

```
In [1]: import pandas as pd
```

```
In [2]: df=pd.read_csv("aerofit.csv")
```

```
In [3]: df
```

```
Out[3]:
```

	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

```
In [4]: df.shape
```

```
Out[4]: (180, 9)
```

```
In [5]: df.describe()
```

```
Out[5]:
```

	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

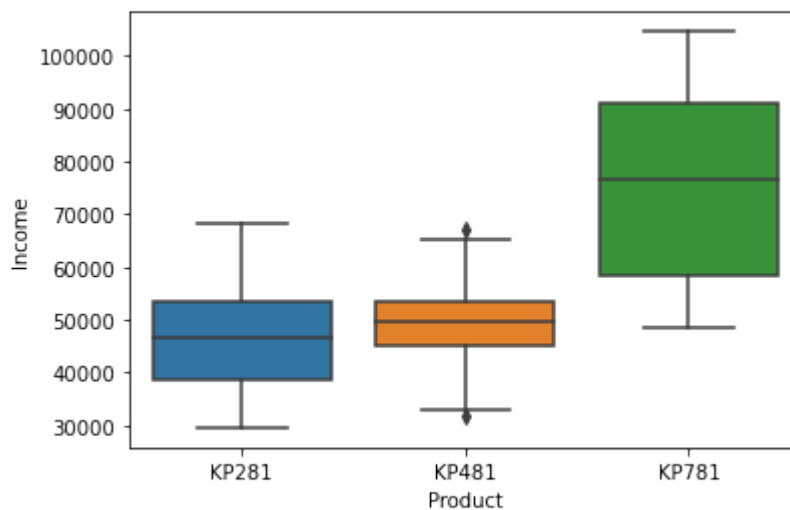
```
In [6]: df.isnull().sum()/ len(df) *100
```

```
Out[6]: Product      0.0  
Age      0.0  
Gender    0.0  
Education 0.0  
MaritalStatus 0.0  
Usage     0.0  
Fitness   0.0  
Income    0.0  
Miles     0.0  
dtype: float64
```

```
In [7]: import seaborn as sbn
```

```
sbn.boxplot(x='Product', y='Income', data =df)
```

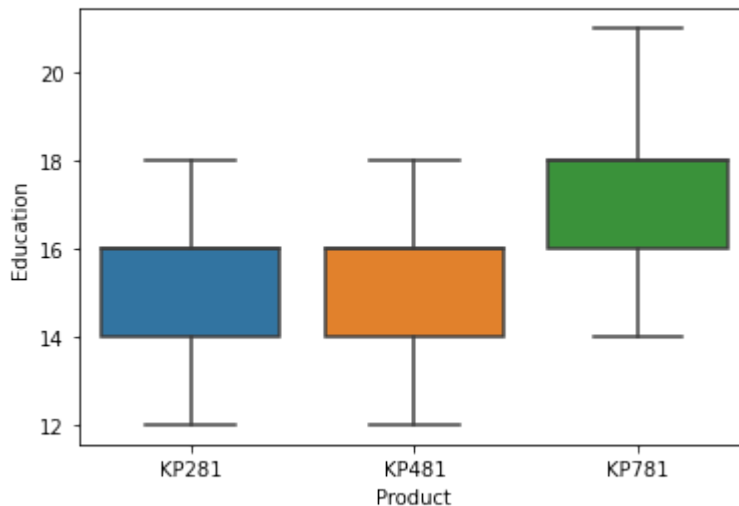
```
Out[7]: <AxesSubplot:xlabel='Product', ylabel='Income'>
```



```
In [ ]: #insight
```

```
In [8]: sbn.boxplot(x='Product', y='Education', data =df)
```

```
Out[8]: <AxesSubplot:xlabel='Product', ylabel='Education'>
```



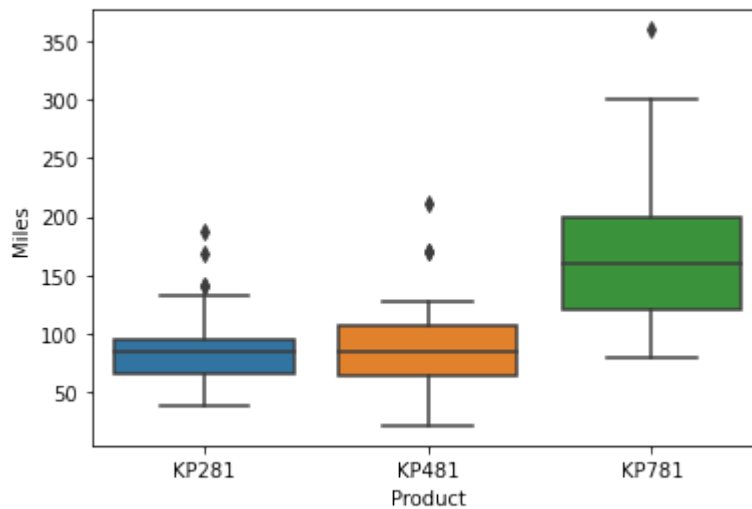
```
In [10]: df.groupby('Product')['Education'].describe()
```

```
Out[10]:
```

	count	mean	std	min	25%	50%	75%	max
Product								
KP281	80.0	15.037500	1.216383	12.0	14.0	16.0	16.0	18.0
KP481	60.0	15.116667	1.222552	12.0	14.0	16.0	16.0	18.0
KP781	40.0	17.325000	1.639066	14.0	16.0	18.0	18.0	21.0

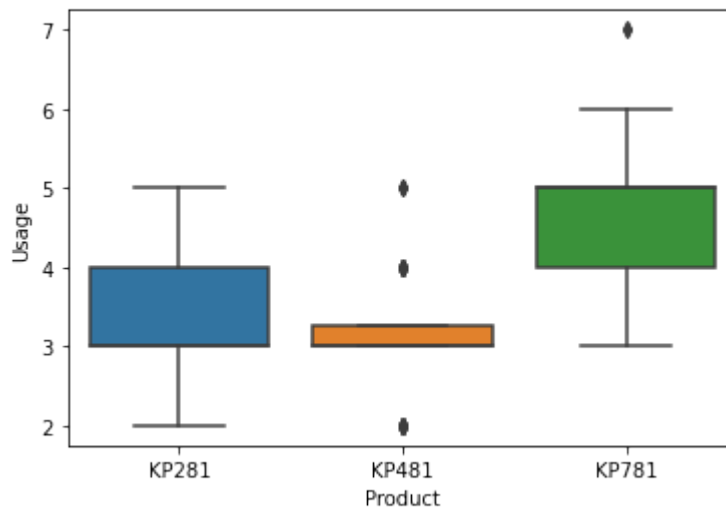
```
In [11]: sns.boxplot(x='Product', y='Miles', data =df)
```

```
Out[11]: <AxesSubplot:xlabel='Product', ylabel='Miles'>
```



```
In [12]: sns.boxplot(x='Product', y='Usage', data =df)
```

```
Out[12]: <AxesSubplot:xlabel='Product', ylabel='Usage'>
```



```
In [13]: df.groupby('Product')['Usage'].describe()
```

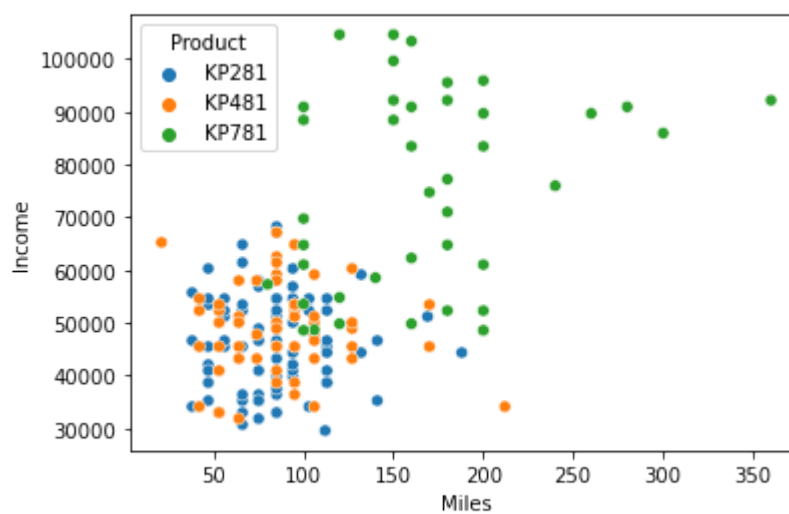
```
Out[13]:
```

	count	mean	std	min	25%	50%	75%	max
Product								
KP281	80.0	3.087500	0.782624	2.0	3.0	3.0	4.00	5.0
KP481	60.0	3.066667	0.799717	2.0	3.0	3.0	3.25	5.0
KP781	40.0	4.775000	0.946993	3.0	4.0	5.0	5.00	7.0

```
In [ ]: #insights
```

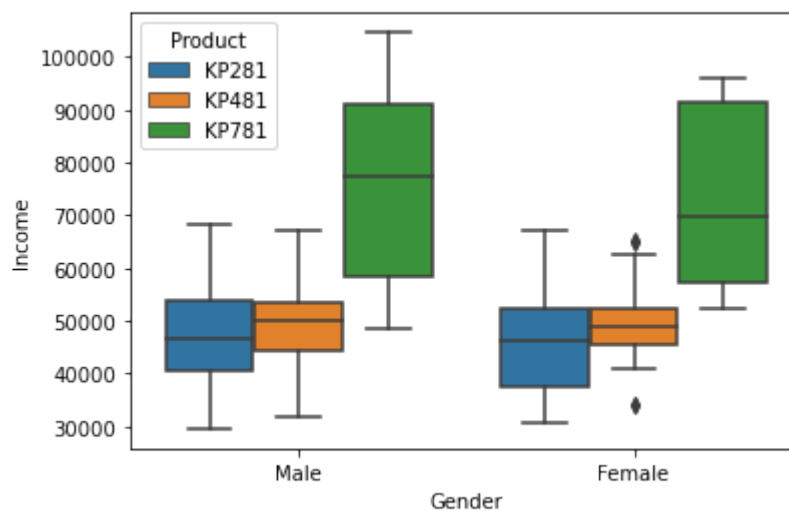
```
In [15]: sns.scatterplot(x='Miles', y='Income', data=df, hue='Product')
```

```
Out[15]: <AxesSubplot:xlabel='Miles', ylabel='Income'>
```



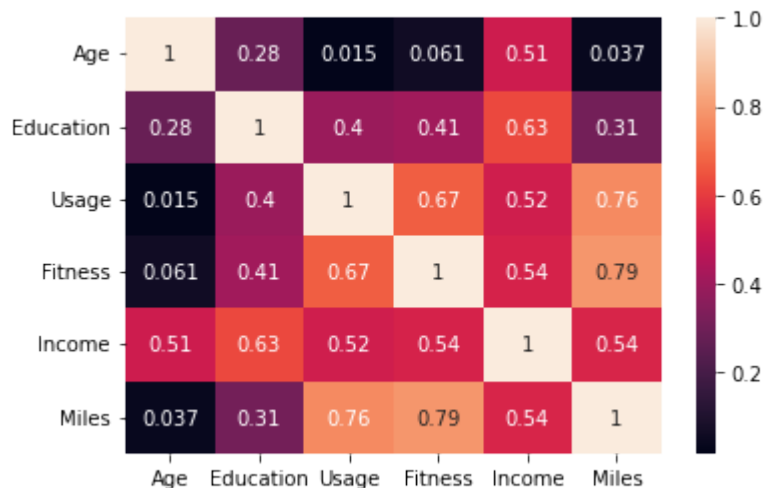
```
In [17]: sns.boxplot(x='Gender', y='Income', data=df, hue='Product')
```

```
Out[17]: <AxesSubplot:xlabel='Gender', ylabel='Income'>
```



```
In [19]: sbn.heatmap(df.corr(), annot=True)
```

```
Out[19]: <AxesSubplot:>
```



```
In [20]: #insights
```

```
In [21]: # Probability to extract more insights from it
#1000 pieces of KP781, HOW MANY OF THEM WILL BE BOUGHT BY FEMALES?
```

```
In [22]: #2 CATEGORICAL SERIES
```

```
pd.crosstab(index=df['Gender'], columns=df['Product'])
```

```
Out[22]:
```

Product	KP281	KP481	KP781
Gender			
Female	40	29	7
Male	40	31	33

```
In [23]: pd.crosstab(index=df['Gender'], columns=df['Product'], margins=True)
```

```
Out[23]:
```

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

```
In [25]: #1000 pieces of KP781, HOW MANY OF THEM WILL BE BOUGHT BY FEMALES?
7/40 * 1000
```

```
Out[25]: 175.0
```

```
In [27]: #If 200 males walk into my store, what should be my MIN inventory for
round(31/104 *200)
```

Out[27]: 60

```
In [28]: pd.crosstab(index=df['Gender'], columns=df['Product'], margins=True, r
```

Out[28]:

Product	KP281	KP481	KP781
Gender			
Female	52.631579	38.157895	9.210526
Male	38.461538	29.807692	31.730769
All	44.444444	33.333333	22.222222

```
In [ ]: #next week is big billion day sales- How much should be the inventory
#a- I am expecting around 2000 potential customers on my website
# b. How many males & females will buy kp281
```

```
In [29]: pd.crosstab(index=df['Gender'], columns=df['Product'], margins= True)
```

Out[29]:

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

```
In [31]: #a
round(80/180*2000)
```

Out[31]: 889

In []: