

Customer Profile Report for Marketing

Author: Mishal Fatima

Contents

| | |
|---------------------------------------------------|----|
| List of Figures | 3 |
| List of Tables | 3 |
| Introduction | 4 |
| Data Cleaning & Exploration | 5 |
| Outlier Analysis | 6 |
| Customer Profile/Characterization..... | 8 |
| Purchase Trends driven by Customer Behavior | 12 |
| Pairplot Analysis | 15 |
| Correlation Matrix | 17 |
| Actionable Insights and Recommendations | 18 |

List of Figures

| | |
|-------------------------------------------------------------------------------------|-----------|
| Figure 1: Box Plot for Outlier Analysis..... | 6 |
| Figure 2: Client breakdown with respect to Gender and Marital Status | 8 |
| Figure 3: Client breakdown by Age, Education, Fitness Level and Income | 10 |
| Figure 4: Sales Trend | 11 |
| Figure 5: Product Purchase Trend by Gender | 12 |
| Figure 6: Product Purchase Trend by Marital Status | 12 |
| Figure 7: Purchase Trends by Age | 13 |
| Figure 8: Purchase Trends by Customer Fitness | 13 |
| Figure 9: Purchase Trends by Income Status..... | 14 |
| Figure 10: Pairplot for Multivariate Analysis..... | 15 |
| Figure 11: Pair Plot categorized by Product Type | 16 |
| Figure 12: Correlation Matrix | 17 |

List of Tables

| | |
|-------------------------------------------------------------|----------|
| Table 1: Outlier Summary by Tukey's IQR Method | 6 |
|-------------------------------------------------------------|----------|

Introduction

AeroFit is a fitness company that sells treadmills to health-conscious individuals. They have three major products (treadmills): KP281 valued at 1500\$, KP481 valued at 1750\$ and their most prestigious KP781 which is valued at 2500\$. The company aims to understand its audience to better cater to their needs and requirements out of their products.

Hence, a customer profiling is done to better understand the audience and their preferences when selecting AeroFit treadmills. The data used is the company collected data of individuals who have purchased from AeroFit stores in the past three months. Categorization of the client is done based on their gender, age, education, marital status, fitness level and income. Furthermore, their expectations on how many days they will use the treadmill and how many miles they are planning to walk/run is also recorded.

The impact of all of the above is then used to analyze how these characteristics lead them to the selection of the treadmill.

The report discusses in detail the data exploration & cleaning, outlier analysis, insights into the audience, their product bias and finally the steps that would help the marketing team in better recommendation of company product to the customer.

Data Cleaning & Exploration

The data provided was checked & cleaned prior to its use to ensure the data integrity and accuracy of results. After importing necessary libraries and dataset, the data exploration process is explained in following steps:

1. The shape of the dataframe is checked i.e. the number of rows and columns (180x9).
2. Datatypes for each column is checked to be consistent.
3. Data set does not have any duplicate values.
4. There are no missing values in each column.
5. The statistical summary of numerical and categorical data is analyzed to rule out any discrepancy in data.
6. Unique values for both numerical and categorical data and counts for the latter are checked to make sure there is no incorrect data.

Overall, the data seemed to be accurate without requiring much cleaning at all. Bins were used during data analysis but were not made a part of the dataframe.

Outlier Analysis

Boxplots and Tukeys IQR is used to identify the outliers in the dataframe for the numerical data. Figure 4 shows the data distribution and deviation. Whereas Table 1 gives the outlier values for Age, Education and Income.

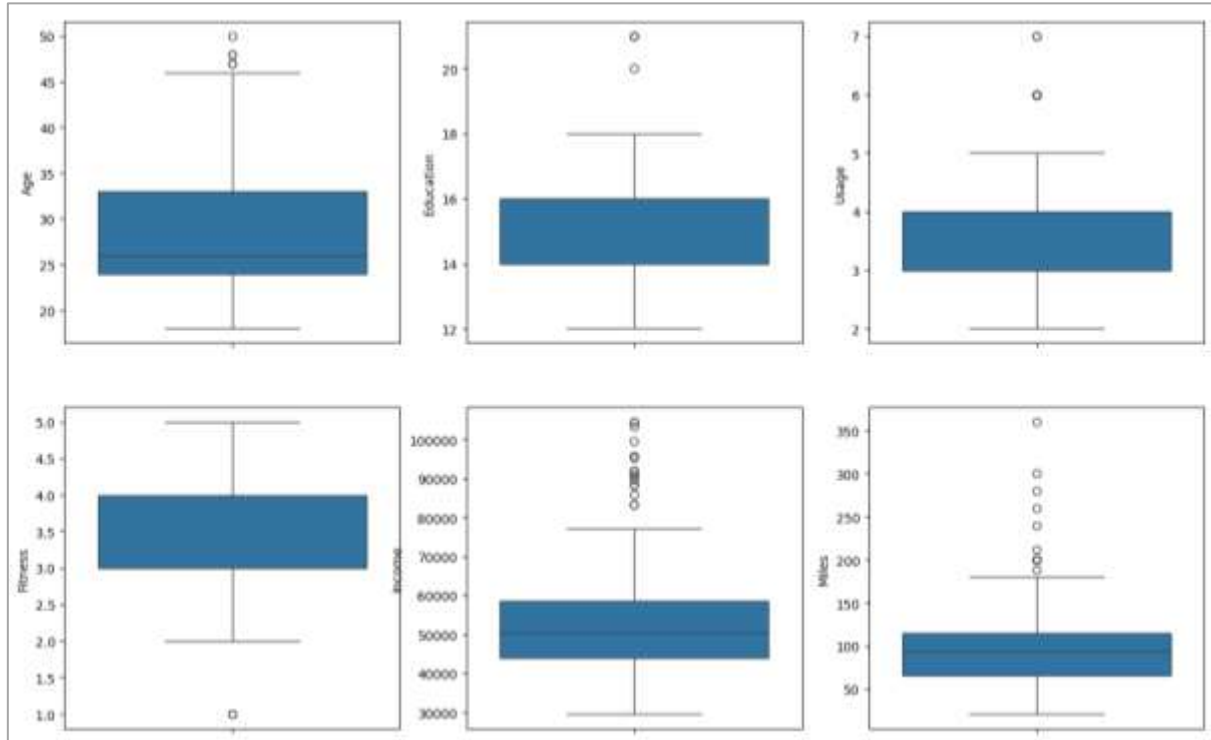


Figure 1: Box Plot for Outlier Analysis

Table 1: Outlier Summary by Tukey's IQR Method

| | |
|-----------------------|---------|
| Age: | |
| Lower Bound | 10.5 |
| Upper Bound | 46.5 |
| Outliers in Age | 5 |
| Education: | |
| Lower Bound | 11.0 |
| Upper Bound | 19.0 |
| Outliers in Education | 4 |
| Income: | |
| Lower Bound | 22144\$ |
| Upper Bound | 80581\$ |
| Outliers in Income | 19 |

AeroFit

No action was taken to remove or discard the outlier data as it would have impacted the insights drawn from the results which would make it impossible to recommend treadmill to a customer. Furthermore, the size of the dataset is too small to categorize these outliers as having inconsequential affect.

Customer Profile/Characterization

Following figures show the general demographic of the clientele of AeroFit. Figure 1 and 2 show the count of each parameter whereas Figure represents their percentages. It can be summarized that:

1. Most of the customer-base of AeroFit prefers KP281 with KP481 having second highest demand leaving KP781 behind. Which can also be seen from Figure 3.
2. 58% of the buyers are male and 59% of the buyer are partnered.
3. Majority of the clientele 60.4% and 26.7% approx. is in their twenties and thirties respectively.

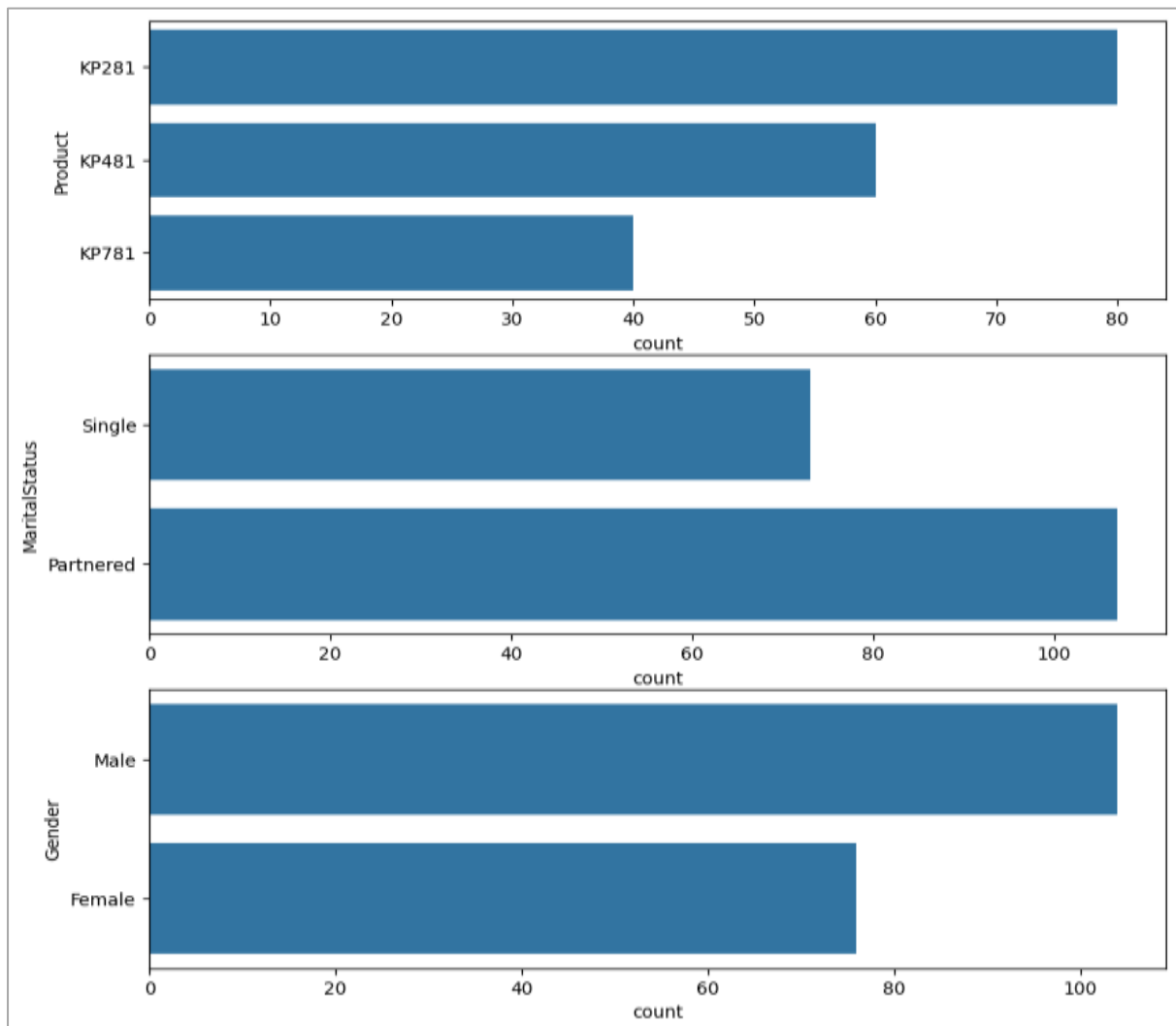


Figure 2: Client breakdown with respect to Gender and Marital Status

AeroFit

4. AeroFit customers have 16, 14 and 18 years of education in order of their counts. With 47% having 16 years of education and 30% having 14 years of education. In short most of the clients are highly educated.
5. According to the data, most individuals claim that they are at fitness level 3, where 1 means poor shape and 5 means good shape.
6. Annual Income for majority of AeroFit customers ranges from 30,000\$ to 70,000\$. Hence medium income earners make up 74% of the client-base.
7. Finally, most of the treadmill buyers claim that they will use it 2-4 days and cover 40-200 miles each week.

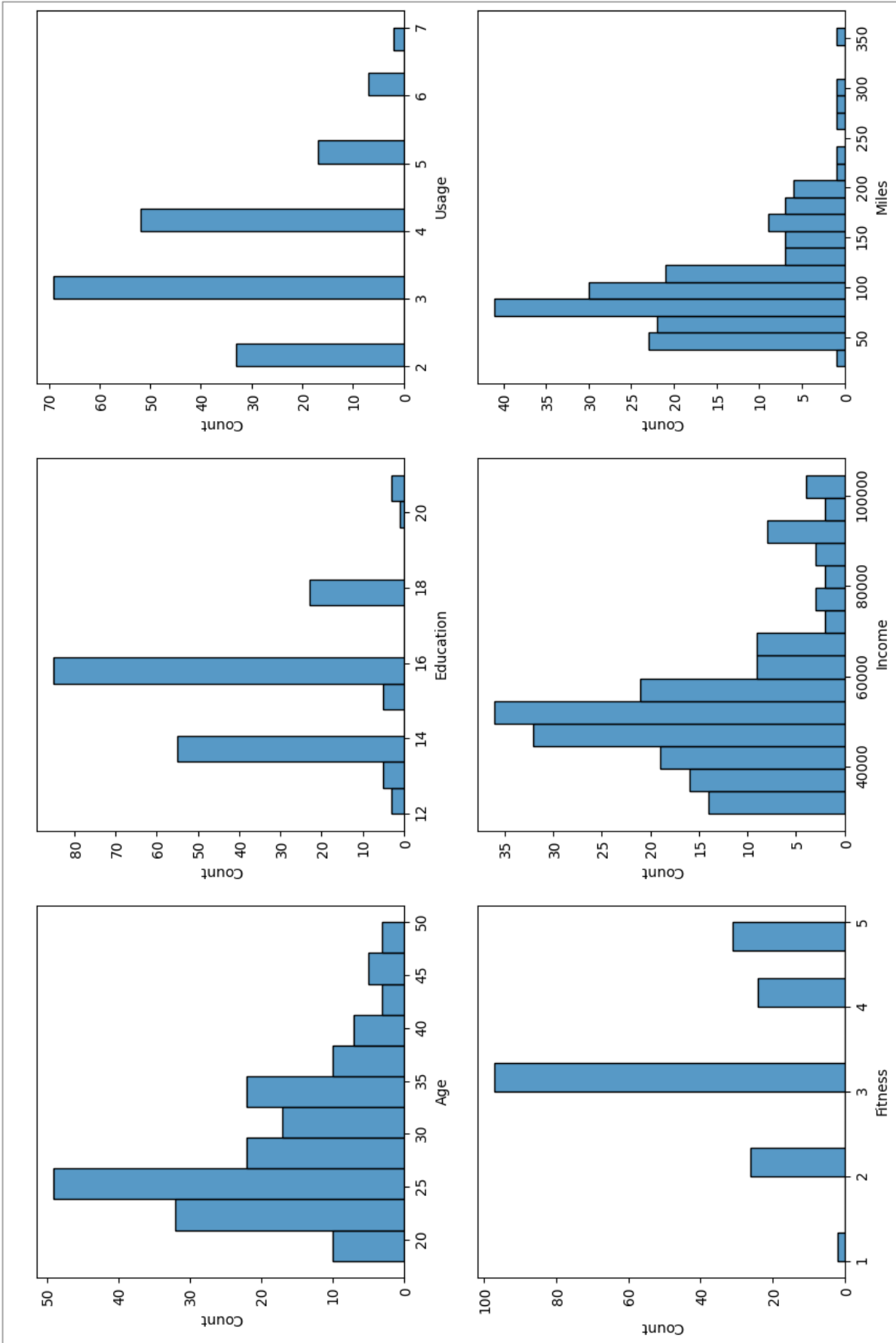


Figure 3: Client breakdown by Age, Education, Fitness Level and Income

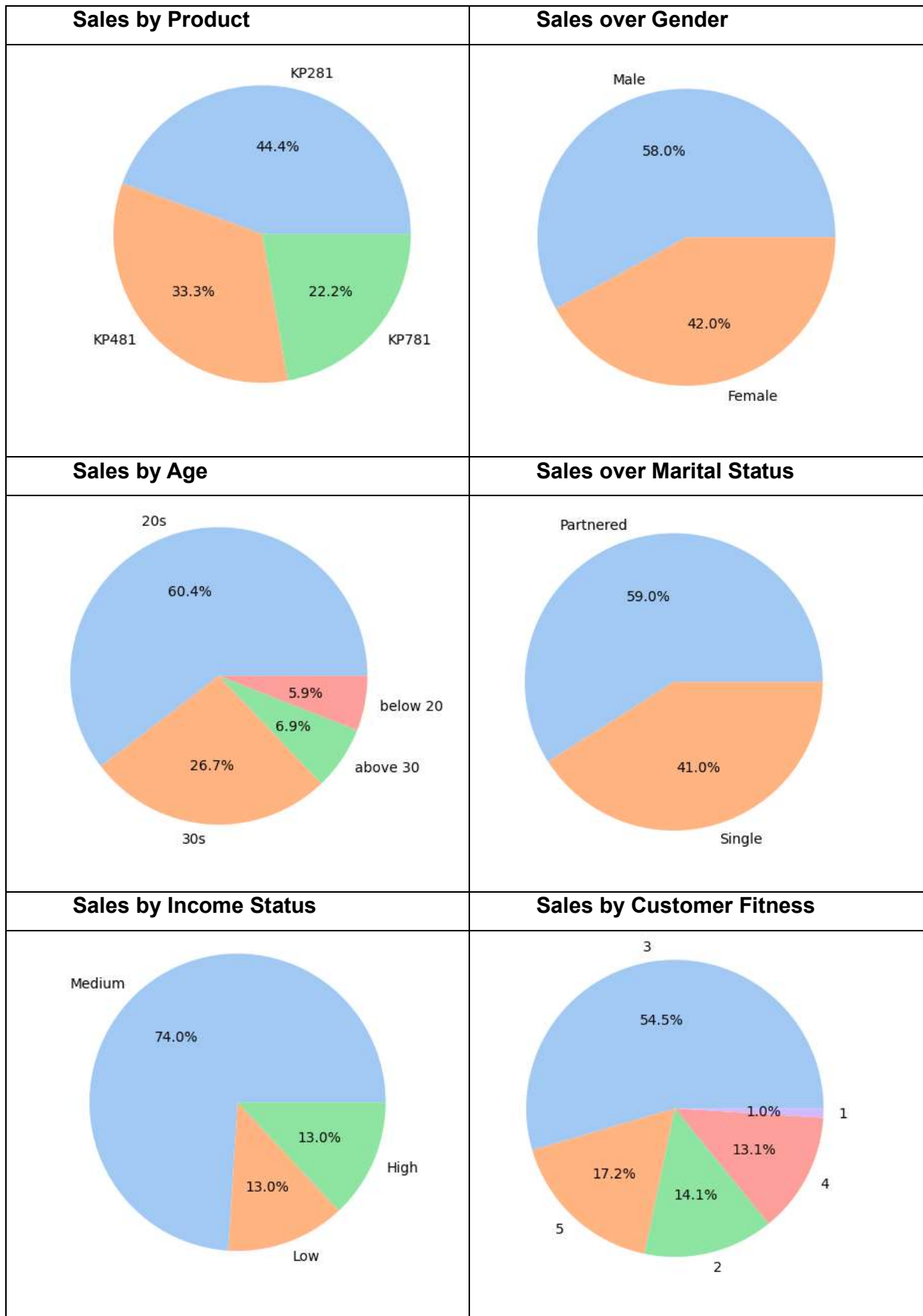


Figure 4: Sales Trend

Purchase Trends driven by Customer Behavior

This section correlates how the factors discussed in the previous section impact the final product selection by the customer.

- Purchase trend for KP281 and KP481 is approx. equal amongst both genders. There is equal probability that the buyer of these two products could either be male or female. However, most buyers of KP781 are men.

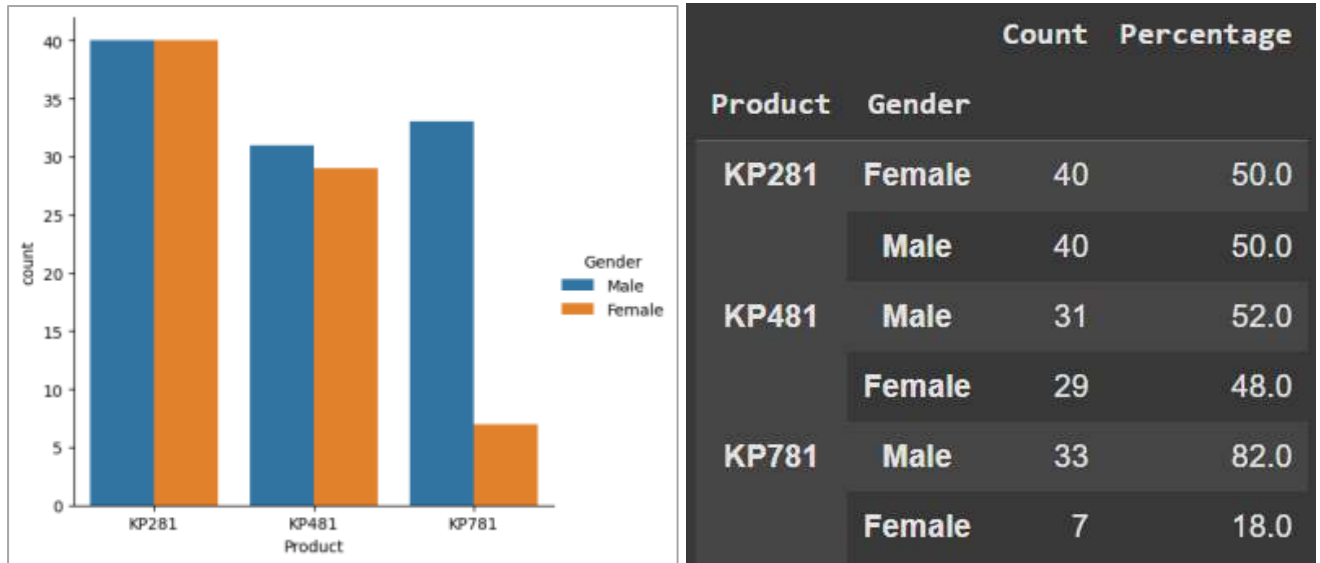


Figure 5: Product Purchase Trend by Gender

- As can be seen by the following figure, the majority of customers are partnered and the trend remains consistent for all products.

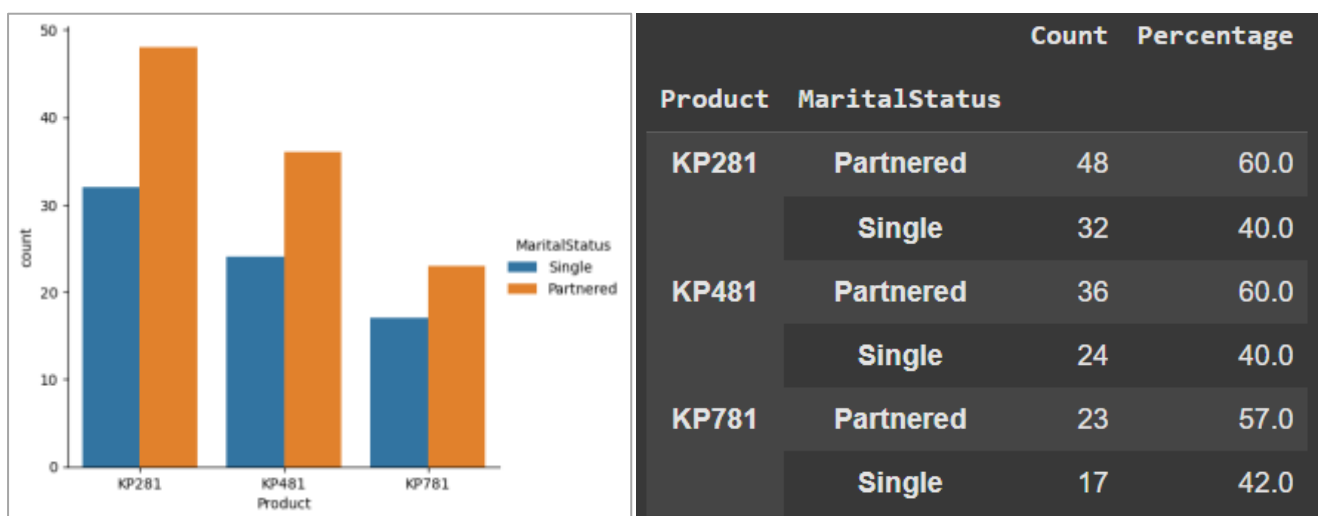


Figure 6: Product Purchase Trend by Marital Status

3. Majority of the buyers for each product are between ages 20-39 years old. However, it is worth noting that 25-30 years old prefer KP781 over KP481 and 30-35 years old prefer KP481 over the other products.

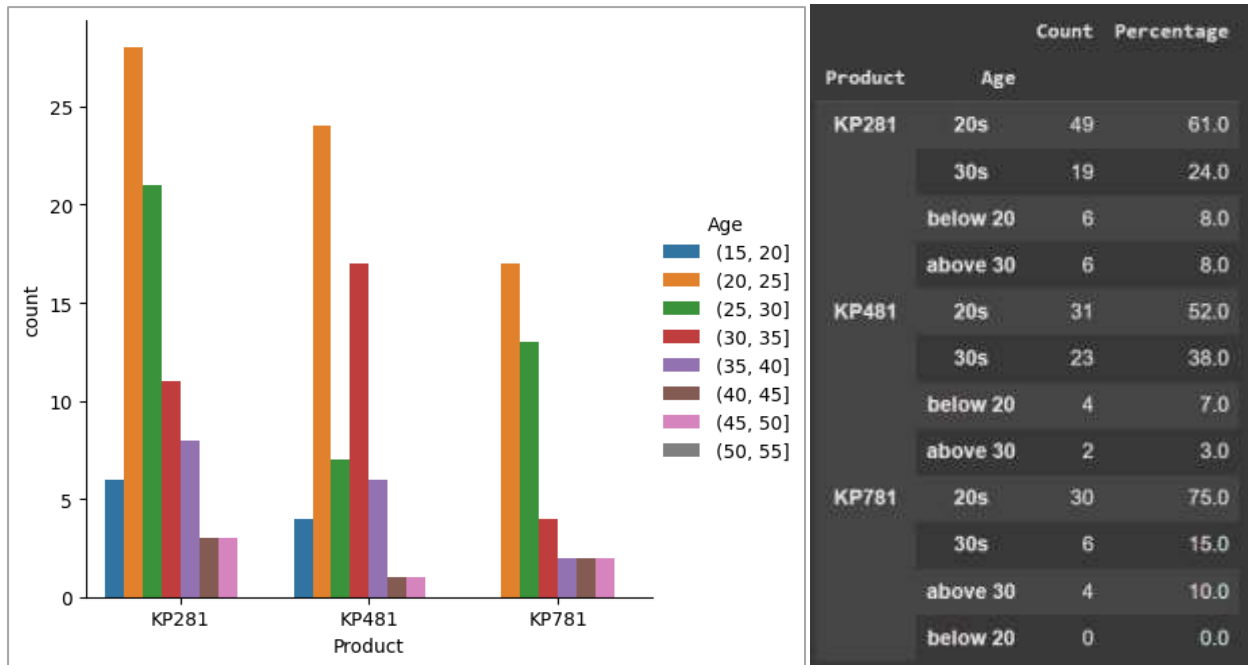


Figure 7: Purchase Trends by Age

4. Customers who purchased KP781 were more likely to be in good shape as compared to KP481 and KP281. 16% amongst the total 17% [Figure 4] of buyers with fitness level of 5 purchase KP781.

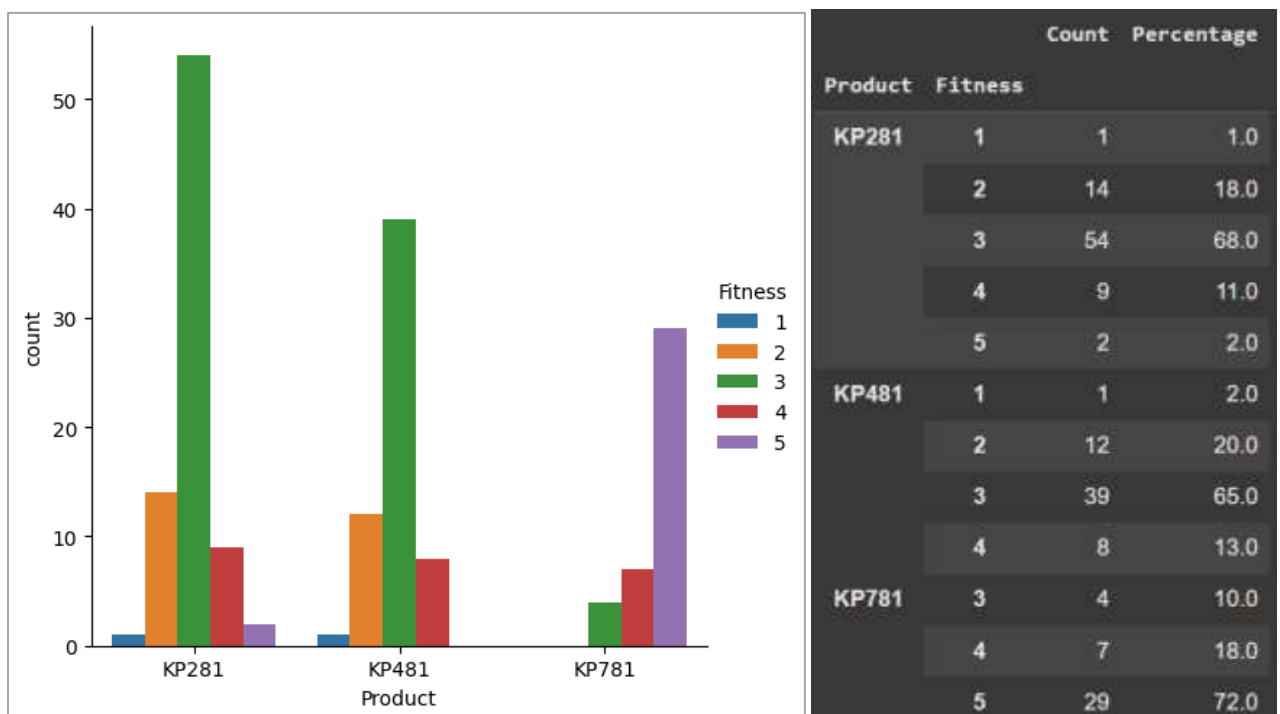


Figure 8: Purchase Trends by Customer Fitness

5. As is evident from Figure 9 only customers with high income purchase KP781 whereas most medium income individuals prefer both KP281 and KP481. Amongst the 13% [Figure 4] with high income status 100% have purchased KP781.

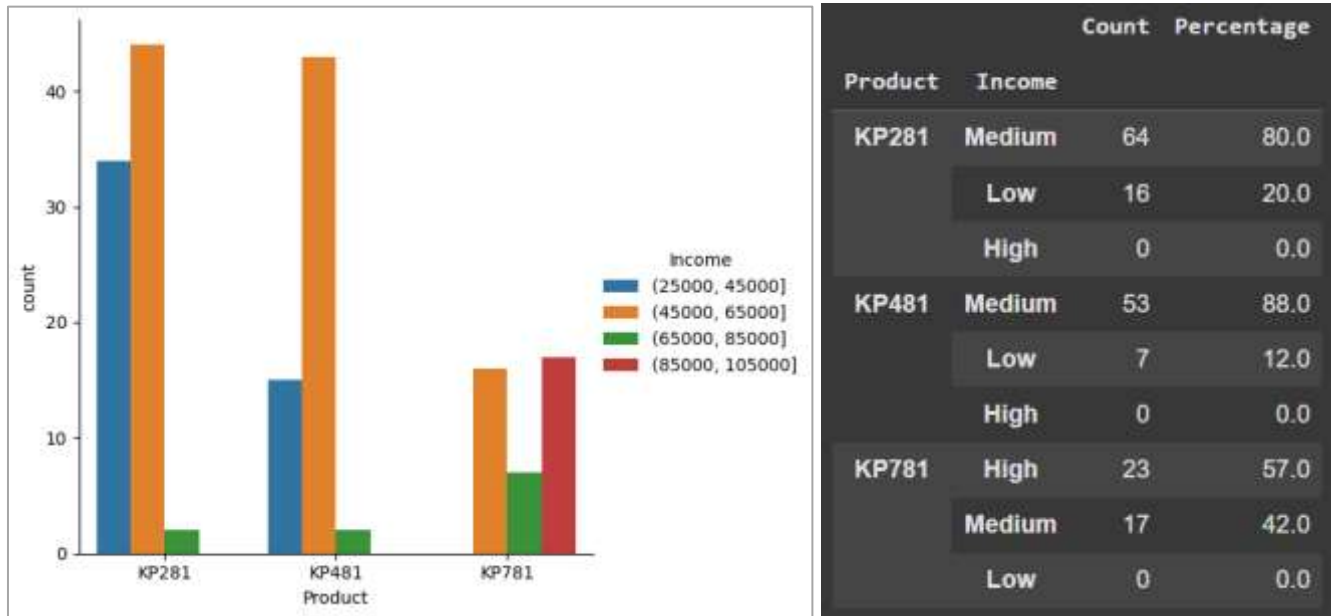


Figure 9: Purchase Trends by Income Status

Pairplot Analysis

Pair plots show a multi-variate relationship between numerical data which can be further broken down with categorical data as shown in Figure 11 where the plot in Figure 10 is further categorized with respect to product.

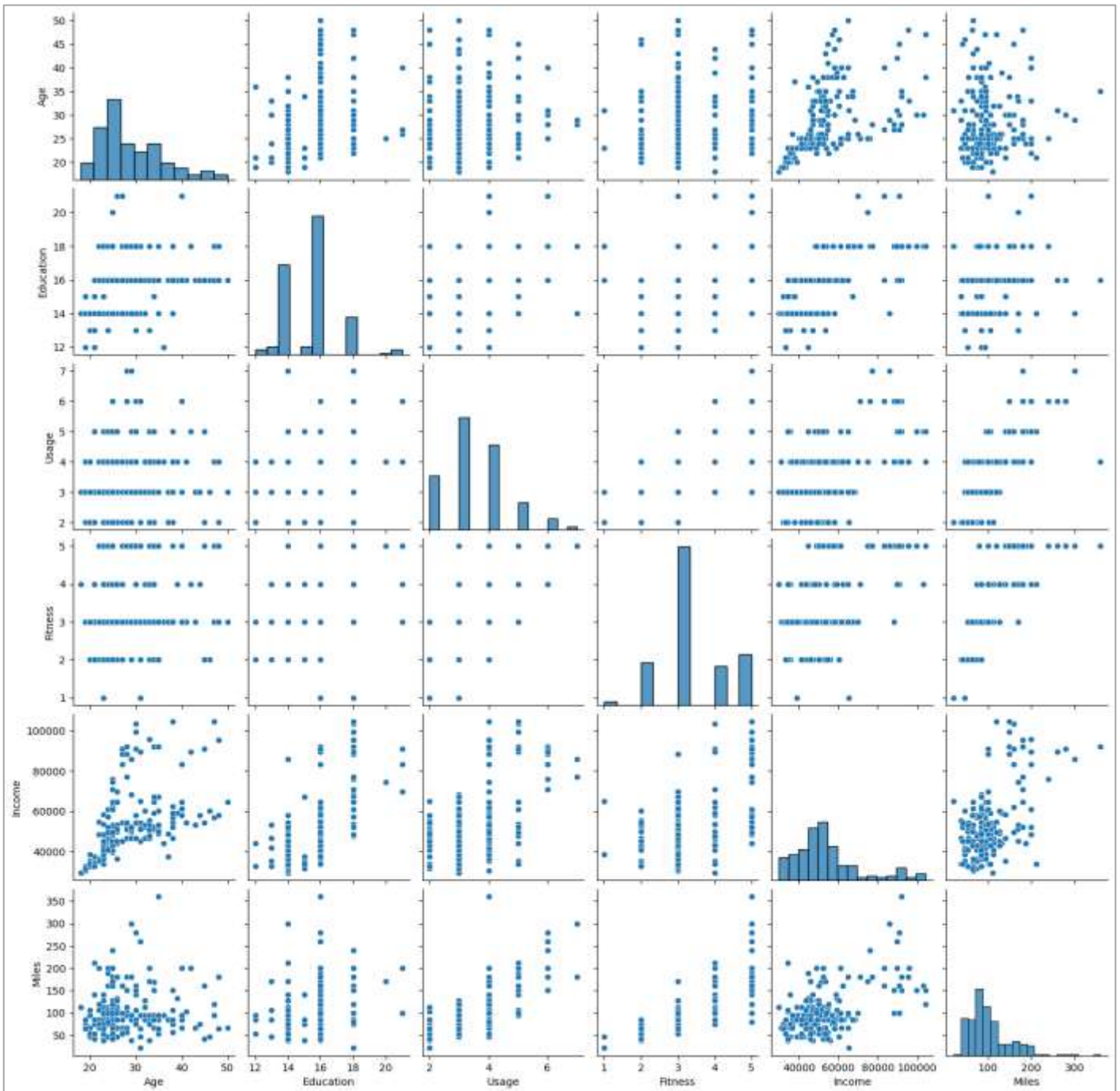


Figure 10: Pairplot for Multivariate Analysis

From Figure 10, it can be seen that the people above 20 years are more likely to be highly educated and income also increases with the age. Similarly, people with higher education are more likely to earn more etc.

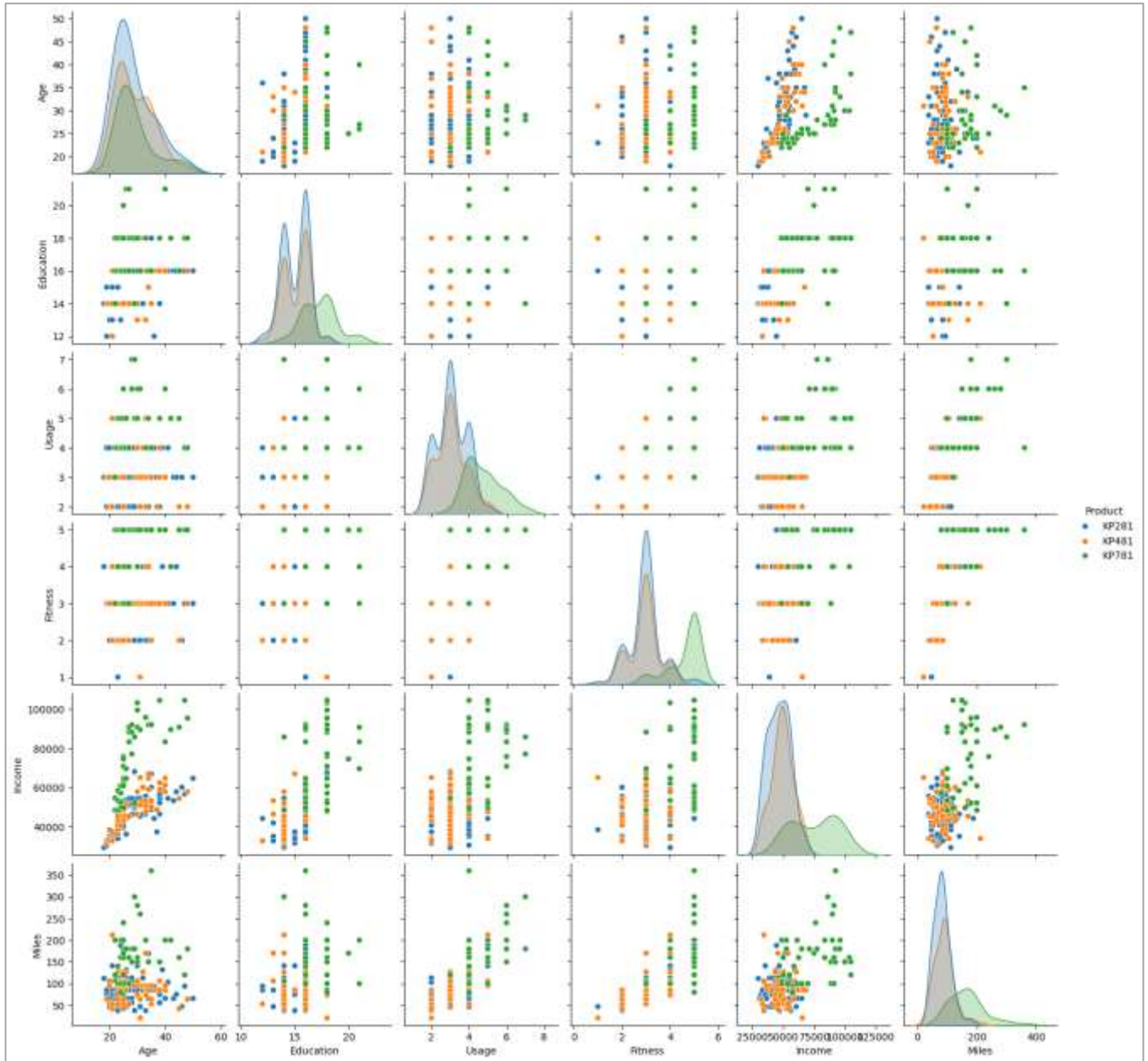


Figure 11: Pair Plot categorized by Product Type

Correlation Matrix

A correlation matrix presents the relationship amongst the variables itself. As can be seen from the following figure that age greatly impacts income meaning that with age income also increases. Similarly, education also has significant correlation with income. High usage of the treadmill is directly related to higher fitness, income and miles covered each week. In short, a correlation matrix gives the same information as pairplots but in a much simpler fashion.



Figure 12: Correlation Matrix

Actionable Insights and Recommendations

From the above analysis, it can be concluded that education, fitness level and income greatly influence the purchase of the customer. To improve the recommendation system these factors must be taken into account.

1. As middle class makes the biggest clientele, they should be educated about how KP781 has users with higher fitness levels.
2. People with higher education are most likely to choose KP781.
3. KP481 should be recommended as a starting point for females.
4. People in 20s and 30s are the biggest demographic and since age is directly correlated to income, they are more likely to purchase KP281 and KP481.
5. People who seem to be in good shape are more inclined to buy KP781.
6. Live demo classes can be arranged to demonstrate the superiority of KP781 over other products.