```
import pandas as pd
import numpy as np


from google.colab import files
uploaded = files.upload()
```

```
Choose Files marketing_campaign.csv
• marketing_campaign.csv(text/csv) - 220188 bytes, last modified: 28/5/2025 - 100% done
Saving marketing_campaign.csv to marketing_campaign (1).csv
```

Load the dataset

```
import io
df = pd.read_csv(io.BytesIO(uploaded['marketing_campaign (1).csv']), sep="\t")


df.head()
```

	ID	Year_Birth	Education	Marital_Status	Income	Kidhome	Teenhome	Dt_Customer	Recency	MntWines	...	NumWebVisitsMonth	A
0	5524	1957	Graduation	Single	58138.0	0	0	04-09-2012	58	635	...	7	
1	2174	1954	Graduation	Single	46344.0	1	1	08-03-2014	38	11	...	5	
2	4141	1965	Graduation	Together	71613.0	0	0	21-08-2013	26	426	...	4	
3	6182	1984	Graduation	Together	26646.0	1	0	10-02-2014	26	11	...	6	
4	5324	1981	PhD	Married	58293.0	1	0	19-01-2014	94	173	...	5	

5 rows × 29 columns

```
print("Original Columns:\n", df.columns.tolist())
```

```
Original Columns:
 ['ID', 'Year_Birth', 'Education', 'Marital_Status', 'Income', 'Kidhome', 'Teenhome', 'Dt_Customer', 'Recency', 'MntWines', 'MntFrui
```

Clean column names (lowercase, remove spaces)

```
df.columns = df.columns.str.strip().str.lower().str.replace(" ", "_")
```

Check for missing values

```
print("\nMissing Values:\n", df.isnull().sum())
```

```
Missing Values:
 id                     0
year_birth              0
education               0
marital_status          0
income                 24
kidhome                 0
teenhome                0
dt_customer             0
recency                 0
mntwines                0
mntfruits               0
mntmeatproducts         0
mntfishproducts         0
mntsweetproducts        0
mntgoldprods            0
numdealspurchases       0
numwebpurchases         0
numcatalogpurchases     0
numstorepurchases       0
numwebvisitsmonth       0
acceptedcmp3            0
acceptedcmp4            0
acceptedcmp5            0
acceptedcmp1            0
acceptedcmp2            0
complain                0
z_costcontact           0
```

```
z_revenue               0
response                0
dtype: int64
```

Fill missing values in 'income' with median

```
if 'income' in df.columns:
    df['income'] = df['income'].fillna(df['income'].median())
```

Drop duplicate rows

```
df.drop_duplicates(inplace=True)
```

Convert 'dt_customer' to datetime format

```
if 'dt_customer' in df.columns:
    df['dt_customer'] = pd.to_datetime(df['dt_customer'], errors='coerce')
```

Create 'age' column from 'year_birth'

```
if 'year_birth' in df.columns:
    df['age'] = 2025 - df['year_birth']
```

Standardize text columns

```
text_columns = df.select_dtypes(include='object').columns
for col in text_columns:
    df[col] = df[col].str.strip().str.lower()
```

Display cleaned dataset info

```
print("\nCleaned Dataset Info:\n")
df.info()
```

```
Cleaned Dataset Info:

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2240 entries, 0 to 2239
Data columns (total 30 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   id                   2240 non-null   int64
 1   year_birth           2240 non-null   int64
 2   education            2240 non-null   object
 3   marital_status       2240 non-null   object
 4   income               2240 non-null   float64
 5   kidhome              2240 non-null   int64
 6   teenhome             2240 non-null   int64
 7   dt_customer          916 non-null    datetime64[ns]
 8   recency              2240 non-null   int64
 9   mntwines             2240 non-null   int64
 10  mntfruits            2240 non-null   int64
 11  mntmeatproducts      2240 non-null   int64
 12  mntfishproducts      2240 non-null   int64
 13  mntsweetproducts     2240 non-null   int64
 14  mntgoldprods         2240 non-null   int64
 15  numdealspurchases    2240 non-null   int64
 16  numwebpurchases      2240 non-null   int64
 17  numcatalogpurchases  2240 non-null   int64
 18  numstorepurchases    2240 non-null   int64
 19  numwebvisitsmonth    2240 non-null   int64
 20  acceptedcmp3         2240 non-null   int64
 21  acceptedcmp4         2240 non-null   int64
 22  acceptedcmp5         2240 non-null   int64
 23  acceptedcmp1         2240 non-null   int64
 24  acceptedcmp2         2240 non-null   int64
 25  complain             2240 non-null   int64
 26  z_costcontact        2240 non-null   int64
 27  z_revenue            2240 non-null   int64
 28  response             2240 non-null   int64
 29  age                  2240 non-null   int64
dtypes: datetime64[ns](1), float64(1), int64(26), object(2)
memory usage: 525.1+ KB
```

Save cleaned dataset

```
df.to_csv("cleaned_marketing_campaign.csv", index=False)
print("\n✅ Cleaned dataset saved as 'cleaned_marketing_campaign.csv'")
```

```
✅ Cleaned dataset saved as 'cleaned_marketing_campaign.csv'
```

Download the cleaned CSV to your computer

```
files.download("cleaned_marketing_campaign.csv")
```