"""a)Load the ‘Black Friday’ dataset into one of the data structures (NumPy or Pandas).\n",

"b)Display header rows and description of the loaded dataset.\n “”””

import pandas as pd

import numpy as np

df = pd.read\_csv("blackfri.csv")

print(\"<-----Data Information----->\")

print("Head of Dataset")

print(df.head(5))

print("Head of Dataset")

print(df.describe())

print(df.info())

"""c) Remove unnecessary features (E.g. drop unwanted columns) from the dataset such as ‘User\_ID’, ‘Product\_ID ‘ ‘Stay\_In\_Current\_City\_Years’"""

df.drop=df.drop(['User\_ID','Product\_ID','Stay\_In\_Current\_City\_Years'], axis=1)

print(df.head(5))

"""d) Manipulate data by replacing empty column values in ‘City\_Category’ with a default value for the city. """

print("Filling empty values")

df['City\_Category'] = df['City\_Category'].fillna(\"A\")

print(df.head(5))

"""e) Convert the attribute ‘City\_Category’ to have ‘A’ to be ‘Metro Cities’, ‘B’ to be ‘Small Towns’ , ‘C’ to be ‘Villages’."""

print("Mapping values/attributes in City\_Category to types")

df['City\_Category'] = df['City\_Category'].map({'A':'Metro cities','B':'Small Towns','C':'Villages'})

print(df.head(5))

"""f) Rename the attribute ‘Product\_Category\_1’ to have ‘Baseball Caps’, \n",

"‘Product\_Category\_2’ to have ‘Wine Tumblers’ and ‘Product\_Category\_3’ to \n",

"have ‘Pet Raincoats’\n",

"""

print("Renaming the column names")

df.rename(columns={'Product\_Category\_1':'Baseball\_Caps','Product\_Category\_2':'Wine\_Tumblers','Product\_Category\_3':'Pet\_Raincoats'},inplace=True)

print(df.head(5))

"""g) Convert the attribute ‘Marital\_Status’ to have ‘1:Married’ and ‘0:Un-Married’\n","""

print("Mapping values/attributes in Marital Status to types")

df['Marital\_Status'] = df['Marital\_Status'].map({1:'Married',0:'Un-Married'})

print(df.head(5))

"""h) Perform the following visualizations on the loaded dataset:\n",

"i) Tally of the Number of Male & Female who bought ‘Product\_Category\_3(Pet\_Raincoats)’. \n",

"""

import matplotlib.pyplot as plt

import seaborn as sns

print("<-------Data Visualisation------->")

print(pd.crosstab(df.Gender,df.Baseball\_Caps))

print(pd.crosstab(df.Gender,df.Pet\_Raincoats))

ax = sns.countplot(data=df,x='Gender',hue='Pet\_Raincoats',palette='Set2')

ax.set(title='Male and Female who bought Pet\_Raincoats',xlabel='Gender',ylabel='Count')

plt.show()

"""h) Perform the following visualizations on the loaded dataset:\n",

"ii) Total Number of Male & Female persons belonging to each city category\n",

"""

ax = sns.countplot(data=df,x='Gender',hue='City\_Category',palette='Set1')

ax.set(title='Male and Female belonging to each city',xlabel='Gender',ylabel='Count')

plt.show()