Task 1: Data Cleaning and Preprocessing

📌 Objective

To clean and prepare the raw **Customer Personality Analysis** dataset by handling missing values, duplicates, inconsistent formats, and incorrect data types using Python (Pandas) in Google Colab.

🛠 Tools Used

- Python

- Pandas

- Google Colab

🗁 Dataset

- Name: Customer Personality Analysis

- Source: [Kaggle](https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis)

- File Used: `marketing_campaign.csv`

🔍 Data Cleaning Steps

1. Handled Missing Values

   - Removed rows with null values to maintain data integrity.

2. Removed Duplicates

   - Dropped duplicate rows to avoid redundancy and inconsistencies.

3. Standardized Text Columns

   - Converted values in columns like `Marital_Status` and `Education` to lowercase and removed extra spaces for consistency.

4. Formatted Dates

   - Converted the `Dt_Customer` column to `datetime` format (dd-mm-yyyy).

5. Renamed Columns

   - Cleaned column names by converting them to lowercase and replacing spaces with underscores.

6. Corrected Data Types

   - Ensured numerical fields like `income`, `kidhome`, and `teenhome` have the appropriate data types.

✅ Output

- Cleaned Dataset: `cleaned_marketing_data.csv`

- Final Row Count: (after removing nulls and duplicates)_

🧑‍💻 How to Run

1. Upload the original CSV in Google Colab.

2. Run the provided Python script.

3. Cleaned dataset will be downloadable as `cleaned_marketing_data.csv`.

📑 Author

- Name: Pallavi Chandapure

- Project: B.Tech (CSE) Data Preprocessing Task