University of South Australia

Supermarket Customers Analysis:

Behavior Analysis, Sales Forecast and Price Promotion Strategy Evaluation

Wangjun Shen

College of Science, Engineering,and Technology

The University of South Australia

Adelaide South Australia Australia

shewy009@mymail.unisa.edu.au

University of South Australia

Content Table

Introduction

This retail industry report analyzes three years of transaction data from a supermarket chain to reveal changes in customer purchasing habits, predict future sales trends, and evaluate the effectiveness of promotional activities.

The key objectives are to conduct a loyal customer analysis and understand their purchasing patterns, informing customized marketing strategies. Additionally, the report will forecast sales for the first three months of 2016 to aid inventory management and marketing strategy adjustments. Finally, the report will explore the relationship between price changes and sales volumes, providing data-driven support to optimize pricing and promotional strategies.

By leveraging these insights, the report aims to help supermarkets better understand customer needs, optimize their product supply and marketing, drive sales growth, and enhance customer satisfaction in the highly competitive market.

Missing Values Analysis

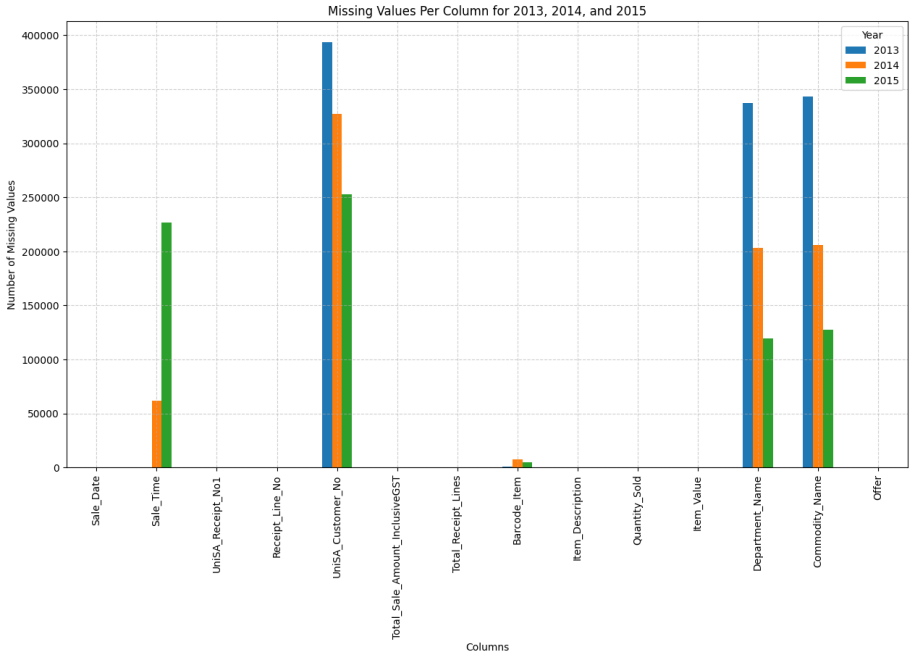

Figure 1: Missing Values for Different Variables in each Year

According to the distribution of missing values in supermarket data from 2013-2015 shown in Figure 1, a few key points emerge: UniSA_Customer_No, Department_Name and Commodity_Name have significant missing values, but this decreases each year,

likely due to improved employee skills or system stability. In contrast, Sale_Time has relatively few missing values, though 2014 saw a spike compared to 2013, possibly indicating temporary issues with recording transaction times that were resolved by 2015. Barcode_Item has negligible missing data, suggesting strong product coding management.

Overall, the trends indicate the supermarket is making progress in improving data quality and completeness, which should enable more robust analysis to support business decision-making going forward.

Feature	2013	2014	2015
Sale_Date	0.00%	0.00%	0.00%
Sale_Time	0.00%	0.49%	1.84%
UniSA_Receipt_No1	0.00%	0.00%	0.00%
Receipt_Line_No	0.00%	0.00%	0.00%
UniSA_Customer_No	3.14%	2.60%	2.06%
Total_Sale_Amount_InclusiveGST	0.00%	0.00%	0.00%
Total_Receipt_Lines	0.00%	0.00%	0.00%
Barcode_Item	0.01%	0.06%	0.04%
Item_Description	0.00%	0.00%	0.00%
Quantity_Sold	0.00%	0.00%	0.00%
Item_Value	0.00%	0.00%	0.00%
Department_Name	2.69%	1.62%	0.97%
Commodity_Name	2.74%	1.64%	1.04%
Offer	0.00%	0.00%	0.00%

Table 1: Missing Variable Percentage (%) for each Variable in each Year

Table 1 show the percentage of each variable in each year. The variables with the most missing values are UniSA_Customer_No, Department_Name, and Commodity_Name. Their missing rates decreased from 3.14%, 2.69%, and 2.74% in 2013 to 2.06%, 0.97%, and 1.04% in 2015 respectively. Since these are crucial identifiers, their missing data could affect further analysis. Conversely, Sale_Time and Barcode_Item have minimal missing values. To handle these missing values, rows containing missing data are deleted to maintain data integrity and avoid information distortion since the overall missing rate is not very high.

Customer Behavior Analysis and Segmentation Based on 2014 Data

Identification of Regular Customers in 2014

In the analysis, customers who had purchase records in 2013, 2014, and 2015 were identified as regular customers. The rationale behind this approach was to focus the analysis on loyal patrons, excluding occasional visitors. For these regular customers, the total number of visits, total purchases, and total spending were calculated. This data was then utilized to gain insights into the shopping habits of these customers, such as their purchase frequency and typical expenditure levels.

Before conducting the analysis, it is important to address a key consideration regarding the dataset. Customers without loyalty cards were assigned generic cards by staff at the register. As a result, some customers may appear to have exceptionally high purchase volumes, which does not reflect their true shopping behaviors. To ensure the accuracy of the analysis, these customers using generic cards will be identified as outliers and excluded, as their data could skew the insights on regular customer behaviors.

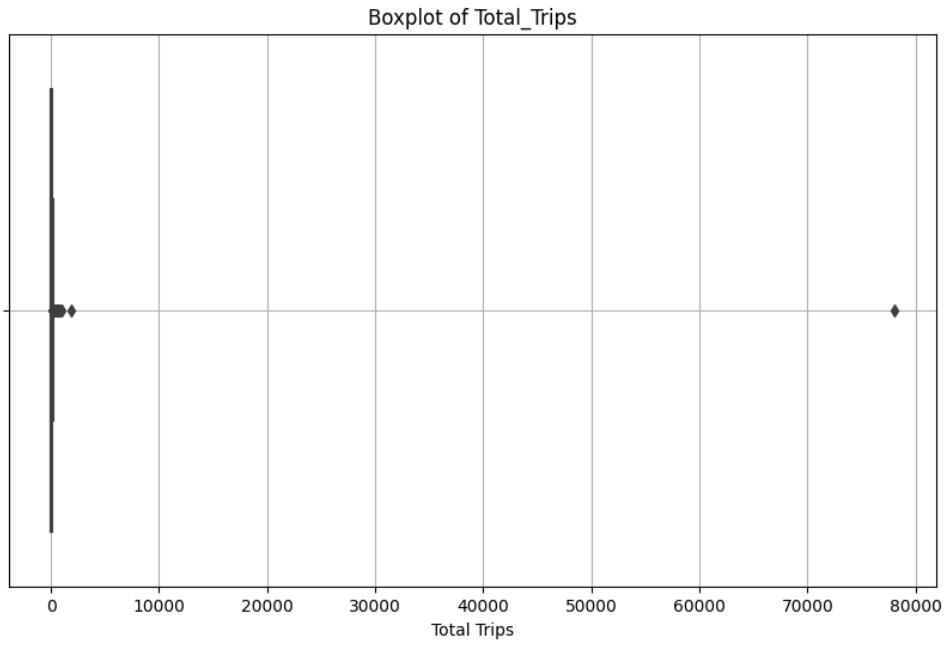

Figure 2: Distribution of Number of Trips with Potential Outliers

The visual results indicate that there are outliers in Number of Trips, and there is a particularly large outlier. The size of this outlier is around 7900, so that other outliers

and IQR boxes are almost squeezed into a line in the visualization results.

Given the presence of a particularly large outlier in the "Number of Trips" variable, the decision was made to use the interquartile range (IQR) method for outlier identification and removal, rather than the 3-sigma approach. The IQR method is more robust to extreme outliers, as it relies on the median and percentiles rather than being influenced by the skewed mean and standard deviation that would occur in this case.

Exploratory Data Analysis of Regular in 2014

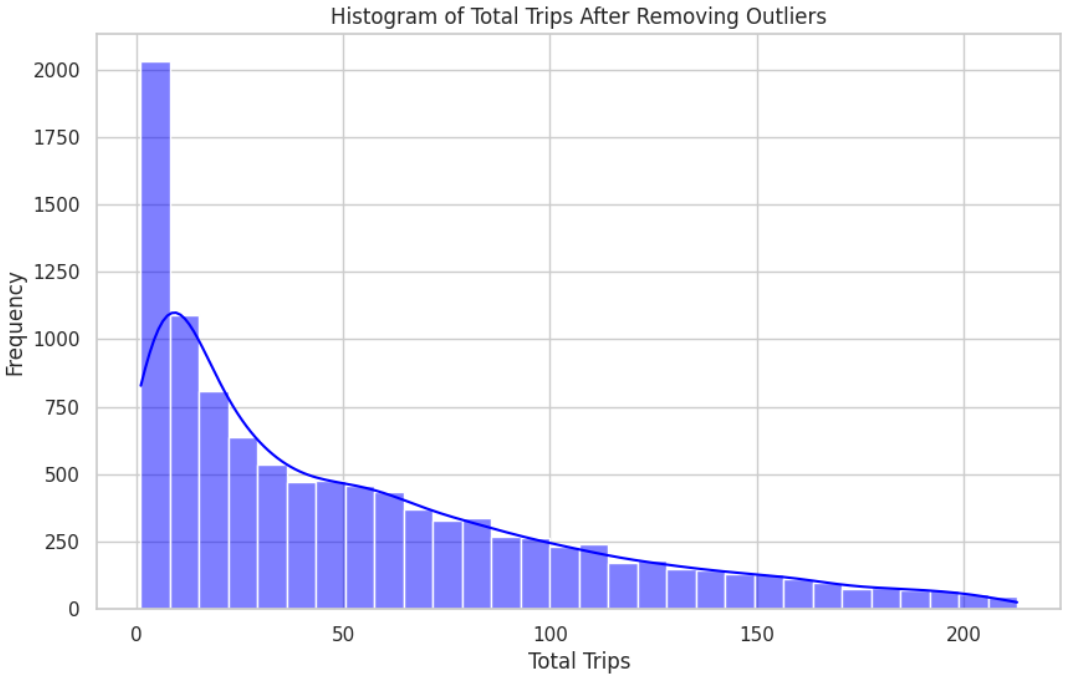

Figure 3: Distribution of Number of Trips without Outliers

Figure 3 shows the histogram of customer visit frequency, with the identified outliers removed, reveals a skewed distribution with a concentration of customers making a moderate number of trips. There is a clear gradual decrease in the frequency as the number of trips increases, a pattern typical in retail settings where a larger group of customers shop regularly but not excessively, while a smaller subset exhibits higher visit frequency. From a business perspective, this suggests that the supermarket's consistent customer base who shop regularly could be the primary targets for loyalty programs and targeted marketing, while the higher-frequency customers in the distribution tail may represent an opportunity for special promotions or bulk buying incentives.

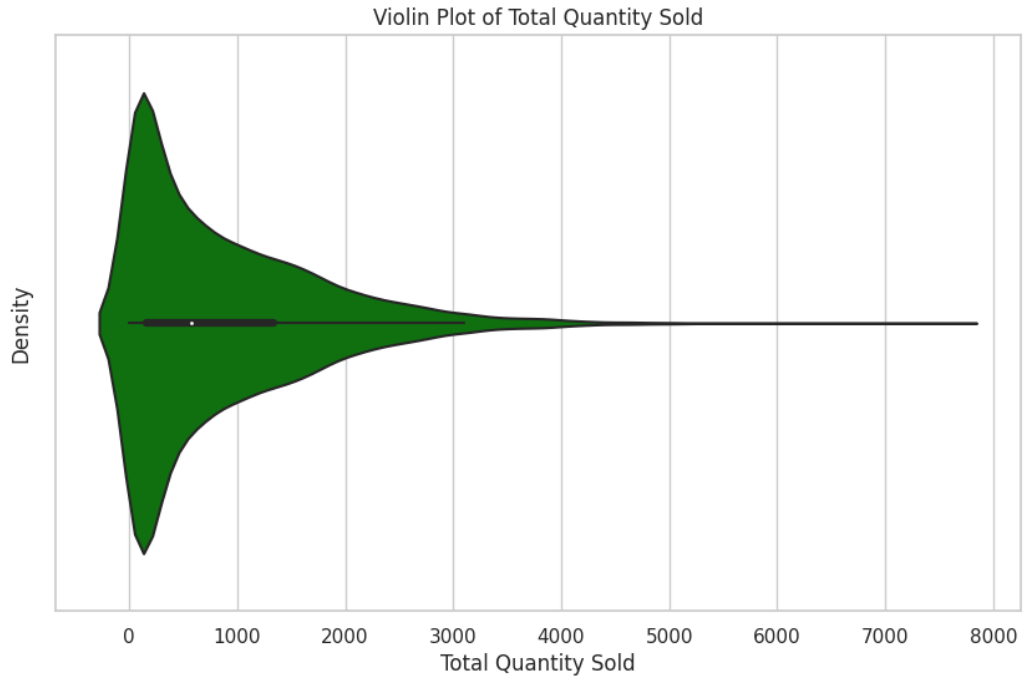

Figure 4: Distribution of Total Quantity Purchased Without Outliers

The violin plot of figure 4 shows the distribution of the total number of items purchased by customers after removing outliers. As can be seen from the figure, the main part of the distribution (the wider area in the middle of the violin plot) is very dense, indicating that most customers make relatively small purchases. The graph extends to the right into a long tail, indicating that as the number of purchases increases, the number of customers purchasing these quantities decreases.

This distribution is common in the retail industry, indicating that there is a large group of customers who frequently make medium-sized transactions, while a smaller group of customers tend to buy in bulk. From a commercial perspective, this insight suggests opportunities for supermarkets to cater to different shopping needs, such as offering volume discounts or targeted promotions to encourage customers in the long tail of the distribution to make larger purchases.

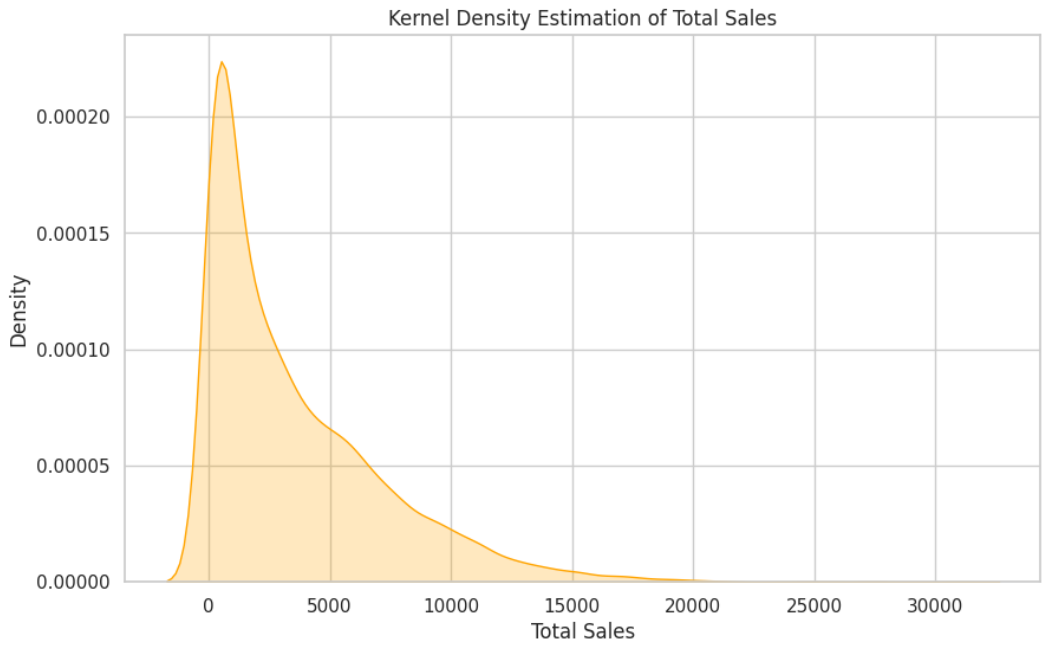

Figure 5: Kernel Density Estimate of Total Money Spent by Customers

The kernel density plot shows the right-skewed distribution of total sales. The density peaks in the lower sales range, indicating most transactions are relatively small. However, the long right tail suggests high-value transactions, though fewer in number, still account for a significant portion of total sales.

This distribution likely reflects a supermarket sales model with two components - frequent small-value transactions and less common but high-value purchases. To optimize sales, the business may need differentiated strategies. For the high-volume, low-value transactions, increasing promotions and customer frequency could boost sales. Maintaining loyalty of high-value customers may require excellent service and personalized experiences.

	mean	std	min	25%	50%	75%	max
BAKERY	6.21%	5.78%	0.00%	3.01%	5.13%	7.76%	100.00%
COOP GIFT CARDS	0.01%	0.39%	0.00%	0.00%	0.00%	0.00%	32.06%
COOP XMAS CLUB	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	2.33%
Corporate Merch	0.00%	0.02%	0.00%	0.00%	0.00%	0.00%	0.95%
DAIRY	11.01%	6.62%	0.00%	7.26%	10.52%	13.86%	100.00%
DELI	6.03%	6.48%	0.00%	2.21%	4.85%	7.98%	100.00%
Dept:0	0.00%	0.01%	0.00%	0.00%	0.00%	0.00%	0.51%
EPAY	0.20%	1.33%	0.00%	0.00%	0.00%	0.00%	38.58%
EXPENSE	0.02%	0.68%	0.00%	0.00%	0.00%	0.00%	69.21%

FRESH MEAT	9.68%	8.17%	0.00%	4.29%	8.55%	13.28%	100.00%
FROZEN	4.62%	5.24%	0.00%	1.53%	3.53%	6.14%	100.00%
FRUIT & VEG	15.36%	9.05%	0.00%	9.48%	14.45%	20.09%	100.00%
GROCERY	38.28%	13.16%	11.24%	31.06%	37.94%	44.92%	100.00%
SEAFOOD AND POULTRY	2.64%	4.23%	0.00%	0.00%	1.38%	3.58%	79.51%
TOBACCO	2.81%	9.69%	0.00%	0.00%	0.00%	0.04%	100.00%
VARIETY	3.12%	3.92%	30.68%	1.24%	2.36%	3.89%	100.00%

Table 3: Customer Purchasing Behavior Metrics

The Grocery department, with the highest average proportion (38.28%) and a standard deviation of 13.16%, is the main contributor to customer consumption, though varying among customers. The Fruit & Veg department also has a high average proportion (15.36%) with a high standard deviation (9.05%), indicating significant customer purchase variations. Dairy is also crucial in customer baskets with an average proportion of 11.01%. Bakery and Deli departments have similar average proportions and standard deviations, showing uniform purchasing behavior. Notably, COOP GIFT CARDS, COOP XMAS CLUB, Corporate Merch, and Dept:0 departments have negligible average values, indicating their minor contribution to total sales. Despite low averages, EXPENSE and EPAY have high maximum values, suggesting special transactions or activities. Data may include negative minimum values due to errors or returns.

From a business perspective, these statistical results provide important insights into customer purchasing behavior. For example, since the Grocery and Fruit & Veg departments have relatively large proportions, the supermarket may need to ensure sufficient inventory in these departments and consider promotions to increase sales. Meanwhile, for the departments with extremely small proportions, the supermarket management may need to evaluate whether to change the marketing strategy or adjust the product mix to better meet market demand.

Clustering of Regular Customers in 2014

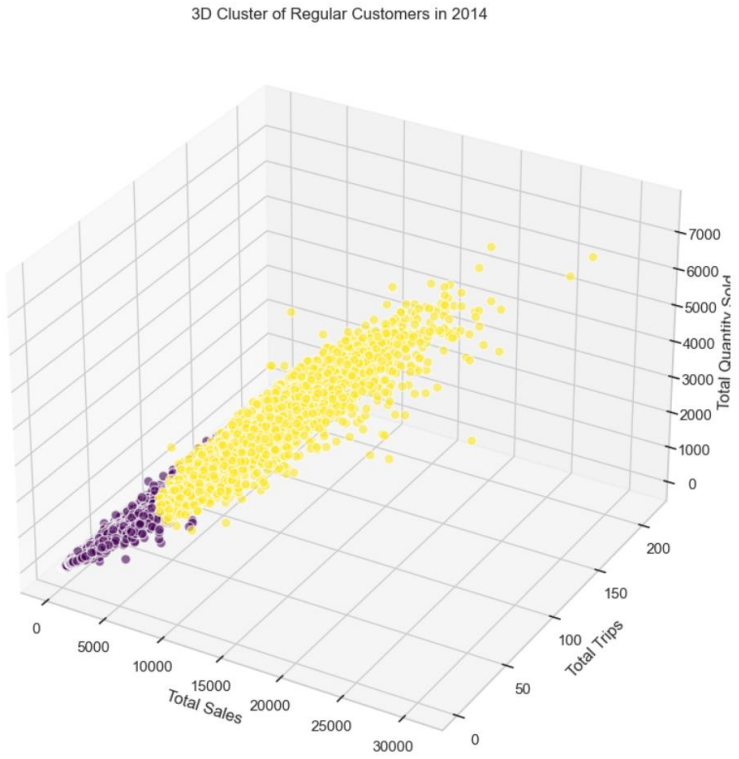

Figure 6: 3D Visualization of Regular Customers in 2014

According to the 2014 supermarket regular customer data clustering results shown in Figure 6, different colors represent different customer groups.

The purple cluster represents low-value regular customers whose total sales, purchase times, and purchase quantities at the supermarket in 2014 were all low. They may be occasional customers that year or only purchase low-value items. This group has great potential, and supermarkets can work hard to increase their purchase frequency through promotional activities and other methods.

The yellow cluster represents high-value regular customers who had high purchase frequency, large single purchase volume, and high total sales at the supermarket in 2014. They may be loyal customers to the supermarket brand, or repeat customers attracted by factors such as product convenience and quality services. For this key group of regular customers, supermarkets should provide an excellent shopping experience and maintain a high level of customer satisfaction, thereby maintaining a stable source of revenue.

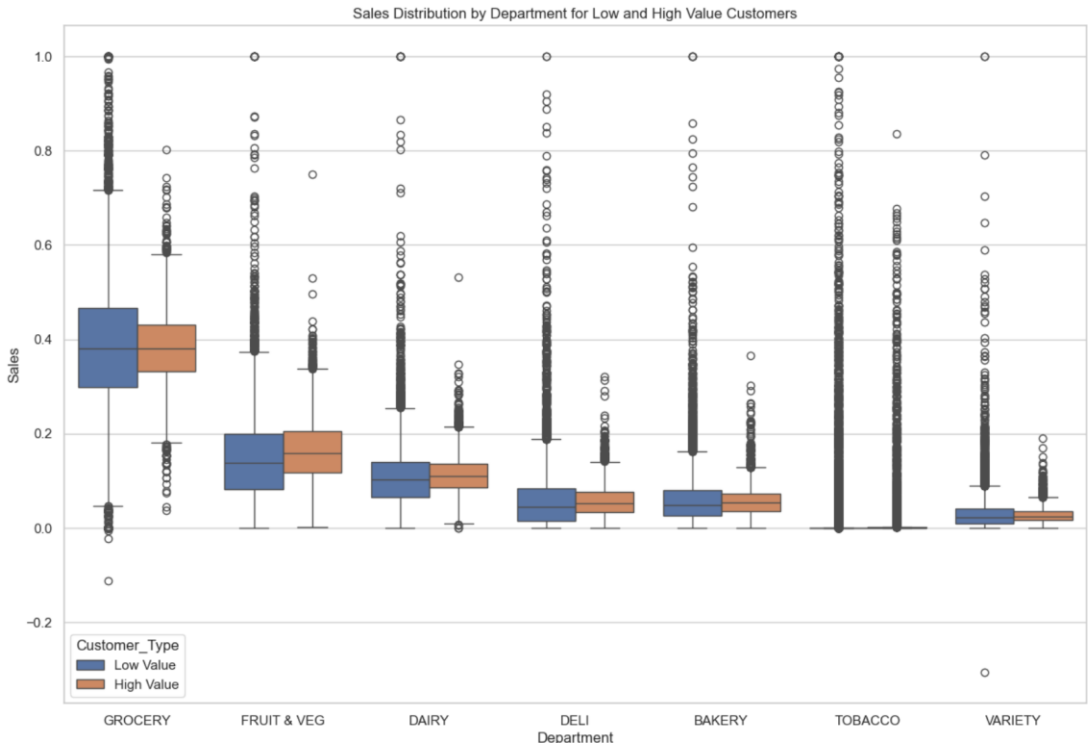

Figure 7: Main Difference of Department Proportion Distribution for Low and High Value Customers

It can be clearly seen from the Figure 7 that there is a significant difference in the sales proportion of low-value regular customer groups and high-value regular customer groups in each commodity department.

Low-value regular customer groups may represent ordinary families or individuals with average income levels, and their spending is mainly concentrated on daily commodities. Their consumption frequency in the grocery (GROCERY) department is higher, and the median is slightly higher than that of high-value regular customers, indicating that they are more dependent on basic commodities. In contrast, high-value customer segments are likely to include middle- and upper-income households and individuals who are higher in the median and upper quartiles in the fresh produce (FRUIT & VEG) sector, indicating that they spend more in this category diversification. At the same time, their spending in the dairy (DAIRY) sector is broader, reflecting greater choice.

However, low-value regular customers have lower median spend in departments such as DELI and BAKERY, possibly choosing more affordable options for financial reasons, as opposed to their purchases being influenced by price and promotions Behavior matches. Additionally, low-value regular customers spend the least in the tobacco (TOBACCO) segment, possibly due to lower preference for this category of goods or due to budget constraints.

Finally, the two regular customer groups have similar median spend in the grocery

(VARIETY) department, but low-value regular customers are more concentrated, which may mean their purchasing patterns are more consistent and predictable, while high-value regular customers have more spread-out spending and scattered.

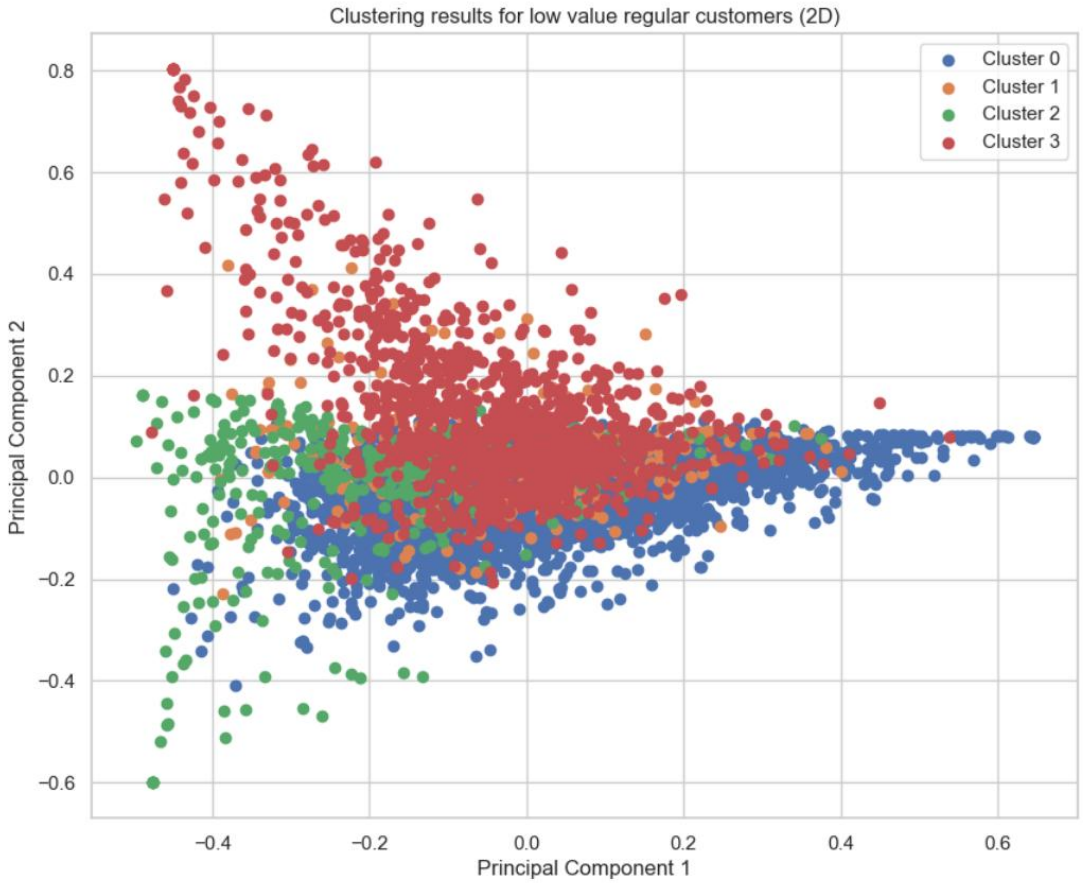

Figure 8: Visualization of PCA Components after Clustering for Low Value Regular Customers

For the low value group among regular customers, another clustering was performed. The clustering results are shown in the figure 8. They can be divided into 4 clusters. However, because they contain too many variables, the visualization in 2 dimensions is difficult. The results did not demonstrate the distinction between the four.

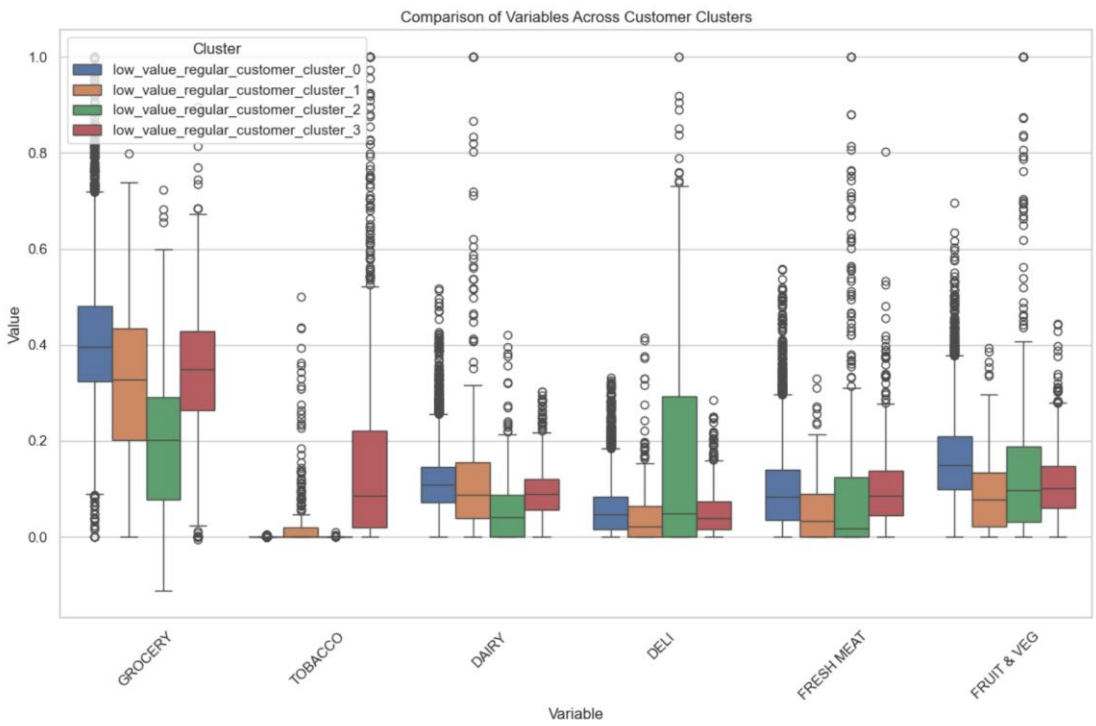

Figure 9: Boxplot of Main Different Department Proportion Distribution for 4 Clusters

For tobacco products, we note that cluster_3 has significantly different consumption patterns. Although the median consumption ratio of this group is 0.1, the upper quartile reaches 0.21, and there are many outliers above 0.5. This shows that in cluster_3, more low-value customers tend to purchase other small-value commodities when buying tobacco. For tobacco retailers, this is a potential value-added opportunity, through cross-category promotions or packaged sales, which can effectively increase the consumption of this group of customers.

Within the grocery category, cluster_0 stands out with the highest median consumption ratio, indicating that its consumers may focus primarily on daily necessities. At the same time, this group has a large number of data points with abnormally high consumption ratios, which means for retailers that they can attract this group of customers by promoting high-volume or wholesale-priced products. In contrast, cluster_2 has the lowest median, even lower than the lower quartile of cluster_1, implying that this group of customers has lower willingness to spend on groceries and may focus more on other consumption areas or have a limited budget.

In the analysis of cooked food products, cluster_2 shows extremely unique high demand. Its median consumption ratio and upper quartile are higher than other groups, and there are many high-value outliers. This data characteristic implies that customers in this group may choose fast and convenient cooked food products more frequently because of their busy work or accelerated pace of life. Targeting this market,

retailers can consider increasing the variety and quality of ready-to-eat meals and providing more healthy options to meet the needs of these customers.

In the consumption of fresh meat products, although cluster_0 and cluster_2 show more abnormally high values in the range of 0.3 to 1, the abnormal values of cluster_0 are mainly concentrated between 0.3 and 0.55, while those of cluster_2 are between 0.3 and 0.8. between. This suggests there is a demand for high quality, or large quantities of fresh meat purchased within both groups, possibly due to larger household size or a preference for fresh food.

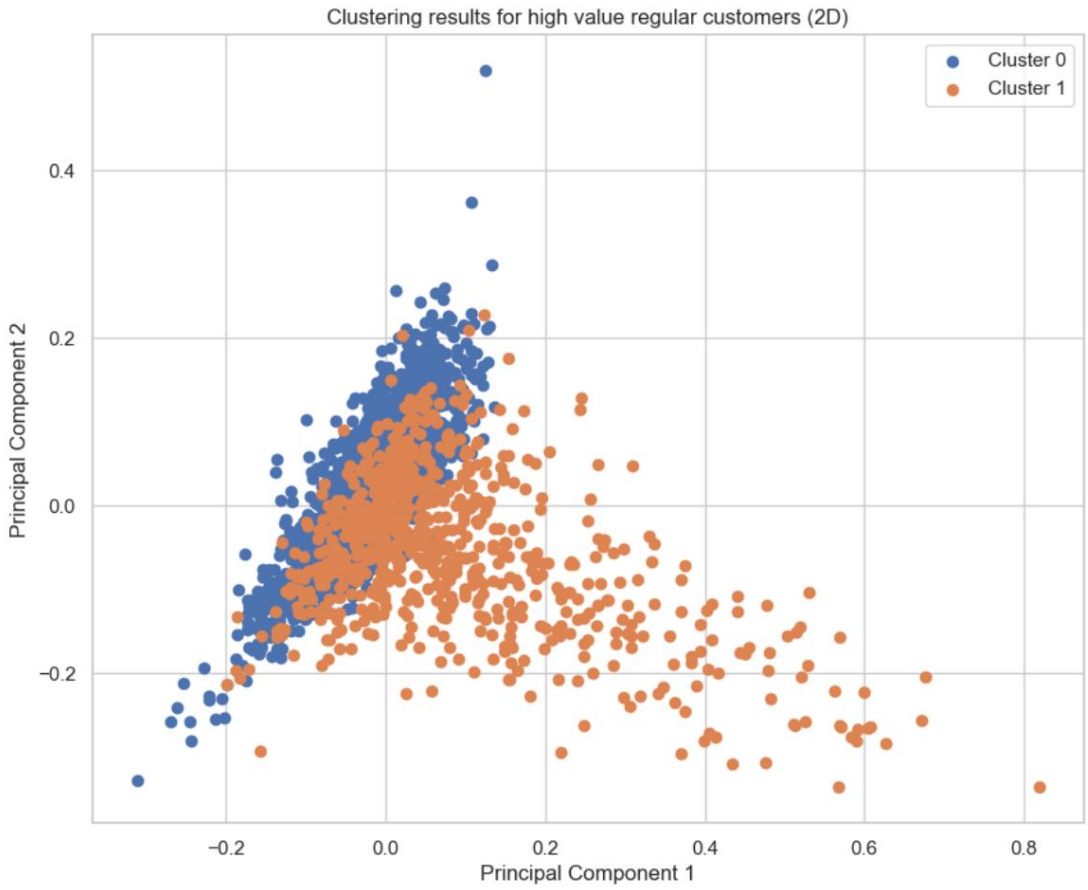

Figure 10: Visualization of PCA Components after Clustering for High Value Regular Customers

For the high-value segment within the regular customer base, an additional clustering analysis was conducted, the outcomes of which are depicted in figure 10. This analysis categorized the segment into 2 distinct clusters. However, the complexity of incorporating numerous variables made it challenging to effectively visualize these results in a two-dimensional space. Consequently, the visual representation did not clearly delineate the differences among the four clusters.

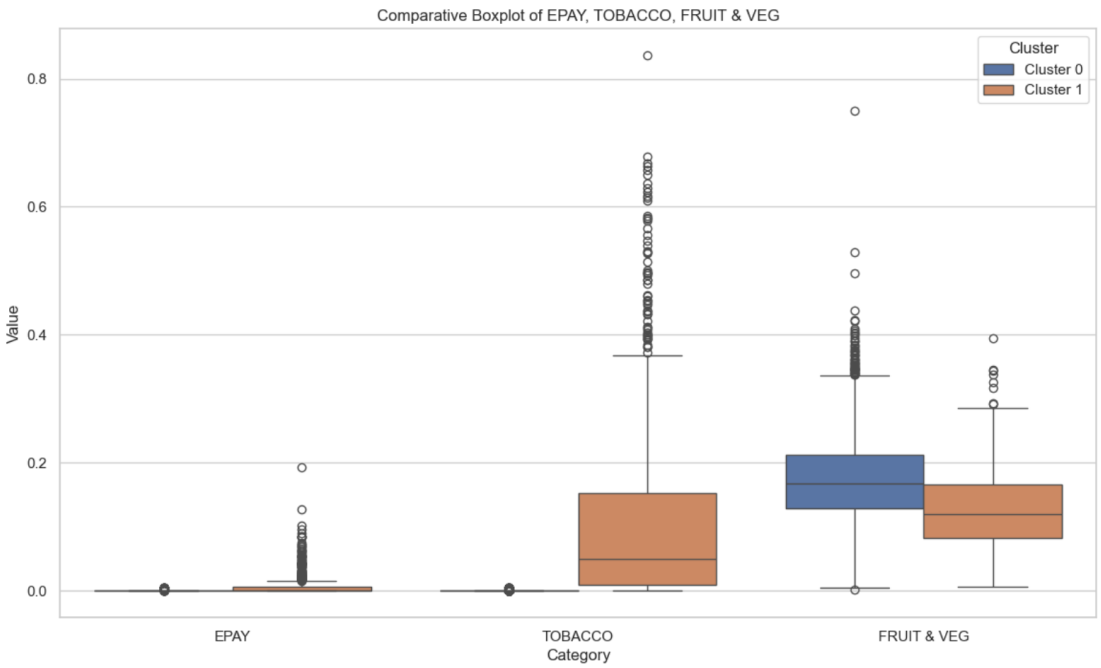

Figure 11: Boxplot of Main Different Department Proportion Distribution for 2 Clusters

Based on the provided box plot visualization results in Figure 11, we can make a detailed interpretation of the consumption patterns of customers in the two clusters. In the consumption category EPAY, we observe that the medians of Cluster 0 and Cluster 1 are almost close to 0, which indicates that EPAY consumption is not the main expenditure of customers in either Cluster 0 or Cluster 1. However, Cluster 0 shows some high outliers, which may mean that although most customers rarely use EPAY, there is still a small group of customers who spend relatively high amounts on this category.

In the TOBACCO consumption category, the difference between the two clusters is more significant. The median TOBACCO consumption of Cluster 1 is significantly higher than that of Cluster 0, which indicates that customers in Cluster 1 generally spend more on tobacco products. In addition, Cluster 1's box plot on TOBACCO consumption also shows wider boxes, which reflects that its consumers' consumption in this category is more volatile and there are more customers spending higher. Although the overall consumption level of Cluster 0 is low, the number of outliers indicates that some customers consume far more tobacco than most customers in the same group.

As for FRUIT & VEG, the consumption category of fruits and vegetables, the medians of the two clusters are similar, indicating that the median consumption amount of the two groups of customers in this category is close. However, the box plot of Cluster 1 is wider, indicating that the consumption distribution range of customers in this cluster is wider and the consumption amount fluctuates greatly. Cluster 0 shows more outliers,

indicating that some customers in this cluster spend far more than the normal level in the FRUIT & VEG categories.

From these observations, we can conclude that although the medians of consumers in both clusters are comparable on the FRUIT & VEG categories, Cluster 1 has a wider distribution of consumption. In the TOBACCO category, customers in Cluster 1 show higher consumption levels and greater fluctuations in consumption amounts. Although the EPAY category is not the main consumption point in both groups, there are several customers in Cluster 0 whose consumption is significantly higher than the group average.

Sales Forecast for the First Quarter of 2016 Based on ARIMA Model

Forecast Total Sales for January 2015

During the data preparation phase, missing values and rows in the data set were deleted. However, due to systemic issues such as data export and system upgrades, there may be deficiencies in the data set itself.

Month	Days in 2013	Days in 2014	Days in 2015
Jan	30	30	17
Feb	28	28	28
Mar	30	31	31
Apr	30	29	29
May	31	31	31
Jun	30	30	30
Jul	31	31	31
Aug	31	31	31
Sep	30	30	30
Oct	31	31	31
Nov	30	30	30
Dec	30	30	30

Table 4: Monthly Days in 2013, 2014, and 2015

The table's data shows consistent recording across 2013-2015, except for January 2015 with 17 days recorded, suggesting possible data loss or an event like a system upgrade

or natural disaster. Simply copying data to fill missing days isn't accurate or scientific. Instead, using the relationship between total sales and specific dates in January 2013 and 2014, alongside the 17 days of data in 2015, could provide a more accurate estimate of January 2015 sales.

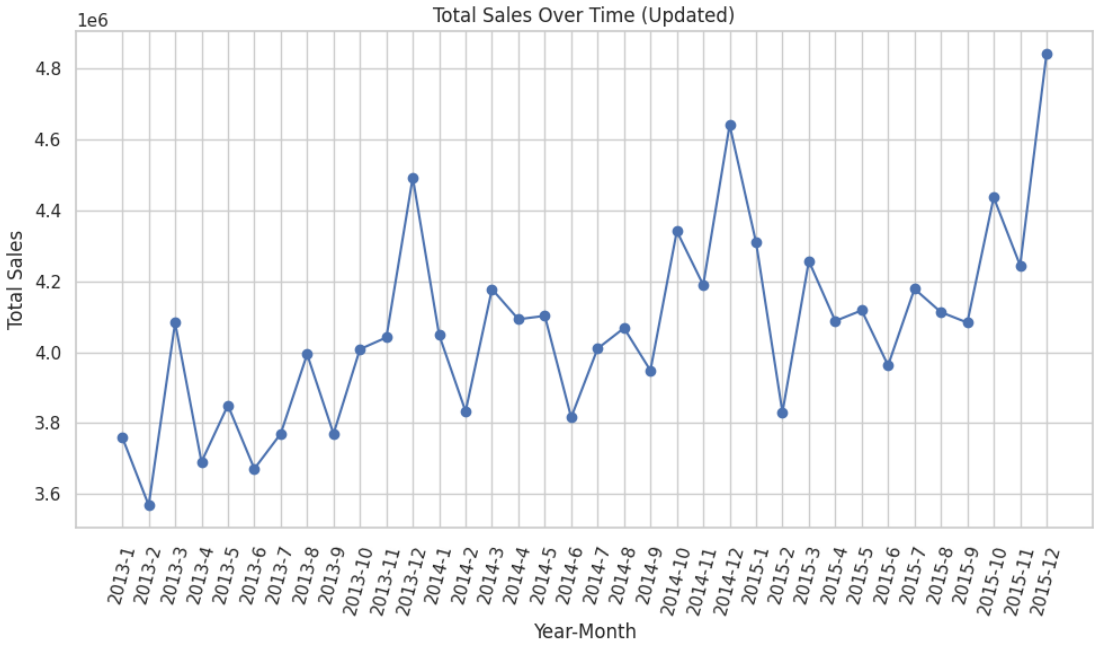

Figure 12: Total Monthly Sales each Month over Years

As can be seen from Figure 12, supermarket sales performance from 2013 to 2015 reflects two key aspects: seasonal trends and overall growth.

December, the holiday season, marks peak sales, likely due to increased consumption during Christmas and New Year. Higher sales are also noted in summer and holidays, reflecting Australian consumption habits. Middle-year months like July and August sometimes see sales declines, possibly due to seasonal changes and lack of promotions.

Beyond seasonal trends, there's an overall sales increase from 2013-2015, particularly in 2015, especially year-end. This could suggest market share expansion or sales efficiency improvement, signaling positive prospects.

Overall, the supermarket's sales show seasonal trends but are generally growing. By maintaining year-end peak sales and managing mid-year fluctuations, the supermarket can continue to improve and prepare for future growth.

Sales Forecast for the First Quarter of 2016

When dealing with data exhibiting significant temporal dependence, such as supermarket sales data, time series models serve as a powerful alternative to traditional

regression techniques. In the retail sector, ARIMA models have proven particularly valuable, especially in the context of supermarket operations. Previous visual analysis has demonstrated the suitability of time series analysis to help further understand seasonal patterns, trends, and periodicity in sales data.

Using the ARIMA (1, 0, 2) model is a reasonable choice. Let me further explain what this model parameter means.

First, the autoregressive (AR) term is set to 1. This means that each month's sales forecast relies on the previous month's sales data. This helps capture trends in sales data from one month to the next, especially when holidays or seasonal factors cause changes in consumption patterns. Second, the difference (I) term is set to 0. This means we don't need to differentiate the time series. This prevents data stability from being affected by outliers or noise, thereby improving the reliability of the model. Finally, the moving average (MA) term is set to 2. This can help the model adjust for random fluctuations in past forecasts, helping to better adapt to unusually high or low sales in certain months.

In general, the ARIMA (1, 0, 2) model can better capture the seasonal characteristics and overall growth trend of the supermarket sales data. This model setup is also better able to handle the impact of outliers and noise on predictions.

Date	Predicted Sales (Millions)	Total Error Margin (Millions)
2016-01	4.367	±0.297
2016-02	4.423	±0.297
2016-03	4.364	±0.297

Table 5: Sales Forecast and Error Range

Table 5 shows the forecast results of total monthly sales for the three months of the first quarter of 2016 based on the data of the past three years and using the time series model and shows the error margin.

For the first three months of 2016, the time series forecast provides detailed sales forecasts and their error margins. The forecast sales total for January is A$4.367 million, with a margin of error of ±A$0.297 million. As January is Australia's summer holiday season, this is typically a peak season for sales, particularly due to increased spending by families and tourists. As a result, sales could vary between A$4.070 million and A$4.664 million. Heading into February, despite seasonal activity that may have slowed sales, total forecast sales are still at A$4.423 million, with the margin of error

unchanged. That suggests sales remain strong, likely helped by school reopening and holiday-specific promotions. By March, the total sales forecast was A$4.364 million, with an error margin of ±A$0.297 million. Sales may slow a bit as summer winds down, but forecasts still point to high sales potential.

The forecast sales from January to March 2016 are higher than the actual data for the same period in the past three years, which is consistent with the overall trend of the data, indicating that the accuracy of the forecast is relatively high.

In practical applications, supermarkets should take this error range into account and prepare corresponding strategies to deal with possible sales fluctuations. For example, if actual sales are close to the lower limit of the forecast, the supermarket may need to increase promotional efforts; if sales are close to the upper limit, it may need to ensure the efficiency of the supply chain to avoid out-of-stock situations. In addition, supermarkets should monitor how actual sales figures compare to forecasts to adjust their strategies for subsequent months. Integration and analysis in this way can help supermarkets better understand market dynamics and make effective business decisions.

Analyzing the Impact of Price Changes on Sales Volume by Using 2013 Data

Analysis of the Relationship between Selling Price and Sales Volume of BANANAS in 2013

BANANAS has the highest value in both sales volume and total sales.

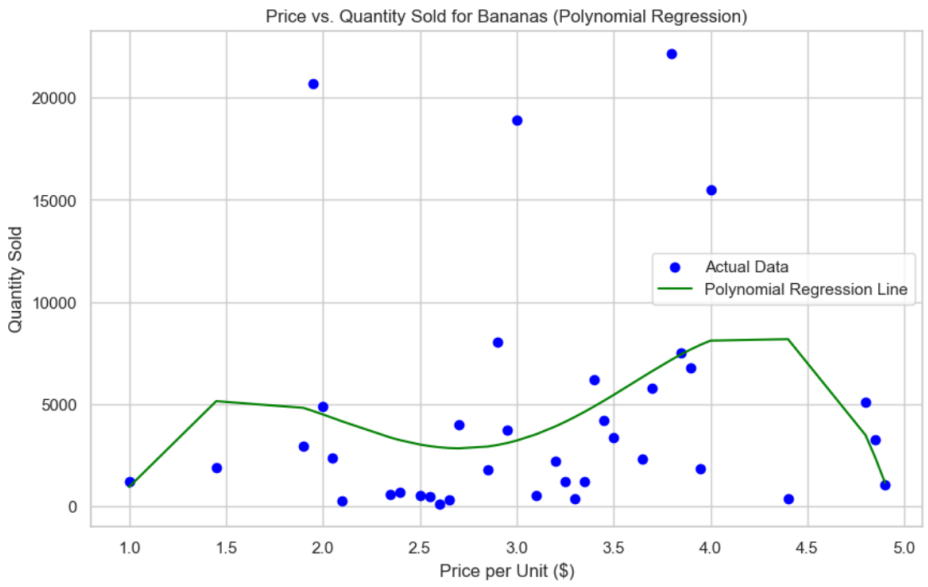

Figure 13: Price vs. Quantity Sold for Banans in 2013 with Polynomial Regression

According to the analysis of the fourth-order polynomial regression model in Figure 13, the relationship between banana prices and sales shows different dynamics in different price ranges. For different price ranges, supermarkets can adopt corresponding business strategies to optimize profits and sales.

In the low-price range ($1.00 to $2.00), sales increased by about 1.64% on average. This reflects the high elasticity of consumer demand for low-priced bananas. Supermarkets can consider slightly raising prices to increase revenue while maintaining sales. In addition, low-priced bananas can be used as store attractions to guide customers into the store and potentially increase sales of other products.

In the mid-price range ($2.00 to $3.00), sales volume decreased by about 0.61% on average. Price sensitivity is beginning to show, with price increases leading to a slight decrease in sales. Supermarkets need to be more cautious when setting banana prices to avoid significant price increases. The pricing strategy in this range should pay more attention to the balance between profit and sales and maximize total revenue through subtle price adjustments.

In the high price range ($3.00 to $4.90), sales volume decreased by about 3.86% on average. Each increase in price brings a more pronounced decrease in sales. Supermarkets should promote high-quality or specialty bananas to such consumers and maintain a stable price strategy to maintain customer loyalty and avoid a sharp decline in sales.

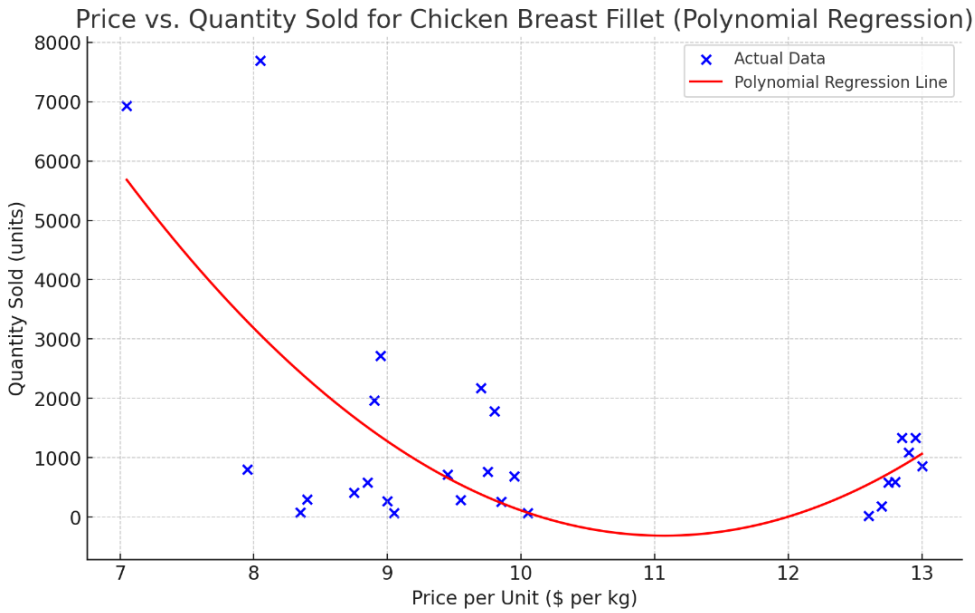

Figure 14: Price vs. Quantity Sold for Chicken Breast Fillet with Polynomial Regression

In the price range of 8 to 10, the sales volume of chicken breast is relatively high and fluctuates greatly. Especially when the price is close to the low of 8 yuan, the sales volume reaches its peak. As prices rise, sales volume gradually declines, but overall remains at a relatively stable level. For this price range, supermarkets can adjust prices to attract more consumers, such as moderate price reductions to increase sales during periods of greater demand, or appropriate price increases to maintain profit margins when demand is low. Additionally, running promotions when prices are lower may further increase sales.

In the price range of 12 to 13, the sales volume of chicken breast is generally low, and the higher the price, the lower the sales volume. This shows that in the higher price range, consumers' willingness to buy significantly decreases. Therefore, supermarkets may need to carry out specific market positioning for products in this price range, such as emphasizing the quality, health benefits or other unique selling points of chicken breasts to attract consumer groups with higher quality requirements. Considering the low sales volume at high price points, supermarkets should be cautious when setting higher prices to avoid losing potential consumers due to excessive prices.

There are some differences in model predictions for the change in sales volume caused by a 1% price increase in the two price ranges. In the range of 8 to 10 , sales volume decreased by about 15.28% on average, which shows that within this range, a small increase in price has a significant negative impact on sales, and consumers are more sensitive to price. In the range of 12 to 13, sales increased by about 76.36% on average. This result may reflect the model's inaccurate prediction of sales changes in the high

price range.

According to the above analysis results, for the price range of 8 to 10, supermarkets should consider price sensitivity and avoid easily raising prices. Instead, use promotions and discount strategies to maintain or increase sales. For the range of 12 to 13, although the model forecast shows huge sales growth, this may be an over-forecast. Therefore, supermarkets should cautiously raise prices within this price range and closely monitor market response and sales data to ensure the effectiveness of the strategy.

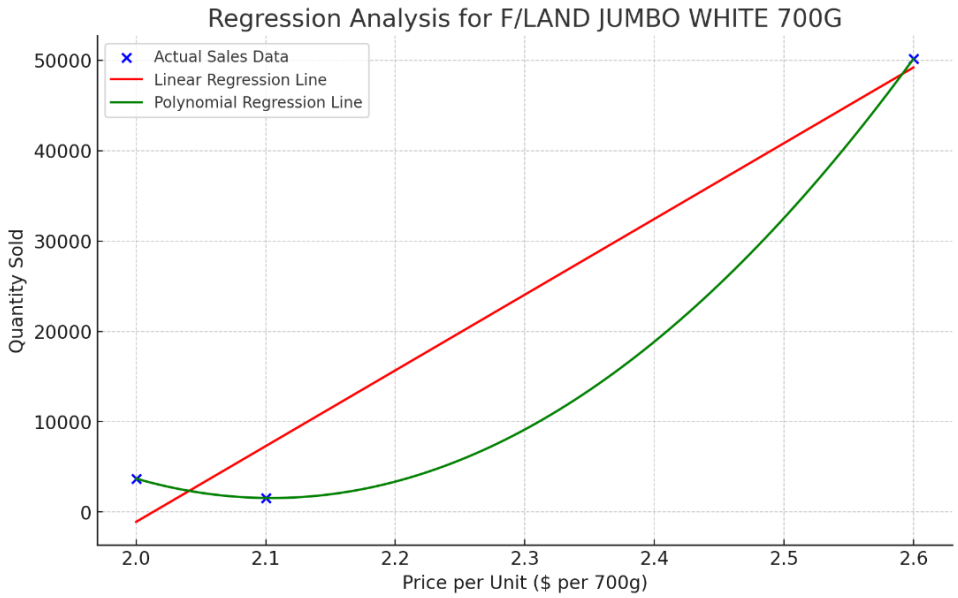

Figure 15: Price vs. Quantity Sold for F/Land Jumbo White 700G with Polynomial Regression and Linear Regression

Visual analysis in Figure 15 shows that there is a significant surge in sales when the price reaches $2.60. This may indicate that consumers are very sensitive to this price point, or there may be promotions that attract consumers. The polynomial regression model shows that for every 1% increase in price, sales volume decreases by approximately 2.12% on average. This suggests that overall, sales volume is relatively sensitive to price changes. Abnormal increases in sales at higher prices may be due to the effects of promotions, such as discounts or bundles, especially if these are widely publicized before consumers make a purchase.

Figure 16: Price vs. Quantity Sold for Chicken BBQ with Polynomial Regression

According to the prediction of the polynomial regression model, for the df_2012_chicken_BBQ product, when the price increases by 1%, sales volume is expected to increase by approximately 2.53% on average. This result shows that in this case, price increases appear to be associated with increases in sales. If sales are low at certain price points, this may be due to discounted sales of products that are about to expire, in which case the low sales may not reflect normal market demand. Discounting may temporarily increase sales, especially at the lowest price points.

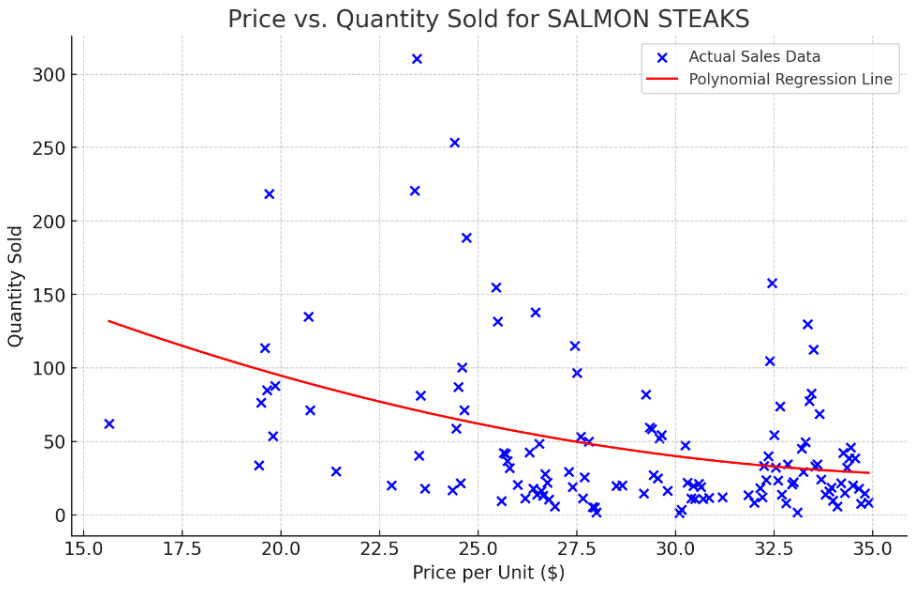

Figure 17: Price vs. Quantity Sold for SALMON STEAKS with Polynomial Regression

The scatter points in the chart show salmon fillet sales at different price points. The

polynomial regression line running through these points shows the overall trend of sales volume as a function of price. In general, as prices rise, sales of salmon steaks show a downward trend, but this decline is not linear, and changes in sales at certain price points may be affected by other market factors. According to polynomial regression analysis of the data, when the price increases by 1%, the sales volume of salmon steaks is expected to decrease by about 2.12% on average. This shows that there is a negative correlation between salmon steak sales and price, and an increase in price will lead to a decrease in sales, reflecting consumers' sensitivity to price.

This negative response means that consumers of salmon steaks are quite sensitive to price changes. An increase in price may significantly reduce sales, so you need to be very cautious when considering a price increase to avoid a sharp decline in sales due to excessive price increases. In addition to price, there may be other market factors that affect the sales of salmon steaks, causing abnormal fluctuations in sales at individual price points. Overall, the results of this analysis provide a basis for the pricing strategy of salmon steak products, reflect the price sensitivity of consumers, and help to formulate reasonable prices to balance profits and sales.

Figure 18: Price vs. Quantity Sold for ESCORT BLUE with Polynomial Regression

Figure 18 clearly points out that Quantity Sold shows a trend of first rising and then falling as the price rises. 125 is a price point. Before that, an increase in price and an increase in sales occurred at the same time, but after that, an increase in price led to a decrease in sales.

When the price range is between 118 and 125, a 1% increase in price results in an average increase in sales volume of about 7.99%. This suggests that within this price range, sales respond positively to price increases, suggesting a potential value perception or promotional effect at these price points. When the price range is between 126 and 132, a 1% increase in price will lead to an average decrease in sales volume of about -16.97%. This contrasts with the first range, which shows a significant negative sensitivity to price increases, suggesting typical consumer behaviour where price increases reduce demand.

Conclusion

The analysis of three years of transaction data from the supermarket chain has provided significant insights into customer behaviors, sales forecasting, and price sensitivity. A crucial element of our findings is the identification of 10,487 regular customers who consistently engaged with the supermarket across the observed period. These loyal patrons primarily contribute to sales in the Grocery, Fruit & Veg, and Dairy departments, highlighting areas for focused marketing and inventory management.

The clustering of regular customers has revealed distinct groups with varying purchasing behaviors and needs. The segmentation into low-value and high-value customer clusters allows for differentiated marketing strategies tailored to each group's characteristics. Low-value regular customers, who frequent basic commodity purchases, could be targeted with promotions aimed at increasing their purchase frequency and transitioning them to higher-value segments. Conversely, high-value regular customers, demonstrating high purchase frequency and larger transaction volumes, should be engaged with personalized experiences and loyalty programs to maintain their satisfaction and spending.

The predictive sales forecast for the first quarter of 2016, using the ARIMA model, shows a slight decline in total sales, with January forecasted at $168.94 million, February at $163.54 million, and March at $163.39 million. This indicates the necessity for robust promotional strategies to bolster sales during this period, particularly in January, and adjusting marketing tactics as the quarter progresses.

Additionally, the analysis of price changes across various product categories indicates a need for strategic pricing adjustments. The relationship between price fluctuations and sales volumes has shown that price sensitivity varies significantly among different departments and must be managed to avoid adverse impacts on sales.

To maintain competitiveness, the supermarket should leverage these analytical insights to refine its marketing strategies, enhance customer satisfaction, and optimize operations. By doing so, it can better serve its established customer base, attract new customers, and maximize profitability in a challenging retail environment.