Market Segmentation Analysis of Transactions in National Convenience Stores

Analytics Specialization and Application
Customer Segmentation

SAMIKSHA KAMATH
20703562
BUSI4370
2024-25

Executive summary

This report presents a comprehensive market segmentation analysis of 3,000 customers transactional behaviour over six months for a national convenience store chain, aiming to create data-driven customer profiles to enhance marketing strategies. The approach began with data preparation, where missing values were handled, duplicate entries removed for consistency. Feature engineering was performed to extract key behavioural metrics, including Recency-Frequency-Monetary (RFM) variables, category-level spending, basket size trends, and temporal purchasing patterns. Standardization technique were applied to ensure comparability across features. A correlation analysis and Principal Component Analysis (PCA) were conducted to reduce dimensionality and identify the most informative features. The clustering process involved K-Means clustering, where the Elbow Method and Silhouette Score were used to determine the optimal number of clusters.

The segmentation identified five distinct customer groups, ranging from infrequent, low-spending shoppers to high-frequency, high-value customers with category-specific preferences. Statistical analysis of each cluster provided quantitative insights into spending behaviours and product affinities, while pen profiles were developed to create relatable customer archetypes. The study identified two high-value segments, for which tailored loyalty programs, frequency-based incentives, and category-specific promotions are recommended to enhance customer retention and increase spending. The results provide a strategic foundation for targeted marketing, pricing optimization, and personalized engagement initiatives, with opportunities for future refinement through predictive modelling, seasonality analysis, and customer lifetime value forecasting to further optimize the retailer's strategy.

Data Preparation – Data Cleaning Steps

All datasets were checked for missing values, and no null values were found across any of the provided tables. However, several data cleaning transformations were necessary to ensure consistency and accuracy before clustering:

Customer Table
This table summarizes spending habits, including total spend, average spend per visit, and number of baskets (visits). Both total spend and average spend were stored as object types due to the presence of the "£" symbol. These were converted to floating-point values to enable numerical analysis. No duplicate records were found, and the dataset showed high variance in visit frequency and spending, making standardization necessary before clustering.

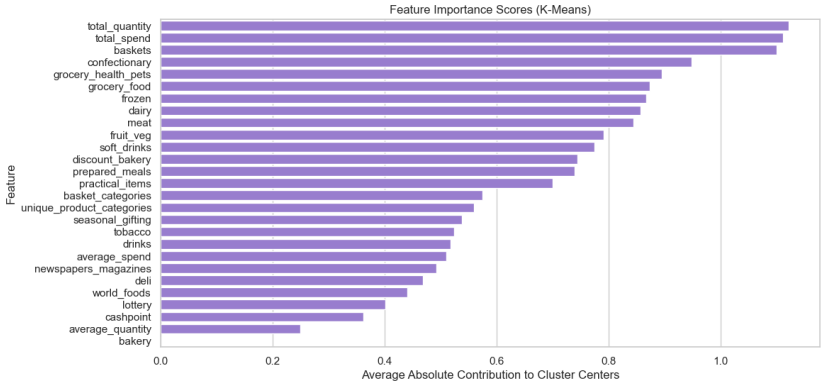
Category Spends Table
This table records spending across 20 product categories. All spend-related columns contained the "£" symbol and were initially stored as objects. To ensure accurate numerical analysis, all category spend columns were cleaned by removing the "£" symbol and converted to float values.

Baskets Table
This table provided insights into purchase timestamps, basket size, and spend per visit. The basket spend column was stored as an object due to the "£" symbol and required conversion to float. Additionally, 19 records had negative basket quantity values, likely due to refunds or data entry errors.The purchase time column was also converted to datetime format to facilitate time-based trend analysis.

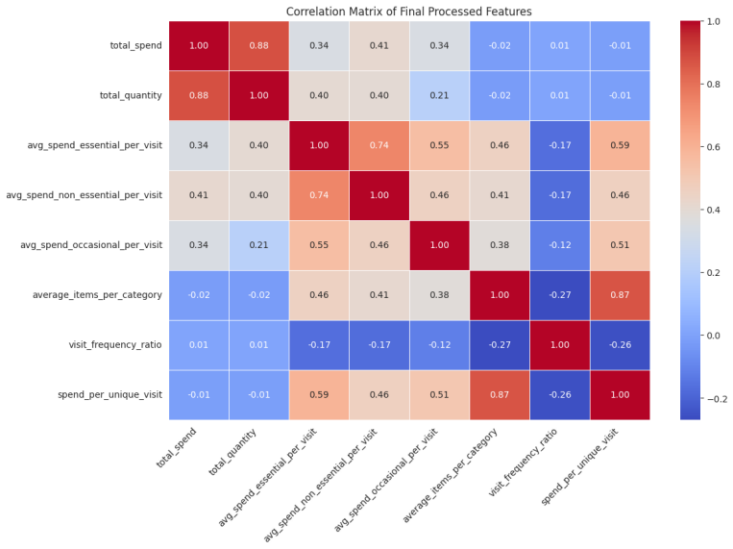
Line Items Table
This table contained granular purchase records at the product level. The spend column was stored as an object and required conversion to a numeric format for accurate calculations. Additionally, 864 records had negative quantity values, possibly due to refunds .Furthermore, 114 duplicate rows were detected and removed to maintain data integrity before analysis.

Feature Selection & Engineering

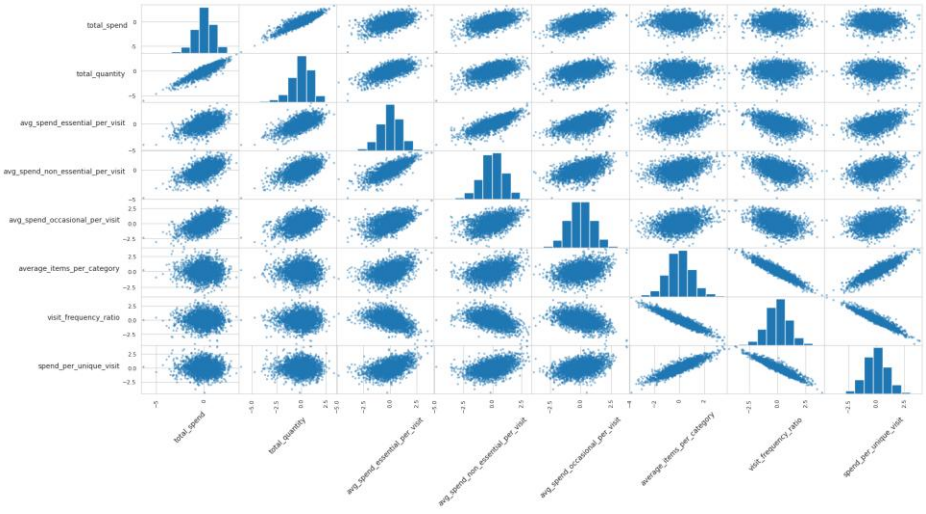
A **feature importance** was calculated by measuring each feature's contribution to the cluster centers in K-Means clustering, using absolute differences. This helped us identify key customer behavior drivers, ensuring our engineered features (e.g., total spend, basket quantity, and category preferences) were the most relevant for segmentation.



The feature-engineered variables provide key insights into customer spending behavior, shopping frequency, and purchase diversity. **Total spend and total quantity purchase** help distinguish high-value customers from low-value ones, highlighting revenue concentration and spending patterns. **Spend per unique visit** differentiates premium shoppers, who make large transactions, from budget-conscious customers, who shop frequently but spend less per trip. **Average spend across essential, non-essential, and occasional categories** reveals whether customers prioritize daily necessities, indulge in discretionary spending, or engage in impulse-driven purchases. **Visit frequency ratio and average items per category** further segment customers based on their shopping consistency and category engagement, helping identify those with broad vs. focused purchasing habits. These engineered features enable more effective segmentation and targeted marketing strategies.

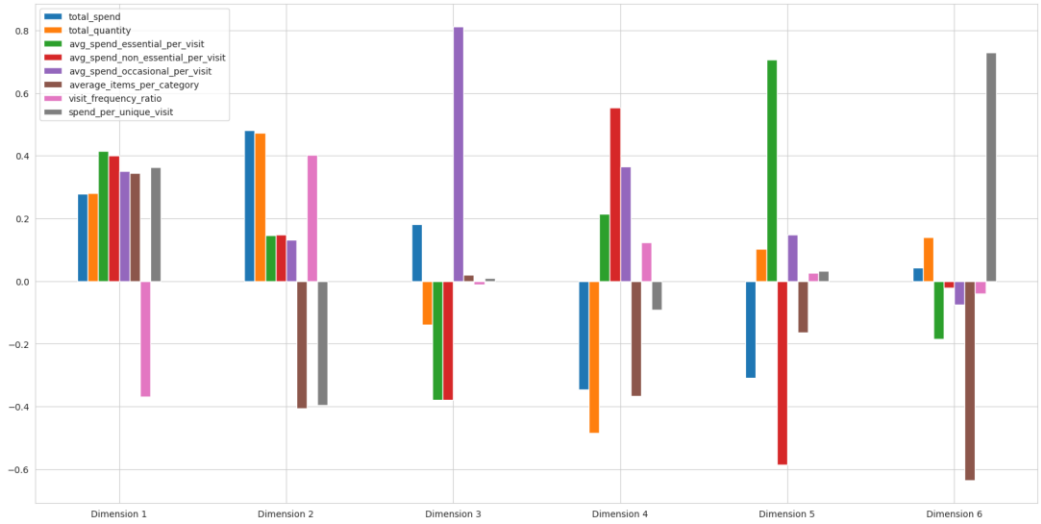


The **correlation matrix** revealed strong relationships between '**total_spend' and 'total_quantity'**, suggesting that higher-spending customers also tend to purchase more items, reinforcing the link between spending power and purchase volume. Similarly, the correlation between **'spend_per_unique_visit' and 'visit_frequency_ratio'** indicated two distinct shopping behaviors—frequent, low-spend shoppers vs. infrequent, high-spend buyers, suggesting different spending strategies among customers. The **log transformation** was necessary due to the right-skewed distribution of spend-related features, where a small number of high-spending customers disproportionately influenced the data, highlighting the need for normalization. After transformation, scaling using **StandardScaler** ensured that all features contributed equally to PCA, preventing variables like total spend from overpowering others, allowing for a more balanced and meaningful dimensionality reduction.



Principal Component Analysis (PCA) was applied to reduce dimensionality while retaining key behavioral patterns. The first two components were chosen as they capture **82.87% of the total variance** (PC1: **52.03%**, PC2: **30.65%**), ensuring that the majority of the dataset's information is

preserved. The explained variance curve shows diminishing returns beyond the second component, making additional components less valuable for interpretation. The **feature loadings indicate that total spend, total quantity, and spend per unique visit** are highly influential, summarizing overall shopping activity and purchase frequency. By selecting two principal components, the analysis achieves an optimal balance between **dimensionality reduction** and information retention, allowing for efficient customer segmentation without significant data loss.



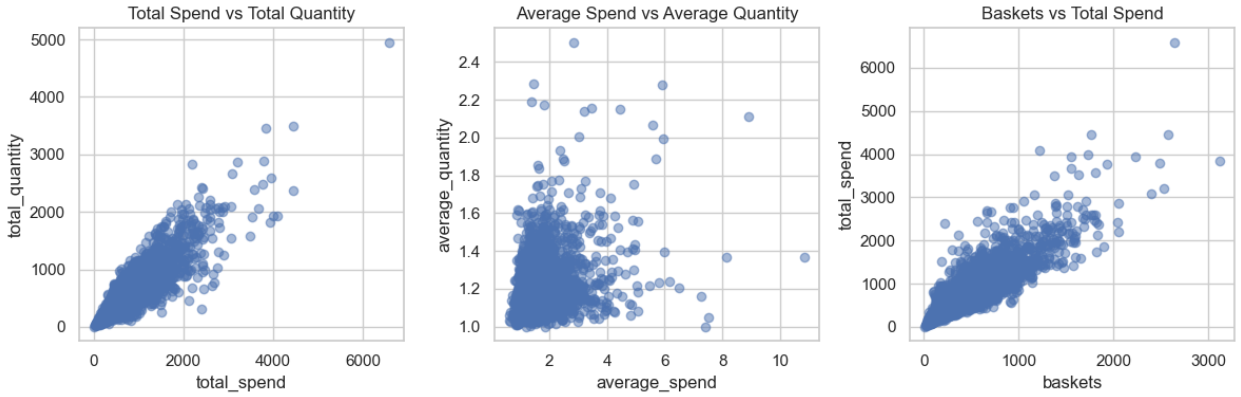
The **RFM analysis** from the baskets table evaluated **Recency (last visit), Frequency (visits), and Monetary Value (spending)** to segment customers.It identified distinct behavioral patterns, helping define **high-value, frequent, occasional, and churning customer segments**. These segments were later refined using **K-Means clustering**, ensuring more precise and actionable group classifications for targeted marketing strategies.

Customer Base Summary

The customer base was analysed to identify key shopping behaviours and patterns before applying segmentation. This section explores spending habits, visit frequency, and purchasing diversity to establish a foundation for clustering.

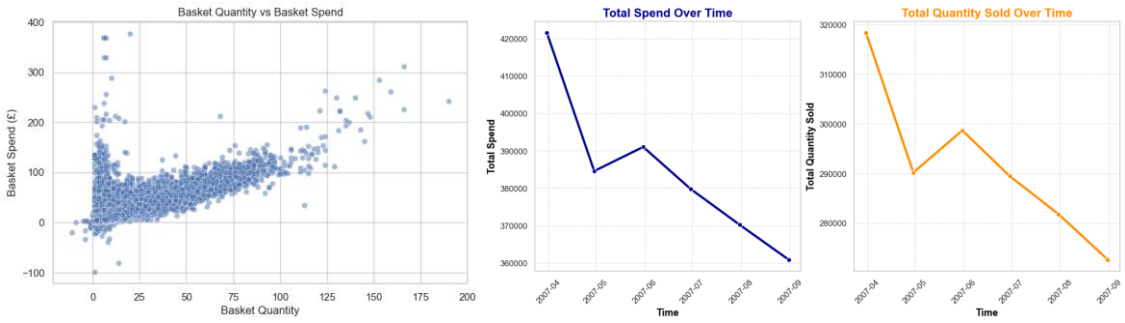
Spending Behaviour

The customer base consists of diverse shopping behaviors, ranging from high-spending, premium buyers to budget-conscious, bulk purchasers. A positive correlation between total spend and quantity suggests that most customers who buy more also spend more, with frequent shoppers contributing significantly to overall revenue. However, distinct outlier groups emerge—premium buyers who spend heavily on fewer items and bulk budget shoppers who prioritize high quantities at lower costs. The right-skewed spending distribution highlights that while the majority of customers have relatively low spending, a small percentage of high-value shoppers drive a disproportionate share of revenue.This segmentation highlights the need to retain high spenders while boosting engagement among lower-spending shoppers.



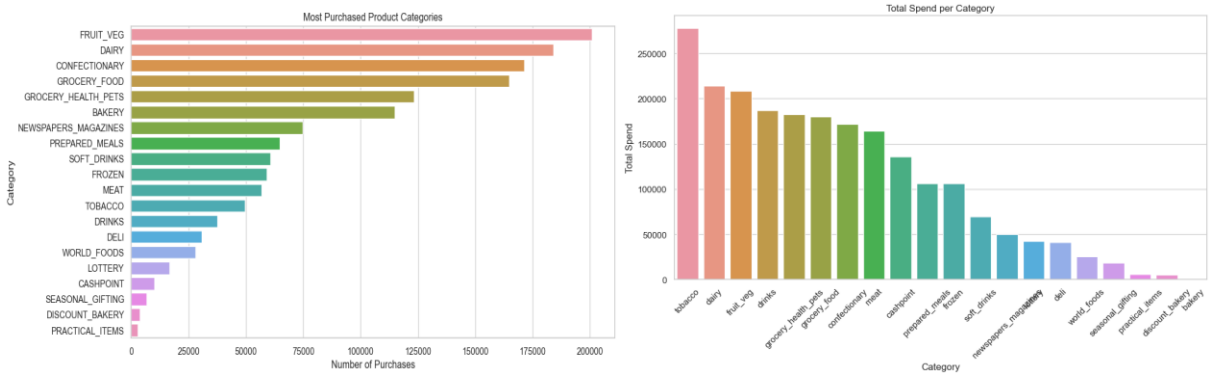
Shopping Frequency and Basket Size

The basket spend analysis reveals varying shopping behaviors, with a positive correlation between basket quantity and spend, meaning customers generally spend more as they purchase more items. However, significant variance exists, with some customers making high-value purchases despite buying fewer items, likely indicating premium product buyers. The dense clustering at lower basket quantities suggests that many customers engage in small, low-spend transactions, often for essentials or impulse buys. The right-skewed distribution highlights that while most baskets have low spend values, a few high-value transactions significantly impact total revenue. Additionally, the presence of negative spend values suggests refunds or corrections, reflecting adjustments in transaction records.



Product Category Preferences

The category spend analysis highlights a clear preference for essential goods, with tobacco, dairy, and fruit & vegetables emerging as the top revenue drivers. These categories reflect consistent demand, reinforcing their role as core contributors to overall sales. In contrast, discount bakery and practical items see minimal spending, suggesting they appeal to a niche segment of budget-conscious shoppers. The most frequently purchased items are fresh essentials like fruit, vegetables, and dairy, emphasizing habitual shopping patterns driven by necessity. Meanwhile, non-essential and occasional purchases, such as seasonal gifting, lottery, and practical items, show significantly lower demand, indicating that most customers prioritize daily necessities over discretionary spending.

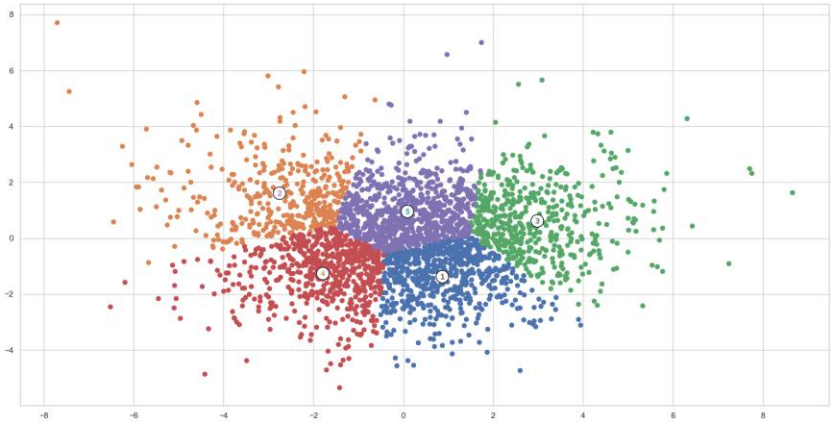


Segmentation Methodology

For segmentation, **K-Means clustering** was employed, as it is well-suited for customer segmentation based on purchasing behavior.The **silhouette scores** indicate how well-separated the clusters are, with higher values suggesting better-defined groups.The Score **k=5 was chosen** because it balances interpretability and segmentation quality, providing distinct customer groups while maintaining a reasonable silhouette score (0.317), making it ideal for actionable business insights.

We performed K-Means **clustering with k=5**, as it provided a balanced segmentation while maintaining a reasonable silhouette score (0.317). This choice ensures meaningful customer differentiation for effective marketing insights.The final model effectively grouped customers based on their spending behavior, shopping frequency, and category preferences, leading to distinct, data-driven customer segments.

After segmenting customers using K-Means, we applied a **reverse transformation** to interpret cluster centers in their original scale. This involved undoing PCA, standardization, and log transformation, restoring the true feature values for meaningful customer profiling.

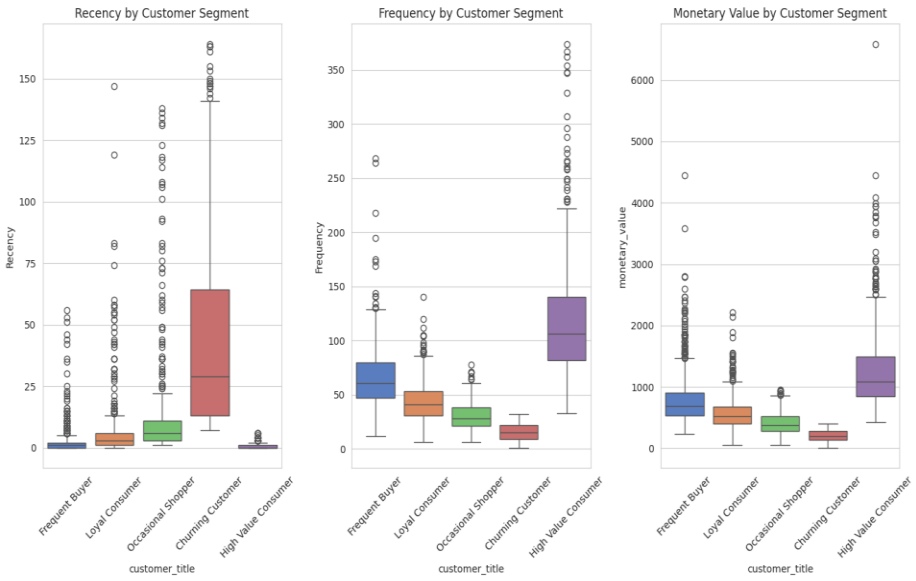


Result

The segmentation approach combined **clustering and RFM analysis**, identifying five distinct consumer groups. This method offers a **detailed understanding of customer behavior**, enabling targeted marketing strategies that align with varying preferences and shopping patterns.

RFM analysis results:

Customer Segment	**Recency (days)**	**Frequency (visits)**	**Monetary Value (£)**	**Percentage of Customer Base**	**Insights**
High Value Customer	0.5	116	1262.8	29.6%	Very high frequency & spending
Frequent Buyer	2.6	64.8	800.8	24.8%	Shops regularly, good spending
Loyal Customer	5.9	43.6	567.3	22.2%	Engaged and loyal, shops often
Occasional Spender	13.3	29.7	403.6	14.9%	Shops sometimes, moderate spending
Churning Customer	46.8	16	209.8	8.4%	Low activity, may stop buying soon



RFM Segment Analysis with Segments:

1. High Value Consumer (Segment 1) – Most Profitable Group

High Value Consumers are the highest spending and most engaged customers, contributing significantly to overall revenue. Their total spend ranges between £1,267.05 and £1,730.24, with an average purchase quantity of 857 to 1,273 items. Their visit frequency ratio is between 0.10 and 0.25, meaning they shop more often than most other segments. While their spend per visit (£7.38 - £15.61) is not the highest, their consistency and diverse category engagement make them the most valuable group.They shop frequently across essential categories like fruit & veg, dairy, and groceries while also purchasing non-essentials such as meat, frozen foods, and prepared meals. Additionally, they contribute to occasional purchases such as seasonal gifting and tobacco, suggesting a mix of both planned and impulse spending.

From an RFM perspective, these customers have the highest frequency and monetary value while maintaining low recency scores, meaning they return often and spend consistently. Their broad spending patterns confirm their importance in driving sustained revenue for the business.

2. Churning Customer (Segment 3) – Previously Loyal but Declining

Churning Customers were once highly engaged and high-spending but have drastically reduced their shopping frequency. Their total spend ranges from £376.32 to £617.32, with an average purchase quantity of 201 to 609 items. However, their visit frequency ratio is extremely low (0.016 - 0.037), indicating that they rarely shop now. Despite this, their spend per unique visit (£33.92 - £148.50) is among the highest across all segments, meaning when they do shop, they make large purchases.

Previously, these customers were engaged in both essential and non-essential categories, but their engagement with occasional purchases has significantly declined, showing that they are no longer making impulse-driven purchases. Their reduced frequency suggests a shift in shopping behavior, potentially due to switching to competitors, price sensitivity, or changes in lifestyle.

According to the RFM model, these customers were previously strong in monetary value and frequency but now have a high recency score, meaning they haven’t shopped in a long time.

While they still contribute significantly per visit, their decreased engagement makes them a high-risk segment for revenue loss.

3. Occasional Shopper (Segment 2) – Selective, Infrequent Buyers

Occasional Shoppers engage with the store sporadically, making purchases only when necessary. Their total spend is between £84.46 and £447.11, with a low total purchase quantity of 69 to 303 items per customer. Their visit frequency ratio varies significantly from 0.10 to 1.22, meaning some customers rarely shop, while others show irregular bursts of activity. Their spend per unique visit is between £0.83 and £17.22, indicating inconsistent spending patterns.

These shoppers prioritize essential categories such as groceries and dairy but also show above-average engagement in occasional purchases such as tobacco, lottery, and seasonal gifting. This suggests they may shop based on convenience or impulse rather than as part of a routine. However, their engagement with non-essential categories remains relatively low, meaning they are less likely to explore additional products.

RFM categorizes them as low-frequency shoppers with moderate monetary value and high recency scores, meaning they go long periods without shopping. Their unpredictable shopping behavior makes them a difficult segment to rely on for consistent revenue.

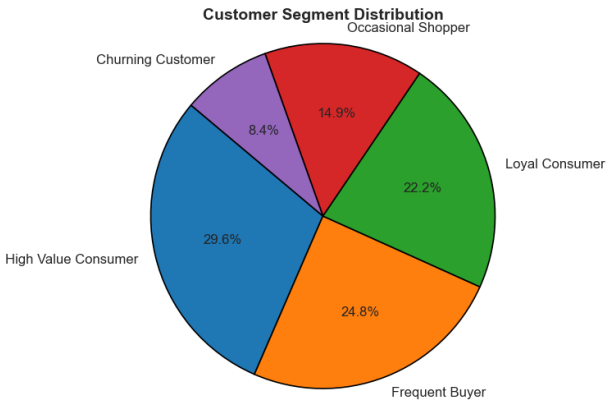
4. Frequent Buyer (Segment 4) – Regular but Low-Spend Shoppers

Frequent Buyers shop regularly but spend less per visit compared to other segments. Their total spend is between £561.56 and £1,354.40, with a total purchase quantity of 443 to 990 items. Their visit frequency ratio ranges from 0.21 to 0.72, meaning they shop often, but their spend per unique visit (£1.60 - £6.75) is the lowest of all segments.These customers primarily focus on essential categories such as fruit & veg, dairy, and groceries, showing lower engagement with non-essentials and occasional purchases. Their small and frequent transactions suggest they may be budget-conscious, buying only what they need rather than making bulk purchases.

According to the RFM model, they rank high in frequency but have moderate monetary value and low recency scores, meaning they visit consistently but do not spend as much per trip. Their consistent shopping behavior ensures a stable, albeit lower-margin, revenue stream.

5. Loyal Consumer (Segment 5) – Reliable and Balanced Shoppers

Loyal Consumers maintain a steady shopping pattern, making them a dependable revenue source. Their total spend is between £441.80 and £1,074.51, with an average purchase quantity of 214 to 346 items. Their visit frequency ratio is low (0.06 - 0.20), meaning they do not shop frequently, but their spend per visit (£9.17 - £26.64) is relatively high, indicating planned and efficient shopping trips.These customers engage in both essential and non-essential categories, demonstrating a balanced shopping approach. They spend moderately on bakery, frozen foods, soft drinks, and practical items, suggesting that they are not impulse-driven but rather plan their purchases carefully. Unlike occasional shoppers, they are more predictable in their spending habits.The RFM model classifies them as moderate in frequency, monetary value, and recency, meaning they contribute to a steady revenue stream without major fluctuations. Their stable engagement ensures ongoing but moderate profitability.



Summary

Upon analysis we have chosen **High Value Consumers** and **Churning Customers** as the two key segments to focus on, as they offer the greatest impact on revenue growth and retention. High Value Consumers drive sustained profitability, while Churning Customers represent potential revenue loss that can be recovered with the right strategies. Additionally, the spending distribution across categories plays a crucial role in understanding their behaviors and targeting them effectively.

High Value Consumers are the most profitable shoppers, consistently spending across multiple categories with high total spend, high quantity purchased, and strong engagement in both essentials and discretionary items. They purchase frequently and contribute significantly to high-revenue categories such as tobacco, dairy, and fruit & vegetables, ensuring consistent business growth. Their strong presence in both non-essential (bakery, frozen foods, and soft drinks) and occasional (seasonal gifting, lottery) categories indicates a balanced mix of planned and impulse shopping. Retaining this segment is critical, as they not only sustain overall revenue but also drive high-margin purchases in discretionary categories.

Churning Customers, on the other hand, were once high-value but have significantly reduced their shopping frequency. While their spend per visit remains high, their decreasing engagement suggests a risk of losing them to competitors or changing shopping preferences. Their past purchasing patterns indicate a preference for essentials like dairy and groceries, but they have also engaged in non-essentials such as soft drinks and prepared meals. However, their lower interaction with occasional categories suggests that impulse-driven purchases have declined. By focusing on re-engaging them with incentives tied to their preferred product categories, there is an opportunity to recover lost revenue and encourage them to return to consistent shopping habit**s**.

Marketing Strategy Suggestions:

High Value Customers (Segment 1)	Churning Customers (Segment 3)
Threshold-Based Free Delivery: Encourage higher spending by offering free delivery on orders that exceed a specified purchase threshold, motivating customers to add more to their carts.	**Urgency-Driven Comeback Offers:**Send time-sensitive discounts or "One-Time Comeback Deals" to reignite interest and encourage immediate purchases.
Premium Memberships: Introduce a VIP program with exclusive benefits like free delivery, extra discounts, or cashback.	**Omnichannel Retargeting Campaign:**Use a mix of personalized emails, SMS, and social media ads to re-establish brand presence and remind them of ongoing promotions.
Cross-Selling & Upselling: Since they already buy a wide variety of products, suggest premium versions or complementary items to increase basket size.	**Subscribe & Save program:**This offers automatic reorders with discounts or bonus loyalty points on essential items. This strategy enhances convenience, encourages long-term retention while ensuring consistent revenue.
Early Access to Seasonal & Trend-Based Products:Provide first access to new product launches, seasonal items, and limited-time gourmet selections in categories they regularly purchase.	**Targeted Category Discounts:**Offer special deals on categories they previously engaged with (e.g., "Your favorite frozen meals are 20% off this week!").

Suggestions for Further Analysis:

For further analysis, a deeper investigation into customer lifetime value (CLV), price sensitivity, and churn prediction can help refine segmentation and marketing strategies. Understanding category affinities and seasonal purchase trends can improve product bundling and promotional timing. Additionally, analyzing customer journey touchpoints and sentiment from reviews can provide insights into improving engagement and satisfaction. A data-driven approach to evaluating marketing campaign effectiveness and competitive positioning can further optimize business decisions.