**Objective** - Preliminary Data Analysis. Explore the dataset and practice extracting basic observations about the data. The idea is for you to get comfortable working in Python.

You are expected to do the following :

1. Come up with a customer profile (characteristics of a customer) of the different products
2. Perform uni-variate and multi-variate analyses
3. Generate a set of insights and recommendations that will help the company in targeting new customers

**Context** - The data is for customers of the treadmill product(s) of a retail store called Cardio Good Fitness. It contains the following variables

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

Explore the dataset to identify differences between customers of each product. You can also explore relationships between the different attributes of customers. You can approach it from any other line of questioning that you feel could be relevant for the business.

**Best Practices for Notebook :**

* The notebook should be well-documented, with inline comments explaining the functionality of code and markdown cells containing comments on the observations and insights.
* The notebook should be run from start to finish in a sequential manner before submission.
* It is preferable to remove all warnings and errors before submission.
* The notebook should be submitted as an HTML file (.html) and as a notebook file (.ipynb)

**Submission Guidelines :**

1. There are two parts to the submission:
   1. A well commented Jupyter notebook [format - .ipynb]
   2. File converted to HTML format
2. Any assignment found copied/ plagiarized with other groups will not be graded and awarded zero marks
3. Please ensure timely submission as any submission post-deadline will not be accepted for evaluation
4. Submission will not be evaluated if,
   1. it is submitted post-deadline, or,
   2. more than 2 files are submitted

Happy Learning!!

**Scoring guide (Rubric) - Cardio Good Fitness Rubric**

| **Criteria** | **Points** |
| --- | --- |
| **Understanding the structure of the data**  - Overview of the dataset shape, datatypes - Statistical summary and check for missing values | 3 |
| **Univariate Data Analysis**  Analysis of spread and distribution of every feature in the dataset. | 12 |
| **Multivariate Data Analysis**  Analysis of interaction between features, in the dataset | 15 |
| **Quality & Use of visualisations**  - The details in Visualization used. | 9 |
| **Conclusion and Recommendations**  - Conclude with the key insights/observations | 15 |
| **Well commented Python Code**  - Structure and flow - Well commented code | 6 |
| Points | 60 |