EDA Portfolio Project - Treadmill Buyer Profile

Project Details:

The market research team at 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 new customers. The team decides to investigate whether there are differences across the product with respect to customer characteristics.

Product Portfolio:

- The KP281 is an entry-level treadmill that sells for $1,500;
- The KP481 is for mid-level runners and sells for $1,750;
- The KP781 treadmill have advanced features, and it sells for $2,500.

Data Description:

The company collected data on individuals who purchased a treadmill from the AeroFit stores during the prior three months. The dataset in aerofit_treadmill_data.csv has the following features:

- Product - product purchased: KP281, KP481, or KP781
- Age - in years
- Gender - male/female
- Education - in years
- MaritalStatus - single or partnered
- Usage - the average number of times the customer plans to use the treadmill each week
- Fitness - self-rated fitness on a 1-5 scale, where 1 is the poor shape and 5 is the excellent shape
- Income - annual income in US dollars
- Miles - the average number of miles the customer expects to walk/run each week

Practicalities:

Analyse the provided data and provide insights to the best of your abilities. Include the relevant tables/graphs/visualization to explain what you have learned about the data.

You may structure your EDA/Business Analysis according to these steps:

1. Data Exploration and Processing:
    - Importing data
    - Reading dataframe
    - Check the shape of the dataframe
    - Datatype of each column
    - Missing value detection
    - Checking duplicate values in the dataset

2. Statistical Summary:
    - Provide an analysis of the statistical summary in few lines for both categorical and numerical features.

3. Non-Graphical Analysis:
    - Value Counts for all categorical features
    - Unique Attributes for all categorical features

4. Graphical Analysis:
    - Univariate Analysis - Numerical features:
        - Distribution Plot
        - Count Plot
        - Box Plot pending
    - Univariate Analysis - Categorical features:
        - Count Plot
    - Bivariate Analysis:
        - Check features effect on the product purchased e.g.
            - Product vs Gender
            - Product vs MaritalStatus
            - Product vs Age
    - Multivariate Analysis:
        - Create pairplots to show relationship of features

5. Correlation Analysis:
    - Show the correlation matrix on heatmap and write your observation of findings in few lines.

6. Outlier Detection:
    - Check for the outliers by using the IQR method.

7. Conditional Probabilities:
    - What percent of customers have purchased KP281, KP481, or KP781?
    - Create frequency tables and calculate the percentage as follows
        - Product – Gender
            - Percentage of a Male customer purchasing a treadmill
            - # Percentage of a Female customer purchasing KP781 treadmill
            - Probability of a customer being a Female given that Product is KP281
        - Product – Age
            - Percentage of customers with Age between 20s and 30s among all customers
        - Product – Income
            - Percentage of a low-income customer purchasing a treadmill
            - Percentage of a high-income customer purchasing KP781 treadmill
            - Percentage of customer with high-income salary buying treadmill given that Product is KP781
        - Product – Fitness
            - Percentage of customers that have fitness level 5
            - Percentage of a customer with Fitness Level 5 purchasing KP781 treadmill
            - Percentage of customer with fitness level 5 buying KP781 treadmill
        - Product - Marital Status
            - Percentage of a customers who are partnered using treadmills

8. Actionable Insights & Recommendations:

    Provide detailed report on the actionable insights and recommendations according to your observations.

Important: Make sure that the solution reflects your entire thought process including the preparation of data - it is more important how the code is structured rather than just the final result or plot. Good Luck