

# Computer Assignment: Clustering

Submitted by- Vandit Sharma

Roll No.- 17EC10060

Aim- To cluster the provided dataset using K-means clustering

***Case-I: Combination of age and annual income parameters***

Results-

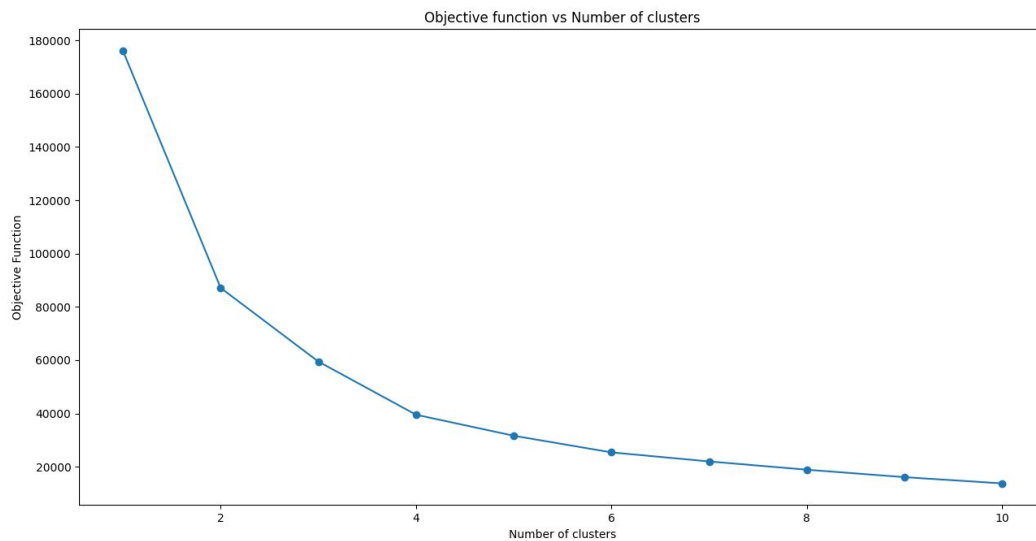


Figure 1: Objective function vs Number of clusters curve

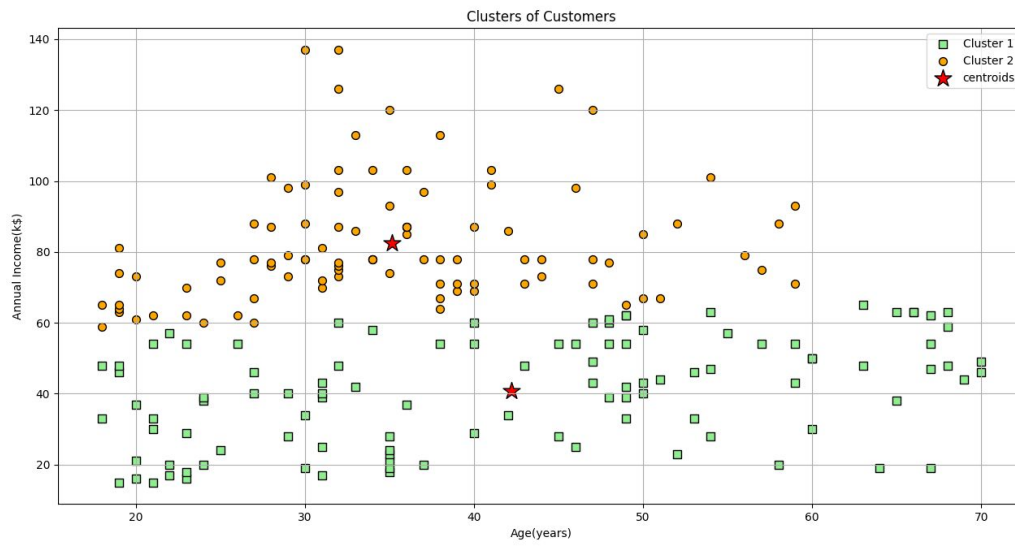


Figure 2: Plot of clustering output when  $K = 2$

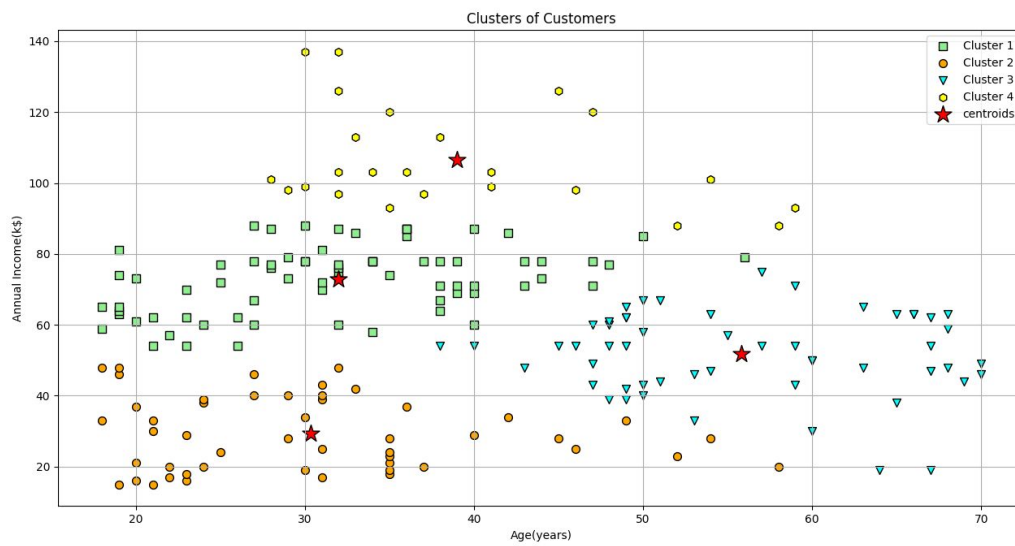


Figure 3: Plot of clustering output when  $K = 4$

## Observations-

- An abrupt change in the objective function vs the number of clusters plot (Figure 1) was observed at  $K = 2$  and  $K = 4$ .
- These data points did not seem to have a clear formation of clusters (like Figure 5).
- The K-means algorithm was able to identify clusters for both  $K = 2$  and  $K = 4$ , but since there was no clear separation, this classification might be inaccurate.

## ***Case-II: Combination of annual income and spending score parameters***

## Results-

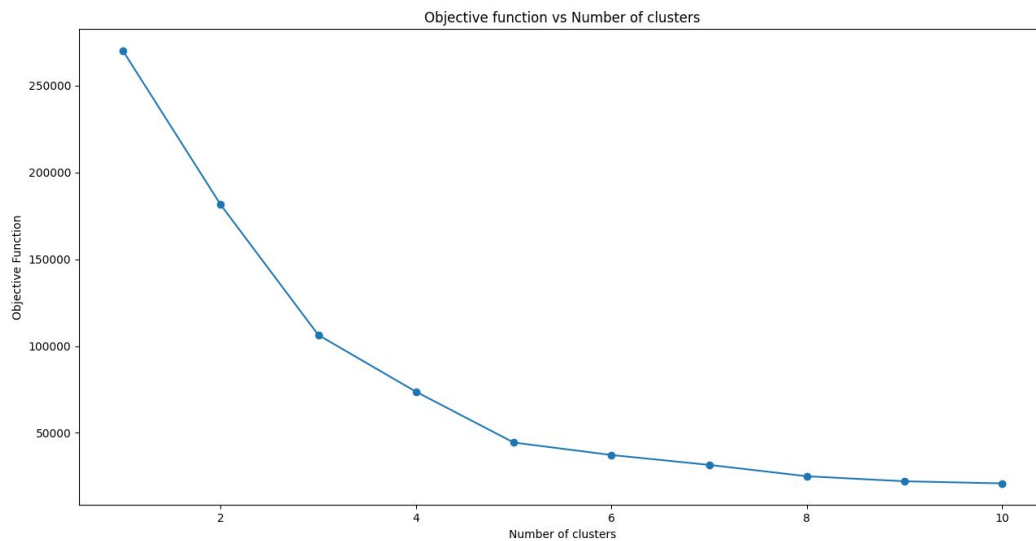


Figure 4: Objective function vs Number of clusters curve

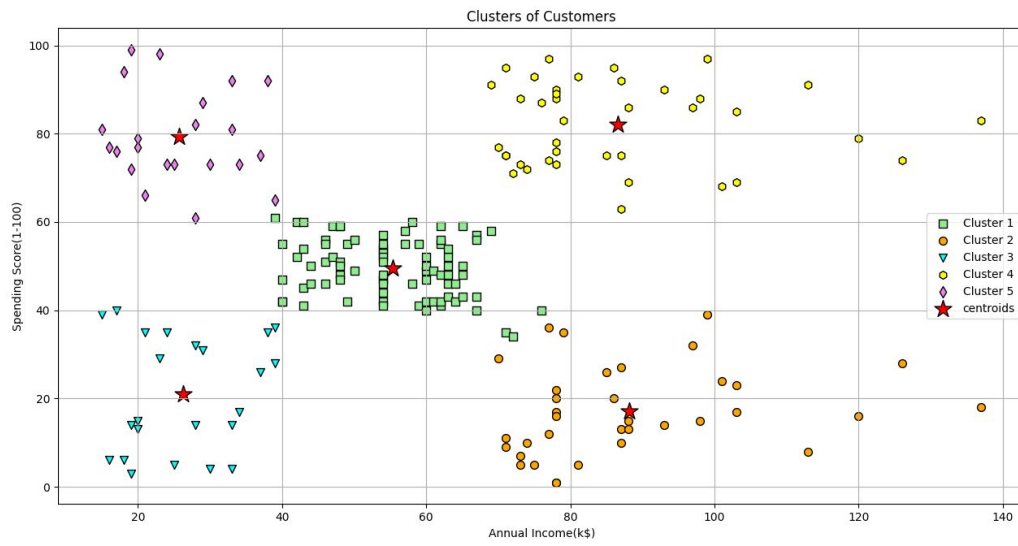


Figure 5: Plot of clustering output when K = 5

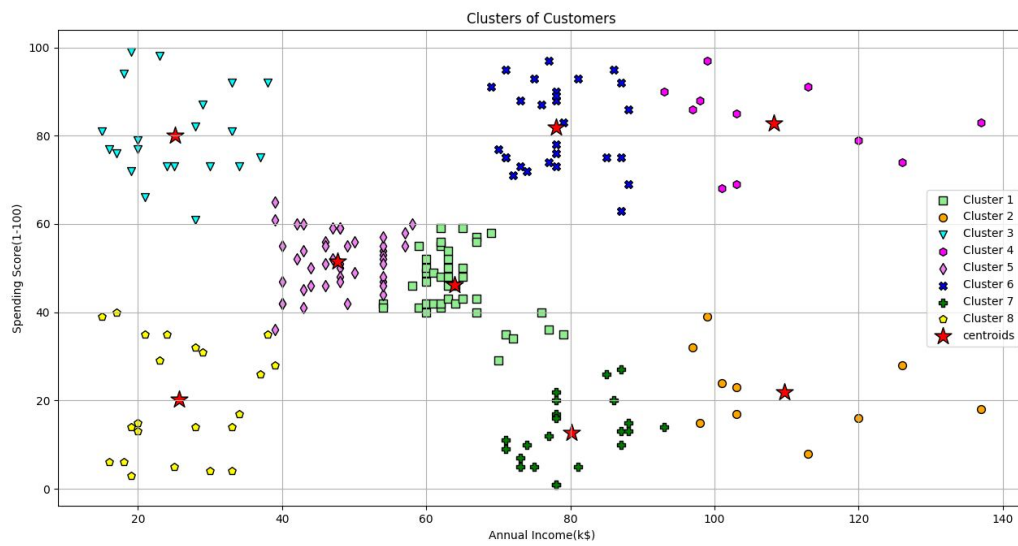


Figure 6: Plot of clustering output when K = 8

## Observations-

- An abrupt change in the objective function vs the number of clusters plot (Figure 4) was observed at  $K = 3$  and  $K = 5$ .
- These data points had a clear formation of clusters (unlike Figure 2,3).
- I obtained clusters for  $K = 5$  and  $K = 8$ . (I decided to go for  $K = 8$  instead of  $K = 3$  to observe the effect of increasing the number of clusters). The K-means algorithm was able to identify clusters for both  $K = 5$  and  $K = 8$ . The results seemed to be visually more accurate for  $K = 5$ , as  $K = 8$  forced the algorithm to break some good clusters.

### ***Case-III: Combination of age and spending score parameters***

## Results-

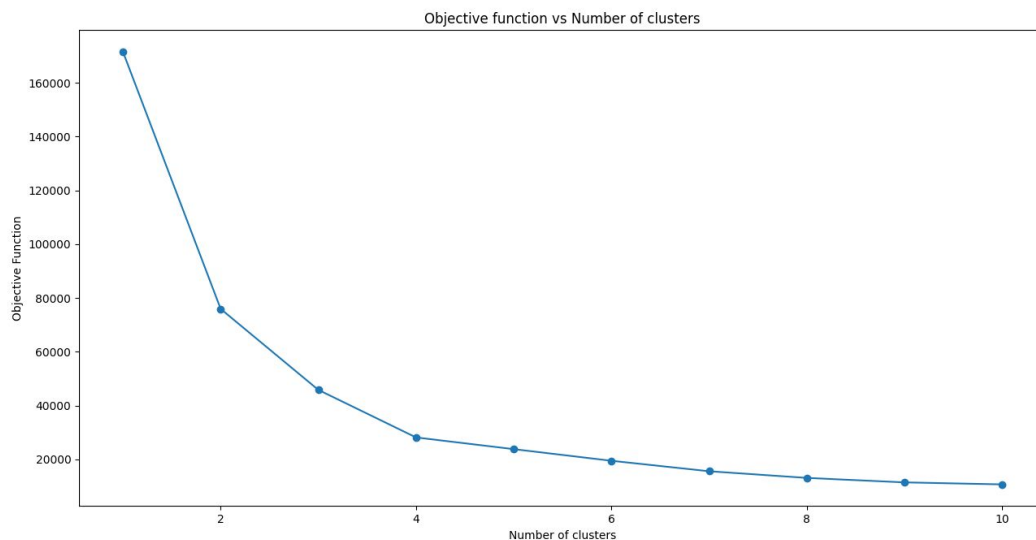


Figure 7: Objective function vs Number of clusters curve

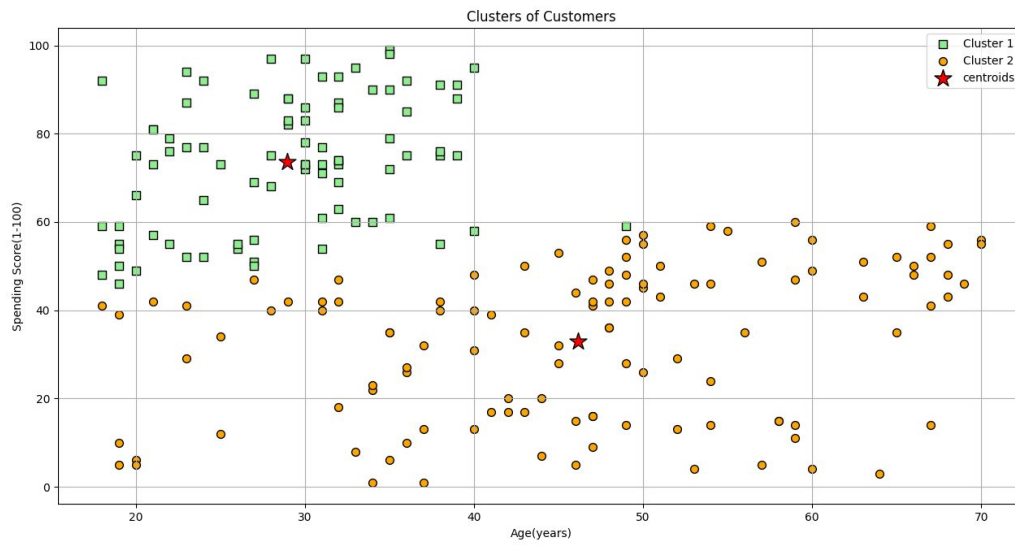


Figure 8: Plot of clustering output when  $K = 2$

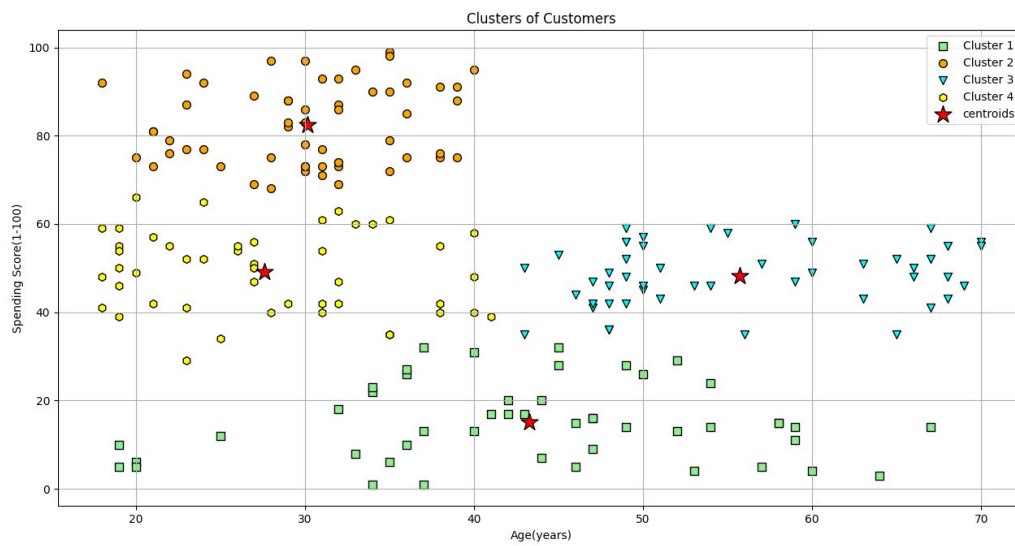


Figure 9: Plot of clustering output when  $K = 4$

## Observations-

- An abrupt change in the objective function vs the number of clusters plot (Figure 7) was observed at  $K = 2$  and  $K = 4$ .
- Visually, these data points did not seem to have a clear formation of clusters (like Figure 5).
- However, the K-means algorithm was able to identify clusters for both  $K = 2$  and  $K = 4$  reasonably well. Especially when  $K = 4$ , we were able to see the clusters quite clearly.