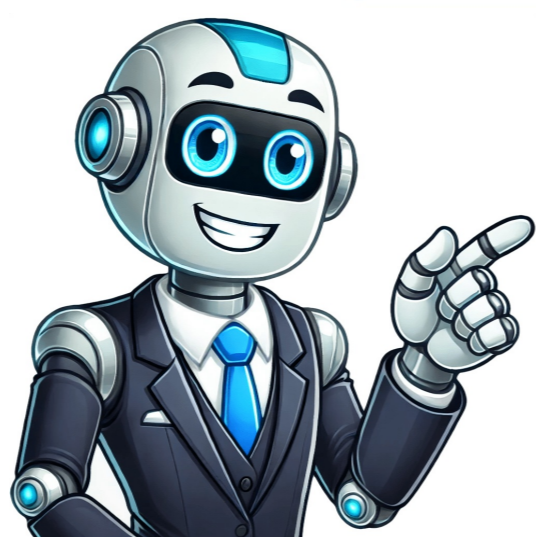


I'm not a robot































The plot shows the silhouette score against the number of clusters. In my opinion, increasing sample size is reducing the noise is a normal behavior. But given that I have 1 million, a heterogenous vector, 2 or 3 clusters is the "best" number of clusters seems unintuitive. In other words, I would expect to find a monotonic decreases in silhouette score as I increase the number of clusters.