

## Course 4: Feature Selection & Dimensionality Reduction-Project Problem Statement

### About the Dataset:

A retail company “ABC Private Limited” wants to understand the customer purchase behaviour (specifically, purchase amount) against various products of different categories. They have shared purchase summaries of various customers for selected high volume products from last month.

The data set also contains customer demographics (age, gender, marital status, city\_type, stay\_in\_current\_city), product details (product\_id and product category) and Total purchase\_amount from last month.

### Data

| Variable                   | Definition  |
|----------------------------|---|
| User_ID                    | User ID   |
| Product_ID                 | Product ID  |
| Gender                     | Sex of User   |
| Age                        | Age in bins   |
| Occupation                 | Occupation (Masked)                                 |
| City_Category              | Category of the City (A,B,C)                        |
| Stay_In_Current_City_Years | Number of years stay in current city                |
| Marital_Status             | Marital Status                                      |
| Product_Category_1         | Product Category (Masked)                           |
| Product_Category_2         | Product may belongs to other category also (Masked) |
| Product_Category_3         | Product may belongs to other category also (Masked) |
| Purchase                   | Purchase Amount (Target Variable)                   |

Answer the following questions:

1. Check the correlation matrix and identify most correlated features.
2. Find out the Covariance Matrix? Also Identify eigen values and eigen vectors of covariance matrix.
3. Find out the two most important Principal Components.
4. Select only 8 features using PCA.