# Aerofit Business case study

- ❖ **Topic:** EDA
- ❖ **Duration:** 1 week

---

## Why this case study?

**From the company's perspective:**

- Aerofit is a leading brand in the field of fitness equipment. Aerofit provides a product range including machines such as treadmills, exercise bikes, gym equipment, and fitness accessories to cater to the needs of all categories of people.
- 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.
- Perform descriptive analytics **to create a customer profile** for each AeroFit treadmill product by developing appropriate tables and charts.
- Construct two-way contingency tables for each AeroFit treadmill product and compute all **conditional and marginal probabilities** and their **insights/impact** on the business.

**From the learner's perspective:**

- Solving this business case holds immense importance for aspiring data analysts and scientists.
- Through the process of engaging with this case study, individuals acquire practical knowledge and develop skills in Data Analysis, which are essential for deriving meaningful insights from data.
- Engaging in the resolution of this specific case study will aid in cultivating expertise in identifying connections between input variables and output variables.

---

**Dataset link:** [aerofit_data.csv]

- **Product:** Product Purchased KP281, KP481, or KP781
- **Age:** In years
- **Gender:** Male/Female
- **Education:** in years
- **MaritalStatus:** single or partnered
- **Usage:** average number of times the customer plans to use the treadmill each week
- **Income:** annual income (in $)
- **Fitness:** self-rated fitness on a 1-to-5 scale, where 1 is poor shape and 5 is the excellent shape.
- **Miles:** average number of miles the customer expects to walk/run each week

**Product Portfolio:**

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

## What is expected?

As a data analyst/scientist at Aerofit, I have been tasked with the responsibility of analyzing the provided dataset to extract valuable insights and deliver actionable recommendations.

**Submission Process:**

- Type your insights and recommendations in the text editor.
- Convert your jupyter notebook into PDF (Save as PDF using Chrome browser's Print command), upload it in your Google Drive (set the permission to allow public access), and paste that link in the text editor.
- Optionally, you may add images/graphs in the text editor by taking screenshots or saving matplotlib graphs using plt.savefig(...).
- After submitting, you will not be allowed to edit your submission.

### General Guidelines:

- Evaluation will be kept lenient, so make sure you attempt this case study. ● It is understandable that you might struggle with getting started on this or feel stuck at some point.

  In such cases:

a. Read the question carefully and try to understand what exactly is being asked.
b. Brainstorm a little. If you're getting an error, remember that Google is your best friend.
c. You can watch the lecture recordings or go through your lecture notes once again if you feel like you're getting confused over some specific topics.
d. Discuss your problems with your peers. Make use of the Slack channel and WhatsApp group.
e. Only if you think that there's a major issue, you can reach out to your Instructor via Slack or Email.

---

## What does 'good' look like?

1. **Import the dataset and do usual data analysis steps like checking the structure & characteristics of the dataset**

    ○ The data type of all columns in the "customers" table.

    **Hint:** We want you to display the data type of each column present in the dataset.

    ○ You can find the number of rows and columns given in the dataset

    **Hint**: We want you to find the shape of the dataset.

    ○ Check for the missing values and find the number of missing values in each column

---

2. **Detect Outliers**

    ○ Find the outliers for every continuous variable in the dataset

    **Hint:** We want you to use boxplots to find the outliers in the given dataset

    ○ Remove/clip the data between the 5 percentile and 95 percentile

    **Hint:** We want You to use np.clip() for clipping the data

---

3. **Check if features like marital status, Gender, and age have any effect on the product purchased**

- Find if there is any relationship between the categorical variables and the output variable in the data.
  **Hint**: We want you to use the count plot to find the relationship between categorical variables and output variables.

- Find if there is any relationship between the continuous variables and the output variable in the data.

  **Hint**: We want you to use a scatter plot to find the relationship between continuous variables and output variables.

---

**4. Representing the Probability**

- Find the marginal probability (what percent of customers have purchased KP281, KP481, or KP781)

  **Hint:** We want you to use the pandas crosstab to find the marginal probability of each product.

- Find the probability that the customer buys a product based on each column.

  **Hint:** Based on previous crosstab values you find the probability.

- Find the conditional probability that an event occurs given that another event has occurred. (Example: given that a customer is female, what is the probability she'll purchase a KP481)

  **Hint:** Based on previous crosstab values you find the probability.

---

**5. Check the correlation among different factors**

- Find the correlation between the given features in the table.

  **Hint:** We want you can use the heatmap and corr function to find the correlation between the variables

---

**6. Customer profiling and recommendation**

- Make customer profilings for each and every product.
  **Hint**: We want you to find at What age, gender, and income group but product the KP281

- Write a detailed recommendation from the analysis that you have done.

---

–