



MARKETING AND RETAIL ANALYSIS

PROJECT MRA

Abstract

In the ever-evolving landscape of marketing and retail, understanding customer behavior and preferences is paramount for driving business growth and customer satisfaction. This analysis delves into two pivotal techniques: Recency, Frequency, Monetary (RFM) analysis and Market Basket analysis. These methodologies offer insightful perspectives into customer segmentation, buying patterns, and strategic decision-making.

Sudheendra K
PGP.DSBA.OCT.B.22
Sudhi0404@gmail.com

Table of Contents

<u>PART A.....</u>	<u>2</u>
<u>PROBLEM STATEMENT:.....</u>	<u>2</u>
<u>DATASET.....</u>	<u>3</u>
<u>AGENDA & EXECUTIVE SUMMARY OF THE DATA.....</u>	<u>3</u>
<u>EXPLORATORY ANALYSIS AND INFERENCES.....</u>	<u>6</u>
<u>CUSTOMER SEGMENTATION USING RFM ANALYSIS:</u>	<u>12</u>
<u>ASSUMPTIONS:.....</u>	<u>12</u>
<u>BEST CUSTOMERS (TOP 5):.....</u>	<u>14</u>
<u>CUSTOMERS ON THE VERGE OF CHURNING (TOP 5):</u>	<u>14</u>
<u>LOST CUSTOMERS (TOP 5):.....</u>	<u>15</u>
<u>LOYAL CUSTOMERS (TOP 5):</u>	<u>15</u>
<u>INTERPRETATION AND SUMMARY OF THE RFM ANALYSIS.....</u>	<u>15</u>
<u>1. HHH SEGMENT (HIGH RECENCY, HIGH FREQUENCY, HIGH MONETARY):</u>	<u>15</u>
<u>2. HMH SEGMENT (HIGH RECENCY, MEDIUM FREQUENCY, HIGH MONETARY):.....</u>	<u>15</u>
<u>3. MMH SEGMENT (MEDIUM RECENCY, MEDIUM FREQUENCY, HIGH MONETARY):.....</u>	<u>15</u>
<u>4. MLL SEGMENT (MEDIUM RECENCY, LOW FREQUENCY, LOW MONETARY):.....</u>	<u>16</u>
<u>5. LLL SEGMENT (LOW RECENCY, LOW FREQUENCY, LOW MONETARY):.....</u>	<u>16</u>
<u>CONCLUSION</u>	<u>16</u>
<u>PART B</u>	<u>16</u>
<u>PROBLEM STATEMENT</u>	<u>16</u>
<u>EXPLORATORY ANALYSIS</u>	<u>16</u>
<u>ASSOCIATION RULES AND THEIR RELEVANCE:</u>	<u>18</u>
<u>IN THE CASE OF THE GROCERY STORE, ASSOCIATION RULES CAN HELP IN SEVERAL WAYS:</u>	<u>18</u>
<u>THRESHOLD VALUES OF SUPPORT AND CONFIDENCE:</u>	<u>19</u>

<u>ASSOCIATIONS IDENTIFIED AND THEIR SIGNIFICANCE:</u>	<u>20</u>
<u>UNDERSTANDING SUPPORT, CONFIDENCE, AND LIFT VALUES:</u>	<u>20</u>
<u>SUPPORT:</u>	<u>20</u>
<u>CONFIDENCE:</u>	<u>20</u>
<u>LIFT:</u>	<u>21</u>
<u>INTERPRETATION</u>	<u>21</u>
<u>SUGGESTIONS FOR LUCRATIVE COMBOS WITH DISCOUNT OFFERS:</u>	<u>21</u>
<u>COMBO: BREAKFAST BOOST.....</u>	<u>21</u>
<u>COMBO: FAMILY PASTA NIGHT</u>	<u>21</u>
<u>COMBO: SNACK DELIGHT</u>	<u>21</u>
<u>COMBO: CLEANING ESSENTIALS.....</u>	<u>22</u>
<u>COMBO: REFRESHING BEVERAGE PACK</u>	<u>22</u>
<u>COMBO: WEEKEND INDULGENCE.....</u>	<u>22</u>
<u>ADDITIONAL STRATEGIES:.....</u>	<u>22</u>
<u>CONCLUSION</u>	<u>22</u>

Part A

Problem Statement:

An automobile parts manufacturing company has collected data on transactions for 3 years. They do not have any in-house data science team, thus they have hired you as their consultant. Your job is to use your data science skills to find the underlying buying patterns

of the customers, provide the company with suitable insights about their customers, and recommend customized marketing strategies for different segments of customers.

Dataset

Auto Sales Data: Sales_Data.xlsx

Agenda & Executive Summary of the data

- From the below snippet we could observe a few rows and attributes of the Sales dataset

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	DAYS_SINCE_LASTORDER	STATUS	PRODUCTLI
0	10107	30	95.70	2	2871.00	1970-01-01 00:00:00.000043155		828 Shipped	Motorcyc
1	10121	34	81.35	5	2765.90	1970-01-01 00:00:00.000043227		757 Shipped	Motorcyc
2	10134	41	94.74	2	3884.34	1970-01-01 00:00:00.000043282		703 Shipped	Motorcyc
3	10145	45	83.26	6	3746.70	1970-01-01 00:00:00.000043337		649 Shipped	Motorcyc
4	10168	36	96.66	1	3479.76	1970-01-01 00:00:00.000043401		586 Shipped	Motorcyc

Snippet 1. 1 Head of the dataset

```

Dataset information:
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2747 entries, 0 to 2746
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  -
0   ORDERNUMBER           2747 non-null   int64
1   QUANTITYORDERED       2747 non-null   int64
2   PRICEEACH             2747 non-null   float64
3   ORDERLINENUMBER       2747 non-null   int64
4   SALES                 2747 non-null   float64
5   ORDERDATE             2747 non-null   int64
6   DAYS_SINCE_LASTORDER  2747 non-null   int64
7   STATUS                2747 non-null   object
8   PRODUCTLINE           2747 non-null   object
9   MSRP                 2747 non-null   int64
10  PRODUCTCODE           2747 non-null   object
11  CUSTOMERNAME          2747 non-null   object
12  PHONE                 2747 non-null   object
13  ADDRESSLINE1          2747 non-null   object
14  CITY                  2747 non-null   object
15  POSTALCODE            2747 non-null   object
16  COUNTRY               2747 non-null   object
17  CONTACTLASTNAME       2747 non-null   object
18  CONTACTFIRSTNAME      2747 non-null   object
19  DEALSIZE              2747 non-null   object
dtypes: float64(2), int64(6), object(12)
memory usage: 429.3+ KB
None

```

Snippet 1. 2 Information of the dataset

Here's a summary of the key details:

- *Number of Entries: 2747*
- *Columns: 20*
- *Data Types: The dataset contains attributes of three main types:*
 - *Numerical: There are 8 numerical attributes, including integer and float types. These attributes can be used for quantitative analysis and calculations.*
 - *Categorical: There are 12 categorical attributes, represented as objects. These attributes represent categories or labels and may require encoding or transformation for analysis.*
 - *String/Object: The STATUS, PRODUCTLINE, PRODUCTCODE, CUSTOMERNAME, PHONE, ADDRESSLINE1, CITY, POSTALCODE, COUNTRY, CONTACTLASTNAME, CONTACTFIRSTNAME, and DEALSIZE attributes contain textual information.*
- *Missing Data: The dataset appears to have no missing values, as indicated by the "Non-Null Count" for each attribute.*

Summary statistics:				
	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER
count	2747.000000	2747.000000	2747.000000	2747.000000
mean	10259.761558	35.103021	101.098951	6.491081
std	91.877521	9.762135	42.042548	4.230544
min	10100.000000	6.000000	26.880000	1.000000
25%	10181.000000	27.000000	68.745000	3.000000
50%	10264.000000	35.000000	95.550000	6.000000
75%	10334.500000	43.000000	127.100000	9.000000
max	10425.000000	97.000000	252.870000	18.000000

	SALES	DAYS_SINCE_LASTORDER	MSRP
count	2747.000000	2747.000000	2747.000000
mean	3553.047583	1757.085912	100.691664
std	1838.953901	819.280576	40.114802
min	482.130000	42.000000	33.000000
25%	2204.350000	1077.000000	68.000000
50%	3184.800000	1761.000000	99.000000
75%	4503.095000	2436.500000	124.000000
max	14082.800000	3562.000000	214.000000

Snippet 1.3 Statistical Summary of the Numerical Attributes

- *The ORDERNUMBER ranges from 10100 to 10425.*
- *The QUANTITYORDERED ranges from 6 to 97.*
- *PRICEEACH ranges from 26.88 to 252.87.*
- *The mean price per item is approximately 101.10, with a standard deviation of 42.04.*
- *SALES ranges from 482.13 to 14082.80.*
- *The mean sales amount is approximately 3553.05, with a standard deviation of 1838.95.*
- *There are some higher sales values that might be considered as potential outliers.*
- *DAYS_SINCE_LASTORDER ranges from 42 to 3562.*
- *The mean days since the last order is approximately 1757.09, with a standard deviation of 819.28.*
- *There are no apparent outliers based on the summary statistics.*
- *MSRP (Manufacturer's Suggested Retail Price) ranges from 33 to 214.*
- *The mean MSRP is approximately 100.69, with a standard deviation of 40.11.*
- *The data appears to be normally distributed, with no extreme outliers.*

```
Missing values:
ORDERNUMBER      0
QUANTITYORDERED  0
PRICEEACH        0
ORDERLINENUMBER  0
SALES            0
ORDERDATE        0
DAYS_SINCE_LASTORDER  0
STATUS          0
PRODUCTLINE      0
MSRP            0
PRODUCTCODE      0
CUSTOMERNAME     0
PHONE           0
ADDRESSLINE1     0
CITY            0
POSTALCODE      0
COUNTRY         0
CONTACTLASTNAME  0
CONTACTFIRSTNAME 0
DEALSIZE        0
dtype: int64
```

Snippet 1. 4 Missing Values

- *The data doesn't have any missing values*

Exploratory Analysis and Inferences

- *Univariate, Bivariate, and multivariate analysis using data visualization (Weekly, Monthly, Quarterly, Yearly Trends in Sales and Sales Across different Categories of different features in the given data) -> Summarise the inferences*

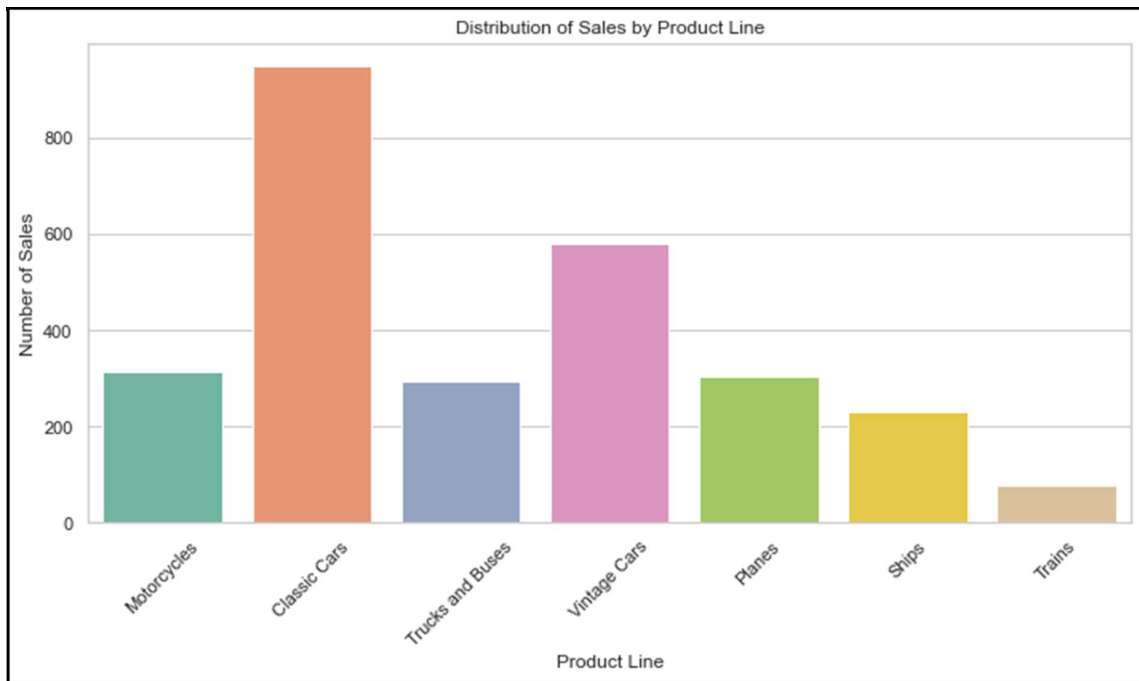


Figure 1. 1 Distribution of Sales by Product Line

- *Distribution of Sales by Product classic cars has the highest sales among the total 7 product lines and lowest being Trains*

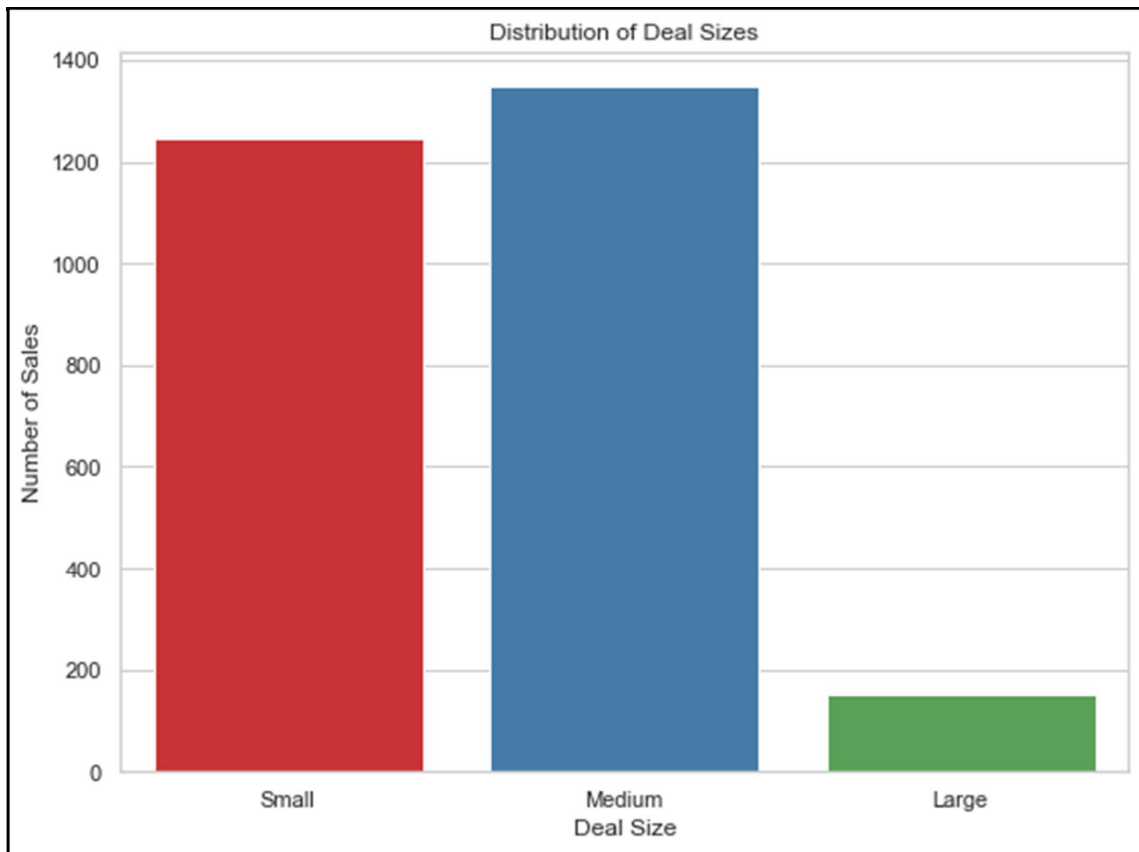


Figure 1. 2 Distribution of Deal Sizes

- *Distribution of Deal Sizes, here we could see that medium sized category stands on top of small and large type.*

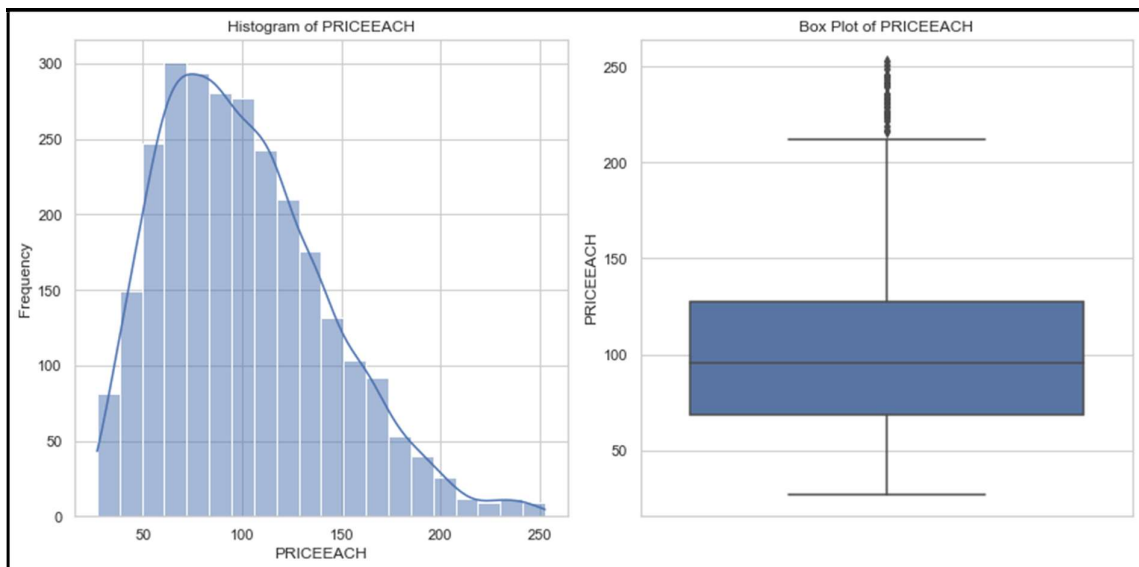


Figure 1. 3 Univariate Analysis of PriceEach

- *PRICEEACH ranges from 26.88 to 252.87.*

- The mean price per item is approximately 101.10, with a standard deviation of 42.04.
- There are apparent outliers based on the boxplot of this attribute.

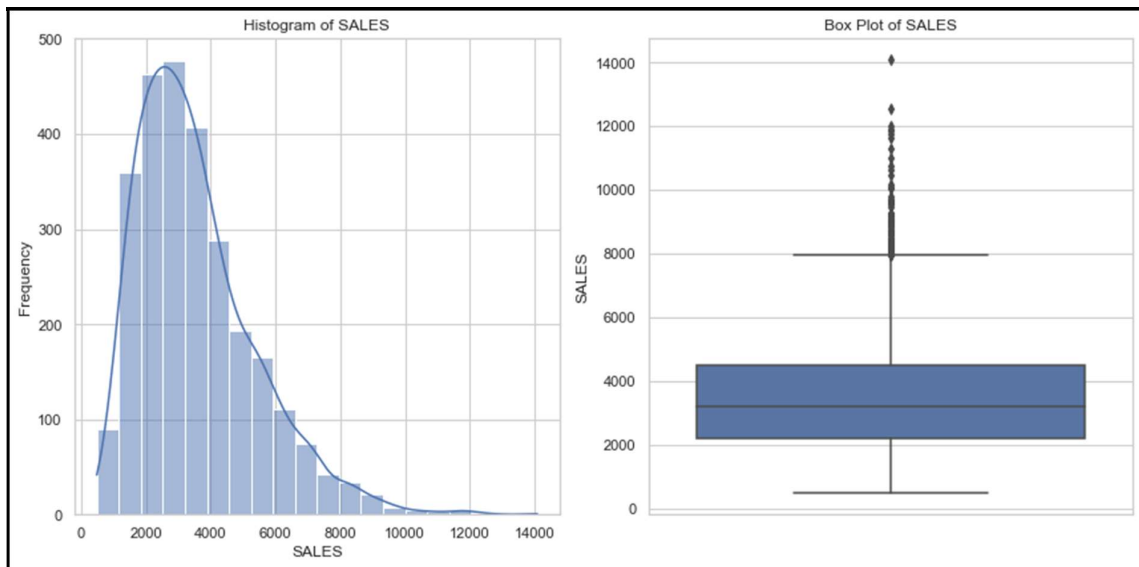


Figure 1. 4 Univariate Analysis of Sales

- SALES ranges from 482.13 to 14082.80.
- The mean sales amount is approximately 3553.05, with a standard deviation of 1838.95.
- There are some higher sales values that might be considered as potential outliers.

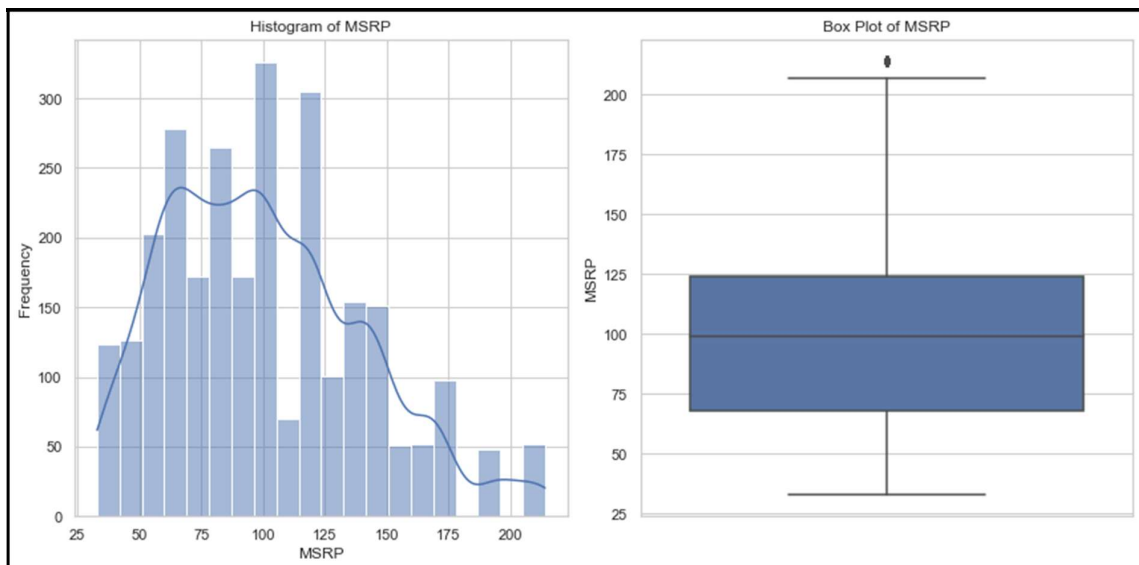


Figure 1. 5 Univariate Analysis of MSRP

- MSRP (Manufacturer's Suggested Retail Price) ranges from 33 to 214.

- The mean MSRP is approximately 100.69, with a standard deviation of 40.11.
- The data appears to be not normally distributed, with no extreme outliers.

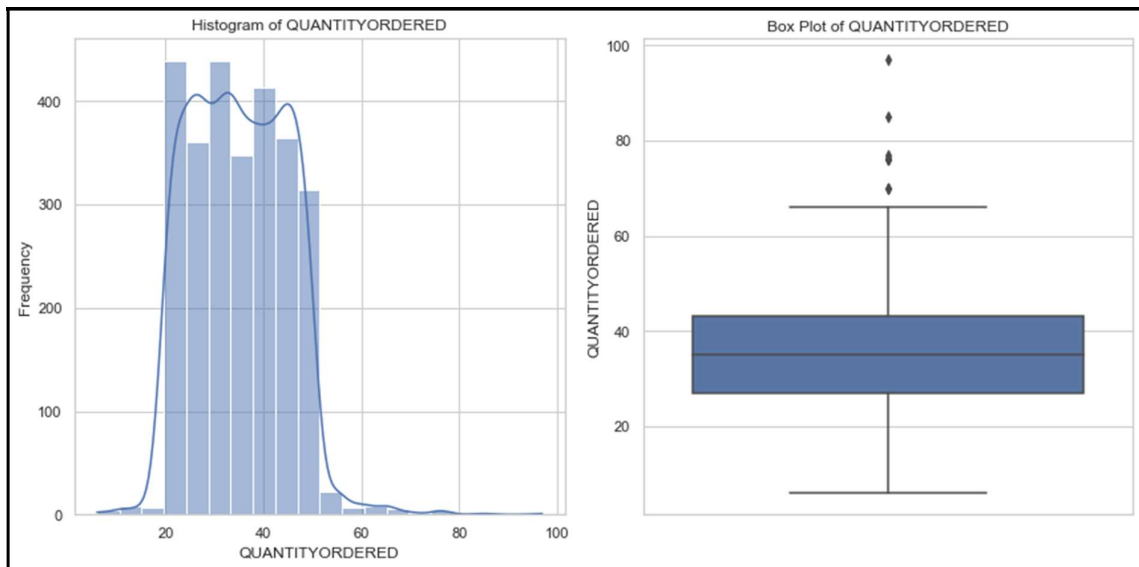


Figure 1. 6 Univariate Analysis of Quantity Order

- The QUANTITYORDERED ranges from 6 to 97.
- The average quantity ordered is around 35.10, with a standard deviation of 9.76.
- The data appears to be normally distributed, with outliers.

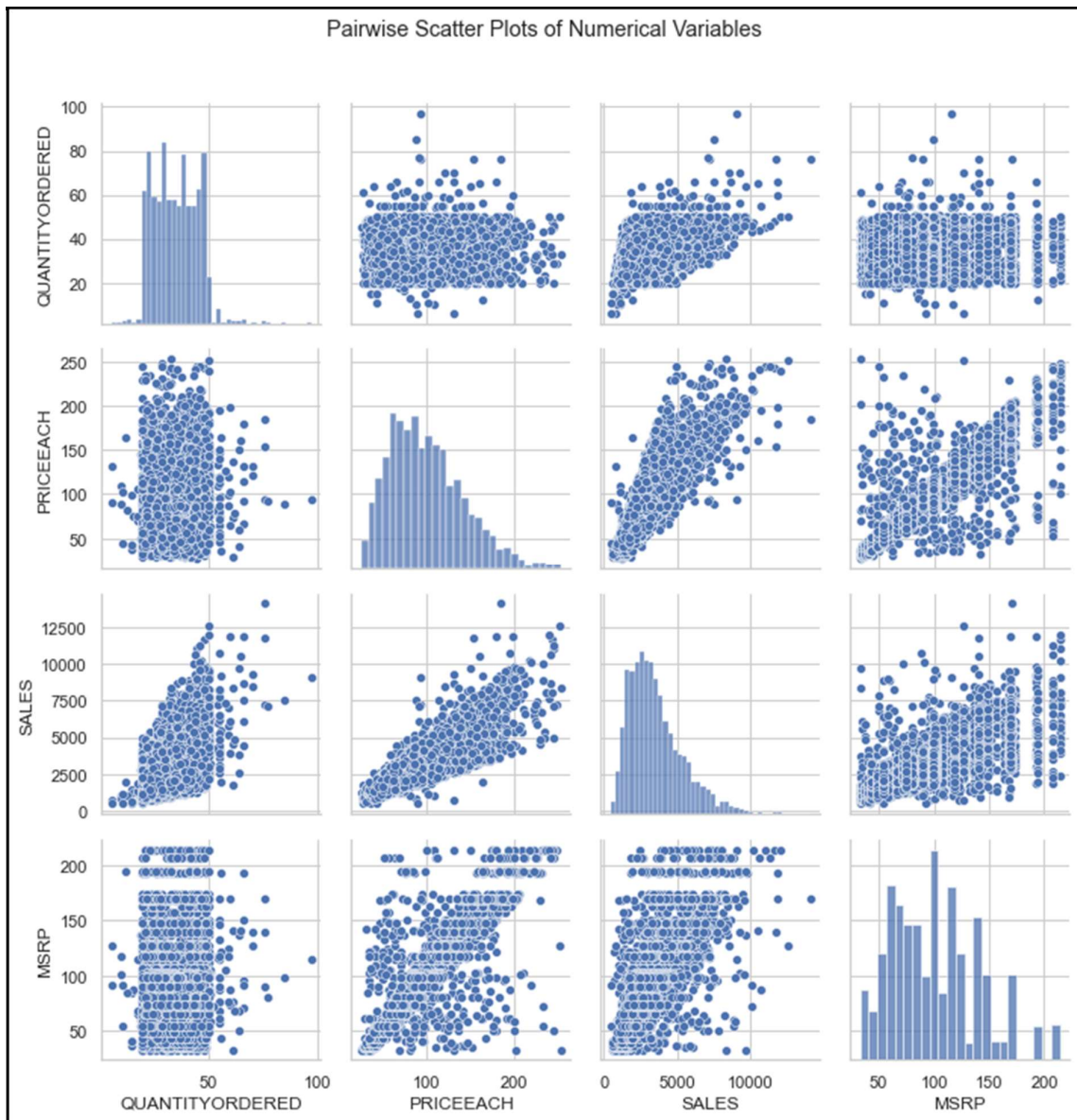


Figure 1. 7 Pair wise Scatter Plots of Numerical Variables

- The *PRICEEACH* attribute has a moderate positive correlation (0.78) with *SALES*, indicating that higher prices are associated with higher sales.
- The *QUANTITYORDERED* attribute has a moderate positive correlation (0.55) with *SALES*, implying that higher quantities ordered are linked to higher sales.

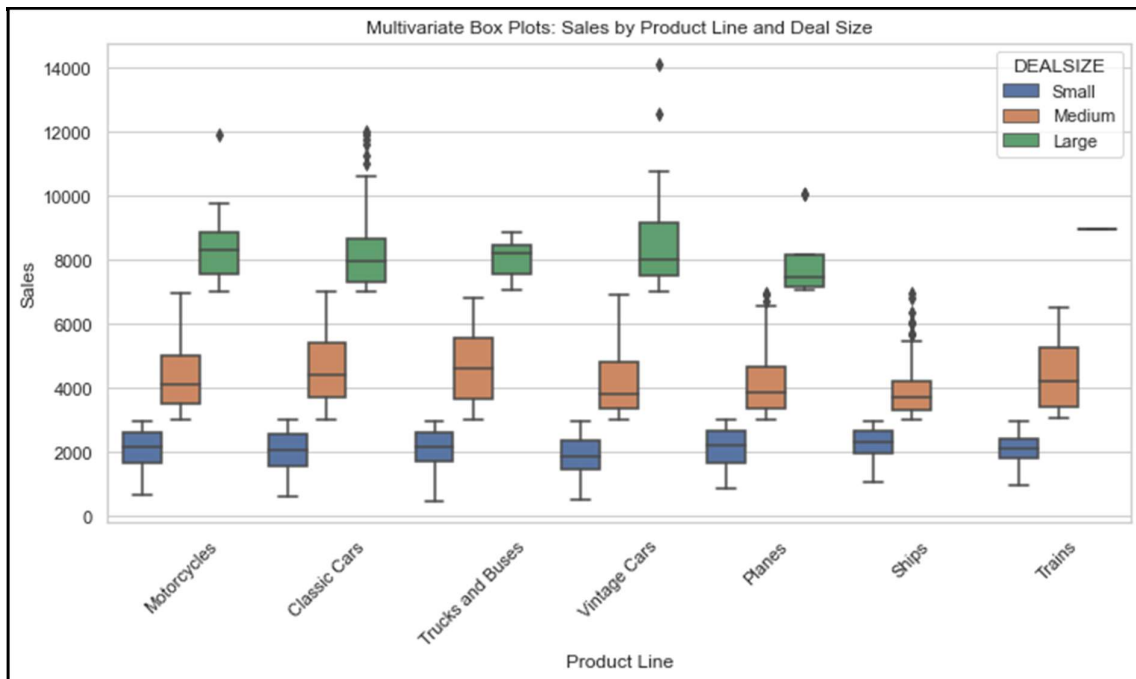


Figure 1. 8 Multivariate Analysis: Sales by Product Line and Deal Size

Customer Segmentation using RFM Analysis:

RFM stands for Recency, Frequency, and Monetary. It's a technique used to segment customers based on their purchasing behaviour. Each letter of RFM represents a key parameter:

- *Recency (R):* How recently a customer made a purchase.
- *Frequency (F):* How often a customer makes a purchase.
- *Monetary (M):* How much money a customer spends on purchases.

The RFM analysis involves assigning a score to each customer for each parameter, typically on a scale of 1 to 5 (with 5 being the highest). Then, customers are segmented based on these scores to create groups with similar behaviours. Common segmentation schemes include "Best Customers," "Loyal Customers," "Churn Risk," and "Low-Value Customers."

Assumptions:

- *Recency* is measured from the last transaction date to the current date.
- *Frequency* is calculated as the total number of transactions for a customer.
- *Monetary* is calculated as the total amount spent by a customer.

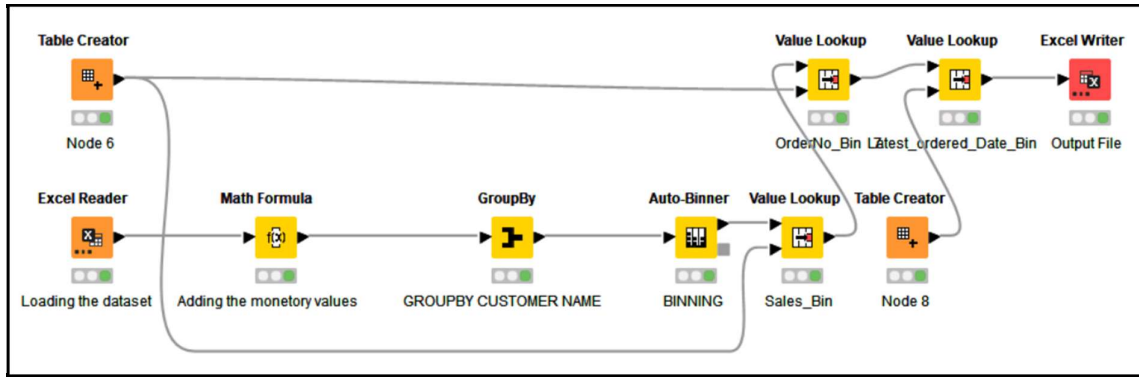


Figure 1. 9 KNIME Workflow Image

CUSTOMER NAME	QUANTITY ORDERED	PRICE EACH	COUNTRY	SALES	DAYS_SINCE_LAST ORDER	R_LABEL	F_LABEL	M_LABEL
AV Stores, Co.	34.8627	91.085	UK	157807.8	421	H	M	M
Alpha Cognac	34.35	101.16	France	70488.44	675	L	M	L
Amica Models & Co.	32.4231	110.85	Italy	94117.26	328	M	L	M
Anna's Decorations, Ltd	31.9348	106.42	Australia	153996.1	131	H	H	H
Atelier graphique	38.5714	92.239	France	24179.96	312	L	M	M
Australian Collectables, Ltd	30.6522	90.042	Australia	64591.46	1018	L	M	L
Australian Collectors, Co.	35.0182	104.59	Australia	200995.4	229	H	H	H
Australian Gift Network, Co	36.3333	110.55	Australia	59469.12	190	L	M	H
Auto Assoc. & Cie.	35.3889	99.488	France	64834.32	275	L	L	M
Auto Canal Petit	37.0741	94.255	France	93170.66	127	M	M	H
Auto-Moto Classics Inc.	35.875	92.8	USA	26479.26	1353	L	M	L

Table 1 Results in the output head of the data

- The final output of the RFM analysis is as above and the pivotal table for the segmentation could be found below.

Status	Recency	Frequency	Monetary		
			H	M	L
Active	H	H	11	1	0
		M	1	9	1
		L	0	0	0
At Risk	M	H	3	3	0
		M	5	11	1
		L	0	11	10
Churn	L	H	1	1	1
		M	1	6	5
		L	0	2	5

Table 2 Customer Segmentation with RFM Analysis

Best Customers (Top 5):

- **Anna's Decorations, Ltd** - High Monetary (M_LABEL), High Frequency (F_LABEL), High Recency (R_LABEL)
- **Australian Collectors, Co.** - High Monetary (M_LABEL), High Frequency (F_LABEL), High Recency (R_LABEL)
- **Euro Shopping Channel** - High Monetary (M_LABEL), High Frequency (F_LABEL), High Recency (R_LABEL)
- **Mini Gifts Distributors Ltd.** - High Monetary (M_LABEL), High Frequency (F_LABEL), High Recency (R_LABEL)
- **Rovelli Gifts** - High Monetary (M_LABEL), High Frequency (F_LABEL), High Recency (R_LABEL)

Customers on the Verge of Churning (Top 5):

- **Clover Collections, Co.** - Low Monetary (M_LABEL), Low Frequency (F_LABEL), Moderate Recency (R_LABEL)
- **Gift Ideas Corp.** - Low Monetary (M_LABEL), Low Frequency (F_LABEL), Low Recency (R_LABEL)
- **Muscle Machine Inc** - High Monetary (M_LABEL), High Frequency (F_LABEL), Moderate Recency (R_LABEL)
- **Quebec Home Shopping Network** - High Monetary (M_LABEL), High Frequency (F_LABEL), Low Recency (R_LABEL)
- **Signal Collectibles Ltd.** - Low Monetary (M_LABEL), Low Frequency (F_LABEL), Low Recency (R_LABEL)

Lost Customers (Top 5):

- **AV Stores, Co.** - High Monetary (M_LABEL), Low Frequency (F_LABEL), High Recency (R_LABEL)
- **Auto Assoc. & Cie.** - Low Monetary (M_LABEL), Low Frequency (F_LABEL), Low Recency (R_LABEL)
- **Herkku Gifts** - High Monetary (M_LABEL), High Frequency (F_LABEL), Moderate Recency (R_LABEL)
- **Super Scale Inc.** - Low Monetary (M_LABEL), Low Frequency (F_LABEL), Moderate Recency (R_LABEL)
- **Tokyo Collectables, Ltd** - High Monetary (M_LABEL), High Frequency (F_LABEL), Moderate Recency (R_LABEL)

Loyal Customers (Top 5):

- **Oulu Toy Supplies, Inc.** - High Monetary (M_LABEL), High Frequency (F_LABEL), Moderate Recency (R_LABEL)
- **Signal Gift Stores** - Moderate Monetary (M_LABEL), Moderate Frequency (F_LABEL), Moderate Recency (R_LABEL)
- **Tekni Collectables Inc.** - High Monetary (M_LABEL), High Frequency (F_LABEL), High Recency (R_LABEL)
- **The Sharp Gifts Warehouse** - High Monetary (M_LABEL), High Frequency (F_LABEL), High Recency (R_LABEL)
- **Vitachrome Inc.** - High Monetary (M_LABEL), High Frequency (F_LABEL), High Recency (R_LABEL)

Interpretation and Summary of the RFM Analysis

1. HHH Segment (High Recency, High Frequency, High Monetary):

- There are 11 customers who have recently made frequent purchases with high monetary spending.
- These are your most valuable and active customers.

2. HMH Segment (High Recency, Medium Frequency, High Monetary):

- There is 1 customer who has recently made moderate-frequency purchases with high monetary spending.
- This customer might be someone who has made several high-value purchases recently but is not as frequent as the HHH segment.

3. MMH Segment (Medium Recency, Medium Frequency, High Monetary):

- There are 11 customers who have made purchases with medium frequency and high monetary spending.

- *These customers might be consistently spending good amounts.*

4. MLL Segment (Medium Recency, Low Frequency, Low Monetary):

- *There are 10 customers who have made purchases with medium recency and low frequency and monetary spending.*
- *These customers might have made a few small purchases and haven't returned recently.*

5. LLL Segment (Low Recency, Low Frequency, Low Monetary):

- *There are 5 customers who have low recency, low frequency, and low monetary spending.*
- *These are potentially inactive customers.*

Conclusion

- *These inferences provide insights into different customer segments based on their recency, frequency, and monetary scores.*
- *Each segment has unique characteristics, allowing the store to tailor marketing strategies.*
- *For example,*
 - *loyal customers might be targeted with loyalty rewards,*
 - *at-risk customers could receive special offers,*
 - *churn-risk customers could be engaged with retention campaigns.*

Part B

Problem Statement

A grocery store shared the transactional data with you. Your job is to conduct a thorough analysis of Point of Sale (POS) data, identify the most commonly occurring sets of items in the customer orders, and provide recommendations through which a grocery store can increase its revenue by popular combo offers & discounts for customers.

Dataset

Grocery Store Data: dataset_group.csv.

Exploratory Analysis

	Date	Order_id	Product
0	01-01-2018	1	yogurt
1	01-01-2018	1	pork
2	01-01-2018	1	sandwich bags
3	01-01-2018	1	lunch meat
4	01-01-2018	1	all-purpose

Snippet 2. 1 Head of the Dataset

- Above, you could see the head of the dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20641 entries, 0 to 20640
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  ---
0   Date        20641 non-null  object
1   Order_id    20641 non-null  int64
2   Product     20641 non-null  object
dtypes: int64(1), object(2)
memory usage: 483.9+ KB
```

Snippet 2. 2 Information of the dataset

- We have total 3 columns with no null values
- There are two objective and integer type of attributes

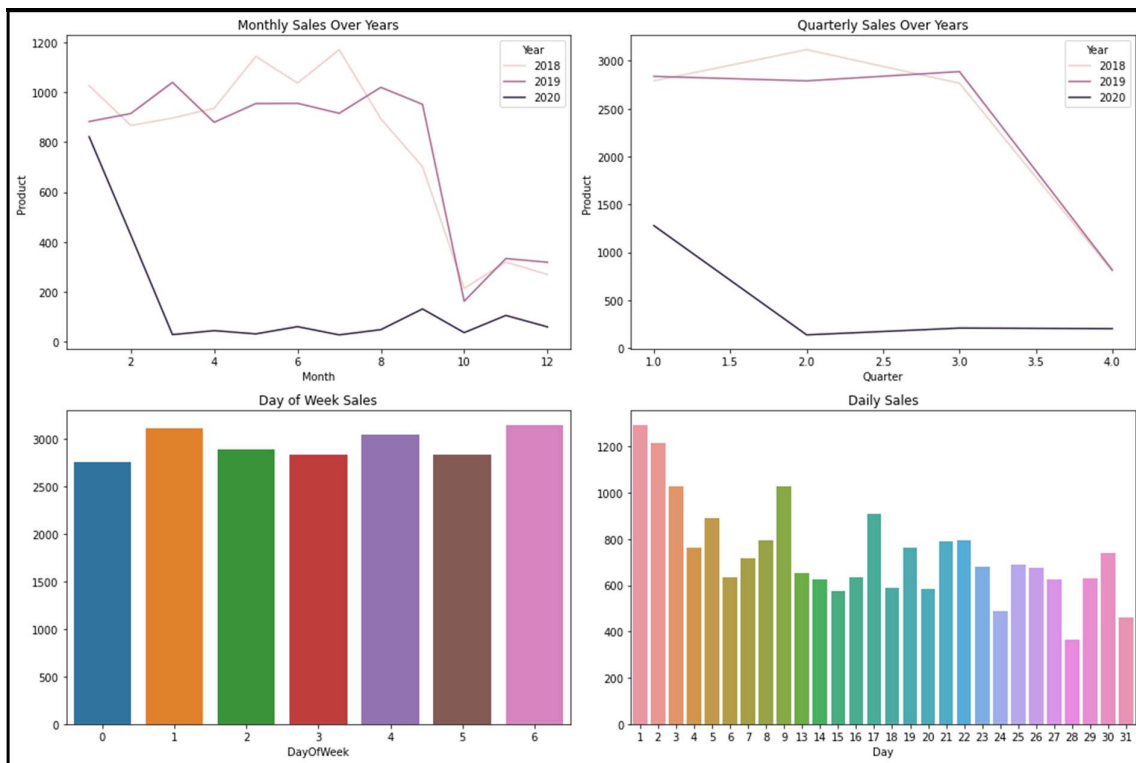


Figure 2. 1 Daily, Weekly, Quarterly and Yearly Sales Visualization

- *The Monthly sales over years doesn't show any regular trend or seasonality*
- *In 2020, the monthly sales were the lowest compared to 2018-19*
- *Monday and Saturday shall have the highest sales over the days*
- *Every month's starting days would have the higher sales and gradually decreases when see at the end of the month*

Association Rules and Their Relevance:

- *Association rules are a fundamental concept in data mining and analytics, commonly used in Market Basket Analysis.*
- *They reveal interesting relationships and patterns within large datasets by identifying co-occurring items in transactions.*
- *In the context of a grocery store's transactional data, association rules can provide insights into how different products are often purchased together.*
- *This information is invaluable for optimizing product placement, devising strategic cross-promotions, and enhancing customer experience.*

In the case of the grocery store, association rules can help in several ways:

- *Product Bundling: Identifying items frequently purchased together can lead to bundling products into combo offers, enhancing customer value and boosting sales.*
- *Store Layout: Placing co-occurring products near each other can increase convenience and encourage additional purchases.*
- *Targeted Marketing: Leveraging associations can facilitate targeted marketing campaigns to specific customer segments.*
- *Inventory Management: Understanding product relationships aids in efficient stock management and supply chain optimization.*

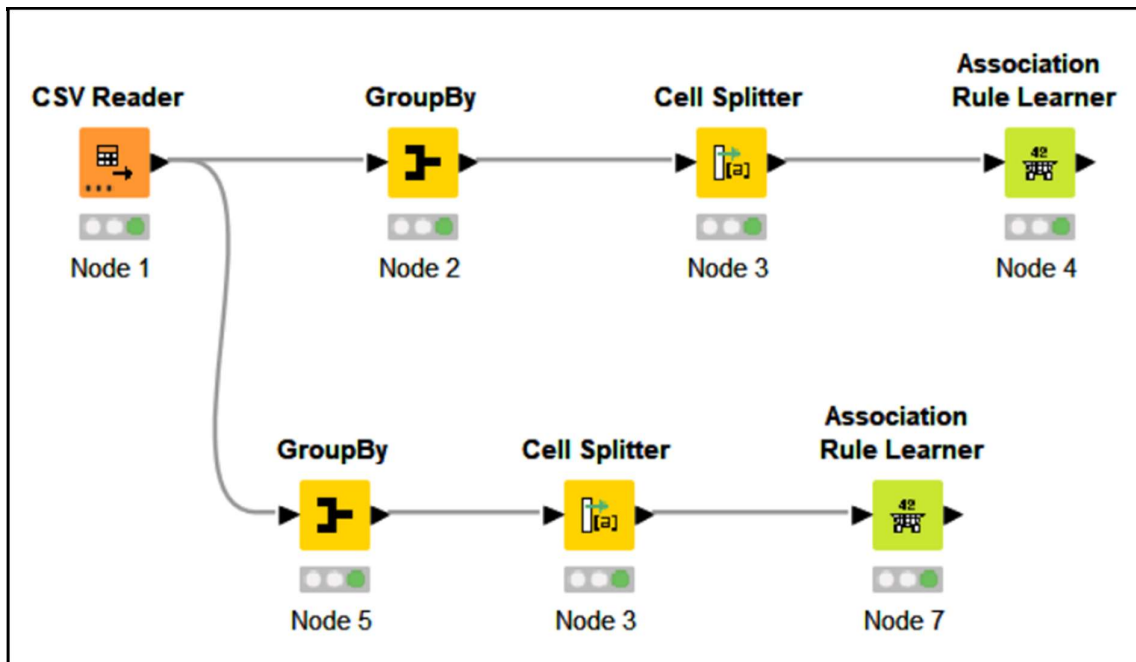


Figure 2. 2 KNIME workflow Image

Threshold Values of Support and Confidence:

- *Support and confidence are critical parameters in association rule analysis. They define the significance of the discovered patterns:*
- *Support: These measures how frequently a set of items appears in the dataset. A higher support value implies that the rule is more common. A low support threshold might lead to rules that are too specific, while a high threshold might miss interesting associations.*
- *Confidence: This quantifies how often the consequent (right-hand side) of the rule occurs when the antecedent (left-hand side) is present. A high confidence suggests a strong association. Too low a confidence might lead to too many false-positive associations.*
- *Selecting appropriate threshold values involves a trade-off between the number of discovered rules and their significance. It's often a trial-and-error process. Start with higher thresholds and gradually lower them until you achieve a balance between relevant rules and manageable results.*
- *In the KNIME workflow, you can adjust these threshold values based on the nature of the data and the goals of your analysis. Iterative experimentation can help you identify the thresholds that provide the most valuable insights.*

Associations Identified and Their Significance:

- Below is the tabular representation of the identified associations along with their corresponding support, confidence, and lift values:

Support	Confidence	Lift	Consequent	Implications	Items
0.050	0.640	1.700	juice	<---	[yogurt, toilet paper, aluminium foil]
0.050	0.620	1.645	juice	<---	[yogurt, poultry, aluminium foil]
0.050	0.613	1.616	coffee/tea	<---	[yogurt, cheeses, cereals]
0.050	0.600	1.424	poultry	<---	[dishwashing liquid/detergent, laundry detergent, mixes]
0.051	0.630	1.678	mixes	<---	[yogurt, poultry, aluminium foil]
0.051	0.611	1.660	sandwich bags	<---	[cheeses, bagels, cereals]
0.051	0.674	1.726	cheeses	<---	[bagels, cereals, sandwich bags]
0.051	0.617	1.558	cereals	<---	[cheeses, bagels, sandwich bags]
0.051	0.630	1.621	dinner rolls	<---	[spaghetti sauce, poultry, cereals]
0.051	0.637	1.512	poultry	<---	[dinner rolls, spaghetti sauce, cereals]
0.051	0.604	1.589	milk	<---	[poultry, laundry detergent, cereals]
0.052	0.628	1.610	eggs	<---	[dinner rolls, poultry, soda]
0.052	0.641	1.649	dinner rolls	<---	[spaghetti sauce, poultry, ice cream]
0.052	0.686	1.628	poultry	<---	[dinner rolls, spaghetti sauce, ice cream]
0.052	0.628	1.614	dinner rolls	<---	[spaghetti sauce, poultry, juice]
0.052	0.602	1.429	poultry	<---	[dinner rolls, spaghetti sauce, juice]
0.052	0.634	1.627	eggs	<---	[paper towels, dinner rolls, pasta]
0.052	0.602	1.621	pasta	<---	[paper towels, eggs, dinner rolls]
0.054	0.642	1.651	dinner rolls	<---	[spaghetti sauce, poultry, laundry detergent]
0.054	0.656	1.556	poultry	<---	[dinner rolls, spaghetti sauce, laundry detergent]
0.055	0.624	1.565	ice cream	<---	[paper towels, eggs, pasta]
0.055	0.630	1.616	eggs	<---	[paper towels, ice cream, pasta]
0.055	0.643	1.731	pasta	<---	[paper towels, eggs, ice cream]
0.055	0.649	1.791	paper towels	<---	[eggs, ice cream, pasta]

Understanding Support, Confidence, and Lift Values:

Support:

- Support indicates the frequency of an itemset in the dataset. For example, a support of 0.050 means that the itemset appears in approximately 5% of all transactions.
- Higher support values imply that the association is relatively common.

Confidence:

- Confidence measures how often the consequent occurs when the antecedent is present. For instance, a confidence of 0.640 means that juice is purchased along with the other items in about 64% of the cases where those items are present.

- Higher confidence values imply a stronger association between the antecedent and the consequent.

Lift:

- Lift quantifies how much more likely it is that the two items will be bought together compared to if they were bought independently. Lift values greater than 1 suggest a positive association between the items.
- A lift of 1 implies independence, while a lift greater than 1 implies a positive relationship.

Interpretation

- For instance, the association rule "juice <--- [yogurt, toilet paper, aluminium foil]" has a support of 0.050, suggesting that this combination is present in 5% of transactions.
- The confidence value of 0.640 indicates that juice is purchased along with the other items in about 64% of cases where yogurt, toilet paper, and aluminium foil are purchased.
- A lift of 1.700 suggests that the purchase of juice is 1.7 times more likely when yogurt, toilet paper, and aluminium foil are purchased together than when they are bought independently.

Suggestions for Lucrative Combos with Discount Offers:

Based on the associations identified through the Association Rule analysis, we can recommend the following potential combos with lucrative offers for the grocery store. These offers are designed to capitalize on the relationships between products and encourage customers to make additional purchases, ultimately increasing revenue.

Combo: Breakfast Boost

- Offer: Buy a pack of yogurt, a box of cereal, and a carton of milk together at a discounted price.
- Rationale: The association rules show strong connections between yogurt, cereals, and milk. By bundling these items, customers are more likely to make a complete breakfast purchase.

Combo: Family Pasta Night

- Offer: Purchase a jar of spaghetti sauce, a pack of pasta, and a poultry product to get a special discount.
- Rationale: The analysis reveals associations between spaghetti sauce, poultry, and pasta. Offering a combo deal for a family pasta night can entice customers to buy all the necessary components.

Combo: Snack Delight

- Offer: Combine cheeses, bagels, and sandwich bags in a single purchase for a reduced price.

- *Rationale: The strong associations between cheeses, bagels, and sandwich bags suggest that customers often buy these items together. Providing them as a combo can enhance convenience and save money for the shoppers.*

Combo: Cleaning Essentials

- *Offer: Bundle dishwashing liquid/detergent, laundry detergent, and cleaning mixes at a special offer price.*
- *Rationale: The association rule linking these cleaning items together can prompt customers to stock up on their household cleaning essentials.*

Combo: Refreshing Beverage Pack

- *Offer: Purchase juice, coffee/tea, and soda together with a discount.*
- *Rationale: The associations between these beverages indicate potential for cross-promotions, targeting customers who enjoy a variety of drinks.*

Combo: Weekend Indulgence

- *Offer: Buy a variety of products like ice cream, dinner rolls, and other treats as a weekend package at a lower cost.*
- *Rationale: Combining indulgent items associated with weekend enjoyment can drive increased sales, especially during leisure days.*

Additional Strategies:

- *"Buy Two, Get One Free" Deals: Implementing "buy two, get one free" or similar promotions on products that have strong associations can incentivize customers to purchase more and capitalize on the products' relationships.*
- *Promotion Visibility: Display products in these combos together, either physically in-store or digitally in an online platform, to remind customers of the enticing deals.*
- *Loyalty Programs: Introduce loyalty programs that offer exclusive discounts and rewards to customers who consistently purchase products from specific combos.*
- *Limited-Time Offers: Creating a sense of urgency with limited-time offers can encourage customers to make quicker purchasing decisions.*

Conclusion

- *Remember that the success of these offers lies in effective communication, clear value propositions, and aligning the combos with customer preferences.*
- *Regularly analysing sales data and customer feedback will allow for fine-tuning these strategies over time.*