor w/ 3 levels "TM195","TM498",...: 1 1 1 1 1 1 1
## $ Age          : int 18 19 19 19 20 20 21 21 21 21 ...
## $ Gender       : Factor w/ 2 levels "Female","Male": 2 2 1 2 2 1 1 2 2
## $ Education    : int 14 15 14 12 13 14 14 13 15 15 ...
## $ MaritalStatus: Factor w/ 2 levels "Partnered","Single": 2 2 1 2 1 1
## $ Usage        : int 3 2 4 3 4 3 3 3 5 2 ...
## $ Fitness      : int 4 3 3 3 2 3 3 3 4 3 ...
## $ Income       : int 29562 31836 30699 32973 35247 32973 35247 32973
## $ Miles        : int 112 75 66 85 47 66 75 85 141 85 ...

#
.
.
.

#=====
#
# THE - END
#=====
```