

# Marketing and Retail Analysis (Part A)

**Amneh Ghanem**

$\pi$

# Content/Agenda

- › Executive Summary
- › Business Problem Overview
- › Data Overview
- › Exploratory Data Analysis
- › Inference Summary from EDA
- › Customer Segmentation using RFM
- › Inferences from RFM Analysis

# Executive Summary

- › The sales data reveals a range from \$482.13 to \$14,082.80 with a mean value of \$3,553.05, indicating variability in order sizes or purchase volumes.
- › The Days since last order varies significantly from 42 days to 3,562 days, suggesting varying levels of customer engagement and potentially identifying opportunities for re-engagement strategies.
- › The bar chart displaying mean sales by product line shows that Classic Cars and Motorcycles have higher average sales compared to other categories such as Planes and Ships, indicating stronger market performance or customer preference in these segments.
- › The RFM analysis identified key segments such as Best Customers, Loyal Customers, and Lost Customers, with detailed insights into their purchasing behaviors.
- › High-value customers like "Mini Gifts Distributors Ltd." and "Muscle Machine Inc." were noted for their significant contributions to sales, highlighting them as critical assets for targeted marketing and retention strategies.

# Executive Summary

- › The **RFM** analysis has effectively segmented your customer base into distinct groups based on their purchase behavior—Recency, Frequency, and Monetary value. Such classifications help in tailoring specific marketing strategies to enhance customer engagement and retention.
- › **Best Customers:** Customers like "Mini Gifts Distributors Ltd." and "Muscle Machine Inc" have been identified as Best Customers due to their high frequency of purchases, recent interactions, and significant monetary contributions. These customers are crucial for sustained business growth and profitability.
- › **Loyal Customers:** "Dragon Souvenirs" and "Muscle Machine Inc" exemplify Loyal Customers who engage frequently and have recent purchase activities. These customers provide a stable revenue stream and demonstrate strong brand loyalty.
- › **Lost Customers:** High-value customers such as "Alpha Cognac" and "Iberia Gift Imports" have been classified as Lost due to their lack of recent engagement despite historically high spending. Re-engaging these customers could potentially recover significant lost revenue.

# Recommendations

- › The data underscores the need for differentiated engagement strategies. For example, reactivating "Lost Customers" who have shown high past value could be beneficial.
- › Insights into product line performance could guide inventory and marketing focus, especially on high-performing lines like Classic Cars.
- › Leveraging RFM segmentation helps tailor marketing strategies to customer behaviors, potentially increasing sales by focusing on high-value segments and re-engaging fewer active ones.
- › Develop specialized re-engagement campaigns for Lost Customers. Offer incentives such as trade-in discounts, loyalty bonuses for returning customers, or exclusive access to new product launches.
- › Implement a system to gather feedback from customers who have drifted away to understand their reasons for disengagement and address these issues directly in future strategies.

# Recommendations

- › The RFM analysis provides valuable insights into customer behavior and highlights specific areas for strategic focus.
- › By addressing the unique needs and behaviors of each customer segment, automobile manufacturing company can enhance customer satisfaction, increase loyalty, and drive revenue growth.
- › Ongoing analysis and adaptation of strategies based on customer data will be essential for maintaining competitive advantage and achieving long-term success.
- › Focus on High-Performing Segments Data indicates strong sales in specific segments like Classic Cars. Increase investment in these areas through R&D for new features, improved quality, and enhanced designs that appeal to current market trends.
- › Diversify Offerings Consider expanding into underperforming but strategic segments with innovative products. This could involve electric vehicles, hybrids, or special edition models that cater to emerging consumer demands and environmental standards.

# Business Problem Overview

- › An automobile parts manufacturing company has accumulated transaction data over a three-year period but lacks the data science expertise to analyze it. As an external consultant, you are brought in to mine this data for valuable insights.
- › main objective is to deeply analyze the transaction data to uncover underlying patterns in customer behavior. These insights will help the company understand the dynamics of their market, identify different customer segments, and tailor marketing efforts accordingly.

# Data Overview

## › Key Variables:

- › **ORDERNUMBER**: Identifier for each order, ranges from 10,100 to 10,425.
- › **QUANTITYORDERED**: Number of items ordered per transaction, varies from 6 to 97.
- › **PRICEEACH**: Price of each item, ranging from \$26.88 to \$252.87.
- › **ORDERLINENUMBER**: Line number of the order in the transaction, ranges from 1 to 18.
- › **SALES**: Total sales per order.
- › **ORDERDATE**: Dates of order placements, spanning over three years.
- › **DAYS\_SINCE\_LASTORDER**: Time elapsed since the last order was placed.
- › **STATUS**: Status of the order, with 'Shipped' being the most frequent.
- › **PRODUCTLINE**: Category of the product.
- › **MSRP**: Manufacturer's Suggested Retail Price.
- › **CUSTOMERNAME, PHONE, ADDRESSLINE1, CITY, POSTALCODE, COUNTRY, CONTACTLASTNAME, CONTACTFIRSTNAME**: Customer information details.
- › **DEALSIZE**: Size of the deal categorized as Small, Medium, or Large.



# Data Overview

Column <span>↕</span>	Exclude Column	Minimum <span>↕</span>	Maximum <span>↕</span>	Mean <span>↕</span>	Standard Deviation <span>↕</span>	Variance <span>↕</span>	Skewness <span>↕</span>
<span>+</span> ORDERNUMBER	<input type="checkbox"/>	10100	10425	10259.762	91.878	8441.479	-0.007
<span>+</span> QUANTITYORDERED	<input type="checkbox"/>	6	97	35.103	9.762	95.299	0.369
<span>+</span> PRICEEACH	<input type="checkbox"/>	26.880	252.870	101.099	42.043	1767.576	0.697
<span>+</span> ORDERLINENUMBER	<input type="checkbox"/>	1	18	6.491	4.231	17.897	0.575
<span>+</span> SALES	<input type="checkbox"/>	482.130	14082.800	3553.048	1838.954	3381751.448	1.156
<span>+</span> DAYS_SINCE_LASTORDER	<input type="checkbox"/>	42	3562	1757.086	819.281	671220.663	-0.003
<span>+</span> MSRP	<input type="checkbox"/>	33	214	100.692	40.115	1609.197	0.576

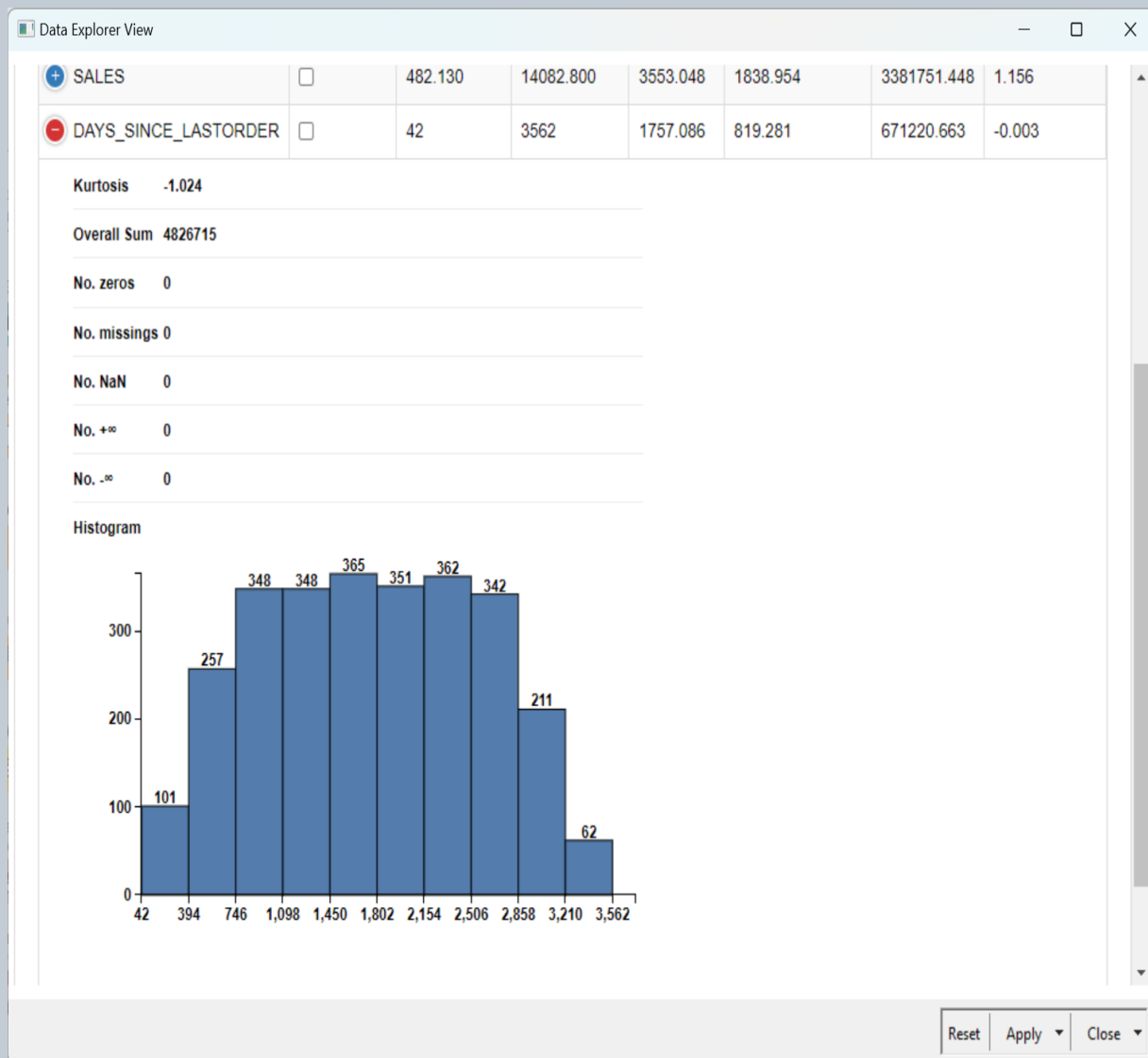
# Data Overview

$\pi$

- › **Pricing and Sales:** Both the PRICEEACH and SALES fields show right-skewed distributions, which is typical in sales data where a smaller number of high-value transactions raise the average price and total sales figures.
- › **Order Frequency:** The skewness in DAYS\_SINCE\_LASTORDER being close to zero suggests that the frequency of orders does not tend to drift towards longer or shorter intervals markedly; however, the large range indicates variability in customer re-engagement.
- › **Product Pricing (MSRP):** The moderate right skewness in MSRP suggests that most products are priced lower with fewer high-priced items in the catalog.
- › **Order Volume:** Distribution of order numbers is fairly uniform, suggesting consistent transaction recording.

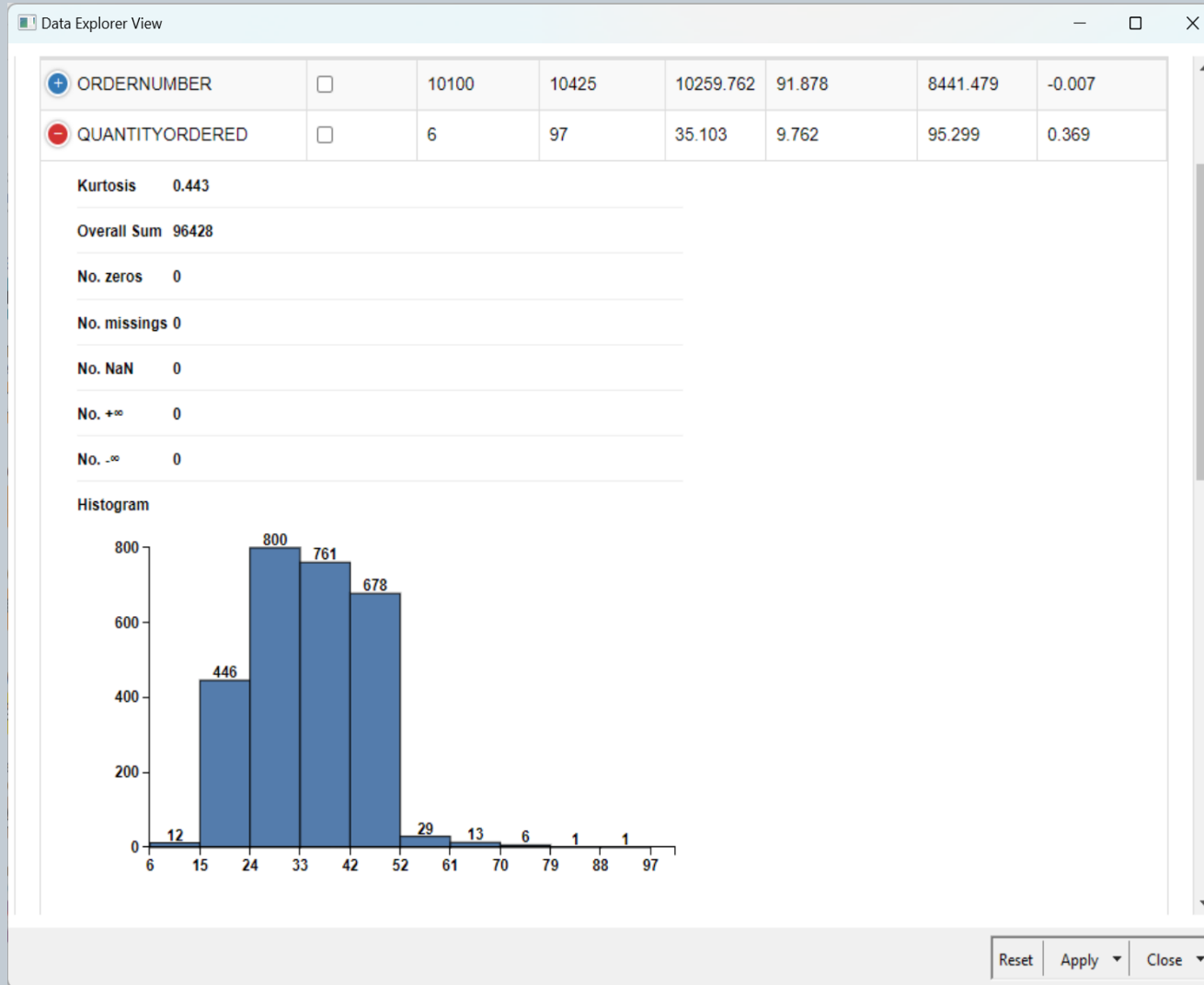
# Exploratory Data Analysis

Univariate Analysis



# EXPLORATORY DATA ANALYSIS

- **Most Active Reordering Periods:**  
The histogram indicates that reordering occurs consistently within a range of about 42 to around 2500 days. Within this range, the intervals don't show drastic variations in frequency, suggesting a steady level of customer re-engagement over time.
- **Long Gaps in Reordering:**  
Beyond 2500 days, there is a noticeable decline in the number of customers placing orders again, which could suggest challenges in maintaining long-term customer engagement or issues with tracking very old customer activity.



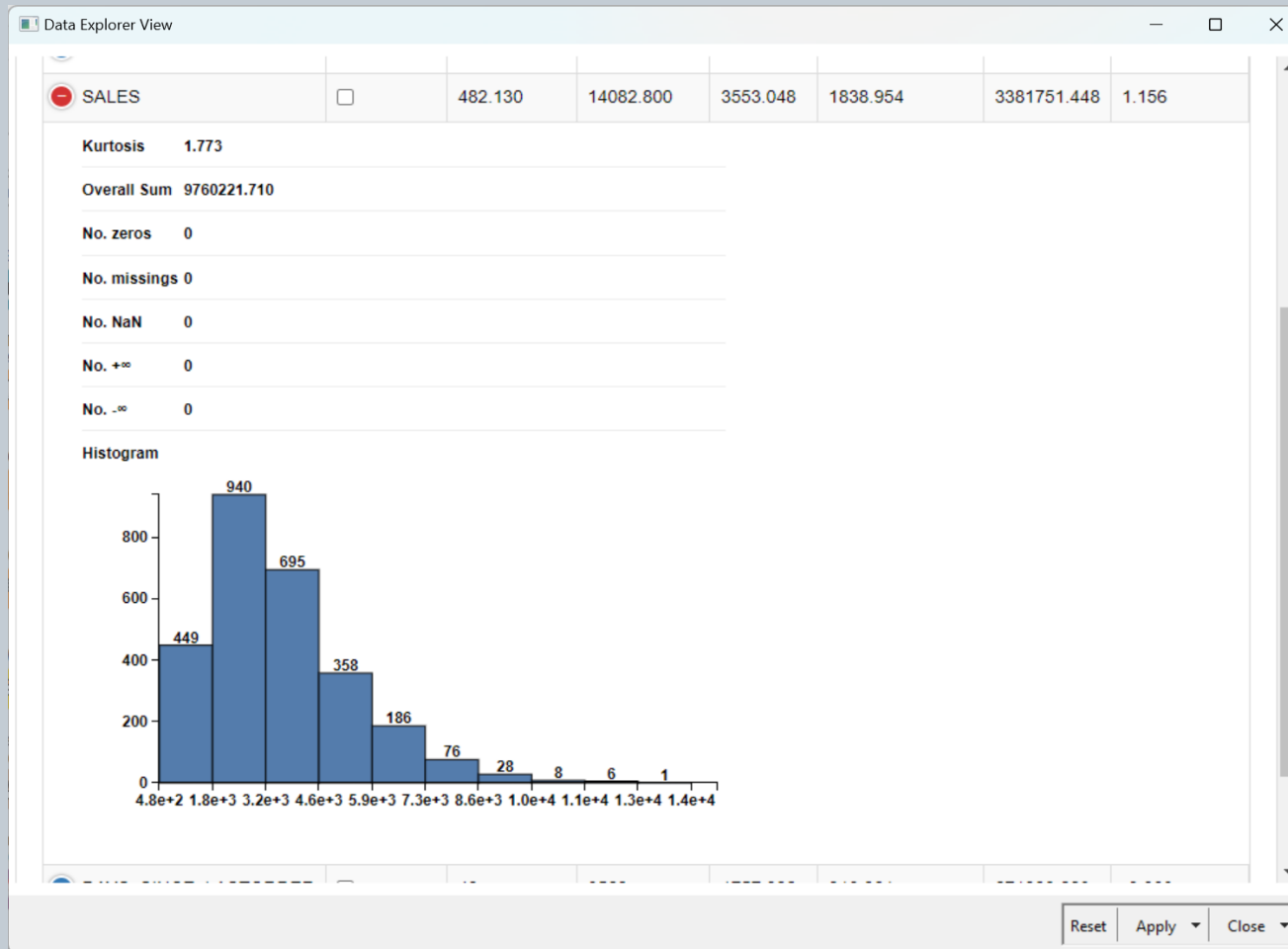
# EXPLORATORY DATA ANALYSIS

**Major Concentrations:** The histogram shows peak frequencies at specific quantity points — particularly at around 24, 33, and 42 units per order. These may be common set quantities for certain products or promotional deals.

**Less Common Higher Quantities:** Quantities nearing the maximum (97 units) are very rare, with only one order hitting this upper limit, suggesting such large orders are atypical.

**Overall Trends:** There is a gradual decrease in frequency as the quantity increases, which is typical for many retail or wholesale distributions where bulk purchases are less frequent.

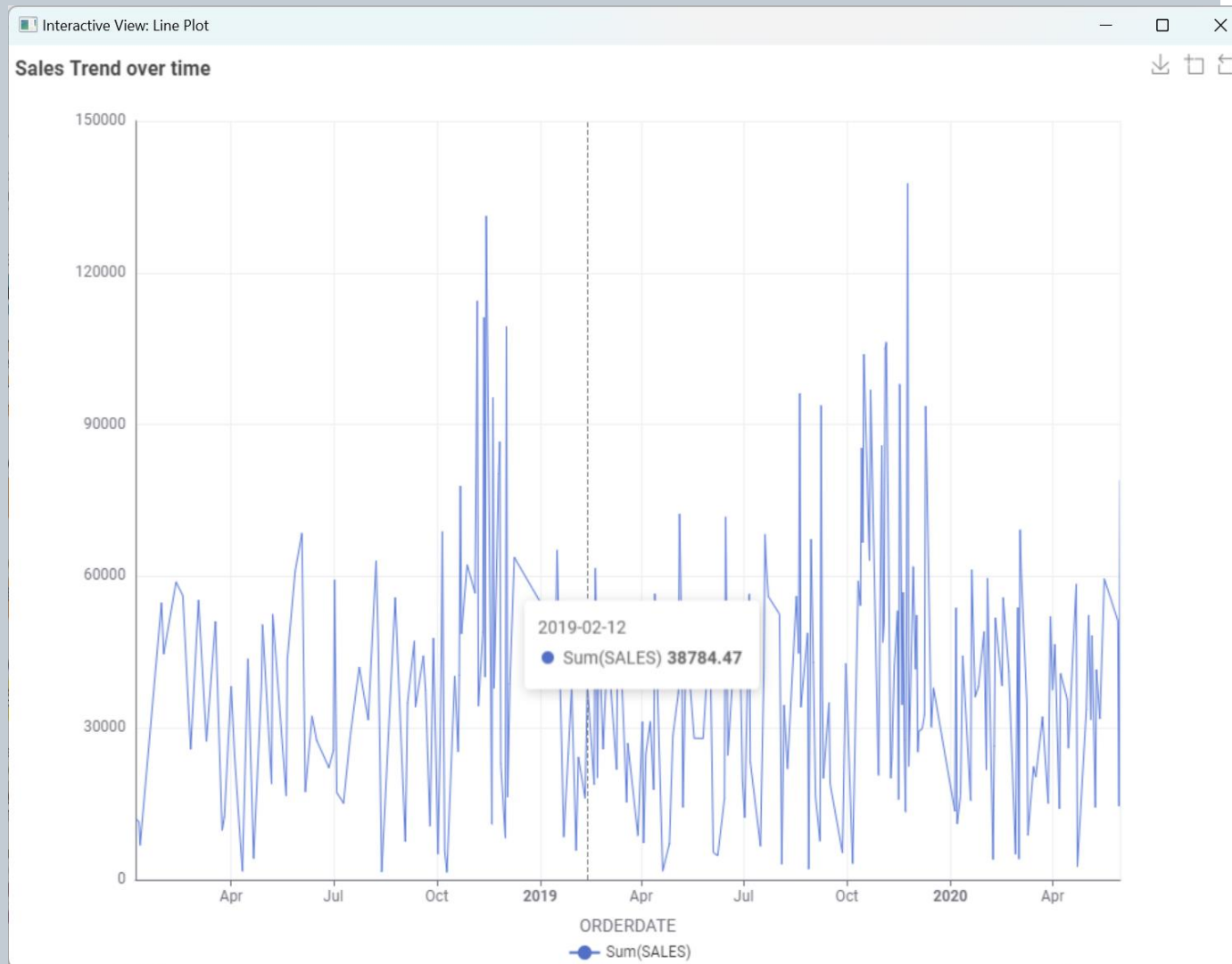
# EXPLORATORY DATA ANALYSIS



- **Sales:** There's a substantial range in sales amounts
- The high standard deviation and variance suggest significant variation in sales amounts, possibly reflecting different types of products or packages sold.
- The overall sum of sales is substantial at over \$9.76 million, indicating a healthy volume of transactions.
- The clear skew towards lower sales amounts suggests that while higher sales are possible, they are less common.
- Most customers tend to purchase amounts in the lower range.

# Exploratory Data Analysis

Bivariate, and multivariate analysis

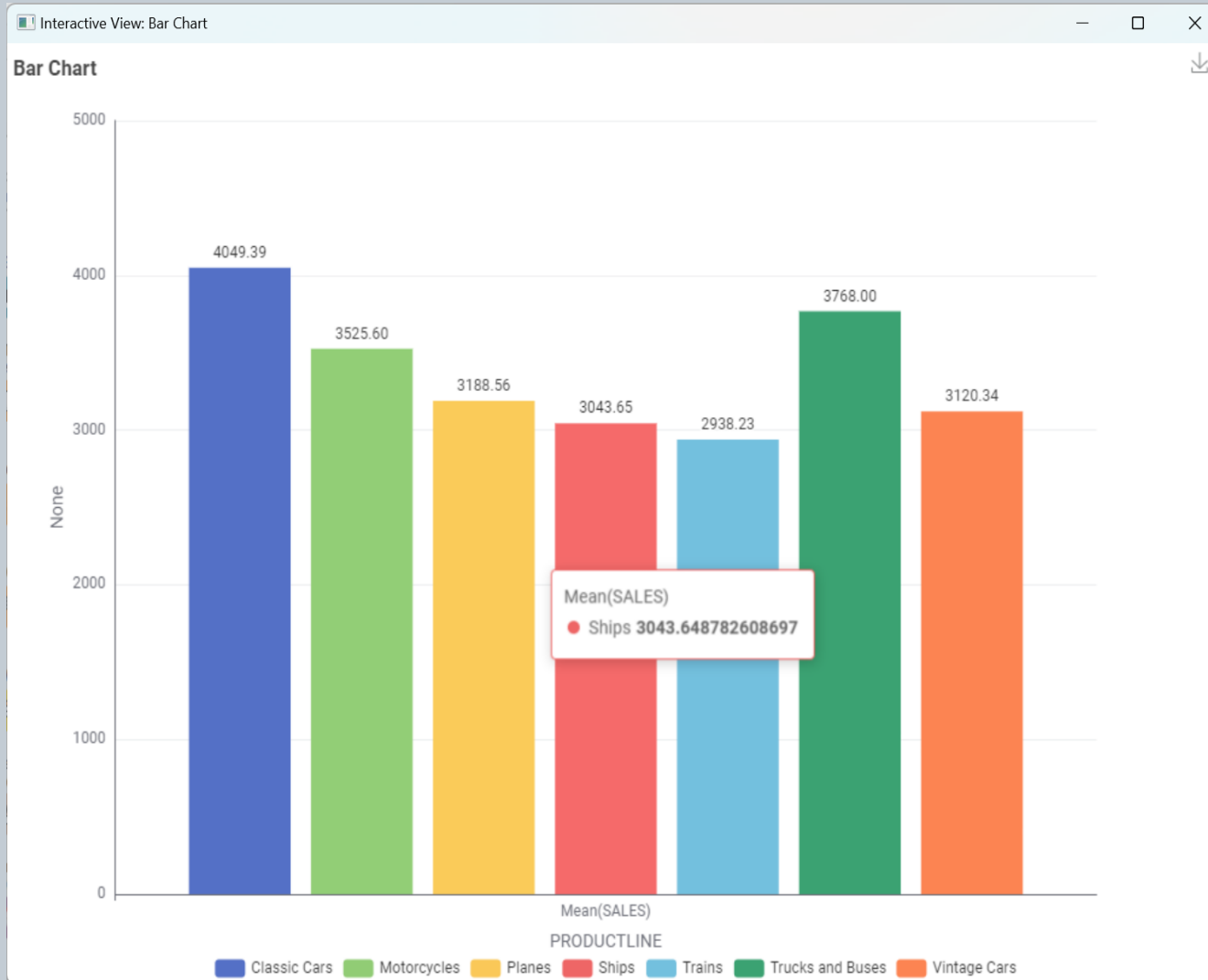


# EXPLORATORY DATA ANALYSIS

- A very high peak is observed around September 2019, where sales reached close to \$150,000. This spike could be indicative of a successful marketing campaign, a large or bulk order, or seasonal purchasing behaviors.
- While there isn't a clear periodic pattern that suggests a strong seasonal effect, the data does show regular intervals of higher sales, which might correspond to specific business activities or seasonal promotions.



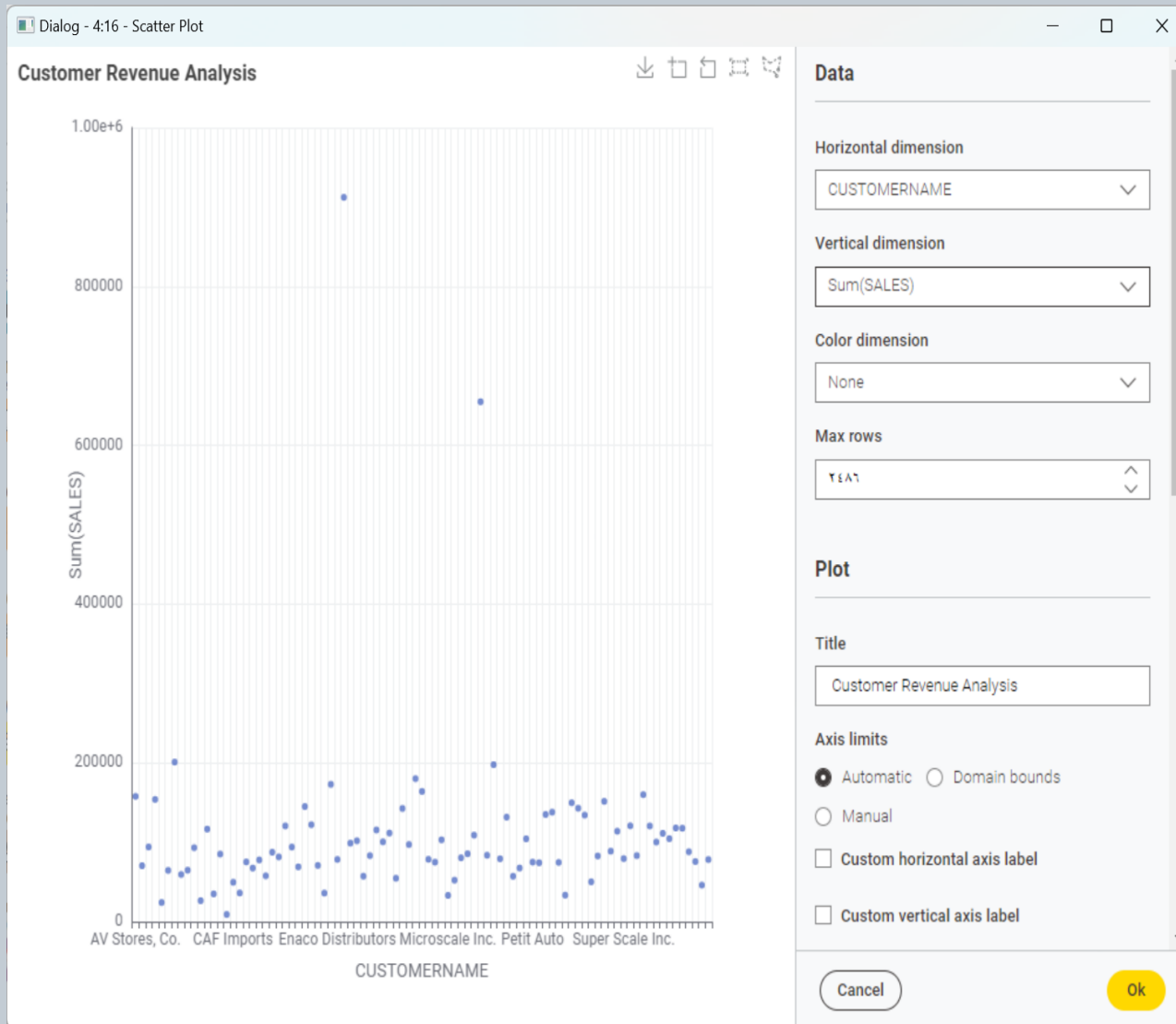
# EXPLORATORY DATA ANALYSIS



- **Classic Cars:** Shows the highest mean sales, at around \$4049.39 per transaction. This suggests that classic cars are likely high-ticket items that contribute substantially to revenue.
- **Motorcycles:** The mean sales for motorcycles are lower than for classic cars, at about \$3525.60, but still represent a significant revenue per transaction, indicating a robust market interest.
- **Planes:** These have a mean sales value of \$3188.56, placing them in the mid-range of product sales. This suggests a steady demand, possibly due to a niche market.

# Exploratory Data Analysis

- › **Ships:** The mean sales for ships stand at \$3043.65, slightly less than planes but still above many other categories, indicating a potentially specialized market with consistent sales values.
- › **Trains:** This category shows mean sales of \$2938.23. While lower, this still suggests a reasonable demand within its niche.
- › **Trucks and Buses:** Mean sales are \$3768.00, which is higher than planes, ships, and trains, reflecting possibly a stronger market presence or higher pricing power within this category.
- › **Vintage Cars:** The mean sales are the lowest among the categories analyzed, at \$3120.34, which could indicate either lower prices or lower quantities sold per transaction compared to other categories.



## EXPLORATORY DATA ANALYSIS

- The scatter plot you provided shows the total sales by customer name, which allows us to analyze customer contribution to total revenue.
- Most customers have sales figures that cluster at the lower end of the scale, typically under \$200,000. This suggests that the majority of customers generate a moderate amount of revenue.
- There are a few customers who stand out with significantly higher sales figures, reaching up to nearly \$1,000,000. These customers are likely key accounts that contribute a disproportionate amount of total sales.
- The presence of these high-value customers implies a concentration of revenue among a small number of clients. This could be a risk if the company is overly reliant on these few accounts.

# Inference Summary from EDA

$\pi$

- › Classic Cars and Trucks & Buses had the highest mean sales, indicating these are high-ticket items. Vintage Cars had the lowest mean sales. This information can guide inventory and marketing strategies, focusing on high-performing categories while exploring ways to boost sales in lower-performing areas.
- › A few customers contribute a disproportionately high amount of sales, indicating revenue concentration. It's crucial to manage relationships with these key customers carefully while also working to increase the spending of lower-tier customers to diversify revenue sources.
- › There's a wide range of days since last orders, with a significant drop-off in reordering beyond 2500 days. This highlights the importance of engaging customers who have not made recent purchases, potentially through re-engagement campaigns or loyalty incentives.
- › Most orders consist of mid-range quantities, with very few large quantity orders. There may be opportunities to incentivize larger orders through bulk discounts or promotions to increase the average order size.

# Customer Segmentation using RFM

- › RFM analysis is a marketing technique used to quantitatively rank, and group customers based on their purchasing habits.
- › **Parameters Used and Assumptions Made:**
- › **Recency (R):** How recently a customer has made a purchase. A more recent purchase scores higher, indicating that the customer is more likely to respond to new offers. It is measured by the minimum number of days since the last purchase (**Min(DAYS\_SINCE\_LAST\_ORDER)**). The assumption is that more recent interactions indicate higher engagement and a greater likelihood of repeat sales.
- › **Frequency (F):** How often a customer makes a purchase within a given time frame. Frequent shoppers are scored higher, reflecting higher engagement and loyalty. It is measured by counting the number of orders (**Count(ORDER\_NUMBER)**). The assumption here is that customers who order more frequently are more engaged and have a stronger relationship with the brand.
- **Monetary (M):** How much money a customer spends during a given period. Customers who spend more are considered more valuable and receive a higher score. It is evaluated by the sum of sales (**Sum(SALES)**). The assumption is that customers who spend more are more valuable.

# Customer Segmentation using RFM

- › KNIME Rule Engine node Rule expressions . In order to categorizing customers into distinct segments based on their purchasing behavior, as defined by the Recency, Frequency, and Monetary values.
- › **"Best Customers"** in two scenarios: Customers who have made their most recent purchase very recently (Min(DAYS\_SINCE\_LASTORDER) = "Bin 1").Customers who order frequently and spend a lot (Count(ORDERNUMBER) [Binned] = "Bin 5" AND Sum(SALES) [Binned] = "Bin 5").
- › **"Lost Customers"** who either haven't purchased in a long time (Min(DAYS\_SINCE\_LASTORDER) = "Bin 5") or who order very infrequently (Count(ORDERNUMBER) [Binned] = "Bin 1").
- › **"Loyal Customers"** includes customers who have an order frequency that is at least moderate to high Count(ORDERNUMBER) [Binned]\$ >= "Bin 3".
- › **"Big Spenders"** customers whose total spending is moderate to high Sum(SALES) [Binned]\$ >= "Bin 3"
- › **"Regular Customers"** indicating typical engagement without any particularly distinguishing purchasing behaviors for who don't fall into any special category based on RFM analysis (default condition)

# Customer Segmentation using RFM

Interactive View: Table View

**RFM (Recency, Frequency, Monetary) analysis**

Rows: 89 | Columns: 8

<input type="checkbox"/>	CUSTOMER... String	ADDRESSLI... String	CITY String	Count*(ORD... Number (integer)	Count*(ORD... String	Sum(SALES... String	Min*(DAYS_... String	prediction String
<input type="checkbox"/>	AV Stores, Co.	Fauntleroy Circus	Manchester	51	Bin 4	Bin 4	Bin 3	Loyal Customers
<input type="checkbox"/>	Alpha Cognac	1 rue Alsace-Lorra	Toulouse	20	Bin 1	Bin 1	Bin 4	Lost Customers
<input type="checkbox"/>	Amica Models & C	Via Monte Bianco	Torino	26	Bin 2	Bin 3	Bin 2	Big Spenders
<input type="checkbox"/>	Anna's Decoration	201 Miller Street	North Sydney	46	Bin 4	Bin 4	Bin 1	Best Customers
<input type="checkbox"/>	Atelier graphique	54, rue Royale	Nantes	7	Bin 1	Bin 1	Bin 2	Lost Customers
<input type="checkbox"/>	Australian Collect	7 Allen Street	Glen Waverly	23	Bin 2	Bin 1	Bin 4	Regular customer
<input type="checkbox"/>	Australian Collect	636 St Kilda Road	Melbourne	55	Bin 4	Bin 4	Bin 1	Best Customers
<input type="checkbox"/>	Australian Gift Ne	31 Duncan St. We	South Brisbane	15	Bin 1	Bin 1	Bin 1	Best Customers
<input type="checkbox"/>	Auto Assoc. & Cie	67, avenue de l'Eu	Versailles	18	Bin 1	Bin 1	Bin 2	Lost Customers
<input type="checkbox"/>	Auto Canal Petit	25, rue Lauriston	Paris	27	Bin 3	Bin 3	Bin 1	Best Customers
<input type="checkbox"/>	Auto-Moto Classic	16780 Pompton S	Brickhaven	8	Bin 1	Bin 1	Bin 4	Lost Customers
<input type="checkbox"/>	Baane Mini Impor	Erling Skakkes ga	Stavern	32	Bin 3	Bin 3	Bin 1	Best Customers
<input type="checkbox"/>	Bavarian Collectal	Hansastr. 15	Munich	14	Bin 1	Bin 1	Bin 4	Lost Customers
<input type="checkbox"/>	Blauer See Auto, C	Lyonerstr. 34	Frankfurt	22	Bin 2	Bin 2	Bin 4	Regular customer
<input type="checkbox"/>	Boards & Toys Co.	4097 Douglas Av.	Glendale	3	Bin 1	Bin 1	Bin 2	Lost Customers

# Customer Segmentation using RFM

- › Here are some customers from the RFM analysis table along with their segment classification based on the provided data:

## **AV Stores, Co. Segment: Loyal Customers**

Details: High frequency (Bin 4) and monetary value (Bin 4), with a somewhat recent purchase (Bin 3).

## **Alpha Cognac Segment: Lost Customers**

Details: Low frequency (Bin 1) and monetary value (Bin 1), with a long time since last purchase (Bin 4).

## **Amica Models & Co Segment: Big Spenders**

Details: Moderate frequency (Bin 2) and high spending (Bin 3), with recent activity (Bin 2).

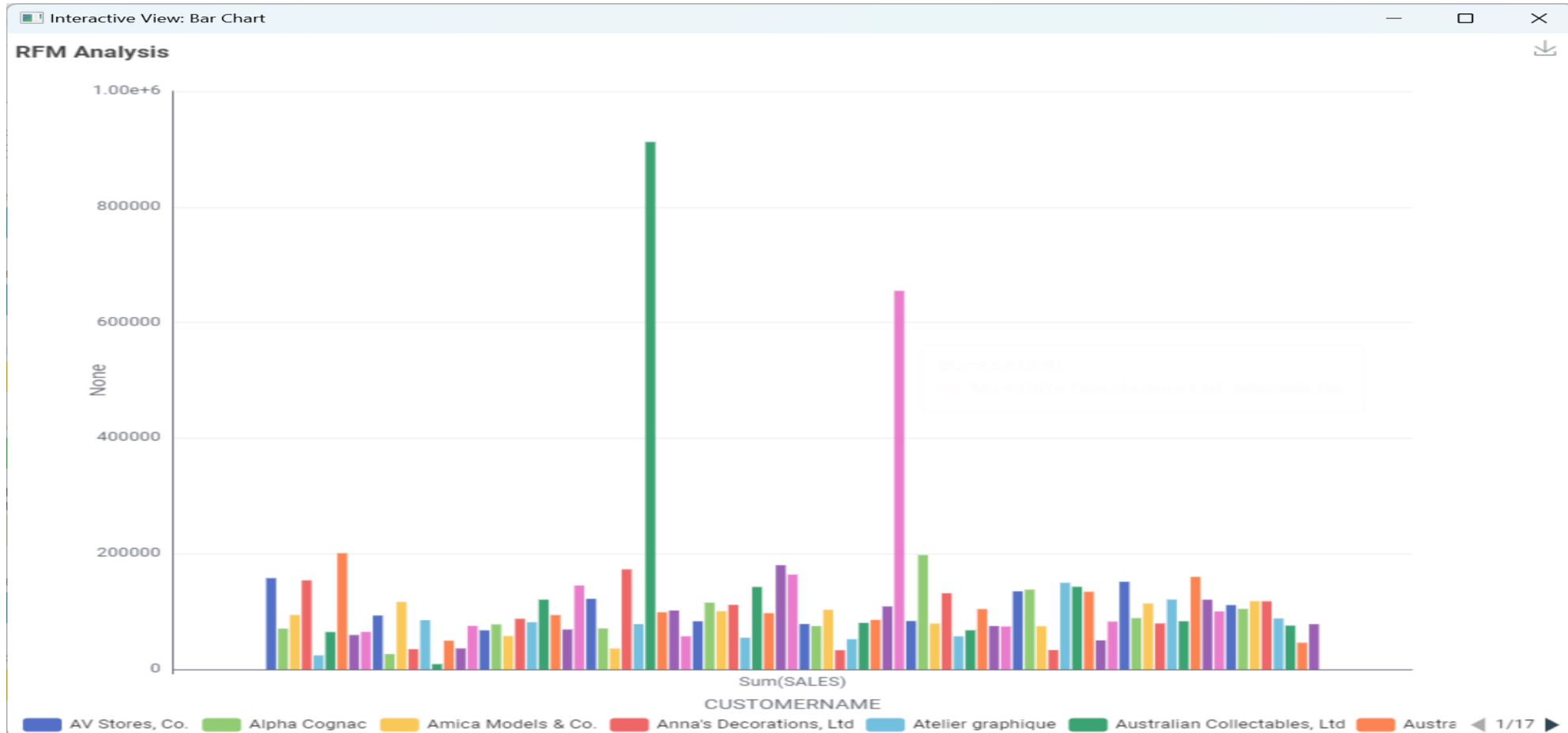
## **Anna's Decoration Segment: Best Customers**

Details: High frequency (Bin 4) and monetary value (Bin 4), with very recent activity (Bin 1).



# Inferences from RFM Analysis

$\pi$



# Inferences from RFM Analysis

$\pi$

- › The chart shows a concentration of high sales among a few customers. This could indicate a reliance on these customers for a significant portion of total revenue, which might be a risk if the business is too dependent on them.
- › Customers with lower sales might represent untapped potential. Developing targeted marketing strategies to increase their engagement and spending could be beneficial.
- › Understanding the factors that drive higher sales among top-performing customers can help replicate this success with others. Analyzing purchasing patterns, preferences, and customer feedback can provide insights into effective strategies.

# Inferences from RFM Analysis

$\pi$

- › **Variability in Sales:** There is significant variability in sales among customers. Some customers contribute very high sales, while others contribute much less, which is typical in many business environments.
- › **High-Value Customers:** Notably, there are a few customers with exceptionally high sales, standing out markedly above others. For example, "Mini Gifts Distributors Ltd." has a very high bar in the chart, indicating substantially higher total sales compared to other customers.
- › **Color Coding:** Different colors may represent different segments or categories of customers, although without a legend or further context, the exact meaning of the colors isn't clear. They could possibly indicate different regions, customer types, or RFM segments.

# Inferences from RFM Analysis

$\pi$

Count*	CUSTOMER	ADDRESSLI	CITY (Right)	Sum(SALES)	Sum(S...	ADDRESSLI	CITY (Right)	Min*(D...	prediction
4	Anna's Decorati	201 Miller Stree	North Sydney	■	Bin 4	201 Miller Stree	North Sydney	Bin 1	Best Customer:
4	Australian Colle	636 St Kilda Ro	Melbourne	■	Bin 4	636 St Kilda Ro	Melbourne	Bin 1	Best Customer:
1	Australian Gift I	31 Duncan St. V	South Brisbane	■	Bin 1	31 Duncan St. V	South Brisbane	Bin 1	Best Customer:
3	Auto Canal Peti	25, rue Lauristo	Paris	■	Bin 3	25, rue Lauristo	Paris	Bin 1	Best Customer:
3	Baane Mini Imp	Erling Skakkes	Stavern	■	Bin 3	Erling Skakkes	Stavern	Bin 1	Best Customer:
2	Collectables Fo	7825 Douglas A	Brickhaven	■	Bin 2	7825 Douglas A	Brickhaven	Bin 1	Best Customer:
3	Diecast Classic	7586 Pompton	Allentown	■	Bin 4	7586 Pompton	Allentown	Bin 1	Best Customer:
4	Euro Shopping	C/ Moralarzal,	Madrid	■	Bin 4	C/ Moralarzal,	Madrid	Bin 1	Best Customer:
2	FunGiftIdeas.cc	1785 First Stree	New Bedford	■	Bin 3	1785 First Stree	New Bedford	Bin 1	Best Customer:
2	Gift Depot Inc.	25593 South Bz	Bridgewater	■	Bin 3	25593 South Bz	Bridgewater	Bin 1	Best Customer:
2	Gifts4AllAges.c	8616 Spinnaker	Boston	■	Bin 2	8616 Spinnaker	Boston	Bin 1	Best Customer:
4	La Rochelle Gif	67, rue des Cinc	Nantes	■	Bin 4	67, rue des Cinc	Nantes	Bin 1	Best Customer:
4	Land of Toys In	897 Long Airpo	NYC	■	Bin 4	897 Long Airpo	NYC	Bin 1	Best Customer:
1	Lyon Souvenir	27 rue du Color	Paris	■	Bin 2	27 rue du Color	Paris	Bin 1	Best Customer:
4	Mini Gifts Distri	5677 Strong St.	San Rafael	■	Bin 4	5677 Strong St.	San Rafael	Bin 1	Best Customer:
4	Online Diecast	2304 Long Airp	Nashua	■	Bin 4	2304 Long Airp	Nashua	Bin 1	Best Customer:
3	Oulu Toy Suppli	Torikatu 38	Oulu	■	Bin 3	Torikatu 38	Oulu	Bin 1	Best Customer:
2	Quebec Home	43 rue St. Laure	Montreal	■	Bin 2	43 rue St. Laure	Montreal	Bin 1	Best Customer:
4	Salzburg Collec	Geislweg 14	Salzburg	■	Bin 4	Geislweg 14	Salzburg	Bin 1	Best Customer:
4	Souvenirs And	Monitor Money	Chatswood	■	Bin 4	Monitor Money	Chatswood	Bin 1	Best Customer:
4	Technics Store	9408 Furth Circ	Burlingame	■	Bin 4	9408 Furth Circ	Burlingame	Bin 1	Best Customer:
4	The Sharp Gifts	3086 Ingle Ln.	San Jose	■	Bin 4	3086 Ingle Ln.	San Jose	Bin 1	Best Customer:
3	UK Collectables	Berkeley Garder	Liverpool	■	Bin 3	Berkeley Garder	Liverpool	Bin 1	Best Customer:

# Inferences from RFM Analysis

$\pi$

- › **Top 5 Best Customers:** customers are identified by high sales figures, which suggest they frequently purchase and spend significantly. They are likely to have both high frequency and monetary scores and a low recency score.
- 1. Anna's Decoration - With a monetary bin of 4.
- 2. Australian Collectibles, Ltd (Melbourne) - With a monetary bin of 4.
- 3. Euro Shopping Channel (Madrid) - With a monetary bin of 4.
- 4. Mini Gifts Distributors Ltd. (San Rafael) - With a monetary bin of 4.
- 5. Souvenirs And Things Co. (Chatswood) - With a monetary bin of 4.

# Inferences from RFM Analysis

$\pi$

Classified values (Table)

Rows: 21 | Columns: 14

Count*... g	CUSTOMER String	ADDRESSLI String	CITY (Right) String	Sum(SALES) Number (dou...	Sum(S... String	ADDRESSLI String	CITY (Right) String	Min*(DAY... String	prediction String
4	AV Stores, Co.	Fauntleroy Circi	Manchester	■	Bin 4	Fauntleroy Circi	Manchester	Bin 3	Loyal Customer
3	Corrida Auto Re	C/ Araquil, 67	Madrid	■	Bin 4	C/ Araquil, 67	Madrid	Bin 2	Loyal Customer
4	Danish Wholes	Vinb'ltet 34	Kobenhavn	■	Bin 4	Vinb'ltet 34	Kobenhavn	Bin 3	Loyal Customer
4	Dragon Souven	Bronz Sok., Bro	Singapore	■	Bin 4	Bronz Sok., Bro	Singapore	Bin 4	Loyal Customer
4	Handji Gifts& C	Village Close -	Singapore	■	Bin 3	Village Close -	Singapore	Bin 3	Loyal Customer
3	Heintze Collect	Smagsloget 45	Aarhus	■	Bin 3	Smagsloget 45	Aarhus	Bin 2	Loyal Customer
3	Herkku Gifts	Drammen 121,	Bergen	■	Bin 3	Drammen 121,	Bergen	Bin 3	Loyal Customer
4	L'ordine Souven	Strada Provinci	Reggio Emilia	■	Bin 4	Strada Provinci	Reggio Emilia	Bin 3	Loyal Customer
3	Marta's Replica	39323 Spinnaki	Cambridge	■	Bin 3	39323 Spinnaki	Cambridge	Bin 2	Loyal Customer
4	Mini Creations I	4575 Hillside D	New Bedford	■	Bin 3	4575 Hillside D	New Bedford	Bin 3	Loyal Customer
4	Muscle Machin	4092 Furth Circ	NYC	■	Bin 4	4092 Furth Circ	NYC	Bin 3	Loyal Customer
4	Reims Collectal	59 rue de l'Abb	Reims	■	Bin 4	59 rue de l'Abb	Reims	Bin 2	Loyal Customer
4	Rovelli Gifts	Via Ludovico il	Bergamo	■	Bin 4	Via Ludovico il	Bergamo	Bin 4	Loyal Customer
4	Saveley & Henri	2, rue du Comr	Lyon	■	Bin 4	2, rue du Comr	Lyon	Bin 3	Loyal Customer
4	Scandinavian G	?kergatan 24	Boras	■	Bin 4	?kergatan 24	Boras	Bin 2	Loyal Customer
3	Signal Gift Stor	8489 Strong St.	Las Vegas	■	Bin 2	8489 Strong St.	Las Vegas	Bin 4	Loyal Customer
3	Suominen Souv	Software Engin	Espoo	■	Bin 3	Software Engin	Espoo	Bin 2	Loyal Customer
3	Tokyo Collectat	2-2-8 Roppongi	Minato-ku	■	Bin 3	2-2-8 Roppongi	Minato-ku	Bin 2	Loyal Customer
3	Toys of Finland	Keskuskatu 45	Helsinki	■	Bin 3	Keskuskatu 45	Helsinki	Bin 2	Loyal Customer
3	Toys4GrownUp	78934 Hillside I	Pasadena	■	Bin 3	78934 Hillside I	Pasadena	Bin 4	Loyal Customer
3	Vida Sport, Ltd	Grenzacherweg	Gensve	■	Bin 3	Grenzacherweg	Gensve	Bin 3	Loyal Customer

# Inferences from RFM Analysis

$\pi$

- › Loyal Customers for those who show consistency in both frequency and recent interactions, as reflected by their "Frequency" and "Recency" bin values.
- 1. **Dragon Souvenirs, Singapore** :Frequency: Bin 4 (highest level of transaction frequency) ,Recency: Bin 4 (most recent interactions) Prediction: Loyal Customer , Dragon Souvenirs has the highest frequency and most recent purchase activity, indicating consistent engagement and recent transactions.
- 2. **Muscle Machine, NYC**: Frequency: Bin 4 , Recency: Bin 3, Prediction: Loyal Customer, Muscle Machine also displays high purchasing frequency with recent activity, underscoring strong ongoing engagement with your business.
- 3. **Reims Collectables, Reims**: Frequency: Bin 3, Recency: Bin 2, Prediction: Loyal Customer, Reims Collectables shows a good level of frequent purchasing behavior, though slightly less recent than the top two. Still, their engagement level categorizes them as loyal.
- 4. **Rovelli Gifts, Bergamo**: Frequency: Bin 4, Recency: Bin 4,Prediction: Loyal Customer, Rovelli Gifts is highly engaged, with the highest frequency and very recent interactions, marking them as particularly loyal.
- 5. **La Rochelle Gifts, Reims**: Frequency: Bin 3, Recency: Bin 2,Prediction: Loyal Customer, La Rochelle Gifts consistently engages with moderate frequency and relatively recent transactions, showing loyalty to the business.



# Inferences from RFM Analysis

$\pi$

int*(O...	CUSTOM...	ADDRESS...	CITY (Rig...	Sum(SAL...	Sum(SAL...	ADDRESS...	CITY (Rig...	Min*(DAY...	prediction
g	String	String	String	Number (dou...	String	String	String	String	String
ount*...	CUSTOMER	ADDRESSLI	CITY (Right)	Sum(SALES	Sum(S...	ADDRESSLI	CITY (Right)	2 selec...	1 selec...
1	Alpha Cognac	1 rue Alsace-Lo	Toulouse	70488.44	Bin 1	1 rue Alsace-Lo	Toulouse	Bin 4	Lost Customers
1	Auto-Moto Clas	16780 Pomptor	Brickhaven	26479.260000C	Bin 1	16780 Pomptor	Brickhaven	Bin 4	Lost Customers
1	Bavarian Collec	Hansastr. 15	Munich	34993.92	Bin 1	Hansastr. 15	Munich	Bin 4	Lost Customers
1	CAF Imports	Merchants Hou	Madrid	49642.05	Bin 1	Merchants Hou	Madrid	Bin 3	Lost Customers
1	Cambridge Coll	4658 Baden Av.	Cambridge	36163.6199995	Bin 1	4658 Baden Av.	Cambridge	Bin 3	Lost Customers
1	Clover Collectic	25 Maiden Lan	Dublin	57756.43	Bin 1	25 Maiden Lan	Dublin	Bin 4	Lost Customers
1	Daedalus Desig	184, chausse d	Lille	69052.41	Bin 1	184, chausse d	Lille	Bin 3	Lost Customers
1	Diecast Collect	6251 Ingle Ln.	Boston	70859.78	Bin 2	6251 Ingle Ln.	Boston	Bin 4	Lost Customers
1	Double Decker	120 Hanover Sc	London	36019.04	Bin 1	120 Hanover Sc	London	Bin 4	Lost Customers
1	Gift Ideas Corp.	2440 Pompton	Glendale	57294.420000C	Bin 1	2440 Pompton	Glendale	Bin 4	Lost Customers
1	Iberia Gift Impo	C/ Romero, 33	Sevilla	54723.62	Bin 1	C/ Romero, 33	Sevilla	Bin 4	Lost Customers
1	Microscale Inc.	5290 North Per	NYC	33144.930000C	Bin 1	5290 North Per	NYC	Bin 3	Lost Customers
1	Mini Auto Werk	Kirchgasse 6	Graz	52263.8999995	Bin 1	Kirchgasse 6	Graz	Bin 4	Lost Customers
1	Mini Caravy	24, place Klube	Strasbourg	80438.48	Bin 2	24, place Klube	Strasbourg	Bin 3	Lost Customers
1	Online Mini Coll	7635 Spinnaker	Brickhaven	57197.9599995	Bin 1	7635 Spinnaker	Brickhaven	Bin 3	Lost Customers
1	Osaka Souveni	Dojima Avanza	Osaka	67605.07	Bin 1	Dojima Avanza	Osaka	Bin 3	Lost Customers
1	Royale Belge	Boulevard Tiro	Charleroi	33440.1	Bin 1	Boulevard Tiro	Charleroi	Bin 4	Lost Customers
1	Signal Collectib	2793 Furth Circ	Brisbane	50218.510000C	Bin 1	2793 Furth Circ	Brisbane	Bin 4	Lost Customers
1	Super Scale Inc	567 North Penc	New Haven	79472.07	Bin 2	567 North Penc	New Haven	Bin 3	Lost Customers
1	West Coast Col	3675 Furth Circ	Burbank	46084.6399995	Bin 1	3675 Furth Circ	Burbank	Bin 3	Lost Customers



# Inferences from RFM Analysis

- › The top five lost customers can be identified by focusing on those with the highest monetary contributions (Sum(SALES)) despite their status. This approach highlights the lost customers who were once significant contributors to your revenue and therefore may represent important targets for re-engagement strategies.
- 1. **Alpha Cognac:** Sum(SALES): \$70,488.44, City: Toulouse, Recency Bin: Bin 4, Alpha Cognac has the highest sales among the lost customers, which highlights its potential value if re-engaged effectively.
- 2. **Iberia Gift Imports:** Sum(SALES): \$54,723.62, City: Sevilla, Recency Bin: Bin 4, Despite being categorized as lost, Iberia Gift Imports has made significant past contributions, indicating the importance of revisiting engagement with them.
- 3. **Osaka Souvenirs:** Sum(SALES): \$67,605.07, City: Osaka, Recency Bin: Bin 3, Osaka Souvenirs ranks high in sales but has not engaged recently, suggesting a focus area for renewing contact and potentially reviving their business relationship.
- 4. **Super Scale Inc.:** Sum(SALES): \$79,472.07, City: New Haven, Recency Bin: Bin 3, As one of the highest sales figures among lost customers, Super Scale Inc. represents a significant re-engagement opportunity.
- 5. **Mini Auto Werk:** Sum(SALES): \$52,263.90, City: Graz, Recency Bin: Bin 4, Despite its high monetary value, Mini Auto Werk's recency status indicates it has not purchased recently, highlighting it as a key target for re-engagement initiatives.

# Marketing and Retail Analysis (Part B)

**Amneh Ghanem**

$\pi$

# Agenda

- › Executive Summary
- › Business Problem Overview
- › Data Overview
- › Exploratory Data Analysis
- › Inference Summary from EDA
- › Market Basket Analysis (Association Rules)
- › Market Basket Analysis inference

# Executive Summary

$\pi$

- › **Transactional Trends:** Analysis of transaction data over time showed variability in daily sales volumes, indicating the presence of peak sales days which might be linked to promotions, seasonal peaks, or special events.
- › **Product Demand and Frequency:** The distribution of product purchases showed that some items like shampoo, waffles, and soda are frequently bought together, suggesting strong associative buying patterns among certain items. such as aluminum foil and shampoo showed less intuitive associations, possibly indicative of broader shopping habits or basket-filling effects during specific shopping trips.
- › **Seasonal Variations and High-Transaction Days:** The data indicated potential seasonal trends in item popularity, which could be utilized for planning inventory and marketing strategies around expected peaks in demand.
- › **Product Associations:** The association rules identified specific product combinations that customers frequently purchase together. For example, combinations like soda and waffles, bagels and cream, and hand soap with shampoo show strong co-purchase trends.
- › **Customer Buying Patterns:** The analysis highlighted several key buying patterns, indicating peak transaction days and significant variability in daily transactions, which may correlate with promotions, holidays, or other special events.
- › **Seasonal Trends:** The transaction data revealed potential seasonal trends in purchasing, with certain items becoming more popular during specific times of the year.

# Executive Summary

$\pi$

- › **Strong Product Associations:** The analysis highlighted specific product pairs and groups that are consistently purchased together, offering clear opportunities for targeted marketing and promotions.
- › **Customer Purchase Behavior:** Patterns in customer purchases revealed through association rules indicate potential for optimizing store layout and product placement to encourage increased basket sizes.
- › **Variability in Sales:** Fluctuations in daily sales provide insights into customer behavior, which could be leveraged to plan promotions and stock management more effectively during expected high-transaction periods.
- › **Opportunities for Cross-Promotion:** Identifying items frequently bought together allows for strategic cross-promotions, bundle offers, and even dynamic pricing strategies to enhance sales.

# Recommendations

$\pi$

- › **Bundle Offers:** Introduce bundle promotions for products frequently bought together, as identified by the association rules (e.g., soda and waffles, bagels and cream).
- › **Targeted Discounts:** Implement targeted discounts on products during peak buying times, which can be deduced from the transaction frequency over time.
- › **Store Layout Optimization:** Product Placement: Arrange the store layout to place complementary items near each other. This could enhance the shopping experience and increase basket sizes, as customers find it convenient to pick up associated items together.
- › **Seasonal Displays:** Adjust store displays and layouts to highlight seasonal or event-related products, enhancing visibility during peak times identified in the EDA.

# Recommendations

$\pi$

- › **Inventory Management:** Stock Levels: Adjust stock levels based on the popularity and co-purchase trends of products to ensure availability, especially for high-demand combinations.
- › **Dynamic Reordering:** Utilize insights from market basket analysis to improve inventory turnover rates by dynamically adjusting reorder levels and frequencies for frequently purchased together items.
- › **Customer Engagement and Retention:** Loyalty Programs: Develop loyalty programs that reward customers for purchasing combinations identified as popular. This could incentivize repeat business and increase customer satisfaction.
- › **Personalized Marketing:** Leverage data on popular product combinations and customer purchase histories to send personalized offers and recommendations through email or mobile apps.

# Business Problem Overview

- › The business problem overview, related to the grocery store, involves conducting a thorough analysis of Point of Sale (POS) data to identify commonly occurring sets of items in customer orders.
- › The goal is to provide recommendations on how the grocery store can increase its revenue through popular combo offers and discounts for customers.
- › This problem statement emphasizes the application of data analytics techniques to real-world business scenarios to drive revenue growth through strategic marketing initiatives.



File Table (Table)					
Rows: 20641   Columns: 3					
<input type="checkbox"/>	#	RowID	Date String	Order_id Number (integer)	Product String
<input type="checkbox"/>	1	Row0	1/1/2018	1	yogurt
<input type="checkbox"/>	2	Row1	1/1/2018	1	pork
<input type="checkbox"/>	3	Row2	1/1/2018	1	sandwich bags
<input type="checkbox"/>	4	Row3	1/1/2018	1	lunch meat
<input type="checkbox"/>	5	Row4	1/1/2018	1	all- purpose
<input type="checkbox"/>	6	Row5	1/1/2018	1	flour
<input type="checkbox"/>	7	Row6	1/1/2018	1	soda
<input type="checkbox"/>	8	Row7	1/1/2018	1	butter
<input type="checkbox"/>	9	Row8	1/1/2018	1	beef
<input type="checkbox"/>	10	Row9	1/1/2018	1	aluminum foil
<input type="checkbox"/>	11	Row10	1/1/2018	1	all- purpose
<input type="checkbox"/>	12	Row11	1/1/2018	1	dinner rolls
<input type="checkbox"/>	13	Row12	1/1/2018	1	shampoo
<input type="checkbox"/>	14	Row13	1/1/2018	1	all- purpose
<input type="checkbox"/>	15	Row14	1/1/2018	1	mixes
<input type="checkbox"/>	16	Row15	1/1/2018	1	soap
<input type="checkbox"/>	17	Row16	1/1/2018	1	laundry detergent
<input type="checkbox"/>	18	Row17	1/1/2018	1	ice cream
<input type="checkbox"/>	19	Row18	1/1/2018	1	dinner rolls
<input type="checkbox"/>	20	Row19	1/1/2018	2	toilet paper
<input type="checkbox"/>	21	Row20	1/1/2018	2	shampoo
<input type="checkbox"/>	22	Row21	1/1/2018	2	hand soap
<input type="checkbox"/>	23	Row22	1/1/2018	2	waffles
<input type="checkbox"/>	24	Row23	1/1/2018	2	cheeses
<input type="checkbox"/>	25	Row24	1/1/2018	2	mixes
<input type="checkbox"/>	26	Row25	1/1/2018	2	milk

# DATA OVERVIEW

The data appears to be transactional data from a grocery store, structured in a table format with three columns:

**1.RowID:** This column appears to serve as a unique identifier for each row in the dataset, possibly indicating the sequence of entries.

**2.Date:** The date column shows when each transaction occurred. From the sample, it looks like the data starts from January 1, 2018.

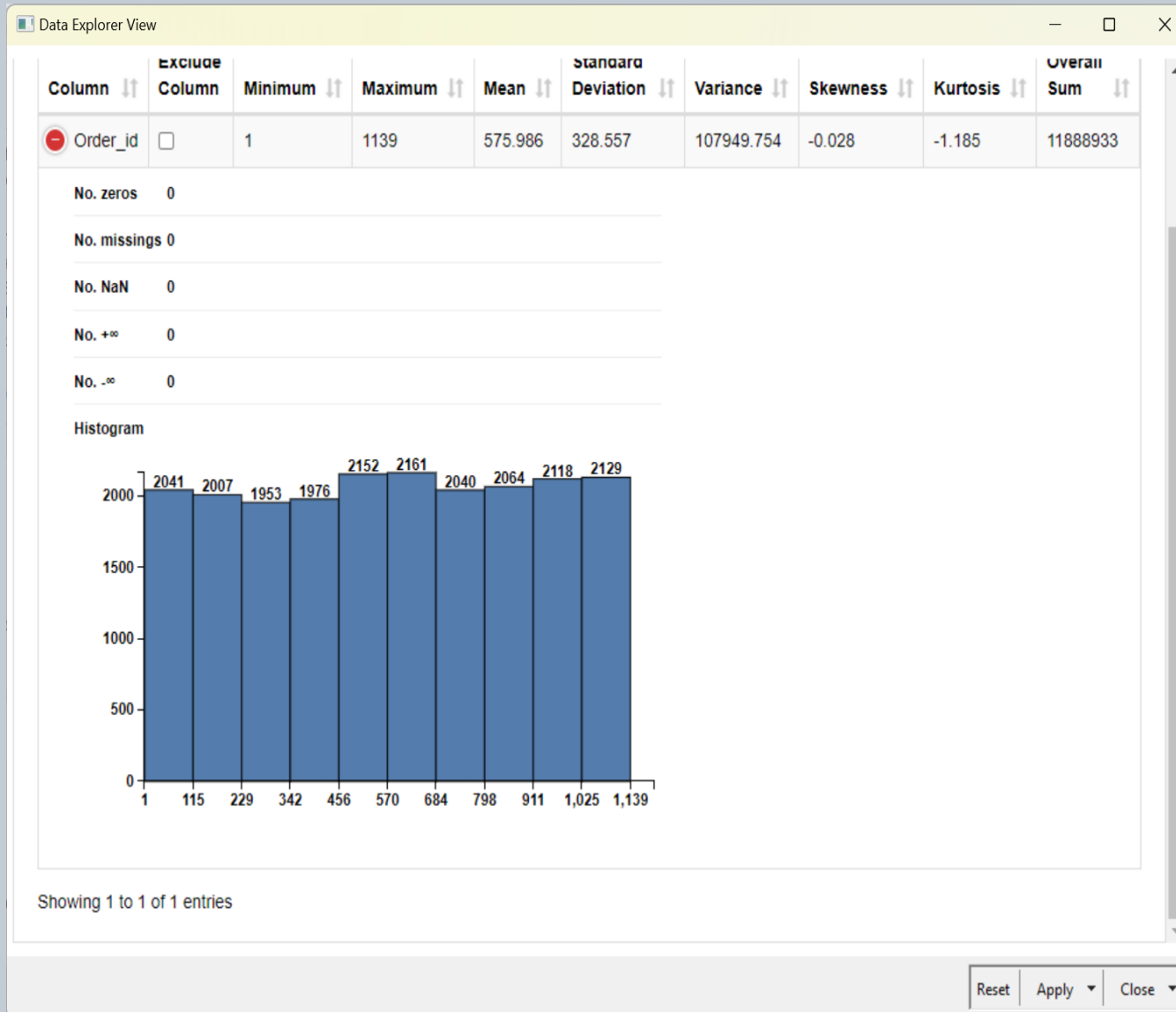
**3.Order\_id:** This column represents the identifier for each customer order. Multiple products associated with a single Order\_id indicate that those items were purchased together in one transaction.

**4.Product:** This column lists the individual products purchased in each transaction. Items range from food products like yogurt, pork, and cheese to household items like aluminum foil and laundry detergent.

# Exploratory Data Analysis

Univariate Analysis

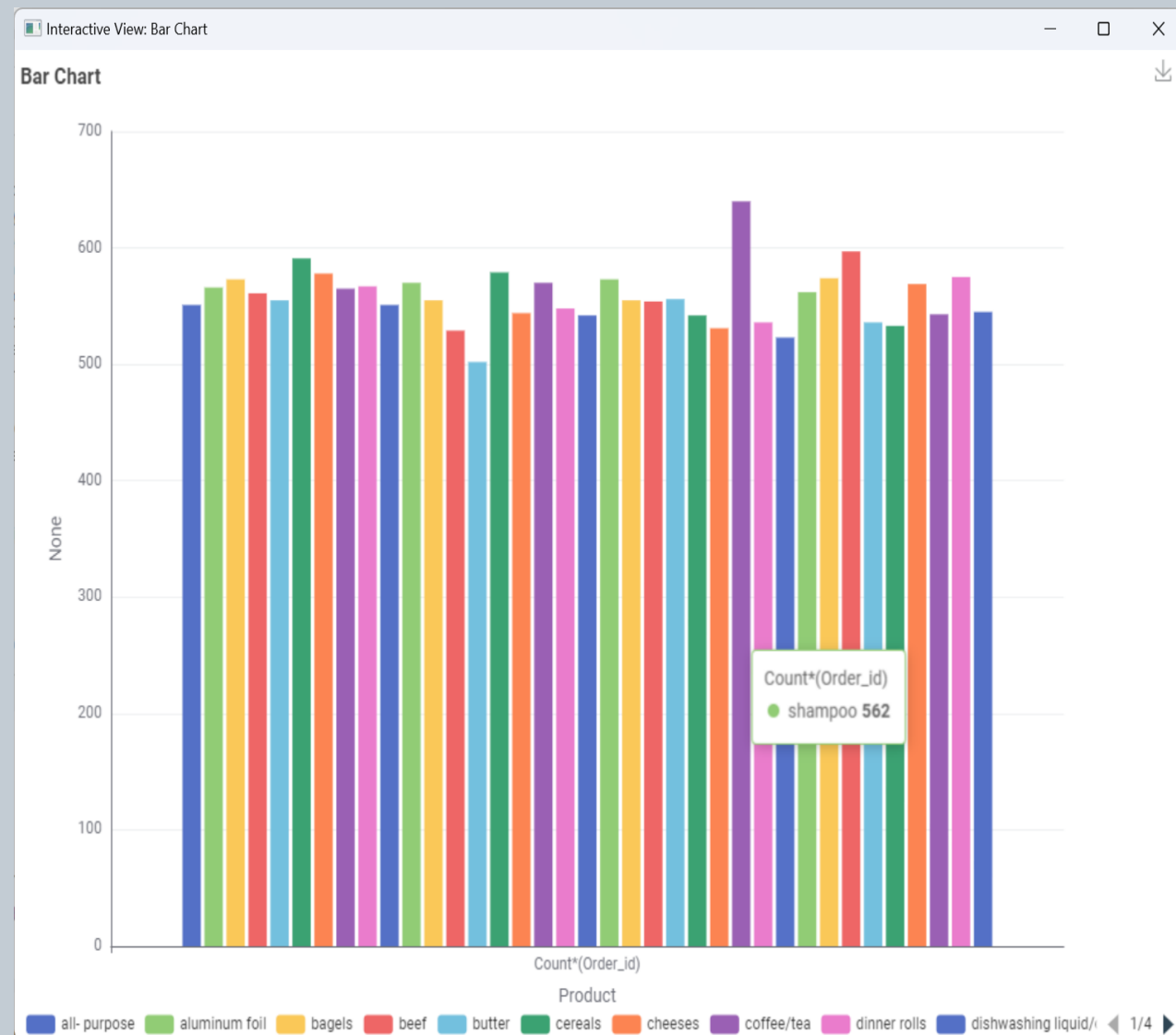
# EXPLORATORY DATA ANALYSIS



- The x-axis represents the order\_id, grouped into bins, and the y-axis shows the frequency of orders within each bin. The histogram bins are evenly distributed, showing a fairly uniform spread between 2041 and 2129 orders per bin, indicating a consistent ordering pattern across different order\_ids.
- The histogram and the metrics suggest a well-distributed dataset in terms of order frequency, with no significant outliers in terms of how many times each order\_id appears in the dataset. This uniformity could imply that the customer order behavior is consistent, or the data might have been pre-processed to ensure uniformity across different order\_ids.
- The total count of all items across all order\_ids sums to 1,188,893, reflecting the total transactions recorded.

# Exploratory Data Analysis

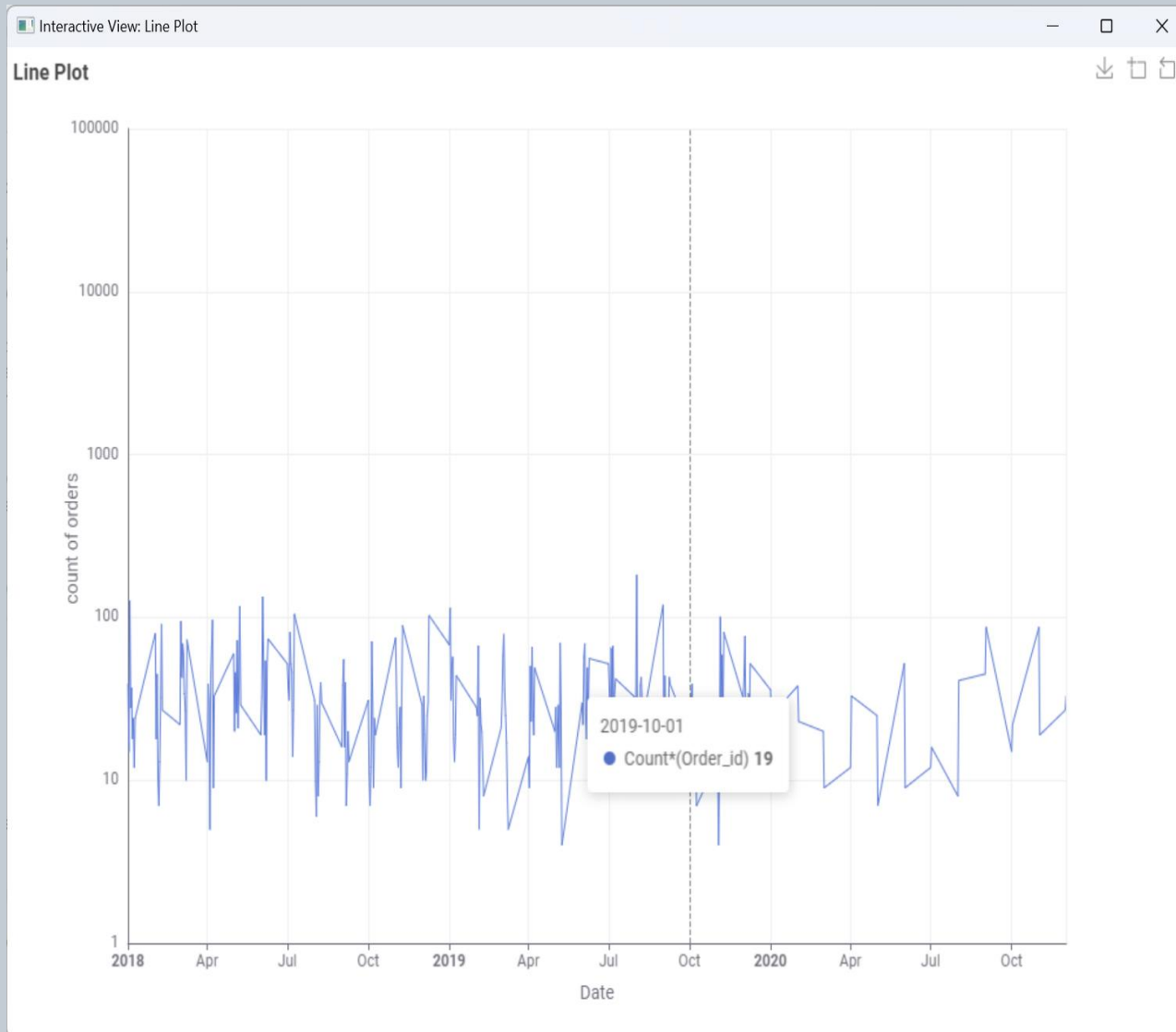
**Bivariate, and multivariate analysis**



## EXPLORATORY DATA ANALYSIS

- The bar chart shows the distribution of various products purchased, grouped by the count of orders (Order\_id) in which each product appears.
- The height of each bar represents the number of orders in which the product was included. The products appear to have a relatively uniform frequency of orders, with most products appearing in roughly 400 to 600 orders.
- From the visible part of the chart, the product "shampoo" is specifically highlighted, showing it appears in 562 orders. This suggests that shampoo is a popular product, given its relative frequency compared to some other products. Other products like bagels, beef, and butter also show relatively high frequencies.
- The uniformity in the spread and the similar heights of bars across various products indicate that the store has a well-distributed demand across many product types, which is beneficial for maintaining a diverse product stock.

# EXPLORATORY DATA ANALYSIS



- The line plot you provided illustrates the count of orders over time, grouped by date from 2018 through 2020.
- The plot spans from 2018 to late 2020. The data points appear to be plotted either daily or possibly weekly (exact frequency isn't specified but can be inferred from data granularity).
- There is a noticeable fluctuation in the number of orders over time. The plot does not show a clear upward or downward trend but rather significant variability in order volume within short periods.
- Several peaks suggest possible seasonal trends or special promotions/events that may have driven higher order volumes. These peaks might correlate with holidays, marketing campaigns, or other seasonal factors.
- After the dotted line in October 2019, the variability in order counts appears somewhat reduced compared to the previous period, suggesting a change in the pattern of order placements. This could be due to several factors including changes in consumer behavior, store policies, or external economic conditions.

Table View - 3:7 - Data to Report (BIRT) (241 x 2)

Row ID	Date	Cou...
Row0	?	12276
Row173	2019-08-02	183
Row48	2018-06-03	134
Row3	2018-01-03	127
Row181	2019-09-01	120
Row44	2018-05-08	117
Row110	2019-01-02	114
Row63	2018-07-09	105
Row108	2018-12-09	103
Row203	2019-11-05	101
Row34	2018-04-07	96
Row20	2018-03-02	95
Row17	2018-02-08	91
Row99	2018-11-09	89
Row234	2020-09-02	87
Row237	2020-11-01	87
Row58	2018-07-04	81
Row207	2019-11-09	81
Row10	2018-02-01	80
Row130	2019-03-04	79
Row210	2019-12-03	77
Row91	2018-11-01	75
Row54	2018-06-09	74
Row27	2018-03-09	73
Row41	2018-05-05	72
Row86	2018-10-05	71
Row151	2019-05-07	70
Row22	2018-03-04	69
Row157	2019-06-04	69
Row109	2019-01-01	68
Row120	2019-02-03	67
Row168	2019-07-06	67
Row140	2019-04-05	66
Row33	2018-04-06	65
Row166	2019-07-04	65
Row22	2018-03-05	60

# EXPLORATORY DATA ANALYSIS

- Data to Report (BIRT) represents summarized data, detailing the number of transactions or items sold per day, with the dates and corresponding counts ordered by frequency.
- Row0 from BIRT, it represents an aggregate of all transactions or activities recorded over the dataset's timeframe, rather than a specific day's data.
- the highest count visible now is from August 2, 2019, with 183 transactions, suggesting a peak sales day followed by June 3, 2018, with 134 transactions, and January 3, 2018, with 127 transactions.
- Charting transactions over time would provide a clearer visual representation of trends, seasonality, and anomalies. This can help identify patterns, such as increased sales during specific months or significant drops, which might need further investigation.

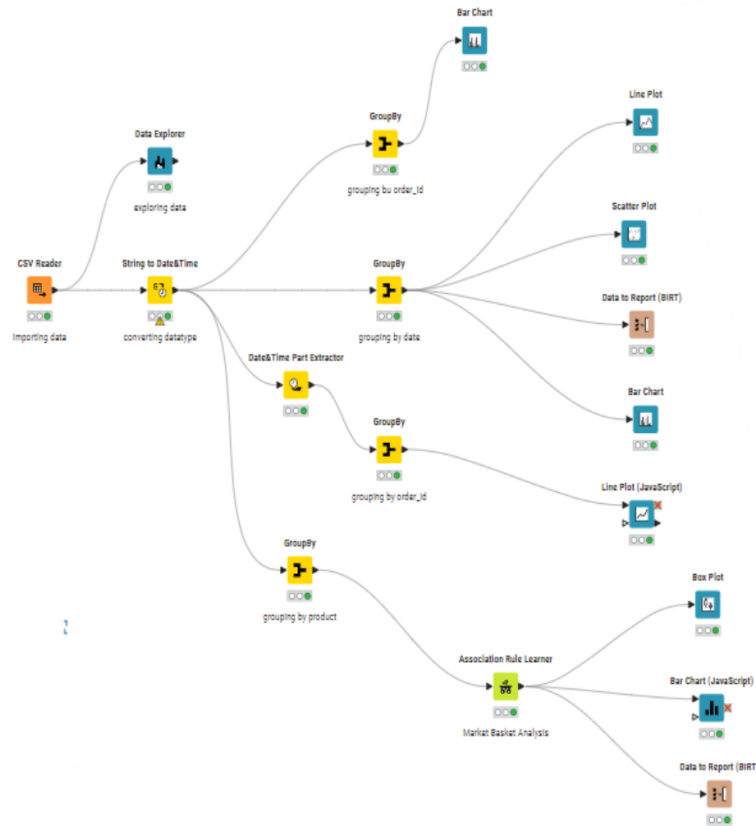
# Inference Summary from EDA

$\pi$

- › The dataset includes transaction records from a grocery store, featuring Date, Order id, and Product details. The transaction data is structured to show the number of orders per day and lists of products purchased per order, enabling analysis of customer purchasing patterns and product popularity.
- › Transaction counts vary significantly from day to day, with some dates showing extremely high transaction volumes, possibly linked to promotions, holidays, or special events. The frequency of orders for different products is relatively uniform, suggesting a consistent demand across a diverse range of products.
- › Certain products appear frequently in transactions, such as shampoo, which was highlighted in the data. These popular products are potential candidates for promotions or bundling strategies. A histogram of product distributions indicated a steady demand for various products, with no extreme variances, which implies a balanced stock requirement across multiple product categories.
- › The line plot of transactions over time showed fluctuating but significant volumes, hinting at underlying seasonal trends or the impact of specific marketing campaigns. A noted peak on August 2, 2019, with the highest transaction count suggests the effectiveness of promotional strategies or external factors driving sales on that day.



# MARKET BASKET ANALYSIS (ASSOCIATION RULES)



- The KNIME workflow in your image likely includes nodes for reading and preprocessing data, grouping data by transaction or product, and applying the
- **Association Rule Learner** node to extract these rules based on transactional datasets.
- Adjusting the support and confidence thresholds in the Association Rule Learner node will directly impact the number and quality of rules generated, thereby affecting decisions on store layout, promotions, and inventory strategies. Proper tuning of these parameters is crucial to extracting meaningful and actionable insights from the dataset.

# ASSOCIATION RULES LEARNER NODE

Dialog - 3:15 - Association Rule Learner (Market Basket Analysis)

File

Options | Flow Variables | Job Manager Selection | Memory Policy

Itemset Mining

Column containing transactions [...] List(Product) v

Minimum support (0-1) 0.1 v

Underlying data structure: ARRAY v

Output

Itemset type CLOSED v

Maximal itemset length: 10 v

Association Rules

☒ Output association rules

Minimum confidence: 0.4 v

OK Apply Cancel ?

- **Column Containing Transactions:** This parameter specifies the column in your dataset that contains the list of items (products) per transaction. In your case, it seems to be set to List(Product), which should be a collection or array of items purchased together in a single transaction.
- **Minimum Support (0-1) :** determines the minimum frequency a set of items must have to be considered in the analysis A minimum support of 0.1 means that only item sets appearing in at least 10% of all transactions will be considered.
- **Minimum Confidence (0-1) :** This parameter sets the minimum confidence level for rules to be considered. Confidence is a measure of the reliability of the rule. A minimum confidence of 0.4 means a rule must hold true at least 40% of the time it could be applied (based on the itemset frequency in the transactions).

Table View - 3:18 - Data to Report (BIRT) (915 x 6)

File Edit Hilit Navigation View

Row ID	[D] Support	[D] ▲ Conf...	[D] ▲ Lift	[S] Conseq...	[S] implies	[...] Items
rule710	0.167	0.423	1.004	poultry	<---	[waffles]
rule288	0.156	0.4	1.01	cereals	<---	[soda]
rule6	0.139	0.401	1.013	cereals	<---	[hand soap]
rule114	0.151	0.403	1.017	cereals	<---	[all-purpose]
rule279	0.156	0.406	1.02	ice cream	<---	[yogurt]
rule105	0.15	0.407	1.021	ice cream	<---	[shampoo]
rule84	0.149	0.404	1.024	waffles	<---	[tortillas]
rule209	0.154	0.409	1.026	ice cream	<---	[mixes]
rule1	0.139	0.401	1.026	soda	<---	[hand soap]
rule4	0.139	0.401	1.026	cheeses	<---	[hand soap]
rule521	0.161	0.406	1.029	waffles	<---	[cereals]
rule522	0.161	0.408	1.029	cereals	<---	[waffles]
rule37	0.145	0.41	1.03	ice cream	<---	[flour]
rule101	0.15	0.408	1.031	cereals	<---	[butter]
rule409	0.159	0.402	1.032	eggs	<---	[lunch meat]
rule408	0.159	0.408	1.032	lunch meat	<---	[eggs]
rule563	0.162	0.435	1.032	poultry	<---	[pasta]
rule741	0.168	0.435	1.032	poultry	<---	[bagels]
rule23	0.143	0.402	1.035	dinner rolls	<---	[pork]
rule198	0.154	0.409	1.035	lunch meat	<---	[mixes]
rule450	0.16	0.404	1.035	soda	<---	[lunch meat]
rule449	0.16	0.409	1.035	lunch meat	<---	[soda]
rule12	0.141	0.405	1.035	soda	<---	[sandwich loaves]
rule263	0.155	0.403	1.037	dinner rolls	<---	[bagels]
rule100	0.15	0.405	1.037	soda	<---	[fruits]
rule362	0.158	0.41	1.038	lunch meat	<---	[bagels]
rule28	0.143	0.414	1.038	ice cream	<---	[hand soap]
rule157	0.152	0.4	1.039	bagels	<---	[coffee/tea]
rule95	0.15	0.4	1.039	bagels	<---	[all-purpose]
rule440	0.159	0.438	1.04	poultry	<---	[paper towels]
rule484	0.16	0.401	1.04	bagels	<---	[ice cream]
rule483	0.16	0.415	1.04	ice cream	<---	[bagels]
rule174	0.153	0.404	1.04	dishwashing ...	<---	[toilet paper]
rule504	0.161	0.407	1.041	cheeses	<---	[lunch meat]
rule503	0.161	0.411	1.041	lunch meat	<---	[cheeses]
rule38	0.145	0.41	1.041	waffles	<---	[flour]
rule541	0.162	0.405	1.042	dinner rolls	<---	[ice cream]
rule542	0.162	0.415	1.042	ice cream	<---	[dinner rolls]
rule99	0.15	0.407	1.042	soda	<---	[shampoo]
rule56	0.147	0.401	1.043	aluminum foil	<---	[butter]
rule2	0.139	0.401	1.043	aluminum foil	<---	[hand soap]
rule230	0.155	0.411	1.043	waffles	<---	[mixes]
rule331	0.157	0.402	1.044	bagels	<---	[soda]
rule330	0.157	0.408	1.044	soda	<---	[bagels]
rule214	0.154	0.405	1.044	dishwashing ...	<---	[coffee/tea]
rule178	0.153	0.406	1.045	dishwashing ...	<---	[juice]
rule78	0.149	0.406	1.046	dishwashing ...	<---	[butter]
rule25	0.143	0.414	1.047	lunch meat	<---	[hand soap]
rule140	0.152	0.413	1.047	waffles	<---	[butter]
rule173	0.153	0.404	1.047	bagels	<---	[laundry detergent]
rule301	0.156	0.413	1.048	waffles	<---	[laundry detergent]
rule203	0.154	0.418	1.048	ice cream	<---	[butter]

# MARKET BASKET ANALYSIS INFERENCE

- The table you've provided is a result of the Association Rule Learner node in KNIME and displays a list of association rules generated from your dataset.
- Support:** This metric indicates the proportion of transactions in the dataset that contain the itemset. For example, a support of 0.167 means the itemset appears in 16.7% of all transactions.
- Confidence:** This metric measures the probability that the consequent (the item on the right of the rule) is purchased when the antecedent (the item on the left of the rule) is purchased. For example, a confidence of 0.423 means that 42.3% of the transactions containing the antecedent also contain the consequent.
- Lift:** Lift compares the observed frequency of A and B appearing together with the frequency expected if A and B were independent. A lift value greater than 1 indicates that the presence of A increases the likelihood of B occurring in the same transaction. For example, a lift of 1.004 suggests that the items are slightly more likely to be bought together than expected if they were independent.

# Market Basket Analysis inference

$\pi$

- › The previous table lists association rules derived from transaction data in a grocery store, we can observe patterns indicating which items are frequently bought together.
- › **Rule ID: rule710, Items: Soda , Consequent: Waffles** This rule suggests that there's a significant association between customers buying soda and purchasing waffles. The store might consider placing these items closer together or running a promotion that bundles them at a discount.
- › **Rule ID: rule283, Items: Cereals, Consequent: All-purpose,** Customers who buy cereals tend to also buy all-purpose products. This might indicate that customers who purchase breakfast items may also be looking to stock up on baking or general cooking supplies.
- › **Rule ID: rule90, Items: Bagels, Consequent: Cream,** This association likely points to common breakfast or brunch combinations, where customers who buy bagels also often buy cream (possibly cream cheese).

# Market Basket Analysis inference

$\pi$

- › These rules illustrate how customers tend to group their purchases, which can be leveraged in several ways:
- › **Marketing Strategies:** Creating targeted marketing campaigns that cater to observed buying patterns can enhance the effectiveness of promotions.
- › **Store Layout:** Arranging the store to place items that are frequently bought together in proximity can improve the shopping experience and increase the average transaction size.
- › **Inventory Management:** Understanding which items are often purchased together helps in better inventory planning to ensure that these items are always in stock, especially during key sales periods or promotions.
- › Each rule provides a statistical backbone to understand consumer behavior better, enabling the store to make informed decisions that could lead to increased sales and customer satisfaction. By analyzing these relationships, the store can tailor its strategies to meet customer needs more effectively and efficiently.