

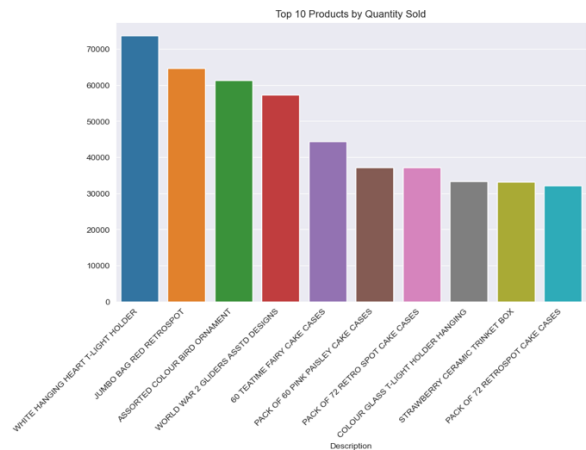
DSAA 5002 - Data Mining and Knowledge Discovery in Data Science

Final Exam Report – Q4 Recommendation and Business Analysis

50015940 Jiaxiang Gao

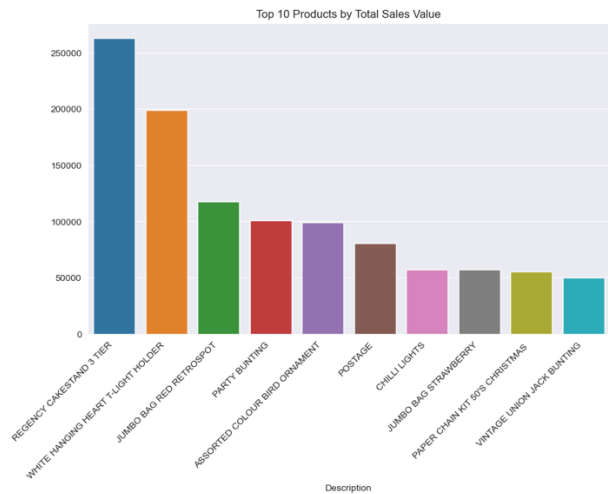
1 Visualization

1.1 Top 10 Products by Quantity Sold



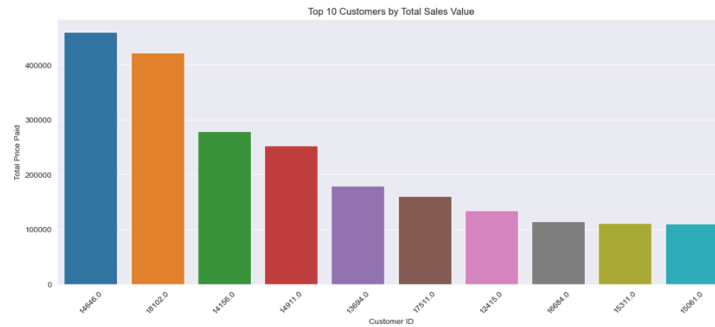
This Figure shows the top 10 products by quantity sold. They are white hanging heart t-light holder, jumbo bag red retrospot, assorted colour bird ornament, world war 2 gliders asstd designs, 60 teatime fairy cake cases, pack of 60 pink paisley cake cases, pack of 72 retro spot cake cases, colour glass t-light holder hanging, strawberry ceramic trinket box, pack of 72 retrospot cake cases.

1.2 Top 10 Products by Total Sales Value



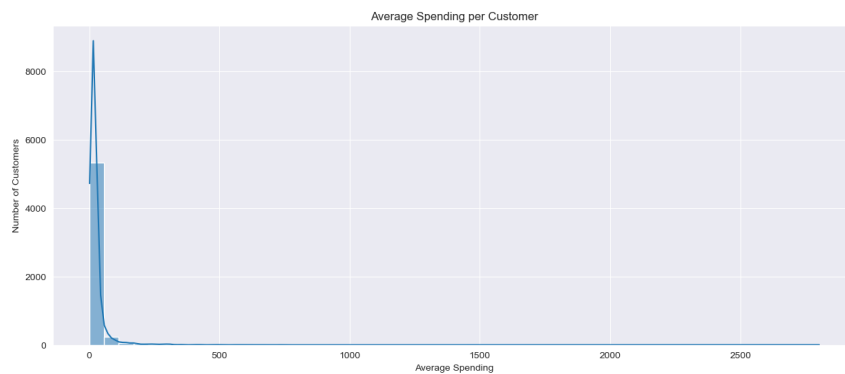
This Figure shows the top 10 products by total sales value. They are regency cakestand 3 tier, white hanging heart t-light holder, jumbo bag red retrospot, party bunting, assorted colour bird ornament, postage, chilli lights, jumbo bag strawberry, paper chain kit 50's christmas, vintage union jack bunting.

### 1.3 Top 10 Customers by Total Sales Value



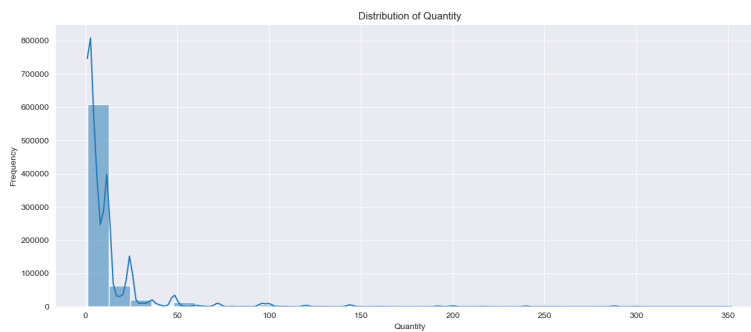
This Figure shows the top 10 Customers by Total Sales Value.

### 1.4 Average Spending per Customer



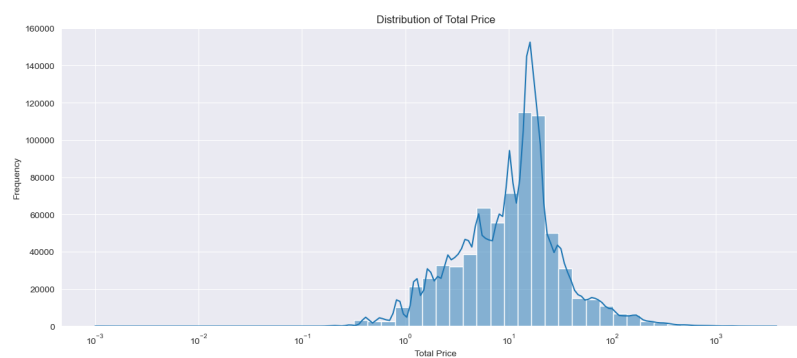
This Figure shows average spending per customer. We can see that the average cost of most customers is less than 250.

### 1.5 Distribution of Quantity



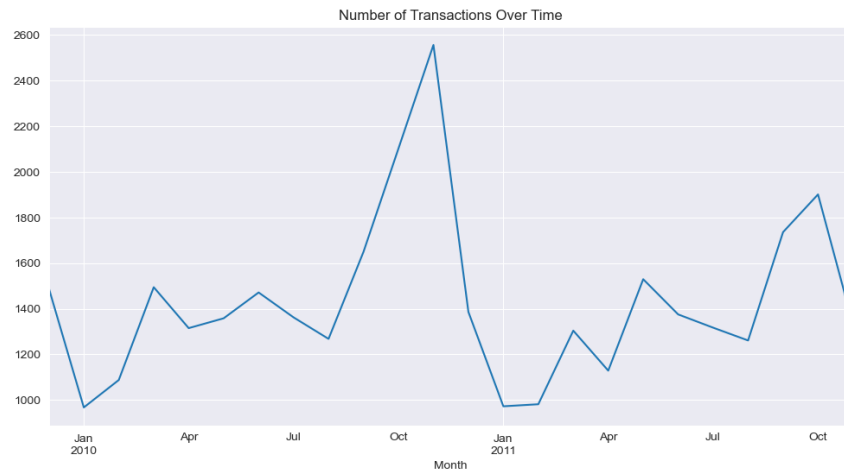
This Figure shows distribution of quantity. We can see that most quantities are less than 50.

### 1.6 Distribution of Total Price



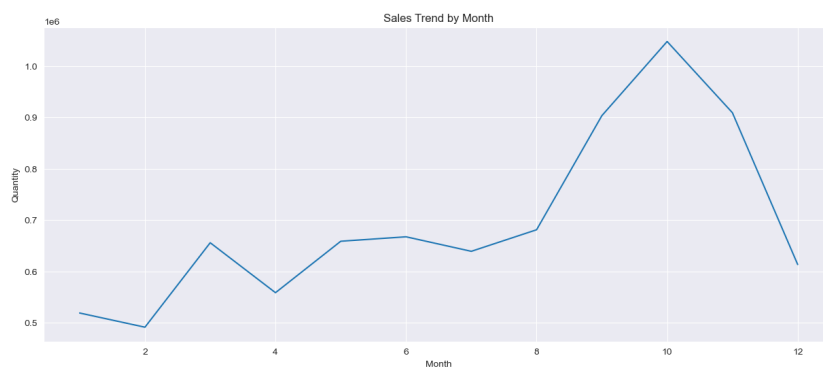
This Figure shows distribution of total price. We can see that most price is concentrated between 0 and 100.

## 1.7 Number of Transactions Over Time



This Figure shows the number of transactions over time. There were higher Transactions in October 2010 and October 2011, while there were lower Transactions in January 2010 and January 2011.

## 1.8 Sales Trend by Month



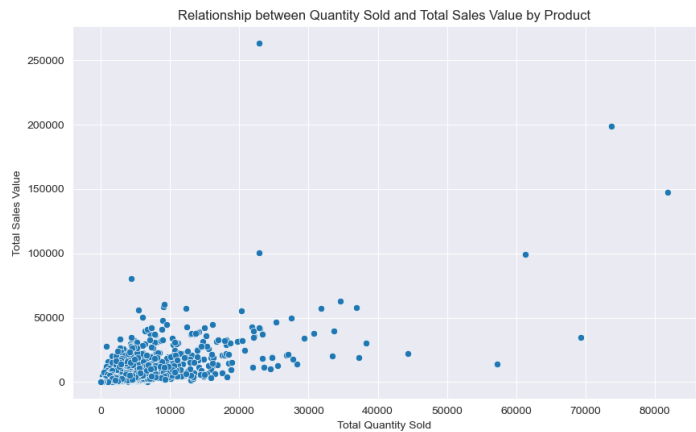
This Figure shows the sales trend by month. There are higher transactions from September to November and lower transactions from December to February.

## 1.9 Sales Trend by Hour



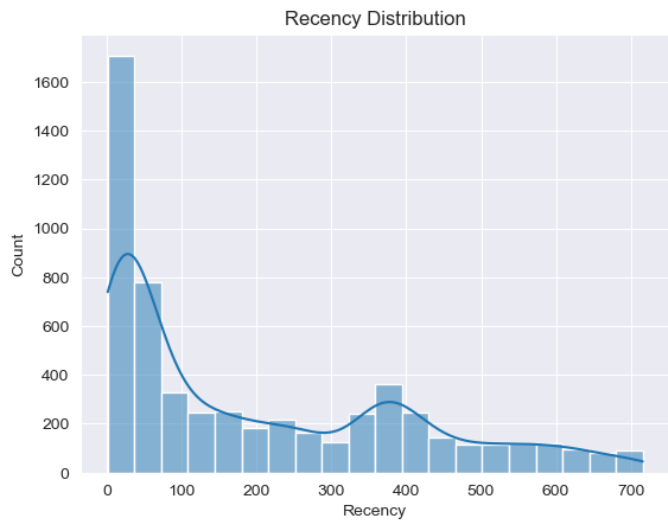
This Figure shows the sales trend by month. There are a relatively large number of transactions from 10:00 to 15:00, and there are relatively few other times.

1.10 Relationship between Quantity Sold and Total Sales Value by Product



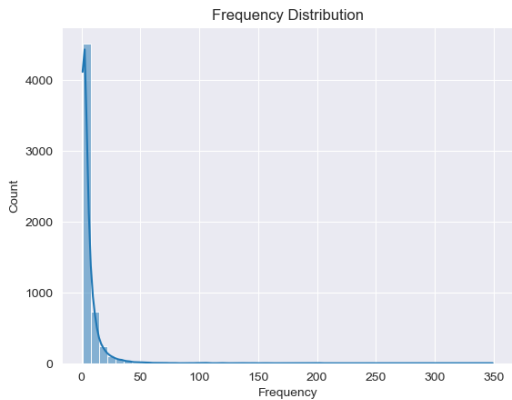
This figure illustrates that most of the data is concentrated at the lower end of the two axes, and as the sales volume and sales value increase, the data points gradually become sparse.

1.11 Recency Distribution from RFM Analysis



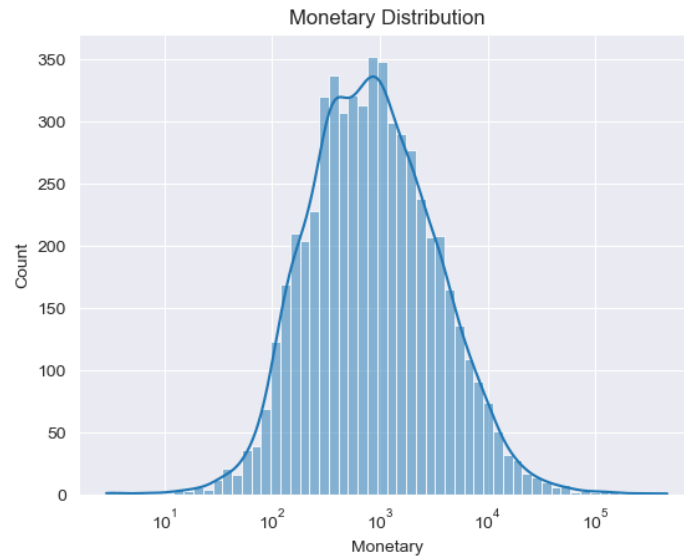
This figure shows the recency distribution from RFM analysis. We can see that the most recent purchase time for most customers is concentrated within 100 days. As time increases, the number of customers who made purchases sharply decreases and maintains a lower level over longer time intervals.

1.12 Frequency Distribution from RFM Analysis



This figure depicts the frequency of transactions per customer. Most customers seem to have a low frequency of transactions, with a steep drop-off as frequency increases.

### 1.13 Monetary Distribution from RFM Analysis



This figure shows the monetary distribution from RFM analysis. We can see that the consumption amount of most customers is concentrated in the lower amount range, and as the amount increases, the number of customers who consume that amount rapidly decreases.

## 2 Business Insights

### 2.1 Seasonal Trends in Sales

There is a clear seasonal pattern in sales, with higher transactions from September to November and lower transactions from December to February. This suggests the need for seasonal marketing strategies and stock adjustments.

### 2.2 Popular Products

The white hanging heart t-light holder, jumbo bag red retro spot, and assorted colour bird ornament are consistently top sellers both in terms of quantity and sales value. Focusing on these high-demand products in marketing and stock can boost sales.

### 2.3 High-Value Items

The regency cakestand 3 tier and party bunting are among the top products by total sales value, indicating that they are high-ticket items. Promotions or bundle deals involving these items might increase their sales further.

### 2.4 Customer Spending Patterns

Most customers spend less than 250, suggesting that the supermarket primarily attracts budget-conscious shoppers. Tailoring promotions and stock to this demographic could be beneficial.

### 2.5 Quantity Distribution

Most sales quantities are less than 50 units. This could indicate that bulk purchases are rare, and the supermarket may benefit from targeting individual or small-scale buyers.

## **2.6 Price Sensitivity**

The concentration of most sales prices between 0 and 100 suggests that customers are price-sensitive. Competitive pricing and discounts could be effective strategies.

## **2.7 Transactional Peaks and Troughs**

The spikes in transactions in October each year and the dips in January suggest the influence of holiday shopping patterns. Preparing for these fluctuations can optimize inventory and staffing.

## **2.8 Hourly Sales Patterns**

The increase in transactions between 10:00 and 15:00 indicates peak shopping hours. Staffing and in-store promotions could be optimized around these times.

## **2.9 Product-Sales Correlation**

The concentration of data at the lower end of the axes in the relationship between quantity sold and total sales value suggests that most products have low to moderate sales volumes and values. Identifying and promoting underperforming products with high potential could be beneficial.

## **2.10 RFM Analysis Insights**

**2.10.1 Recency:** Most customers make purchases within 100 days. Engaging customers shortly after their last purchase might encourage repeat sales.

**2.10.2 Frequency:** The low transaction frequency of most customers suggests a potential to increase customer loyalty programs.

**2.10.3 Monetary:** The concentration of customer spending in the lower range indicates a market for affordable goods, and there might be an opportunity to upsell or cross-sell to increase average spending.

### 3 Generating 10 Sales Boosting Recommendations through Association Rule Analysis

The table below shows the top 10 association rules by lift. Merchants can consider placing them in similar locations or bundling them for promotional sales according to these 10 rules. If you want to view all the rules, you can check the rules.csv file.

antecedent s	consequent s	antecedent support	consequent support	support	confidence	lift
('22746',)	('22745',)	0.01278392545136870 0	0.01537565521258010 0	0.01065812463599300 0	0.833712984054670 0	54.22292400082830
('22748',)	('22746',)	0.01697728596389050 0	0.01278392545136870 0	0.01132789749563190 0	0.667238421955403 0	52.19354763086230 0
('22748',)	('22745',)	0.01697728596389050 0	0.01537565521258010 0	0.01310425160163080 0	0.771869639794168 0	50.20076407297680
('21094',)	('21086', '21080')	0.01956901572510190	0.01147350029120560 0	0.01007571345369830 0	0.514880952380952 0	44.87566473289820
('21240',)	('21239',)	0.01627839254513690 0	0.01560861968549800 0	0.01112405358182880 0	0.683363148479428 0	43.78113902758120
('21900',)	('21899',)	0.01820034944670940 0	0.01447291788002330 0	0.01126965637740240 0	0.6192	42.78335613682090
('23301',)	('23300',)	0.01779266161910310 0	0.01485148514851490	0.01094933022714040	0.615384615384615 0	41.43589743589740
('21094',)	('21086',)	0.01956901572510190	0.01747233546884100 0	0.01371578334304020 0	0.700892857142857 0	40.11443452380950
('22577',)	('22578',)	0.01770529994175890 0	0.01560861968549800 0	0.01100757134536980	0.621710526315789 0	39.83123036135110
('21901',)	('21900',)	0.01592894583576000 0	0.01820034944670940 0	0.01144437973209090 0	0.718464351005484 0	39.47530530164530