Marketing and Retail Analytics Project

Tableau links:
Part A
Part B

Name: Aishwariya Hariharan
PGP-DSBA Online September’ 23
Date: 07 June 2024

Contents

List of Figures

List of Tables

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 customised marketing strategies for different segments of customers.

Data Description

Sample of the Dataset:

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	DAYS_SINCE_LASTORDER	STATUS	PRODUCTLINE	MSRP
0	10107	30	95.70	2	2871.00	43155	828	Shipped	Motorcycles	95
1	10121	34	81.35	5	2765.90	43227	757	Shipped	Motorcycles	95
2	10134	41	94.74	2	3884.34	43282	703	Shipped	Motorcycles	95
3	10145	45	83.26	6	3746.70	43337	649	Shipped	Motorcycles	95
4	10168	36	96.66	1	3479.76	43401	586	Shipped	Motorcycles	95

Table 1: Sample dataset Part A

Data Dictionary

The detailed data dictionary description is given below:

1. **ORDERNUMBER** - This column represents the unique identification number assigned to each order.
2. **QUANTITYORDERED** - It indicates the number of items ordered in each order.
3. **PRICEEACH** - This column specifies the price of each item in the order.
4. **ORDERLINENUMBER** - It represents the line number of each item within an order.
5. **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.
6. **ORDERDATE** - It denotes the date on which the order was placed.
7. **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.
8. **STATUS** - It indicates the status of the order, such as "Shipped," "In Process," "Cancelled," "Disputed," "On Hold," or "Resolved"
9. **PRODUCTLINE** - This column specifies the product line categories to which each item belongs.
10. **MSRP** - It stands for Manufacturer's Suggested Retail Price and represents the suggested selling price for each item.
11. **PRODUCTCODE** - This column represents the unique code assigned to each product.
12. **CUSTOMERNAME** - It denotes the name of the customer who placed the order.
13. **PHONE** - This column contains the contact phone number for the customer.
14. **ADDRESSLINE1** - It represents the first line of the customer's address.
15. **CITY** - This column specifies the city where the customer is located.
16. **POSTALCODE** - It denotes the postal code or ZIP code associated with the customer's address.
17. **COUNTRY** - This column indicates the country where the customer is located.
18. **CONTACTLASTNAME** - It represents the last name of the contact person associated with the customer.
19. **CONTACTFIRSTNAME** - This column denotes the first name of the contact person associated with the customer.
20. **DEALSIZE** - It indicates the size of the deal or order, which are the categories "Small," "Medium," or "Large."

Data Types

```
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)
```

Figure 1: Data description Part A

About data:

- 20 columns - 12 object type, 6 integer type, 2 float type
- 2747 rows
- No duplicate rows
- No null values

Statistical Summary of the Dataset

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	DAYS_SINCE_LASTORDER	MSRP
count	2747.000000	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	43598.914088	1757.085912	100.691664
std	91.877521	9.762135	42.042548	4.230544	1838.953901	230.231295	819.280576	40.114802
min	10100.000000	6.000000	26.880000	1.000000	482.130000	43106.000000	42.000000	33.000000
25%	10181.000000	27.000000	68.745000	3.000000	2204.350000	43412.000000	1077.000000	68.000000
50%	10264.000000	35.000000	95.550000	6.000000	3184.800000	43640.000000	1761.000000	99.000000
75%	10334.500000	43.000000	127.100000	9.000000	4503.095000	43786.000000	2436.500000	124.000000
max	10425.000000	97.000000	252.870000	18.000000	14082.800000	43982.000000	3562.000000	214.000000

Table 2: Statistical Summary Part A

- The data is not scaled.

Exploratory Data Analysis

Univariate Analysis

- **Distribution of quantity ordered**

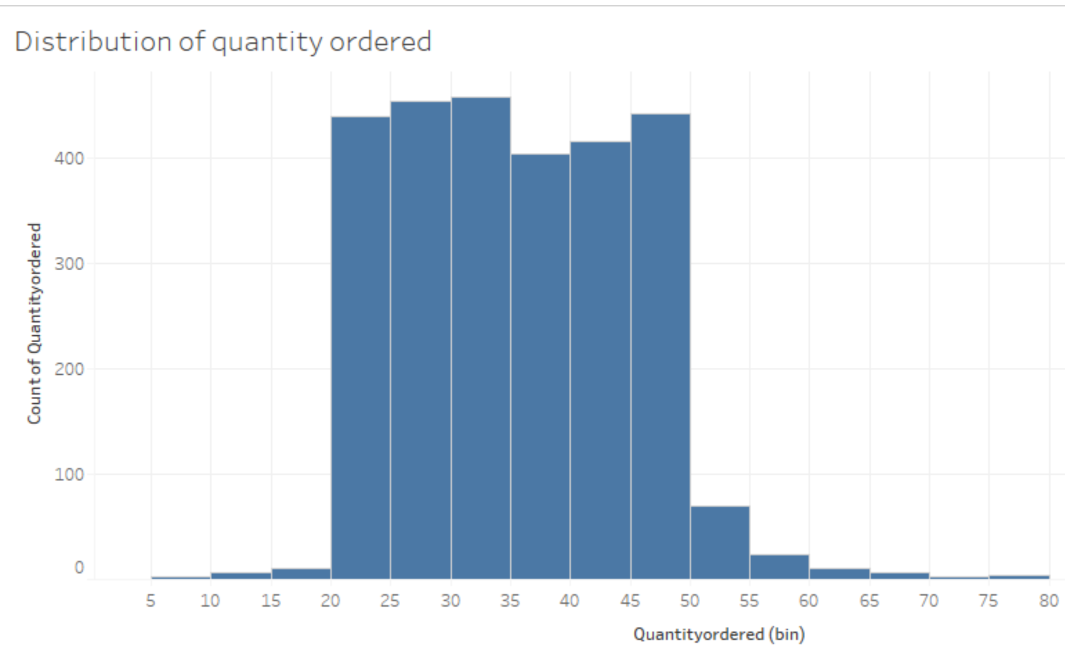


Figure 2: Distribution of quantity ordered

- Most customers' order quantity is between 20 to 50.
- The median order quantity is about 35.

- **Distribution of the price of each item**

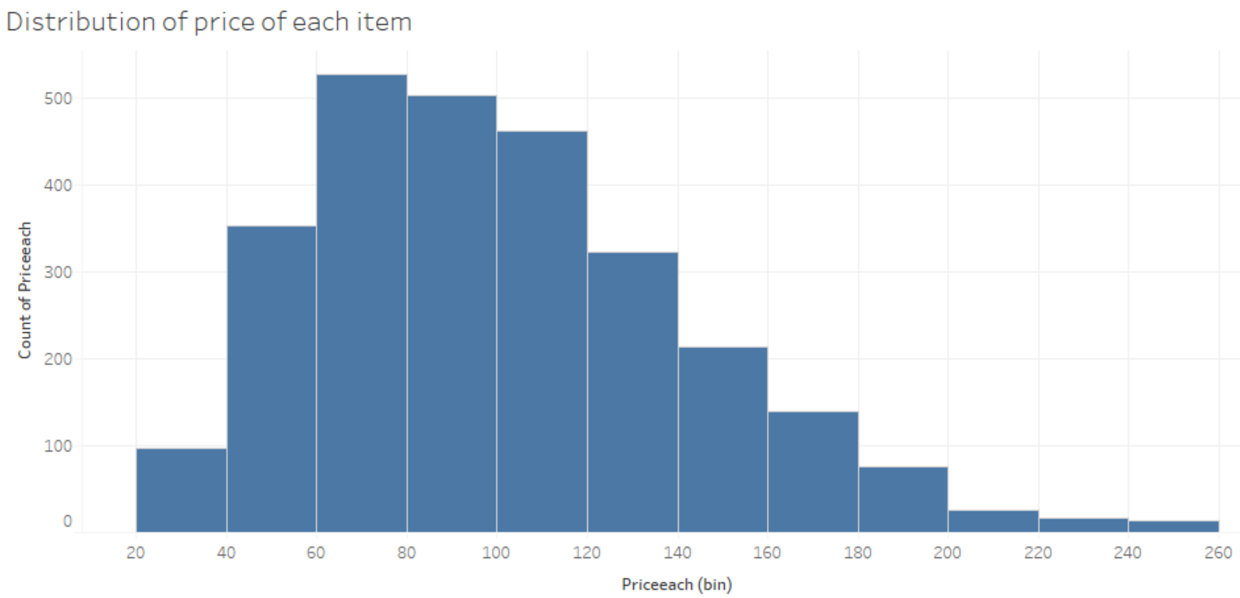


Figure 3: Distribution of the price of each item

- Most items are priced between $40 and $140.
- The median price of each item is about $95.
- The data is skewed to the right.

- **Distribution of sales**

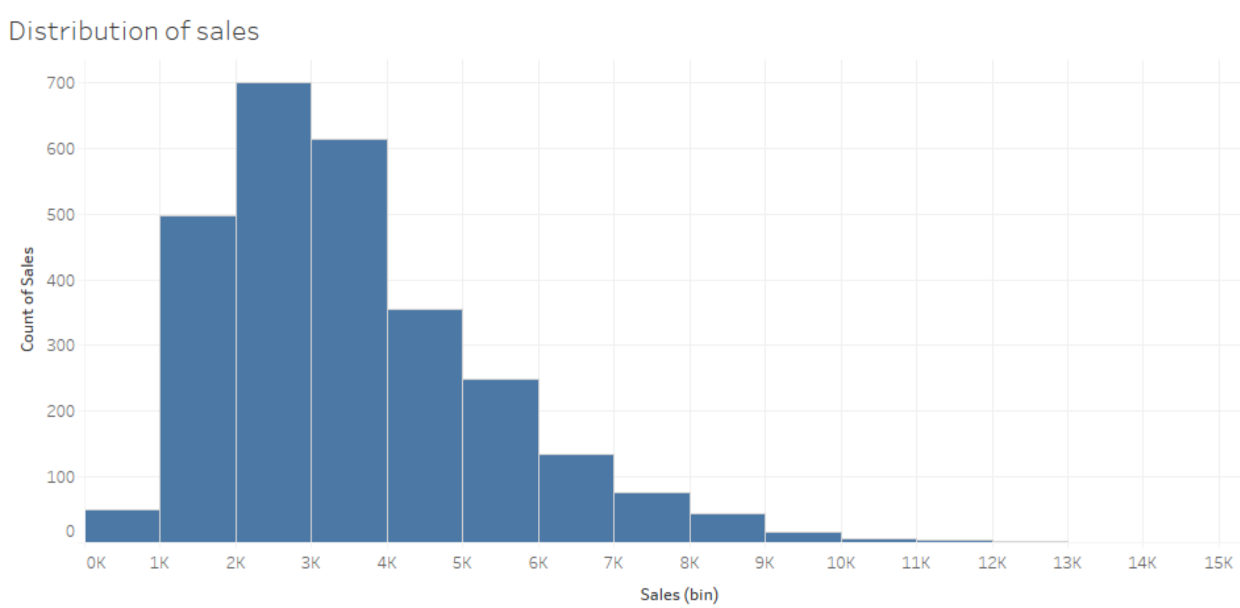


Figure 4: Distribution of sales

- The highest sales frequency is between 2k to 3k.
- The median of sales amount is $3184.
- The data is skewed to the right.

- **Days since the last order distribution**

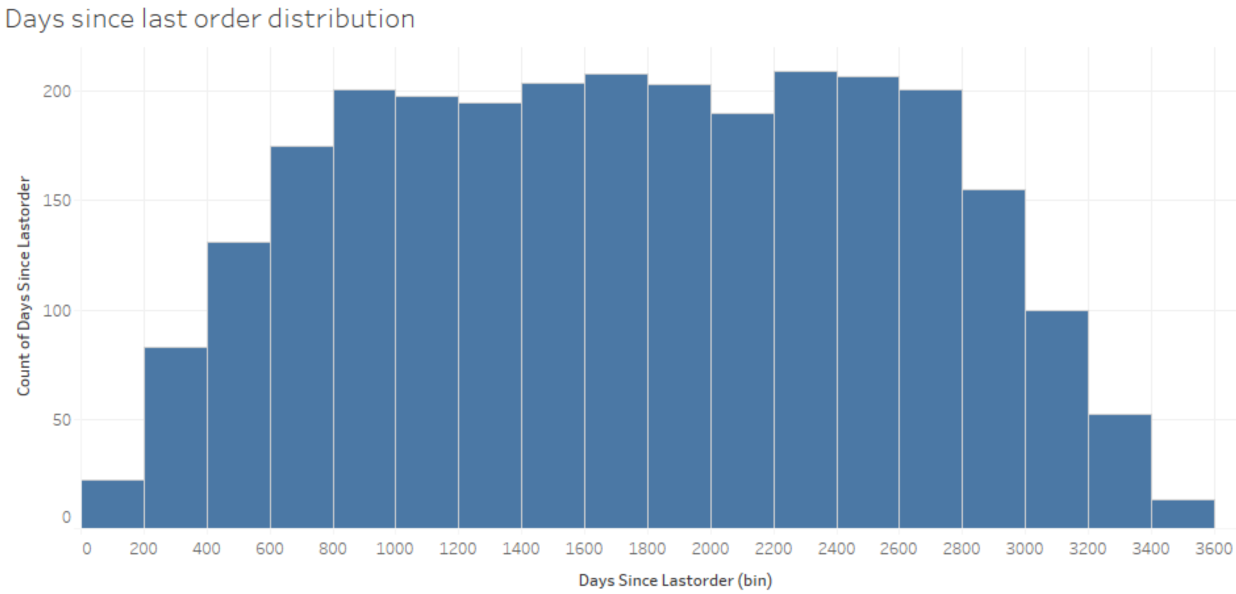


Figure 5: Days since last order distribution

- Most of the customers’ last order was about 900 to 2800 days ago.
- The median days since last order is 1761 days.

- **Distribution of MSRP**

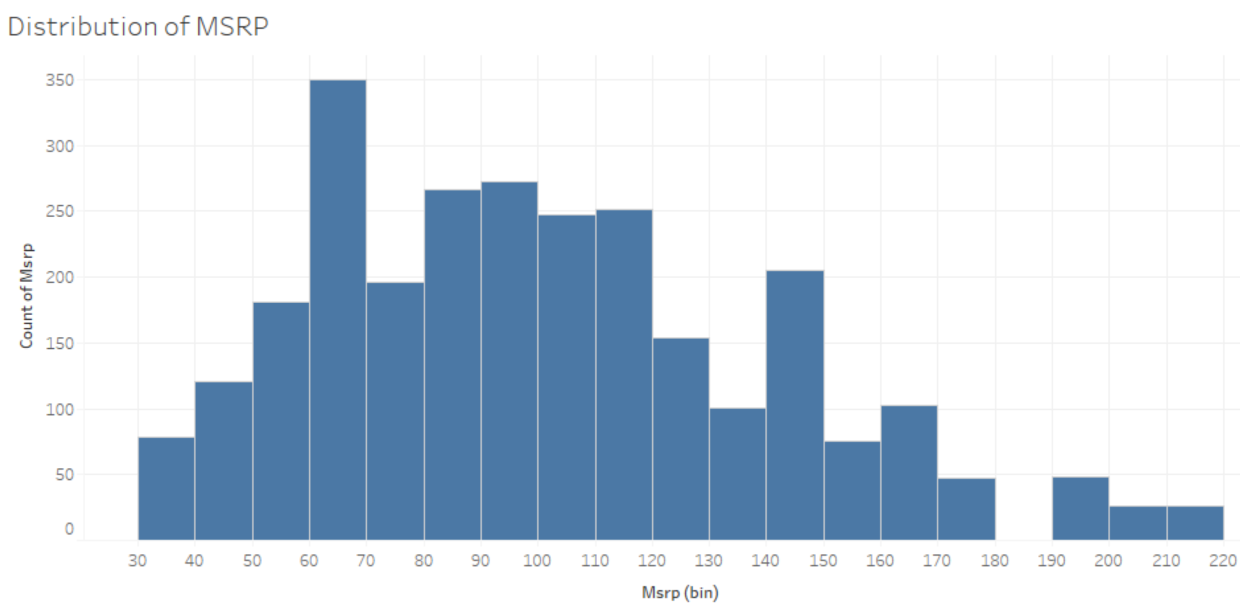


Figure 6: Distribution of MSRP

- For most items, manufacturers' suggested retail price is about $60 to $70.
- However, the median MSRP is $99.

Bivariate and Multivariate Analysis

- **Average sales over time**

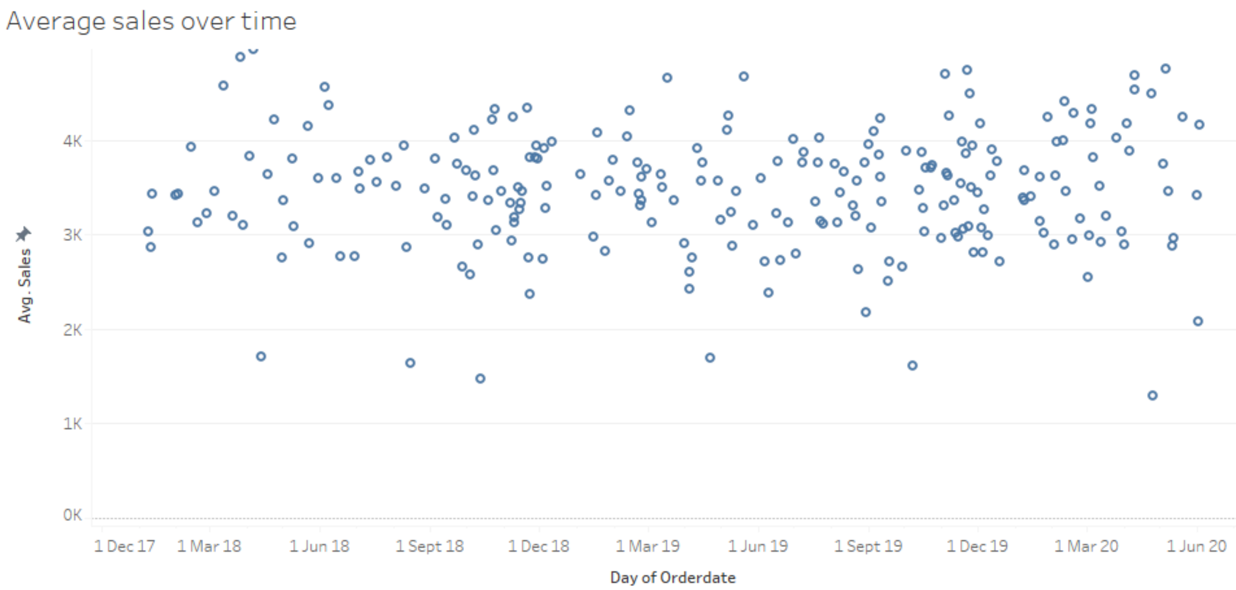


Figure 7: Average sales over time

- The average sales from Jan 20188 to May 2020 are between the range of $2000 to $4500.

- **Sum of sales over time**

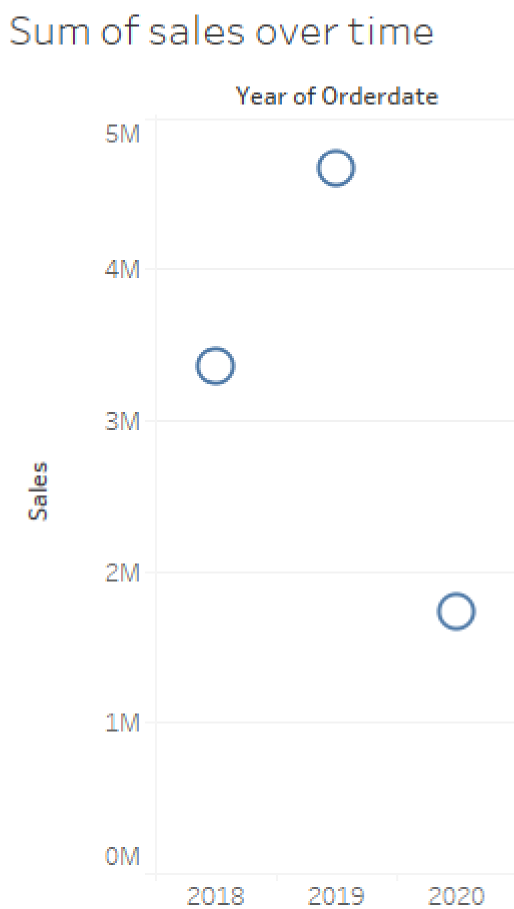


Figure 8: Sum of sales over time

- The total sales is highest for the year 2019 compared to 2018.
- THe year 2020 has record only till the month of May.

- **Sales trend over time**

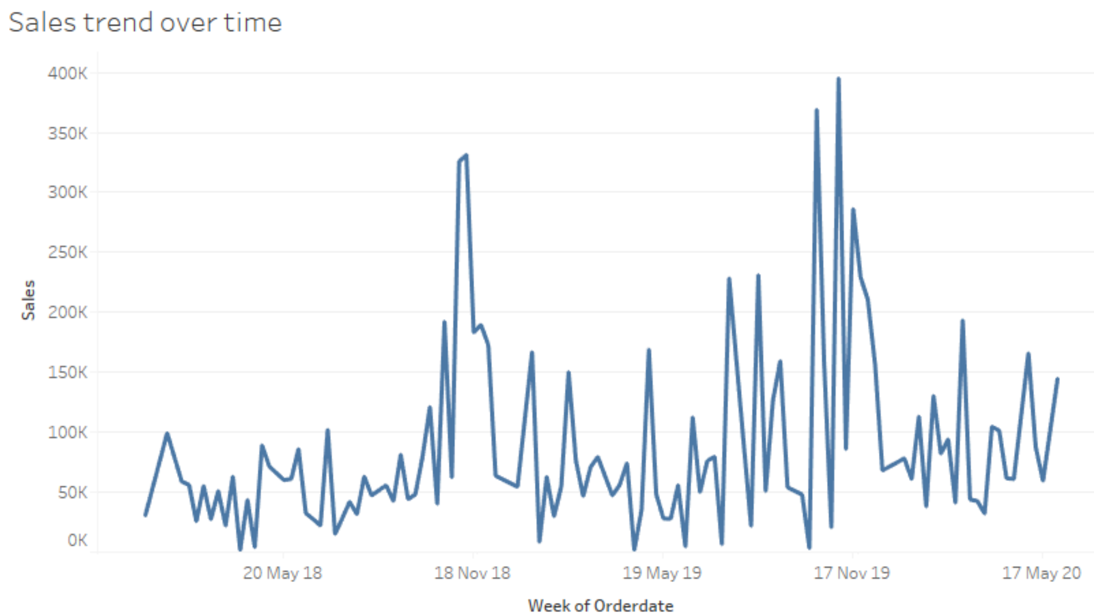


Figure 9: Sales trend over time

- The sales was the highest during Nov 2018 and Oct-Nov 2019,

- **Productline vs Sales**

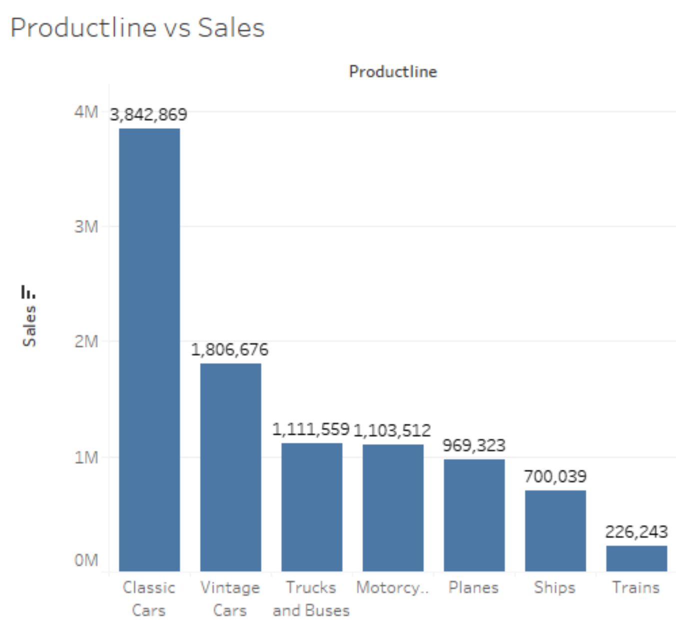


Figure 10: Productline vs Sales

- The highest sales are for classic cars, followed by vintage cars which is half the sales of classic cars

- **Deal size vs Sales**

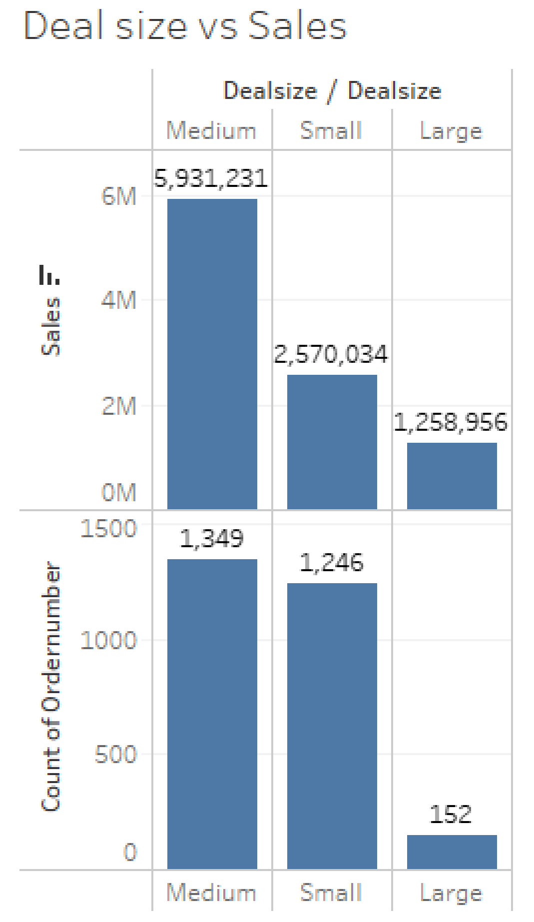


Figure 11: Deal size vs Sales

- Medium-sized deals are the ones generating the highest sales.
- Although small-sized deals are almost as much in number as medium-sized, the sales values is half that of medium-sized sales values.

- **Distinct Orderno. vs Country**

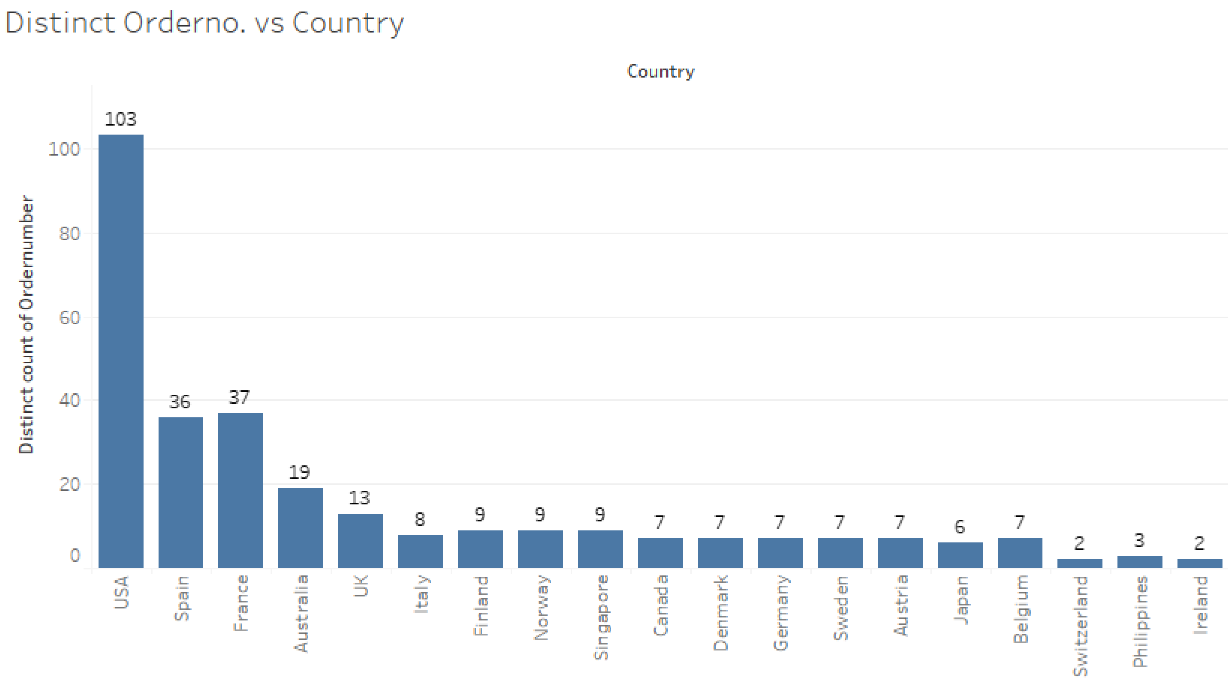


Figure 12: Distinct Orderno. vs Country

- **Sales in countries**

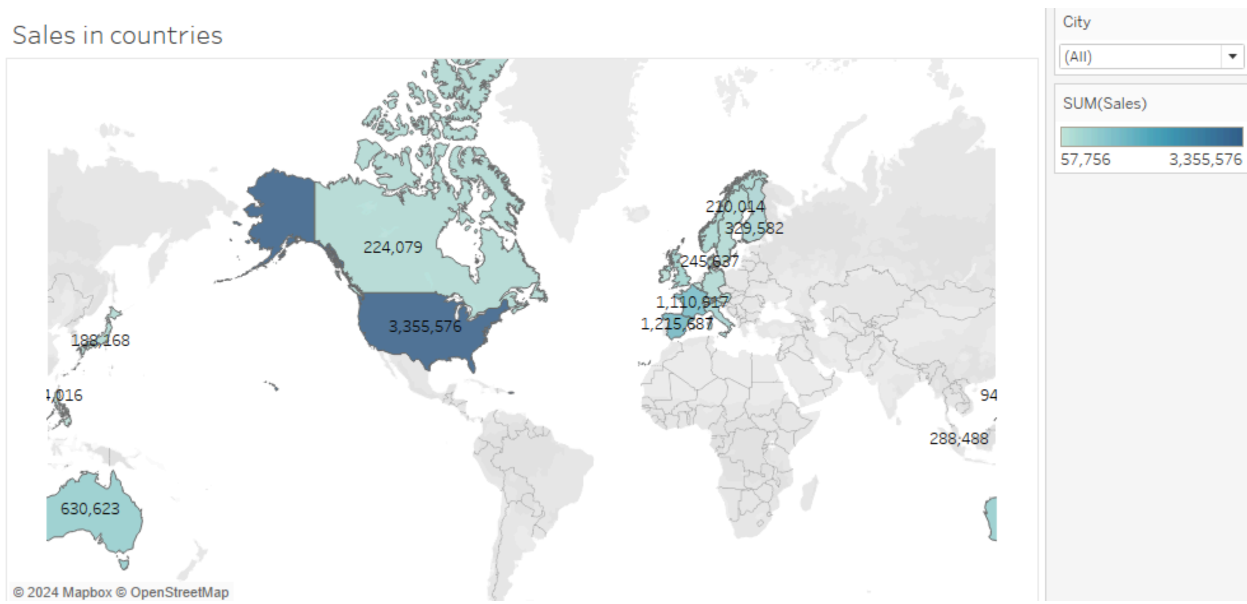


Figure 13: Sales in countries

- USA has the highest sales across all countries.
- Spain has the second-highest sales value, followed by France and Australia.

- **Deal size vs Country**

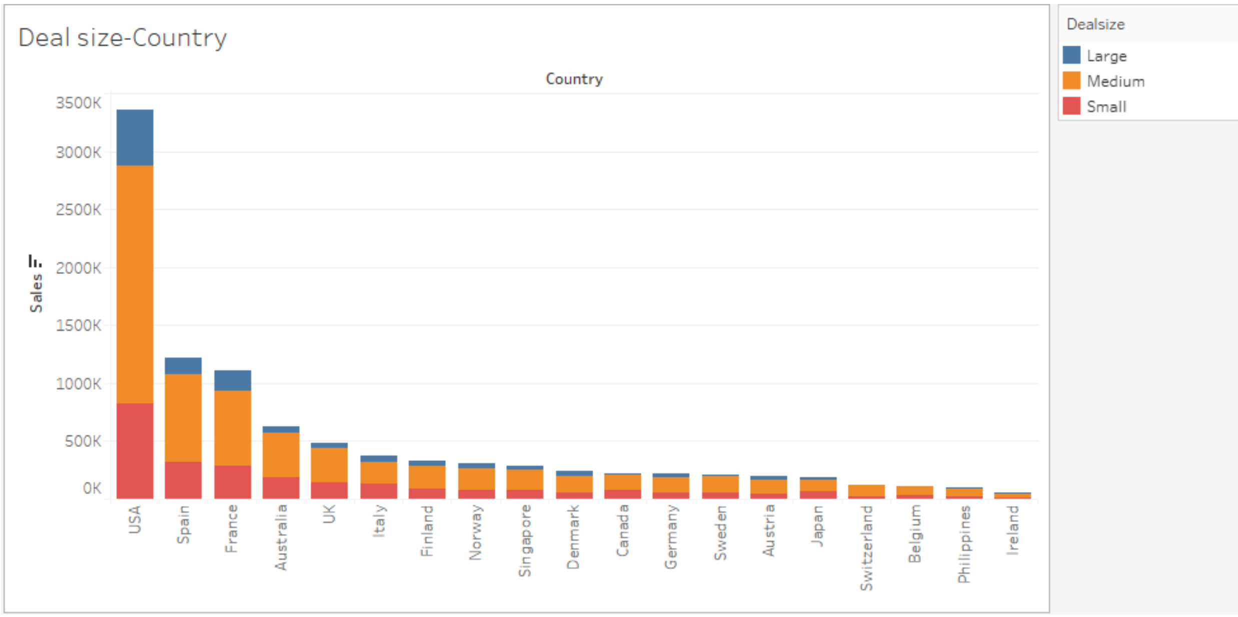


Figure 14: Deal size vs Country

  - In all the countries, medium-sized deals are the highest in number.

- **Avg sales vs Status**

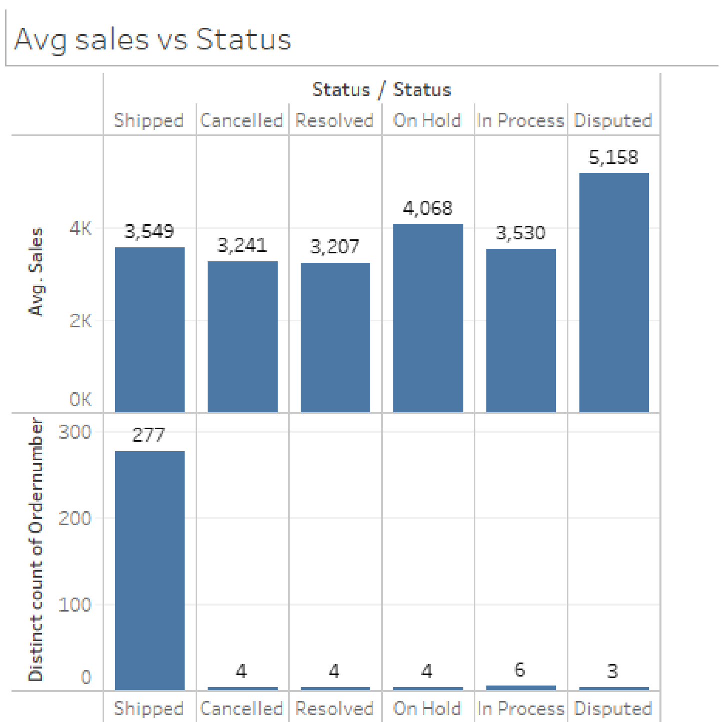


Figure 15: Avg sales vs Status

- Most of the orders are shipped.
- Although there are only 3 disputed orders, their average sales is significantly higher than those of shipped items.

- **MSRP vs Productline**

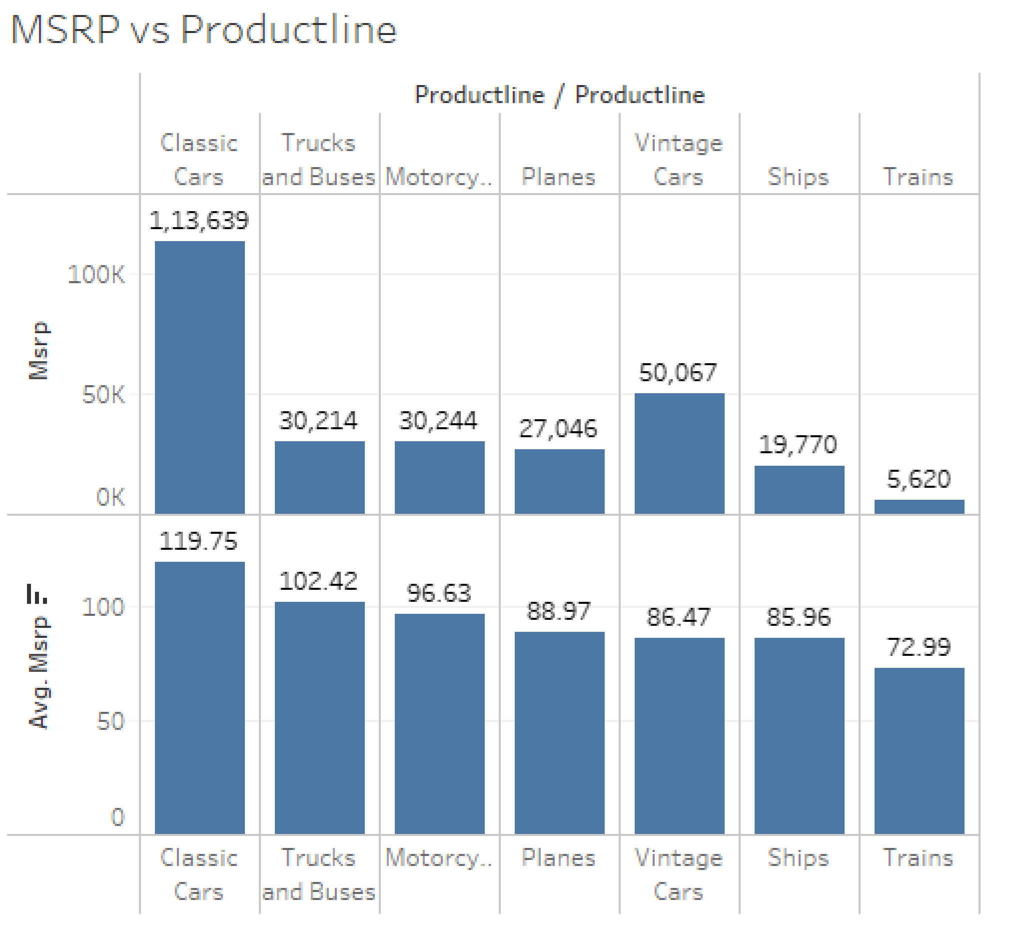


Figure 16: MSRP vs Productline

- The MSRP for classic cars is the highest, which is obvious since the sales of classic cars are the highest.

- **Correlation**

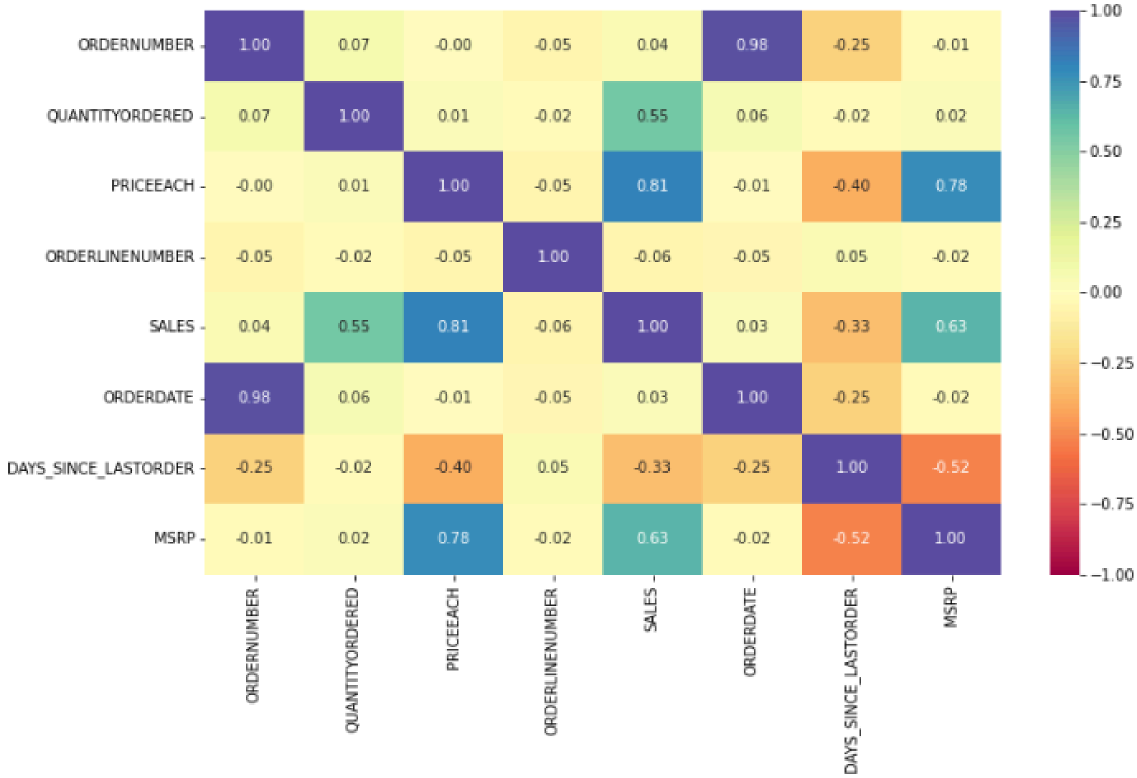


Figure 17: Correlation plot

- There is a strong correlation between
  - Price of each item and sales
  - Price of each item and MSRP
  - Sales and MSRP
- We do not consider the order number column since it's technically categorical.

What is RFM analysis?

RFM analysis evaluates customers based on three key metrics:

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

Customer Segmentation using RFM Analysis into 4 segments

Based on the RFM scores, customers can be segmented into four key groups:

1. **Champions**:
    - **Characteristics:** High Recency, High Frequency, High Monetary Value.
    - **Description:** These are your best customers who buy often, spend the most, and have made a purchase recently.

2. **Loyal Customers**:
    - **Characteristics:** Medium Recency, High Frequency, Medium to High Monetary Value.
    - **Description:** These customers buy often but may not have the highest spending.

3. **Potential Loyalists**:
    - **Characteristics:** High Recency, Medium Frequency, Medium Monetary Value.
    - **Description:** These customers have made recent purchases but are not frequent buyers yet.

4. **At-Risk Customers**:
    - **Characteristics:** Low Recency, Medium to Low Frequency, Medium to Low Monetary Value.
    - **Description:** These customers haven’t purchased recently and may be at risk of churning.

What parameters were used and assumptions made?

- [ ] We group the customers based on their names so that we can later determine our top customers and the potential ones to churn.

- [ ] We use the following parameters and the corresponding aggregations for our analysis:

I	QUANTITYORDERED	Mean	☐	
D	PRICEEACH	Mean	☐	
D	SALES	Mean	☐	
31	ORDERDATE	Count	☑	
I	DAYS_SINCE_LASTORDER	Minimum	☐	
S	STATUS	List (sorted)	☑	
S	PRODUCTLINE	Unique concatenate with count	☑	
I	MSRP	Mean	☐	
S	PHONE	First	☑	
S	ADDRESSLINE1	First	☑	
S	CITY	First	☑	
S	POSTALCODE	First	☑	
S	COUNTRY	First	☑	
S	CONTACTLASTNAME	First	☑	
S	CONTACTFIRSTNAME	First	☑	

Figure 18: Aggregations for RFM Analysis

- [ ] Here, DAYS_SINCE_LASTORDER is the Recency metric.
    - [ ] Since we need the most recent purchase date, we use the minimum aggregation.

- [ ] ORDERDATE is the Frequency metric.
    - [ ] Since we need to find how many times (one order is taken for one day) the customer making the purchases, we take the count of the dates on which the customer places an order.
    - [ ] More the better.

- [ ] SALES is the Monetary metric.
    - [ ] Sales is the product of the price of each item and the quantity ordered.
    - [ ] We take the mean of the sales to find how much each customer spends on an average.

KNIME workflow image

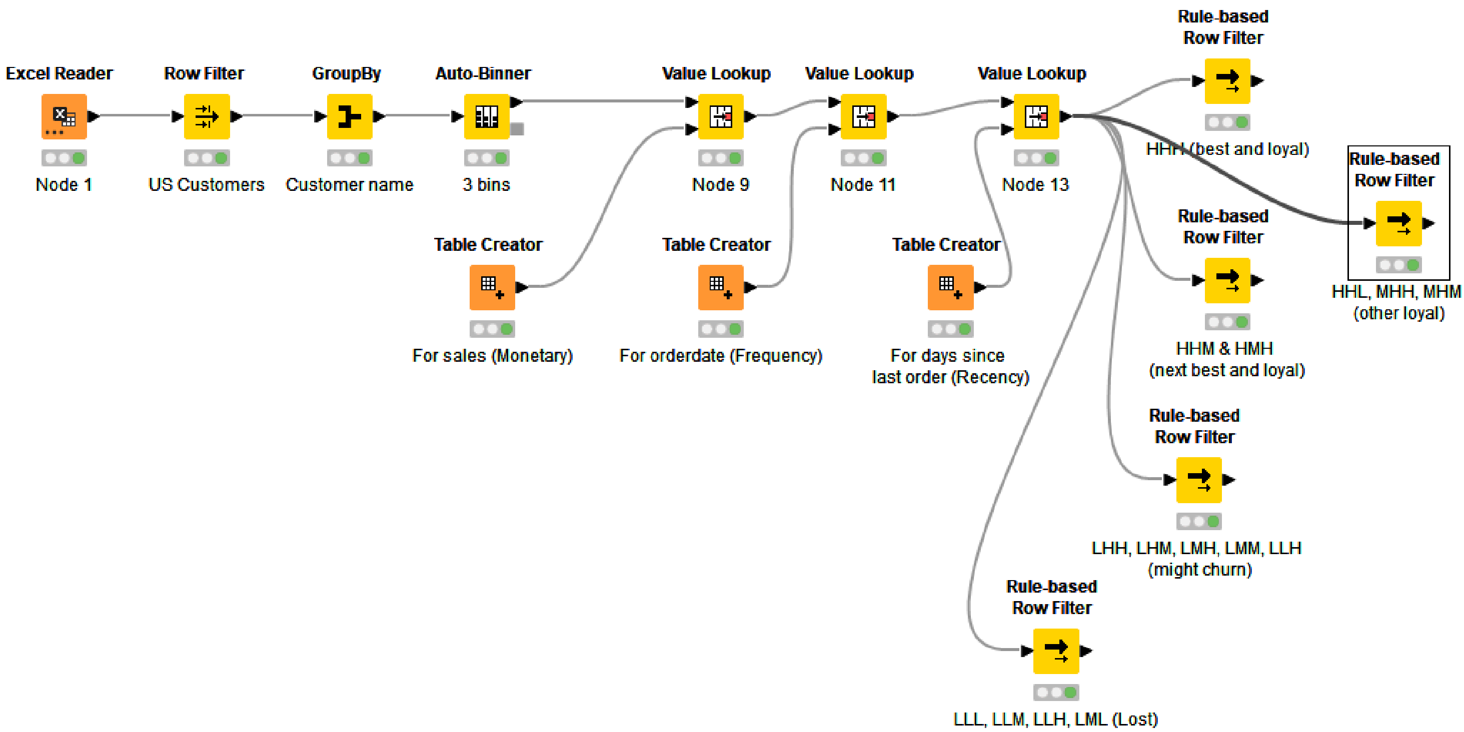


Figure 19: KNIME workflow for RFM Analysis

What results are there in the output table head?

CUSTOMERNAME
QUANTITYORDERED (Mean)
PRICEEACH (Mean)
SALES (Mean)
ORDERDATE (Count)
DAYS_SINCE_LASTORDER (Min*)
STATUS (Sorted list)
PRODUCTLINE (Unique concatenate with count)
MSRP (Mean)
PHONE (First)
ADDRESSLINE1 (First)
CITY (First)
POSTALCODE (First)
COUNTRY (First)
CONTACTLASTNAME (First)
CONTACTFIRSTNAME (First)
SALES (Mean) [Binned]
ORDERDATE (Count) [Binned]
DAYS_SINCE_LASTORDER (Min*) [Binned]
Sales (Monetary)
Orderdate (Frequency)
Days_since_last_order (Recency)

Figure 20: RFM output columns

- The 4 bins (**Bin 1, Bin 2, and Bin 3,**) that the RFM columns were binned into were converted into the labels **Low, Medium, and High**.
- This is because we need to reverse the ranking of bins for the recency metric. The lowest recency value gets the label High.

Table "default" - Rows: 89 | Spec - Columns: 20 | Properties | Flow Variables

Row ID	S CUSTOMERNAME	D QUANT...	D PRICEE...	D SALES (...	I ORDERDATE ...	I DAYS_...	[...] STATUS (Sor...	S PRODUCTLINE (Unique co...	D MSRP (...	S PHO...	S AD...	S CITY (F...	S POSTA...	S COUNTRY...
Row0	AV Stores, Co.	34.863	91.085	3,094.271	51	421	[Shipped,Shipped...	Classic Cars(19), Vintage Cars(...	92.843	(171) 55...	Fauntler...	Manchester	EC2 5NT	UK
Row1	Alpha Cognac	34.35	101.16	3,524.422	20	675	[Shipped,Shipped...	Classic Cars(4), Planes(6), Ship...	97.15	61.77.6555	1 rue Al...	Toulouse	31000	France
Row2	Amica Models & Co.	32.423	110.853	3,619.895	26	328	[Shipped,Shipped...	Classic Cars(4), Trucks and Bus...	107.654	011-498...	Via Mon...	Torino	10100	Italy
Row3	Anna's Decorations, Ltd	31.935	106.424	3,347.742	46	131	[Shipped,Shipped...	Classic Cars(24), Trucks and Bu...	104.717	02 9936 ...	201 Mill...	North Sydney	2060	Australia
Row4	Atelier graphique	38.571	92.239	3,454.28	7	312	[Shipped,Shipped...	Motorcycles(2), Classic Cars(4)...	95.571	40.32.2555	54, rue ...	Nantes	44000	France
Row5	Australian Collectables, Ltd	30.652	90.042	2,808.324	23	1018	[Disputed,Dispute...	Vintage Cars(17), Classic Cars(...	88.13	61-9-384...	7 Allen ...	Glen Waverly	3150	Australia
Row6	Australian Collectors, Co.	35.018	104.59	3,654.462	55	229	[Shipped,Shipped...	Motorcycles(14), Classic Cars(1...	103.527	03 9520 ...	636 St K...	Melbourne	3004	Australia
Row7	Australian Gift Network, Co	36.333	110.554	3,964.608	15	190	[Shipped,Shipped...	Classic Cars(4), Motorcycles(4)...	111.533	61-7-384...	31 Dunc...	South Brisbane	4101	Australia
Row8	Auto Assoc. & Cie.	35.389	99.488	3,601.907	18	275	[Shipped,Shipped...	Classic Cars(2), Trucks and Bus...	100.389	30.59.8555	67, ave...	Versailles	78000	France
Row9	Auto Canal Petit	37.074	94.255	3,450.765	27	127	[Shipped,Shipped...	Motorcycles(16), Classic Cars(11)	94.852	(1) 47.55...	25, rue ...	Paris	75016	France
Row10	Auto-Moto Classics Inc.	35.875	92.8	3,309.908	8	1353	[Shipped,Shipped...	Ships(2), Vintage Cars(3), Plan...	87.375	6175558...	16780 P...	Brickhaven	58339	USA
Row11	Baane Mini Imports	33.812	108.574	3,643.725	32	245	[Shipped,Shipped...	Motorcycles(7), Classic Cars(7)...	107.469	07-98 9555	Erling Sk...	Stavern	4110	Norway
Row12	Bavarian Collectables Impo...	28.643	84.289	2,499.566	14	801	[Shipped,Shipped...	Planes(8), Ships(2), Vintage Ca...	82.714	+49 89 6...	Hansast...	Munich	80686	Germany
Row13	Blauer See Auto, Co.	36.864	108.031	3,871.436	22	705	[Shipped,Shipped...	Classic Cars(13), Vintage Cars(...	105.818	+49 69 6...	Lyoners...	Frankfurt	60528	Germany
Row14	Boards & Toys Co.	34	89.807	3,043.117	3	410	[Shipped,Shipped...	Classic Cars(1), Vintage Cars(1...	92.333	3105552...	4097 Do...	Glendale	92561	USA
Row15	CAF Imports	36	104.963	3,818.619	13	625	[Shipped,Shipped...	Classic Cars(2), Vintage Cars(3...	106.923	+34 913 ...	Mercha...	Madrid	28023	Spain
Row16	Cambridge Collectables Co.	32.455	101.329	3,287.602	11	484	[Shipped,Shipped...	Classic Cars(1), Trucks and Bus...	97.364	6175555...	4658 Ba...	Cambridge	51247	USA
Row17	Canadian Gift Exchange N...	31.955	105.341	3,419.951	22	364	[Shipped,Shipped...	Classic Cars(5), Trucks and Bus...	106.409	(604) 55...	1900 O...	Vancouver	V3F 2K1	Canada
Row18	Classic Gift Ideas, Inc	31.81	103.32	3,214.618	21	344	[Shipped,Shipped...	Classic Cars(4), Trucks and Bus...	102.476	2155554...	782 Firs...	Philadelphia	71270	USA
Row19	Classic Legends Inc.	36	109.803	3,889.76	20	309	[Shipped,Shipped...	Classic Cars(6), Trucks and Bus...	106.65	2125558...	5905 Po...	NYC	10022	USA
Row20	Clover Collections, Co.	30.625	112.87	3,609.777	16	659	[Shipped,Shipped...	Classic Cars(6), Planes(4), Mot...	106.875	+353 18...	25 Maid...	Dublin	2	Ireland
Row21	Collectable Mini Designs Co.	38.16	91.535	3,499.569	25	575	[Shipped,Shipped...	Classic Cars(8), Planes(6), Ship...	93.12	7605558...	361 Fur...	San Diego	91217	USA
Row22	Collectables For Less Inc.	33.125	97.237	3,399.083	24	179	[Shipped,Shipped...	Classic Cars(12), Vintage Cars(...	99.458	6175558...	7825 Do...	Brickhaven	58339	USA
Row23	Corrida Auto Replicas, Ltd	36.344	105.175	3,769.227	32	407	[Shipped,Shipped...	Classic Cars(6), Trucks and Bus...	102.625	(91) 555 ...	C/ Araq...	Madrid	28023	Spain
Row24	Cruz & Sons Co.	36.962	96.08	3,615.99	26	971	[Shipped,Shipped...	Classic Cars(13), Planes(5), Mo...	97.962	+63 2 55...	15 McC...	Makati City	1227 MM	Philippines
Row25	Daedalus Designs Imports	34.95	95.474	3,452.62	20	573	[Shipped,Shipped...	Motorcycles(18), Classic Cars(2)	94.5	20.16.1555	184, ch...	Lille	59000	France
Row26	Danish Wholesale Imports	36.528	108.038	4,028.933	36	499	[Disputed,Dispute...	Classic Cars(14), Vintage Cars(...	106.417	31 12 3555	Vinb'ltet...	Kobenhavn	1734	Denmark
Row27	Diecast Classics Inc.	35.839	108.566	3,939.94	31	228	[In Process,In Pr...	Motorcycles(7), Classic Cars(8)...	106.581	2155551...	7586 Po...	Allentown	70267	USA
Row28	Diecast Collectables	38.611	101.783	3,936.654	18	672	[Shipped,Shipped...	Classic Cars(7), Trucks and Bus...	103.722	6175552...	6251 In...	Boston	51003	USA
Row29	Double Decker Gift Stores,...	29.75	99.108	3,001.587	12	670	[Shipped,Shipped...	Classic Cars(3), Planes(3), Ship...	93.25	(171) 55...	120 Han...	London	WA1 1DP	UK
Row30	Dragon Souveniers, Ltd.	35.442	113.106	4,023.016	43	649	[Shipped,Shipped...	Classic Cars(17), Trucks and Bu...	113.442	+65 221 ...	Bronz S...	Singapore	79903	Singapore
Row31	Enaco Distributors	38.348	88.783	3,409.211	23	659	[Shipped,Shipped...	Classic Cars(1), Ships(9), Vinta...	87.087	(93) 203 ...	Rambla ...	Barcelona	8022	Spain
Row32	Euro Shopping Channel	36.012	97.383	3,522.371	259	42	[Cancelled,Cancel...	Motorcycles(17), Classic Cars(1...	97.015	(91) 555 ...	C/ Moral...	Madrid	28034	Spain
Row33	FunGiftIdeas.com	34.731	109.587	3,804.759	26	111	[Shipped,Shipped...	Motorcycles(4), Classic Cars(15...	108	5085552...	1785 Fir...	New Bedford	50553	USA
Row34	Gift Depot Inc.	36.12	108.932	4,075.792	25	226	[Shipped,Shipped...	Motorcycles(8), Classic Cars(12...	110.92	2035552...	25593 S...	Bridgewater	97562	USA
Row35	Gift Ideas Corp.	35.053	87.6	3,015.496	19	947	[Shipped,Shipped...	Planes(12), Motorcycles(2), Cla...	86.526	2035554...	2440 Po...	Glendale	97561	USA
Row36	Gifts4AllAges.com	35.885	91.564	3,200.38	26	148	[On Hold,On Hold...	Classic Cars(4), Ships(11), Vint...	90.731	6175559...	8616 Sp...	Boston	51003	USA
Row37	Handji Gifts& Co	34.333	95.593	3,208.298	36	488	[Shipped,Shipped...	Classic Cars(15), Trucks and Bu...	97.222	+65 224 ...	Village C...	Singapore	69045	Singapore

Table 3: Output table of RFM Analysis

Inferences from RFM Analysis and identified segments

1. Who are your best customers?

Best customers are those who have either high recency, frequency, and monetary metrics, or high recency and frequency, but medium monetary metrics, or high recency and monetary and medium frequency metrics.

CUSTOMERNAME	QUANT…	PRICEE…	SALES (…	ORDER…	DAYS_…	STATUS (Sorted list)	PRODUCTLINE (Uniq…	MSRP (…	PHON…	ADDRESSL…	CITY (F…	POSTA…	COUNT…
Online Diecast Creations Co.	36.706	108.302	3,873.097	34	253	[Shipped,Shipped,Ship…	Classic Cars(18), Trucks a…	105	6035558647	2304 Long Airp…	Nashua	62005	USA
The Sharp Gifts Warehouse	41.4	93.376	4,000.257	40	182	[On Hold,On Hold,On …	Classic Cars(10), Planes(1…	92.775	4085553659	3086 Ingle Ln.	San Jose	94217	USA

CUSTOMERNAME	QUANT…	PRICEE…	SALES (…	ORDERDAT…	DAYS_…	STATUS (Sor…	PRODUCTLINE (Unique concatenate with count)	MSRP (…	PHO…	ADD…	CITY (F…	..	COUNT..
Anna's Decorations, Ltd	31.935	106.424	3,347.742	46	131	[Shipped,Shipped…	Classic Cars(24), Trucks and Buses(9), Motorcycles(6…	104.717	02 9936 8…	201 Miller…	North Sydney	2060	Australia
Australian Collectors, …	35.018	104.59	3,654.462	55	229	[Shipped,Shipped…	Motorcycles(14), Classic Cars(12), Trucks and Buses(…	103.527	03 9520 4…	636 St Kil…	Melbourne	3004	Australia
Diecast Classics Inc.	35.839	108.566	3,939.94	31	228	[In Process,In Pr…	Motorcycles(7), Classic Cars(8), Trucks and Buses(9)…	106.581	2155551555	7586 Po…	Allentown	70…	USA
Euro Shopping Channel	36.012	97.383	3,522.371	259	42	[Cancelled,Cance…	Motorcycles(17), Classic Cars(106), Trucks and Buse…	97.015	(91) 555 …	C/ Moralz…	Madrid	28…	Spain
Gift Depot Inc.	36.12	108.932	4,075.792	25	226	[Shipped,Shipped…	Motorcycles(8), Classic Cars(12), Planes(3), Trains(2)	110.92	2035552570	25593 So…	Bridgewater	97…	USA
La Rochelle Gifts	34.566	97.046	3,398.583	53	139	[In Process,In Pr…	Motorcycles(14), Classic Cars(6), Trucks and Buses(8…	96.151	40.67.8555	67, rue d…	Nantes	44…	France
Land of Toys Inc.	33.286	104.121	3,348.356	49	216	[Cancelled,Cance…	Motorcycles(16), Classic Cars(13), Trucks and Buses(…	102.98	2125557818	897 Long…	NYC	10…	USA
Mini Gifts Distributors …	35.367	102.696	3,638.1	180	219	[In Process,In Pr…	Classic Cars(67), Trucks and Buses(36), Vintage Cars…	102.506	4155551450	5677 Str…	San Rafael	97…	USA
Salzburg Collectables	36.05	101.398	3,744.966	40	188	[Shipped,Shipped…	Motorcycles(5), Classic Cars(19), Planes(6), Vintage …	102.925	6562-9555	Geislweg 14	Salzburg	5020	Austria
Technics Stores Inc.	34.676	104.914	3,552.443	34	241	[Shipped,Shipped…	Motorcycles(11), Classic Cars(3), Trucks and Buses(7…	102.294	6505556809	9408 Fur…	Burlingame	94…	USA
UK Collectables, Ltd.	36.069	108.536	4,069.251	29	76	[Cancelled,Cance…	Motorcycles(9), Classic Cars(18), Vintage Cars(2)	110.276	(171) 555…	Berkeley …	Liverpool	WX…	UK

Table 4: Best customers

- **Marketing Strategy:** Reward them with exclusive offers, early access to new products, and personalized communications.
- **Fix issues (if any):** If the order status is on hold or disputed, it is important to ensure that they are resolved as soon as possible.

2. Which customers are on the verge of churning?

These are customers who have a low recency metric but medium to high frequency and monetary metrics. The low recency might be a concern for churn because maybe they might have found a better offer for the automobile products elsewhere.

CUSTOMERNA…	QUANT…	PRICEE…	SAL…	ORDERDAT…	DAYS_SINC…	STATUS (Sorted …	PRODUCTLINE (Unique concatenate …	MSRP (…	PH…	AD…	CITY (F…	POSTA…	COUNT.
Dragon Souveniers…	35.442	113.106	4,023.016	43	649	[Shipped,Shipped,Shi…	Classic Cars(17), Trucks and Buses(10), Vi…	113.442	+65 221…	Bronz S…	Singapore	79903	Singapore
Blauer See Auto, Co.	36.864	108.031	3,871.436	22	705	[Shipped,Shipped,Shi…	Classic Cars(13), Vintage Cars(5), Trucks …	105.818	+49 69 …	Lyoners…	Frankfurt	60528	Germany
Cruz & Sons Co.	36.962	96.08	3,615.99	26	971	[Shipped,Shipped,Shi…	Classic Cars(13), Planes(5), Motorcycles(7…	97.962	+63 2 5…	15 McCa…	Makati City	1227 MM	Philippines
Toys4GrownUps.com	35.333	97.225	3,485.399	30	649	[Resolved,Resolved,…	Motorcycles(13), Classic Cars(6), Vintage …	97.267	626555…	78934 H…	Pasadena	90003	USA
Stylish Desk Decors…	36.038	96.993	3,415.558	26	702	[Shipped,Shipped,Shi…	Classic Cars(5), Trucks and Buses(8), Plan…	96.038	(171) 55…	35 King …	London	WX3 6FW	UK
Enaco Distributors	38.348	88.783	3,409.211	23	659	[Shipped,Shipped,Shi…	Classic Cars(1), Ships(9), Vintage Cars(6),…	87.087	(93) 203…	Rambla …	Barcelona	8022	Spain

Table 5: Customers in verge of churning

- **Marketing Strategy:** Win them back with re-engagement campaigns, special discounts, and personalized incentives to encourage them to return.

3. **Who are your lost customers?**

These are customers who have low recency, frequency, and monetary metrics, or low recency and medium to low frequency and monetary metrics.

CUSTOMERNAME	QUANT…	PRICEE…	SAL…	ORDER…	DAYS_…	STATUS (Sor…	PRODUCTLINE (Unique co…	MSRP (M…	PHONE …	ADDRESSLINE…	CITY (F…	POSTA…	COUNT…
Iberia Gift Imports, Corp.	39.267	93.283	3,648.241	15	904	[Shipped,Shipped…	Trucks and Buses(7), Classic C…	93.133	(95) 555 82 82	C/ Romero, 33	Sevilla	41101	Spain
Clover Collections, Co.	30.625	112.87	3,609.777	16	659	[Shipped,Shipped…	Classic Cars(6), Planes(4), Mo…	106.875	+353 1862 …	25 Maiden Lane	Dublin	2	Ireland
Alpha Cognac	34.35	101.16	3,524.422	20	675	[Shipped,Shipped…	Classic Cars(4), Planes(6), Shi…	97.15	61.77.6555	1 rue Alsace-Lorraine	Toulouse	31000	France
Mini Auto Werke	35.467	98.083	3,484.26	15	717	[Resolved,Resolv…	Classic Cars(6), Trucks and Bu…	103	7675-3555	Kirchgasse 6	Graz	8010	Austria
Signal Collectibles Ltd.	34.267	95.396	3,347.901	15	836	[Shipped,Shipped…	Trucks and Buses(2), Vintage …	90.933	4155554312	2793 Furth Circle	Brisbane	94217	USA
Auto-Moto Classics Inc.	35.875	92.8	3,309.908	8	1353	[Shipped,Shipped…	Ships(2), Vintage Cars(3), Pla…	87.375	6175558428	16780 Pompton St.	Brickhaven	58339	USA
Norway Gifts By Mail, Co.	32.792	97.954	3,301.01	24	825	[Shipped,Shipped…	Classic Cars(11), Planes(10), …	97.25	+47 2212 1…	Drammensveien 12…	Oslo	N 0106	Norway
Gift Ideas Corp.	35.053	87.6	3,015.496	19	947	[Shipped,Shipped…	Planes(12), Motorcycles(2), Cl…	86.526	2035554407	2440 Pompton St.	Glendale	97561	USA
Double Decker Gift Stores…	29.75	99.108	3,001.587	12	670	[Shipped,Shipped…	Classic Cars(3), Planes(3), Shi…	93.25	(171) 555-7…	120 Hanover Sq.	London	WA1 1DP	UK
Marseille Mini Autos	32.16	92.397	2,997.446	25	757	[Shipped,Shipped…	Classic Cars(17), Vintage Cars…	99.76	91.24.4555	12, rue des Bouchers	Marseille	13008	France
Signal Gift Stores	32.034	91.429	2,853.486	29	657	[Shipped,Shipped…	Classic Cars(20), Vintage Cars(9)	96.724	7025551838	8489 Strong St.	Las Vegas	83030	USA
Australian Collectables, Ltd	30.652	90.042	2,808.324	23	1018	[Disputed,Disput…	Vintage Cars(17), Classic Cars…	88.13	61-9-3844-6…	7 Allen Street	Glen Waverly	3150	Australia
Bavarian Collectables Imp…	28.643	84.289	2,499.566	14	801	[Shipped,Shipped…	Planes(8), Ships(2), Vintage C…	82.714	+49 89 61 0…	Hansastr. 15	Munich	80686	Germany

Table 6: Lost customers

- **Suggestion:**
    - Investing heavily to win back lost customers is not advisable.
    - However, since the RFM values between the lost customers and the customers about to churn are very comparable, it seems that the range of these values are very close.
    - Hence, a little bit of effort such as social media and email retargeting, and informing customers about new product launches, that won't cost too much can be done.

4. **Who are your loyal customers?**

Loyal customers are those who have high frequency and monetary metrics, but might not be the most recent buyers.

It’s important to nurture the relationship to maintain their loyalty and potentially convert them into champions.

Row ID	CUSTOMERNAME	QUANT…	PRICEE…	SALES (…	ORD…	DAYS…	STATUS (Sorted list)	PRODUCTLINE (Unique concatenat…	MSRP (…	PHO…	AD…	CITY (F…	PO…	COUNT…
Row55	Muscle Machine Inc	36.979	111.151	4,119.52	48	502	[Shipped,Shipped,Shipped,…]	Classic Cars(34), Trucks and Buses(7), …	108.396	2125557…	4092 Fu…	NYC	10022	USA
Row72	Souveniers And Things Co.	34.804	95.189	3,295.021	46	186	[In Process,In Process,In Pr…	Motorcycles(2), Classic Cars(10), Truck…	93.087	+61 2 94…	Monitor …	Chatswood	2067	Australia
Row68	Saveley & Henriot, Co.	34.829	100.548	3,484.738	41	586	[Shipped,Shipped,Shipped,…]	Classic Cars(10), Trucks and Buses(7), …	102.024	78.32.5555	2, rue d…	Lyon	69004	France
Row41	L'ordine Souveniers	32.821	111.147	3,656.444	39	493	[Shipped,Shipped,Shipped,…]	Classic Cars(22), Planes(10), Motorcycl…	107.795	0522-556…	Strada …	Reggio Emilia	42100	Italy
Row69	Scandinavian Gift Ideas	35.763	97.597	3,533.14	38	262	[Cancelled,Cancelled,Cancel…	Classic Cars(9), Trucks and Buses(7), Pl…	99.184	0695-34 …	?kergat…	Boras	S-844 67	Sweden
Row26	Danish Wholesale Imports	36.528	108.038	4,028.933	36	499	[Disputed,Disputed,Dispute…	Classic Cars(14), Vintage Cars(6), Plan…	106.417	31 12 3555	Vinb'ltet…	Kobenhavn	1734	Denmark

Table 7: Loyal customers

- **Suggestions**:
    - **Exclusive Discounts and Offers**: Provide special discounts, offers, or cashback rewards exclusive to loyal customers.
    - **Points-Based System**: Implement a points-based system where customers earn points for purchases, which can be redeemed for discounts or free products.
    - **Personalized Emails**: Send personalized emails with product recommendations based on their purchase history.
    - **Anniversary and Birthday Offers**: Celebrate customer anniversaries or birthdays with special discounts or gifts.
    - **Early Access to Sales**: Give loyal customers early access to sales, new product launches, or special events.
    - **Invite-Only Events**: Organize exclusive events, webinars, or in-store experiences for your loyal customers.
    - **Premium Customer Support**: Offer premium customer support or dedicated account managers for loyal customers.
    - **Simplified Purchase Process**: Streamline the purchasing process, offer faster checkout options, or provide free shipping.

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.

Data Description

Sample of the Dataset:

	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

Table 8: Dataset Sample Part B

Data Dictionary

The detailed data dictionary description is given below:

1. **Date:** It denotes the date on which the order was placed.
2. **Order_id: It** represents the unique identification number assigned to each order.
3. **Product:** The item bought by the customer

Data Types

```
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)
```

Figure 21: Data type Part B

About data:

- 3 columns - 2 object type, 1 integer type
- 20640 rows
- No null values

Statistical Summary of the Dataset

	Order_id
count	20641.000000
mean	575.986289
std	328.557078
min	1.000000
25%	292.000000
50%	581.000000
75%	862.000000
max	1139.000000

Table 9: Statistical summary Part B

- The data is not scaled.

Exploratory Data Analysis

- **Number of distinct orders number over the years**

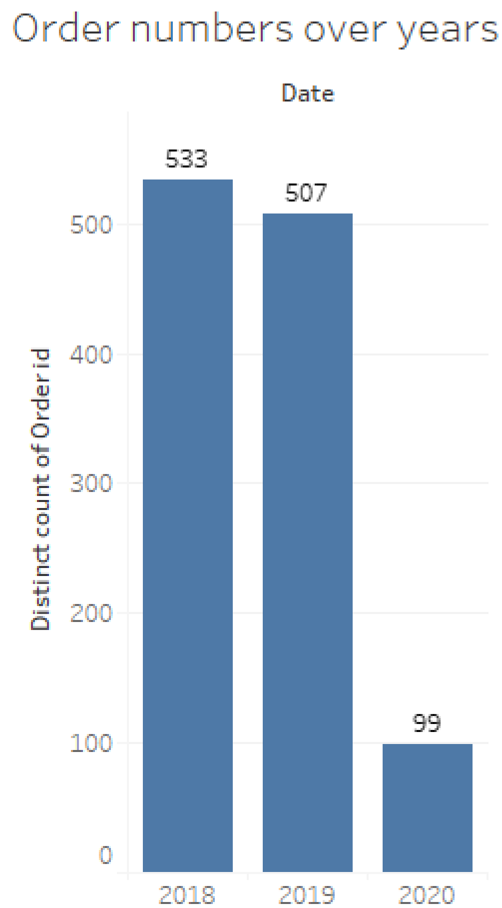


Figure 22: Order numbers over the years

- The number of orders were maximum during 2018. However,there's not much of a difference between 2018 and 2019.
- The year 2020 has records only till February, which is why the order count is low

- **Order Trend Over Years**

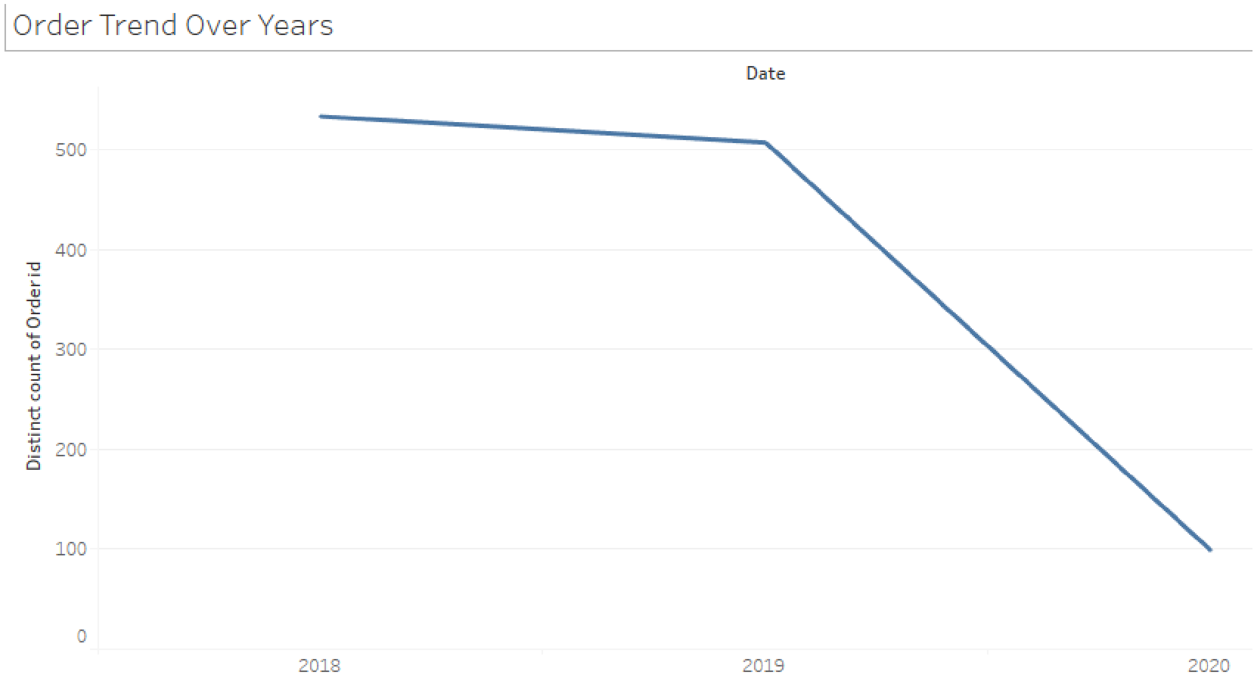


Figure 23: Order trend over years

- **Quarterly trend of orders**

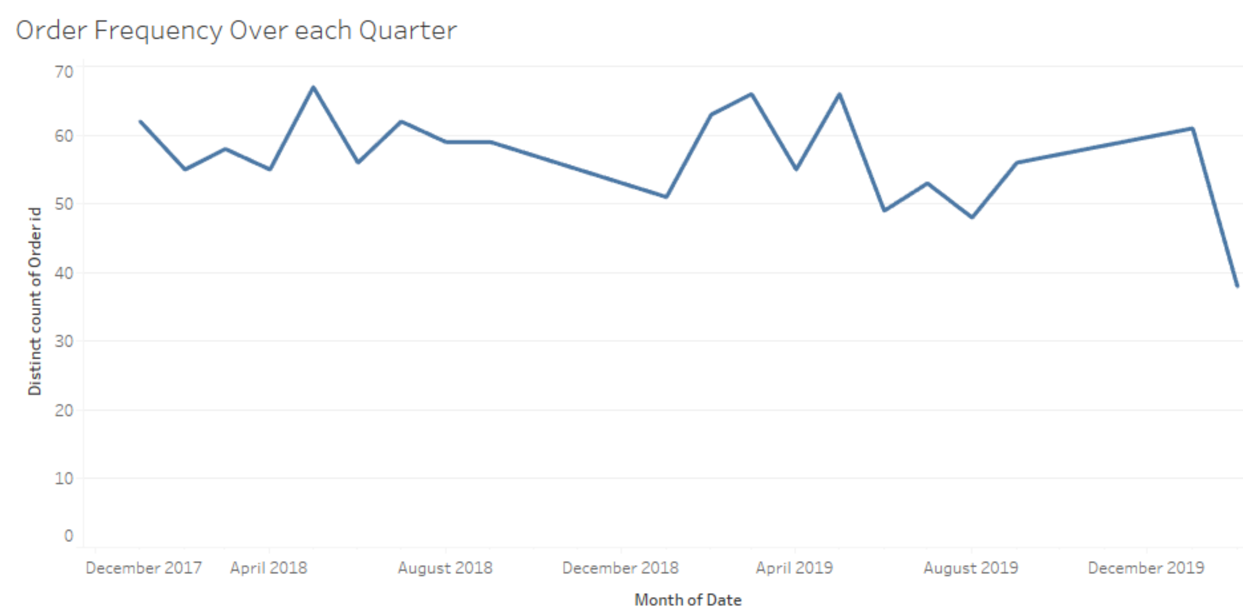


Figure 24: Quarterly trend of orders

- We can see a seasonality increase in the number of orders in both 2018 and 2019 of March. However, since we only have limited data, we cannot say for sure that this is a pattern.

- **Monthly comparison of the number of orders**

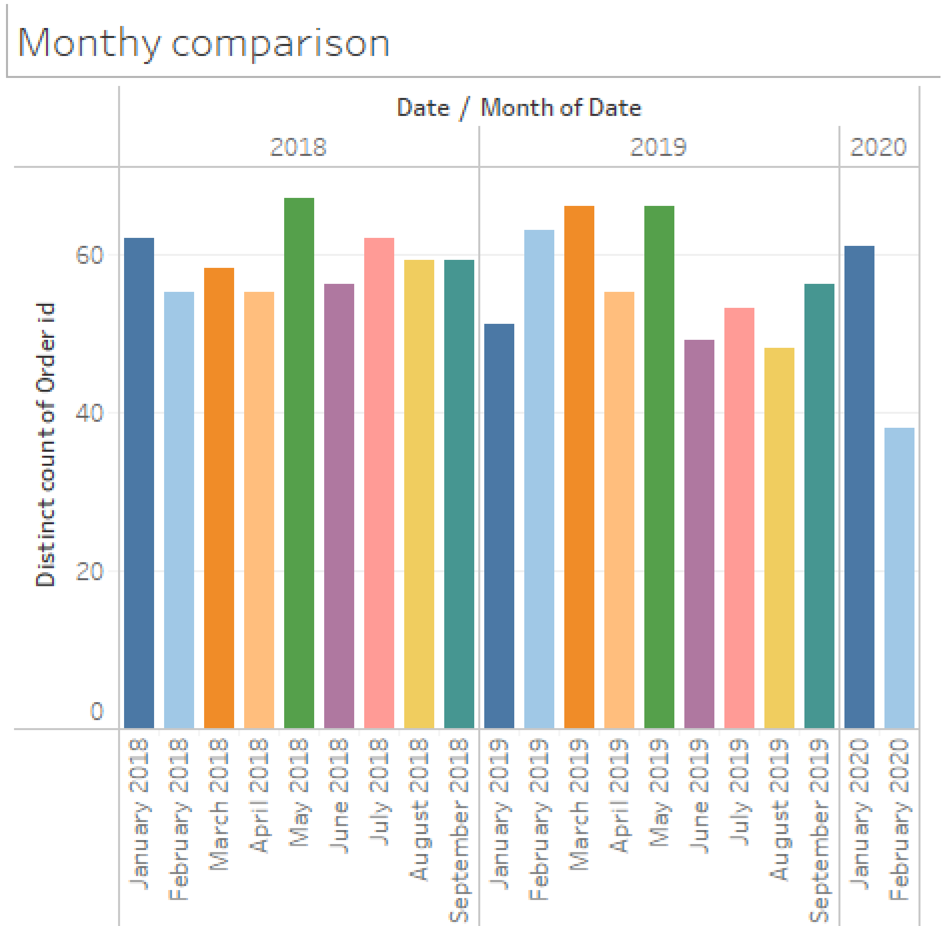


Figure 25: Monthly comparison of number of orders

- **Order Patterns Day Wise**

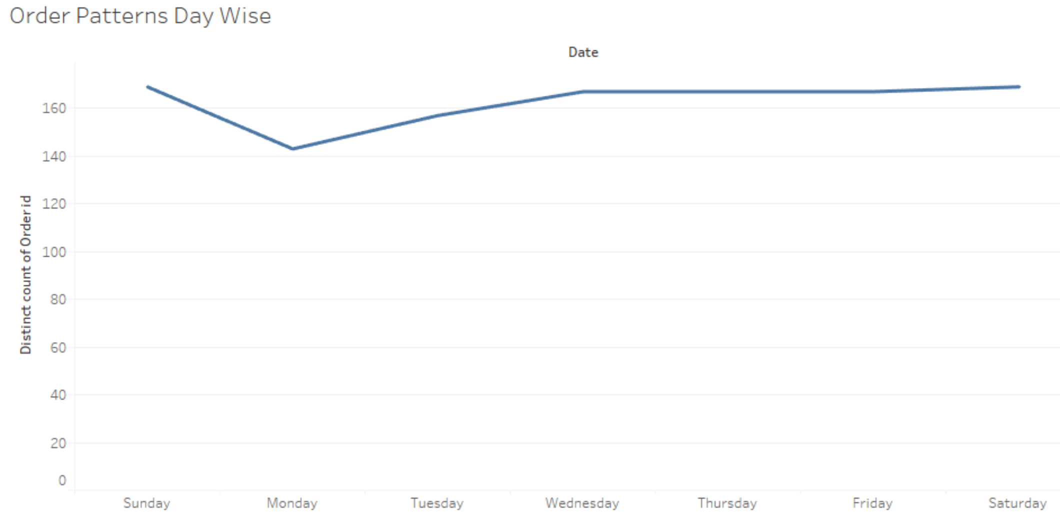
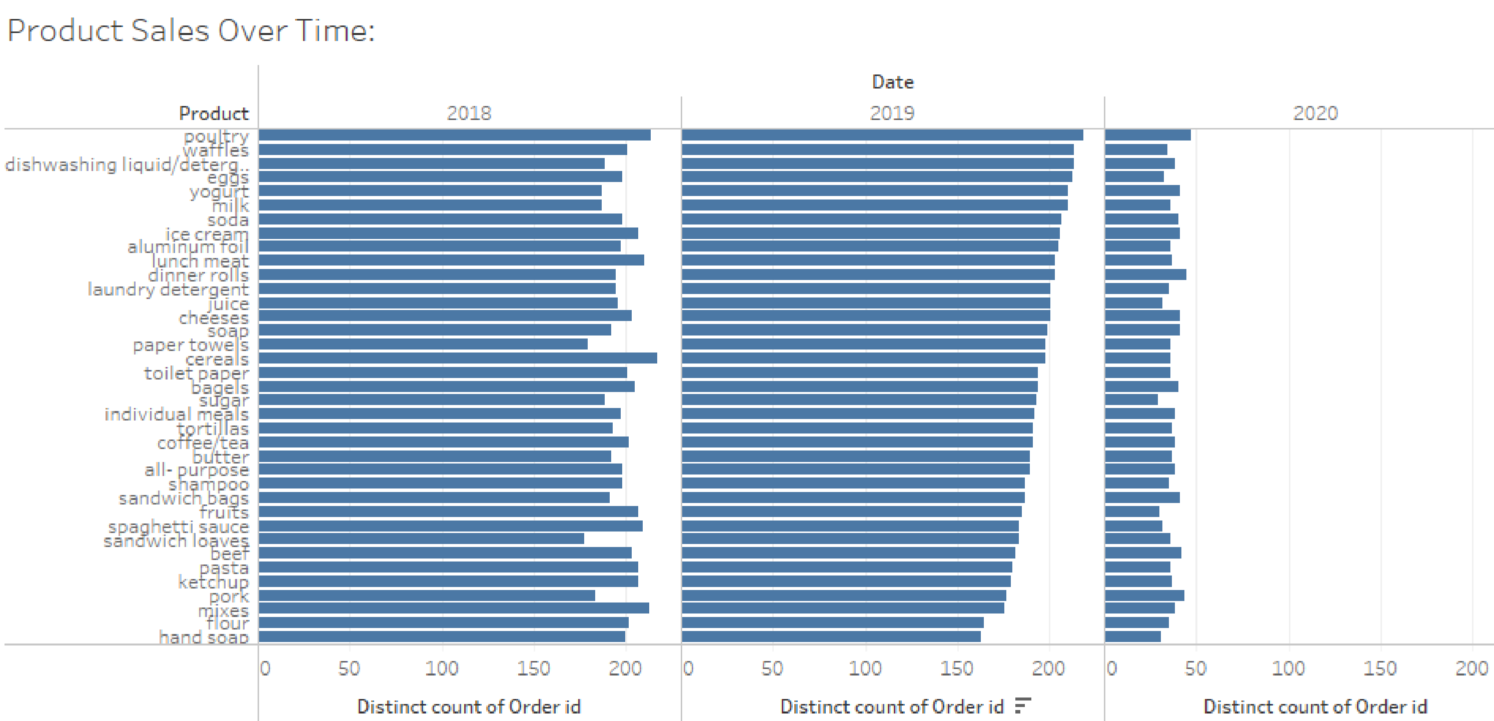


Figure 26: Order patterns day wise

- The number of orders placed is low on Mondays compared to other days of the week

- **Product Sales Over Time:**



•

Figure 27: Product sales over time

- There is no significant change in the demand for the products between 2018 and 2019.

Market Basket Analysis (Association Rules)

Association rules are a fundamental technique used in data mining to uncover interesting relationships, patterns, or associations among a set of items in large datasets. These rules help in understanding how the presence of one item in a transaction relates to the presence of other items. The primary measures used to evaluate these rules are support, confidence, and lift.

1. **Support:** This measure indicates how frequently an itemset appears in the dataset. It is calculated as the proportion of transactions in which the itemset occurs. High support means that the itemset is common across many transactions.
2. **Confidence:** This measure assesses the likelihood that an item B is purchased when item A is purchased. It is the conditional probability of item B given item A. High confidence implies a strong association between the items.
3. **Lift:** This measure evaluates the strength of an association rule by comparing the observed support to the expected support if the items were independent. Lift values greater than 1 indicate a positive association, while values less than 1 suggest a negative association.

Key Benefits of Association Rules in Market Basket Analysis for the given dataset:

1. **Product Placement and Store Layout:**
    - By identifying frequently co-purchased items, supermarkets can optimize store layouts to place these items near each other. This convenience can enhance the shopping experience and potentially increase sales.
    - Example: If association rules reveal that customers who buy bread also frequently purchase butter, placing these items in proximity can encourage additional purchases.
2. **Inventory Management:**
    - Understanding associations between products helps in predicting demand more accurately. This ensures that frequently co-purchased items are stocked adequately, reducing the likelihood of stockouts and improving customer satisfaction.
    - Example: If a supermarket finds a strong association between cereal and milk, they can ensure synchronized stocking to avoid scenarios where customers find one item but not the other.
3. **Targeted Marketing and Promotions:**
    - Association rules enable supermarkets to design targeted marketing campaigns and promotions. By identifying product bundles that customers are likely to buy together, supermarkets can offer discounts or promotions on these bundles to increase sales.
    - Example: If customers often purchase pasta and pasta sauce together, the supermarket can offer a discount on purchasing both items as a bundle, encouraging more sales.

4. **Personalized Recommendations:**
    ○ Using association rules, supermarkets can provide personalized product recommendations to customers based on their purchase history. This can enhance the customer shopping experience and increase the likelihood of repeat purchases.
    ○ Example: If a customer frequently buys a particular brand of coffee, the supermarket can recommend complementary products like coffee filters or creamers.

KNIME Workflow

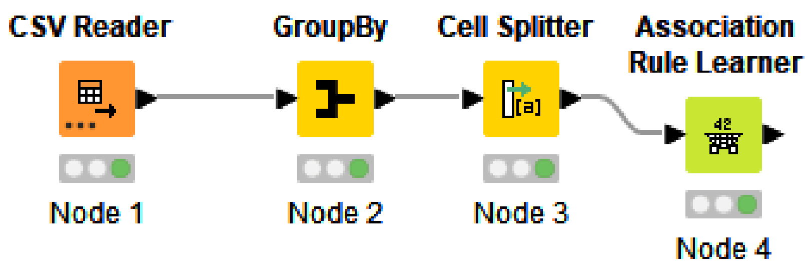


Figure 28: KNIME workflow for Market Basket Analysis

Threshold values of Support and Confidence

- ☐ For the given dataset the minimum support value is 0.05, i.e. 5%. This implies the items in the basket must appear in atleast 5% of the orders.

- ☐ The minimum confidence level for the given data is set to 0.6. This is a medium level of confidence. A conditional probability of 60% that items in basket B is bought when basket A products are bought.

Associations Identified

The following table shows the basket associations.

For example:

- If a customer buys the items [Dinner rolls, Spaghetti sauce, and Ice cream] then there is a 68.6% probability that they will alos buy poultry.
    - The lift value is 1.628, which is greater than 1. This indicates a positive association.
    - I.e. customers who buy [Dinner rolls, Spaghetti sauce, and Ice cream] are 1.6 times (or 60%) more likely to also buy Poultry compared to the probability of buying Poultry independently of buying [Dinner rolls, Spaghetti sauce, and Ice cream].
    - We can notice that [Dinner rolls, Spaghetti sauce, and Laundry Detergent] - - > [Poultry] has 65% confidence with 1.55 lift.
    - Therefore, [Dinner rolls, Spaghetti sauce, and Ice cream] - - -> [Poultry] is better combination than [Dinner rolls, Spaghetti sauce, and Laundry Detergent] - - > [Poultry]

D Support	D Con...	D Lift	S Conseq...	S implies	[...] Items
0.052	0.686	1.628	poultry	<---	[dinner rolls,spaghetti sauce,ice cream]
0.051	0.674	1.726	cheeses	<---	[bagels,cereals,sandwich bags]
0.054	0.656	1.556	poultry	<---	[dinner rolls,spaghetti sauce,laundry detergent]
0.055	0.649	1.791	paper towels	<---	[eggs,ice cream,pasta]
0.055	0.643	1.731	pasta	<---	[paper towels,eggs,ice cream]
0.054	0.642	1.651	dinner rolls	<---	[spaghetti sauce,poultry,laundry detergent]
0.052	0.641	1.649	dinner rolls	<---	[spaghetti sauce,poultry,ice cream]
0.05	0.64	1.7	juice	<---	[yogurt,toilet paper,aluminum foil]
0.051	0.637	1.512	poultry	<---	[dinner rolls,spaghetti sauce,cereals]
0.052	0.634	1.627	eggs	<---	[paper towels,dinner rolls,pasta]
0.051	0.63	1.678	mixes	<---	[yogurt,poultry,aluminum foil]
0.051	0.63	1.621	dinner rolls	<---	[spaghetti sauce,poultry,cereals]
0.055	0.63	1.616	eggs	<---	[paper towels,ice cream,pasta]
0.052	0.628	1.61	eggs	<---	[dinner rolls,poultry,soda]
0.052	0.628	1.614	dinner rolls	<---	[spaghetti sauce,poultry,juice]
0.055	0.624	1.565	ice cream	<---	[paper towels,eggs,pasta]
0.05	0.62	1.645	juice	<---	[yogurt,poultry,aluminum foil]
0.051	0.617	1.558	cereals	<---	[cheeses,bagels,sandwich bags]
0.05	0.613	1.616	coffee/tea	<---	[yogurt,cheeses,cereals]
0.051	0.611	1.66	sandwich bags	<---	[cheeses,bagels,cereals]
0.051	0.604	1.589	milk	<---	[poultry,laundry detergent,cereals]
0.052	0.602	1.429	poultry	<---	[dinner rolls,spaghetti sauce,juice]
0.052	0.602	1.621	pasta	<---	[paper towels,eggs,dinner rolls]
0.05	0.6	1.424	poultry	<---	[dishwashing liquid/detergent,laundry detergent,mixes]

Table 10: Identified associations

Possible Combos with Lucrative Offers

The following are some of the best associations with about 64% confidence and higher lift values.

- Dinner rolls, Spaghetti sauce, and Ice cream - - -> Poultry
- Bagel, Cereals, Sandwich bags - - -> Cheese
- Eggs, Ice cream, Pasta - - -> Paper towels
- Yoghurt, Toilet paper, Aluminium foil - - -> Juice
- Spaghetti sauce, Poultry, Laundry detergent - - -> Dinner rolls

Suggestions:

- Optimize store layouts to place these items near each other. This convenience can enhance the shopping experience and potentially increase sales.
- Offer discounts or promotions such as:
    - Buy any 3 and get a discount on poultry to increase sales.
    - “Buy any bagel and cereal, get a discount on cheese" or "Free sandwich bags with the purchase of cheese."
- Provide personalized product recommendations to customers based on their purchase history.
- Position eggs, ice cream, pasta, and paper towels in adjacent aisles or sections. This can remind customers to purchase paper towels when buying these food items.