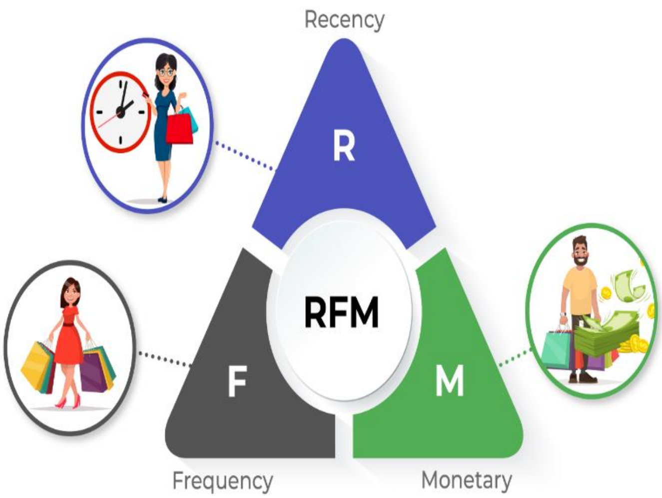
Marketing and Retail Analytics
RFM & MBA Final Project



RFM (Recency, Frequency, Monetary) analysis segments customers based on how recently they purchase, how often they purchase, and how much they spend. This helps in identifying high-value customers and tailoring marketing strategies effectively.

Market Basket Analysis is a data mining technique used to understand the purchase behavior of customers by analyzing the co-occurrence of items in transactions. It identifies patterns and relationships between items frequently bought together, enabling businesses to optimize product placement, create effective

Great Learning

Authored by: Shrinidhi CG

Contents

Part A: Recency Frequency Monetary (RFM) Analysis

Part B : Market Basket Analysis

Part A : Recency Frequency Monetary (RFM) Analysis

Executive Summary

The dataset comprises transaction data from an automobile parts manufacturing company collected over three years. The primary objective of the analysis is to uncover underlying customer buying patterns, provide actionable insights, and recommend customized marketing strategies for the company.

About the Data

1. Overview: *The dataset encompasses 20 columns and 2,757 rows of transactional data from an automobile parts manufacturing company over three years. It captures detailed information on orders, product pricing, customer data, and order statuses, providing a comprehensive view of the company's sales and customer behavior.*

2. Shape of data: *Dataset contains 2757 rows and 20 columns.*

3. Summary Statistics:

Numerical Variable

Column Name	Min	Max	Mean	Standard Deviation	Variance	Skewness	Kurtosis	No. zeros	25 % Quantile	50 % Quantile	75 % Quantile
QUANTITYORDERED	6	97	35.1	9.762	95.299	0.369	0.443	0	27	35	43
PRICEEACH	26.88	252.87	101	42.043	1767.58	0.697	0.229	0	68.71	95.55	127.1
ORDERLINENUMBER	1	18	6.49	4.231	17.897	0.575	-0.591	0	3	6	9
SALES	482.1	14083	3553	1838.954	3381751	1.156	1.773	0	2204.1	3184.8	4508
DAYS_SINCE_LASTORDER	42	3562	1757	819.281	671221	-0	-1.024	0	1077	1761	2437
MSRP	33	214	101	40.115	1609.2	0.576	-0.139	0	68	99	124

Categorical Variable

Column Name	No. Missings	Unique Values
CITY	0	71
COUNTRY	0	19
CUSTOMERNAME	0	89
DEALSIZE	0	3
PRODUCTLINE	0	7
STATUS	0	6

4. Assumptions About the Data:

Data Completeness: *The dataset is comprehensive with no missing values in crucial categorical columns, ensuring reliable categorical analysis.*

Numerical Data Insights:

QUANTITYORDERED: Orders generally fall between 6 and 97 units, with an average of about 35 units. The slight positive skew indicates that lower quantities are more common.

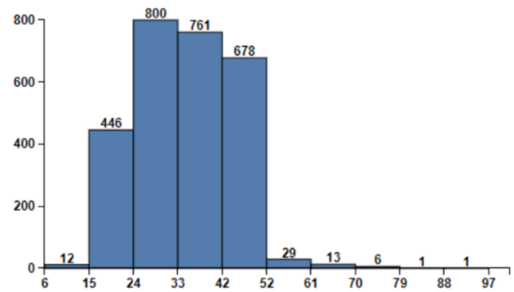


Figure 1: Quantity Ordered Histogram

PRICEEACH: Item prices vary significantly, from $26.88 to $252.87, with an average price of $101.10. The positive skew shows that higher prices are less common but influential.

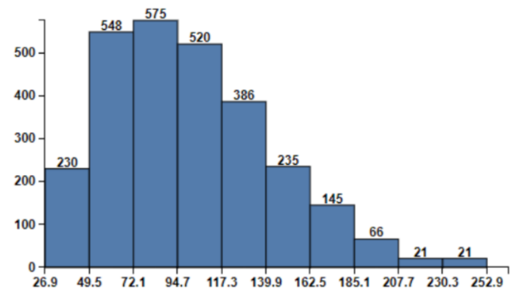


Figure 2: Price Each Histogram

ORDERLINENUMBER: The number of items per order line ranges from 1 to 18, with a mean of 6.49, suggesting moderate order complexity.

DAYS_SINCE_LASTORDER: This metric shows a broad range from 42 to 3,562 days, with a nearly symmetric distribution, suggesting varied customer re-engagement times.

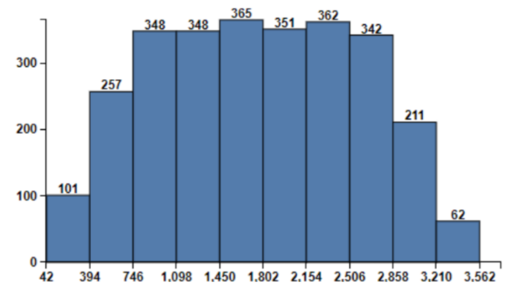


Figure 3: Days since last order Histogram

SALES: Sales figures show high variability, ranging from $482.13 to $14,082.80, indicating a significant impact from high-value transactions.

MSRP: The Manufacturer's Suggested Retail Price ranges from $33 to $214, with an average of $100.69, showing variability in product pricing.

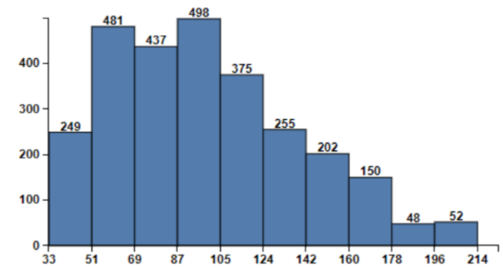


Figure 4: MSRP Histogram

Data Diversity and Segmentation: The diversity in CITY, COUNTRY, CUSTOMERNAME, PRODUCTLINE, and DEALSIZE allows for extensive customer segmentation and geographic analysis. The STATUS column provides insights into different stages of order processing, which could highlight operational efficiencies or issues.

Variability in QUANTITYORDERED, PRICEEACH, and SALES points to different customer buying behaviors, which can inform targeted marketing strategies.

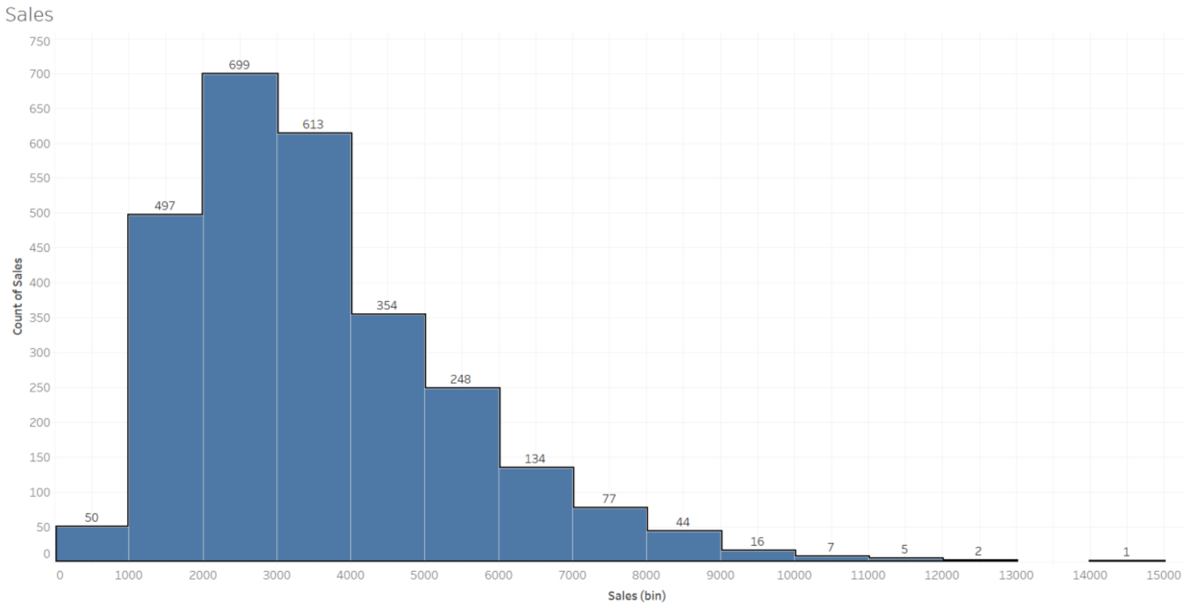
Customer Behavior: The broad range of DAYS_SINCE_LASTORDER indicates varying levels of customer engagement, which can be used to tailor re-engagement strategies.

High variability in SALES and PRICEEACH suggests that a few high-value transactions are significant, making them key targets for strategic focus.

The dataset offers a rich blend of numerical and categorical data, ideal for detailed customer analysis. The variability and range in key metrics like SALES, PRICEEACH, and QUANTITYORDERED provide valuable insights into customer behavior and sales patterns. The data's completeness and diversity support robust segmentation, targeted marketing, and strategic decision-making.

Exploratory Analysis and Inferences

<u>Univariate analysis of sales</u>



The analysis of the SALES variable reveals a right-skewed distribution, indicating that most sales transactions fall within lower sales ranges, with fewer transactions at higher sales values.

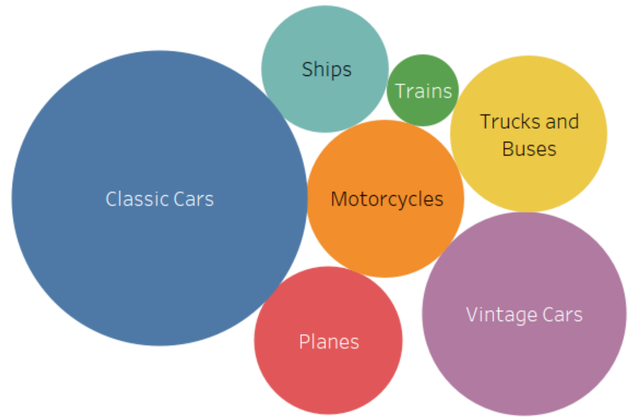
Dominance of Mid-Range Sales:

- *The majority of sales transactions occur in the $1000 to $5000 range, accounting for a significant portion of the dataset.*
- *Specifically, the $2000 to $3000 range has the highest frequency, with 699 transactions, followed by $3000 to $4000 with 613 transactions. This suggests that most sales are concentrated in this middle tier.*

Lower Frequency at Higher Sales Ranges:

- *As sales amounts increase beyond $5000, the frequency of transactions continues to decline.*
- *For example, there are 134 transactions in the $6000 to $7000 range, but only 77 in the $7000 to $8000 range. This pattern continues, with just 3 transactions falling within the $12,000 to $15,000 range.*

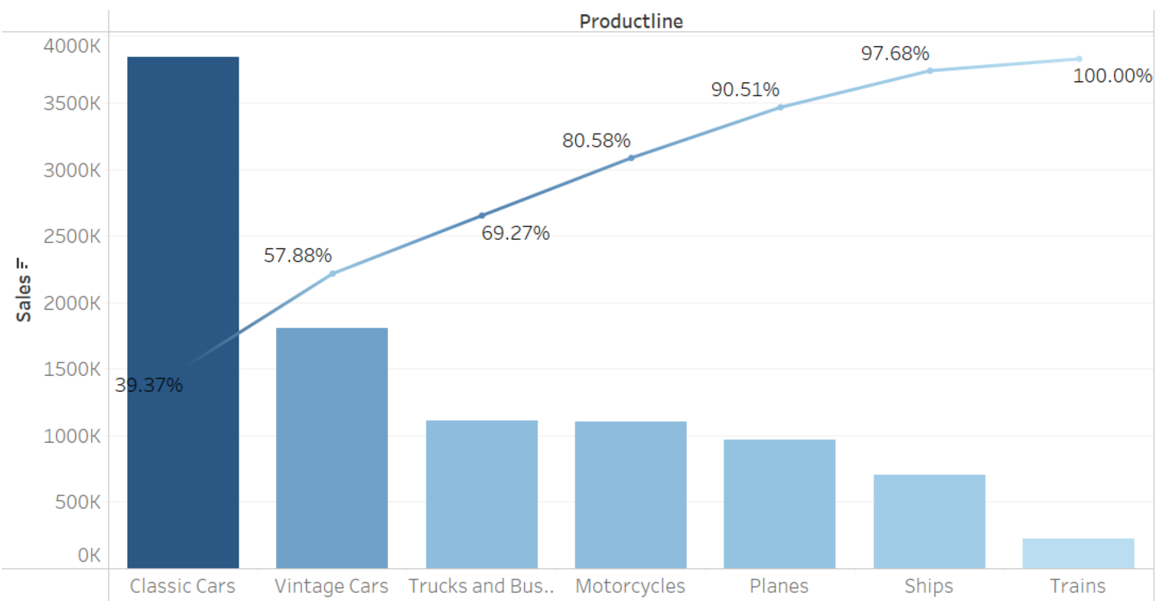
Bivariate Analysis of ProductLine & Price



Classic Cars, priced at $109,321, are the most expensive, likely appealing to collectors and high-end customers who value exclusivity and luxury. Vintage Cars, with an average price of $52,117, also cater to a premium market, though they are more accessible than Classic Cars.

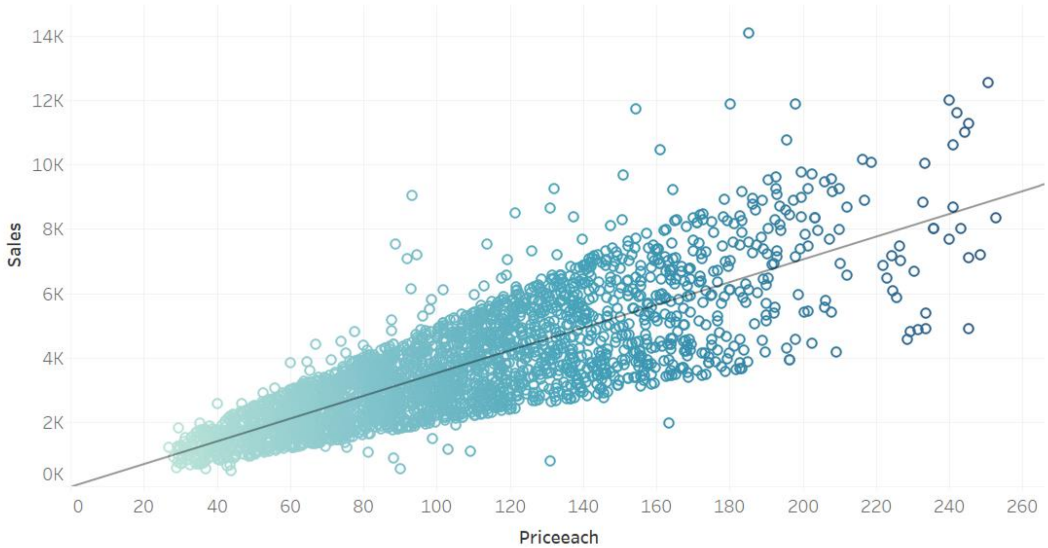
Mid-range products like Motorcycles and Trucks & Buses, priced around $31,000, seem to target practical buyers who seek a balance between cost and functionality. Planes Ships and Trains, priced at $27,517 $20,279 and $6,476respectively, are less expensive, potentially indicating smaller models or niche products tailored for specific interests.

ProductLine Vs Sales Pareto



The Pareto analysis reveals that a few key product lines—Classic Cars, Vintage Cars, Trucks and Buses, and Motorcycles—drive the majority of the company's sales. Classic Cars, contributing 39.37% of total sales, and Vintage Cars, contributing 57.88%, together form a significant portion of the revenue. When combined with Trucks and Buses (69.27%) and Motorcycles (80.58%), these categories *account for over 80% of the company's sales. This indicates that a focused strategy on these top-performing product lines is crucial. Classic Cars and Vintage Cars, in particular, are major revenue drivers with strong market demand, while Trucks and Buses also show robust performance, reflecting their importance in the commercial sector. Motorcycles, although less dominant individually, still represent a significant share of sales, underscoring their relevance in the company's portfolio. Prioritizing and optimizing strategies for these key product lines will be essential for maximizing revenue and sustaining growth.*

<u>Bivariate Analysis of Price & Sales</u>

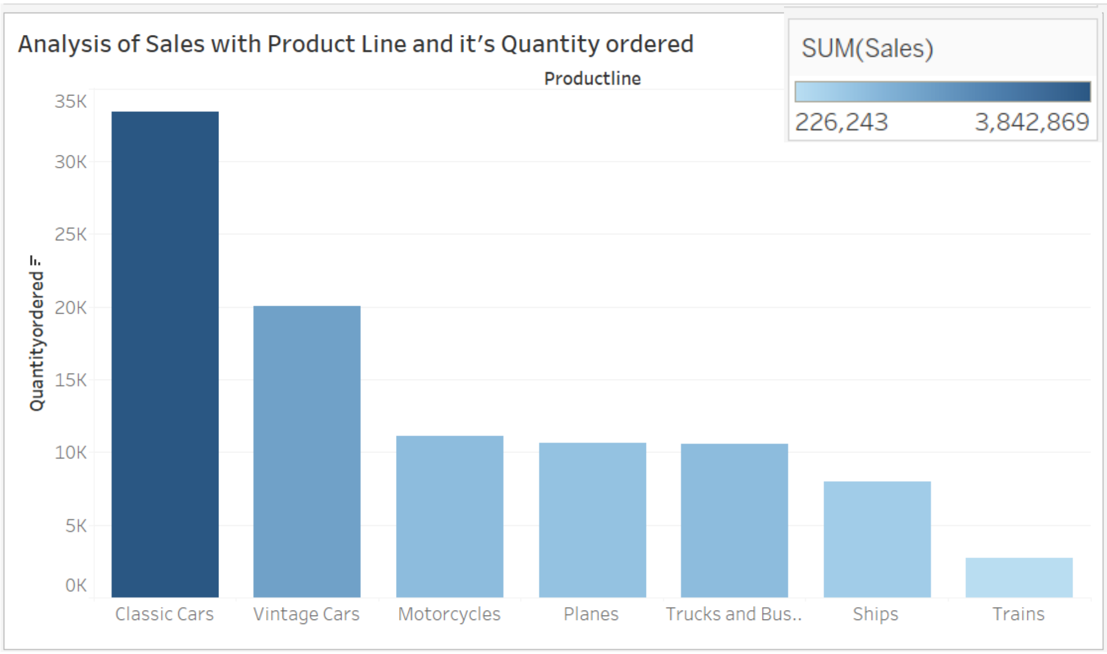


The bivariate analysis of price and sales reveals a positive correlation between these two variables, suggesting that higher-priced product lines tend to generate higher sales. As the price of a product line increases, so does the total sales generated by that line. This trend indicates that customers are willing to invest in more expensive items, possibly due to perceived value, quality, or the exclusivity of the products.

For example, Classic Cars, which have the highest average price of $109,321, also generate significant sales, reflecting strong demand in the premium segment. Similarly, Vintage Cars, while priced lower at $52,117, still contribute notably to overall sales. On the other hand, lower-priced product lines, such as Trains and Ships, tend to generate lower sales, which may reflect a smaller or more niche market.

This relationship suggests that the company's higher-priced products are key drivers of revenue, indicating the potential for focusing marketing efforts on these premium segments to maximize sales. However, the trend also highlights the importance of maintaining a balanced product portfolio to cater to different customer segments and price sensitivities.

<u>Multivariate Analysis of Product Line, Quantity & Sales</u>

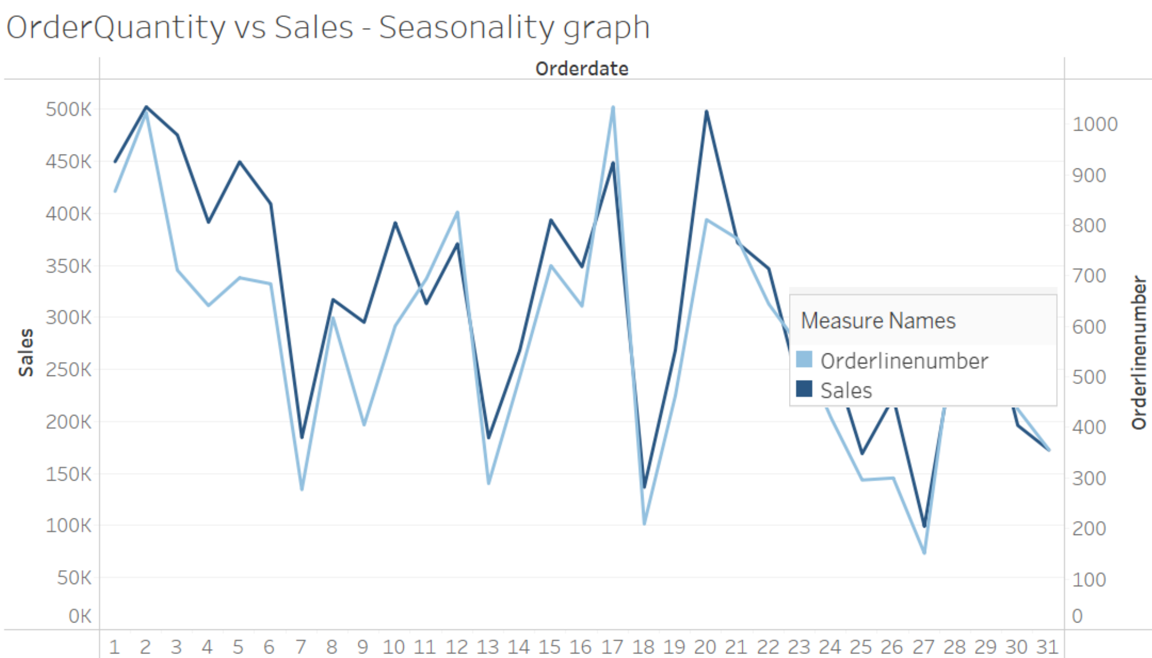


The bivariate analysis of product lines, quantity ordered, and sales reveals a strong positive correlation: product lines with higher prices tend to have higher ordered quantities, which in turn results in higher sales.

For example, Classic Cars, which have a high price, also see a large quantity of orders. This combination of high price and high quantity ordered leads to substantial sales, making Classic Cars a key revenue driver for the company. In contrast, Trains, which are priced lower, have a smaller quantity ordered, resulting in lower overall sales. This suggests that while higher-priced product lines attract more orders, lower-priced items may appeal to a more limited market, leading to lower sales volumes.

This analysis highlights how pricing strategies and product appeal play crucial roles in driving both the quantity ordered and overall sales, with higher-priced, high-demand products generating the most significant revenue.

Analysis of Daily Patterns in Sales and Order Line Numbers



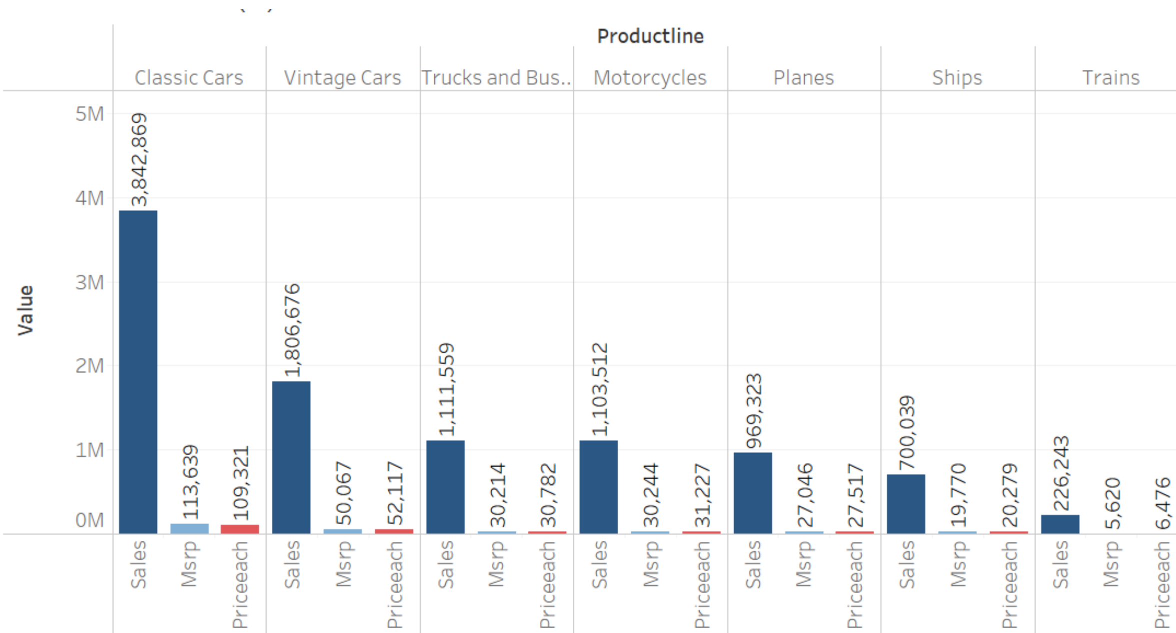
Consistent Daily Peaks: The analysis reveals that if there is a peak in ORDERLINENUMBER (quantity ordered) on a specific day of the month, there is a corresponding peak in SALES on the same day. This indicates a direct relationship between the number of orders and the total sales revenue.

Revenue Correlation: The fact that peaks in order quantity align with peaks in sales suggests that increased order volumes directly contribute to higher sales revenue. In other words, when more items are ordered, the total sales amount increases proportionally.

Demand Patterns: This consistent pattern also reflects that fluctuations in demand are mirrored in both the quantity of orders and the revenue generated. High demand days lead to both high order volumes and high sales, while low demand days result in lower order volumes and sales.

Sales Strategies: Recognizing these patterns allows the company to tailor promotions and marketing efforts to capitalize on peak days and manage resources effectively during lower demand periods.

Analysis of Product Line Performance: Sales, MSRP, and Market Trends

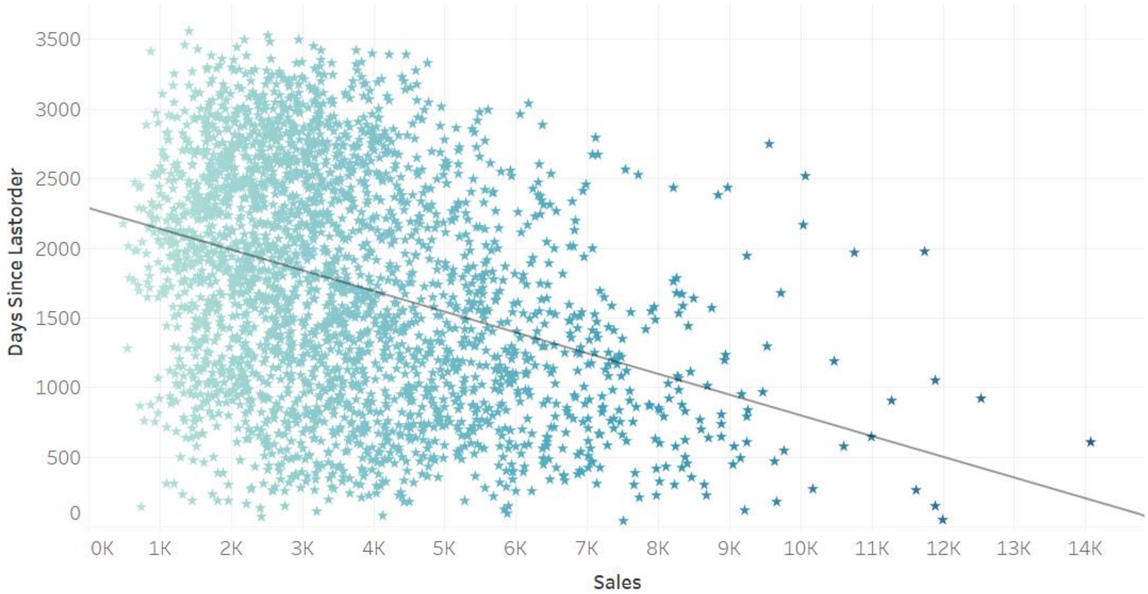


Price vs. Sales Relationship: There is a general trend where higher-priced product lines also have higher total sales. However, this relationship is not uniform across all categories, suggesting that factors such as market demand and product value influence sales more than price alone.

Market Demand: Classic Cars and Vintage Cars lead in sales, suggesting strong market demand and high perceived value in these segments. The pricing of these products closely aligns with their MSRP, indicating well-established markets and customer acceptance.

Premium Pricing: For several product lines, the actual selling price exceeds the MSRP, indicating a premium pricing strategy or high demand. This is particularly noticeable in Vintage Cars and Motorcycles, suggesting that buyers are willing to pay more than the suggested retail price.

Analysis of Recency and Sales



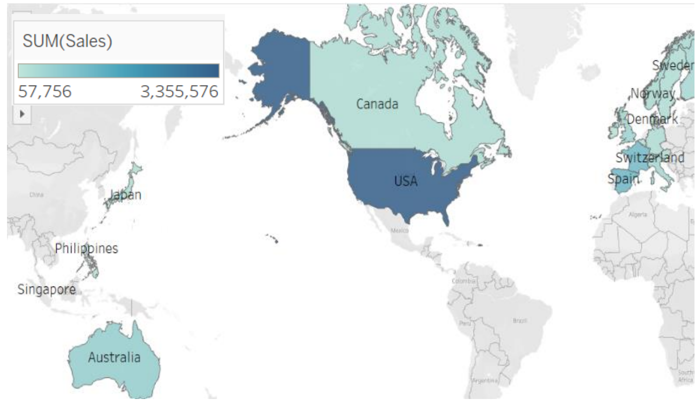
Recency measures how recently a customer has made a purchase. In this context, DAYS_SINCE_LASTORDER represents the number of days since the customer last placed an order. A high value in this metric indicates that a longer time has passed since the last order, suggesting that the customer is less engaged or has not purchased recently.

The observed negative correlation means that as DAYS_SINCE_LASTORDER increases (i.e., as the time since the last purchase grows longer), sales tend to decrease. This implies that customers who haven't ordered recently are less likely to contribute to current sales.

Insights: Customers who have not placed an order for an extended period are likely less engaged or have reduced interest in purchasing. This lower engagement results in fewer sales from these customers.

In RFM terms, Recency is a critical factor influencing sales. Customers with a recent order history are more likely to make additional purchases, whereas those with a longer time since their last order are less active and contribute less to sales.

Analysis of Country, City & Sales



The USA stands out as the dominant market, contributing the largest share of total sales. This suggests a robust market presence and a high level of consumer demand in the region. Spain and France also exhibit significant sales figures, indicating strong market performance.

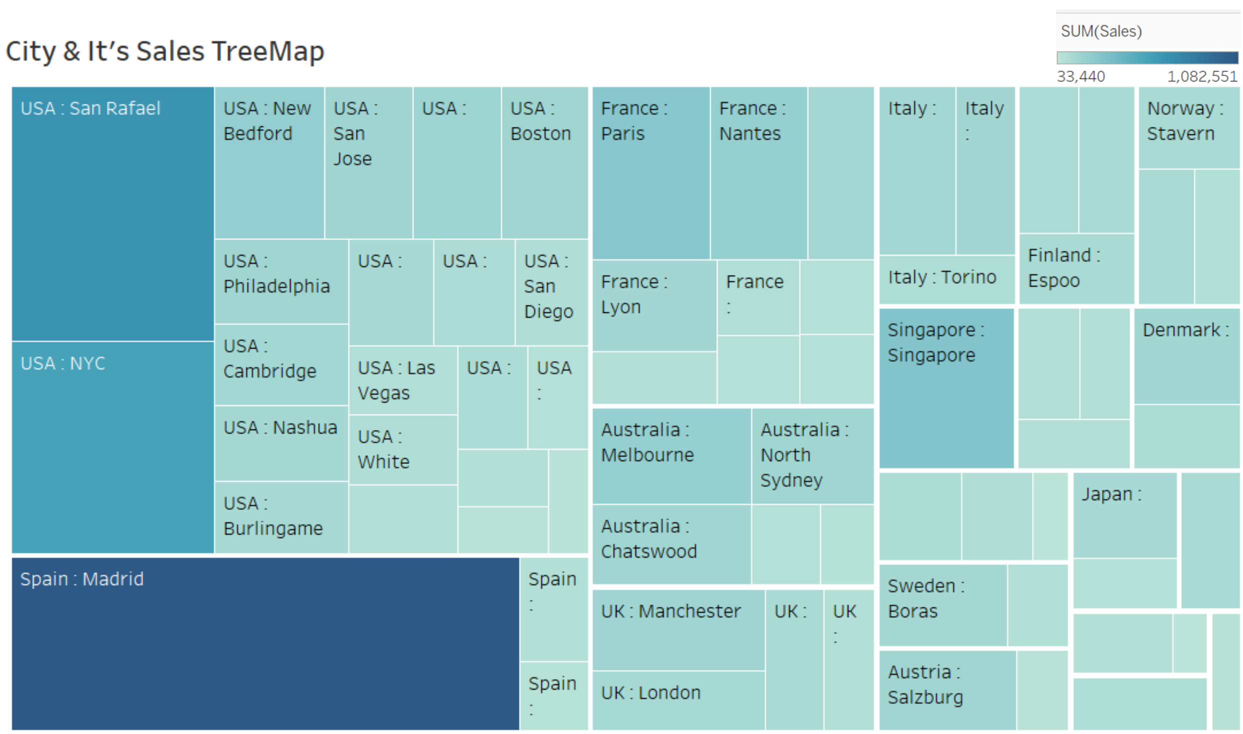
Countries like Italy and the UK also show substantial sales, reflecting their important roles in the company's global market strategy. The high sales in these regions point to successful market penetration and effective sales strategies. Conversely, countries such as Switzerland and the Philippines show comparatively lower sales figures. This may indicate either smaller market sizes or less penetration compared to higher-performing regions.

Overall, the significant sales figures in specific countries like the USA, Spain, and France underscore the importance of these markets to the company's overall revenue. This calls for a focused approach to maintain and grow these key markets, while also considering strategies to enhance performance in regions with lower sales.

Country & It's Sales HeatMap

	Productline						
Country	Classic Cars	Motorcycles	Planes	Ships	Trains	Trucks and B..	Vintage Cars
USA	1,267,891	457,496	322,753	195,290	69,254	381,612	661,281
Spain	476,165	74,635	89,986	124,460	43,370	177,557	229,515
France	388,951	226,390	108,156	66,487	27,341	116,982	176,610
Australia	193,086	89,969	74,854	4,160	1,681	77,319	189,555
UK	159,378	40,803	41,164	72,959	12,636	28,143	123,799
Italy	128,577	7,568	98,186	17,704	6,275	5,915	110,451
Finland	153,552	47,867	34,375	29,808	5,117	40,479	18,383
Norway	134,787	51,769	29,501		11,310	37,076	43,021
Singapore	132,890	4,176		14,156	13,279	89,028	34,960
Denmark	157,182		7,586	38,697	11,476	9,589	21,106
Canada	61,623	4,177	25,510	40,309		51,946	40,513
Germany	148,315	7,498	23,001	5,501	5,043	10,178	20,936
Sweden	69,088	15,567	8,900	30,916	3,808	47,931	33,804
Austria	101,459	26,048	17,860	9,025		20,473	27,197
Japan	47,271	26,536	49,177	18,860	3,524	13,349	29,450
Switzerland	117,714						
Belgium	20,137		5,625	31,708	9,017		41,926
Philippines	53,112	18,062	20,907				1,935
Ireland	31,689	4,953	11,784		3,113	3,983	2,234

The sales data reveals notable regional and product line variations. The USA leads with the highest total sales across all product lines, contributing significantly to the overall revenue with $3,355,575.69. Spain and France also show strong performances, with total sales reaching $1,215,686.92 and $1,110,916.52, respectively. Classic Cars are the top-selling product line overall, particularly in high-sales regions like the USA, Spain, and France. The data indicates that regions such as Australia, Italy, and the UK also exhibit substantial sales in Classic Cars, contributing significantly to their total sales figures. Conversely, countries like Switzerland and Belgium have lower overall sales, with Switzerland contributing $117,713.56 and Belgium $108,412.62. Product lines such as Motorcycles and Trains have less impact, especially in lower-sales regions. This analysis highlights the importance of focusing on high-performing regions and product lines, particularly Classic Cars, which dominate sales in major markets. For effective strategic planning, prioritizing efforts in high-sales regions and expanding product offerings where there's potential for growth could be beneficial.



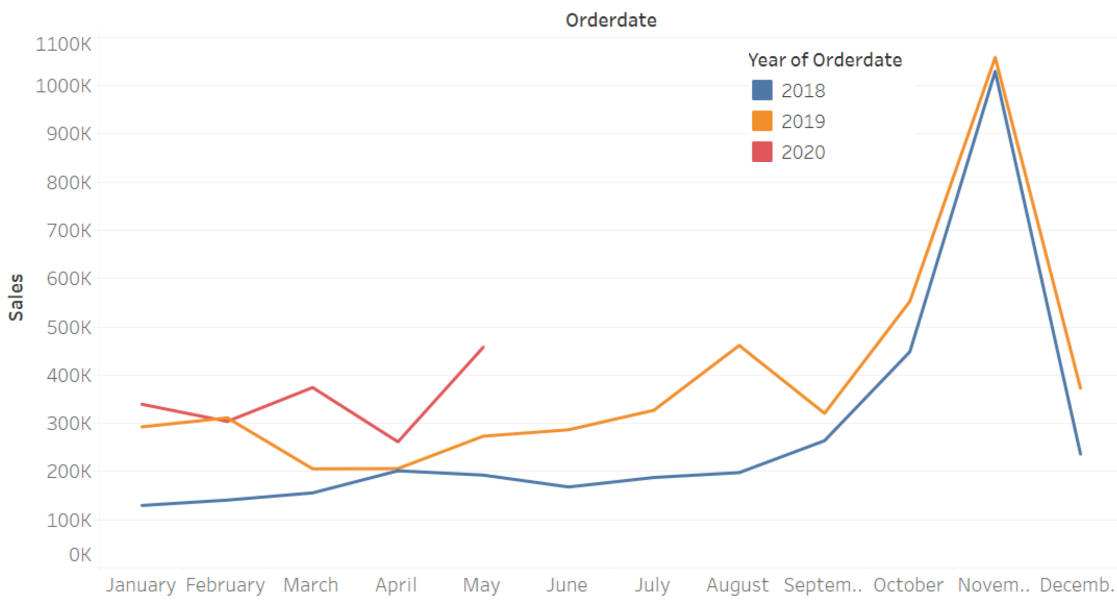
The USA stands out as the leading market, with notable contributions from cities like San Rafael, NYC, and New Bedford. San Rafael, in particular, has a substantial share, indicating it as a key sales hub. New York City and San Rafael are significant contributors, reflecting the strong demand in these areas.

Despite the USA leading the market overall, it's noteworthy that a specific city in Spain, Madrid, contributes more to sales than any individual city in the USA. Madrid's sales figures surpass those of major US cities like San Rafael, NYC, and San Jose, highlighting its exceptional performance. This suggests that Madrid is a critical market for the company, potentially outperforming key US cities in terms of revenue generation.

Australia, with Melbourne and North Sydney, and the UK, with Manchester and London, are notable for their contributions. These regions show strong market presence, with sales figures reflecting their important roles in the company's strategy.

Cities such as Singapore, with its total sales matching the country's overall sales figure, show high market engagement. Finland and Norway also exhibit strong sales, with cities like Helsinki and Oslo contributing significantly.

Weekly, Monthly, Quarterly, Yearly Trends in Sales



Monthly Trends:

- *January to March: Sales seem to start relatively low at the beginning of the year, with a slight increase around February and March. This could indicate a slow start, possibly due to post-holiday season effects, or lower demand during winter months.*

- *April to June: There's a noticeable dip in sales during April, followed by some fluctuations. The sales trend here could be influenced by specific events or seasonal factors, potentially related to the end of the first quarter or changes in consumer behavior as the weather warms.*
- *July to September: Sales show some recovery and stability during these months. This period may reflect steady or moderate demand, possibly influenced by summer activities or mid-year sales events.*
- *October to December: The final quarter sees a significant increase in sales, peaking sharply in November and December. This is typical of the holiday season, where sales often surge due to holiday shopping, end-of-year promotions, and increased consumer spending.*

Quarterly Trends:

- *Q1 (January to March): The first quarter shows a modest performance with sales fluctuating slightly but generally staying within a lower range. This might reflect a post-holiday lull and the beginning of a new fiscal year.*
- *Q2 (April to June): Sales appear to stabilize or slightly decline in the second quarter, with some recovery by the end of June. This might indicate a period of adjustment before entering the busier second half of the year.*
- *Q3 (July to September): The third quarter shows more consistency with moderate sales, which could be tied to summer activities and back-to-school shopping.*
- *Q4 (October to December): The fourth quarter is the strongest, with a clear upward trend culminating in the highest sales figures of the year. This aligns with major holiday shopping periods and end-of-year sales.*

Yearly Trends:

- *General Observations: Sales are highly seasonal, with the most significant increases occurring in the last quarter of the year. The beginning and middle of the year exhibit more modest sales, with some fluctuations likely due to smaller seasonal events or market changes.*

Summary: The graph clearly shows strong seasonality, especially towards the end of the year. This suggests that marketing, promotions, and stock management should focus heavily on the final quarter.

Customer Segmentation using RFM analysis (4 segments)

RFM Analysis is a customer segmentation technique used to identify and target high-value customers by analyzing their purchasing behavior. It stands for Recency, Frequency, and Monetary value, which are three critical metrics used to assess customer value and behavior. Here's a breakdown of each parameter used in RFM analysis:

Parameters Used in RFM Analysis:

Recency (R):

- *Definition: Measures how recently a customer has made a purchase.*
- *Calculation: Typically calculated as the number of days since the last purchase. In this case, ORDERDATE is used to calculate recency.*
- *Assumption: More recent purchases indicate a higher likelihood of the customer engaging with the business again.*

Frequency (F):

- *Definition: Measures how often a customer makes a purchase within a given time period.*
- *Calculation: Count of distinct ORDERNUMBER for each customer.*
- *Assumption: Customers who purchase more frequently are considered more engaged and valuable.*

Monetary Value (M):

- *Definition: Measures the total amount of money a customer has spent over a given period.*
- *Calculation: Total SALES for each customer.*
- *Assumption: Higher spending customers are more valuable to the business.*

Assumptions Made:

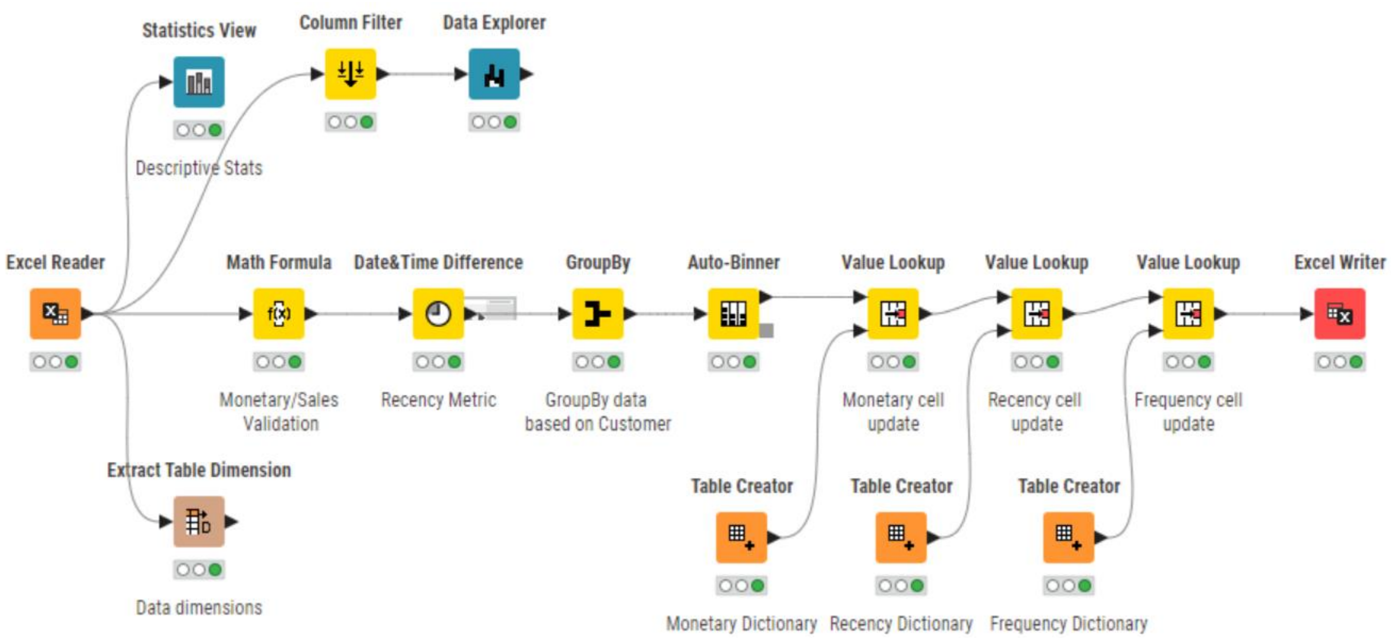
Data Accuracy: The analysis assumes that the ORDERDATE, ORDERNUMBER, and SALES data are accurate and reflect genuine transactions.

Recency Interpretation: More recent purchases are assumed to correlate with higher engagement and a likelihood of future purchases.

Frequency Interpretation: Higher purchase frequency is interpreted as higher customer engagement and loyalty.

Monetary Interpretation: Higher spending is assumed to indicate a higher-value customer.

KNIME workflow image



Columns in the Output Table:

1. *CUSTOMERID: Unique identifier for each customer.*
2. *Recency (Binned): The binned value for recency, which categorizes customers into different segments based on how recent their last purchase was. For example, bins will be "Very High", "High", "Medium", "Low" depending on the number of days since their last purchase.*

3. *Frequency (Binned): The binned value for frequency, categorizing customers based on how often they make purchases. Bins will be "Very High", "High", "Medium", "Low" based on the number of orders placed.*
4. *Monetary (Binned): The binned value for monetary, which categorizes customers based on their total spending. Bins will be "Very High", "High", "Medium", "Low" based on the total amount spent.*
5. *Recency: The actual recency value (e.g., number of days since last purchase).*
6. *Frequency: The actual frequency value (e.g., number of orders placed).*
7. *Monetary: The actual monetary value (e.g., total sales).*

Understanding the Output Table:

- *CUSTOMERID allows for identifying each customer and correlating their RFM metrics.*
- *Recency, Frequency, and Monetary provide the detailed values for each customer, which are useful for further analysis and insights.*
- *Recency (Binned), Frequency (Binned), and Monetary (Binned) help in categorizing customers into different segments, which can be used for targeted marketing strategies.*

Inferences from RFM Analysis

For a focused analysis, it is essential to target key drivers of revenue. Specifically, concentrating products that contribute to 80% of the total sales, which include Classic Cars, Vintage Cars, Trucks and Buses, and Motorcycles. These product lines are pivotal due to their significant impact on overall sales. Additionally, prioritizing analysis in countries with high sales performance: the USA, Spain, France, the UK, and Australia. By honing in on these high-revenue products and regions, we can better understand and leverage the primary factors driving your business's financial success.

RFM Classfication

Country: (Multiple values)
Productline: (Multiple values)

		Monetary HML			
Frequenc..	Recency ..	Very High	High	Moderate	Low
Very High	Very High	6			
	High	4	1		
	Moderate	3			
	Low	1			
High	Very High	1	2		
	High		1	1	
	Moderate	1			
	Low		1		
Moderate	Very High		1	2	1
	High		2	2	
	Moderate		2	4	1
	Low			1	
Low	Very High			2	1
	High				4
	Moderate			1	2
	Low			2	9

1. Best Customers | Criteria: *Customers in the "Very High" segment for all three RFM metrics (Recency, Frequency, and Monetary). These customers frequently purchase, have high monetary values, and have made recent purchases.*

   Recency, Frequency, Monetary: Very High

   ### Customer & Their Total Sales

Customername	
Euro Shopping Channel	726,116
Mini Gifts Distributors Ltd.	602,663
La Rochelle Gifts	111,574
The Sharp Gifts Warehouse	107,234
Souveniers And Things Co.	126,588
Reims Collectables	87,617

2. Customers on the Verge of Churning | *Criteria: Customers with high monetary value and frequency but have not ordered recently (lower Recency score).*

Recency: Moderate or Low

Monetary & Frequency: Very High or High (Based on available customers available post targeting key drivers of revenue)

Customer & Their Total Sales

Customername	
AV Stores, Co.	128,659
Corrida Auto Replicas, Ltd	120,615
Land of Toys Inc.	125,587
Marta's Replicas Co.	97,688
Online Diecast Creations Co.	131,685
Saveley & Henriot, Co.	111,557

3. Lost Customers | Criteria: *Customers with low scores in Recency, Frequency, and Monetary. They haven't ordered recently, purchase infrequently, and spend less.*

   Recency, Frequency, and Monetary: Low

Customer & Their Total Sales

Customername	
Auto Assoc. & Cie.	64,834
CAF Imports	24,883
Cambridge Collectables Co.	29,126
Daedalus Designs Imports	69,052
Double Decker Gift Stores, Ltd	21,780
Iberia Gift Imports, Corp.	52,278
Online Mini Collectables	52,505
Signal Collectibles Ltd.	50,219
West Coast Collectables Co.	38,964

4. Loyal Customers | Criteria: *Customers who show consistent engagement with recent orders, high frequency, and substantial monetary value.*

   Recency & Frequency: High or Very High

Customer & Their Total Sales

Customername	
Euro Shopping Channel	726,116
Mini Gifts Distributors Ltd.	602,663
Australian Collectors, Co.	161,117
Muscle Machine Inc	192,503
La Rochelle Gifts	111,574
The Sharp Gifts Warehouse	107,234
Anna's Decorations, Ltd	153,996
Souveniers And Things Co.	126,588
Reims Collectables	87,617
Technics Stores Inc.	85,653
Mini Creations Ltd.	56,164

Summary

By segmenting customers using the RFM analysis, we can identify:

- *Best Customers: Those who buy frequently, spend a lot, and have recently purchased.*
- *Customers on the Verge of Churning: High-value customers who haven't interacted recently.*
- *Lost Customers: Those with minimal engagement and low spend.*
- *Loyal Customers: Consistent buyers with high recent spend and order frequency.*

Tailoring marketing strategies to each of these segments can help retain high-value customers, re-engage those at risk, and optimize strategies to revive lost customers.

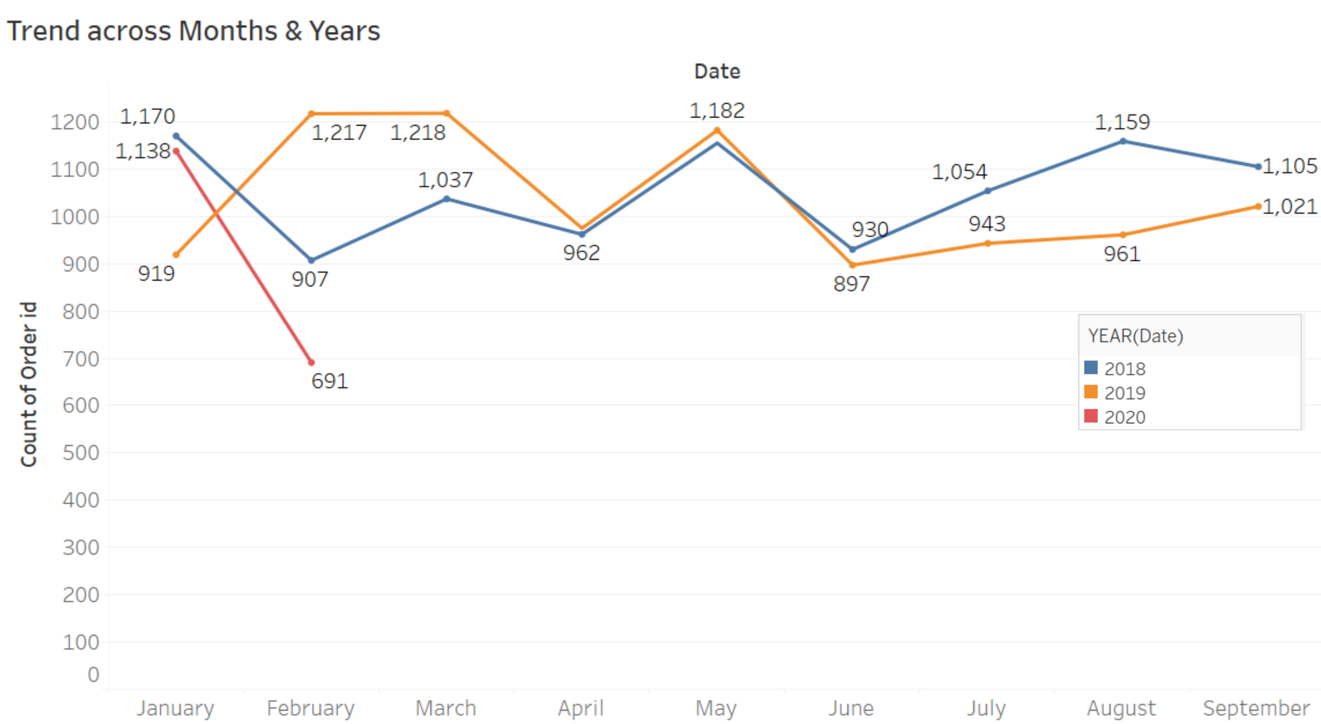
Part B : Market Basket Analysis

Objective: *To analyze the Point of Sale (POS) data of a grocery store to identify the most frequently purchased item sets, understand consumer buying patterns, and develop targeted combo offers and discounts that can increase store revenue and improve customer satisfaction.*

Background: *The grocery store has provided transactional data that includes various product purchases over a specified period. The data reveals individual transactions and product details but lacks insights into how frequently products are bought together and how these patterns can be leveraged to drive sales.*

Exploratory Data Analysis:

Trend across Months & Years:



The line chart displays the monthly trend in the count of orders (Order IDs) across three years: 2018, 2019, and 2020.

1. Overall Trend

- *2018:*
  - *The number of orders fluctuates throughout the year.*
  - *Peaks: February (1,217), May (1,182), and August (1,159).*
  - *Dips: April (907) and June (930).*
- *2019:*
  - *The year shows a more stable trend, with fewer fluctuations compared to 2018.*
  - *Peaks: March (1,218) and May (1,037).*
  - *Dips: February (919) and June (897).*
  - *Noticeable gradual increase from June to September.*
- *2020:*
  - *Steep drop from January (1,138) to March (691), indicating a significant reduction in orders during this period. This could be due to external factors, such as the COVID-19 pandemic.*

2. Yearly Comparison

- *2018 vs. 2019:2018 shows higher variability in order counts, with more significant peaks and troughs. In contrast, 2019 has a more consistent trend, with order counts generally increasing towards the end of the period shown.*
- *2020: The limited data shows a dramatic decrease in orders, which could be due to the impact of the pandemic, leading to lower consumer activity.*

3. Key Observations

- *February 2020: The count of orders drastically drops compared to previous years, likely reflecting the onset of the pandemic and its initial impact on consumer behavior.*
- *Stable Periods: Both 2018 and 2019 show relatively stable order counts from March to May, but with different peaks.*
- *Seasonal Effects: There may be seasonal effects influencing the order trends, such as holidays or other annual events. This is seen in the peaks around February and August in 2018 and 2019.*

4. Implications

- *Impact of External Factors: The significant drop in early 2020 suggests that external events, likely the pandemic, had a major effect on consumer behavior and order volume.*
- *Planning for Variability: The fluctuations in 2018 indicate that the store may need to prepare for periods of high demand followed by drops, while the steadier growth in 2019 suggests more predictable ordering patterns.*
- *Opportunities: Understanding the causes of these fluctuations (e.g., promotions, holidays, economic conditions) can help the store better plan inventory and marketing strategies.*

<u>Top Products Bought Together</u>

Top N Parameter

20

Top 20 Products with good support Bought together

Product (dataset!group1) \ Product	aluminu m foil	bagels	cereals	cheeses	coffee/t..	dinner rolls	dishwas hing liq..	eggs	ice cream	individua l meals	juice	laundry d etergent	lunch meat	milk	poultry	soap	soda	toilet paper	waffles	yogurt
aluminum foil	438																			
bagels	192	439																		
cereals	189	185	451																	
cheeses	193	191	199	445																
coffee/tea	181	173	193	191	432															
dinner rolls	189	177	195	191	181	443														
dishwashing li..	181	191	193	186	175	188	442													
eggs	179	191	199	193	181	198	195	444												
ice cream	201	182	189	204	188	184	188	190	454											
individual meals	185	188	182	183	174	170	181	175	185	428										
juice	196	185	178	171	178	179	174	188	190	177	429									
laundry deterg..	183	174	193	183	166	188	186	186	181	173	177	431								
lunch meat	183	180	192	183	190	185	198	181	198	202	184	179	450							
milk	183	184	198	178	173	178	194	192	188	165	175	189	183	433						
poultry	200	191	206	206	199	222	213	201	204	194	201	199	212	203	480					
soap	179	166	185	185	181	178	182	190	190	182	178	176	189	184	209	432				
soda	184	179	178	192	181	195	190	213	192	189	179	186	182	197	197	199	445			
toilet paper	182	186	189	191	178	171	174	186	182	174	196	174	187	186	195	187	185	431		
waffles	193	182	183	197	196	193	200	193	197	178	190	178	210	163	190	197	202	197	449	
yogurt	202	185	196	196	183	190	180	199	178	180	201	181	184	191	206	171	191	185	198	438

This heatmap visualizes the co-occurrence of the top 20 products frequently bought together in transactions. The intensity of the color (from light blue to dark blue) represents the frequency of these products being purchased together.

1. Top Products with High Co-occurrence

- Cereals and Cheeses (454 transactions): The highest co-occurrence in the dataset, indicating that when customers buy cereals, they often also buy cheese.

- Milk and Poultry (480 transactions): Another strong combination, showing that these items are frequently purchased together.
- Milk and Cheeses (445 transactions): This pairing is also common, suggesting that customers often add cheese to their cart when buying milk.
- Poultry and Cheeses (454 transactions): Indicates a significant relationship between these two items being bought together.

2. Other Notable Pairs

- Eggs and Milk (444 transactions): This is a classic combination, often purchased together in grocery shopping trips.
- Ice Cream and Individual Meals (428 transactions): Suggests that customers buying ready-to-eat meals also treat themselves to ice cream.
- Soap and Laundry Detergent (431 transactions): Both are household cleaning items, making them likely to be purchased together.

3. Implications for Combo Offers

- Meal Preparation Bundles: Given the frequent co-occurrence of cereals, cheeses, eggs, and milk, offering a "breakfast essentials" combo could attract customers looking for convenience and savings.
- Protein Packages: Poultry, milk, and cheese are commonly bought together. A "protein-rich" bundle could be marketed for health-conscious customers or families.
- Household Essentials: Products like soap and laundry detergent are often purchased together, suggesting a potential combo deal for household cleaning supplies.
- Indulgence Packs: The pairing of individual meals and ice cream points to a potential "meal and dessert" combo offer, catering to customers looking for a quick meal and a treat.

4. Strategic Recommendations

- Targeted Promotions: Promotions can be crafted based on these frequent pairs, offering discounts on one item if the other is purchased, thereby driving higher basket values.

- Cross-Promotion Opportunities: Use this data to strategically place these items near each other in the store or highlight them together in online shopping platforms to encourage combined purchases.
- Seasonal Bundles: Identify if any of these pairs have seasonal spikes (e.g., ice cream in summer) and create timely offers.

5. Further Analysis

- Expand Analysis to Other Products: While this heatmap focuses on the top 20 products, expanding the analysis to include more products might reveal additional valuable combinations.
- Customer Segmentation: Analyzing these combinations across different customer segments (e.g., families, singles) could lead to more personalized offers.
- Trend Analysis: Examining these combinations over time can help identify trends or shifts in consumer behavior, allowing for dynamic adjustment of combo offers.

Use of Market Basket Analysis (Association Rules)

Market Basket Analysis (MBA) is a powerful data mining technique used to understand the co-occurrence of products in transactions. Association rules derived from MBA help in uncovering patterns of products that are frequently bought together. Here's how MBA can be applied effectively in a grocery store context:

1. <u>Identifying Product Associations</u>

Objective: Discover which products are commonly purchased together, revealing customer buying patterns.

Application:

- *Frequent Itemset: Determine the combinations of products that appear together most often in transactions.*
- *Association Rules: Generate rules that indicate relationships between products. For example, a rule might state that "If a customer buys bread, they are likely to buy butter."*

Benefits:

- *Targeted Promotions: Create special offers or discounts on product combinations that are frequently purchased together.*
- *Cross-Selling Opportunities: Increase sales by suggesting complementary products during the checkout process.*

2. Enhancing Store Layout and Product Placement

Objective: Optimize the store layout based on product associations to increase the likelihood of additional purchases.

Application:

- *Product Placement: Place frequently co-purchased items near each other to encourage impulse buys. For example, place cheese next to bread and milk if they are often bought together.*
- *Shelf Organization: Use insights from association rules to design store shelves that align with customer purchasing behavior.*

Benefits:

- *Improved Shopping Experience: Make it easier for customers to find related products, enhancing their shopping experience.*
- *Increased Sales: Increase the likelihood of additional purchases by strategically positioning products.*

3. Personalized Marketing and Recommendations

*Objective**:** Tailor marketing efforts to individual customer preferences and purchasing habits.*

Application:

- *Targeted Offers: Use association rules to create personalized offers for customers based on their previous purchases. For example, offer a discount on pasta to customers who frequently buy tomato sauce.*

- *Recommendation Systems: Implement recommendation engines on e-commerce platforms to suggest products based on past purchases and co-occurrence patterns.*

Benefits:

- *Increased Customer Loyalty: Offer personalized deals that resonate with individual preferences, fostering customer loyalty.*
- *Higher Conversion Rates: Improve the effectiveness of marketing campaigns by targeting customers with relevant product recommendations.*

4. <u>Optimizing Inventory Management</u>

Objective: Manage inventory more effectively by understanding product demand patterns.

Application:

- *Stock Levels: Adjust inventory levels based on the frequency of product combinations. Ensure that items frequently bought together are stocked in adequate quantities.*
- *Promotional Planning: Plan inventory for promotional periods based on the predicted demand for popular product combinations.*

Benefits:

- *Reduced Stockouts: Prevent stockouts of products that are frequently bought together by ensuring sufficient inventory levels.*
- *Efficient Restocking: Optimize restocking processes based on product demand patterns.*

5. <u>Enhancing Customer Experience</u>

*Objective**:** Improve the overall customer shopping experience by understanding and anticipating their needs.*

Application:

- *Product Bundles: Create attractive product bundles based on commonly bought together items, offering convenience and savings.*

- *Customized Offers: Use insights from association rules to offer personalized discounts and deals that cater to individual preferences.*

Benefits:

- *Increased Satisfaction: Provide a more personalized and satisfying shopping experience for customers.*
- *Enhanced Value: Offer valuable and relevant promotions that resonate with customers' purchasing habits.*

Association Rules and Their Relevance

Association Rules are a fundamental technique in data mining and market basket analysis used to identify relationships between items in transactional data. These rules help uncover patterns and associations within the data, revealing how frequently items are purchased together and providing insights into customer behavior.

<u>What Are Association Rules?</u>

Association rules are of the form {Antecedent} => {Consequent}, where:

- **Antecedent:** The items that are found together frequently.
- **Consequent:** The item that is likely to be purchased given the presence of the antecedent items.

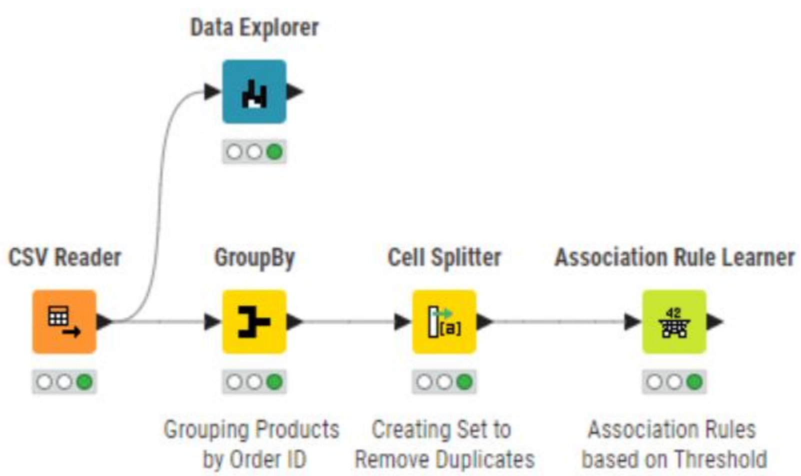
<u>Relevance of Association Rules in This Case</u>

In the context of the grocery store's Point of Sale (POS) data, association rules offer valuable insights into purchasing patterns and consumer behavior.:

1. ***Understanding Customer Buying Patterns:*** *Association rules help identify which products are often bought together. For example, if milk and bread are frequently purchased together, the store can infer that customers often combine these items in their shopping baskets.*
2. ***Optimizing Product Placement:*** *By analyzing these patterns, the store can strategically place products that are commonly bought together in close proximity. This can lead to increased impulse purchases and improved shopping efficiency for customers.*

3. ***Creating Targeted Promotions:*** *The insights from association rules allow for the development of targeted promotional offers. For instance, if the rule {Cereal} => {Milk} shows high confidence, the store can create a discount offer for customers who buy both items together.*
4. ***Enhancing Marketing Strategies:*** *Personalized marketing campaigns can be designed based on the association rules. For example, targeted email offers can be sent to customers suggesting products they frequently buy together or introducing new product combinations.*
5. ***Improving Inventory Management:*** *Understanding which items are often purchased together can aid in managing inventory more effectively. The store can ensure that related products are stocked appropriately to meet demand and avoid stockouts.*
6. ***Boosting Revenue:*** *By leveraging these insights to create combo offers and bundles, the store can increase the average transaction value. Well-placed promotions and product pairings can drive higher sales and enhance overall revenue.*
7. ***Enhancing Customer Experience:*** *Providing relevant recommendations and convenient product pairings based on customer purchasing behavior can improve the shopping experience, making it easier and more enjoyable for customers to find and buy what they need.*

KNIME Workflow



Threshold Values for Support and Confidence in Association Rules

Association Rules are evaluated based on various metrics, with ***support*** *and* ***confidence*** *being two of the most critical. Setting appropriate threshold values for these metrics helps in filtering out relevant rules and focusing on those that provide actionable insights.*

1. Support : ***Support*** *measures the proportion of transactions in the dataset that contain a particular itemset. It indicates how frequently an itemset appears in the overall dataset.*

Threshold Value Set: 0.05

- **Interpretation:** A support threshold of 0.05 means that only those itemsets appearing in at least 5% of the total transactions will be considered. This ensures that the rules generated are based on itemsets that are sufficiently frequent to be of interest.

Relevance:

- **Filtering Out Rare Items:** By setting the support threshold to 0.05, we exclude infrequent itemsets that might not provide significant insights or might be considered noise in the data. This helps in focusing on itemsets that have a substantial presence in transactions.
- **Ensuring Practicality:** This threshold balances between capturing relevant patterns and avoiding the inclusion of itemsets that appear too rarely to be actionable or useful for business decisions.

2. Confidence : ***Confidence*** *measures the likelihood that the consequent of the rule is purchased when the antecedent is present. It indicates the reliability of the rule.*

Threshold Value Set: 0.6

- **Interpretation:** A confidence threshold of 0.6 means that only rules where the consequent is purchased in at least 60% of the transactions where the antecedent is present will be considered. This helps in ensuring that the rules are not only frequent but also reliably associated.

Relevance:

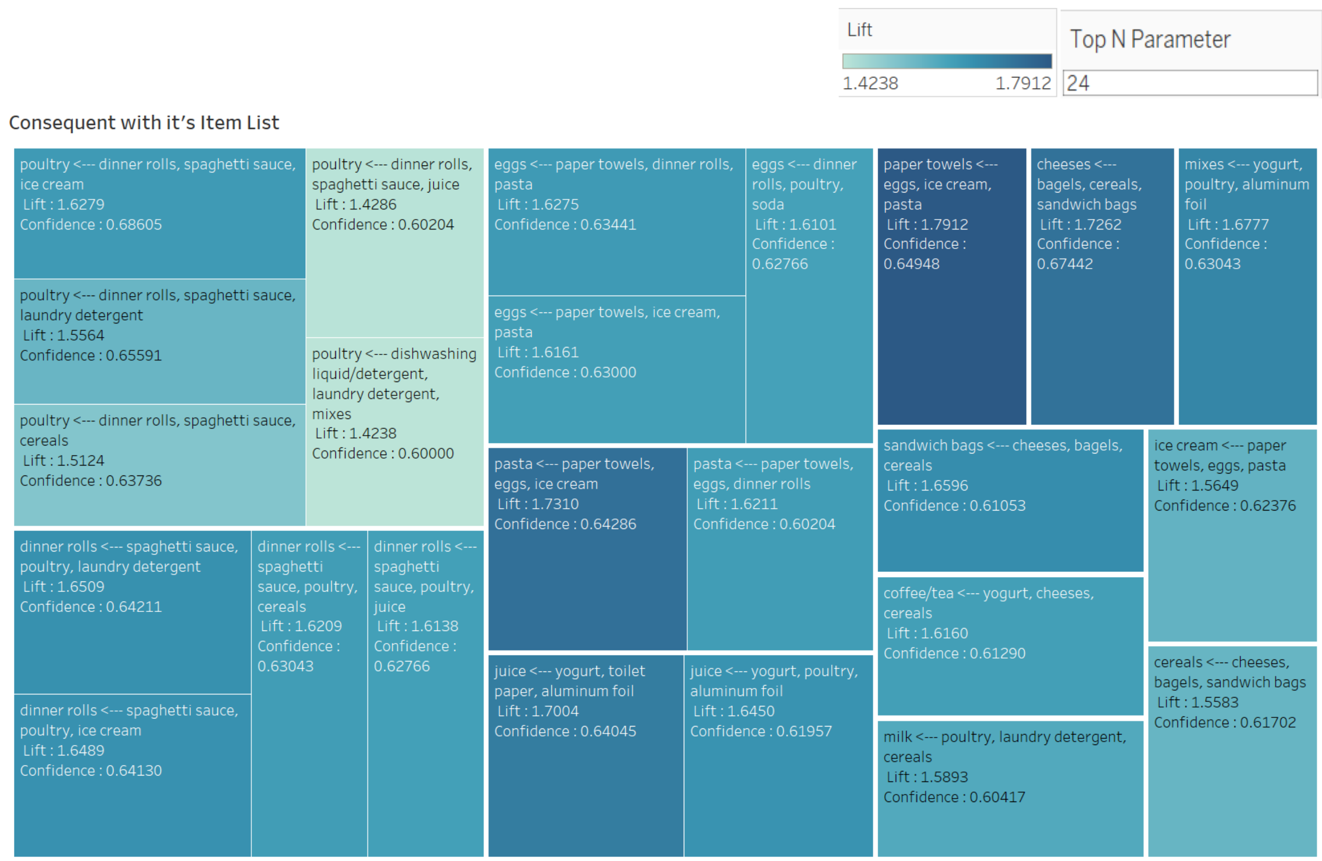
- **Ensuring Rule Reliability:** Setting a confidence threshold of 0.6 ensures that the rules you generate are not only frequent but also have a strong predictive power. This helps in identifying rules that are more likely to be true and actionable.
- **Avoiding Weak Associations:** Rules with lower confidence might indicate weak or unreliable associations that do not offer substantial value for marketing strategies or inventory management.

Association Rules Identified

The association rules identified from the Market Basket Analysis provide valuable insights into customer purchasing patterns. Below is a detailed analysis of these rules based on their support, confidence, and lift metrics:

Support	Confidence	Lift		Consequent	implies	Item List
0.050044	0.640449438	1.700400723		juice	<---	yogurt, toilet paper, aluminum foil
0.050044	0.619565217	1.644952873		juice	<---	yogurt, poultry, aluminum foil
0.050044	0.612903226	1.615964755		coffee/tea	<---	yogurt, cheeses, cereals
0.050044	0.6	1.42375		poultry	<---	dishwashing liquid/detergent, laundry detergent, mixes
0.050922	0.630434783	1.677722471		mixes	<---	yogurt, poultry, aluminum foil
0.050922	0.610526316	1.659640749		sandwich bags	<---	cheeses, bagels, cereals
0.050922	0.674418605	1.726208518		cheeses	<---	bagels, cereals, sandwich bags
0.050922	0.617021277	1.55828655		cereals	<---	cheeses, bagels, sandwich bags
0.050922	0.630434783	1.620914712		dinner rolls	<---	spaghetti sauce, poultry, cereals
0.050922	0.637362637	1.512408425		poultry	<---	dinner rolls, spaghetti sauce, cereals
0.050922	0.604166667	1.589251347		milk	<---	poultry, laundry detergent, cereals
0.0518	0.627659574	1.610144719		eggs	<---	dinner rolls, poultry, soda
0.0518	0.641304348	1.648861517		dinner rolls	<---	spaghetti sauce, poultry, ice cream
0.0518	0.686046512	1.627931202		poultry	<---	dinner rolls, spaghetti sauce, ice cream
0.0518	0.627659574	1.613779357		dinner rolls	<---	spaghetti sauce, poultry, juice
0.0518	0.602040816	1.428592687		poultry	<---	dinner rolls, spaghetti sauce, juice
0.0518	0.634408602	1.627458103		eggs	<---	paper towels, dinner rolls, pasta
0.0518	0.602040816	1.621098085		pasta	<---	paper towels, eggs, dinner rolls
0.053556	0.642105263	1.650920756		dinner rolls	<---	spaghetti sauce, poultry, laundry detergent
0.053556	0.655913978	1.556429211		poultry	<---	dinner rolls, spaghetti sauce, laundry detergent
0.055312	0.623762376	1.564901644		ice cream	<---	paper towels, eggs, pasta
0.055312	0.63	1.616148649		eggs	<---	paper towels, ice cream, pasta
0.055312	0.642857143	1.73100304		pasta	<---	paper towels, eggs, ice cream
0.055312	0.649484536	1.79119343		paper towels	<---	eggs, ice cream, pasta

- Support**:** All rules have a support value around 0.05, indicating that these itemsets are present in at least 5% of the transactions. This threshold ensures that the rules are based on sufficiently frequent itemsets.
- Confidence**:** The confidence values range from approximately 0.60 to 0.69, showing that the rules have a high probability of the consequent item being purchased given the antecedent items.
- Lift**:** Lift values vary, with some rules indicating strong associations (lift > 1.6). Lift values above 1 suggest that the items are bought together more frequently than would be expected by chance.



*In your Market Basket Analysis, the **support** values around 0.05 indicate that the itemsets are present in about 5% of transactions, ensuring they are of moderate frequency. **Confidence** values between 0.60 and 0.69 show a strong likelihood that the consequent items are bought when the antecedent is present, reflecting reliable associations. **Lift** values exceeding 1, such as 1.73, suggest that these item combinations occur together more frequently than expected by chance, highlighting strong and valuable associations. These metrics help in identifying significant item relationships for targeted promotions and effective product bundling.*

Recommendations for Lucrative Offers Based on Association Rules

1. *Breakfast Essentials Combo:*

- **Items:** Coffee/Tea, Yogurt, Cheeses, Cereals
- **Offer:** Buy any 2 items from the combo and get 25% off on the third item.
- **Rationale:** Coffee/Tea is frequently purchased with yogurt, cheeses, and cereals. This combo caters to breakfast needs and encourages customers to buy complementary items together.

2. *Dinner Bundle:*

- **Items:** Dinner Rolls, Spaghetti Sauce, Poultry, Ice Cream
- **Offer:** Buy Dinner Rolls and Spaghetti Sauce together, and get a 50% discount on Poultry or Ice Cream.
- **Rationale:** Dinner Rolls are often bought with Spaghetti Sauce, Poultry, and Ice Cream. This bundle provides a complete meal solution, driving higher sales of related items.

3. *Pasta & Household Essentials Pack:*

- **Items:** Pasta, Paper Towels, Eggs, Ice Cream
- **Offer:** Buy Pasta and Paper Towels together and receive a free pack of Eggs or Ice Cream.
- **Rationale:** Pasta is commonly purchased with Paper Towels, Eggs, and Ice Cream. This offer targets multiple household needs in one purchase, increasing overall basket size.

4. *Snack and Clean-Up Deal:*

- **Items:** Juice, Yogurt, Toilet Paper, Aluminum Foil
- **Offer:** Buy Juice and any 2 items from Yogurt, Toilet Paper, or Aluminum Foil and get 20% off the total purchase.
- **Rationale:** Juice is frequently bought with Yogurt, Toilet Paper, and Aluminum Foil. This offer encourages customers to purchase related snack and household items together.

5. *Ice Cream Delight Pack:*

- **Items:** Ice Cream, Paper Towels, Eggs, Pasta

- **Offer:** Purchase Ice Cream and receive a 30% discount on Paper Towels, Eggs, and Pasta.
- **Rationale:** Ice Cream is bought with Paper Towels, Eggs, and Pasta. This offer promotes a combination of snacks and household essentials, enhancing cross-selling opportunities.

6. *Eggs & Dairy Special:*

- **Items:** Eggs, Milk, Cheese
- **Offer:** Buy Eggs and Milk, and get a free pack of Cheese.
- **Rationale:** Eggs are frequently purchased with Milk and Cheese. This special offer encourages customers to buy all three essential dairy products together.

7. *Poultry Pack:*

- **Items:** Poultry, Dinner Rolls, Spaghetti Sauce
- **Offer:** Buy Poultry and Dinner Rolls together and get a 30% discount on Spaghetti Sauce.
- **Rationale:** Poultry is often bought with Dinner Rolls and Spaghetti Sauce. This offer helps in upselling complementary items for a complete meal.

These recommendations are designed to leverage common purchasing patterns identified in the association rules, driving higher sales through targeted promotions and bundle offers.