Report - task_1 Name- Mukul Parwani

Task 1: Data Cleaning and Preprocessing

Dataset Used: Customer Personality Analysis

Summary of Changes Performed:

1. File Loading Fix:
   - Loaded CSV correctly using a comma separator (pd.read_csv()), as the default loading created one-column issue.
2. Column Name Cleanup:
   - Removed leading/trailing spaces in column names.
   - Converted all column names to lowercase with underscores for consistency.
3. Handled Missing Values:
   - Detected 24 missing values in the income column.
   - Dropped those rows to ensure clean analysis.
4. Removed Duplicates:
   - Checked and removed any duplicate rows using drop_duplicates().
5. Standardized Categorical Text:
   - Converted entries in education and marital_status columns to lowercase and stripped whitespaces for uniformity.
6. Converted Date Format:
   - Transformed dt_customer column to proper datetime type for easier date analysis.
7. Ensured Correct Data Types:
   - Cast numerical columns like income, kidhome, and teenhome to appropriate data types (float, int).
8. Saved Cleaned Dataset:
   - Exported the final cleaned dataset as cleaned_customer_personality.csv.