Experiment-1

IMPLEMENT PROGRAM FOR TIME SERIES DATA CLEANING, LOADING AND HANDLING TIME SERIES DATA AND PREPROCESSING TECHNIQUES

AIM:

TO WRITE A TO IMPLEMENT PROGRAM FOR TIME SERIES DATA CLEANING, LOADING AND HANDLING TIME SERIES DATA AND PREPROCESSING TECHNIQUES

PROCEDURE:

1) Import necessary libraries.

2)Load the necessary libraries

3) view the data & get the information about the features

4)Remove the missing values and duplicate values.

5) split the date and time from date column and convert it to datetime type

6) find the different categories where money is spend

7)Use a bar plot to check category with highest expense

CODE:

import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

df = pd.read\_csv("expense\_data\_1.csv")

df.head()

df.drop\_duplicates()

df.isna().sum()

df.drop(["Subcategory", "Note.1"], axis=1, inplace=True)

df.head()

df.describe()

df.info()

df['time'] = df['Date'].str.split(" ").str[1]

df['Date'] = df['Date'].str.split(" ").str[0]

df.head()

df['Currency'].unique()

df['Amount'] = df.apply(lambda row: row['Amount'] \* 93 if row['Currency'] == 'USD' else row['Amount'], axis=1)

df.drop(['Currency', 'Account.1'], axis=1, inplace=True)

df.head()

df['Category'].unique()

df['Account'].unique()

df['Account'] = df.apply(lambda row: 'online' if row['Account'] == 'CUB - online payment' else row['Account'], axis=1)

df['Account'].unique()

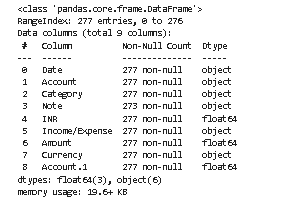
plt.figure(figsize=(16,18))

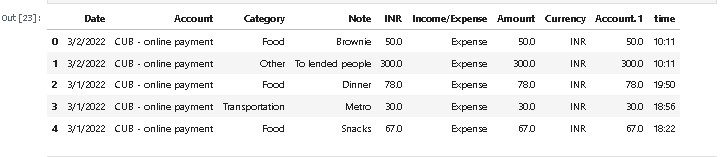
sns.barplot(data=df, x='Category', y='Amount')

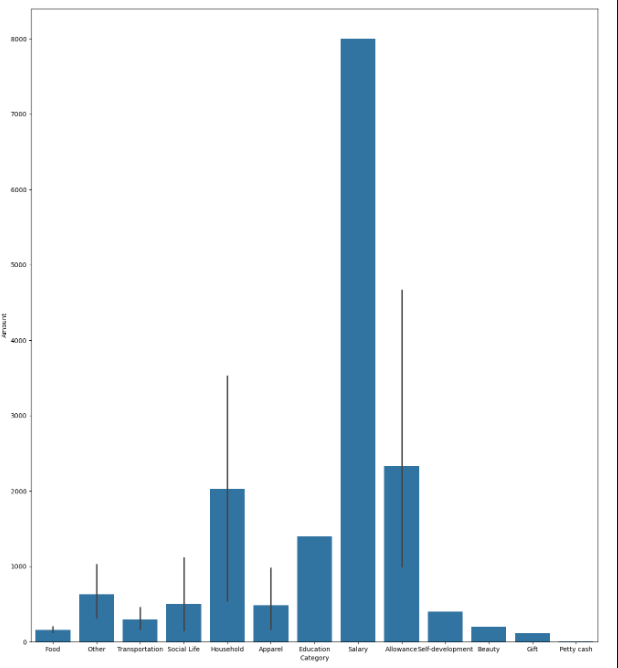
sns.countplot(df['Income/Expense'])

plt.hist(df['Amount'])

OUTPUT:







RESULT:

The program to implementing different preprocessing techniques is implemented successfully.