



REPORT ON

**CUSTOMER SEGMENTATION
AND A/B TESTING ANALYSIS**

Oluwasegun Adegoke



Executive Summary

This project aims to enhance business insights and increase customer engagement by applying customer segmentation and running targeted marketing campaigns. Through K-means clustering, distinct customer segments were identified, followed by A/B testing to determine the effectiveness of offering a discount to specific groups. The results revealed valuable insights into customer behavior and responsiveness to promotional strategies, enabling more efficient allocation of marketing resources.

Project Objective and Requirements

The primary objective of this project is to segment the customer base, identify specific groups, and determine the effectiveness of targeted promotions. We seek to answer whether a discount incentive can effectively boost engagement and sales, specifically for the identified customer segments.

Data Requirements

- Customer demographics (age, gender, location).
- Purchase transaction history (frequency, order value).
- Traffic sources for acquisition (organic search, social media).

Technical Requirements

- Python for data manipulation, statistical analysis, and clustering.
- SQL for querying and extracting data.
- Visualization tool (Tableau) for presenting insights.

Outcome Requirements

- Define and validate customer segments through clustering.
- Execute A/B tests to measure strategy effectiveness.
- Provide actionable marketing insights to enhance customer engagement and optimize retention.

Methodology

The project workflow involved multiple stages: data preparation, feature engineering, customer segmentation, A/B testing, and visualization.

- **Data Sources and Extraction:** The dataset, provided through BigQuery, contained customer transactional and demographic information. Relevant data on customer behavior, transaction history, and traffic sources were extracted, focusing on active purchases and removing records with incomplete critical data points.
- **Data Preparation:** Initial datasets contained customer purchase behavior and demographic information. Data cleaning was performed to handle missing values and ensure consistency.
- **Feature Engineering:** Several key features were engineered, including:
 - **Average Order Value:** Calculated average spends per order for each customer.
 - **Purchase Recency:** Measured days since the last purchase to evaluate customer engagement.
 - **Preferred Product Category & Traffic Source:** Determined the most purchased product categories and how customers accessed the store.
 - **Total Orders & Time of Day Preference:** Counted total orders and noted preferred time of day for purchases.

Clustering Methodology

I used K-Means Clustering to segment customers based on purchasing patterns, including metrics like total orders, average order value, and traffic source. After normalization and determining the optimal number of clusters with the elbow method, we segmented customers into four clusters.

Preprocessing Steps:

- **Feature Scaling:** To prepare the data for clustering, **StandardScaler** was used to normalize the features. This step was crucial as clustering methods are sensitive to feature magnitudes, and feature scaling ensured all attributes contributed equally to the model.
- **Dimensionality Reduction Using PCA: Principal Component Analysis (PCA)** was applied to reduce the high-dimensional feature set into a smaller number of principal components while retaining as much variance as possible. PCA helped reduce computational complexity and made the clustering more interpretable by summarizing the information into two principal components.
- **Determining the Optimal Number of Clusters Using the Elbow Method:** To decide on the number of clusters, the **elbow method** was employed. We plotted the within-cluster sum of squares (WCSS) against the number of clusters and observed the "elbow point" where the decrease in WCSS slowed down. This indicated an optimal balance between the number of clusters and model complexity.

After these steps, clustering was applied to create four distinct customer segments:

1. **Cluster 0:** "Low Value, Dormant Shoppers" - Customers with low average order values and who haven't purchased recently.
2. **Cluster 1:** "High Value, Frequent Shoppers" - Customers with high spending and frequent purchases.
3. **Cluster 2:** "Moderate Value, Engaged Shoppers" - Moderate spenders with recent purchases.
4. **Cluster 3:** "Moderate Value, Highly Dormant Shoppers" - Moderate spenders who haven't engaged recently.

A/B Testing Methodology

Objective: The main goal of the A/B testing was to determine whether offering a 10% discount would lead to increased engagement and spending among specific customer segments, particularly "High Value, Frequent Shoppers" and "Low Value, Dormant Shoppers."

Steps:

1. **Hypothesis:** A 10% discount will have a positive impact on sales and engagement for selected segments.
2. **Experimental Design:**
 - The "High Value, Frequent Shoppers" and "Low Value, Dormant Shoppers" segments were selected for testing.
 - Customers were divided into two groups: a **control group** (no discount) and a **treatment group** (received a 10% discount).
3. **Data Collection:** Customer behavior metrics such as **average order value** was collected for both the control and treatment groups.
4. **Statistical Analysis:**
 - **Independent t-tests** were used to compare the control and treatment groups to determine if the observed differences were statistically significant.
 - Metrics tested included **Average Order Value**, providing insight into how different segments reacted to the discount.

Results and Insights

Average Order Value: The A/B test showed that the discount led to a significant increase in average order value for the "High Value, Frequent Shoppers" segment. The p-value for the t-test was well below the significance threshold, indicating that offering a discount to this group is effective in boosting order values.

Segment Behavior:

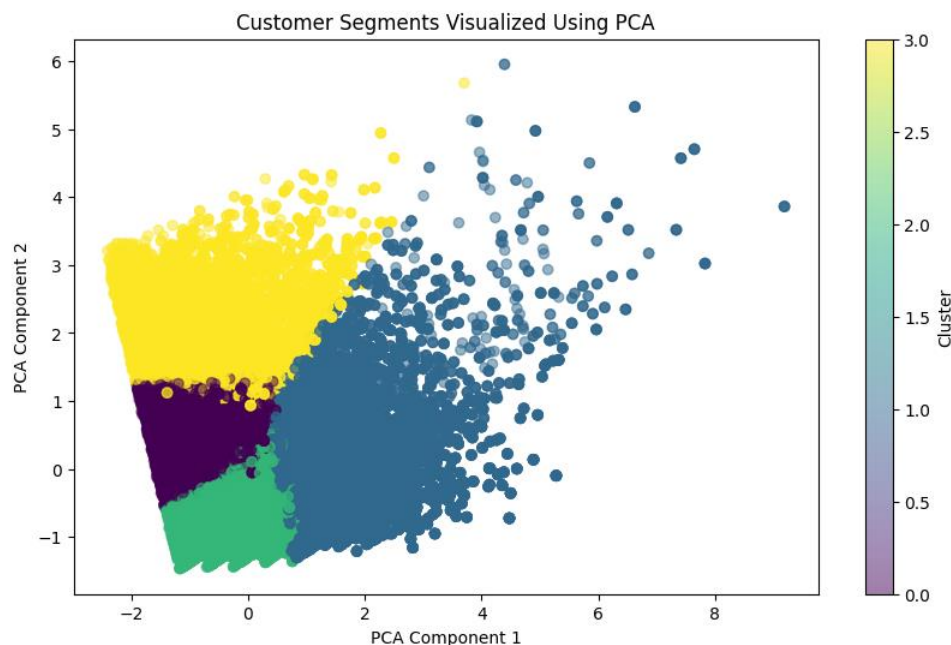
- **High Value, Frequent Shoppers** are highly responsive to discounts, suggesting a strong opportunity to maximize revenue from this group by offering targeted promotions.
- **Low Value, Dormant Shoppers** showed limited responsiveness to the 10% discount, indicating the need for alternative re-engagement strategies, such as loyalty programs or higher incentives.

Recommendations

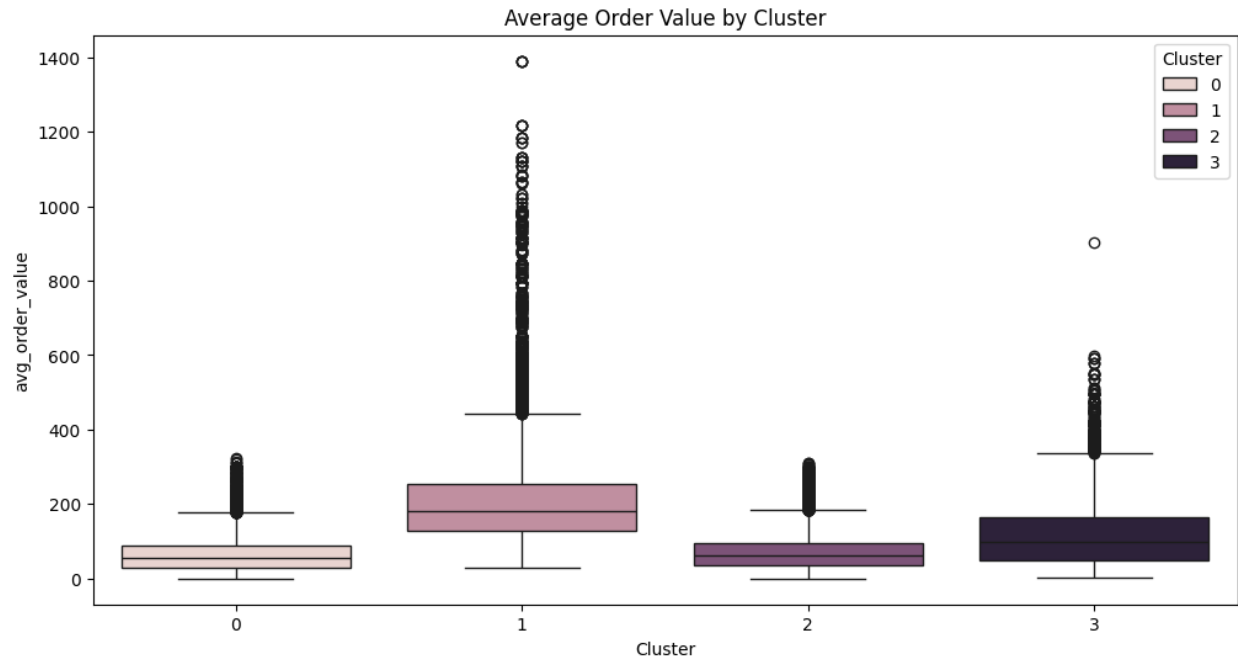
- **Focus Marketing Efforts on High-Value Segments:** Given their responsiveness, more resources should be allocated towards **High Value, Frequent Shoppers**, as they are likely to drive significant revenue through targeted promotions.
- **Revise Strategy for Low Value Segments:** For **Low Value, Dormant Shoppers**, consider different types of incentives or engagement campaigns, such as personalized recommendations or loyalty rewards, to stimulate interest.
- **Further Testing:** Additional tests with varied discount percentages or other incentives (e.g., free shipping) should be conducted to better understand the thresholds that effectively engage **Low Value** customers.

Visualizations and Reporting

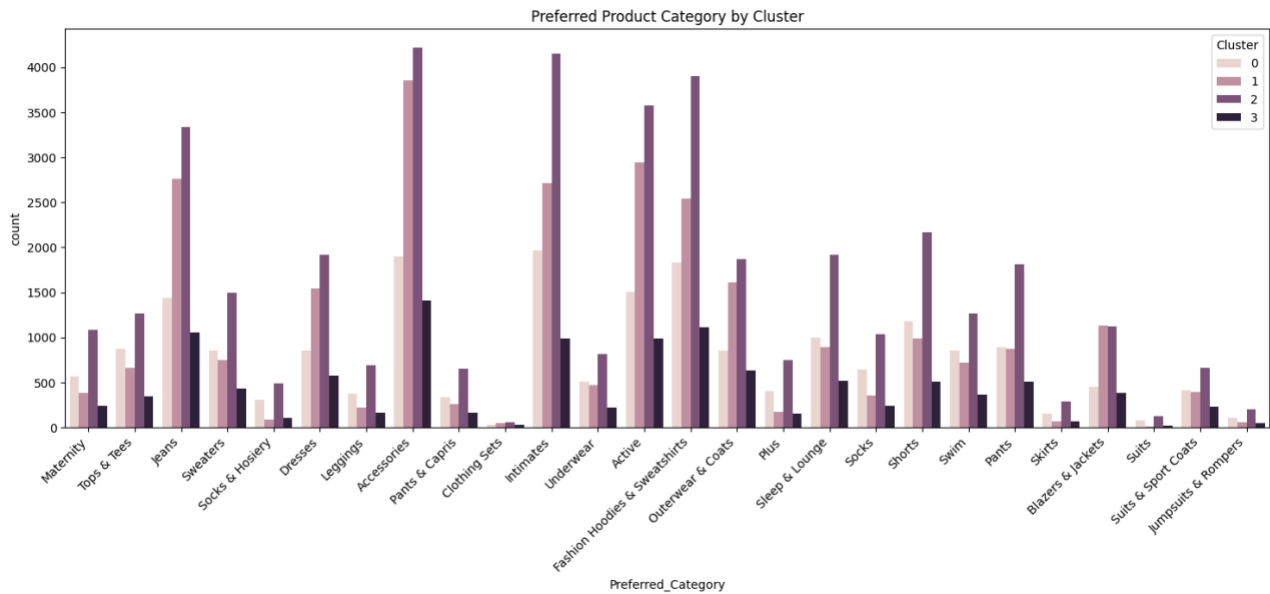
1. PCA Customer Segments



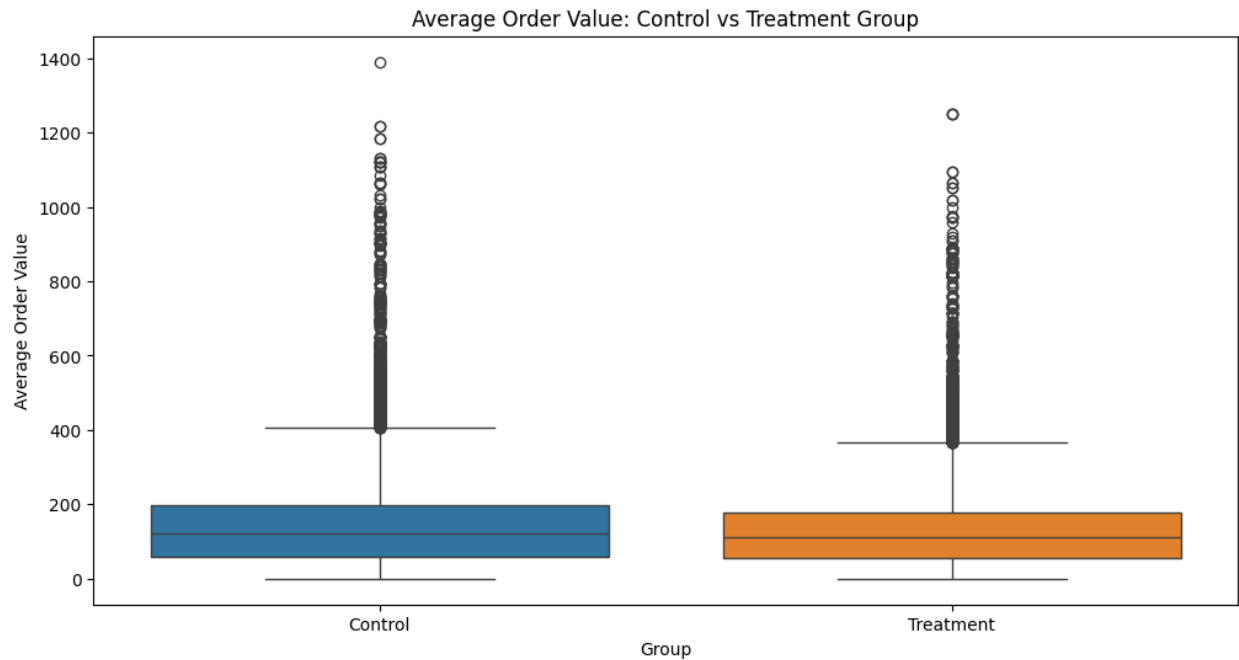
2. Average Value by Cluster



3. Preferred Product Category by Cluster



4. Average Order Value: Control vs Treatment



5. Dashboard Visualization



Conclusion

This project successfully segmented the customer base and demonstrated the value of targeted marketing. By focusing on key customer groups, the company can maximize the effectiveness of its promotional activities, increasing both sales and customer satisfaction. The analysis showed that certain groups are highly responsive to discounts, while others may require different approaches to drive engagement.

Future efforts should include a deeper dive into demographic features like age and gender, more complex segmentation models (e.g., hierarchical clustering), and broader A/B testing with different types of incentives.