import pandas as pd

import numpy as np

# Load the provided data

data = pd.read\_csv("churn\_raw\_data.csv") # Please replace with the correct file path

# Step 1: Handle missing values

# Check for missing values

missing\_values = data.isnull().sum()

# For example, if 'Income' contains missing values, you can impute them with the mean:

mean\_income = data['Income'].mean()

data['Income'].fillna(mean\_income, inplace=True)

# Step 2: Handle outliers (if necessary)

# Identify and handle outliers (I will not perform this step here).

# Step 3: Standardize or normalize numerical features (if required)

# I will skip standardization for this example, but you can use the StandardScaler if needed.

# Step 4: Encode categorical variables into numerical format

# Let's encode the 'Churn' column (Yes/No) into binary values (1/0).

data['Churn'] = data['Churn'].apply(lambda x: 1 if x == 'Yes' else 0)

# Step 5: Validate data for logical inconsistencies (not shown here)

# You should check the data for logical inconsistencies based on domain knowledge.

# Step 6: Remove duplicates (if present)

data.drop\_duplicates(inplace=True)

# Step 7: Correct data entry errors or formatting issues (not shown here)

# Step 8: Transform and create new features if needed (not shown here)

# Save the cleaned data to a new file

data.to\_csv("cleaned\_churn\_data.csv", index=False) # Adjust the file path as needed