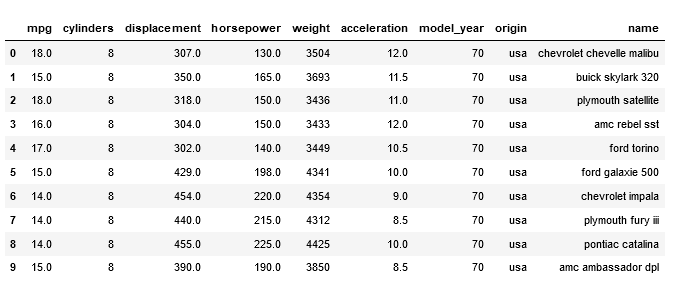
**PANDAS EXERCISE 1**

**Code:**

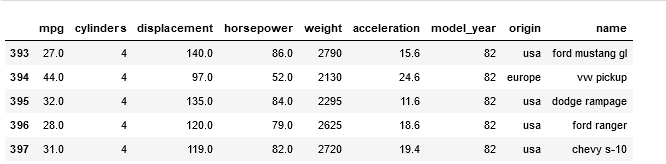
import pandas as pd

cars=pd.read\_csv('https://github.com/YBIFoundation/Dataset/raw/main/MPG.csv')

cars.head(10)

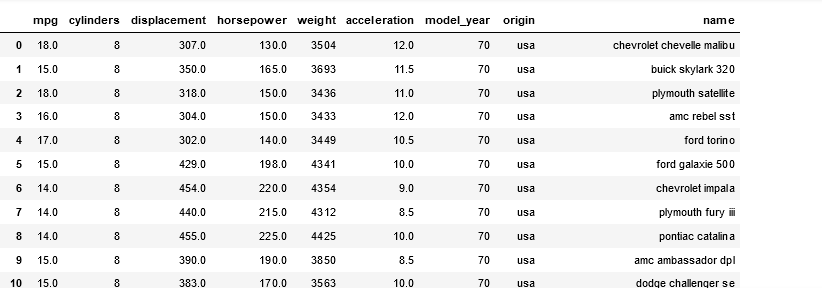


cars.tail()

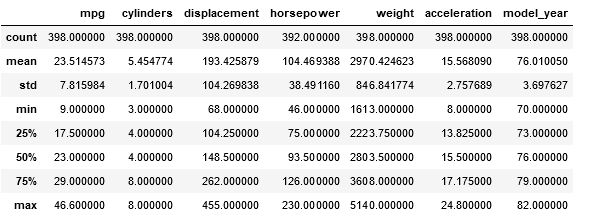


pd.options.display.max\_rows = None

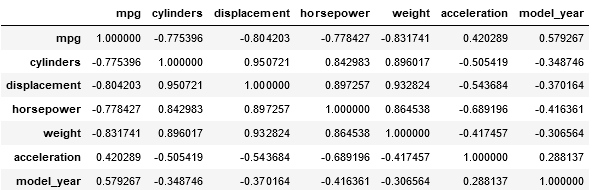
cars



cars.describe()



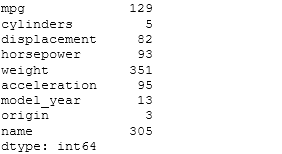
cars.corr()



cars.shape

(398, 9)

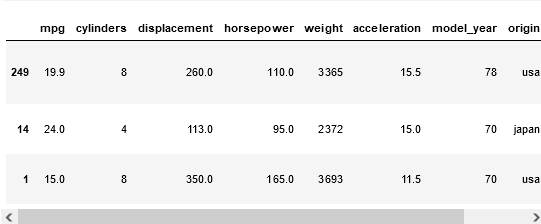
cars.nunique()



cars['origin'].value\_counts()



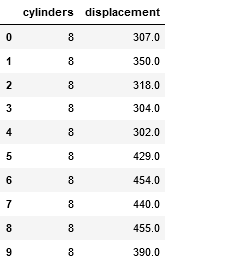
cars.sample(3)



cars.displacement[13]

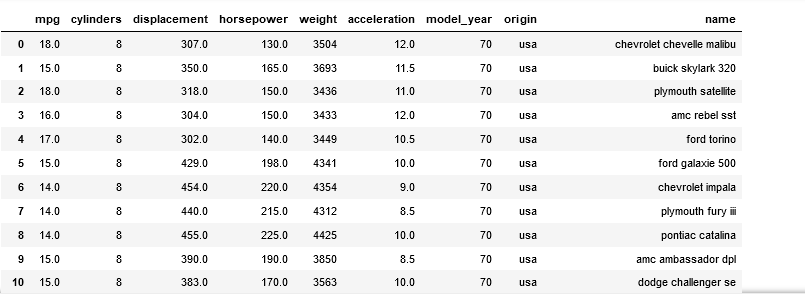
455.0

cars.loc[0:9,['cylinders', 'displacement']]



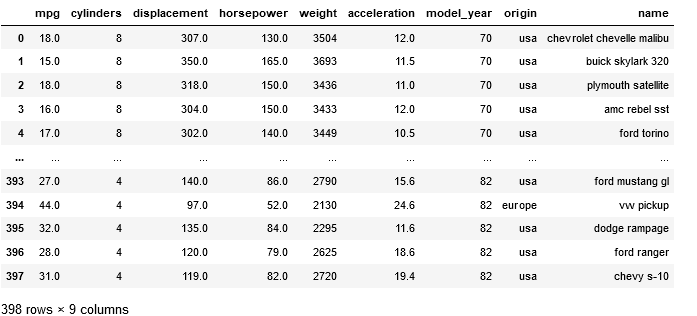
sub\_sample = cars.iloc[:, [1,2]]

sub\_sample



pd.options.display.max\_rows = 10

cars



**LABEL ENCODING**

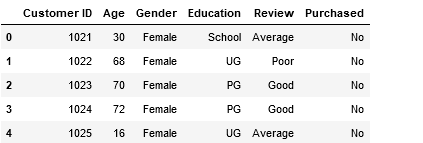
**Code:**

import pandas as pd

import numpy as np

df=pd.read\_csv('https://github.com/YBIFoundation/Dataset/raw/main/Customer%20Purchase.csv')

df.head()



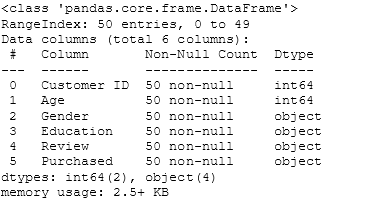
df.shape

(50, 6)

df.columns

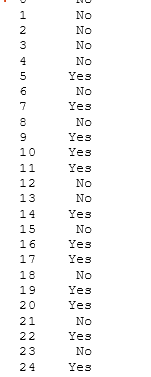
Index(['Customer ID', 'Age', 'Gender', 'Education', 'Review', 'Purchased'], dtype='object')

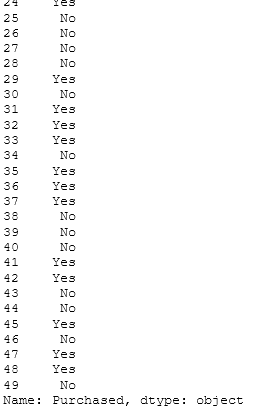
df.info()



y=df['Purchased']

y





from sklearn.preprocessing import LabelEncoder

ls = LabelEncoder()

y=ls.fit\_transform(y)

y



ls.classes\_

array(['No', 'Yes'], dtype=object)

ls.inverse\_transform([0,0,1,1])

array(['No', 'No', 'Yes', 'Yes'], dtype=object)

**ORDINAL ENCODING**

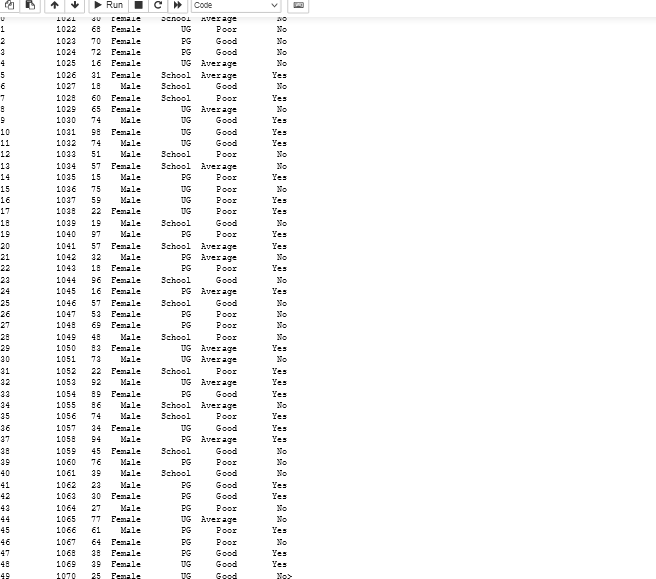
**Code:**

import pandas as pd

import numpy as np

df=pd.read\_csv('https://github.com/YBI-Foundation/Dataset/raw/main/Customer%20Purchase.csv')

df.head



df.shape

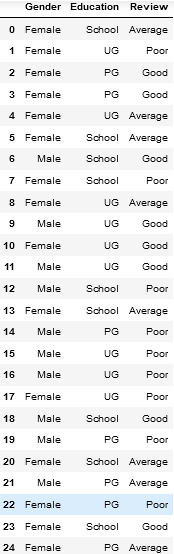
(50, 6)

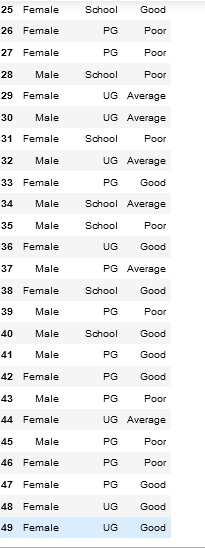
df.columns

Index(['Customer ID', 'Age', 'Gender', 'Education', 'Review', 'Purchased'], dtype='object')

x=df[['Gender','Education','Review']]

x





from sklearn.preprocessing import OrdinalEncoder

oe=OrdinalEncoder()

x=oe.fit\_transform(x)

x

array([[0., 1., 0.],

[0., 2., 2.],

[0., 0., 1.],

[0., 0., 1.],

[0., 2., 0.],

[0., 1., 0.],

[1., 1., 1.],

[0., 1., 2.],

[0., 2., 0.],

[1., 2., 1.],

[0., 2., 1.],

[1., 2., 1.],

[1., 1., 2.],

[0., 1., 0.],

[1., 0., 2.],

[1., 2., 2.],

[1., 2., 2.],

[0., 2., 2.],

[1., 1., 1.],

[1., 0., 2.],

[0., 1., 0.],

[1., 0., 0.],

[0., 0., 2.],

[0., 1., 1.],

[0., 0., 0.],

[0., 1., 1.],

[0., 0., 2.],

[0., 0., 2.],

[1., 1., 2.],

[0., 2., 0.],

[1., 2., 0.],

[0., 1., 2.],

[1., 2., 0.],

[0., 0., 1.],

[1., 1., 0.],

[1., 1., 2.],

[0., 2., 1.],

[1., 0., 0.],

[0., 1., 1.],

[1., 0., 2.],

[1., 1., 1.],

[1., 0., 1.],

[0., 0., 1.],

[1., 0., 2.],

[0., 2., 0.],

[1., 0., 2.],

[0., 0., 2.],

[0., 0., 1.],

[0., 2., 1.],

[0., 2., 1.]])

oe.categories\_

[array(['Female', 'Male'], dtype=object),

array(['PG', 'School', 'UG'], dtype=object),

array(['Average', 'Good', 'Poor'], dtype=object)]

oe.inverse\_transform([[0,0,0]])

array([['Female', 'PG', 'Average']], dtype=object)

oe.inverse\_transform([[1,1,1]])

array([['Male', 'School', 'Good']], dtype=object)

oe.inverse\_transform([[1,2,2]])

array([['Male', 'UG', 'Poor']], dtype=object)

**FEATURE STANDARD SCALING**

**Code:**

import pandas as pd

import numpy as np

df=pd.read\_csv('https://github.com/YBI-Foundation/Dataset/raw/main/Boston.csv')

df.head()



df.shape

(506, 14)

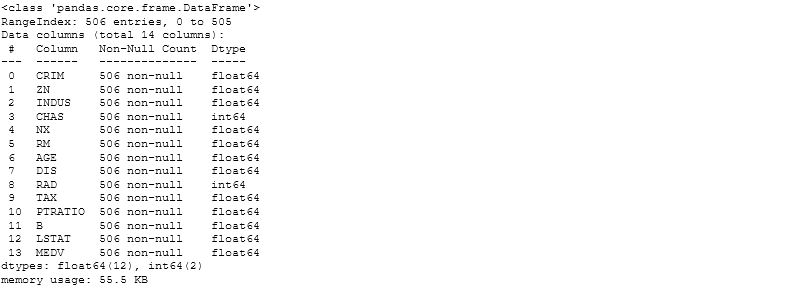
df.columns

Index(['CRIM', 'ZN', 'INDUS', 'CHAS', 'NX', 'RM', 'AGE', 'DIS', 'RAD', 'TAX',

'PTRATIO', 'B', 'LSTAT', 'MEDV'],

dtype='object')

df.info()



df.describe()



y=df['MEDV']

y.shape

(506,)

X=df[['CRIM','ZN','INDUS','CHAS','NX','RM','AGE','DIS','RAD','TAX','PTRATIO','B','LSTAT']]

X.shape

(506, 13)

from sklearn.model\_selection import train\_test\_split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size = 0.3, random\_state = 2529)

X\_train.shape, X\_test.shape, y\_train.shape, y\_test.shape

((354, 13), (152, 13), (354,), (152,))

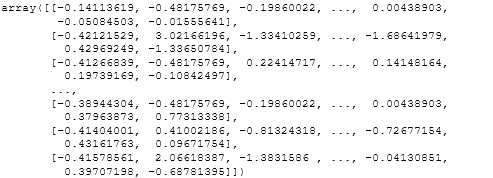
from sklearn.preprocessing import StandardScaler

ss = StandardScaler()

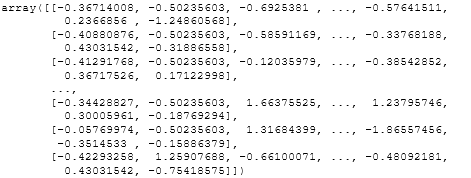
X\_train\_ss = ss.fit\_transform(X\_train)

X\_test\_ss = ss.fit\_transform(X\_test)

X\_train\_ss



X\_test\_ss



X\_train\_ss.mean(axis=0)



X\_test\_ss.mean(axis=0)



X\_train\_ss.std(axis=0)

array([1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.])

X\_train\_ss.std(axis=0)

array([1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1., 1.])

**FEATURE SCALING**

**Code:**

import pandas as pd

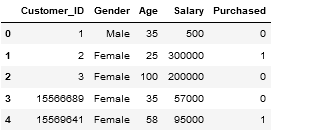
import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

df=pd.read\_csv('https://github.com/YBI-Foundation/Dataset/raw/main/Online%20Purchase.csv')

df.head()



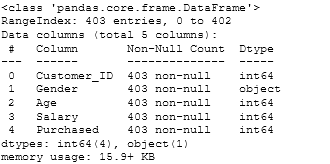
df.shape

(403, 5)

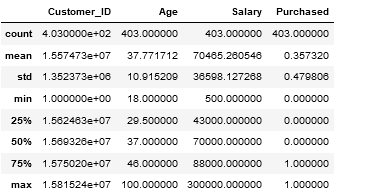
df.columns

Index(['Customer\_ID', 'Gender', 'Age', 'Salary', 'Purchased'], dtype='object')

df.info()



df.describe()





y=df['Purchased']

y.shape

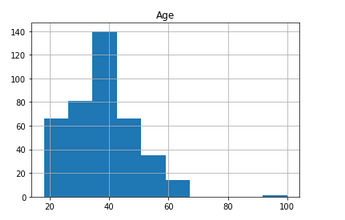
(403,)

X=df[['Age','Salary']]

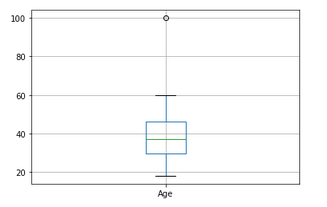
X.shape

(403, 2)

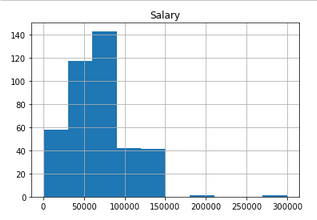
df[['Age']].hist();



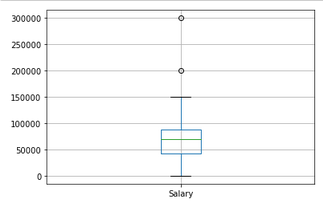
df[['Age']].boxplot()



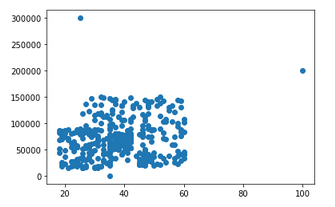
df[['Salary']].hist();



df[['Salary']].boxplot();



plt.scatter(df['Age'], df['Salary']);





from sklearn.model\_selection import train\_test\_split

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size = 0.3, random\_state = 2529)

X\_train.shape, X\_test.shape, y\_train.shape, y\_test.shape

((282, 2), (121, 2), (282,), (121,))

from sklearn.preprocessing import StandardScaler

ss = StandardScaler()

X\_train\_ss = ss.fit\_transform(X\_train)

X\_test\_ss = ss.fit\_transform(X\_test)

X\_train\_ss = pd.DataFrame(X\_train\_ss, columns=X\_train.columns)

X\_test\_ss = pd.DataFrame(X\_test\_ss, columns=X\_test.columns)

fig, (ax1, ax2) = plt.subplots(ncols=2, figsize=(12,5))

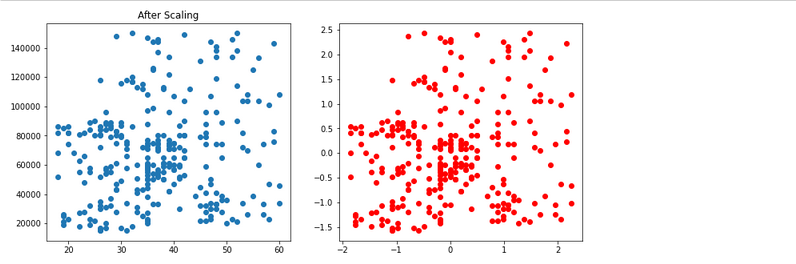
ax1.scatter(X\_train['Age'],X\_train['Salary'])

ax1.set\_title('Before Scaling')

ax2.scatter(X\_train\_ss['Age'],X\_train\_ss['Salary'],color='red')

ax1.set\_title('After Scaling')

plt.show()



**ONE HOT ENCODING**

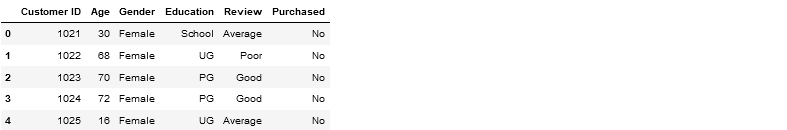
**CODE:**

import pandas as pd

import numpy as np

df=pd.read\_csv('https://github.com/YBI-Foundation/Dataset/raw/main/Customer%20Purchase.csv')

df.head()



df.shape

(50, 6)





df.columns

Index(['Customer ID', 'Age', 'Gender', 'Education', 'Review', 'Purchased'], dtype='object')

df.info()



X=df[['Gender','Education','Review']]

X

|  | **Gender** | **Education** | **Review** |
| --- | --- | --- | --- |
| **0** | Female | School | Average |
| **1** | Female | UG | Poor |
| **2** | Female | PG | Good |
| **3** | Female | PG | Good |
| **4** | Female | UG | Average |
| **5** | Female | School | Average |
| **6** | Male | School | Good |
| **7** | Female | School | Poor |
| **8** | Female | UG | Average |
| **9** | Male | UG | Good |
| **10** | Female | UG | Good |
| **11** | Male | UG | Good |
| **12** | Male | School | Poor |
| **13** | Female | School | Average |
| **14** | Male | PG | Poor |
| **15** | Male | UG | Poor |
| **16** | Male | UG | Poor |
| **17** | Female | UG | Poor |
| **18** | Male | School | Good |
| **19** | Male | PG | Poor |
| **20** | Female | School | Average |
| **21** | Male | PG | Average |
| **22** | Female | PG | Poor |
| **23** | Female | School | Good |
| **24** | Female | PG | Average |
| **25** | Female | School | Good |
| **26** | Female | PG | Poor |
| **27** | Female | PG | Poor |
| **28** | Male | School | Poor |
| **29** | Female | UG | Average |
| **30** | Male | UG | Average |
| **31** | Female | School | Poor |
| **32** | Male | UG | Average |
| **33** | Female | PG | Good |
| **34** | Male | School | Average |
| **35** | Male | School | Poor |
| **36** | Female | UG | Good |
| **37** | Male | PG | Average |
| **38** | Female | School | Good |
| **39** | Male | PG | Poor |
| **40** | Male | School | Good |
| **41** | Male | PG | Good |
| **42** | Female | PG | Good |
| **43** | Male | PG | Poor |
| **44** | Female | UG | Average |
| **45** | Male | PG | Poor |
| **46** | Female | PG | Poor |
| **47** | Female | PG | Good |
| **48** | Female | UG | Good |
| **49** | Female | UG | Good |



from sklearn.preprocessing import OneHotEncoder

ohe = OneHotEncoder()

X\_Gender = ohe.fit\_transform(X[['Gender']])

ohe.categories\_

[array(['Female', 'Male'], dtype=object)]

X\_Gender.toarray().shape

(50, 2)

X\_Gender.toarray()

array([[1., 0.],

[1., 0.],

[1., 0.],

[1., 0.],

[1., 0.],

[1., 0.],

[0., 1.],

[1., 0.],

[1., 0.],

[0., 1.],

[1., 0.],

[0., 1.],

[0., 1.],

[1., 0.],

[0., 1.],

[0., 1.],

[0., 1.],

[1., 0.],

[0., 1.],

[0., 1.],

[1., 0.],

[0., 1.],

[1., 0.],

[1., 0.],

[1., 0.],

[1., 0.],

[1., 0.],

[1., 0.],

[0., 1.],

[1., 0.],

[0., 1.],

[1., 0.],

[0., 1.],

[1., 0.],

[0., 1.],

[0., 1.],

[1., 0.],

[0., 1.],

[1., 0.],

[0., 1.],

[0., 1.],

[0., 1.],

[1., 0.],

[0., 1.],

[1., 0.],

[0., 1.],

[1., 0.],

[1., 0.],

[1., 0.],

[1., 0.]])



X\_Education = ohe.fit\_transform(X[['Education']])

ohe.categories\_

[array(['PG', 'School', 'UG'], dtype=object)]



X\_Education.toarray().shape

(50, 3)



X\_Education.toarray()

array([[0., 1., 0.],

[0., 0., 1.],

[1., 0., 0.],

[1., 0., 0.],

[0., 0., 1.],

[0., 1., 0.],

[0., 1., 0.],

[0., 1., 0.],

[0., 0., 1.],

[0., 0., 1.],

[0., 0., 1.],

[0., 0., 1.],

[0., 1., 0.],

[0., 1., 0.],

[1., 0., 0.],

[0., 0., 1.],

[0., 0., 1.],

[0., 0., 1.],

[0., 1., 0.],

[1., 0., 0.],

[0., 1., 0.],

[1., 0., 0.],

[1., 0., 0.],

[0., 1., 0.],

[1., 0., 0.],

[0., 1., 0.],

[1., 0., 0.],

[1., 0., 0.],

[0., 1., 0.],

[0., 0., 1.],

[0., 0., 1.],

[0., 1., 0.],

[0., 0., 1.],

[1., 0., 0.],

[0., 1., 0.],

[0., 1., 0.],

[0., 0., 1.],

[1., 0., 0.],

[0., 1., 0.],

[1., 0., 0.],

[0., 1., 0.],

[1., 0., 0.],

[1., 0., 0.],

[1., 0., 0.],

[0., 0., 1.],

[1., 0., 0.],

[1., 0., 0.],

[1., 0., 0.],

[0., 0., 1.],

[0., 0., 1.]])



X\_Review = ohe.fit\_transform(X[['Review']])

ohe.categories\_

[array(['Average', 'Good', 'Poor'], dtype=object)]



X\_Review.toarray().shape

(50, 3)



X\_Review.toarray()

array([[1., 0., 0.],

[0., 0., 1.],

[0., 1., 0.],

[0., 1., 0.],

[1., 0., 0.],

[1., 0., 0.],

[0., 1., 0.],

[0., 0., 1.],

[1., 0., 0.],

[0., 1., 0.],

[0., 1., 0.],

[0., 1., 0.],

[0., 0., 1.],

[1., 0., 0.],

[0., 0., 1.],

[0., 0., 1.],

[0., 0., 1.],

[0., 0., 1.],

[0., 1., 0.],

[0., 0., 1.],

[1., 0., 0.],

[1., 0., 0.],

[0., 0., 1.],

[0., 1., 0.],

[1., 0., 0.],

[0., 1., 0.],

[0., 0., 1.],

[0., 0., 1.],

[0., 0., 1.],

[1., 0., 0.],

[1., 0., 0.],

[0., 0., 1.],

[1., 0., 0.],

[0., 1., 0.],

[1., 0., 0.],

[0., 0., 1.],

[0., 1., 0.],

[1., 0., 0.],

[0., 1., 0.],

[0., 0., 1.],

[0., 1., 0.],

[0., 1., 0.],

[0., 1., 0.],

[0., 0., 1.],

[1., 0., 0.],

[0., 0., 1.],

[0., 0., 1.],

[0., 1., 0.],

[0., 1., 0.],

[0., 1., 0.]])



X=df[['Gender','Education','Review']]

X=ohe.fit\_transform(X)

X.toarray().shape

(50, 8)



X.toarray()

array([[1., 0., 0., 1., 0., 1., 0., 0.],

[1., 0., 0., 0., 1., 0., 0., 1.],

[1., 0., 1., 0., 0., 0., 1., 0.],

[1., 0., 1., 0., 0., 0., 1., 0.],

[1., 0., 0., 0., 1., 1., 0., 0.],

[1., 0., 0., 1., 0., 1., 0., 0.],

[0., 1., 0., 1., 0., 0., 1., 0.],

[1., 0., 0., 1., 0., 0., 0., 1.],

[1., 0., 0., 0., 1., 1., 0., 0.],

[0., 1., 0., 0., 1., 0., 1., 0.],

[1., 0., 0., 0., 1., 0., 1., 0.],

[0., 1., 0., 0., 1., 0., 1., 0.],

[0., 1., 0., 1., 0., 0., 0., 1.],

[1., 0., 0., 1., 0., 1., 0., 0.],

[0., 1., 1., 0., 0., 0., 0., 1.],

[0., 1., 0., 0., 1., 0., 0., 1.],

[0., 1., 0., 0., 1., 0., 0., 1.],

[1., 0., 0., 0., 1., 0., 0., 1.],

[0., 1., 0., 1., 0., 0., 1., 0.],

[0., 1., 1., 0., 0., 0., 0., 1.],

[1., 0., 0., 1., 0., 1., 0., 0.],

[0., 1., 1., 0., 0., 1., 0., 0.],

[1., 0., 1., 0., 0., 0., 0., 1.],

[1., 0., 0., 1., 0., 0., 1., 0.],

[1., 0., 1., 0., 0., 1., 0., 0.],

[1., 0., 0., 1., 0., 0., 1., 0.],

[1., 0., 1., 0., 0., 0., 0., 1.],

[1., 0., 1., 0., 0., 0., 0., 1.],

[0., 1., 0., 1., 0., 0., 0., 1.],

[1., 0., 0., 0., 1., 1., 0., 0.],

[0., 1., 0., 0., 1., 1., 0., 0.],

[1., 0., 0., 1., 0., 0., 0., 1.],

[0., 1., 0., 0., 1., 1., 0., 0.],

[1., 0., 1., 0., 0., 0., 1., 0.],

[0., 1., 0., 1., 0., 1., 0., 0.],

[0., 1., 0., 1., 0., 0., 0., 1.],

[1., 0., 0., 0., 1., 0., 1., 0.],

[0., 1., 1., 0., 0., 1., 0., 0.],

[1., 0., 0., 1., 0., 0., 1., 0.],

[0., 1., 1., 0., 0., 0., 0., 1.],

[0., 1., 0., 1., 0., 0., 1., 0.],

[0., 1., 1., 0., 0., 0., 1., 0.],

[1., 0., 1., 0., 0., 0., 1., 0.],

[0., 1., 1., 0., 0., 0., 0., 1.],

[1., 0., 0., 0., 1., 1., 0., 0.],

[0., 1., 1., 0., 0., 0., 0., 1.],

[1., 0., 1., 0., 0., 0., 0., 1.],

[1., 0., 1., 0., 0., 0., 1., 0.],

[1., 0., 0., 0., 1., 0., 1., 0.],

[1., 0., 0., 0., 1., 0., 1., 0.]])



X=df[['Gender','Education','Review']]

ohe = OneHotEncoder(categories=[['Male','Female'],['School','UG','PG'],['Poor','Average','Good']])

X = ohe.fit\_transform(X)

ohe.categories\_

[array(['Male', 'Female'], dtype=object),

array(['School', 'UG', 'PG'], dtype=object),

array(['Poor', 'Average', 'Good'], dtype=object)]



X.toarray().shape

(50, 8)



X.toarray()

array([[0., 1., 1., 0., 0., 0., 1., 0.],

[0., 1., 0., 1., 0., 1., 0., 0.],

[0., 1., 0., 0., 1., 0., 0., 1.],

[0., 1., 0., 0., 1., 0., 0., 1.],

[0., 1., 0., 1., 0., 0., 1., 0.],

[0., 1., 1., 0., 0., 0., 1., 0.],

[1., 0., 1., 0., 0., 0., 0., 1.],

[0., 1., 1., 0., 0., 1., 0., 0.],

[0., 1., 0., 1., 0., 0., 1., 0.],

[1., 0., 0., 1., 0., 0., 0., 1.],

[0., 1., 0., 1., 0., 0., 0., 1.],

[1., 0., 0., 1., 0., 0., 0., 1.],

[1., 0., 1., 0., 0., 1., 0., 0.],

[0., 1., 1., 0., 0., 0., 1., 0.],

[1., 0., 0., 0., 1., 1., 0., 0.],

[1., 0., 0., 1., 0., 1., 0., 0.],

[1., 0., 0., 1., 0., 1., 0., 0.],

[0., 1., 0., 1., 0., 1., 0., 0.],

[1., 0., 1., 0., 0., 0., 0., 1.],

[1., 0., 0., 0., 1., 1., 0., 0.],

[0., 1., 1., 0., 0., 0., 1., 0.],

[1., 0., 0., 0., 1., 0., 1., 0.],

[0., 1., 0., 0., 1., 1., 0., 0.],

[0., 1., 1., 0., 0., 0., 0., 1.],

[0., 1., 0., 0., 1., 0., 1., 0.],

[0., 1., 1., 0., 0., 0., 0., 1.],

[0., 1., 0., 0., 1., 1., 0., 0.],

[0., 1., 0., 0., 1., 1., 0., 0.],

[1., 0., 1., 0., 0., 1., 0., 0.],

[0., 1., 0., 1., 0., 0., 1., 0.],

[1., 0., 0., 1., 0., 0., 1., 0.],

[0., 1., 1., 0., 0., 1., 0., 0.],

[1., 0., 0., 1., 0., 0., 1., 0.],

[0., 1., 0., 0., 1., 0., 0., 1.],

[1., 0., 1., 0., 0., 0., 1., 0.],

[1., 0., 1., 0., 0., 1., 0., 0.],

[0., 1., 0., 1., 0., 0., 0., 1.],

[1., 0., 0., 0., 1., 0., 1., 0.],

[0., 1., 1., 0., 0., 0., 0., 1.],

[1., 0., 0., 0., 1., 1., 0., 0.],

[1., 0., 1., 0., 0., 0., 0., 1.],

[1., 0., 0., 0., 1., 0., 0., 1.],

[0., 1., 0., 0., 1., 0., 0., 1.],

[1., 0., 0., 0., 1., 1., 0., 0.],

[0., 1., 0., 1., 0., 0., 1., 0.],

[1., 0., 0., 0., 1., 1., 0., 0.],

[0., 1., 0., 0., 1., 1., 0., 0.],

[0., 1., 0., 0., 1., 0., 0., 1.],

[0., 1., 0., 1., 0., 0., 0., 1.],

[0., 1., 0., 1., 0., 0., 0., 1.]])



ohe.inverse\_transform([[0.,1.,1.,0.,0.,0.,1.,0.]])

array([['Female', 'School', 'Average']], dtype=object)



X = df[['Gender','Education','Review']]

ohe = OneHotEncoder(drop='first')

X = ohe.fit\_transform(X)

ohe.categories\_

[array(['Female', 'Male'], dtype=object),

array(['PG', 'School', 'UG'], dtype=object),

array(['Average', 'Good', 'Poor'], dtype=object)]



X.toarray().shape

(50, 5)



X.toarray()

array([[0., 1., 0., 0., 0.],

[0., 0., 1., 0., 1.],

[0., 0., 0., 1., 0.],

[0., 0., 0., 1., 0.],

[0., 0., 1., 0., 0.],

[0., 1., 0., 0., 0.],

[1., 1., 0., 1., 0.],

[0., 1., 0., 0., 1.],

[0., 0., 1., 0., 0.],

[1., 0., 1., 1., 0.],

[0., 0., 1., 1., 0.],

[1., 0., 1., 1., 0.],

[1., 1., 0., 0., 1.],

[0., 1., 0., 0., 0.],

[1., 0., 0., 0., 1.],

[1., 0., 1., 0., 1.],

[1., 0., 1., 0., 1.],

[0., 0., 1., 0., 1.],

[1., 1., 0., 1., 0.],

[1., 0., 0., 0., 1.],

[0., 1., 0., 0., 0.],

[1., 0., 0., 0., 0.],

[0., 0., 0., 0., 1.],

[0., 1., 0., 1., 0.],

[0., 0., 0., 0., 0.],

[0., 1., 0., 1., 0.],

[0., 0., 0., 0., 1.],

[0., 0., 0., 0., 1.],

[1., 1., 0., 0., 1.],

[0., 0., 1., 0., 0.],

[1., 0., 1., 0., 0.],

[0., 1., 0., 0., 1.],

[1., 0., 1., 0., 0.],

[0., 0., 0., 1., 0.],

[1., 1., 0., 0., 0.],

[1., 1., 0., 0., 1.],

[0., 0., 1., 1., 0.],

[1., 0., 0., 0., 0.],

[0., 1., 0., 1., 0.],

[1., 0., 0., 0., 1.],

[1., 1., 0., 1., 0.],

[1., 0., 0., 1., 0.],

[0., 0., 0., 1., 0.],

[1., 0., 0., 0., 1.],

[0., 0., 1., 0., 0.],

[1., 0., 0., 0., 1.],

[0., 0., 0., 0., 1.],

[0., 0., 0., 1., 0.],

[0., 0., 1., 1., 0.],

[0., 0., 1., 1., 0.]])