CUSTOMER SEGMENTATION AND PURCHASE BEHAVIOR ANALYSIS FOR A REGIONAL SUPERMARKET CHAIN

1. Background and Problem Statement:

A regional supermarket chain operating across three states in India faced challenges in understanding its diverse customer base. While loyalty programs were active, they were generic and yielded low redemption rates. The marketing team lacked clarity on spending behavior, frequency of visits, and category preferences across different customer types. This project was initiated to perform **customer segmentation** using purchase data and to provide strategic insights for differentiated loyalty offers and personalized promotions.

2. Objectives:

- Segment the customer base based on actual purchase behavior
- Identify distinct clusters with actionable characteristics (frequency, value, category preference)
- Provide marketing and operations teams with data-driven profiles for targeted campaign design
- Recommend loyalty program improvements based on behavioral insights

3. Methodology:

3.1 Data Description

- **Data Source:** Point-of-sale (POS) system and loyalty card database
- **Size:** 58,920 customer records, 1.2 million transactions over 12 months
- **Key Fields:**
    - Customer ID
    - Transaction date
    - Total basket value
    - Items purchased per category (e.g., dairy, snacks, frozen foods)
    - Loyalty points earned/redeemed

- Store location

3.2 Data Cleaning and Feature Engineering

- Handled nulls in loyalty activity and filtered out one-time test or employee purchases
- Aggregated data per customer into:
  - **Avg. basket size**
  - **Transaction frequency per month**
  - **Share of spend per category**
  - **Recency of last purchase**
- Normalized variables for clustering using StandardScaler in Python

3.3 Segmentation Technique

- Applied **K-Means clustering** on engineered features
- Used Elbow Method and Silhouette Score to select optimal number of clusters (**k=4**)
- Post-clustering validation done via boxplots and category heatmaps

4. Results: Customer Segments Identified

Cluster	Label	Size	Characteristics
C1	**Value-Focused Shoppers**	34%	Low spend, high frequency, high coupon usage, snacks/dairy heavy
C2	**Bulk Monthly Buyers**	22%	High basket size, low frequency, buy in bulk, dry goods and household products
C3	**Premium Category Loyalists**	18%	Medium-high spend, regular buyers of frozen, gourmet, and imported products
C4	**Occasional Deal Hunters**	26%	Irregular frequency, high discount redemption, mix of impulse category purchases

- **C3** customers were the most profitable, showing high loyalty and category stability
- **C1 and C4** were the most price-sensitive and highly influenced by discounts
- **C2** showed low engagement with loyalty programs but contributed significant revenue per visit

5. Recommendations:

5.1 Loyalty Program Personalization

- **C1:** Weekly cashback incentives on repeat purchases, category-based milestone rewards
- **C2:** Monthly reminder campaigns with personalized basket refill checklists
- **C3:** Exclusive premium member offers, invite-only tasting events, bundled gourmet deals
- **C4:** Time-limited deep discount offers to increase visit frequency

5.2 In-Store Promotion Strategy

- Target C3 at gourmet aisle and checkout areas with bundled premium offers
- Place QR codes with discounts near dairy/snack shelves for C1
- Use store-level CRM to automate SMS reminders for C2 stock-up weeks

5.3 Email Campaigns and Retargeting

- Segment email list based on clusters
- Test cluster-specific creatives and offers through A/B testing
- Sync cluster segments with Facebook/Instagram ad audiences for remarketing

6. Outcome

- Predicted a **15–22% increase in redemption rate** of loyalty program within 3 months
- C4 reactivation projected to improve footfall by **up to 9%** in lower-performing stores
- Tailored campaign ROI expected to increase from **1.4x to 2.3x** over 6 months based on uplift models

7. Future Work:

- Integrate demographic data (age, family size, pin code) to build hybrid clusters
- Apply Market Basket Analysis per cluster to recommend bundle offers
- Develop RFM score model to layer on top of behavioral clusters

8. Stakeholder Relevance:

Academic:

- Case study in unsupervised learning (K-Means), suitable for data mining and customer analytics courses
- Demonstrates practical application of feature engineering and cluster profiling

Corporate:

- Equips retail managers and CRM teams with actionable segmentation for loyalty, promotions, and assortment planning
- Enables transition from generic offers to **data-driven, segment-specific personalization**