Customer Segmentation Using Clustering (Python)

1. Project Overview

This project applies K-Means Clustering to segment customers based on demographic and behavioral attributes.
The aim is to enable targeted marketing strategies by identifying distinct customer groups.

2. Problem Statement

The company needs to better understand its customers to optimize marketing efforts.
By clustering customers into meaningful groups, campaigns can be tailored to improve engagement, conversion, and retention.

3. Dataset Details

Dataset: Customer Personality Analysis (Kaggle)
Rows: 2,240
Columns: 29

Key Features:

- **Demographic:** Year_Birth, Education, Marital_Status, Income, Kidhome, Teenhome
- **Behavioral:** Spending on products (Wine, Meat, Gold, etc.), Purchase counts, Recency
- **Campaign Response:** AcceptedCmp1–5, Response
- **Others:** Customer loyalty, Complaints

4. Tools & Libraries Used

- **Python:** Data processing, clustering, visualization
- **Libraries:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn
- **Jupyter Notebook:** Development environment

5. Methodology

Step 1: Data Loading & Preview

- Imported CSV, checked shape, columns, data types, found missing values in Income.

Step 2: Data Cleaning

- Dropped nulls in Income, converted Dt_Customer to datetime, created Customer_Since_Days, dropped unnecessary columns.

Step 3: Encoding

- Applied one-hot encoding to Education and Marital_Status.

Step 4: Scaling

- Standardized features using StandardScaler.

Step 5: Optimal Clusters

- Used **Elbow Method** and chose k = 3.

Step 6: Clustering

- Applied K-Means, added cluster labels.

Step 7: Visualization

- Created PCA 2D scatter plot to visualize clusters.

Step 8: Profiling

- Compared average features across clusters to create customer personas.

6. Cluster Profiles & Marketing Recommendations

Cluster 0 – Luxury Loyalists

- High income, high premium spending, loyal & engaged.
- **Strategy:** VIP programs, exclusive previews, premium upsells.

Cluster 1 – Budget Buyers

- Lowest income, least spending, low engagement.
- **Strategy:** Discounts, bundles, re-engagement campaigns.

Cluster 2 – Value Seekers

- Moderate income & spending, responds to deals.
- **Strategy:** Loyalty rewards, cross-selling, seasonal promos.

After applying K-Means clustering, three distinct customer segments were identified.
The following table summarizes their key characteristics and the suggested marketing strategies for each group.

Cluster	Persona Name	Key Characteristics	Marketing Strategies
0	**Luxury Loyalists**	-Highest income -High spending on premium items -Very loyal and engaged	- VIP loyalty programs - Exclusive product previews - Premium upselling

1	**Budget Buyers**	- Lowest income - Minimal spending - Low engagement	- Discounts and special offers - Bundle promotions - Re-engagement campaigns
2	**Value Seekers**	- Moderate income - Good spending, especially on wines/meats - Responsive to deals	- Loyalty rewards - Cross-selling opportunities - Seasonal promotional offers

These profiles allow the marketing team to tailor campaigns according to each segment's preferences, improving engagement, conversion, and overall customer satisfaction.

7. Conclusion

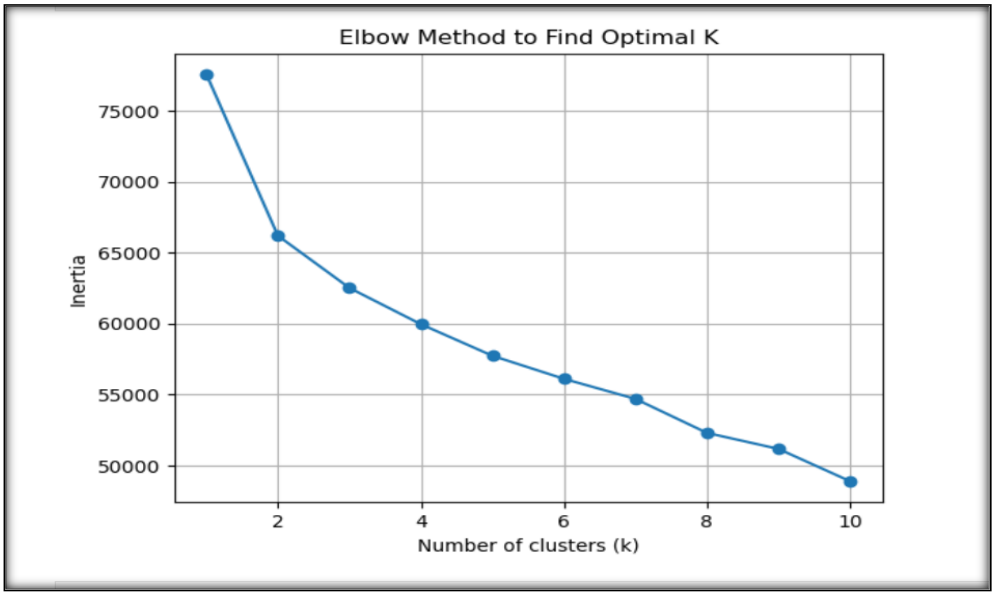
Segmentation provides clear, data-driven customer groups.
Targeted strategies can increase ROI, satisfaction, and retention.

8. Future Work

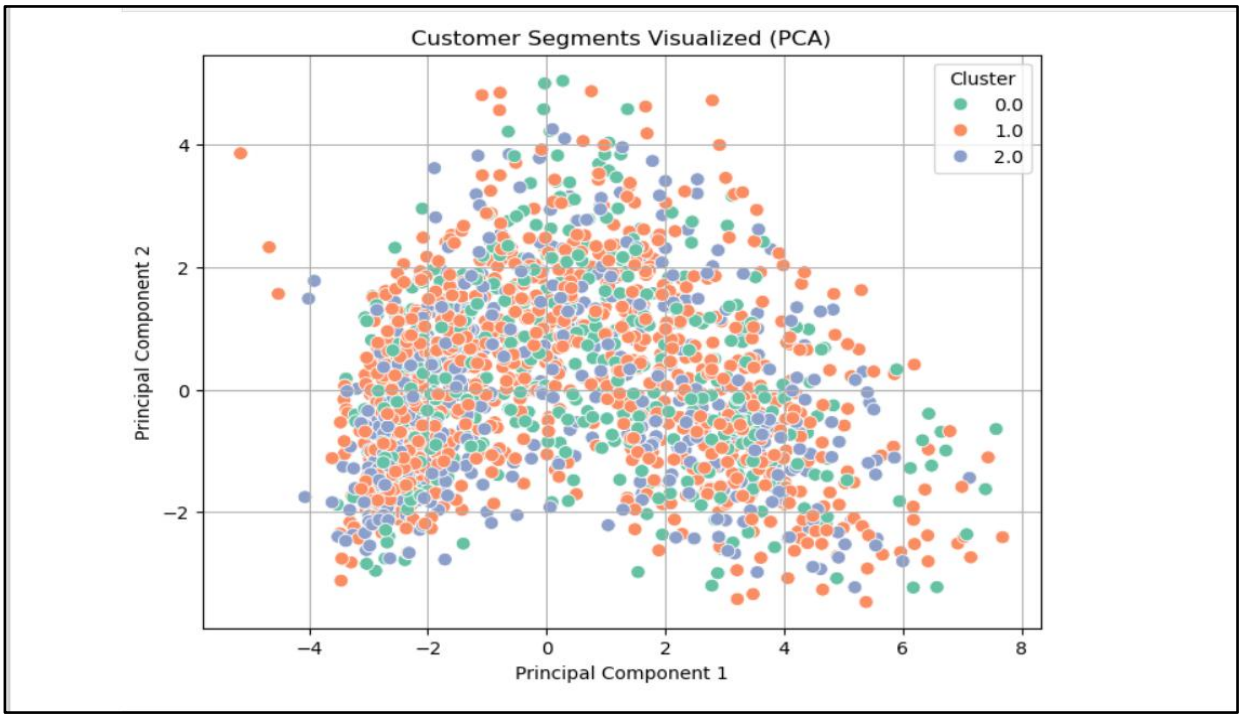
- Add RFM analysis for deeper insights.
- Test other clustering algorithms.
- Track changes in clusters over time.

9. Visualizations

Elbow Method Plot



PCA Cluster Scatter Plot



Cluster-wise Comparison Chart