**Placement Test**

**in Big Data Analytics**

Full Name: Blerta Hoxha

Student id:

**Question 1**

Write a Python program to create a Histogram by using the data that are provided in the following array:

data = [21,22,23,4,5,6,77,8,9,10,31,32,33,34,35,36,37,18,49,50,100]

Tips: a.Use matplotlib for creating the Histogram, b. Use 5 bins (intervals / classes)

**Answer**

import numpy as np

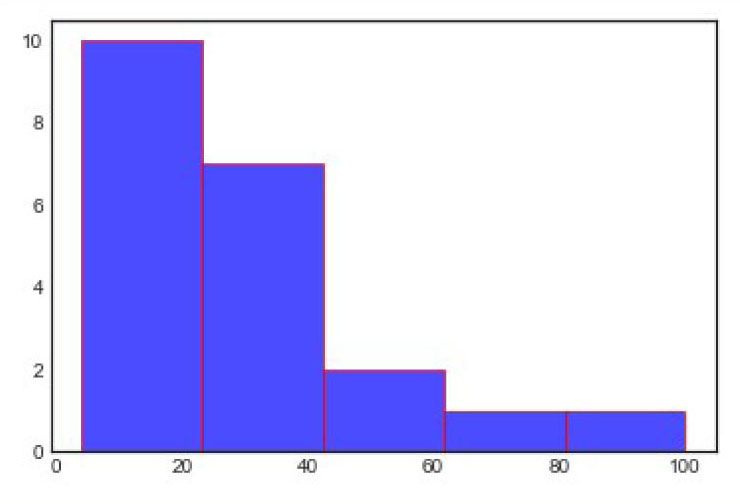
import matplotlib.mlab as mlab

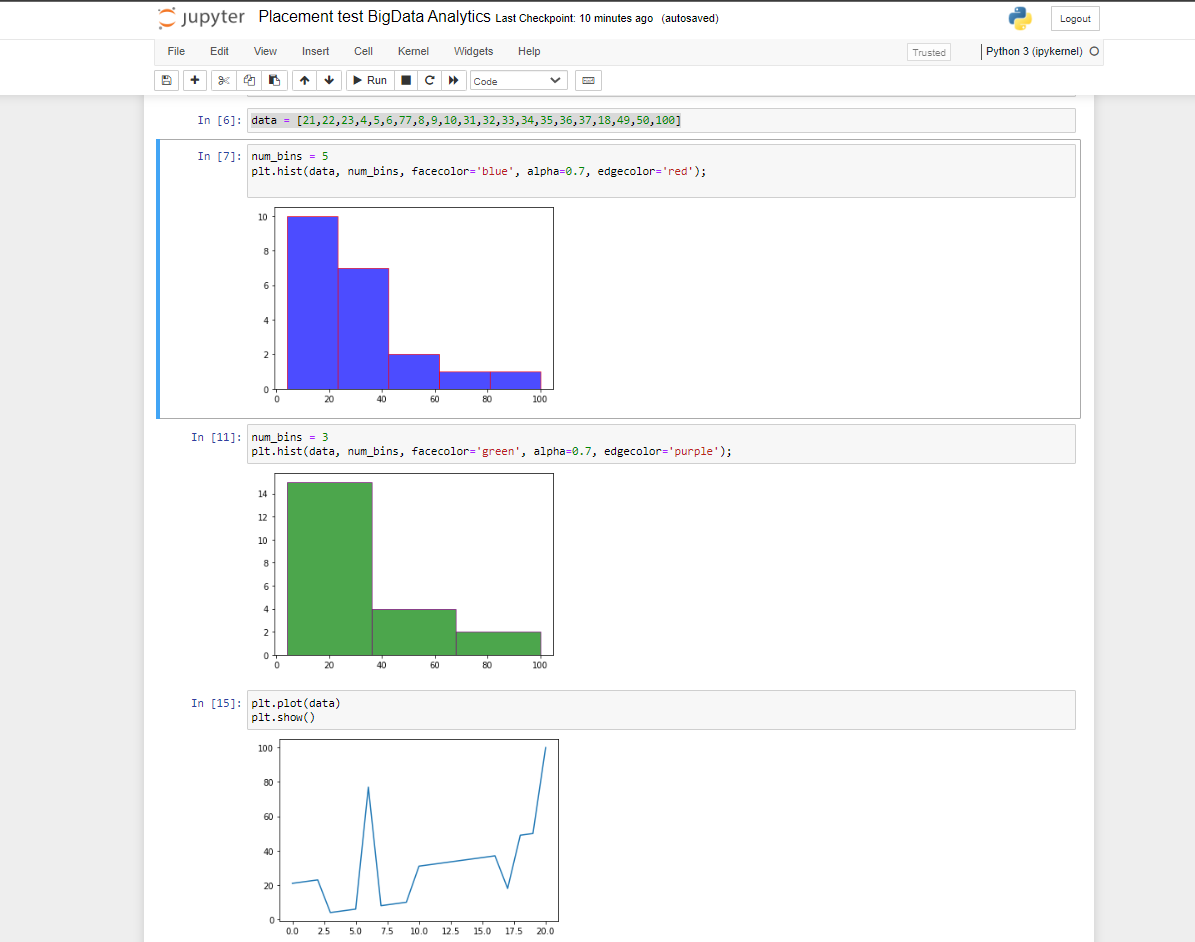
import matplotlib.pyplot as plt

data = [21,22,23,4,5,6,77,8,9,10,31,32,33,34,35,36,37,18,49,50,100]

num\_bins = 5

plt.hist(data, num\_bins, facecolor='blue', alpha=0.7, edgecolor='red');

****

****

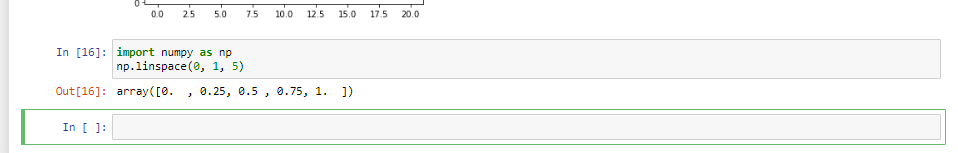
**Question 2**

Create an array of 5 values evenly spaced between 0 and 1 by using NumPy and linspace

**Answer**

import numpy as np

np.linspace(0, 1, 5)

****

**Question 3**

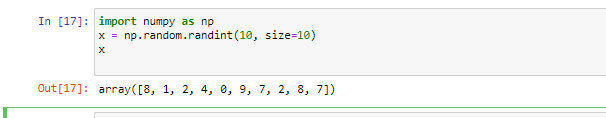
Create a NumPy array of size 10, populated with random integer numbers that take value in the interval [0, 9]

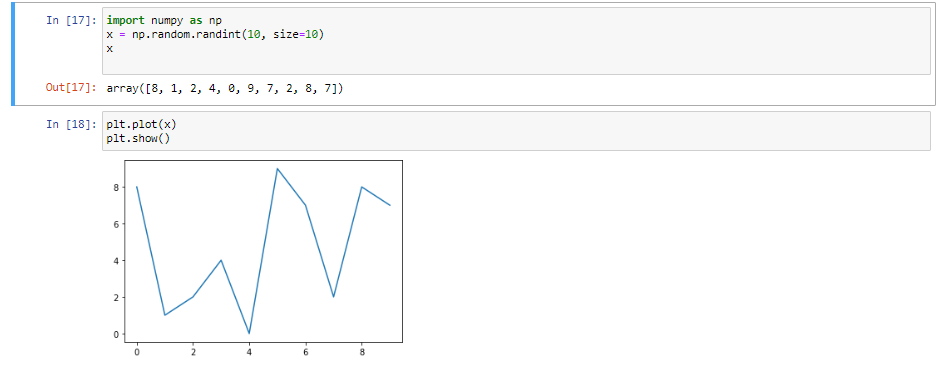
**Answer**

import numpy as np

x = np.random.randint(10, size=10)

x





**Question 4**

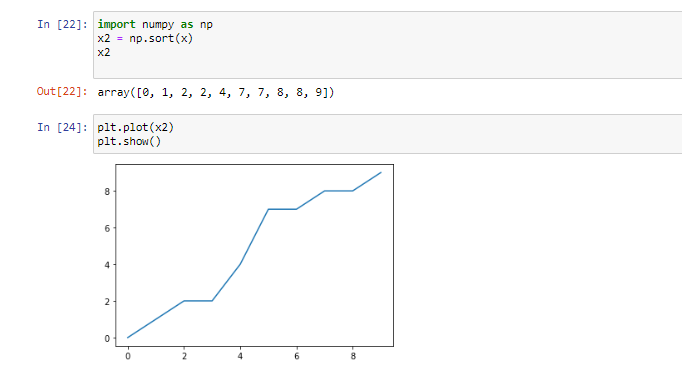
Sort from min to max value the previously generated in Question 3, NumPy array

**Answer**

import numpy as np

x2 = np.sort(x)

x2



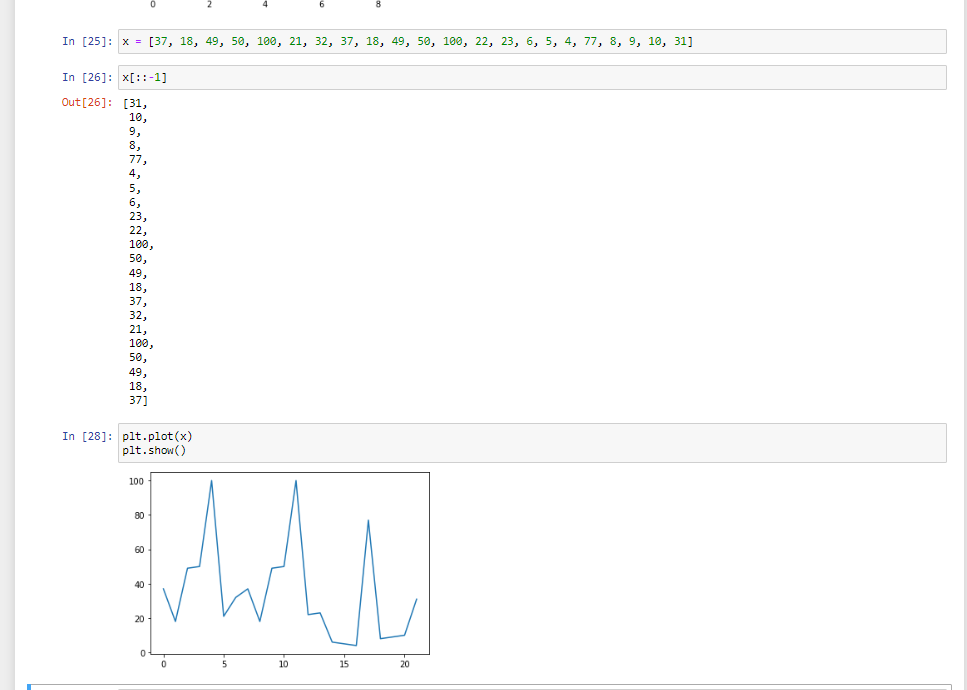
**Question 5**

Reverse the elements of the following array:

x = [37, 18, 49, 50, 100, 21, 32, 37, 18, 49, 50, 100, 22, 23, 6, 5, 4, 77, 8, 9, 10, 31]

**Answer**

x[::-1]



**Question 6**

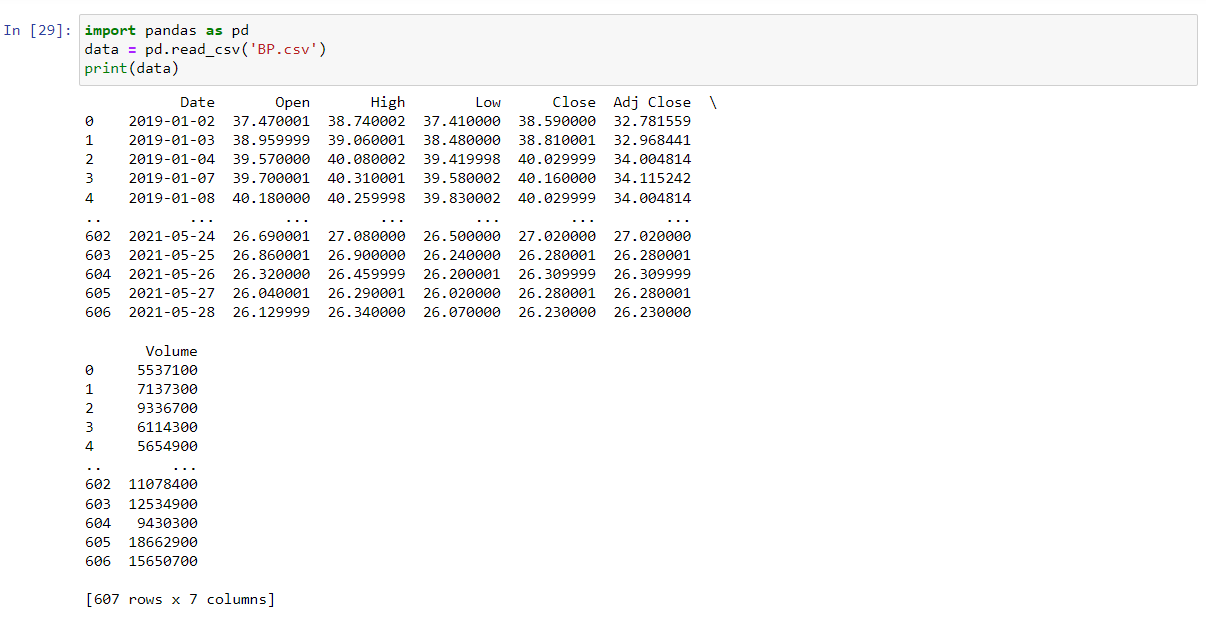
Read the csv file that is provided (BP.csv) using Pandas

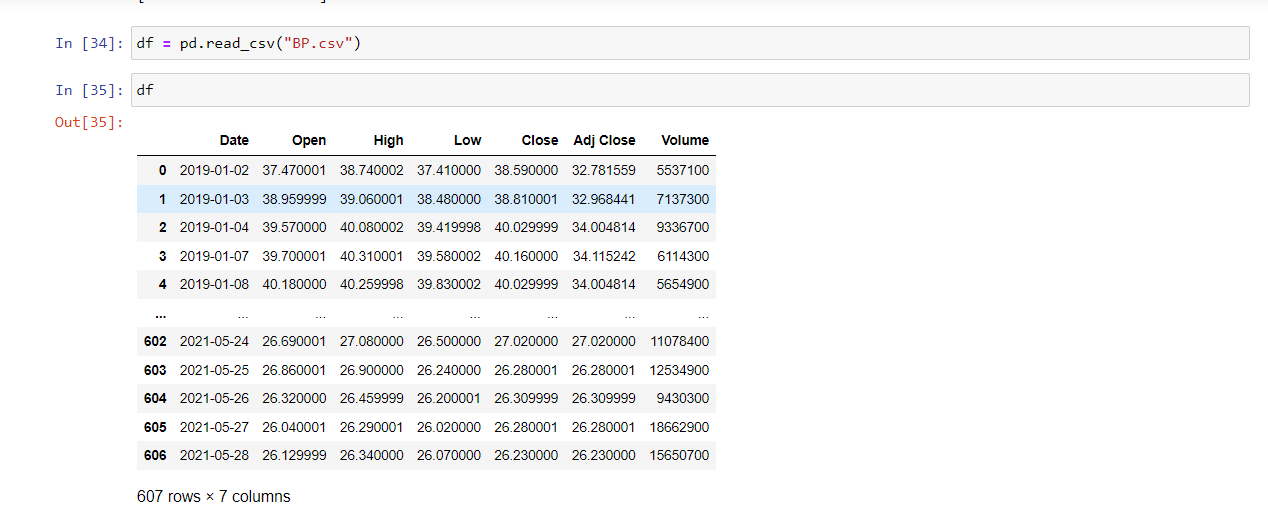
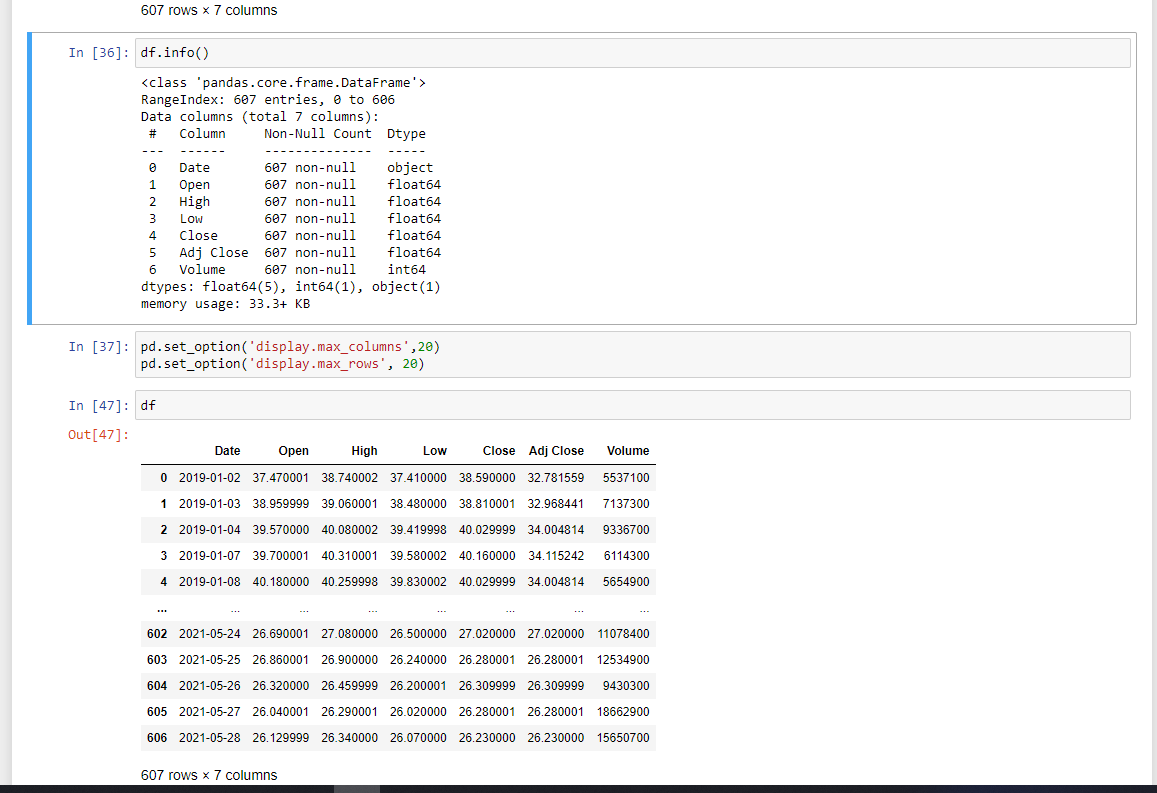
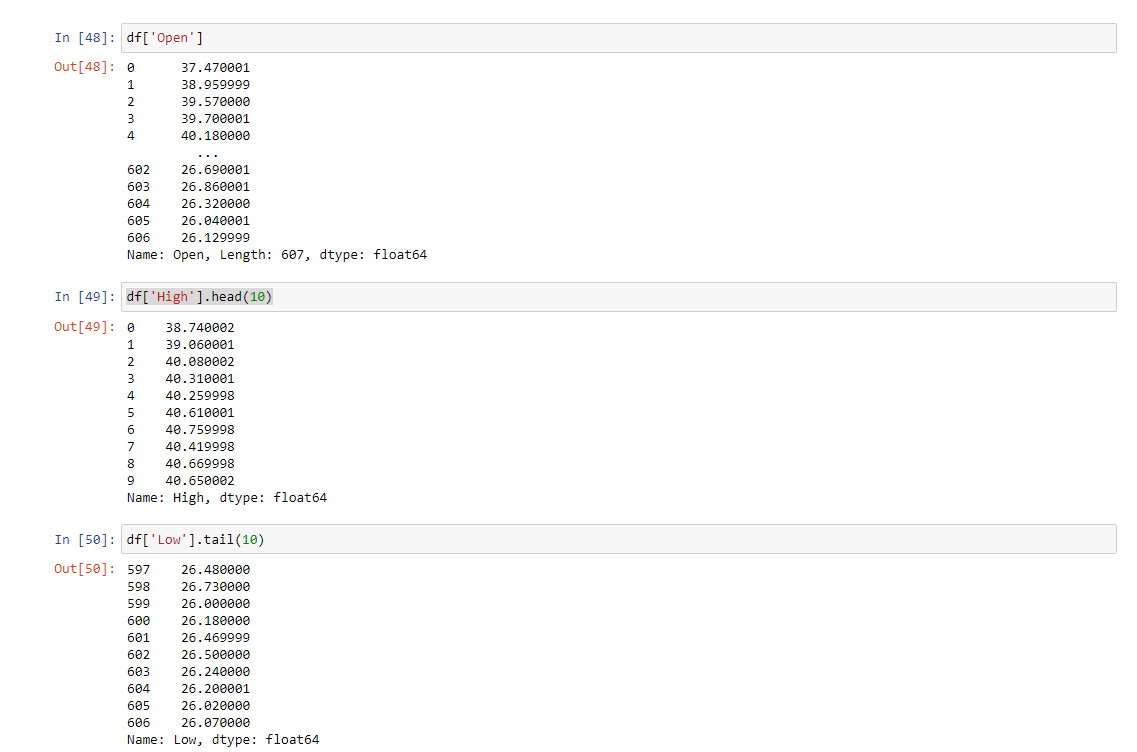
**Answer**

import pandas as pd

data = pd.read\_csv('c:\python\_code\class\_material\BP.csv')

print(data)



**Question 7**

Read only the following 2 columns: ‘Date, Adj Close from the csv file that is provided (BP.csv) using Pandas

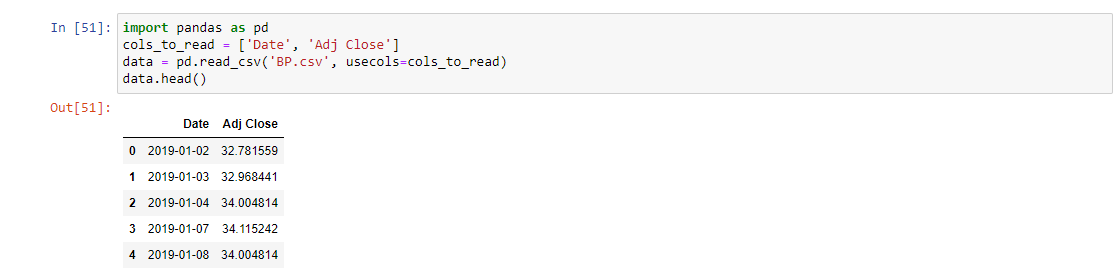
**Answer**

import pandas as pd

cols\_to\_read = ['Date', 'Adj Close']

data = pd.read\_csv('c:\python\_code\class\_material\BP.csv', usecols=cols\_to\_read)

data.head()

****

**Question 8**

Read the first 10 lines of the csv file that is provided (BP.csv) using Pandas

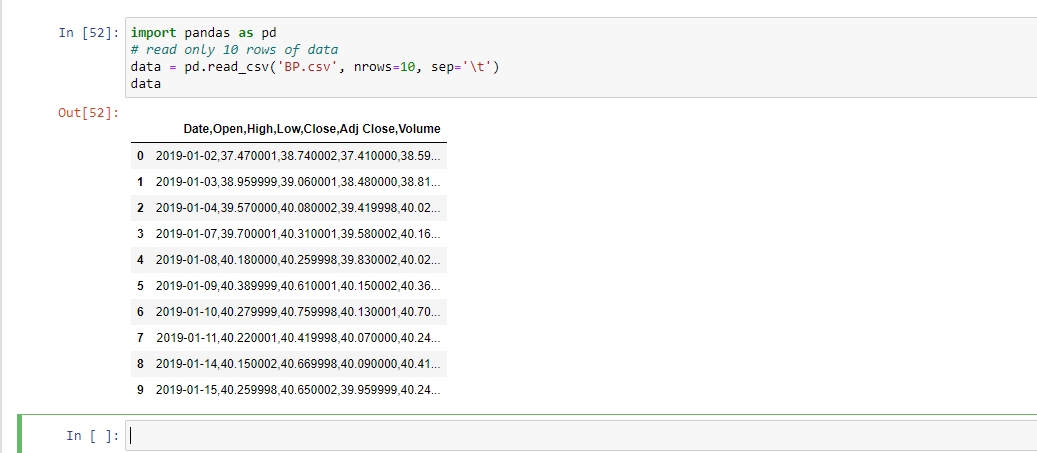
**Answer**

import pandas as pd

# read only 10 rows of data

data = pd.read\_csv('c:\python\_code\class\_material\BP.csv', nrows=10, sep='\t')

data



**Question 9**

Set column “High” as index in the csv file that is provided (BP.csv) using Pandas

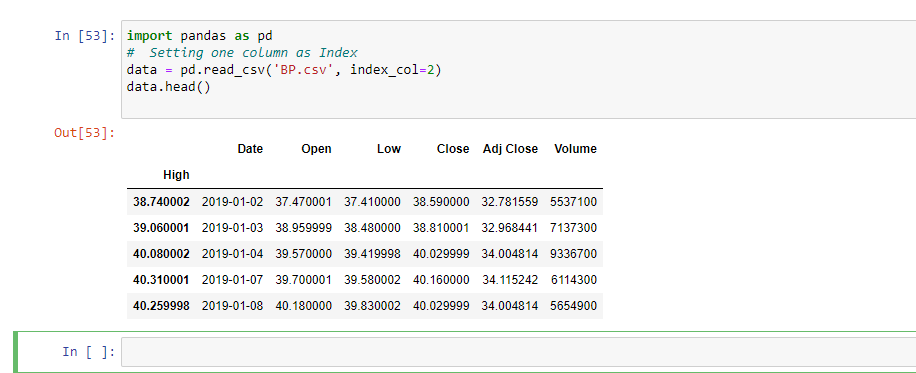
**Answer**

import pandas as pd

# Setting one column as Index

data = pd.read\_csv('c:\python\_code\class\_material\BP.csv', index\_col=2)

data.head()



**Question 10**

Read the columns “Open”, “Close” and “Volume” of the csv file that is provided (BP.csv) using Pandas. Furthermore, calculate the following statistics, only for the column “Open”

* Mean
* Standard deviation
* Minimum
* Maximum
* Index of Minimum
* Index of Maximum
* 25th percentile
* Median
* 75th percentile

**Answer**

import pandas as pd

import numpy as np

# only read selected columns

cols\_to\_read = ['Open', 'Close', 'Volume']

data = pd.read\_csv('c:\python\_code\class\_material\BP.csv', usecols=cols\_to\_read)

open\_price = np.array(data['Open'])

# print(open\_price)

print("Mean open\_price: ", open\_price.mean())

print("Standard deviation:", open\_price.std())

print("Minimum open\_price: ", open\_price.min())

print("Maximum open\_price: ", open\_price.max())

print("Index of Minimum open\_price: ", open\_price.argmin())

print("Index of Maximum open\_price::", open\_price.argmax())

print("25th percentile: ", np.percentile(open\_price, 25))

print("Median: ", np.median(open\_price))

print("75th percentile: ", np.percentile(open\_price, 75))



**Question 11**

Read the “Open” prices of the csv file that is provided (BP.csv) using Pandas and create the Histogram of “Open” prices by using matplotlib. Moreover, provide a title for the histogram and name the x and y labels

**Answer**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn; seaborn.set() # set plot style

data = pd.read\_csv('c:\python\_code\class\_material\BP.csv')

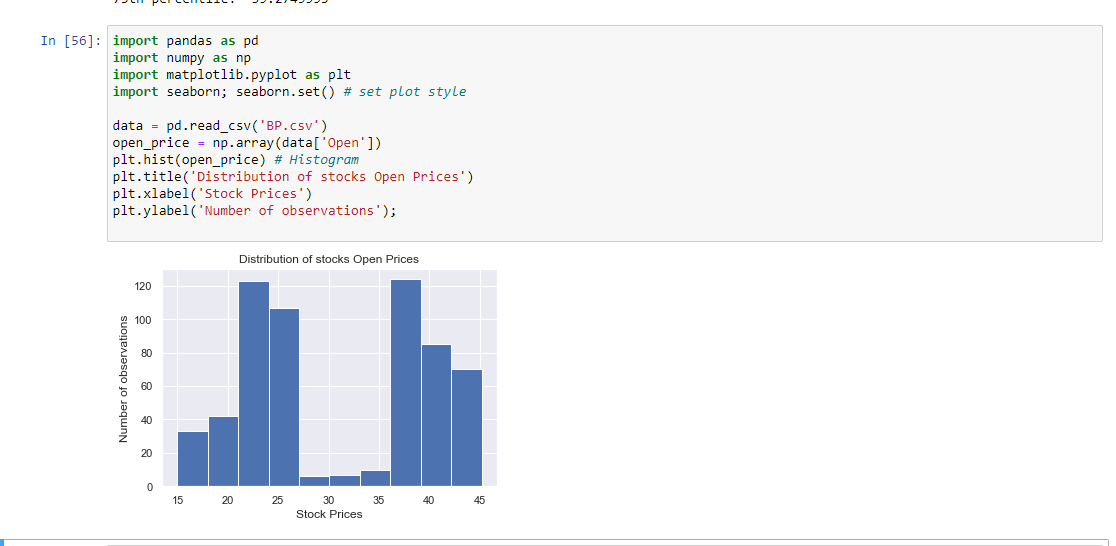
open\_price = np.array(data['Open'])

plt.hist(open\_price) # Histogram

plt.title('Distribution of stocks Open Prices')

plt.xlabel('Stock Prices')

plt.ylabel('Number of observations');



**Question 12**

Remove the 4th item from following list and add the same item in 2nd index

x = [34, 54, 67, 89, 11, 43, 94]

**Answer**

print("Original list ", x)

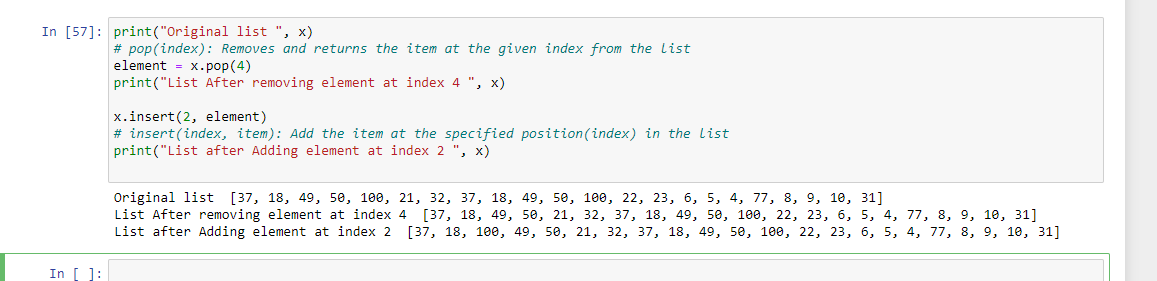
# pop(index): Removes and returns the item at the given index from the list

element = x.pop(4)

print("List After removing element at index 4 ", x)

x.insert(2, element)

# insert(index, item): Add the item at the specified position(index) in the list

print("List after Adding element at index 2 ", x) 

**Question 13**

Count the occurrence of each element from the list x = [11, 45, 8, 11, 23, 45, 23, 45, 89]

**Answer**

# Count the occurrence of each element from a list

x = [11, 45, 8, 11, 23, 45, 23, 45, 89]

print("Original list ", x)

# create a dictionary to show the count of each element

count\_occurrence = dict()

for item in x:

if item in count\_occurrence:

count\_occurrence[item] += 1

else:

count\_occurrence[item] = 1

print("Printing count of each item ", count\_occurrence)

