Introduction

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

The market research team at AeroFit wants 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 new customers. The team decides to investigate whether there are differences across the product with respect to customer characteristics

In [ ]:
```
#Importing libraries,loading the dataset and performing EDA
```

In [ ]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from prettytable import prettytable
```

In [ ]:
```
df=pd.read_csv('aerofit_treadmill.txt')
```

In [ ]:
```
df.head()
```

Out[ ]:

	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

In [ ]:
```
df.shape
```

Out[ ]:
```
(180, 9)
```

There are 180 rows and 9 columns i dataset

Data Preprocessing

In [ ]:
```
df.isna().sum()
```

Out[ ]:
```
Product          0
Age              0
Gender           0
Education        0
MaritalStatus    0
Usage            0
Fitness          0
Income           0
Miles            0
Age_group        0
dtype: int64
```

In [ ]:
```
df['Age_group'] = pd.cut(df['Age'], 3, labels=['young-adult','middleaged-adult','se
```

In [ ]:
```
df.head()
```

Out[ ]:

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

Note

- We do not have any NAN values in the data set
- We have created an extra column for better Analysis

In [ ]:
```

```

Statistical Description of Categorical variable

In [ ]:
```
df.describe(include='object')
```

Out[ ]:

	Product	Gender	MaritalStatus
count	180	180	180
unique	3	2	2
top	KP281	Male	Partnered
freq	80	104	107

Statistical Description of Numerical variables

In [ ]:
```
df.describe()
```

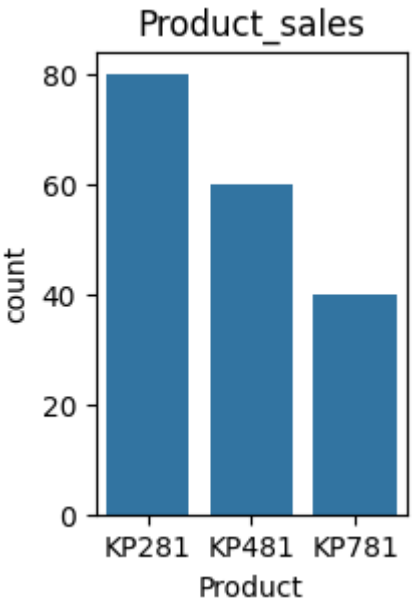
Out[ ]:

	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

#UNIVARIENT ANALYSIS

In [ ]:
```
#Product vise sales
plt.figure(figsize=(2,3))
sns.countplot(x=df['Product'])
plt.title('Product_sales')
```

Out[ ]:
```
Text(0.5, 1.0, 'Product_sales')
```

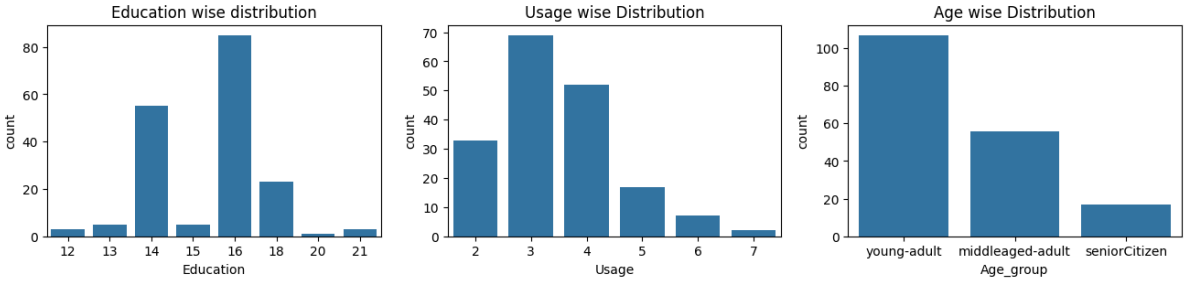


Observations

1. KP281 has more sales compared to other varients this might be due to low price
2. Lets understand this further detail in upcoming

Distribution of Product over Discrete variables

In [ ]:
```
figure,axes=plt.subplots(nrows=1,ncols=3,figsize=(16,3))
sns.countplot(x=df['Education'],ax=axes[0])
sns.countplot(x=df['Usage'],ax=axes[1])
sns.countplot(x=df['Age_group'],ax=axes[2])
axes[0].set_title('Education wise distribution')
axes[1].set_title('Usage wise Distribution')
axes[2].set_title('Age wise Distribution')
plt.show()
```

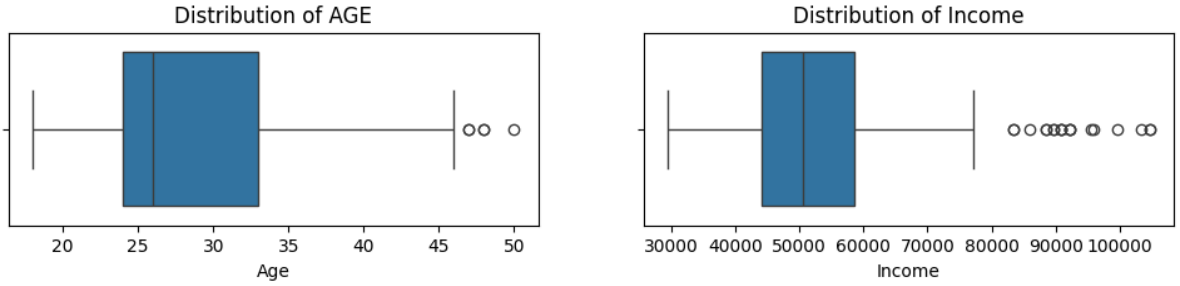


Observation

- Young adults are more likely to purchase the product followed by midaged-adults and senior citizens
- Mostly customers are fond of using these threadmills 3 times a week

Distribution for Continous Variables

In [ ]:
```
fig,axes=plt.subplots(nrows=1,ncols=2,figsize=(12,2))

sns.boxplot(x=df['Age'],ax=axes[0])
sns.boxplot(x=df['Income'],ax=axes[1])
axes[0].set_title('Distribution of AGE')
axes[1].set_title('Distribution of Income')
plt.show()
```



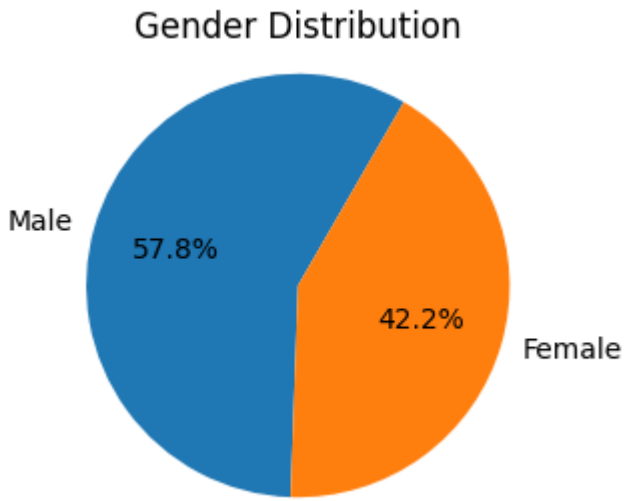
Observations

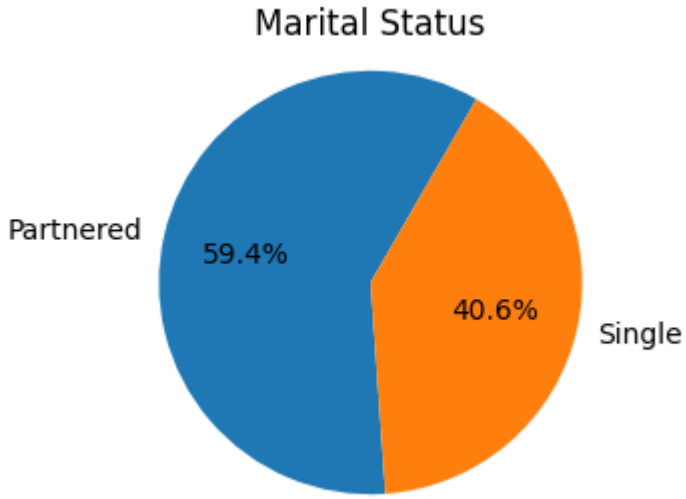
- Majority of the customers are in the age group(23-33)
- Majority of customers have an income level between (45,000-60,000)

In [ ]:
```
gender_counts=df['Gender'].value_counts()
Single_Married=df['MaritalStatus'].value_counts()
```

In [ ]:
```
#Gender Wise distribution

gender_counts=df['Gender'].value_counts()
Single_Married=df['MaritalStatus'].value_counts()


plt.figure(figsize=(3, 3))
plt.pie(gender_counts,labels=gender_counts.index,autopct='%1.1f%%', startangle=60)
plt.axis('equal')
plt.title('Gender Distribution')
plt.show()


#Martialstatus wise Customers
plt.figure(figsize=(3, 3))
plt.pie(Single_Married,labels=Single_Married.index,autopct='%1.1f%%', startangle=60
```

```
plt.axis('equal')
plt.title('Marital Status')
plt.show()
```



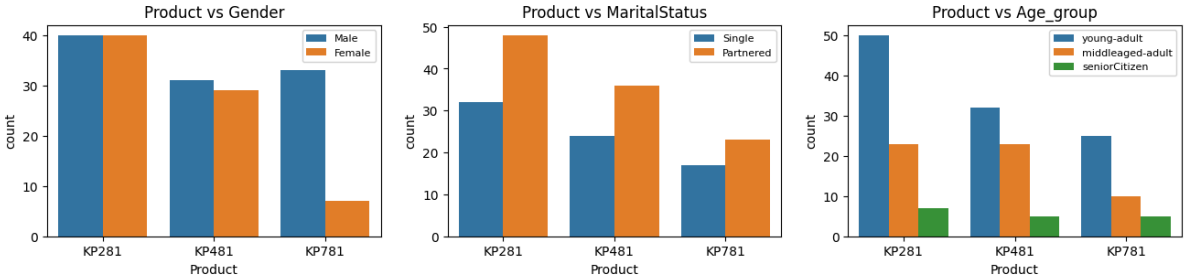


Observations\

- Majority of the customers are male
- The probability that a randomly choosen cuatomer will be Maried is almost 60%

'Bi-varient Analysis'

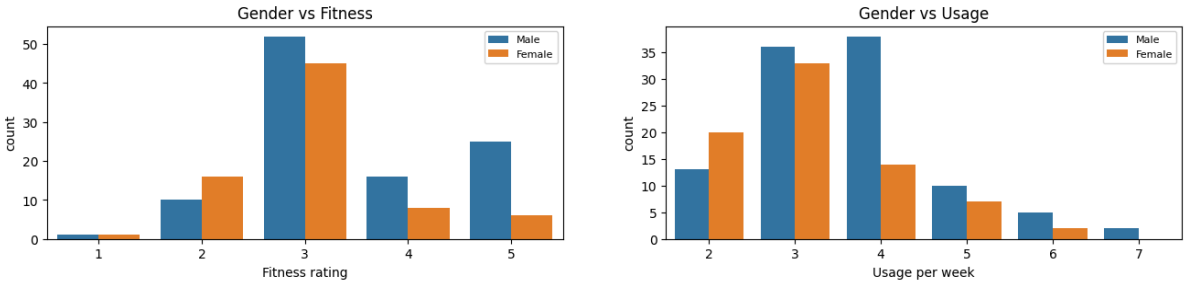
In [ ]:
```
fig,axes=plt.subplots(nrows=1,ncols=3,figsize=(16,3))
sns.countplot(x=df['Product'],hue=df['Gender'],ax=axes[0])
sns.countplot(x=df['Product'],hue=df['MaritalStatus'],ax=axes[1])
sns.countplot(x=df['Product'],hue=df['Age_group'],ax=axes[2])
axes[0].set_title('Product vs Gender')
axes[1].set_title('Product vs MaritalStatus')
axes[2].set_title('Product vs Age_group')
axes[0].legend(prop={'size': 8})
axes[1].legend(prop={'size': 8})
axes[2].legend(prop={'size': 8})
plt.show()
```



Observations

- It appears sales for KP781 product is highest among Male gender. This might be because KP781 has some extra features which is attarcting male gender customers or the market strategies are more leaned towards male.
- It appears sales for all three products is highes amoong the Married people.

In [ ]:
```
fig,axes=plt.subplots(nrows=1,ncols=2,figsize=(16,3))
sns.countplot(x=df['Fitness'],hue=df['Gender'],ax=axes[0])
sns.countplot(x=df['Usage'],hue=df['Gender'],ax=axes[1])
axes[0].set_title('Gender vs Fitness')
axes[1].set_title('Gender vs Usage')
axes[0].legend(prop={'size': 8})
axes[0].set_xlabel('Fitness rating')
axes[1].set_xlabel('Usage per week')
axes[1].legend(prop={'size': 8})
plt.show()
```



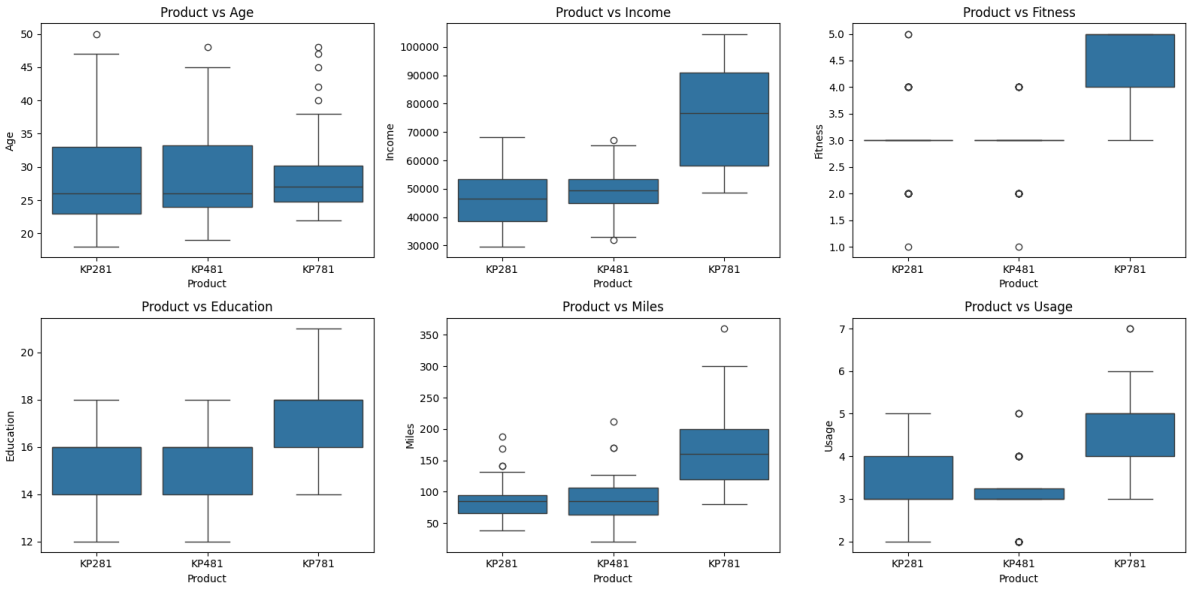
Observations

- Almost **75%** of Male customers have **(2-4) Fitness rating**.
- Almost **84%** of Femal customers use treadmill for **2 to 4 times a week**.

In [ ]:
```
#Distribution of Product amaong diffrent Variables
```

In [ ]:
```
fig, axes = plt.subplots(nrows=2, ncols=3, figsize=(16,8))
axes = axes.flatten()

sns.boxplot(x=df['Product'],y=df['Age'],ax=axes[0])
sns.boxplot(x=df['Product'],y=df['Income'],ax=axes[1])
sns.boxplot(x=df['Product'],y=df['Fitness'],ax=axes[2])
sns.boxplot(x=df['Product'],y=df['Education'],ax=axes[3])
sns.boxplot(x=df['Product'],y=df['Miles'],ax=axes[4])
sns.boxplot(x=df['Product'],y=df['Usage'],ax=axes[5])

axes[0].set_title('Product vs Age')
axes[1].set_title('Product vs Income')
axes[2].set_title('Product vs Fitness')
axes[3].set_title('Product vs Education')
axes[4].set_title('Product vs Miles')
axes[5].set_title('Product vs Usage')
```

```
plt.tight_layout()
```



Observation

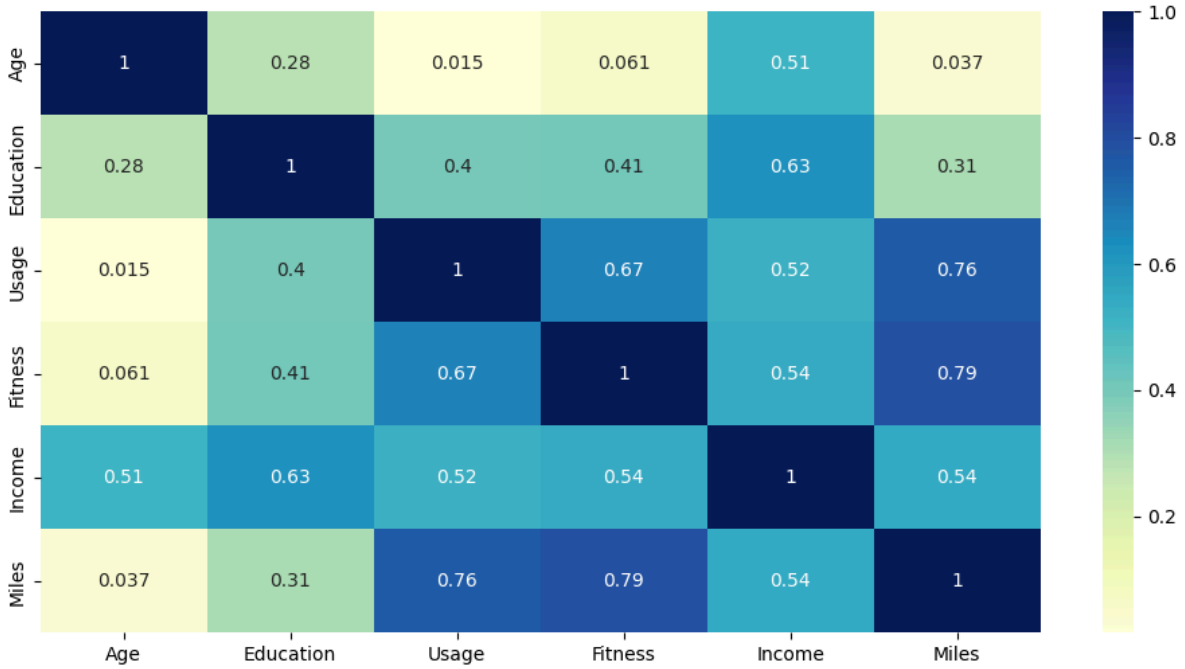
The analysis clearly shows that preference for the KP781 model among the customers who are highly educated, having high income and are engaged in running activities greater then **150 miles per week**.

Multi Varient Analysis

In [ ]:
```
sns.pairplot(data=df)
plt.show()
```

In [ ]:
```
correlation_matrix=df.corr()
plt.figure(figsize=(12,6))
sns.heatmap(correlation_matrix,annot=True,cmap="YlGnBu")
plt.show()
```

```
<ipython-input-23-78d437092379>:1: FutureWarning: The default value of numeric_onl
y in DataFrame.corr is deprecated. In a future version, it will default to False.
Select only valid columns or specify the value of numeric_only to silence this war
ning.
  correlation_matrix=df.corr()
```



Observations

- Age and income are psoitive corelated
- Education and income are also having postive co relation, which means higher the education higher the income levels
- High corelation between fitness and miles as well, whcih tells us people who have do more activity on tredmill are high rated in fitness

Computing prababilty-Marginaland conditional Probility

In [ ]:
```
Marginal_Probability=pd.crosstab(index=df['Product'],columns='Marginal probability'
np.round((Marginal_Probability),2)
```

Out[ ]:

col_0	Marginal probability
Product	
KP281	80
KP481	60
KP781	40

In [ ]:
```
Probability_of_each_product=pd.crosstab(index=df['Product'],columns=' probability')
print(np.round((Probability_of_each_product),2))
```

```
col_0     probability
Product
KP281            0.44
KP481            0.33
KP781            0.22
```

Observations

- The probability that the customer will purchase KP281 treadmill is **44%**
- The probability that the customer will purchase KP481 treadmill is **33%**.
- The probability that the customer will ourchase KP781 treadmill is **22%.**

In [ ]:
```
pd.crosstab(index=df['Product'],columns=df['Gender'],margins=True,normalize=True)
```

Out[ ]:

Gender	Female	Male	All
Product			
KP281	0.222222	0.222222	0.444444
KP481	0.161111	0.172222	0.333333
KP781	0.038889	0.183333	0.222222
All	0.422222	0.577778	1.000000

The **Probability** that the given customer is **male** would be **57%**

The **conditional probabilty** that given customer is male, would purchase treadmill

- For KP281 is 22%
- For KP481 is 17%.
- For KP781 is 18%.

The **Probability** that the given customer is **Female** would be **42%**

The **conditional probabilty** that given customer is female, would purchase treadmill

- For KP281 is 22%
- For KP481 is 16%.
- For KP781 is 3%.

In [ ]:
```
pd.crosstab(index=df['Product'],columns=df['MaritalStatus'],margins=True,normalize=
```

Out[ ]:

MaritalStatus	Partnered	Single	All
Product			
KP281	0.266667	0.177778	0.444444
KP481	0.200000	0.133333	0.333333
KP781	0.127778	0.094444	0.222222
All	0.594444	0.405556	1.000000

The **Probability**that the given customer is **Married** would be **59%**

The **conditional probabilty** that given customer is married, would purchase treadmill moder

- For KP281 is 26%
- For KP481 is 20%.
- For KP781 is 12%.

The **Probability** that the given customer is **Single** would be **40%**

The **conditional probabilty** that given customer is Single, would purchase treadmill

- For KP281 is 17%
- For KP481 is 13%.
- For KP781 is 9%.

In [ ]:
```
pd.crosstab(index=df['Product'],columns=df['Age_group'],margins=True,normalize=True
```

Out[ ]:

Product	young-adults	middleaged-adult	seniorCitizen	All
KP281	0.277778	0.127778	0.038889	0.444444
KP481	0.177778	0.127778	0.027778	0.333333
KP781	0.138889	0.055556	0.027778	0.222222
All	0.594444	0.311111	0.094444	1.000000

The **Probability** that a **young adult** would purchase threadmill is **59%.**

The **conditional probabilty** that a young_adult, would purchase treadmill

- For KP281 is 27%
- For KP481 is 17%.
- For KP781 is 13%.

The **Probability** that a **middle-aged adult** would purchase threadmill is **31%.**

The **conditional probabilty** that a middle-aged adult, would purchase treadmill

- For KP281 is 12%
- For KP481 is 12%.
- For KP781 is 5%.

The **Probability** that a **Senior citizen** would purchase threadmill is **9%.**

The **conditional probabilty** that a Senior citizen, would purchase treadmill

- For KP281 is 3%
- For KP481 is 2%.
- For KP781 is 2%.

In [ ]:
```
pd.crosstab(index=df['Product'],columns=df['Fitness'],margins=True,normalize=True)
```

Out[ ]:

Product	1	2	3	4	5	All
KP281	0.005556	0.077778	0.300000	0.050000	0.011111	0.444444
KP481	0.005556	0.066667	0.216667	0.044444	0.000000	0.333333
KP781	0.000000	0.000000	0.022222	0.038889	0.161111	0.222222
All	0.011111	0.144444	0.538889	0.133333	0.172222	1.000000

Observations

- The probabilty that given customer uses KP281 treadmill 3 times a week is 30%
- The probabilty that given customer uses KP481 treadmill 3 times a week is 21%
- The probabilty that given customer uses KP781 treadmill 5 times a week is 16%

----------**Insights**----------

- Most of the customers are young adults and midaged adults
- The probaility of customer being a sinior citizen is less.
- There is not much difference in Distribution of data between Single and Partnered
- Most of customers use treadmill 2-4 times a week and the have rated fitness 3 to 4
- The analysis clearly shows that preference for the KP781 model among the customers who are highly educated, having high income and are engaged in running activities greater then 150 miles per week.

----------**Customer Profiling**----------

KP281 Treadmill

- Targetting young- adult (18-24) & Mid-aged adults
- Whether Married or un-married
- Equal representation of Male & Female
- Income (40,000-52,000)
- Usage 3-4 times a week
- Gerneral fittness rating 3

KP481 Treadmill

- Suitable for both Young-adults and Middle aged adults \
- Married or single
- Income(45,000-50,000)
- Average usage (3 times a week)

KP781 Treadmill

- Trgetting Premium customers,mainly young adults.
- Having high income level(>60,000)
- often Highly educated and Fitness Enthusiats

----------**Recomendations**----------

Imrovement of KP281 treadmill.

- Conduct a thorough analysis to identify the specific features or technologies that customers expect more in KP481 from KP281
- Based on customer feedback and market research, enhance the features or technology of KP481 to justify the higher price point

- Consider offering additional value-added services or accessories with KP481 to make it more appealing to customers in the mid-level market segment.

Building New Marketing Strategies for KP781

- Highlight the premium features, high-quality construction, and advanced technology of KP781 in marketing materials to attract affluent customers
- Since the usage levels is high for this model. Branding this through reputed athelits will increase the sales.

In [ ]: