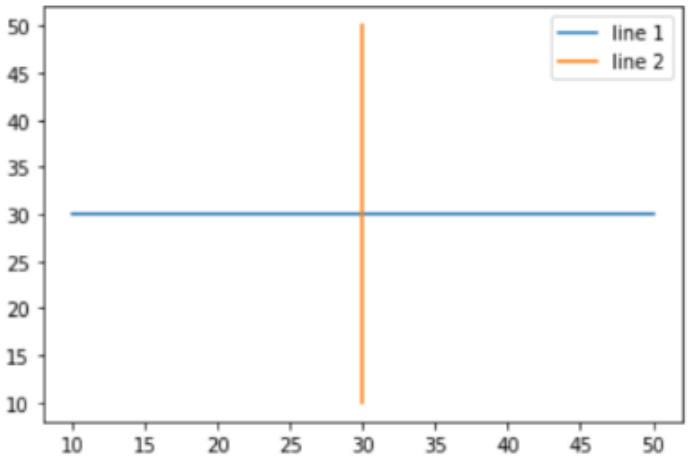
Year: B. Tech III (Semester VI)
Data Analysis and Visualization (BTIT13601)
List of Experiments

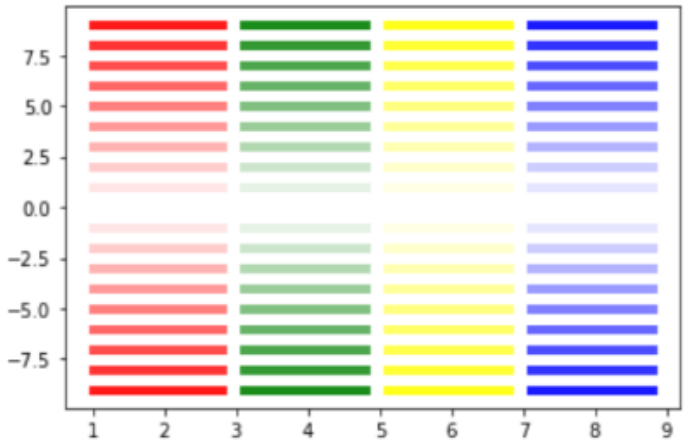
1. Create .xls file as dataset containing student records. Dataset should contain following information.
    a. Sr. No.
    b. Enrolment Number
    c. Department Name
    d. Student Name
    e. Current semester
    f. Email ID
    g. Mobile Number
    h. Current SPI
    i. CPI
    j. Professional Elective
    k. TOE Name
    l. CNS Marks
    m. SE Marks
    n. DMBI Marks
    o. WT Marks
    p. EO Marks

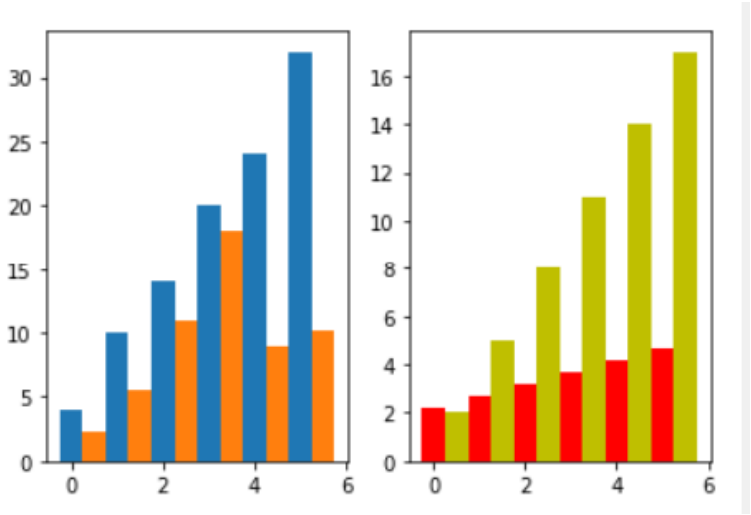
    Create at least 20 records for the dataset and perform following operations.
    a. Use Sum function to get **Grand Total** of all subject marks.
    b. Find percentage with your formula for all 5 subjects.
    c. Find students is PASS or FAIL with appropriate function from percentage.
    d. Find Class with Nested if function (use marks > 75 = “distinction”, >=60=” FirstClass”, >=35 = “Second Class” else “Fail”

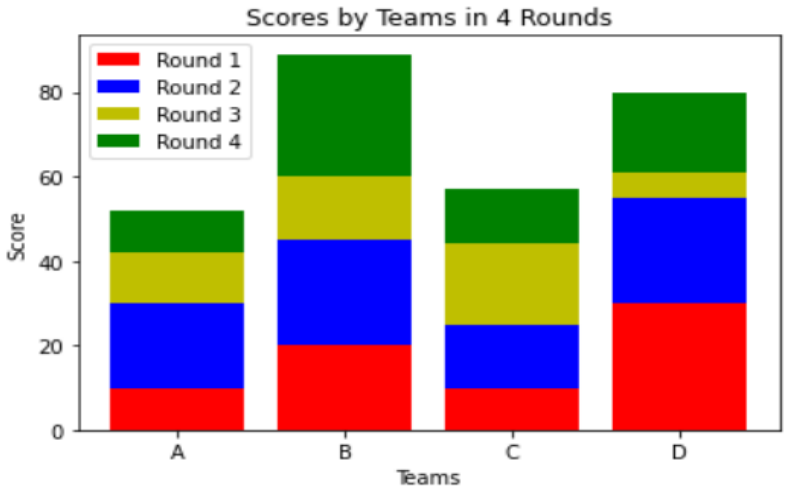
2. Create and initialize lists, tuples, sets, and dictionaries in python also perform operations like indexing, slicing, adding, removing, and sorting elements.
3. Perform experiments using NumPy.
4. Explore Pandas Data Structures.
5. Write a program for Data Loading, Storage and File Formats.
6. Perform experiments based on Interacting with Web APIs.

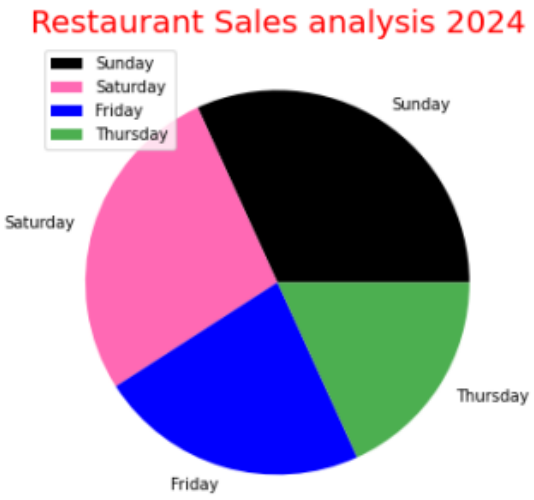
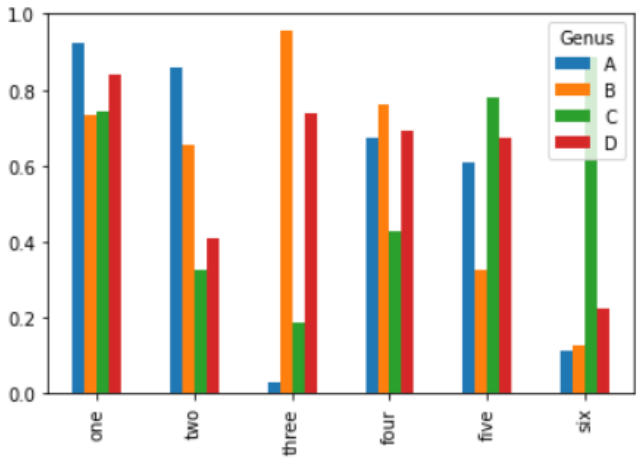
7. For the given input values of x and y, compute coefficient values to build linear regression model and plot it. Use this model to predict the value of y for the new value of x. Also compute MSE and RSS. (Note: Use only numpy library)
8. For the given input values of x and y, build a linear regression model using scipy library and plot it.
9. Generate at least 100 random values of x and compute the corresponding y values with appropriate linear function. For these pairs of (X,Y), generate model using both of the above technique and compare them. Also find co-relation coefficient (R-value), Probability value (P-value), and standard error. Try to modify the values in (X, Y) and identify its effect on the R-value, P-value and standard error.
10. Build a linear regression model that predicts house prices based on square footage.
11. Build a linear regression model using SKlearn library for the following :
    a. Take input values in terms of lists (X,Y).
    b. Generate training dataset and testing dataset from the input lists using train_test_split() from SKlearn.
    c. Train the model using training dataset.
    d. Predict Y for test dataset.
    e. Draw the plot.
    f. Repeat step (b)-(e) for random pairs of (X,Y).
12. Train a multiple linear regression model from SKlearn library for the California housing dataset. Also perform the steps (b)-(e) from the above practical(5).
13. Train a logistic regression model from SKlearn library for classification of data of the Iris dataset. Also perform the steps (b)-(e) from the above practical(5). Also display various metrics – Accuracy, Precision_score, recall_score, f1_score.
14. Perform Bias Variance decompositions for the datasets – California housing, Iris and randomly generated data.
15. Implement k-nearest neighbour algorithm.
16. Generate the following plots using matplotlib library of python.


line 1
line 2
50
45
40
35
30
25
20
15
10
10
15
20
25
30
35
40
45
50


7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
1
2
3
4
5
6
7
8
9


30
25
20
15
10
5
0
0
2
4
6
16
14
12
10
8
6
4
2
0
0
2
4
6





17. Generate the following plots from the dataframe/series.