

The Aero fit data is given to identify the characteristics of the target audience for each type of treadmill offered by the company, to provide a better recommendation of the treadmills to the new customers. The team decides to investigate whether there are differences across the product with respect to customer characteristics. The data is given to find, how various aspects of customer effect the purchase of treadmill and provide valuable insights from the findings that would help company to target the customers better and increase the sales.

The top product sold is calculated. How different income ranges purchases the products is calculated. The relation between income and education is calculated how it effects in product purchase is calculated. The Age preference of product purchase is calculated.

The given data set has 180 rows and 9 columns. The following are nine columns: Product, Age, Gender, Education, Marital status, Usage, Fitness, Income, and Miles. Each columns provides which different inputs of the customers like Product column is to show type of product chosen. Age column specifies the different ages of the customers. Gender of the customers is shown in Gender column. Education columns shows the years spent by the customer for education. Marital status column shows whether is partnered or single. Usage columns how frequently the product is used. Fitness column the fitness level of the customers. Income columns different income ranges of the customers. The Miles shows the Miles completed by the customers.

The data type of the given file is:

#	Column	Non-Null Count	Dtype
0	Product	180 non-null	object
1	Age	180 non-null	int64
2	Gender	180 non-null	object
3	Education	180 non-null	int64
4	MaritalStatus	180 non-null	object
5	Usage	180 non-null	int64
6	Fitness	180 non-null	int64
7	Income	180 non-null	int64
8	Miles	180 non-null	int64

Table 1: Data type of the file

From Table 1, we can see that columns which are object are object type, including product, Gender and Martial status. The Age, Education, Usage, Fitness, Income and miles are in integer type. The file has zero null values.

1. Initial analysis data

The observations made from describe table are as follows:

There are no missing values in the data. There are 3 unique products in the dataset. KP281 is the most frequent product. 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. Most of the people are having 16 years of education i.e. 75% of persons are having education ≤ 16 years. Out of 180 data points, 104's gender is Male and rest are the female. Standard deviation for Income & Miles is very high. These variables might have the outliers in it.

1.1 Analysis of all the integer columns

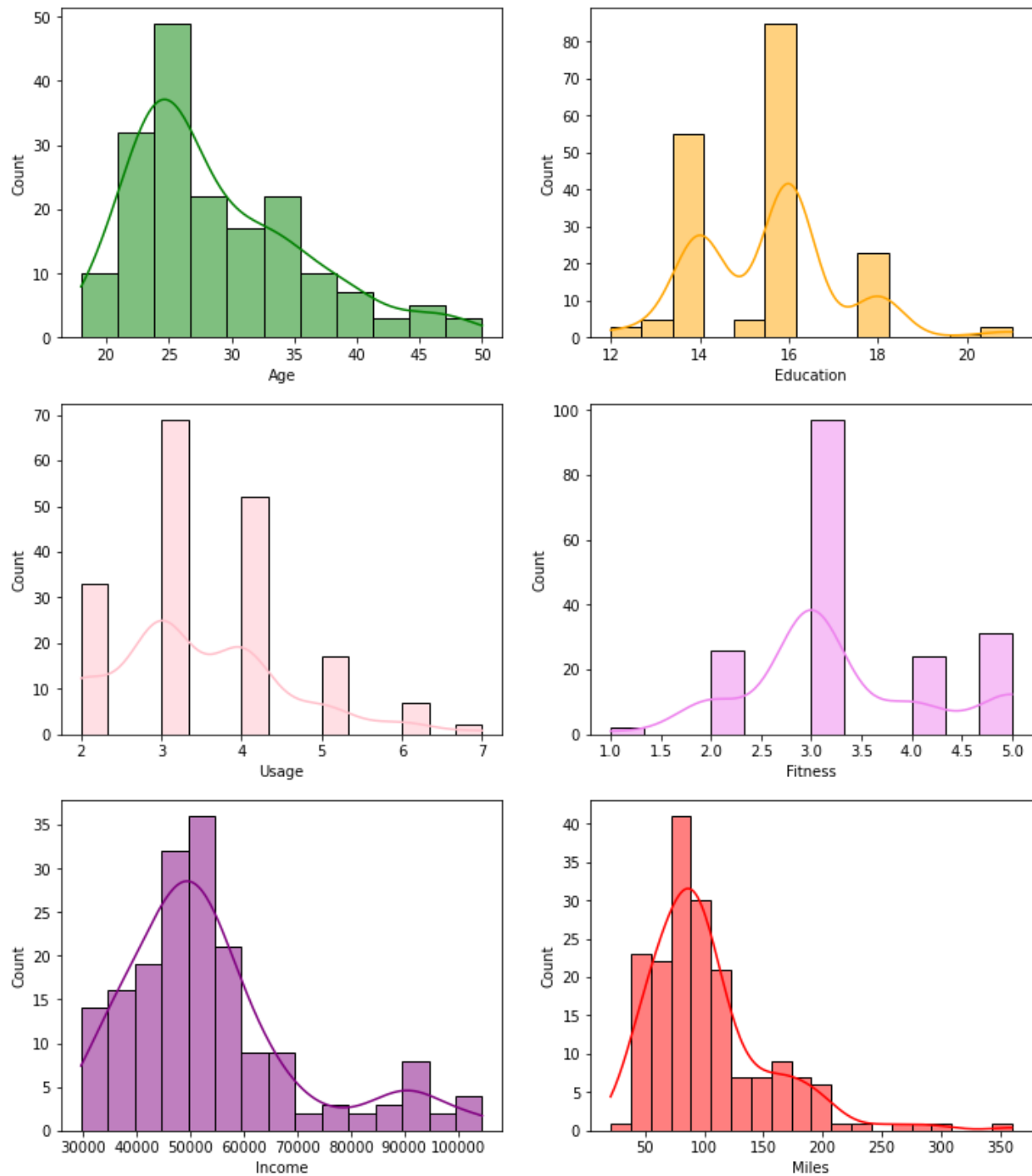


Figure 1: count of all the integer columns

Understanding the distribution of the data for the quantitative attributes:

1. Age
2. Education

3. Usage
4. Fitness
5. Income
6. Miles

The age graph shows that the age group from 18 to 28 yrs. are keener about fitness and are more inclined to buy the products. As the age increases the interest towards fitness is also lowering.

The education graph shows 16 years was the maximum years spent by the clients for the education then follows 14 years in second place and 18 years in third place for years spent in education.

The Income graph shows that the income range from 30,000 to 65,000 are keener in purchasing the products. The graph clearly shows medium income range customers are who buy our products second place is for low income range customers

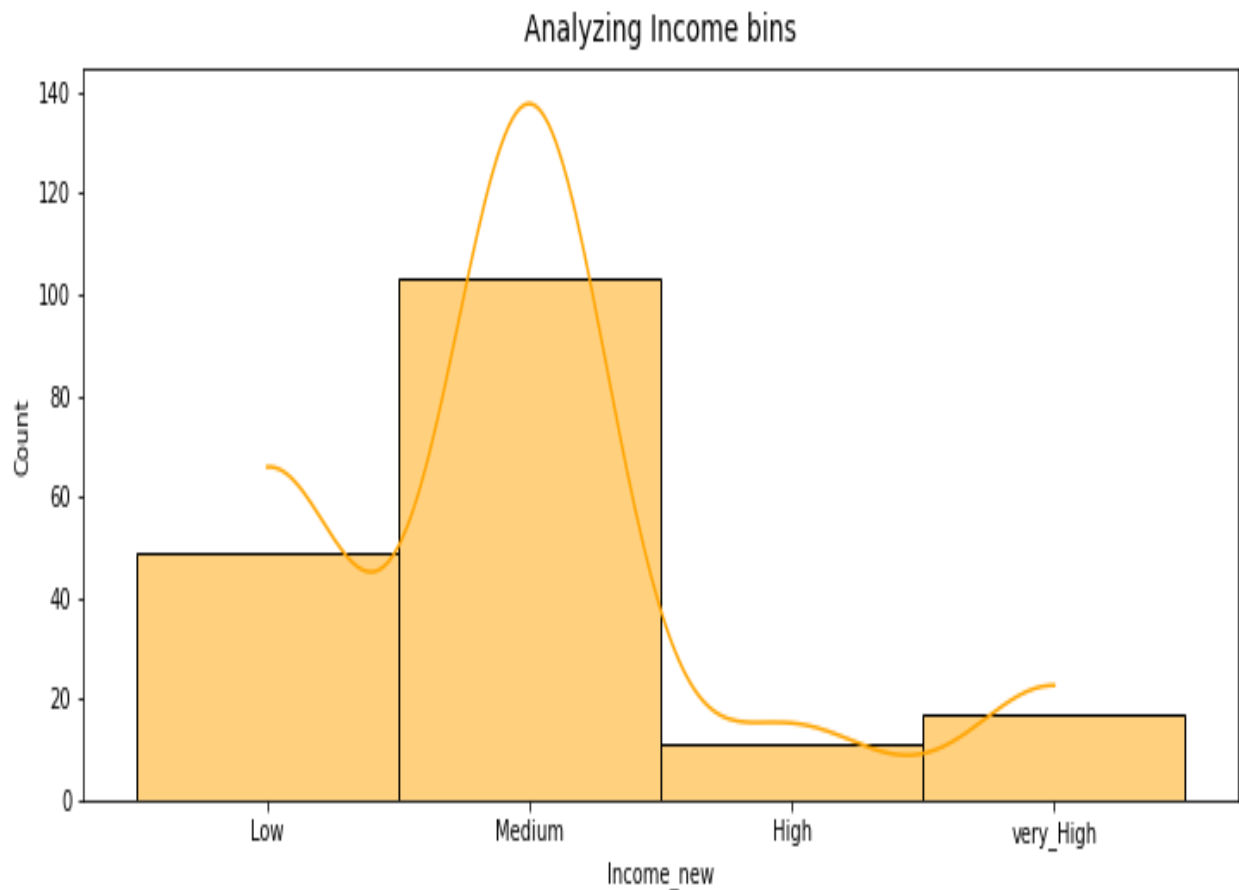


Figure 2: Analyzing the income ranges using bins

1.2categorical columns analyzation

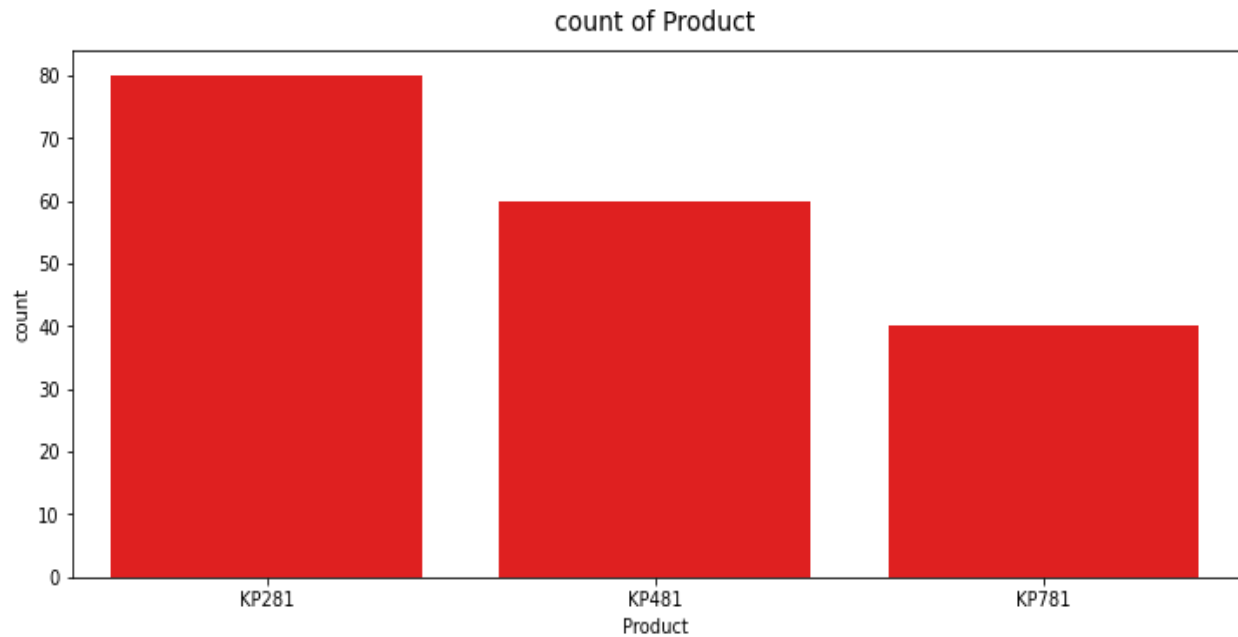


Figure 3 : product Frequencies

From the figure 3 we can draw the insight that KP281 was the most frequent product sold. The KP281 is the cheapest product of all the three and the highest priced product was least sold

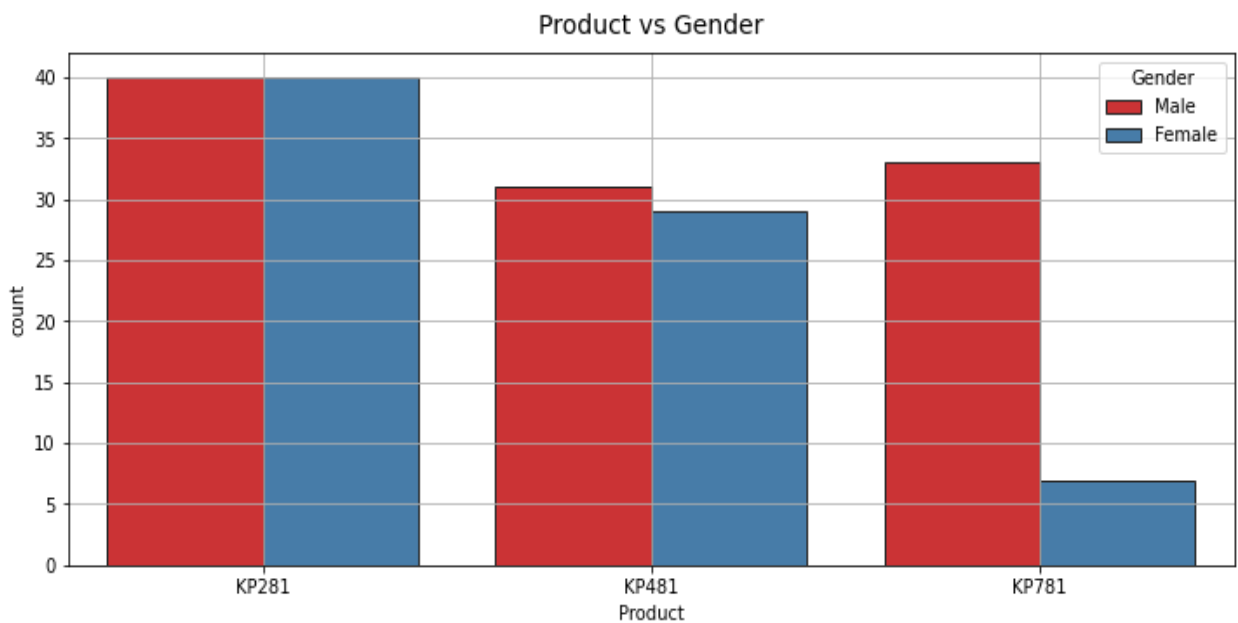


Figure 4: Gender preference for product

From figure 4 we can say that for KP281 the male and female purchases are equal, for KP481 the male purchases are slightly greater than the female purchases and for the KP781 the male purchases are higher, higher that of KP481 male purchases.

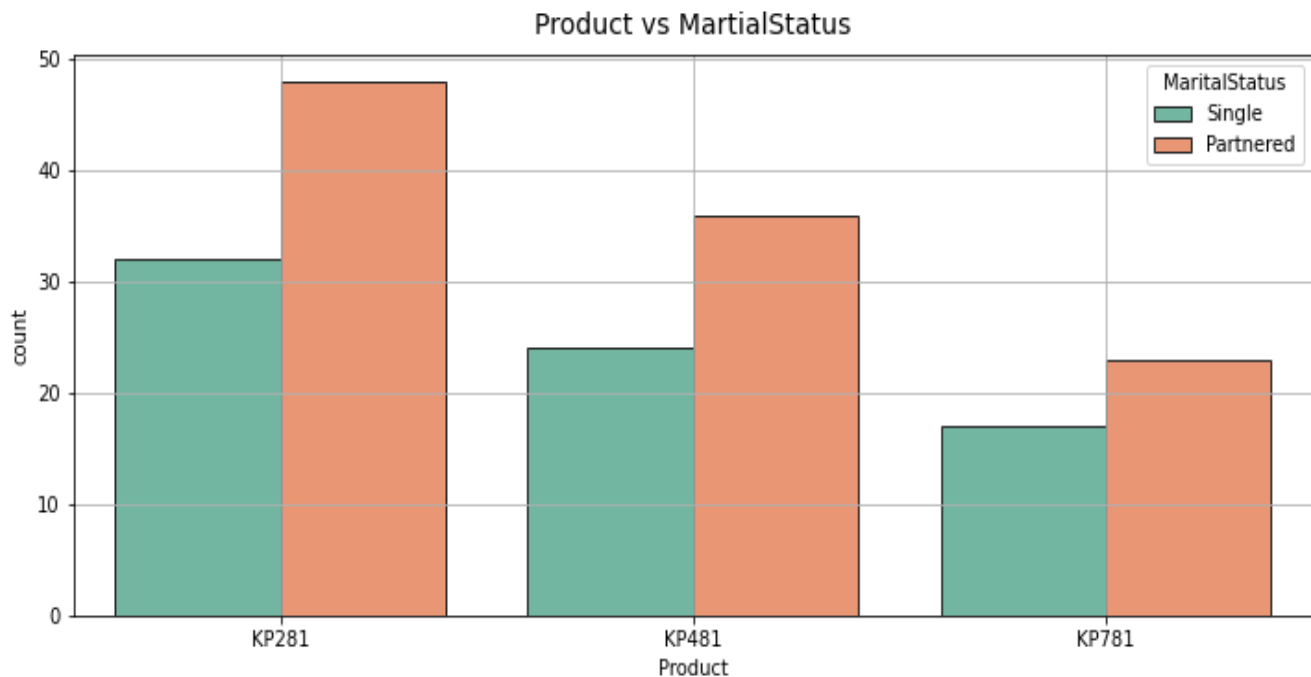


Figure 5: Marital status influence on product purchase

The figure 5 shows that Partnered customers are more interested to by the products. It also says the partnered customers are more fitness oriented. In all three 3 products the partnered customer are high and comparing with in these in the KP281 product range the purchase of partnered customers are higher. According to single customers count was more in KP281 product compare to all the three, but lower than the partnered customers count.

According to the figure 6 we can clearly observe customers with very_high range of income Purchase only high end product KP781. The medium income range customers prefer both Low end KP281 and middle range product KP481. The Low income range customers prefer Mostly on low end product kp281. The high income range customers are not keen intereste d in the product purchases they purchases of all three products is lower than all income Range customer's purchases.

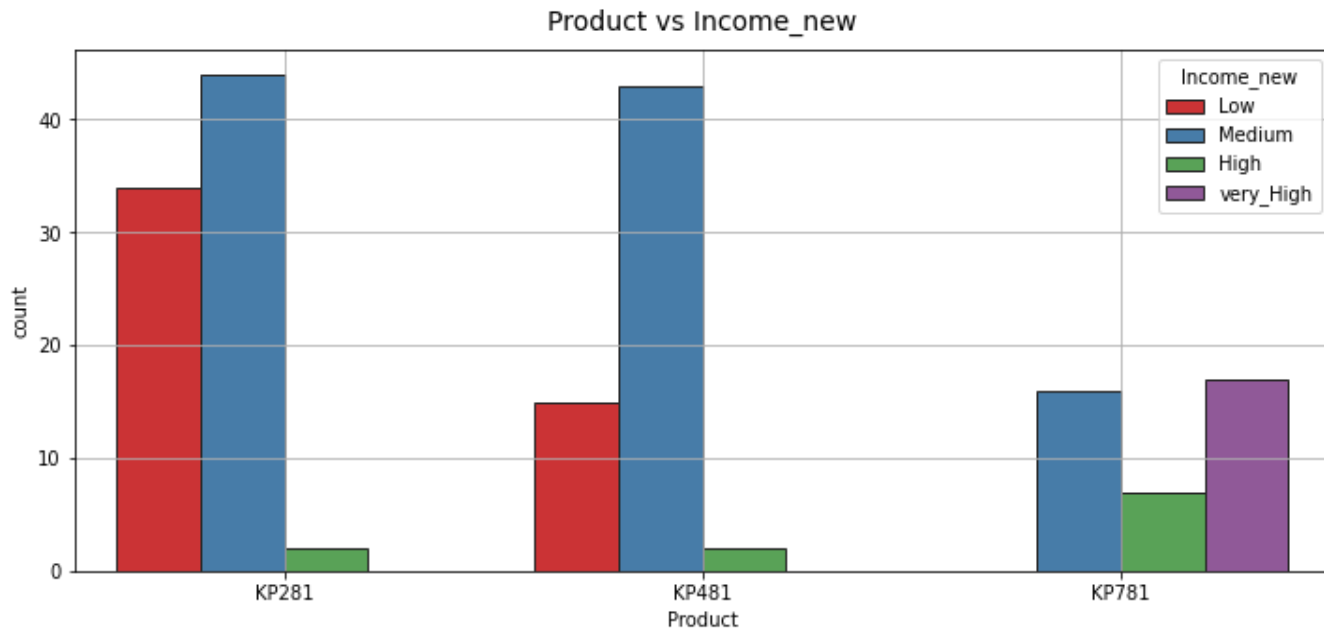


Figure 6: Income ranges purchase of products

2. Analyzing quantitative attributes influence on product purchase

From figure 7 observation made are:

2.1 Product vs Age

- Customers purchasing products KP281 & KP481 are having same Age median value.
- The spread of KP281 is more and it has highest maximum value.
- Customers whose age lies between 25-30, are more likely to buy KP781 product
- There are no negative outliers.

2.2 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.

2.3 Product vs Usage

- Customers who are planning to use the treadmill greater than 4 times a week, are more likely to purchase the KP781 product.

- While the other customers are likely to purchasing KP281 or KP481.
- There is negative outlier in KP481.

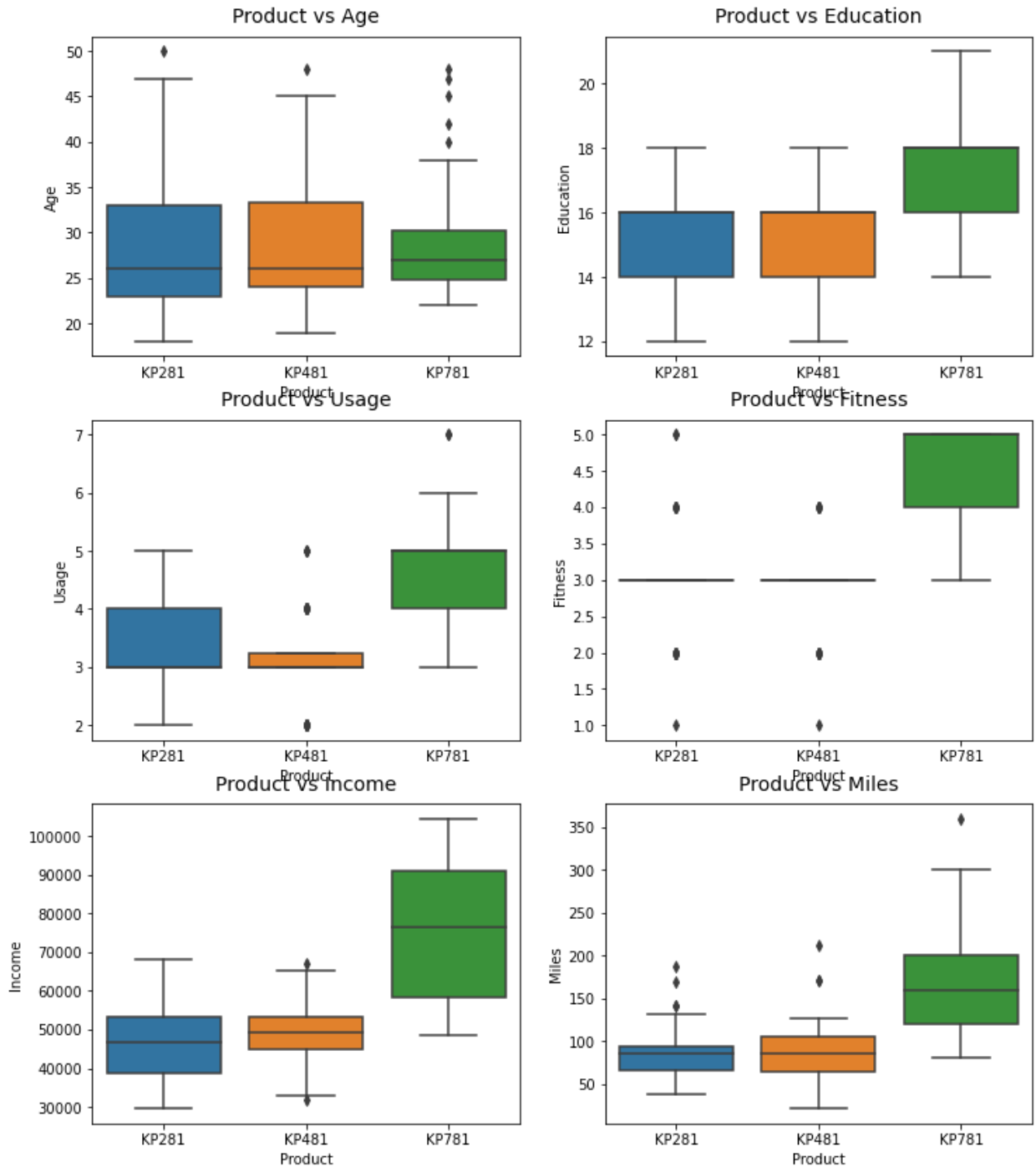


Figure 7: Quantitative attributes effect on product purchase

2.4 Product vs Fitness

- The more the customer is fit (fitness ≥ 3), higher the chances of the customer to purchase the KP781 product.

2.5 Product vs Income

- Higher the Income of the customer (Income ≥ 60000), higher the chances of the customer to purchase the KP781 product.
- The median value for KP781 is high. For the remaining both products the median value is most same.

2.6 Product vs Miles

- If the customer expects to walk/run greater than 120 Miles per week, it is more likely that the customer will buy

3. Understanding through income through bins

3.1 Income Vs Education

In the figure 8 graph we observe that customers who have spent many of years on education have high income and vey highest income ranges. There both ranges are almost same value of education.

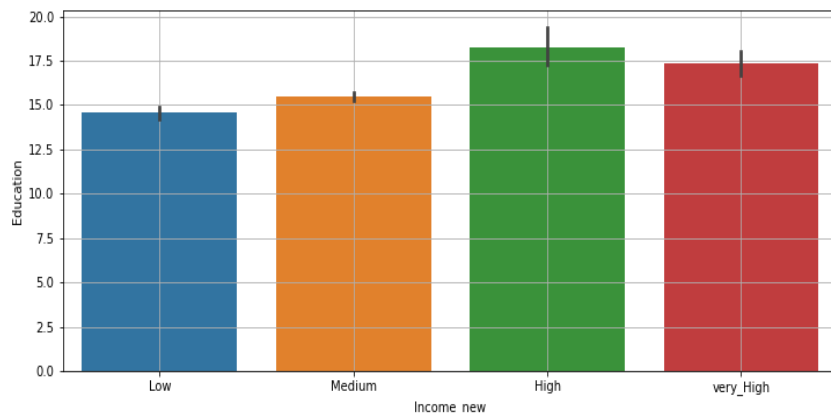


Figure 8: Income_new vs Education

3.2. Income_new vs Gender

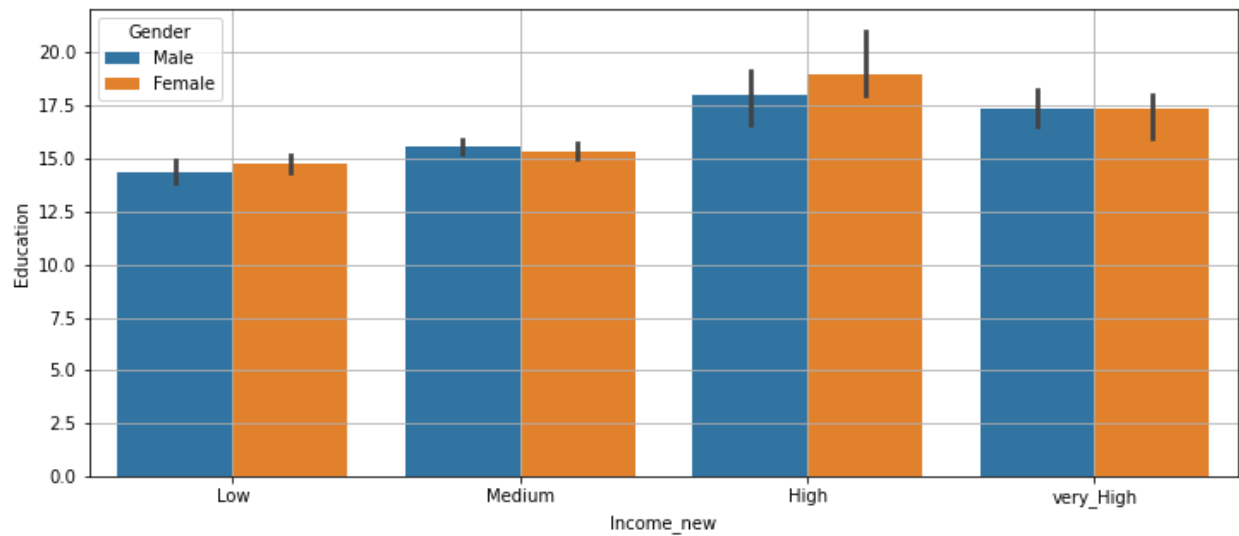


Figure 9: Analyzing the Gender wise education and income range

In the Figure 9 we can observe that Females have spent more years in education in high income range. In remaining three bins the education levels are equal in both male and female.

3.3. Income ranges vs product vs education

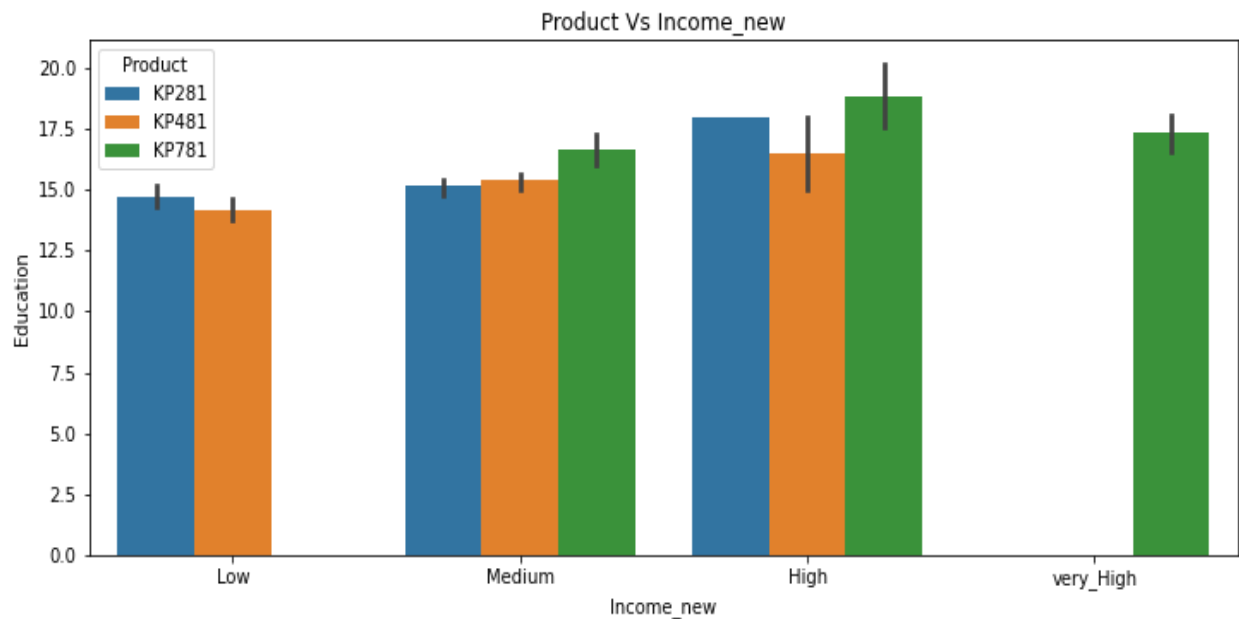


Figure 10: detailing of product purchase within the income ranges

In the Figure 10 clearly shows in very_high income range only KP781 is purchased and in low income range only two products are purchased KP281 and KP481.

Business insights:

1. The product KP281 is more purchased overall sales percentage of it high with 44 % followed by KP481 with 33 % and least sales is by company's highest price product KP781 with 22 % sales.
2. Males are more likely to purchase a treadmill with 58% ratio than Females. Both are likely to purchase equal number of KP281 and KP481, but Males have high chances of purchasing KP781 as 82% of total sale of KP781 is purchased by Males.
3. People with Education levels less than or equal to 16 are likely to purchase KP281 and KP481. And people with Education levels greater than or equal to 16 are likely to purchase KP781.
4. People with Usage less than or equal to 4 are likely to purchase KP281 and KP481. And people with Usage greater than or equal to 4 are likely to purchase KP781.
5. People with Income less than 6000 are likely to purchase KP281 and KP481. And people with Income greater than 6000 are likely to purchase KP781.
6. People with Fitness Level 3 or less are likely to purchase KP281 and KP481. And with Fitness Level 5 are likely to purchase KP781.
7. According to Income_new which divides the salaries ranges using bins shows that medium income customers (45000<customers income <65000) range are more. The medium range customers purchase KP281 and KP481.
8. The customers whose education is greater than 16 earn high and very high income. Females have spent more time of education in high income range.
9. The very high income range customers are keener to buy the most expensive product KP781. The low income range customers buy only KP281 and KP481.
10. People who use the treadmill more are more likely to purchase KP781. As, buying the treadmill is directly proportional to its usage

Recommendations:

1. Customers who are having age between 18 to 35 should targeted more as more than 85% of the customers who bought treadmill lie in this range.
2. Medium income range customers are to be targeted more and should be encouraged to purchase KP781 product. Because these are more into KP281 and KP481
3. Though very_high income customers buy only KP781 still there purchase percentage is only 9 % which very low. Can observe in figure no: 6 .Need to increase it.
4. High income customer's purchases total purchases is only 6% which very minimal. If concentrated it can increase the purchases of KP781 product.
5. Partnered customers should be targeted more as 59% of the customers who brought treadmill lie in the range.

