NoA connect recruitment task

Customer segmentation with unsupervised machine learning

Methods

1. Clean and preprocess transaction data (handle missing values, standardize date formats)
2. Feature engineering from transaction data:
    - Recency: Days since last purchase
    - Frequency: Number of transactions per customer
    - Monetary: Average purchase amount
    - Purchase variability: Standard deviation of purchase amounts
    - Currency diversity: Number of different currencies used

Exploratory Data Analysis

1. Univariate analysis of each engineered feature
2. Correlation analysis between features
3. Check for skewness and outliers that might affect clustering

Dimensionality Reduction

Standardizing features and performs PCA analysis with these key steps:

1. Selects core RFM features for clustering
2. Standardizes the features using StandardScaler
3. Applies PCA to reduce dimensionality
4. Creates visualizations to understand:
    - Explained variance by each principal component
    - Customer distribution in the first two principal components
    - A biplot showing how original features contribute to the principal components

K-means Path:

1. Determine optimal number of clusters using:

    - Elbow method (inertia vs. number of clusters)
    - Silhouette analysis
    - Gap statistic
2. Run K-means with optimal k
3. Evaluate stability of clusters with bootstrapping
4. Validate clusters using silhouette coefficient and Davies-Bouldin index

K-means clustering with comprehensive validation, including:

1. Determining optimal cluster count using:
    - Elbow method (inertia plot)
    - Silhouette score analysis
    - Davies-Bouldin index
    - Gap statistic
2. Cluster stability analysis through bootstrapping that:
    - Runs K-means on multiple bootstrap samples
    - Measures how consistently customers are assigned to the same cluster
    - Calculates stability score for each cluster
3. Cluster quality validation with:
    - Silhouette coefficient (both overall and per-cluster)
    - Davies-Bouldin index
    - Silhouette plot visualization
4. Visualization of clusters:
    - PCA-based 2D scatter plot of customer segments
    - Radar chart showing cluster profile characteristics

Customer Segment Analysis

Based on your clustering results, I can identify two distinct customer segments:

Cluster 0: "Occasional Low-Value Shoppers" (83.56% of customers)

- **High recency** (255 days since last purchase)
- **Very low frequency** (1.7 purchases)
- **Low monetary value** ($62.65 average order)
- **Low purchase variability** ($11.23)
- **Very short tenure** (13.4 days)

This is your largest segment by far, representing customers who made very few purchases in a short period and haven't returned in a long time.

Cluster 1: "High-Value Loyal Customers" (16.44% of customers)

- **Moderate recency** (94 days since last purchase)
- **High frequency** (34 purchases)
- **High monetary value** ($188.47 average order)
- **High purchase variability** ($153.01)
- **Long tenure** (248 days)

This smaller but valuable segment represents your loyal customers who purchase frequently, spend more per order, and have maintained a relationship with your business for a much longer time.

The stark contrast between these segments suggests a significant opportunity to develop strategies for moving customers from Cluster 0 to Cluster 1.

Strategies to Convert Occasional Shoppers to Loyal Customers

Win-Back & Reactivation

1. **Targeted win-back campaigns** for Cluster 0 customers inactive for >200 days
2. **Time-limited discounts** on products similar to their past purchases
3. **"We miss you" personalized communications** highlighting new offerings

Value Enhancement

1. **Free shipping thresholds** set just above their average order value
2. **Bundle discounts** to encourage larger purchases
3. **Product recommendations** based on past purchases to increase basket size

Loyalty Development

1. **Early loyalty program enrollment** with immediate benefits
2. **Second purchase incentives** (critical for moving to repeat customer status)
3. **Post-purchase follow-ups** to establish engagement patterns

Retention Tactics

1. **Subscription options** for frequently purchased items
2. **Milestone rewards** to encourage reaching the tenure of Cluster 1
3. **Educational content** about products to increase customer investment

Monitoring & Measurement

1. **Transition scoring** to track movement from Cluster 0 to 1
2. **Churn prediction models** to identify at-risk customers before they lapse
3. **CLV projection updates** as customers show loyalty signals

These strategies should be A/B tested and refined based on conversion performance metrics.