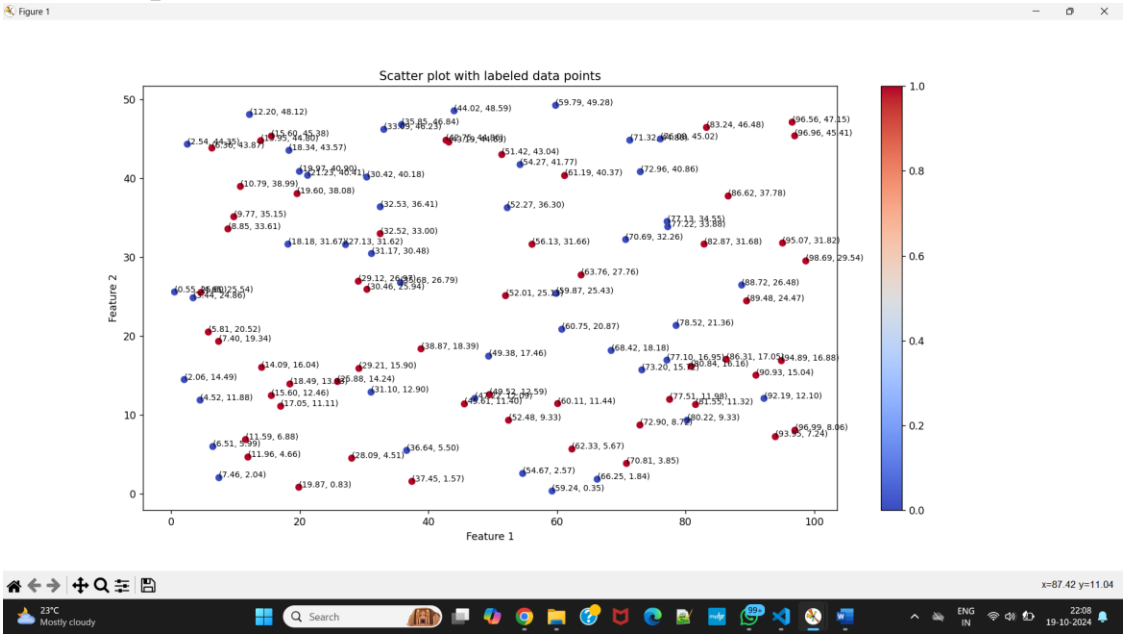
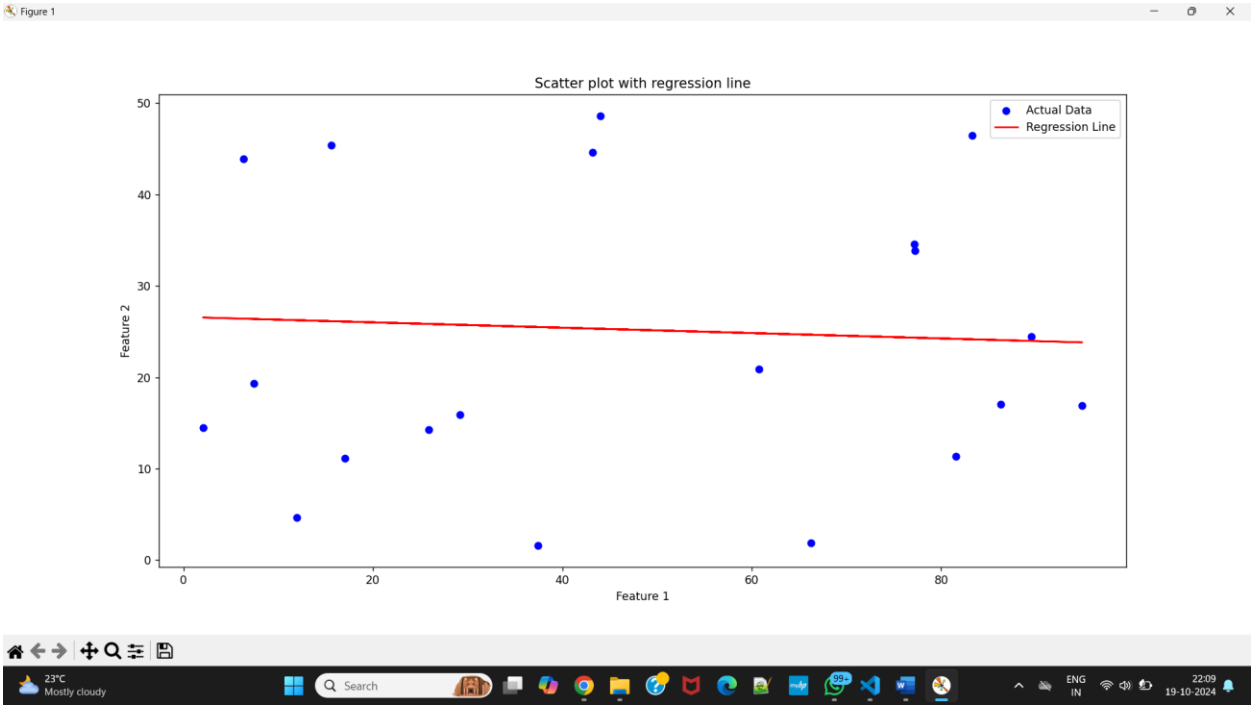
1. Task Description

Task : Advanced Data Visualization with Matplotlib

Combine Matplotlib with pandas to create advanced data visualizations with labeled data points and statistical overlays.

2. Task Output Screenshot





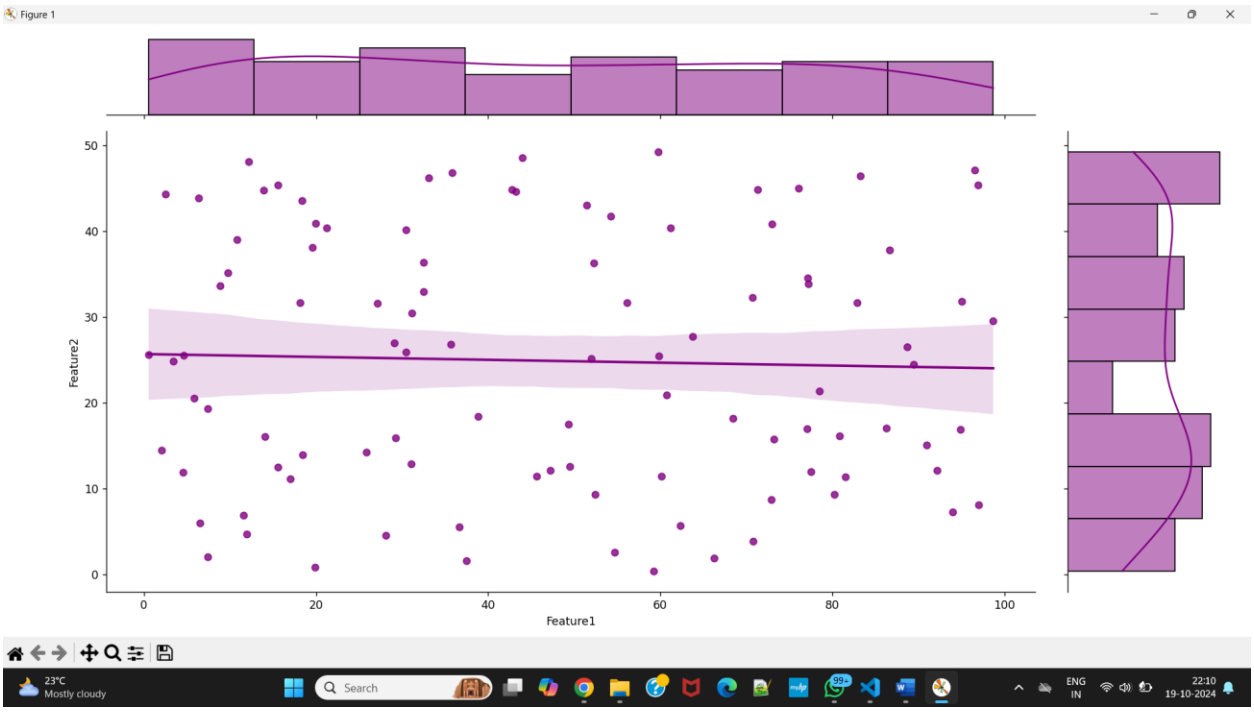


Figure 1
Feature2
Feature1
0
10
20
30
40
50
0
20
40
60
80
100
23°C
Mostly cloudy
Search
ENG
IN
22:10
19-10-2024

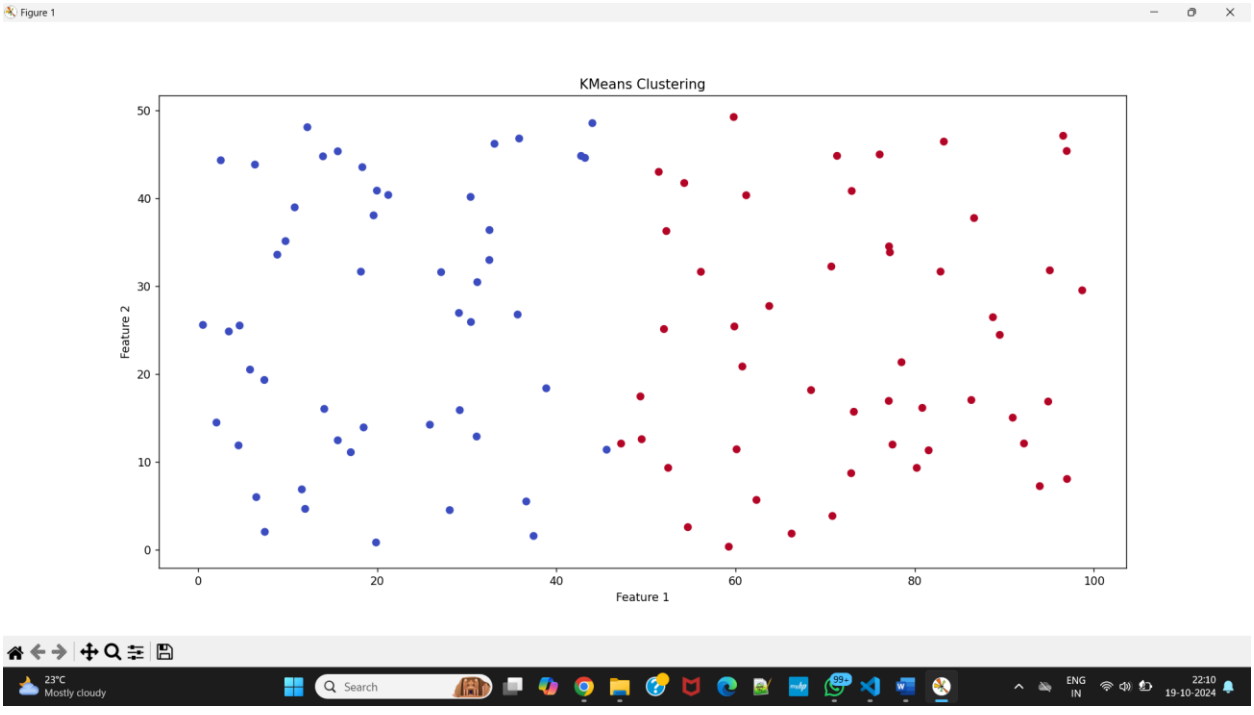


Figure 1
KMeans Clustering
Feature 2
Feature 1
0
10
20
30
40
50
0
20
40
60
80
100
23°C
Mostly cloudy
Search
ENG
IN
22:10
19-10-2024

3. Algorithms Used In Task :

1) Linear Regression (From scikit-learn)

Purpose: Linear regression is a supervised learning algorithm used to model the relationship between a dependent variable (target) and one or more independent variables (features). The goal is to find the best-fit line that minimizes the error between predicted and actual values.

Usage in the Code:

a) The algorithm is used to fit a linear regression model that predicts Feature2 based on Feature1.

b) We split the data into training and testing sets, train the model, and then plot the regression line along with the actual data points to visualize the relationship.

2) **KMeans Clustering (From scikit-learn)**

Purpose: KMeans is an unsupervised learning algorithm that partitions data into k clusters, where each data point belongs to the cluster with the nearest mean (centroid). The algorithm iteratively refines these clusters by minimizing the distance between data points and the centroid.

Usage in the Code:

a) The algorithm is applied to cluster the data into 2 groups based on Feature1 and Feature2.

b) We then plot the data points with colors corresponding to their cluster, visualizing the separation of clusters.