ANALYSING CUSTOMER BEHAVIOUR IN A MULTICHANNEL GIFT RETAILER

A report on multinomial predictive modelling, geographic trade area analysis, and customer segmentation to understand loyalty, gifting behavior, and off-season purchasing trends (2004–2007)

AUGUST 31, 2025

PRATIK GANGULI

UIVERSITY OF AUCKLAND

Table of Contents

EXECUTIVE SUMMARY

This report directly addresses these challenges by focusing on four specific analytical objectives. To improve customer relationship management, the first question examines the extent to which a customer’s initial purchase channel can predict their subsequent channel usage patterns. This analysis is grounded in the 2001–2007 period, when the Internet channel evolved into a significant driver of revenue, and considers the degree to which retail channel engagement is influenced by a customer’s proximity to a physical store. Complementing this, the second question under this challenge explores store trade area effects and their impact on all channels, seeking to determine how distance from the nearest store shapes retail spending levels, channel preferences across Retail, Internet, and Catalog, and the demographic and psychographic composition of the customer base. Together, these two lines of inquiry provide a robust framework for understanding channel loyalty, switching behavior, and the geographic factors that inform long-term customer relationship strategies

To strengthen competitiveness in the highly saturated Christmas gifting market, this analysis investigates the **consistency and predictability of customer gifting behavior**. Since the Christmas season represents the company’s peak sales period, the study leverages detailed transactional data that separates gift versus non-gift revenue across Internet and Catalog channels, along with seasonal counts of distinct and new gift recipients. Customers were segmented according to their overall **gifting intensity and engagement patterns**, which revealed clear distinctions between high-value gifters, occasional gifters, and low-engagement segments. These insights enable the development of **targeted retention strategies to sustain loyalty among top-tier gifters**, while also highlighting opportunities to **nurture and convert lower-value customers into more consistent contributors** during the peak season.

To address the challenge of generating consistent business during the non-Christmas months, the fourth question posed examines purchasing behavior in the Spring season compared to the Fall/Christmas period, with a focus on identifying distinct channel preferences and demographic characteristics among Spring buyers. This analysis evaluates whether customers who engage in Spring purchases differ in meaningful ways, such as proximity to stores, age distribution, or digital adoption tendencies, from those whose purchasing is concentrated solely in the holiday season. By understanding these differences, the study highlights targetable customer segments and proposes strategies to encourage off-season purchases, thereby reducing the company’s reliance on the peak holiday period and creating a more balanced, year-round revenue stream.

Statement of problem

The focal organization is a nationally recognized multichannel gift company specializing in premium food products, with sales heavily concentrated during the Fall and Christmas season. Operating across three primary sales channels, Retail, Internet, and Catalog, the company enjoys strong brand loyalty but faces persistent challenges in managing **customer relationships** across multiple channels, **competing effectively** in a saturated Christmas gifting market, and generating consistent business during the non-Christmas months to **reduce seasonality risks.** While the company has collected extensive transactional, demographic, psychographic, and geographic data on over 100,000 customers from 2001 to 2007, it has not fully leveraged this information for advanced analytics-driven decision-making. Basic segmentation and channel tracking have been conducted in the past, but there has been limited application of predictive modelling or in-depth behavioural analysis to guide targeted strategies across channels and seasons. This report addresses these gaps by examining whether customer data can be used to predict purchasing behavior, assess channel loyalty or switching patterns, and analyze how store proximity influences spending, channel preferences, and customer profiles. It further investigates the consistency and predictability of gift-giving behavior to strengthen retention in the highly competitive holiday market and explores seasonal purchasing patterns to identify opportunities for boosting off-season sales. By answering these questions, the analysis aims to transform raw customer data into actionable insights, enabling the company to strengthen loyalty, defend its competitive position, and create sustainable revenue streams year-round.

Research Questions

This project directly addresses these challenges by focusing on four specific analytical objectives. To improve customer relationship management, it examines how customer data across channels can be leveraged to predict purchasing behavior, build effective segments, and guide marketing strategies, with success measured through higher customer retention, increased campaign response, and sustainable revenue growth. It also investigates how a customer's first purchase channel is predictive of channel usage for subsequent purchases, particularly during the period when the Internet channel matured into a major sales driver, and how retail channel usage depends on proximity to a store.

To strengthen competitiveness in the Christmas gift market, the analysis explores the consistency and predictability of gift-giving behavior, using separate measures for gift and non-gift revenue in Internet and catalog channels, as well as counts of distinct and new gift recipients by season, to segment customers based on their propensity for gift purchases. To reduce reliance on holiday season revenue, the project assesses how much customers buy in Spring compared to Fall and whether Spring buyers display distinct channel preferences or demographic characteristics that can be targeted to boost off-season sales. Together, these analyses aim to generate actionable insights that strengthen loyalty, defend market position, and create sustainable, year-round revenue streams.

This report directly addresses these challenges by focusing on four specific analytical objectives. To improve customer relationship management, the first question posed examines the extent to which a customer's initial purchase channel can serve as a predictor of their subsequent channel usage patterns. This analysis is grounded in the 2001–2007 period, a time when the Internet channel evolved into a significant driver of revenue and considers the degree to which retail channel engagement is influenced by a customer's proximity to a physical store. Complementing this, the second question under this challenge explores store trade area effects and their impact on all channels, seeking to determine how distance from the nearest store shapes retail spending levels, channel preferences across Retail, Internet, and Catalog, and the demographic and psychographic composition of the customer base. Together, these two lines of inquiry provide a robust framework for understanding channel loyalty, switching behavior, and the geographic factors that inform long-term customer relationship strategies

To strengthen competitiveness in the highly saturated Christmas gifting market, this analysis investigates the **consistency and predictability of customer gifting behavior**. Since the Christmas season represents the company's peak sales period, the study leverages detailed transactional data that separates gift versus non-gift revenue across Internet and Catalog channels, along with seasonal counts of distinct and new gift recipients. Customers were segmented according to their overall **gifting intensity and engagement patterns**, which revealed clear distinctions between high-value gifters, occasional gifters, and low-engagement segments. These insights enable the development of **targeted retention strategies to sustain loyalty among top-tier gifters**, while also highlighting opportunities to **nurture and convert lower-value customers into more consistent contributors** during the peak season.

To address the challenge of generating consistent business during the non-Christmas months, the fourth question posed examines purchasing behavior in the Spring season compared to the Fall/Christmas period, with a focus on identifying distinct channel preferences and demographic characteristics among Spring buyers. This analysis evaluates whether customers who engage in Spring purchases differ in meaningful ways, such as proximity to stores, age distribution, or digital adoption tendencies, from those whose purchasing is concentrated solely in the holiday season. By understanding these differences, the study highlights targetable customer segments and proposes strategies to encourage off-season purchases, thereby reducing the company's reliance on the peak holiday period and creating a more balanced, year-round revenue stream.

General Approach for Research Question 1

Research Question 1 focuses on understanding how a customer's initial purchase channel influences the channel they use for later purchases. The aim is to determine whether customers remain loyal to their original channel Retail, Internet, or Catalog or if they switch to a different one over time.

To address Research Question 1, the analysis follows a two-step approach. The first half of the question is answered using a **transition matrix**, which is a framework that shows how customers move from one primary purchase channel in a base period to another in a later period. Customers who remain in the same channel appear on the diagonal, indicating loyalty, while those who switch are shown in the off-diagonal cells. This provides a clear view of migration patterns, showing which channels gain or lose customers over time and whether certain channels act as feeders to others. By summarising these movements, the transition matrix framework effectively addresses the first part of the question, understanding the extent and direction of loyalty and switching between channels.

The second half of the research question focuses on identifying the factors that influence a customer's final purchase channel in 2007, when the Internet channel had matured among customers, considering their purchasing history and store distance. This is addressed using a **multinomial logistic regression model**, a statistical method suitable when the outcome variable is categorical with more than two possible outcomes in this case, predicting the final purchase channel in 2007 (Retail, Internet, or Catalogue) based on customers' previous purchase channels from 2004 to 2006 and their distance from the store (StoreDist). The model estimates the probability of each customer ending up in each channel and quantifies how earlier channel choices and distance influence these probabilities. By incorporating predictive performance measures such as multiclass AUC, the model not only explains how past behaviour affects final channel choice but also provides a robust way to validate these predictions. This addresses the second part of the question, offering insight into the drivers of channel outcomes over time.

Model Setup

The dataset used for the transition analysis contains **100,051 rows** and **184 columns**. Before modelling, missing values relevant to this task were addressed, with particular focus on gaps in StoreDist. These variable smeasuresthe customer's distance from the nearest retail store and waerecritical for later analysis, making accurate imputation essential. The missing values were handled using a **tiered median back-off approach**. First, the median StoreDist was calculated within a narrowly defined peer group customers sharing the same **IncCode**, **AgeCode**, and **HomeValue**. If no match was found, the grouping was broadened to **IncCode** and **AgeCode**, and if still unavailable, to **AgeCode** alone. As a final fallback, the global median StoreDist across all customers was used. This method ensured that imputations were as contextually relevant as possible, preserving realistic patterns in the data while eliminating missingness.

For **feature engineering**, yearly total spend was created by summing all spending across Retail, Internet (gift and non-gift), and Catalog (gift and non-gift) for each calendar year. For example, the **2004 total spend** included all purchases made in both Spring and Fall of 2004 across all three channels; the same method was applied for 2005, 2006, and 2007. Data from

pre-2004 was excluded from this analysis, as it represented an accumulated total from 2001–2003, making it impossible to determine the exact year of purchase and risking inaccurate allocation of spending behaviour. These yearly totals were then used to determine each customer's **primary purchase channel** for that year, i.e., the channel with the **highest total spend**. This structure allowed the analysis to track customer movement from the **base year 2004** to the **final year 2007**, while also capturing intermediate transitions in 2005 and 2006, providing a complete picture of channel switching and loyalty over the entire period. After data cleaning and variable preparation, the transition matrix for **2004 to 2007** was constructed to track customers' primary purchase channel changes over time.

The next step from the transition analysis involved building a **multinomial logistic regression model** to predict each customer's final purchase channel in 2007 Retail, Internet, or Catalog. The modelling dataset was prepared by joining each customer's primary purchase channel for 2004, 2005, and 2006 (calculated earlier from yearly total spend) with their known channel for 2007, which served as the target variable. Only customers with a non-missing 2007 channel were retained. Categorical variables, such as the primary channels for earlier years, were converted into factors, while **StoreDist** (distance to the nearest store) was retained as a numeric predictor. The final model specification was chosen to capture two main influences: state dependence, where prior-year channel choices indicate loyalty or switching patterns, and spatial effects, where proximity to a store influences the likelihood of preferring Retail over Internet or Catalog.

The model estimated the probability of each customer ending in each channel based on these predictors, providing a predicted most likely channel and a quantified influence of earlier purchase behaviour and store proximity on the final 2007 outcome. This modelling approach complemented the transition matrix by moving beyond a descriptive view of channel movement to a predictive framework, offering deeper insight into how past purchase patterns and store distance drive final channel choice.

Analysis & Interpretation

After setting up both models, they were run to generate insights for the research question. For the **transition matrix[Figure 1.1]**, the results showed that loyalty was highest among Retail customers, with **88.3%** of those whose primary channel was Retail in 2004 remaining in the same channel by 2007. Internet customers also showed relatively strong loyalty, with **74.8%** staying with Internet, while Catalog customers had a loyalty rate of **78.1%**. Cross-channel switching was less common, with the largest migration flows being from **Catalog to Internet (14.3%)** and from **Internet to Catalog (14.9%)**. These findings highlight both the stability of customer channel preferences and the modest but notable movement between Catalog and Internet channels over the three-year span.

For the **multinomial logistic regression[Figure 1.2]**, the coefficients showed that prior-year channel choices were strong predictors of the final 2007 channel, confirming the effect of state dependence. For example, having Internet as the primary channel in 2006 significantly increased the likelihood of being in Internet in 2007, while a consistent history of Retail increased the likelihood of staying in Retail. Store distance had a smaller effect, with a slight negative relationship between distance and choosing Retail, indicating that customers farther from a store were marginally more inclined toward Internet or Catalog. The model achieved strong predictive performance,[Figure 1.3] with a **Hand & Till multiclass AUC of 0.947**, per-class AUCs of **0.950 for Catalog**, **0.936 for Internet**, and **0.966 for Retail**, a macro-average AUC of **0.951**, and a micro-average AUC of **0.957**. These high AUC scores demonstrate that the model could reliably distinguish between customers likely to choose each channel. Together, these results provide both a descriptive view of past customer behaviour and a predictive framework for anticipating future channel choices.

Conclusion

In summary, the analysis showed that customer loyalty was strongest in the Retail channel, with 88.3% of Retail customers in 2004 remaining in Retail by 2007, compared to 78.1% for Catalog and 74.8% for Internet. Most switching occurred between Catalog and Internet, with 14.3% of Catalog customers moving to Internet and 14.9% of Internet customers moving to Catalog, indicating a modest but notable exchange between these two channels. Internet customers had the highest overall switching rate, with one in four leaving for another channel. The multinomial logistic regression confirmed that past behaviour is a powerful predictor, as prior-year channel choices strongly influenced the final channel in 2007, reflecting state dependence in customer behaviour. Store distance played a more minor role, with greater distances slightly reducing the likelihood of remaining in Retail and nudging customers toward Internet or Catalog. The regression model also showed high predictive accuracy, demonstrating strong performance in distinguishing between customers likely to choose each channel.

General Approach for Research Question 2

Research Question 2 examines how a customer's distance from the nearest store influences their purchasing behaviour across different sales channels, Retail, Internet, and Catalog, as well as their demographic and psychographic profiles. The goal is to identify whether proximity to a store encourages customers to spend more in retail and whether those living farther away shift towards alternative channels such as catalogs or the Internet. Additionally, by linking distance to customer profiles, the analysis aims to reveal how characteristics such as age, income, and lifestyle preferences vary across distance groups.

To answer this question, the analysis uses two main techniques. First, **RFM analysis,** which evaluates Recency (how recently a customer purchased), Frequency (how often they purchase), and Monetary value (total spend), is applied to segment customers by engagement and value. This provides a clear behavioural snapshot for each store distance group. Second, an **ANOVA (Analysis of Variance) test** is used to statistically determine whether the differences in RFM scores and spending patterns between distance groups are significant. Together, these methods provide both descriptive insights into behavioural trends and statistical confirmation of whether store proximity meaningfully impacts channel choice, spending habits, and customer profiles.

Model Setup

The dataset used for this analysis contained over 99,000 customer records, each including purchase history from 2004 to 2007, demographic details, psychographic traits, and the customer's distance to the nearest store. Customers were grouped into four distance bins: 0–10 km, 11–20 km, 21–30 km, and 31+ km. This binning allowed for clear comparisons of spending behaviour and profile characteristics across distinct store trade areas.

The dataset used for this analysis contained over 99,000 customer records, including purchase history from 2004 to 2007, demographic information, psychographic traits, and each customer's distance to the nearest store. Customers were grouped into four distance bins: **0–10 km**, **11–20 km**, **21–30 km**, and **31+ km**, enabling clear comparisons of spending behaviour and customer profiles across different store trade areas. Lifetime spend variables were created for each channel—Retail, Internet, and Catalog by summing purchases across all seasons and years. These values were then aggregated by distance group to calculate the share of spend by channel, providing an initial view of how purchasing patterns shift with distance from the store.

Several feature engineering steps were applied to refine and simplify the demographic and psychographic data for analysis. A **HasChild** variable was created by combining five separate child age range indicators (0–2, 3–5, 6–11, 12–16, and 17–18 years) into a single flag, marking "Yes" if a household had a "Y" in any category and "No" otherwise. An **Outdoorsy** variable was created to capture customers with outdoor-oriented interests, combining Hunting, Camping, Boating, and Travel, while an **Indoorsy** variable was created to represent customers with indoor-oriented hobbies, such as Sewing, Needlework, Collecting, Fine Arts, and Cooking. **PetOwner** merged Pets and DogOwner into one category, **LuxuryLifestyle** grouped customers with interests in premium or high-end activities, such as Wines, Fine Arts, and Fashion, while **WellnessLifestyle** combined customers focused on personal health and self-improvement, through interests like Exercise and Self Help. These transformations made customer profiles more interpretable, highlighted broader lifestyle patterns, and reduced complexity in the data.

For the behavioural analysis, **RFM scores** were calculated for each customer. Recency scores were based on the most recent purchase season, Frequency scores on the total number of purchases, and Monetary scores on total lifetime spend. The scores were weighted out of 100, with Recency = 25, Frequency = 25, and Monetary = 50. This weighting was chosen because, for a gift-oriented business model, the total value a customer spends (Monetary) directly impacts revenue, so it receives the highest weight. Recency and Frequency are equally crucial for understanding customer engagement and purchasing habits, which is why they share the remaining weight equally. An overall RFM score was then calculated by summing the three components. Finally, **ANOVA tests** were run to compare RFM scores and spending behaviour across the four distance groups, providing statistical confirmation of whether store distance significantly influences customer engagement and purchasing patterns.

For the behavioural analysis, RFM scores were calculated for each customer. Recency scores were based on the most recent purchase season, Frequency scores were determined from total purchase counts, and Monetary scores were derived from lifetime spend, with weights of 25 for Recency, 25 for Frequency, and 50 for Monetary to reflect the greater business impact of spend value. An overall RFM score was then computed by summing these three components.

Finally, ANOVA tests were run to compare spending and RFM metrics across the four distance bins. This statistical step was essential to confirm whether observed differences in engagement and spend were significant, and to quantify the impact of store distance on customer behaviour.

Analysis & Interpretation

After setting up both models, the second research question was addressed by first examining how customers' channel preferences change depending on how far they live from a store. The first stage of the analysis examined channel mix by store distance. The results[Figure 2.1] showed a clear and consistent trend: customers living closer to a store allocated a significantly larger share of their spending to Retail, with nearly 46% of spend from those within 0–10 km occurring in this channel. As distance increased, retail's share declined steadily, reaching just 28.7% for customers living 31 km or more away. In contrast, Catalog spending rose with distance from 29.1% in the nearest group to 43.8% in the farthest group indicating that customers further away relied more on remote purchasing. Internet spending remained relatively stable across all distance groups, fluctuating only between 24% and 31%, suggesting that its role is consistent regardless of proximity. These findings confirm a shift from Retail to Catalog as distance increases, while Internet remains a supplementary but steady channel.

Next, to address the second half of the research question, **psychographic profiles by store distance[Figure 2.2]** to see how customer lifestyles change with location. Some traits stayed consistent, like families with children, which made up about 23% of customers in every distance group, suggesting family status is not location-dependent. Indoorsy interests, such as sewing, fine arts, and cooking, rose modestly from 16.4% in the nearest group to 18.1% in the farthest. LuxuryLifestyle interests, including wines and fashion, increased slightly from 15.8% to 16.2%. More notable changes were seen in Outdoorsy activities (e.g., camping, boating, travel), which rose from 17.8% to 20.0%, and PetOwner households grew from 13.5% to 16.4% with distance. Wellness lifestyle traits, such as exercise and self-help, peaked at 20.6% for customers living more than 31 km from a store. These patterns suggest that customers further from stores tend to show greater interest in outdoor activities, pet ownership, and wellness-oriented lifestyles.

Demographic profiles provided further context. [Figure 2.3]Age distribution was stable, with middle-aged customers (35–54 years) making up roughly half of all groups. However, there was a gradual increase in older customers (55–74 years) in the farther distance bins.[Figure 2.4] Income showed a more apparent effect: higher-income households ($150k+) were more common close to stores (33.7%) but dropped to 22.1% in the farthest group, while lower- and middle-income groups increased with distance. This shift implies that premium, higher-value products may be better targeted to customers within 20 km of a store, while more value-oriented offers may appeal to those further away.

Finally, the RFM & ANOVA analysis quantified engagement differences across distance groups. Customers within 0–10 km recorded the most recent purchases, with an **average** recency score of 13.4[Figure 2.8], purchased most frequently with an **average** frequency score of 14.0, and achieved the highest **average** monetary value at 31.4. These factors combined to give them the strongest overall **average** RFM score of 58.8. In comparison, customers living 31 km or more away had an **average** recency score of 12.9, an **average** frequency score of 13.1, and the lowest **average** monetary value at 30.5, resulting in an overall **average** RFM score of 56.5.[Figure 2.5 - Figure 2.7] ANOVA results confirmed that these differences were statistically significant ($p < 0.001$) and unlikely to have occurred due to random variation. Importantly, nearby customers spent up to $28 more in Retail than the farthest group, while distant customers spent up to $10 more in Catalog. This reinforces the earlier finding from the channel mix analysis that distance influences not just the preferred channel but also the customer's overall value and engagement level.

Conclusion

The analysis found that channel preferences vary notably with store distance. Customers closer to a store allocate a larger share of spending to retail, while those farther away increasingly rely on catalogs, with the Internet maintaining a steady share across all distances. Psychographic profiling showed stable family presence across distance groups but revealed that customers farther from stores are more inclined toward outdoor activities, pet ownership, and wellness-focused lifestyles. Demographic trends indicated a stable middle-aged majority, a higher proportion of older customers in distant areas, and a decline in high-income households with increasing distance. RFM analysis demonstrated that proximity to stores is linked to more recent, frequent, and higher-value purchases, while ANOVA testing confirmed these differences as statistically significant, highlighting that location is a meaningful driver of customer behaviour and channel choice.

General Approach for Research Question 3

Research Question 3 applies the Segmentation–Targeting–Positioning (STP) framework to analyse customers engaged in Christmas and Fall gifting purchases, with the aim of identifying distinct behavioural groups and designing strategies tailored to their needs. By uncovering meaningful differences within the gifting customer base, the business can replace broad, uniform campaigns with targeted initiatives that align with the unique behaviours, preferences, and value potential of each segment.

Segmentation was performed using an unsupervised k-means clustering approach, with cluster formation based on key transactional gifting variables: total gift spends across Internet and Catalog channels, order volumes, number of product lines purchased, and the number of gift recipients (including new recipients). This ensured that each cluster captured genuine, behaviour-driven distinctions in gifting patterns rather than superficial or one-dimensional differences.

Once the clusters were identified, the Targeting phase assessed their relative size, value contribution, and demographic characteristics to determine where marketing resources could be most effectively deployed. For example, high-value Internet-dominant gifters may present strong growth opportunities, while Catalog-loyal customers may require retention-focused strategies to preserve their contribution.

The Positioning stage overlaid demographic and psychographic data, including household income, home value, age, and dwelling type to transform clusters into actionable customer personas. This made it possible to develop precise positioning statements and craft marketing communications that resonate with each segment's characteristics, purchase motivations, and preferred channels, ensuring that each group receives the most relevant and effective messaging.

Model Setup

The dataset for this analysis was drawn from the DMEF extract, which contains longitudinal transaction-level purchase records from 2004 to 2007. To prepare the data for clustering, a structured sequence of cleaning, transformation, and feature engineering steps was applied, with the aim of isolating customer behaviours specific to Fall/Christmas gifting activity.

The preparation began by limiting the scope to Fall seasons only, removing all Spring-related variables to maintain a seasonal focus. Retail channel variables were excluded to simplify the analysis and concentrate on Internet and Catalog gifting behaviours. Within these channels, only gift-related features were retained which were approximately 30,000 gifters [Figure 3.5], including gift dollars (total spend on purchases marked as gifts), number of gift recipients, and number of new gift recipients from 2004 to 2007, including gift dollars (total spend on purchases marked as gifts), number of gift recipients (unique individuals receiving gifts), and number of new gift recipients (first-time gift recipients) for each year from 2004 to 2007. To capture broader behavioural depth, Internet and Catalog order counts and line counts were also included. These yearly measures were aggregated to the customer level, producing key features such as *Int_TotalGift, Cat_TotalGift, Int_TotalOrders, Cat_TotalOrders, TotalGiftRec*, and *TotalNewGR*. A *StoreDist* variable, representing proximity to the nearest store, was also retained to incorporate geographic context into the clustering.

To address data integrity and ensure balanced influence of all records, Extreme outliers were capped so that the very top spenders those whose activity was far beyond what most customers spend did not distort the clustering results. Customers above these caps were flagged with a VIP indicator, allowing them to remain in the dataset for strategic insight without distorting clustering distances. Missing values in StoreDist were imputed using the median. Because gift spend, order counts, and recipient counts were heavily right-skewed, a log(1+x) transformation was applied to reduce skewness. All features were then scaled using a robust scaler based on median and interquartile range (IQR), minimising the influence of any residual outliers. The variable All_TotalGift was dropped from the clustering set to avoid overweighting total gift spend, as it was a linear combination of Internet and Catalog spending.

Clustering was performed using the k-means algorithm with Euclidean distance on the scaled feature set. Multiple random starting points (nstart = 200) were used to determine the initial cluster centroids, ensuring that the final customer groups were stable and not dependent on a single, potentially unrepresentative starting position. In business terms, this means the segmentation results are reliable and repeatable, rather than being influenced by random chance in how the groups were first drawn. We tested different options for the number of customer groups (k), ranging from 2 groups up to 8 groups, and compared them using the silhouette score, a measure that tells how well each customer fits within their assigned group compared to other groups. In this context, k simply refers to the number of clusters or segments we are trying to create. [Figure 3. 3] The k = 4 solution achieved an average **silhouette score of 0.514[Figure 3.4]**, a notably strong result in consumer retail datasets, where behavioural overlap often suppresses silhouette values. The four-cluster solution not only demonstrated statistical robustness but also offered clear managerial interpretability, mapping directly to four intuitive segments: **Casual Gifters, Internet Gifters, Catalog Gifters, and Omni-Channel VIP Gifters**. This alignment of statistical performance and business relevance made k = 4 the optimal choice for segmentation.

Analysis & Interpretation

[Figure 3.1] The four-cluster segmentation achieved in this analysis provided a balanced and actionable breakdown of the customer base. Each cluster exhibited distinct gifting behaviors and spending patterns, allowing for clear business differentiation. The **first segment**, *Casual Gifters*, represents occasional holiday shoppers with below-median Internet and Catalog spend, low order and line counts, and a limited number of gift recipients or new recipients acquired each year. This group forms the baseline, lower-value tier in the seasonal customer hierarchy.

The **second segment**, *Internet-Dominant Gifters*, displayed consistently higher spend, order volume, and line items on the Internet channel compared to Catalog, coupled with above-average numbers of gift recipients and new gift recipients. Their profile suggests a younger, more digitally oriented customer who is comfortable transacting online and represents the growth potential of gifting sales through digital channels.

The **third segment**, *Catalog-Dominant Gifters*, was comprised of traditional buyers whose spend, and order volume were concentrated in the Catalog channel, with minimal Internet activity. While they maintained a stable recipient base, they acquired fewer new gift recipients than their Internet-heavy counterparts. These customers are typically older, more settled, and exhibit loyalty to the Catalog format.

The **fourth** and most **lucrative segment**, *Omni-Channel Heavy Gifters (VIP-rich)*, combined high engagement across both Internet and Catalog channels with the largest gift recipient bases and the highest number of new gift recipients. A substantial proportion carried a VIP flag, marking them as disproportionately valuable revenue contributors. Their willingness to engage with multiple touchpoints underscores their strategic importance for both retention and cross-channel marketing.

Silhouette analysis[Figure 3.4] confirmed that customers were more like those in their assigned clusters than to those in other clusters, even if the overall separation was moderate, a typical outcome in retail consumer datasets where behaviours overlap. The cluster size distribution was healthy, with no single group dominating[Figure 3.3], ensuring each segment represents a meaningful portion of the base.

Demographic [Figure 3.2] enrichment further sharpened the business relevance of each cluster. *Casual Gifters* were linked to lower home values, shorter residence lengths, and younger renter profiles, suggesting less disposable income and stability. *Internet Gifters* skewed toward younger, higher-income professionals with urban, mobile lifestyles, matching their affinity for digital shopping. *Catalog Gifters* were generally older homeowners with higher home values and longer residence tenure, reflecting established, traditional buying patterns. *Omni-Channel VIP Gifters* spanned demographics but leaned toward affluence and homeownership, aligning with their premium omni-channel shopping behavior.

By combining behavioural segmentation with demographic insight, this clustering delivers not only statistical validity but also a clear roadmap for differentiated targeting and positioning strategies that can maximise both seasonal sales and long-term customer value.

Conclusion

In summary, the Christmas/Fall gifting customer base is far from homogeneous, with four distinct and actionable segments emerging from the k-means clustering within the STP framework. These segments differ not only in transactional behaviours such as spend levels, order volumes, recipient counts, and channel preferences but also in demographics, including age, income, home value, and residential stability. Casual Gifters are low-spend, occasional buyers who present an upsell opportunity through tailored seasonal promotions. Internet Gefters are younger, more affluent, and digitally inclined, making them prime candidates for aggressive cultivation through personalised online experiences and fast, convenient fulfilment. Catalog Gifters are older, traditional customers loyal to the catalog channel, and should be retained through trust-based, familiar marketing while being gradually migrated toward digital channels. Omni-Channel VIP Gefters are affluent, multi-channel shoppers who contribute disproportionately to seasonal revenue and require premium, high-touch engagement to maintain loyalty. The strategic imperative is to protect and reward the high-value omni-channel segment, grow the digital-heavy group, sustain catalog loyalty while encouraging channel diversification, and convert occasional buyers into more consistent contributors moving from one-size-fits-all campaigns to precision targeting that maximises both seasonal and long-term customer value.

General Approach for Research Question 4

Research Question 4 explores purchasing behavior across seasons, aiming to compare sales in the dominant Fall/Christmas period with those in Spring and determine how the **STP (Segmentation–Targeting–Positioning)** framework can be used to reduce dependency on the peak holiday season. In this context, **Segmentation** involves dividing customers into groups based on their seasonal purchasing patterns: Fall-only buyers, Spring-only buyers, Mixed buyers who shop in both seasons, and customers with no seasonal spend [Figure 4.4]. This breakdown allows the business to clearly see where current revenue is concentrated and where there is potential to grow Spring sales.

Once segments are defined, **Targeting** identifies which groups present the greatest commercial opportunity. Fall-only buyers may be the primary focus for Spring conversion, Spring-only buyers require retention strategies, and Mixed buyers offer opportunities for cross-season cross-selling. **Positioning** then shapes the messaging, promotional offers, and channel strategies tailored to each segment, for example, using online promotions and in-store events for Fall-only buyers, loyalty programs for Spring-only buyers, and bundled offers for Mixed buyers. By applying STP in this way, strategies can be customised for each group, helping to smooth seasonal sales fluctuations and build steadier, year-round revenue streams.

Model Setup

The dataset for Research Question 4 was prepared using the same cleaning and imputation procedures applied in Research Question 1 to ensure consistency across analyses. This included handling missing values for key predictor variables in the logistic regression model: **AgeCode** (customer age group), **IncCode** (household income group), and **StoreDist** (distance to the nearest store). Missing entries in these fields were resolved using a **tiered peer-group median imputation** strategy, starting with narrowly matched groups (e.g., similar age, income, and home value), relaxing the matching criteria step-by-step, and falling back to the overall dataset median only when no suitable peer group was found. This approach preserved realistic demographic and geographic distributions while ensuring that no observations were dropped from the Spring-purchase propensity model.

In addition to cleaning, Seasonality features are then engineered to align analysis to the business question. Several key derived variables were created to capture customer spending patterns across seasons. **Seasonal spend totals** were calculated by summing all Spring purchases across the years 2004–2007 to form **Spend_Spring**, and all Fall purchases over the same period to form **Spend_Fall**. These were then combined to produce **TotalSpend**, representing each customer's total expenditure across both seasons. From these measures, a **Seasonality** index was computed as the share of total spend occurring in Spring, providing a direct measure of seasonal purchase bias. Binary flags for **SpringBuyer** and **FallBuyer** were also generated to identify whether a customer made any purchases in each season. These engineered variables formed the foundation for both descriptive seasonal analyses and the logistic regression model predicting Spring-buying likelihood.

To connect seasonality with channel behavior, channel-usage indicators are created for the full 2004–2007 window: Use_Retail, Use_Internet, and Use_Catalog mark whether a customer ever purchased through each channel. These flags enable like-for-like comparisons of channel mix between Spring buyers and Fall-only buyers and serve as descriptive inputs to the STP layer.

With the seasonality index in hand, the customer base is segmented into action-oriented groups consistent with STP thinking Fall-only (no Spring spend), Spring-only (no Fall spend), Mixed (Fall-leaning), Mixed (Spring-leaning), and No Spend (no seasonal activity). For each segment, simple KPIs are prepared segment size and share, average seasonality, and the share using each channel to inform sizing, targeting, and positioning decisions in later strategy.

A predictive layer is then specified to quantify what drives Spring purchasing. A logistic regression is fit with SpringBuyer as the outcome. Predictors include StoreDist (numeric distance to the nearest store) and FirstChannel (the first recorded primary channel), alongside demographic controls (AgeCode and IncCode) to account for observable differences in the base. This model estimates the probability that a customer will purchase in Spring and isolates the incremental impact of proximity and digital/retail starting points after controlling for demographics as shown in [Figure 4.6].

Overall, the model setup integrates cleaned and imputed demographic variables, engineered seasonal spend measures, and channel usage indicators to create a robust analytical base for exploring seasonal purchasing behaviour. By aligning data preparation with the methodology used in earlier questions, the model ensures comparability while retaining full sample coverage. The logistic regression specification, incorporating Age, Income, Store Distance, and First Purchase Channel, is designed to quantify the relative influence of customer profile and channel history on the likelihood of Spring purchasing. This preparation stage ensures that the subsequent analysis can directly link statistically validated drivers to actionable strategies within the STP framework.

Analysis & Interpretation

To address Research Question 4, [Figure 4.3] a logistic regression model was applied to estimate the likelihood of a customer making a Spring purchase, using Age, Income, Store Distance, and First Purchase Channel as predictors. The analysis revealed that the customer's first purchase channel was the strongest driver of Spring purchasing[Figure 4.5]. Customers whose first recorded transaction was through Retail were almost twice as likely to make a Spring purchase compared to those who started in the Catalog channel, while those starting via Internet were also noticeably more likely to purchase in Spring. This shows a clear advantage for customers acquired through store or digital channels when it comes to encouraging off-season buying.

Store distance emerged as the second most important factor. Customers living closer to a store had a higher probability of purchasing in Spring, reinforcing the idea that Spring shoppers tend to be convenience-driven, making use of nearby store access for non-holiday shopping. Age had a small but positive effect, with slightly older customers showing marginally higher Spring engagement. Income, however, did not have a statistically significant impact, suggesting that affluence alone does not explain seasonal buying patterns instead, channel experience and physical proximity to a store are more decisive.

The seasonality index segmentation provided further context.[Figure 4.2] Out of the 100,051 customers, [Figure 4.7]47% purchased only in Fall, 16% purchased only in Spring, 10% were mixed but Fall-leaning, 5% were mixed but Spring-leaning, and 21% made no seasonal purchases at all. This means that nearly half the customer base generates revenue only during the peak holiday season, creating a heavy seasonal dependency. However, around one-third of the base already contributes to Spring sales, representing an immediate opportunity for expansion.

Channel usage patterns confirmed these differences.[Figure 4.1] In Fall, customers split their spend more evenly across Retail (42%), Internet (32%), and Catalog (27%). In Spring, however, Retail's share jumps to 61%, while Internet drops slightly to 26% and Catalog declines sharply to 19%. This shift indicates that Spring buyers are far more store-oriented than Fall buyers.

Demographic analysis revealed that Spring buyers live closer to stores, with an average distance of around 18 miles, and tend to have slightly higher home values, suggesting they are more urban and somewhat more affluent. STP segmentation analysis showed that Fall-only and Fall-leaning customers together make up 57% of the base and have higher reliance on Catalog, while Spring-leaning and Spring-only segments favour Retail and Internet channels. Moving even 10% of Fall-only customers into Spring activity could add roughly 6,000 Spring buyers without acquiring new customers.

Conclusion

In summary, the analysis shows that seasonal purchasing patterns are heavily skewed toward the Fall period, with nearly half of the customer base buying only during this time. However, there is significant untapped potential in Spring, as one-third of customers already make purchases outside the holiday season. The logistic regression results highlight that a customer's first purchase channel and their proximity to a store are the most influential drivers of Spring engagement, while age has a minor positive effect and income is not a significant factor. STP segmentation reveals distinct behavioural profiles, Fall-only and Fall-leaning customers rely more on Catalog, while Spring-leaning and Spring-only customers prefer Retail and Internet. This suggests that targeted interventions such as shifting Catalog customers toward digital and in-store channels, leveraging proximity with geo-targeted promotions, and creating retention incentives for existing Spring customers can meaningfully reduce seasonal dependency. By combining predictive modelling with actionable segmentation, the findings offer a clear, data-driven blueprint for building more balanced, year-round sales.

Recommendation

The analysis provides a clear and actionable roadmap for addressing the company's key challenges, offering targeted strategies to strengthen loyalty, improve channel performance, and drive sustainable growth.

Nurturing Loyalty Based on First Purchase Channel:
It begins with where each customer's journey starts. The data shows that a customer's first purchase channel whether Retail, Internet, or Catalog sets the tone for their future relationship with the brand. Retail starters are the most loyal, with nearly nine out of ten staying in-store over three years. Internet starters are also relatively loyal, while Catalog customers tend to drift between channels, especially toward Internet. This means loyalty is not a mystery it can be nurtured. To keep the Retail group engaged, the business can lean on what makes the in-store experience special: exclusive events, personalised service, and rewards for repeat visits. For Internet starters, the key is to keep the digital experience sharp personalised recommendations, subscription-style gifting plans, and smooth checkout. Catalog starters, while more traditional, can be introduced to digital touchpoints without alienating them think scannable codes in catalogs or easy online order tracking. By treating each starter group according to its strengths, the company can strengthen retention and gently guide customers toward the channels that deliver the highest return.

Targeting Strategies Based on Geographic Proximity:
Where customers live also shapes how they buy. Those living within 10 kilometres of a store spend more, shop more often, and choose Retail far more frequently than their distant counterparts. Further out, Catalog becomes the channel of choice, with Internet staying steady across all distances. This insight makes targeting more efficient. Local customers can be drawn in with in-store exclusives, same-day pickup, and community-based promotions. Those further away will respond better to convenience-free delivery thresholds, extended return windows, and bundled offers that make remote shopping as attractive as a store visit. Matching the offer to the geography means every marketing dollar works harder.

Tailoring Engagement to Gifting Segments:
The gifting market the company's crown jewel also isn't one-size-fits-all. The segmentation revealed four distinct personas: Casual Gifters, Internet Gifters, Catalog Gifters, and Omni-Channel VIP Gifters. From a positioning standpoint, *Casual Gifters* require low-barrier holiday promotions and simple gift guides to increase seasonal participation. *Internet Gifters* should be targeted as digital natives with mobile-first experiences, personalized online recommendations, and expedited delivery. *Catalog Gifters* benefit from traditional loyalty programs, print marketing, and an emphasis on trust and familiarity. *Omni-Channel VIP Gifters* demand exclusivity, early access to promotions, and premium loyalty rewards to sustain their high engagement and revenue contribution.

Unlocking Off-Season Revenue Potential:
Finally, there's the challenge of seasonality. Right now, nearly half of all customers buy only in Fall, but one-third already make purchases in Spring often the same customers who started in Retail or Internet, and those living closer to stores. This is an untapped goldmine. By targeting Fall-only and Fall-leaning customers with Spring-specific campaigns tied to events like Easter, Mother's Day, and graduations, the company can bring more buyers into the off-

season. Spring buyers are more store- and Internet-oriented, so campaigns should focus on those channels, while using catalog communications to nudge Catalog-heavy customers into trying them. Converting just 10% of Fall-only customers into Spring buyers could mean thousands of new orders without spending a dollar on customer acquisition.

Integrated Growth Strategy:
Together, these insights form a clear growth strategy. By nurturing customers based on their first channel, tailoring offers to their location, personalising engagement for distinct gifting segments, and unlocking Spring's potential, the company can address its three biggest challenges: reduce seasonal dependency, strengthen loyalty, and improve channel performance. This isn't about casting a wider net, it's about fishing in the right waters, with the right bait, at the right time. The data shows exactly where those waters are.

Reference & Exhibits

[1] OpenAI. (2025). *ChatGPT* (Feb 2025 version) [Large language model]. OpenAI. https://chat.openai.com/

[2] Grammarly, Inc. (2025). *Grammarly* [Writing assistance software]. Grammarly https://www.grammarly.com/

[3] DecisionPro, Inc. (2025). *Enginius* [Marketing analytics software]. Enginius. https://www.enginius.biz/

[4] Canva Pty Ltd. (2025). *Canva* [Graphic design platform]. Canva. https://www.canva.com/

EXHIBITS:

	CATALOG	INTERNET	RETAIL
CATALOG	1858 (78.1%)	340 (14.3%)	181 (7.6%)
INTERNET	191 (14.9%)	960 (74.8%)	132 (10.3%)
RETAIL	161 (6.3%)	139 (5.4%)	2267 (88.3%)

Figure 1.1 Customer Channel Loyalty and Switching Patterns, 2004–2007

```
> summary(multi_model)
Call:
multinom(formula = Purchase_07 ~ Purchase_04 + Purchase_05 +
    Purchase_06 + StoreDist, data = model_data)

Coefficients:
         (Intercept) Purchase_04Internet Purchase_04Retail Purchase_05Internet
Internet   -2.882585            1.676691          1.124815           1.5694020
Retail     -3.979052            1.090304          2.283065           0.9761141
         Purchase_05Retail Purchase_06Internet Purchase_06Retail     StoreDist
Internet         0.3663055            2.744254          1.108190  0.0001066745
Retail           2.1842350            1.685709          3.213777 -0.0004051550
```

Figure 1.2 Multinomial Logistic Regression Coefficients for Predicting 2007 Primary Purchase Channel

```
Hand & Till multiclass AUC: 0.947
> cat("Per-class AUC:\n")
Per-class AUC:
> print(round(res_auc$per_class_auc, 3))
 Catalog Internet   Retail
   0.950    0.936    0.966
> cat("Macro-average AUC:", round(res_auc$macro_auc, 3), "\n")
Macro-average AUC: 0.951
> cat("Micro-average AUC:", round(res_auc$micro_auc, 3), "\n")
Micro-average AUC: 0.957
```

Figure 1.3 Multiclass AUC Performance Metrics for Channel Prediction Model

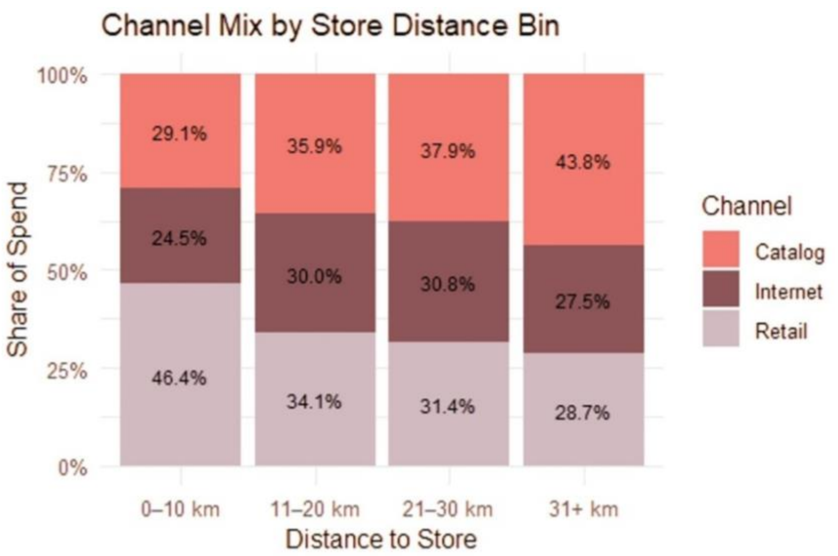

Figure 2.1 Channel Spending Share by Store Distance Bin

PROFILE VARIABLES	0-10 KM	11-20 KM	21-30 KM	31+ KM
HasChild	23.8	23.6	22.9	23.2
Indoorsy	16.4	16.1	17.0	18.1
Luxury Lifestyle	15.8	15.3	15.5	16.2
Outdoorsy	17.8	17.8	18.1	20.0
PetOwner	13.5	14.0	13.9	16.4
WellnessLifestyle	20.2	19.3	19.6	20.6

Figure 2.2 Psychographic Profile Distribution by Store Distance Bin

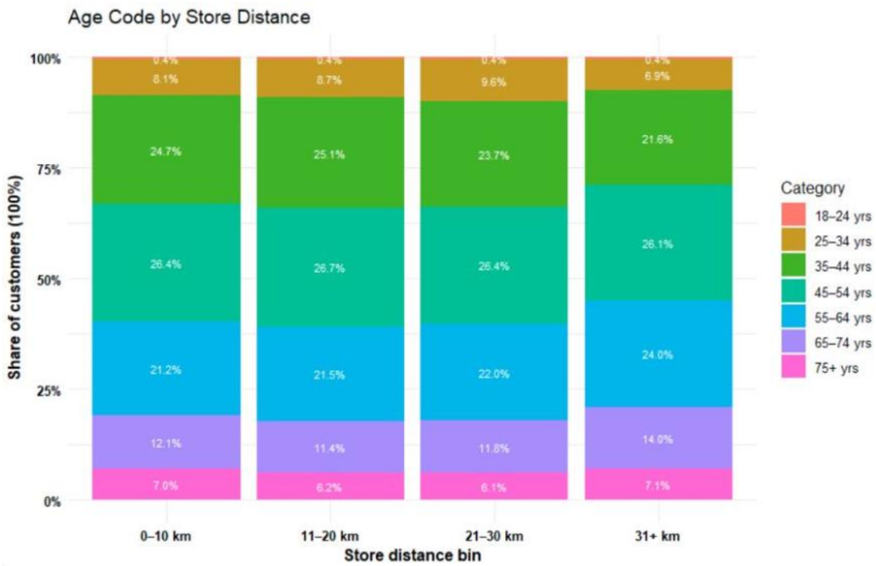

Figure 2.3 Age Distribution of Customers by Store Distance Bin

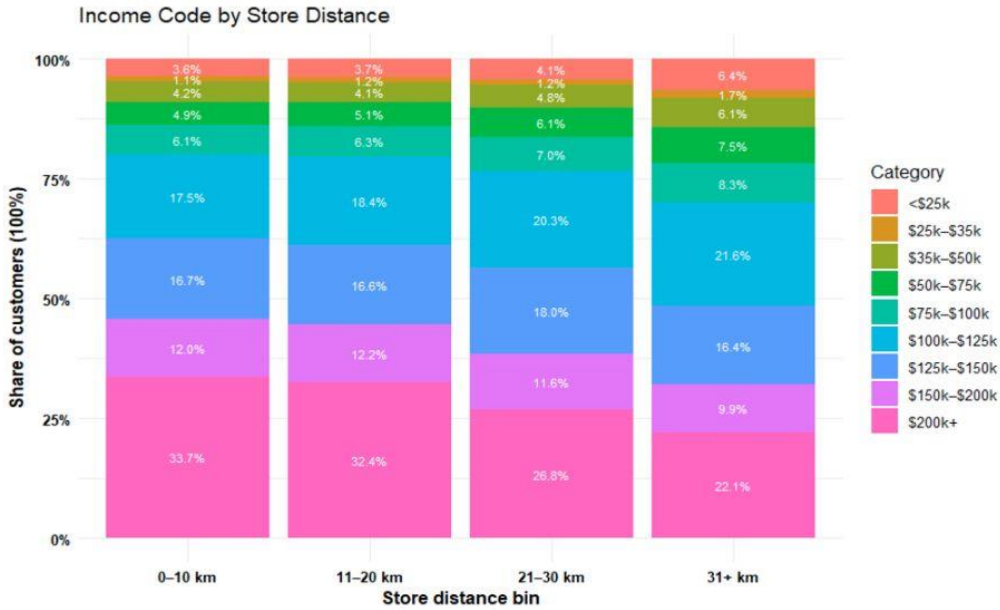

Figure 2.4 Household Income Distribution by Store Distance Bin

```
============================
Variable: retail_lifetime_dollars
                 Df    Sum Sq Mean Sq F value Pr(>F)
StoreDist_bin     3 1.241e+07 4137238   215.3 <2e-16 ***
Residuals     99079 1.904e+09   19212
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
968 observations deleted due to missingness

Post-hoc Tukey Test:
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = formula, data = df_Trade_test)

$StoreDist_bin
                       diff        lwr         upr     p adj
11-20 km-0-10 km  -19.037232 -22.149028 -15.9254349 0.0000000
21-30 km-0-10 km  -24.671286 -28.378080 -20.9644931 0.0000000
31+ km-0-10 km    -27.566975 -30.490527 -24.6434233 0.0000000
21-30 km-11-20 km  -5.634055  -9.388162  -1.8799475 0.0006673
31+ km-11-20 km    -8.529744 -11.513058  -5.5464299 0.0000000
31+ km-21-30 km    -2.895689  -6.495300   0.7039221 0.1641084
```

Figure 2.5 ANOVA and Tukey Post-hoc Test Results for Retail Lifetime Spend by Store Distance Bin

```
==============================
Variable: internet_lifetime_dollars
                 Df    Sum Sq Mean Sq F value   Pr(>F)
StoreDist_bin     3 4.105e+05  136836   8.006 2.48e-05 ***
Residuals     99079 1.694e+09   17093
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
968 observations deleted due to missingness

Post-hoc Tukey Test:
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = formula, data = df_Trade_test)

$StoreDist_bin
                       diff        lwr         upr     p adj
11-20 km-0-10 km   3.828738  0.8936064  6.76386876 0.0044721
21-30 km-0-10 km   2.264735 -1.2316128  5.76108348 0.3428564
31+ km-0-10 km    -1.212282 -3.9698554  1.54529199 0.6713235
21-30 km-11-20 km -1.564002 -5.1049783  1.97697379 0.6680097
31+ km-11-20 km   -5.041019 -7.8549620 -2.22707657 0.0000248
31+ km-21-30 km   -3.477017 -6.8722680 -0.08176613 0.0423353
```

Figure 2.6 ANOVA and Tukey Post-hoc Test Results for Internet Lifetime Spend by Store Distance Bin

```
==============================
Variable: catalog_lifetime_dollars
                 Df    Sum Sq Mean Sq F value   Pr(>F)
StoreDist_bin     3 1.647e+06  549164   17.31 3.12e-11 ***
Residuals     99079 3.143e+09   31725
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
968 observations deleted due to missingness

Post-hoc Tukey Test:
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = formula, data = df_Trade_test)

$StoreDist_bin
                        diff        lwr       upr     p adj
11-20 km-0-10 km   4.7094495  0.7106915  8.708208 0.0132458
21-30 km-0-10 km   4.0347019 -0.7286458  8.798050 0.1298984
31+ km-0-10 km    10.4125983  6.6557407 14.169456 0.0000000
21-30 km-11-20 km -0.6747476 -5.4988953  4.149400 0.9841197
31+ km-11-20 km    5.7031488  1.8694952  9.536802 0.0007640
31+ km-21-30 km    6.3778964  1.7522814 11.003511 0.0022441
```

Figure 2.7 ANOVA and Tukey Post-hoc Test Results for Catalogue Lifetime Spend by Store Distance Bin

StoreDist_bin	avg_recency	avg_frequency	avg_monetary	avg_RFM_Total
0-10 km	13.4	14.0	31.4	58.8
11-20 km	13.0	13.3	31.1	57.3
21-30 km	13.0	12.9	30.7	56.6
31+ km	12.9	13.1	30.5	56.5

Figure 2.8 Average RFM Scores by Store Distance Bin

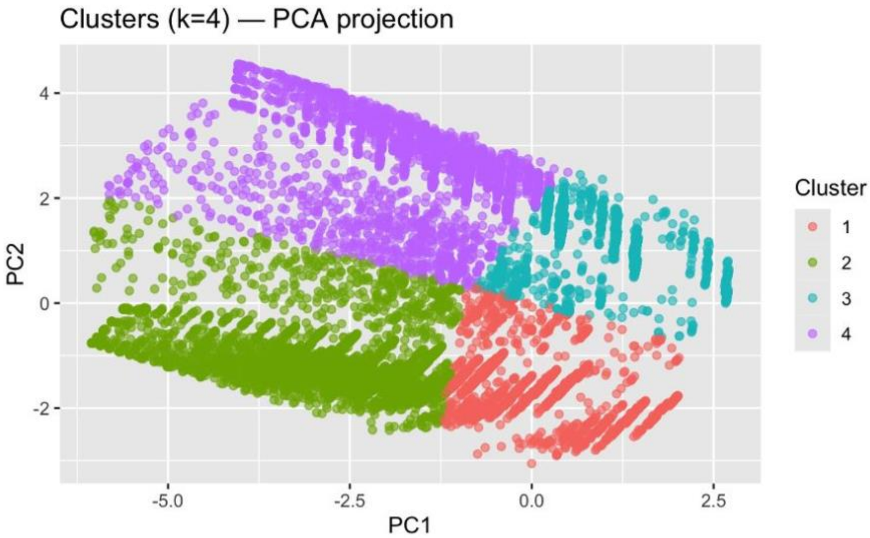

Figure 3.1 PCA Projection of Customer Clusters (k = 4)

Cluster <fctr>	LengthRes <dbl>	StoreDist <dbl>	AgeCode <int>
1	10	20.25	5
2	11	20.54	5
3	7	18.30	3
4	7	17.90	4

Figure 3.2 Cluster Averages for Residence Length, Store Distance, and Age Code

1	2	3	4
10795	4569	11564	3940

Figure 3.3 Customer Counts by Cluster (k = 4)

k <int>	avg_sil <dbl>
3	0.4830754
4	0.5142461

Figure 3.4 Average Silhouette Scores for k = 3 and k = 4 Cluster Solutions

GifterFlag <chr>	n <int>	Percent <dbl>
Gifters	30868	30.85227
Non-Gifters	69183	69.14773

Figure 3.5 Distribution of Gifters and Non-Gifters in the Customer Base

Channel Usage

SEASON	RETAIL	INTERNET	CATALOG
Fall	42%	32%	27%
Spring	61%	26%	19%

Figure 4.1 Channel Usage Share in Fall and Spring Seasons

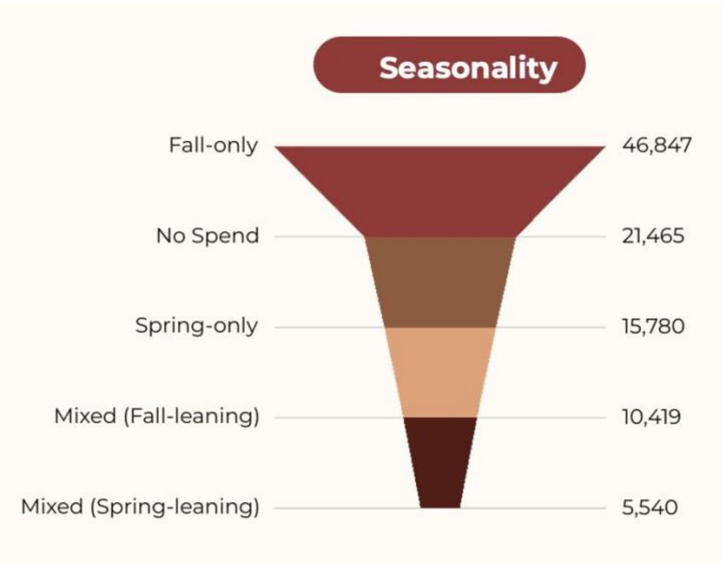

Figure 4.2 Seasonal Customer Segmentation Based on Purchase Timing

```
=== Logistic Model: SpringBuyer ~ Age + Income + Distance + FirstChannel ===
> print(summary(model))

Call:
glm(formula = SpringBuyer ~ suppressWarnings(as.numeric(AgeCode)) +
    suppressWarnings(as.numeric(IncCode)) + suppressWarnings(as.numeric(StoreDist)) +
    FirstChannel, family = binomial, data = dat)

Coefficients:
                                         Estimate Std. Error z value Pr(>|z|)
(Intercept)                            -1.379e+00  4.514e-02 -30.537  < 2e-16 ***
suppressWarnings(as.numeric(AgeCode))   2.042e-02  6.098e-03   3.348 0.000815 ***
suppressWarnings(as.numeric(IncCode))   3.163e-03  3.959e-03   0.799 0.424331
suppressWarnings(as.numeric(StoreDist)) -2.133e-04  4.998e-05  -4.268 1.97e-05 ***
FirstChannelInt                         3.439e-01  1.955e-02  17.593  < 2e-16 ***
FirstChannelRet                         9.547e-01  1.685e-02  56.647  < 2e-16 ***
---
```

Figure 4.3 Logistic Regression Coefficients for Predicting Spring Purchasing Likelihood

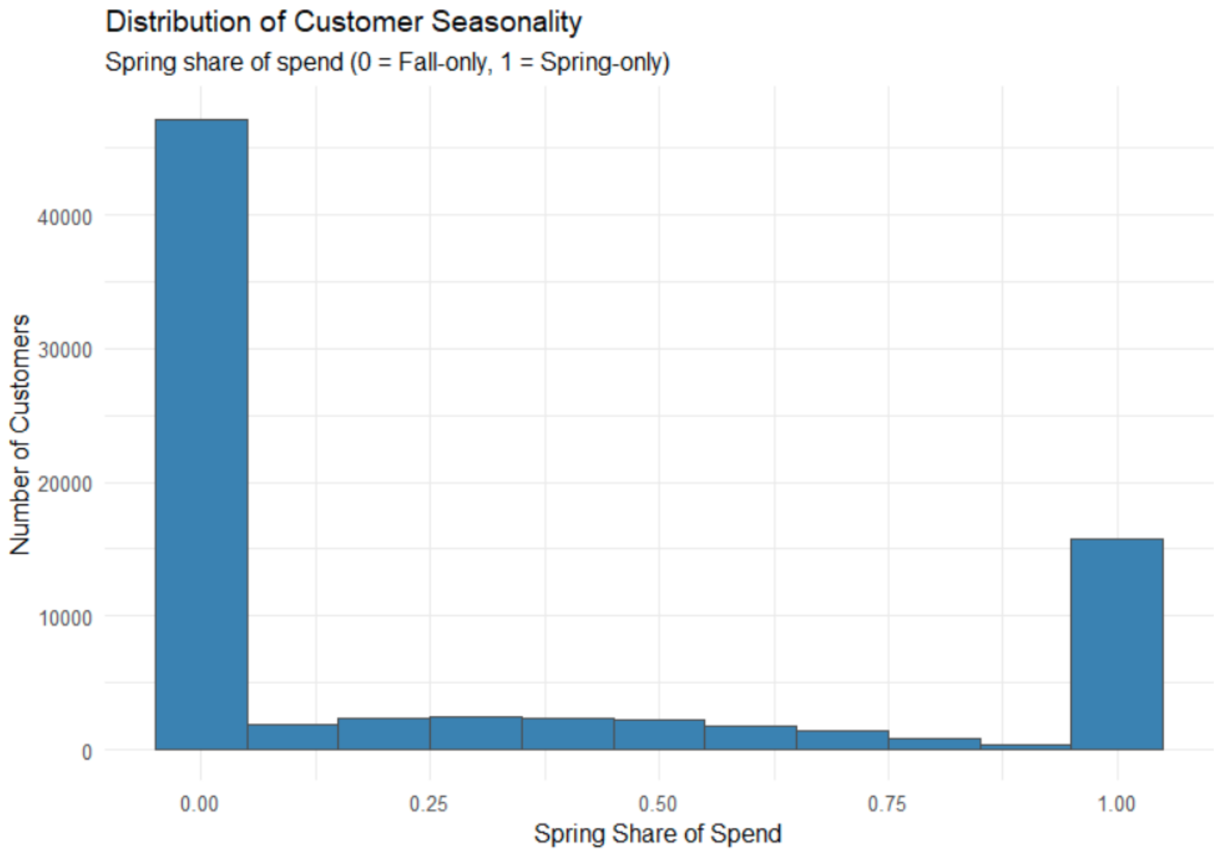

Figure 4.4 Distribution of Customer Seasonality Based on Spring Share of Spend

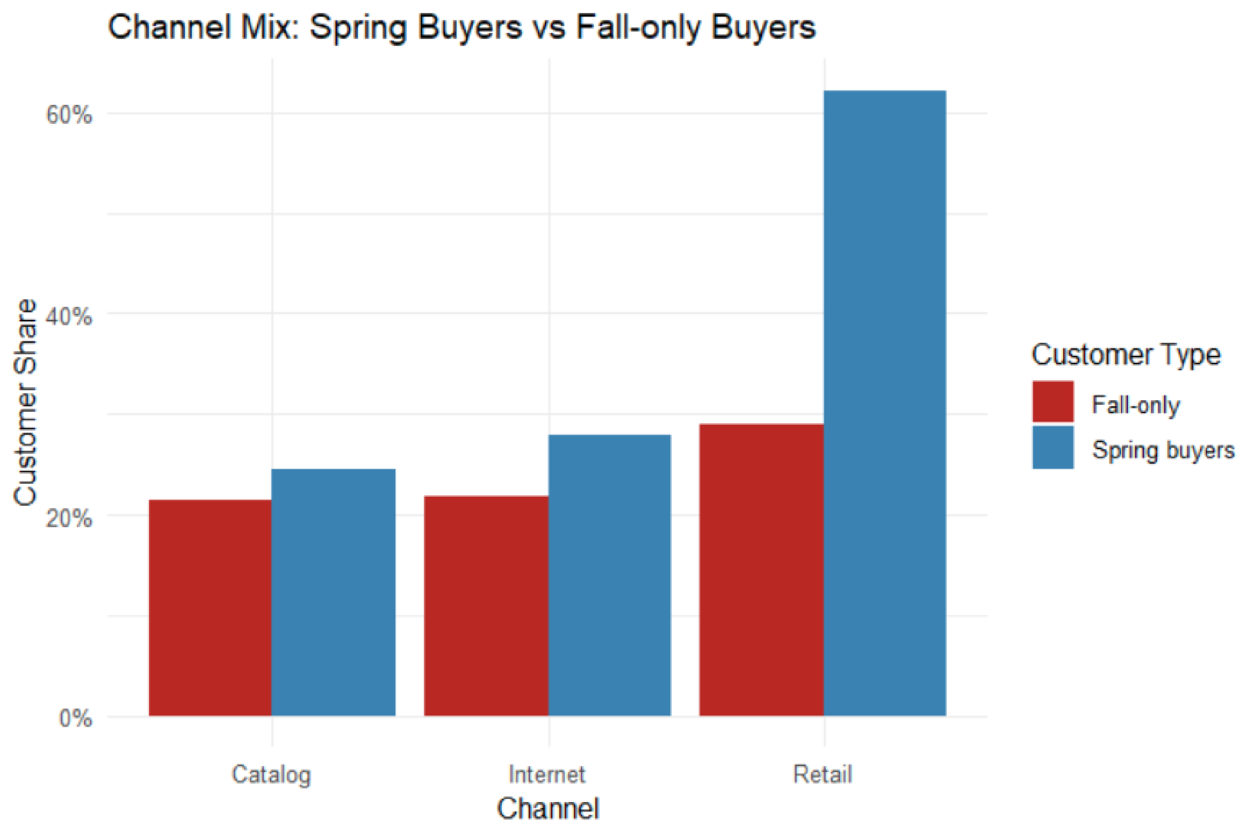

Figure 4.5 Channel Usage Comparison Between Spring Buyers and Fall-only Buyers

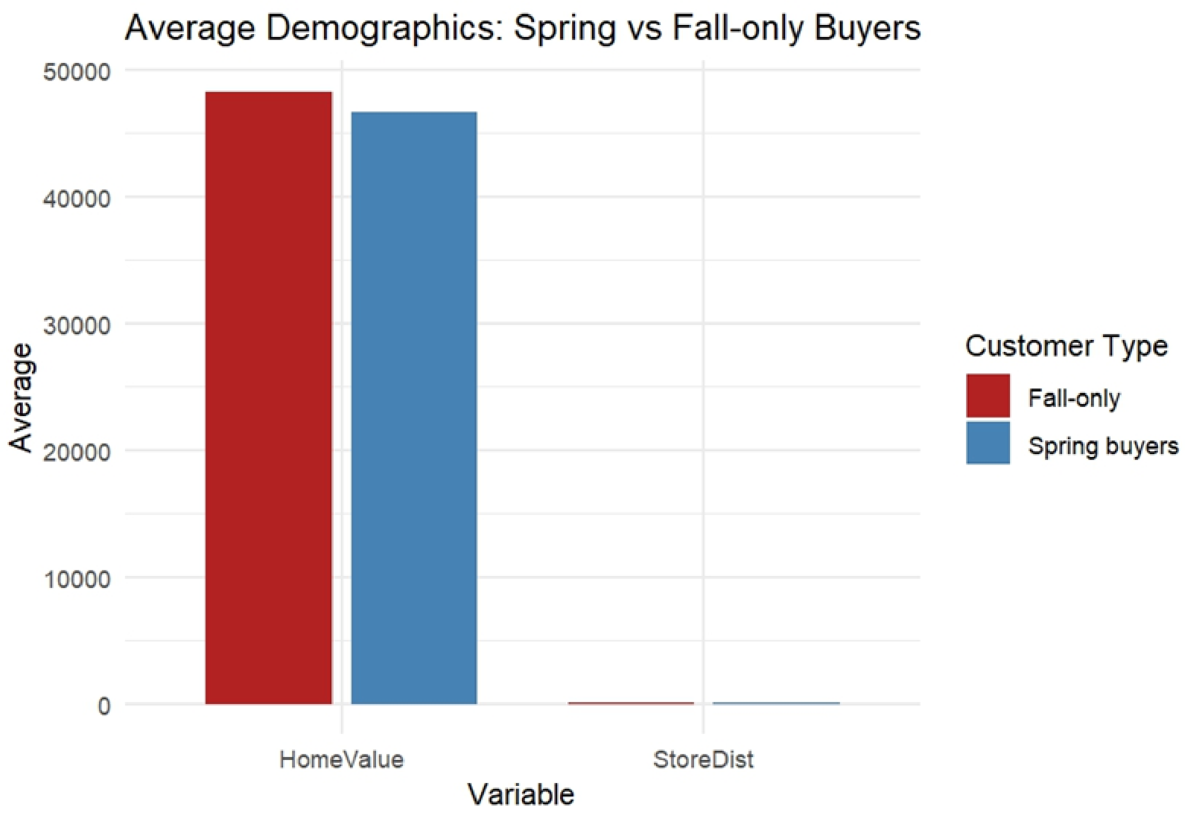

Figure 4.6 Average Demographics

```
> print(seg_kpi)
# A tibble: 5 × 7
  Segment        customers  share avg_seasonality retail_share internet_share catalog_share
  <fct>              <int>  <dbl>           <dbl>        <dbl>          <dbl>         <dbl>
1 Fall-only          46847 0.468            0            0.423          0.319         0.314
2 Mixed (Fall-…      10419 0.104            0.274        0.620          0.313         0.359
3 Mixed (Sprin…       5540 0.0554           0.657        0.671          0.266         0.273
4 Spring-only        15780 0.158            1            0.605          0.260         0.159
5 No Spend           21465 0.215          NaN            0              0             0
```

Figure 4.7 Key Metrics by Seasonal Customer Segment