**#project4**

#Example data has been downloaded from the open access Human Gene Expression Atlas and represents typical data bioinformaticians work with.

#It is "Transcription profiling by array of brain in humans, chimpanzees and macaques, and brain, heart, kidney and liver in orangutans" experiment in a tab-separated format.

**#importing numpy package**

import numpy as np

**#import matplotlib.pyplot**

#matplotlib.pyplot is a collection of command style functions that make matplotlib work like MATLAB. Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc.

import matplotlib.pyplot as plt

**#from scipy.spatial.distance import pdist, squareform**

from scipy.cluster.hierarchy import dendrogram

from fastcluster import \*

**#load data which is in textfile format columes from 1 to 31, data type as float and delimiter as new line**

data = np.genfromtxt("ExpRawData-E-TABM-84-A-AFFY-44.tab",names=True,usecols=tuple(range(1,32)),dtype=float, delimiter="\t")

**#printing the length of data**

print (len(data))

**#printing the length of data.dtype.names**

print (len(data.dtype.names))

**#printing labels data.dtype.names**

print (data.dtype.names)

**#creating a new view of data**

data\_array = data.view((np.float, len(data.dtype.names)))

**#transposing the data\_array matrix**

data\_array = data\_array.transpose()

**#printing the data\_array**

print (data\_array)

**#data\_dist = pdist(data\_array)**

**# computing the distance**

data\_link = linkage(data\_array,method="single",metric="euclidean")

**#print the data linkage**

print (data\_link)

**# computing the linkage**

**#creating a dendrogram with lables data.dtype.names**

dendrogram(data\_link,labels=data.dtype.names)

**#plot X-axis name as samples**

plt.xlabel('Samples')

**#plot Y-axis name as distance**

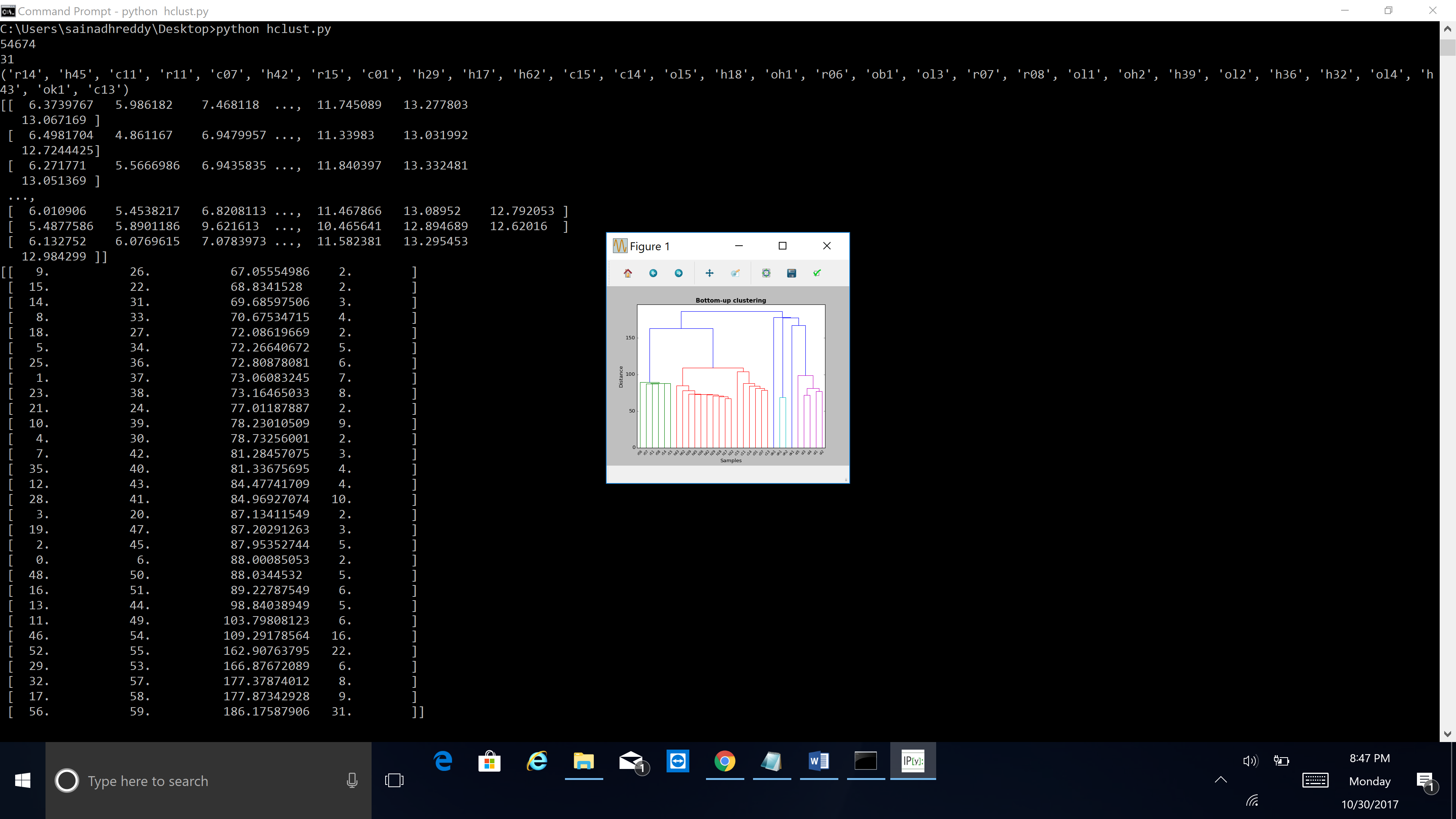
plt.ylabel('Distance')

**#title up the bottom of clustering with fontwieght and fontsize**

plt.title('Bottom-up clustering', fontweight='bold', fontsize=14);

**#display the figure**

plt.show()



4.

**output explanation**

(a). h17 and h32 are grounded together, 2 samples are in this group and 67.05554986 is the distance between the 2 groups.

(b). oh1 and oh2 are grouped together, 2 samples are in this new group and 68.8341528 is the distance between the 2 groups.

(c). h18 and h17h32 are grouped together, 3 samples are in this new group and 69.68597506 is the distance between 2 groups.

(d). h29 and h18h17h32 are grouped together, 4 samples are in this new group and 70.67534715 is the distance between 2 groups.