```
# Task 1 - Data Cleaning & Preprocessing
# Dataset: marketing_campaign.csv
# Author: Bishal Kumar Mishra
```

```
import pandas as pd
```

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

```
Choose files No file chosen Upload widget is only available when the cell has been executed in the current browser session. Please rerun this cell to enable.
Saving marketing_campaign.csv to marketing_campaign.csv
```

```
df = pd.read_csv("marketing_campaign.csv")
```

```
!ls
```

```
marketing_campaign.csv  sample_data
```

```
df = pd.read_csv("marketing_campaign.csv")
print("\n✅ File uploaded and loaded successfully!")
print("Initial shape:", df.shape)
```

```
✅ File uploaded and loaded successfully!
Initial shape: (2240, 1)
```

```
#Quick look at the data

print("\nMissing values per column:\n", df.isnull().sum())
print("Duplicate rows:", df.duplicated().sum())
```

```
Missing values per column:
 ID\tYear_Birth\tEducation\tMarital_Status\tIncome\tKidhome\tTeenhome\tDt_Customer\tRece
dtype: int64
Duplicate rows: 0
```

```
#Remove duplicates

df.drop_duplicates(inplace=True)
print("\nRemoved duplicate rows.")
```

```
Removed duplicate rows.
```

```
#Handle missing values
# Filing missing Income with median

if "Income" in df.columns:
    df["Income"].fillna(df["Income"].median(), inplace=True)
    print("Filled missing 'Income' values with median.")
```

```
# Fill missing categorical columns with mode
```

```
for col in df.select_dtypes(include=["object"]).columns:
    if df[col].isnull().sum() > 0:
        df[col].fillna(df[col].mode()[0], inplace=True)
```

```
#Clean text columns

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

```
 #Convert date column

if "Dt_Customer" in df.columns:
    df["Dt_Customer"] = pd.to_datetime(df["Dt_Customer"], errors="coerce", dayfirst=True
    print("Converted 'Dt_Customer' to datetime format.")
```

```
#Convert numeric columns

numeric_cols = [
    "Year_Birth", "Income", "Kidhome", "Teenhome", "Recency",
    "MntWines", "MntFruits", "MntMeatProducts", "MntFishProducts",
    "MntSweetProducts", "MntGoldProds", "NumDealsPurchases",
    "NumWebPurchases", "NumCatalogPurchases", "NumStorePurchases",
    "NumWebVisitsMonth", "Complain", "Response"
]
```

```
for col in numeric_cols:
    if col in df.columns:
        df[col] = pd.to_numeric(df[col], errors="coerce")
```

```
#Rename columns

df.columns = [c.strip().lower().replace(" ", "_") for c in df.columns]
```

```
#Final checks

print("\n Cleaning complete!")
print("Final shape:", df.shape)
print("Remaining missing values:", df.isnull().sum().sum())
```

```
 Cleaning complete!
Final shape: (2240, 1)
Remaining missing values: 0
```

```
 #Save results

df.to_csv("cleaned_marketing_campaign.csv", index=False)
```

```
summary = """
Task 1 - Data Cleaning and Preprocessing

Steps completed:
1. Removed duplicate rows
2. Filled missing numeric values (median)
3. Filled missing categorical values (mode)
4. Cleaned text columns
5. Converted date columns to datetime
```

```
6. Renamed columns for consistency
"""

with open("changes_summary.txt", "w") as f:
    f.write(summary)

print("\nFiles generated:")
print(" - cleaned_marketing_campaign.csv")
print(" - changes_summary.txt")
```

```
Files generated:
 - cleaned_marketing_campaign.csv
 - changes_summary.txt
```

```
#Download cleaned files

from google.colab import files
files.download("cleaned_marketing_campaign.csv")
files.download("changes_summary.txt")
```

```
```

```
print("\n All done! Files are ready for GitHub upload.")
```

```
 All done! Files are ready for GitHub upload.
```