Business Case: Aerofit_treadmill

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 the new customers

Product Portfolio:

KP281 is an entry-level treadmill that sells for $1,500. KP481 is for mid-level runners that sell for $1,750. KP781 treadmill is having advanced features that sell for $2,500.

In [1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```
df = pd.read_csv('D:\\aerofit_treadmill.csv')
```

Defining Problem Statement and Analysing basic metrics

In [3]:
```
#Observations on shape of data
df.shape
```

Out[3]:
```
(180, 9)
```

In [4]:
```
#data types of all the attributes
df.info()
```

Out[4]:
```
<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
```

Conversion of categorical attributes to 'category'

In [5]:
```
df = df.astype({'Product':'category','Gender':'category','MaritalStatus':'category'})
```

In [6]:
```
#statistical summary
df.info()
```

Out[6]:
```
<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    category
 1   Age            180 non-null    int64   
 2   Gender         180 non-null    category
 3   Education      180 non-null    int64   
 4   MaritalStatus  180 non-null    category
 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: category(3), int64(6)
memory usage: 9.5 KB
```

Non-Graphical Analysis: Value counts and unique attributes

In [7]:
```
df.value_counts()
```

Out[7]:
```
Product  Age  Gender  Education  MaritalStatus  Usage  Fitness  Income  Miles
KP281    18   Female  14         Single         3      4        29562   112    1
KP481    19   Male    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        51195   169    1
         35   Female  16         Partnered      3      3        68261   94     1
              Male    18         Partnered      4      5        95508   188    1
KP781    40   Male    18         Partnered      4      5        95508   188    1
Length: 180, dtype: int64
```

In [8]:
```
df.nunique()
```

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

Visual Analysis

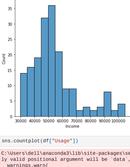
In [9]:
```
df.head(20)
```

Out[9]:

	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
5	KP281	20	Female	14	Partnered	3	3	32973	66
6	KP281	21	Female	14	Partnered	3	3	35247	75
7	KP281	21	Male	13	Single	3	3	32973	85
8	KP281	21	Male	15	Single	5	4	35247	141
9	KP281	21	Female	15	Partnered	2	3	37521	85
10	KP281	22	Male	14	Single	3	3	36384	85
11	KP281	22	Female	14	Partnered	3	2	35247	66
12	KP281	22	Female	16	Single	4	3	36384	75
13	KP281	22	Female	14	Single	3	3	35247	75
14	KP281	23	Male	16	Partnered	3	1	38658	47
15	KP281	23	Male	16	Partnered	3	3	40932	75
16	KP281	23	Female	14	Single	2	3	34110	103
17	KP281	23	Male	16	Partnered	4	3	39795	94
18	KP281	23	Female	16	Single	4	3	39658	113
19	KP281	23	Female	15	Partnered	2	2	34110	38

In [10]:
```
sns.displot(df["Income"])
```

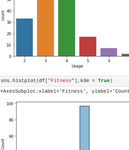
Out[10]:
```
<seaborn.axisgrid.FacetGrid at 0x2650fab9b00>
```



In [11]:
```
sns.countplot(df["Usage"])
```

```
C:\Users\dell\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[11]:
```
<AxesSubplot:xlabel='Usage', ylabel='count'>
```



In [12]:
```
sns.histplot(df["Fitness"],kde = True)
```

Out[12]:
```
<AxesSubplot:xlabel='Fitness', ylabel='Count'>
```

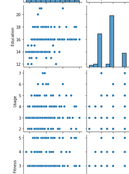
In [13]:
```
sns.boxplot(df["Age"])
```

```
C:\Users\dell\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[13]:
```
<AxesSubplot:xlabel='Age'>
```

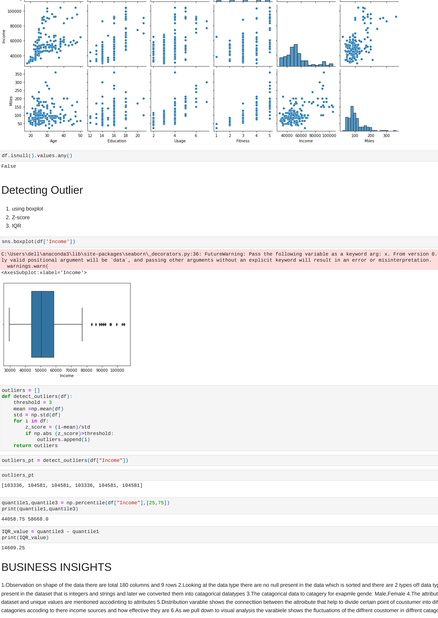
In [14]:
```
sns.heatmap(df.corr(),annot = True)
```

Out[14]:
```
<AxesSubplot:>
```



In [15]:
```
sns.pairplot(data=df)
```

Out[15]:
```
<seaborn.axisgrid.PairGrid at 0x265a0857610>
```



In [16]:
```
df.isnull().values.any()
```

Out[16]:
```
False
```

Detecting Outlier

1. using boxplot
2. Z-score
3. IQR

In [18]:
```
sns.boxplot(df['Income'])
```

```
C:\Users\dell\anaconda3\lib\site-packages\seaborn\_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.
  warnings.warn(
```

Out[18]:
```
<AxesSubplot:xlabel='Income'>
```


In [20]:
```
outliers = []
def detect_outliers(df):
    threshold = 3
    mean = np.mean(df)
    std = np.std(df)
    for i in df:
        z_score = (i-mean)/std
        if np.abs (z_score)>threshold:
            outliers.append(i)
    return outliers
```

In [23]:
```
outliers_pt = detect_outliers(df["Income"])
```

In [24]:
```
outliers_pt
```

Out[24]:
```
[103336, 104581, 104581, 103336, 104581, 104581]
```

In [34]:
```
quantile1,quantile3 = np.percentile(df["Income"],[25,75])
print(quantile1,quantile3)
```

```
44058.75 58668.0
```

In [37]:
```
IQR_value = quantile3 - quantile1
print(IQR_value)
```

```
14609.25
```

BUSINESS INSIGHTS

1.Observation on shape of the data there are total 180 columns and 9 rows 2.Looking at the data type there are no null present in the data which is sorted and there are 2 types off data type present in the dataset that is integers and strings and later we converted them into categorical datatypes 3.The categorical data to categorise for example gender: Male,Female 4.The attribute of dataset and unique values are mentioned accoderting to attributes 5.Distribution variable shows the connection between the attributute that help to divide certain point of coustumer into different catagories according to there income sources and how effective they are 6.As we pull down to visual analysis the variabele shows the fluctuations of the diffrent coustumer in diffrent catagories

Recommendations

As per the insight the mid-end (i.e) kp781 is most purchased by the married catagory at some level the basic variant of the tredmill is used by the unmarried. The mid variant came be cope up with the more features in that price that the might attract both kinds of coustomers and most useage also and major helath maintaing catagory are rise not invest much more in the treadmill and launching one more variant with more feature like a breed of low-end and mid-end which may attrat more coustomers

In [ ]:
```

```