

Customer Segmentation using K-Means Clustering (Mall Customers)

Unsupervised segmentation of 200 mall customers on income and spending behaviour — the classic retail-analytics project: elbow-method model selection, five interpretable segments, and a live segment finder.

1. Introduction / Background

Every retailer knows customers are not all alike, but few can say precisely how their shoppers group — who spends freely, who buys carefully, who needs a nudge. Customer segmentation turns that intuition into data: group customers by behaviour so marketing, inventory and loyalty programs can target each group differently.

This project does the classic version properly on the Mall Customers dataset (200 shoppers): standardized features, k-means++ initialization, the elbow method to justify $k=5$, silhouette-score validation, and five named, interpretable segments with business profiles — plus an interactive demo with a live segment map and segment finder.

2. Problem Statement

Blanket marketing wastes budget on customers who will never respond, while high-value segments go unnurtured. Without segmentation, promotions are guesswork. Students usually run K-Means with an arbitrary k and stop; this project shows the full discipline — justifying k , validating without labels, and translating clusters into business actions.

3. Objectives

- Segment 200 mall customers by annual income and spending score with K-Means.
- Justify the number of segments with the elbow method and validate with the silhouette score.
- Profile each segment with a business interpretation and marketing action.
- Build an interactive segment map and live segment finder.

4. Proposed Methodology

Annual Income and Spending Score are selected as clustering features and standardized (zero mean, unit variance) so income scale cannot dominate distances. K-Means is run for $k=1..10$ with k-means++ initialization; the WCSS elbow justifies $k=5$. The final model is fitted with $k=5$ (300 max iterations) and the silhouette score is computed. Centroids are translated into five business profiles: Target, Careful, Spendthrift, Sensible and Standard.

5. System Architecture / Working

- Input: (annual income, spending score), standardized with the saved scaler.
- Model: K-Means with $k=5$, k-means++ seeds, converged centroids.
- Profiling: each centroid → segment name + business action.

- Demo: 200 customers plotted coloured by segment; the finder assigns new inputs to the nearest centroid.

6. Hardware / Software Requirements

- Python 3, scikit-learn, Pandas, NumPy, Matplotlib, Jupyter Notebook.
- Any laptop — the notebook runs in under two minutes.
- The web demo runs offline in any modern browser after download.

7. Expected Outcomes

- Five interpretable customer segments with centroid profiles and marketing actions.
- Elbow and silhouette analysis (silhouette ≈ 0.55 design target) justifying and validating the segmentation.
- An interactive demo, complete report, PPT and viva Q&A; on unsupervised methodology.

8. Applications

- Teaching unsupervised learning: model selection and validation without labels.
- Retail-analytics demonstration: targeted campaigns, loyalty tiers, inventory planning.
- A template for segmentation on richer customer data.

9. Future Scope

- Richer features: recency, frequency, monetary (RFM) segmentation.
- Comparison with DBSCAN and Gaussian mixtures.
- Deployment on a live retail dataset with A/B-tested campaigns.

10. Conclusion

The project shows unsupervised learning done rigorously — k chosen by evidence, quality quantified without labels, and clusters translated into decisions — the complete journey from raw customer records to actionable segments.