Aerofit
FROM FIT-LESS TO FITNESS

1. Problem Statement & Basic Matrix

Problem Statement: Aerofit aims to identify the characteristics of the target audience for each type of treadmill offered by the company. The objective is to analyse customer data to understand differences in customer characteristics across different treadmill products, and provide tailored recommendations to new customers.

Dataset Overview:

- Product Purchased: KP281, KP481, or KP781
- Age: In years
- Gender: Male/Female
- Education: In years
- Marital Status: Single or partnered
- Usage: Average number of times the customer plans to use the treadmill each week
- Income: Annual income (in $)
- Fitness: Self-rated fitness on a 1-to-5 scales
- Miles: Average number of miles the customer expects to walk/run each week

2. Observations and Data cleaning

Row data Shape - `(180, 9)`
Missing/Null values - 0

Observations:

- The dataset has 180 entries with no missing values.
- Categorical attributes: Product, Gender, Marital Status.
- Numerical attributes: Age, Education, Usage, Income, Fitness, Miles
- Range of Age: (18, 50)
- Range of Income: (29562, 104581)

DataTypes (After category conversion)

```
Product        category
Age             int64
Gender         category
Education         int64
MaritalStatus  category
```

Usage int64
Fitness int64
Income int64
Miles int64

3. Non-Graphical Analysis: Value counts and unique attributes

Value counts
From the value counts of each category, I can infer that most numbers are from top 40-80 categories only for Products 76-104 for gender & 73-107 for Marital status, data spread is not even for all categories it's skewed for age, income and miles.
For the value counts for each attribute please refer PDF of python file.

Unique Attributes -

Product
KP281 80
KP481 60
KP781 40

Gender
Male 104
Female 76

MaritalStatus
Partnered 107
Single 73

Summary statistics

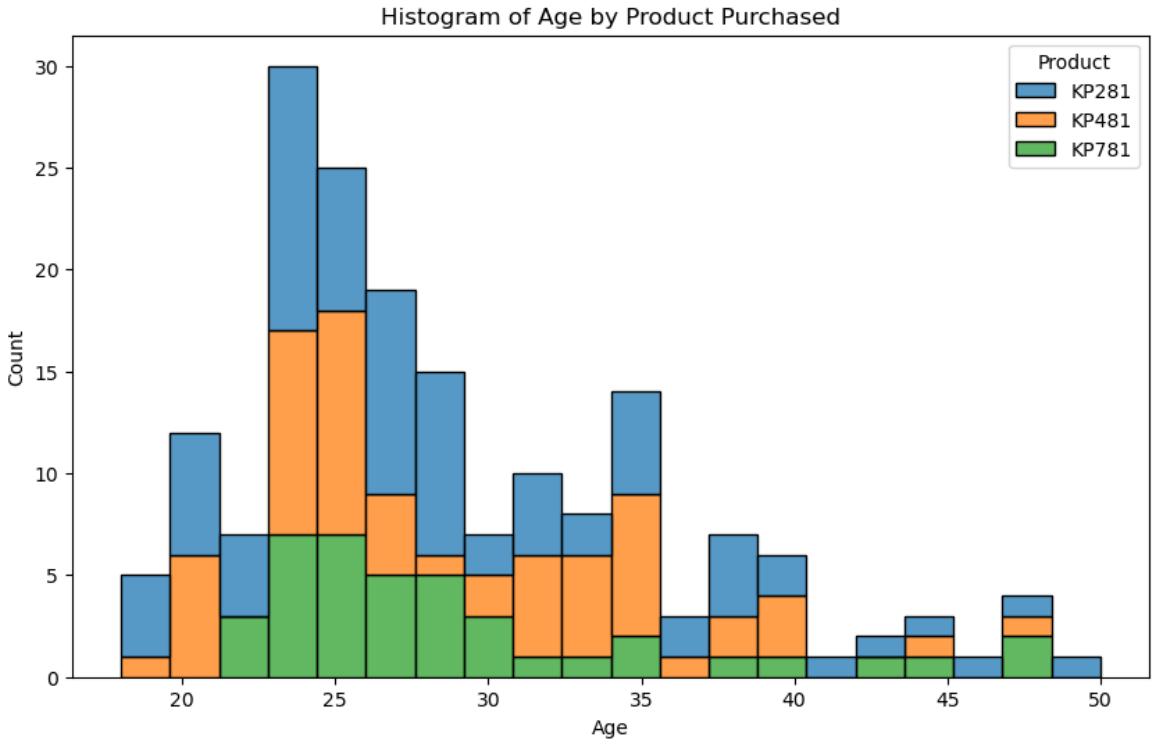
Kindly refer to the python file. Data format is not proper to show here.

4. Visual Analysis

4.1 Histogram

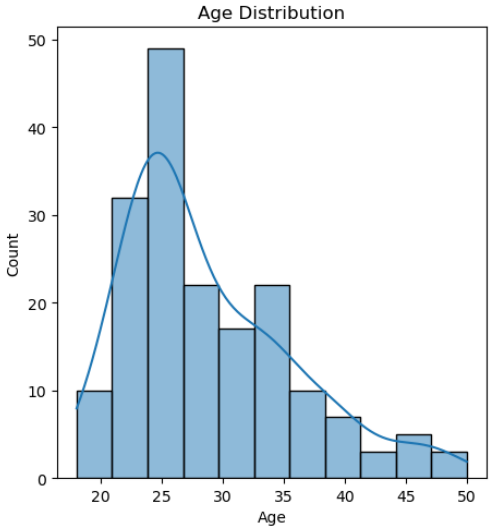
After data cleaning and creating Histplot(Histogram) with setting the bins to 50,
Here is what I observed

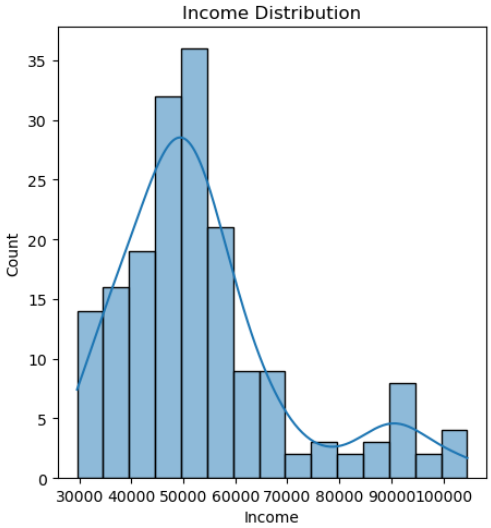
1. Product KP281 an entry-level treadmill that sells for $1,500 sells more than other two products
2. Age 22-35 has the maximum purchase of the advance treadmill KP781
3. Young generation from age 24-35 has purchased most treadmills.

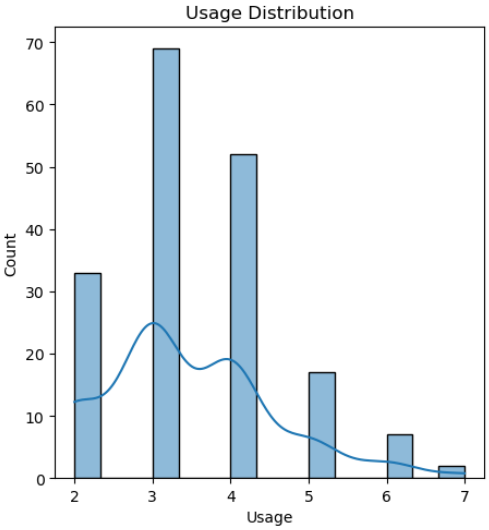


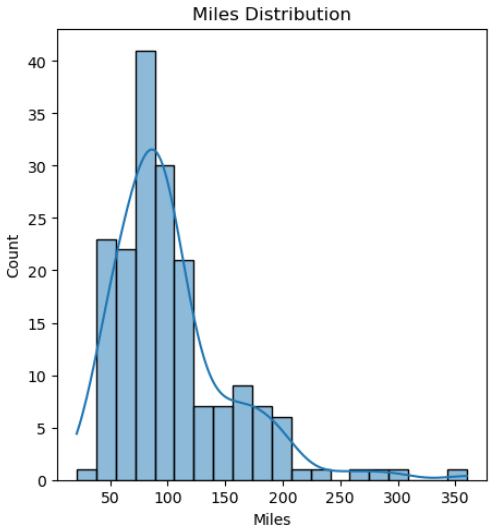
4.2 Histogram - column wise

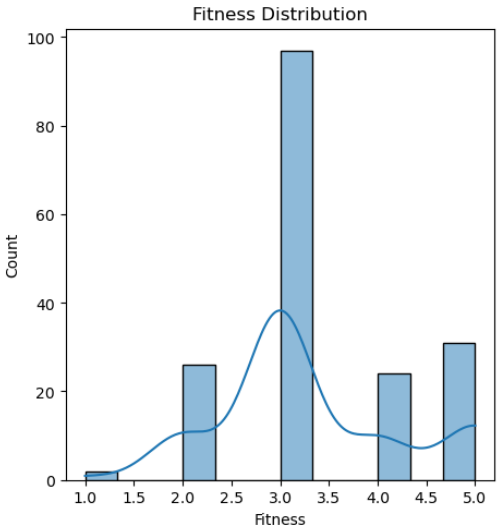
We can observe that Age, Income and Miles column has skewed data.





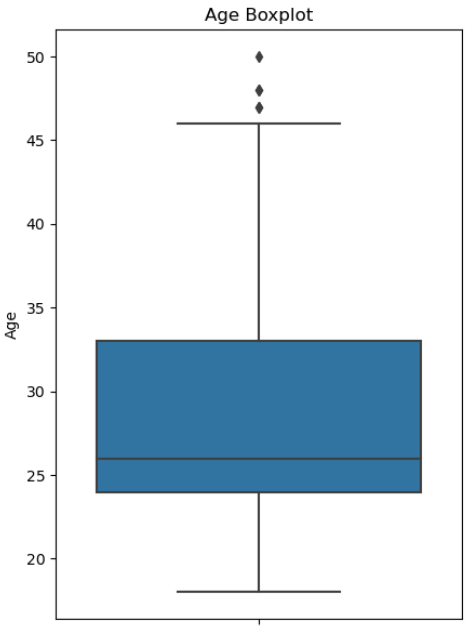


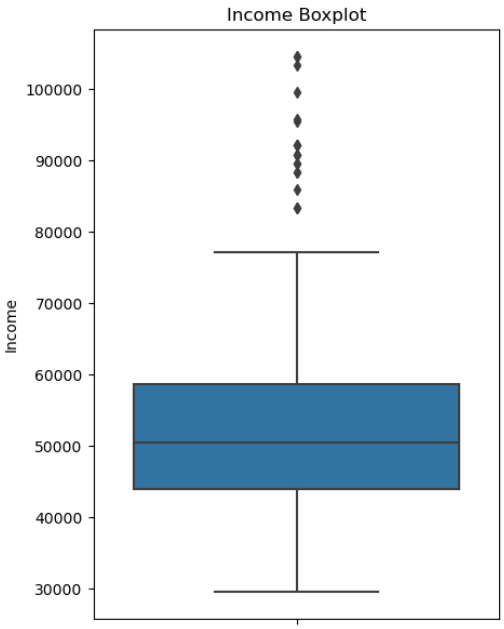


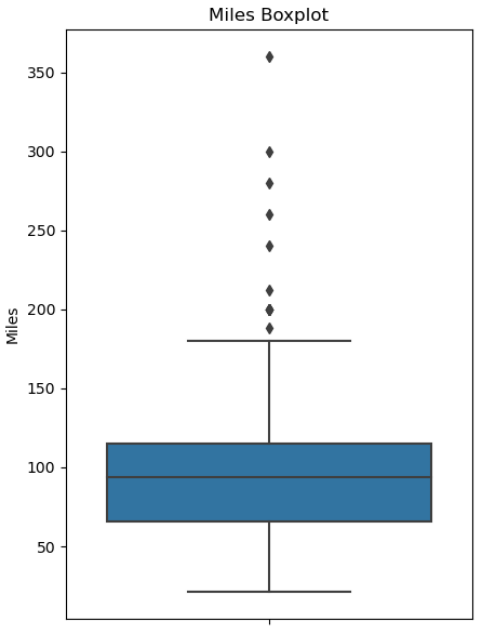


4.3 Boxplot for Age, Income and Miles

Box plot Infers the Age, Income and Miles columns data spread, as we can see Income has the most outliers.

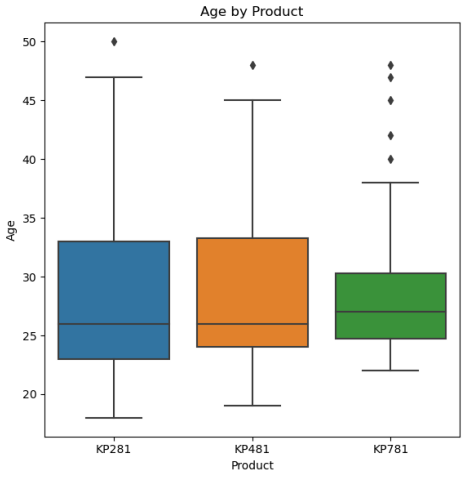


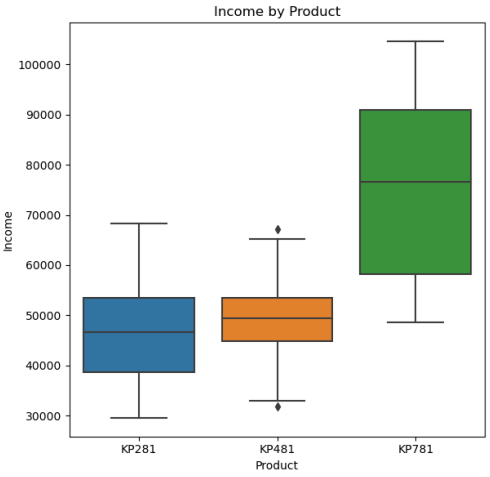


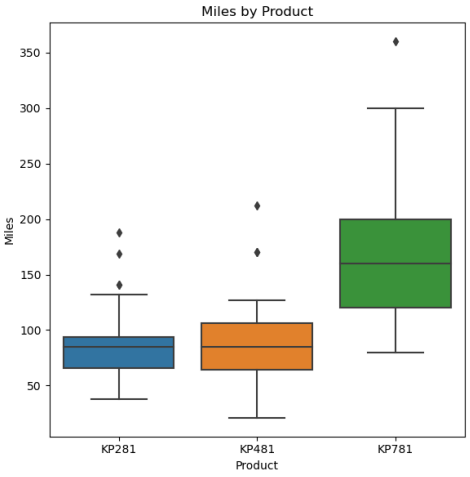


This is an interesting visual to get clear understanding of the data, we can infer that

- Higher the income higher the chance of product KP781 to be sold
- People who walk more are tends to buy advance featured trade mill KP781

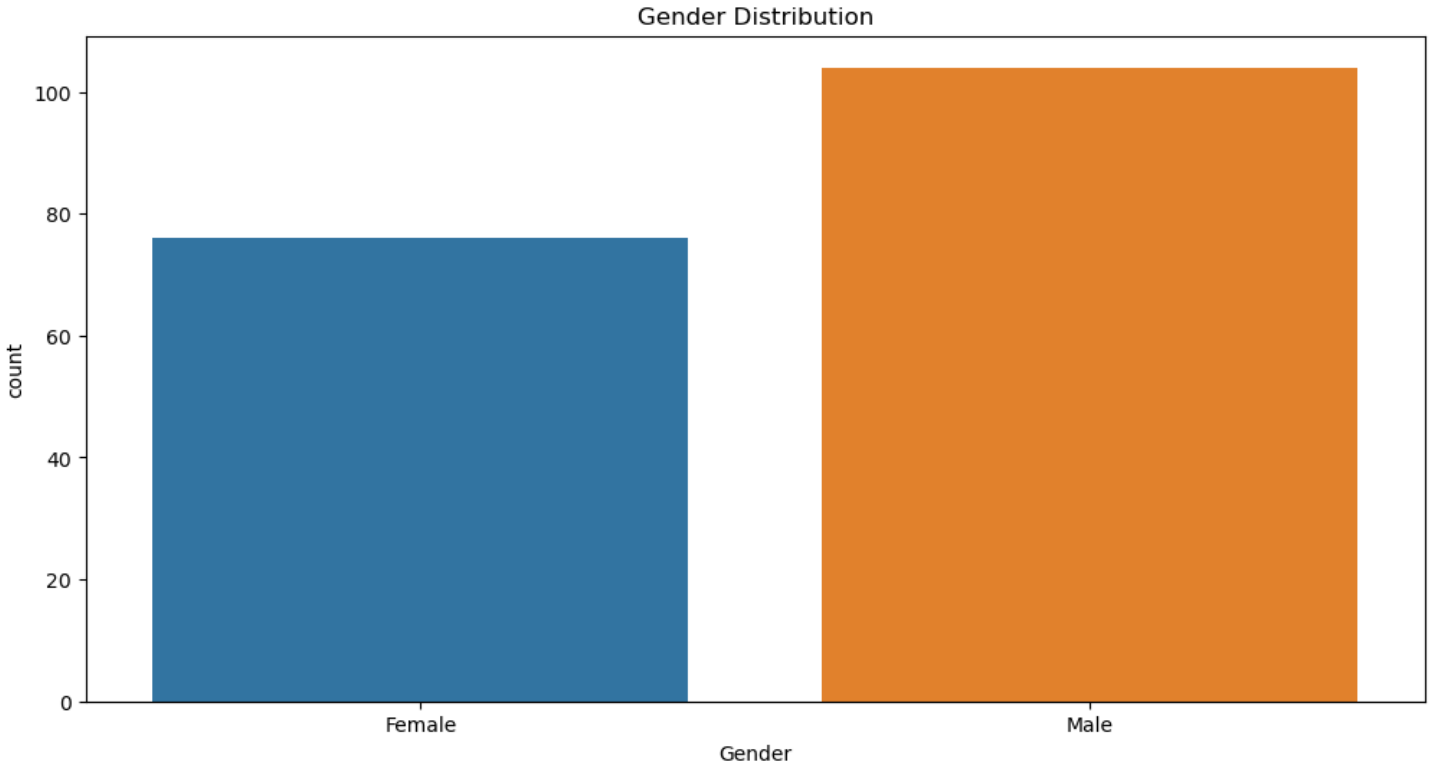


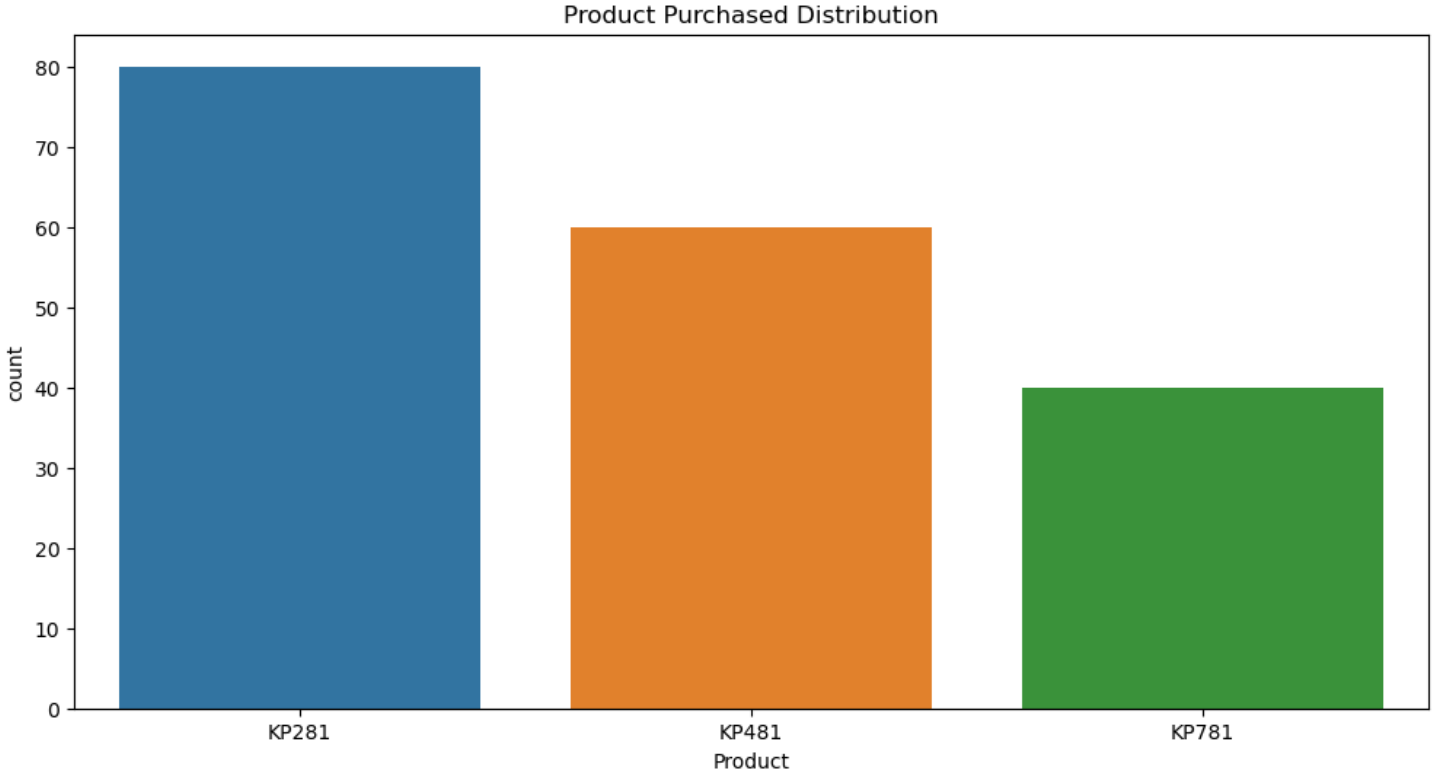


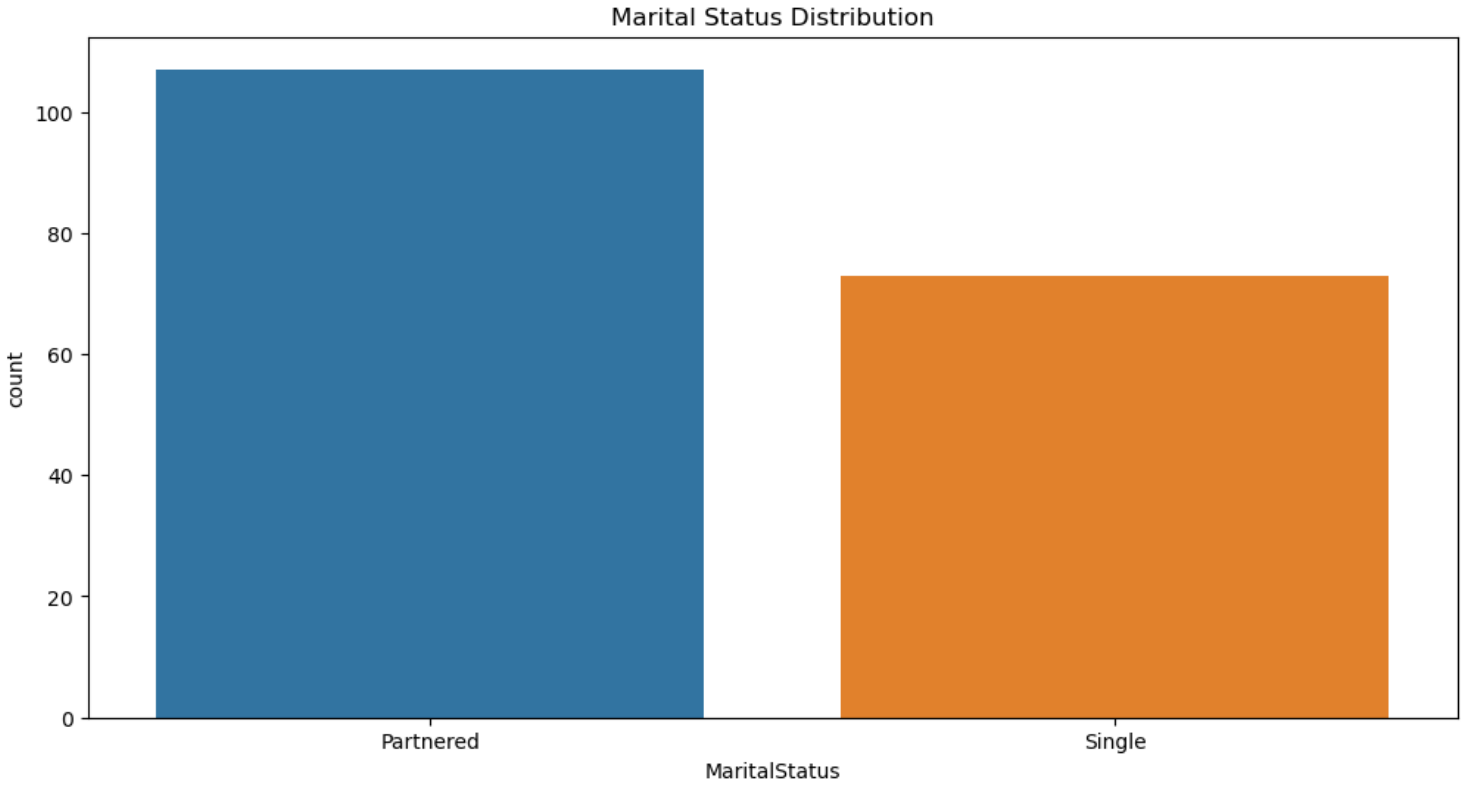


4.4 Count plots for Categorical Variables

1. Count plot shows the speed of gender, male customers has purchased more treadmills.
2. Count plot shows the speed of Marital status, partnered customers has purchased more treadmills.


Gender Distribution
100
80
60
40
20
0
count
Female
Male
Gender


Product Purchased Distribution
80
70
60
50
40
30
20
10
0
count
KP281
KP481
KP781
Product



5. Missing Values & Outliers check

There are no missing values in the dataset, Outlier's check has been done in above box plot analysis.

6. Insights based on Non-Graphical and Visual analysis

6.1 Comments on the Range of Attributes

- Range of Age: (18, 50)
- Range of Income: (29562, 104581)

6.2 Comments on the correlation heatmap

- The correlation matrix shows a positive correlation between fitness, usage and miles
- Age and usage of the trade mill has negative corelation
- Similarly, age and miles have negative correlation

7. Business Insights based on Non-Graphical and Visual Analysis

Comments on Range of Attributes:

- Age ranges from 18 to 50.
- Income ranges from $29,000 to $105,000.
- Usage ranges from 2 to 7 times per week.
- Miles ranges from 10 to 35 miles per week.
- Fitness levels range from 1 to 5.

Distribution Comments:

- Age: Mostly normally distributed with a peak around 30-40 years.
- Income: Right-skewed distribution.
- Usage: Most customers plan to use the treadmill around 3-5 times per week.

- Miles: Left-skewed distribution, with most customers expecting to run less than 20 miles per week.
- Fitness: Majority rate their fitness between 3 and 4.

Relationship Comments:

- Product vs Age: Older customers prefer higher-end models like KP781.
- Product vs Income: Customers with higher incomes tend to purchase the more expensive KP781 model.
- Product vs Miles: Customers planning to run more miles per week tend to choose the higher-end KP781 treadmill.
- Product vs Fitness: Higher fitness levels correlate with the purchase of more advanced treadmill models (KP781).

Other insights:

- Probability of a male customer buying a KP781 treadmill: **31.73%**
- Probability of a Partnered customer buying a KP781 treadmill: **21.50%**

Marginal Probabilities:

Product
KP281 44.444444
KP481 33.333333
KP781 22.222222

MaritalStatus	Partnered	Single
Product		
KP281	0.600	0.400
KP481	0.600	0.400
KP781	0.575	0.425

8. Recommendations

Marketing Strategy:

- KP281: Target younger, lower-income individuals through social media and affordable payment plans.
- KP481: Focus on mid-level fitness enthusiasts with moderate income through fitness magazines and online communities.
- KP781: Aim at high-income, high-fitness level individuals with advanced features through luxury fitness clubs and high-end marketing channels.

Product Development:

- KP281: Introduce affordable, entry-level features attractive to younger customers.
- KP481: Enhance features to appeal to mid-level runners seeking better performance.
- KP781: Innovate with cutting-edge technology and advanced features to attract high-end customers.

Customer Segmentation:

- Develop tailored marketing campaigns based on customer profiles.
- Offer personalized promotions and discounts based on usage patterns and fitness levels.