Marketing And Retail Analysis Project

By

Arjun anil kumar

Part A:

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.

	0	1	2	3	4
ORDERNUMBER	10424	10425	10424	10425	10424
QUANTITYORDERED	50	38	49	33	54
PRICEEACH	240.02	155.13	162.64	142.2	133.0
ORDERLINENUMBER	6	12	3	4	5
SALES	12001.0	5894.94	7969.36	4692.6	7182.0
ORDERDATE	2020-05-31 00:00:00	2020-05-31 00:00:00	2020-05-31 00:00:00	2020-05-31 00:00:00	2020-05-31 00:00:00
DAYS_SINCE_LASTORDER	50	150	226	376	427
STATUS	In Process	In Process	In Process	In Process	In Process
PRODUCTLINE	Classic Cars	Classic Cars	Trucks and Buses	Trucks and Buses	Trucks and Buses
MSRP	214	147	136	118	116
PRODUCTCODE	S10_1949	S10_4962	S12_1666	S12_4473	S18_1097
CUSTOMERNAME	Euro Shopping Channel	La Rochelle Gifts	Euro Shopping Channel	La Rochelle Gifts	Euro Shopping Channel
PHONE	(91) 555 94 44	40.67.8555	(91) 555 94 44	40.67.8555	(91) 555 94 44
ADDRESSLINE1	C/ Moralzarzal, 86	67, rue des Cinquante Otages	C/ Moralzarzal, 86	67, rue des Cinquante Otages	C/ Moralzarzal, 86
CITY	Madrid	Nantes	Madrid	Nantes	Madrid
POSTALCODE	28034	44000	28034	44000	28034
COUNTRY	Spain	France	Spain	France	Spain
CONTACTLASTNAME	Freyre	Labrune	Freyre	Labrune	Freyre
CONTACTFIRSTNAME	Diego	Janine	Diego	Janine	Diego
DEALSIZE	Large	Medium	Large	Medium	Large

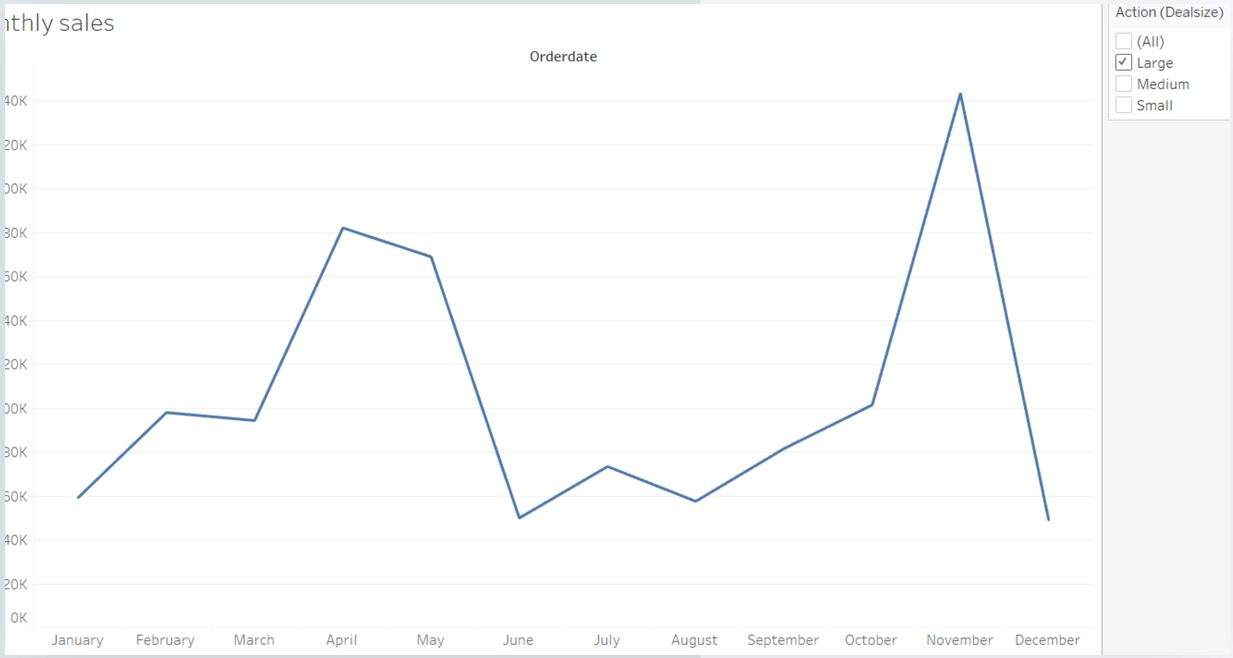
There are no null values in the dataset.

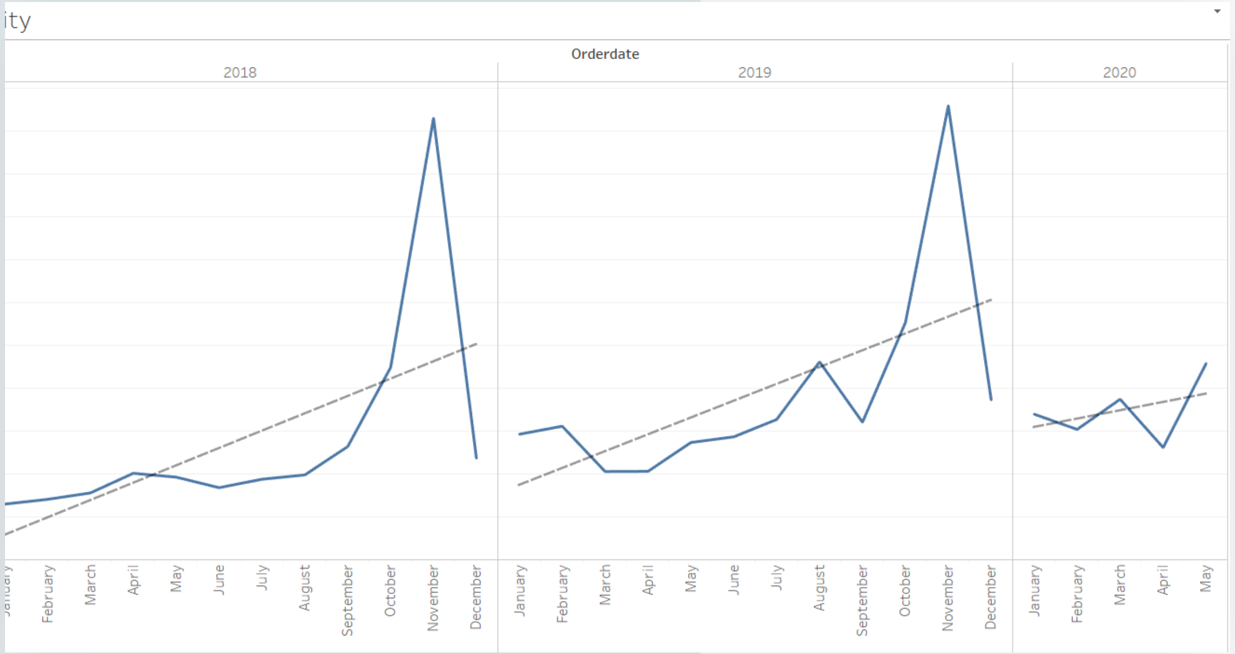
Order date is the only datetime variable

There are 2747 rows of data.

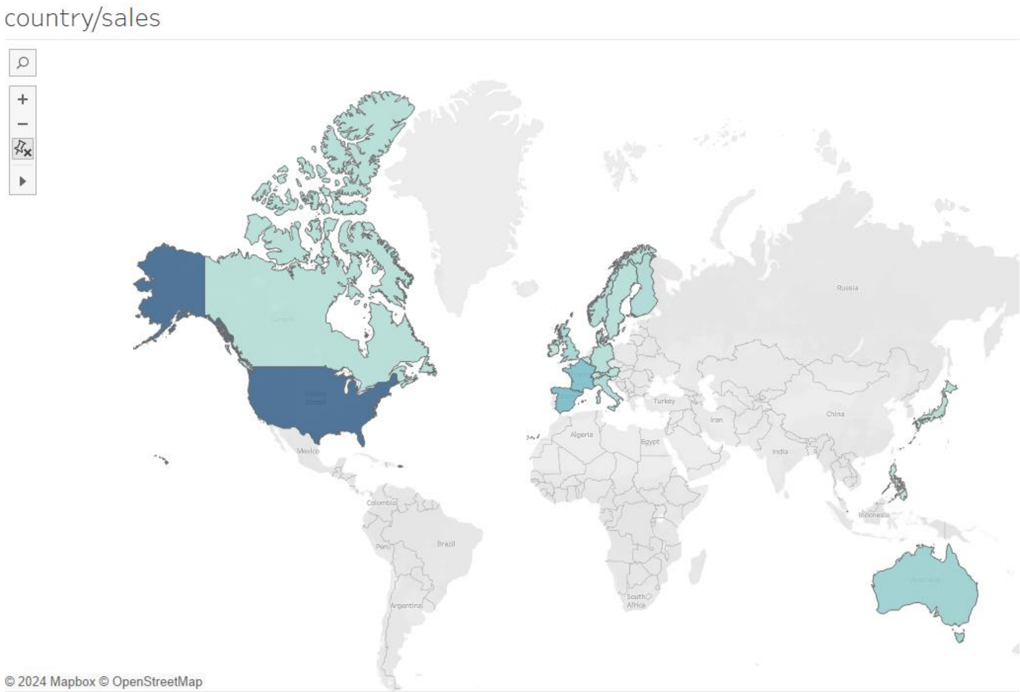
```
RangeIndex: 2747 entries, 0 to 2746
Data columns (total 20 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   ORDERNUMBER           2747 non-null   int64
 1   QUANTITYORDERED       2747 non-null   int64
 2   PRICEEACH             2747 non-null   float64
 3   ORDERLINENUMBER       2747 non-null   int64
 4   SALES                 2747 non-null   float64
 5   ORDERDATE             2747 non-null   datetime64[ns]
 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: datetime64[ns](1), float64(2), int64(5), object(12)
```

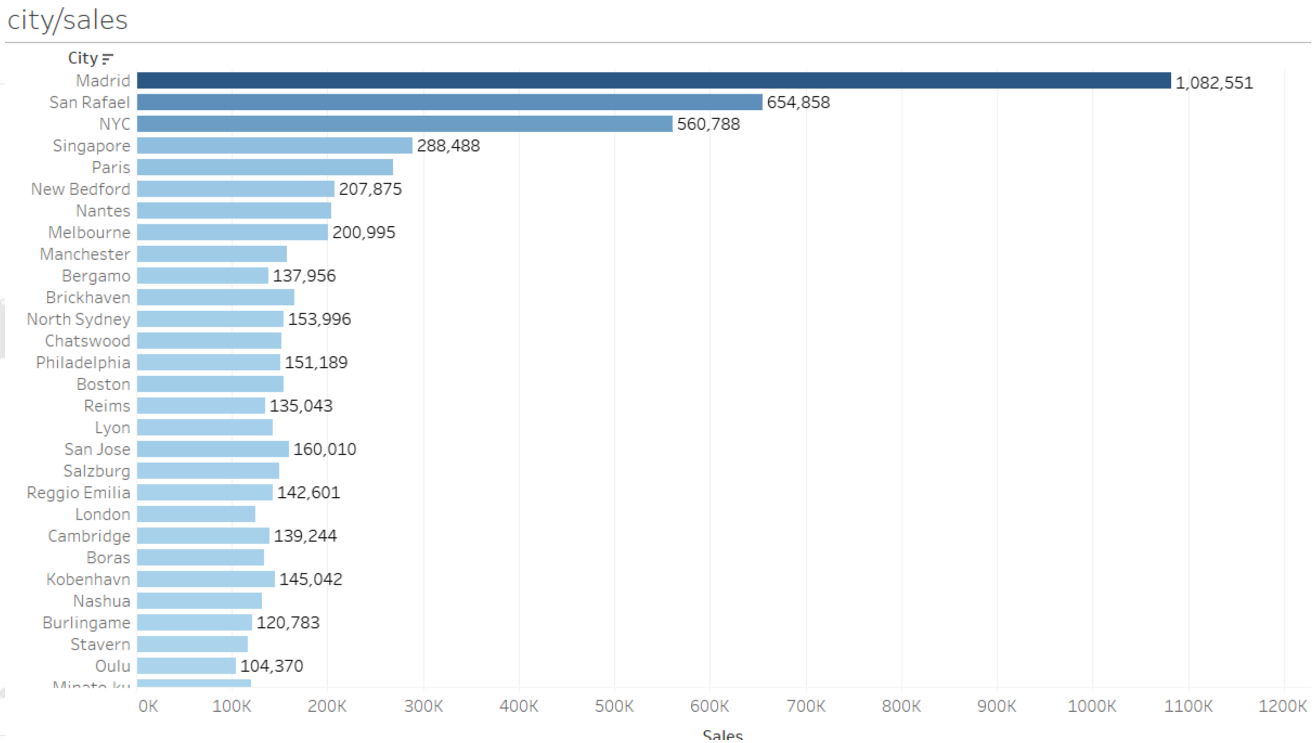
	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	DAYS_SINCE_LASTORDER	MSRP
count	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000
mean	10259.761558	35.103021	101.098951	6.491081	3553.047583	1757.085912	100.691664
std	91.877521	9.762135	42.042548	4.230544	1838.953901	819.280576	40.114802
min	10100.000000	6.000000	26.880000	1.000000	482.130000	42.000000	33.000000
25%	10181.000000	27.000000	68.745000	3.000000	2204.350000	1077.000000	68.000000
50%	10264.000000	35.000000	95.550000	6.000000	3184.800000	1761.000000	99.000000
75%	10334.500000	43.000000	127.100000	9.000000	4503.095000	2436.500000	124.000000
max	10425.000000	97.000000	252.870000	18.000000	14082.800000	3562.000000	214.000000



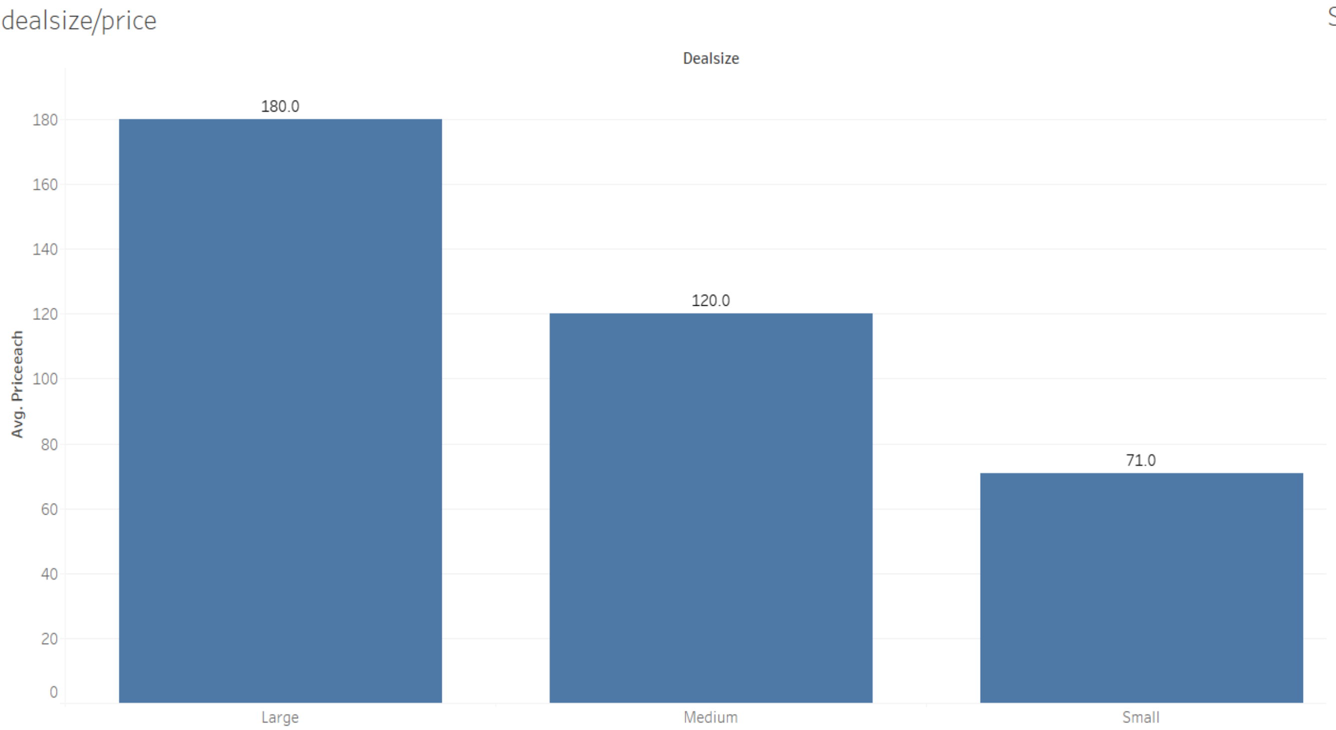


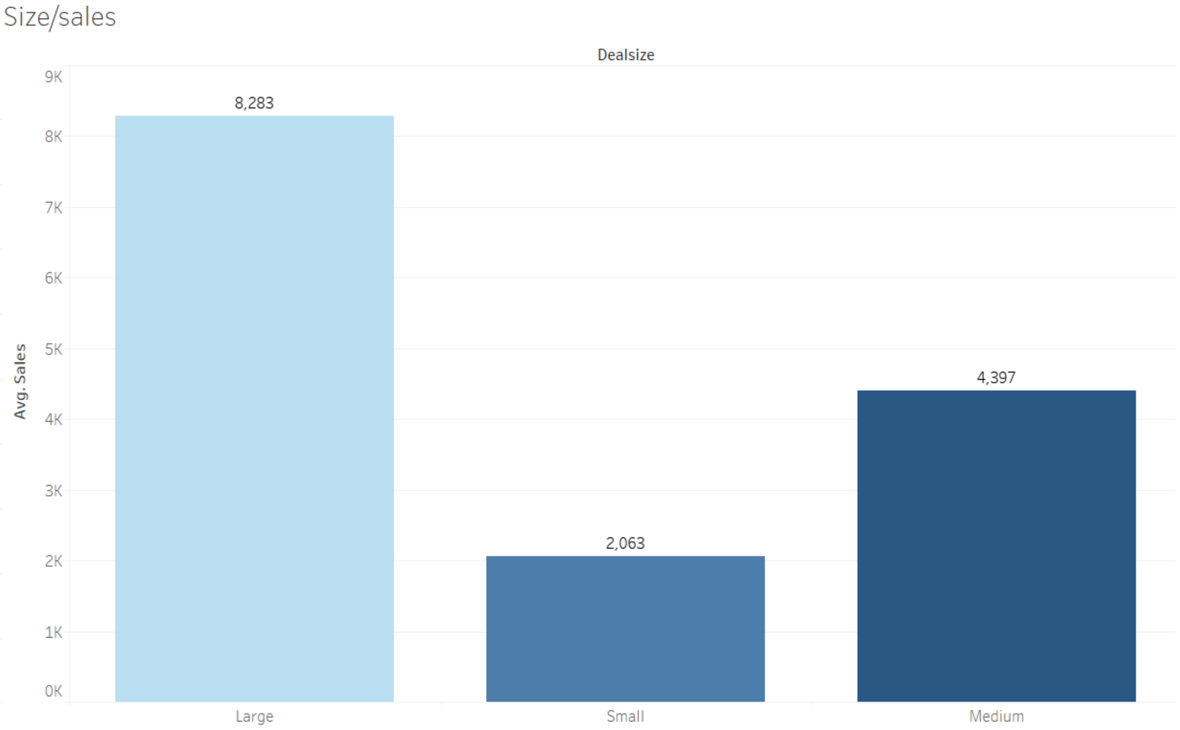
- From September to November the sales get very high and as the year ends the sales decrease.
- Overall trend of the sales is going upwards.
- For large deals March to May is showing an increase in sales.



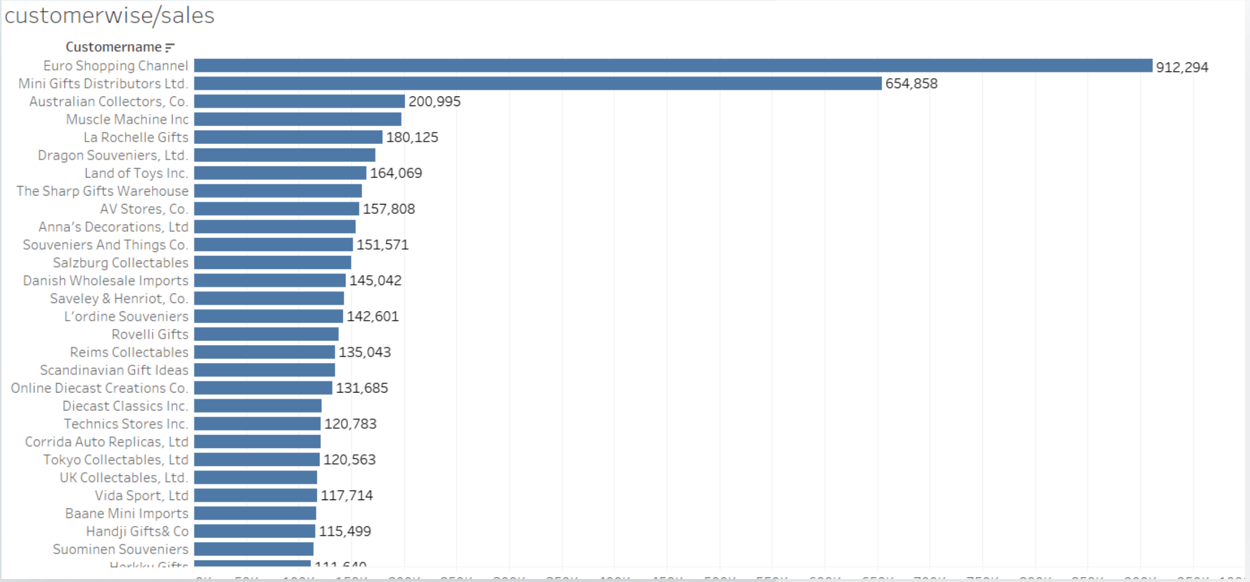


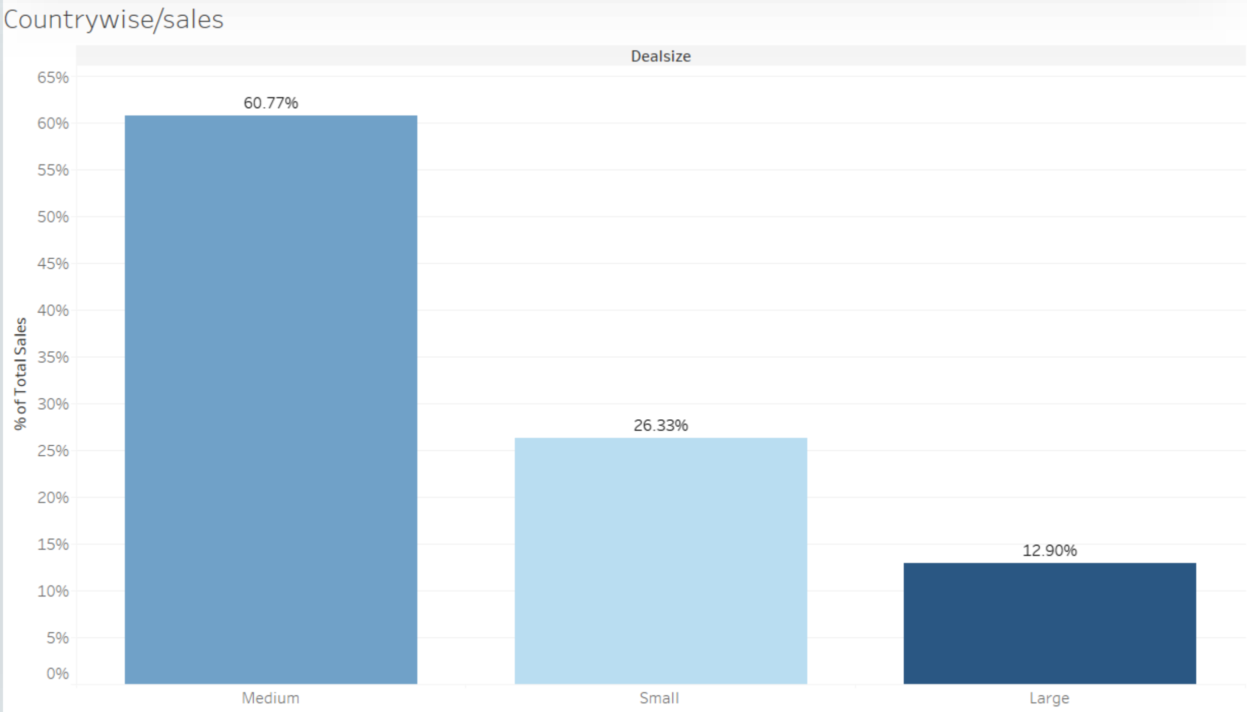
- USA is the largest consumer.
- As sales increase the quantity of orders also increases.
- Madrid records the highest sales among cities.
- Australia, Canada, Finland, France, Germany, Italy, Norway, the UK, and the USA have wider outreach and high sales among the countries.





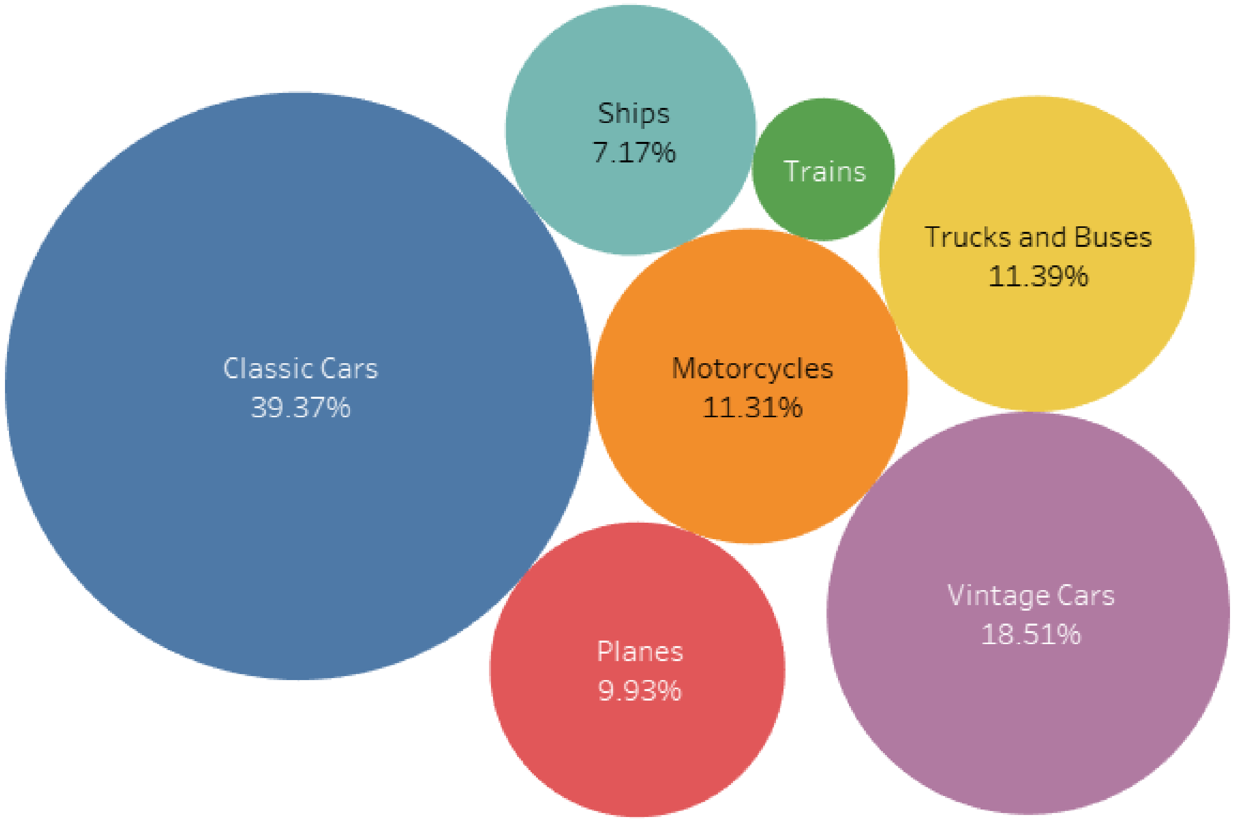
- For large deals the average price of the products is also high and for small it is less.
- The average sales of large deals is 8200, medium 4400 and for small it is 2000.
- Medium deals have the highest amount of quantity ordered about 52000, small it is 38000 and for large it is 7200.





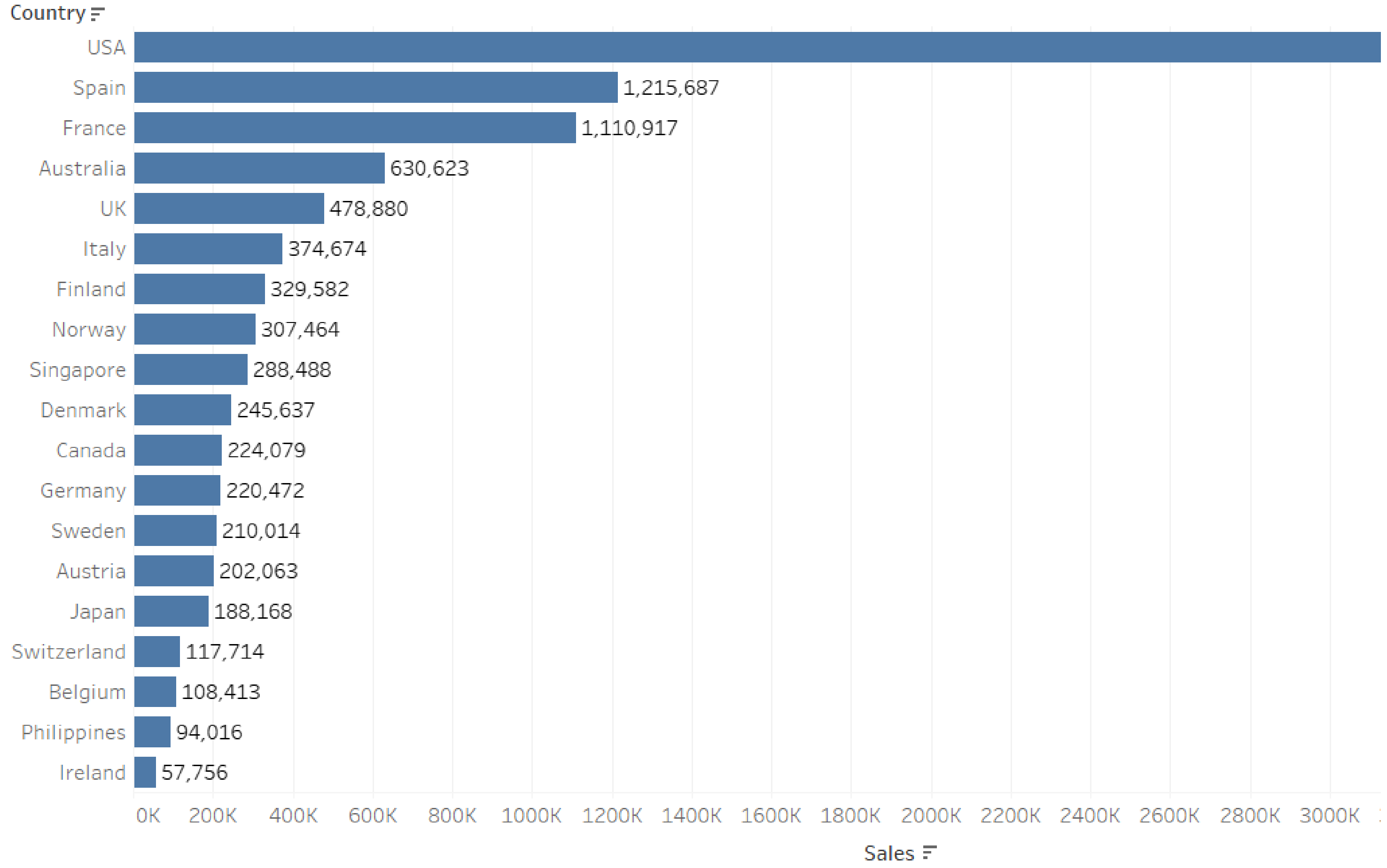
- Medium deals generate 60% of sales while small produce 27% and large produce 13%.
- Euro shopping channel, Mini gifts distributors, Australian collectors, Muscle machine, and La Rochelle gifts are the top 5 consumers in terms of sales generated.
- Royal Belge, Microscale INC, Auto-moto Classics INC, Atelier Graphique, Boards, and Toys CO are the bottom 5 customers in terms of sales generated.

product/sales



- Classic Cars generate the maximum amount of sales about 40% and Trains contribute the lowest about 2.3%.

country/product



Top Countries having the highest sales among different products:

Classic cars: USA, Spain, France, Australia, and UK

Motorcycles: USA, France, Australia and Spain

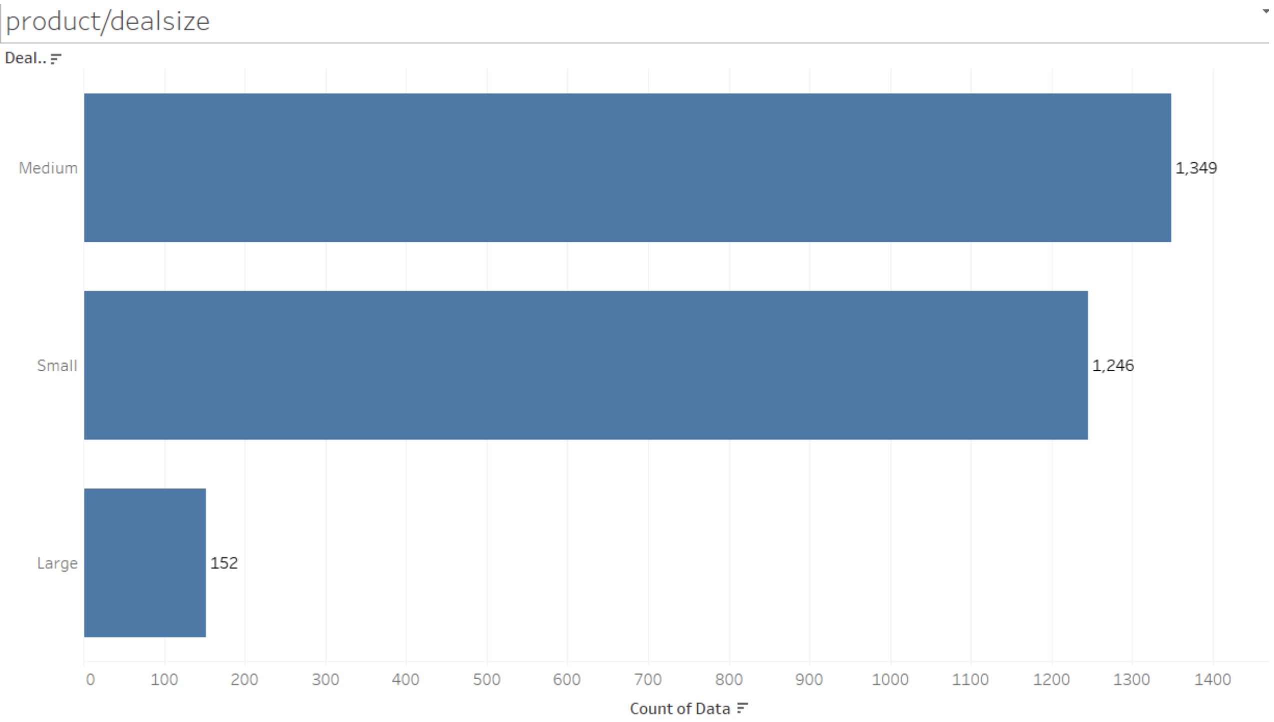
Planes: USA, France, Australia, Spain, and Italy

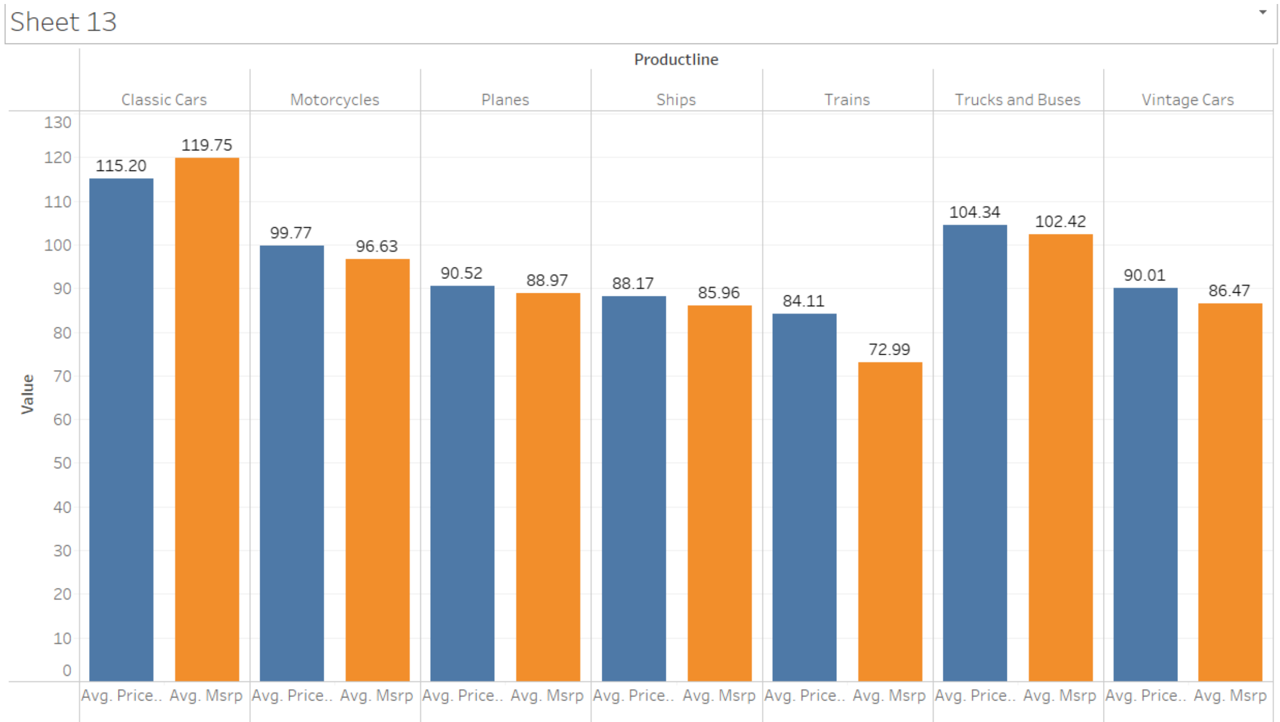
Ships: USA, Spain, UK and France

Trains: USA, Spain and France

Trucks and Buses: USA, Spain, France, Singapore and Australia

Vintage cars: USA, Spain, UK, France and Italy





- •For motorcycles, planes, ships, Trains, and vintage cars Small deals are higher in amount.
- •Except for classic cars the average price of each product line is higher than the average price of its MSRP.
- •Classic cars, Trucks & Buses have the highest average price for their parts.

Inference

Sales Trends and Patterns:

1. **Seasonal Sales Trends**:
    1. High sales from September to November.
    2. Sales decline at the end of the year.
    3. Overall sales trend is upward.
2. **Large Deal Sizes**:
    1. Increased sales from March to May.
    2. Large deals have higher average prices.
3. **Geographic Insights**:
    1. USA is the largest consumer.
    2. Madrid records the highest sales among cities.
    3. High sales and outreach in the USA, Australia, Canada, Finland, France, Germany, Italy, Norway, and the UK.
4. **Order Quantity**:
    1. Sales increase correlates with increased order quantities.
    2. Medium deals have the highest quantity ordered (52,000), followed by small (38,000) and large (7,200).

Deal Sizes:

1.Sales Contribution:

1. Medium deals generate 60% of sales.
2. Small deals produce 27%.
3. Large deals produce 13%.

2.Average Sales by Deal Size:

1. Large deals: $8,200
2. Medium deals: $4,400
3. Small deals: $2,000

Top and Bottom Consumers:

1.Top Consumers:

1. Euro Shopping Channel
2. Mini Gifts Distributors
3. Australian Collectors
4. Muscle Machine
5. La Rochelle Gifts

2.Bottom Consumers:

1. Royal Belge
2. Microscale INC
3. Auto-Moto Classics INC
4. Atelier Graphique
5. Boards and Toys CO

Product Line Insights:

1. **Sales Distribution by Product Line**:

    1. Classic Cars generate 40% of sales.
    2. Trains contribute the lowest at 2.3%.

1. **Top Countries by Product Line**:

    1. **Classic Cars**: USA, Spain, France, Australia, UK
    2. **Motorcycles**: USA, France, Australia, Spain
    3. **Planes**: USA, France, Australia, Spain, Italy
    4. **Ships**: USA, Spain, UK, France
    5. **Trains**: USA, Spain, France
    6. **Trucks and Buses**: USA, Spain, France, Singapore, Australia
    7. **Vintage Cars**: USA, Spain, UK, France, Italy

1. **Deal Sizes by Product Line**:

    1. Small deals are higher in amount for motorcycles, planes, ships, trains, and vintage cars.
    2. Except for classic cars, the average price of each product line is higher than the MSRP.
    3. Classic cars and trucks & buses have the highest average prices for parts.

RFM Analysis

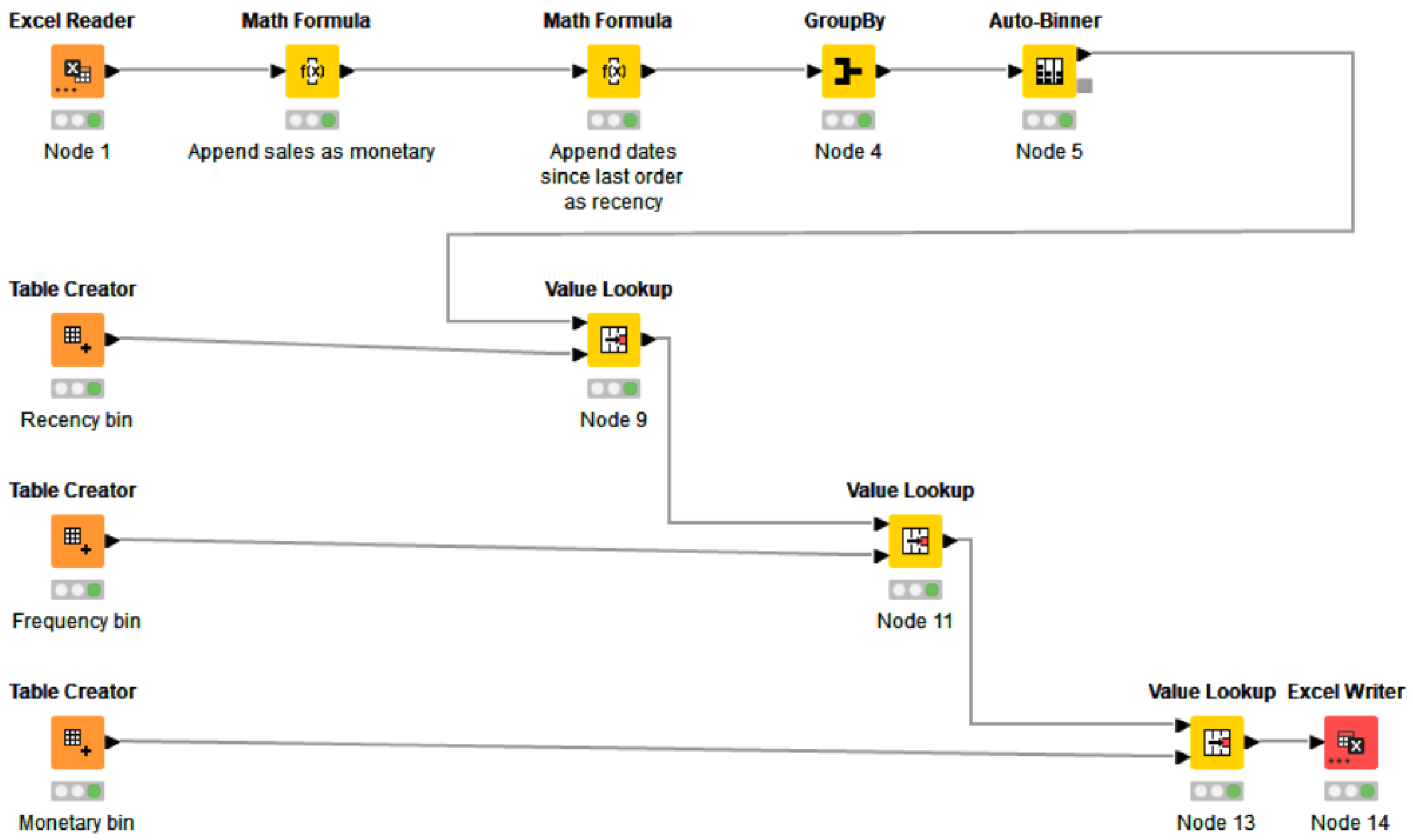
RFM (Recency, Frequency, Monetary) analysis is a customer segmentation method that evaluates how recently, how often, and how much customers purchase to identify and target key customer groups effectively.

Parameters and Assumptions made:

- Recency: The minimum of dates since the last order.
- Frequency: The total count of number of orders of each customer.
- Monetary: The total amount of sales generated by each customer.

- Based on the customer name the groups were created, which reduced the total number of datasets to 89 rows.
- For Recency, frequency, and monetary columns 4 bins were created which range as:
  [0-0.25, 0.25-0.5, 0.5-0.75, 0.75-1].
- For each value corresponding to different bins each row was under Recency, Frequency, and monetary columns were classified as Very low, Low, High, and Very high.

KNIME Workflow



Final Output sample

Table "default" - Rows: 89 | Spec - Columns: 16 | Properties | Flow Variables

Row ID	S CUSTO...	I ORDER...	I QUANT...	D PRICEE...	D SALES	S CITY	S COUNTRY	D RECENCY	D MONET...	I PRODU...	S ORDERN...	S RECENCY [...	S MONETA...	S Recency_RFL	S Frequency_RFL	S Monetary_RFL
Row0	AV Stores, Co.	3	1778	91.085	157,807.81	Manchester	UK	421	157,807.81	37	Bin 2	Bin 3	Bin 4	Low	Low	Very high
Row1	Alpha Cognac	3	687	101.16	70,488.44	Toulouse	France	675	70,488.44	20	Bin 2	Bin 4	Bin 1	Very low	Low	Very low
Row2	Amica Model...	2	843	110.853	94,117.26	Torino	Italy	328	94,117.26	26	Bin 1	Bin 2	Bin 3	High	Very low	High
Row3	Anna's Decor...	4	1469	106.424	153,996.13	North Sydney	Australia	131	153,996.13	41	Bin 4	Bin 1	Bin 4	Very high	Very high	Very high
Row4	Atelier graph...	3	270	92.239	24,179.96	Nantes	France	312	24,179.96	7	Bin 2	Bin 2	Bin 1	High	Low	Very low
Row5	Australian C...	3	705	90.042	64,591.46	Glen Waverly	Australia	1,018	64,591.46	23	Bin 2	Bin 4	Bin 1	Very low	Low	Very low
Row6	Australian C...	5	1926	104.59	200,995.41	Melbourne	Australia	229	200,995.41	40	Bin 4	Bin 1	Bin 4	Very high	Very high	Very high
Row7	Australian Gi...	3	545	110.554	59,469.12	South Brisbane	Australia	190	59,469.12	15	Bin 2	Bin 1	Bin 1	Very high	Low	Very low
Row8	Auto Assoc. ...	2	637	99.488	64,834.32	Versailles	France	275	64,834.32	17	Bin 1	Bin 2	Bin 1	High	Very low	Very low
Row9	Auto Canal P...	3	1001	94.255	93,170.66	Paris	France	127	93,170.66	18	Bin 2	Bin 1	Bin 3	Very high	Low	High
Row10	Auto-Moto Cl...	3	287	92.8	26,479.26	Brickhaven	USA	1,353	26,479.26	8	Bin 2	Bin 4	Bin 1	Very low	Low	Very low
Row11	Baane Mini I...	4	1082	108.574	116,599.19	Stavern	Norway	245	116,599.19	32	Bin 4	Bin 1	Bin 3	Very high	Very high	High
Row12	Bavarian Coll...	1	401	84.289	34,993.92	Munich	Germany	801	34,993.92	14	Bin 1	Bin 4	Bin 1	Very low	Very low	Very low
Row13	Blauer See A...	4	811	108.031	85,171.59	Frankfurt	Germany	705	85,171.59	22	Bin 4	Bin 4	Bin 2	Very low	Very high	Low
Row14	Boards & To...	2	102	89.807	9,129.35	Glendale	USA	410	9,129.35	3	Bin 1	Bin 2	Bin 1	High	Very low	Very low
Row15	CAF Imports	2	468	104.963	49,642.05	Madrid	Spain	625	49,642.05	13	Bin 1	Bin 3	Bin 1	Low	Very low	Very low
Row16	Cambridge C...	2	357	101.329	36,163.62	Cambridge	USA	484	36,163.62	11	Bin 1	Bin 3	Bin 1	Low	Very low	Very low
Row17	Canadian Gif...	2	703	105.341	75,238.92	Vancouver	Canada	364	75,238.92	19	Bin 1	Bin 2	Bin 2	High	Very low	Low
Row18	Classic Gift I...	2	668	103.32	67,506.97	Philadelphia	USA	344	67,506.97	21	Bin 1	Bin 2	Bin 1	High	Very low	Very low
Row19	Classic Lege...	3	720	109.804	77,795.2	NYC	USA	309	77,795.2	20	Bin 2	Bin 2	Bin 2	High	Low	Low
Row20	Clover Collec...	2	490	112.87	57,756.43	Dublin	Ireland	659	57,756.43	16	Bin 1	Bin 4	Bin 1	Very low	Very low	Very low
Row21	Collectable M...	2	954	91.535	87,489.23	San Diego	USA	575	87,489.23	25	Bin 1	Bin 3	Bin 2	Low	Very low	Low
Row22	Collectables ...	3	795	97.237	81,577.98	Brickhaven	USA	179	81,577.98	24	Bin 2	Bin 1	Bin 2	Very high	Low	Low
Row23	Corrida Auto...	3	1163	105.175	120,615.28	Madrid	Spain	407	120,615.28	26	Bin 2	Bin 2	Bin 4	High	Low	Very high
Row24	Cruz & Sons ...	3	961	96.08	94,015.73	Makati City	Philippines	971	94,015.73	22	Bin 2	Bin 4	Bin 3	Very low	Low	High
Row25	Daedalus De...	2	699	95.474	69,052.41	Lille	France	573	69,052.41	14	Bin 1	Bin 3	Bin 1	Low	Very low	Very low
Row26	Danish Whol...	5	1315	108.038	145,041.6	Kobenhavn	Denmark	499	145,041.6	31	Bin 4	Bin 3	Bin 4	Low	Very high	Very high
Row27	Diecast Class...	4	1111	108.566	122,138.14	Allentown	USA	228	122,138.14	31	Bin 4	Bin 1	Bin 4	Very high	Very high	Very high
Row28	Diecast Colle...	2	695	101.783	70,859.78	Boston	USA	672	70,859.78	18	Bin 1	Bin 4	Bin 2	Very low	Very low	Low
Row29	Double Deck...	2	357	99.108	36,019.04	London	UK	670	36,019.04	12	Bin 1	Bin 4	Bin 1	Very low	Very low	Very low
Row30	Dragon Souv...	5	1524	113.106	172,989.68	Singapore	Singapore	649	172,989.68	37	Bin 4	Bin 4	Bin 4	Very low	Very high	Very high
Row31	Enaco Distrib...	3	882	88.783	78,411.86	Barcelona	Spain	659	78,411.86	22	Bin 2	Bin 4	Bin 2	Very low	Low	Low
Row32	Euro Shoppin...	26	9327	97.383	912,294.11	Madrid	Spain	42	912,294.11	106	Bin 4	Bin 1	Bin 4	Very high	Very high	Very high
Row33	FunGiftIdeas...	3	903	109.587	98,923.73	New Bedford	USA	111	98,923.73	26	Bin 2	Bin 1	Bin 3	Very high	Low	High
Row34	Gift Depot Inc.	3	903	108.932	101,894.79	Bridgewater	USA	226	101,894.79	25	Bin 2	Bin 1	Bin 3	Very high	Low	High
Row35	Gift Ideas Co...	3	666	87.6	57,294.42	Glendale	USA	947	57,294.42	19	Bin 2	Bin 4	Bin 1	Very low	Low	Very low
Row36	Gifts4AllAges...	3	933	91.564	83,209.88	Boston	USA	148	83,209.88	21	Bin 2	Bin 1	Bin 2	Very high	Low	Low
Row37	Handji Gifts&...	4	1236	95.593	115,498.73	Singapore	Singapore	488	115,498.73	29	Bin 4	Bin 3	Bin 3	Low	Very high	High

Top 5 Best customers

Customers with very high Recency, Frequency, and Monetary value

CUSTOMER NAME	ORDERNUMBER	CITY	COUNTRY	RECENCY	MONETARY	ORDERNUMBER [Binned]	RECENCY [Binned]	MONETARY [Binned]	Recency_RFL	Frequency_RFL	Monetary_RFL
Anna's Decorations, Ltd	4	North Sydney	Australia	131	153996.13	Bin 4	Bin 1	Bin 4	**Very high**	**Very high**	**Very high**
Australian Collectors, Co.	5	Melbourne	Australia	229	200995.41	Bin 4	Bin 1	Bin 4	**Very high**	**Very high**	**Very high**
Diecast Classics Inc.	4	Allentown	USA	228	122138.14	Bin 4	Bin 1	Bin 4	**Very high**	**Very high**	**Very high**
Euro Shopping Channel	26	Madrid	Spain	42	912294.11	Bin 4	Bin 1	Bin 4	**Very high**	**Very high**	**Very high**
La Rochelle Gifts	4	Nantes	France	139	180124.9	Bin 4	Bin 1	Bin 4	**Very high**	**Very high**	**Very high**

Customers about to churn

- Customers who have Recency in Bin 2 are considered to be about to churn.

CUSTOMER NAME	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	SALES	CITY	COUNTRY	**RECENCY**	MONETARY	PRODUCTCODE	ORDERNUMBER [Binned]	RECENCY [Binned]	MONETARY [Binned]	**Recency_RFL**	Frequency_RFL	Monetary_RFL
Tokyo Collectables, Ltd	4	1150	101.1828	120562.7	Minato-ku	Japan	**259**	120562.7	32	Bin 4	Bin 2	Bin 3	**High**	Very high	High
Toys of Finland, Co.	3	1051	105.7523	111250.4	Helsinki	Finland	**259**	111250.4	30	Bin 2	Bin 2	Bin 3	**High**	Low	High
Scandinavian Gift Ideas	3	1359	97.59737	134259.3	Boras	Sweden	**262**	134259.3	38	Bin 2	Bin 2	Bin 4	**High**	Low	Very high
Auto Assoc. & Cie.	2	637	99.4878	64834.32	Versailles	France	**275**	64834.32	17	Bin 1	Bin 2	Bin 1	**High**	Very low	Very low
Reims Collectables	5	1433	94.34293	135042.9	Reims	France	**287**	135042.9	37	Bin 4	Bin 2	Bin 4	**High**	Very high	Very high

Lost Customers

- Customers who have not made any purchases for quite a long period.

CUSTOMER NAME	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	SALES	CITY	COUNTRY	**RECENCY**	MONETARY	PRODUCTCODE	ORDERNUMBER [Binned]	RECENCY [Binned]	MONETARY [Binned]	**Recency_RFL**	Frequency_RFL	Monetary_RFL
Auto-Moto Classics Inc.	3	287	92.8	26479.26	Brickhaven	USA	**1353**	26479.26	8	Bin 2	Bin 4	Bin 1	**Very low**	Low	Very low
Rovelli Gifts	3	1650	85.67729167	137955.72	Bergamo	Italy	**1032**	137955.7	34	Bin 2	Bin 4	Bin 4	**Very low**	Low	Very high
Australian Collectables, Ltd	3	705	90.04173913	64591.46	Glen Waverly	Australia	**1018**	64591.46	23	Bin 2	Bin 4	Bin 1	**Very low**	Low	Very low
Cruz & Sons Co.	3	961	96.08	94015.73	Makati City	Philippines	**971**	94015.73	22	Bin 2	Bin 4	Bin 3	**Very low**	Low	High
Gift Ideas Corp.	3	666	87.6	57294.42	Glendale	USA	**947**	57294.42	19	Bin 2	Bin 4	Bin 1	**Very low**	Low	Very low

Loyal Customers

- Customers who have very recent purchases and frequency

CUSTOMER NAME	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	SALES	CITY	COUNTRY	RECENCY	MONETARY	PRODUCTCODE	ORDERNUMBER [Binned]	RECENCY [Binned]	MONETARY [Binned]	Recency_RFL	Frequency_RFL	Monetary_RFL
Anna's Decorations, Ltd	4	1469	106.4241	153996.1	North Sydney	Australia	131	153996.1	41	Bin 4	Bin 1	Bin 4	Very high	Very high	Very high
Australian Collectors, Co.	5	1926	104.5902	200995.4	Melbourne	Australia	229	200995.4	40	Bin 4	Bin 1	Bin 4	Very high	Very high	Very high
Baane Mini Imports	4	1082	108.5738	116599.2	Stavern	Norway	245	116599.2	32	Bin 4	Bin 1	Bin 3	Very high	Very high	High
Diecast Classics Inc.	4	1111	108.5658	122138.1	Allentown	USA	228	122138.1	31	Bin 4	Bin 1	Bin 4	Very high	Very high	Very high
Euro Shopping Channel	26	9327	97.3832	912294.1	Madrid	Spain	42	912294.1	106	Bin 4	Bin 1	Bin 4	Very high	Very high	Very high

RFM Inferences

Geographic Insights:

• Top customers and loyal customers are spread across various countries, including Australia, USA, Spain, and France, suggesting a broad geographic distribution of high-value customers.

Engagement Patterns:

• Customers who are about to churn show high monetary values but have recently reduced their purchase frequency, indicating a need for targeted retention strategies.

• Lost customers have not made recent purchases despite some having high historical spending, indicating potential for re-engagement campaigns.

High-Value Segments:

• The RFM analysis highlights the importance of recency and frequency in maintaining customer value, as evidenced by the consistently high scores of loyal and top customers.

• Focus on maintaining frequent and recent interactions with high-value customers can sustain and potentially increase their lifetime value.

Part B:

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.

- **Data:** from 01-01-2018 to 26-02-2020
- **Objective:** The project involves conducting a thorough analysis of Point of Sale (POS) Data for providing recommendations through which a grocery store can increase its revenue by popular combo offers & discounts for customers.
- **Dataset:** 20641 Rows, 3 columns,
- **Missing values:** None
- **Duplicate values:** 4730
- The exploratory analysis and insights provide a clear understanding of the data and highlight the key trends and patterns in sales.
- **Market Basket Analysis** using association rules was performed to identify the relationships between the products purchased by the customers.
- This analysis helped to identify the products that are frequently purchased together, which can be used to create lucrative offers for the customers.

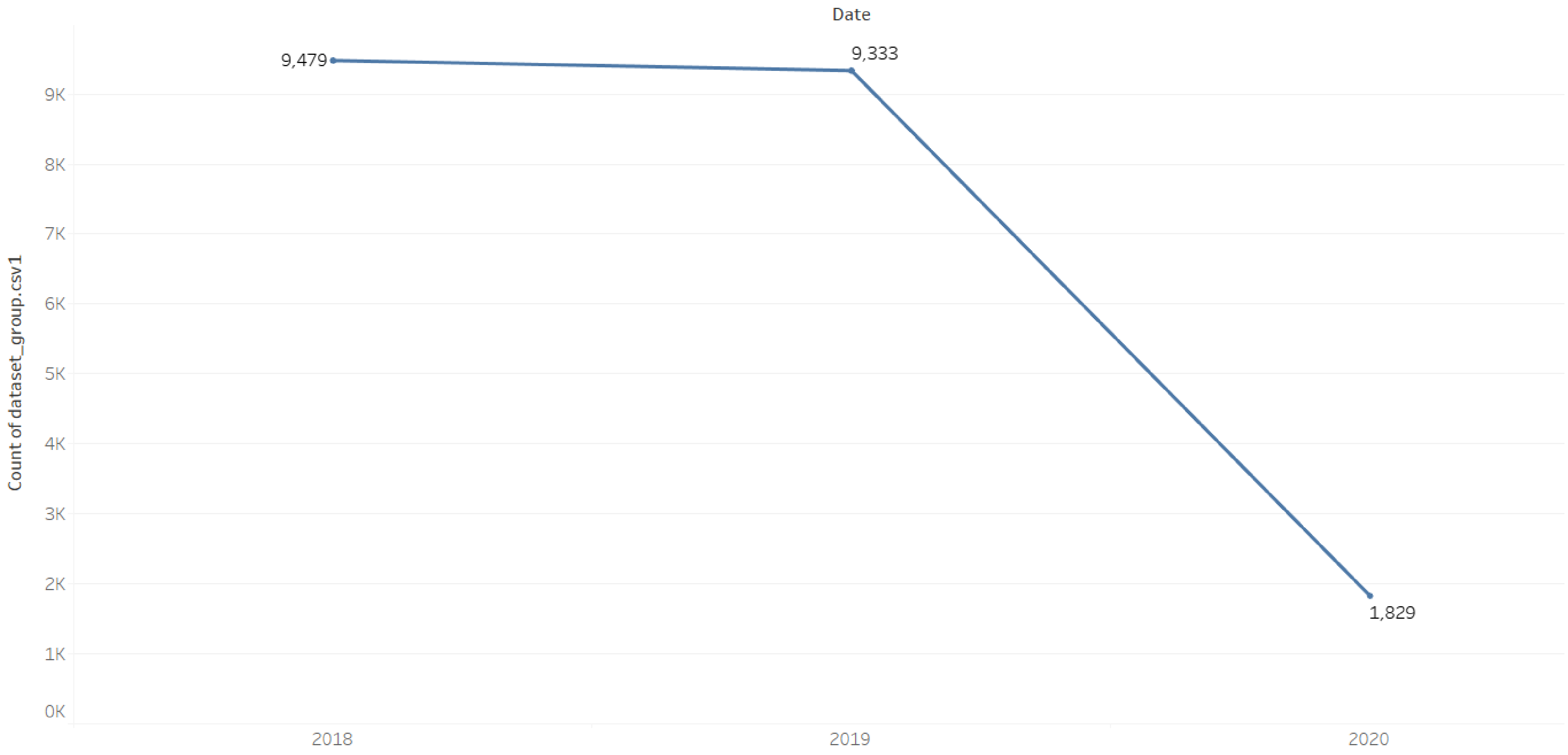
Data Dictionary

Date	date of product sold
Order_id	Id of the order
Product	Name of the Product sold

Assumptions

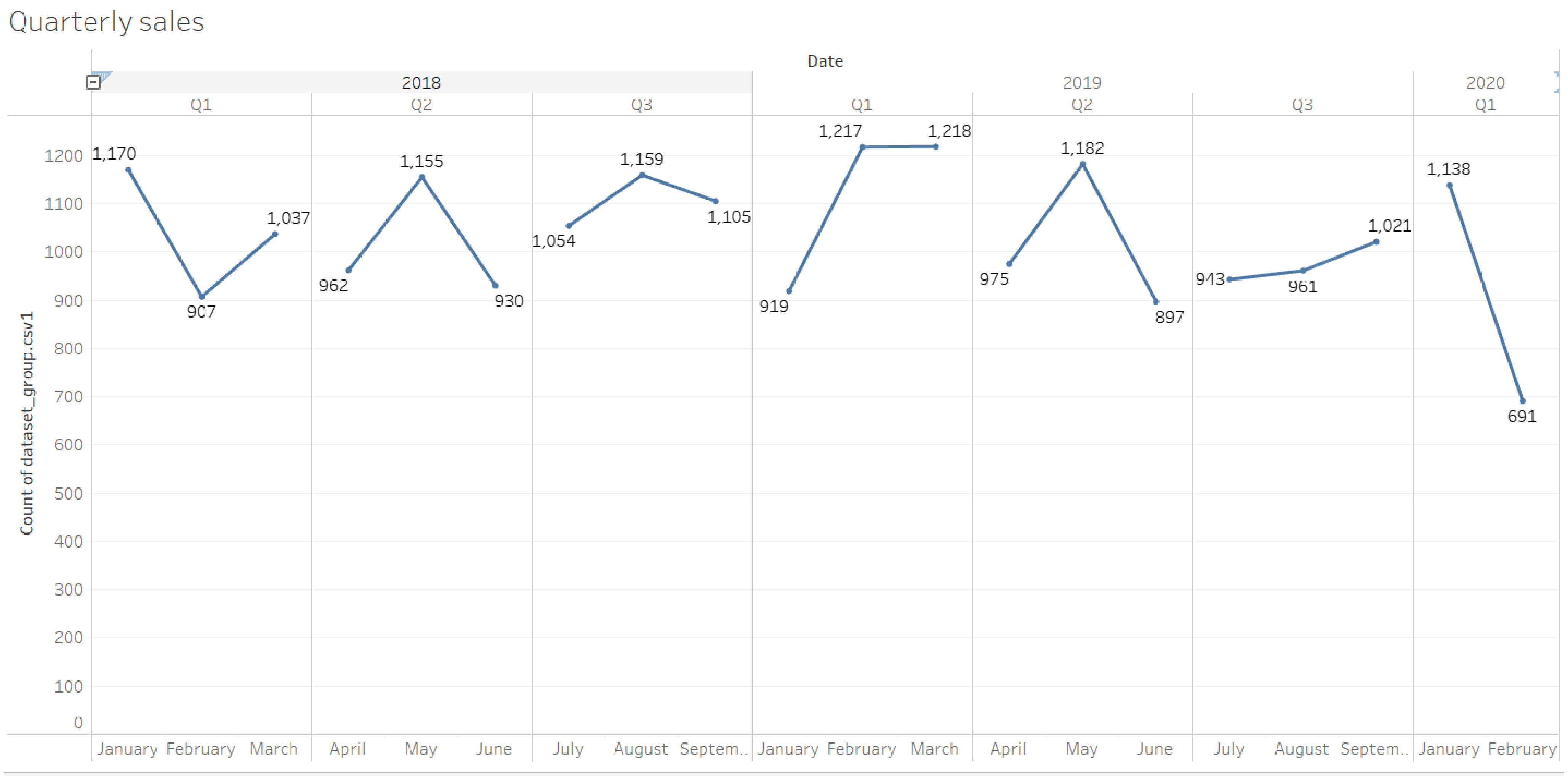
- The data represents a list of items purchased at a grocery store on various dates.
- Each entry in the data represents a single item purchased.
- The first column in the data represents the date the item was purchased.
- The second column represents the customer who made the purchase.
- The third column represents the item purchased.
- The same item can be purchased by multiple customers on different dates.
- There is no information provided about the quantity or price of each item.
- We have not dropped the duplicated values.

Yearly sales



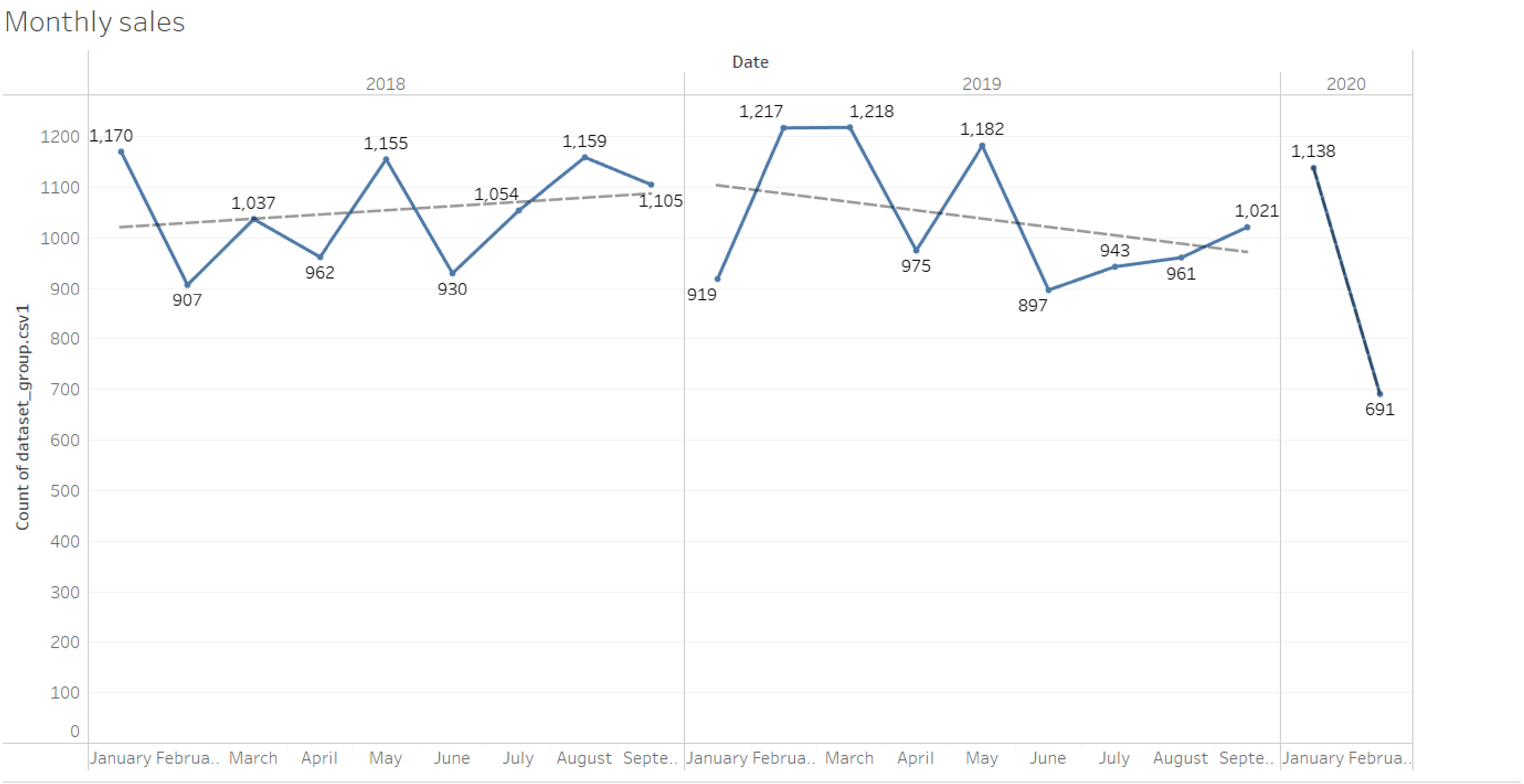
Yearly sales:

- During the year 2018, 9500 products have been purchased.
- During the year 2019, 9400 products have been purchased.



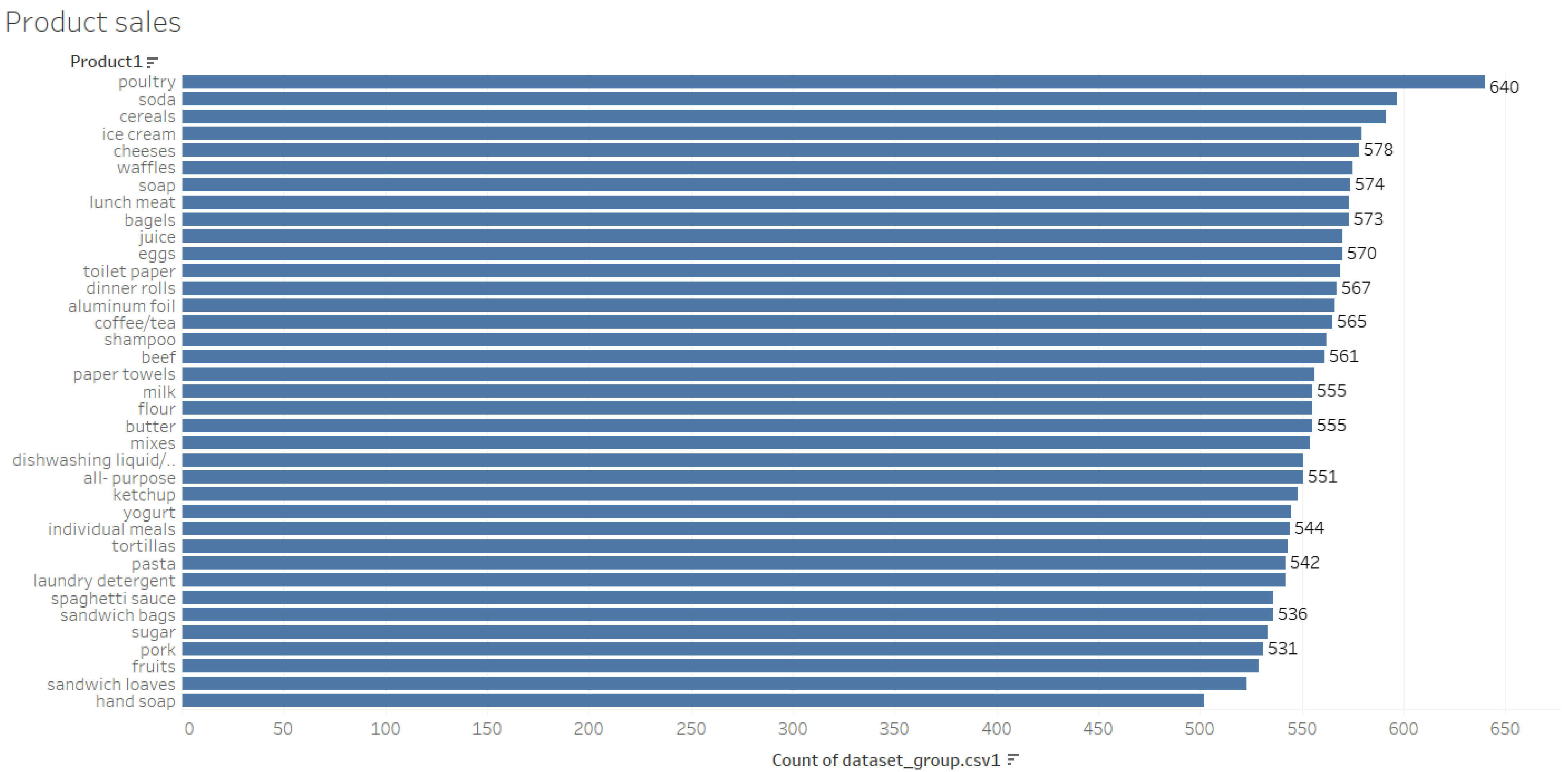
Quarterly sales:

- The sales pattern for 2018 is followed by 2019 during the first 2 quarters of the year 2019
- The first quarter of the year had higher sales compared to the first quarter of 2018
- During the 3rd quarter of 2019, the sales have decreased very drastically.
- The first quarter of the year 2020 also resulted in very low sales.



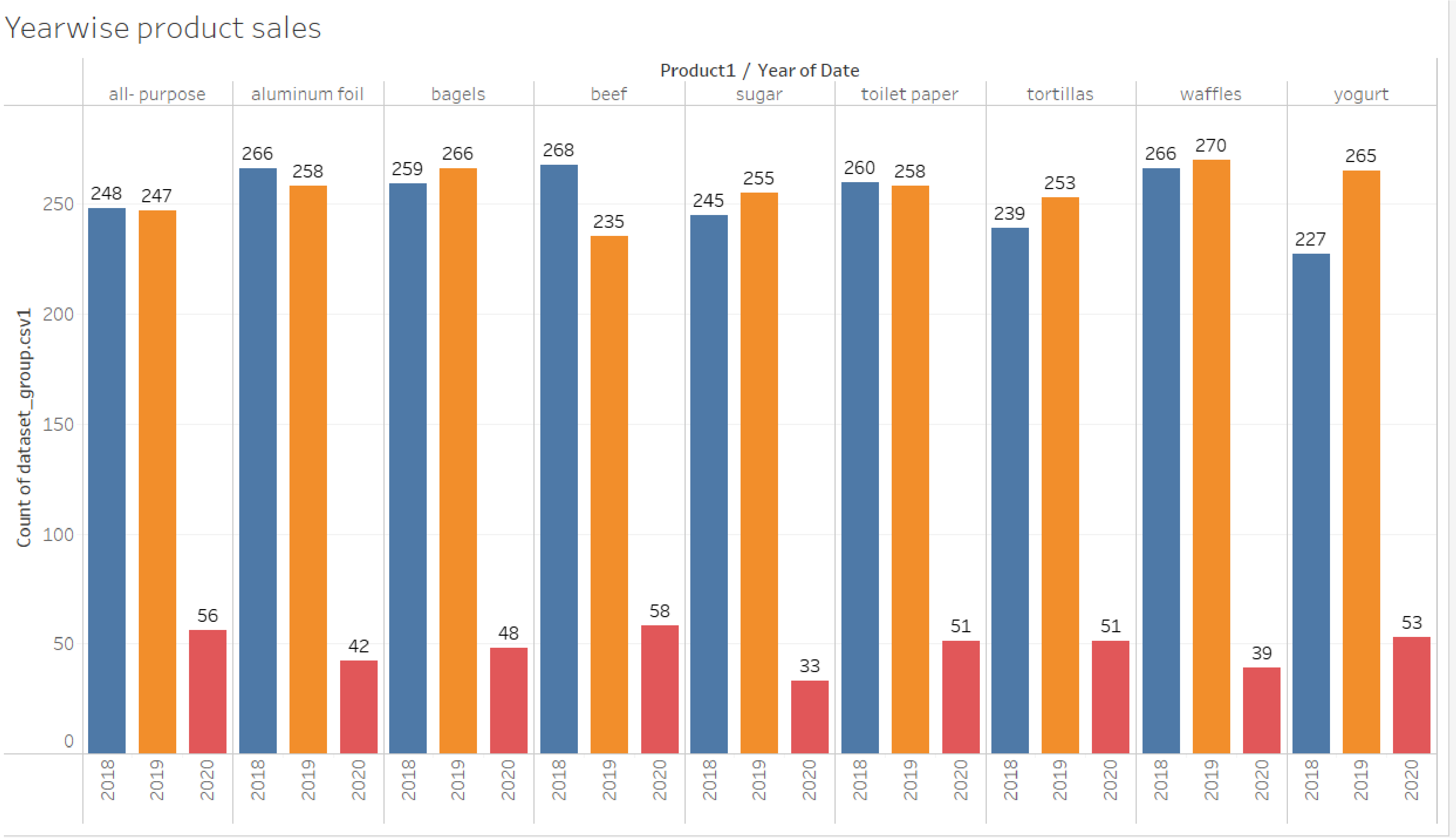
Monthly sales:

- The sales were trending upwards during the year 2018
- The sales went downwards towards the end of 2019.
- The start of the next year was also affected and the sales were very low.



Product sales:

- Poultry, soda, and cereals are the highest-selling commodities with an average of 600 items sold.
- Fruits, sandwich loaves, and hand soaps are the lowest-selling commodities with an average of 520 items sold.



Years of product sales:

- The buying pattern observed is the same for all the products for both years.

Inference

- There was a slight decline in yearly sales from 2018 to 2019, with a decrease from 9500 to 9400 products sold. This indicates a need to investigate factors that may have contributed to this decline and to implement strategies to boost sales.
- The sales pattern for the first two quarters of 2019 followed the same trend as 2018, suggesting that these periods have consistent demand year over year.
- The drastic decrease in sales during the third quarter of 2019 is a red flag. This could be due to external factors such as market conditions, competition, or internal issues like stock shortages or ineffective promotions.
- The low sales in the first quarter of 2020, continuing the trend from late 2019, indicate a prolonged issue that began in the third quarter of 2019. Immediate corrective measures are necessary to prevent further decline.
- Sales were trending upwards throughout 2018, showing positive growth and potentially reflecting successful marketing and sales strategies.
- Poultry, soda, and cereals are the highest-selling commodities, averaging 600 items sold. These products are key drivers of sales and should be prioritized in inventory management, promotions, and marketing strategies to maximize revenue.
- Fruits, sandwich loaves, and hand soaps are the lowest-selling commodities, averaging 520 items sold. These products may need targeted marketing efforts to boost sales or a reevaluation of their product lines to understand why they are underperforming.
- The buying patterns for all products are consistent across 2018 and 2019, indicating stable consumer preferences. This consistency can be leveraged to predict future sales trends and optimize inventory levels.

Market Basket Analysis

Definition:

Market Basket Analysis is a statistical technique that analyzes customer purchase patterns to identify associations between different products. It helps businesses understand which products are frequently purchased together and how customers' buying habits affect sales.

Data:

To conduct market basket analysis, businesses need transactional data that includes details such as customer ID, product ID, and transaction date. This data is then used to create a matrix that represents the relationships between different products.

Association Rules:

Association rules are used to identify the strength of the relationship between different products. These rules are expressed in terms of support, confidence, and lift. Support refers to the frequency of co-occurrence of items in a transaction, while confidence measures the probability that if a customer buys one item, they will also buy another. Lift measures the degree of correlation between two items.

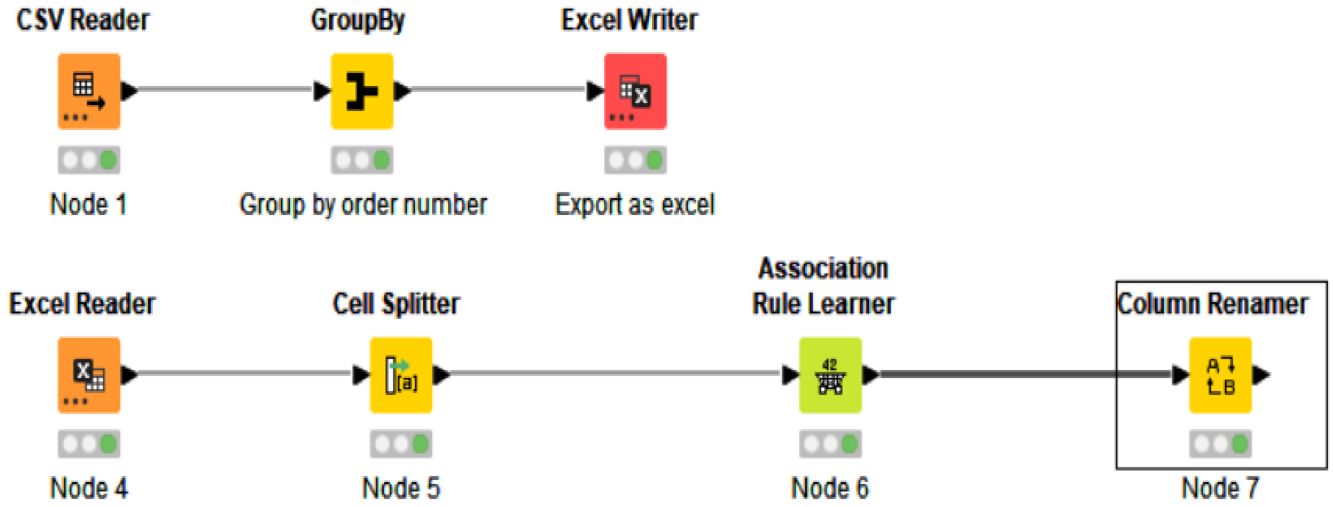
Applications:

Market Basket Analysis is used in a variety of industries, including retail, e-commerce, and marketing. Retailers use this technique to optimize product placement and promotions. E-commerce companies use it to personalize product recommendations, and marketers use it to develop targeted advertising campaigns.

Benefits:

Market Basket Analysis helps businesses increase revenue by identifying cross-selling opportunities and developing targeted promotions. It also helps improve customer satisfaction by providing personalized recommendations and improving the overall shopping experience

KNIME Workflow



Output Table

Table "default" - Rows: 24 | Spec - Columns: 6 | Properties | Flow Variables

Row ID	D Support	D Confide...	D Lift	S Rec_items	S implies	[...] Items
rule0	0.05	0.64	1.7	juice	<---	[yogurt,toilet paper,aluminum foil]
rule1	0.05	0.62	1.645	juice	<---	[yogurt,poultry,aluminum foil]
rule2	0.05	0.613	1.616	coffee/tea	<---	[yogurt,cheeses,cereals]
rule3	0.05	0.6	1.424	poultry	<---	[dishwashing liquid/detergent,laundry detergent,mixes]
rule4	0.051	0.63	1.678	mixes	<---	[yogurt,poultry,aluminum foil]
rule5	0.051	0.611	1.66	sandwich bags	<---	[cheeses,bagels,cereals]
rule6	0.051	0.674	1.726	cheeses	<---	[bagels,cereals,sandwich bags]
rule7	0.051	0.617	1.558	cereals	<---	[cheeses,bagels,sandwich bags]
rule8	0.051	0.63	1.621	dinner rolls	<---	[spaghetti sauce,poultry,cereals]
rule9	0.051	0.637	1.512	poultry	<---	[dinner rolls,spaghetti sauce,cereals]
rule10	0.051	0.604	1.589	milk	<---	[poultry,laundry detergent,cereals]
rule11	0.052	0.628	1.61	eggs	<---	[dinner rolls,poultry,soda]
rule12	0.052	0.641	1.649	dinner rolls	<---	[spaghetti sauce,poultry,ice cream]
rule13	0.052	0.686	1.628	poultry	<---	[dinner rolls,spaghetti sauce,ice cream]
rule14	0.052	0.628	1.614	dinner rolls	<---	[spaghetti sauce,poultry,juice]
rule15	0.052	0.602	1.429	poultry	<---	[dinner rolls,spaghetti sauce,juice]
rule16	0.052	0.634	1.627	eggs	<---	[paper towels,dinner rolls,pasta]
rule17	0.052	0.602	1.621	pasta	<---	[paper towels,eggs,dinner rolls]
rule18	0.054	0.642	1.651	dinner rolls	<---	[spaghetti sauce,poultry,laundry detergent]
rule19	0.054	0.656	1.556	poultry	<---	[dinner rolls,spaghetti sauce,laundry detergent]
rule20	0.055	0.624	1.565	ice cream	<---	[paper towels,eggs,pasta]
rule21	0.055	0.63	1.616	eggs	<---	[paper towels,ice cream,pasta]
rule22	0.055	0.643	1.731	pasta	<---	[paper towels,eggs,ice cream]
rule23	0.055	0.649	1.791	paper towels	<---	[eggs,ice cream,pasta]

Support, confidence, and lift values

Support:

It is the probability of observing the items together in a transaction. It is calculated as the number of transactions that contain both items divided by the total number of transactions. It measures how frequently the itemset occurs in the dataset. High support indicates that the itemset is popular and should be considered for promotion or placement together.

Confidence:

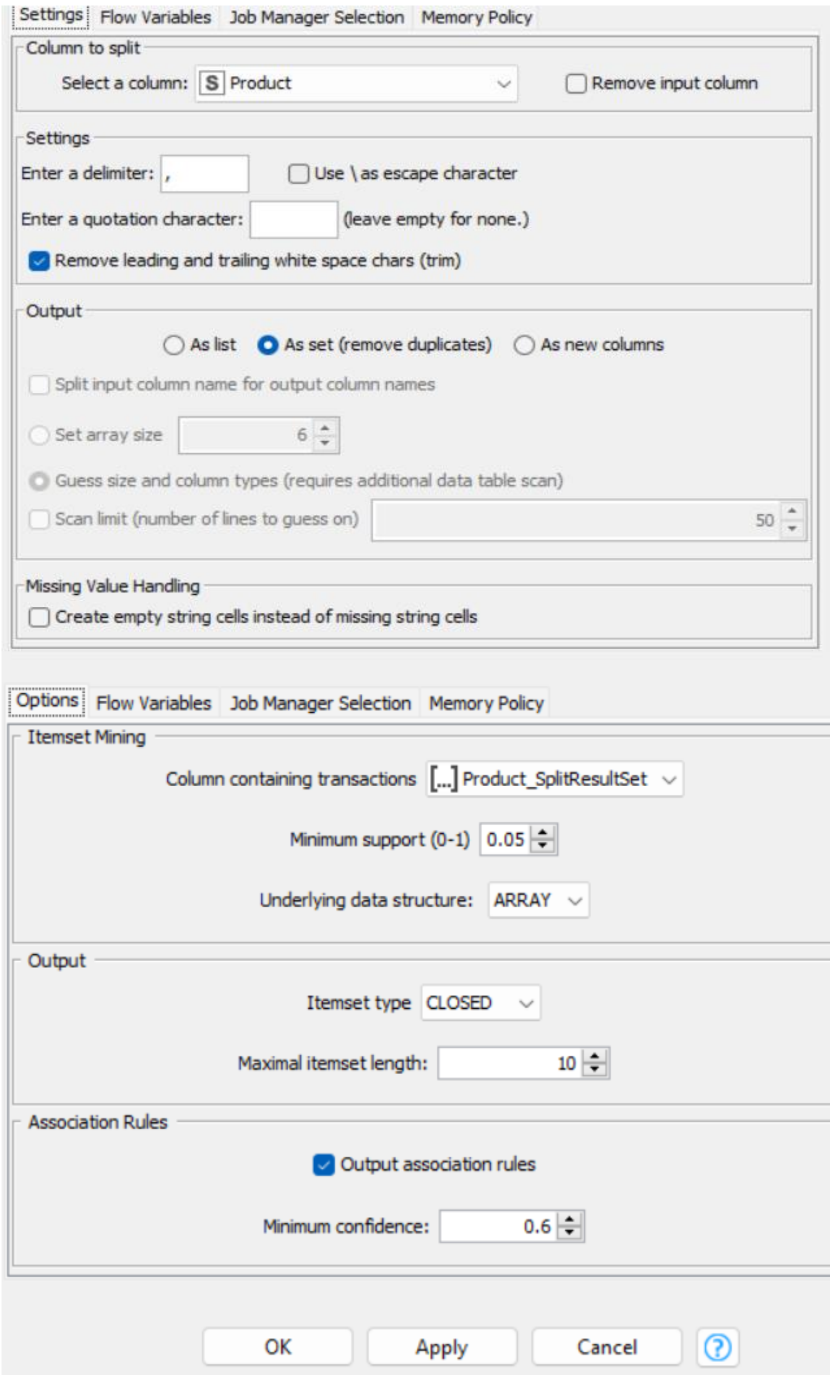
It is the conditional probability that a transaction containing one item also contains another item. It is calculated as the number of transactions containing both items divided by the number of transactions containing the first item. It measures the strength of the association between two items. High confidence indicates that the items are likely to be bought together, and can be used to recommend or suggest items to customers.

Lift:

It is the measure of how much more often two items occur together than expected if they were independent of each other. It is calculated as the support of the itemset divided by the product of the individual supports of the items. A lift value of 1 indicates that the items are independent, while a value greater than 1 indicates a positive association between the items. A lift value of less than 1 indicates a negative association between the items. A high lift indicates that the items have a strong association and can be used for cross-selling or bundling.

Parameters used

- Here, the support level is kept to 0.05 which means a probability of 5%
- The confidence is kept at 0.6 which means, it has an occurrence of 60%.
- For this dataset having only 37 different products generating a rule of more than 100's or 1000's will be misleading from a business perspective.
- The above parameter generates only 24 rules which has higher lift values than other parameter rules.



Insights and Recommendations

1) Cross-Promotional Strategies:

Rule 1: Juice is associated with purchasing yogurt, toilet paper, and aluminum foil. Consider placing these items together in promotions or displays to encourage cross-category purchases.

2) Bundle Offers:

Rule 3: Coffee/tea is associated with yogurt, cheeses, and cereals. Create bundle offers where customers can purchase these items together at a discounted price, promoting complementary products.

3) Product Placement Optimization:

Rule 6: Cheeses are associated with bagels, cereals, and sandwich bags. Place these items near each other in-store to increase visibility and encourage impulse buys.

4) Seasonal Promotions:

Rule 10: Eggs are associated with dinner rolls, poultry, and soda. Promote these combinations during relevant seasons or events (e.g., dinner rolls and poultry for holiday dinners).

5) Promotion of Convenience Products:

Rule 20: Paper towels are associated with eggs, ice cream, and pasta. Promote these items together as a convenience bundle, appealing to shoppers looking for easy meal solutions or household essentials.

6) Targeted Marketing Campaigns:

Analyze these associations to tailor marketing campaigns. For instance, focus on advertising juice alongside yogurt, toilet paper, and aluminum foil during breakfast hours or promote pasta with paper towels, eggs, and ice cream as a quick dinner solution.

Thank you