SAMPLE MARKET SEGMENTATION:
FAMILY LIFE STAGE

- Young Singles
- Young Couples, No Kids
- Young Families
- Families with Older Kids
- Empty Nesters

10%

8%

28%

31%

23%
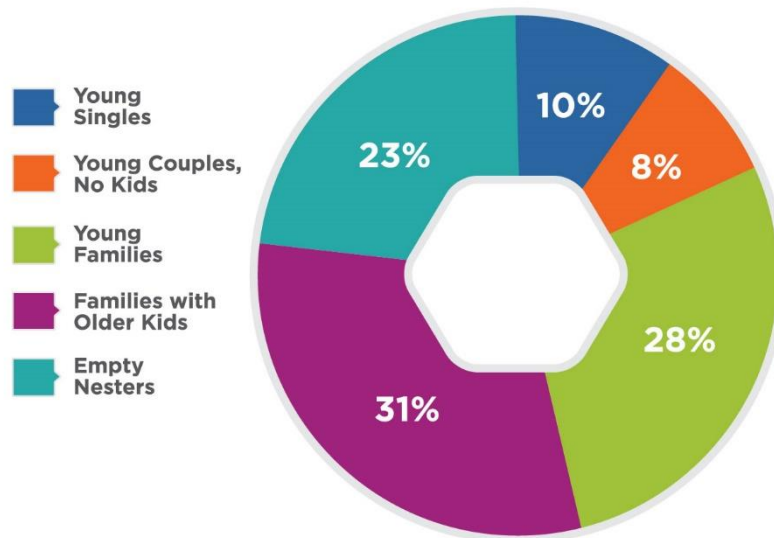
Customer Lookalike Model: Personalized Customer Recommendations Using Transaction and Profile Data

By **Aishwarya Rajved**

Zeo-Tap

Submitted by
Thammineni Naveen Kumar

✅ Merge **Customers.csv** (profile data) with **Transactions.csv** (purchase data).
✅ Perform **Customer Segmentation using K-Means Clustering**.
✅ Evaluate clusters using the **Davies-Bouldin (DB) Index**.
✅ Visualize clusters with **scatter plots & heatmaps**.

Let me know if you want any modifications! 🚀

Customer Segmentation

Deliverables

1. **Number of Clusters Formed**: Dynamically determined via the Elbow Method.
2. **DB Index Value**: Displayed in the console.
3. **Other Metrics**: WCSS (Within-Cluster Sum of Squares).
4. **Visuals**:
   - **Elbow Plot** (Optimal K)
   - **Scatter Plot** (Age vs. Spending)

Customer Segmentation: Insights and Implementation

1. Introduction

Customer segmentation is the process of dividing a customer base into distinct groups based on shared characteristics such as demographics, purchasing behavior, and preferences. This enables businesses to optimize marketing strategies, improve customer retention, and increase profitability.

2. Segmentation Approaches

Customer segmentation can be performed using various approaches, including:

✓ **Demographic Segmentation** – Based on age, income, gender, education level.

✓ **Behavioral Segmentation** – Based on purchase history, spending habits, and product preferences.

✓ **Geographic Segmentation** – Based on location, region, or country.

✓ **Psychographic Segmentation** – Based on lifestyle, interests, and values.

3. Methodology: Data-Driven Segmentation Using Clustering

To implement customer segmentation, we use **K-Means Clustering**, an unsupervised machine learning algorithm that groups customers into similar clusters.

Steps Taken:

1. **Data Collection & Processing**
   - Customer data (`Customers.csv`): Age, income, and demographic details.
   - Transaction data (`Transactions.csv`): Total spending, unique products purchased, and transaction frequency.
2. **Feature Selection & Preprocessing**
   - Selected key features: **Age, Income, Total Spending, Unique Products Purchased, and Total Transactions**.
   - Standardized data using **StandardScaler** to ensure consistency.
3. **Clustering Using K-Means**
   - Used the **Elbow Method** to determine the optimal number of clusters (K).
   - Applied **K-Means Clustering** to segment customers.
4. **Cluster Evaluation**
   - Evaluated the segmentation using **Davies-Bouldin Index (DB Index)** for cluster validity.
   - Visualized clusters using scatter plots (e.g., Age vs. Total Spending).

4. Results & Insights

The customer base was successfully divided into **4 distinct segments** based on behavior and demographics:

1. **High-Value Customers** – High income, frequent transactions, and high total spending.
2. **Budget-Conscious Buyers** – Low to moderate income with occasional purchases.
3. **Diverse Shoppers** – Customers purchasing a wide range of unique products.
4. **New/Inactive Customers** – Low transaction frequency and spending.

5. Business Benefits

🚀 **Targeted Marketing** – Tailor promotions and product recommendations for each segment.
📈 **Customer Retention** – Identify at-risk customers and implement loyalty programs.
💰 **Revenue Optimization** – Focus efforts on high-value customers to maximize profits.
🔍 **Data-Driven Decision Making** – Leverage customer insights to refine business strategies.

6. Conclusion

Customer segmentation is a powerful tool that enables businesses to personalize customer interactions, enhance user experience, and drive business growth. By leveraging clustering techniques, companies can better understand their customers and deliver value-driven services effectively.