EXPERIMENT 3A

Pandas Library – Data Preprocessing

Aim:

To understand the importance of data preprocessing in data science.

Algorithm:

1. Import the required libraries — numpy and pandas.

2. Load the dataset using pd.read\_csv() and display it.

3. Find the mode of the Country column and fill missing values with it.

4. Fill missing values in the Age column with its median.

5. Fill missing values in the Salary column with its mean (rounded).

6. Display the updated dataset after handling missing data.

7. Apply one-hot encoding to the Country column using pd.get\_dummies().

8. Concatenate the encoded columns with the remaining dataset.

9. Convert the Purchased column values from ‘Yes/No’ to 1/0.

Program:

import numpy as np

import pandas as pd

df=pd.read\_csv("C:\pre\_process\_data.csv")

print("Original Dataset\n")

print(df)

print("\n")

print(df.Country.mode())

print("\n")

print(df.Country.mode()[0])

print("\n")

print(type(df.Country.mode()))

df.Country.fillna(df.Country.mode()[0],inplace=True)

df.Age.fillna(df.Age.median(),inplace=True)

df.Salary.fillna(round(df.Salary.mean()),inplace=True)

print(df)

print("\n")

print(pd.get\_dummies(df.Country))

print("\n")

updated\_dataset=pd.concat([pd.get\_dummies(df.Country),df.iloc[:,[1,2,3]]],axis=1)

print(updated\_dataset)

print("\n")

print(df.info())

print("\n")

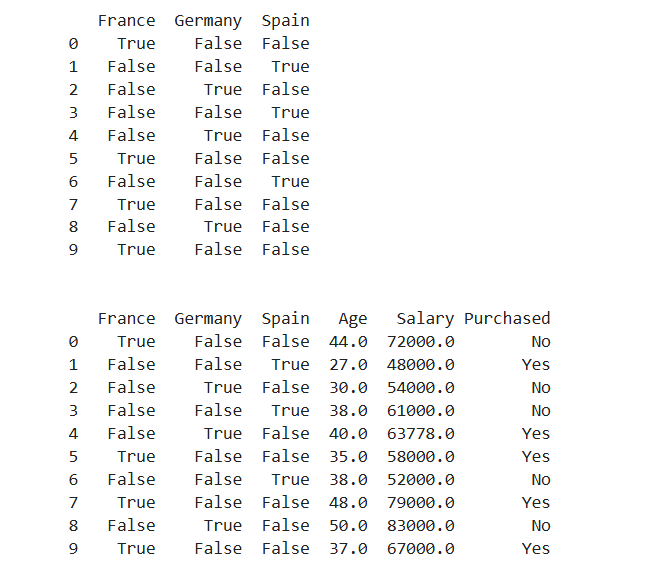
updated\_dataset.Purchased.replace(['No','Yes'],[0,1],inplace=True)

print(updated\_dataset)

Output:

A screenshot of a computer

AI-generated content may be incorrect.



A screenshot of a computer

AI-generated content may be incorrect.

Result:

Hence the python program for data preprocessing is written and executed successfully.