# **Sales Data Analysis:**

ABC Private Limited, a retail company, wants to gain insight into its customers' purchasing habits, particularly their spending on different product categories. To do this, they have provided a summary of the purchase history of a number of high-volume products from the previous month. This data includes information on the customer demographics, such as age, gender, marital status, city type, and length of stay in their current city, as well as details on the products themselves, including product ID and product category.

# **Question 1: (60 Marks)**

**DATA PREPROCESSING**

* Check descriptive statistics of the dataset
* Check for missing values in the data
* check for unique values in the data
* Perform EDA
* Purchase Distribution with respect to other variables in the dataset.
* check for outliers
* Analysis by Gender, Marital Status, occupation, occupation vs purchase, purchase by city, purchase by age group, etc
* Drop unnecessary fields
* Convert categorical data into integer using map function (e.g 'Gender' column)
* missing value treatment
* Rename columns
* fill nan values
* map range variables into integers (e.g 'Age' column)

**Data Visualization**

* visualize an individual column
* Age vs Purchased
* Occupation vs Purchased
* Product\_category\_1 vs Purchased
* Product\_category\_2 vs Purchased
* Product\_category\_3 vs Purchased
* City category pie chart
* check for more possible plots

**Perform all of the above analysis in python and provide clear interpretation along with the visuals wherever necessary. The codes should be commented, along with concise headings.**

# **Question:2 (40 Marks)**

The company is looking to build a model that can predict how much a customer is likely to spend on a given product, which will enable them to tailor personalized offers to each customer for different products.