MRA Project

Anirudh Sardiwal
Aug. 18, 2024

Great Learning | PGP DSBA | O.OCT23.A

Back

Table of Contents

1.0 Part A
Auto Sales

Problem Statement

An automobile parts manufacturing company has collected data on transactions for 3 years. They do not have any in-house data science team, thus they have hired you as their consultant. Your job is to use your data science skills to find the underlying buying patterns of the customers, provide the company with suitable insights about their customers, and recommend customized marketing strategies for different segments of customers.

Back

1.1 EDA

The data has 2747 rows and 20 Columns.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2747 entries, 0 to 2746
Data columns (total 20 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   ORDERNUMBER           2747 non-null   int64
 1   QUANTITYORDERED       2747 non-null   int64
 2   PRICEEACH             2747 non-null   float64
 3   ORDERLINENUMBER       2747 non-null   int64
 4   SALES                 2747 non-null   float64
 5   ORDERDATE             2747 non-null   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)
```

1.1 - Dataframe general information

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	DAYS_SINCE_LASTORDER	MSRP
count	2747.0	2747.0	2747.0	2747.0	2747.0	2747	2747.0	2747.0
mean	10260.0	35.0	101.0	6.0	3553.0	2019-05-13 21:56:17.211503360	1757.0	101.0
min	10100.0	6.0	27.0	1.0	482.0	2018-01-06 00:00:00	42.0	33.0
25%	10181.0	27.0	69.0	3.0	2204.0	2018-11-08 00:00:00	1077.0	68.0
50%	10264.0	35.0	96.0	6.0	3185.0	2019-06-24 00:00:00	1761.0	99.0
75%	10334.0	43.0	127.0	9.0	4503.0	2019-11-17 00:00:00	2436.0	124.0
max	10425.0	97.0	253.0	18.0	14083.0	2020-05-31 00:00:00	3562.0	214.0
std	92.0	10.0	42.0	4.0	1839.0	NaN	819.0	40.0

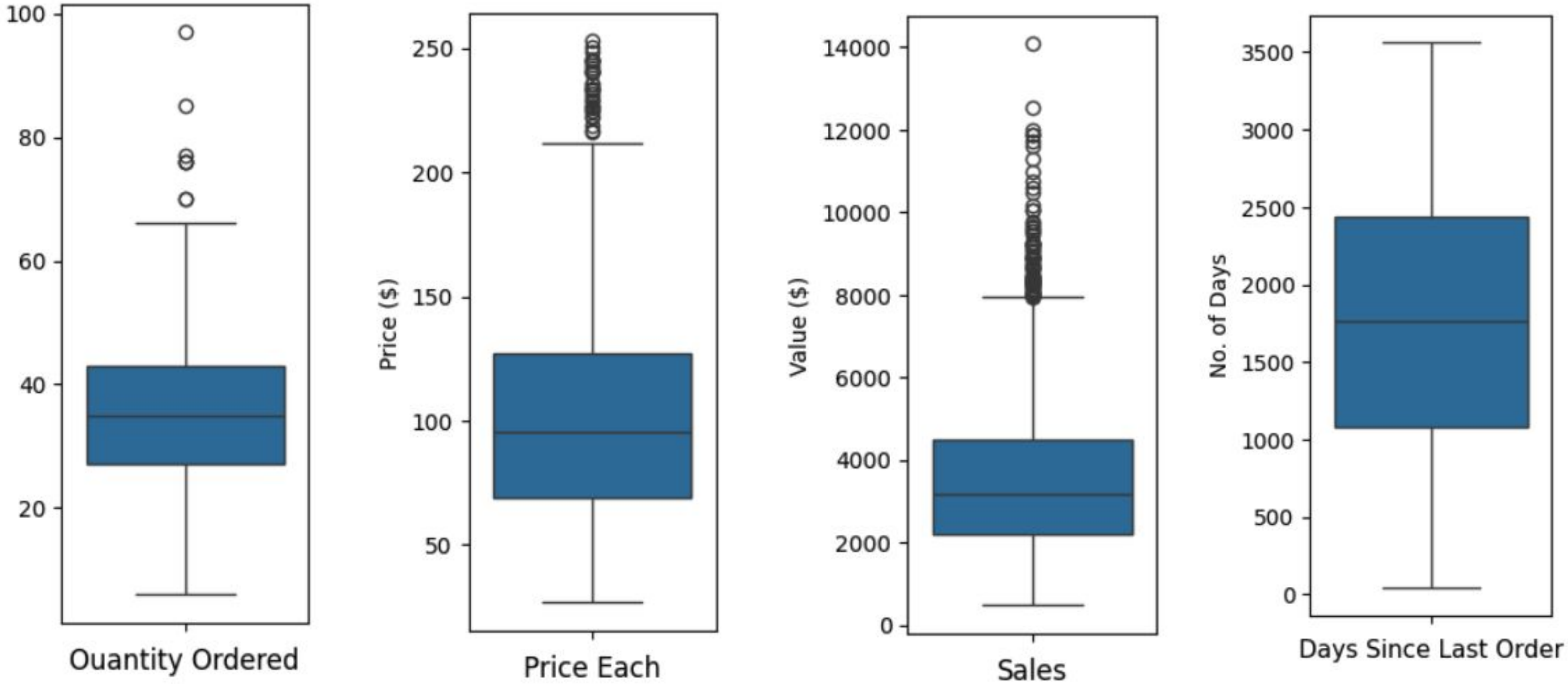
1.2 - Statistics of Numeric features

Back

ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	DAYS_SINCE_LASTORDER	STATUS	PRODUCTLINE	MSRP	PRODUCTCODE	CUSTOMERNAME	PHONE	ADDRESSLINE1	CITY	POSTALCODE	COUNTRY	CONTACTLASTNAME	CONTACTFI STNAME
10107	30	95.70	2	2871	2/24/2018	828	Shipped	Motorcycles	95	S10_1678	Land of Toys Inc.	2125557818	897 Long Airport Avenue	NYC	10022	USA	Yu	Kwai
10121	34	81.35	5	2765.9	5/7/2018	757	Shipped	Motorcycles	95	S10_1678	Reims Collectables	26.47.1555	59 rue de l'Abbaye	Reims	51100	France	Henriot	Paul
10134	41	94.74	2	3884.34	7/1/2018	703	Shipped	Motorcycles	95	S10_1678	Lyon Souveniers	+33 1 46 62 7555	27 rue du Colonel Pierre Avia	Paris	75508	France	Da Cunha	Daniel
10145	45	83.26	6	3746.7	8/25/2018	649	Shipped	Motorcycles	95	S10_1678	Toys4GrownUps.com	6265557265	78934 Hillside Dr.	Pasadena	90003	USA	Young	Julie
10168	36	96.66	1	3479.76	10/28/2018	586	Shipped	Motorcycles	95	S10_1678	Technics Stores Inc.	6505556809	9408 Furth Circle	Burlingame	94217	USA	Hirano	Juri
10180	29	86.13	9	2497.77	11/11/2018	573	Shipped	Motorcycles	95	S10_1678	Daedalus Designs Imports	20.16.1555	184, chausse de Tournai	Lille	59000	France	Rance	Martine
10188	48	114.84	1	5512.32	11/18/2018	567	Shipped	Motorcycles	95	S10_1678	Herkku Gifts	+47 2267 3215	Drammen 121, PR 744 Sentrum	Bergen	N 5804	Norway	Oeztan	Veysel
10211	41	114.84	14	4708.44	1/15/2019	510	Shipped	Motorcycles	95	S10_1678	Auto Canal Petit	(1) 47.55.6555	25, rue Lauriston	Paris	75016	France	Perrier	Dominique
10223	37	107.18	1	3965.66	2/20/2019	475	Shipped	Motorcycles	95	S10_1678	Australian Collectors, Co.	03 9520 4555	636 St Kilda Road	Melbourne	3004	Australia	Ferguson	Peter
10237	23	101.44	7	2333.12	4/5/2019	432	Shipped	Motorcycles	95	S10_1678	Vitachrome Inc.	2125551500	2678 Kingston Rd.	NYC	10022	USA	Frick	Michael
10251	28	113.88	2	3188.64	5/18/2019	390	Shipped	Motorcycles	95	S10_1678	Tekni Collectables Inc.	2015559350	7476 Moss Rd.	Newark	94019	USA	Brown	William
10263	34	108.14	2	3676.76	6/28/2019	350	Shipped	Motorcycles	95	S10_1678	Gift Depot Inc.	2035552570	25593 South Bay Ln.	Bridgewater	97562	USA	King	Julie

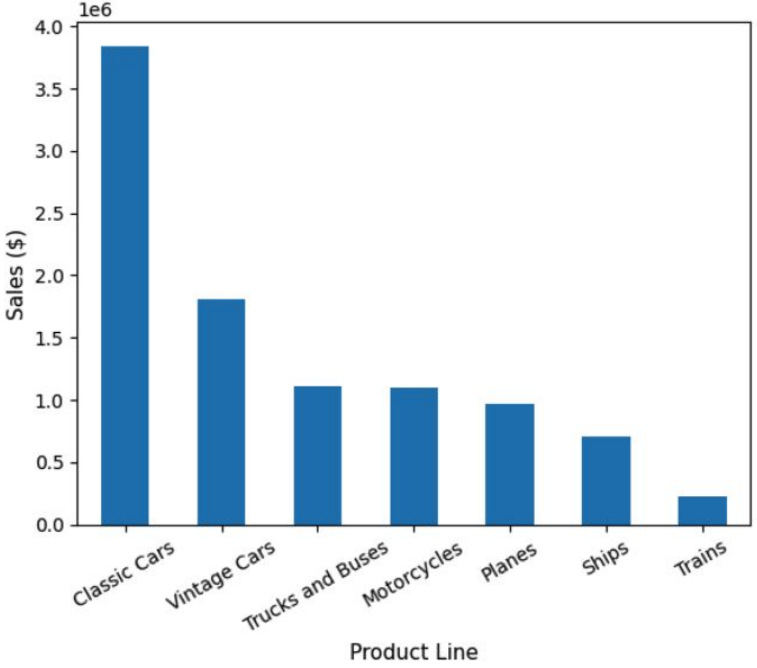
1.2 - Head of Data

Back

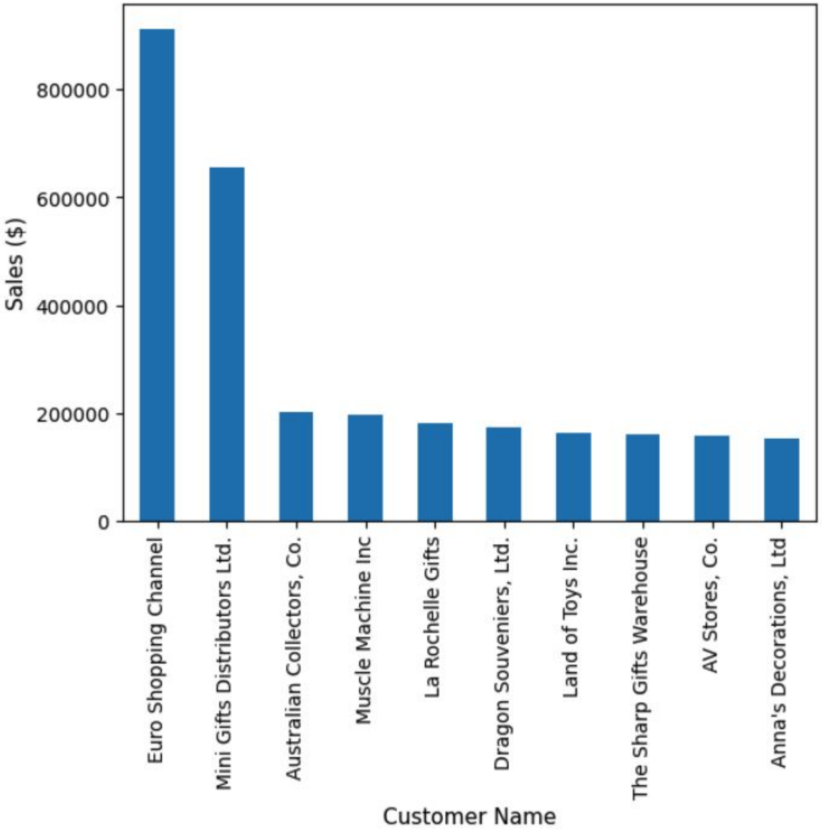


1.3 - Boxplots of Various features

Back

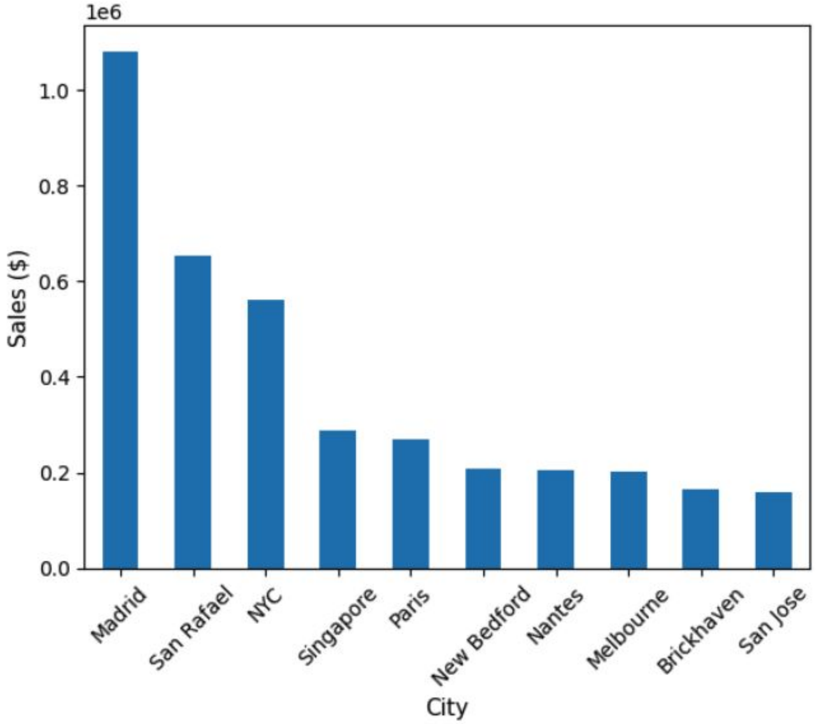


1.4 - Product Line items by Sales amount

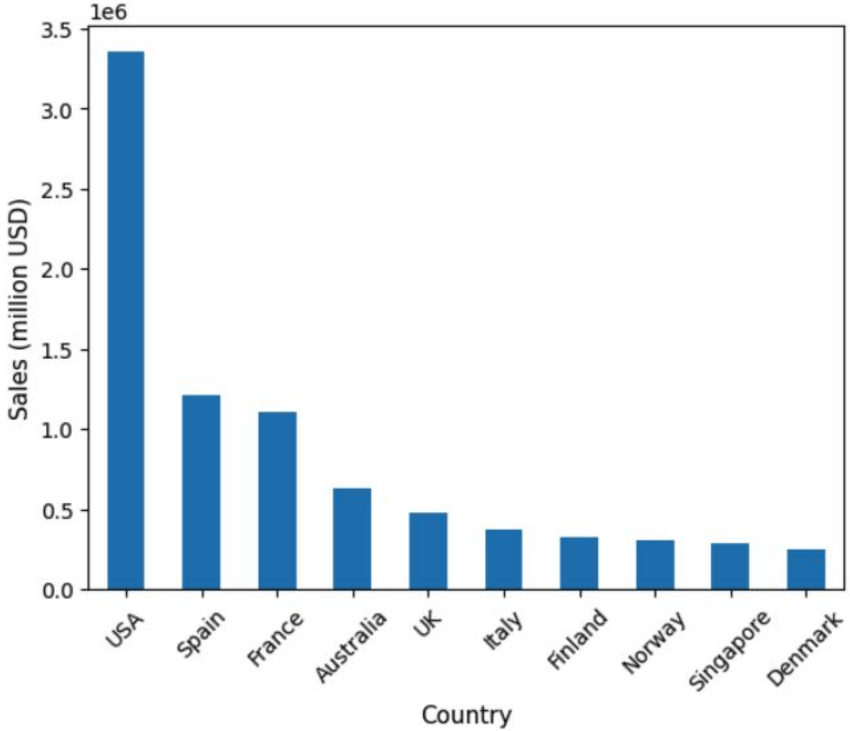


1.5 - Top Customers by Sales

Back

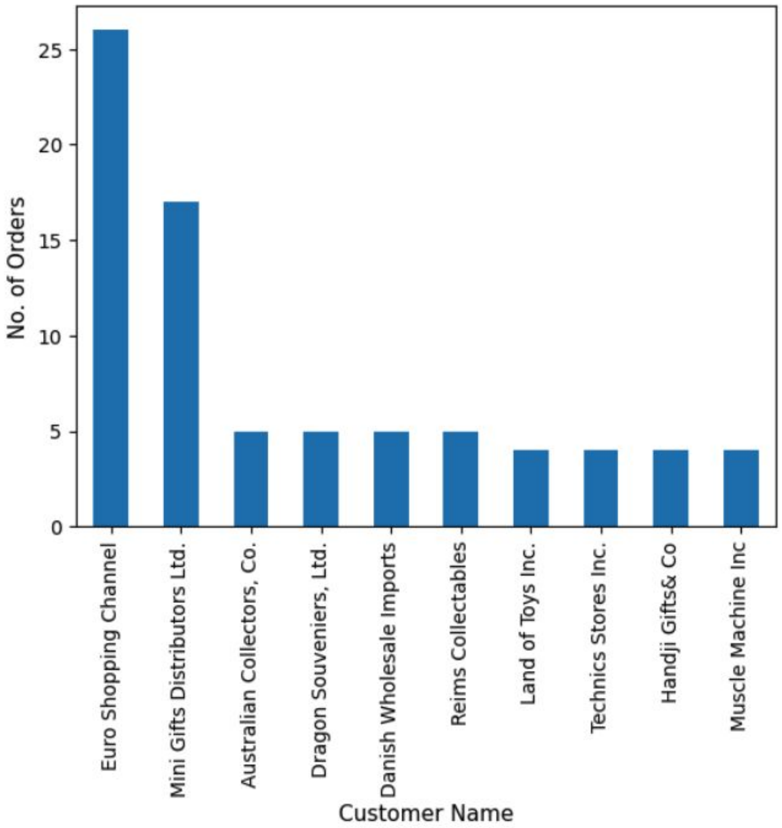


1.6 - Cities by Sales amount

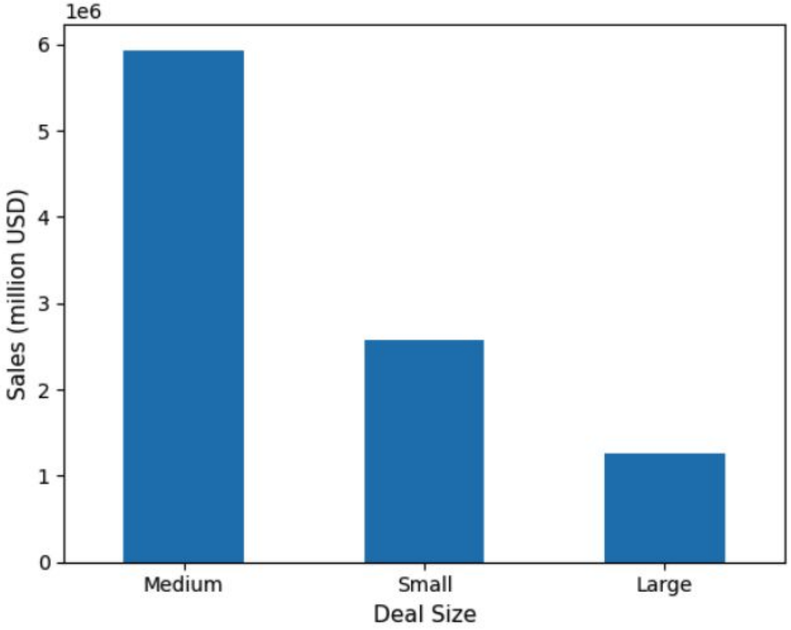


1.7 - Countries by Sales amount

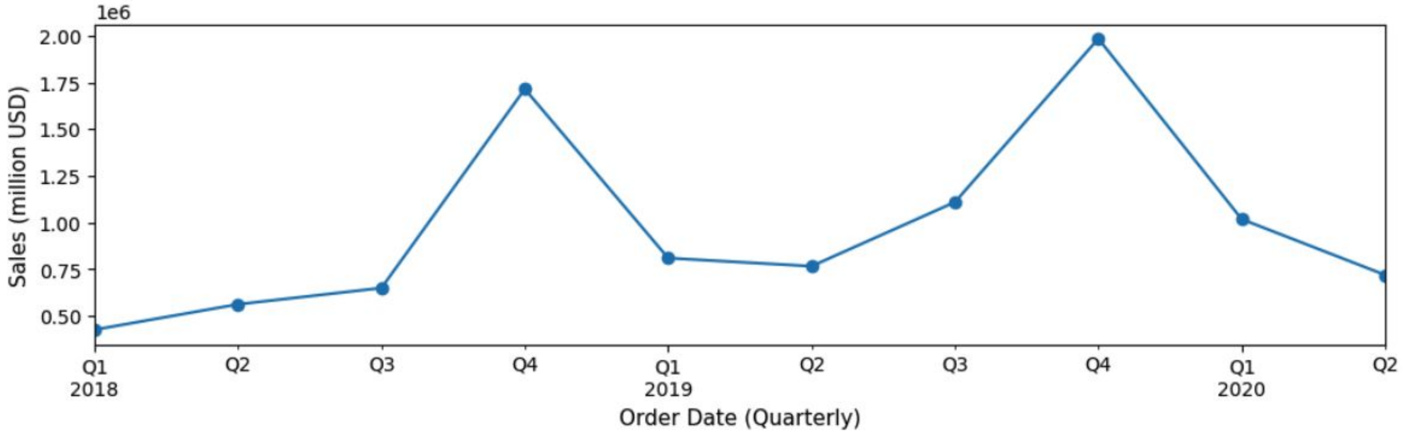
Back



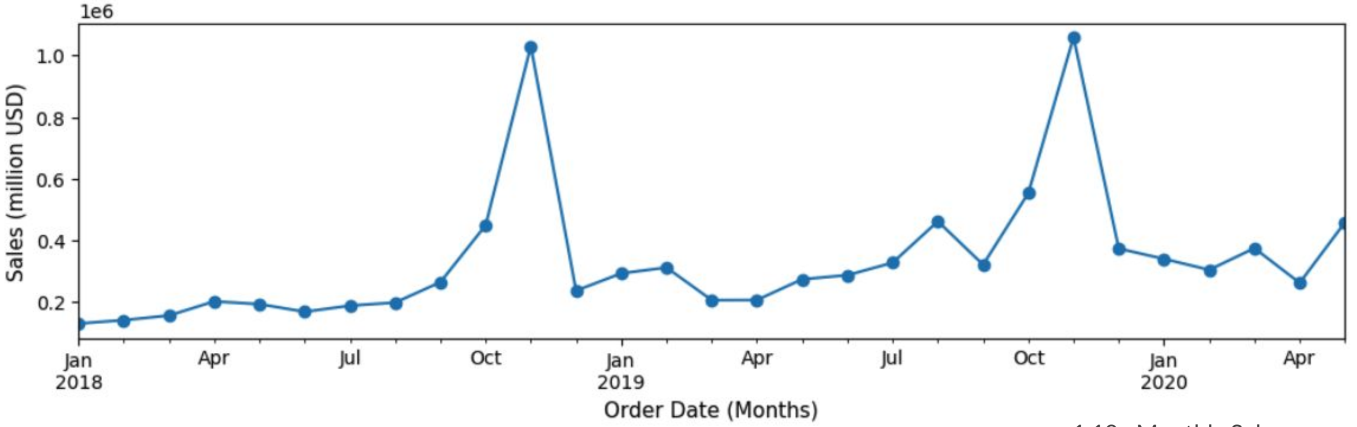
1.8 - Customers with maximum no. of orders



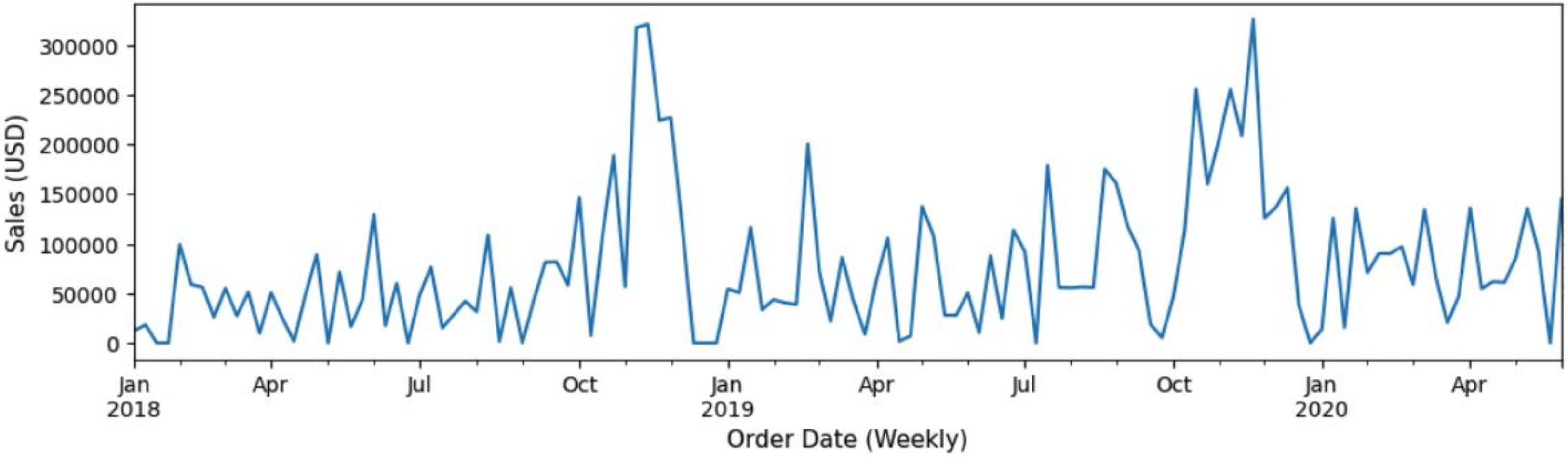
1.9 - Deal sizes by Sales amount



1.9 - Quarterly Sales



1.10 - Monthly Sales



1.11 - Weekly Sales

Back

1.2 Customer Segmentation

What is RFM?

RFM stands for Recency, Frequency, Monetary.

Recency - How recently has the customer purchased? (Unit: Days)

Frequency - How many times has a customer purchased in a given time period. (Unit: Count)

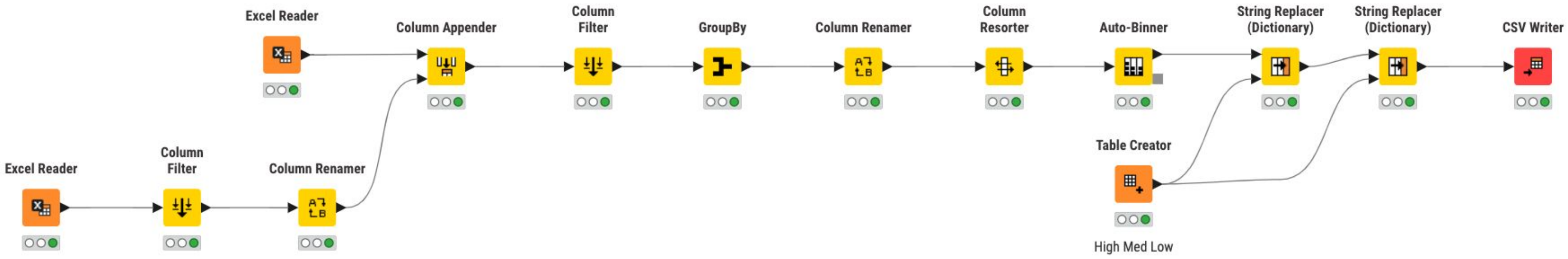
Monetary - What is the total amount of Sales generated by a customer over a given time period. (Unit: Currency)

It is a customer segmentation technique and helps in deciding the direction of Marketing that should be applied to various segments of customers.

Assumptions:

- Past behaviour is indicative of future behaviour.
- All 3 parameters have equal weight.
- Customer will continue to purchase in the future.
- Recent activity indicates interest in the brand.
- Customers who buy more frequently are more loyal and will continue to do so.
- High frequency indicates satisfaction and strong relationship with the brand.
- Customers who spend more are more valuable to the company.
- More spending means the customer finds significant value in the products or services.

Back



1.2.1 - Knime Workflow diagram

RowID	CUSTOM... String	QUANTIT... Number (dou...	PRICEEA... Number (dou...	ORDERDA... Local Date	PRODUC... String	CITY String	COUNTRY String	DEALSIZE String	Recency Number (inte...	Frequency Number (inte...	Monetary Number (dou...	Recency [... String	Frequenc... String	Monetary... String
Row0	AV Stores, Co.	34.863	91.085	2019-10-14	Classic Cars, ...	Manchester	UK	Medium, Med...	421	3	157,807.81	2 Med	2 Med	1 High
Row1	Alpha Cognac	34.35	101.16	2020-03-28	Classic Cars, ...	Toulouse	France	Medium, Med...	675	3	70,488.44	3 Low	2 Med	3 Low
Row2	Amica Model...	32.423	110.853	2019-09-09	Classic Cars, ...	Torino	Italy	Large, Large, ...	328	2	94,117.26	2 Med	3 Low	2 Med
Row3	Anna's Decor...	31.935	106.424	2018-11-04	Classic Cars, ...	North Sydney	Australia	Small, Small, ...	131	4	153,996.13	1 High	1 High	1 High
Row4	Atelier graphi...	38.571	92.239	2019-11-25	Motorcycles, ...	Nantes	France	Medium, Med...	312	3	24,179.96	2 Med	2 Med	3 Low
Row5	Australian Col...	30.652	90.042	2020-05-09	Vintage Cars, ...	Glen Waverly	Australia	Medium, Sma...	1018	3	64,591.46	3 Low	2 Med	3 Low
Row6	Australian Col...	35.018	104.59	2019-02-20	Motorcycles, ...	Melbourne	Australia	Medium, Med...	229	5	200,995.41	1 High	1 High	1 High
Row7	Australian Gif...	36.333	110.554	2018-09-25	Classic Cars, ...	South Brisbane	Australia	Large, Mediu...	190	3	59,469.12	1 High	2 Med	3 Low

1.2.1 - Output table head

Back

1.3 Inferences

Count of Customers		Monetary			
Recency	Frequency	1 High	2 Med	3 Low	Grand Total
1 High	1 High	11	1		12
	2 Med	1	9	1	11
1 High Total		12	10	1	23
2 Med	1 High	3	3		6
	2 Med	5	11	1	17
	3 Low		11	10	21
2 Med Total		8	25	11	44
3 Low	1 High	1	1	1	3
	2 Med	1	6	5	12
	3 Low		2	5	7
3 Low Total		2	9	11	22
Grand Total		**22**	**44**	**23**	**89**

1.3.1 - Pivot table of results

Legend

Best Customers - R: High, F: High, M: High

Loyal Customers - R: High, F: Med to High, M: Med to High

On Verge of Churning - R: Low, F: Med to Low, M: Med to Low

Lost Customers - R: Low, F: Low, M: Low

Back

Best Customers

CUSTOMERNAME	QUANTITYORDERED	PRICEEACH	ORDERDATE	PRODUCTLINE	CITY	COUNTRY	DEALSIZE	Recency values	Frequency values	Monetary values
Anna's Decorations, Ltd	32	106.42	2018-11-04	Classic Cars, Classi	North Sydney	Australia	Small, Small, Small	131	4	153996.13
Australian Collectors, Co.	35	104.59	2019-02-20	Motorcycles, Classic	Melbourne	Australia	Medium, Medium, S	229	5	200995.41
Diecast Classics Inc.	36	108.57	2019-11-02	Motorcycles, Motorc	Allentown	USA	Medium, Medium, L	228	4	122138.14
Euro Shopping Channel	36	97.38	2020-03-01	Motorcycles, Classic	Madrid	Spain	Large, Large, Medi	42	26	912294.11
La Rochelle Gifts	35	97.05	2019-10-29	Motorcycles, Motorc	Nantes	France	Medium, Small, Sm	139	4	180124.9

Loyal Customers

Auto Canal Petit	37	94.26	2019-05-26	Motorcycles, Motorc	Paris	France	Medium, Mediun	127	3	93170.66
Collectables For Less Inc.	33	97.24	2019-07-21	Classic Cars, Classi	Brickhaven	USA	Medium, Large,	179	3	81577.98
FunGiftIdeas.com	35	109.59	2020-03-03	Motorcycles, Motorc	New Bedford	USA	Medium, Small,	111	3	98923.73
Gift Depot Inc.	36	108.93	2019-06-28	Motorcycles, Motorc	Bridgewater	USA	Medium, Mediun	226	3	101894.79
Gifts4AllAges.com	36	91.56	2020-05-06	Classic Cars, Classi	Boston	USA	Medium, Small,	148	3	83209.88

On Verge of Churning

Cruz & Sons Co.	37	96.08	2018-11-27	Classic Cars, Cl	Makati City	Philippines	Medium, Mediun	971	3	94015.73
Enaco Distributors	38	88.78	2018-11-26	Classic Cars, Sh	Barcelona	Spain	Medium, Mediun	659	3	78411.86
Marseille Mini Autos	32	92.40	2020-01-06	Classic Cars, Cl	Marseille	France	Large, Medium,	757	3	74936.14
Signal Gift Stores	32	91.43	2019-11-29	Classic Cars, Vir	Las Vegas	USA	Medium, Small,	657	3	82751.08
Stylish Desk Decors, Co.	36	96.99	2018-06-12	Classic Cars, Cl	London	UK	Medium, Mediun	702	3	88804.5
Toys4GrownUps.com	35	97.22	2018-08-25	Motorcycles, Mo	Pasadena	USA	Medium, Mediun	649	3	104561.96

Lost Customers

Bavarian Collectables Imports, Co.	29	84.29	2019-09-15	Planes, Ships, Vin	Munich	Germany	Medium, Small,	801	1	34993.92
Clover Collections, Co.	31	112.87	2019-09-16	Classic Cars, Clas	Dublin	Ireland	Large, Medium,	659	2	57756.43
Double Decker Gift Stores, Ltd	30	99.11	2018-11-14	Classic Cars, Plan	London	UK	Medium, Mediun	670	2	36019.04
Iberia Gift Imports, Corp.	39	93.28	2018-11-14	Trucks and Buses,	Sevilla	Spain	Medium, Large,	904	2	54723.62
Signal Collectibles Ltd.	34	95.40	2019-02-10	Trucks and Buses,	Brisbane	USA	Medium, Mediun	836	2	50218.51

2.0 Part B
Grocery Store

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.

Back

2.1 EDA

The data has 20,641 rows and 3 Columns. It has no null values. There are 37 unique products in the data.

	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

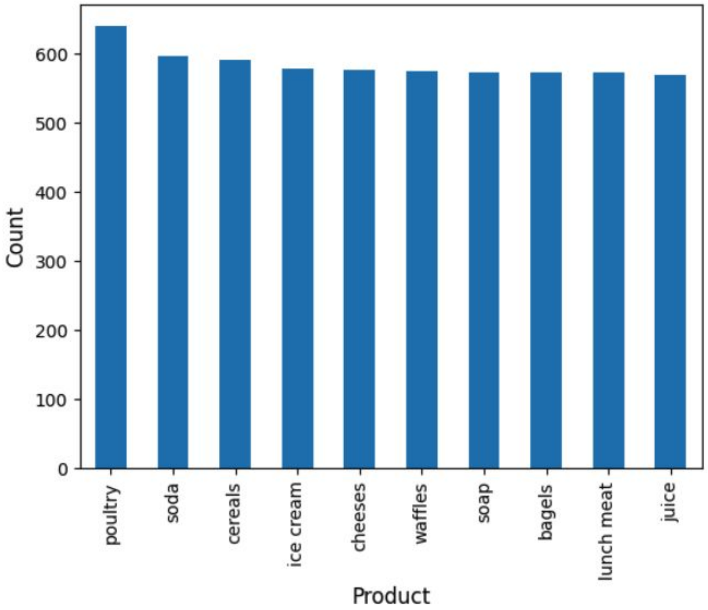
2.1.1 - Data Head

	Date	Order_id	Product
20636	2020-02-25	1138	soda
20637	2020-02-25	1138	paper towels
20638	2020-02-26	1139	soda
20639	2020-02-26	1139	laundry detergent
20640	2020-02-26	1139	shampoo

2.1.2 - Data Tail

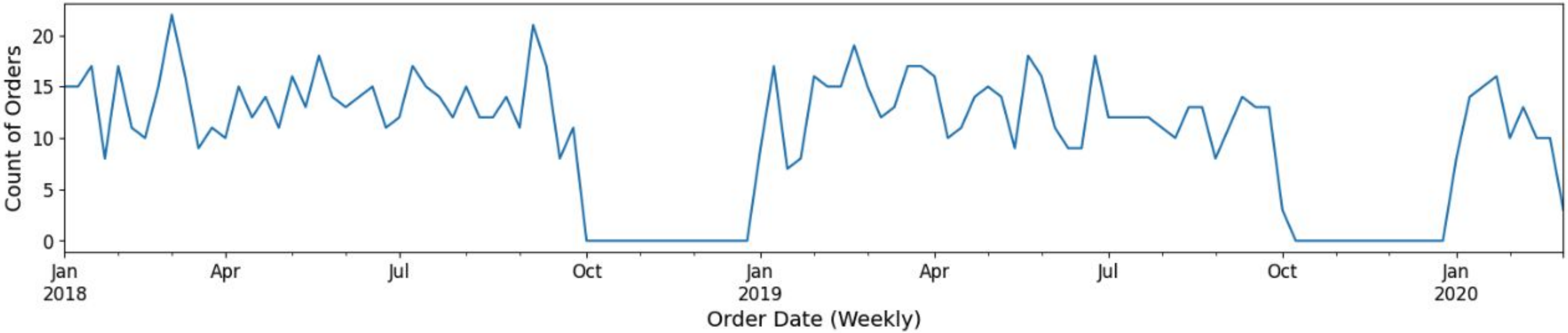
```
['yogurt', 'pork', 'sandwich bags', 'lunch meat', 'all- purpose',
 'flour', 'soda', 'butter', 'beef', 'aluminum foil', 'dinner rolls',
 'shampoo', 'mixes', 'soap', 'laundry detergent', 'ice cream',
 'toilet paper', 'hand soap', 'waffles', 'cheeses', 'milk',
 'dishwashing liquid/detergent', 'individual meals', 'cereals',
 'tortillas', 'spaghetti sauce', 'ketchup', 'sandwich loaves',
 'poultry', 'bagels', 'eggs', 'juice', 'pasta', 'paper towels',
 'coffee/tea', 'fruits', 'sugar'], dtype=object)
```

2.1.3 - All products

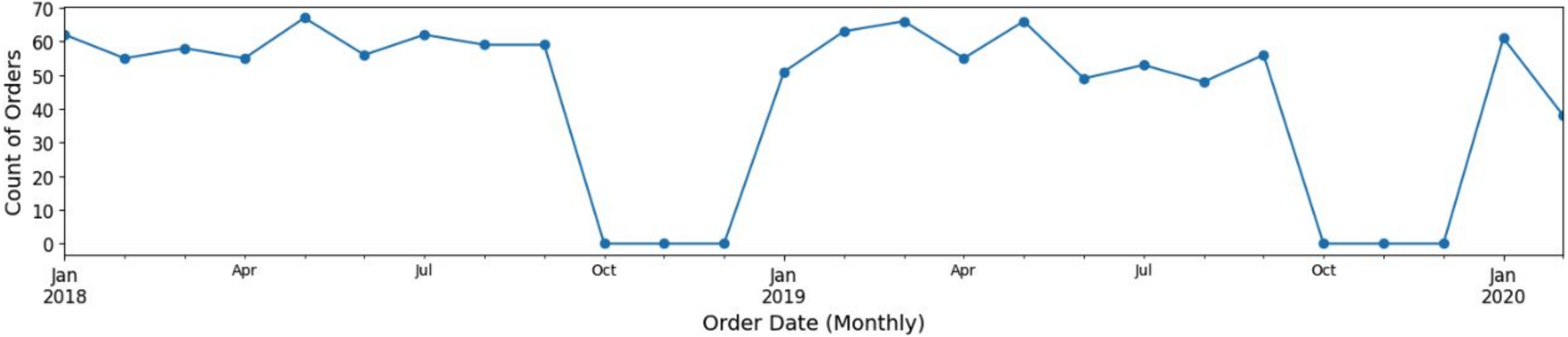


2.1.4 - Top products by Occurrence

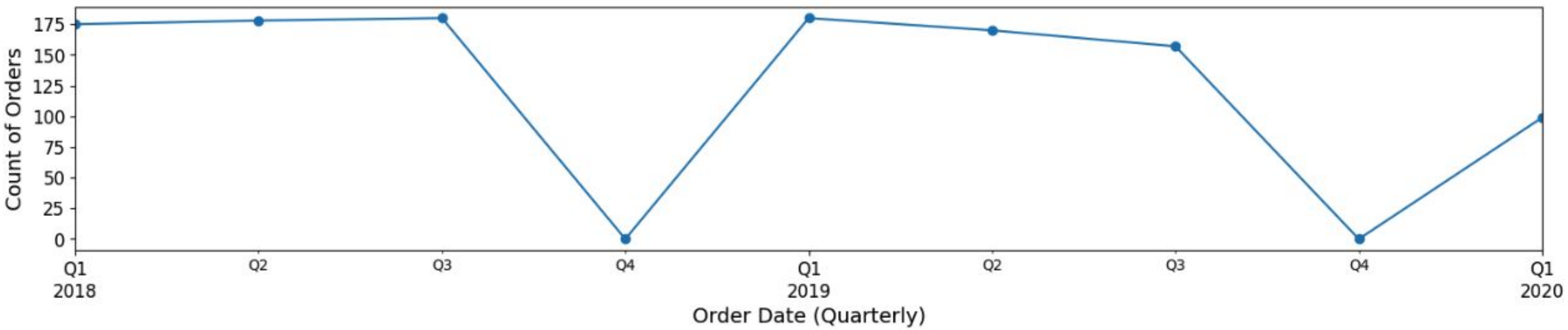
Back



2.1.5 - Sales pattern - Weekly



2.1.5 - Sales pattern - Monthly



2.1.5 - Sales pattern - Quarterly

Back

2.2 Use of Association Rules

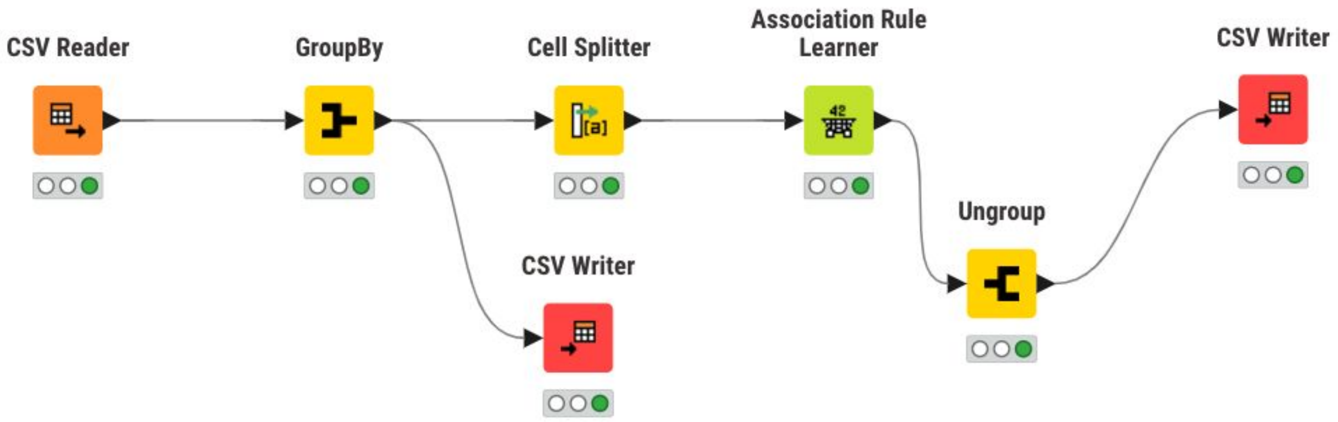
Association Rules are patterns or relationships between items of a dataset. In the context of a grocery store, they help uncover how products are related to each other based on customer purchase behavior. These rules are derived from techniques like Market Basket Analysis and can be helpful in making decisions for marketing and promotions.

The metrics used for this are Support, Confidence, and Lift, when talking about the relationship between the main Item (Antecedent) and the Recommended Item (Consequent).

Support is the probability of the presence of the Antecedent in the data.

Confidence is the probability that Antecedent and Consequent are together in a basket given that the Antecedent is already in it.

Lift measures how much more likely the consequent is to occur when the antecedent is present, compared to its independent occurrence.



2.2.1 - Knime Workflow image

Support below 0.1 was giving 930 rules, and over 0.2 it was giving 0 rules. So it was set at 0.17 which gave 130 rules.

That means the minimum probability of occurrence of the Antecedent in a basket is 17%.

Confidence above 0.5 was giving 0 rules, at 0.45 was giving 59 rules and at and below 0.4 was giving 130 rules. Therefore it was set at 0.4.

This means given that the Antecedent is in the basket, the probability that the Consequent is also in the basket is a minimum of 40%.

Back

2.3 Associations Identified

Support	Confidence	Lift	Consequent	implies	Items
0.17	0.46	1.23	juice	<---	shampoo
0.17	0.45	1.23	shampoo	<---	juice
0.17	0.47	1.18	ice cream	<---	paper towels
0.17	0.43	1.18	paper towels	<---	ice cream
0.17	0.45	1.15	dishwashing liquid/de	<---	milk
0.17	0.44	1.15	milk	<---	dishwashing liquic
0.17	0.45	1.08	poultry	<---	beef
0.17	0.40	1.08	beef	<---	poultry
0.17	0.45	1.08	poultry	<---	individual meals
0.17	0.40	1.08	individual meals	<---	poultry
0.17	0.44	1.13	eggs	<---	dishwashing liquic
0.17	0.44	1.13	dishwashing liquid/de	<---	eggs
0.17	0.44	1.13	soda	<---	dinner rolls
0.17	0.44	1.13	dinner rolls	<---	soda
0.17	0.44	1.11	cereals	<---	dinner rolls
0.17	0.43	1.11	dinner rolls	<---	cereals
0.17	0.46	1.10	poultry	<---	shampoo
0.17	0.41	1.10	shampoo	<---	poultry
0.17	0.46	1.10	poultry	<---	tortillas

2.3.1 - Associations in tabular form

Support: Goes up to a maximum of 0.19. This means no Antecedent has a probability of being present in the basket more than 19%.

Confidence: Goes up to a maximum of 0.5. This means with the Antecedent already present in a basket, the chances that the Consequent is also present are never more than 50%.

Lift: Ranges from 1.05 to 1.23. Those closer to 1 means there is no association between the Antecedent and Precedent. While a Lift of 1.23 does not imply a significant association between the items.

2.4 Possible Combos & Offers

2.4.1 Recommendations: Rules with High Confidence

Place next to each other

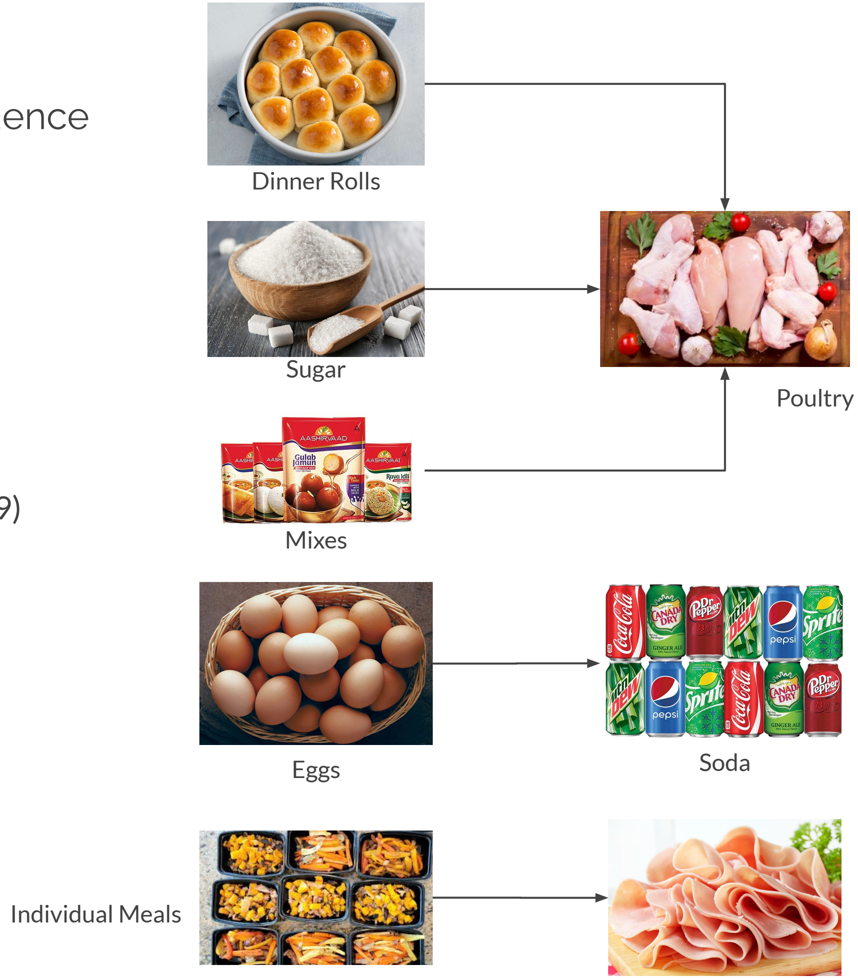
Dinner Rolls → Poultry (Conf: 0.5, Lift: 1.19)

Sugar → Poultry (Conf: 0.5, Lift: 1.18)

Mixes → Poultry (Conf: 0.48, Lift: 1.15)

Eggs → Soda (Conf: 0.48, Lift: 1.23)

Individual Meals → Lunch Meats (Conf: 0.47, Lift: 1.19)



Back

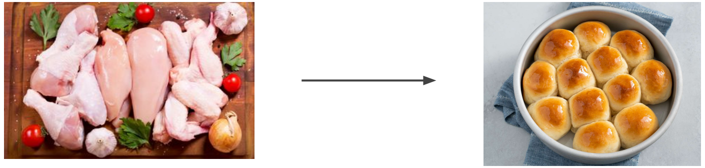
2.4.2 **Promos & Bundles:** Rules with High Lift

1. Buy 2 Juice, get Spaghetti Sauce free (Conf: 0.46, Lift: 1.22)
2. Dinner Roll + Pasta Bundle (25% off)(Conf: 0.44, Lift: 1.20)
3. Buy X amount of Poultry get 1 packet Dinner Rolls (Conf: 0.50, Lift: 1.19)
4. Buy 2 Individual Meals, get 1 packet Lunch Meat free (Conf: 0.45, Lift: 1.19)
5. Buy 2 Juice, get 1 Yogurt free.

2 x Juice

Spaghetti Sauce free

Dinner Rolls + Pasta Bundle



Buy X amount of Poultry get Dinner Rolls free

2 x Individual Meals

1 packet Lunch Meat free

2 x Juice

1 packet Yogurt free