household_total_assets

cluster	mean	median	min	max	count
0	2,498,101	2,549,036	1,684,695	3,303,162	293
1	4,110,073	4,093,173	3,305,163	4,988,522	375
2	857,386	866,920	1,276	1,671,133	332

annual_household_income

cluster	mean	median	min	max	count
0	152,771	154,430	955	298,080	293
1	155,496	156,565	32	299,648	375
2	147,379	146,450	85	299,598	332

Cluster centers:

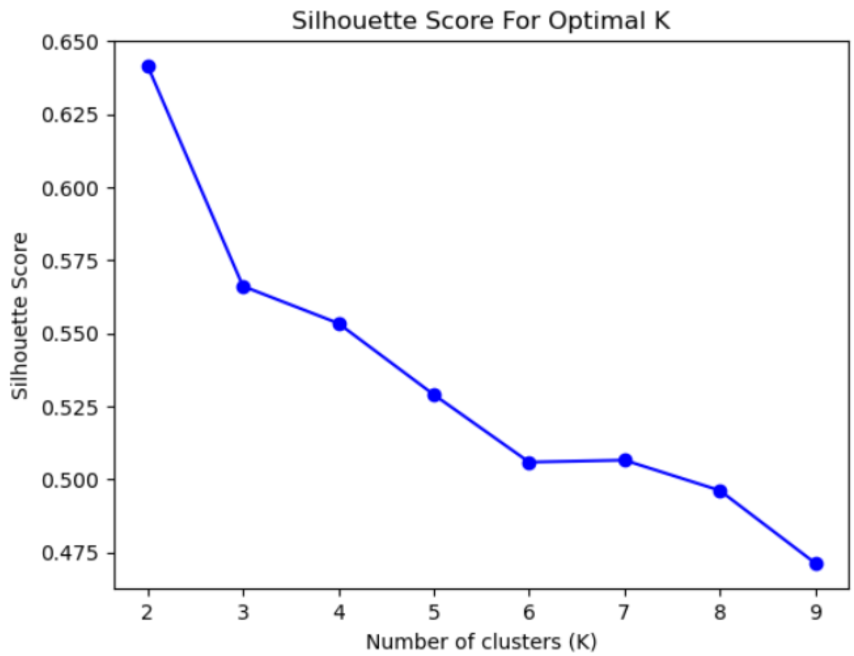
[[2498101.02730375 152771.49829352]

[4110073.37333333 155496.32533333]

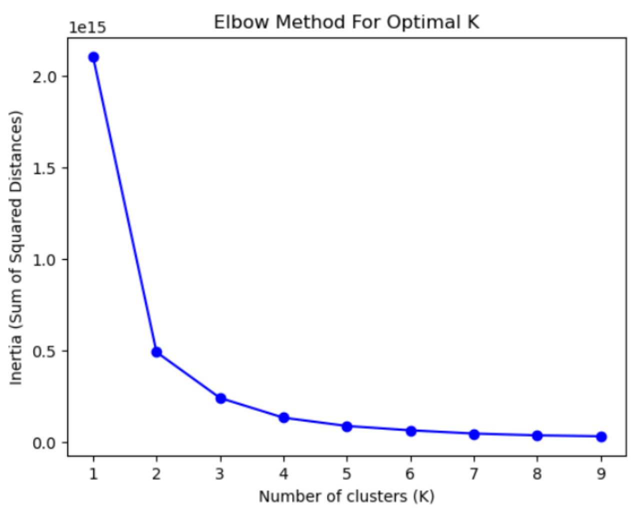
[857386.78915663 147379.93072289]]

From the clustering done with K-means algorithm we can see 3 district groups. Group 0 has 293 potential customers that have household assets from $1,684,695 to $3,303,162 which is a mean and median of around $2,500,000. This puts the group in the mid-range of the total customers. They also have a mean and median annual income of $150,000. We can tell from this information that they invest a medium amount in their household. Group 1 has 375 potential customers making it the largest cluster. They have household assets from $3,305,163 to $4,988,522 with a mean and median of around $4,110,073. This group has an annual income of around $155,496. Group 1 spends the most on household assets so is most likely to buy something. Finally, group 2 has 332 potential customers in the cluster. They have household assets from $1,276 to $1,671,133 with a mean and median of around $857,386. Their mean annual income is $147,379. This group is least likely to spend money of household assets. All 3 groups have similar mean annual incomes just the spending habits differ.

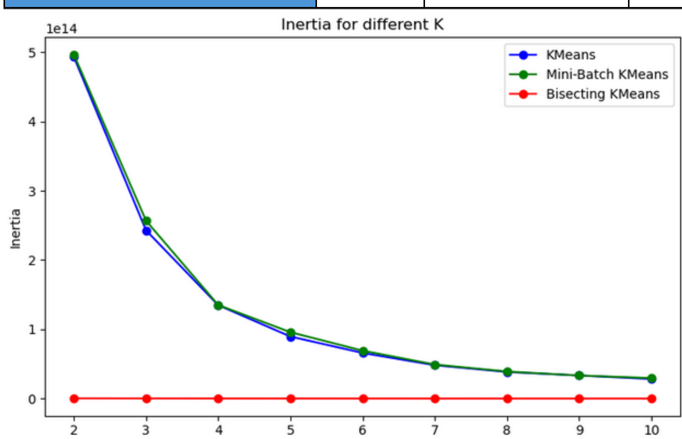
One way to find a good clustering k number is using the silhouette method. Silhouette score gives a number from –1.0 to 1.0 where bigger the number the better and more likely the clustering is best.

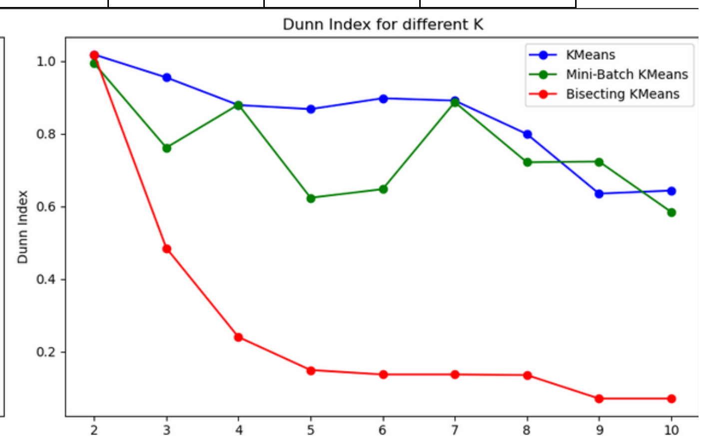


The elbow method is also another way to try gauge which number of clusters to use. This method works by calculating the distances between data points and their respective cluster centroids. This method doesn't work very well on this particular problem due to the cluster groups being very similar.



K	K-Means		Bisecting K-Means		Mini-Batch K-Means	
	Dunn Index	Inertia	Dunn Index	Inertia	Dunn Index	Inertia
2	1.017	493 Trillion	1.017	400 Million	0.992	496 Trillion
3	0.954	242 Trillion	0.484	242 Million	0.761	256 Trillion
4	0.878	134 Trillion	0.239	110 Million	0.879	135 Trillion
5	0.867	89 Trillion	0.149	84 Million	0.623	95 Trillion
6	0.897	65 Trillion	0.136	64 Million	0.646	69 Trillion
7	0.890	48 Trillion	0.136	46 Million	0.886	49 Trillion
8	0.798	38 Trillion	0.135	30 Million	0.720	39 Trillion
9	0.634	33 Trillion	0.070	27 Million	0.722	33 Trillion
10	0.642	28 Trillion	0.070	23 Million	0.584	29 Trillion





Metric	K-Means	Agglomerative	DBSCAN
Dunn Index	1.017	0.8118	inf
Inertia	493 Trillion	594 Trillion	2.1 Quadrillion

KNN

Steps	N neighbors	Error
	1	0.50
	2	0.51
	3	0.53
	5	0.51
	7	0.51
	9	0.50