

🧼 Data Cleaning & EDA: Customer Personality Analysis

This project is part of a Data Analyst Internship task focusing on cleaning and preprocessing a real-world dataset from Kaggle – the **Customer Personality Analysis** dataset. It includes demographic and marketing data of a retail company's customers.

✅ Summary of Work Done

📊 Dataset Overview

- **Source**: Kaggle (Customer Personality Analysis)

- **Records**: ~2,200 customers

- **Attributes**: Demographics, product spendings, campaign responses, web/catalog/store behavior

🛠️ Cleaning & Preprocessing Steps Performed

Step	Description
1. Loaded dataset	Imported the raw <code>.csv</code> file with <code>sep='\t'</code>
2. Renamed columns	Changed all column names to lowercase, underscore-separated, readable names
3. Handled missing values	Filled missing values in <code>income</code> column with the median
4. Removed duplicates	Dropped all duplicate rows using <code>drop_duplicates()</code>
5. Handled infinite values	Replaced <code>inf</code> and <code>-inf</code> with <code>NaN</code> then treated
6. Standardized categorical values	Cleaned and normalized text like <code>marital_status</code>
7. Converted date columns	Parsed <code>enrollment_date</code> to proper <code>datetime</code> object
8. Added derived columns	Created <code>age</code> , <code>total_spent</code> , <code>kids_total</code> , and <code>enrollment_year</code>
9. Fixed indexing issues	Prevented extra <code>Unnamed</code> column by using <code>index=False</code> in <code>to_csv()</code>
10. Verified data types	Ensured correct typing for numeric, string, and datetime fields

📈 Basic EDA Highlights

- Majority of customers are between 30-60 years of age.
- Income varies widely, but high income doesn't always correlate with high spending.
- Single and married individuals dominate the customer base.
- Highest customer acquisition occurred between 2012-2015.

📁 Repository Contents

| File | Description |

| `marketing_campaign_jaya.ipynb` | Complete code: data cleaning + EDA |

| Raw dataset file

| `README.md` | This file – contains summary and details of the project |

🚀 How to Run

Install required packages:

```
```bash
```

```
pip install pandas matplotlib seaborn
```

```
```
```

Open the notebook:

```
```bash
```

```
jupyter notebook marketing_campaign_jaya.ipynb
```

```
```
```

📩 For queries or contributions, reach out via GitHub or LinkedIn.

<https://www.linkedin.com/in/jaya-bijore-ab2114289/>

<https://github.com/jayabijore20>