**Detailed Report on Customer Segmentation and Analysis**

**1. Data Overview**

The dataset customer\_orders.csv contains information about customers' orders and spending. The key columns include:

* customer\_name: Anonymized customer identifier
* total\_spent: Total amount spent by the customer
* total\_orders: Total number of orders placed by the customer
* average\_purchase: Average amount spent per purchase
* average\_monthly\_spend: Average monthly spending
* average\_monthly\_orders: Average number of monthly orders
* last\_ordered: Date of the last order
* customer\_since: Date since the customer has been with the company

**2. Data Cleaning and Preparation**

* **Column Standardization:** Column names were standardized to lowercase with underscores replacing spaces.
* **Anonymization:** Customer names were anonymized for privacy.
* **Type Conversion:** Numeric columns were cleaned of special characters and converted to floats. Date columns were converted to datetime format.
* **Feature Engineering:** Two new features were created:
  + customer\_tenure: Number of days between the customer's first and last order.
  + recency: Number of days since the customer's last order.

**3. Summary Statistics**

The dataset includes 5025 records with the following summary statistics:

| **Metric** | **Mean** | **Median** | **Std Dev** | **Min** | **Max** |
| --- | --- | --- | --- | --- | --- |
| Total Spent | $91.95 | $35.51 | $192.98 | $4.45 | $2995.82 |
| Total Orders | 3.32 | 1.0 | 8.34 | 1 | 220 |
| Average Purchase | $34.85 | $22.95 | $41.09 | $4.45 | $911.02 |
| Average Monthly Spend | $4.66 | $1.49 | $13.41 | $0.05 | $529.28 |
| Average Monthly Orders | 0.03 | 0.0 | 0.22 | 0 | 5 |
| Customer Tenure (days) | 819.21 | 530.0 | 898.43 | 0 | 3845 |
| Recency (days) | 1106.43 | 922.0 | 777.02 | 192 | 3513 |

**4. Outlier Analysis**

* Boxplots for numeric columns indicated the presence of outliers, particularly in total\_spent and total\_orders.

**5. Clustering Analysis**

* **Standardization:** The numeric features were standardized using StandardScaler.
* **Elbow Method and Silhouette Score:** To determine the optimal number of clusters, both the Elbow Method and Silhouette Score were used. The optimal number of clusters was found to be 4.
* **KMeans Clustering:** KMeans clustering was performed with 4 clusters.

**6. Cluster Analysis**

**Cluster Summary:**

| **Cluster Label** | **Total Spent (Mean)** | **Total Orders (Mean)** | **Recency (Mean)** | **Customer Tenure (Mean)** |
| --- | --- | --- | --- | --- |
| 0 | $65.31 | 2.28 | 697.33 | 974.16 |
| 1 | $45.83 | 1.54 | 2059.21 | 362.26 |
| 2 | $1725.35 | 84.48 | 426.36 | 1978.44 |
| 3 | $557.33 | 19.64 | 519.35 | 1547.97 |

**Detailed Descriptive Statistics for Clusters:**

**Cluster 0:**

* Lower spenders with low total orders.
* Moderate recency and customer tenure.

**Cluster 1:**

* Lowest spenders with the least number of orders.
* Highest recency indicating they haven't ordered recently.
* Short customer tenure.

**Cluster 2:**

* Highest spenders with the highest number of orders.
* Lower recency indicating recent orders.
* Long customer tenure.

**Cluster 3:**

* Medium to high spenders with moderate number of orders.
* Moderate recency and customer tenure.

**Visualization of Clusters:**

* **Total Spent by Cluster:**
* **Total Orders by Cluster:**
* **Pairplot for Relationships between Variables in Clusters:**
* **PCA for Clusters:**

**7. Correlation Analysis**

* The correlation matrix highlighted the relationships between different numeric features. Notably:
  + Total Spent had a strong positive correlation with Total Orders (0.69).
  + Average Monthly Spend had a moderate correlation with Total Spent (0.44).

**8. Cluster Profiles**

**Cluster 2 Profile:**

* **Total Spent:** Mean: $1725.35, Median: $1522.04
* **Total Orders:** Mean: 84.48, Median: 73.0
* **Recency:** Mean: 426.36 days
* **Customer Tenure:** Mean: 1978.44 days

**Cluster 3 Profile:**

* **Total Spent:** Mean: $557.33, Median: $466.16
* **Total Orders:** Mean: 19.64, Median: 18.0
* **Recency:** Mean: 519.35 days
* **Customer Tenure:** Mean: 1547.97 days

**Cluster 0 Profile:**

* **Total Spent:** Mean: $65.31, Median: $37.27
* **Total Orders:** Mean: 2.28, Median: 1.0
* **Recency:** Mean: 697.33 days
* **Customer Tenure:** Mean: 974.16 days

**Cluster 1 Profile:**

* **Total Spent:** Mean: $45.83, Median: $28.23
* **Total Orders:** Mean: 1.54, Median: 1.0
* **Recency:** Mean: 2059.21 days
* **Customer Tenure:** Mean: 362.26 days

**9. Distribution of Total Spent**

* The distribution of total spent showed a right-skewed distribution, indicating most customers spend a relatively low amount, with a few high spenders.

**10. Recommendations**

**Target High-Value Customers (Cluster 2):**

* Implement loyalty programs and exclusive offers to retain high spenders.
* Personalize marketing campaigns to encourage repeat purchases.

**Engage Medium-Value Customers (Cluster 3):**

* Identify opportunities to upsell and cross-sell.
* Provide incentives for increased spending and orders.

**Re-engage Low-Value Customers (Clusters 0 and 1):**

* Analyze the reasons for low engagement and address pain points.
* Offer promotions and discounts to stimulate purchases.

**Additional Insights and Recommendations:**

**Customer Segmentation:**

K-means clustering was employed to segment customers into distinct groups based on their spending behavior, frequency of orders, and recency of purchases. The distribution of customers across clusters is as follows:

* **Cluster 0:** 3204 customers
* **Cluster 1:** 1547 customers
* **Cluster 2:** 25 customers
* **Cluster 3:** 249 customers

Each cluster represents a different type of customer, ranging from new customers to high-value infrequent customers.

**Customer Characteristics by Cluster:**

Detailed characteristics of customers in each cluster include:

* Total Spent
* Total Orders
* Average Purchase
* Average Monthly Spend
* Average Monthly Orders
* Customer Tenure (days)
* Recency (days)

**Retention Strategies by Cluster:**

Based on the characteristics of each cluster, here are some potential strategies:

**Cluster 0 (New Customers):**

* **Customer Education:** Provide resources or tutorials to help them understand and get the most out of your product or service.
* **Early Engagement:** Engage them early with excellent customer service and support to build a strong relationship.
* **Follow-up Communication:** Regularly check in with personalized messages or offers to keep your brand at the top of their mind.

**Cluster 1 (Inactive Customers):**

* **Win-back Campaigns:** Implement campaigns specifically designed to win back lapsed customers. This could involve special offers or highlighting new features or products since they last shopped.
* **Feedback Loop:** Reach out to them to understand why they stopped purchasing. Their feedback can provide valuable insights to prevent other customers from becoming inactive.

**Cluster 2 (Loyal Customers):**

* **Loyalty Programs:** Implement a loyalty program that rewards them for their repeat business. This could involve discounts, exclusive offers, or early access to new products.
* **Customer Appreciation:** Show appreciation for their loyalty through personalized thank you messages, gifts, or special recognition.

**Cluster 3 (High-Value Infrequent Customers):**

* **Personalized Communication:** Use the data you have about their purchases to personalize your communication and make them feel valued.
* **Exclusive Offers:** Provide exclusive offers or early access to new products or services that match their purchasing habits.
* **Outstanding Customer Service:** High-value customers expect high-quality service. Ensure they have a dedicated point of contact or priority support to keep them satisfied.

**Action Steps for Stakeholders:**

Implement the recommended retention strategies tailored to each customer cluster. This involves:

* Monitoring the effectiveness of these strategies through metrics such as customer retention rate and CLV.
* Continuously analyzing customer data and refining strategies based on evolving customer behavior.

**Data Export:**

For further analysis and targeted actions, the segmented customer data has been exported to an Excel file (customer\_segments.xlsx), allowing stakeholders to delve deeper into each cluster.

**Conclusion:**

Understanding customer behavior and preferences is crucial for devising effective retention strategies. By leveraging data-driven insights and segmentation techniques, businesses can optimize their efforts to retain customers and foster long-term relationships.

**Next Steps:**

* Implement the proposed retention strategies and monitor their impact on customer retention metrics.
* Continuously analyze customer feedback and behavior to refine strategies and adapt to changing market dynamics.
* Explore advanced analytics techniques to gain deeper insights into customer segments and enhance personalization efforts.

This report provides a comprehensive overview of customer segmentation and insights that can guide strategic decisions to enhance customer engagement and business growth.

A graph of a chart

Description automatically generated with medium confidence

A graph with a line

Description automatically generatedA screenshot of a graph

Description automatically generatedA graph showing a number of clusters

Description automatically generated