Overview | Customer Analysis | Customer Segmentation

# Customer Segmentation 👥

After evaluating various models( kmeans, ap, meanshift, dbscan, kmodes, etc.), the best silhouette score is shown in kmeans. With high dimension data as ours, we use PCA to reduced dimension of data before clustering process using kmeans.

## K-Mean with PCA Clustering

Original Data (21 dimensions)

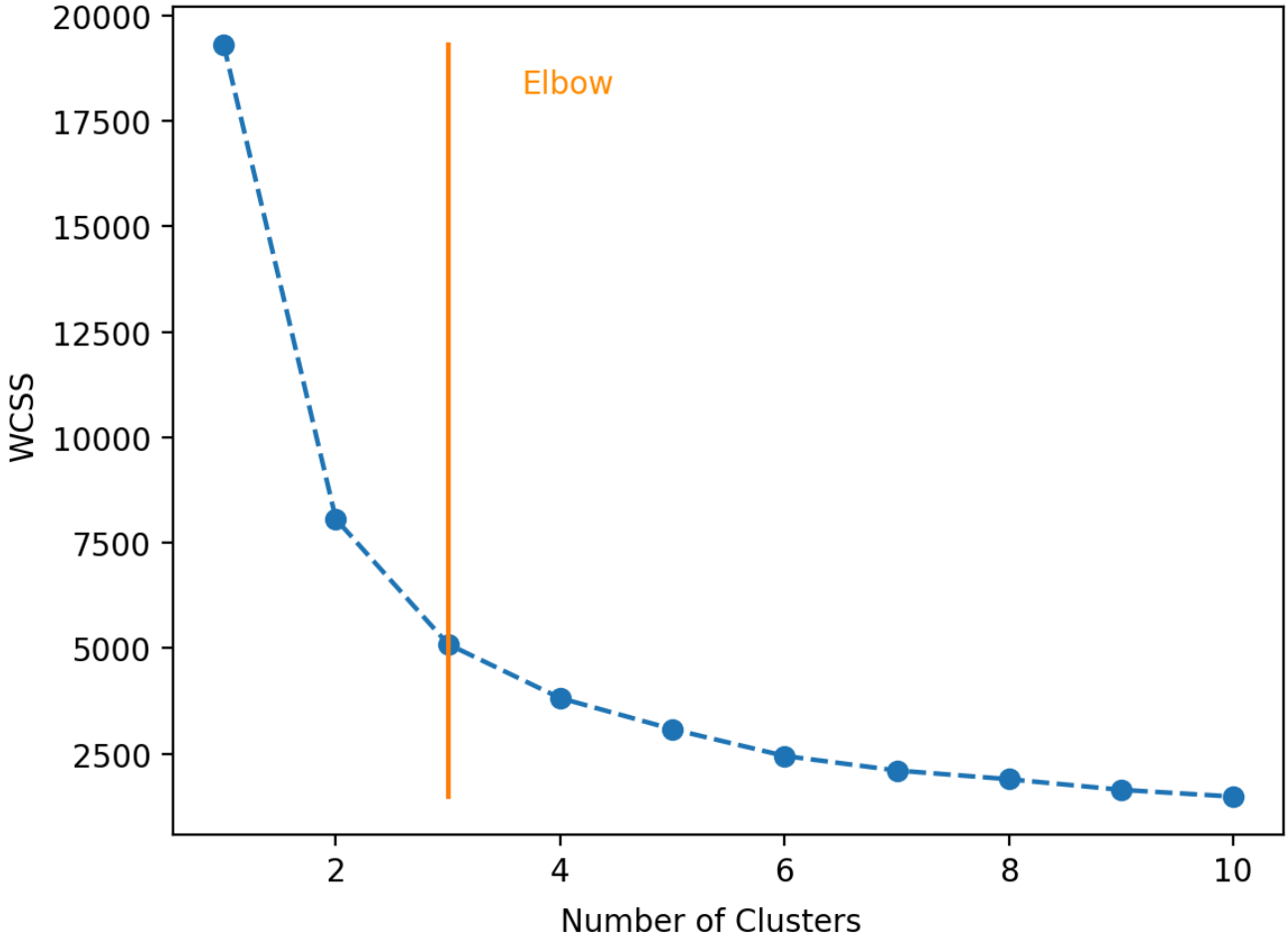
| | Income | Kidhome | Teenhome | Recency | MntWines | MntFruits |
|---|---|---|---|---|---|---|
| 0 | 58,138 | 0 | 0 | 58 | 635 | 88 |
| 1 | 46,344 | 1 | 1 | 38 | 11 | 1 |
| 2 | 71,613 | 0 | 0 | 26 | 426 | 49 |
| 3 | 26,646 | 1 | 0 | 26 | 11 | 4 |
| 4 | 58,293 | 1 | 0 | 94 | 173 | 43 |
| 5 | 62,513 | 0 | 1 | 16 | 520 | 42 |
| 6 | 55,635 | 0 | 1 | 34 | 235 | 65 |
| 7 | 33,454 | 1 | 0 | 32 | 76 | 10 |
| 8 | 30,351 | 1 | 0 | 19 | 14 | 0 |
| 9 | 5,648 | 1 | 1 | 68 | 28 | 0 |

Reduced-dimension (2 PCs)

| | PC1 | PC2 |
|---|---|---|
| 0 | 3.9061 | -1.3603 |
| 1 | -2.2067 | -0.6226 |
| 2 | 1.9702 | -0.7049 |
| 3 | -2.5575 | -1.6737 |
| 4 | -0.2907 | 0.9669 |
| 5 | 0.9335 | 2.3464 |
| 6 | 0.9001 | -0.4491 |
| 7 | -2.3709 | 0.6265 |
| 8 | -2.9557 | 0.681 |
| 9 | -4.8077 | 1.9284 |

## WCSS vs Number of Clusters

The “elbow” point indicates a strong candidate for the optimal number of clusters at 3.




## Customer Segments

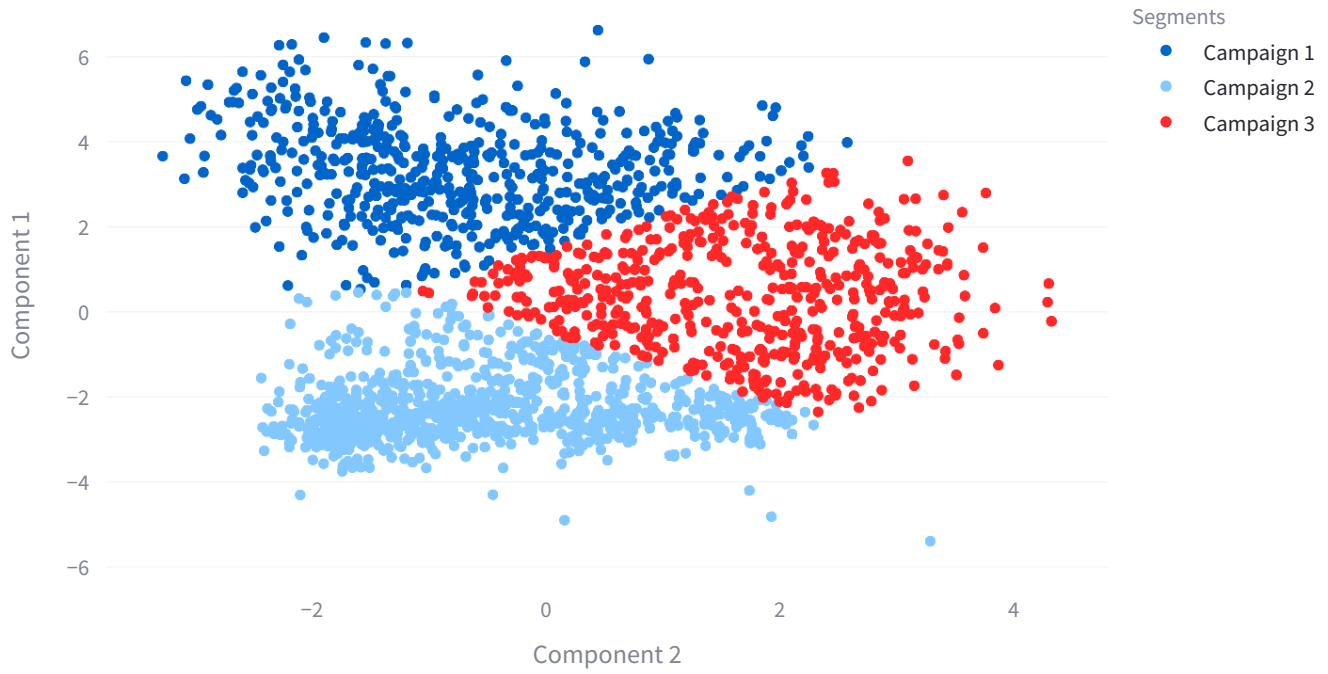
Merge the 2 PCA components with original data. Apply number of Cluster to generate segments.

WCSS shows elbow at K=3

3

2 5

| | Income ($) | Kidhome | Teenhome | Recency | MntWines | MntFruits | MntMeatProducts | MntFishProduc |
|---|---|---|---|---|---|---|---|---|
| 0 | 58,138 | 0 | 0 | 58 | 635 | 88 | 546 | 1 |
| 1 | 46,344 | 1 | 1 | 38 | 11 | 1 | 6 | |
| 2 | 71,613 | 0 | 0 | 26 | 426 | 49 | 127 | 1 |
| 3 | 26,646 | 1 | 0 | 26 | 11 | 4 | 20 | |
| 4 | 58,293 | 1 | 0 | 94 | 173 | 43 | 118 | |
| 5 | 62,513 | 0 | 1 | 16 | 520 | 42 | 98 | |
| 6 | 55,635 | 0 | 1 | 34 | 235 | 65 | 164 | |
| 7 | 33,454 | 1 | 0 | 32 | 76 | 10 | 56 | |
| 8 | 30,351 | 1 | 0 | 19 | 14 | 0 | 24 | |
| 9 | 5,648 | 1 | 1 | 68 | 28 | 0 | 6 | |




Each segmentation contains customer with different behavior. We can drive further to find the significant factors in which campaign can be launched based upon.