Homework 2:

Answer:

```
import pandas as pd
from sklearn import preprocessing
import numpy as np
pd.set_option('display.precision',4)
from scipy.spatial import distance
from scipy.spatial.distance import pdist, squareform
#data extraction and describe the properties of the data
df_red = pd.read_csv('E:\ECE 657A\winter 2019\homework\hw2\winequality-red.csv',sep=";")
df_white = pd.read_csv('E:\ECE 657A\winter 2019\homework\hw2\winequality-white.csv', sep=";")
des_red= df_red.describe()
des_white= df_white.describe()

#saving the properties in excel
writer = pd.ExcelWriter('output.xlsx')
des_red.to_excel(writer,'Sheet1')
des_white.to_excel(writer,'Sheet2')
writer.save()

#histogram the data
#df_red.hist()
#df_white.hist()

#extracting first 10 rows
df_red2=df_red[0:10]
df1 = df_red2.loc[:,'fixed acidity':'alcohol']
print(df1)
```

```
x = df1.values #returns a numpy array
```

#anwer to ques no: 1

```
#z-core scaling of data
std_scale = preprocessing.StandardScaler().fit(x)
df_std = std_scale.transform(x)
df_std =pd.DataFrame(df_std)
des1=df_std.describe()
print(des1)
des1.to_excel(writer,'Sheet3')
writer.save()

#analysis only showing for wine red
#min-max scaling of data
minmax_scale = preprocessing.MinMaxScaler().fit(x)
df_minmax = minmax_scale.transform(x)
df_minmax = pd.DataFrame(df_minmax)
des2=df_minmax.describe()
print(des2)
des2.to_excel(writer,'Sheet4')
writer.save()


#mean-subtracted scaling of data
meansub_scale=df1-df1.mean()
des3=meansub_scale.describe()
```

```
print(des3)
des3.to_excel(writer,'Sheet5')
writer.save()

#similar analysis can be carried out for wine white
```

#answer to question no:2

```
#red wine data
#converting data into matrices
x1=df_std.values
x2=df_minmax.values
x3= meansub_scale.values

#calculating distance matrix
y1= pdist(x1,'euclidean')
y2=squareform(y1)

#minimum of each array from distance matrix
t1= y2.min(axis=1)

#maximum of each array from distance matrix
t2= y2.max(axis=1)

#similar can be done for manhattan distance and cosine distance

#repeat the same for white wine data
```