

---

## Problem Statement:

You have been given the role of a business analyst for an E-Commerce company and have been asked to prepare a basic report on the data. Here, we are analysing different aspects of E-Commerce Company by doing Exploratory Data Analysis.

**Domain:** Sales Analytics

## Tasks/ Questions to be Answered:

1. To get familiar with the data (9 marks)
  - a. Print out the first 10 and the last 10 records of the data. (2 marks)
  - b. How many rows and columns are present in the dataset? Use any two different methods to extract this information. (2 + 2 + 1 marks)
  - c. How many object data types are there? (1 mark)
  - d. Is there any Boolean data type? (1 mark)
2. Eliminating the non-informative columns. (4 marks)
  - a. Drop the columns product\_specifications and description. (2 marks)
  - b. Which method or function is used to permanently delete the columns mentioned in part (b)? Write the code explicitly (2 marks)
3. Here we summarize the data at brand level. (3 Marks)
  - a. How many unique Brands are there? (1 mark)
  - b. Note that each brand contains multiple products. Show the average rating of the products within each Brand (2 marks)
4. Next, we study the main categories of the products. (9 Marks)
  - a. Create an appropriate plot to show the count of items ordered for each product\_main\_category. (5 Marks)
  - b. From the plot identify for which two product\_main\_category(s) the maximum and the minimum orders were placed (2 Marks)
  - c. Write code to print out the top 5 product\_main\_category(s) in descending order? (2 Marks)
5. Find the net revenue generated by the E-Commerce company over all orders placed. (6 Marks)
6. Calculate the BrandRevenue for each brand and list the top 10 brands having maximum revenue in descending order (5 Marks)
  - a. Calculate BrandRevenue for each brand (2 Marks)
  - b. List the top 10 brands having maximum revenue in descending order (3 Marks)

7. Compare prices for each product (5 Marks)
  - a. Draw boxplots of retail\_price & discount\_price. (2 Marks)
  - b. Are there any outliers? (Yes/No) (1 Mark)
  - c. Create a scatterplot retail\_price (x-axis) & discounted\_price (y-axis) (2 Marks)
8. Create a new dataframe to include the brand specific information as stated (3 Marks)
  - a. Total number of orders placed per brand
  - b. Total retail\_price per brand
  - c. Total discount\_price per brand
  - d. Total BrandRevenue generated per brand.
  - e. Pairplot using these four features.
9. Compare performance regionwise (6 Marks)
  - a. Draw a lineplot for the monthly Revenue of ECom Company for each region separately. (4 Marks)
  - b. Identify the best and the worst performing months for each region. (2 Marks)