Q1. What is the probability of a male customer buying a KP781 Treadmill?

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

Insights

- Probability of a Male customer buying **KP781** is ***31.73%***.

Q2. What is the total count of each product present in the dataset?

Product	KP281	KP481	KP781
count	80.0	60.0	40.0
P-Val	44.44%	33.33%	22.22%

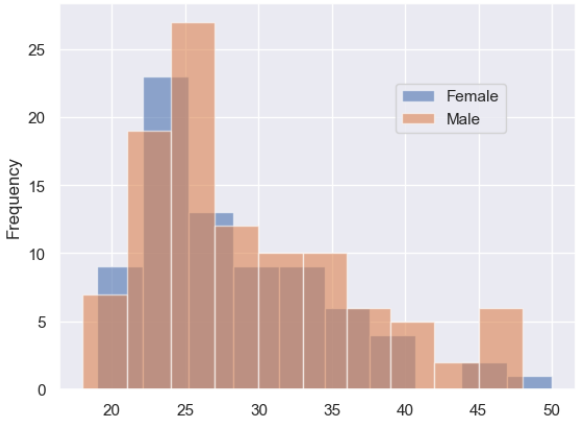
Insights

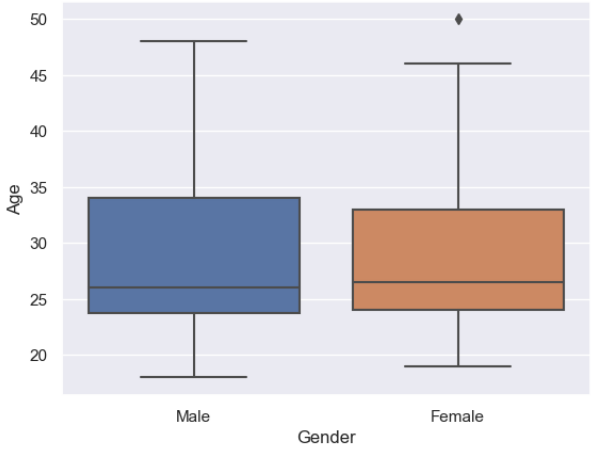
- **KP281** have count **80** with **44.44%** probability.
- **KP481** have count **60** with **33.33%** probability.
- **KP781** have count **40** with **22.22%** probability.

Q3. Describe the Age & Gender distribution of all the customers?

Insights

Gender	count	mean	std	min	25%	50%	75%	max
Female	76.0	28.565789	6.342104	19.0	24.00	26.5	33.0	50.0
Male	104.0	28.951923	7.377978	18.0	23.75	26.0	34.0	48.0



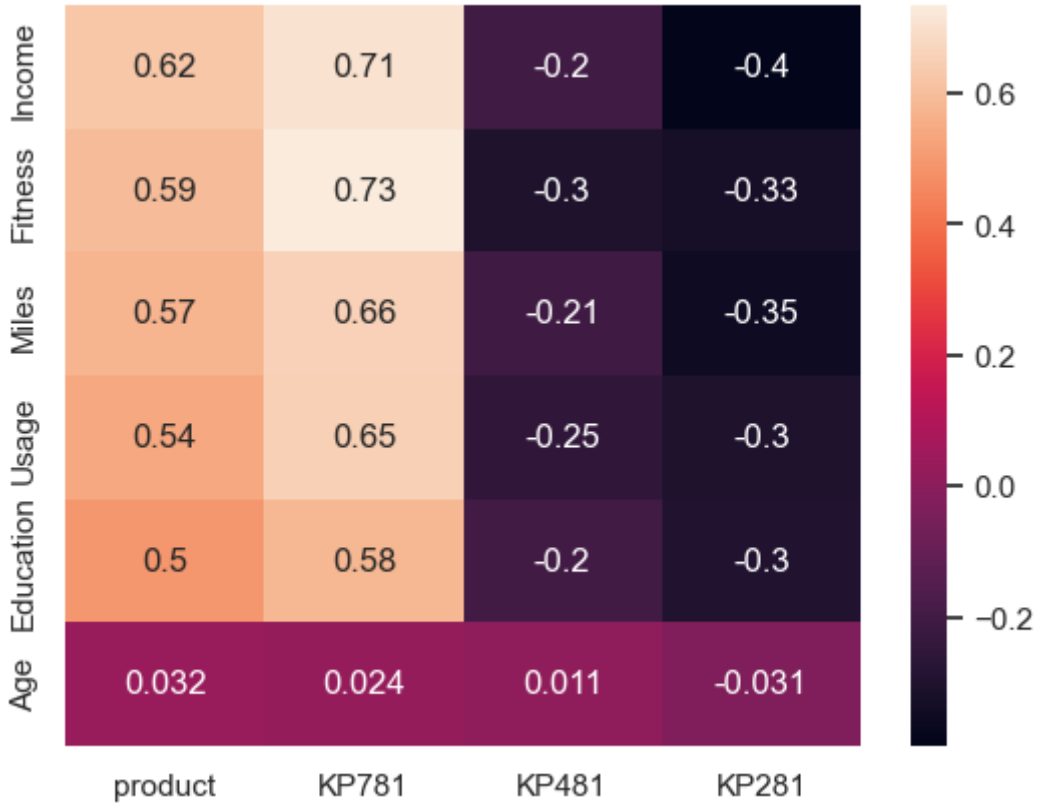


Gender	Age
Female	28.565789
Male	28.951923

Insights

- Using **T-Test,** Probability of *Male* age higher than *Female* age have a p-Vale of `35.679%.`
- There is a small difference between mean Age of Male & Female.
- When following **95%** Confidence Interval. So, we failed to prove that Male age are Higher than Female Age. among Customers.

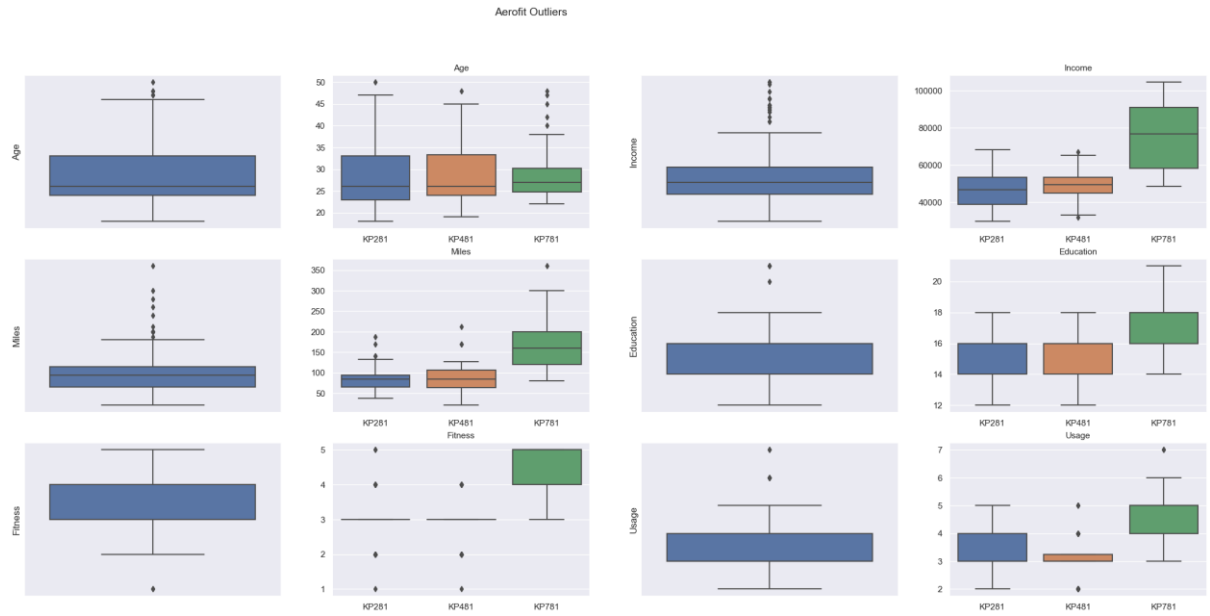
Q4. Top 3 features having the highest correlations with the Product column. and why?



Insights

- Product Have higher co-relation with **Income**, **Fitness** & **Miles**.
  - For **KP781** we have observed same co-relation.
  - But, for **KP481** Top 3 co-relation are `Age`, `Education` & `Income`.
  - But, for **KP281** Top 3 co-relation are `Age`, `Education` & `Usage`.
- As, we know **KP781** is the expansive & better one, mostly preferred by Athletes or Fitness enthusiast, who's also have higher Income.
- But, for **KP481, KP281** people with good Education, who are health conscious, want to have a Treadmill but can't afford an expansive one. customers.

Q5. Were there any outliers present in the Data? If yes, suggest suitable method for their treatment?

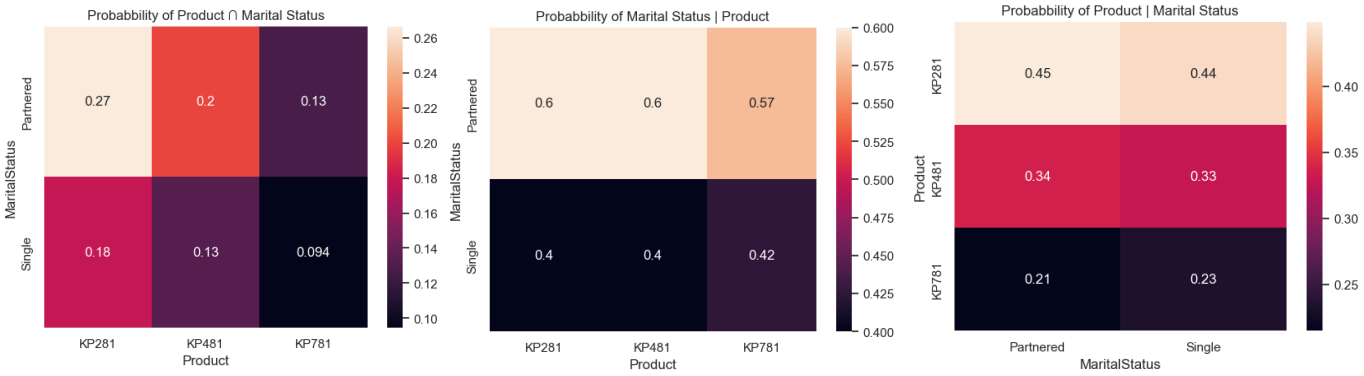



```
def check_outlier(df, x):
    Q1 = df[x].quantile(0.25)
    Q3 = df[x].quantile(0.75)
    IQR = Q3 - Q1
    lower = Q1 - 1.5*IQR
    upper = Q3 + 1.5*IQR
    lower_outlier = df[x][df[x] < lower]
    upper_outlier = df[x][df[x] > upper]

    return {
        'lower': {'list': lower_outlier, 'length': len(lower_outlier)},
        'upper': {'list': upper_outlier, 'length': len(upper_outlier)}
      }

for i in aerofit[['Age', 'Income', 'Miles', 'Education', 'Fitness',
'Usage']].columns:
    outlier = check_outlier(aerofit, i)
    print("{} : ({}, {})".format(i, outlier['lower']['length'],
outlier['upper']['length']))
```

```
Age : (0, 5)
Income : (0, 19)
Miles : (0, 13)
Education : (0, 4)
Fitness : (2, 0)
Usage : (0, 9)
```

- When, we remove these outliers it might effect credibility of `Product` specific analysis. Because Some of the product dependent on the data.

Q6. Marital Status Implies no significant information on the usages of different Treadmills? (T/F)



Insights

- From Above Probability Heatmap, observed that for each product given marital status have nearly same probability.

Thus, Statement is True

Q7. The variance of Income in lower ages is smaller as compare to the variance in higher ages. In Statistics , this is known as

```
low_income_age =
aerofit.copy().loc[aerofit['Age']<aerofit['Age'].mean()][['Income', "Age"]]
high_income_age =
aerofit.copy().loc[aerofit['Age']>aerofit['Age'].mean()][['Income', "Age"]]
```

```
low_income_age['Income'].var() < high_income_age['Income'].var()
```
True

```
low_income_age['Income'].std() < high_income_age['Income'].std()
```
True

Insights

- Here Variance & Standard Deviation in low Age Income Group is less than Higher Age Income Group.
- Thus, it results a funnel shape & It's known as **Heteroscedasticity.**

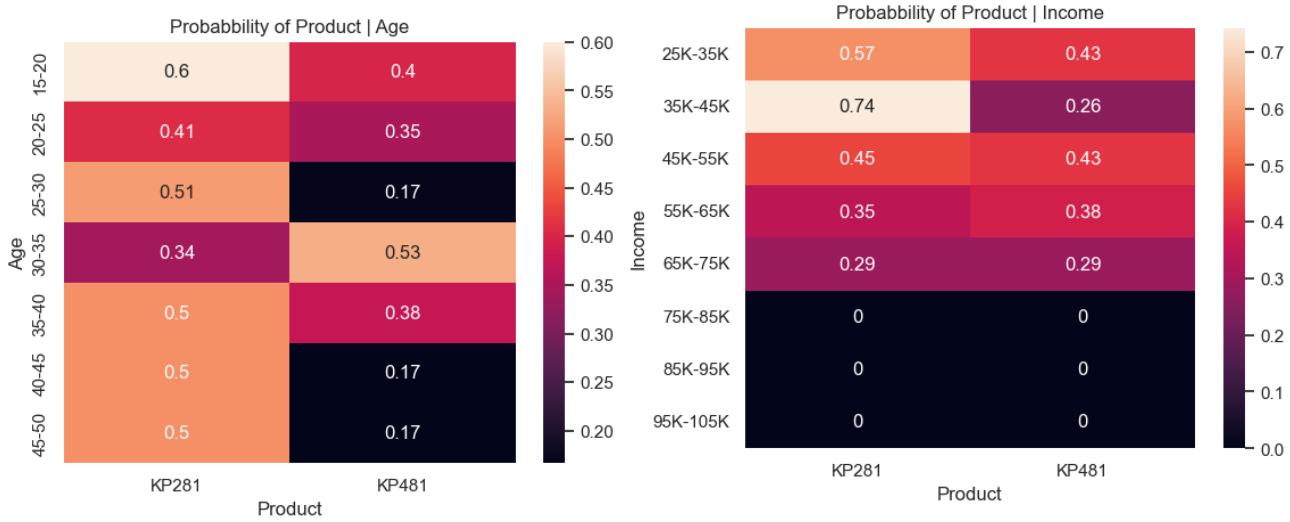
Q8. What proportion of woman have brought the KP781 Treadmill? Provide reason of Answer.

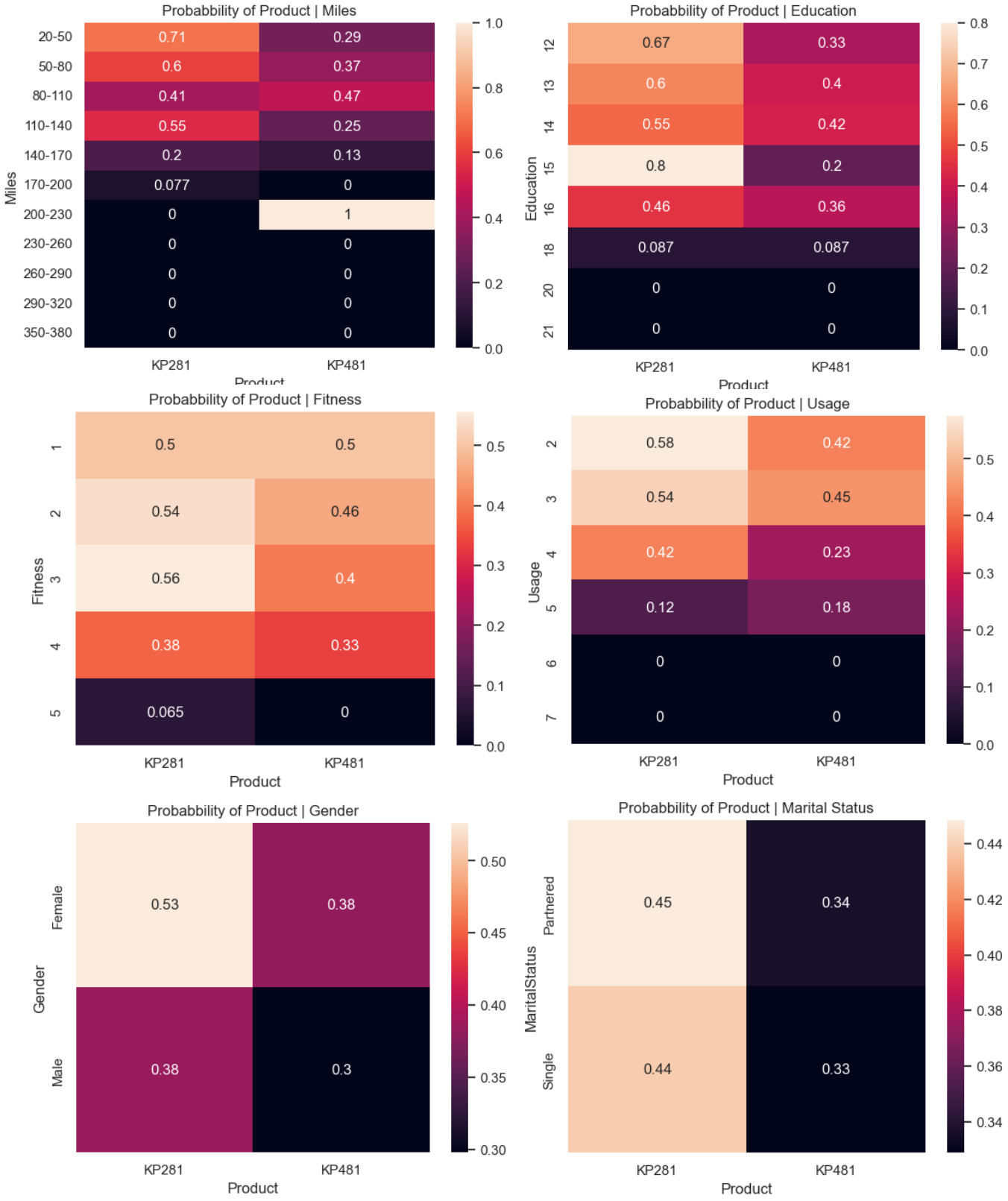
Product	KP281	KP481	KP781
Gender			
Female	52.631579	38.157895	9.210526
Male	38.461538	29.807692	31.730769

Insights

- Probability of a Female customer buying **KP781** is ***9.21%***.

Q9. Distinguish between Customer Profiles for KP281 and KP481 Treadmill.





Insights

- Most of **30-35** year old People most likely to prefer **KP481** over KP281.
- People have Income Between **25K** to **75K** most likely to prefer **KP281** over KP481**.**
- People who covered **80-110 & 200-230** Miles most likely to prefer **KP481** but **20-80** & **110-200** prefer **KP481.**

- People who have Education level among **12-16** most likely to prefer **KP281** over KP481**.**
- Among all Fitness people are most likely to prefer **KP281** over KP481**.**
- People with Usage level among **2-4** prefer **KP281** but **level 5** people prefer **KP481**.

Q10. The overall Probability of purchase for KP281, KP481 & KP781 Treadmill is __ , ___ , ___.

```
Product

KP281  44.44%
KP481  33.33%
KP781  22.22%
```

Q11. Give conditions when you will and when you 'll not recommended KP781 Treadmill to a Customer?

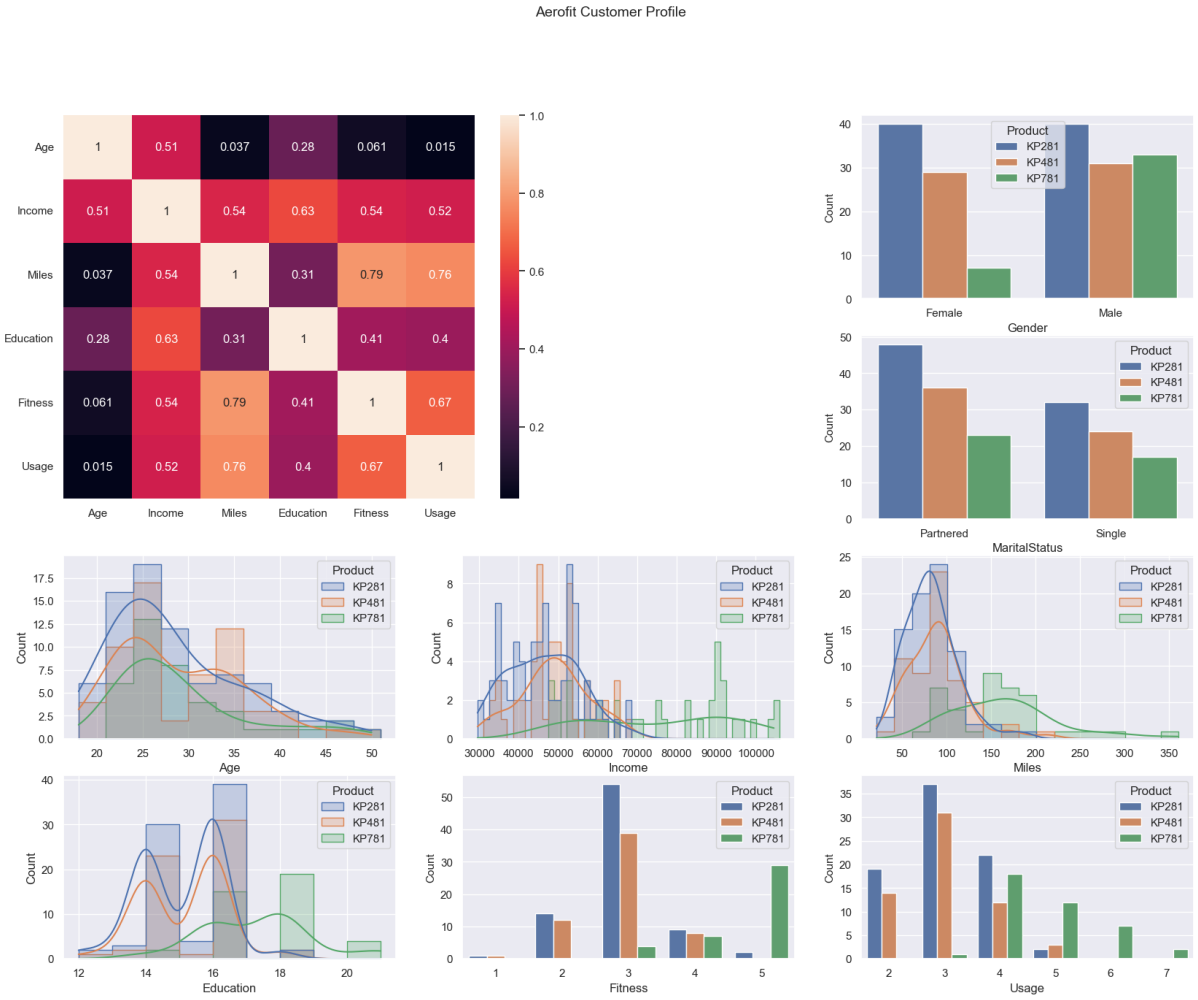
When to recommend KP781

- To Male's
- Age between **20-30**.
- Income with **90K**.
- Who intended or already covered **150-200** Miles.
- Have Education between **16-19**.
- Have Fitness of **level5**.
- Have Usages level of **4-5**.

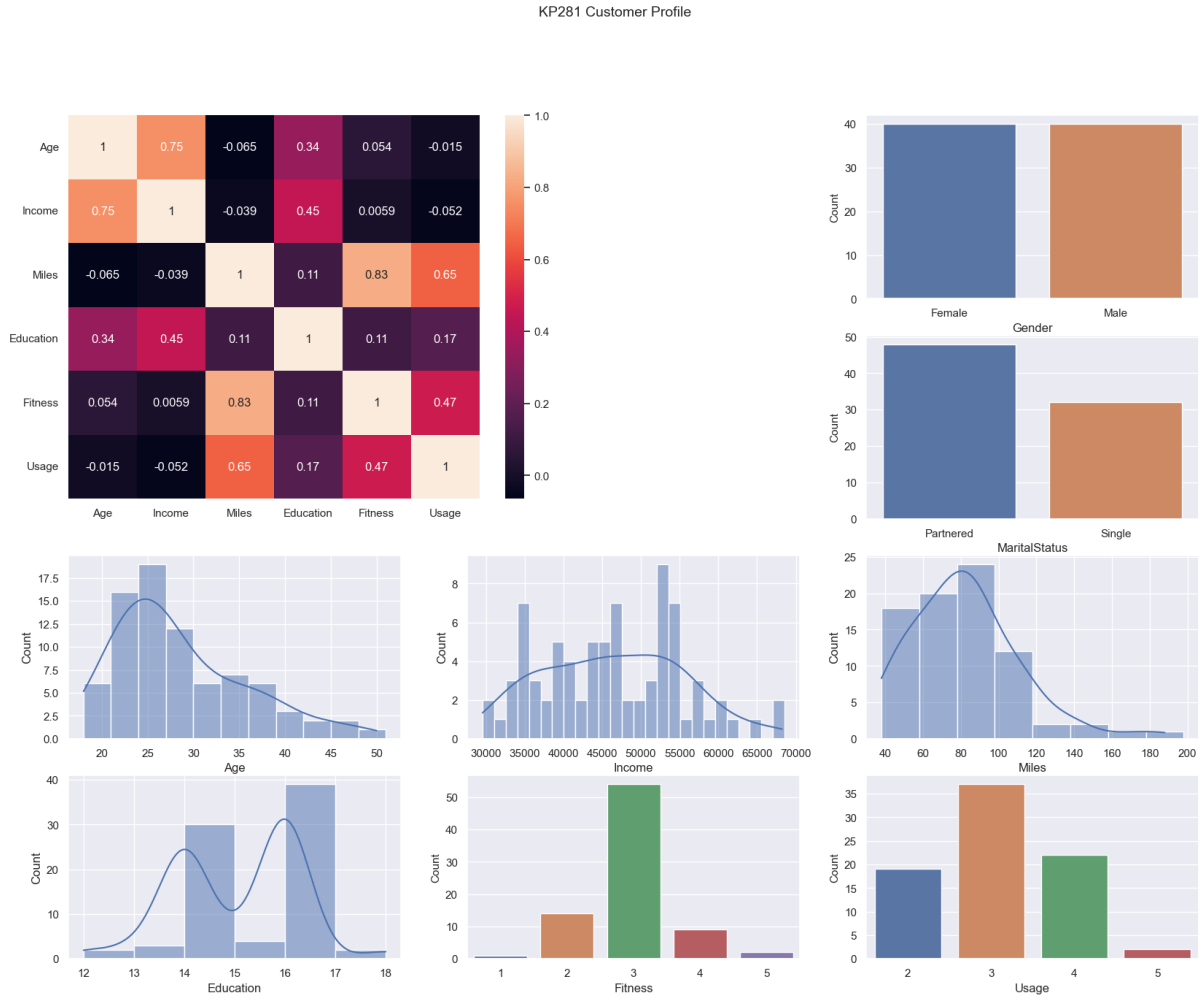
When not to recomend KP781

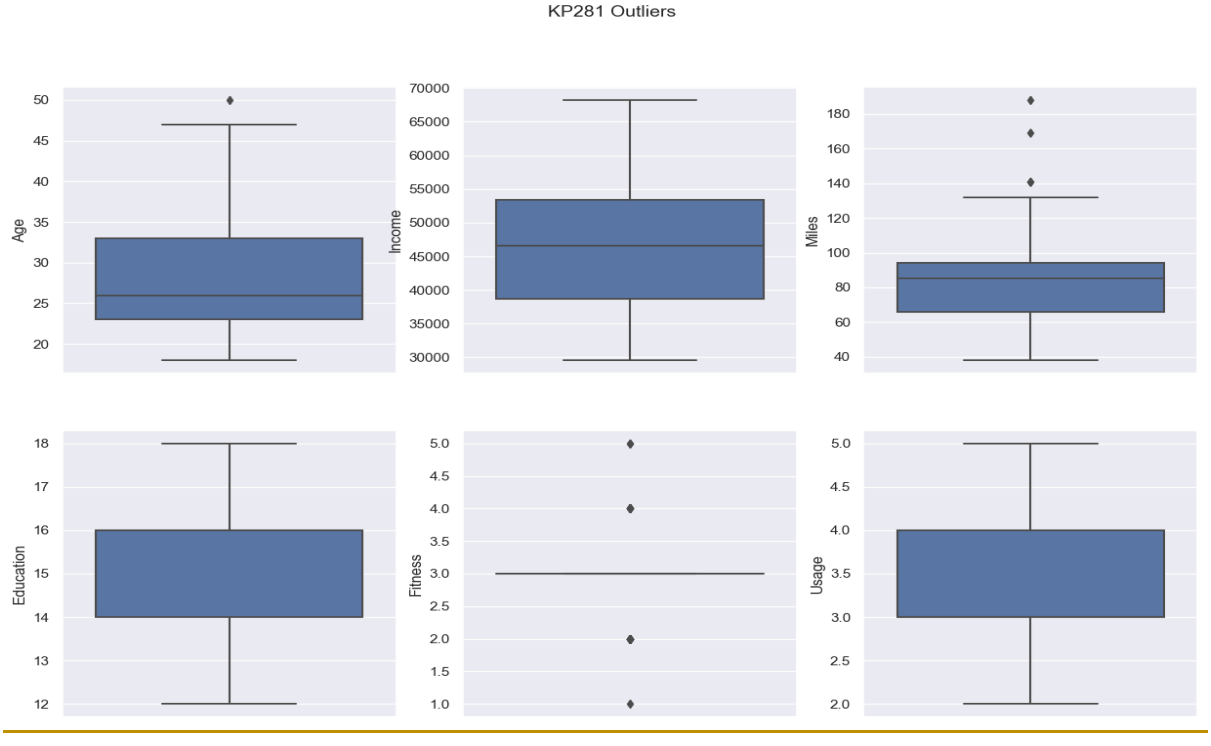
* To Female's
* Age above **35**.
* Income below **75K**.
* Have Fitness level less than **3**.
* Have Usages level less than **4**.

Customer Profile of Different Product

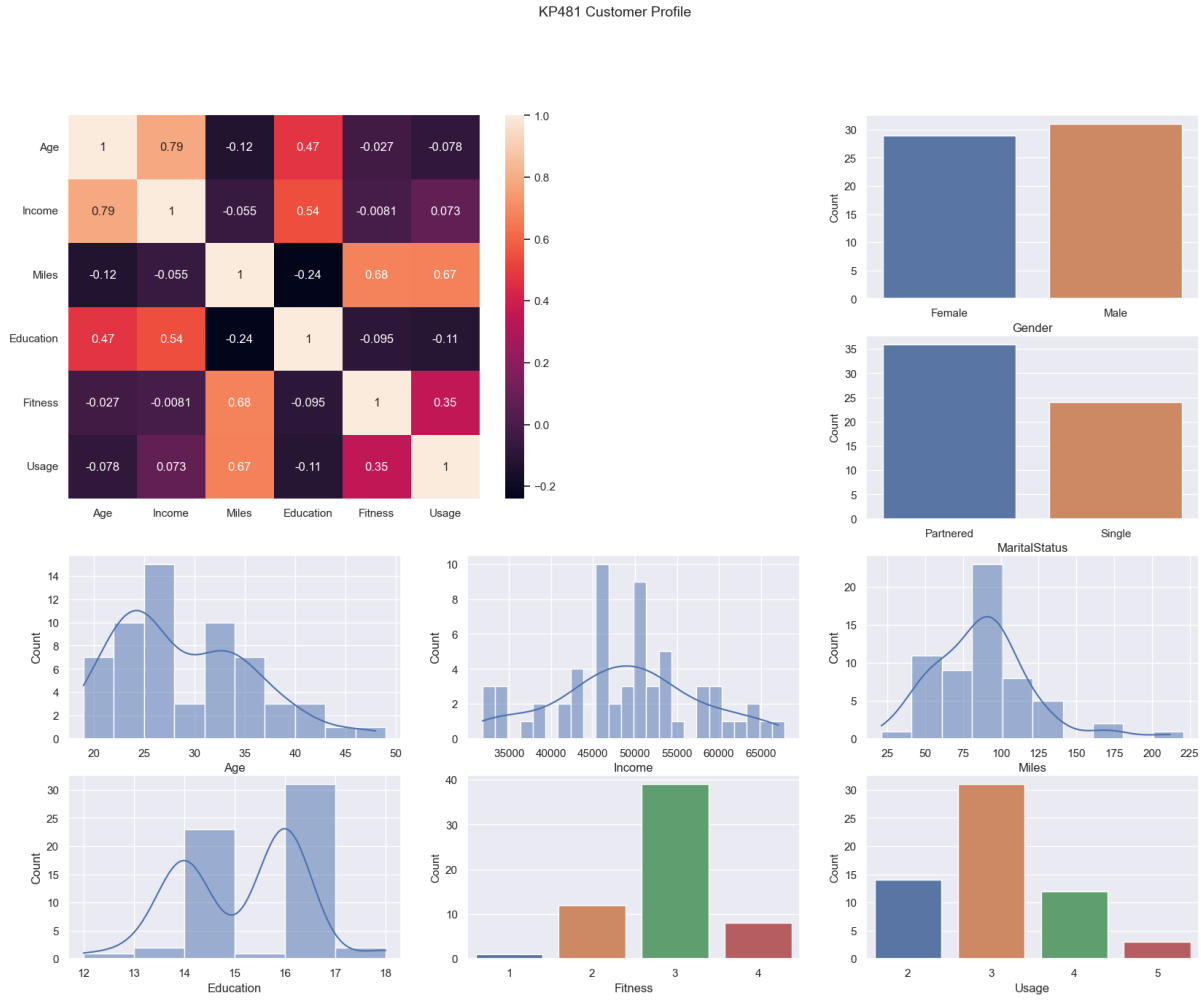


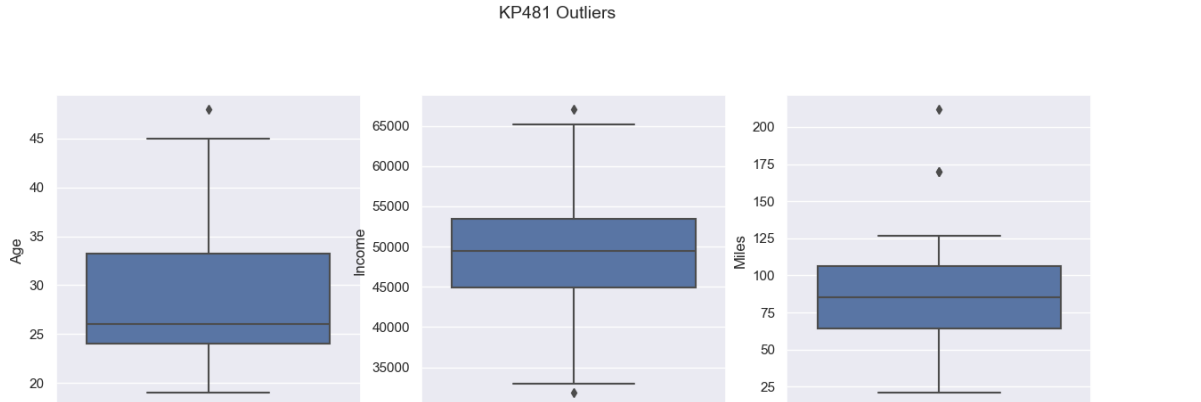
Profile of KP281





Profile of KP481





Profile of KP781