End-to-End Customer Segmentation Project

Data Analytics & Unsupervised Machine Learning (Grocery Retail Case)

Tools: Python (Pandas, Scikit–learn), PCA, Clustering

| Presenter: Wonjun Ryhu

Why Customer Segmentation?

Targeted Marketing

Deliver personalized promotions and messaging to the right customers.

Budget Efficiency

Allocate marketing resources where they'll have the greatest impact.

Higher Returns

Increase conversion rates and customer lifetime value through precision.

Understanding who your customers are enables smarter business decisions.

Made with GAMMA

Business Problem & Objective

The Challenge

Can we automatically group customers based on behavior, demographics, and household structure?

Our Goal

Identify meaningful customer segments and translate them into actionable marketing strategies.

The Solution

Unsupervised Machine Learning–based customer segmentation.

🗄 DATA FOUNDATION

Dataset & Preparation

01

Dataset Overview

2,000+ grocery store customers with demographics, household info, product spending, and campaign responses.

02

Data Cleaning

Removed missing values and extreme outliers to ensure data quality.

03

Feature Creation

Converted dates, engineered customer tenure, created Age, Total Spend, Family Size, and Is Parent features.

Feature Engineering & ML Preprocessing
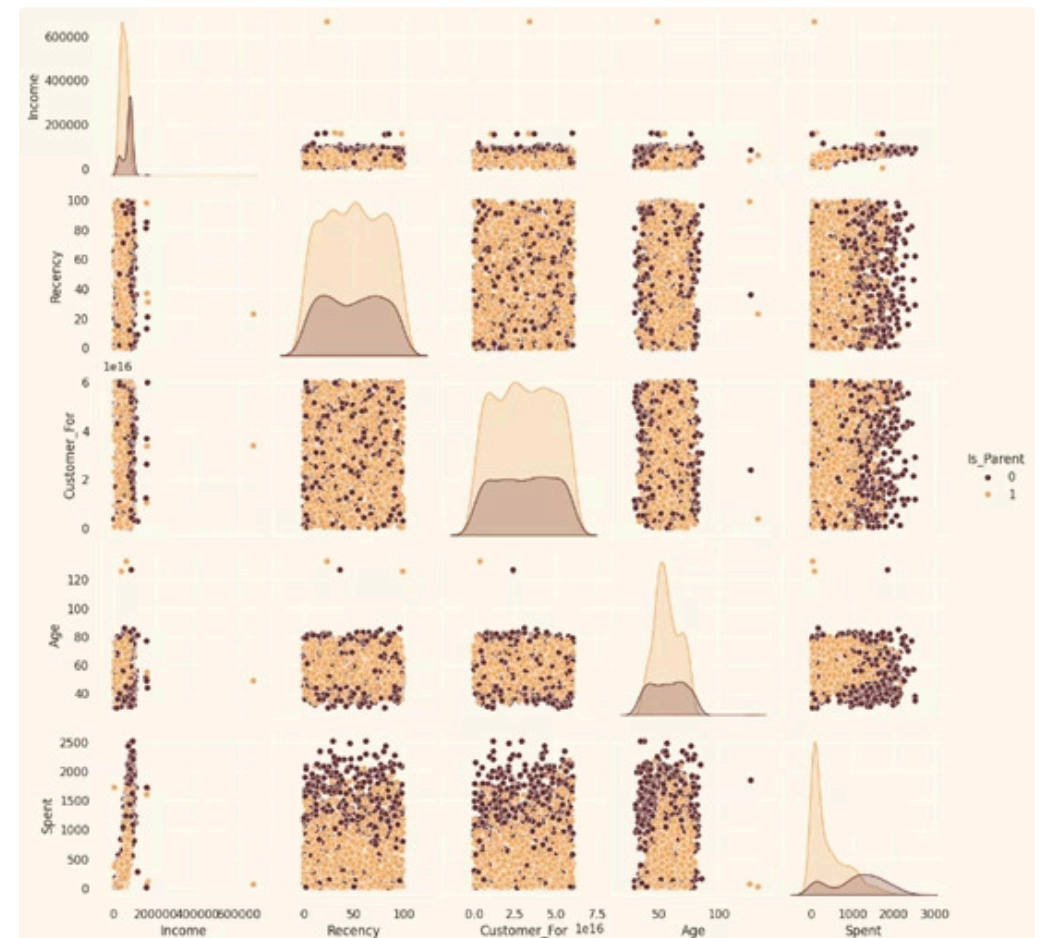
Engineered Features

- Age and Total Spend
- Children and Family Size
- Living With status

Encoding & Scaling

- Label Encoding for Education and Living With
- StandardScaler for equal feature importance

Result

Clean, fully numerical dataset ready for machine learning algorithms.

Dimensionality Reduction & Clustering

PCA Applied

Reduced features to 3 principal components while preserving underlying information.
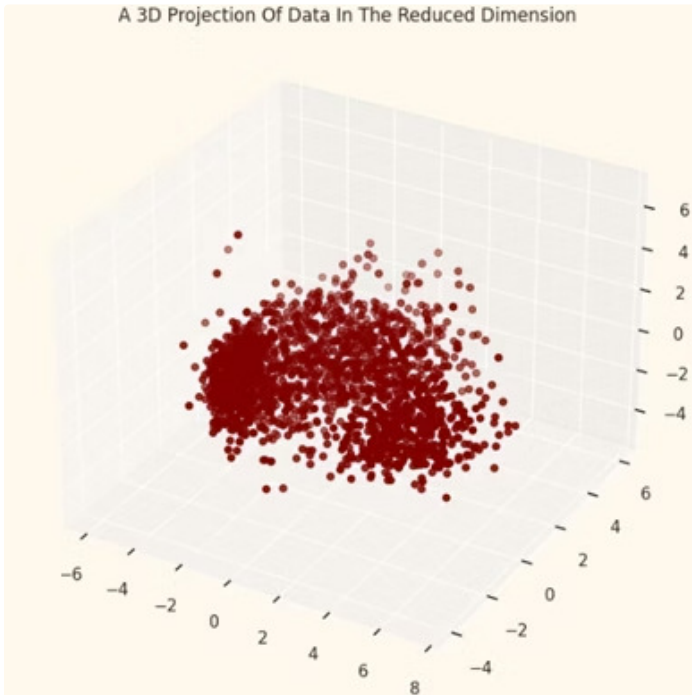
Elbow Method

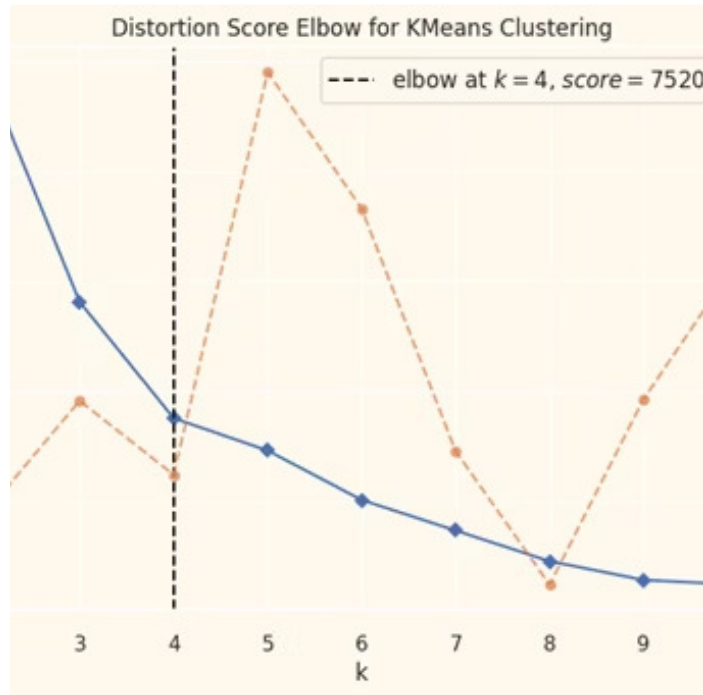Determined optimal number of clusters: 4 distinct groups.

Clustering

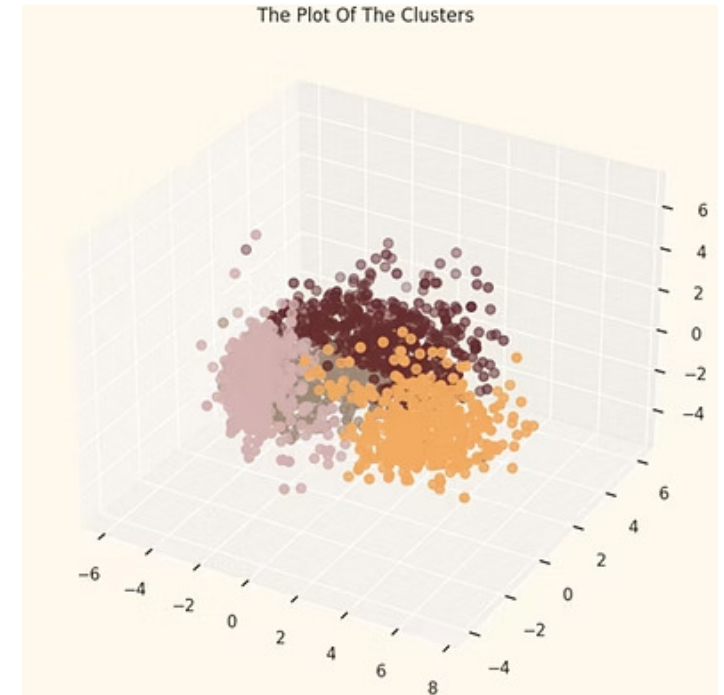Agglomerative Hierarchical Clustering algorithm applied.

A 3D Projection Of Data In The Reduced Dimension

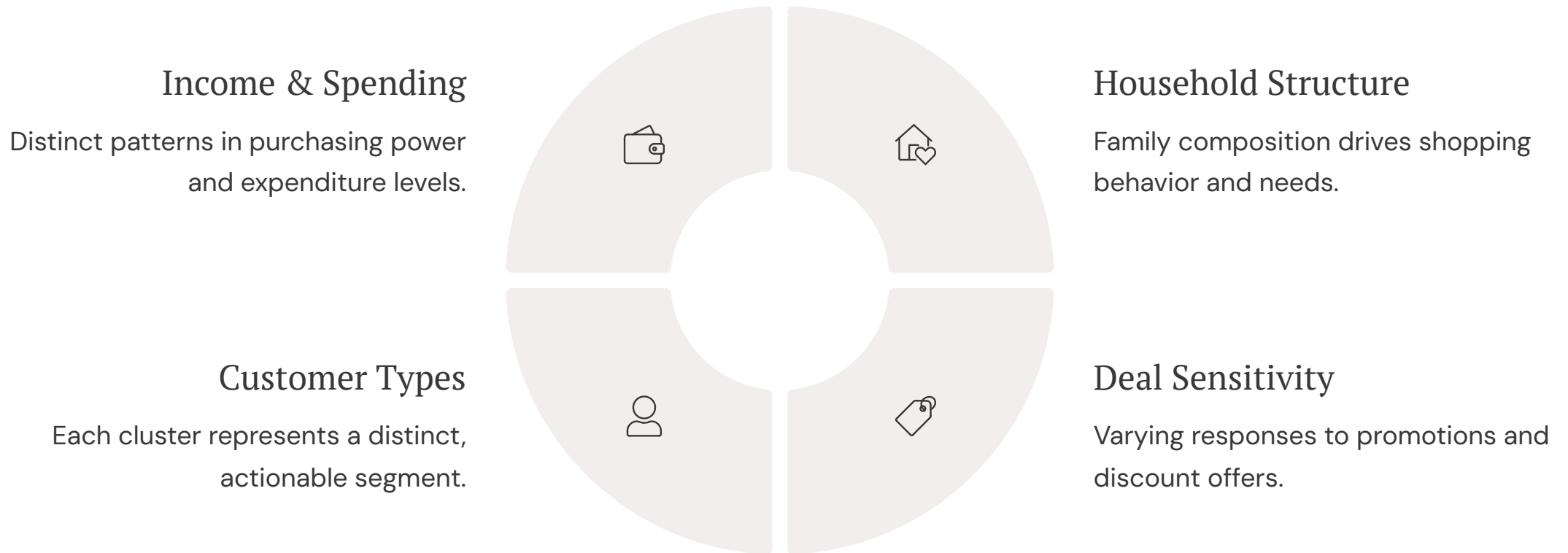Distortion Score Elbow for KMeans Clustering

--- elbow at $k = 4$, $score = 7520$

The Plot Of The Clusters

Cluster Overview

Customers are clearly separated based on income, spending behavior, household structure, and promotion sensitivity.

Income & Spending

Distinct patterns in purchasing power and expenditure levels.

Household Structure

Family composition drives shopping behavior and needs.

Customer Types

Each cluster represents a distinct, actionable segment.

Deal Sensitivity

Varying responses to promotions and discount offers.

Cluster 0: Deal-Driven Family Shoppers

Deal-Driven Family Shoppers

Profile: Parents in 2–4 person households, typically older with teenagers. Highly responsive to discounts.

Marketing Strategy

- Multi-buy offers (Buy 2 Get 1 Free)
- Family-size bundles and meal kits
- Traditional flyers + digital campaigns

Cluster 1: Young Families, Low Spending

Young Families, Low Spending

Profile: Parents with one young child, smaller households, younger age group with lower overall spending.

Marketing Strategy

- Child-focused convenience promotions
- App-based coupons and social media
- Fresh, organic, baby-related products

Cluster 2: High-Income, Non-Parent Segment

High-Income, Non-Parent Segment

Profile: Couples or individuals with high income and low discount sensitivity.

Strategy

- Premium and specialty products
- Experience–focused marketing
- Curated gourmet meal kits

Cluster 3: High-Spending Parents, Budget Constraints

High-Spending Parents, Budget Constraints

Profile: Larger families with teenagers, lower income but high spending needs.

Strategy

- Private-label products and bulk discounts
- Strong loyalty programs and weekly deals
- Emphasize cost savings and value

Key Takeaways

ML-Powered Insights

Unsupervised machine learning enables powerful, data-driven customer segmentation.

Technical Framework

PCA + clustering transforms complex customer data into actionable insights.

Business Impact

Real value comes from connecting clusters to targeted marketing actions.

Reusable Solution

Framework applies across retail, e-commerce, and subscription services.

Outcome: A portfolio-ready, end-to-end analytics project with real business impact.