Case Study 1: - Customer Analysis for Retail

Data available: -

1. *Customer*: Customers demographic information
2. *Transaction*: Transactions of customers
3. *Product Hierarchy*: Product Information (Category, sub-category etc)

Problems: -

1. Merge the available datasets and name the consolidated dataset – 'Customer_Final'. Ensure to keep all the customers who have done transactions and select the join type accordingly.

2. Prepare a summary report for the merged dataset:
    a. Get the column names and their corresponding datatypes
    b. Top and Bottom 20 observations
    c. Total no. of rows in the data
    d. Total no. of missing in all the columns
    e. "Five-Number summary" for continuous variables (Min, Max, Median, Q1, Q3)
    f. Frequency tables for all the categorical variables

3. Calculate the following information using the merged dataset: -
    a. Time period of the available transaction data
    b. Count of transactions where the total amount of transaction was negative

4. Analyze which product categories are more popular among male vs female customers

5. Which city code has the maximum customers and what is the percentage of customers from that city?

6. Which store type sells the maximum products by value and by quantity?

7. What is the total amount earned from the “Electronics” and “Clothing” categories from Flagship stores?

8. What is the total amount earned from “Male” customers under the “Electronics” category?

9. How many customers have more than 10 unique transactions, after removing all transactions which have any negative amounts?

10. For all customers aged between 25-35, find out the following:
    a. What is the total amount spent for “Electronics” and “Books” product categories?
    b. What is the total amount spent by these customers between 1st January, 2014 to 1st March, 2014?