Business Case
Aerofit - Descriptive Statistics & Probability

Dataset Overview

Colab link to check :
https://colab.research.google.com/drive/1b1firw19RijgNCpBwABwC-xAafFK6Slb?usp=sharing

Shape:

The dataset contains 180 rows and 9 columns.

Data Types:

The dataset comprises a mix of object (string) and integer data types:
Product: object
Age: int64
Gender: object
Education: int64
MaritalStatus: object
Usage: int64
Fitness: int64
Income: int64
Miles: int64

Missing Values: There are no missing values in any of the columns.

Duplicates: There are no duplicate rows in the dataset.

Summary Statistics:
Here are the main points that can be inferred from the dataset:

Products: The dataset includes information on multiple products identified by their codes, such as KP281, KP481, and KP781.

Age: The age of customers ranges from young adults to older adults. Detailed statistics (e.g., mean, median) can provide more insights.

Gender: The dataset captures gender information with values such as Male and Female.

Education: The dataset records the education level as an integer, likely representing the number of years of education or a categorical level.

Marital Status: The marital status is captured with values such as Single and Partnered.

Usage: Usage is recorded as an integer, potentially representing the frequency of product usage a week.

Fitness: Fitness is measured on a numeric scale, indicating customers' fitness level.

Income: The income of customers is recorded in numerical format, providing insights into the financial background of the customers.

Miles: Miles might represent the distance customers are willing to travel, perhaps related to the usage of the product.

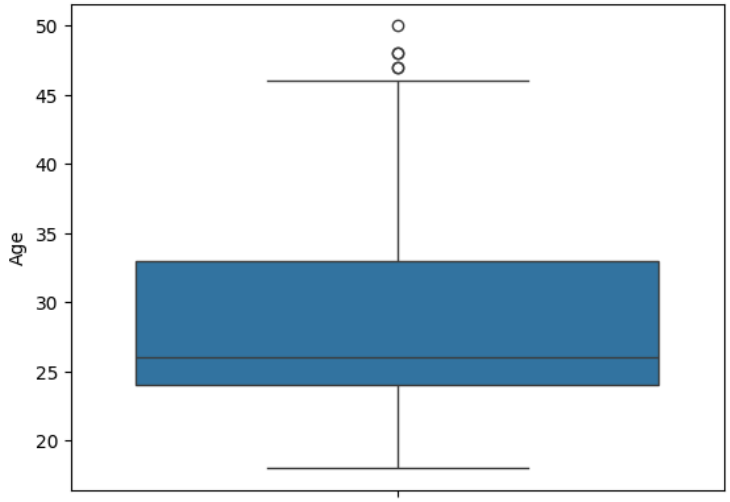
Outcome:

Box Plot of Age:

Code:

```
#Making box plot of Age column to get the outliers in Customer Age
sns.boxplot(df["Age"])
```

Output:



Insight: The box plot highlights any outliers in the `Age` column.

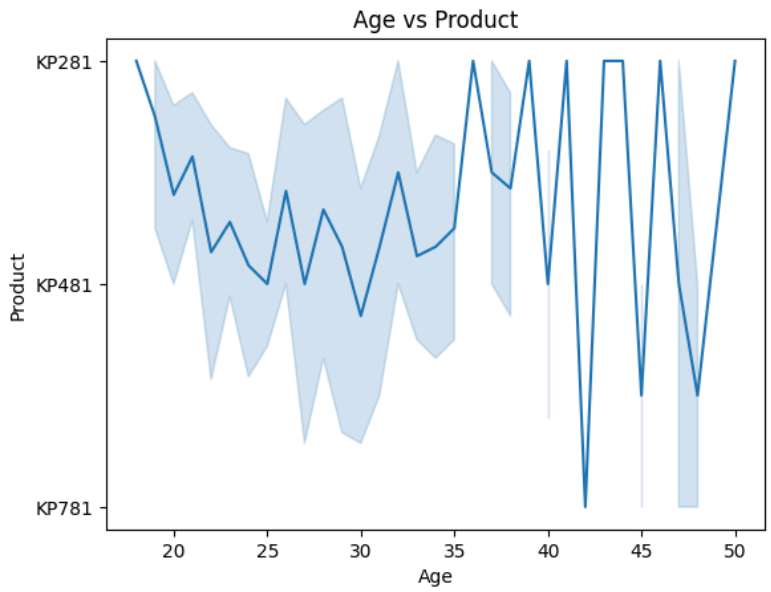
Recommendation:

Age vs. Product:

Code:

```
# visualizing the people which age group people buy which products
sns.lineplot(x = 'Age', y = 'Product', data = df)
plt.title("Age vs Product")
plt.show()
```

Output:

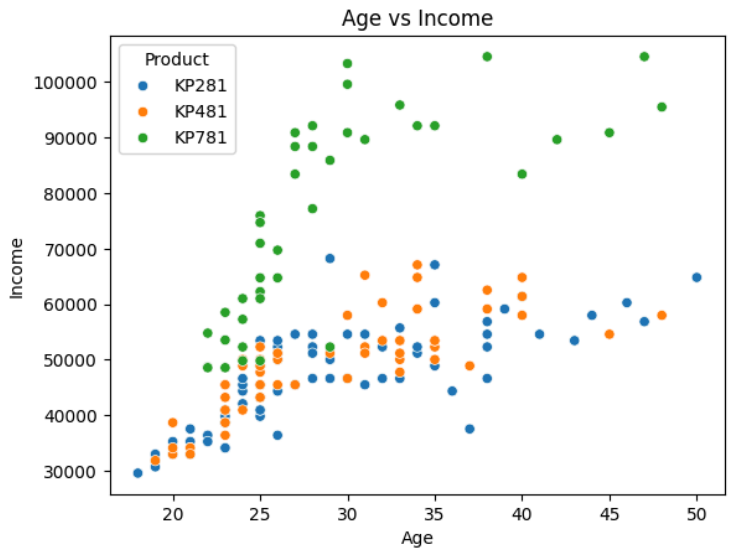


Insight: Line plot showing the relationship between customer age and the products they purchase.
Recommendation: Use this insight to target specific age groups with marketing campaigns for certain products.

Age vs. Income by Product:
Code:

```
# Checking the difference in customer's age and income with varities of products
sns.scatterplot(x = 'Age', y = 'Income',hue='Product', data = df)
plt.title("Age vs Income")
plt.show()
```

Output:

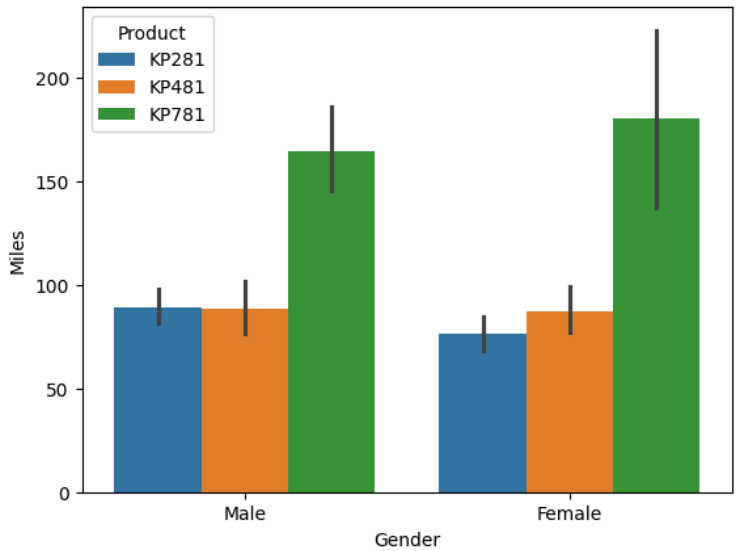


Insight: Scatter plot showing the distribution of income across different ages and products.
Recommendation: Identify high-income customer segments and tailor product features or advertising to appeal to these groups.

Miles by Gender and Product:
Code:

```
# Checking how Product purchase varies depending on Gender
sns.barplot(x = 'Gender',y = 'Miles',hue='Product',data=df)
plt.show()
```

Output:



Insight: Bar plot showing how the miles traveled varies by gender and product.
Recommendation: Consider gender-specific marketing strategies based on how far customers are willing to travel for products.

Probability:
Code:

```
# Probability of choosing a Married person and he/she purchase a kp481 treadmill
df.loc[df['Product']=='KP481'].shape[0]/df.loc[df['MaritalStatus'] == 'Partnered'].shape[0]
```

Output: 0.5607476635514018
Insight: Partnered individuals might prioritize certain aspects of the treadmill that align with their lifestyle. This could include shared workout routines, features that support couples' fitness goals, or attributes that enhance the home workout experience.
Recommendation: Develop marketing campaigns that target partnered or married couples. Highlight the benefits of the KP481 treadmill for joint workouts, family fitness goals, and its ability to cater to different fitness levels.

Code:

```
# Percentage of marginal probability for the Product based on Gender of customers
pd.crosstab(df['Product'],df['Gender'],margins=True,normalize=True) * 100
```

Output:

Gender	Female	Male	All
Product			
KP281	22.2222	22.222222	44.444444
KP481	16.11111	17.222222	33.333333
KP781	3.88888	18.333333	22.222222
All	42.2222	57.777778	100.000000

Insight:
42.22% of the customers are female, and 57.78% are male. This indicates a higher proportion of male customers in the dataset.
KP281 is evenly distributed among male and female customers, each constituting 22.22% of the total customer base. This means KP281 has a balanced appeal across genders.

KP481 is slightly more popular among males (17.22%) compared to females (16.11%). However, the difference is not very significant, suggesting a fairly balanced gender appeal

Recommendation:

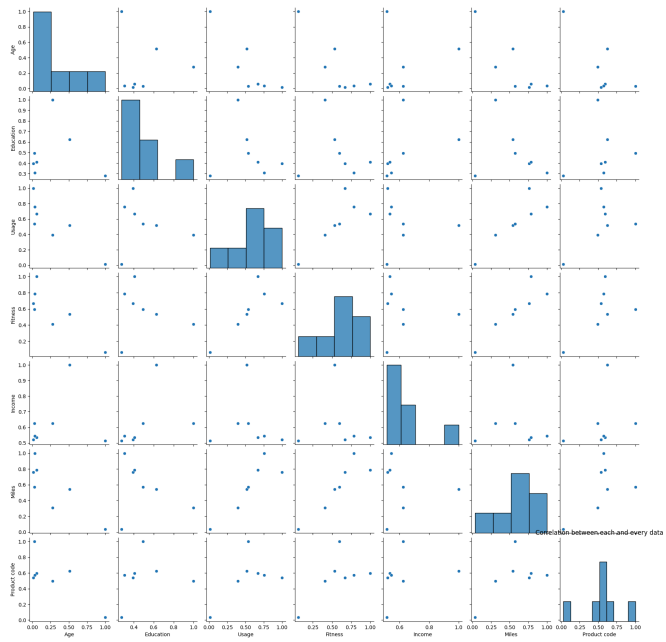
Since KP781 is significantly more popular among male customers, marketing campaigns for this product should be tailored to appeal to male customers. Highlight features that resonate with male fitness goals and preferences. Implement promotions and discounts that cater to the preferences of each gender. For example, offer bundled deals with accessories that are popular among male customers for KP781 and similar deals for female customers for other products.

Pairplot

Code:

```
# Pairplot for correlation matrix
sns.pairplot(corr_matrix)
plt.title("Correlation between each and every data")
plt.show()
```

Output:



Insight: Pairplots can provide a more detailed view of the relationships between features.

Recommendation: Explore pair plots to identify trends and patterns that are not immediately obvious from the correlation matrix.

Probability of Male purchasing a device

Probability of a male purchasing KP781
Code:

```
# Probability of a male purchasing kp781 treadmill is
df_kp781 = df.loc[df['Product'] == 'KP781']
len(df_kp781[df_kp781['Gender'] == 'Male'])/len(df_kp781)
```

Output: 0.825
Insight: 82.5% of KP781 treadmill purchasers are male.
Recommendation: Focus marketing efforts for the KP781 treadmill on male customers. Consider male-oriented advertising channels and messaging.

Probability of a male purchasing KP481
Code:

```
# Probability of a male purchasing kp481 treadmill is
df_kp481 = df.loc[df['Product'] == 'KP481']
len(df_kp481[df_kp481['Gender'] == 'Male'])/len(df_kp481)
```

Output: 0.5166666666666667
Insight: 51.67% of KP481 treadmill purchasers are male.
Recommendation: This product has a more balanced gender distribution. Tailor marketing efforts to appeal to both male and female customers equally.

Probability of a male purchasing KP281
Code:

```
# Probability of a male purchasing kp281 treadmill is
df_kp281 = df.loc[df['Product'] == 'KP281']
len(df_kp281[df_kp281['Gender'] == 'Male'])/len(df_kp281)
```

Output: 0.5
Insight: 50% of KP281 treadmill purchasers are male.
Recommendation: Like the KP481, the KP281 has a balanced gender distribution. Ensure that advertising campaigns for this product are gender-neutral to attract a diverse customer base.

Probability of selecting a person with Age = 34
Code:

```
# Probability of selecting a person with Age = 34
df['Age'].value_counts(normalize=True)[34]
```

Output: 0.03333333333333333

Insight: 3.33% of the customers are aged 34.

Recommendation: Although this age group is small, ensure that marketing strategies are inclusive of all age groups to avoid missing potential customers.

Probability of purchasing KP781 given ‘Marital Status’ is ‘Single’

Code:

```
# Probability of purchasing KP781 given 'Marital Status' is 'Single'
df[df['MaritalStatus'] ==
'Single']['Product'].value_counts(normalize=True)['KP781']
```

Output:

```
Product
KP281    0.438356
KP481    0.328767
KP781    0.232877
Name: proportion, dtype: float64
```

Insight: Given that a customer is single, the probabilities of purchasing each product are:

KP281: 43.84%

KP481: 32.88%

KP781: 23.29%

Recommendation: Single customers are more likely to purchase KP281. Focus marketing campaigns for KP281 on single customers.

Probability of purchasing KP781 given ‘Usage’ is 5 times a week

Code:

```
# Probability of purchasing KP781 given 'Usage' is 5 times a week
df[df['Usage'] == 5]['Product'].value_counts(normalize=True)['KP781']
```

Output: 0.7058823529411765

Insight: 70.59% of customers who use the treadmill 5 times a week, purchase the KP781.

Recommendation: Highlight the durability and features of the KP781 that appeal to frequent users in marketing materials. This could include endorsements from fitness enthusiasts or highlighting the treadmill's performance in high-usage scenarios.