Colab Notebook link:

INSIGHTS:

Analyzing the dataset:

There are total 180 rows and 9 columns in the dataset.

Columns are: ['Product', 'Age', 'Gender', 'Education', 'MaritalStatus', 'Usage', 'Fitness', 'Income', 'Miles']

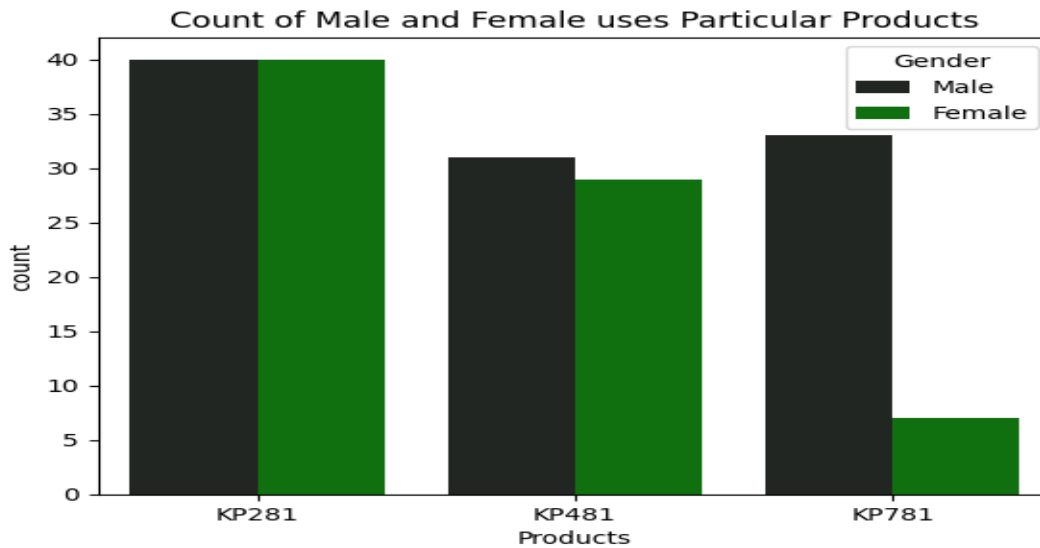Data type of columns: 6 integer type columns, 3 object type columns in the dataset.
Object type columns are: ['Product', 'Gender', 'MaritalStatus']
Integer type columns are: ['Age', 'Education', 'Usage', 'Fitness', 'Income', 'Miles']
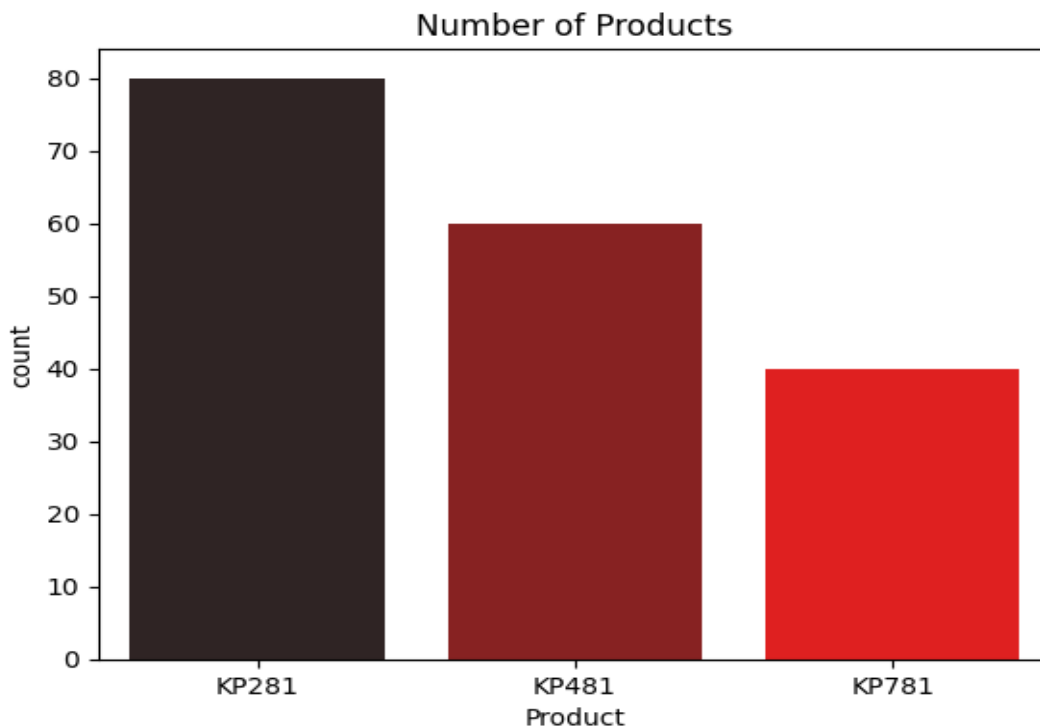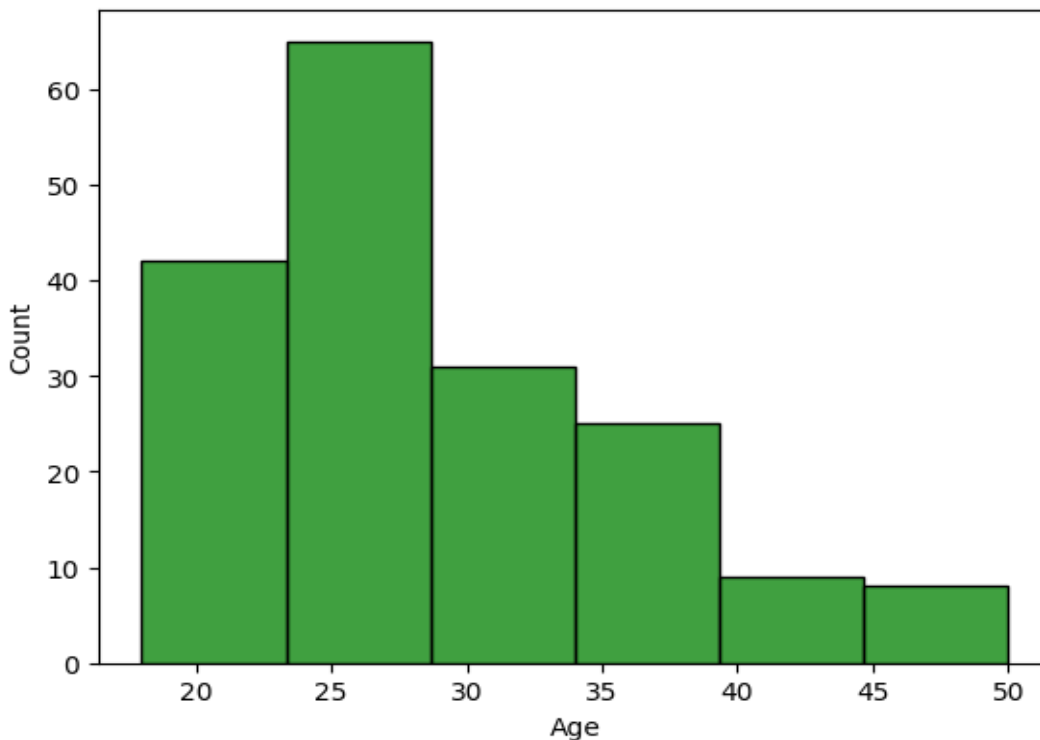

Total number of products is 3 in the dataset.
Products are: KP281, KP481 and KP781. KP281 is most frequent orders, followed by KP481 and KP781 respectively.

Total 104 Males and 76 Females are in the customers list.

Most products purchased by males, females are less interested in the product compared to males.
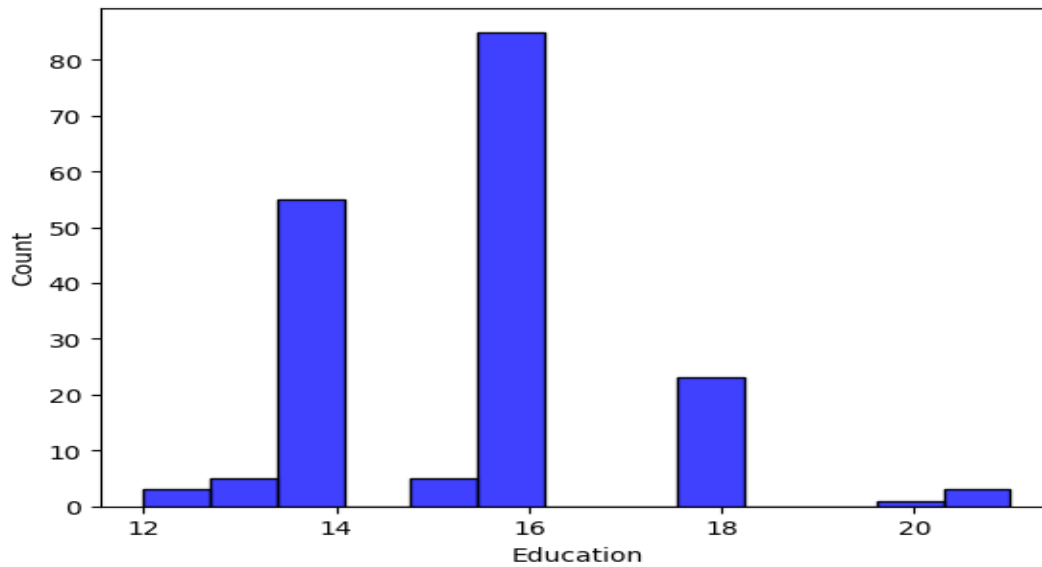

Total no of unique age is: 32

List of unique ages are: [18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 43, 44, 46, 47, 50, 45, 48, 42]

Minimum and Maximum age is 18 and 50 respectively in the dataset.

We can observe from the above chart that, the age group 24 to 29 customers are more likely to use treadmill among other age group customers.
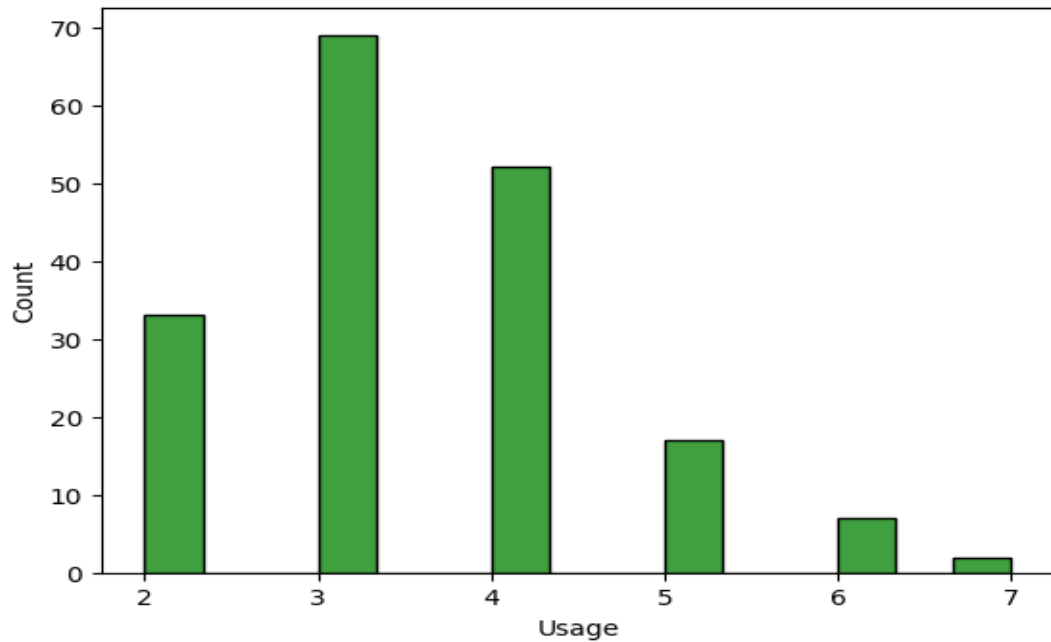
Total no. of unique education is: 8
Educations are: [14, 15, 12, 13, 16, 18, 20, 21]
Highest number of customers have 16 as their Education, 14 is the second highest education among the customers.
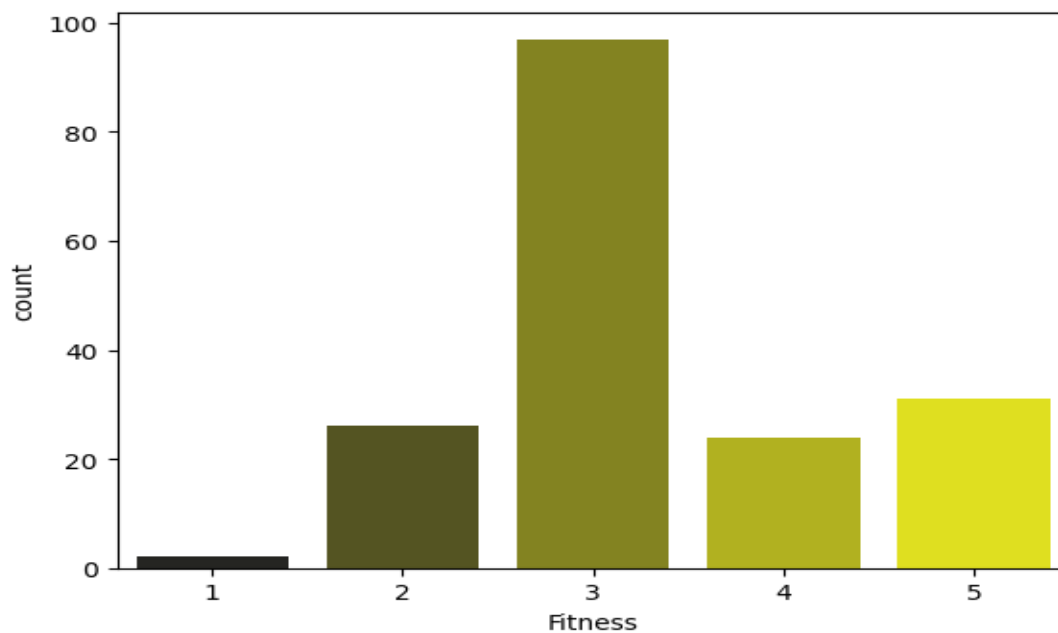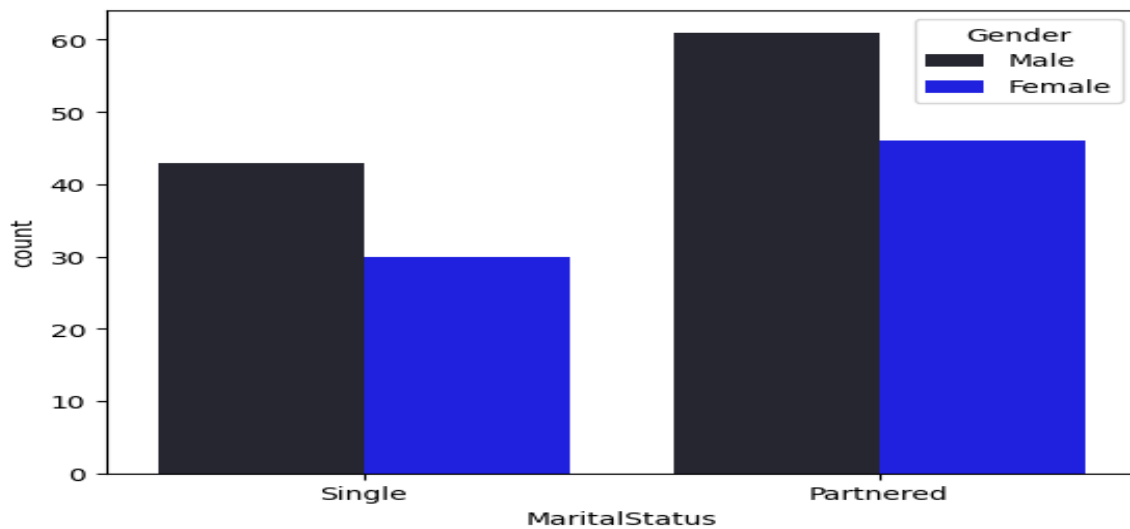

Highest rated Fitness rating is 3.
5 is the second highest rating provided by the customers.

Most customer usage treadmill at least 3 days per week.
4 days and 2 days per week is the second and third highest usage among the customers.
Very few customers use product 7 days per week.

Majority of the customers who have purchased are Partnered.
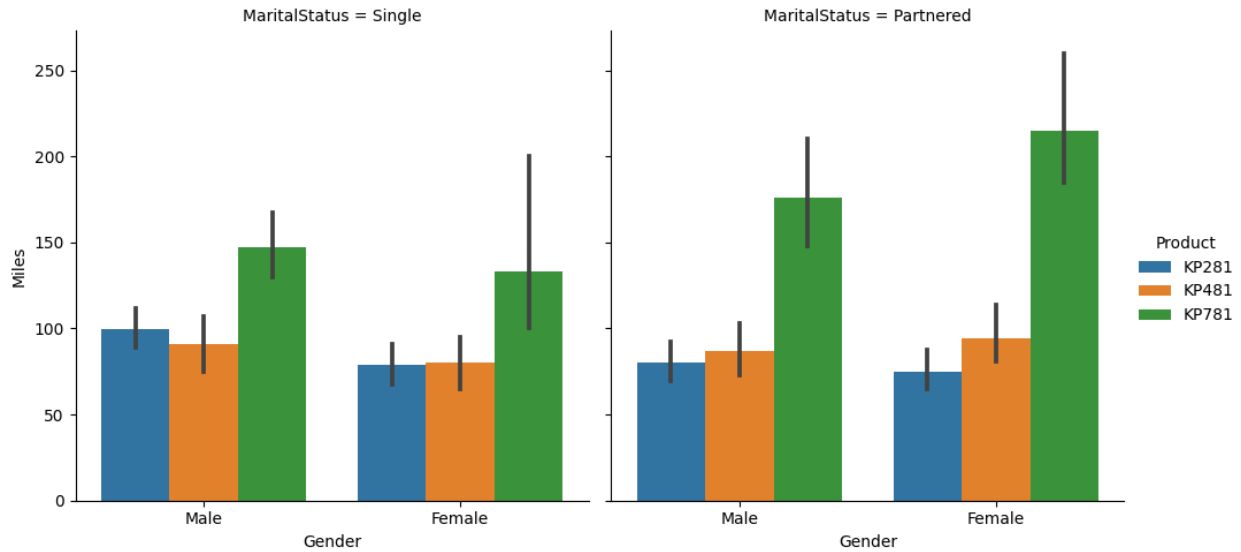
Maximum miles covered by partnered female in KP781 product.
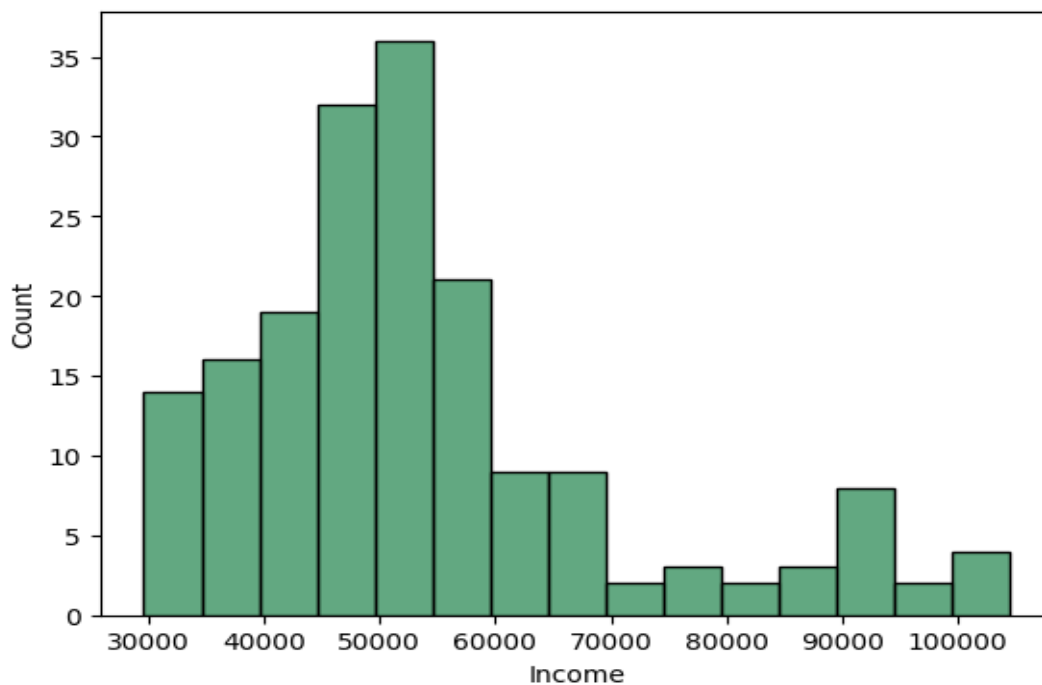
Minimum Income recorded for customer is: 29562
Average Income of the purchased customer is: 53719.58
Highest Income recorded for customer is: 104581
More than 35 customers earn 50-55K per year.
More than 30 customers earn 45-50K per year

Conversion of Categorical attributes to Category:

Convert Fitness Rating to following descriptive categories
1. Poor Shape
2. Bad Shape
3. Average Shape
4. Good Shape
5. Excellent Shape.

Missing Values & Duplicates:

There are no missing values in the dataset.

There are no duplicates in the dataset.

Detect Outliers:

Minimum and Maximum age of users: 18 and 50 respectively. Difference between mean age and median age is (28.79-26) or 2.79.

Minimum and Maximum education of users: 12 and 21 respectively. Difference between mean value and median value of education is (16-15.57) or 0.43.

Minimum and Maximum usage (average number of times the customer plans to use the treadmill each week) of users: 2 and 7 respectively. Difference between mean and median value of usage is (3.45-3) or 0.45.

Minimum and Maximum fitness (Self-rated fitness on a 1-to-5 scale) of users: 1 and 5 respectively. Difference between mean value and median value of fitness rating is (3.31-3) or 0.31.

Minimum and Maximum Income of users: 29562.00 and 104581.00 respectively. Difference between mean value and median value of income is (53719.58-50596.50) or 3123.08.

Minimum and Maximum miles (the average number of miles the customer expects to walk/run each week) of users: 21 and 360 respectively. Difference between mean value and median value of miles is (103.19-94) or 9.19.

Conclusion: Age, Education and Usage are having very few outliers. While Income and Miles are having more outliers.

Better understanding of outliers by box-plots:
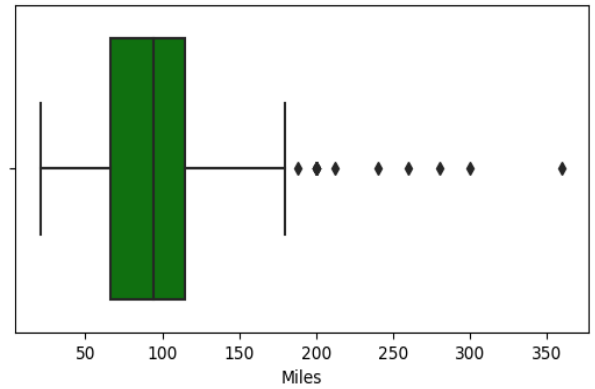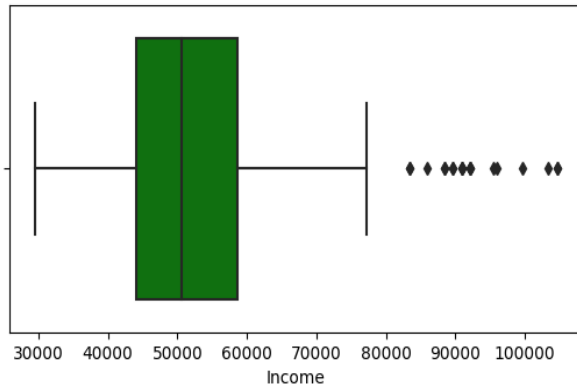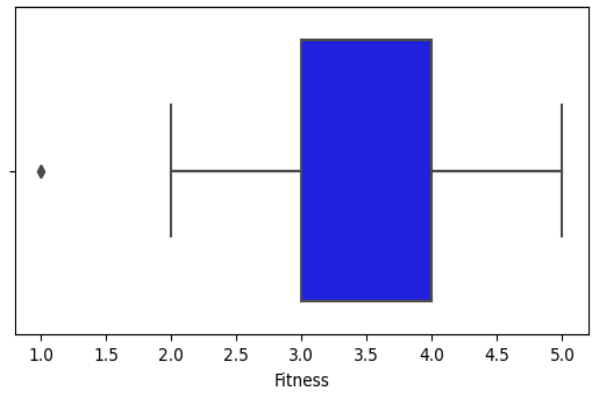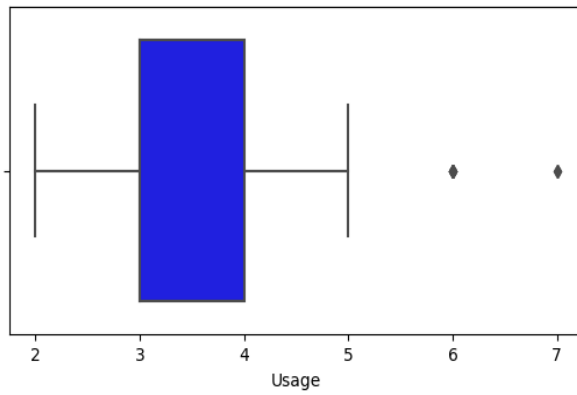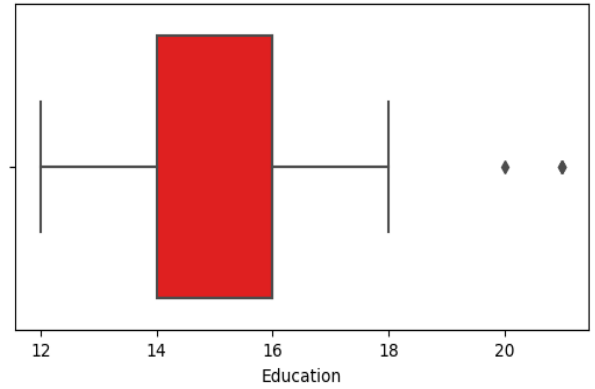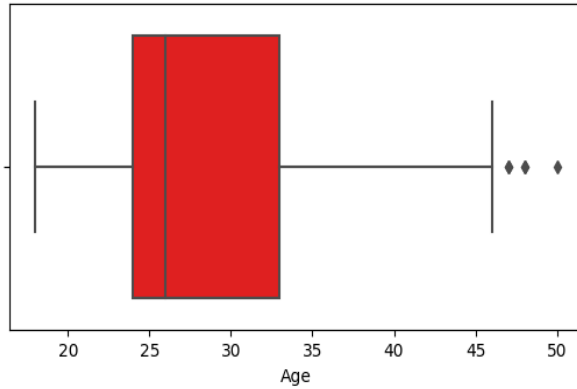
Features like gender, marital status, age, education, usage, fitness, income and miles have any effect on the product purchased:

Gender, Marital Status and Products are categorical variables. So, by the Bivariate(categorical-categorical) Analysis, we can find out the relation between them.
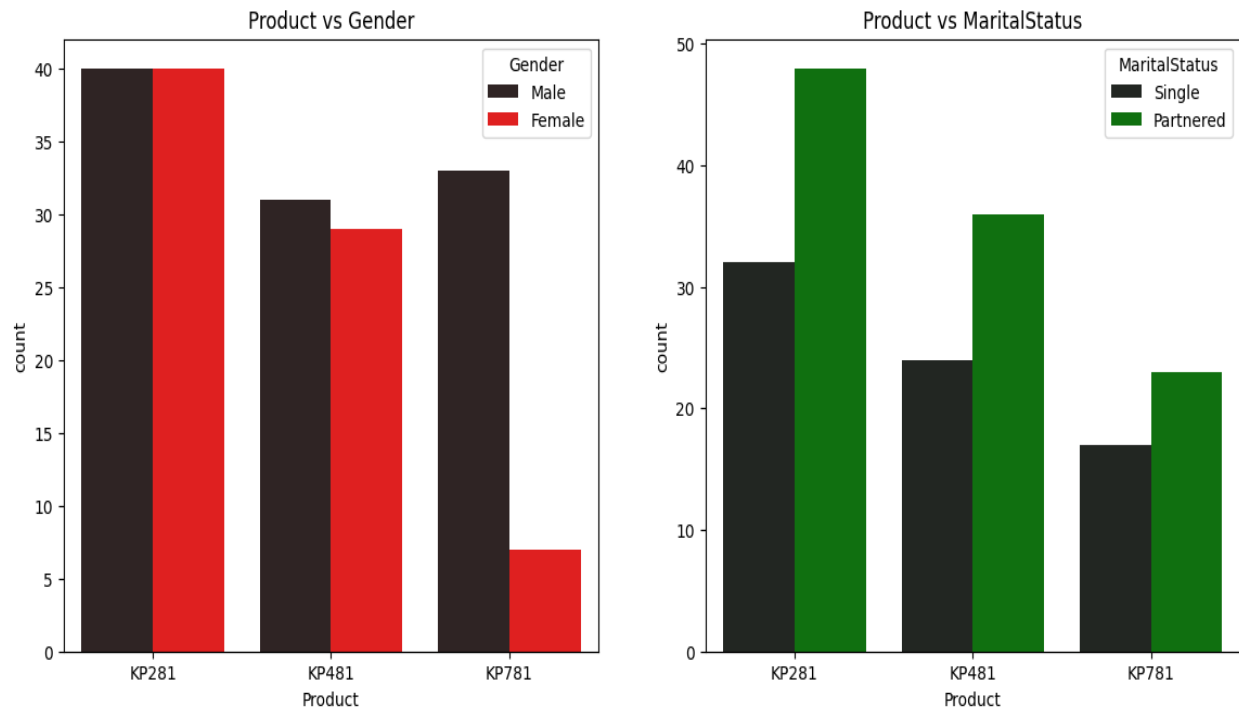
1. **Product vs. Gender**
   - Equal number of males and females have purchased KP281 product and for the product KP481, almost same.
   - But for KP781 product, most of the male customers have purchased the product.

2. **Product vs. Marital Status**
   - Customer, who is partnered, is purchased more products than Single. The difference is greater than 10 for products KP281 and KP481. But the difference is less than 10 for product KP781.

Age, Education, Usage, Fitness, Income and Miles are continuous variable. But Product is categorical variable. So, by the Bivariate(categorical-continuous) Analysis, we can find out the relation between them.

Product vs Age

Product vs. Age
- ➢ Most of the customers are comfortable with KP281 product type.
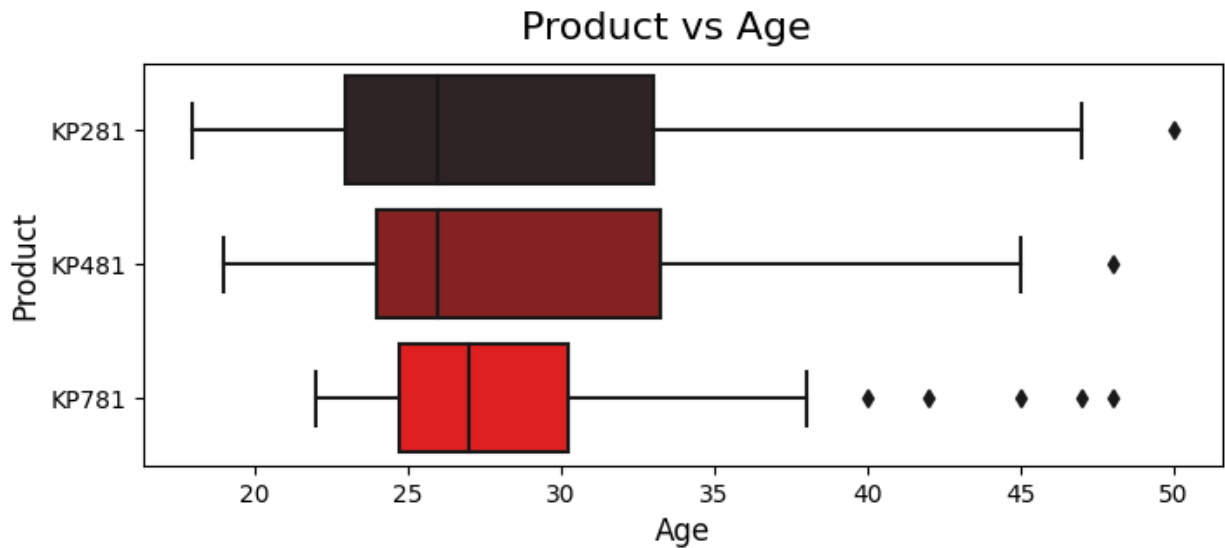- ➢ KP481 is the second highest popular product among the younger side of the customer.
- ➢ Few customers are likely to use KP781 product as compare to other products.
- ➢ And very few customers with age above 40 use product KP781.

Product vs Education
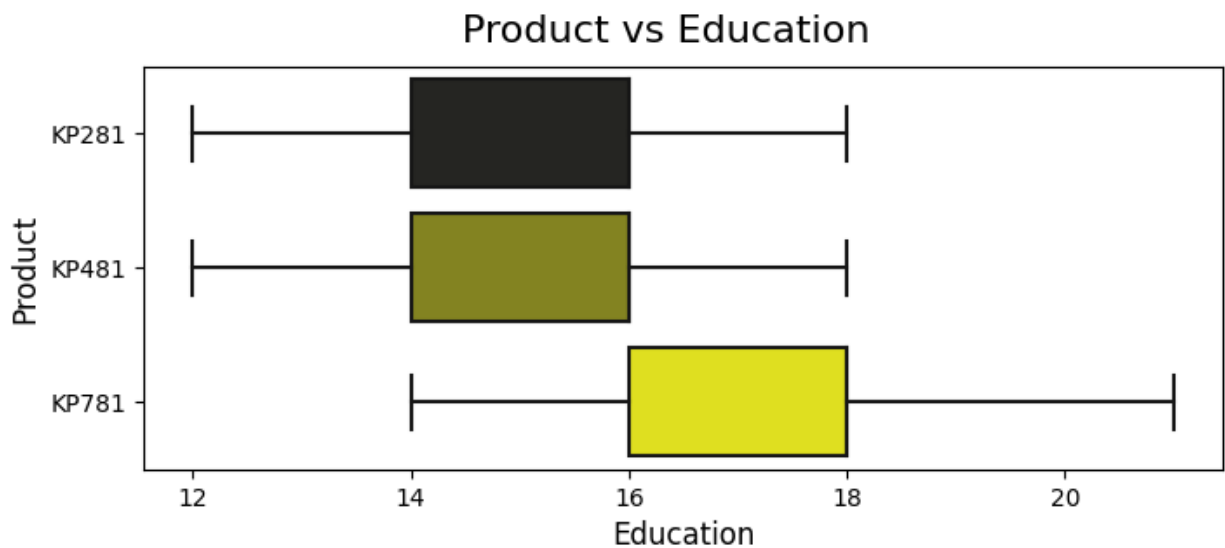
Product vs. Education
- ➢ Customers whose Education is greater than 16, have more chances to purchase the KP781 product.
- ➢ While the customers with Education less than 16 have equal chances of purchasing KP281 or KP481.
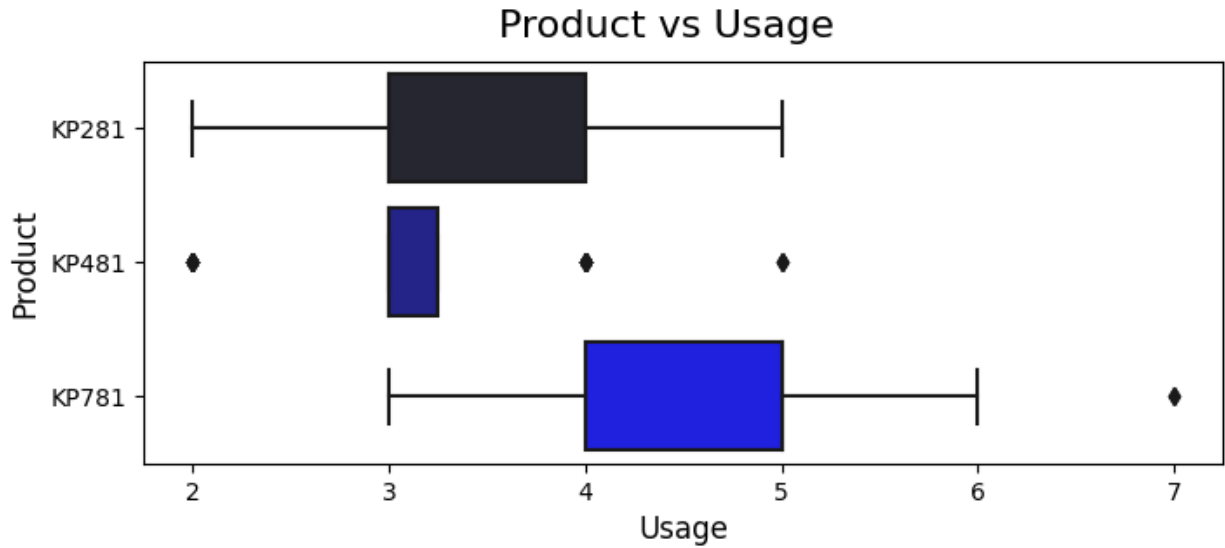
Product vs Usage

Product vs. Usage
➢ Customers, who are planning to use the treadmill greater than 4 times a week, are more likely to use the product KP781.
➢ While the other customers are likely to purchasing KP281 or KP481.
➢ Maximum customers, who are purchasing KP481, are planning to use the treadmill 3 times a week.
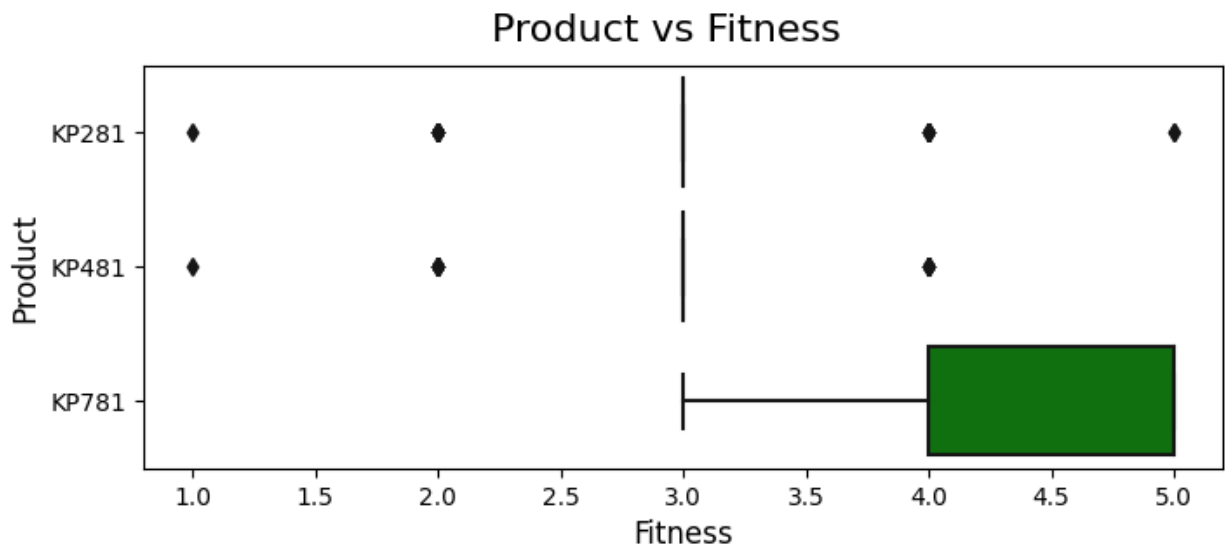
Product vs Fitness

Product vs. Fitness
➢ The more the customer is fit (fitness >= 3), higher the chances of the customer to purchase the KP781 product
➢ KP481 and KP281 product type are scattered across the fitness rating.
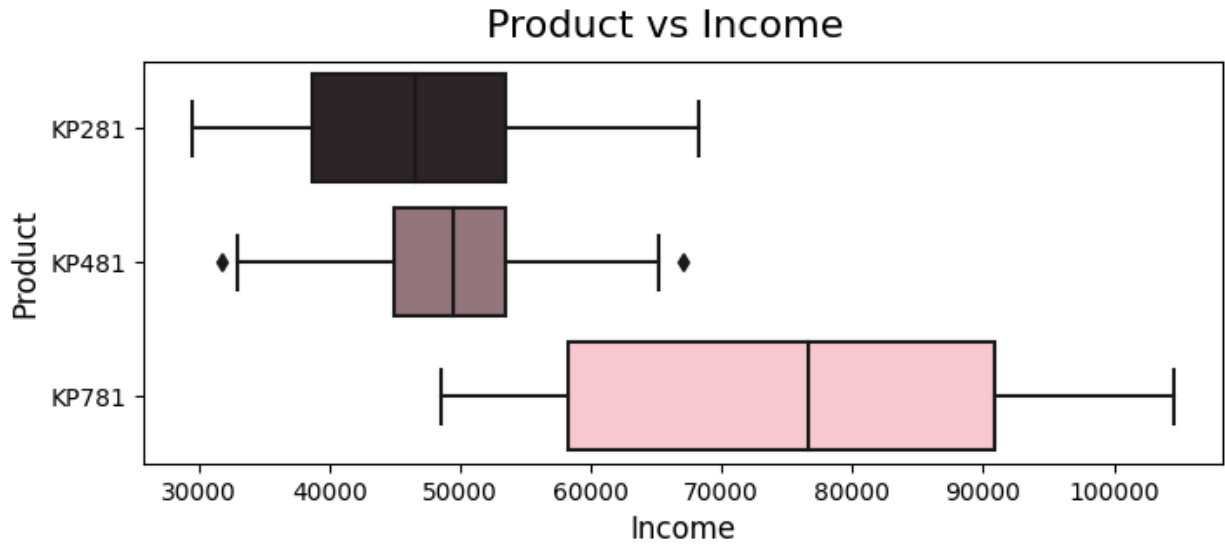
Product vs Income

Product vs. Income

➢ Income greater than or equal to 59000 means higher income customers, higher chances to purchase the KP781 product.

➢ Middle income customers are likely to purchase the KP281 as well as KP481 products.
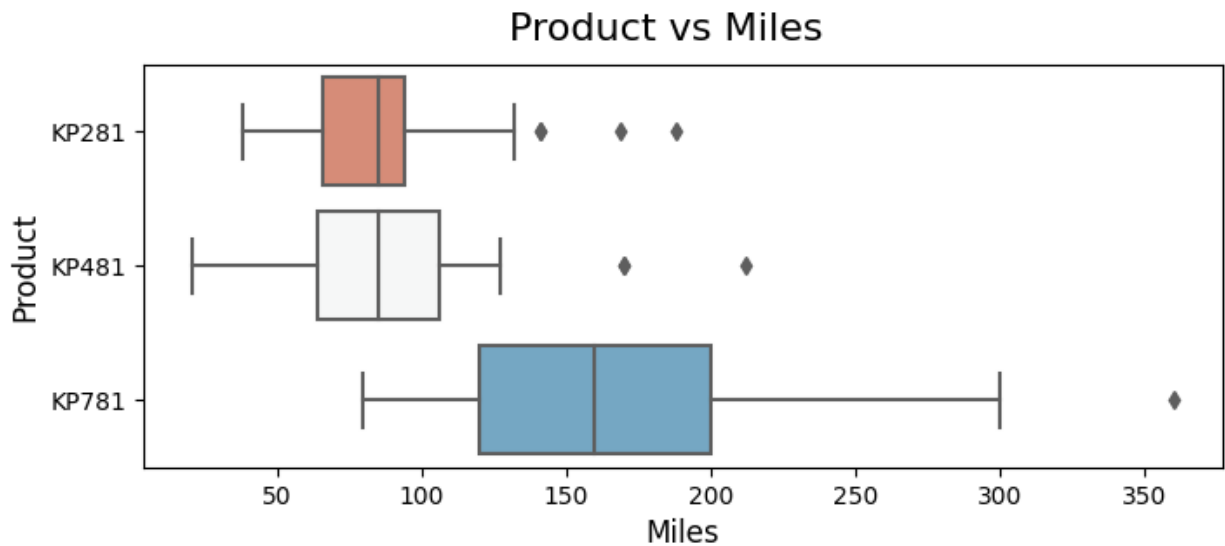
Product vs Miles

Product vs. Miles

➢ If the customer expects to walk/run greater than 120 Miles per week, it is more likely that the customer will purchase KP781 product.

➢ KP481 product is the second highest miles covering product among the customers.

➢ KP281 product customer had covered less distance compared with other two products.

Check correlation among different factors:

Correlation Heatmap Among Different Factors
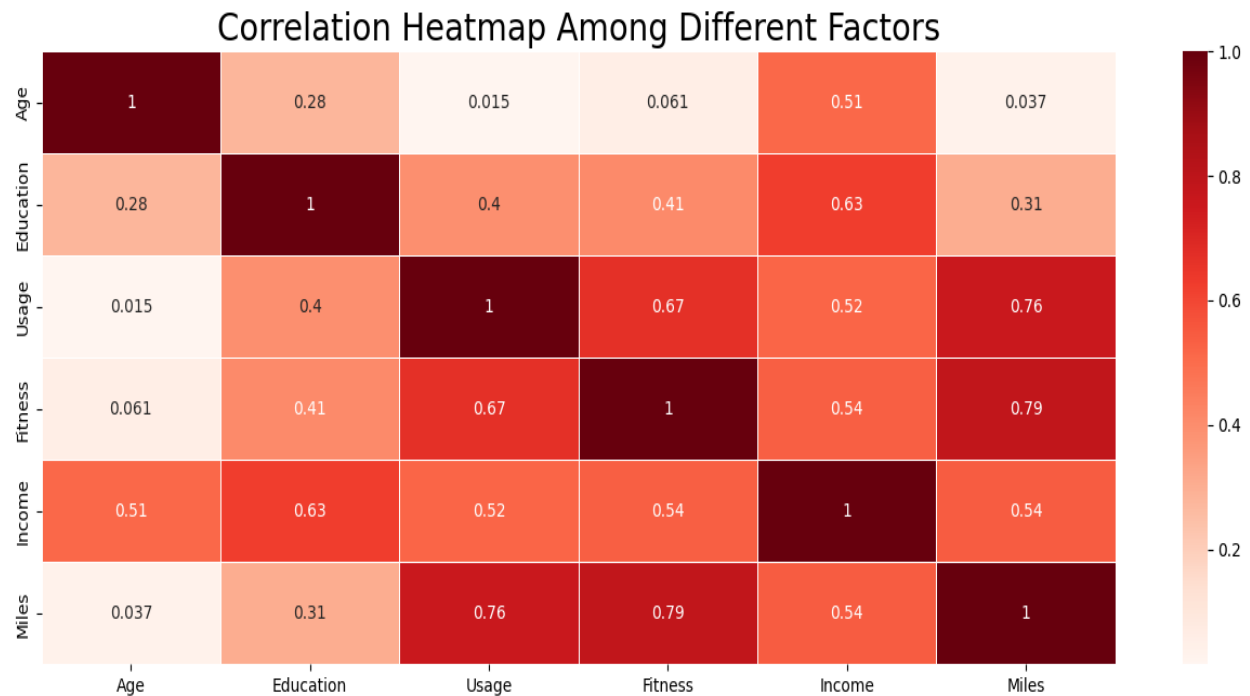
From the above heatmap, we can evaluate the linear relationship between different factors:
- ➢ Correlation between Age and Usage is 0.015, which is lowest compare to others.
- ➢ Correlation between Age and Miles is 0.037, which is 2nd lowest correlation.
- ➢ And the 3rd lowest correlation is 0.061, between Fitness and Age.
- ➢ Correlation between Education and Age is 0.28, Education and Miles is 0.31, Education and Usage is 0.4, Education and Fitness is 0.41.
- ➢ Correlation between Miles and Fitness is 0.79, which is highest in this dataset.
- ➢ 2nd highest correlation is 0.76 in between Miles and Usage.
- ➢ 3rd highest correlation is 0.67 in between Fitness and Usage.
- ➢ Notable point is correlation between Income and other factors individually is greater than 50%.
  - • Correlation between Income and Education is 0.63.
  - • Correlation between Income and Fitness is 0.54
  - • Correlation between Income and Miles is 0.54
  - • Correlation between Income and Usage is 0.52
  - • Correlation between Income and Age is 0.51

Marginal Probability:

Product purchased percentage as per different Age group
- ➢ Grouping the age in 4 different categories as per following range:
    - Age between (0 to 21) is 'Teen' category
    - Age between (22 to 35) is 'Adult' category
    - Age between (36 to 45) is 'Middle Aged' category
    - Age between (46 to 60) is 'Elder' category.
- ➢ The probability of purchasing the treadmill by Adult age category is 75% of total.
- ➢ Other than this, the probability of purchasing KP281 product by Adult is 31.11%.
- ➢ KP481 product purchasing probability by Adult is 25%.
- ➢ And KP781 product purchasing probability by Adult is 18.89%.

Product purchased percentage as per Gender-wise
- ➢ Probability of Male customer purchasing any product is : 57.78 %
- ➢ Probability of Female customer purchasing any product is : 42.22 %
- ➢ The probability of purchasing KP281 product by Male and Female is equal.
- ➢ Also, for KP481 product, the probability of purchasing by Male and Female is nearly equal.
- ➢ And for KP781 product, the probability of purchasing by Male (18.33%) is greater than Female (3.89%).

Product Purchased percentage as per Fitness category
- ➢ Probability of purchasing by Average shape category customer is higher for products KP281 (30%) and for KP481 (21.67%) compare to other fitness category customers.
- ➢ Excellent shape category customers have more probability to purchase KP781 product than other products.

Product purchased percentage as per Marital Status
- ➢ Probability of purchasing the treadmill by Partnered category is higher than Single for any product.

Conditional Probability:

Among Age group, the probability of Product purchasing
- ➢ P(KP281 | Teen) = 58.82%,   P(KP481 | Teen) = 41.18%,   P(KP781 | Teen) = 0%.
- ➢ P(KP281 | Adult) = 41.18%,   P(KP481 | Adult) = 33.33%,   P(KP781 | Adult) = 25.19%.
- ➢ P(KP281 | Middle Aged)= 50%,   P(KP481 | Middle Aged)= 31.82%,
    P(KP781 | Middle Aged)= 18.18%.
- ➢ P(KP281 | Elder) = 50%,   P(KP481 | Elder)= 16.67%,   P(KP781 | Elder)= 33.33%.

Among Gender, the probability of Product purchasing

➢ P(KP281 | Female) = 52.63%, P(KP481 | Female) = 38.16%, P(KP781 | Female) = 9.21%.
➢ P(KP281 | Male) = 38.46%, P(KP481 | Male) = 29.81%, P(KP781 | Male) = 31.73%.

Among Fitness category, the probability of Product purchasing

➢ P(KP281 | Average Shape) = 55.67%, P(KP481 | Average Shape) = 40.21%, P(KP781 | Average Shape) = 4.12%.
➢ P(KP281 | Bad Shape) = 53.85%, P(KP481 | Bad Shape) = 46.15%, P(KP781 | Bad Shape) = 0%.
➢ P(KP281 | Excellent Shape) = 6.45%, P(KP481 | Excellent Shape) = 0%, P(KP781 | Excellent Shape) = 93.55%.
➢ P(KP281 | Good Shape) = 37.50%, P(KP481 | Good Shape) = 33.33%, P(KP781 | Good Shape) = 29.17%.
➢ P(KP281 | Poor Shape) = 50%, P(KP481 | Poor Shape) = 50%, P(KP781 | Poor Shape) = 0%.

Among Marital Status, the probability of Product purchasing

➢ P(KP281 | Partnered) = 44.86%, P(KP481 | Partnered) = 33.64%, P(KP781 | Partnered) = 21.50%.
➢ P(KP281 | Single) = 43.84%, P(KP481 | Single) = 32.88%, P(KP781 | Single) = 23.29%.

Customer Profiling:

Customer Profiling based on product categories:

KP281

- KP281 is maximum selling product and also it is easily affordable entry level product.
- Maximum entry level customers are likely to purchase this product.
- KP281 is popular among Male as well as Female customers.
- Partnered female as well as partnered male customers purchased this product more than single female or single male respectively.
- Younger to Elder beginner level customers prefers this product.
- Product is mostly used by customers, 3 to 4 times a week.
- Average distance covered in this product is around 70 to 90 miles per week.
- As per fitness rating, Average shape customers are likely to purchase this product.
- Income ranges between 35k to 55 k have preferred to purchase this product.

KP481

- KP481 is an intermediate level product and also second most popular product among the customers.
- Popularity of KP481 among Male as well as Female customers is nearly same.
- More Partnered customers prefer this product.
- Probability of Female customer buying KP481 is significantly higher than male.
- The age range of KP481 treadmill customers is roughly between 24-34 years.
- Three different age groups prefer this product - Teen, Adult and middle aged
- Product is mostly used by customers, 3 times a week.
- Average distance covered in this product is around 70 to 110 miles per week.
- KP481 is also the second most distance travelled model.
- Fitness Level of this product user varies from Bad to Average Shape depending on their usage.
- Average Income of the customer who buys KP481 is 49K.

KP781

- KP781 is the advance level product and mostly the higher income persons prefer this product.
- Due to the High Price & being the advanced type, customer prefers less of this product.
- More male customers preferred this product.
- Male single as well as male partnered prefer to purchase this product.
- Elder aged customers are likely to use this product.
- Product is mostly used by customers, 4 to 5 times a week.
- Customers use this product mainly to cover more distance.
- Average distance covered in this product is around 120 to 200 miles per week.
- Customers who use this product have rated excelled shape as fitness rating.
- Average Income of KP781 buyers are over 75K per annum.
- Single male, who have higher income with excellent shape more likely to purchase this product.
- This product is preferred by the customer where the correlation between Education and Income is high.

Colab Notebook link:
https://colab.research.google.com/drive/17O48U4d4iyVmv0OjXwnAfP65skREd1Dd?usp=sharing