NAME : MAHESH DARAKHE.

BATCH : DSML Nov22 Beginner Morning Tue

```
import numpy as np
import pandas as pd
import seaborn as sns
from matplotlib import pyplot as plt

df=pd.read_csv('https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/001/125/original/aerofit_treadmill.csv?1639992749')
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 9 columns):
 #   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 
dtypes: int64(6), object(3)
memory usage: 12.8+ KB
```

Insight:

Given database contains 180 rows and 9 columns

given database does not have any null values

```
df.shape
```

```
(180, 9)
```

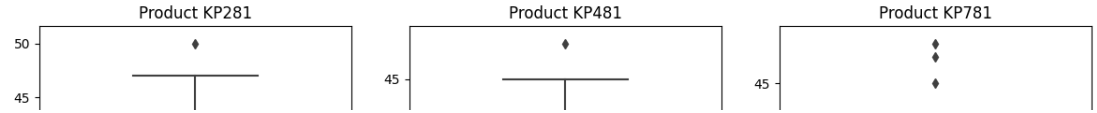
```
df.nunique()
```

```
Product           3
Age              32
Gender            2
```

```
Education         8
MaritalStatus     2
Usage             6
Fitness           5
Income           62
Miles            37
dtype: int64
```

Given chart shows number of unique elements for each columns from this we can say columns Age,Income and Miles are catagorical columns.

```
fig, axes = plt.subplots(1, 3, figsize=(12, 5))
sns.boxplot(x='Product', y='Age', data=df[df['Product'] == 'KP281'], showmeans=True, ax=axes[0])
axes[0].set_title('Product KP281')
sns.boxplot(x='Product', y='Age', data=df[df['Product'] == 'KP481'], showmeans=True, ax=axes[1])
axes[1].set_title('Product KP481')
sns.boxplot(x='Product', y='Age', data=df[df['Product'] == 'KP781'], showmeans=True, ax=axes[2])
axes[2].set_title('Product KP781')
plt.tight_layout()
plt.show()
```

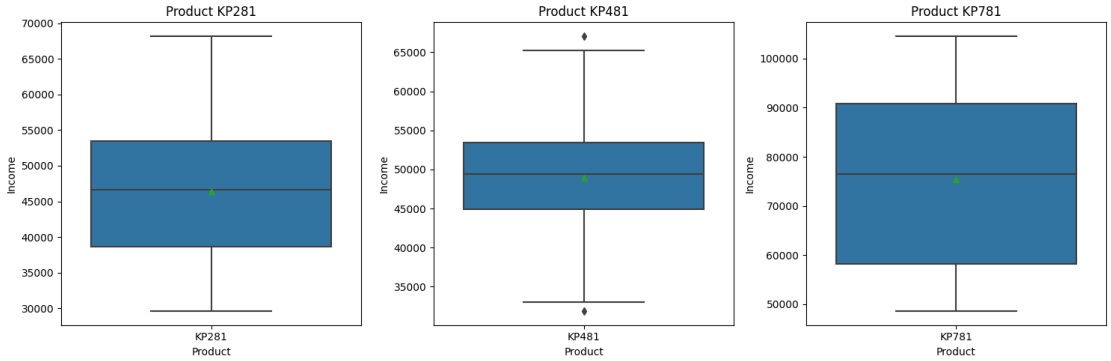


Insight:From the above plots if a user age is above 30 it is more likely that they will buy KP281 or KP481

Also there is a difference Between Mean and Median Values so there are outliers present in the given data

For KP781 there are a lot of outliers present above age 40 so that can also mean a lot of aged people prefer to buy KP781 insted of KP281 and KP481


```
fig, axes = plt.subplots(1, 3, figsize=(15, 5))
sns.boxplot(x='Product', y='Income', data=df[df['Product'] == 'KP281'], showmeans=True, ax=axes[0])
axes[0].set_title('Product KP281')
sns.boxplot(x='Product', y='Income', data=df[df['Product'] == 'KP481'], showmeans=True, ax=axes[1])
axes[1].set_title('Product KP481')
sns.boxplot(x='Product', y='Income', data=df[df['Product'] == 'KP781'], showmeans=True, ax=axes[2])
axes[2].set_title('Product KP781')
plt.tight_layout()
plt.show()
```



Insights:

If a customer has income Between 6 to 9 lakhs that Customer is Highly Likely to buy KP781

Outliers can be seen for KP481 mening people with very high income and low income are also buying KP481

There are no visible outliers for KP281

```
df.describe()
```

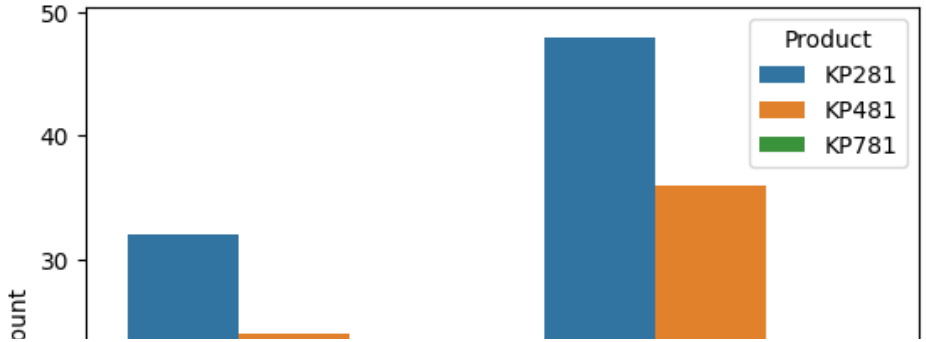
	Age	Education	Usage	Fitness	Income	Miles
count	180.000000	180.000000	180.000000	180.000000	180.000000	180.000000
mean	28.788889	15.572222	3.455556	3.311111	53719.577778	103.194444
std	6.943498	1.617055	1.084797	0.958869	16506.684226	51.863605
min	18.000000	12.000000	2.000000	1.000000	29562.000000	21.000000
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

Insight:

There is a little difference in mean and median values so there are a few outliers in every column

```
sns.countplot(df,x='MaritalStatus',hue='Product')
```

```
<Axes: xlabel='MaritalStatus', ylabel='count'>
```



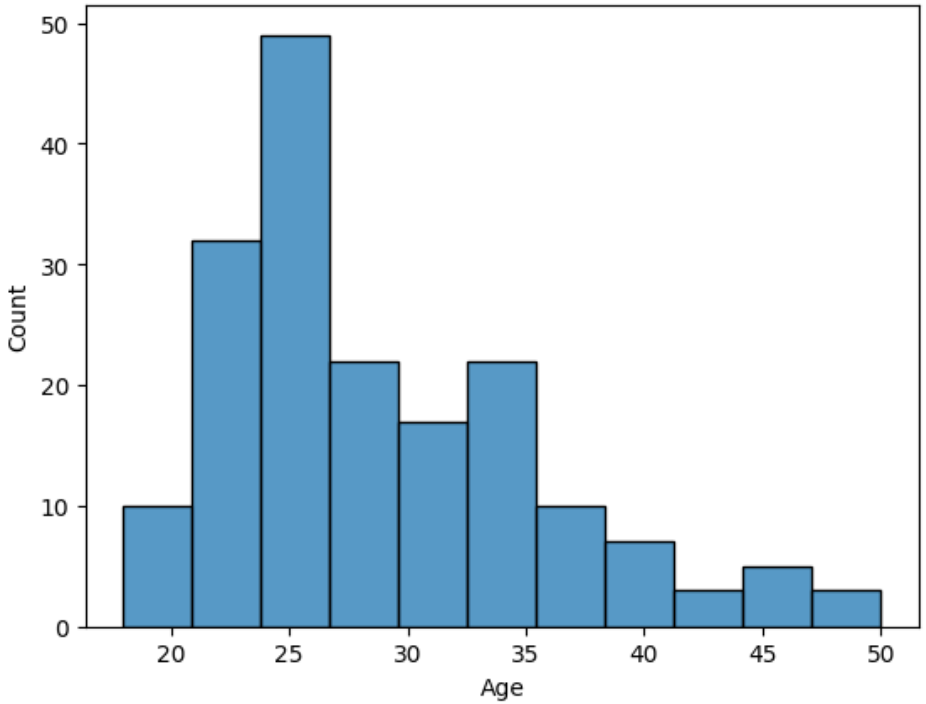
Insight:

Single People are less likely to buy one of the products.

More people prefer buying KP281 than KP481 than KP781

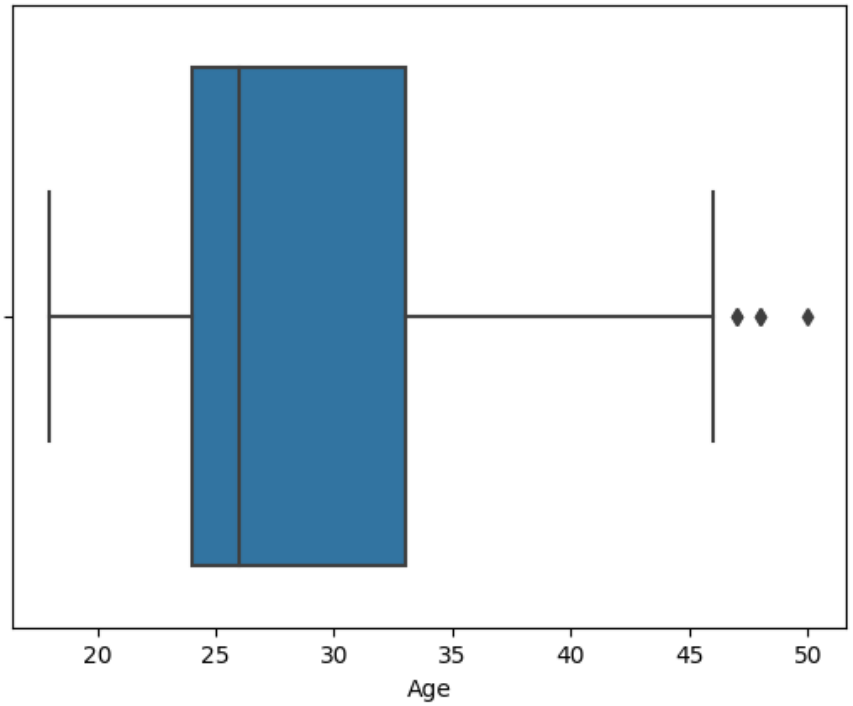

```
sns.histplot(df,x='Age')
```

```
<Axes: xlabel='Age', ylabel='Count'>
```



```
sns.boxplot(df,x='Age')
```

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



Insight:

Above plots shows the density of ages of customes.With a few outliers present and age of customers is in between 24 to 33

```
sns.distplot(df['Income'])
plt.show()
```

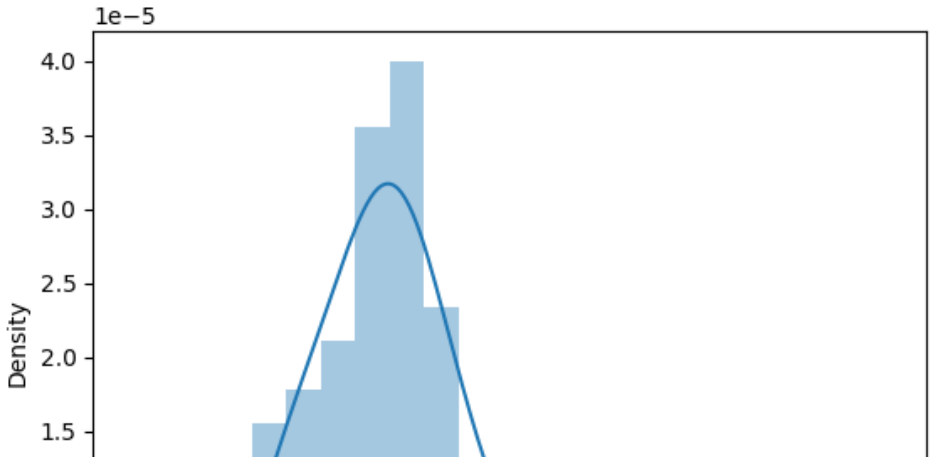
```
<ipython-input-29-50bd26e4ed8e>:1: UserWarning:

`distplot` is a deprecated function and will be removed in seaborn v0.14.0.

Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).

For a guide to updating your code to use the new functions, please see
https://gist.github.com/mwaskom/de44147ed2974457ad6372750bbe5751
```

```
sns.distplot(df['Income'])
```



Insight:

Above plot shows customers with income in range of 30000 rs to 60000 rs buy one of the products


```
cross_tab = pd.crosstab(index=df['Gender'], columns=df['Product'])
marginal_probabilities = cross_tab / cross_tab.sum().sum()
print('Marginal Probabilities :')
print(marginal_probabilities)
```

```
Marginal Probabilities :
Product     KP281     KP481     KP781
Gender
Female   0.222222  0.161111  0.038889
Male     0.222222  0.172222  0.183333
```

Above table shows probabilities of products with respect to males and females.

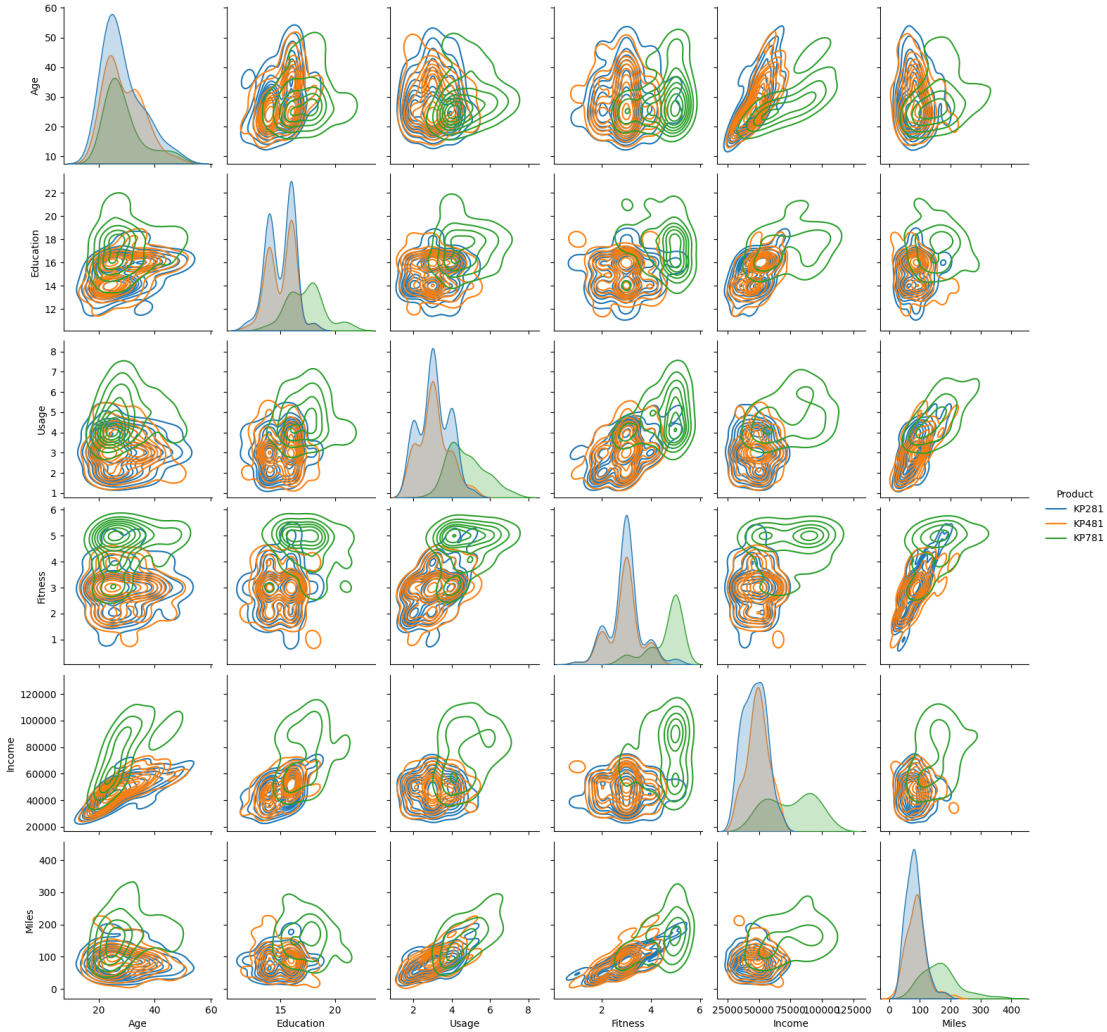
```
cross_tab = pd.crosstab(index=df['Gender'].count(), columns=df['Product'])
marginal_probabilities = cross_tab / cross_tab.sum().sum()
print('Marginal Probabilities :')
print(marginal_probabilities)
```

```
Marginal Probabilities :
Product     KP281     KP481     KP781
row_0
180      0.444444  0.333333  0.222222
```

Above table shows probabilities of the products.

```
sns.pairplot(df,hue='Product',kind='kde')
```

```
<seaborn.axisgrid.PairGrid at 0x7eea4d768e50>
```



Recommendations:

1. There's a high preference for KP281 and KP481 among younger customers, you could focus on marketing these products to that demographic.
2. you can create campaigns that specifically target individuals in the age range where KP281 and KP481 are preferred, and another campaign for those with higher incomes who are more likely to purchase KP781.

3. since customers with incomes between 6 to 9 lakhs are more likely to buy KP781, you can offer pricing plans that cater to this income range.
4. customers purchasing KP481 are found to have diverse income levels, consider introducing bundle offers. This might appeal to both high and low-income customers who prefer KP481.