Aerofit Case Study

Descriptive Statistics

Jupyter Notebook link:
https://drive.google.com/file/d/13kGIBA_8faF5bzmt2vb0SGTBTcDje2Ej/view?usp=sharing

1. Import the dataset and do usual data analysis steps like checking the structure & characteristics of the dataset

a) Columns in the Dataframe
['Product', 'Age', 'Gender', 'Education', 'MaritalStatus', 'Usage', 'Fitness', 'Income', 'Miles']

b) Details of the datatype in the columns

```
Product          object
Age               int64
Gender           object
Education         int64
MaritalStatus    object
Usage             int64
Fitness           int64
Income            int64
Miles             int64
```

c) The Dataframe contains 180 rows and 9 columns

d) No of unique values in the Dataframe:
Product : 3
Age : 32
Gender : 2
Education : 8
MaritalStatus : 2
Usage : 6
Fitness : 5
Income : 62
Miles : 37

e) No of the values in each column
Product :
{'KP281': 80, 'KP481': 60, 'KP781': 40}
Age :
{25: 25, 23: 18, 24: 12, 26: 12, 28: 9, 35: 8, 33: 8, 30: 7, 38: 7, 21: 7, 22: 7, 27: 7, 31: 6, 34: 6, 29: 6, 20: 5, 40: 5, 32: 4, 19: 4, 48: 2, 37: 2, 45: 2, 47: 2, 46: 1, 50: 1, 18: 1, 44: 1, 43: 1, 41: 1, 39: 1, 36: 1, 42: 1}
Gender :
{'Male': 104, 'Female': 76}
Education :
{16: 85, 14: 55, 18: 23, 15: 5, 13: 5, 12: 3, 21: 3, 20: 1}

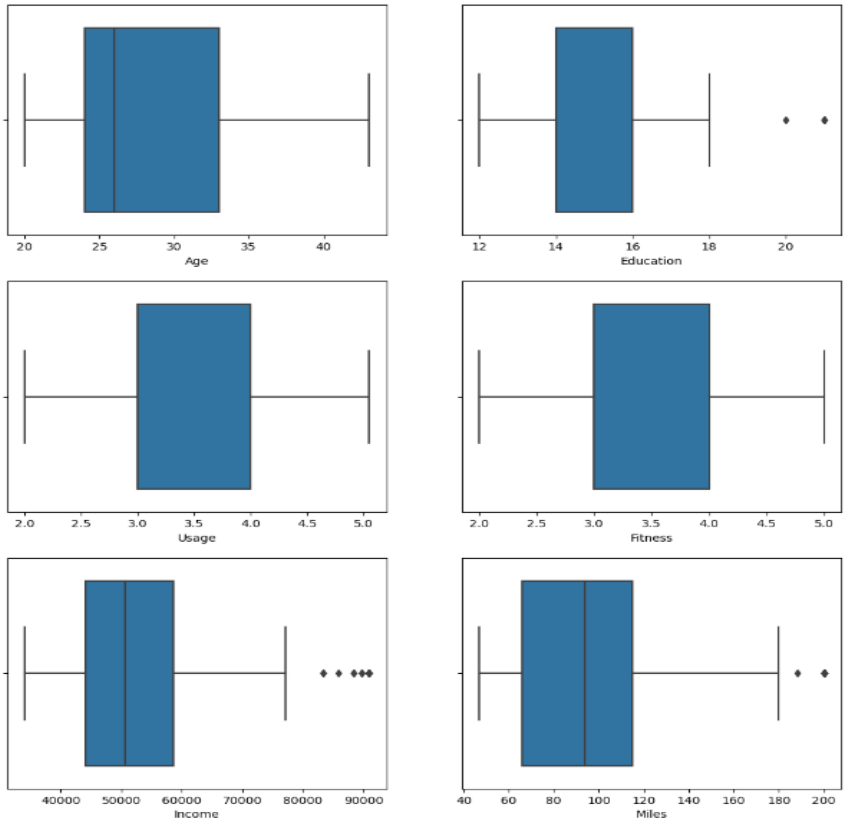
MaritalStatus :
{'Partnered': 107, 'Single': 73}

Usage :
{3: 69, 4: 52, 2: 33, 5: 17, 6: 7, 7: 2}
Fitness :
{3: 97, 5: 31, 2: 26, 4: 24, 1: 2}
Income :
{45480: 14, 52302: 9, 46617: 8, 54576: 8, 53439: 8, 50028: 7, 51165: 7, 40932: 6, 48891: 5, 32973: 5, 35247: 5, 38658: 5, 34110: 5, 43206: 5, 57987: 4, 44343: 4, 36384: 4, 64809: 3, 90886: 3, 60261: 3, 92131: 3, 59124: 3, 88396: 2, 49801: 2, 83416: 2, 64741: 2, 61006: 2, 48556: 2, 31836: 2, 47754: 2, 89641: 2, 61398: 2, 56850: 2, 67083: 2, 42069: 2, 39795: 2, 37521: 2, 104581: 2, 99601: 1, 103336: 1, 85906: 1, 52290: 1, 77191: 1, 74701: 1, 95866: 1, 69721: 1, 29562: 1, 75946: 1, 70966: 1, 62251: 1, 52291: 1, 57271: 1, 53536: 1, 58516: 1, 54781: 1, 48658: 1, 62535: 1, 65220: 1, 55713: 1, 68220: 1, 30699: 1, 95508: 1}
Miles :
{85: 27, 95: 12, 66: 10, 75: 10, 47: 9, 106: 9, 94: 8, 113: 8, 53: 7, 100: 7, 180: 6, 200: 6, 56: 6, 64: 6, 127: 5, 160: 5, 42: 4, 150: 4, 38: 3, 74: 3, 170: 3, 120: 3, 103: 3, 132: 2, 141: 2, 280: 1, 260: 1, 300: 1, 240: 1, 112: 1, 212: 1, 80: 1, 140: 1, 21: 1, 169: 1, 188: 1, 360: 1}

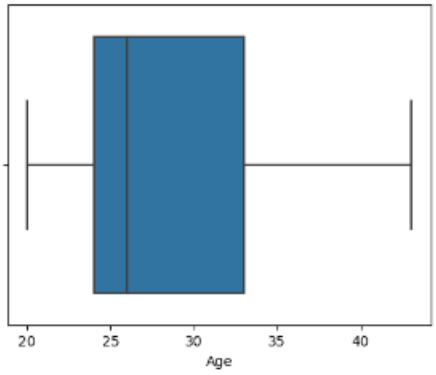
2. Detect Outliers

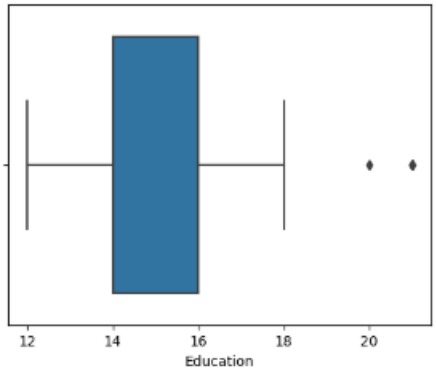
1. Find the outliers for every continuous variable in the dataset

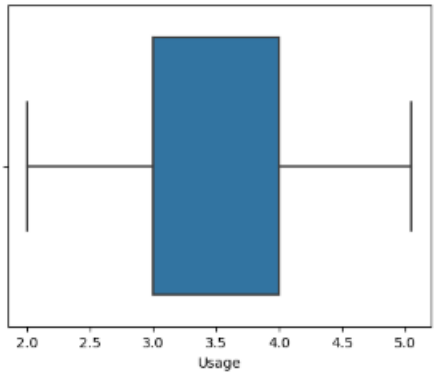
   The continuous variable in the dataset is ‘Age’, ‘Usage’, ’Fitness’, ‘Income’ and ‘Miles’

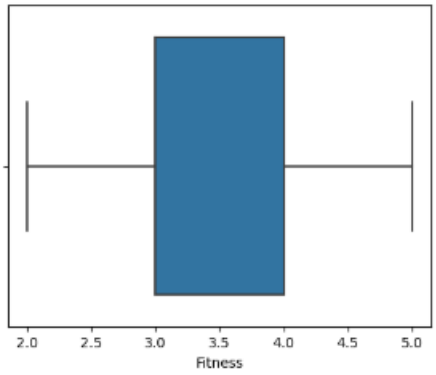


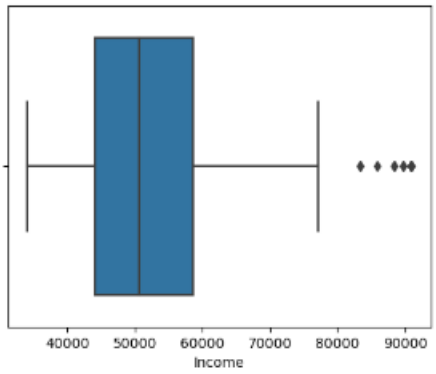
2. Remove/clip the data between the 5 percentile and 95 percentiles.

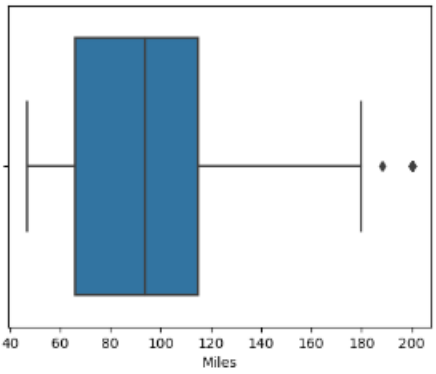




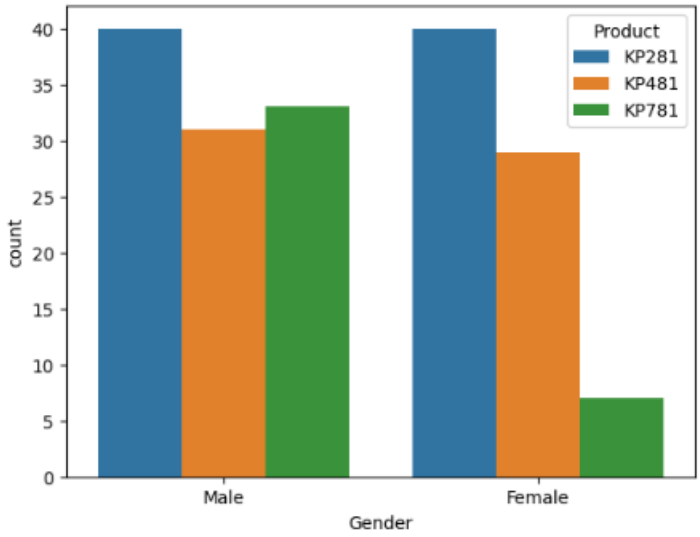


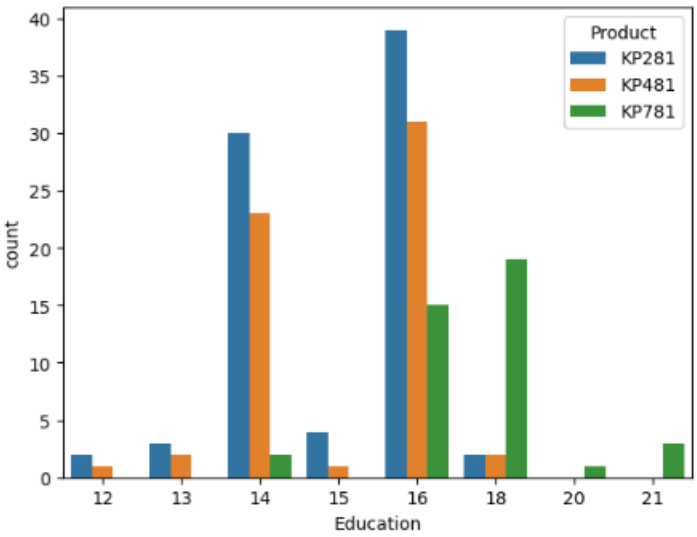


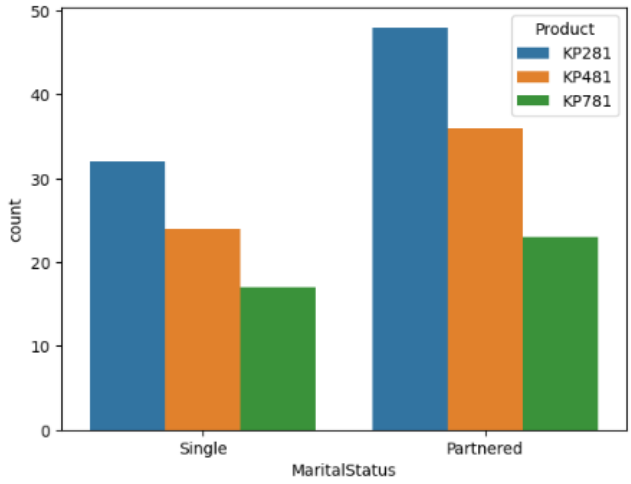




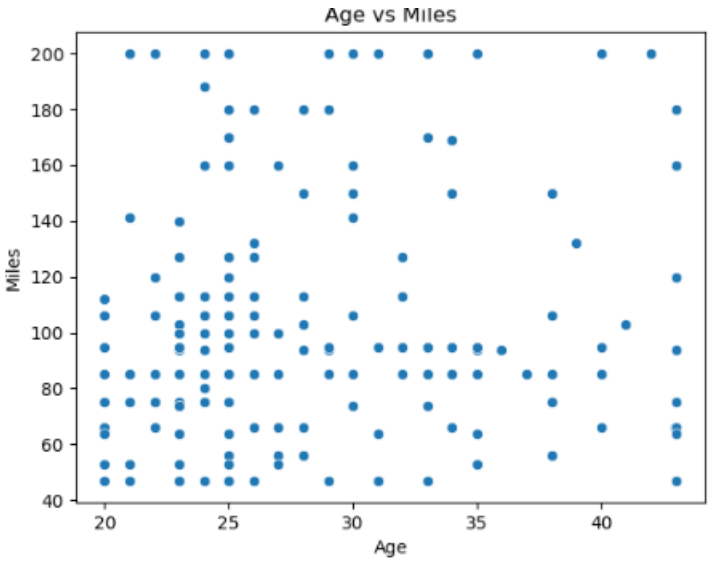
3. **Check if features like marital status, Gender, and age have any effect on the product purchased**
    a. Find if there is any relationship between the categorical variables and the output variable in the data.

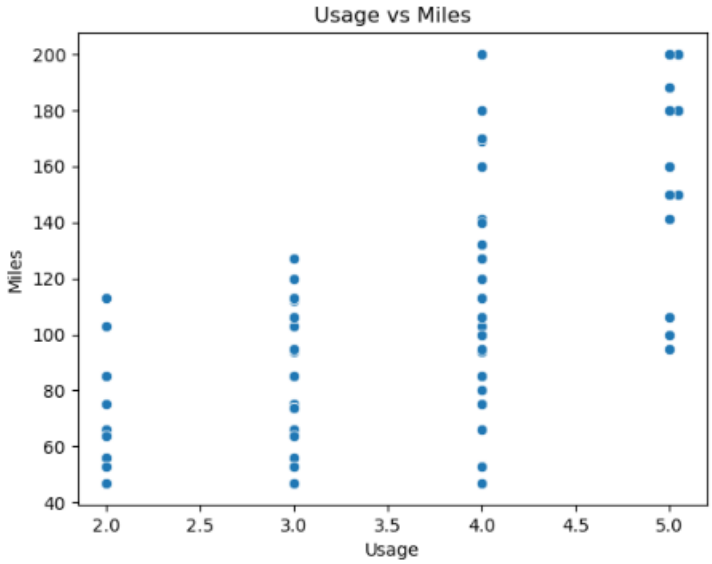


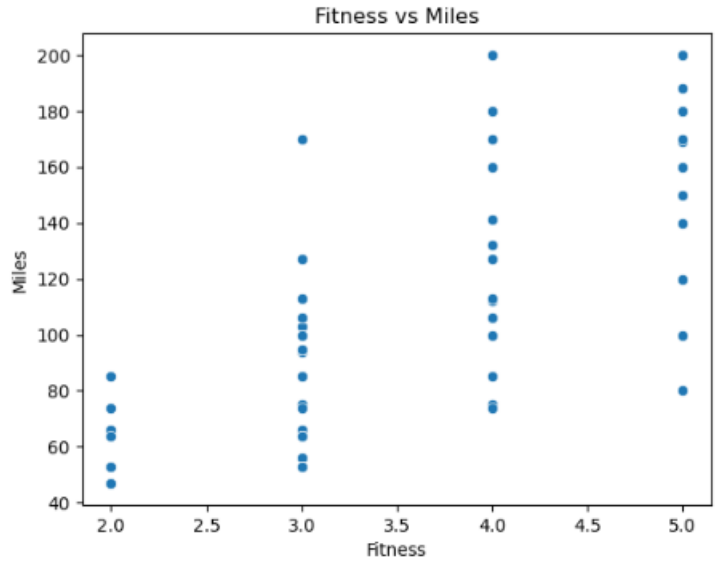


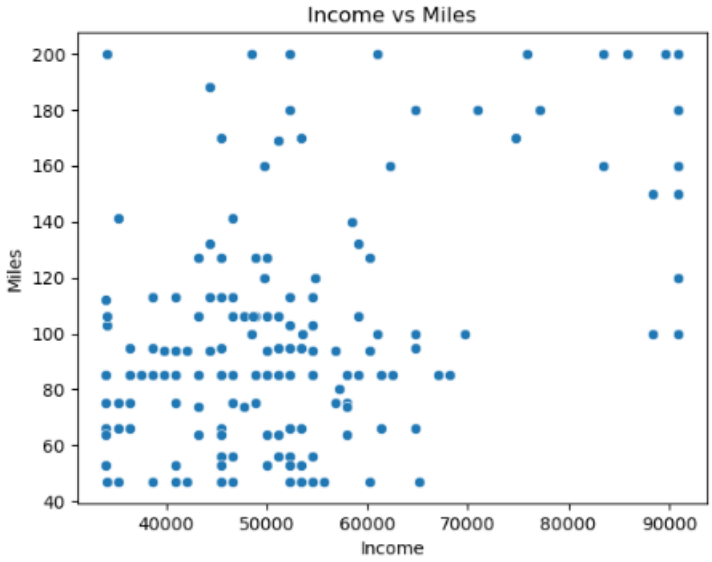


b. Find if there is any relationship between the continuous variables and the output variable in the data.









4. **Representing the Probability**
    - **a.** Find the marginal probability (what percent of customers have purchased KP281, KP481, or KP781)

    - **b.** Find the probability that the customer buys a product based on each column.
    - **c.** Find the probability that an event occurs given that another event has occurred. (Example: given that a customer is female, what is the probability she'll purchase a KP481)
        i. Conditional Probability of Buying KP481 given Female: 0.43859649122807026
        ii. Conditional Probability of Buying KP481 given Male: 0.3205128205128205

5. **Check the correlation among different factors**
   Find the correlation between the given features in the table.

6. **Customer profiling and recommendation**
   i. Make customer profilings for each and every product.

	Product	Age	Gender	Income
0	KP281	28.427500	Female	46584.31125
1	KP481	28.801667	Male	49046.60750
2	KP781	28.828750	Male	73908.28125

ii. Write a detailed recommendation from the analysis that you have done.

Insights:

- From the boxplots it is quite clear that:
  Age, Education and Usage are having very few outliers.
  While Income and Miles are having more outliers.
- KP281 is the most frequent bought product
- Male customers are more compared to Females
- More number of partnered customers buying the products
- The people having the education of 16 years buying the more products
- 57.78% Customers are Male.
- 59.44% Customers are Partnered.
- Most sold product KP281, its 44.44% of sales out of overall Aerofit Treadmill sale.
- Minimum & Maximum age of the person is 18 & 50, mean is 28.79 and 75% of persons have age less than or equal to 33.

Recommendations:

The recommendations based on the customer profiling

- The product KP281 is bought by the Female customer having the average age of 28 and having the salary of average salary of 46000.
- The product KP481 is bought by the Male customer having the average age of 29 and having the salary of average salary of 49000.
- The product KP781 is bought by the Male customer having the average age of 29 and having the salary of average salary of 74000.
- So, the respective products may be advertised for the person having the similar characteristics as in profiling.
- The extra offers like no cost EMI, offers can be included so that other people also buy the product.
- Hence, the product KP281 is the most bought product in the beginner level, so the product is likely to be sold more than the KP781. We can recommend KP781 only to the people having the similar characteristics as in the customer profiling.