Assignment 4:

Githublink :

https://github.com/hasomaiy/machine_learning_problems/blob/cf317d3dc6eabfe97dccba497d1b44f045713e90/Assignment4.ipynb

Video Link:-

https://drive.google.com/drive/folders/1YKNgD7AvnJUVD_S0cJF4GGHso5yOp4Et?usp=sharing

Programming elements:

Linear Regression, K-Means Clustering and Data Analysis In class programming:

1. Apply Linear Regression to the provided dataset using underlying steps.
a. Import the given “Salary_Data.csv”
b. Split the data in train_test partitions, such that 1/3 of the data is reserved as test subset.
c. Train and predict the model.
d. Calculate the mean_squared error
e. Visualize both train and test data using scatter plot.

2. Apply K means clustering in the dataset provided:
• Remove any null values by the mean.
• Use the elbow method to find a good number of clusters with the K-Means algorithm
• Calculate the silhouette score for the above clustering

3. Try feature scaling and then apply K-Means on the scaled features. Did that improve the Silhouette score? If Yes, can you justify why