

Customer Segmentation using RFM and Market Basket Analysis: Hospital Pharmacy

Submitted by

HARISH G D

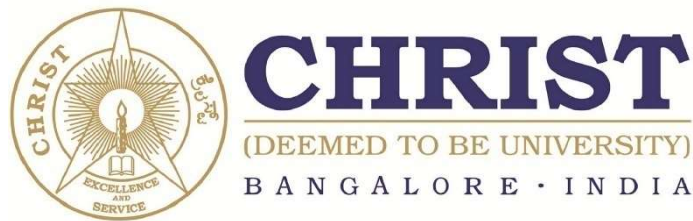
2327422

SUBJECT: Machine Learning Algorithms (MLA 2)

CIA 2 Assignment

Guided by

Prof. Nagendra B V



CHRIST SCHOOL OF BUSINESS AND MANAGEMENT

BENGALURU

October-2024

Domain Overview: Hospital-Pharmacy Sector

The **Hospital-pharmacy sector** represents a unique convergence of healthcare and retail industries. Pharmacies operating within or in association with hospitals serve not only the medical staff and in-patients but also out-patients, visitors, and general customers. Unlike typical retail environments, these pharmacies manage a mix of prescription medications, over-the-counter (OTC) products, personal care items, and health-related supplements, creating a specialized demand pattern. The primary goal of this sector is to ensure that patients receive timely access to medications while maintaining optimal inventory levels and customer satisfaction.

Pharmacies in hospital settings face distinct challenges in managing inventory, as they must balance the need for medical urgency with business objectives. Effective stock management is crucial, particularly for fast-moving and essential items. Additionally, customer segmentation and tailored marketing become vital in understanding and serving diverse needs—ranging from frequent, loyal hospital visitors to one-time customers.

The sector has also seen a growing emphasis on customer feedback and satisfaction. With customer loyalty programs and personalized promotions, pharmacies aim to enhance engagement, ensuring that patients and visitors prefer their services over external retail outlets. This has led to the adoption of advanced analytics to optimize inventory, identify product demand trends, and tailor promotions, enabling hospital pharmacies to serve both medical and business goals.

About the Dataset:

In the hospital-pharmacy sector, understanding customer purchasing behavior is crucial for optimizing inventory management, enhancing customer satisfaction, and driving sales growth. However, the diverse nature of customers—ranging from patients, caregivers, and hospital staff to visitors—creates challenges in segmenting and targeting marketing efforts effectively.

The key challenges include:

1. **Customer Segmentation:** The inability to accurately segment customers based on their purchasing behavior hinders targeted marketing efforts. Identifying high-value customers, potential loyalists, and at-risk customers is essential for personalized marketing and retention strategies.

2. **Inventory Management:** With a wide range of products, including prescription medications and OTC items, effective inventory management is crucial to avoid stockouts or overstock situations. Understanding purchasing patterns can help forecast demand and optimize stock levels.
3. **Market Basket Analysis:** Insights into frequently purchased items together can inform promotional strategies and improve cross-selling efforts. Identifying these patterns helps enhance customer experience and increases average transaction values.

Data Dictionary

Column Name	Data Type	Description
CustomerID	Integer	Unique identifier for each customer.
TransactionDate	Date	Date on which the transaction occurred.
AmountSpent	Float	Total amount spent by the customer in a single transaction, indicating monetary value.
ItemsBought	String	List of items purchased in each transaction, separated by commas.
CustomerAge	Integer	Age of the customer, providing demographic information.
CustomerGender	String	Gender of the customer (e.g., Male, Female, Other), used for demographic analysis.
StoreID	Integer	Unique identifier for the store where the transaction took place, useful for multi-store analysis.
PaymentMethod	String	Method of payment used by the customer (e.g., Credit Card, Cash, Insurance), indicating customer preferences.
CustomerRegion	String	Region or geographical area where the customer resides (e.g., Urban, Rural), providing context for marketing strategies.
FeedbackRating	Integer	Customer feedback rating on a scale of 1 to 5, measuring satisfaction with the purchase experience.

Problem Statement:

The Objective is to identify customer segments and uncover purchasing patterns that can drive personalized marketing strategies and improve overall service delivery in the hospital-pharmacy sector.

Data Preparation:

Data Cleaning:

Since there are no missing values in the Dataset, The Data Cleaning part is not necessary.

```
Missing values in each column:
```

```
CustomerID      0
TransactionDate  0
AmountSpent      0
ItemsBought      0
CustomerAge      0
CustomerGender   0
StoreID          0
PaymentMethod    0
CustomerRegion   0
FeedbackRating   0
dtype: int64
```

Exploratory Data Analysis (EDA):

	AmountSpent	CustomerAge	FeedbackRating
count	5000.000000	5000.000000	5000.000000
mean	103.701062	50.970000	2.99060
std	54.117208	19.639619	1.41722
min	10.000000	18.000000	1.00000
25%	57.277500	34.000000	2.00000
50%	103.850000	51.000000	3.00000
75%	149.372500	68.000000	4.00000
max	199.990000	84.000000	5.00000

AmountSpent: Customers spend an average of \$103.70, with a range from \$10 to \$199.99. The distribution is right-skewed, meaning most customers spend less than the average.

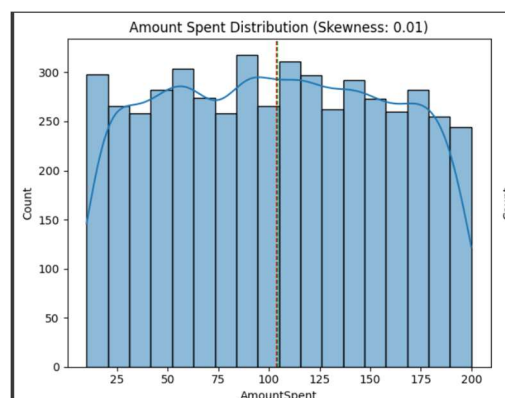
CustomerAge: The average customer is 50.97 years old, with a range from 18 to 84. The distribution of ages is roughly symmetrical.

FeedbackRating: Customers give an average rating of 2.99 out of 5. The distribution is right-skewed, indicating more lower ratings than higher ones.

Overall:

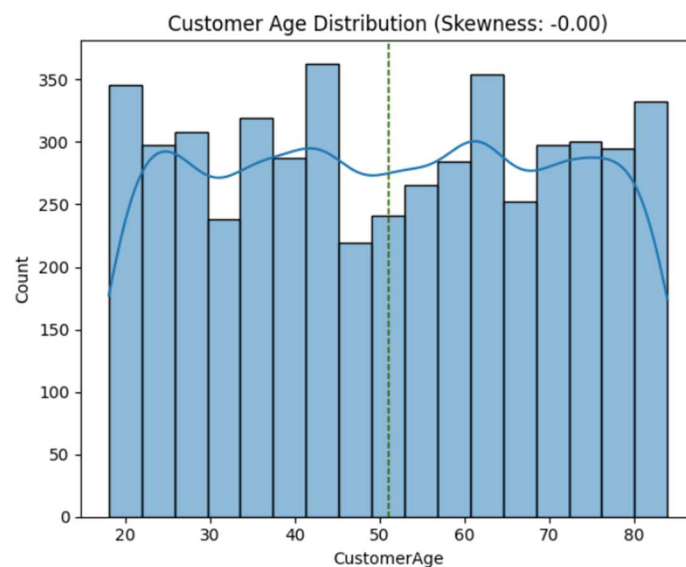
- Customers tend to spend less than the average.
- Customer ages are spread out fairly evenly.
- Most customers are not very satisfied with the product or service.

Distribution of AmountSpent:



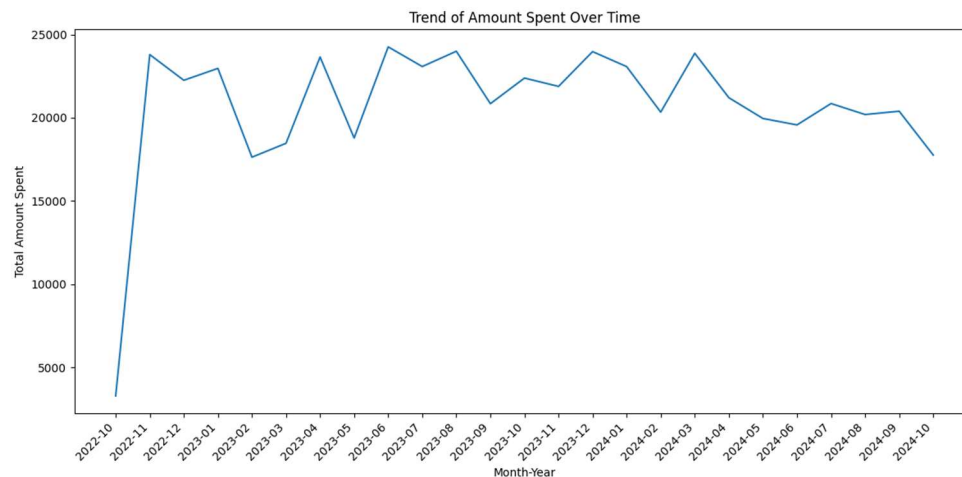
The histogram shows the distribution of customer spending. Most customers spend a moderate amount, with a few outliers who spend significantly more or less. The distribution is slightly right-skewed, meaning there are more customers spending less than the average.

Distribution of Customer Age:



The distribution of customer ages is almost perfectly symmetrical. There are no significant outliers, and the customer base is evenly distributed across different age groups.

Trend of Income:



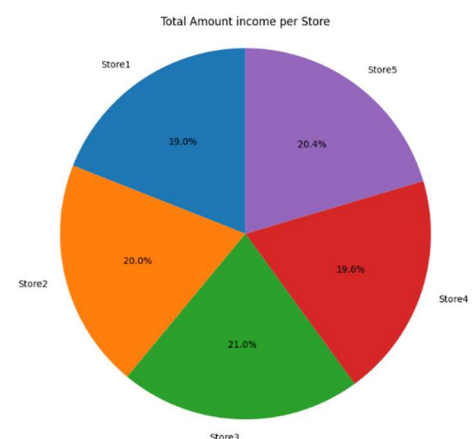
Key Observations:

- There are several peaks and troughs in the spending pattern.
- The highest spending occurred in May 2023, and the lowest spending was in February 2023.
- The spending in the recent months (September and October 2024) seems to be on the lower side compared to earlier months.

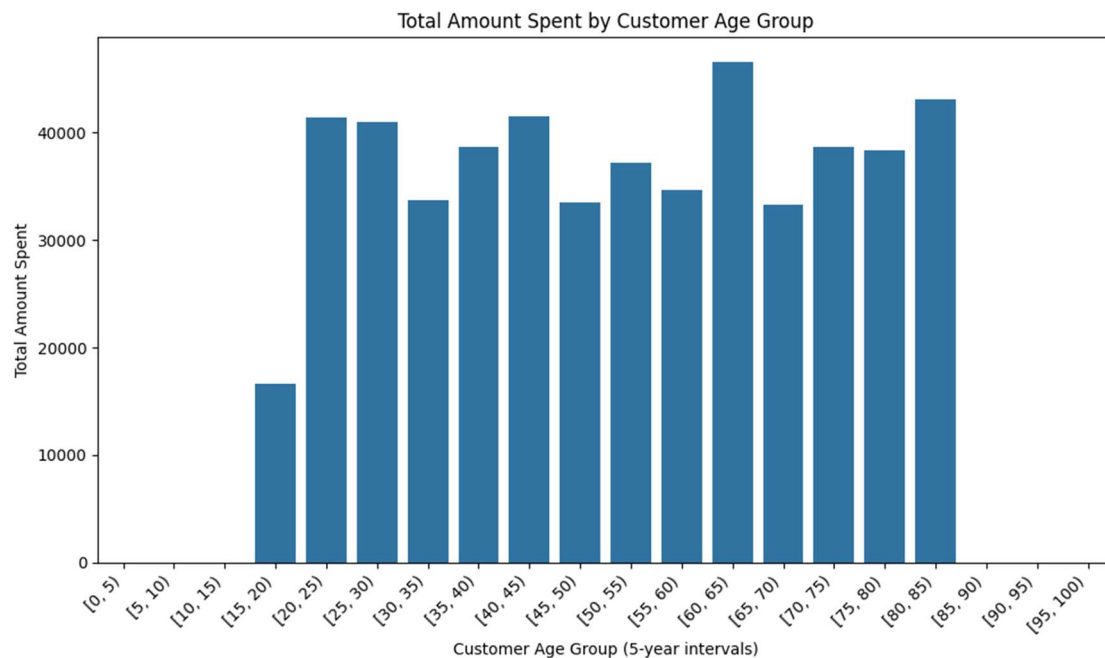
The fluctuations in spending could be due to various factors such as seasonal changes, marketing campaigns, economic conditions, or changes in product offerings. The recent dip in spending could be a temporary trend or might indicate a longer-term change in customer behavior..

Total Income by stores:

Store 3 has More Income than other Stores and store 1 is least in the sales.



Amount Spent by Age Group:



The total amount spent varies significantly across different age groups. The highest spending is observed in the 60-65, 80-85, 25-30, and 35-40 age groups. In contrast, the lowest spending is seen in the 15-20, 0-5, and 10-15 age groups.

RFM Analysis:

Business Inference:

Elevated RFM ratings indicate the pharmacy's most valued customers, characterized by recent transactional activity, frequent purchases, and substantial spending levels. This segment represents a critical source of revenue for the pharmacy.

- **High Recency, Moderate Frequency, and High Monetary:** These customers have recently made purchases and exhibit high spending, albeit with moderate transaction frequency. They may be less frequent shoppers but tend to purchase high-value items, such as specialty medications or health products.
- **High Frequency but Lower Monetary Value:** Customers in this category engage consistently with the pharmacy, purchasing lower-cost items regularly. Their loyalty reflects a stable relationship with the pharmacy, even if their overall spending is lower.

```

RFM Metrics:
  Recency  Frequency  Monetary
CustomerID
CUST0001    144         9   1020.10
CUST0002    329         1    50.20
CUST0003    118         9   1103.80
CUST0004    169         6    650.22
CUST0005    158         7    636.10

RFM Table:
  Recency  Frequency  Monetary  R_Score  F_Score  M_Score  RFM_Score
CustomerID
CUST0001    144         9   1020.10         2         5         5        255
CUST0002    329         1    50.20         1         1         1        111
CUST0003    118         9   1103.80         2         5         5        255
CUST0004    169         6    650.22         2         4         4        244
CUST0005    158         7    636.10         2         4         4        244

Customer Segmentation:
Segment
At Risk                24
Lost Customers        191
Loyal Customers       190
Neutral               235
New Customers         165
Potential Loyalists   188
dtype: int64

```

Customer Segmentation Interpretation

The customer segmentation derived from the RFM analysis categorizes the customer base into five distinct segments based on their purchasing behavior:

1. At Risk (24 customers):

- These customers have not engaged in recent transactions and show signs of declining loyalty. Their low recency and frequency scores indicate that they may be on the verge of churning.

2. Lost Customers (191 customers):

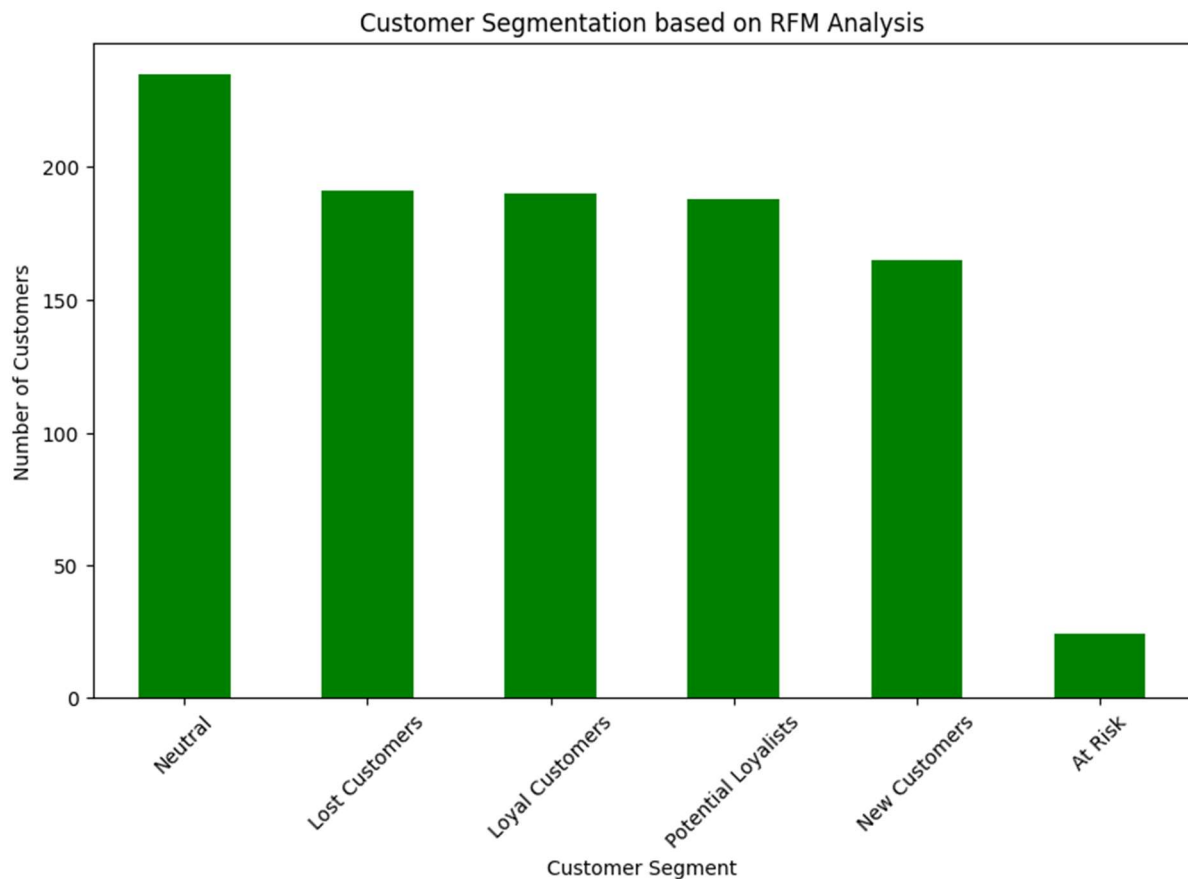
- This segment represents a significant number of customers who have not transacted for a considerable period. They are likely no longer utilizing the pharmacy's services or products, presenting a major concern for customer retention.

3. Loyal Customers (190 customers):

- These customers demonstrate high levels of engagement, with recent transactions, frequent purchases, and significant spending. They are likely to be repeat customers who contribute significantly to revenue.
- 4. **Neutral (235 customers):** Neutral customers demonstrate moderate engagement but do not fall into the loyal or at-risk categories. Their spending and frequency indicate a stable relationship but lack the high engagement seen in loyal customers.
- 5. **New Customers (165 customers):**
 - This segment consists of customers who have recently made their first purchase. Their low frequency and recency indicate that they are still in the early stages of their relationship with the pharmacy.
- 6. **Potential Loyalists (188 customers):**
 - These customers show moderate frequency but may have a lower monetary contribution. They have made recent purchases but do not yet exhibit the spending levels seen in loyal customers. They represent an opportunity for growth.

Recommendations

To enhance customer engagement and retention across various segments, the pharmacy should implement tailored strategies for each group. For **At Risk** customers, targeted re-engagement campaigns with personalized discounts and feedback solicitation can rekindle interest. **Lost Customers** should receive compelling win-back offers and regular reminders of new products to entice their return. To reward **Loyal Customers**, enhanced loyalty programs with exclusive perks and personalized communications will foster continued engagement. **Neutral Customers** can be encouraged to increase their spending through engagement initiatives and incentivized offers, while **New Customers** should be welcomed with onboarding campaigns and regular follow-ups to build familiarity with the pharmacy's services. Lastly, **Potential Loyalists** can be converted into loyal customers through targeted promotions and relationship-building efforts that highlight the value of increased spending. By adopting these tailored approaches, the pharmacy can effectively boost customer loyalty, drive sales, and improve overall satisfaction.



Cluster Analysis:

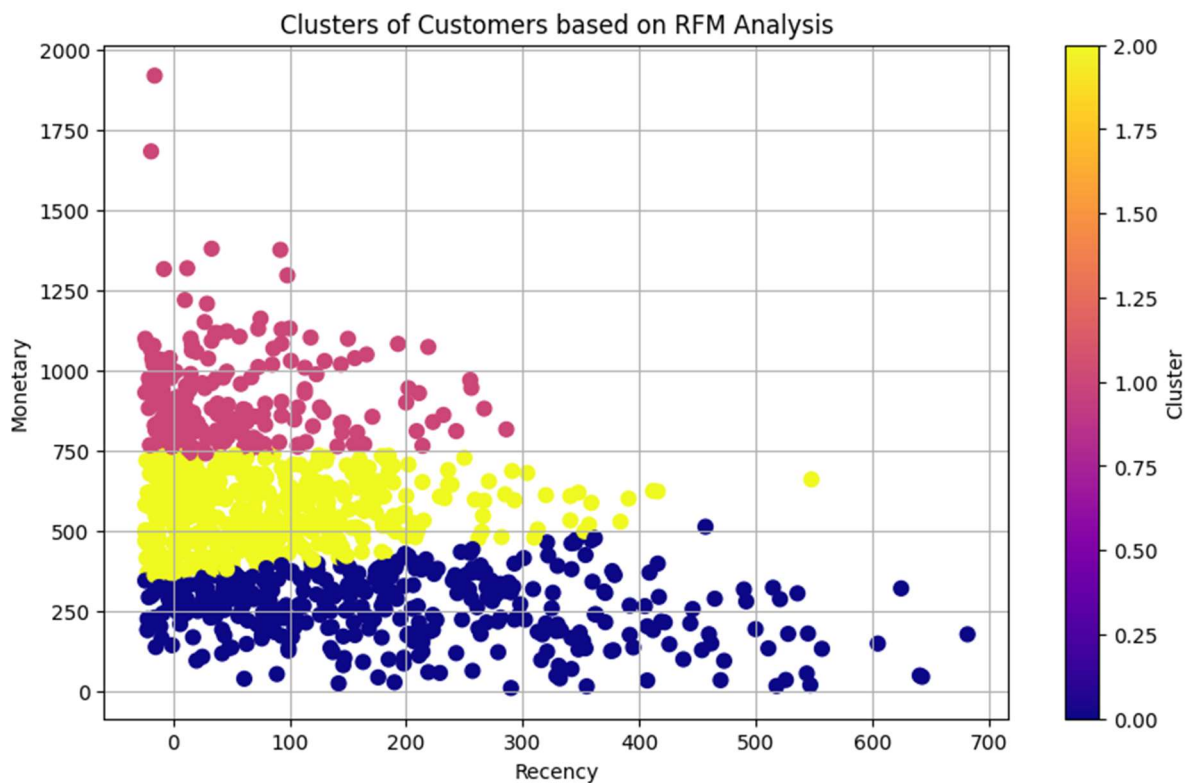
Average RFM Metrics by Cluster:				
	Cluster	Recency	Frequency	Monetary
0	0	193.203857	3.038567	263.602893
1	1	52.825641	7.856410	922.205487
2	2	81.381609	5.436782	558.591701

The average RFM metrics by cluster reveal distinct customer behaviors within each group:

- **Cluster 0:** This cluster has a high average recency (193.20 days), indicating that these customers haven't interacted with the pharmacy for a while. They show a low frequency of transactions (3.04), suggesting infrequent purchases, but their monetary value is moderate (approximately \$263.60). These customers are likely **At Risk** or **Lost Customers**, highlighting the need for re-engagement strategies to encourage them to return.
- **Cluster 1:** Customers in this cluster have the lowest average recency (52.83 days), indicating recent transactions. They exhibit the highest frequency (7.86) and substantial monetary spending (approximately \$922.21). This group can be classified

as **Loyal Customers**, who frequently engage with the pharmacy and contribute significantly to revenue. Strategies aimed at rewarding and maintaining their loyalty should be prioritized.

- **Cluster 2:** With a moderate recency (81.38 days) and a frequency of 5.44, these customers display decent engagement levels. Their monetary value (approximately \$558.59) is also noteworthy. They may be seen as **Potential Loyalists**; therefore, initiatives to convert them into loyal customers through targeted marketing efforts and personalized incentives would be beneficial.



In summary, the pharmacy should focus on re-engagement tactics for Cluster 0, loyalty programs for Cluster 1, and targeted promotions for Cluster 2 to optimize customer relationships and revenue.

Market Basket Analysis:

Association Rules:						
	antecedents \					
100717	('Insulin', 'Insulin')					
100706	('Multivitamin', 'Antiseptic Lotion', 'Thermom...')					
131930	('Antibiotic Cream', 'Antiseptic Lotion', 'Cou...')					
131913	('Bandages', 'Paracetamol', 'Thermometer')					
140192	('Antiseptic Lotion', 'Calcium Supplement', 'C...')					
...	...					
5652	('Bandages', 'Vitamin C')					
5639	('Bandages', 'Insulin')					
625	('Vitamin C')					
2009	('Antibiotic Cream', 'Vitamin C')					
1335	('Antibiotic Cream', 'Cough Syrup')					
	consequents		antecedent support \			
100717	('Multivitamin', 'Antiseptic Lotion', 'Thermom...')		0.143001			
100706	('Insulin', 'Insulin')		0.162135			
131930	('Bandages', 'Paracetamol', 'Thermometer')		0.148036			
131913	('Antibiotic Cream', 'Antiseptic Lotion', 'Cou...')		0.161128			
140192	('Blood Pressure Monitor', 'Multivitamin', 'Th...')		0.149043			
...	...					
5652	('Vitamin C')		0.154079			
5639	('Vitamin C')		0.161128			
625	('Vitamin C')		0.289023			
2009	('Vitamin C')		0.165156			
1335	('Antibiotic Cream')		0.152064			
	consequent support	support	confidence	lift	leverage \	
100717	0.162135	0.054381	0.380282	2.345464	0.031195	
100706	0.143001	0.054381	0.335404	2.345464	0.031195	
131930	0.161128	0.055388	0.374150	2.322066	0.031535	
131913	0.148036	0.055388	0.343750	2.322066	0.031535	
140192	0.161128	0.055388	0.371622	2.306377	0.031373	

Frequent Itemsets: The frequent itemsets reveal which products are commonly purchased together, with the support value indicating the proportion of transactions that include each item or set of items. For example:

1. **Antibiotic Cream:** Support of approximately 49.45% suggests that nearly half of all transactions include this item, making it a significant product for the pharmacy.
2. **Antiseptic Lotion:** At 48.54%, this item is also highly prevalent, indicating strong customer preference.
3. **Aspirin:** With a support of about 51.56%, this item is a staple purchase, possibly related to common health needs.

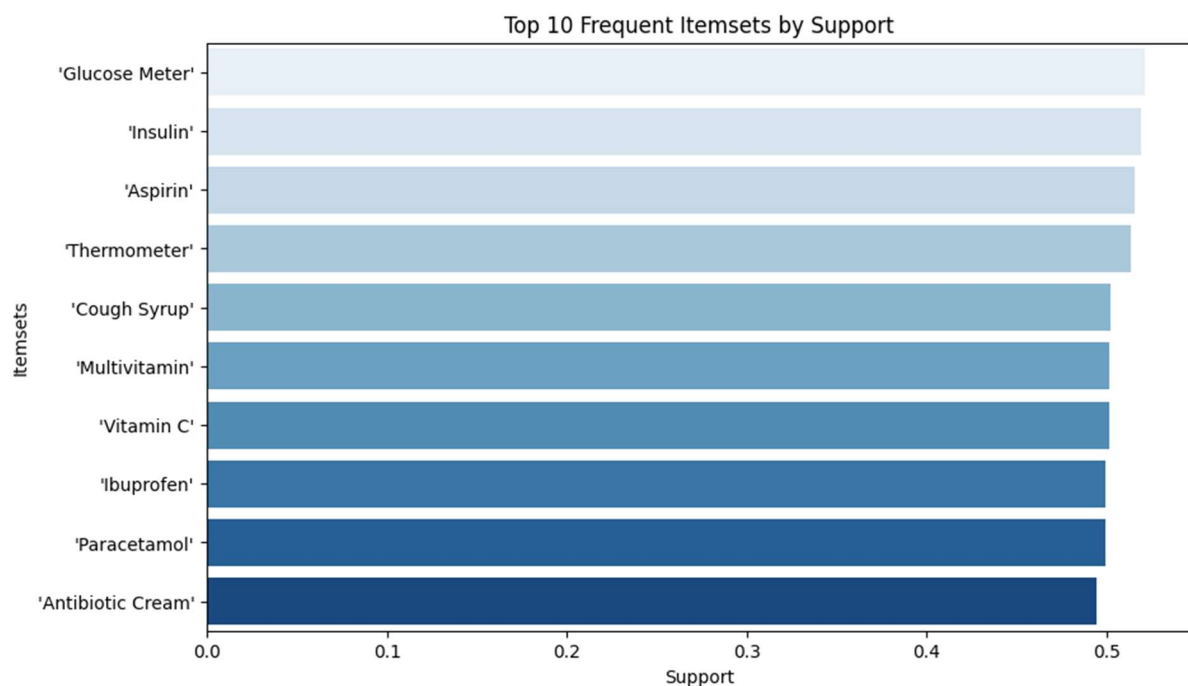
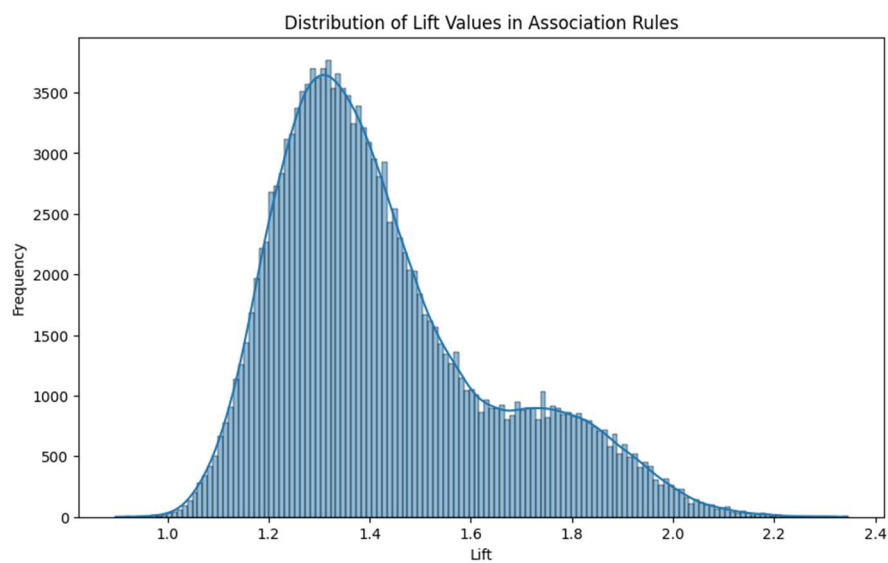
These insights can guide inventory management, promotional strategies, and product placement in-store to enhance customer convenience.

Association Rules: The association rules indicate relationships between products that customers frequently buy together, providing deeper insights into consumer behavior. Key metrics for interpretation include confidence and lift:

- **Confidence:** This metric shows the likelihood that a customer who buys the antecedent will also purchase the consequent. For example, a confidence of 38.03%

for the rule involving **Insulin** indicates a significant chance that customers who buy Insulin will also purchase **Multivitamin**, **Antiseptic Lotion**, and **Thermometer**.

- **Lift:** This metric compares the observed support of a rule with the expected support if the antecedents and consequents were independent. A lift greater than 1 indicates a strong positive association. For instance, a lift of 2.35 for the **Insulin** rule suggests that the combination is over twice as likely to occur than would be expected by chance, highlighting a strong correlation between those products.



Business Recommendations:

To enhance sales and customer satisfaction, the pharmacy should implement several strategic recommendations based on frequent itemsets and association rules. Promotional bundles can encourage customers to buy complementary items, such as pairing **Antibiotic Cream** with other first aid products. Targeted marketing campaigns should focus on specific customer segments that frequently purchase related health products, like **Insulin**. Ensuring high-support items, such as **Aspirin** and **Antiseptic Lotion**, are well-stocked is crucial for customer retention. Additionally, marketing materials should highlight associations between popular items, encouraging related purchases, while customer education through signage or flyers can inform shoppers about the benefits of combined purchases, ultimately driving sales growth.

Conclusion and Strategic Plan:

To optimize operations and enhance customer engagement in the hospital-pharmacy sector, the following strategic recommendations are proposed:

1. **Customer Segmentation:** Implement robust segmentation to tailor marketing strategies for distinct groups such as At Risk, Loyal, and New Customers, focusing on personalized communication and targeted promotions.
2. **Re-engagement Campaigns:** Develop targeted campaigns for At Risk and Lost Customers, using personalized discounts and feedback solicitation to rekindle interest.
3. **Enhanced Loyalty Programs:** Create exclusive rewards for Loyal Customers, encouraging repeat business and referrals to expand the customer base.
4. **Market Basket Analysis:** Utilize insights to develop promotional bundles and strategically place related products together to encourage cross-selling..