MRA Sale Report

BHUPESH UPADHYAY

Summery

- The data is about the sales of Automobile parts to the vendor since last 3 years.

- Agenda of the this report is get insight from the data and recommend client better solution for the business decision.

- The given data has sales details with the sales date, sales figure and other demographic details.

Contents of the PPT

- Exploratory analysis and inferences of the data given part A business problem with the suitable graphs and diagram related to the problem statement.
- Customer segmentation using RMF Analysis with the parameter used during the analysis. Knime workflow has been shown in the process of part A.
- Result of RFM analysis with Top 5 customers list, top 5 customers on the verge of churning, top 5 lost customers and top 5 loyal customers of part A.
- Exploratory analysis and inferences of the data given part B business problem with the suitable graphs and diagram related to the problem statement.

Contents of the PPT

- Market Basket Analysis (Association Rules) of the part B data with explanation of association rule relevance to this case and knime workflow presentation.

- Identification of the associations with tabular form presentation and explanation about support, confidence & & lift values that are calculated

- Suggestion of Possible Combos with Lucrative Offers with the recommendations and discount offers or combos (or buy two get one free) based on the associations.

Data Info.

- The dataset has 20 columns and 2746 rows.
- The dataset contain 2 floats, 6 integers and 12 objects data types.
- There is no null values.

#	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

Summary Stats

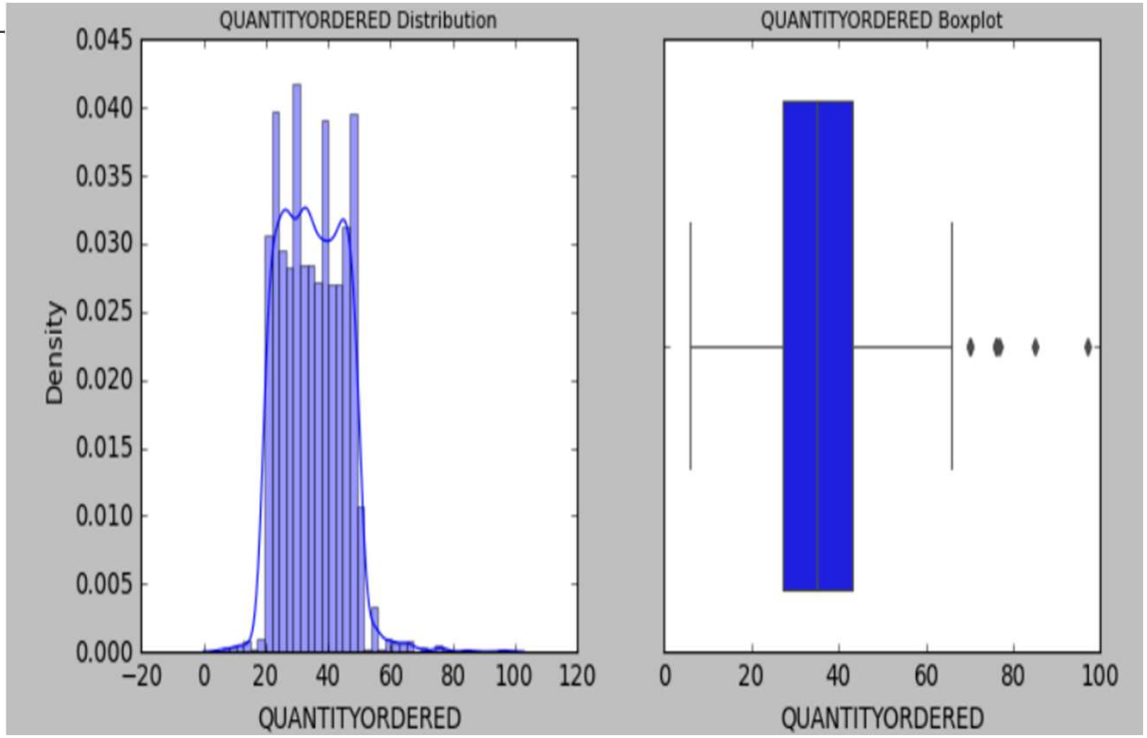
	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	DAYS_SINCE_LASTORDER	MSRP
count	2747	2747	2747	2747	2747	2747	2747	2747
mean	10259.76	35.10302	101.099	6.491081	3553.048	43598.91	1757.086	100.6917
std	91.87752	9.762135	42.04255	4.230544	1838.954	230.2313	819.2806	40.1148
min	10100	6	26.88	1	482.13	43106	42	33
25%	10181	27	68.745	3	2204.35	43412	1077	68
50%	10264	35	95.55	6	3184.8	43640	1761	99
75%	10334.5	43	127.1	9	4503.095	43786	2436.5	124
max	10425	97	252.87	18	14082.8	43982	3562	214

- Total Maximum no of orders are 97 units and average is 35 units
- The maximum unit price is 252 and average is 101.
- The most recent sale has been done 42 days back and average recency is 1757 days.

Exploratory Analysis and Inferences

UNIVARIATES

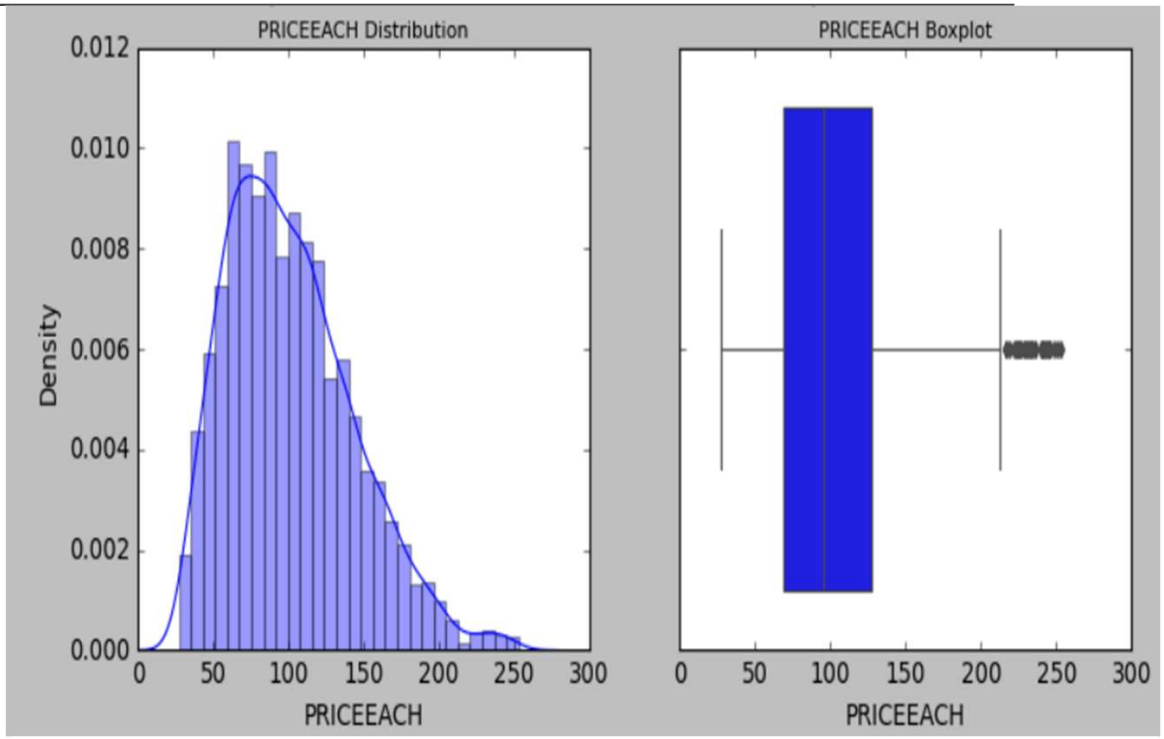
- As per the distribution images we can conclude that the most of the concentration 20 to 60 units
- As per the boxplot there is outlier in the dataset.



Exploratory Analysis and Inferences

UNIVARIATES

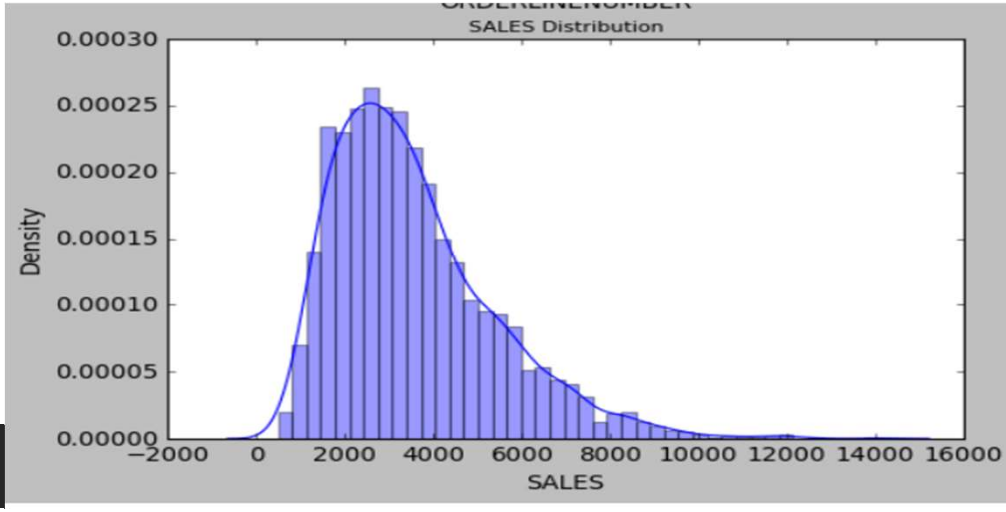
- As per the distribution images we can conclude that the price of the items are mostly 75 to 120. distribution inclined towards right side.
- As per the boxplot there is outlier in the dataset.

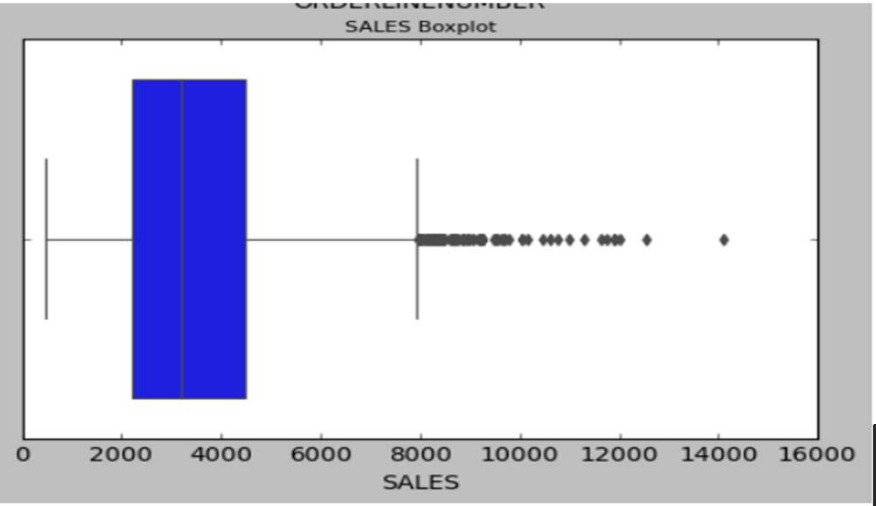


Exploratory Analysis and Inferences

UNIVARIATES

- As per the distribution images we can conclude that the sales of the items are mostly 2000 to 4000. distribution inclined towards right side.
- As per the boxplot there is outlier in the dataset.

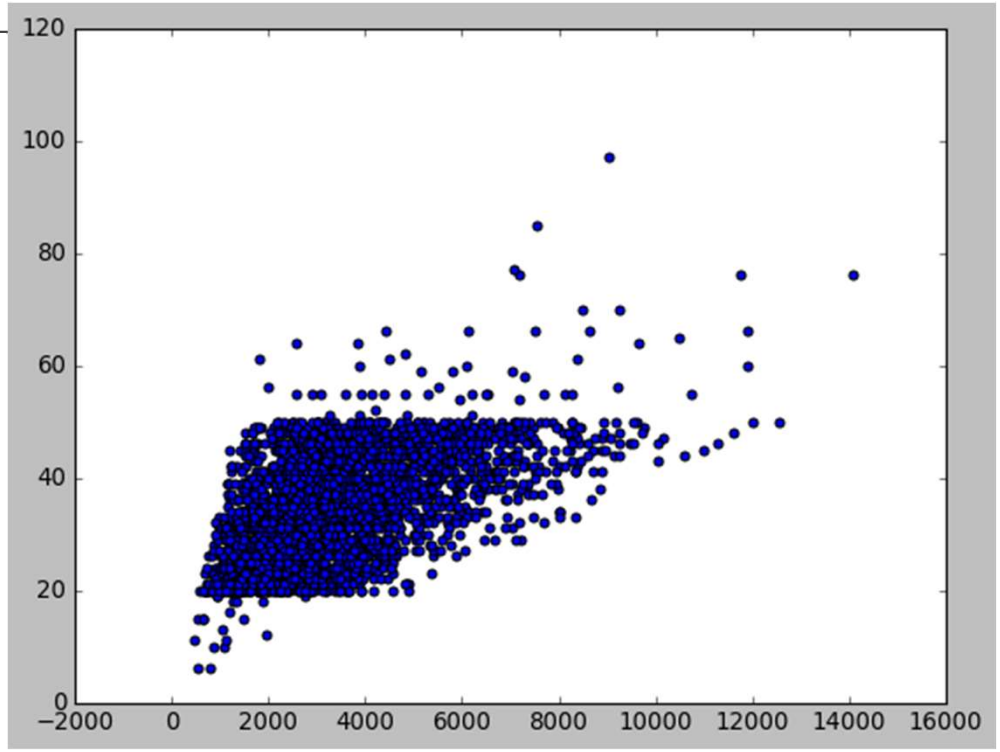




Exploratory Analysis and Inferences

BIVARIATE

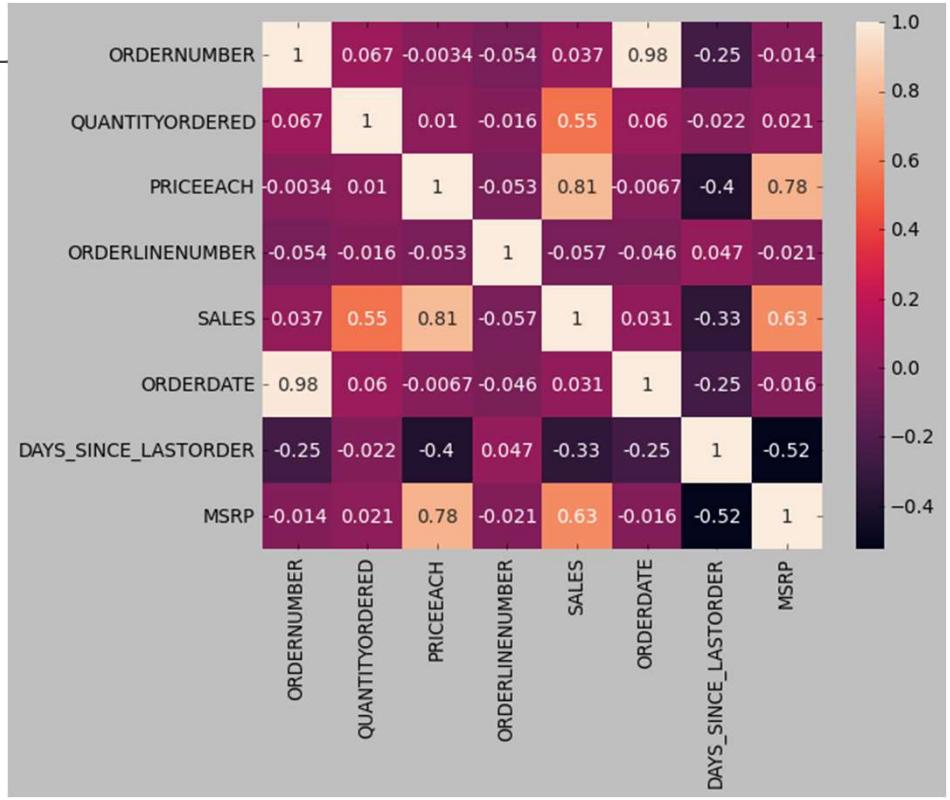
- The below is the relationship b/w Sales and the Quantity ordered.
- In the image it showing that the raltionship b/w sales and quantity is positive.



Exploratory Analysis and Inferences

MULTIVARIATE

- Unit price and MRP has the positive relation.
- Sales and unit price has the positive relation.
- Sales and MRP has the positive relation.



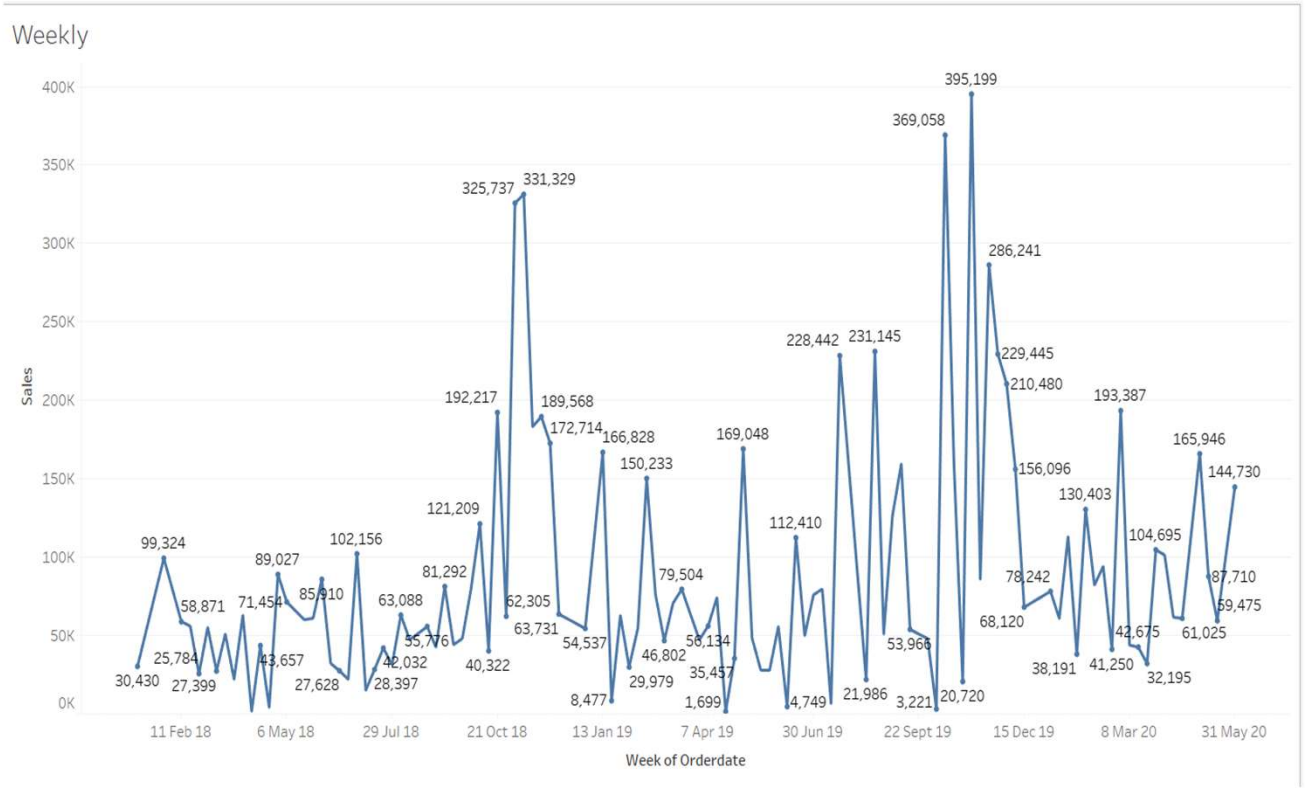
Exploratory Analysis and Inferences

WEEKLY SALES

- In weekly sales data it is depicted that week by week sales are increasing.
- In first week of the November 2019 is highest 395,199. In third week April 2019 sale is lowest 1699.

TABLEAU ACCESS

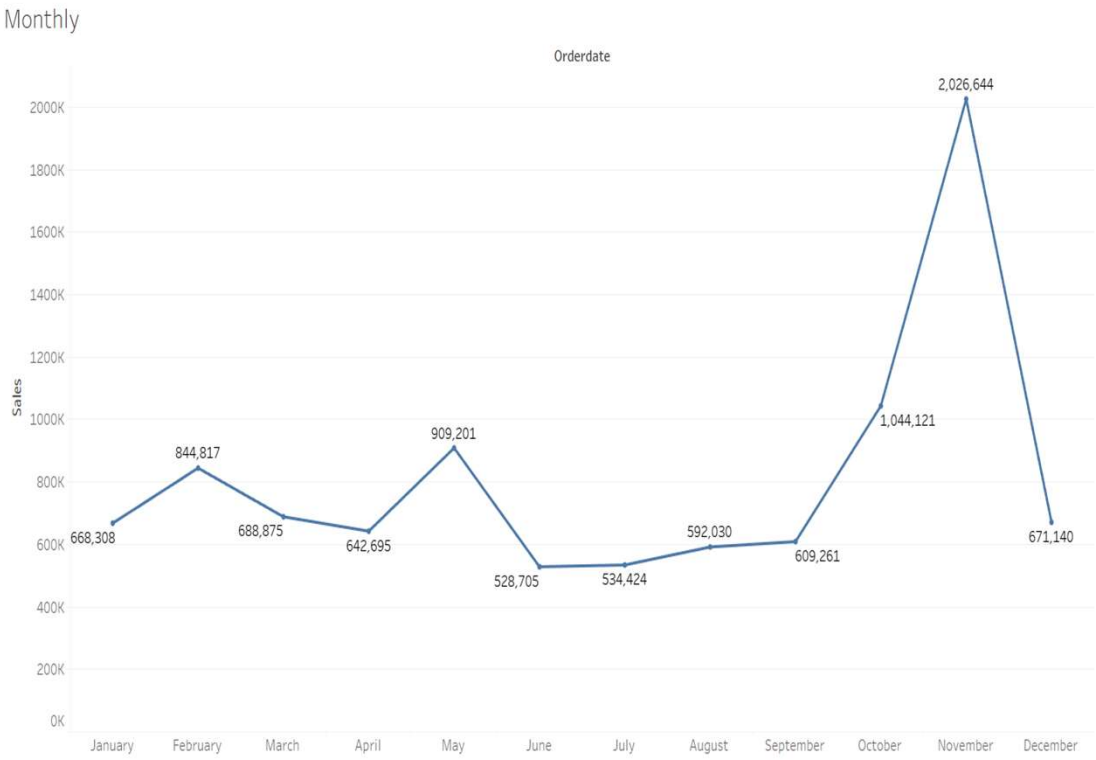
- https://public.tableau.com/app/profile/bhupesh.upadhyay/viz/MRAProjectBhupeshUpadhyay/TrendMStatus?publish=yes



Exploratory Analysis and Inferences

MONTHLY SALES

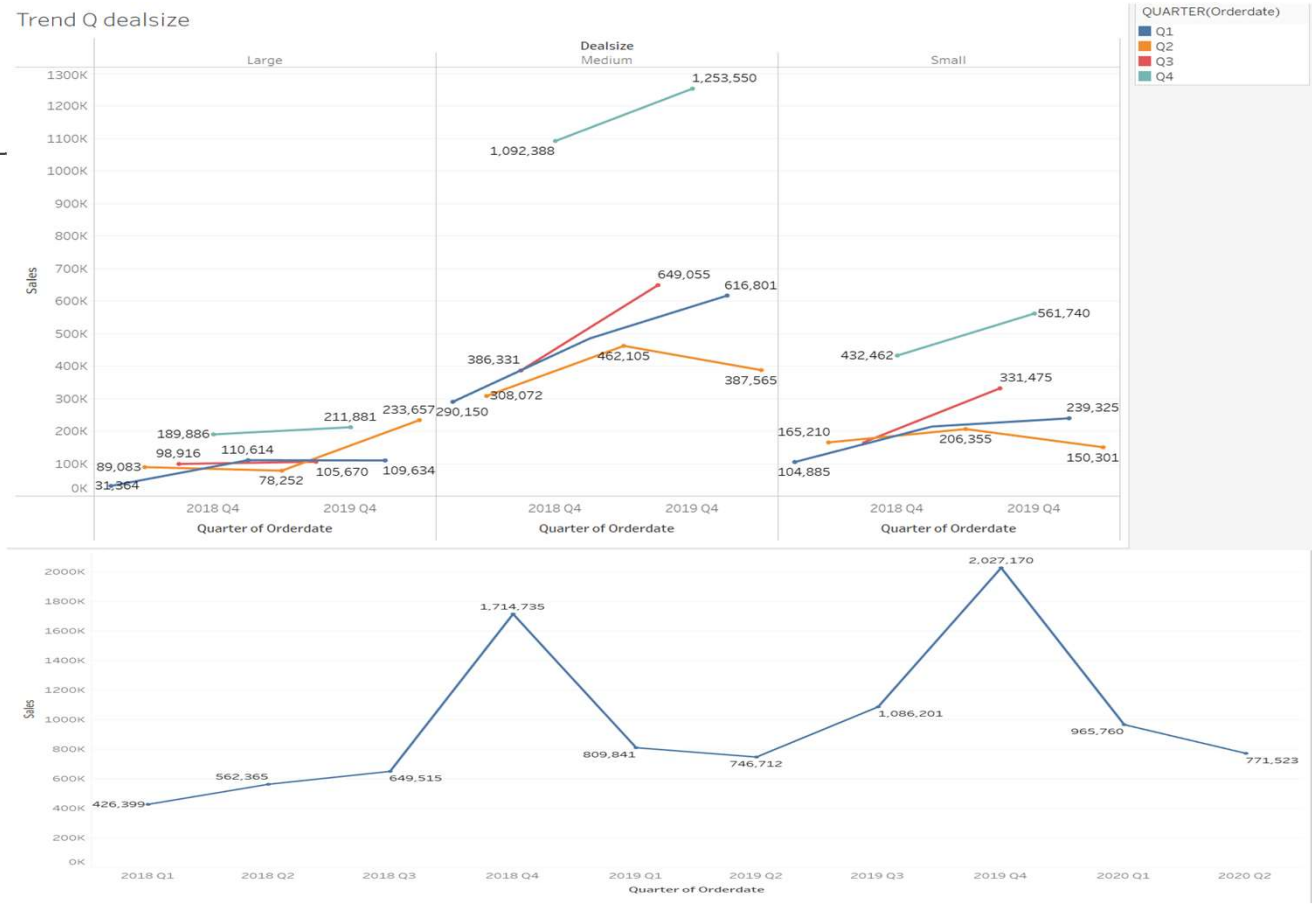
- In monthly sales data it is depicted that month by month the sales is uneven.
- In the month of the November is highest 2,026,644. In the month of June sale is lowest 528,705.



Exploratory Analysis and Inferences

QUARTERLY SALES

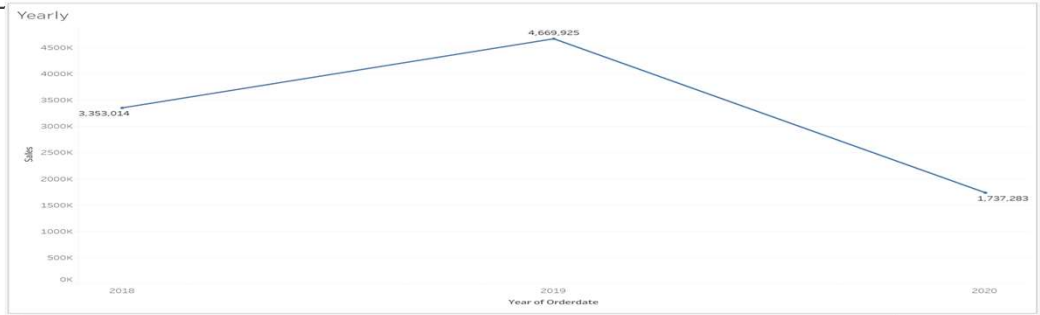
- In Quarterly sales data it is depicted that some year is high and in some year is low.
- In Q4 2018 and in Q4 2019 the sales is highest 1,714,735 and 2,027,170.
- in Q4 2019 in medium deal size has the highest sales of 1,253,550

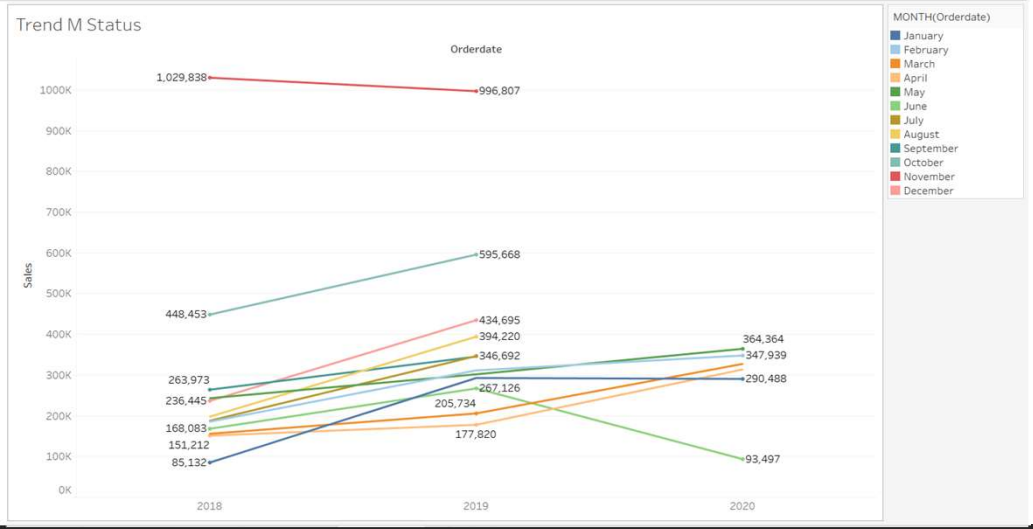


Exploratory Analysis and Inferences

YEARLY SALES

- In yearly sales data it is depicted that sale is going down.
- In 2018 the sale is highest of 4,669,925.
- Year on year in almost every month sales has been increased

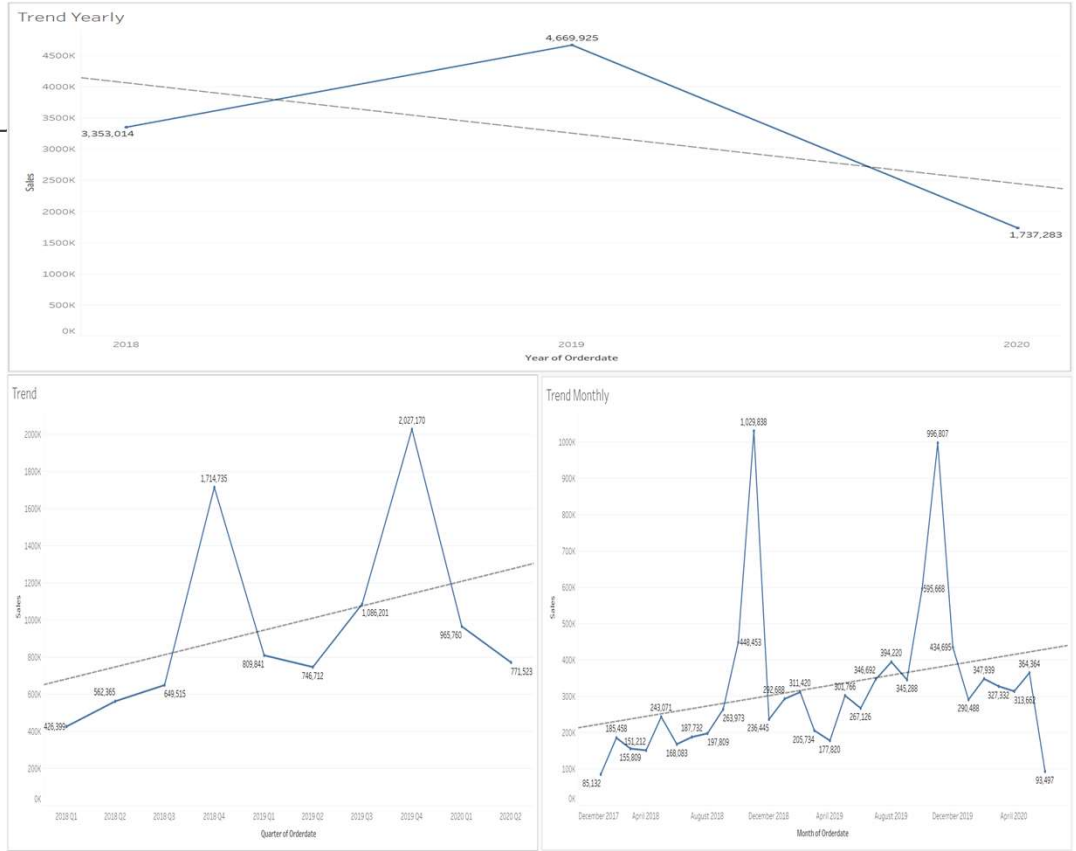




Exploratory Analysis and Inferences

- Monthly and Quarterly showing upward trend.

- But yearly trend is showing downward trend.



Customer Segmentation using RFM analysis

- **RFM stands for Recency, Frequency, and Monetary Value.** It is a marketing technique used to quantitatively rank and group customers based on the recency, frequency, and monetary total of their recent transactions to identify the best customers and perform targeted marketing campaigns.
- **Recency:** This measures how recently the customer has made a purchase. A customer who has made a purchase recently is more likely to be interested in making another purchase than a customer who has not made a purchase in a long time.
- **Frequency:** This measures how often the customer makes purchases. A customer who makes purchases frequently is more likely to be a loyal customer than a customer who makes purchases only occasionally.
- **Monetary value:** This measures how much money the customer spends each time they make a purchase. A customer who spends more money each time they make a purchase is more valuable to the company than a customer who spends less money.

Customer Segmentation using RFM analysis

PARAMETERS USED AND ASSUMPTIONS MADE

- **Price each** - This column specifies the price of each item in the order which will be used in the calculation of the Monetary.
- **Quantity Ordered** - It indicates the number of items ordered in each order which will be used in the calculation of the Monetary.
- **SALES** - This column denotes the total sales amount for each order, which is calculated by multiplying the quantity ordered by the price of each item.

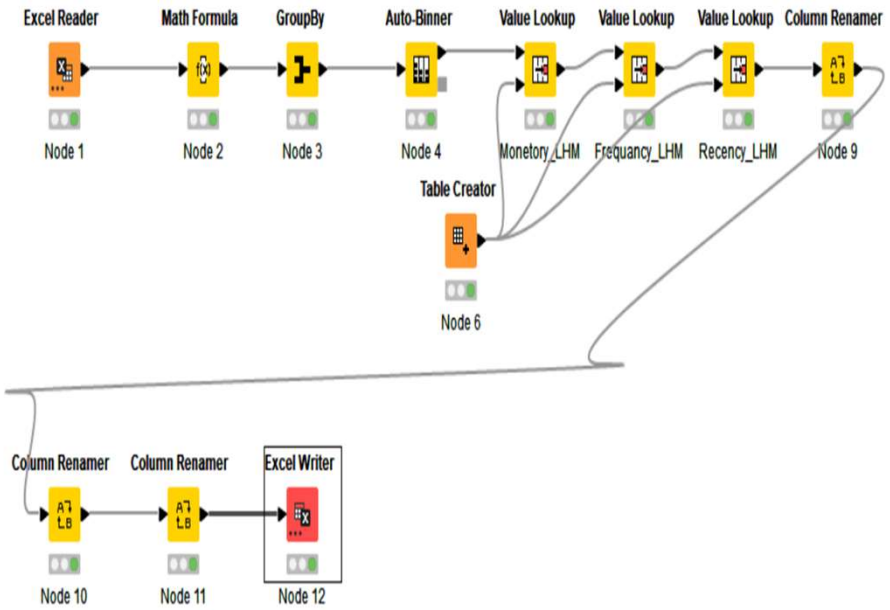
Customer Segmentation using RFM analysis

PARAMETERS USED AND ASSUMPTIONS MADE

- ORDERNUMBER - This column represents the unique identification number assigned to each order which will use for calculating Frequency.
- DAYS_SINCE_LASTORDER - This column represents the number of days that have passed since the last order for each customer. It can be used to analyze customer purchasing patterns.

Customer Segmentation using RFM analysis

Node	output
Excel Reader	File uploaded to the Knime
Math Formula	Monetary column
Group by	new hile crerated with new columns
Auto - binner	4 new bins created
Table creater	bins interchanged with ABCD columns name
Value lookup	to execute the ABCD columns
Column Renamer	RFM columns renamed
Excer writer	File saved to the local directory



Inferences from RFM Analysis and identified segments

BEST CUSTOMERS

CUSTOMERNAME	Monetary	Monatory_ABCD	Frequancy_ABCD	Recency_ABCD
Euro Shopping Channel	912294.11	D	D	A
Mini Gifts Distributors Ltd.	654858.06	D	D	A
Australian Collectors, Co.	200995.41	D	D	A
Muscle Machine Inc	197736.94	D	D	C
La Rochelle Gifts	180124.9	D	D	A

PARAMETERS INDICATORS ABCD

- D denotes the highest value A denotes lowest value.
- In the table we can see that the customer highest value of Monetary and frequency and lowest recency.
- Lowest recency indicate that customer take small break to purchase the which is the good sign.

Inferences from RFM Analysis and identified segments

CUSTOMERS ARE ON THE VERGE OF CHURNING

CUSTOMERNAME	Monetary	Monatory_LMH	Frequancy_LMH	Recency_LMH
Collectable Mini Designs Co.	87489.23	B	B	C
Motor Mint Distributors Inc.	83682.16	B	B	C
Mini Caravy	80438.48	B	A	C
Super Scale Inc.	79472.07	B	A	C
Petit Auto	74972.52	B	B	C

PARAMETERS INDICATORS ABCD

- D denotes the highest value A denotes lowest value.
- In the table we can see that the customer has the low frequency, Minatory and high recency value which indicated that customer is on verge of churning
- Lowest recency indicate that customer take long break to purchase the which is a bad sign.

Inferences from RFM Analysis and identified segments

LOST CUSTOMERS

CUSTOMERNAME	Monatory	Monatory_ABCD	Frequancy_ABCD	Recency_ABCD
Auto-Moto Classics Inc.	26479.26	A	A	D
Royale Belge	33440.1	A	A	D
Bavarian Collectables Imports, Co.	34993.92	A	A	D
Double Decker Gift Stores, Ltd	36019.04	A	A	D
Signal Collectibles Ltd.	50218.51	A	A	D

PARAMETERS INDICATORS ABCD

- D denotes the highest value A denotes lowest value.
- In the table we can see that the customer has lowest minatory value lowest frequency value and highest recency
- It indicates these customers are not purchasing items and they are under the list of lost customers.

Inferences from RFM Analysis and identified segments

LOYAL CUSTOMERS

CUSTOMERNAME	Frequancy	Monatory_LMH	Frequancy_LMH	Recency_LMH
Euro Shopping Channel	42	D	D	A
Mini Gifts Distributors Ltd.	219	D	D	A
Australian Collectors, Co.	229	D	D	A
La Rochelle Gifts	139	D	D	A
AV Stores, Co.	421	D	D	C

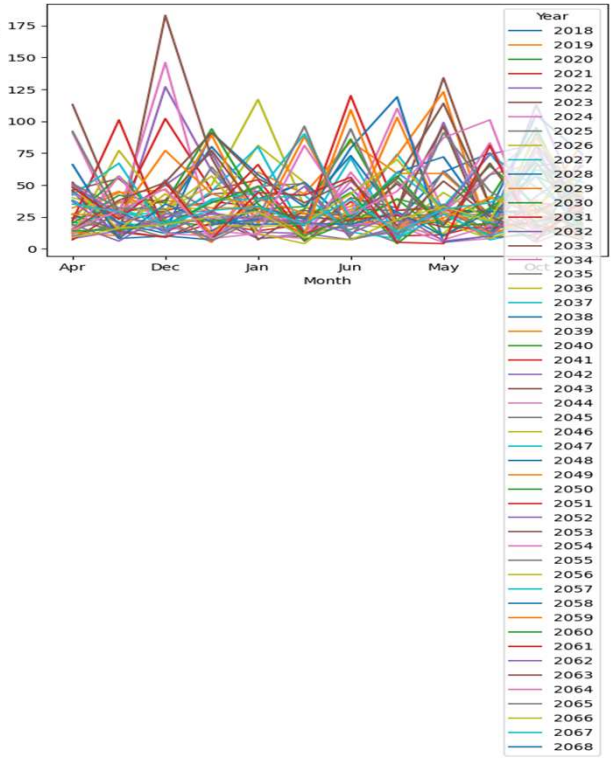
PARAMETERS INDICATORS ABCD

- D denotes the highest value A denotes lowest value.
- In the table we can see that the customer has highest monetary value highest frequency value and lowest recency
- It indicates these customers are the most loyal customers who frequently buys items. For proving the loyalty frequency must be high. May be monetary value would be low.

Exploratory Analysis of data & an executive summary

- To calculate the trend and insights from the data, we can group the data by product and use the count formula to get the exact value sold for each product. This will give us a detailed overview of how many times each product has been sold and how much revenue it has generated.
- This information can be used to identify popular products, products that are sold together, and products that are not selling well. This information can then be used to improve marketing campaigns and product placement.
- For example, if we see that a particular product is not selling well, we can offer discounts or promotions to encourage customers to buy it. We can also place the product in a more prominent location in the store.

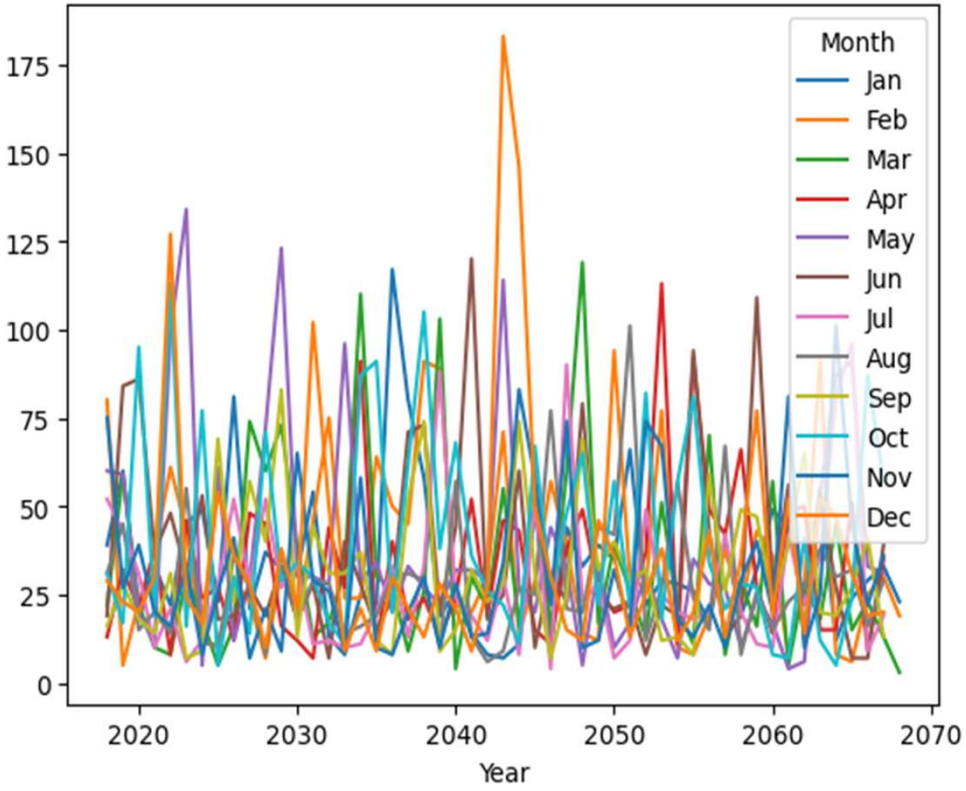
Exploratory Analysis of data & an executive summary



- The graph shows that the sales of all goods have been increasing over time. However, the rate of increase has varied from year to year. For example, the sales of goods increased significantly from 2018 to 2019, but the rate of increase slowed down from 2020 to 2021.
- The graph also shows that there are some goods that are more popular than others. For example, the sales of goods have been consistently high for goods like "all-purpose", "aluminum foil", "bagels", and "beef". However, the sales of goods like "poultry" and "spaghetti sauce" have been more volatile, with periods of high and low sales.
- Overall, the graph shows that the sales of all goods have been increasing over time. However, the rate of increase has varied from year to year, and some goods are more popular than others.

Exploratory Analysis of data & an executive summary

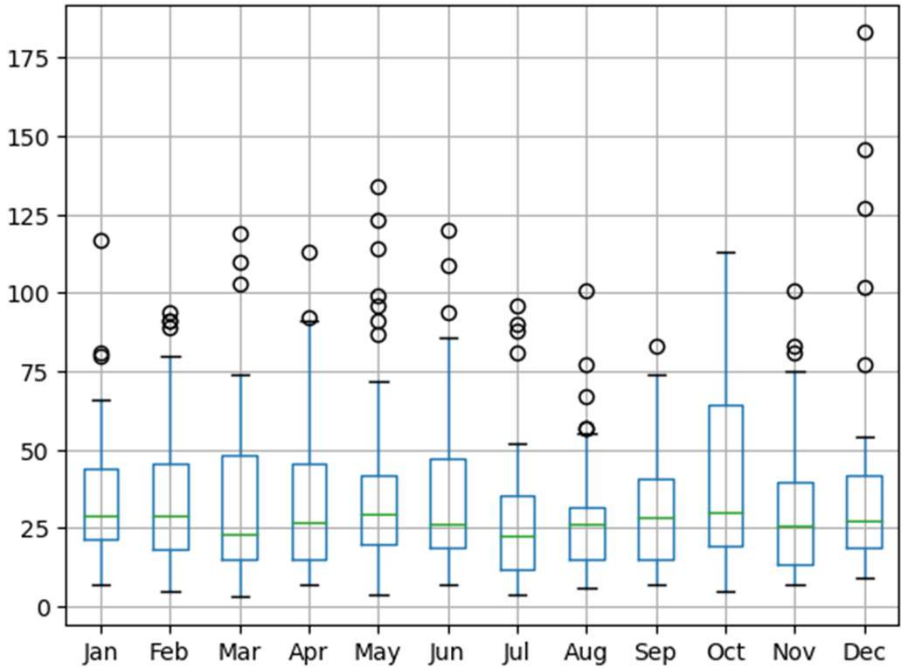
- January: The sales of all goods were relatively low in January. The only goods that had significant sales were "all-purpose" and "aluminum foil".
- February: The sales of all goods increased in February, with the exception of "poultry".
- March: The sales of all goods continued to increase in March, with the sales of "beef" and "butter" being particularly high.
- April: The sales of all goods peaked in April, with the sales of "all-purpose" being the highest.
- May: The sales of all goods declined in May, but remained relatively high.
- June: The sales of all goods continued to decline in June, but remained above the levels of January and February.
- July: The sales of all goods rebounded in July, with the sales of "all-purpose" and "aluminum foil" being particularly high.
- August: The sales of all goods continued to increase in August, with the sales of "beef" and "butter" being particularly high.
- September: The sales of all goods peaked again in September, with the sales of "all-purpose" being the highest.
- October: The sales of all goods declined in October, but remained relatively high.
- November: The sales of all goods continued to decline in November, but remained above the levels of January and February.
- December: The sales of all goods rebounded in December, with the sales of "all-purpose" and "aluminum foil" being particularly high.



Exploratory Analysis of data & an executive summary

- Every month have the outliers .

- The graph shows that the sales of all goods have been increasing over time. However, the rate of increase has varied from month to month. For example, the sales of goods increased significantly from January to February, but the rate of increase slowed down from March to April..

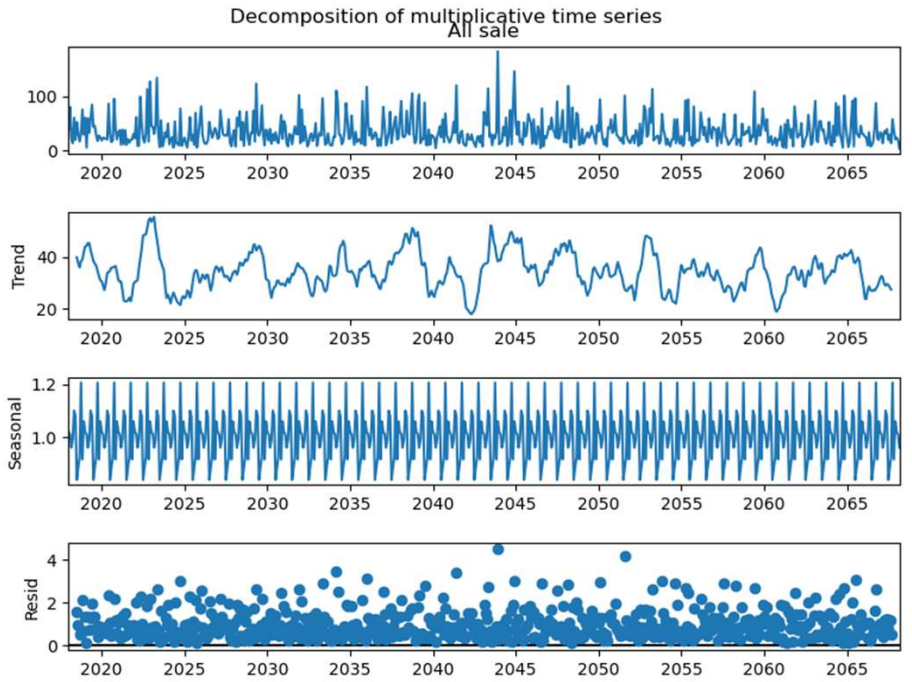


Exploratory Analysis of data & an executive summary

- After the decomposing the data we can make some occlusion on the basis of that.

- **Trend** - There is no sign of the trend.

- **Seasonality** – The image is depicting that there is a sign of seasonality.



Use of Market Basket Analysis

- Association rules are a rule-based machine learning method for discovering interesting relations between variables in large databases. It is intended to identify strong rules discovered in databases using some measures of interestingness. In any given transaction with a variety of items, association rules are meant to discover the rules that determine how or why certain items are connected.

- The 41 items that were sold multiple times can be helpful in association rules by identifying products that are frequently purchased together. This information can be used to improve product placement in stores and online. For example, if the items "milk" and "bread" are frequently purchased together, then they can be placed near each other in the store. This will make it more likely that customers who buy milk will also buy bread.

- The association rules can also be used to identify customers who are likely to be interested in new products or services. For example, if a customer has previously purchased milk, bread, and eggs, then they are likely to also be interested in buying coffee. This information can be used to target marketing campaigns more effectively.

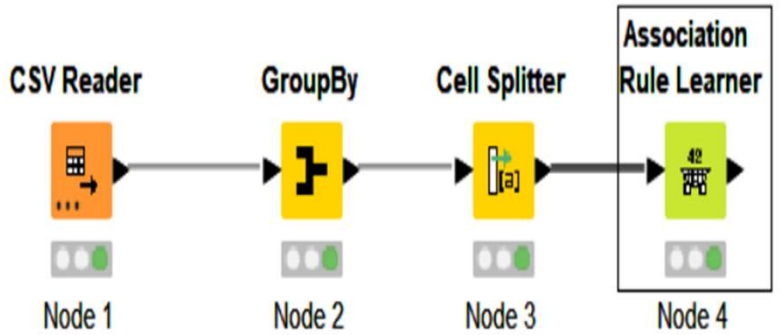
Use of Market Basket Analysis

- In the image you sent, the text says "Inde n dychw" which is Welsh for "I need you". This could be a marketing campaign for a new product or service that is targeted at people who are likely to be interested in it. The association rules can be used to identify these people by finding patterns in their past purchases.
- Here are some other ways that the association rules can be used:

- To detect fraudulent transactions.
- To improve customer service.
- To diagnose diseases.

Use of Market Basket Analysis

Node	output
Excel Reader	File uploaded to the Knime
Group by	new File crerated with new columns
Cell Spiltter	Removing the duplicate
Association rule learner	to execute association rule and defining the catogries.



Use of Market Basket Analysis

- The threshold values of support and confidence are used to determine which association rules are considered to be significant. A rule is considered to be significant if its support and confidence are both above the threshold values.
- The threshold value of support is 0.0025, which means that an association rule must occur in at least 0.25% of the transactions to be considered significant. The threshold value of confidence is 0.806, which means that an association rule must be true in at least 80.6% of the transactions where the antecedent is true to be considered significant.
- An association rule would only be considered significant if it occurred in at least 0.25% of the transactions and was true in at least 80.6% of the transactions where the antecedent was true.
- The threshold values of support and confidence can be adjusted to control the number of association rules that are generated. A lower threshold value will result in more association rules being generated, while a higher threshold value will result in fewer association rules being generated.

Row ID	D Support	D Confide...	D Lift	S Conseq...	S implies	[...] Items
rule0	0.025	0.806	2.021	ice cream	<---	[paper towels,yogurt,pasta,...]
rule1	0.025	0.806	1.912	poultry	<---	[dinner rolls,spaghetti sauce,ice cream,...]
rule2	0.025	0.829	2.13	dinner rolls	<---	[spaghetti sauce,poultry,waffles,...]
rule3	0.025	0.829	1.966	poultry	<---	[shampoo,hand soap,juice,...]
rule4	0.026	0.857	2.194	cheeses	<---	[paper towels,cereals,sandwich bags,...]
rule5	0.028	0.842	1.998	poultry	<---	[dinner rolls,spaghetti sauce,beef,...]
rule6	0.028	0.821	1.947	poultry	<---	[dinner rolls,spaghetti sauce,sandwich loaves,...]
rule7	0.029	0.846	2.008	poultry	<---	[dinner rolls,spaghetti sauce,hand soap,...]

Associations Identified

- Support: It is the percentage of transactions that contain all the items in the rule. In this case, the support for the rule "paper towels, yogurt, pasta" is 0.0025, which means that it occurs in 0.25% of the transactions.
- Confidence: It is the percentage of transactions that contain the antecedent (paper towels, yogurt) that also contain the consequent (pasta). In this case, the confidence for the rule is 0.806, which means that 80.6% of the transactions that contain paper towels and yogurt also contain pasta.
- Lift: It is a measure of how much more likely it is for a customer to buy pasta if they have already bought paper towels and yogurt. In this case, the lift is 2.021, which means that customers are 2.021 times more likely to buy pasta if they have already bought paper towels and yogurt.

Row ID	D Support	D Confide...	D Lift	S Conseq...	S implies	[...] Items
rule0	0.025	0.806	2.021	ice cream	<---	[paper towels,yogurt,pasta,...]
rule1	0.025	0.806	1.912	poultry	<---	[dinner rolls,spaghetti sauce,ice cream,...]
rule2	0.025	0.829	2.13	dinner rolls	<---	[spaghetti sauce,poultry,waffles,...]
rule3	0.025	0.829	1.966	poultry	<---	[shampoo,hand soap,juice,...]
rule4	0.026	0.857	2.194	cheeses	<---	[paper towels,cereals,sandwich bags,...]
rule5	0.028	0.842	1.998	poultry	<---	[dinner rolls,spaghetti sauce,beef,...]
rule6	0.028	0.821	1.947	poultry	<---	[dinner rolls,spaghetti sauce,sandwich loaves,...]
rule7	0.029	0.846	2.008	poultry	<---	[dinner rolls,spaghetti sauce,hand soap,...]

Associations Identified

- Support: It is the percentage of transactions that contain all the items in the rule. In this case, the support for the rule "dinner rolls, spaghetti, sauce, ice cream" is 0.0025, which means that it occurs in 0.25% of the transactions.
- Confidence: It is the percentage of transactions that contain the antecedent (dinner rolls, spaghetti, sauce) that also contain the consequent (ice cream). In this case, the confidence for the rule is 0.829, which means that 82.9% of the transactions that contain dinner rolls, spaghetti, and sauce also contain ice cream.
- Lift: It is a measure of how much more likely it is for a customer to buy ice cream if they have already bought dinner rolls, spaghetti, and sauce. In this case, the lift is 2.13, which means that customers are 2.13 times more likely to buy ice cream if they have already bought dinner rolls, spaghetti, and sauce.

Suggestion of Possible Combos with Lucrative Offers

Support	Confidence	lift	Consequent	implies	Items
0.025460931	0.805555556	2.020986295	ice cream	<---	[paper towels, yogurt, pasta, lunch meat]
0.025460931	0.805555556	1.911516204	poultry	<---	[dinner rolls, spaghetti sauce, ice cream, beef]
0.025460931	0.828571429	2.13034505	dinner rolls	<---	[spaghetti sauce, poultry, waffles, laundry detergent]
0.025460931	0.828571429	1.966130952	poultry	<---	[shampoo, hand soap, juice, sugar]
0.026338894	0.857142857	2.193900482	cheeses	<---	[paper towels, cereals, sandwich bags, sugar]
0.02809482	0.842105263	1.998245614	poultry	<---	[dinner rolls, spaghetti sauce, beef, sugar]
0.02809482	0.820512821	1.947008547	poultry	<---	[dinner rolls, spaghetti sauce, sandwich loaves, soap]
0.028972783	0.846153846	2.007852564	poultry	<---	[dinner rolls, spaghetti sauce, hand soap, sugar]

Suggestion of Possible Combos with Lucrative Offers

- Spaghetti sauce, poultry, waffles, laundry detergent: This rule has a high support and confidence, which means that it is frequently purchased together. You could offer a combo deal where customers get a free bottle of laundry detergent when they buy spaghetti sauce, poultry, and waffles.

- Shampoo, hand soap, juice, sugar: This rule also has a high support and confidence. You could offer a buy-two-get-one-free deal on shampoo, hand soap, juice, and sugar.

Suggestion of Possible Combos with Lucrative Offers

- Paper towels, cereals, sandwich bags, sugar: This rule has a lower support than the previous two rules, but it still has a high confidence. You could offer a discount on paper towels, cereals, sandwich bags, and sugar when customers buy a certain amount of these items.
- Poultry: This rule has the lowest support of the four rules, but it still has a high confidence. You could offer a loyalty program where customers earn points for every purchase of poultry. These points could then be redeemed for discounts on other products.

These are just a few ideas for combos with lucrative offers. You can also get creative and come up with your own ideas. The most important thing is to offer deals that are relevant to your customers and that will encourage them to buy more products.