

## Aerofit Treadmill Project

There may be some unique reasons why customers choose any specific product. Businesses must continue adapting to customer expectations and mindset to succeed long-term. The customer product choice allows the business to stay relevant and ahead of the competition.

### Problem Statement:

Create a customer profile based on the criteria of age, marital status, gender ,education and income that which type of treadmill offered is better.

### Observations:

Total number of data entries =180

#### Attributes given in data set:

1. **Products:** Treadmill models KP281, KP481, or KP781
2. **Age :** In years
3. **Gender:** Male/Female
4. **Education:** In years
5. **MaritalStatus:** Single or partnered
6. **Usage:** The average number of times the customer plans to use the treadmill each week.
7. **Income:** Annual income (in \$)
8. **Fitness:** Self-rated fitness on a 1-to-5 scale, where 1 is poor shape and 5 is excellent shape.
9. **Miles:** The average number of miles the customer expects to walk/run each week.

### **Numerical Attributes:** Age, Education, Usage, Income, Fitness, Miles

The descriptive analysis of numerical attributes is shown in the table below.

	Age	Education	Usage	Fitness	Income	Miles
count	180.000000	180.000000	180.000000	180.000000	180.000000	180.000000
mean	28.788889	15.572222	3.455556	3.311111	53719.577778	103.194444
std	6.943498	1.617055	1.084797	0.958869	16506.684226	51.863605
min	18.000000	12.000000	2.000000	1.000000	29562.000000	21.000000
25%	24.000000	14.000000	3.000000	3.000000	44058.750000	66.000000
50%	26.000000	16.000000	3.000000	3.000000	50596.500000	94.000000
75%	33.000000	16.000000	4.000000	4.000000	58668.000000	114.750000
max	50.000000	21.000000	7.000000	5.000000	104581.000000	360.000000

We can categorize Income and education in

- i) High Income, Low income
- ii) High Education, Low education

Fitness can be categorized as

Physically Fit (Highest), Moderately Fit, Need Improvement(Lowest)

### **Categorical Attributes:** Product, Gender, MaritalStatus

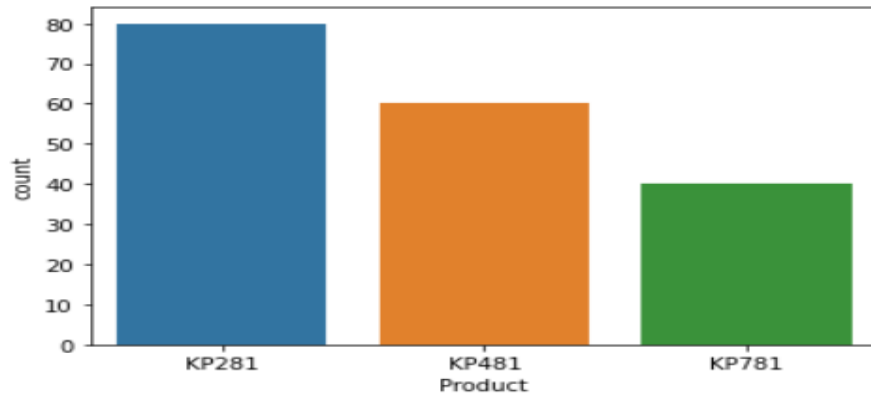
**Product Categories:** Treadmill models KP281, KP481, or KP781

**Gender categories:** Male and Female

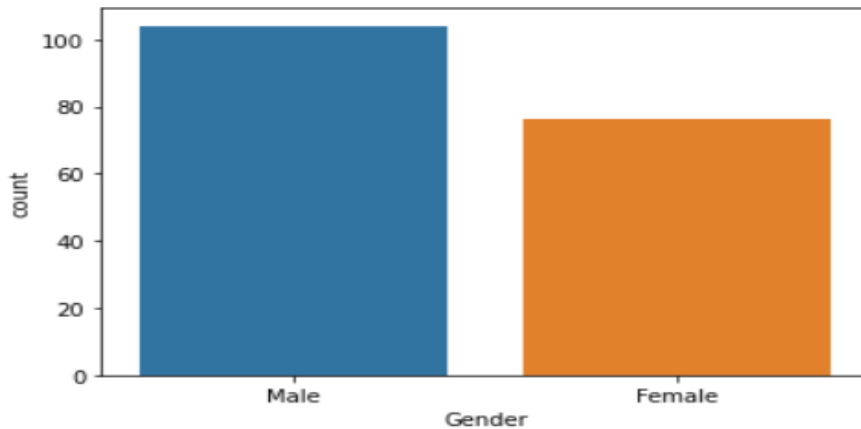
**Marital Status:** Partnered and Single

## Univariate Analysis:

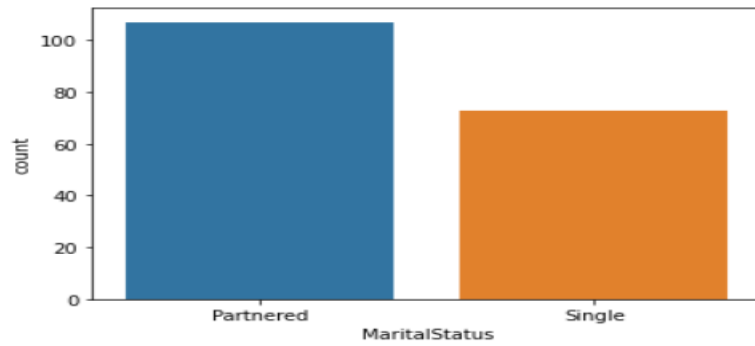
1. The countplot of products show that count of KP281 is more or it is mostly selling article.



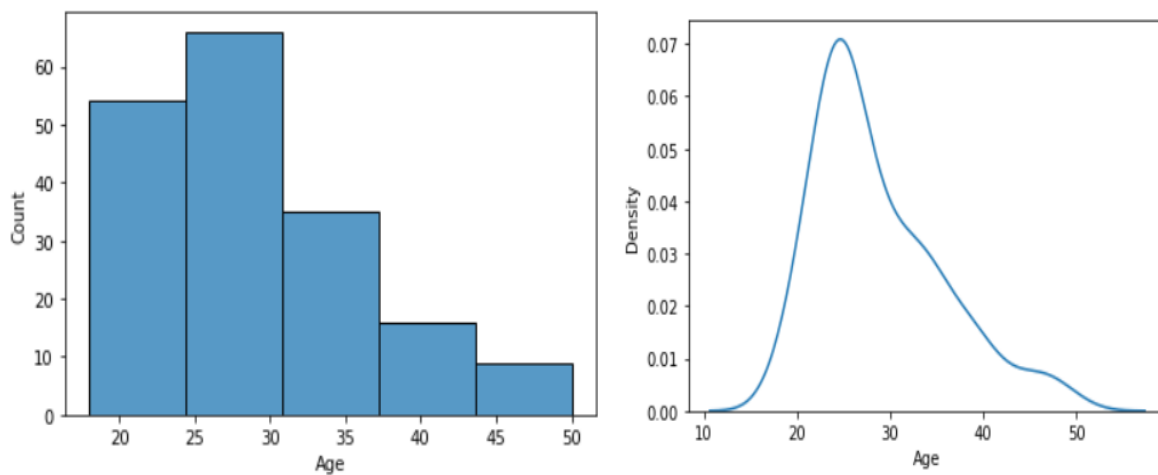
2. According to gender analysis **more male use treadmills.**



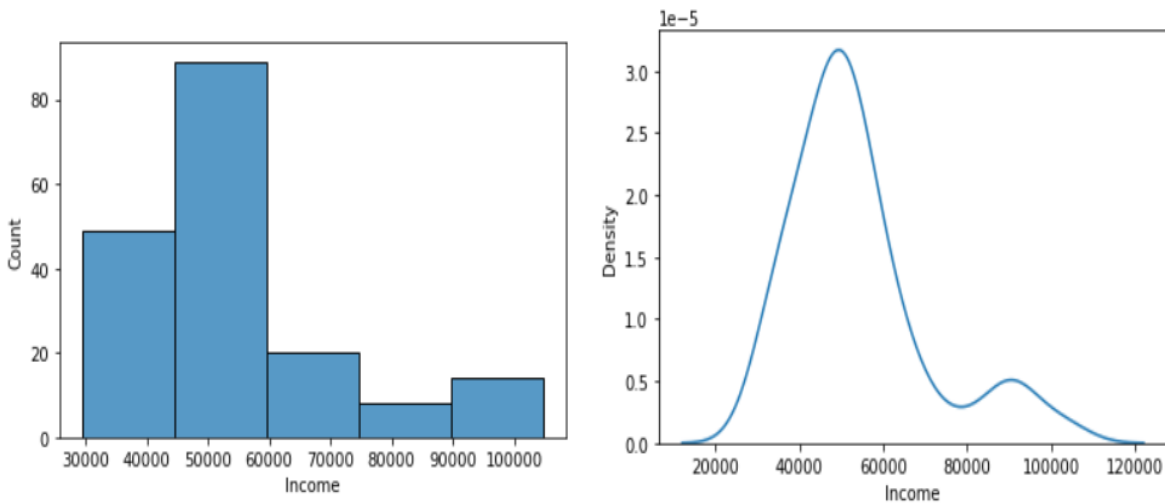
3.. Married people use treadmill more as compared to single.



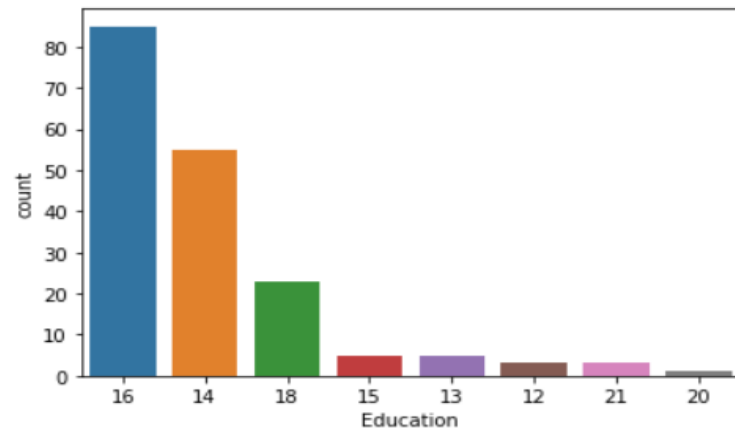
4. this plot represents that most of the people who use treadmill has age between 20 to 40 years. And mean value is around 29 years.



5. Majority people who use treadmill lies in the range of 20K to 80K. Mean is 53K.



6. Most of the highly educated people use treadmills.



## Bivariate Analysis:

### Outlier analysis:

#### 1. outlier analysis using python libraries for KP-781 with respect to “Age”

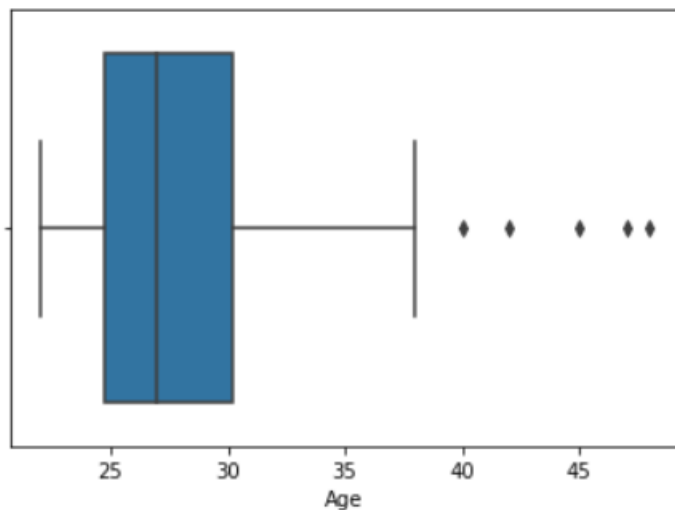
```
KP781=df[df["Product"]=="KP781"]
k7_25=np.percentile(KP781["Age"],25)
k7_50=np.percentile(KP781["Age"],50)
k7_75=np.percentile(KP781["Age"],75)
IQR_k7=k7_75-k7_25
upper_whisker_k7=k7_75+1.5*IQR_k7
lower_whisker_k7=k7_25-1.5*IQR_k7
Age_outliers_in_K781=KP781.loc[KP781["Age"]>upper_whisker_k7]
len(Age_outliers_in_K781) # count of age outliers
```

5

Outlier

identified =5

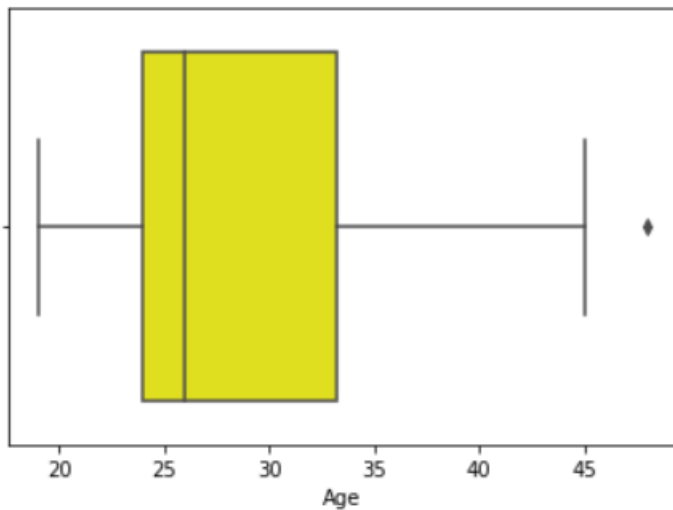
As shown in boxplot below:



Similarly for KP481 w.r.t. “Age”

```
sns.boxplot(data=KP481,x=KP481["Age"],color="yellow")
```

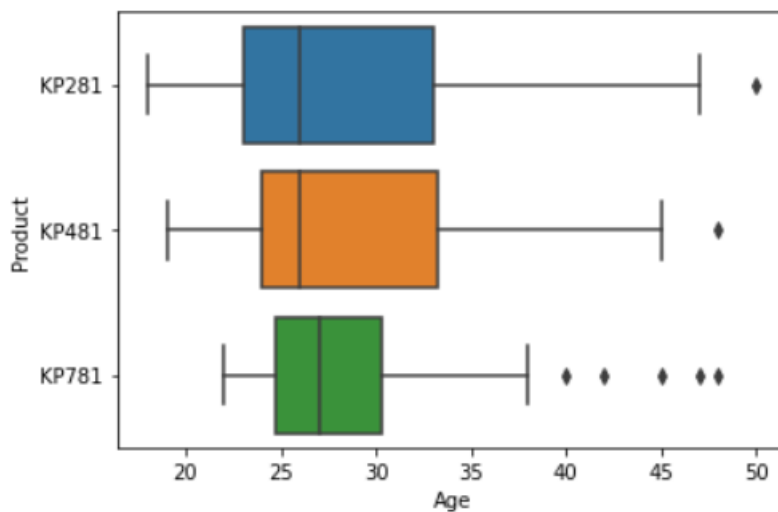
```
<AxesSubplot:xlabel='Age'>
```



```
#outlier detection in Age
KP481=df[df["Product"]=="KP481"]
k4_25=np.percentile(KP481["Age"],25)
k4_50=np.percentile(KP481["Age"],50)
k4_75=np.percentile(KP481["Age"],75)
IQR_k4=k4_75-k4_25
upper_whisker_k4=k4_75+1.5*IQR_k4
lower_whisker_k4=k4_25-1.5*IQR_k4
Age_outliers_in_K481=KP481.loc[KP481["Age"]>upper_whisker_k4]
len(Age_outliers_in_K481) # count of age outliers
```

1

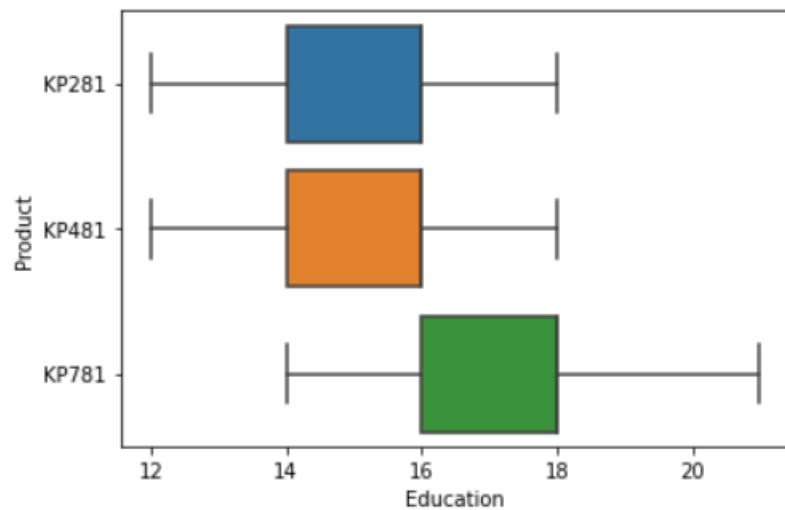
Analysis of “Age” in all three types of treadmills



It shows that KP281 and KP481 have 1 outliers but KP781 has 5(more) outliers.

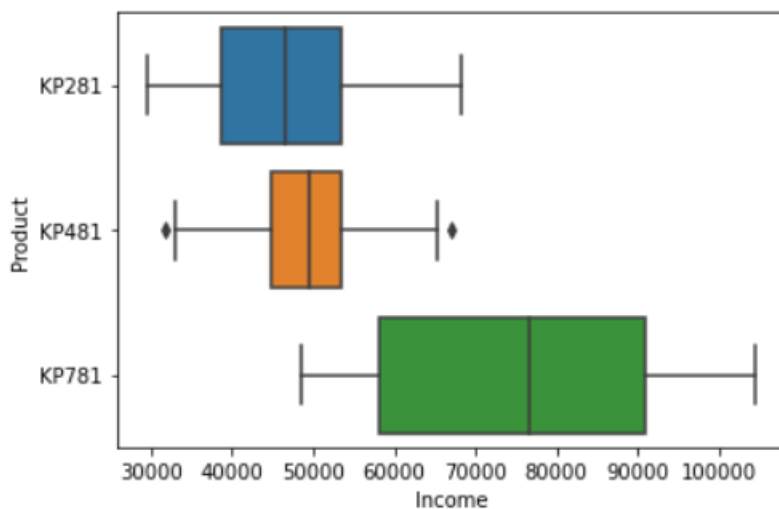
Target customers for KP281 and KP481 are having almost the same age group. But KP781 spread is more so high standard deviation as compared to the other two.

### Analysis on the Basis of “Education”:



There are no outliers in all three product categories. People having education between 14 to 16 years prefer KP481 and KP 281. Whereas those who have education in years between 16 to 18 prefer KP781.

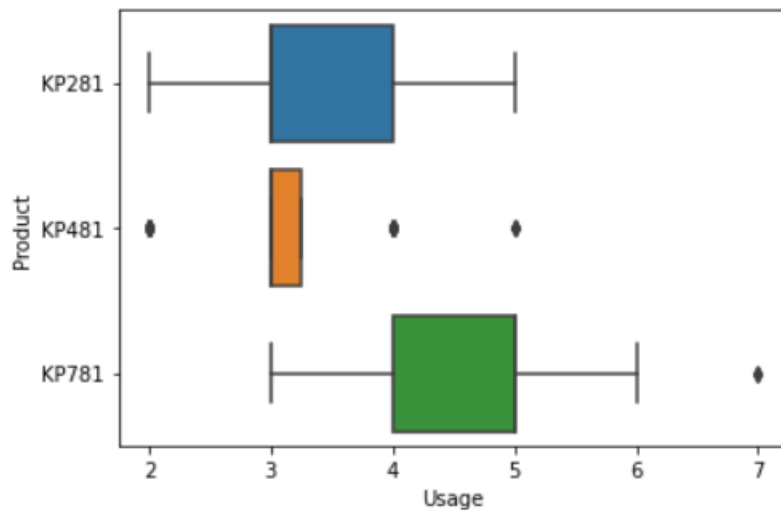
### Analysis on the Basis of “Income”:





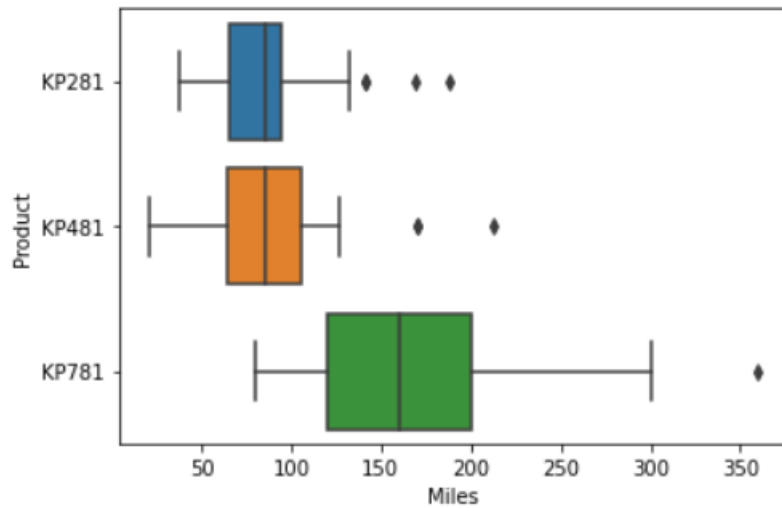
- This plot clearly shows that the lower income group prefers KP281 and the higher income group prefers KP-781 due to its pricing.
- Average income of people buying KP781(\$75441) is much higher than KP-281(\$46419) and KP-481(\$48973).
- **Income range(aprox.) - Product**
  - \$29K to \$68K - KP281
  - \$31K to \$68K - KP481
  - \$48K to \$104K - KP781

### Analysis on the Basis of “Usage”:



- People who use treadmills 3 to 4 times prefer KP281 where as KP781 is preferred by those people who use treadmills more than 4 times.
- KP481 is used by all “usage” categories of people.

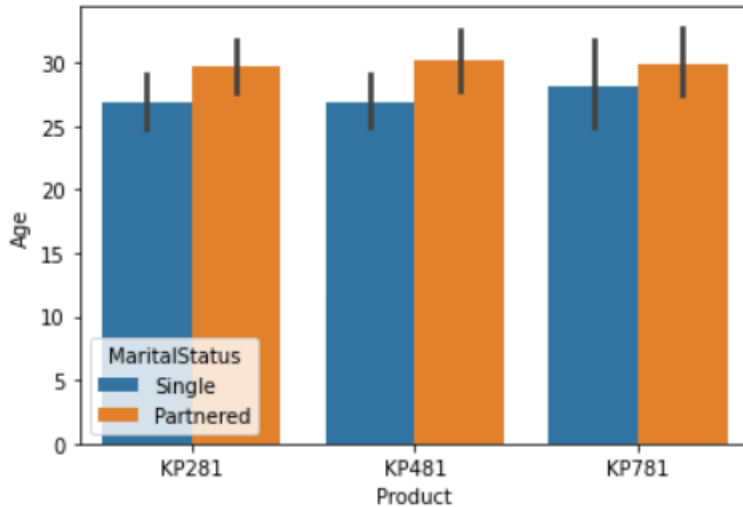
**Analysis on the basis of “Miles” average number of miles the customer expects to walk**



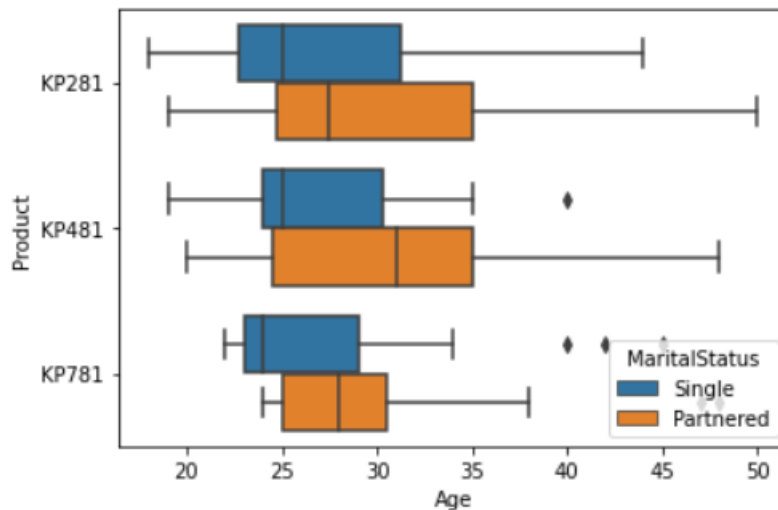
- From the box plot it is clear that people who run more than average 160miles per week prefer KP781.
- There are outliers in all three categories.
  - KP281-3 outliers
  - KP481-2 outliers
  - KP781-1 outliers

## Multivariate Analysis:

Customer age and marital status for different product categories:

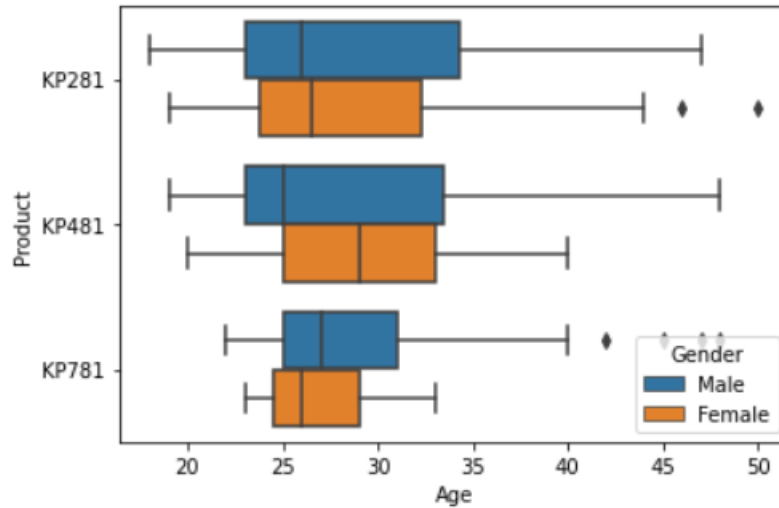


This graph shows that all three products have a similar “Age” range and people who are married prefer treadmills with greater “age” values. But there are few outliers for single people.

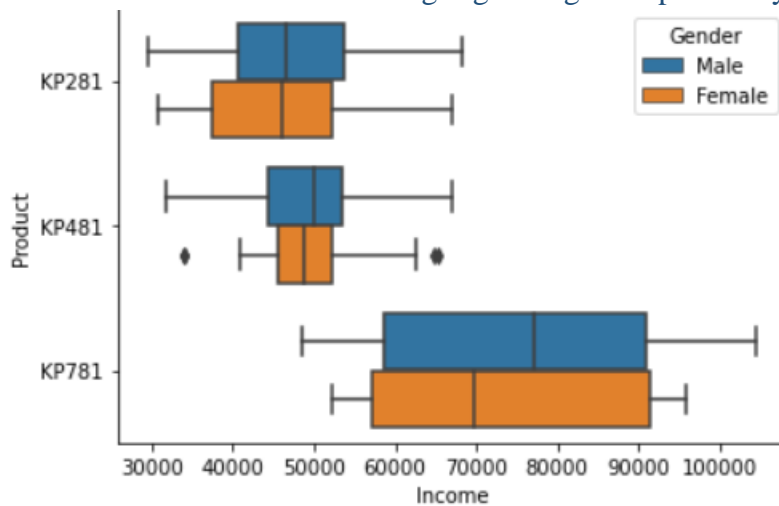


This boxplot concludes that the age group between 25 to 30 prefers KP781 irrespective of their marital status.

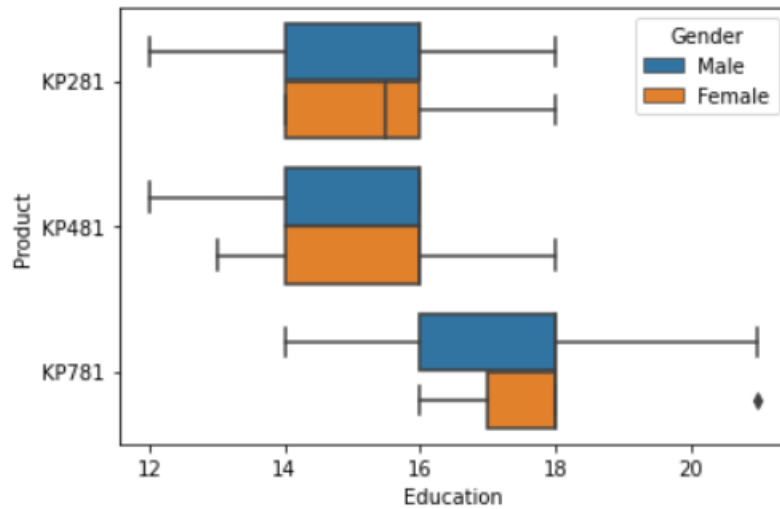
**Analysis of different product categories with respect to gender for different attributes:**



- It shows that the IQR of all products is almost the same for both genders.
- Average age of males and females is almost equal for KP-281 and KP-781 but for KP-481 the difference in average age is high comparatively.

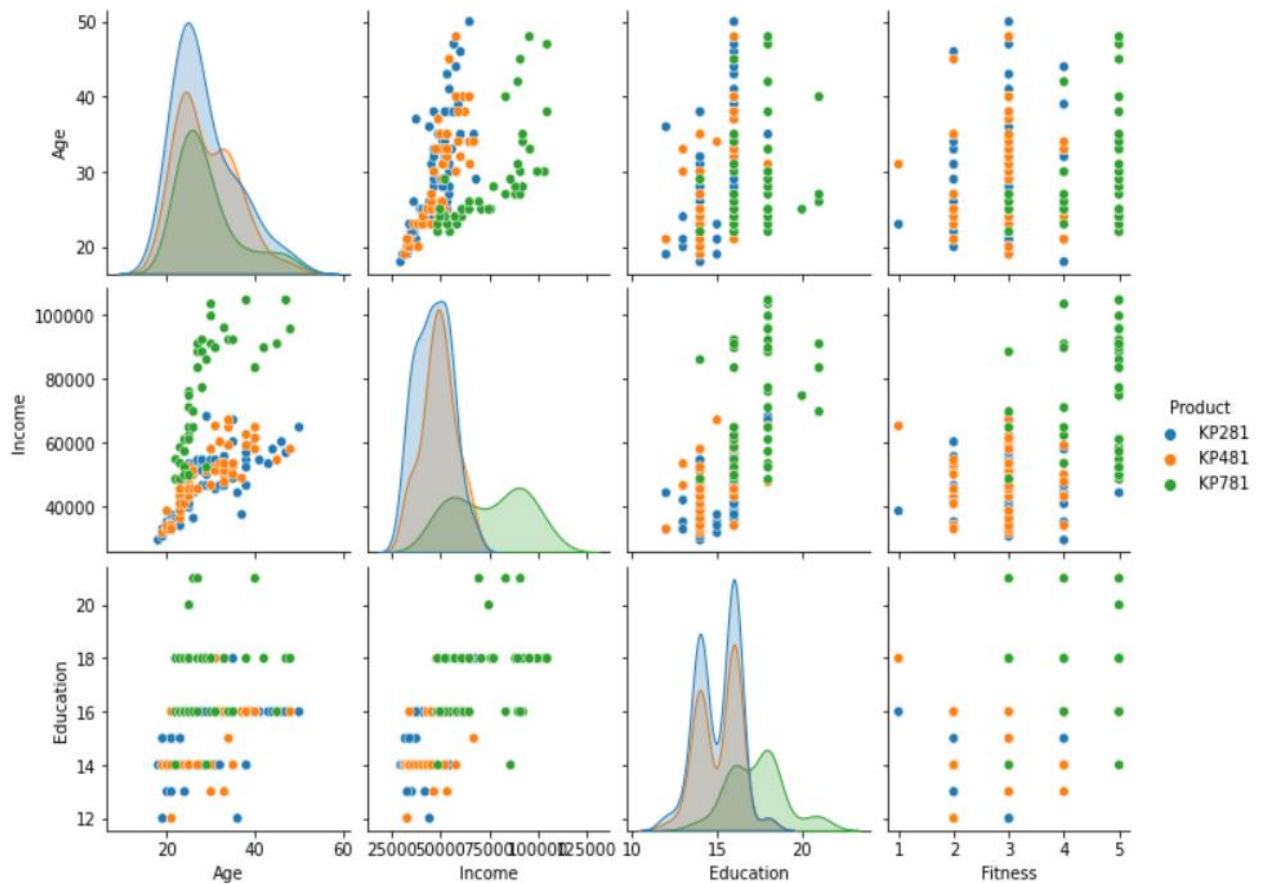


- As seen in previous analysis that high income range people prefer KP-781 the same can be seen from this plot.
- Otherwise gender does not show very impact on income range of product. It is almost same for both males and females.



- It shows that KP481 and KP281 has male female IQR is same.
- But in KP781 females education range in years is 17 to 18 years (IQR=1 year ) and for males age range 16 to 18 years(IQR=2 years)

## Pairplot:



From this pair plot we can conclude that:

- Higher income people have higher education and they prefer KP\_781 and fitness level is also good.
- People with moderate income choose all three categories of treadmills.

# Probability Analysis:

## For “Income”

Let low-income <\$50K and high-income >=\$50K

```
df_low_income=df[df["Income"]<50000]  
df_high_income=df[df["Income"]>=50000]
```

```
pd.crosstab(df_low_income["Product"],df_low_income["Income"],margins=True)
```

Income	29562	30699	31836	32973	34110	35247	36384	37521	38658	39795	...	43206	44343	45480	46617	47754	48556	48658	48891	49801	All
Product																					
KP281	1	1	1	3	2	5	3	2	3	2	...	1	4	5	7	0	0	0	2	0	48
KP481	0	0	1	2	3	0	1	0	2	0	...	4	0	9	1	2	0	0	3	0	30
KP781	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	2	1	0	2	5
All	1	1	2	5	5	5	4	2	5	2	...	5	4	14	8	2	2	1	5	2	83

4 rows × 22 columns

```
pd.crosstab(df_high_income["Product"],df_high_income["Income"],margins=True)
```

Income	50028	51165	52290	52291	52302	53439	53536	54576	54781	55713	...	88396	89641	90886	92131	95508	95866	99601	103336	104581	All
Product																					
KP281	2	3	0	0	6	3	0	7	0	1	...	0	0	0	0	0	0	0	0	0	32
KP481	5	4	0	0	3	5	0	1	0	0	...	0	0	0	0	0	0	0	0	0	30
KP781	0	0	1	1	0	0	1	0	1	0	...	2	2	3	3	1	1	1	1	2	35
All	7	7	1	1	9	8	1	8	1	1	...	2	2	3	3	1	1	1	1	2	97

4 rows × 42 columns

From above analysis results the following data

Product	Low-Income	High-Income	Total
KP-281	48	32	80
KP-481	30	30	60
KP-781	5	35	40
Total	83	97	180

## Marginal Probability

$P(\text{High-Income}) = 97/180 = 0.54$

$P(\text{Low-Income}) = 83/180 = 0.46$

54% people are in high income group as per our assumption

**Conditional probability:**

$$P(\text{High-Income/KP-281}) = 32/80 = 0.40$$

$$P(\text{Low-Income/KP-281}) = 48/80 = 0.60$$

$$P(\text{High-Income/KP-481}) = 30/60 = 0.5$$

$$P(\text{Low-Income/KP-481}) = 30/60 = 0.5$$

$$P(\text{High-Income/KP-781}) = 35/40 = 0.875$$

$$P(\text{Low-Income/KP-781}) = 5/40 = 0.125$$

From this analysis, it is concluded that

- KP-281 is preferred by low income group.
- KP-481 is preferred by all income groups
- KP-781 is highly preferred by high income group.

**For “Education”**

Assumption: People with more than 15 years of experience are highly educated



```
In [6]: pd.crosstab(df["Product"],df["Education"])
```

```
Out[6]:
```

Education	12	13	14	15	16	18	20	21
Product								
KP281	2	3	30	4	39	2	0	0
KP481	1	2	23	1	31	2	0	0
KP781	0	0	2	0	15	19	1	3

```
In [65]: High_edu=df[df["Education"]>15]  
Low_edu=df[df["Education"]<=15]
```

```
In [66]: a=pd.crosstab(High_edu["Product"],High_edu["Education"],margins=True)  
a
```

```
Out[66]:
```

Education	16	18	20	21	All
Product					
KP281	39	2	0	0	41
KP481	31	2	0	0	33
KP781	15	19	1	3	38
All	85	23	1	3	112

```
In [67]: b=pd.crosstab(Low_edu["Product"],Low_edu["Education"],margins=True)
b
```

```
Out[67]:
```

	Education	12	13	14	15	All
Product						
KP281		2	3	30	4	39
KP481		1	2	23	1	27
KP781		0	0	2	0	2
All		3	5	55	5	68

```
In [68]: c=pd.DataFrame({"High_Edu":a["All"],"Low_Edu":b["All"]})
c
```

```
Out[68]:
```

	High_Edu	Low_Edu
Product		
KP281	41	39
KP481	33	27
KP781	38	2
All	112	68

```
In [69]: c['Sum_of_row'] = c.loc[:, "High_Edu": "Low_Edu"].sum(axis = 1)
c.reset_index(inplace=True)
c
```

```
Out[69]:
```

	Product	High_Edu	Low_Edu	Sum_of_row
0	KP281	41	39	80
1	KP481	33	27	60
2	KP781	38	2	40
3	All	112	68	180

### Marginal Probability:

$P(\text{High\_Edu}) = 112/180 = 0.622$

$P(\text{Low\_Edu}) = 68/180 = 0.377$

62 % of people are highly educated (has an education more than 15 years)

$$P(KP281) = 80/180 = 0.44$$

$$P(KP481) = 60/180 = 0.33$$

$$P(KP781) = 40/180 = 0.22$$

Most of the people by the KP281 model of treadmill about 44%

### **Conditional Probability:**

$$P(\text{Low\_Edu} \mid KP281) = 39/80 = 0.49$$

$$P(\text{High\_Edu} \mid KP281) = 41/80 = 0.51$$

In the sale of KP281 there is no effect of education.

$$P(\text{Low\_Edu} \mid KP481) = 27/60 = 0.45$$

$$P(\text{High\_Edu} \mid KP481) = 33/60 = 0.55$$

55% highly educated people prefer KP481/

$$P(\text{Low\_Edu} \mid KP781) = 2/40 = 0.05$$

$$P(\text{High\_Edu} \mid KP781) = 38/40 = 0.95$$

From above analysis it is clear that highly educated people prefer KP781.

$$P(KP281 \mid \text{High\_Edu}) = 41/112 = 0.37$$

$$P(KP481 \mid \text{High\_Edu}) = 33/112 = 0.29$$

$$P(KP781 \mid \text{High\_Edu}) = 38/112 = 0.34$$

### **Recommendation:**

- People with an income more than \$50000 and an education in years is more than 15 prefer KP-781.
- KP281 and KP-481 are preferred by almost same income(<\$48K) and education (<15years) group.