Customer Segmentation Using K-Means Clustering

1. Project Overview

This project aims to segment customers into distinct groups based on their age, annual income, and spending score. The insights derived can help businesses tailor marketing strategies, personalize offers, and improve customer engagement.

2. Dataset Overview

Dataset Name: Mall Customers Dataset
Key Features:

- Age: Customer's age in years.
- Annual Income: Customer's income in $1000s.
- Spending Score: A score assigned based on customer behavior (1–100).

3. Problem Statement

The goal is to:
1. Group customers into clusters based on spending behavior and demographics.
2. Analyze and interpret the clusters to inform business strategies.

4. Data Preprocessing Steps

1. Handling Missing Values: Verified that there were no missing values.
2. Feature Scaling: Applied StandardScaler to normalize the data for optimal performance.
3. Dimensionality Reduction: Used PCA (Principal Component Analysis) to reduce dimensions while preserving data variance.

5. Model Selection

Algorithm Used: K-Means Clustering
Model Parameters:
- K = 4 clusters selected based on the Elbow Method and Silhouette Score.
- Random State: 42 for reproducibility.

6. Model Evaluation

Key Metrics:

Metric	Value
Silhouette Score	0.412
Inertia	126.259
Cluster Sizes	58, 47, 60, 35

Alternative Models Tried:

- DBSCAN: Discarded due to poor silhouette score (0.012).
- Gaussian Mixture Model: Did not improve performance significantly.
- K-Means with PCA performed the best.

7. Cluster Analysis

Cluster	Age Range	Average Income ($1000s)	Average Spending Score	Key Insights
0	19–70	70	28	Customers in this group have a wide age range and an average income, but they show low spending behavior.
1	20–68	45	38	This group includes older customers with a moderate income with conservative spending behavior.
2	18–40	49	67	Younger, moderate-income customers who spend a lot.
3	18–40	83	72	High-income young customers with high spending behavior.

8. Business Recommendations

Cluster 0:

Recommendation: This group shows a wide age range with moderate income but low spending behavior, suggesting price sensitivity or a focus on needs over wants.

Action:

Value-focused promotions: Offer discounts, bundle deals, or loyalty programs to encourage more spending.

Targeted campaigns: Promote essential or functional products/services that align with their needs.

Flexible payment plans: Introduce flexible payment options to ease financial burden.

Cluster 1:

Recommendation: Customers in this group have moderate income and slightly above-average spending behavior. They might be willing to spend on non-essentials if the offering is appealing.

Action:

Mid-range product offerings: Tailor products/services to appeal to middle-income earners, focusing on value for money.

Seasonal promotions: Introduce time-limited discounts or seasonal offers to boost their engagement.

Upsell opportunities: Promote higher-value products/services with a reasonable price increase.

Cluster 2:

Recommendation: This group is young, with moderate income and high spending behavior, suggesting they are open to spending on experiences and lifestyle products.

Action:

> Exclusive promotions: Offer personalized promotions or loyalty programs to build brand loyalty.

> Social media and influencer campaigns: Focus on digital marketing to reach them via platforms they frequent (Instagram, TikTok).

> Product bundles: Offer bundled deals that include lifestyle products and experiences at a value price.

Cluster 3:

> **Recommendation:** Young customers with high income and high spending behavior, suggesting a desire for premium and luxury products.

> **Action:**

>> Premium product offerings: Develop and promote high-end, exclusive products that align with their aspirations.

>> Personalized experiences: Offer bespoke services, VIP programs, or loyalty rewards for high-value customers.

>> Luxury marketing strategies: Use targeted ads and influencer partnerships to appeal to their lifestyle choices and social status.

>> Exclusive events: Create exclusive events or early access to new products to cater to their premium tastes.

9. Visualizations
Include visualizations like:
1. Cluster Plots (Age vs Income, Income vs Spending Score).
2. Bar Graphs showing cluster sizes.
3. Scatter Plots showing PCA-reduced data with cluster labels.

10. Final Deliverables
1. Python Notebook (.ipynb) with code, results, and visualizations.
2. Cluster Profiles Report (.pdf) summarizing insights and recommendations.
3. Dataset with Cluster Labels (.csv) for further use.