**TASK 1.**

**1.Load the Data (Python)**

import pandas as pd

# Load the dataset

df = pd.read\_csv('marketing\_campaign.csv', sep='\t') # Assuming it's tab-separated

**2. Initial Inspection**

# Check the first few rows

print(df.head())

# Check data types and nulls

print(df.info())

# Check for duplicates

print("Number of duplicate rows:", df.duplicated().sum())

# Check unique values in categorical columns

for col in df.select\_dtypes(include='object').columns:

print(f"\nUnique values in {col}: {df[col].unique()}")

**3. Handle Missing Values**

# Check for nulls

print(df.isnull().sum())

# Handle nulls (example: fill 'Income' with median)

df['Income'].fillna(df['Income'].median(), inplace=True)

# Drop any row if there are nulls in other important columns (example, date)

df.dropna(subset=['Dt\_Customer'], inplace=True)

**4. Remove Duplicates**

# Remove duplicate rows

df.drop\_duplicates(inplace=True)

print("Number of duplicate rows after removal:", df.duplicated().sum())

**5. Standardize Text Values**

# Standardize 'Education'

df['Education'] = df['Education'].str.lower().str.capitalize()

# Standardize 'Marital\_Status'

df['Marital\_Status'] = df['Marital\_Status'].replace({

'Married': 'Married',

'Together': 'Together',

'Absurd': 'Other',

'Widow': 'Widow',

'YOLO': 'Other',

'Divorced': 'Divorced',

'Alone': 'Single',

'Single': 'Single'

})

**6. Convert Date Formats**

# Convert 'Dt\_Customer' to datetime

df['Dt\_Customer'] = pd.to\_datetime(df['Dt\_Customer'])

# Extract year, month, day

df['Year\_Customer'] = df['Dt\_Customer'].dt.year

df['Month\_Customer'] = df['Dt\_Customer'].dt.month

df['Day\_Customer'] = df['Dt\_Customer'].dt.day

**7. Rename Column Headers**

# Rename columns (lowercase, replace spaces with underscores)

df.columns = df.columns.str.lower().str.replace(' ', '\_')

**8. Check and Fix Data Types**

# Check data types again

print(df.dtypes)

# Example: Convert 'ID' to integer

df['id'] = df['id'].astype(int)

**9. Summary of Changes**

# Summary of changes

summary = """

Dataset Cleaning Summary:

1. Handled missing values in 'Income' by filling with the median.

2. Removed duplicate rows.

3. Standardized 'Education' and 'Marital\_Status' columns.

4. Converted 'Dt\_Customer' to datetime and extracted year, month, and day.

5. Renamed columns to lowercase with underscores.

6. Verified and adjusted data types.

"""

print(summary)

**Python Code (Complete)**

import pandas as pd

# Load the dataset

df = pd.read\_csv("C:/Users/hp/Downloads/marketing\_campaign.csv", sep='\t')

# Handle missing values (fixing Future Warning)

df['Income'] = df['Income'].fillna(df['Income'].median())

# Convert date formats (fixing ValueError)

df['Dt\_Customer'] = pd.to\_datetime(df['Dt\_Customer'], format='%d-%m-%Y')

# Extract year, month, day

df['Year\_Customer'] = df['Dt\_Customer'].dt.year

df['Month\_Customer'] = df['Dt\_Customer'].dt.month

df['Day\_Customer'] = df['Dt\_Customer'].dt.day

# Remove duplicates

df.drop\_duplicates(inplace=True)

# Standardize text values

df['Education'] = df['Education'].str.lower().str.capitalize()

df['Marital\_Status'] = df['Marital\_Status'].replace({

'Married': 'Married',

'Together': 'Together',

'Absurd': 'Other',

'Widow': 'Widow',

'YOLO': 'Other',

'Divorced': 'Divorced',

'Alone': 'Single',

'Single': 'Single'

})

# Rename column headers

df.columns = df.columns.str.lower().str.replace(' ', '\_')

# Check and fix data types

df['id'] = df['id'].astype(int)

# Summary of changes

summary = """Dataset Cleaning Summary:

1. Handled missing values in 'Income' by filling with the median.

2. Correctly converted 'Dt\_Customer' to datetime.

3. Extracted year, month, and day.

4. Removed duplicate rows.

5. Standardized 'Education' and 'Marital\_Status' columns.

6. Renamed columns to lowercase with underscores.

7. Verified and adjusted data types.

"""

print(summary)

# Save the cleaned dataset

df.to\_csv('cleaned\_customer\_data.csv', index=False)