Fall 2022: CS5710

Machine Learning In-Class

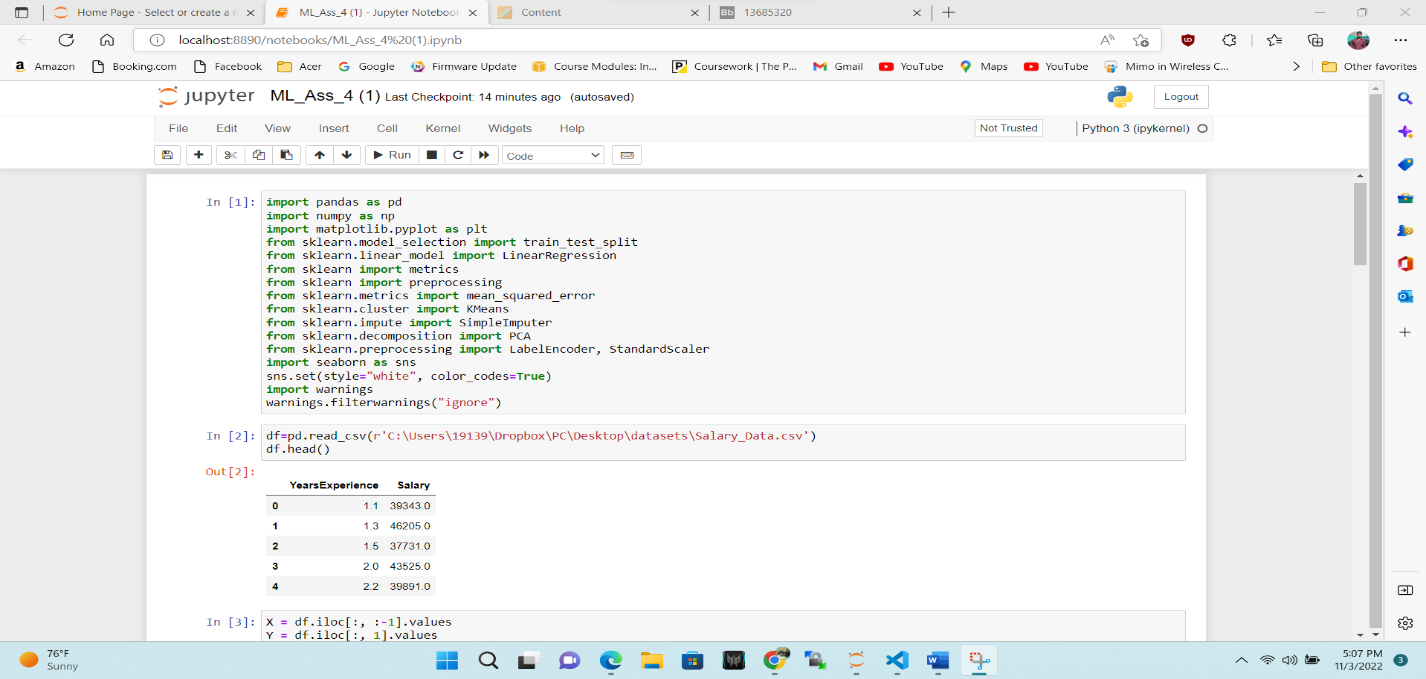
Programming Assignment-2

NAME: NEERAJ REDDY ALLAM

ID:700729065

1. Apply Linear Regression to the provided dataset using underlying steps.

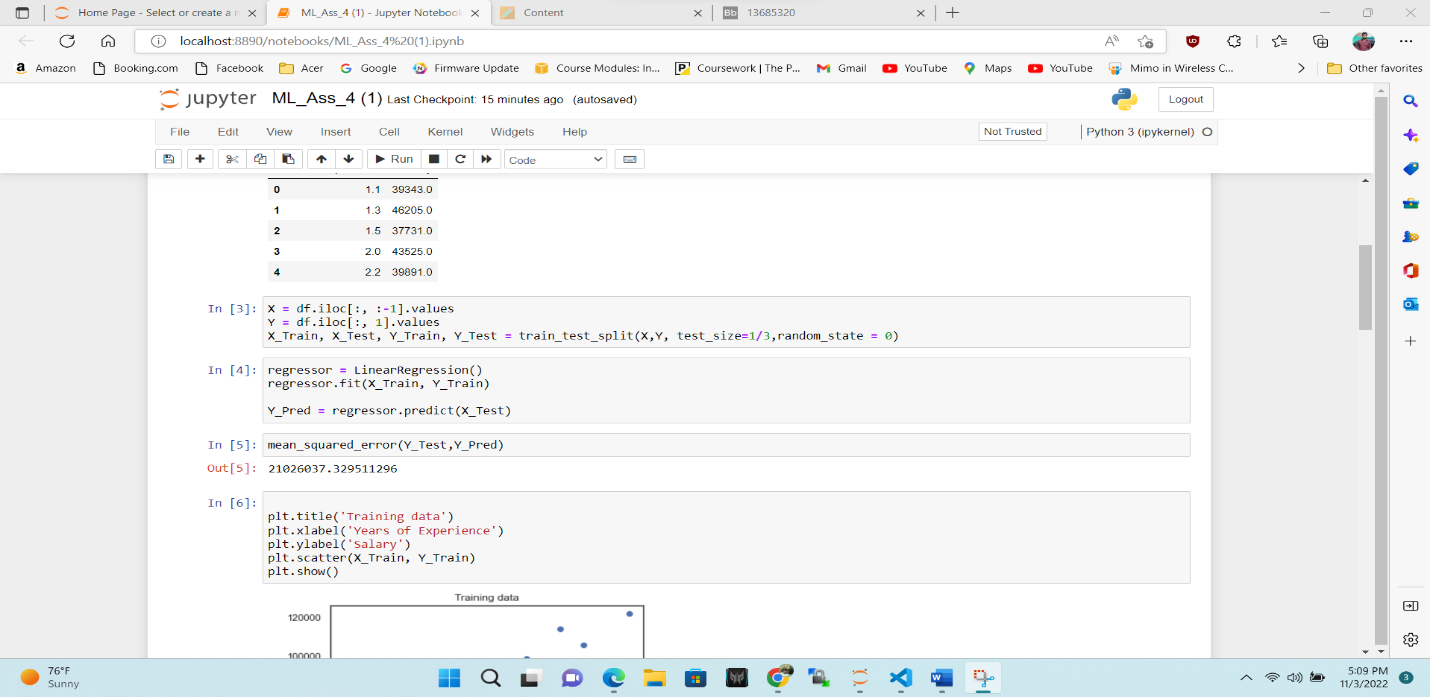
a. Import the given “Salary\_Data.csv”



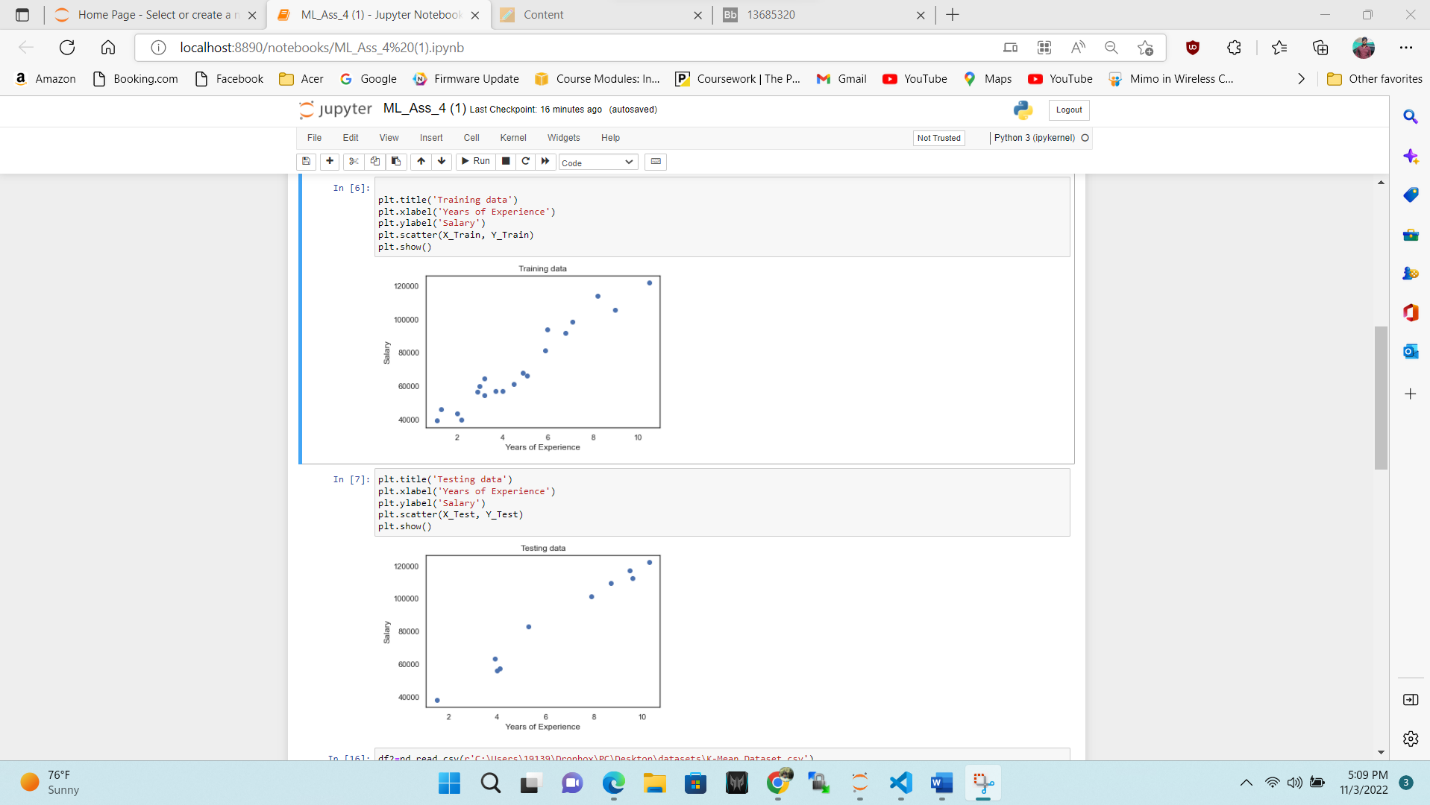
b. Split the data in train\_test partitions, such that 1/3 of the data is reserved as test subset.

c. Train and predict the model.

d. Calculate the mean\_squared error

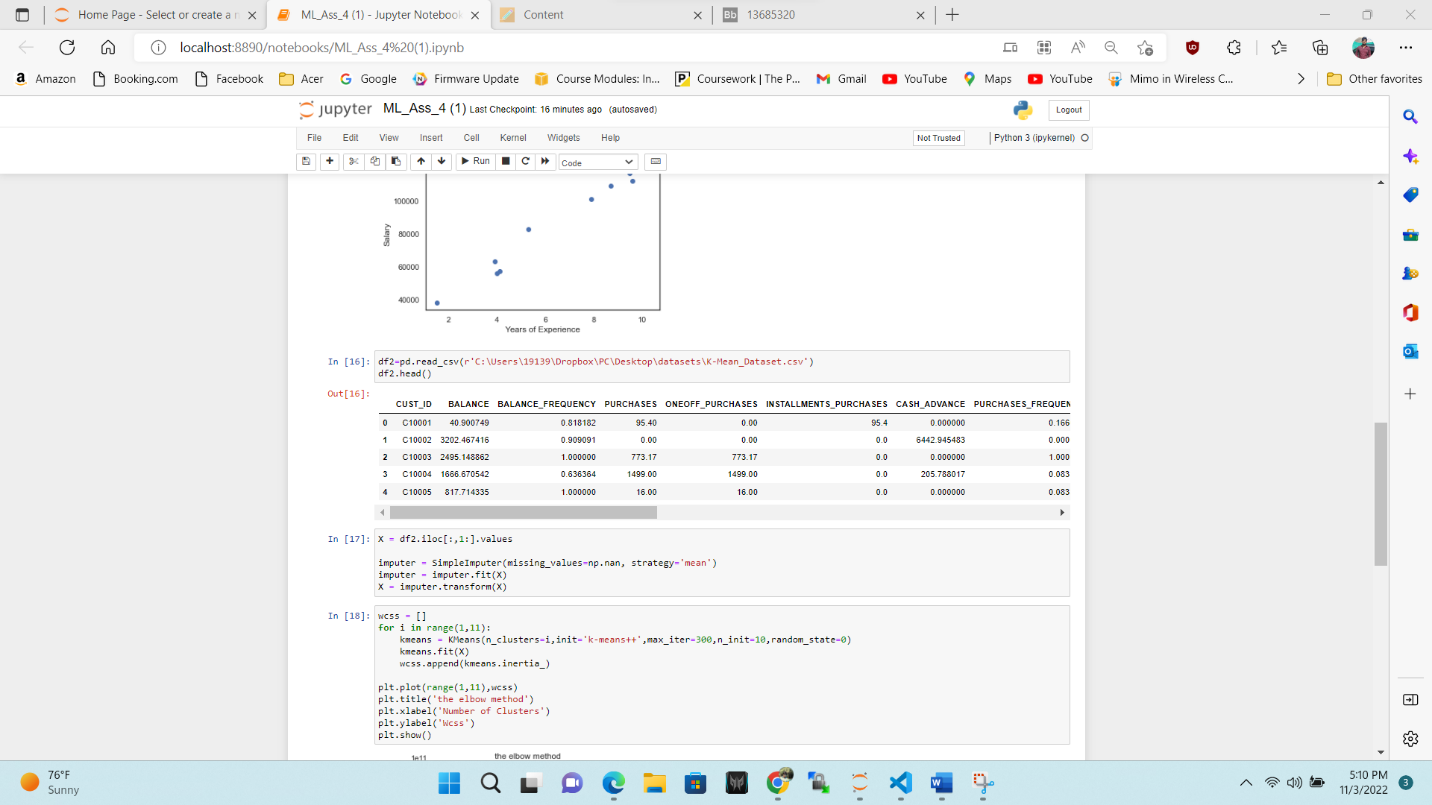


e. Visualize both train and test data using scatter plot.



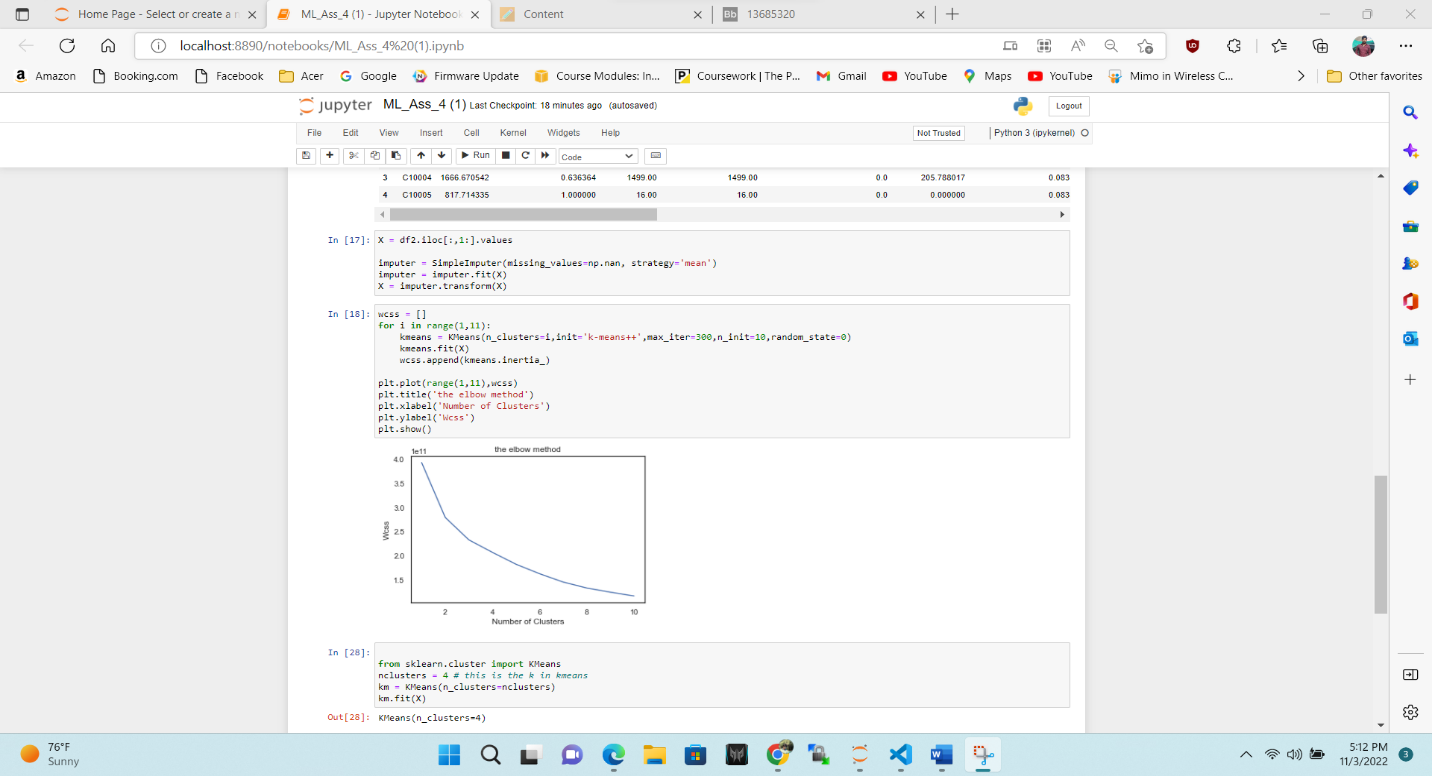
2. Apply K means clustering in the dataset provided:

• Remove any null values by the mean.



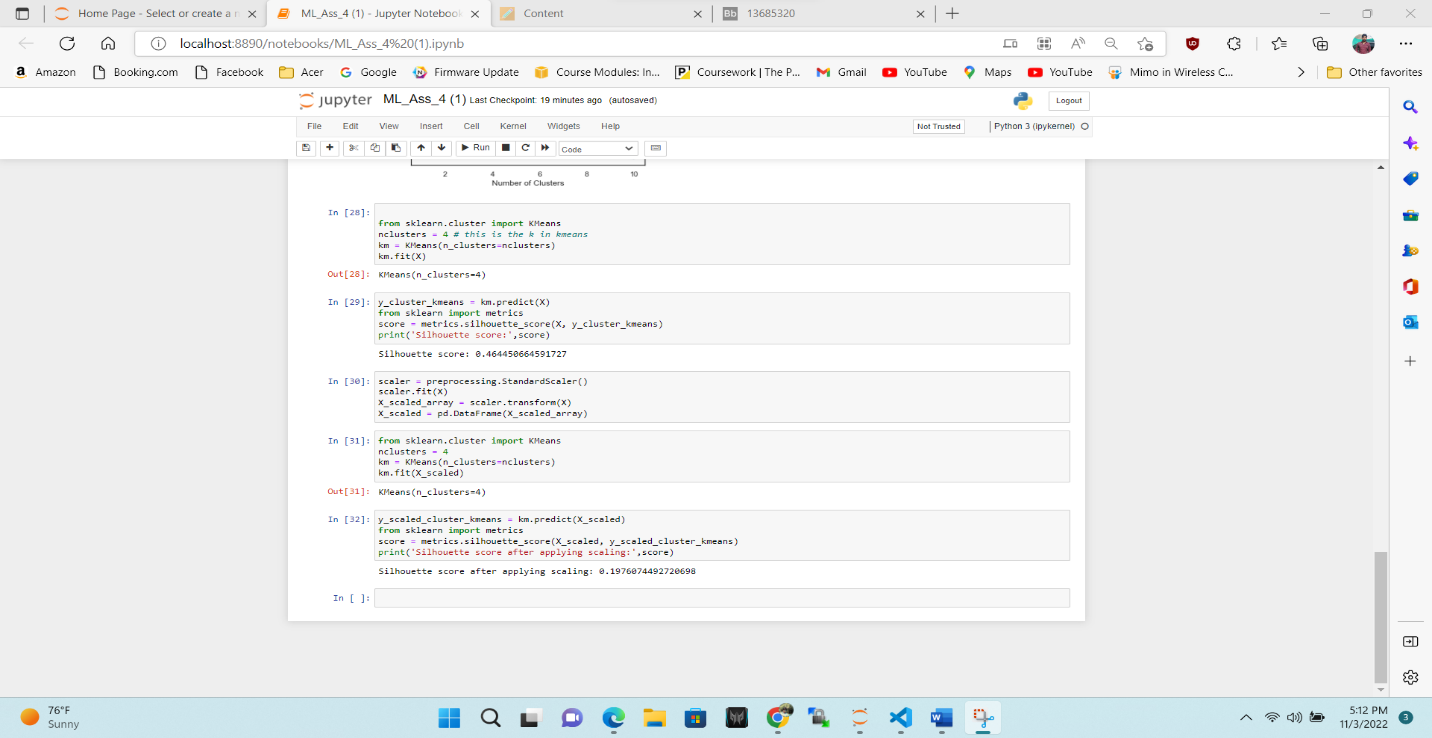
• Use the elbow method to find a good number of clusters with the K-Means algorithm

• 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



**Silhouette score before applying scaling:** 0.464450664591727

**Silhouette score after applying scaling:** 0.1976074492720698