Experiment 7

The aim of the experiment is to understand the use of Elbow plot for identifying appropriate number of clusters in clustering.

1. Read the file 'facebooklive.csv' using read_csv() in pandas.

2. Extract two features 'num_shares' and num_loves' and scale them using StandardScalar() in sklearn.

3. Create a function that calculates the squared of the euclidean distance between 2 points.

4. Vary number of clusters from 1 to 21 and evaluate total within-cluster sum of squares (WSS) for each case. Draw a plot between WSS and k. The point at which decrease in WSS with respect to k starts to flatten is known as the elbow of the plot. The x-index at this point is considered as appropriate number of clusters for a given application.