DAY-2



 PRESENTED BY:

* CH. ALEKHYA GAYATHRI
* E. HIMA BINDHU
* D. TRISHA
* A. HEMA
* I. MANISHA
* K. ASHWINI

TEAM LEADER: CH. ALEKHYA GAYATHRI

\*WHY WE USE NUMPY AND PANDAS:

NUMPY: Numpy is a library which is used for perform advance mathematical calculations in python.

PANDAS: Pandas is a library which is used to access the data in present in the file.

* Pandas is power full python library used for data manipulation and data

analysis

\*HOW TO CHECK INSTALLATION LIBRARY:

COMMAND PROMPT: pip list ---> which shows the library are related to python.

PROCESS:

* Open chrome and open google collab click on first link
* Open new notebook
* In upside one connected button and click it.
* CSV file is downloaded by data using numpy,pandas.
* Then open new notebook and add it in the new notebook.
* In right side open file select the folder and upload the session data
* And there we can see the file in screen in data CSV.
* Search the Kaggle and click it.
* Now we can start the coding with generative with AI

KAGGLE: So many data sets are available or not available in the website we have to check is known as Kaggle.

\*WE HAVE TO IMPORT THE DATA SET IN COLAB NOTEBOOK:

step-1: we have to import the dataset in colab notebook

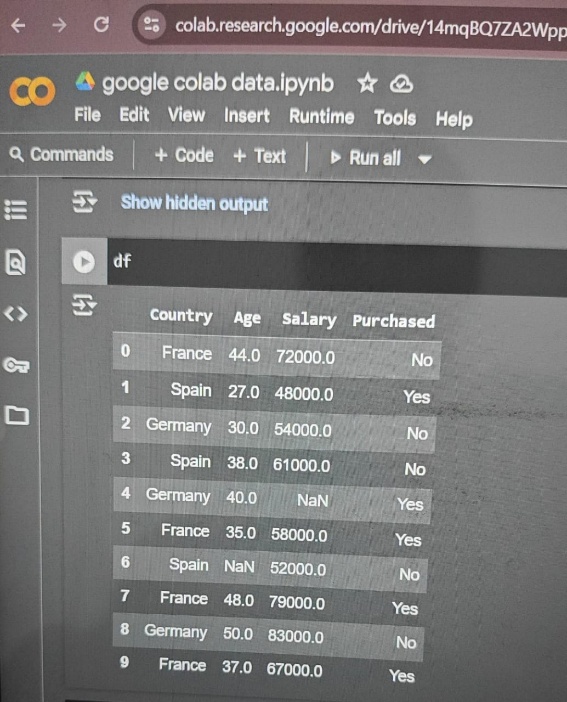
#we have to import pandas library

import pandas as pd #pd is a neck name for pandas

df=pd.read\_csv('Data.csv')#df is a variable, read\_csv is a predifined function from pandas package

type(df)

OUTPUT:



\*GROUPING THE INPUT COLUMNS AND OUTPUT COLUMNS:

step-2:grouping the input columns and output columns

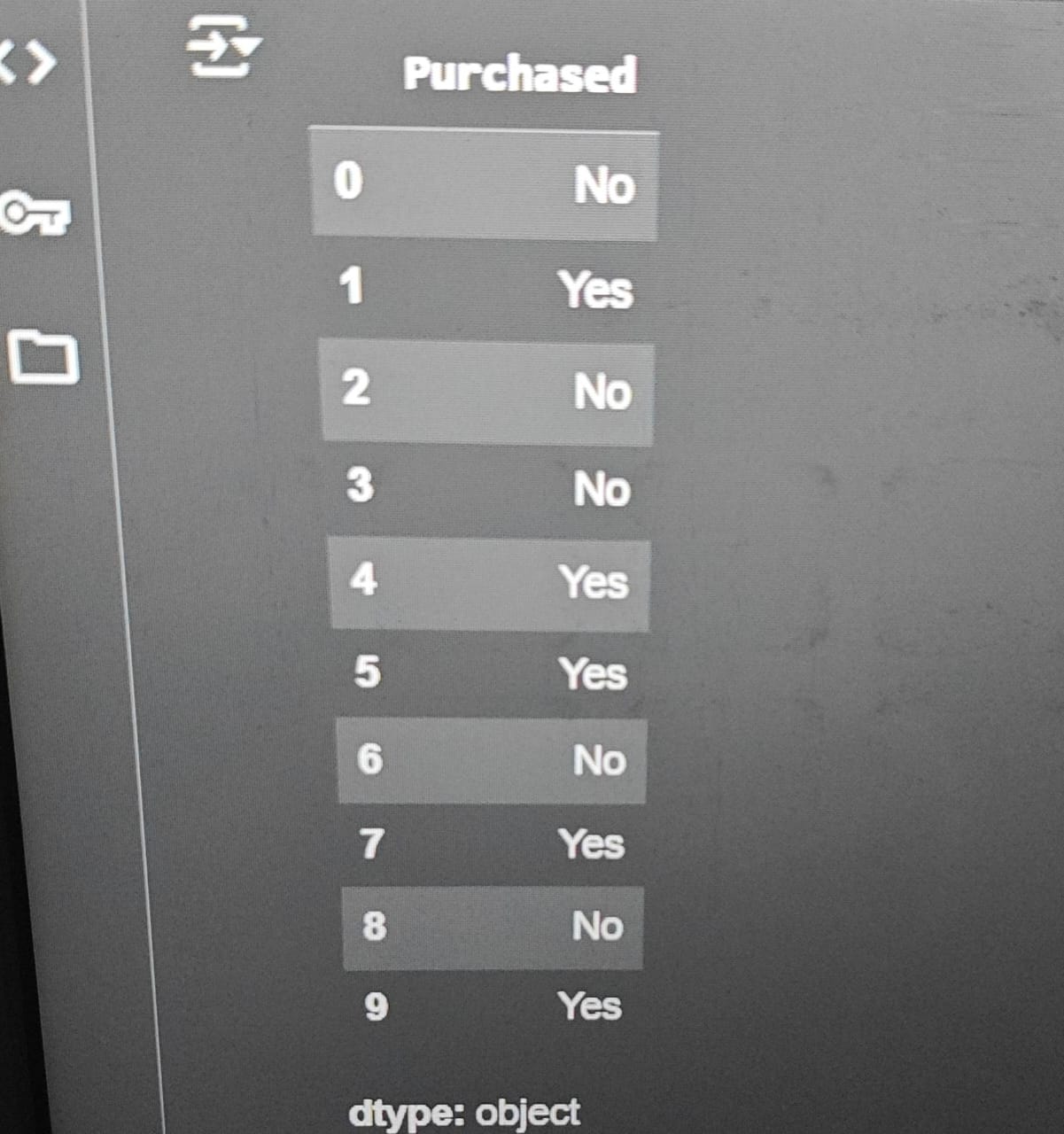
X=df.iloc[:,:-1]

OUTPUT:



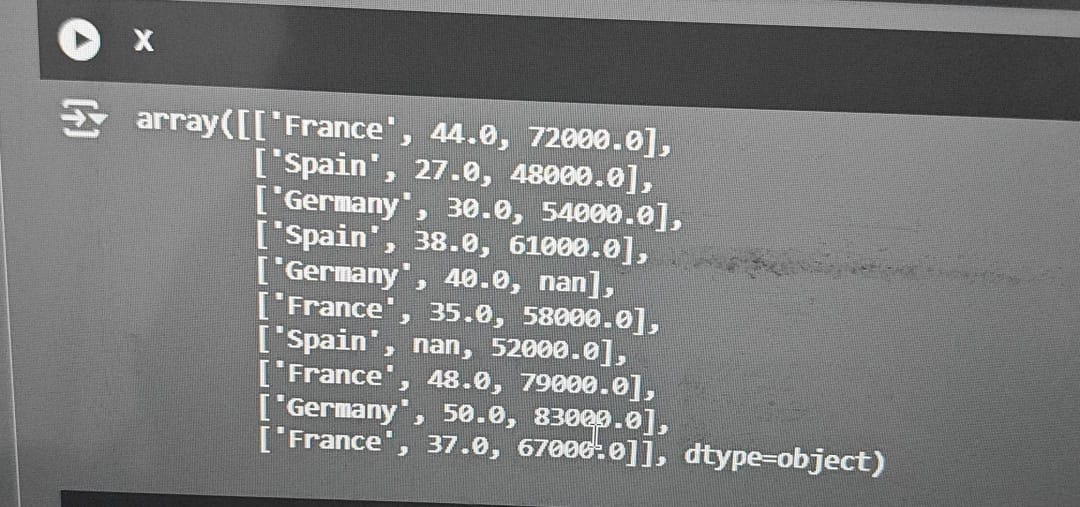
Y=df.iloc[:,-1]#slicing the output column

OUTPUT:



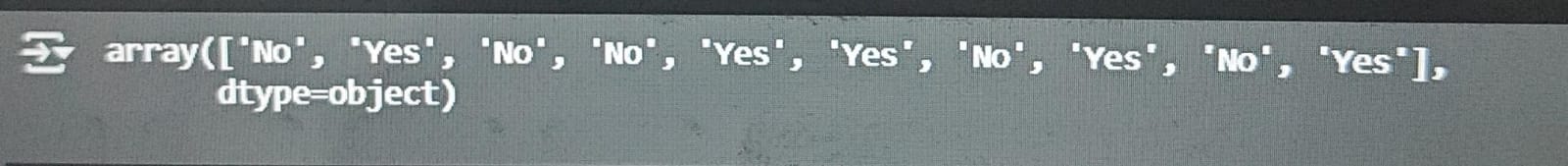
X=X.values

OUTPUT:



Y=Y.values

OUTPUT:



\*HANDLING MISSING DATA OR VALUES:

#step-3:handling missing data or values

#scinario-1:#calculate mean or average of that column and replace the empty values

#scinario-2:#missing values in non numerical columns

#scinario-3:if we have missing values in huge data set ex:2 rows as empty values out of 2000

import numpy as np

from sklearn.impute import SimpleImputer #it is a class which was importing from sklearn.impute

imputer=SimpleImputer(missing\_values=np.nan,strategy='mean')

imputer.fit(X[:,1:3])

X[:,1:3]=imputer.transform(X[:,1:3])#which performs the actions

\*HANDLING THE CATEGORICAL DATA CONVERT THE CATEGORICAL DATA INTO NUMERIC DATA:

#step-4:handling the categorical data,convert the categorical data into numeric data

#which converts the text data into numeric data

from sklearn.preprocessing import OneHotEncoder #label encoder is a class from sk.preprocessing or libray

from sklearn.compose import ColumnTransformer #column transfromer is a class imported from sk.compose

#column transformer is uesd to apply changes of numerical data into categorical data

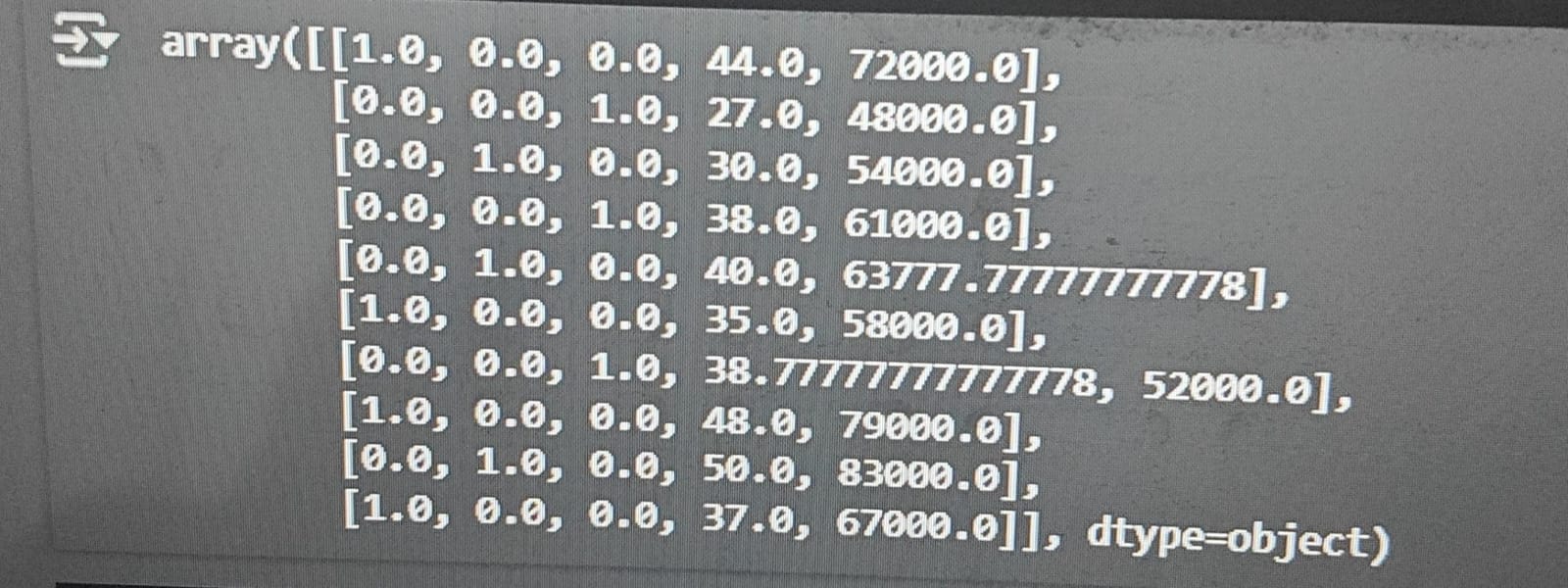
ct=ColumnTransformer(transformers=[('encoder',OneHotEncoder(),[0])],remainder='passthrough')#encoder is a method which converts text to binary

X=ct.fit\_transform(X)

#onehot encoder is only used for inpuits

X

OUTPUT:



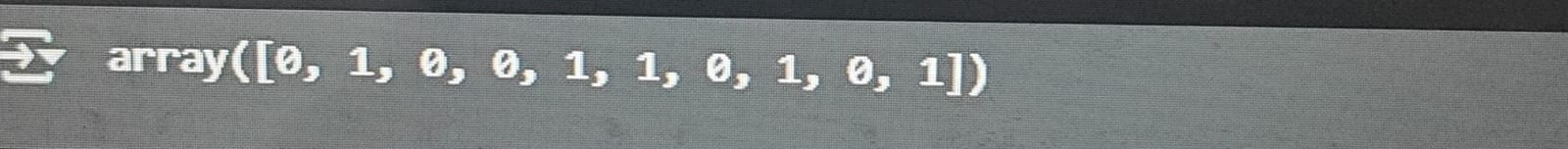
from sklearn.preprocessing import LabelEncoder

le=LabelEncoder()

Y=le.fit\_transform(Y)

Y

OUTPUT:



\*SPLIT THE DATA INTO TRAINING AND TESTING DATA:

#step-5:split the data into training and testing data

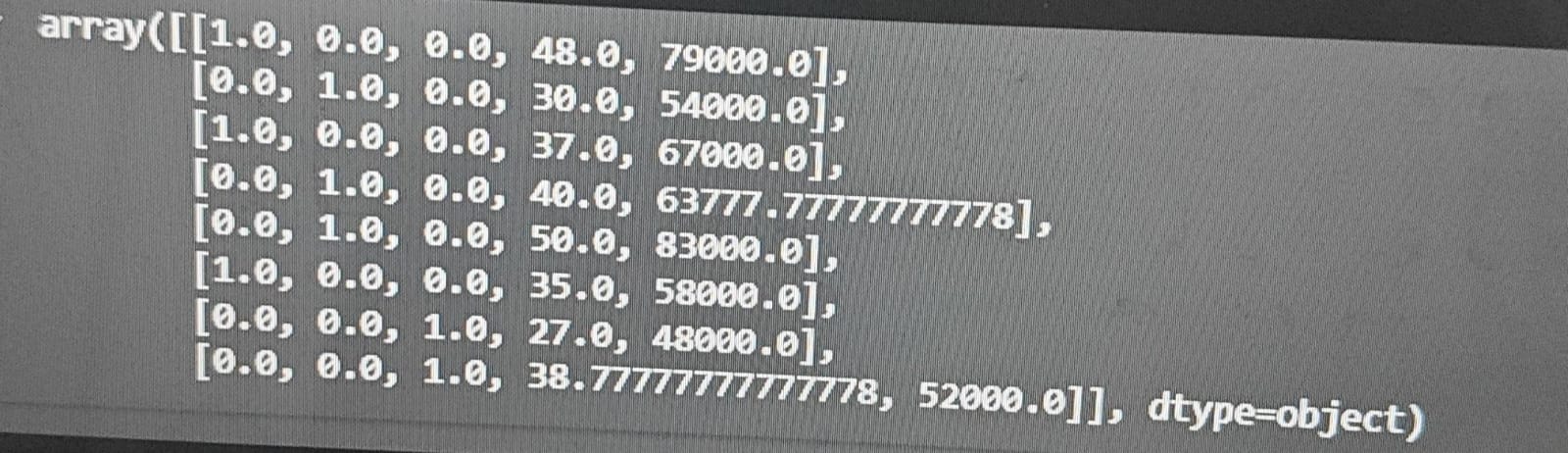
from sklearn.model\_selection import train\_test\_split# it is a class which was imported from sklearn.model\_selection

#it is was used for splitting the data for training and selecting the data

X\_train,X\_test,Y\_train,Y\_test=train\_test\_split(X,Y,test\_size=0.2)

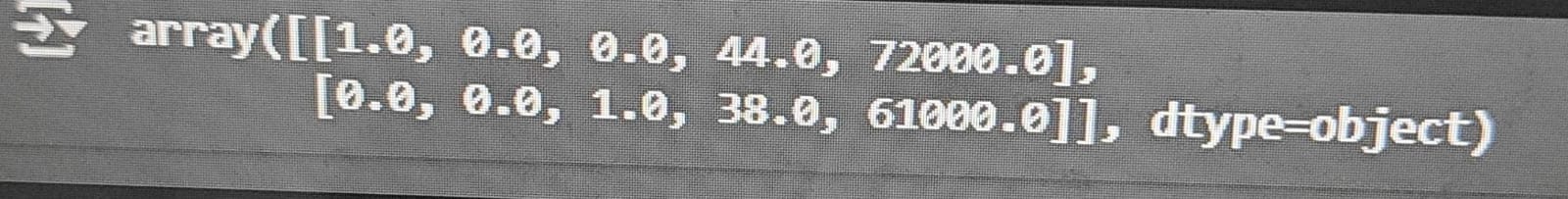
X\_train

OUTPUT:

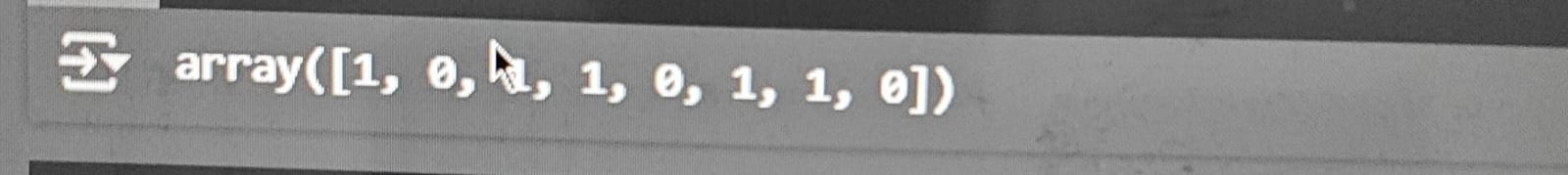


X\_test

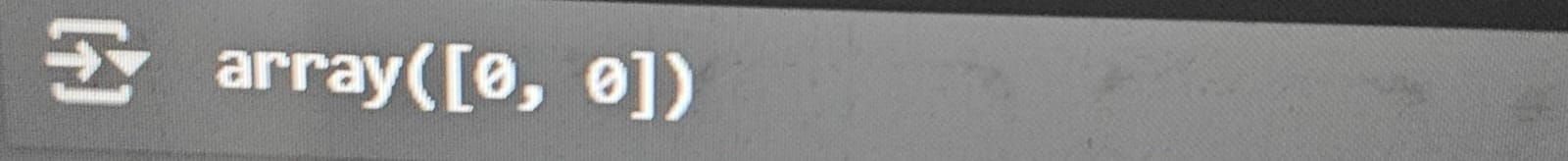
OUTPUT:



Y\_train



Y\_test



THANK YOU