AEROFIT BUSINESS CASE STUDY

About Aerofit:

Aerofit is a leading brand in the field of fitness equipment. Aerofit provides a product range including machines such as treadmills, exercise bikes, gym equipment, and fitness accessories to cater to the needs of all categories of people.

Business Problem to solve:

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. To investigate whether there are differences across the product with respect to customer characteristics.

Perform descriptive analytics to create a customer profile for each AeroFit treadmill product by developing appropriate tables and charts. For each AeroFit treadmill product, construct two-way contingency tables and compute all conditional and marginal probabilities along with their insights/impact on the business.

The company collected the data on individuals who purchased a treadmill from the AeroFit stores during the prior three months.

The dataset has the following features:

Product Purchased: KP281, KP481, or KP781

Age: In years

Gender: Male/Female

Education: In years

MaritalStatus: Single or partnered

Usage: The average number of times the customer plans to use the treadmill each week.

Income: Annual income (in $)

Fitness: Self-rated fitness on a 1-to-5 scale, where 1 is the poor shape and 5 is the excellent shape.

Miles: The average number of miles the customer expects to walk/run each week

Important Python Liberaries to Import | Reading Files | Basic Data Exploration

```
#IMPORTING LIBERARIES:
import numpy as numpy
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')

#READING THE FILE:
from google.colab import drive
drive.mount('/content/drive')
df = pd.read_csv('/content/drive/MyDrive/python files/aerofit_treadmill.csv')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

```
#EXPLORAING THE DATA FOR THE FIRST TIME:
df.head()
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles
0	KP281	18	Male	14	Single	3	4	29562	112
1	KP281	19	Male	15	Single	2	3	31836	75
2	KP281	19	Female	14	Partnered	4	3	30699	66
3	KP281	19	Male	12	Single	3	3	32973	85
4	KP281	20	Male	13	Partnered	4	2	35247	47

Next steps: Generate code with df | View recommended plots | New interactive sheet

Basic Observation on the Data

```
#DATA SHAPE:

df.shape
#Our data has 180 Rows and 9 Columns
```

```
(180, 9)
```

```
df.columns
```

```
Index(['Product', 'Age', 'Gender', 'Education', 'MaritalStatus', 'Usage',
       'Fitness', 'Income', 'Miles'],
      dtype='object')
```

```
# DATA INFO:

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
```

```
# FINDING TOTAL NULL VALUES IN THE DATA:

print('Total Null Values:',df.isna().sum().sum())
```

```
Total Null Values: 0
```

```
# FINDING STAISTICAL DATA FOR THE GIVEN NETFLIX DATA:

df.describe(include='all')
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles
count	180	180.000000	180	180.000000	180	180.000000	180.000000	180.000000	180.000000
unique	3	NaN	2	NaN	2	NaN	NaN	NaN	NaN
top	KP281	NaN	Male	NaN	Partnered	NaN	NaN	NaN	NaN
freq	80	NaN	104	NaN	107	NaN	NaN	NaN	NaN
mean	NaN	28.788889	NaN	15.572222	NaN	3.455556	3.311111	53719.577778	103.194444
std	NaN	6.943498	NaN	1.617055	NaN	1.084797	0.958869	16506.684226	51.863605
min	NaN	18.000000	NaN	12.000000	NaN	2.000000	1.000000	29562.000000	21.000000
25%	NaN	24.000000	NaN	14.000000	NaN	3.000000	3.000000	44058.750000	66.000000
50%	NaN	26.000000	NaN	16.000000	NaN	3.000000	3.000000	50596.500000	94.000000
75%	NaN	33.000000	NaN	16.000000	NaN	4.000000	4.000000	58668.000000	114.750000
max	NaN	50.000000	NaN	21.000000	NaN	7.000000	5.000000	104581.000000	360.000000

KEY INSIGHTS:

Average Age: ~29 years — a young demographic.

Income: Ranges from $29.5k to $104.5k. Skewed toward middle income, with mean ~$53.7k.

Miles (Usage): Highly variable; some users travel very high distances (~360 miles), suggesting some heavy fitness users.

Dataset is dominated by Male Consumers (104)

Most of the customers in the dataset are Married (107)

```
# FINDINNG THE NUMBER OF UNIQUE VALUES IN ALL THE COLUMNS IN THE DATA:

df.nunique()
```

	0
Product	3
Age	32
Gender	2
Education	8
MaritalStatus	2
Usage	6
Fitness	5
Income	62
Miles	37

dtype: int64

Since the Data given is clean, now we can analyse the data accordingly and we will create additional columns for our understanding if necessary.

We have created 4 different columns namely:

1. Income Range: Different income ranges are created to place different income values in a particular range for better analytical understanding.

2. Income Class: Here we have taken assumption and divided the incomes into 3 different classes:

> Low Level Income: Income ranges between $0 and $30000

> Mid Level Income: Income ranges between $30000 and $60000

> High Level Income: Income ranges between $60000 and above

3. Price of the product: we have given prices of the product:

> The KP281 is an entry-level treadmill that sells for $1,500.

> The KP481 is for mid-level runners that sell for $1,750.

> The KP781 treadmill is having advanced features that sell for $2,500.

4. Age Group: Since we have all the customers of different ages, we have categorised these ages in different ranges i.e. 18-30, 30-40, 40-50 and above 50.

```
df['IncomeRange'] = pd.cut(df['Income'],
                           bins = [20000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 110000],
                           labels = ['20000-30000', '30000-40000', '40000-50000', '50000-60000', '60000-70000', '70000-80000', '80000-90000', '90000-100000', '100000-110000'],
                           right = False)
df['IncomeClass'] = pd.cut(df['Income'],
                           bins = [0,30000,60000,110000],
                           labels = ['Low Level Income','Mid Level Income','High Level Income'],
                           right = False)

df['AgeGroup'] = pd.cut(df['Age'],
```

```
                      bins = [18,30,40,50,100],
                      labels = ['18-30','30-40','40-50','50 and above'],
                      right = False)

price_map = {'KP281': 1500, 'KP481': 1750, 'KP781': 2500}
df['Price'] = df['Product'].map(price_map)

df
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles	IncomeRange	IncomeClass	AgeGroup	Price
0	KP281	18	Male	14	Single	3	4	29562	112	20000-30000	Low Level Income	18-30	1500
1	KP281	19	Male	15	Single	2	3	31836	75	30000-40000	Mid Level Income	18-30	1500
2	KP281	19	Female	14	Partnered	4	3	30699	66	30000-40000	Mid Level Income	18-30	1500
3	KP281	19	Male	12	Single	3	3	32973	85	30000-40000	Mid Level Income	18-30	1500
4	KP281	20	Male	13	Partnered	4	2	35247	47	30000-40000	Mid Level Income	18-30	1500
...	...	...	...	...	...	...	...	...	...	...	...	...	...
175	KP781	40	Male	21	Single	6	5	83416	200	80000-90000	High Level Income	40-50	2500
176	KP781	42	Male	18	Single	5	4	89641	200	80000-90000	High Level Income	40-50	2500
177	KP781	45	Male	16	Single	5	5	90886	160	90000-100000	High Level Income	40-50	2500
178	KP781	47	Male	18	Partnered	4	5	104581	120	100000-110000	High Level Income	40-50	2500
179	KP781	48	Male	18	Partnered	4	5	95508	180	90000-100000	High Level Income	40-50	2500

180 rows × 13 columns

Next steps: Generate code with df | View recommended plots | New interactive sheet

```
df['IncomeClass'].value_counts()
```

IncomeClass	count
Mid Level Income	137
High Level Income	42
Low Level Income	1

dtype: int64

> Most of the buyers are under Mid Level Income Class and some of the buyers are earning a High Level Income.

> Also in our data only 1 Buyer is under the Low Level Income Category. This is probably due to the lack of proper data. But we need to consider that this data is for the last 3 months only.

Gender-wise Income Information in the Data:

```
df.groupby('Gender')['Income'].aggregate(['mean','median','min','max'])
```

Gender	mean	median	min	max
Female	49828.907895	47754.0	30699	95866
Male	56562.759615	52302.0	29562	104581

KEY INSIGHTS:

Average (Mean) Income:

Males earn significantly more on average (₹56,563) compared to females (₹49,829).

This indicates a gender income gap of about ₹6,734 on average.

Median Income:

The median income for males is ₹52,302, while for females it's ₹47,754.

Since the median is less affected by extreme values, this further confirms that males consistently earn more than females across the distribution.

Minimum and Maximum Incomes:

Minimum income is similar for both: ₹29,562 (male) vs ₹30,699 (female).

Maximum income for males is ₹104,581, which is noticeably higher than the female maximum of ₹95,866.

GENDER WISE PERCENTAGE SHARE IN THE PURCHASE OF PRODUCT IN THE LAST 3 MONTHS

```
pd.crosstab([df['Gender']],df['Product'], margins=True)
```

Gender	Product KP281	KP481	KP781	All
Female	40	29	7	76
Male	40	31	33	104
All	80	60	40	180

```
print('Male:', round(104/180*100,2),'%')
print('Female:', round(76/180*100,2),'%')
```

```
Male: 57.78 %
Female: 42.22 %
```

Probability of Male buying the Most Premium Product of Aerofit i.e. KP781:

```
print('Probability of Male buying the Most Premium Product of Aerofit i.e. KP781:', round(33/180,2))
```

```
Probability of Male buying the Most Premium Product of Aerofit i.e. KP781: 0.18
```

Probability of Female buying the Most Premium Product of Aerofit i.e. KP781:

```
print('Probability of Male buying the Most Premium Product of Aerofit i.e. KP781:', round(7/180,2))
```

```
Probability of Male buying the Most Premium Product of Aerofit i.e. KP781: 0.04
```

EFFECT OF INCOME ON THE PERFORMANCE OF THE PRODUCT:

CONDITIONAL PROBABILITIES:

```
pd.crosstab(df['IncomeClass'],df['Product'], margins=True)
```

Product	KP281	KP481	KP781	All
IncomeClass				
Low Level Income	1	0	0	1
Mid Level Income	73	53	11	137
High Level Income	6	7	29	42
All	80	60	40	180

Probability of Mid Level Income Class customers given that the product purchased is KP281:

```
print('Probability ( Mid-Level Income | KP281 ) :', round(73/80, 2))
```

```
Probability ( Mid-Level Income | KP281 ) : 0.91
```

Probability of Mid Level Income Class customers given that the product purchased is KP481:

```
print('Probability ( Mid-Level Income | KP481 ) :', round(53/60, 2))
```

```
Probability ( Mid-Level Income | KP481 ) : 0.88
```

Probability of Mid Level Income Class customers given that the product purchased is KP781:

```
print('Probability ( Mid-Level Income | KP781 ) :', round(11/40, 2))
```

```
Probability ( Mid-Level Income | KP781 ) : 0.28
```

KEY INSIGHTS:

KP281 and KP481 Are Strongly Preferred by Mid-Level Income Customers:

> A very high proportion of KP281 (91%) and KP481 (88%) buyers belong to the Mid-Level Income group ($30k to $60k)

> These products are well-aligned with affordability and perceived value for this income segment.

KP781 Appeals Less to Mid-Level Income Group:

> Only 28% of KP781 buyers are from the mid-level income group.

> This indicates KP781 is more attractive to High-Income customers (above $60k), likely due to premium pricing or advanced features.

Probability of Low Level Income Class Customers given that product purchased is KP281:

```
print('Probability ( Low-Level Income | KP281 ) :', round(1/80, 2))
```

```
Probability ( Low-Level Income | KP281 ) : 0.01
```

KEY INSIGHTS:

- KP281 Has Minimal Appeal Among Low-Income Customers (below $30k):

> Only 1% of KP281 buyers are from the Low-Level Income group.

- Even though KP281 is likely the cheapest model, it still doesn't attract low-income customers, possibly due to:

> Limited purchasing power,

> Preference for no equipment, or

> Lack of awareness/access.

EFFECT OF AGE ON THE PERFORMANCE OF THE PRODUCT:

```
pd.crosstab([df['AgeGroup']], df['Product'], margins=True)
```

Product	KP281	KP481	KP781	All
AgeGroup				
18-30	53	33	27	113
30-40	20	22	8	50
40-50	6	5	5	16
50 and above	1	0	0	1
All	80	60	40	180

Probability of Younger Generation i.e. 18-30 age group purchasing the product given that the product is entry level i.e. KP281

```
print('Probability ( Age Group (18-30) | KP281 ) :', round(53/80, 2))
```

```
Probability ( Age Group (18-30) | KP281 ) : 0.66
```

Probability of Middle Aged Generation i.e. 30-50 age group purchasing the product given that the product is entry level i.e. KP281

```
print('Probability ( Age Group (30-50) | KP281 ) :', round((20+6)/80, 2))
```

```
Probability ( Age Group (30-50) | KP281 ) : 0.33
```

Probability of Younger Generation i.e. 18-30 age group purchasing the product given that the product is middle variant i.e. KP481

```
print('Probability ( Age Group (18-30) | KP481 ) :', round(33/60, 2))
```

```
Probability ( Age Group (18-30) | KP481 ) : 0.55
```

Probability of Middle aged Generation i.e. 30-40 age group purchasing the product given that the product is middle variant i.e. KP481

```
print('Probability ( Age Group (30-50) | KP481 ) :', round((22+5)/60, 2))
```

```
Probability ( Age Group (30-50) | KP481 ) : 0.45
```

KEY INSIGHTS:

KP281:

- Most popular among young adults (18–30) — two-thirds of KP281 buyers are in this group.
- Slightly lower appeal to older age brackets.

KP481:

- Has a more balanced age distribution between 18–30 and 30–50.
- This suggests KP481 may be appealing to a broader age range, potentially due to features, price, or perceived value.

EFFECT OF FITNESS LEVEL ON THE PERFORMANCE OF THE PRODUCT:

```
pd.crosstab([df['Fitness']], df['Product'], margins=True)
```

Product	KP281	KP481	KP781	All
Fitness				
1	1	1	0	2
2	14	12	0	26
3	54	39	4	97
4	9	8	7	24
5	2	0	29	31
All	80	60	40	180

Probability that a random selected customer bought KP781:

```
print('Probability ( KP781 ) :', round(40/180, 2))
```

```
Probability ( KP781 ) : 0.22
```

Probability that a random selected customer bought KP781 has a Fitness Level 5:

```
print('Probability ( Fitness Level 5 | KP781 ) :', round(29/40, 2))
```

```
Probability ( Fitness Level 5 | KP781 ) : 0.72
```

Probability that the customer has a Fitness Level 3:

```
print('Probability ( Fitness Level 3 ) :', round(97/180, 2))
```

```
Probability ( Fitness Level 3 ) : 0.54
```

Probability that the customer who has a fitness level 3 bought KP481::

```
print('Probability ( KP481 | fitness level 3 ) :', round(39/97, 2))
```

```
Probability ( KP481 | fitness level 3 ) : 0.4
```

Probability that the customer who has a fitness level 3 bought KP281::

```
print('Probability ( KP281 | fitness level 3 ) :', round(54/97, 2))
```

```
Probability ( KP281 | fitness level 3 ) : 0.56
```

Probability that the customer who bought KP281 has a fitness level 3::

```
print('Probability ( fitness level 3 | KP281 ) :', round(54/80, 2))
```

```
Probability ( fitness level 3 | KP281 ) : 0.68
```

Probability that a customer with Fitness level 5 bought KP781:

```
print('Probability ( KP781 | fitness level 5 ) :', round(29/31, 2))
```

```
Probability ( KP781 | fitness level 5 ) : 0.94
```

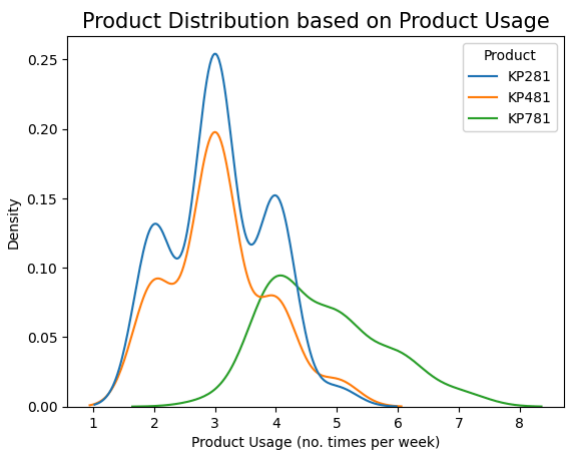
KEY INSIGHTS:

- The majority of customers (**54%**) have a **moderate fitness level (Fitness Level 3)**.
- **KP281** is the most preferred product among Fitness Level 3 customers, with **56%** of them choosing it.
- **40%** of Fitness Level 3 customers also purchased **KP481**, making it the second most popular product for this group.
- Among all customers who bought **KP281**, **68%** have Fitness Level 3 — showing a strong alignment between the product and the average fitness profile.
- **KP781** is purchased by only **22%** of all customers, making it the **least popular** product overall.
- However, **72%** of those who bought **KP781** have a **high fitness level (Fitness Level 5)**, indicating it appeals to **advanced fitness users**.
- Overall, **KP281** aligns best with the most common customer fitness level, while **KP781** caters to a **niche, highly fit** customer segment.

DISTRIBUTION OF VARIOUS ASPECTS AND MAKING CUSTOMER PROFILE IN AEROFIT DATASET:

PRODUCT DISTRIBUTION BASED ON PRODUCT USAGE:

```
sns.kdeplot(x=df['Usage'], hue=df['Product'])
plt.title('Product Distribution based on Product Usage', size=15)
plt.xlabel('Product Usage (no. times per week)')
plt.show()
```



KEY INSIGHTS:

KP281:

> Most Frequently Used Product Peak usage around 3 times per week, with a secondary bump near 4.
>
> Shows the highest peak density, suggesting it is the most consistently used product among customers.
>
> Usage is tightly concentrated between 2 and 4 times per week.
>
> Indicates a strong user base with regular fitness routines.

KP481 - Moderate Usage

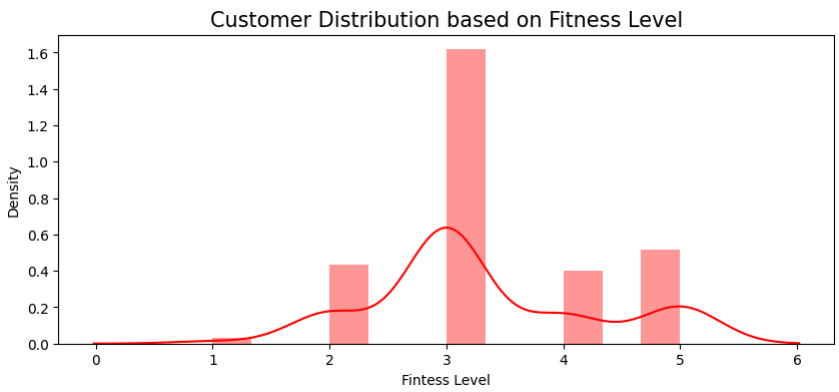
> Peaks similarly around 3 times per week, but with a lower peak than KP281.
>
> Less density overall, suggesting it's less popular or used less regularly than KP281.
>
> Also concentrated in the 2–4 usage range.

KP781:

> High Variability, Less Frequent Use Flatter and wider distribution, indicating greater variability in usage frequency.
>
> Peak is broader around 4–5 times/week, with usage extending beyond 6 times per week.
>
> Lower peak density suggests less frequent or consistent use, but it appeals to users with more intensive fitness routines.

CUSTOMER DISTRIBUTION BASED ON FITNESS LEVEL:

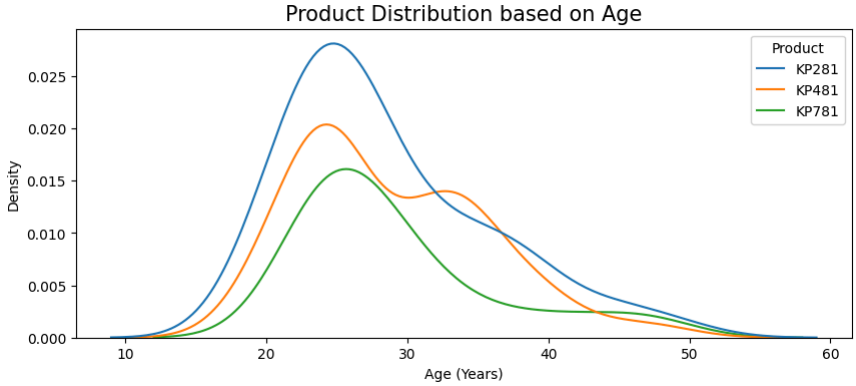
```
plt.figure(figsize=(10,4))
sns.distplot(x=df['Fitness'], color='red')
plt.title('Customer Distribution based on Fitness Level', size=15)
plt.xlabel('Fintess Level')
plt.show()
```



KEY INSIGHTS:

- The majority of customers have rated themselves at **Fitness Level 3 (out of 5)**, indicating an **average level of fitness**. This suggests that many customers are likely purchasing fitness products to **incorporate regular physical activity** into their daily routines and improve their overall health.
- A significant portion of customers have rated themselves at **Fitness Levels 4 and 5**, reflecting a **high level of fitness and health consciousness**. These individuals are likely using the products to **maintain their physique**, enhance their performance, and stay committed to a **healthy and active lifestyle**.

```
plt.figure(figsize=(10,4))
sns.kdeplot(x=df['Age'], hue=df['Product'])
plt.title('Product Distribution based on Age', size=15)
plt.xlabel('Age (Years)')
plt.show()
```



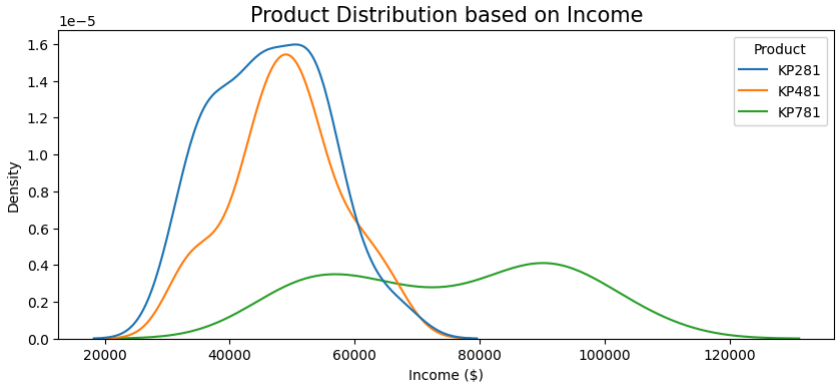
KEY INSIGHTS:

KP281 peaks around **age 25**, confirming it's most popular with **younger** users.

KP481 has a slightly **older peak**, around **28–30 years**, suggesting it's preferred by users with a bit more fitness experience.

KP781 shows broader usage across **mid-20s to late 40s**, appealing to a wider but more niche, older user base.

```
plt.figure(figsize=(10,4))
sns.kdeplot(x=df['Income'],hue=df['Product'])
plt.title('Product Distribution based on Income', size=15)
plt.xlabel('Income ($)')
plt.show()
```



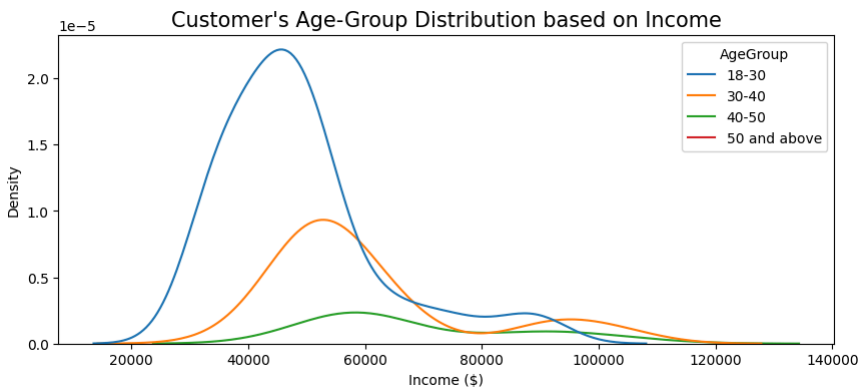
KEY INSIGHTS:

KP281 is most dense around the **$30,000 – $40,000** income range, indicating strong appeal to middle-income consumers.

KP481 is also concentrated in the **$35,000–$45,000** range but with slightly lower density, showing it's a natural next step after KP281.

KP781 has a distinct peak at higher incomes (**$85,000–$100,000**), confirming it's favored by **high-income individuals**, likely due to being a premium product.

```
plt.figure(figsize=(10,4))
sns.kdeplot(x=df['Income'],hue=df['AgeGroup'])
plt.title("Customer's Age-Group Distribution based on Income", size=15)
plt.xlabel('Income ($)')
plt.show()
```



KEY INSIGHTS:

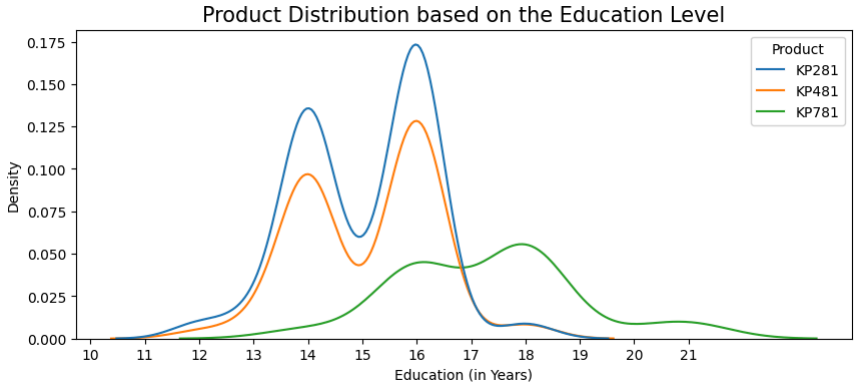
The **18–30 age group** dominates the income distribution, mostly in the **$25,000–$45,000** range, aligning with KP281's user base.

30–40 age group has a broader income spread, from **$40,000 to ~$60,000**, suggesting this group may favor **KP481.**

40–50 and 50+ groups are very sparse and extend into **higher income brackets** i.e. more than $60,000, aligning more with **KP781**.

PRODUCT DISTRIBUTION BASED ON CUSTOMER'S EDUCATION LEVEL (in Years)

```
plt.figure(figsize=(10,4))
sns.kdeplot(x=df['Education'], hue=df['Product'])
plt.title("Product Distribution based on the Education Level", size=15)
plt.xlabel('Education (in Years)')
plt.xticks(range(10,22,1))
plt.show()
```



KEY INSIGHTS:

> KP281 and KP481 are most popular among customers with 14-16 years of education. This suggests these products appeal to those with college-level education.

> KP781 customers tend to have higher education levels, peaking around 17-18 years. This indicates the high-end product attracts a more educated customer base.

> There's a clear shift: lower-end products (KP281, KP481) are popular with mid-level education, while the high-end product (KP781) appeals to higher education levels.

CUSTOMER DISTRIBUTION BASED ON THE EXPECTED AVERAGE MILES PER WEEK FOR DIFFERENT PRODUCTS

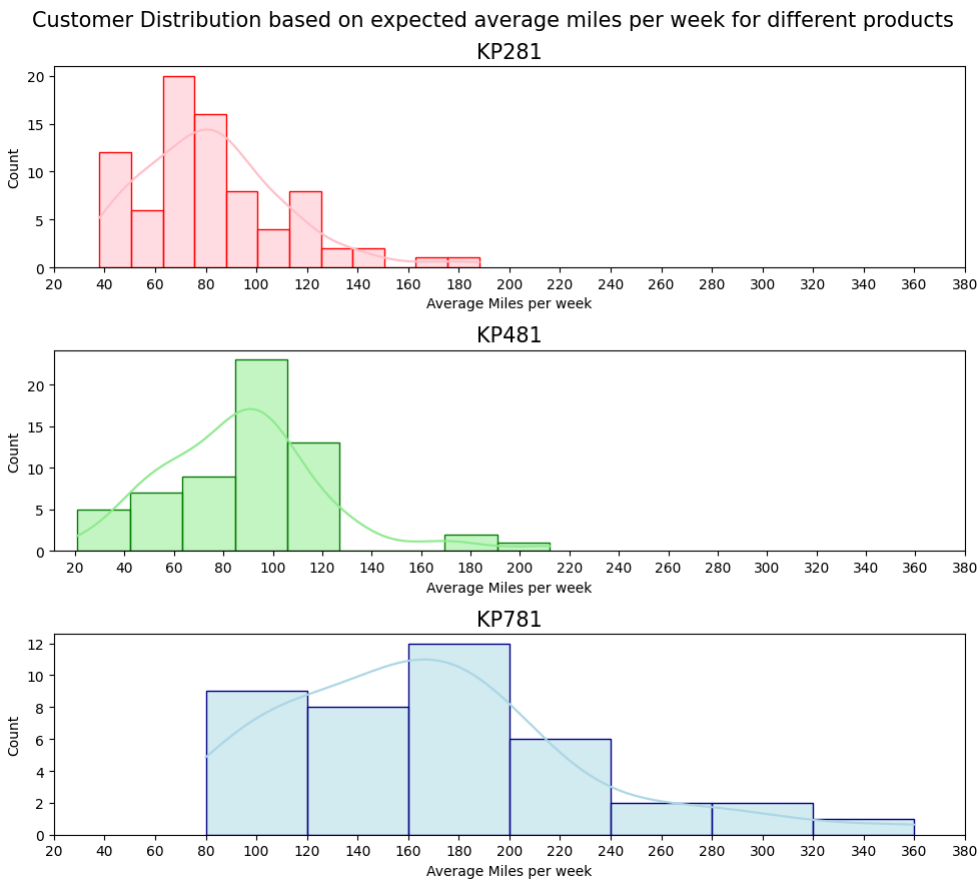
```
fig = plt.figure(figsize=(10,9))

plt.subplot(3,1,1)
sns.histplot(x=df.groupby('Product')['Miles'].get_group('KP281'), color='pink',edgecolor='red', kde=True)
plt.title('KP281', size=15)
plt.xticks(range(20,400,20))
plt.xlabel('Average Miles per week', size=10)

plt.subplot(3,1,2)
sns.histplot(x=df.groupby('Product')['Miles'].get_group('KP481'), color='lightgreen',edgecolor='green', kde=True)
plt.title('KP481', size=15)
plt.xticks(range(20,400,20))
plt.xlabel('Average Miles per week', size=10)

plt.subplot(3,1,3)
sns.histplot(x=df.groupby('Product')['Miles'].get_group('KP781'), color='lightblue',edgecolor='darkblue', kde=True)
plt.title('KP781', size=15)
plt.xticks(range(20,400,20))
plt.xlabel('Average Miles per week', size=10)

fig.suptitle('Customer Distribution based on expected average miles per week for different products', size=15)
plt.tight_layout()
plt.show()
```



KEY INSIGHTS:

KP281 customers typically run between 40-120 miles per week, with a peak around 80 miles. This product is for moderate mileage users.

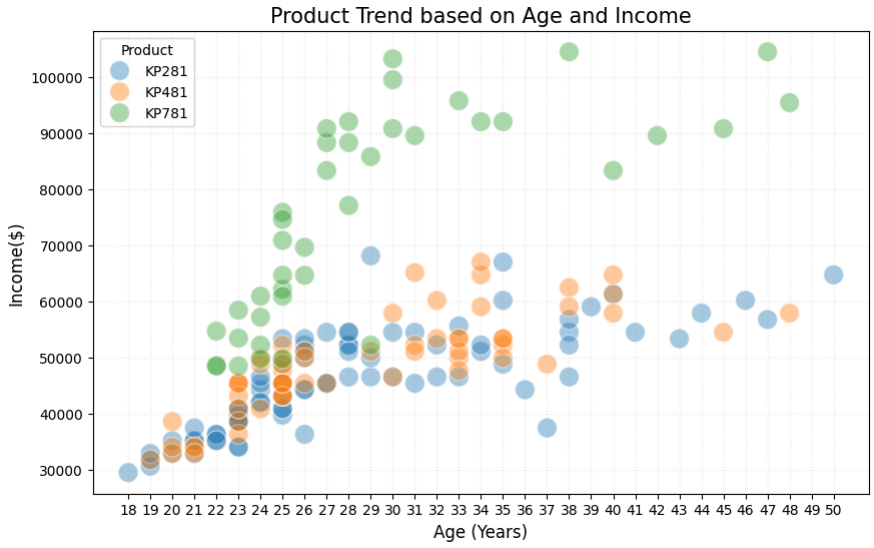
KP481 customers generally run more, mostly between 60-140 miles, peaking around 100 miles. This is for slightly more active users.

KP781 customers are high-mileage users, mostly between 140-220 miles, with a peak around 180 miles. This product is for serious runners.

Each product targets a distinct range of weekly mileage.

PRODUCT TREND BASED ON CUSTOMER'S AGE AND INCOME

```
plt.figure(figsize=(10,6))
sns.scatterplot(data=df, x='Age', y='Income', hue='Product', s=200, alpha=0.4)
plt.xticks(range(18,51))
plt.xlabel('Age (Years)', size=12)
plt.ylabel('Income($)', size=12)
plt.title('Product Trend based on Age and Income', size=15)
plt.grid(True, linestyle=':', alpha=0.2)
plt.show()
```



KEY INSIGHTS

KP781 is primarily purchased by customers with higher incomes across all ages.

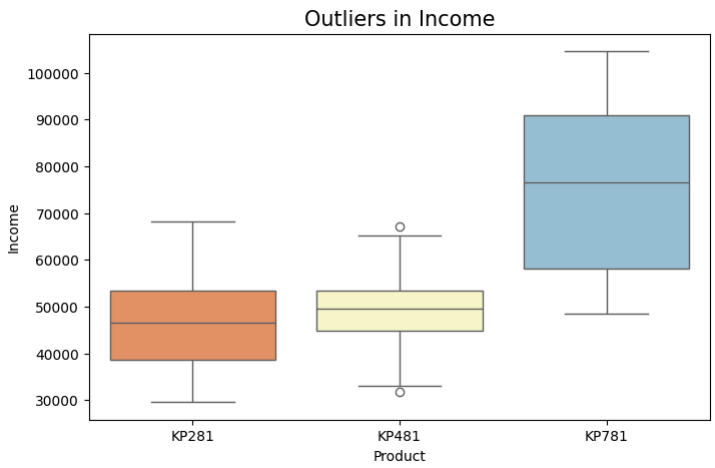
KP281 and KP481 are bought by customers with lower to mid-range incomes.

As age increases, there's a general trend for customers to have higher incomes, especially for KP781.

The high-income, older customers (e.g., 35-50 years old) are key for KP781 sales.

FINDING THE OUTLIERS IN THE DATA

```
plt.figure(figsize=(8,5))
sns.boxplot(data=df, y='Income', x='Product', palette='RdYlBu')
plt.title('Outliers in Income', size=15)
plt.show()
```



KEY INSIGHTS

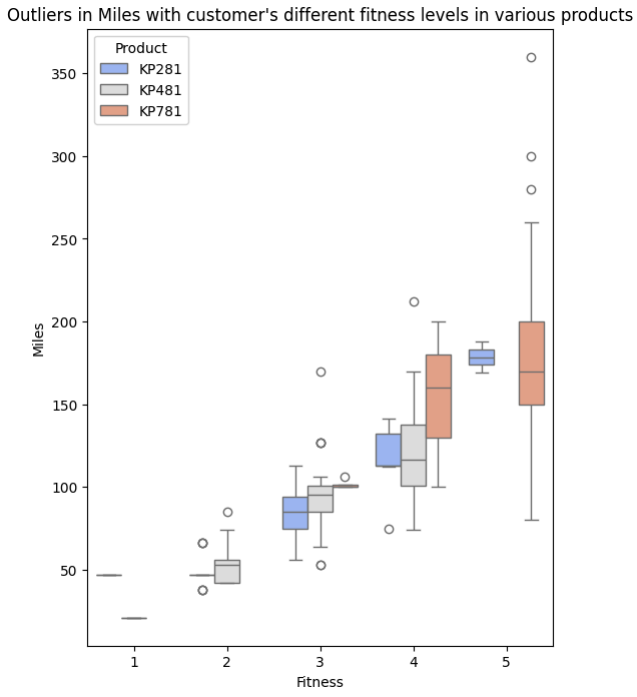
KP781 customers have significantly higher incomes than KP281 and KP481 customers.

The income range for KP781 is much wider and starts at a higher minimum.

KP281 and KP481 customers have similar income levels, with KP481 having a slightly higher median income.

There are a few customers with unusually high or low incomes (outliers) for KP281 and KP481.

```
plt.figure(figsize=(6,8))
sns.boxplot(data=df, y='Miles',x='Fitness', hue='Product', palette='coolwarm')
plt.title("Outliers in Miles with customer's different fitness levels in various products", size=12)
plt.show()
```



KEY INSIGHTS

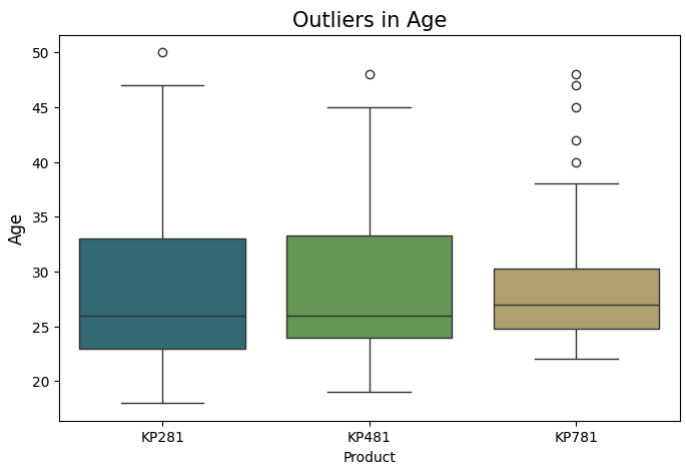
Miles increase with fitness level across all products.

KP781 customers consistently record higher miles at each fitness level compared to KP281 and KP481.

There are customers who run exceptionally high miles (outliers) at higher fitness levels, especially for KP781.

KP281 and KP481 show more variability in miles at lower fitness levels.

```
plt.figure(figsize=(8,5))
sns.boxplot(data=df, y='Age', x='Product', palette='gist_earth')
plt.ylabel('Age',size=12)
plt.title('Outliers in Age', size=15)
plt.show()
```



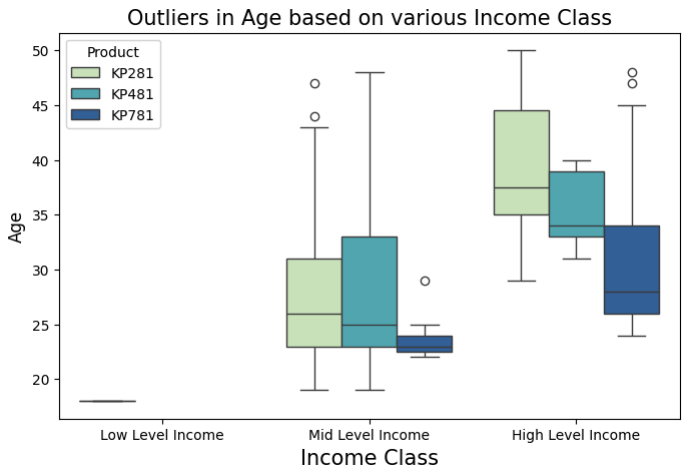
KEY INSIGHTS

KP781 customers tend to be slightly younger on average (median age around 27-28) compared to KP281 and KP481 (median age around 26-27).

The age range for KP781 customers is also slightly tighter, with some older outliers.

All products have customers across a broad age range (18-50), but the core age group for KP781 seems to be slightly younger adults.

```
plt.figure(figsize=(8,5))
sns.boxplot(data=df, y='Age', x='IncomeClass',hue='Product', palette='YlGnBu')
plt.ylabel('Age',size=12)
plt.xlabel('Income Class',size=15)
plt.title('Outliers in Age based on various Income Class', size=15)
plt.show()
```



KEY INSIGHTS

Low-income customers are almost exclusively buying KP281, and they are the youngest group.

Mid-income customers show a mix of products, with KP281 and KP481 being more common among slightly older individuals in this group. KP781 is bought by the youngest in this income class.

High-income customers are generally older, and they predominantly buy KP781. KP281 and KP481 are bought by younger individuals within the high-income group.

This shows that both income and age influence product choice, with higher income generally correlating with older age for KP781, but younger ages for the lower-tier products within the same income bracket.

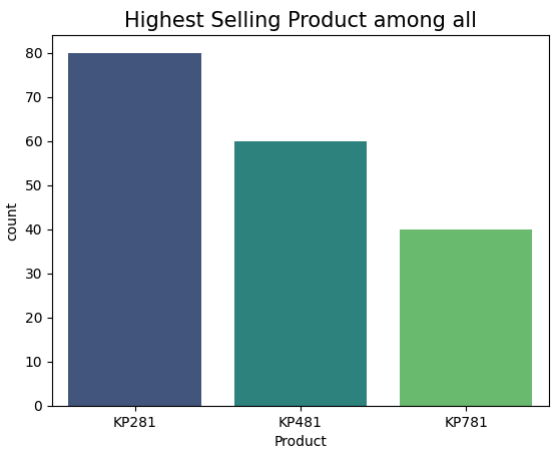
OVERALL HIGHEST SELLING PRODUCT AMONG ALL (in last 3 months):

```
df['Product'].value_counts()
```

Product	count
KP281	80
KP481	60
KP781	40

dtype: int64

```
sns.countplot(data=df, x='Product', palette = 'viridis')
plt.title('Highest Selling Product among all', size=15)
plt.show()
```



KEY INSIGHTS:

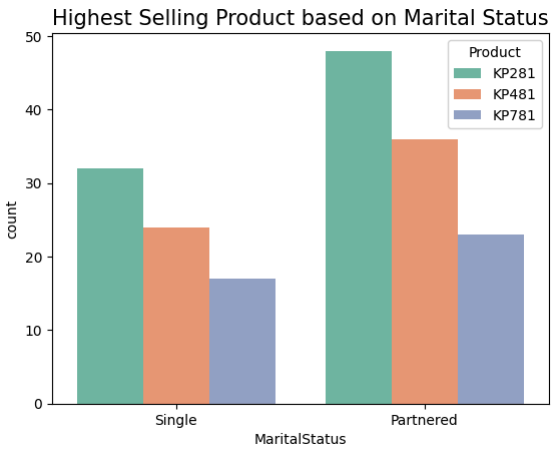
KP281 is the best-selling product overall. It has the highest count of customers.

KP481 is the second most popular product.

KP781 has the lowest customer count, indicating it's the least sold product.

HIGHEST SELLING PRODUCT BASED ON MARITAL STATUS

```
sns.countplot(data=df, x='MaritalStatus',hue='Product', palette='Set2')
plt.title('Highest Selling Product based on Marital Status', size=15)
plt.show()
```



KEY INSIGHTS:

Partnered customers buy more products than single customers.

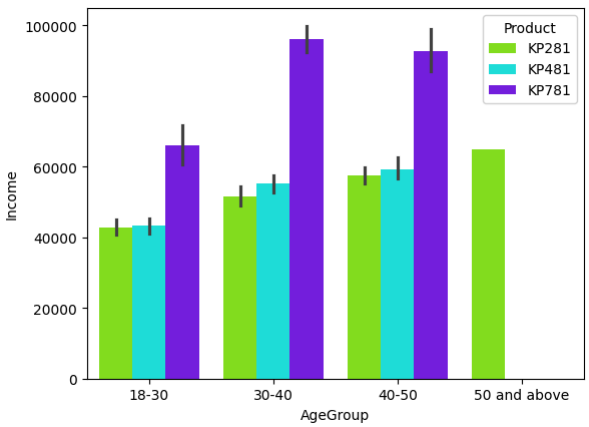
For both single and partnered customers, KP281 is the most purchased product.

Among partnered customers, KP281 sales are significantly higher than for single customers.

KP781 is the least popular product for both marital statuses.

AVERAGE CUSTOMER INCOME BY AGE GROUP AND PRODUCT MODEL

```
#plt.figure(figsize=(15,6))
sns.barplot(data=df, y='Income', x='AgeGroup', hue='Product', palette = 'hsv')
plt.show()
```



KEY INSIGHTS:

Product KP281 (Entry-Level):

Lowest average income across all age groups.

Most popular choice for "50 and above" customers.

Product KP481 (Mid-Range):

Higher income than KP281, lower than KP781.

Income peaks in the 30-50 age groups.

Product KP781 (High-End):

Consistently purchased by customers in the highest income brackets.

Significant income peak in the 30-50 age groups (approaching/exceeding $90,000).

Primary target for premium product due to higher disposable income in these age groups.

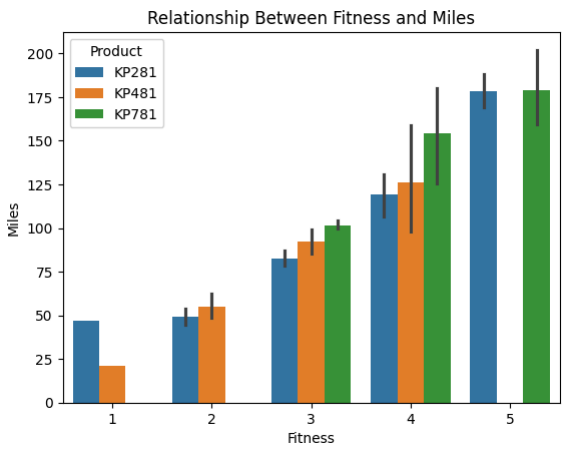
Age and Income Interplay:

30-50 age range: Sweet spot for higher-end products (KP481, KP781) due to higher incomes.

18-30 age group: Prefers KP281/KP481, with KP781 buyers having relatively lower average income than older high-end buyers.

"50 and above" age group: Primarily purchases entry-level KP281.

```
sns.barplot(data=df,x='Fitness',y='Miles', hue='Product')
plt.title("Relationship Between Fitness and Miles")
plt.show()
```



KEY INSIGHTS:

General Trend: Miles covered increase significantly with higher fitness levels across all products.

KP281 (Entry-Level):

Used across all fitness levels (1-5).

Strong performance even at lower fitness levels (e.g., highest miles at Fitness 1).

KP481 (Mid-Range):

Lowest miles at Fitness Level 1.

Steady increase in miles with fitness, but typically below KP781.

KP781 (High-End):

Highest average miles for customers with high fitness levels (3, 4, 5).

Not typically purchased by those at Fitness Level 1.

Clearly targets high-performance, high-mileage users.

Product Positioning: Products align with user fitness and mileage: KP281 (broad appeal), KP481 (mid-tier), KP781 (high-end, serious users).

HOW INCOME, AGE GROUP, FITNESS LEVEL AFFECT THE PERFORMANCE OF THE PRODUCT

```
pd.set_option('display.max_rows', None)
#df.groupby(['Product','IncomeClass','AgeGroup','Fitness'])['Product'].count()
pd.crosstab([df['AgeGroup'],df['IncomeClass'],df['Fitness']],df['Product'],margins=True)
```

AgeGroup	IncomeClass	Product Fitness	KP281	KP481	KP781	All
18-30	**Low Level Income**	**4**	1	0	0	1
	Mid Level Income	**1**	1	0	0	1
		2	9	8	0	17
		3	36	20	1	57
		4	4	5	1	10
		5	1	0	9	10
	High Level Income	**3**	1	0	3	4
		4	0	0	4	4
		5	0	0	9	9
30-40	**Mid Level Income**	**2**	4	3	0	7
		3	10	11	0	21
		4	3	3	0	6
		5	1	0	0	1
	High Level Income	**1**	0	1	0	1
		3	2	4	0	6
		4	0	0	1	1
		5	0	0	7	7
40-50	**Mid Level Income**	**2**	0	1	0	1
		3	3	2	0	5
		4	1	0	0	1
	High Level Income	**2**	1	0	0	1
		3	1	2	0	3
		4	0	0	1	1
		5	0	0	4	4
50 and above	**High Level Income**	**3**	1	0	0	1
All			80	60	40	180

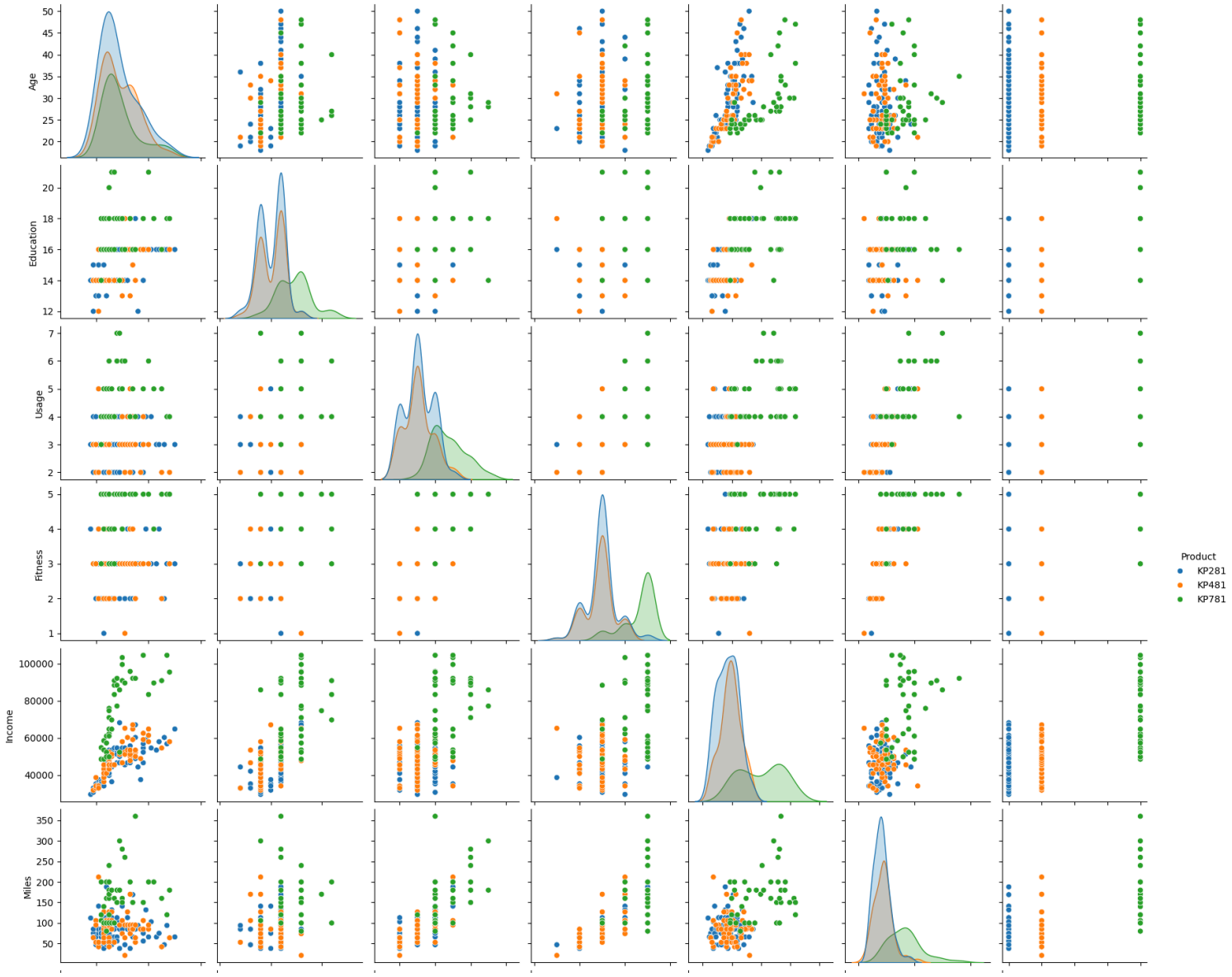
KEY INSIGHTS

- KP281 is predominantly purchased by Mid-Level Income individuals, especially those with Fitness Level 3 in the 18–30 and 30–40 age brackets. This suggests KP281 is affordable and attractive to moderately fit, younger customers looking to stay active.
- KP481 is also favored by Mid-Level Income individuals with Fitness Level 3, but less so than KP281. It appeals to slightly older customers (30–40) and is still perceived as value for money for average fitness users.
- KP781, the premium product, is purchased almost exclusively by High-Income and highly fit (Fitness Levels 4 & 5) individuals.
- It has zero adoption among low-income and low-fitness groups, indicating it is seen as a premium, performance-driven product.
- Fitness Level 3 dominates across most income groups and products, indicating a large average-fitness market segment that drives the bulk of sales for KP281 and KP481.
- Low-Level Income customers make virtually no purchases of KP481 and KP781, and very limited purchases of KP281. This shows a clear affordability barrier, signaling an opportunity if the company plans a lower-cost model for this group.
- There's a positive correlation between higher fitness level and higher-priced product choice: customers with Fitness Level 5 mostly buy KP781, implying these customers are performance-driven and willing to invest more.

CORRELATION BETWEEN VARIOUS NUMERIC ASPECTS IN THE DATASET

```
plt.figure(figsize=(10,10))
sns.pairplot(data=df, hue='Product')
plt.show()
```

<Figure size 1000x1000 with 0 Axes>



KEY INSIGHTS

- **Age vs. Products:**
  - KP281 seems to be popular across a wider age range, with a peak around 25-30 years.
  - KP481 also has a broad age distribution, similar to KP281.
  - KP781 appears to be more concentrated among slightly younger individuals, possibly 20-30 years old, with a tendency for slightly higher age range compared to the other two.
- **Education vs. Products:**
  - All products are purchased across various education levels.
  - KP781 shows a slight tendency towards higher education levels (16+ years of education).
  - KP281 and KP481 are more broadly distributed across education levels, with a concentration around 14-16 years of education.
- **Usage vs. Products:**
  - KP781 users tend to plan for higher weekly usage (4-6 times/week) compared to the other two products.
  - KP281 and KP481 users generally plan for 2-4 times/week usage.
- **Fitness vs. Products:**
  - KP781 users report higher self-rated fitness levels (3-5).
  - KP281 and KP481 users are more concentrated in the 2-3 fitness levels.
- **Income vs. Products:**
  - KP781 purchasers generally have higher income levels (above $50,000 to over $70,000).
  - KP281 and KP481 are bought by individuals across a wider income range, with a concentration in the $30,000 - $60,000 bracket.
- **Miles vs. Products:**
  - KP781 users plan to run more miles per week (100-300 miles) than the other two.
  - KP281 and KP481 users plan for lower mileage (50-150 miles).
- **Price vs. Products:**
  - KP781 is the most expensive product, followed by KP481, and then KP281 which is the most affordable. This is directly reflected in the distinct price clusters.

```
corr_data = df[['Miles','Income', 'Usage','Age','Education', 'Fitness','Price']]
corr_data.corr()
```

	Miles	Income	Usage	Age	Education	Fitness	Price
Miles	1.000000	0.543473	0.759130	0.036618	0.307284	0.785702	0.643923
Income	0.543473	1.000000	0.519537	0.513414	0.625827	0.535005	0.695847
Usage	0.759130	0.519537	1.000000	0.015064	0.395155	0.668606	0.623124
Age	0.036618	0.513414	0.015064	1.000000	0.280496	0.061105	0.029263
Education	0.307284	0.625827	0.395155	0.280496	1.000000	0.410581	0.563463
Fitness	0.785702	0.535005	0.668606	0.061105	0.410581	1.000000	0.696616
Price	0.643923	0.695847	0.623124	0.029263	0.563463	0.696616	1.000000

KEY INSIGHTS

High Correlation between Fitness & Miles (0.79)

Individuals with higher fitness levels tend to travel more miles, indicating active usage of fitness products.

Strong Correlation between Fitness & Price (0.70)

Fitter individuals are more likely to purchase higher-priced products, such as KP781. Suggests that fitness-conscious customers invest more in quality.

Miles & Usage (0.76)

The more someone uses the product, the more miles they log. Indicates the product is being used for its intended physical activity purpose.

Price & Income (0.70)

Higher-income customers tend to purchase more expensive products, validating the price tier strategy across KP281, KP481, and KP781.

Fitness & Usage (0.67)

Fitter individuals use the product more frequently, showing strong engagement with the product.

Education's Moderate Role:

Correlates moderately with Income (0.63), Fitness (0.41), and Price (0.56), implying educated customers may be more health- and quality-conscious.

Age has Very Low Correlation with Fitness (0.06), Price (0.03), and Usage (0.02)

Suggests that age isn't a significant factor in determining fitness levels, product choice, or usage behavior within this dataset.

```
1 Start coding or generate with AI.
```

CUSTOMER PROFILE FOR ALL THE PRODUCT MODELS:

Product KP281 (Entry-Level Treadmill - $1,500)

Overall Popularity:

> This is the best-selling product overall, with the highest number of customers.

Income Level:

> Predominantly purchased by Mid-Level Income individuals ($30,000$–60,000). A very high proportion (91%) of KP281 buyers fall into this group.

> Most dense around the $30,000$–40,000 income range.

> Has minimal appeal among Low-Level Income customers (only 1% of buyers).

Age Group:

> Most popular among young adults (18-30 years old), with two-thirds (66%) of KP281 buyers in this age group.

> Peaks around age 25.

> Also the most popular choice for customers aged "50 and above."

Fitness Level:

> Most preferred product among Fitness Level 3 customers (56% of them choose it).

> Used across all fitness levels (1-5), showing strong alignment with average fitness profiles.

> Customers with Fitness Level 1 show the highest miles for KP281 compared to other products at that fitness level.

Gender:

> Equal purchase count between Male and Female customers (40 each).

Marital Status:

> Most purchased product for both single and partnered customers.

> Sales are significantly higher among partnered customers.

Usage (Times per Week):

> Most frequently used product, with a peak usage around 3 times per week, and a secondary peak near 4.

> Usage is tightly concentrated between 2 and 4 times per week, indicating a strong user base with regular routines.

Miles (Expected per Week):

> Customers typically run between 40-120 miles per week, with a peak around 80 miles. This product is for moderate mileage users.

> Shows more variability in miles at lower fitness levels.

Education:

> Most popular among customers with 14-16 years of education (college-level).

Product KP481 (Mid-Level Treadmill - $1,750)

Overall Popularity:

> The second most popular product overall.

Income Level:

> Also highly favored by Mid-Level Income individuals ($30,000—60,000), with 88% of its buyers in this group.
>
> Concentrated in the $35,000—45,000 income range, slightly higher than KP281.

Age Group:

> Has a more balanced age distribution between 18-30 and 30-50 age groups, suggesting broader appeal.
>
> Slightly older peak around 28-30 years.

Fitness Level:

> 40% of Fitness Level 3 customers purchased KP481, making it the second most popular for this group.
>
> Miles increase steadily with fitness, but generally remain below KP781 miles for higher fitness levels.
>
> At Fitness Level 1, KP481 users record the lowest miles among all products.

Gender:

> Male customers (31) slightly outnumber female customers (29) in purchases.

Marital Status:

> Second most popular product for both single and partnered customers.
>
> More purchased by partnered customers.

Usage (Times per Week):

> Moderate usage, peaking similarly around 3 times per week, but with a lower density than KP281.
>
> Also concentrated in the 2-4 usage range.

Miles (Expected per Week):

> Customers generally run more than KP281 users, mostly between 60-140 miles, peaking around 100 miles. This is for slightly more active users.
>
> Shows more variability in miles at lower fitness levels.

Education:

> Most popular among customers with 14-16 years of education.

Product KP781 (High-End Treadmill - $2,500)

Overall Popularity:

> The least sold product overall.

Income Level:

> Purchased almost exclusively by High-Level Income individuals (above $60,000). Only 28% of its buyers are from the mid-level income group.
>
> Has a distinct peak at higher incomes ($85,000—100,000), confirming it's favored by high-income individuals.

Age Group:

> Customers tend to be slightly younger on average (median age around 27-28) compared to the other products, but with a broader usage across mid-20s to late 40s.
>
> The 30-50 age range represents a significant income peak for this product.
>
> Not typically purchased by the "50 and above" age group.

Fitness Level:

> Purchased almost exclusively by highly fit (Fitness Levels 4 & 5) individuals.
>
> A very high proportion (72%) of KP781 buyers have a Fitness Level 5.
>
> Customers with higher fitness levels (3, 4, 5) consistently record the highest average miles among all products.
>
> Has zero adoption among low-fitness groups (Fitness Level 1 and 2 are not shown, or have very minimal counts).

Gender:

> Significantly more Male customers (33) purchase KP781 compared to Female customers (7).

Marital Status:

> Least popular product for both single and partnered customers, but more purchased by partnered customers.

Usage (Times per Week):

> Shows high variability and less consistent use, with a broader peak around 4-5 times per week, extending beyond 6 times.
>
> Appeals to users with more intensive fitness routines, even if less frequent.

Miles (Expected per Week):

> Customers are high-mileage users, mostly between 140-220 miles, with a peak around 180 miles. This product is for serious runners.
>
> Customers consistently record higher miles at each fitness level compared to KP281 and KP481.

Education:

> Customers tend to have higher education levels, peaking around 17-18 years, indicating it attracts a more educated customer base.

```
Start coding or generate with AI.
```

FINAL BUSINESS INSIGHTS BASED ON THE AEROFIT DATA

Overall Customer Trends

- **KP281 (Entry-Level)** is the **most popular product**, accounting for **44% of total sales**.
- **Customers are mostly young (18–30)**, with a strong skew toward **mid-level income**.
- Majority of customers (54%) have **moderate fitness levels (Level 3)**.
- Customers prefer to **use treadmills 2–4 times per week**.

Product-Wise Insights

KP281 (Entry-Level – $1,500)

- **Best-selling product**, especially popular among **18–30-year-olds** (66%).
- Strong alignment with **mid-income ($30k–$60k)** (91%).
- Preferred by users with **fitness level 3** (68%).
- Most users run **60–120 miles/week**, peaking at ~80 miles.
- Appeals to moderately fit, price-conscious consumers.

KP481 (Mid-Range – $1,750)

- **Second most popular**, purchased by **broader age group (18–50)**.
- Also skewed toward **mid-level income** (88%).
- Balanced appeal across **fitness levels 2 to 4**, especially level 3 (40%).
- Weekly usage and mileage higher than KP281, peaking around 100 miles.
- Ideal for moderately experienced, consistent users.

KP781 (High-End – $2,500)

- **Least sold (22%)**, but **strongest among high-income ($60k+)** (72% of buyers).
- 72% of buyers have **fitness level 5**, showing high alignment with elite fitness users.
- Most customers run **140–220 miles/week**, peaking at ~180 miles.
- Targets a **niche of high-performing, health-conscious individuals**.
- Zero adoption among low-income or low-fitness segments.

Demographic Insights

- **Gender**:
  - **57.78% male** and **42.22% female** customers.
  - Males are **more likely to purchase premium product KP781** (82.5% of its buyers).
- **Marital Status**:
  - **Partnered individuals buy more**, especially KP281.
  - Both partnered and single individuals show minimal interest in KP781.
- **Age & Income Combination**:
  - **Young (18–30) & mid-income** → KP281.
  - **Age 30–50 with higher income** → KP481 and KP781.
  - **50+ segment prefers KP281**, likely due to simplicity and affordability.
- **Education**:
  - KP281/KP481: **14–16 years of education** (college level).
  - KP781: **17–18+ years**, showing a tendency toward higher education.

Behavioral Insights

- **Product Usage**:
  - KP281: Peaks at **3x per week**.
  - KP481: Slightly broader but lower usage.
  - KP781: **4–6x per week**, indicating serious users.
- **Fitness & Miles Correlation**:
  - Fitness strongly correlates with **mileage** and **product price**.
  - Higher fitness → more usage, more miles → higher likelihood of purchasing KP781.

```
Start coding or generate with AI.
```

FINAL BUSINESS RECOMMENDATIONS TO AEROFIT

Product Strategy

- **Continue prioritizing KP281** as the core product, given it is the **highest-selling model** (44% of sales) and **most popular among mid-income, young (18–30) customers**.
- **Position KP481** as the next-step model for those progressing in age and fitness — data shows higher adoption in **30–40 age group** and **Fitness Level 3–4**.
- **Maintain KP781 as a premium-tier product**, as it is **purchased almost exclusively by high-income individuals (69%) with Fitness Level 5 (72%)**.

Customer Segmentation & Targeting

- **18–30 years old + Mid-income (30k–60k)**: Focus on KP281 promotions — this segment forms the **largest customer group**.
- **30–50 years old + Higher income**: Recommend KP481 and KP781 — they **dominate this age-income group** combination.
- **Fitness Level 3 customers**: Promote KP281 and KP481 — **56% of Fitness Level 3 users buy KP281**, and **40% buy KP481**.

Sales & Promotion Strategy

- For **partnered individuals**, prioritize marketing KP281 — they are the **largest buyer group across all products**, especially KP281.
- Use **fitness level data to match product recommendations**:
  - **Level 3** → KP281 or KP481
  - **Level 5** → KP781
- Emphasize **product usage consistency** in promotions — KP281 shows **peak usage at 3x/week**, KP781 peaks **at 4–5x/week**.

Income-Based Positioning

- KP281 should be positioned for **middle-income buyers** ($30k–$60k) — **91% of its buyers are from this income class**.
- KP781 should be reserved for **high-income individuals** — **only 28% of its buyers are mid-income**.
- Avoid targeting **low-income segments**, as **only 1 customer** across all 180 entries is in this group.

Fitness & Mileage Alignment

- Highlight mileage suitability in marketing:
  - **KP281**: Moderate users (~80 miles/week)
  - **KP481**: Slightly more active users (~100 miles/week)
  - **KP781**: High mileage users (~180 miles/week)
- Use **Fitness Level correlation**: Higher fitness customers log more miles and tend to purchase KP781.

Product Education Targeting

- **KP281 & KP481** appeal most to those with **14–16 years of education** (college level).
- **KP781 buyers** show slightly **higher education levels (17–18 years)** — promote it as a high-end, performance-driven treadmill.

Geographic & Store-Level Implications

- Focus KP281 inventory in areas with **younger, mid-income population**.
- Prioritize KP781 availability in areas with **high-income, fitness-conscious consumers** (especially aged 30–50).
- Ensure KP481 is available in **all locations** to support transitioning users from KP281.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

1 Start coding or generate with AI.

```
Start coding or generate with AI.
```

```
Start coding or generate with AI.
```

```
Start coding or generate with AI.
```

```
Start coding or generate with AI.
```