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

```
! wget https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/001/125/original/aerofit_treadmill.csv?1639992749 -O aerofit_
```

```
--2023-03-12 12:14:28--  https://d2beiqkhq929f0.cloudfront.net/public_assets/assets/000/001/125/original/aerofit_treadmill.csv
Resolving d2beiqkhq929f0.cloudfront.net (d2beiqkhq929f0.cloudfront.net)... 108.157.172.176, 108.157.172.183, 108.157.172.10, .
Connecting to d2beiqkhq929f0.cloudfront.net (d2beiqkhq929f0.cloudfront.net)|108.157.172.176|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 7279 (7.1K) [text/plain]
Saving to: 'aerofit_treadmill.csv'

aerofit_treadmill.c 100%[===================>]   7.11K  --.-KB/s    in 0s

2023-03-12 12:14:29 (645 MB/s) - 'aerofit_treadmill.csv' saved [7279/7279]
```

```
data=pd.read_csv('aerofit_treadmill.csv')
data
```

	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
...	...	...	...	...	...	...	...	...	...
175	KP781	40	Male	21	Single	6	5	83416	200
176	KP781	42	Male	18	Single	5	4	89641	200
177	KP781	45	Male	16	Single	5	5	90886	160
178	KP781	47	Male	18	Partnered	4	5	104581	120
179	KP781	48	Male	18	Partnered	4	5	95508	180

180 rows × 9 columns

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

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitne
count	180	180.000000	180	180.000000	180	180.000000	180.0000
unique	3	NaN	2	NaN	2	NaN	Na
top	KP281	NaN	Male	NaN	Partnered	NaN	Na
freq	80	NaN	104	NaN	107	NaN	Na
mean	NaN	28.788889	NaN	15.572222	NaN	3.455556	3.3111
std	NaN	6.943498	NaN	1.617055	NaN	1.084797	0.9588
min	NaN	18.000000	NaN	12.000000	NaN	2.000000	1.0000
25%	NaN	24.000000	NaN	14.000000	NaN	3.000000	3.0000
50%	NaN	26.000000	NaN	16.000000	NaN	3.000000	3.0000
75%	NaN	33.000000	NaN	16.000000	NaN	4.000000	4.0000
max	NaN	50.000000	NaN	21.000000	NaN	7.000000	5.0000

```
data.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
```

```
data.shape
```

```
(180, 9)
```

```
data.value_counts()
```

```
Product  Age  Gender  Education  MaritalStatus  Usage  Fitness  Income  Miles
KP281    18   Male    14         Single         3      4        29562   112      1
KP481    30   Female  13         Single         4      3        46617   106      1
         31   Female  16         Partnered      2      3        51165   64       1
                      18         Single         2      1        65220   21       1
              Male    16         Partnered      3      3        52302   95       1
                                                                                ..
KP281    34   Female  16         Single         2      2        52302   66       1
              Male    16         Single         4      5        51165   169      1
         35   Female  16         Partnered      3      3        60261   94       1
                      18         Single         3      3        67083   85       1
KP781    48   Male    18         Partnered      4      5        95508   180      1
Length: 180, dtype: int64
```

```
data.isnull().sum()
```

```
Product          0
Age              0
Gender           0
Education        0
MaritalStatus    0
Usage            0
Fitness          0
Income           0
Miles            0
dtype: int64
```

```
data.nunique()
```

```
Product           3
Age              32
Gender            2
Education         8
MaritalStatus     2
Usage             6
Fitness           5
Income           62
Miles            37
dtype: int64
```

```
data.Product.value_counts()
```

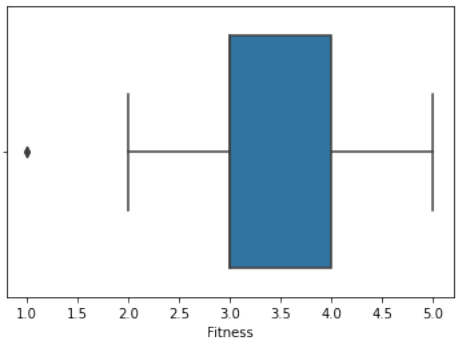
```
KP281    80
KP481    60
KP781    40
Name: Product, dtype: int64
```

```
data.Age.unique()
```

```
array([18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34,
       35, 36, 37, 38, 39, 40, 41, 43, 44, 46, 47, 50, 45, 48, 42])
```

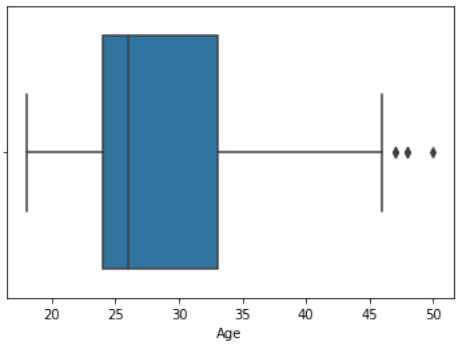
```
#Outliers
sns.boxplot(data.Fitness)
plt.show()
```

```
/usr/local/lib/python3.9/dist-packages/seaborn/_decorators.py:36: FutureWarning: P
  warnings.warn(
```



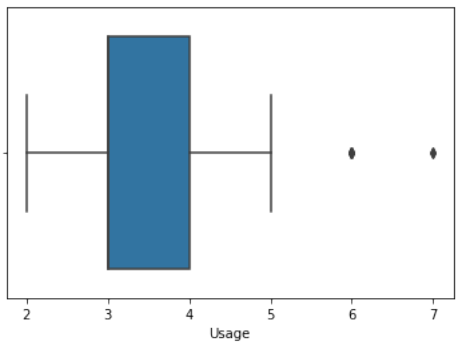

```
sns.boxplot(data.Age)
```

```
/usr/local/lib/python3.9/dist-packages/seaborn/_decorators.py:36: FutureWarning: P
  warnings.warn(
<AxesSubplot:xlabel='Age'>
```



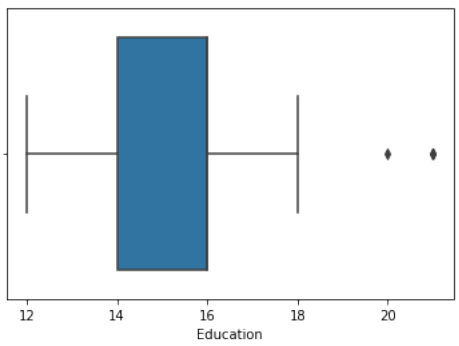

```
sns.boxplot(data.Usage)
```

```
/usr/local/lib/python3.9/dist-packages/seaborn/_decorators.py:36: FutureWarning: P
  warnings.warn(
<AxesSubplot:xlabel='Usage'>
```



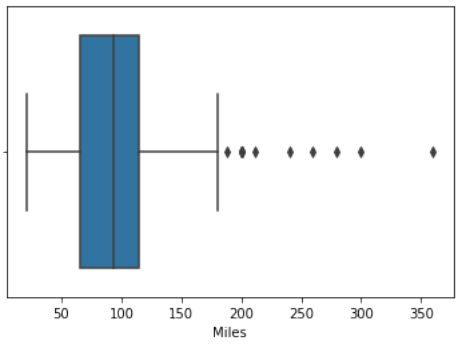

```
sns.boxplot(data.Education)
plt.show()
```

```
/usr/local/lib/python3.9/dist-packages/seaborn/_decorators.py:36: FutureWarning: P
  warnings.warn(
```



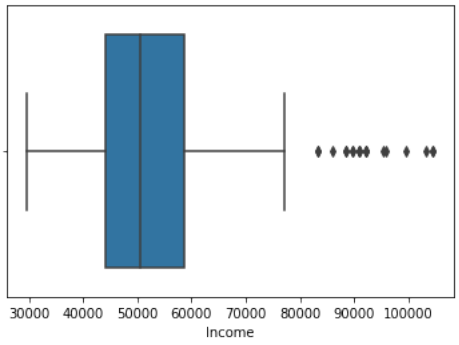
```
sns.boxplot(data.Miles)
plt.show()
```

```
/usr/local/lib/python3.9/dist-packages/seaborn/_decorators.py:36: FutureWarning: P
  warnings.warn(
```

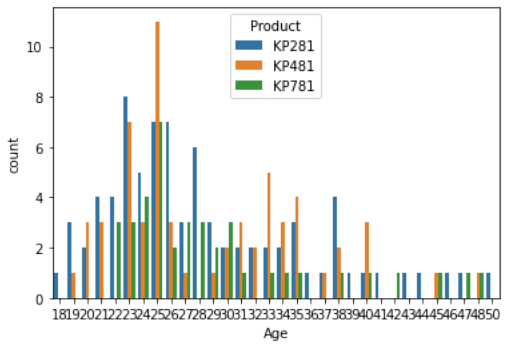



```
sns.boxplot(data.Income)
plt.show()
```

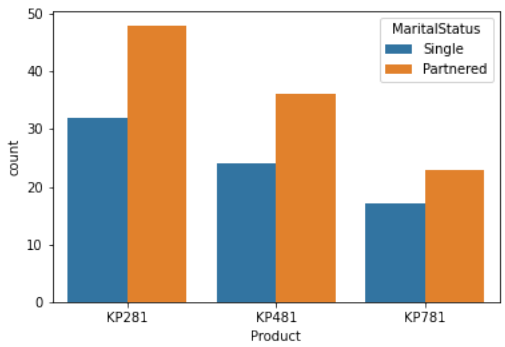
```
/usr/local/lib/python3.9/dist-packages/seaborn/_decorators.py:36: FutureWarning: P
  warnings.warn(
```



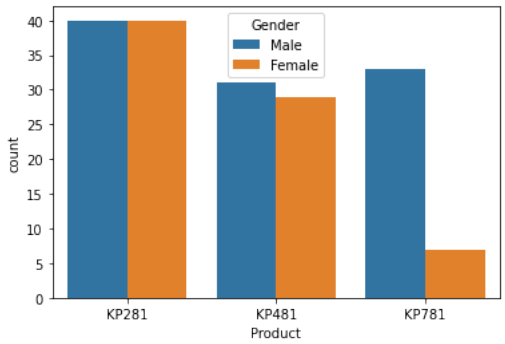

```
#Bivariate analysis
sns.countplot(data=data,x=data['Age'],hue=data['Product'])
plt.show()
```



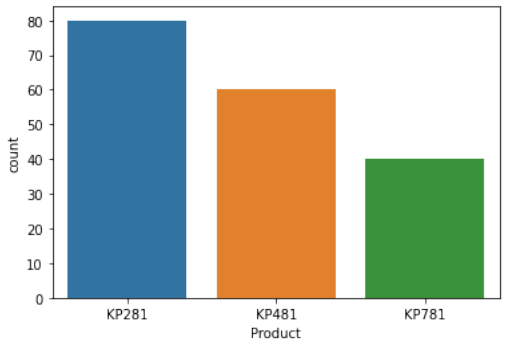

```
sns.countplot(data=data,x=data['Product'],hue=data['MaritalStatus'])
plt.show()
```



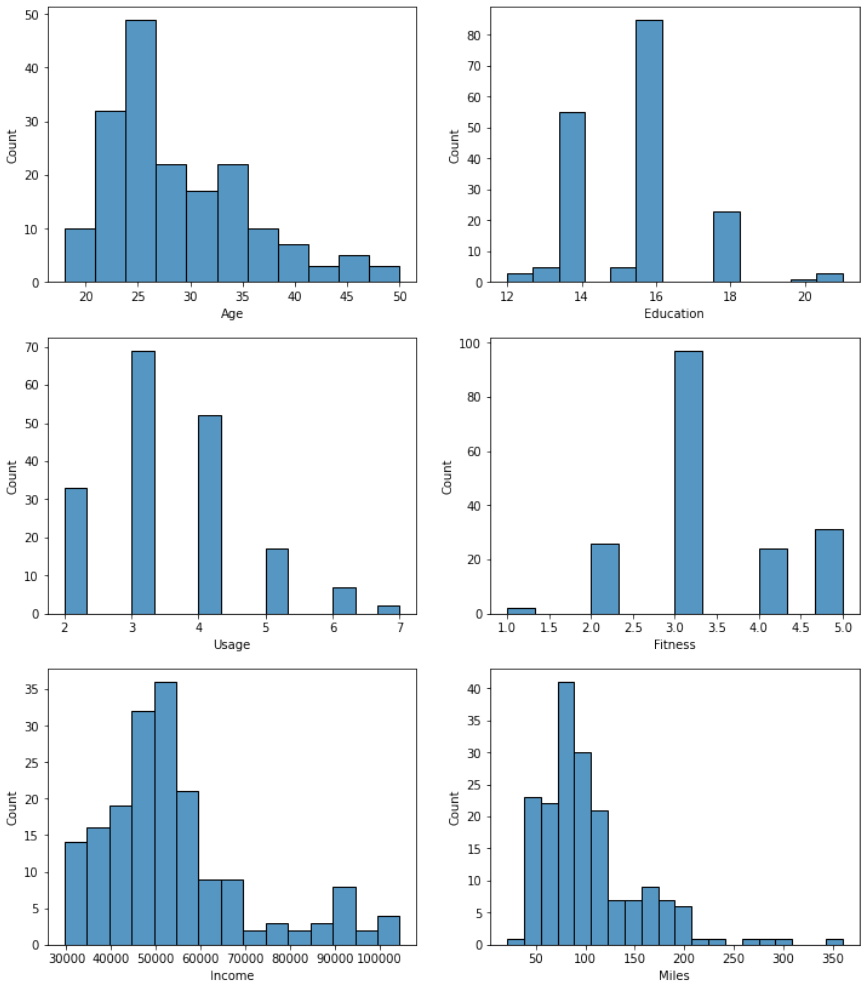
```
sns.countplot(data=data,x=data.Product,hue=data.Gender)
plt.show()
```




```
#Univariate analysis
sns.countplot(data=data,x=data.Product)
plt.show()
```

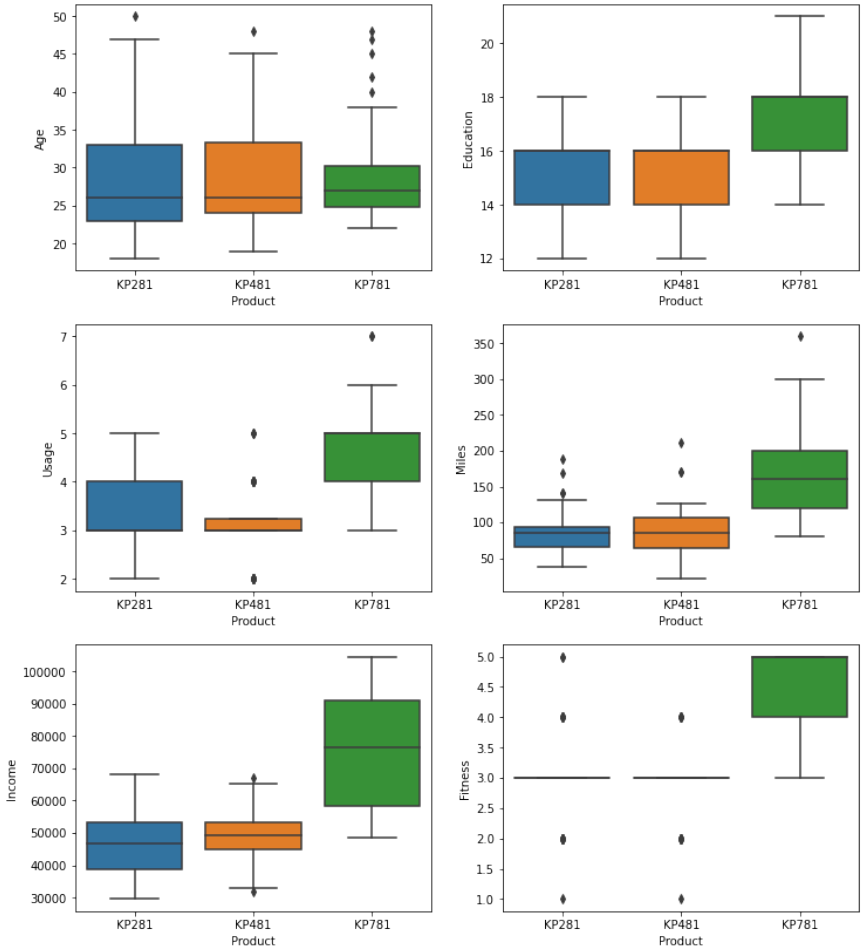



```
fig, axis = plt.subplots(nrows=3, ncols=2, figsize=(12, 10))
fig.subplots_adjust(top=1.2)

sns.histplot(data=data, x="Age", ax=axis[0,0])
sns.histplot(data=data, x="Education", ax=axis[0,1])
sns.histplot(data=data, x="Usage", ax=axis[1,0])
sns.histplot(data=data, x="Fitness", ax=axis[1,1])
sns.histplot(data=data, x="Income", ax=axis[2,0])
sns.histplot(data=data, x="Miles", ax=axis[2,1])
plt.show()
```



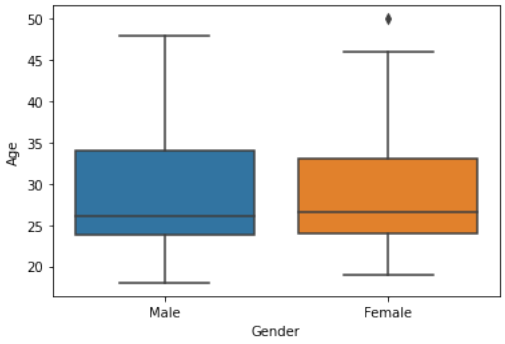

```
#Bivariate Analysis
fig,axis =plt.subplots(nrows=3, ncols=2, figsize=(12, 10))
fig.subplots_adjust(top=1.2)

sns.boxplot(data=data,x=data.Product,y=data.Age,ax=axis[0,0])
sns.boxplot(data=data,x=data.Product,y=data.Education,ax=axis[0,1])
sns.boxplot(data=data,x=data.Product,y=data.Usage,ax=axis[1,0])
sns.boxplot(data=data,x=data.Product,y=data.Miles,ax=axis[1,1])
sns.boxplot(data=data,x=data.Product,y=data.Income,ax=axis[2,0])
sns.boxplot(data=data,x=data.Product,y=data.Fitness,ax=axis[2,1])
```

```
<AxesSubplot:xlabel='Product', ylabel='Fitness'>
```

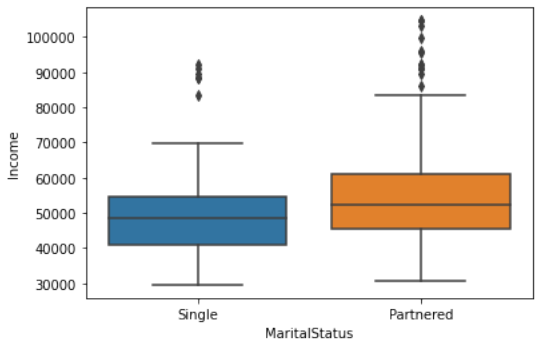



```
sns.boxplot(data=data,x=data.Gender,y=data.Age)
```

```
<AxesSubplot:xlabel='Gender', ylabel='Age'>
```




```
sns.boxplot(data=data,x=data.MaritalStatus,y=data.Income)
plt.show()
```



```
sns.pairplot(data)
plt.show()
```


```
#Marginal Probability
df1=pd.crosstab(index=data.Gender,columns=data.Product,margins=True)
df1
```

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

```
df1['KP281']
```

```
Gender
Female    40
Male      40
All       80
Name: KP281, dtype: int64
```

```
pd.crosstab(index=data.MaritalStatus,columns=data.Product,margins=True,normalize=True)
```

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

Conditional Probability

```
##Prob of purchase by female customer
data_female=data[data.Gender=='Female']
data_female
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles
2	KP281	19	Female	14	Partnered	4	3	30699	66
5	KP281	20	Female	14	Partnered	3	3	32973	66
6	KP281	21	Female	14	Partnered	3	3	35247	75
9	KP281	21	Female	15	Partnered	2	3	37521	85
11	KP281	22	Female	14	Partnered	3	2	35247	66
...	...	...	...	...	...	...	...	...	...
152	KP781	25	Female	18	Partnered	5	5	61006	200
157	KP781	26	Female	21	Single	4	3	69721	100
162	KP781	28	Female	18	Partnered	6	5	92131	180
167	KP781	30	Female	16	Partnered	6	5	90886	280
171	KP781	33	Female	18	Partnered	4	5	95866	200

76 rows × 9 columns

```
round(len(data_female)/len(data),2)
```

```
0.42
```

```
#Prob of buying KP281 given it is a female customer
round(len(data_female[data_female.Product=='KP281'])/len(data_female),2)
```

```
0.53
```

```
#Prob of buying KP481 given it is a female customer
round(len(data_female[data_female.Product=='KP481'])/len(data_female),2)
```

```
0.38
```

```
#Prob of buying KP781 given it is a female customer
round(len(data_female[data_female.Product=='KP781'])/len(data_female),2)
```

```
0.09
```

```
#Probability of male buying any of the product
data_male=data[data.Gender=='Male']
data_male.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
3	KP281	19	Male	12	Single	3	3	32973	85
4	KP281	20	Male	13	Partnered	4	2	35247	47
7	KP281	21	Male	13	Single	3	3	32973	85

```
round(len(data_male)/len(data),2)
```

```
0.58
```

```
#Probability of byuing KP281 given it sis male customer
round(len(data_male[data_male.Product=='KP281'])/len(data_male),2)
```

```
0.38
```

```
#Probability of byuing KP481 given it sis male customer
round(len(data_male[data_male.Product=='KP481'])/len(data_male),2)
```

```
0.3
```

```
#Probability of byuing KP781 given it sis male customer
round(len(data_male[data_male.Product=='KP781'])/len(data_male),2)
```

```
0.32
```

```
#Probability of purchase by partnered customer
data_partner=data[data.MaritalStatus=='Partnered']
data_partner.head()
```

	Product	Age	Gender	Education	MaritalStatus	Usage	Fitness	Income	Miles
2	KP281	19	Female	14	Partnered	4	3	30699	66
4	KP281	20	Male	13	Partnered	4	2	35247	47
5	KP281	20	Female	14	Partnered	3	3	32973	66
6	KP281	21	Female	14	Partnered	3	3	35247	75
9	KP281	21	Female	15	Partnered	2	3	37521	85

```
len(data_partner)/len(data)
```

```
0.5944444444444444
```

```
#Probability of buying KP281 given partnered
round(len(data_partner[data_partner.Product=='KP281'])/len(data_partner),2)
```

```
0.45
```

```
#Probability of buying KP481 given partnered
round(len(data_partner[data_partner.Product=='KP481'])/len(data_partner),2)
```

```
0.34
```

```
#Probability of buying KP781 given partnered
round(len(data_partner[data_partner.Product=='KP781'])/len(data_partner),2)
```

```
0.21
```

```
data_single=data[data.MaritalStatus=='Single']
data_single.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
3	KP281	19	Male	12	Single	3	3	32973	85
7	KP281	21	Male	13	Single	3	3	32973	85