import numpy as np

import pandas as pd

from sklearn.preprocessing import StandardScaler

from sklearn.preprocessing import LabelEncoder

from sklearn.preprocessing import OneHotEncoder

from sklearn.impute import SimpleImputer

data={'marks':[8,4,6,np.nan,8,9],'sec':['a','b','c',np.nan,'a','a'],'color':['red','blue','green','pink','voilet','orange'],'juice':['mango','strawberry','lemon','butterfruit','apple','waterlmelon']}

print(data)

df=pd.DataFrame(data)

print("input dataframe")

print(df)

print("after applying imputer")

 #Imputer for numerical column (marks)

num\_imputer = SimpleImputer(strategy='mean')

df['marks'] =num\_imputer.fit\_transform(df[['marks']])

# Imputer for categorical column (sec)

cat\_imputer =SimpleImputer(strategy='most\_frequent')

df['sec'] =cat\_imputer.fit\_transform(df[['sec']]).ravel()

print(df)

print("after applying label encoder")

le=LabelEncoder()

ledata=le.fit\_transform(df['color'])

print(ledata)

df['color']=ledata

print(df)

print("after applying One Hot encoder")

hec=OneHotEncoder(sparse\_output=False)

trdata=hec.fit\_transform(df[['juice']])

new\_df=pd.DataFrame(trdata,columns=hec.get\_feature\_names\_out(['juice']))

df2=pd.concat([df,new\_df],axis=1)

df2.drop(columns=['juice'],inplace=True)

pd.set\_option('display.expand\_frame\_repr', False)

# Prevent line breaks

print(df2)

#print(df2)

#applying Standard Scaler

ss=StandardScaler()

sdata=ss.fit\_transform(df2[['marks','color']])

print(sdata)

df3=pd.DataFrame(sdata,columns=['marks','color'])

print(df3)

df2.drop(columns=['marks','color'],inplace=True)

df4=pd.concat([df2,df3],axis=1)

print(df4)