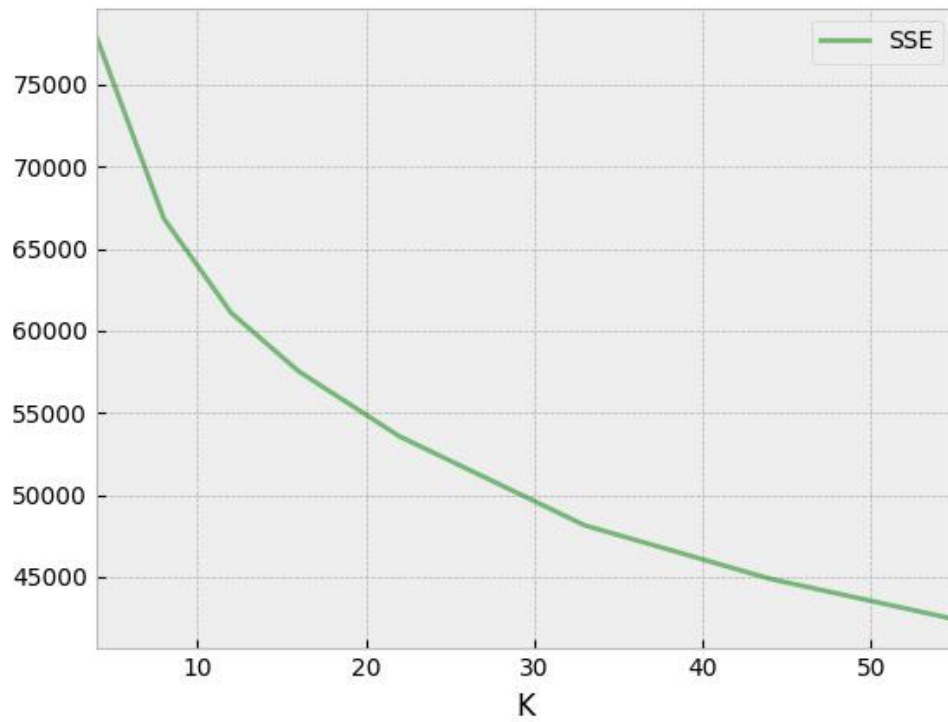
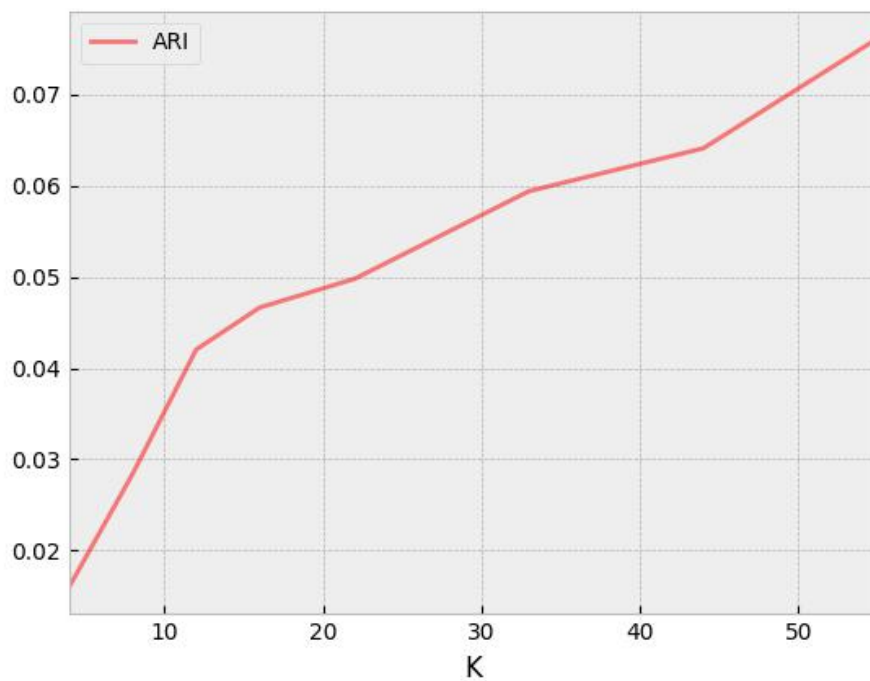


a) Subbha Praveen Tadisetty AI0004

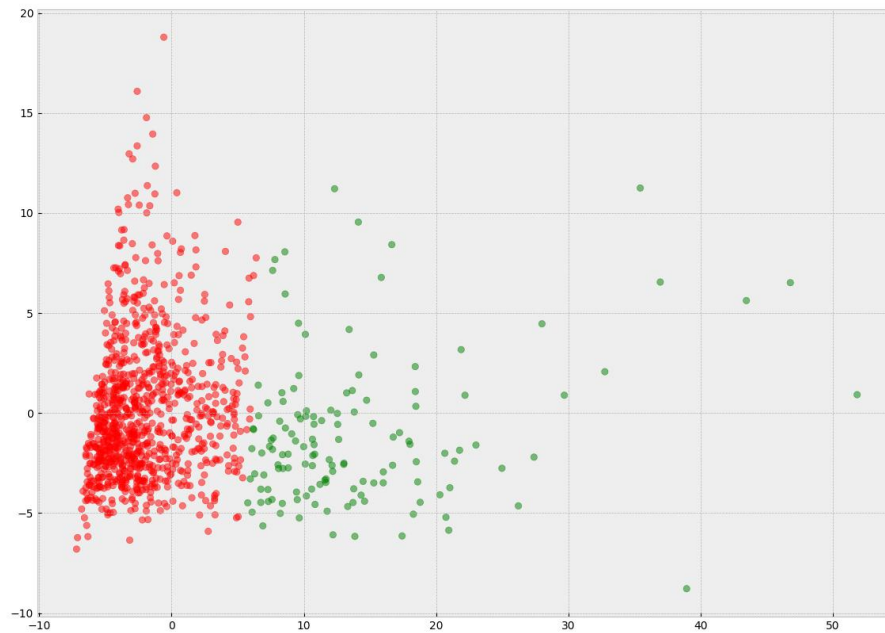
b) SSE vs K



c) ARI vs K



d) 2D PCA feature scatter plot



e) Table1: K-means clustering.

Number of cluster	Rand Index	SSE
4	0.016	77906
8	0.028	66841
12	0.042	61090
16	0.047	57531
22	0.050	53557
33	0.059	48134
44	0.064	44879
55	0.076	42405

f) K=22 clusters Table 2: K-means clustering after dimensionality reduction

Number of components in PCA	Rand Indexusing chosen
N = 2	0.026
N = 4	0.038
N = 8	0.050
N = 16	0.053
N = 32	0.046

K value is selected as 22 as the SSE value is not reducing with increase in K value. Reduction in error is very less for k greater than 22.

