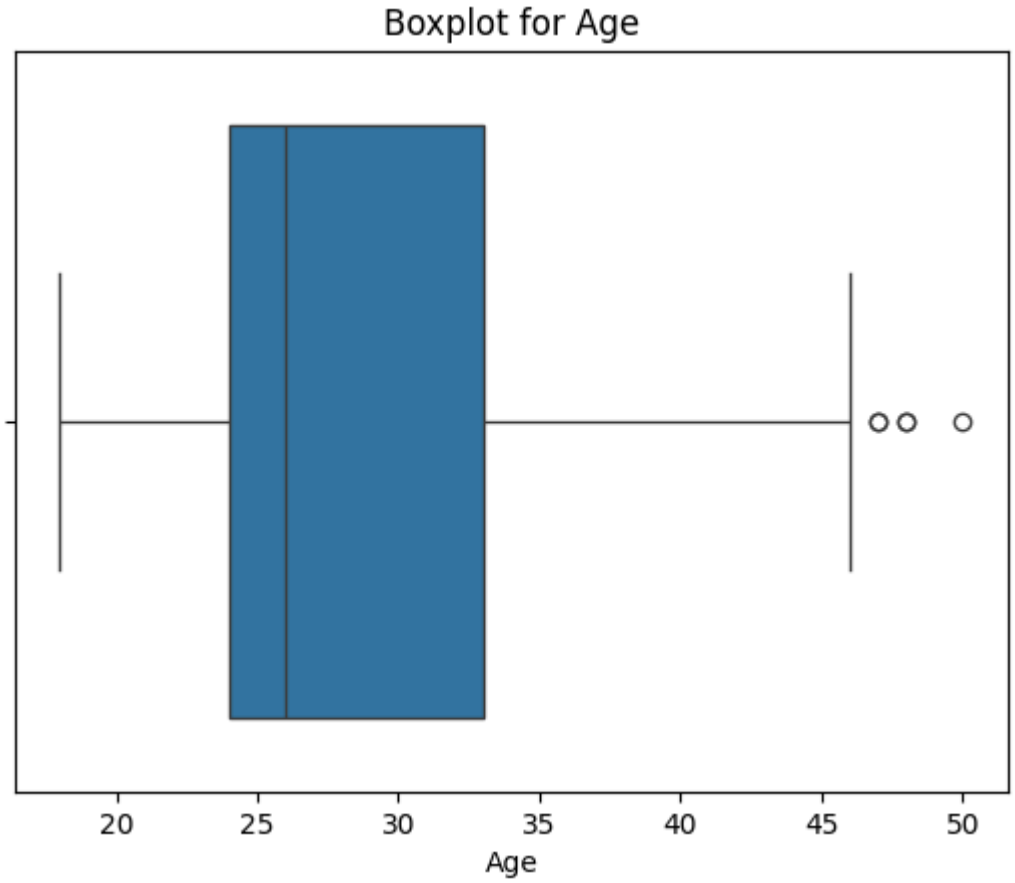
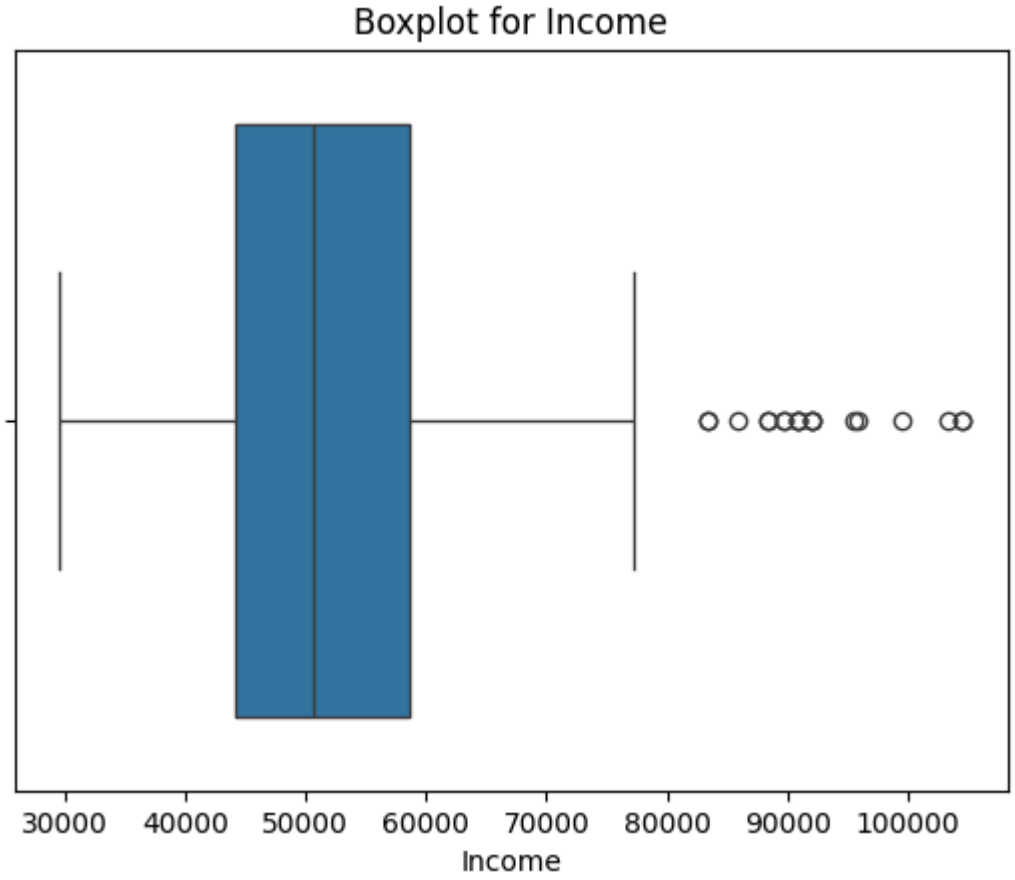
In [4]:
```
# Import Required Libraries
import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Load the dataset
os.chdir("/Users/ayeshasiddiqha/Downloads")
data = pd.read_csv('aerofit_treadmill.csv')

# Display basic details
print(data.info())
print(data.head())

# Check for missing values
print(data.isnull().sum())

# Statistical Summary
print(data.describe())

# Data Types
print(data.dtypes)

# Boxplot for outlier detection
columns = ['Age', 'Income', 'Usage', 'Fitness', 'Miles']
for col in columns:
    sns.boxplot(x=data[col])
    plt.title(f'Boxplot for {col}')
    plt.show()

# Check Skewness and Median vs Mean for Outliers
print(data[columns].describe())

# Distribution plots for continuous variables
for col in columns:
    sns.histplot(data[col], kde=True)
    plt.title(f'Histogram for {col}')
    plt.show()

# Count plot for categorical variables
sns.countplot(x='Gender', data=data)
plt.title('Count of Customers by Gender')
plt.show()

sns.countplot(x='MaritalStatus', data=data)
plt.title('Count of Customers by Marital Status')
plt.show()

sns.countplot(x='Product', data=data)
plt.title('Count of Customers by Product Purchased')
plt.show()

# Boxplots for Age and Income by Product Purchased
sns.boxplot(x='Product', y='Age', data=data)
plt.title('Age Distribution by Product')
plt.show()
```

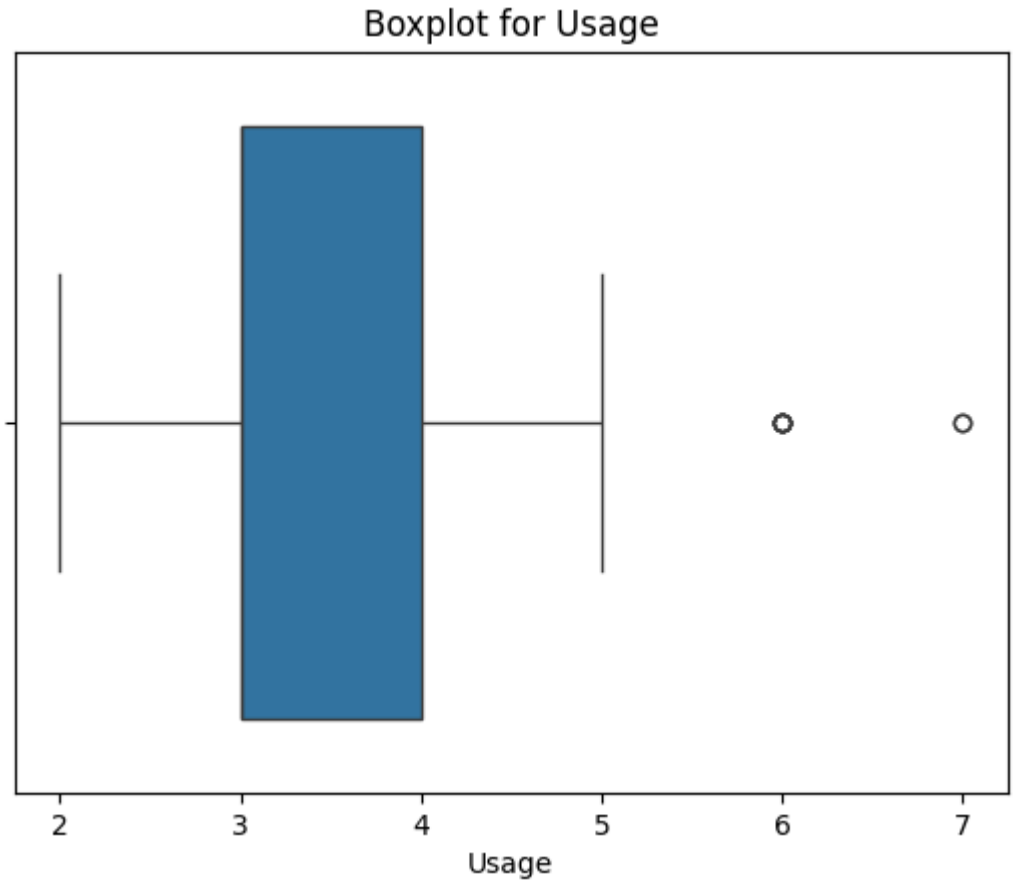
```
sns.boxplot(x='Product', y='Income', data=data)
plt.title('Income Distribution by Product')
plt.show()

# Pairplot for Correlation Analysis
sns.pairplot(data, hue='Product', diag_kind='kde')
plt.show()

# Correlation Heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(data.corr(numeric_only=True), annot=True, cmap='coolwarm')
plt.title('Correlation Heatmap')
plt.show()

# Marginal Probabilities
marginal = data['Product'].value_counts(normalize=True)
print("Marginal Probabilities:")
print(marginal)

# Conditional Probabilities
conditional_gender = pd.crosstab(data['Gender'], data['Product'], normali
print("Conditional Probability by Gender:")
print(conditional_gender)

conditional_marital = pd.crosstab(data['MaritalStatus'], data['Product'],
print("Conditional Probability by Marital Status:")
print(conditional_marital)

# Create customer profiles based on averages per product
# Numeric profiling
numeric_profiles = data.groupby('Product').mean(numeric_only=True)
print("Customer Profiles (Continuous Variables):")
print(numeric_profiles)

# Categorical profiling
categorical_profiles = data.groupby('Product')[['Gender', 'MaritalStatus'
print("Customer Profiles (Categorical Variables):")
print(categorical_profiles)

# Actionable Insights
print("\n--- Actionable Insights ---")
print("1. KP281 is purchased by customers with lower income, younger age,
print("2. KP481 caters to mid-income customers with moderate fitness and
print("3. KP781 is preferred by high-income customers with excellent fitn
print("Recommendations:")
print("1. Target younger demographics with affordable KP281 through socia
print("2. Use fitness-focused campaigns for KP781 targeting high-income g
print("3. Highlight balanced features of KP481 for middle-income, fitness
print("4. Offer KP281 with affordable accessories for beginners (e.g., fi
print("5. Include premium memberships or coaching sessions with KP781.")
print("6. Place KP281 in regions with a younger population and lower aver
print("7. Promote KP781 in urban areas with higher-income groups.")
print("8. Introduce discounts during seasonal promotions for KP281 and KP
print("9. Encourage KP281 customers to upgrade to KP481 with loyalty disc
```

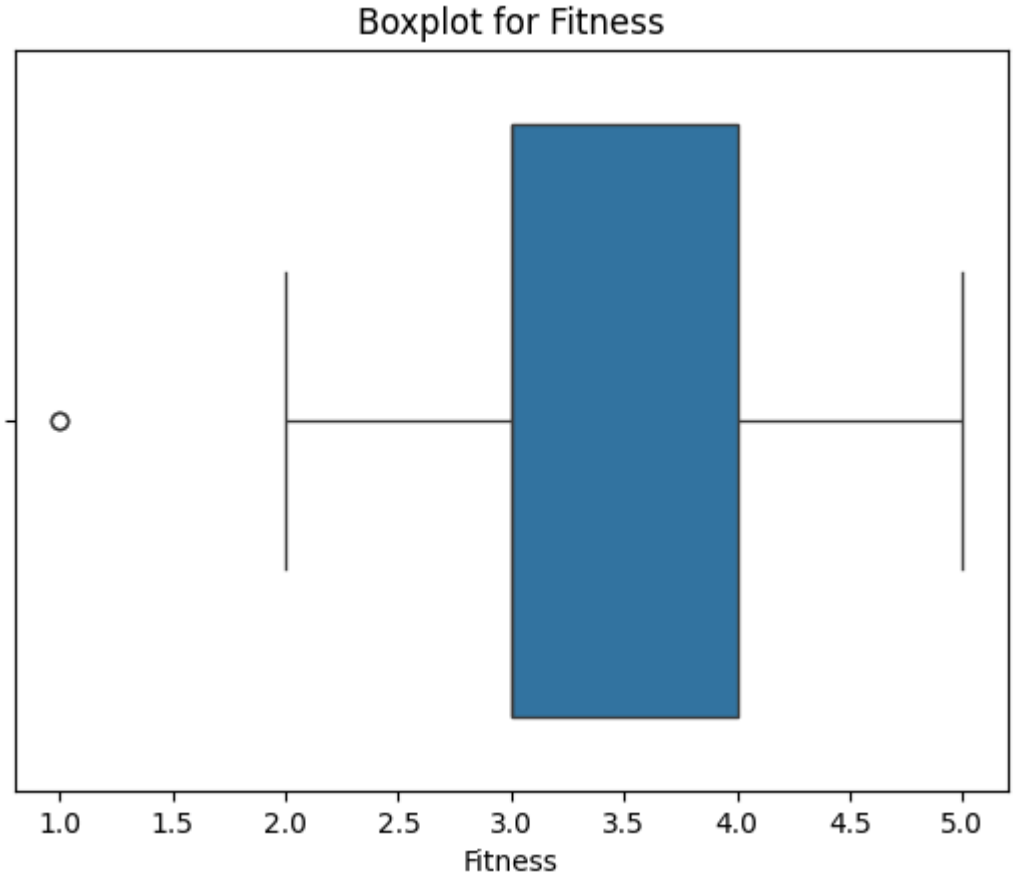
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 9 columns):
 #   Column         Non-Null Count  Dtype 
---  ------         --------------  ----- 
 0   Product        180 non-null    object
 1   Age            180 non-null    int64 
 2   Gender         180 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
None
  Product  Age  Gender  Education MaritalStatus  Usage  Fitness  Income  M
iles
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
Product          0
Age              0
Gender           0
Education        0
MaritalStatus    0
Usage            0
Fitness          0
Income           0
Miles            0
dtype: int64
              Age   Education       Usage     Fitness         Income  \
count  180.000000  180.000000  180.000000  180.000000     180.000000   
mean    28.788889   15.572222    3.455556    3.311111   53719.577778   
std      6.943498    1.617055    1.084797    0.958869   16506.684226   
min     18.000000   12.000000    2.000000    1.000000   29562.000000   
25%     24.000000   14.000000    3.000000    3.000000   44058.750000   
50%     26.000000   16.000000    3.000000    3.000000   50596.500000   
75%     33.000000   16.000000    4.000000    4.000000   58668.000000   
max     50.000000   21.000000    7.000000    5.000000  104581.000000   

            Miles  
count  180.000000  
mean   103.194444  
std     51.863605  
min     21.000000  
25%     66.000000  
50%     94.000000  
75%    114.750000  
max    360.000000  
Product          object
Age               int64
```

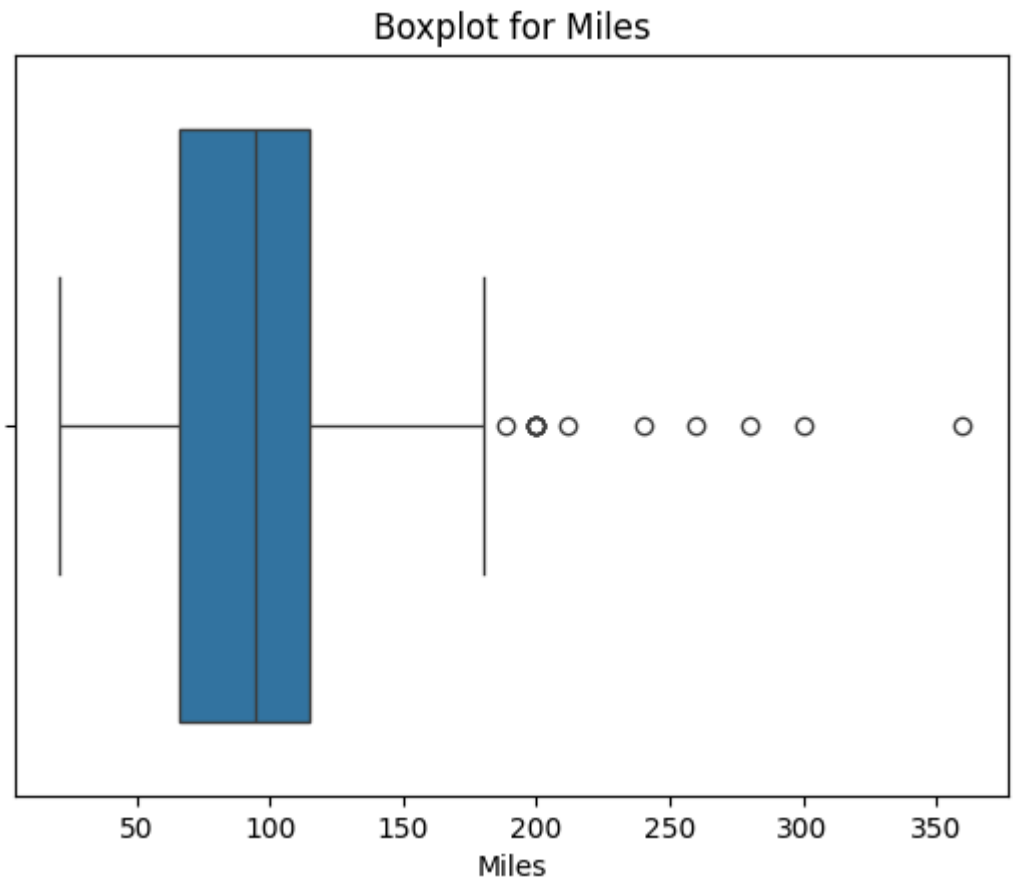
```
Gender           object
Education         int64
MaritalStatus    object
Usage             int64
Fitness           int64
Income            int64
Miles             int64
dtype: object
```



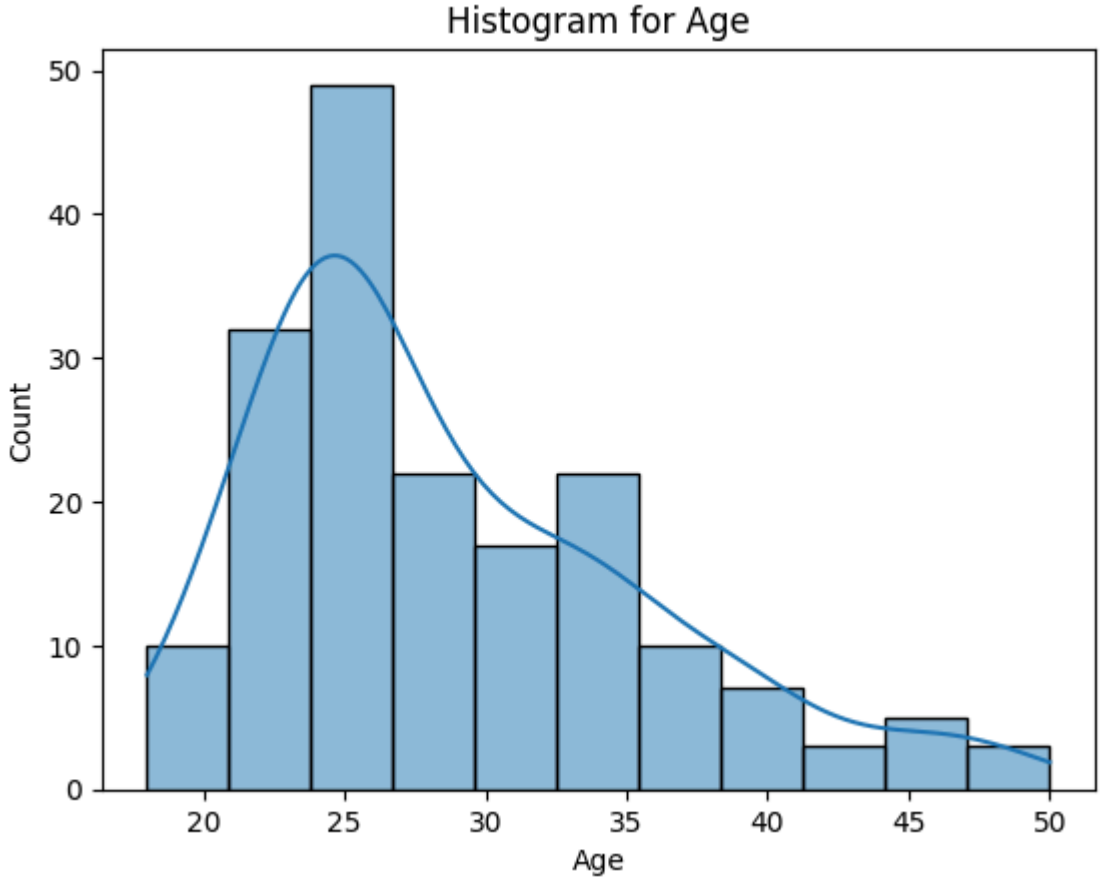

Boxplot for Income
30000
40000
50000
60000
70000
80000
90000
100000
Income

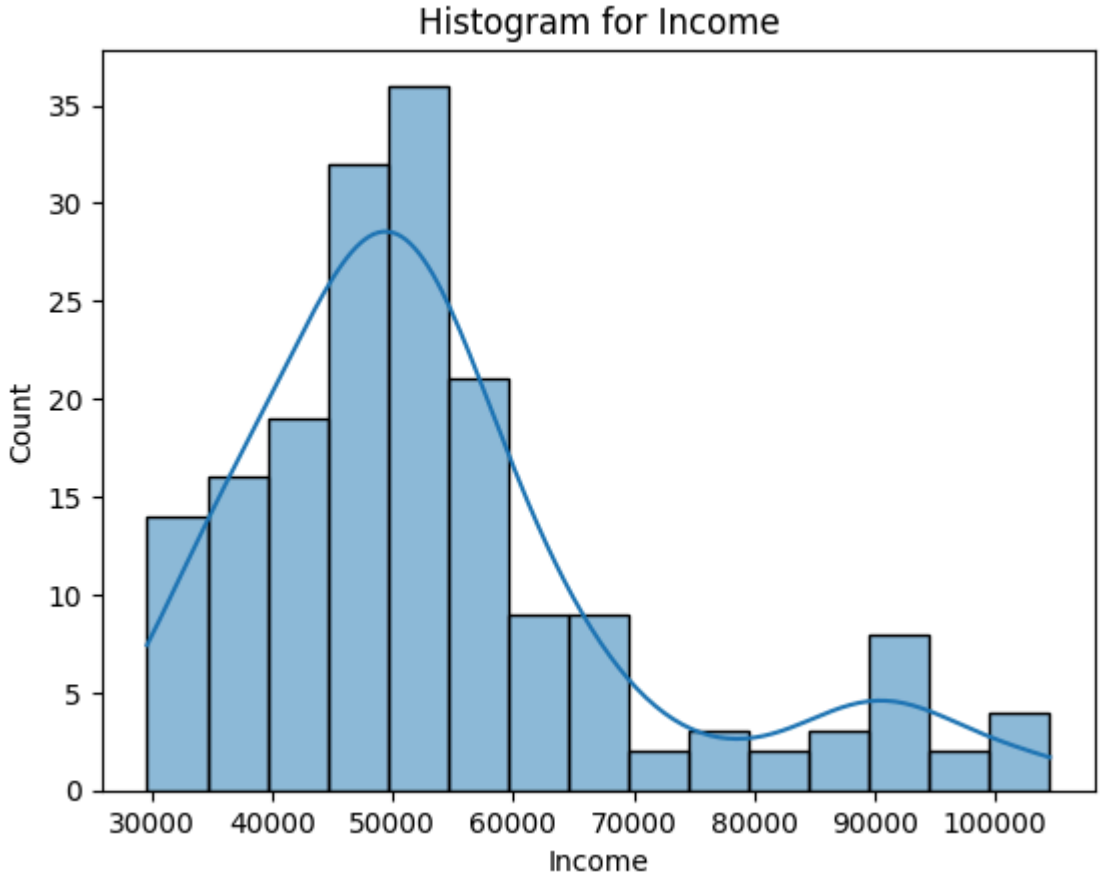

Boxplot for Usage
2
3
4
5
6
7
Usage

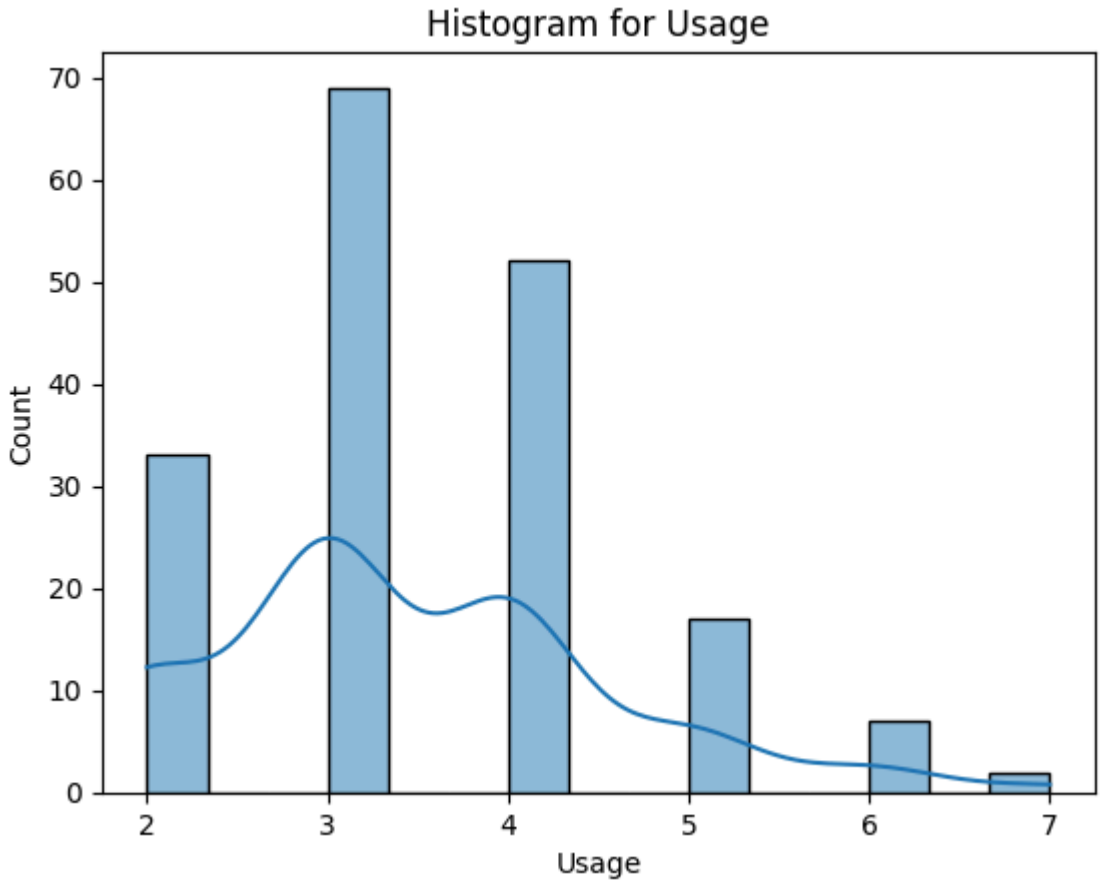

Boxplot for Fitness
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Fitness

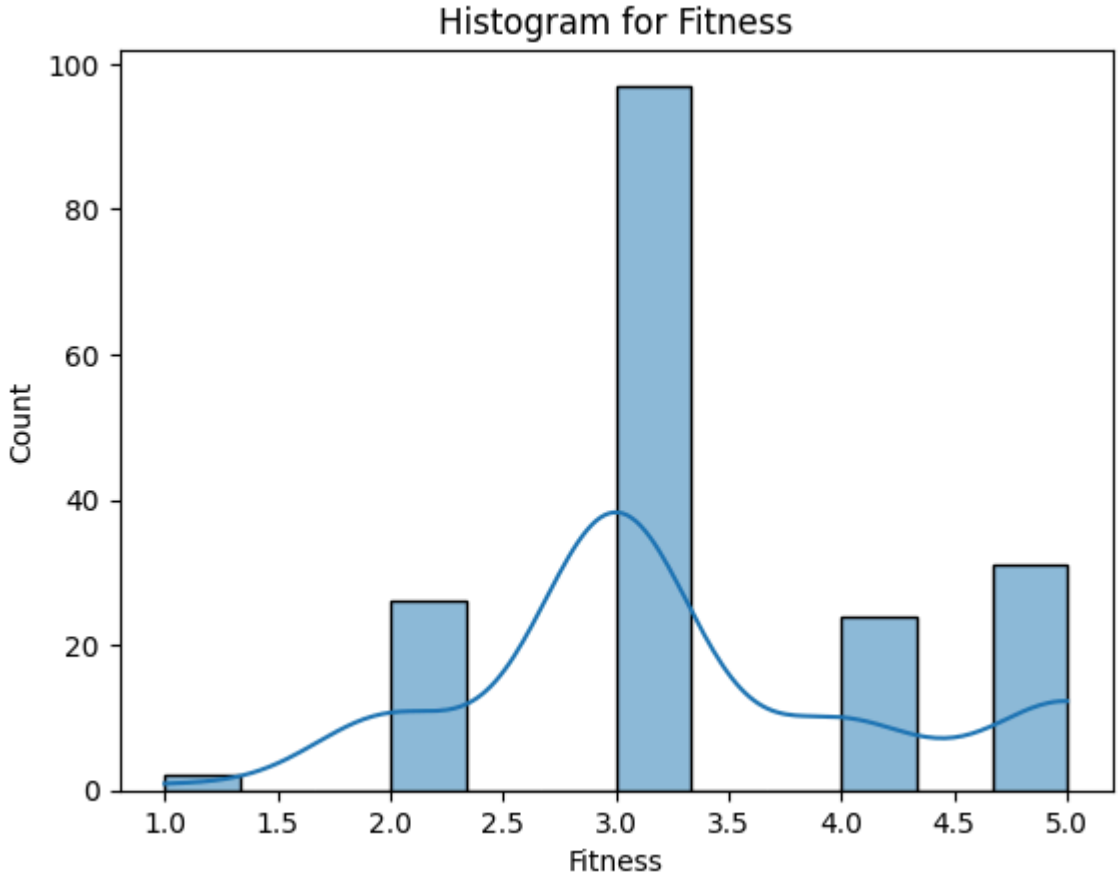

Boxplot for Miles
50
100
150
200
250
300
350
Miles

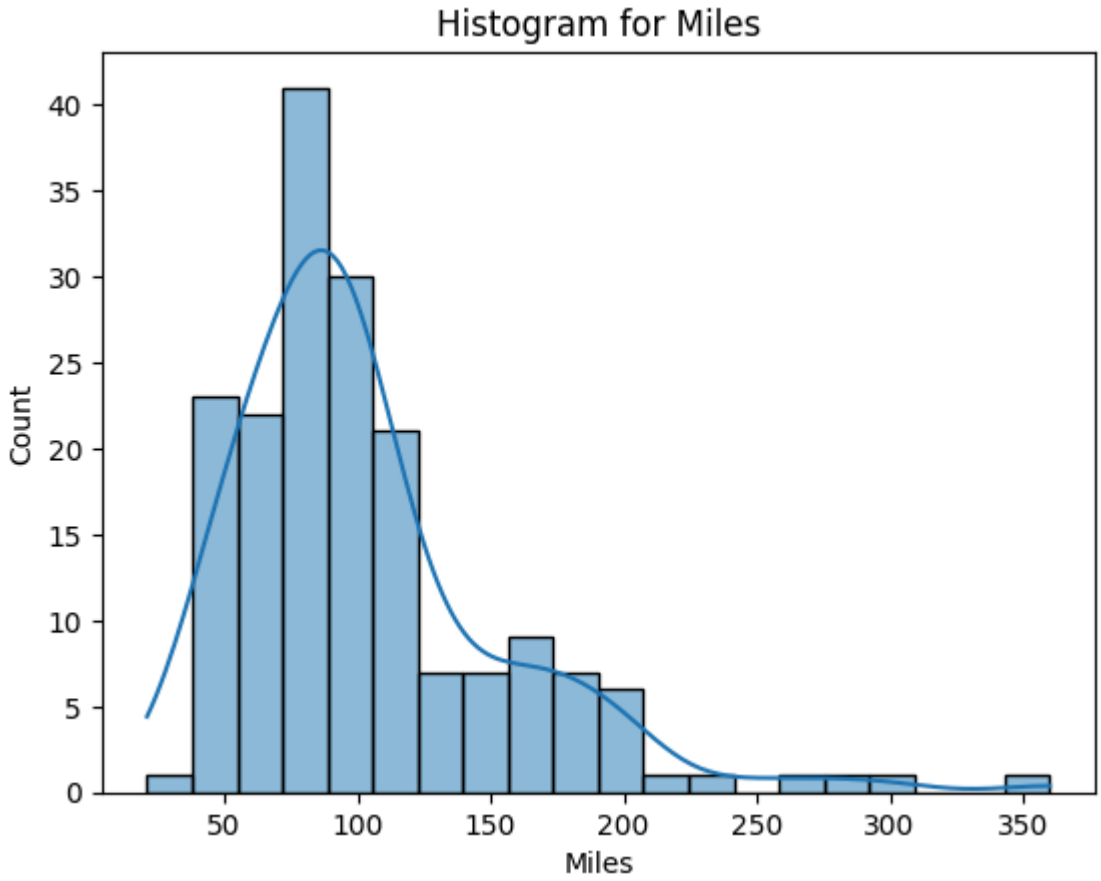
	Age	Income	Usage	Fitness	Miles
count	180.000000	180.000000	180.000000	180.000000	180.000000
mean	28.788889	53719.577778	3.455556	3.311111	103.194444
std	6.943498	16506.684226	1.084797	0.958869	51.863605
min	18.000000	29562.000000	2.000000	1.000000	21.000000
25%	24.000000	44058.750000	3.000000	3.000000	66.000000
50%	26.000000	50596.500000	3.000000	3.000000	94.000000
75%	33.000000	58668.000000	4.000000	4.000000	114.750000
max	50.000000	104581.000000	7.000000	5.000000	360.000000

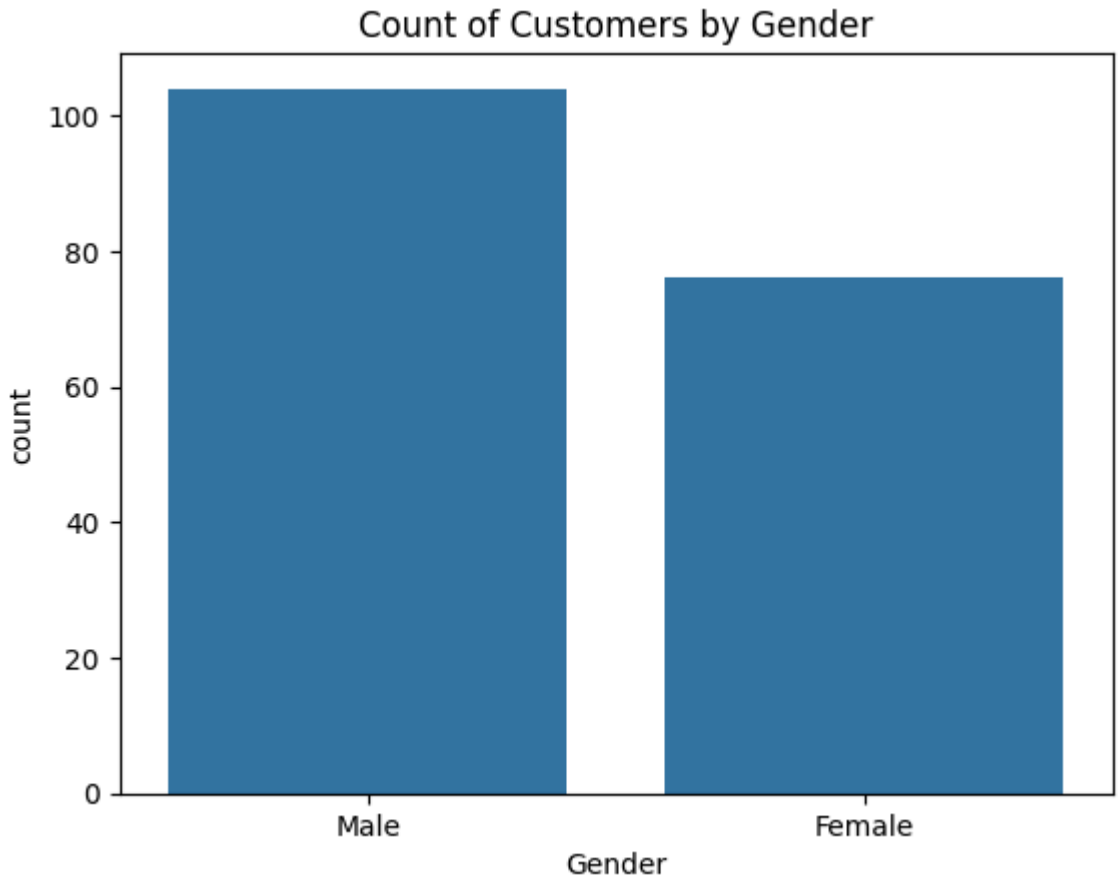


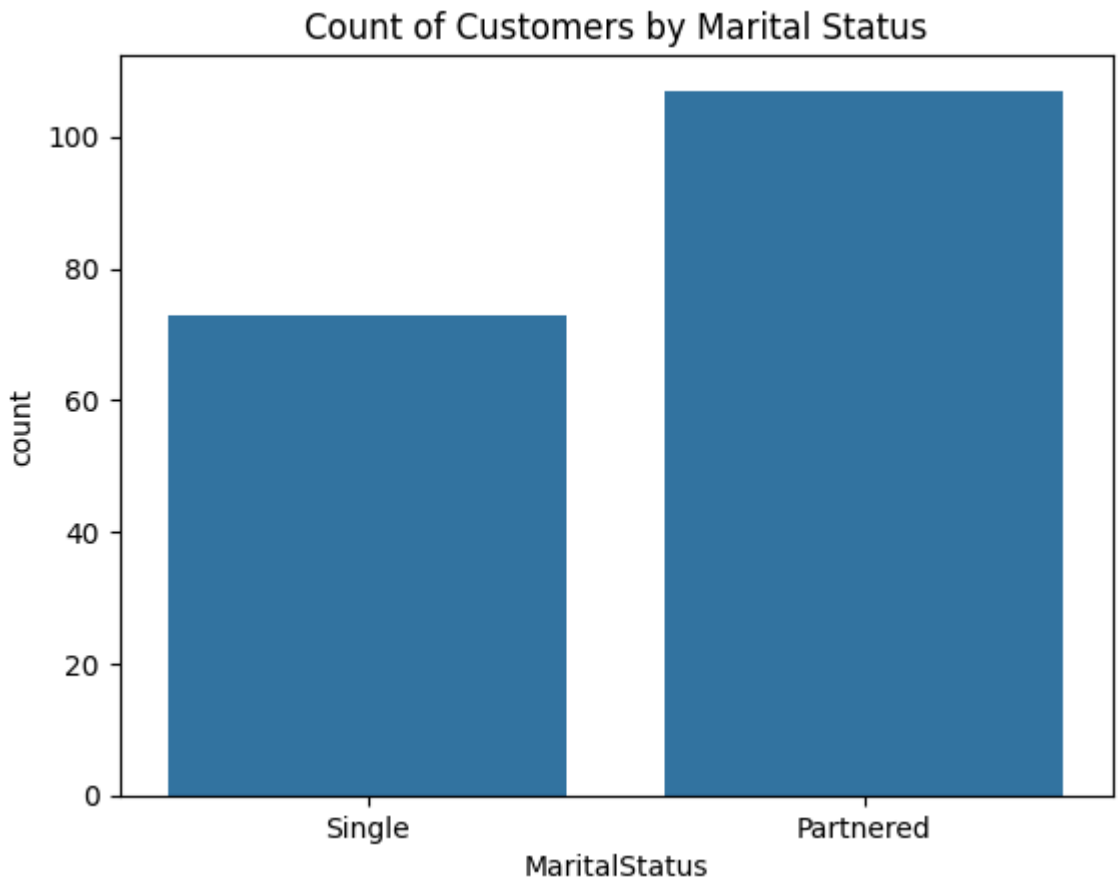

Histogram for Income
Count
Income
0
5
10
15
20
25
30
35
30000
40000
50000
60000
70000
80000
90000
100000

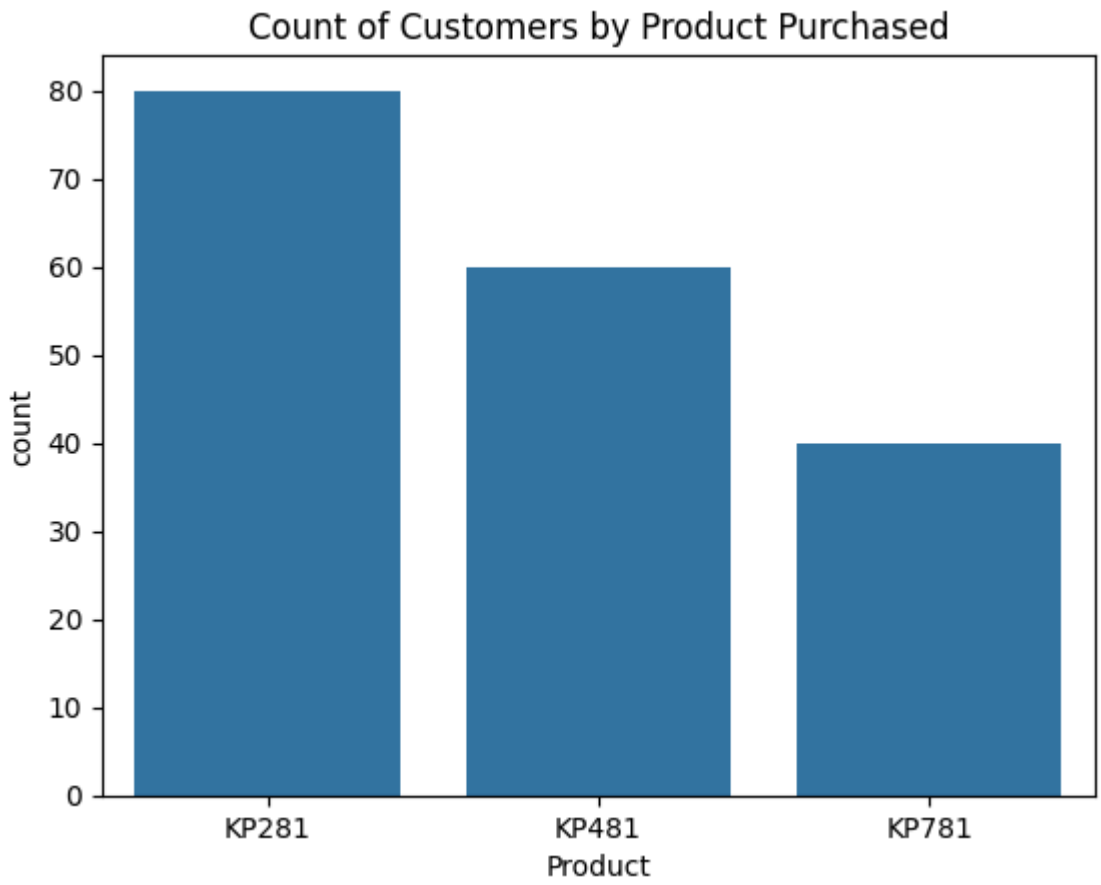

Histogram for Usage
Count
Usage
0
10
20
30
40
50
60
70
2
3
4
5
6
7

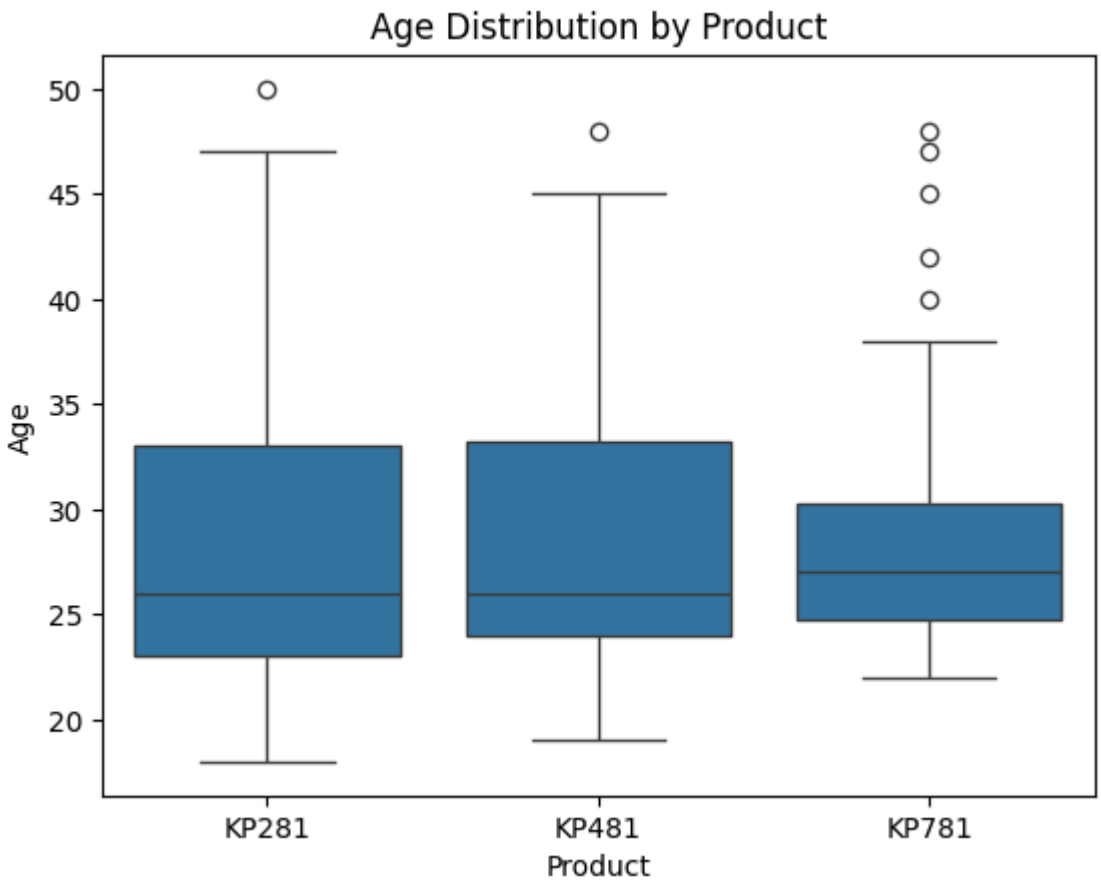

Histogram for Fitness
Count
Fitness
0
20
40
60
80
100
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0

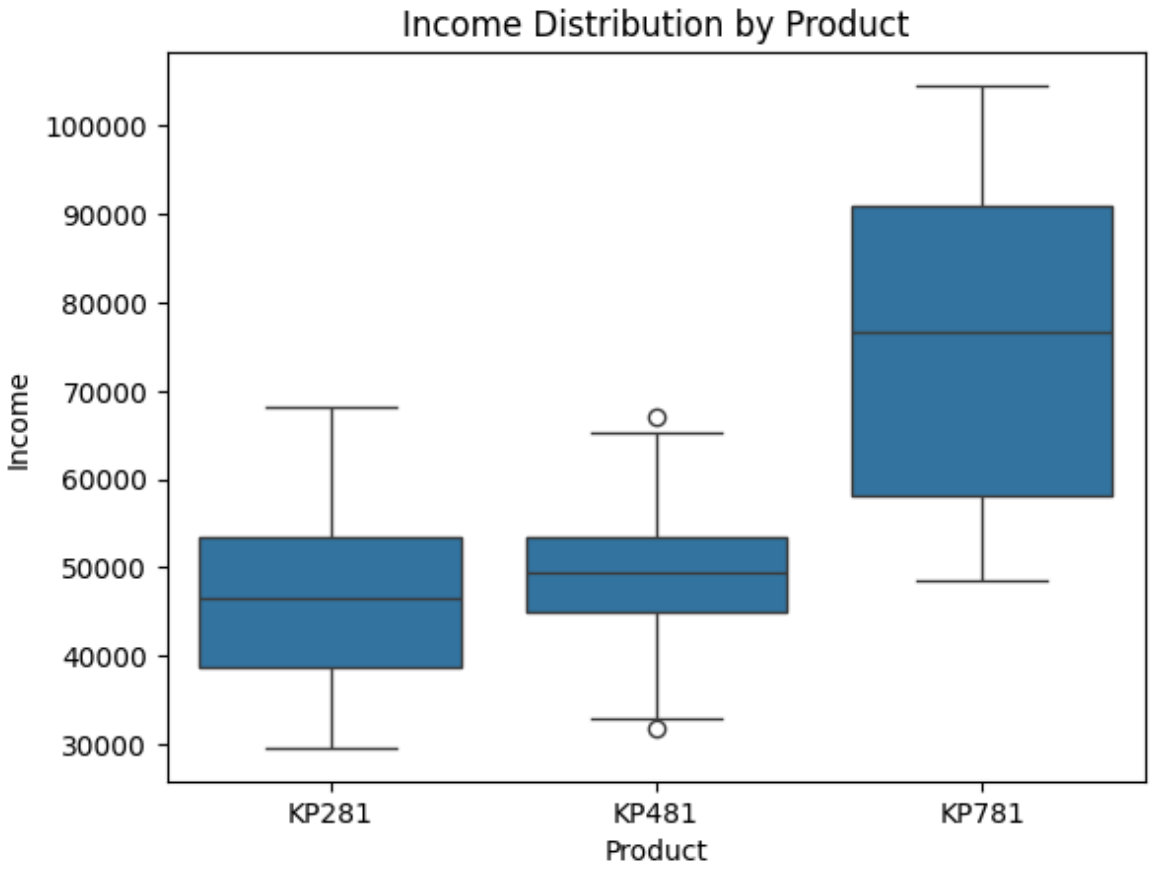

Histogram for Miles
Count
Miles
0
5
10
15
20
25
30
35
40
50
100
150
200
250
300
350

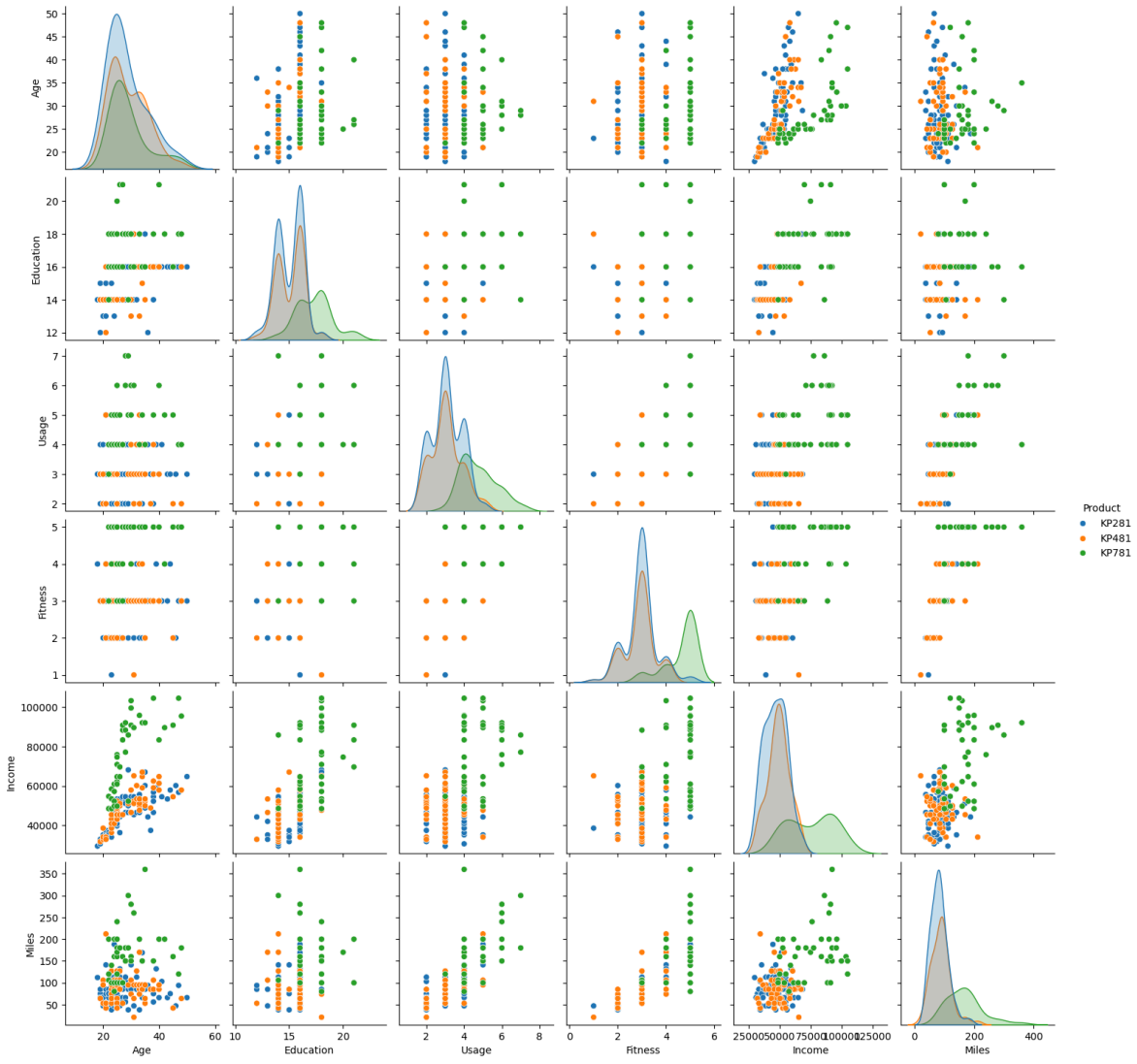

Count of Customers by Gender
count
100
80
60
40
20
0
Male
Female
Gender


Count of Customers by Marital Status
count
100
80
60
40
20
0
Single
Partnered
MaritalStatus


Count of Customers by Product Purchased
count
0
10
20
30
40
50
60
70
80
KP281
KP481
KP781
Product


Age Distribution by Product
Age
20
25
30
35
40
45
50
KP281
KP481
KP781
Product


Income Distribution by Product
Income
100000
90000
80000
70000
60000
50000
40000
30000
KP281
KP481
KP781
Product


Age
Education
Usage
Fitness
Income
Miles
Product
KP281
KP481
KP781


Correlation Heatmap
Age
Education
Usage
Fitness
Income
Miles
1
0.28
0.015
0.061
0.51
0.037
0.28
1
0.4
0.41
0.63
0.31
0.015
0.4
1
0.67
0.52
0.76
0.061
0.41
0.67
1
0.54
0.79
0.51
0.63
0.52
0.54
1
0.54
0.037
0.31
0.76
0.79
0.54
1
Age
Education
Usage
Fitness
Income
Miles
1.0
0.8
0.6
0.4
0.2

```
Marginal Probabilities:
Product
KP281    0.444444
KP481    0.333333
KP781    0.222222
Name: proportion, dtype: float64
Conditional Probability by Gender:
Product     KP281     KP481     KP781
Gender
Female   0.526316  0.381579  0.092105
Male     0.384615  0.298077  0.317308
Conditional Probability by Marital Status:
Product           KP281     KP481     KP781
MaritalStatus
Partnered      0.448598  0.336449  0.214953
Single         0.438356  0.328767  0.232877
Customer Profiles (Continuous Variables):
           Age  Education     Usage  Fitness     Income       Miles
Product
KP281    28.55  15.037500  3.087500   2.9625  46418.025   82.787500
KP481    28.90  15.116667  3.066667   2.9000  48973.650   87.933333
KP781    29.10  17.325000  4.775000   4.6250  75441.575  166.900000
Customer Profiles (Categorical Variables):
                 Gender MaritalStatus
Product
KP281    [Female, Male]     Partnered
KP481              Male     Partnered
KP781              Male     Partnered

--- Actionable Insights ---
1. KP281 is purchased by customers with lower income, younger age, and low
er fitness levels.
2. KP481 caters to mid-income customers with moderate fitness and usage.
3. KP781 is preferred by high-income customers with excellent fitness and
higher usage expectations.

Recommendations:
1. Target younger demographics with affordable KP281 through social media
ads.
2. Use fitness-focused campaigns for KP781 targeting high-income groups in
gyms or premium fitness clubs.
3. Highlight balanced features of KP481 for middle-income, fitness-conscio
us individuals.
4. Offer KP281 with affordable accessories for beginners (e.g., fitness ba
nds).
5. Include premium memberships or coaching sessions with KP781.
6. Place KP281 in regions with a younger population and lower average inco
mes.
7. Promote KP781 in urban areas with higher-income groups.
8. Introduce discounts during seasonal promotions for KP281 and KP481.
9. Encourage KP281 customers to upgrade to KP481 with loyalty discounts.
```

In [ ]: