MCKV INSTITUTE OF ENGINEERING

243, GT ROAD (NORTH), LILUAH, HOWRAH-711204

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

PAPER CODE: - PC-CS591 PAPER: - IT WORKSHOP

Assignment on Matplotlib

1. Take a data set in a data frame and plot the bar chart example may be year of Production and

the Production Quantity. (Take help of Pandas)

2. Take a data set from the above data frame which you have prepared and plot the pie

chart. (Take help of Pandas)

3. Take a csv file with name, employee id and salary. Read the datas of salary and plot the bar

Chart with employee name on the x axis and employee salary on the y-axis. Just take 5 records in the csv files.

4. Take the dirtydata.csv as an input and plot the box-plot for the column duration and find out How many outliers are there?

5. What do you mean by multiple boxplots? Take a data set in a data frame and draw the multiple boxplots. Make a csv file with multiple parameters and then plot multiple box plots for each parameter.

Assignment on Seaborn

1. Write a Python program to create a plot to get a general Statistics of Iris data. (Using the

dataset “iris.csv”).

2. Write a Python program to create a graph to find relationship between the sepal length

and width. (Using the dataset “iris.csv”).

3. Write a Python program to create a join plot using "hexbin" to describe individual

distributions on the same plot between Sepal length and Sepal width. (Using the dataset

“iris.csv”).

4. Write a Python program to create a joinplot and add regression and kernel density fits

using "reg" to describe individual distributions on the same plot between Sepal length

and Sepal width. (Using the dataset “iris.csv”).

5. Write a Python program to find the correlation between variables of iris data. Also create

a heatmap using Seaborn to present their relations (Using the dataset “iris.csv”).

Assignment on NumPy

1. Take a 3X3 matrix randomly and another matrix by assigning elements. Now add the 2 matrix and store it in a separate matrix C.

2. Take 3X3 matrixes add it with another 3x3 of all 1 matrix and convert that to a complex data type matrix.

3. Take a 3X3 matrix and multiply element wise with 3X 3 another matrix.

4. Multiply 2 3X3 matrixes.

5. Randomly generate 10 numbers in between 1 to 10 and convert that to a 5X2 matrix. Take a floating-point matrix and perform the addition. a= floating point matrix and b is an integer matrix of same dimension. Why a+=b is not equal to b+=a. Explain briefly. How can we resolve this problem?

6. Take x values and plot the cos(x) and tan(x) in a graph using matplotlib.

7. Take a multidimensional array (3, 3, 3) and print the last column.

8. Take a function f(x) =x^3 + 5\*y +4\*z and determine the values and store them in a (3,3,3) matrix.

9. Using axis add the column values of a 3X3 matrix and then add it with the maximum values of rows taken in a 3X 3 matrix

10. f(x) = x^3 + 5y and store it in (3,3) matrix.

11. Take a function array 'a' cube with a range of 20 and find out what will be the value of a [[7,8], [9,11]]. If we take values a [[7, 8], [9, 21]] will it take if not why?

12. Take a random number from 0 to 19 and make a 4X5 matrix then find the values of (i) 3rd row and 4th column only and (ii) only 4th column values.

13. Take a matrix of 5X4 randomly and create 2 3x3 values i and j respectively and take a tuple named m with i and j as an argument and generate the values of the matrix for the specified tuple. i<=4 and j<=3.

14. Take a matrix of 3x3 and find out the Eigen vector and Eigen values of that matrix.

15. Take a matrix and by taking the i, j or x, y values implement the hstack and vstack methods.

Assignment on Pandas

1. Create 3 indexes with name, roll number, fees amount in a vector and print the data frame by taking values in indexes.

2. Read mba.csv file and just fetch the name and fees and print it with the help of data frame.

3. Fetch your answer for question 2 with the help of Pandas Series.

4. Copy the datas from csv file dirtydata.csv and replace all the null calories value to 120 and

also try to change the null date format to a valid date.

5. Take a data frame df2 and replace all Nan rows with the mode value for the column calories. Data set is dirtydata.

6. Take df2 and for all the index values for the column calories if it is > 250 replace the calorie value with 250.