import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Ignore warnings

import warnings

warnings.filterwarnings('ignore')

# Load the dataset

data = pd.read\_csv("BankChurners\_Cleaned.csv")

# Inspect data types

print("\nData Types:")

print(data.dtypes)

# Check the number of rows and columns

print("\nDataset Shape:")

print(data.shape)



# Check for missing values

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

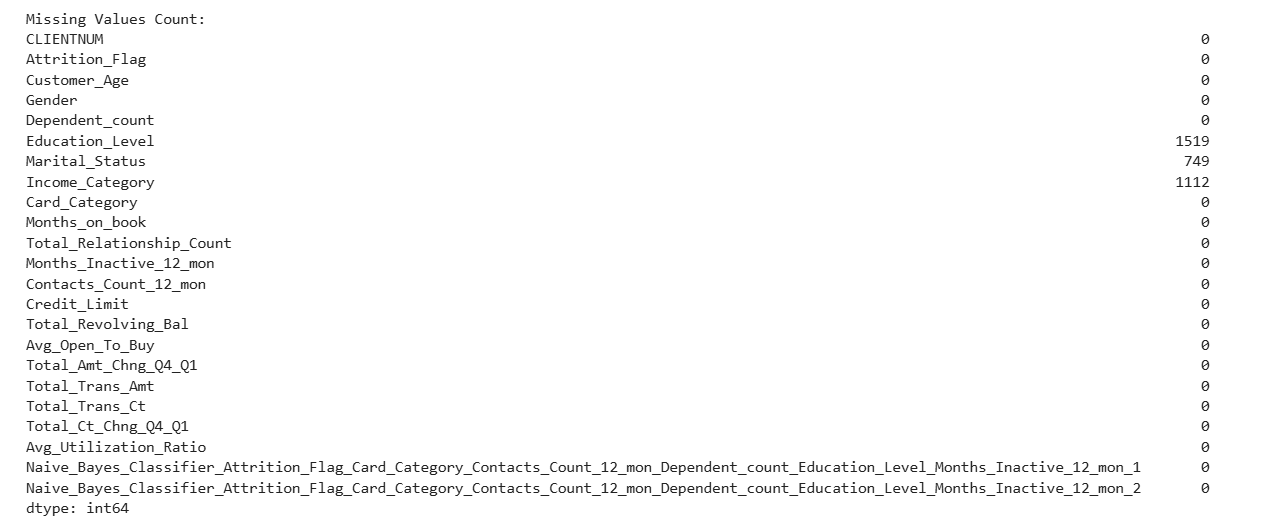
missing\_percentage = (missing\_values / len(data)) \* 100

print("\nMissing Values Count:")

print(missing\_values)

print("\nPercentage of Missing Values:")

print(missing\_percentage)





# List of columns with missing values to be handled

columns\_with\_missing = ['Education\_Level', 'Marital\_Status', 'Income\_Category']

# Step 1: Show missing values before handling them

print("\n--- Missing Values Before Handling ---\n")

print(data[columns\_with\_missing].isnull().sum())

# Step 2: Display the unique values and the mode for each column

print("\n--- Mode and Unique Values for Each Column ---\n")

for col in columns\_with\_missing:

if data[col].isnull().sum() > 0: # Only process columns with missing values

mode\_value = data[col].mode()[0] # Get the mode of the column

print(f"Column: '{col}'")

print(f"Unique Values Before Handling: {data[col].unique()}")

print(f"Mode (most frequent value) Used to Fill Missing Values: {mode\_value}\n")

# Fill missing values with the mode

data[col].fillna(mode\_value, inplace=True)

# Step 3: Verify if all missing values are handled

print("\n--- Missing Values After Handling ---\n")

print(data[columns\_with\_missing].isnull().sum())

# Step 4: Show a few rows to demonstrate how the changes look

print("\n--- Sample Rows After Handling Missing Values ---\n")

print(data[columns\_with\_missing].head(10)) # Show first 10 rows of the columns

