**Practical exercise**

**Run the following lines of code in Jupyter notebook environment using the csv file “salaries” provided. Then Save the file as “data\_preprocessing”**

**#importing libraries**

import numpy as np # contains mathematical tools

import matplotlib.pyplot as plt #for plotting charts

import pandas as pd #for importing and managing datasets

**#importing the datasets**

df=pd.read\_csv(‘salaries.csv’) #imports csv data file

**#DataFrame operations**

data1=df.head() #holds the first 5 rows

data2=df.tail() #holds the last 5 rows of dataframe

data3=df.shape #dimensions

data4=df.dtypes

data5=df.info()

data6=df.isnull().sum()

**#Slicing in Dataframe**

data7=df.iloc[0:5,:]

data8=df.iloc[:,0:5]

data9=df.iloc[:,:-1]

data10=df.iloc[:,0:5]

data11=df.iloc[:,5]

**#Getting the matrix of features (matrix of independent variables and dependent variable vector)**

x=df.iloc[: , :-1].values

y=df.iloc[:,5].values

**#Filling the missing values with mean**

from sklearn.impute import SimpleImputer #SciKit Learn contain many libraries for creating ML models, import a class SimpleImputer

imputer = SimpleImputer(strategy='mean', missing\_values=np.nan) # an object imputer of a class simpleImputer is created

imputer = imputer.fit(X[:,2:3]) #fitting an object imputer to matrix X

X[:,2:3] = imputer.transform(X[:,2:3]) #replaces the missing data of the matrix by mean of the column

**#Filling the missing values with mode**

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

imputer = imputer.fit(X[:,4:5]) #fitting an object imputer to matrix X

X[:,4:5] = imputer.transform(X[:,4:5]) #replaces the missing data of the matrix by mode of the column