

Clustering Analysis Report

Number of Clusters: 7

Based on the Davies-Bouldin Index plot, we selected 7 clusters as our optimal clustering solution. This choice is supported by the following metrics:

- Lower DB Index values indicate better clustering,
- High intra-cluster similarity
- Low inter-cluster similarity
- Better overall cluster separation

Cluster Stability:

- The DB Index shows relative stability around $k=7$
- The small increase in DB Index after $k=7$ suggests that additional clusters may not provide meaningful improvements
- The gradual plateau in the elbow curve around this point supports the choice of 7 clusters.

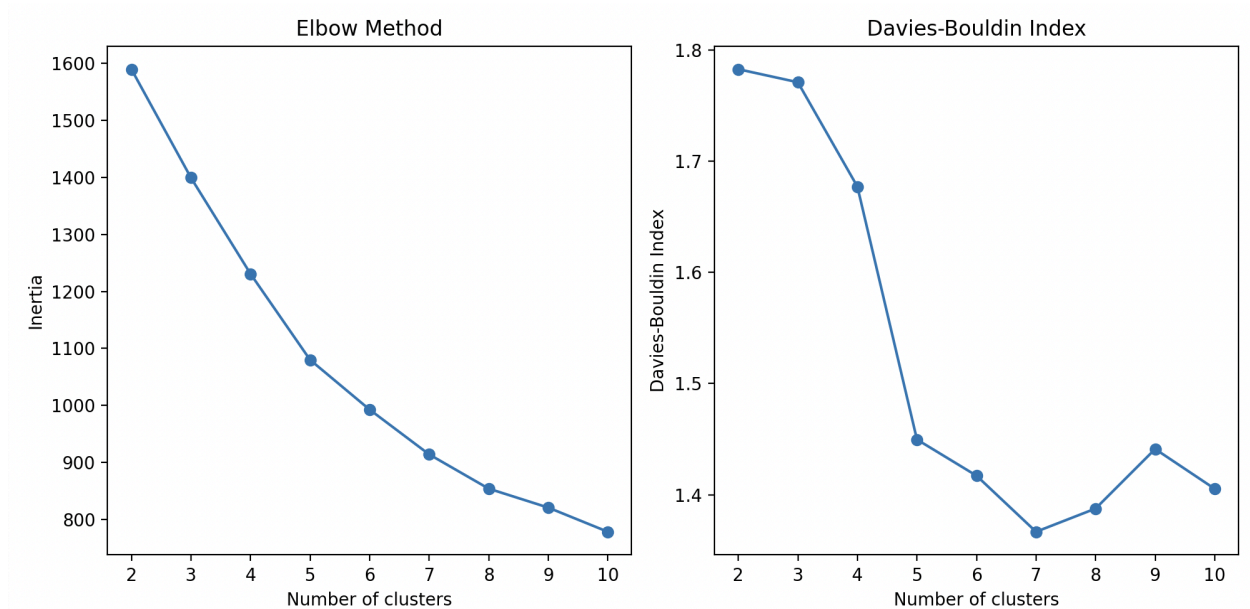
Key Observations:

- The choice of 7 clusters provides a good balance between:
- Cluster granularity
- Model complexity
- Cluster separation
- The relatively stable DB Index around $k=7$ suggests that this clustering solution is robust
- The significant drop in inertia from the initial state indicates that the clusters are well-defined

```
Clusters: 2, DB Index: 1.7826
Clusters: 3, DB Index: 1.7709
Clusters: 4, DB Index: 1.6770
Clusters: 5, DB Index: 1.4497
Clusters: 6, DB Index: 1.4172
Clusters: 7, DB Index: 1.3667
Clusters: 8, DB Index: 1.3876
Clusters: 9, DB Index: 1.4411
Clusters: 10, DB Index: 1.4054
2025-01-27 22:27:52.233 Python[22331:20596590] +[IMKClient subclass]: chose IMKClient_Modern
2025-01-27 22:27:52.234 Python[22331:20596590] +[IMKInputSession subclass]: chose IMKInputSession_Modern

Optimal number of clusters: 7
```

Validation of Choice:



I have tried visualizing the elbow plot but you can clearly observe that there is ambiguity in there



Since the data has 4 dimensions it cannot be visualized. That is why I have used a famous dimensionality reduction technique called PCA (principal component analysis) for 2 dimensional visualization of data