

Customer Segmentation Using K-Means Clustering

This report explains customer segmentation using the K-Means clustering algorithm. Customer segmentation groups customers based on similar characteristics, helping businesses personalize services, increase engagement, and improve decision-making. K-Means is applied to identify hidden patterns in customer behavior.

Customer Segmentation

Group Patterns

Behavior Similarity

Customer segmentation organizes individuals into meaningful groups based on shared behavior or demographic traits. The diagram shows customers forming natural groups based on similarities, allowing businesses to understand needs, preferences, and spending behavior more accurately. This helps companies target customers effectively.

K-Means Algorithm

Initialize

Assign Points

Recalculate

K-Means clustering divides data into groups by placing points into clusters based on similarity. It works by initializing random centers, assigning points based on closeness, and updating the centers repeatedly. This process reveals natural customer groupings that support marketing and product decisions.

Clustering Concept

Cluster A

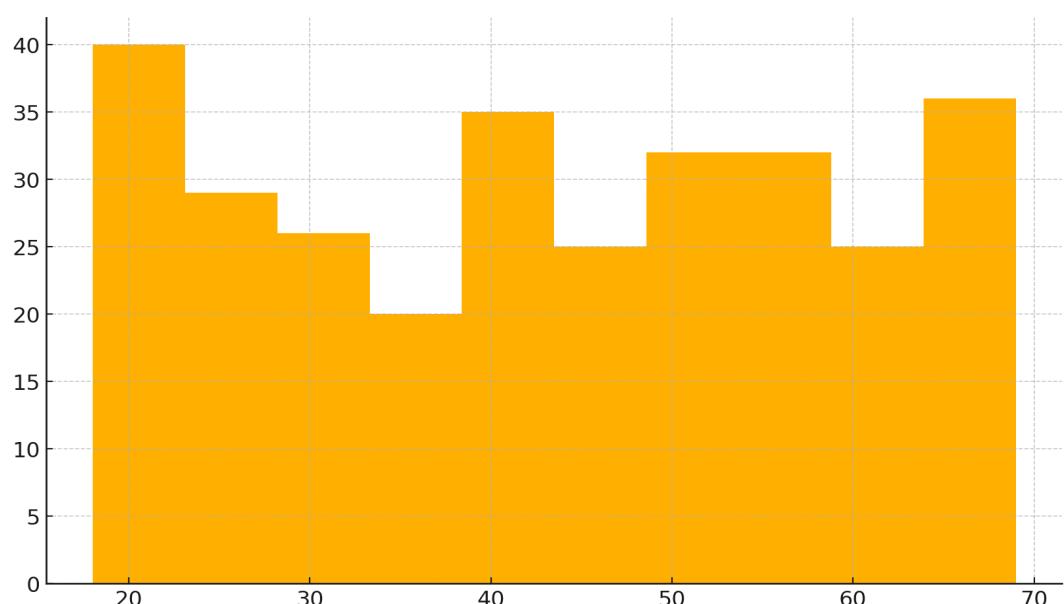
Cluster B

Cluster C

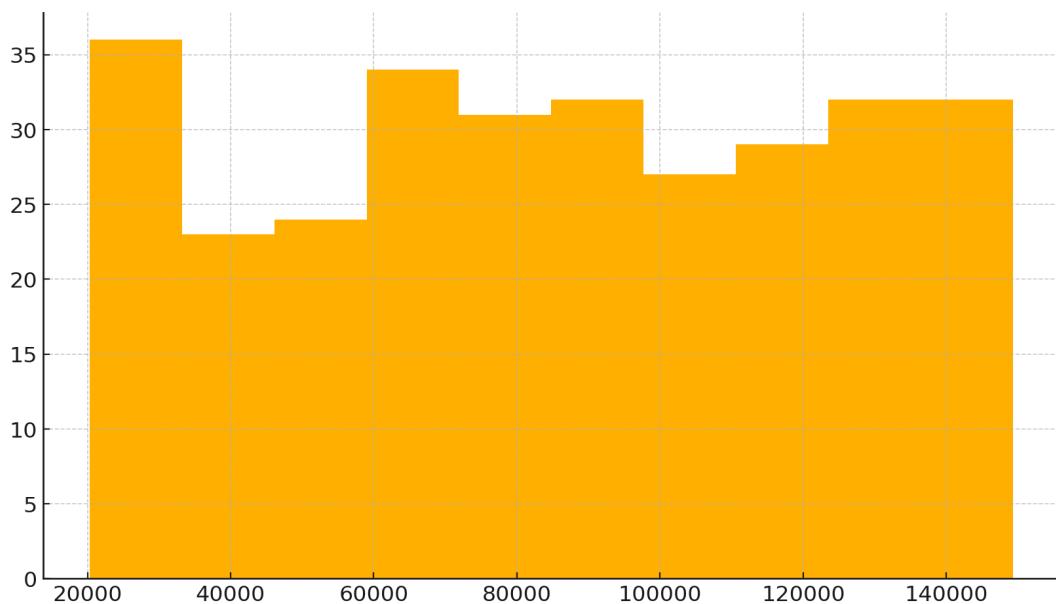
Clustering forms groups of customers who share similar traits. These clusters represent patterns that businesses can use to understand purchasing habits, income levels, or engagement levels. Clustering enables strategic targeting and resource optimization.

Dataset contains: Age, Annual Income, and Spending Score for 300 customers. These attributes reflect demographic, financial, and behavioral characteristics essential for segmentation.

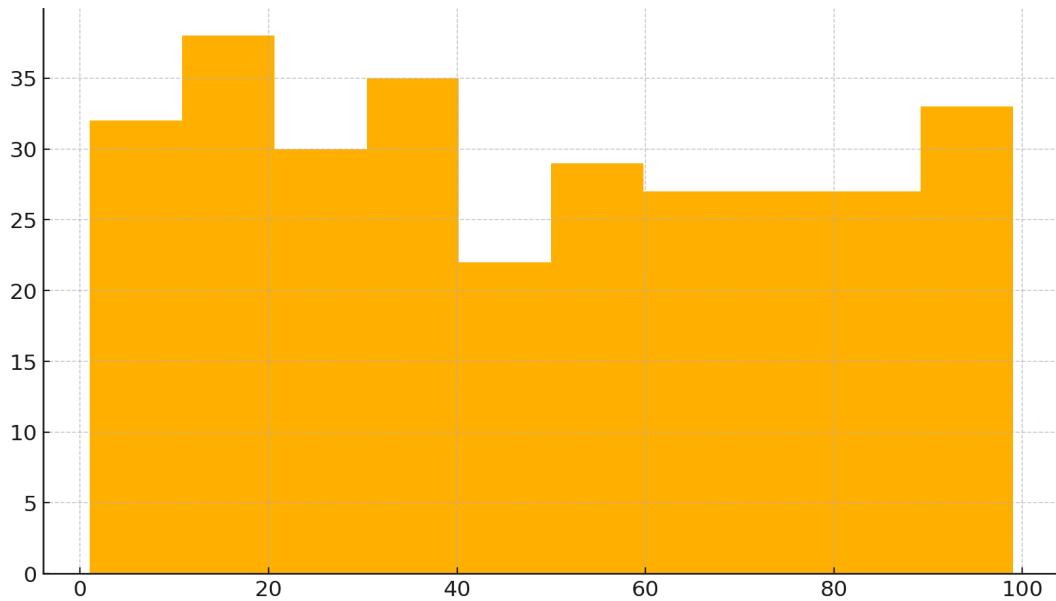
Technologies used: Python, Pandas, NumPy, Matplotlib, Scikit-learn (K-Means), ReportLab for PDF formatting.



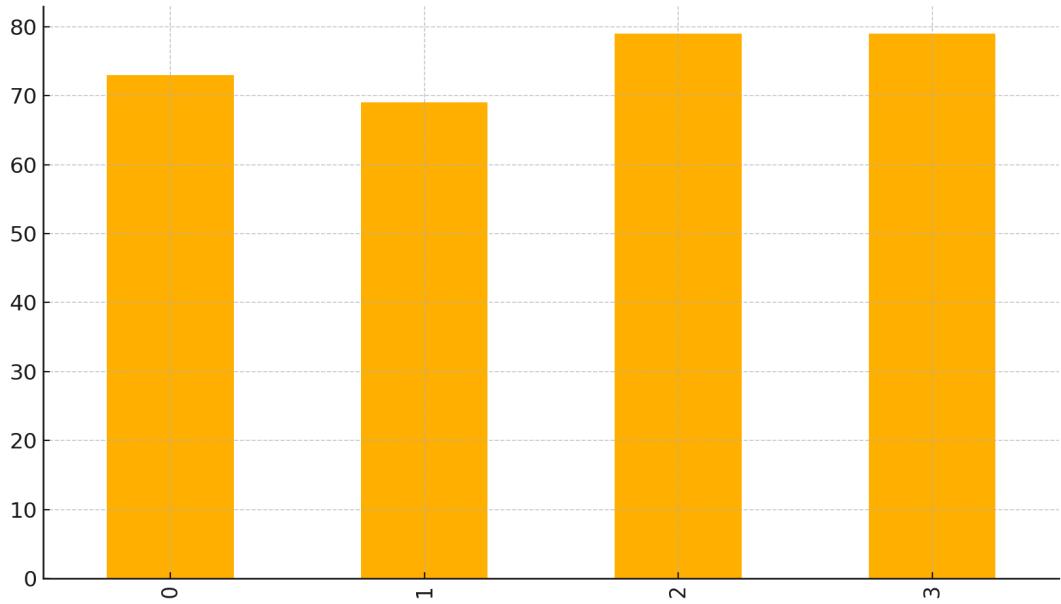
Age distribution helps identify dominant age groups and understand generational behavior. It highlights which age ranges are highly represented and supports demographic-based customer strategies.



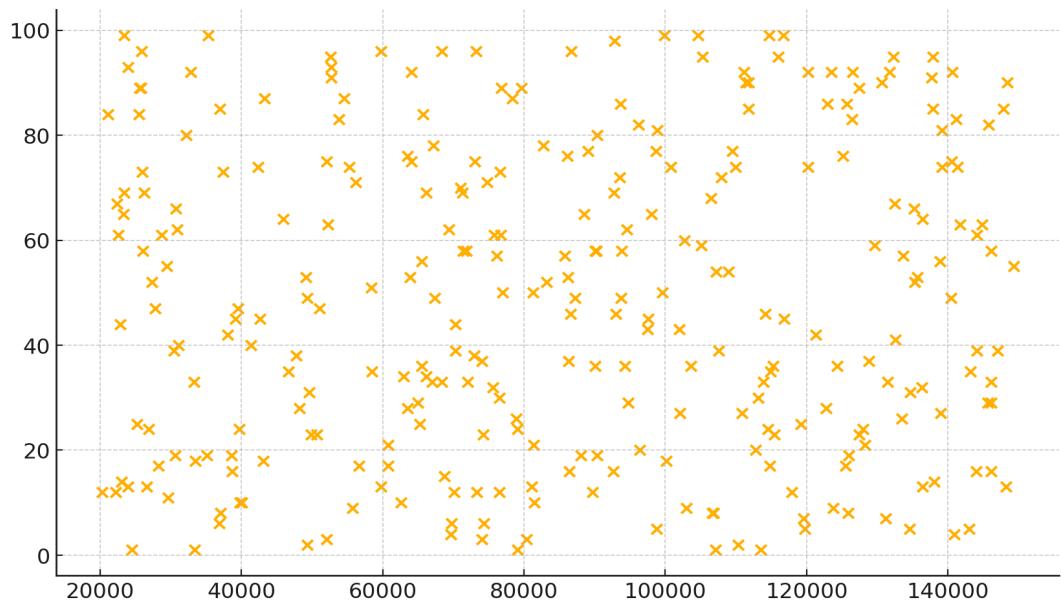
Annual Income distribution reveals purchasing power differences. Knowing income ranges helps businesses plan pricing, promotions, and identify premium or budget-focused customers.



Spending Score distribution represents how actively customers engage in purchases. Higher scores indicate loyal or frequent buyers, while lower scores highlight customers requiring engagement strategies.



Cluster count graph displays the number of customers in each group. Balanced clusters indicate effective segmentation, while imbalances help refine cluster selection or input features.



Scatter plot visualizes how Annual Income and Spending Score work together to form clusters. It shows clear separation between customer groups, confirming meaningful segmentation.

Cluster Insights: Cluster 1 includes premium high-value customers; Cluster 0 contains moderate-income low-spending individuals; Cluster 2 includes younger customers with varied behavior; Cluster 3 consists of low-spending older users.

Business Recommendations: Focus premium offers on high-value clusters, use loyalty programs for low-spending groups, introduce digital campaigns for younger clusters, and apply retention strategies for older customer groups.