1. Executive Summary

This report presents the results of a customer segmentation analysis conducted to refine marketing strategies, enhance customer engagement, and boost sales for a convenience store chain.

Transactional data is transformed into a customer-centric dataset by creating new features that will later on be used for the segmentation of customers into six distinct groups, each exhibiting unique traits and purchasing behaviours, unpacking valuable insights into customer preferences and motivations.

Big Spenders and Bargain Hunters segments were chosen as the most important segments to focus on due to their market potential and distinct purchasing patterns. Tailored strategies, such as emphasising premium products for Big Spenders and offering value-oriented options for Bargain Hunters, were suggested to optimise marketing efforts.

In addition, other strategies such as tailored marketing, recommendation system, loyalty program and segment behavior predictive models were also recommended.

2. Feature Description

15 features were chosen to create a comprehensive dataset, aiming to construct nuanced profiles for each customer segment. This is a relatively high number of features; however they collectively provide a well-rounded understanding of customer engagement and value, transaction behavior, shopping patterns, and product preferences. Furthermore, the high dimensionality inherent in the dataset can be effectively mitigated during model implementation through Principal Component Analysis (PCA).

The data dictionary below outlines the description of each feature along with the rationale behind its inclusion:

Feature	Description	Importance
total_transactions	Total number of transactions made by the customer	Signifies how often he made a purchase within the analysed period. A higher value suggests higher engagement
total_spend	Total amount of money the customer has spent across all transactions	Indicates customer lifetime value, offering insights into the overall financial contribution of the customer to the business
total_products	Total number of products purchased by a customer across all transactions	Helps understanding the breadth of the customer's buying habits and preferences
unique_products	Number of different products the customer has purchased.	Reveals the diversity in the product purchase behavior of customers
days_since_last_purchase	Number of days passed since the customer's last purchase	A shorter duration signifies more recent engagement, suggesting higher levels of customer activity and potential for future purchases

avg_item_spend	Average spending per item, calculated as total spend divided by total number of products purchased	Indicates purchase preferences, whether customers are buying many small items or single higher cost products
avg_basket_spend	Average spending per transaction, calculated as total spend divided by total transactions	Indicates the spending behavior buying power of customers
avg_qty	Average quantity of products purchased per transaction, calculated as total number of products purchased divided by total transactions	Reveals typical purchase volume of the customer and their consumption patterns
avg_days_between_purchases	Average number of days between consecutive purchases made by the customer	Provides insights into customers' shopping frequency and aids in predicting the next purchase
fav_dow	Preferred day of the week for shopping (0 for Monday, 6 for Sunday)	Capture the shopping behaviours of customers and help to infer insights on their habit or lifestyle
fav_hour	Preferred hour of the day for shopping, in a 24-hour format.	
food_and_grocery	Total spending on fruit_veg, dairy grocery_food, bakery, meat, world_foods, deli, discount_bakery	Provide insights on customers' preferences and purchase behaviours regarding different product categories, valuable for tailoring product recommendations
snacks_and_frozen	Total spending on prepared_meals, frozen, confectionary, drinks, soft_drinks	
household_and_convenience	Total spending on practical_items, seasonal_gifting newspapers_magazines, grocery_health_pets	
misc	Total spending on tobacco, lottery, cashpoint	

3. Customer Base Summary

Overall, the company's market consists of a diverse range of individuals, from loyal customers to occasional buyers. Favorite shopping time suggest potential alignment with employed individuals, presenting opportunities for targeted marketing.

On average, customers display high loyalty with approximately 65 purchases over 6 months, ranging from 2 to 374 transactions, showcasing varying activity levels. The average buying frequency is every 4 days, with a mode of 0 days since last purchase indicating significant customer activity.

Customers demonstrate a preference for exploring a wide range of products, purchasing an average of 584 items and 276 unique products. Food and grocery emerge as the top revenue-generating category.

In terms of spending behavior, the average customer allocates £772 across all transactions, translating to an average transaction value of £14.77.

Wednesday and Thursday are peak shopping days, with Saturday and Friday showing lower activity. Peak shopping hours are between 10am and 12pm. This suggests that a considerable portion of customers may be employed or engaged in weekday activities, as they find it convenient to run errands or shop for groceries during their lunch breaks or before and after work hours.

4. Segmentation Methodology

The data points are clustered using Kmeans, with the optimal number of clusters determined using a combination of the Elbow Method with WCSS score and the Silhouette Score Method. The Elbow Method plots inertia/WCSS score (sum of squared distances between each data point and its assigned cluster centroid) against the number of clusters (k), with the optimal k at the "elbow" point where inertia reduction becomes marginal. Meanwhile, the Silhouette Method computes the silhouette coefficient for each data point is calculated across various k values, selecting the k with the highest average silhouette score as the optimal number of clusters.

The KMeans algorithm is initialised across the range of k values from 2 to 11. While the "elbow" point is not too distinct in the WCSS plot, the drop in WCSS becomes more minimal after 6. A peak in the average silhouette score at k = 6 also suggests its potential. Despite not having the highest score, k = 6 offers more granular cluster separation than k = 2 or k = 3, aiding in personalised targeting strategies later.

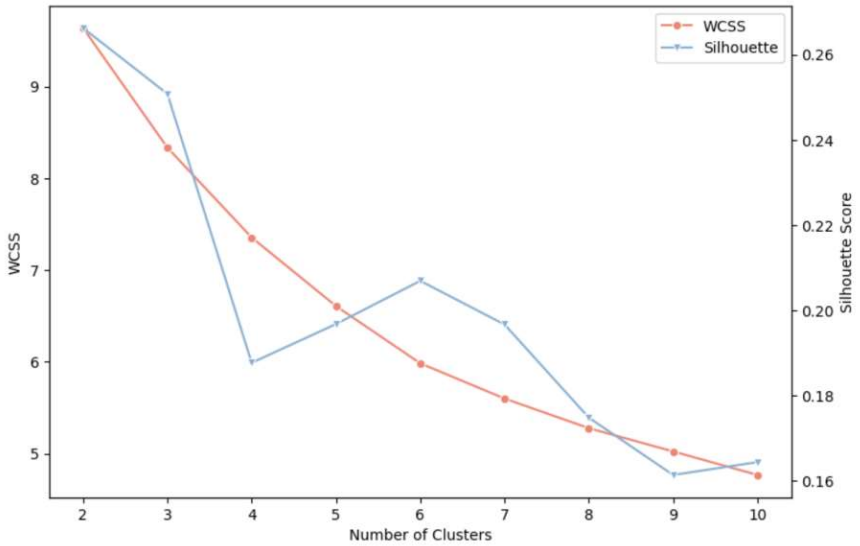
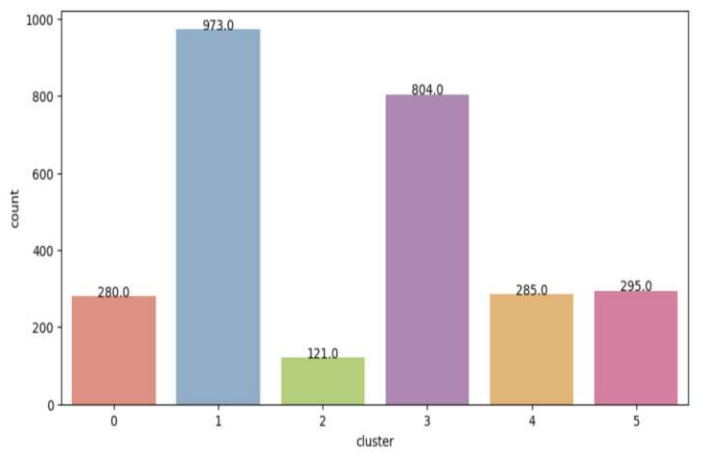


Figure 1: WCSS vs Silhouette Score

Upon applyting the Kmeans algorithim on 6 clusters, the results are shown to be well distributed and separated, as evidenced by the cluster distribution bar plot and the 3D visualisation of customer clusters in PCA space. This means the clustering process successfully identified meaningful patterns within the data

and partitioned them into homogeneous groups. No cluster contains very few customers, indicating that each cluster is significant and does not just include outliers or noise in the data.



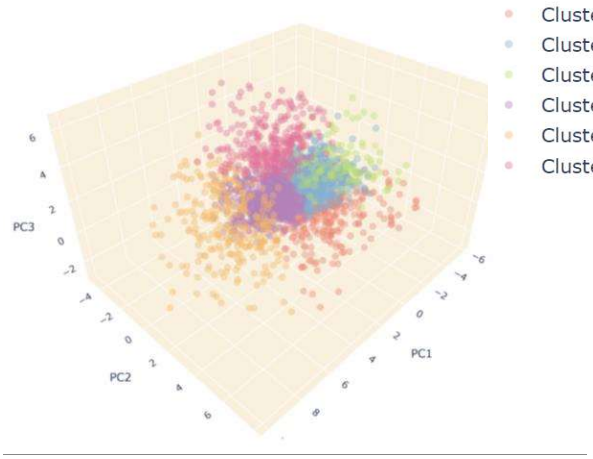


Figure 2: Distribution of Customers across Clusters **Figure 3:** 3D Visualisation of Clusters in PCA Space

5. Results

The clusters are named and profiled for better interpretation.

Cluster 0: The Big Spenders (280 customers)

Customers in this cluster exhibit high-value transactions, characterized by very high average basket spend of £31.4 and average quantity of 25 items. Despite fewer transactions, their substantial spending suggests a high disposable income, possibly demanding careers, and limited time for frequent shopping. They prioritise practicality and household need over discretionary purchases.

Cluster 1: The Practical Shoppers (973 customers)

These customers shop in a balanced manner, with moderately higher intervals between purchases of 5 days on average. Despite shopping across all categories, their total spends, total products purchased, and total number of transactions made are on the lower side. They tend to spend wisely on items, evidenced by their relatively low average spending per item and per transaction. This cluster likely comprises individuals with stable jobs or income, who prioritise simplicity and efficiency over luxury or indulgence. With the highest number of customers, this cluster represents the prototypical shopper.

Cluster 2: The Inactive Customers (121 customers)

This cluster represents customers who have been inactive recently, with a high average of 52 days since their last purchase. Compared to other clusters, they have the lowest number of transactions, spending, and number of products purchased. While their average spending and quantity per transaction are moderate, the long period of inactivity suggests a decline in interest. These customers might be occasional or seasonal shoppers who only make sporadic purchase during specific occasions.

Cluster 3: The Bargain Hunters (804 customers)

These individuals are savvy shoppers who prioritise value in their purchases. Their low average days between purchases and high transaction volume suggest frequent shopping activity across a wide variety of products. Despite their frequent shopping behavior, their low average spends per item and per purchase indicate they may be price-sensitive consumers who prioritise affordability and practicality in their purchases. Customers in this cluster are likely to be young families or budget-conscious individuals who

carefully manage their expenses. This cluster makes the highest revenue contribution to the company, totaling of £597,859.

Cluster 4: The Trend Explorers (285 customers)

Customers in this cluster are very frequent and recent shoppers, with an average of only 1.86 days since their last purchase and 1.34 days between purchases. They demonstrate high engagement with very large total transactions, total spending, and total products. However, they tend to buy small, low-value items, reflected in their low average item spend and average basket spend. Their propensity for frequent shopping and willingness to try different products suggest they may be young professionals who are conscious of trends.

Cluster 5: The Indulgent Seekers (295 customers)

These shoppers enjoy life's pleasures and are not afraid to indulge in their desires. Despite their lower average spend per transaction, they usually go for higher-value items, with an average item spending of £2.27. Their notable spending on tobacco, lottery, and cashpoint suggests they may be more indulgent or impulsive consumers who allocate a portion of their budget to personal enjoyment.

6. Summary

The clustering analysis revealed six distinct customer segments: The Big Spenders, The Practical Shoppers, The Inactive Customers, The Bargain Hunters, The Trend Explorers, and The Indulgent Seekers. This clustering solution provides valuable insights into customer behavior and preferences, which help the company tailor marketing strategies, product offerings, and customer experiences more effectively.

Based on the segmentation results, the company should focus its attention on the Big Spenders and the Bargain Hunters segments.

The Big Spenders comprise customers with high-value transactions and a focus on practical household needs, which are likely to recur. Therefore, targeting this segment can lead to significant revenue generation. Their substantial spending also indicates a higher disposable income compared to other segments. The recommended strategies for this segment are:

- Positioning strategy: Emphasize high-quality, premium products and exclusive offerings that cater to the lifestyle of these customers. Introducing premium services such as personalised shopping assistance or access to exclusive events can further differentiate the brand and attract this affluent customer segment.
- Personalised offers and loyalty programs: As this segment demonstrate lower shopping frequency, offering exclusive rewards and tailoring promotions and discounts to frequently purchased items can incentivise repeat purchases and increase customer loyalty.
- Subscription services: Introduce subscription-based services or automatic delivery at regular intervals for essential items frequently purchased by this cluster can cater to the busy lifestyles of these customers, enhance their overall shopping experience, and encourage ongoing engagement.

The Bargain Hunters comprise customers who are careful with their expenses and often look for budget-friendly options. While individual transaction value may be lower, the high frequency of purchases and focus on essential items indicate a very high level of loyalty. Accounting for 29% of the customer base and the

largest revenue contribution, the Bargain Hunters present a substantial market share for strategic engagement. Some suggested strategies for this group of customers are:

- Positioning strategy: This segment is likely to be receptive to targeted marketing efforts highlighting value and savings. Products should be positioned as budget-friendly, value-oriented solutions that offer practicality, affordability, and savings.
- Exclusive promotions: Introduce time-sensitive promotions like limited-time offers or flash sales to create a sense of urgency and encourage immediate action. Offer exclusive discounts to reward customers' loyalty and incentivise purchases.
- Pricing policies: Implementing price matching policies assures customers of competitive pricing and reinforces trust in the brand, therefore encouraging repeat purchases. In addition, offering discounts or special pricing for bulk purchases can incentivise larger transactions and appeal to the shopping behavior of this segment.

In addition to cluster-specific strategies, the company can adopt other initiatives to make use of the segmentation results and analysis insights:

- Personalise relationship with customers through tailored marketing campaigns, product offerings, vouchers based on their cluster profiles. In addition, a recommendation system could be constructed to suggest popular products within a customer's respective cluster that they have not yet purchased, facilitating cross-selling and boosting sales.
- Develop a loyalty program. Considering the possible absence of a loyalty program, evident from the lack of demographic customer data, implementing one presents an opportunity to incentivise repeat purchases and provide valuable data for further meaningful analysis. For example, exploring demographic or geographic patterns within each cluster can reveal personal and regional preferences and trends for targeted marketing strategies and business expansion initiatives.
- Developing predictive models to forecast segment behavior, such as purchase likelihood or churn probability. The company can leverage these predictive insights to adjust its marketing strategies and retention efforts to stay ahead of the curve.