

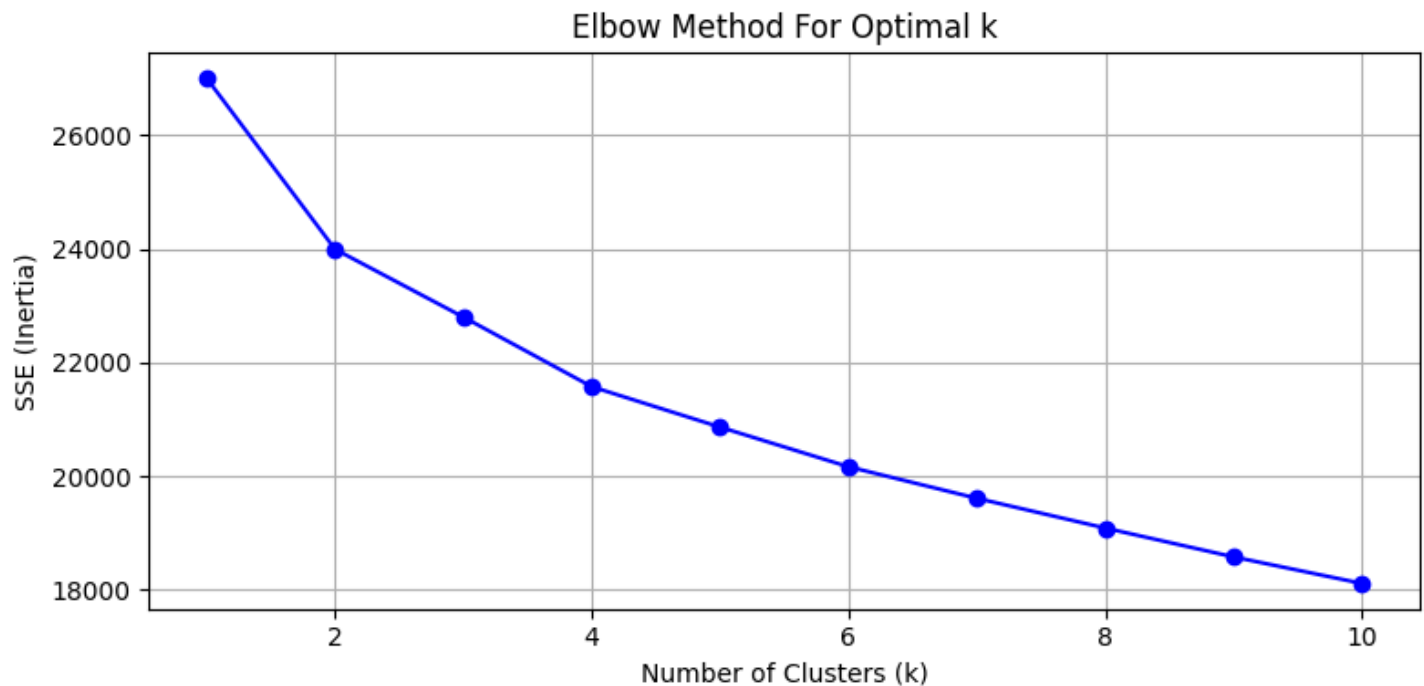
K-Means Clustering Report: Health & Fitness Dataset

Objective

To group individuals based on their health and fitness attributes using unsupervised learning (K-Means clustering), in order to identify distinct patterns or profiles among the population.

Step-by-Step Workflow

1. Data Preprocessing
 - Loaded dataset: health_fitness_dataset.csv
 - Removed 'ID' column as non-informative.
 - Encoded categorical variables: Gender (Male=0, Female=1), Chronic Conditions (categorical codes).
 - Missing values handled using mean imputation.
 - Duplicate entries removed.
2. Feature Scaling
 - StandardScaler applied to normalize feature values for equal contribution in clustering.
3. Elbow Method for Optimal k
 - SSE (inertia) computed for $k = 1$ to 10.
 - Optimal number of clusters determined as $k=4$ based on elbow point.



4. Clustering with $k=4$
 - Applied KMeans clustering with 4 clusters.
 - Cluster labels added to the dataset.

5. Visualization

- PCA (2D) used for cluster visualization.
- Clear visual separation observed in scatterplot with various colors for each cluster.

Cluster Characteristics

Cluster	Age	Gender	Height (cm)	Weight (kg)	BMI	Steps per Day	Hours of Exercise per Week	Diet Quality Score	Chronic Conditions
0	30.16	1.0	169.95	70.45	24.23	10317.65	4.63	5.39	0.41
1	43.16	0.0	170.64	69.80	24.21	10613.76	5.03	5.63	1.51
2	43.43	0.0	169.69	70.72	24.05	10296.81	5.02	5.27	-0.50
3	56.66	1.0	170.11	69.46	23.65	10532.91	5.17	5.89	0.58

Cluster 0: Young females with average fitness, moderate exercise, and low chronic conditions.

Cluster 1: Middle-aged males with high steps, good exercise, and high chronic conditions.

Cluster 2: Middle-aged males with slightly lower steps and diet scores, and fewer chronic conditions.

Cluster 3: Older females with good activity levels, best diet quality, and moderate chronic conditions.