

## Google\_Colab\_Link:

[https://colab.research.google.com/drive/1aLeI0ec6z2xx9dsD7xC7RvxLwo\\_HMf5W?usp=sharing](https://colab.research.google.com/drive/1aLeI0ec6z2xx9dsD7xC7RvxLwo_HMf5W?usp=sharing)

### 1. Defining Problem Statement and Analyzing basic metrics (10 Points)

#### Problem Statement:

- The Aerofit company would like to get the suggestions and recommendations of a particular type of Treadmill model based on the customer demography and requirements.

#### 1. Observations on shape of data, data types of all the attributes, conversion of categorical attributes to 'category' (If required), statistical summary

- The data consists of 180 rows and 9 columns.
- The product, Gender and Marital Status columns are of Object datatype and remaining columns are integer datatype.
- Product column corresponds to the model of the Treadmill.
- Age column refers to the age of the customer.
- Gender column corresponds to the gender of the customer.
- Education column refers to the educational qualification of the customer.
- MaritalStatus column refers to the marital status of the customer.
- Usage column corresponds to the no.of days a customer is using the treadmill in a week.
- Fitness column refers to the shape of the customer where 5 – good fitness, 1-poor fitness.
- Miles column refers to average no.of miles a customer wishes to walk/run on the treadmill.

### 2. Non-Graphical Analysis: Value counts and unique attributes (10 Points)

- The Fitness rating of 3 corresponds to almost 53 % , which means that almost half of customer who comes to buy a treadmill have a moderate fitness, and can be recommended **KP481** or **KP781**.
- The Fitness rating of 1 corresponds to only 1% , which means that only few no.of customers' fitness is very poor, who come to by a treadmill and they can be considered as beginners and can be recommended **KP281** model.

- Among the Partnered and single status customers , the fitness trend is almost same with high no.of customers with moderate fitness and only one customer with poor fitness.
- There are more no.of females with poor fitness of (1 and 2) than males.
- Overall no.of males with moderate and excellent fitness are more.

### 3. Visual Analysis - Univariate & Bivariate (30 Points)

1. For continuous variable(s): Distplot, countplot, histogram for univariate analysis (10 Points)

#### Univariate Analysis:

- The sales of KP281 are more corresponding to other two models.
- There are more no.of customers with age around 25 years compared to other age intervals.
- There are more no.of customers in with education 16 followed by 14, 18.
- More no.of male customers are there compared to female customers.
- The customers who are using the treadmill 3 to 4 times a week are more.
- The customers with moderate fitness of 3 are more in number.

2. For categorical variable(s): Boxplot (10 Points)

#### Bivariate Analysis:

- KP281: Equal no.of males and females have bought this model and the overall count of sold products of this model is high.
- KP481: For this model, the male customer count is slightly higher than the female customer count.
- KP781: For this model, the male customer count is very high compared to female count. The sold product count of this model is less compared to other models.
- For Education with 16, the sold count of models KP281 and KP481 are high and for model KP781, the education with 18 has max sold count.
- For Fitness with 3,4,5, more no.of male customers are there compared to female customers.
- Among male and female customers, Married customers count is more than the single count.

### 3. For correlation: Heatmaps, Pairplots(10 Points)

- From the pair plot and heat map, it is clear that the customers walking more miles per week are more fit and the miles and fitness columns are highly positively correlated.
- Income and Education are also showing the positive correlation.
- There are more no.of customers in the income range less than 65000 \$ than those above 65000 \$.

### 4. Missing Value & Outlier Detection (10 Points)

#### Missing Values:

- There are no missing values in the dataset that is given, which can be known by observing the information of the dataset and from using the isna function.

#### Outliers:

- The median age of **males** for all the models is almost same and there are only few outliers for KP781 for males. The spread of boxplot of KP281 and KP481 is almost same and the spread of box plot for KP781 is small.
- For **Females**, the median age of KP481 is slightly higher than the other models, there are few outliers for KP281. The spread of boxplot of KP281 is more compared to the spread of box plot for KP781 and KP481.

#### Probability of Customer buying a model:

- From the crosstab, its clear that most of the customers with salary range 30K-70K are more likely to purchase KP281 model.
- The customers with the salary range above 70K are clearly opting for KP781 model and the probability of these customers buying other models is evidently zero.

### 5. Business Insights based on Non-Graphical and Visual Analysis (10 Points)

#### 1. Comments on the range of attributes

- The Age column has the values from 18 to 50 years, with mean age of 28.7 years.
- The Education column has the values from 12 to 21 , with mean of 15.5.

- The Usage column has the values from 2 to 7, with mean 3.4.
- The Fitness column has values from 1(least fit , based on the shape of the body) to 5(most fit), with mean 3.3.
- The Income column has the values from 29000 \$ to 105000 \$ , with mean 53000 \$.
- The Miles has the values 21 to 360 which corresponds to the no.of miles the customer would like to complete on treadmill, with mean 103.1.

## 2. Comments on the distribution of the variables and relationship between them

- Miles and Fitness has highest positive correlation.
- Usage and Age has lowest positive correlation.
- No columns are exhibiting negative correlation.

## 3. Comments for each univariate and bivariate plot

- KP281 is the highly sold model among all the models.
- The customers with age around 25 years are more in number.
- The customers with salary more than 7000 \$ and fitness 5 are willing to buy only KP781 model.
- Customers with education 14,16 are more likely to buy KP281/KP481, where as customers with education 18 are more likely to by KP781.

## 6. Recommendations (10 Points) - Actionable items for business. No technical jargon. No complications. Simple action items that everyone can understand

### Customers with income range of less than 50000 \$:

- These customers are more likely to buy a Treadmill and hence they need more attention. Also, they are more probable of buying KP281 model. Hence, they can be recommended this model.
- If any additional offers or campaigns are rolled for these customers, the chances of selling this model will be higher.

### Customers with income range of 50000 \$ - 70000 \$:

- These customers are moderately likely to buy a Treadmill.
- Also, they are equally probable of buying KP281/KP481 model. Hence, they can be recommended these models.

- If any additional offers or campaigns are rolled for these customers, the chances of selling these models will be high.

**Customers with income range of greater than 70000 \$:**

- These customers with in this range are only opting for KP781 model.
- Hence these customers can be recommended this model.

**Fitness:**

- The customers with fitness 1,2,3 is most likely to buy KP281 model, moderately likely to buy KP481, hence they can be recommended these two models.
- The customers with fitness 5 are most likely to buy KP781 model and hence, they can be recommended this model.