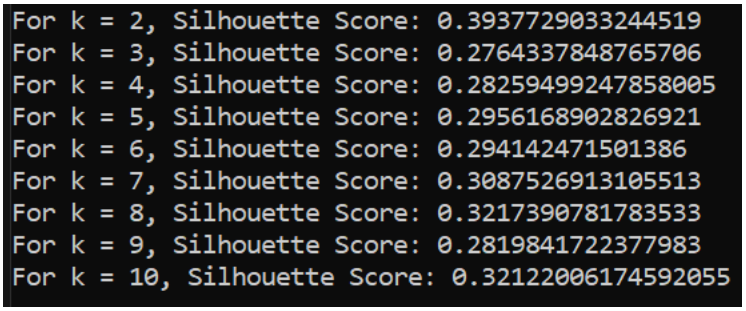
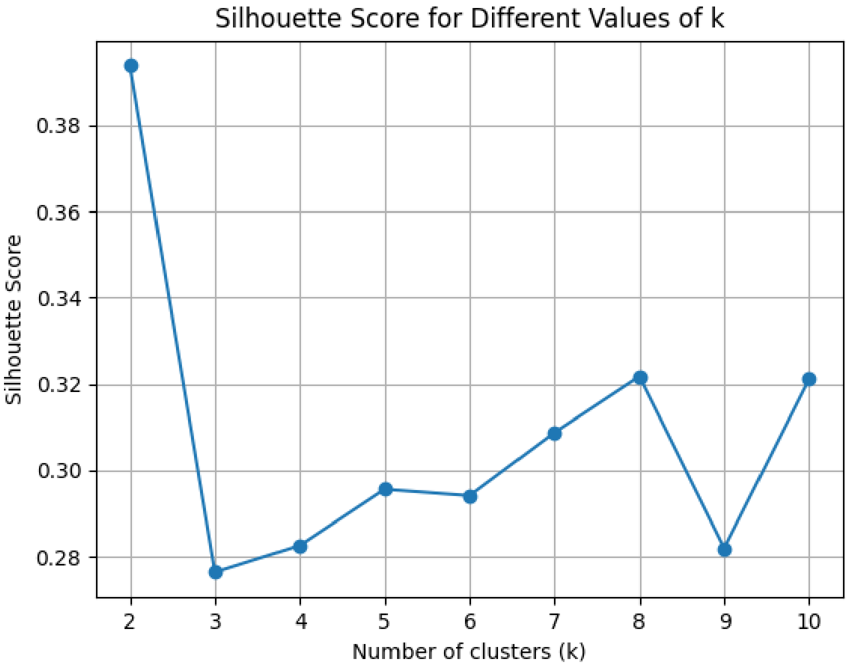
2. Random Initilization

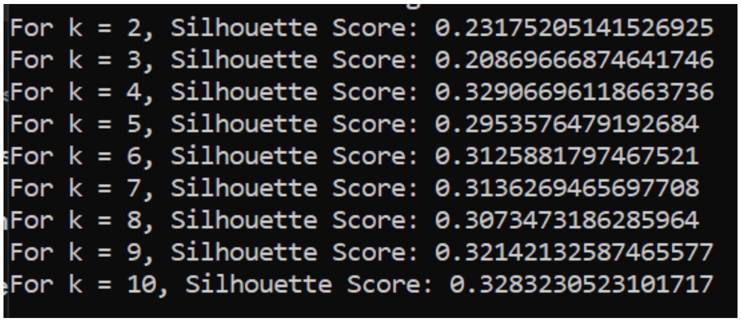


```
For k = 2, Silhouette Score: 0.3937729033244519
For k = 3, Silhouette Score: 0.2764337848765706
For k = 4, Silhouette Score: 0.28259499247858005
For k = 5, Silhouette Score: 0.2956168902826921
For k = 6, Silhouette Score: 0.294142471501386
For k = 7, Silhouette Score: 0.3087526913105513
For k = 8, Silhouette Score: 0.3217390781783533
For k = 9, Silhouette Score: 0.2819841722377983
For k = 10, Silhouette Score: 0.32122006174592055
```

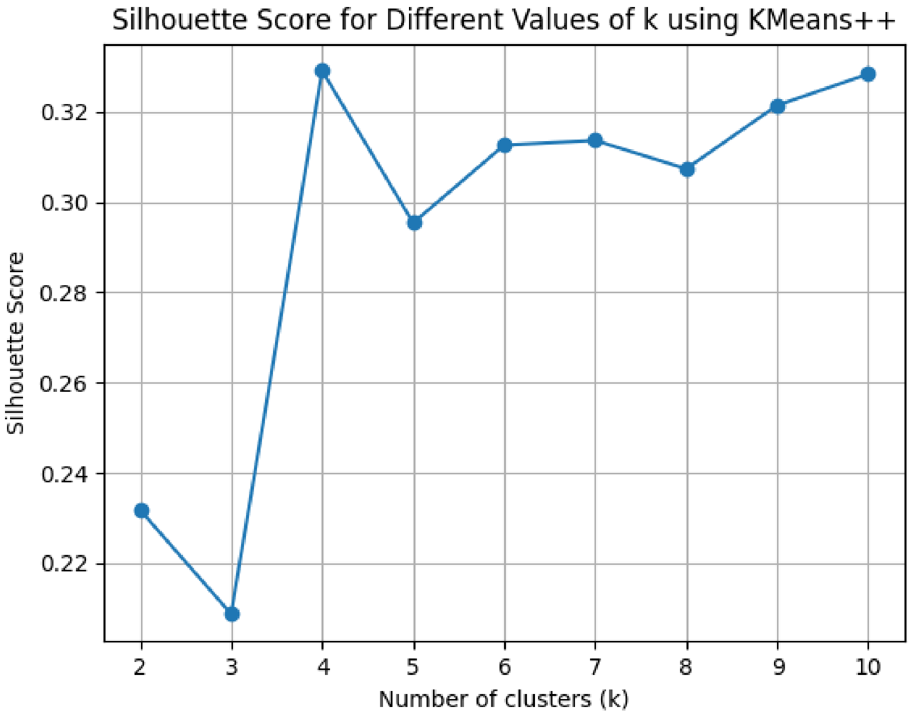


I found k = 2, with silhouette score of 0.39 to be the best.

3. KMeans ++ Initialization



```
For k = 2, Silhouette Score: 0.23175205141526925
For k = 3, Silhouette Score: 0.20869666874641746
For k = 4, Silhouette Score: 0.32906696118663736
For k = 5, Silhouette Score: 0.2953576479192684
For k = 6, Silhouette Score: 0.3125881797467521
For k = 7, Silhouette Score: 0.3136269465697708
For k = 8, Silhouette Score: 0.307347318628596 4
For k = 9, Silhouette Score: 0.3214213258746557 7
For k = 10, Silhouette Score: 0.3283230523101717
```



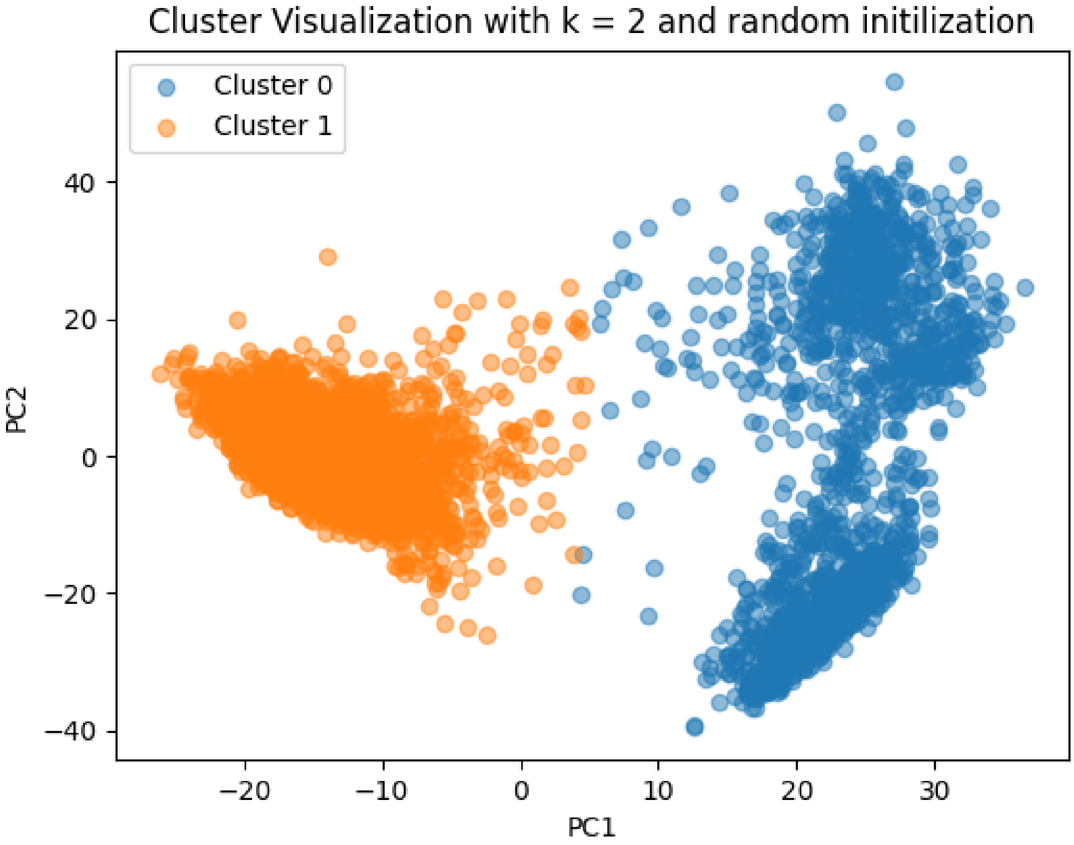
Using KMeans++, I found k = 4, with silhouette score of 0.329 to be the best.

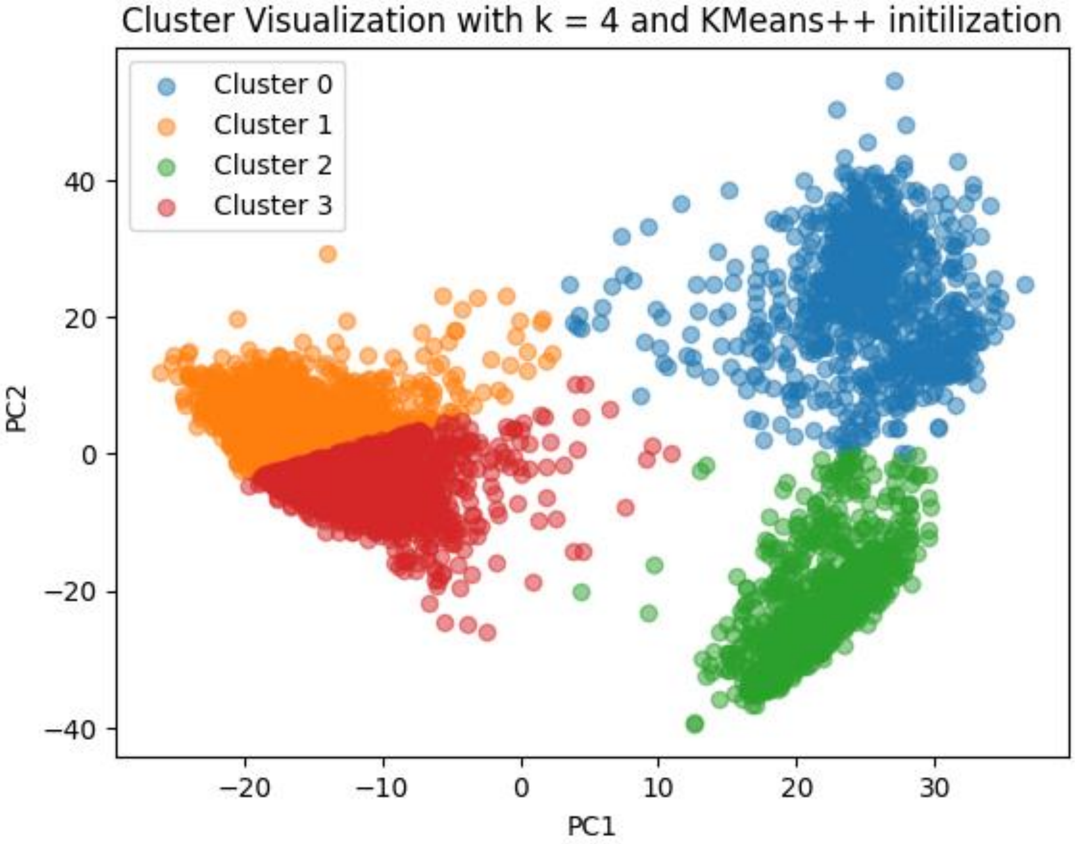
Conclusions:

1. **Effect of Initialization**: KMeans++ had better runtime and therefore converges faster.
2. **Tradeoff for selecting k**: The optimal number of clusters determined by the silhouette score changed between the two scenarios. In the case of random initialization, k = 2 yielded the higher silhouette score, suggesting that the data might have a clear separation into two clusters. While k = 2 might produce higher silhouette scores with random initialization, it doesn't necessarily mean it's the best choice for representing the

underlying structure of the data. k = 4 used with KMeans if faster, and may produce more meaningful and interpretable results depending on the context.

4.Scatter Plots




Cluster Visualization with k = 4 and KMeans++ initilization
● Cluster 0
● Cluster 1
● Cluster 2
● Cluster 3
40
20
0
−20
−40
PC2
−20
−10
0
10
20
30
PC1