Market & Retail Analytics

RFM Analysis

Market Basket Analysis

-Kanchan Narsinghani

Agenda

- RFM Analysis
  - Problem Statement
  - About Data
  - Univariate Analysis
  - Bivariate Analysis
  - Summary
  - RFM Analysis
  - What assumptions and parameters made ?
  - KNIME workflow
  - Output
    - Top Customers
    - Lost Customers
    - Churning Customers
    - Champion Customers

Agenda

- Market Basket Analysis
  - Problem Statement
  - About Data
  - Exploratory Analysis
  - Summary
  - KNIME Workflow
  - Association Rules
  - Discount and Combo Offers

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.

About Data

Name	Type	# Missing valu...	# Unique values	Minimum	Maximum	25% Quantile	50% Quantile (...	75% Quantile	Mean	Mean Absolut...	Standard Devi...	Sum	10 most com...
ORDERNUMBER	Number (integ...	0	298	10,100	10,425	10,181	10,264	10,335	10,259.762	79.559	91.878	28,183,565	10,106 (18; 0.6...
QUANTITYORD...	Number (integ...	0	58	6	97	27	35	43	35.103	8.176	9.762	96,428	34 (111; 4.04%...
PRICEEACH	Number (doubl...	0	1844	26.88	252.87	68.71	95.55	127.1	101.099	33.895	42.043	277,718.819	59.87 (6; 0.22...
ORDERLINENU...	Number (integ...	0	18	1	18	3	6	9	6.491	3.552	4.231	17,831	1 (298; 10.85%...
SALES	Number (doubl...	0	2690	482.13	14,082.8	2,204.1	3,184.8	4,508	3,553.048	1,426.213	1,838.954	9,760,221.71	1,030.44 (2; 0...
ORDERDATE	Local Date	0	246	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	2018-11-14 (3...
DAYS_SINCE_L...	Number (integ...	0	1832	42	3,562	1,077	1,761	2,437	1,757.086	699.963	819.281	4,826,715	2,207 (7; 0.25%...
STATUS	String	0	6	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	Shipped (2541;...
PRODUCTLINE	String	0	7	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	Classic Cars (9...
MSRP	Number (integ...	0	80	33	214	68	99	124	100.692	32.463	40.115	276,600	118 (102; 3.71...
PRODUCTCODE	String	0	109	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	S18_3232 (51; ...
CUSTOMERNA...	String	0	89	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	Euro Shopping ...
PHONE	String	0	88	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	(91) 555 94 44 ...
ADDRESSLINE1	String	0	89	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	C/ Moralzarzal,...
CITY	String	0	71	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	Madrid (304; 1...
POSTALCODE	String	0	73	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	28034 (259; 9...
COUNTRY	String	0	19	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	USA (928; 33.7...
CONTACTLAS...	String	0	76	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	Freyre (259; 9...
CONTACTFIRS...	String	0	72	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	Diego (259; 9.4...
DEALSIZE	String	0	3	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	⊘	Medium (1349;...

There are 2747 rows and 20 columns of data.

There are 5 integer, 2 float, 12 string columns and 1 date column.

There are no missing values.

Total Quantity ordered is 96428 units for all products

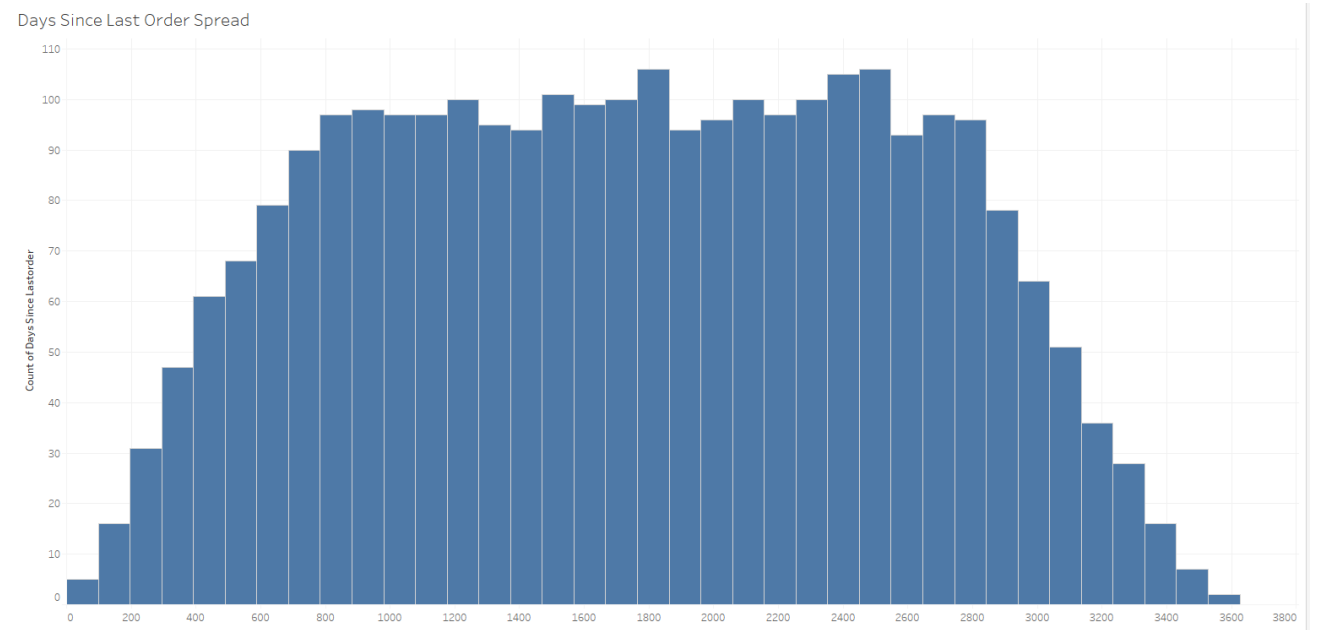Sales for the company is 97,60,221 .

Minimum and maximum difference in days between 2 order is 42 and 3562.

Univariate Analysis

Days Since Last Order Spread

Days since last order is calculated as difference between 2 order dates in terms of days.
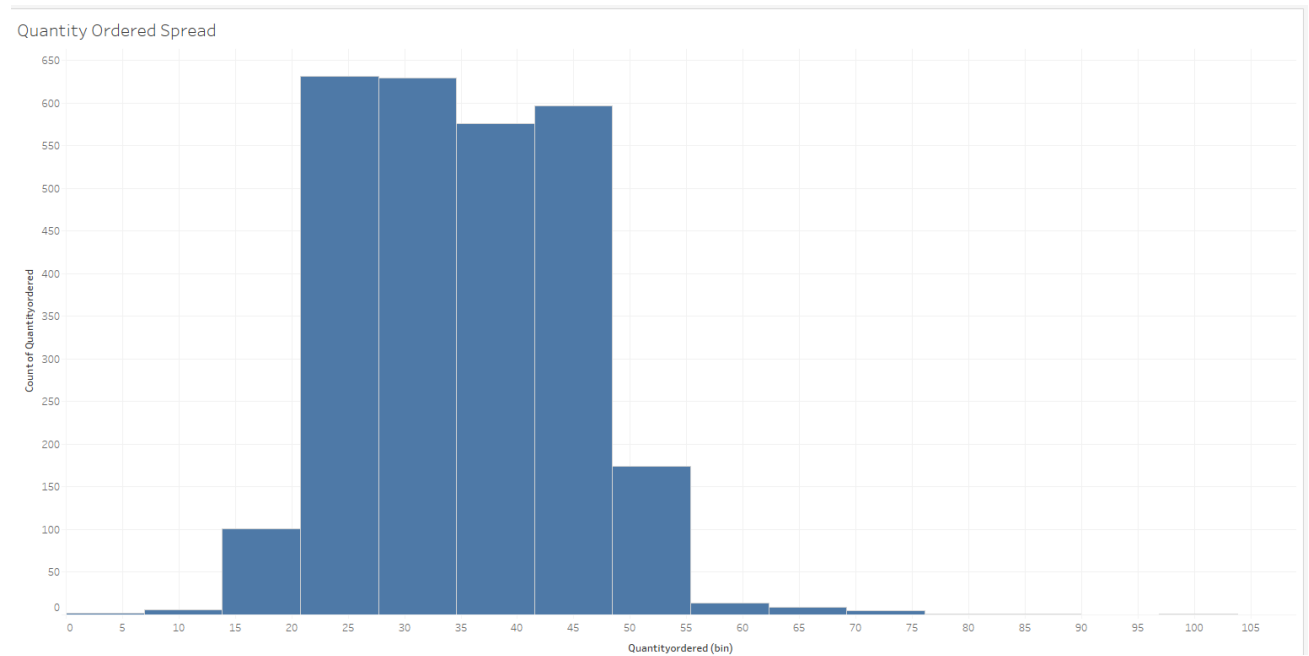
It ranges from 0 to 3600 days.

Days Since Last Order Spread

Quantity Ordered spread

We can see from the chart that minimum ordered quantity can be low as 5 to high as 75.
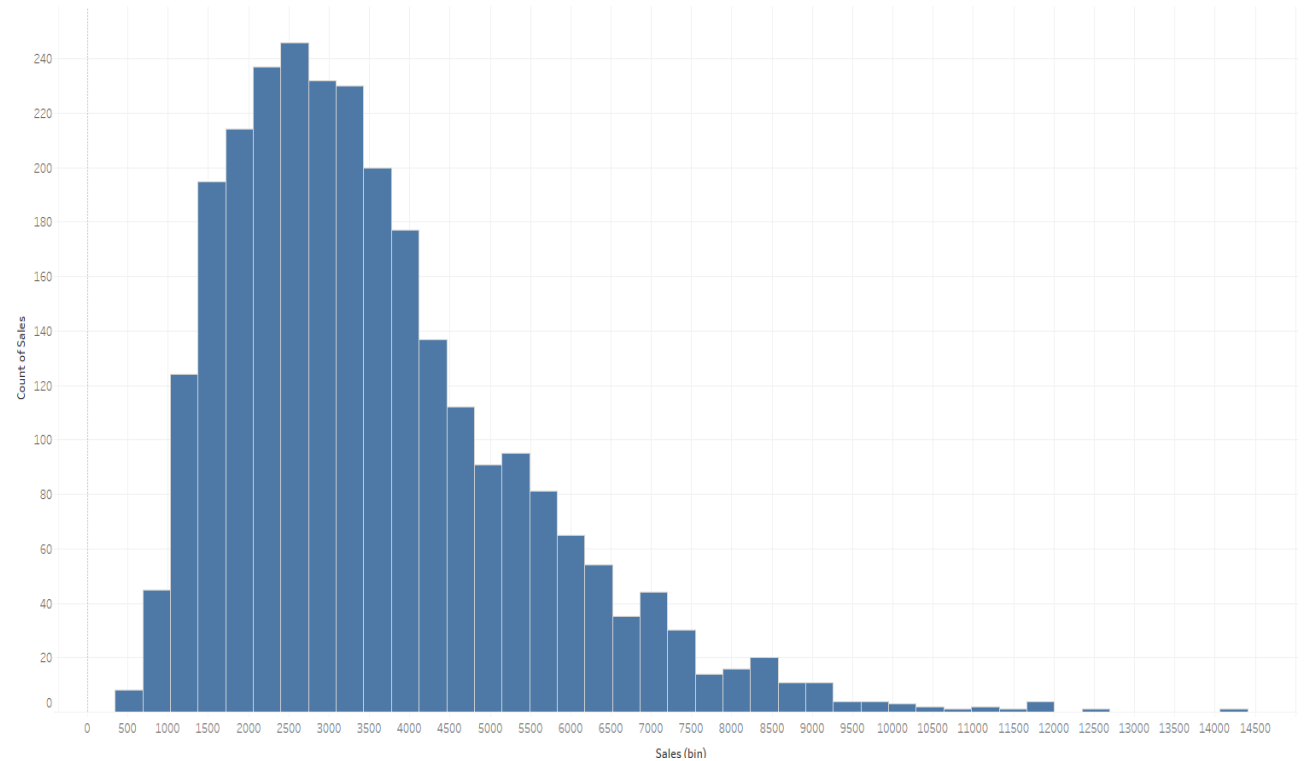
Most of products lie in range of 20 -50.

Sales

We can see from the chart that minimum amount of purchase is 500 and can go as high as 14600.

It shows left skewed distribution of data.

Most of sales are between 2500 to 3000 range.
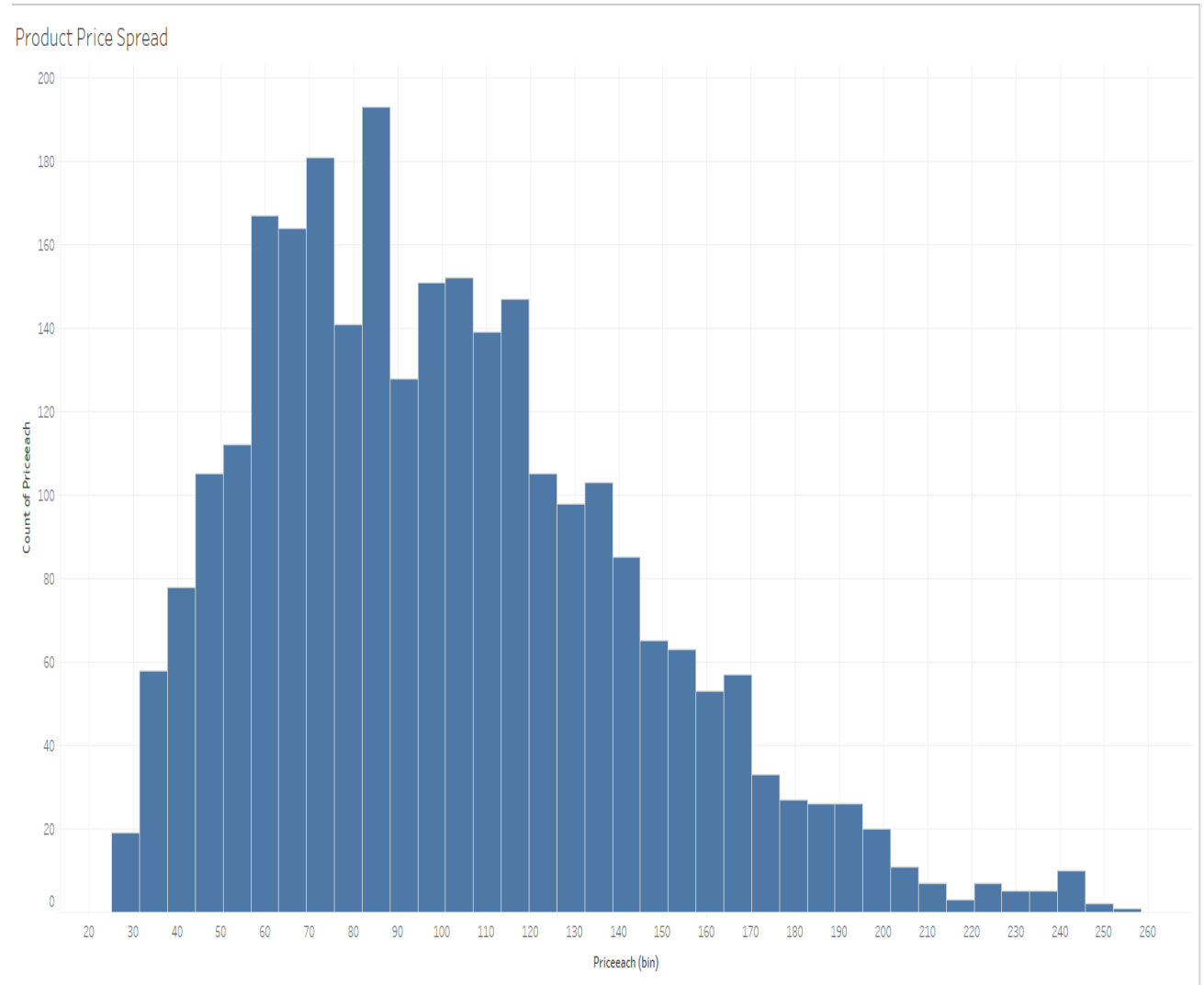

Sales Spread

Product Price Spread

Product price can be low as 25 and high as 260.

Data has left skewed distribution.
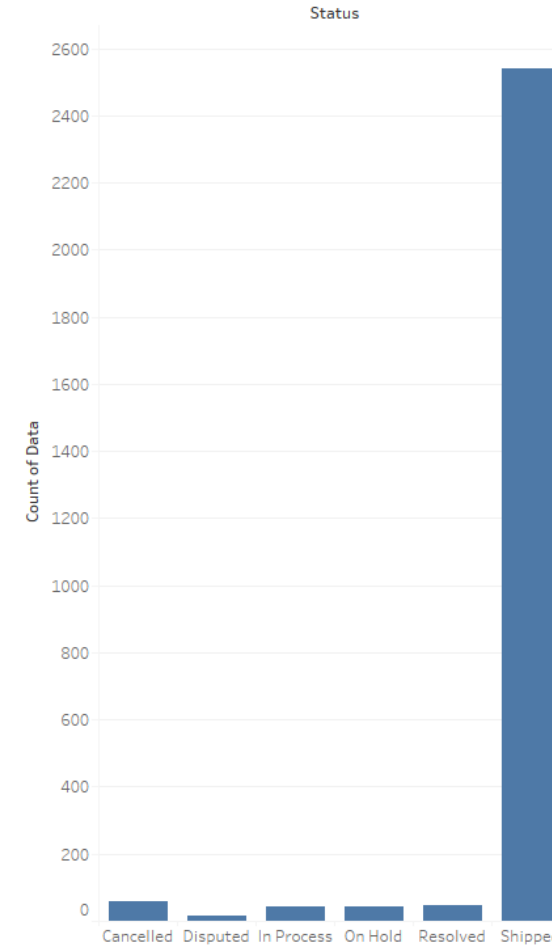
Most of products are priced at 80-90.

Shipping Status wise Orders

We can see there are 6 categories of status: cancelled, disputed, In Process, On hold, Resolved, Shipped.

Most of the orders have shipped status.

Shipping Status Orders

Product Line wise Orders

There are 7 types of products in our transaction data.

Most of the orders are for Classic cars and vintage cars.

Then there are almost equal orders for motorcycles, planes, trucks and buses.

ProductLine Wise Orders
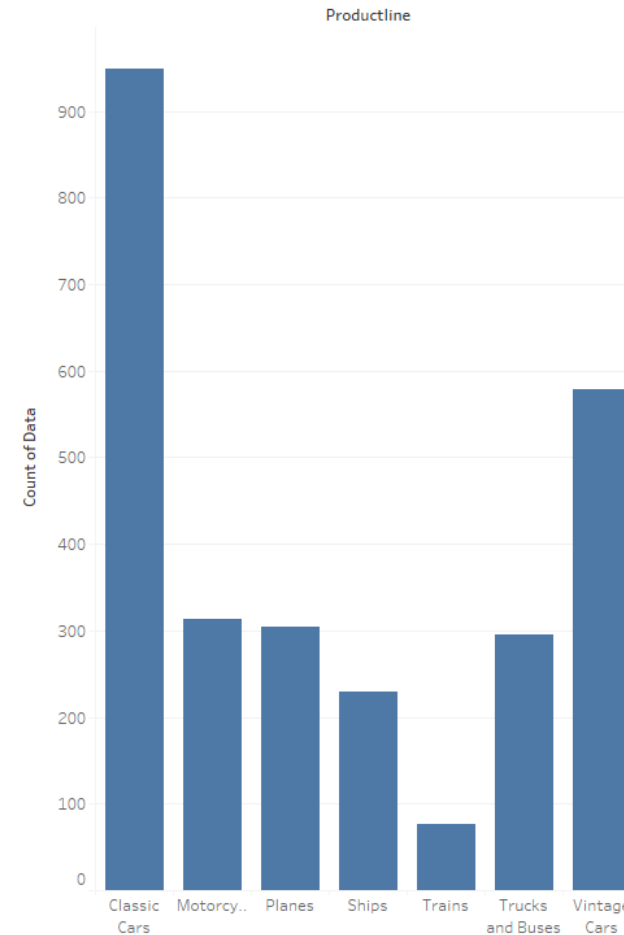
DealSize wise Orders

Our deals from data are categorized into small, medium and large.

Our highest orders range in medium dealsize followed by small then large dealsize.

DealSize Wise Orders
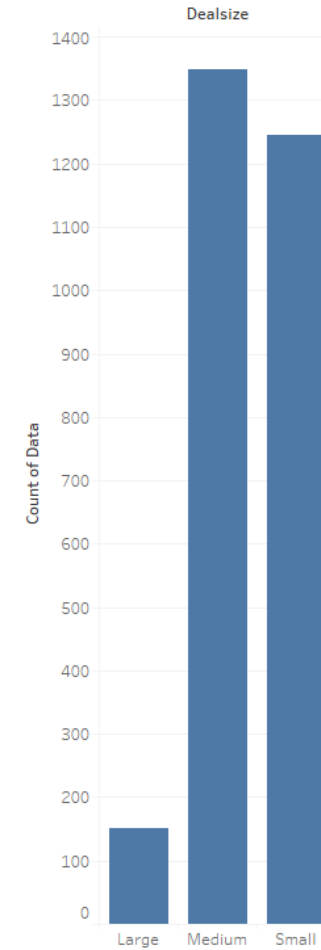
Product Code wise Orders

Our data contains 109 product codes.

Most of our orders are for S18_3232.
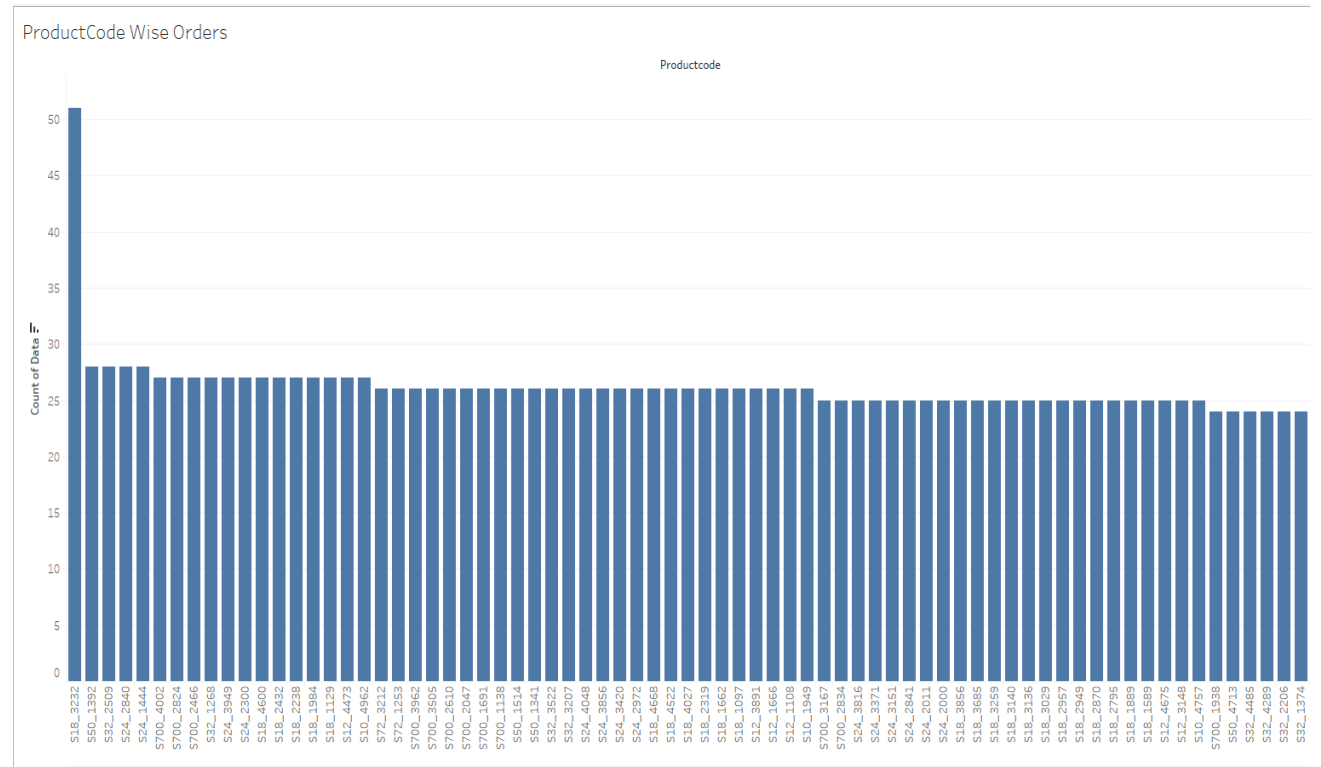
Country wise Orders

We have highest sales in USA followed by Spain and France.

Our lowest sales are in Ireland, Belgium, Philippines and Switzerland.

CountryWise Orders
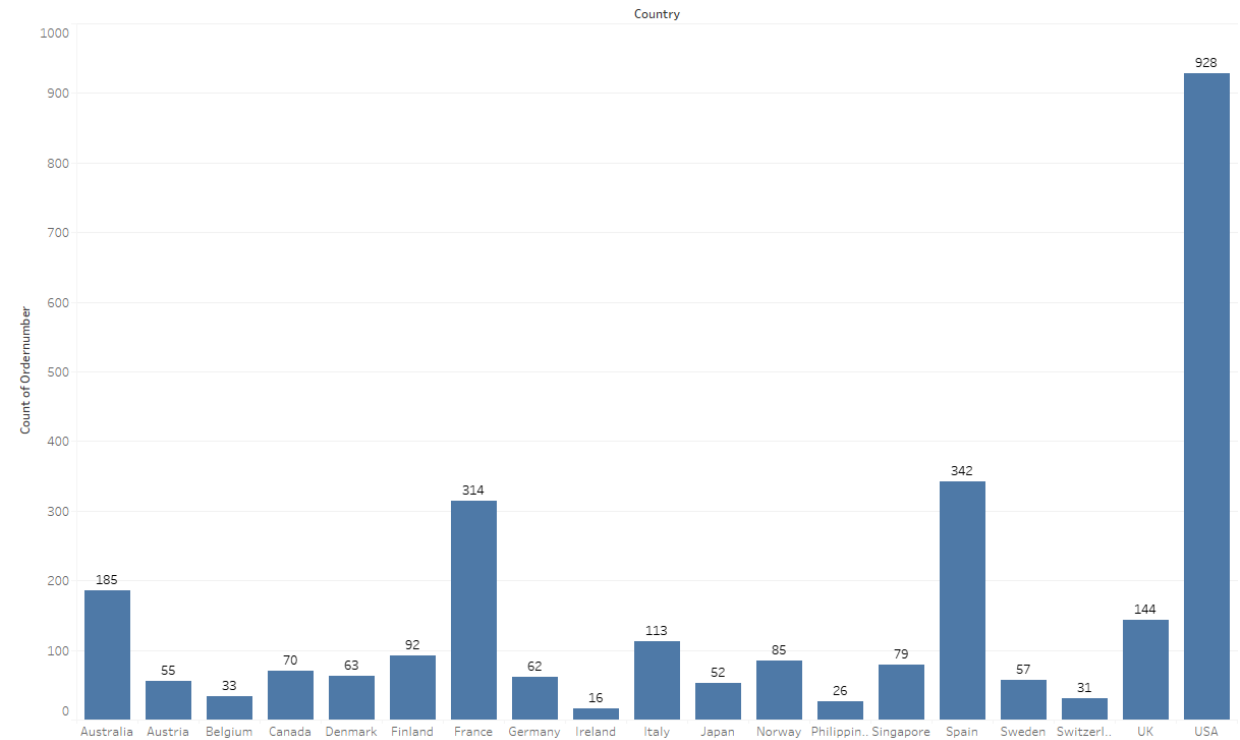
Bivariate Analysis

Sales vs Product Price

We can see positive relation between price and sales.

Most of products sold lie in price range of 20 to 140 leading to sales of 1k to 7k.

Sales vs price of product
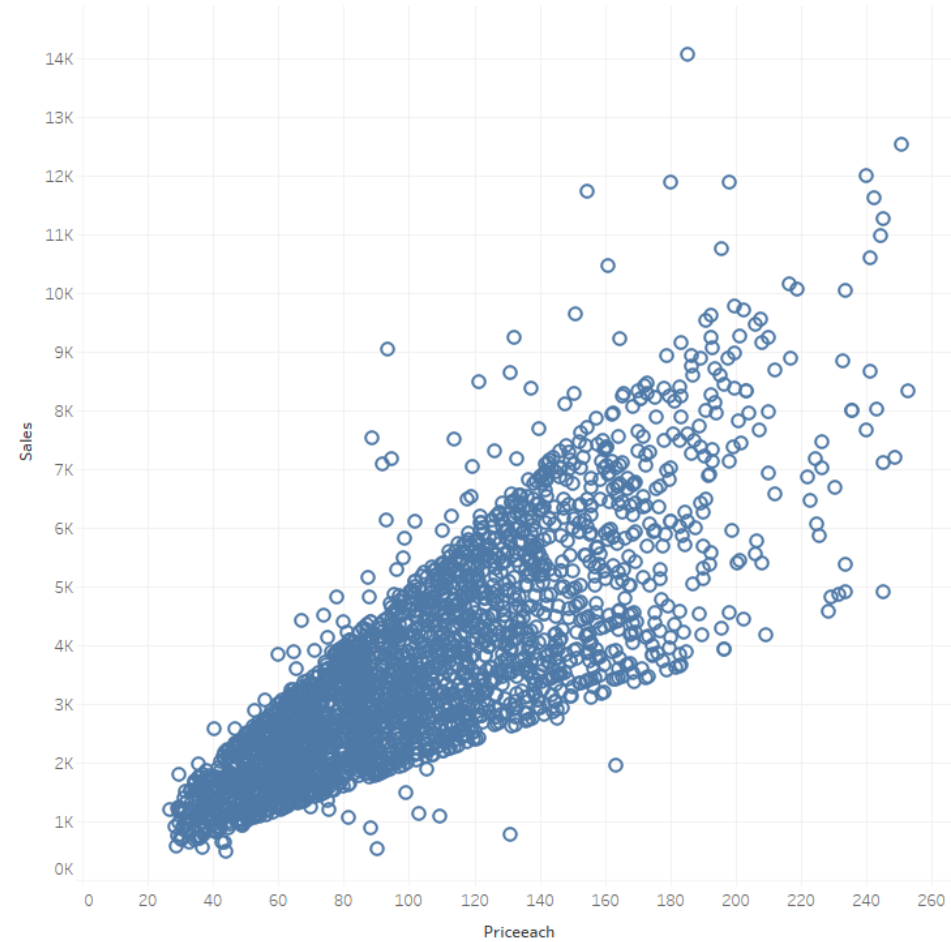
Month and Year wise Sales

We have data for 3 years : 2018, 2019, 2020.

Our lowest sales are in January 2018 at 129,754.

Our Highest sales are in November 2019 at 10,58,699.

Month and Yearwise sales

Month of O..	Orderdate 2018	2019	2020
January	129,754	292,688	339,543
February	140,836	311,420	303,983
March	155,809	205,734	374,263
April	201,610	206,148	261,633
May	192,673	273,438	457,861
June	168,083	286,674	
July	187,732	327,144	
August	197,809	461,501	
September	263,973	320,751	
October	448,453	552,924	
November	1,029,838	1,058,699	
December	236,445	372,803	

Quarterly Sales

We have quarterly data for 3 years.

We have minimum sale in Q1 of 2018 and highest sales in Q4 of 2019.

Quarter and Yearwise sales

Quarter of ..	Orderdate		
	2018	2019	2020
Q1	426,399	809,841	1,017,789
Q2	562,365	766,261	719,494
Q3	649,515	1,109,396	
Q4	1,714,735	1,984,426	

Weekly Sales

We have 51 weeks of data for 2018, 2019, 2020.

We have few weeks with no sales.

Our lowest sales are in week 1 of 2018 and highest sales in week 46 of 2018.

Weekly and Yearwise sales

	Orderdate		
Week of Or..	2018	2019	2020
Week 1	12,133	54,537	
Week 2	18,296	50,491	94,870
Week 3		116,338	44,273
Week 4		33,473	113,153
Week 5	99,324	43,610	87,247
Week 6		24,220	85,335
Week 7	115,052	54,903	78,198
Week 8	25,784	200,665	135,351
Week 9	55,245	72,674	5,098
Week 10	27,399		188,289
Week 11		107,972	43,935
Week 12	60,767	42,237	42,675
Week 13	12,399	8,722	47,334
Week 14	62,995	63,394	136,075
Week 15	1,711	105,599	54,870
Week 16	47,877	1,699	61,693
Week 17	38,629	7,129	61,025
Week 18	50,398	65,301	33,693
Week 19	71,454	180,547	146,631
Week 20		27,987	73,332
Week 21	60,146		59,475
Week 22	61,073	78,339	65,812
Week 23	85,910	10,166	78,918
Week 24	60,005	87,846	
Week 25		24,565	
Week 26	22,168	94,141	
Week 27	102,156	88,305	
Week 28	15,146	23,252	
Week 29	28,397	116,261	
Week 30	42,032	118,874	
Week 31	31,569	52,505	
Week 32	108,827	59,568	
Week 33	1,637	56,078	
Week 34	55,776	175,066	
Week 35		118,283	
Week 36	42,593	66,939	
Week 37	81,292	151,794	
Week 38	81,721	53,966	
Week 39	58,368	5,308	
Week 40	146,359	42,744	
Week 41	7,249	62,296	
Week 42	65,606	309,983	
Week 43	166,933	159,926	
Week 44	62,305	153,558	
Week 45	325,737	282,507	
Week 46	331,329	134,856	
Week 47	183,204	286,512	
Week 48	197,803	221,986	
Week 49	164,479	178,472	
Week 50	63,731	126,211	
Week 51		68,120	

Geographic Sales

Our data shows sales in USA, Canada, Australia, Philippines, Japan and some European Countries.

Our highest sales are in USA.

Geographrically sales
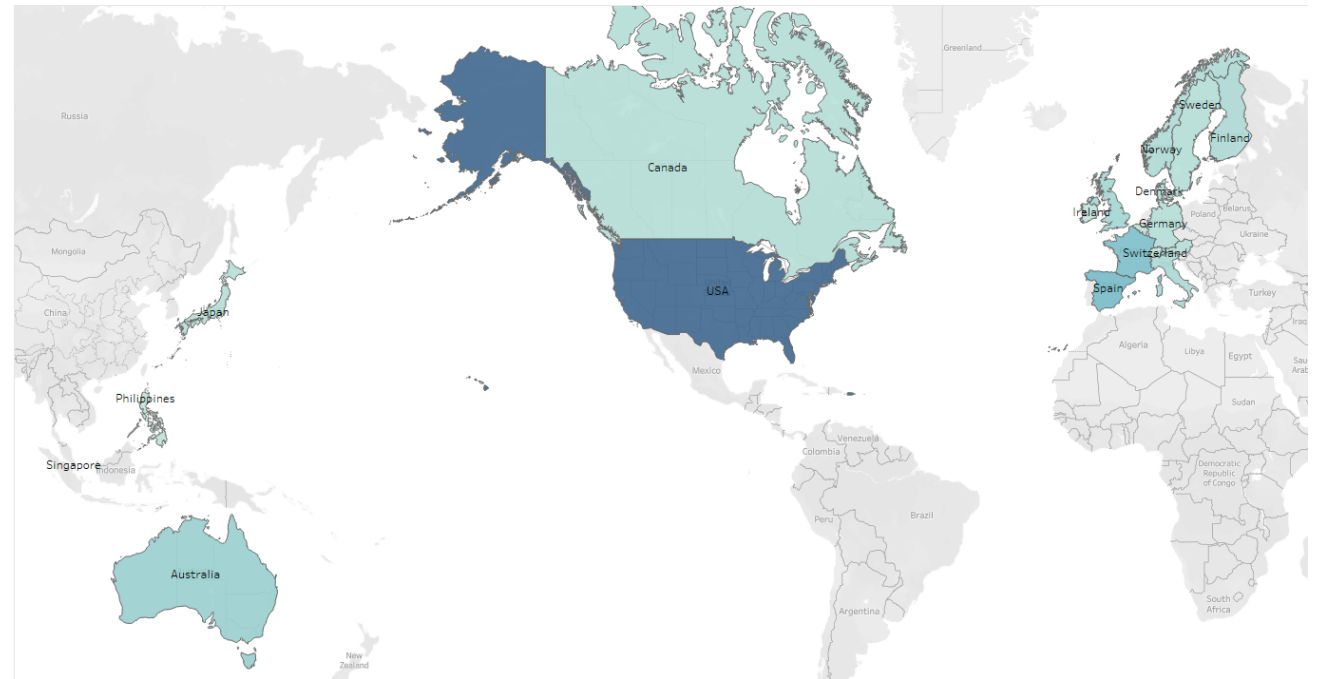
Customer wise Days since last order

Data shows difference in days between 2 orders placed by customer for all orders.

Customers have placed orders at difference of minimum 0 days to maximum 3600 days.
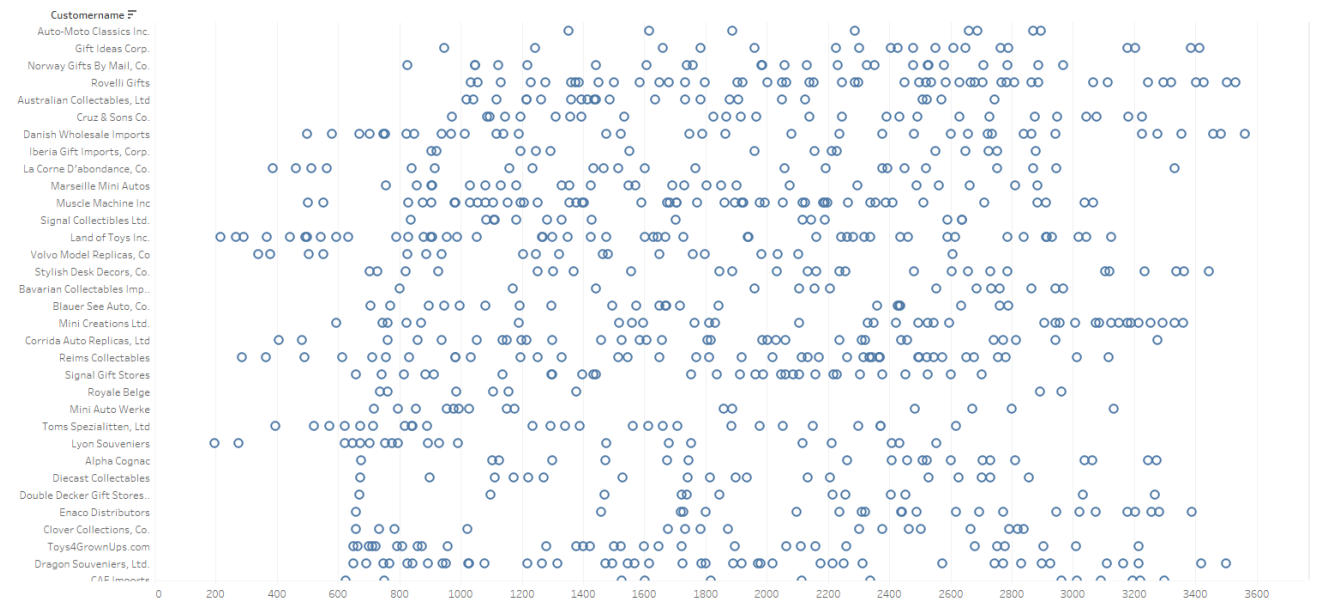
Customer wise Recency

We calculate recency as maximum order date subtracted from today's date.

This gives us a unique value per customer.

Lower the recency, better the customer for the firm.

Customers have recency ranging from 1600 days to 2100 days.
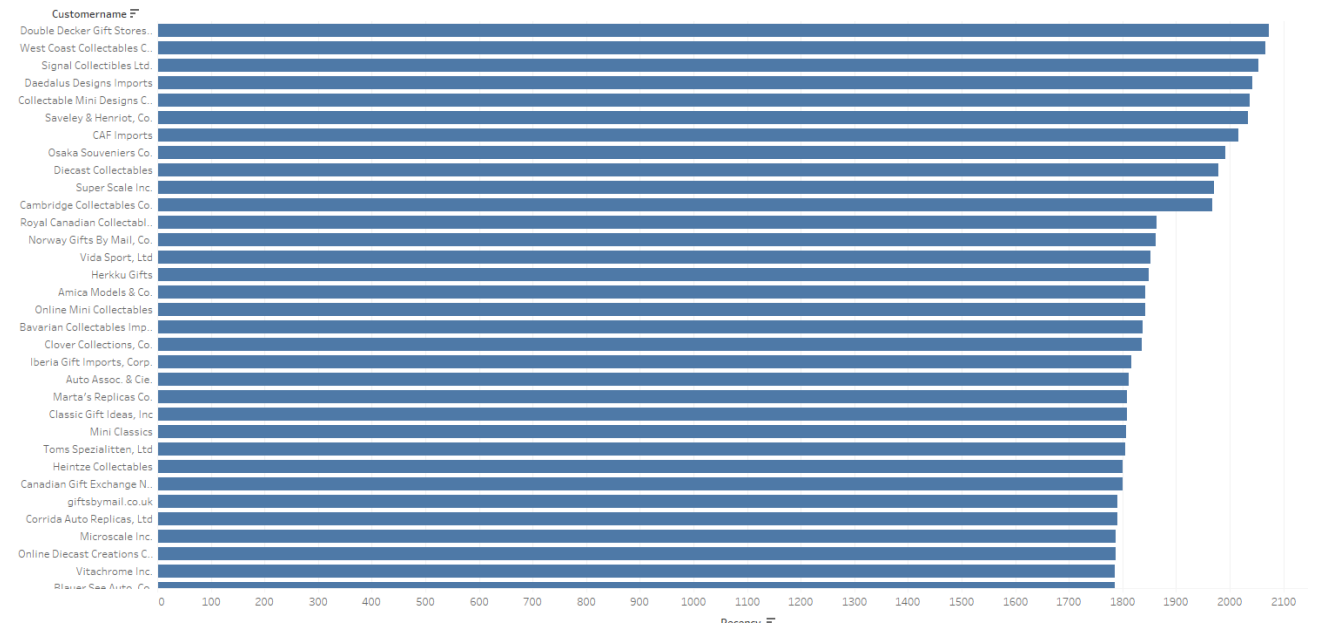
Customer wise Frequency

This shows number of times customer placed an order.

Customers have placed orders from as low as 1 time to 26 times.

Customer wise Frequency

Customer wise Sales

We are calculating total amount of sales per customer.

Our customers have sales ranging from 1k to 950k.

Customer wise sales
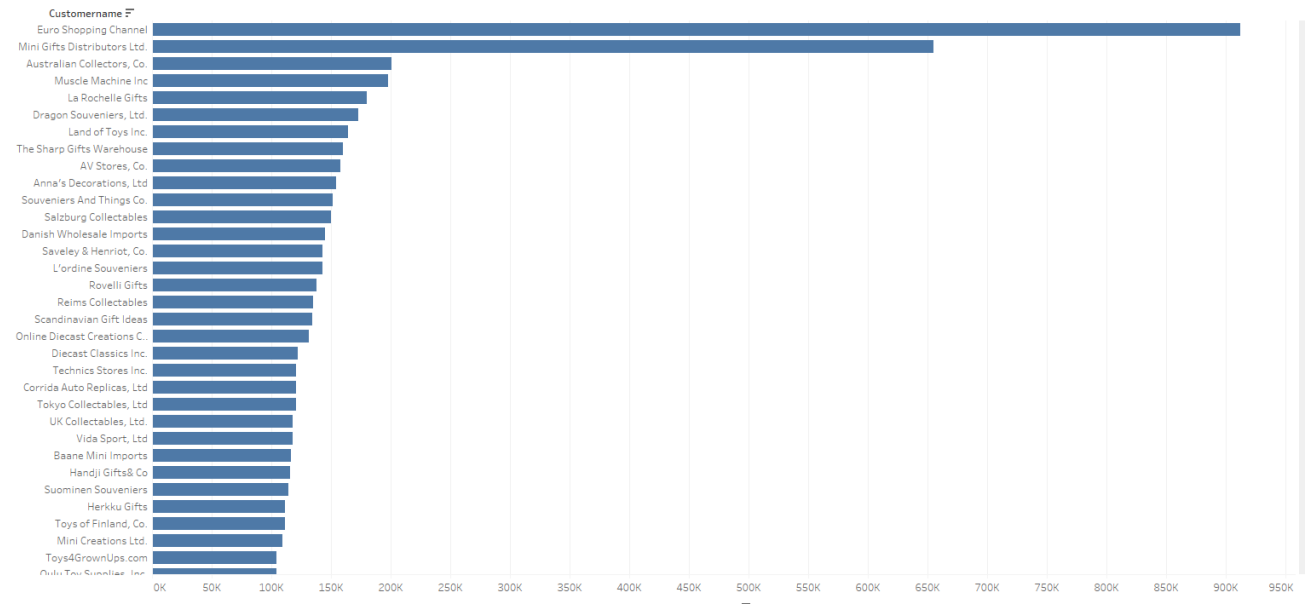
Summary

- We have done exploratory analysis on data and found following facts.

- Starting with numeric variables, Days since last order column has values ranging from 0 to 3600 days.

- minimum amount of purchase is 500 and can go as high as 14600.It shows left skewed distribution of data. Most of sales are between 2500 to 3000 range.

- Product price can be low as 25 and high as 260.Data has left skewed distribution. Most of products are priced at 80-90.

- We can see there are 6 categories of status: cancelled, disputed, In Process, On hold, Resolved, Shipped. Most of the orders have shipped status.

- There are 7 types of products in our transaction data. Most of the orders are for Classic cars and vintage cars then there are almost equal orders for motorcycles, planes, trucks and buses.

- Our deals from data are categorized into small, medium and large.

- Our highest orders range in medium deal size followed by small then large deal size.

- Our data contains 109 product codes. Most of our orders are for S18_3232.

Summary(contd.)

- We have highest sales in USA followed by Spain and France. Our lowest sales are in Ireland, Belgium, Philippines and Switzerland.

- We can see positive relation between price and sales. Most of products sold lie in price range of 20 to 140 leading to sales of 1k to 7k.

- We have data for 3 years : 2018, 2019, 2020. Our lowest sales are in January 2018 at 129,754. Our Highest sales are in November 2019 at 10,58,699. We have minimum sale in Q1 of 2018 and highest sales in Q4 of 2019.

- We have 51 weeks of data for 2018, 2019, 2020. We have few weeks with no sales. Our lowest sales are in week 1 of 2018 and highest sales in week 46 of 2018.

- Customers have placed orders at difference of minimum 0 days to maximum 3600 days.

- We calculate recency as maximum order date subtracted from today's date. This gives us a unique value per customer. Lower the recency, better the customer for the firm. Customers have recency ranging from 1600 days to 2100 days. Customers have placed orders from as low as 1 time to 26 times.
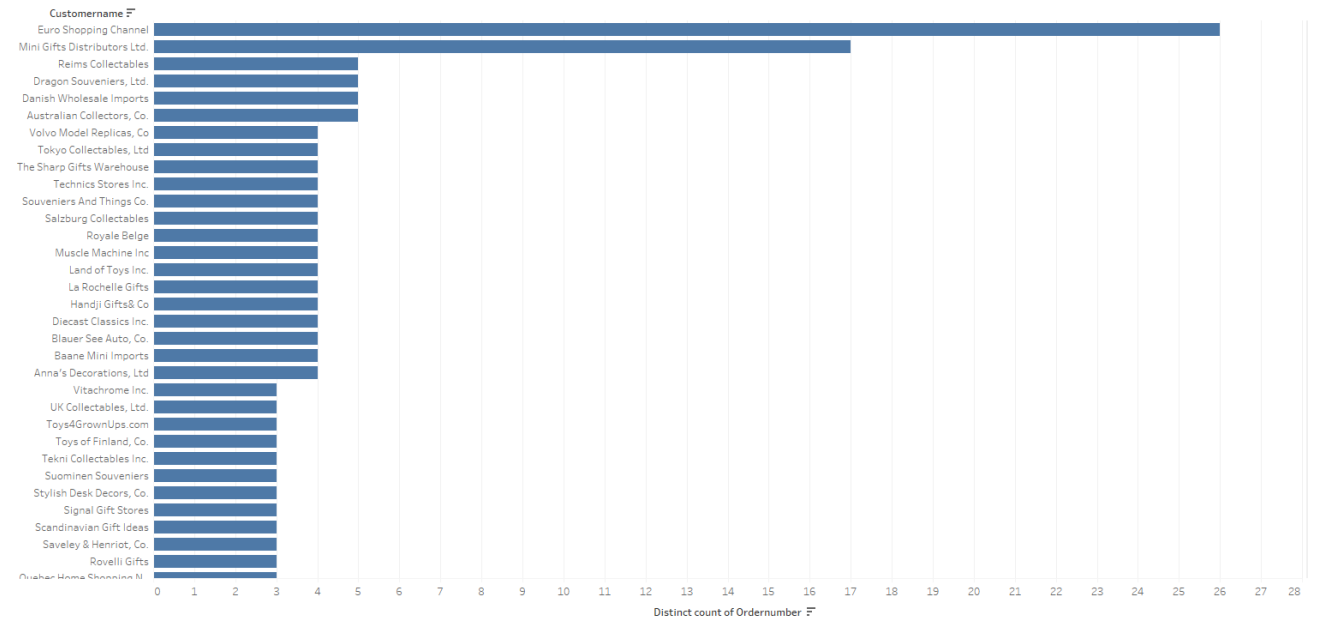
- Our customers have sales ranging from 1k to 950k.

RFM Analysis

- RFM stands for Recency-Frequency-Monetary.

- Recency stands for days since customer last ordered with business.

- Frequency stands for number of times customer placed an order.

- Monetary stands for total sales per customer.

- We calculate RFM score for each customer.

- We define number of segments and assign customers to bins based on the RFM score.

What all parameters used and assumptions made?

- We use customer name, Order date and Sales value for calculating RFM.
- We filtered null values from the data.
- We calculated Recency by deducting order date from current date.
- We aggregated columns based on customer name and took minimum recency and sum of order numbers and sum of sales value per customer.
- Created 4 segments to assign bins to customer based on RFM score.
- Created final RFM Score column to assign aggregated RFM value per customer.

KNIME workflow
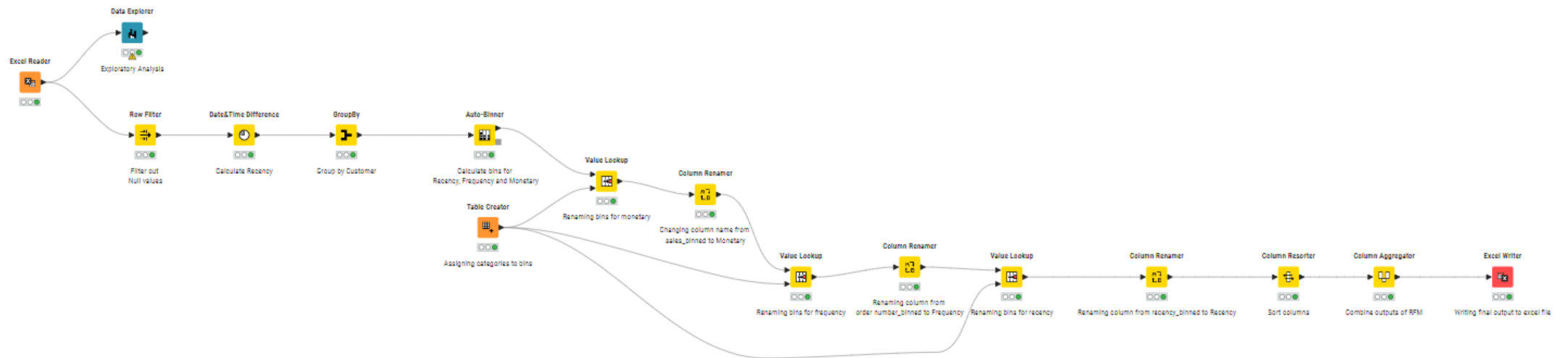
Output Table

CUSTOMERNAME	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	DAYS_SINCE_LASTORDER	STATUS	PRODUCTLINE	MSRP	PRODUCTCODE	PHONE	ADDRESSLINE1	CITY	POSTALCODE	COUNTRY
AV Stores, Co.	51	34.8627451	91.0845098	51	157807.8	2018-03-18	421	Shipped	Classic Cars	92.84314	S12_1108	(171) 555-1	Fauntleroy Circus	Manchest	EC2 5NT	UK
Alpha Cognac	20	34.35	101.16	20	70488.44	2018-07-04	675	Shipped	Classic Cars	97.15	S10_4757	61.77.6555	1 rue Alsace-Lorrain	Toulouse	31000	France
Amica Models & Co.	26	32.42307692	110.8526923	26	94117.26	2019-08-17	328	Shipped	Classic Cars	107.6538	S10_1949	011-498855	Via Monte Bianco 34	Torino	10100	Italy
Anna's Decorations, Lt	46	31.93478261	106.4241304	46	153996.1	2018-09-11	131	Shipped	Classic Cars	104.7174	S10_1949	02 9936 85	201 Miller Street	North Sydr	2060	Australia
Atelier graphique	7	38.57142857	92.23857143	7	24179.96	2018-05-20	312	Shipped	Motorcycles	95.57143	S10_2016	40.32.2555	54, rue Royale	Nantes	44000	France
Australian Collectables	23	30.65217391	90.04173913	23	64591.46	2018-11-21	1018	Disputed	Vintage Cars	88.13043	S18_1342	61-9-3844-(	7 Allen Street	Glen Wave	3150	Australia
Australian Collectors,	55	35.01818182	104.5901818	55	200995.4	2018-04-29	229	Shipped	Motorcycles	103.5273	S10_1678	03 9520 45	636 St Kilda Road	Melbourne	3004	Australia
Australian Gift Networl	15	36.33333333	110.554	15	59469.12	2018-09-25	190	Shipped	Classic Cars	111.5333	S10_1949	61-7-3844-(	31 Duncan St. West	South Bris	4101	Australia
Auto Assoc. & Cie.	18	35.38888889	99.48780142	18	64834.32	2019-02-02	275	Shipped	Classic Cars	100.3889	S10_1949	30.59.8555	67, avenue de l'Euro	Versailles	78000	France
Auto Canal Petit	27	37.07407407	94.25518519	27	93170.66	2019-01-15	127	Shipped	Motorcycles	94.85185	S10_1678	(1) 47.55.6!	25, rue Lauriston	Paris	75016	France

CONTACTLASTNAME	CONTACTFIRSTNAME	DEALSIZE	Recency	ORDERNUMBER [Binned]	SALES [Binned]	Recency [Binned]	Monetary_categories	Frequency_categories	Recency_categories	RFM Score
Ashworth	Victoria	Medium	1772	Bin 4	Bin 4	Bin 3	4	4	2	442
Roulet	Annette	Medium	1640	Bin 1	Bin 1	Bin 1	1	1	4	114
Accorti	Paolo	Large	1841	Bin 2	Bin 3	Bin 4	3	2	1	321
O'Hara	Anna	Small	1659	Bin 4	Bin 4	Bin 2	4	4	3	443
Schmitt	Carine	Medium	1764	Bin 1	Bin 1	Bin 3	1	1	2	112
Connery	Sean	Medium	1598	Bin 2	Bin 1	Bin 1	1	2	4	124
Ferguson	Peter	Medium	1760	Bin 4	Bin 4	Bin 2	4	4	3	443
Calaghan	Tony	Large	1695	Bin 1	Bin 1	Bin 2	1	1	3	113
Tonini	Daniel	Large	1809	Bin 1	Bin 1	Bin 4	1	1	1	111
Perrier	Dominique	Medium	1630	Bin 3	Bin 3	Bin 1	3	3	4	334

Top Customers

These are customers with highest RFM score of 444.

We should always appreciate these customers and provide our best efforts to retain their business.

Euro Shopping Channel is our best customer with lowest recency and highest frequency and monetary.

Top Customers - Customers with Highest RFM score

CUSTOMERNAME	Data Sum of RFM Score	Sum of Recency	Sum of SALES	Max of ORDERNUMBER
Euro Shopping Channel	444	1576	912294.11	259
Mini Gifts Distributors Ltd.	444	1578	654858.06	180
La Rochelle Gifts	444	1576	180124.9	53
The Sharp Gifts Warehouse	444	1615	160010.27	40
Souveniers And Things Co.	444	1578	151570.98	46
Salzburg Collectables	444	1590	149798.63	40
Danish Wholesale Imports	444	1622	145041.6	36
L'ordine Souveniers	444	1597	142601.33	39
Reims Collectables	444	1638	135042.94	41
Grand Total	3996	14370	2631342.82	259

Lost Customers

These customers have lowest RFM score of 111.

These customers can be excluded from company's marketing efforts as they can not be saved.

Daedalus Design Imports has highest recency whereas Bavarian Collectables Imports Co. has lowest frequency and sales figure.

Lost Customers - Customers with lowest RFM score

CUSTOMERNAME	Data Sum of RFM Score	Sum of Recency	Sum of SALES	Max of ORDERNUMBER
Bavarian Collectables Imports, Co.	111	1835	34993.92	14
Double Decker Gift Stores, Ltd	111	2071	36019.04	12
Cambridge Collectables Co.	111	1965	36163.62	11
West Coast Collectables Co.	111	2064	46084.64	13
CAF Imports	111	2015	49642.05	13
Signal Collectibles Ltd.	111	2052	50218.51	15
Iberia Gift Imports, Corp.	111	1814	54723.62	15
Online Mini Collectables	111	1840	57197.96	15
Clover Collections, Co.	111	1834	57756.43	16
Auto Assoc. & Cie.	111	1809	64834.32	18
Osaka Souveniers Co.	111	1990	67605.07	20
Daedalus Designs Imports	111	2041	69052.41	20
Grand Total	1332	23330	624291.59	20

Churning Customers

These are customers with lowest recency.

These customers haven't ordered in recent past with us. We should try to salvage them or find reasons as why they are no more ordering with us.

Double Decker Gift Stores has not ordered with us in recent past.

Churning Customers - Customers with lowest recency

CUSTOMERNAME	Data Sum of RFM Score	Sum of Recency	Sum of SALES	Max of ORDERNUMBER
Double Decker Gift Stores, Ltd	111	2071	36019.04	12
West Coast Collectables Co.	111	2064	46084.64	13
Signal Collectibles Ltd.	111	2052	50218.51	15
Daedalus Designs Imports	111	2041	69052.41	20
Collectable Mini Designs Co.	221	2036	87489.23	25
Grand Total	665	10264	288863.83	25

Loyalist Customers

These are set of customers who have placed most number of orders.

We should reward them for success of the business and maintain their recurrent transactions with us.

Euro Shopping channel is our most loyal customer.

Loyalist Customers - Customers with High Frequency Order

CUSTOMERNAME	Data Sum of RFM Score	Sum of Recency	Sum of SALE	Max of ORDERNUMBER
Euro Shopping Channel	444	1576	912294.11	259
Mini Gifts Distributors Ltd.	444	1578	654858.06	180
Australian Collectors, Co.	443	1760	200995.41	55
La Rochelle Gifts	444	1576	180124.9	53
AV Stores, Co.	442	1772	157807.81	51
Grand Total	2217	8262	2106080.29	259

Market Basket Analysis

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.

About Data

Name	Type	# Missing valu...	# Unique values	Minimum	Maximum	25% Quantile	50% Quantile (...	75% Quantile	Mean	Mean Absolut...	Standard Devi...	Sum	10 most com...
Date	String	0	603	?	?	?	?	?	?	?	?	?	08-02-2019 (1...
Order_id	Number (integ...	0	1139	1	1,139	292	581	862	575.986	283.458	328.557	11,888,933	226 (34; 0.16%...
Product	String	0	37	?	?	?	?	?	?	?	?	?	poultry (640; 3...

We can see there is no missing data.

We have data for 37 products.

1139 orders were placed.

Our data shows highest orders for poultry.

Exploratory Analysis

Number of Products per Order ID

We have orders with as high as 34 products and low as 3 products.
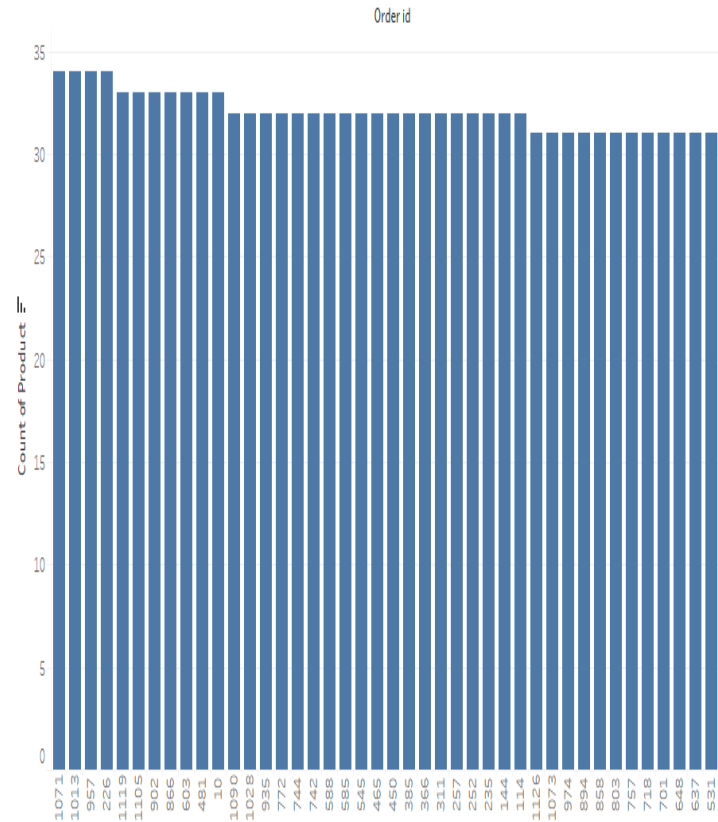
Number of Products per Order ID
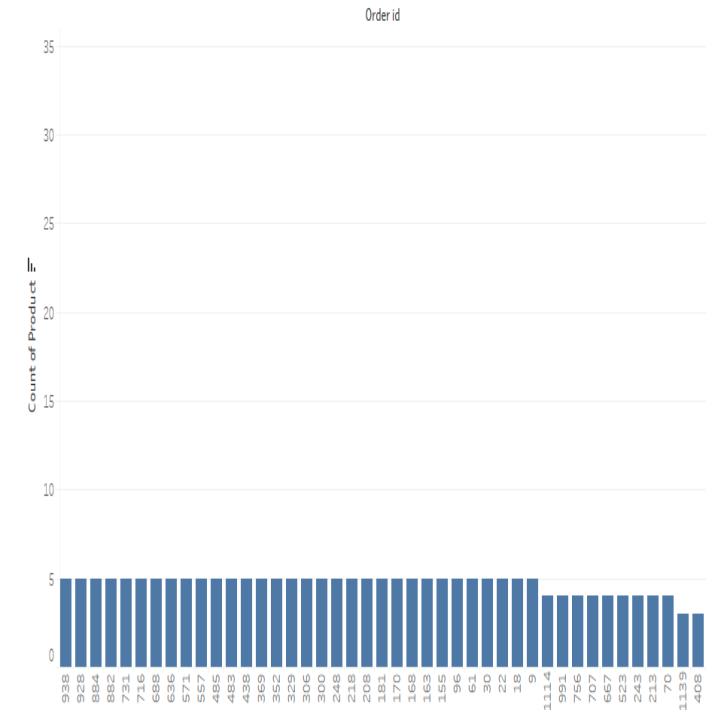
List of Products per Order

We can see that customers have ordered varied products in their orders.

They may order food items or cleaning supplies in the order.

Customers have ordered upto 34 products in single order or only 3 products as well.
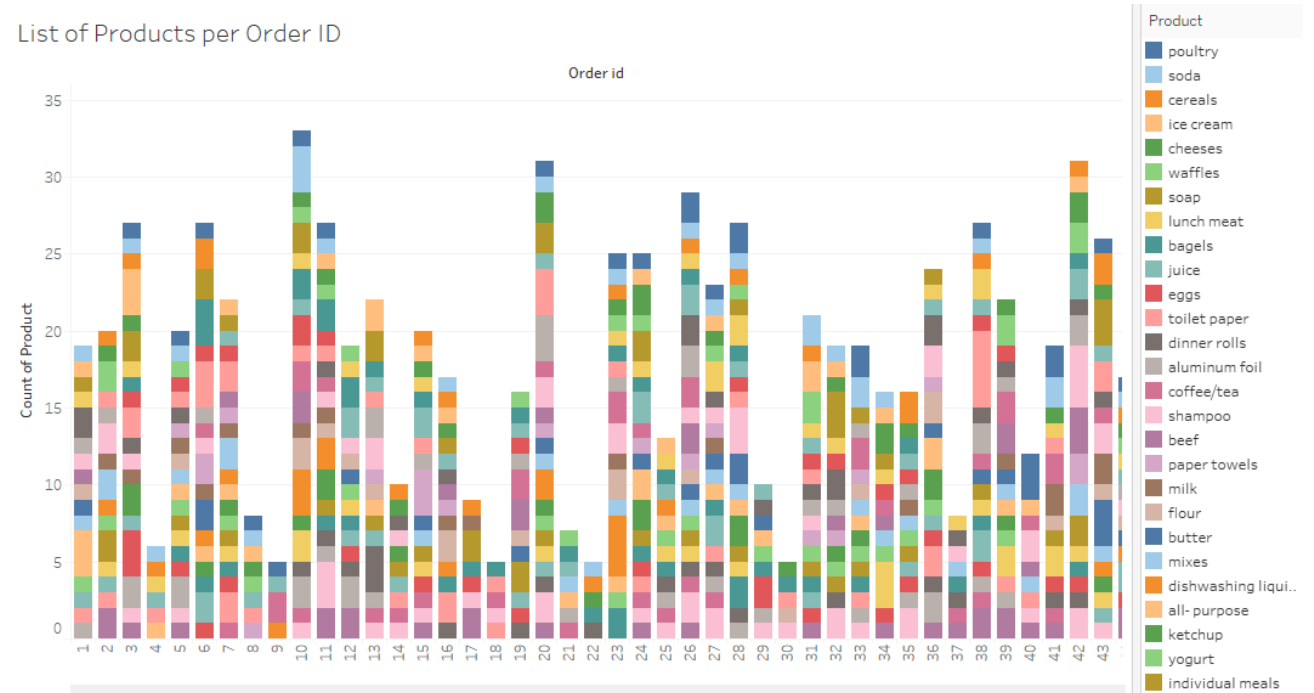
Frequency of Products Ordered

Poultry has been ordered most number of times.

Hand soap has been ordered least number of times.

Number of times Product ordered
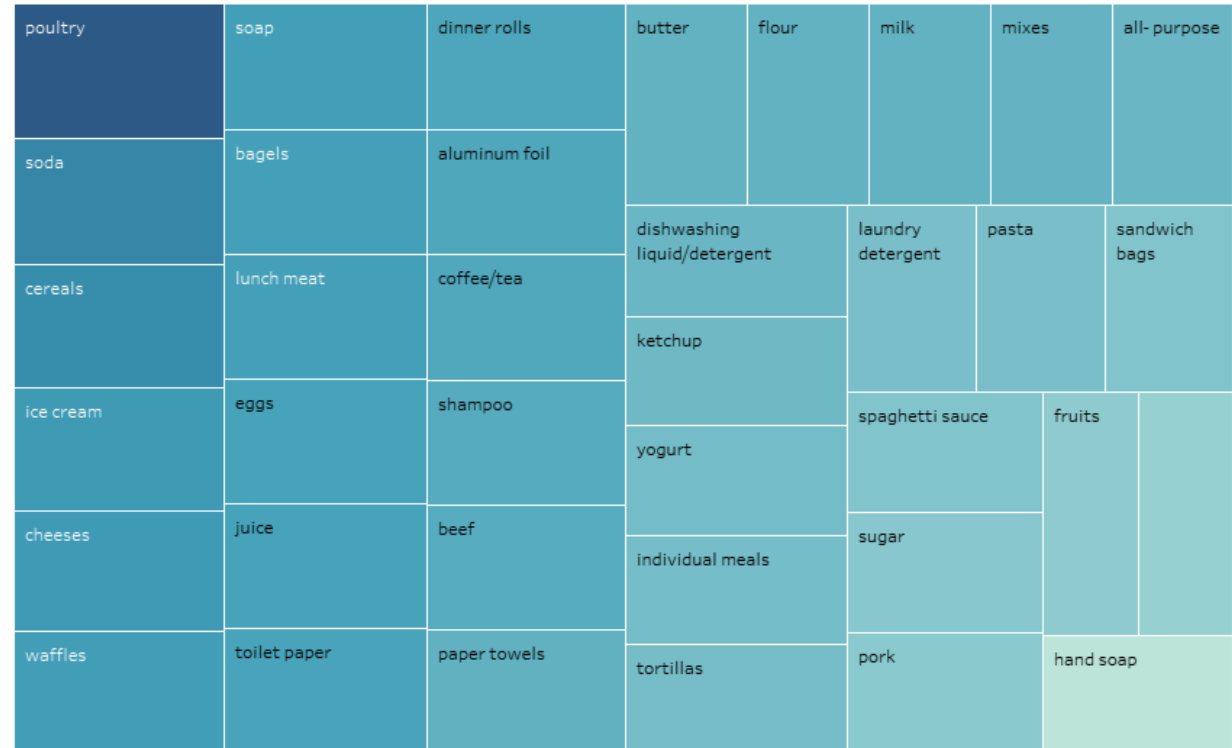
Yearly Sales

Most of our orders are for 2018, 2019 and we have least amount of data for 2020.

Sales in 2018 and 2019 are above 9k and sales in 2020 is around 2k.

Yearly Orders
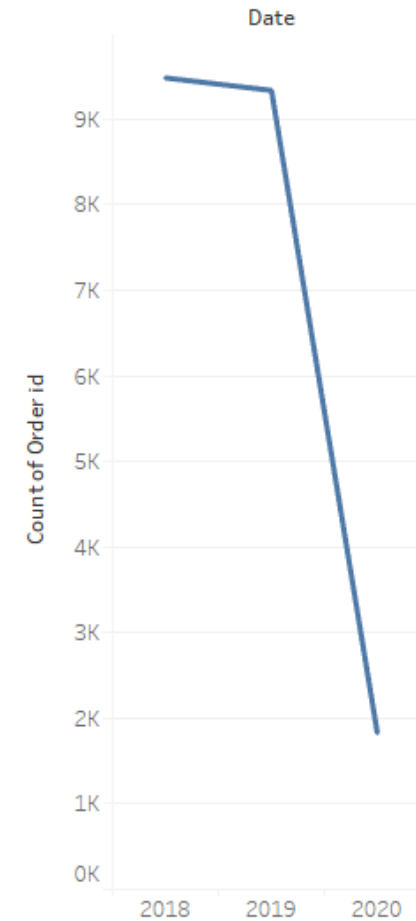
Quarterly Sales

We have data for 3 quarters of 2018 and 2019 and 1 quarter of 2020.

We have similar sales for all quarters of 2018 and 2019 and lowest sales for Q1 of 2020.

QuaarterWise Order Details

Quarter of ..	Date		
	2018	2019	2020
Q1	3,114	3,354	1,829
Q2	3,047	3,054	
Q3	3,318	2,925	

Monthly Sales

We have data of January to September for 2018 and 2019. January and February data for 2020.

Highest sales are in march 2019.

Lowest sales in February 2020.

MonthWise Order Details

| | Date | | |
Month of D..	2018	2019	2020
January	1,170	919	1,138
February	907	1,217	691
March	1,037	1,218	
April	962	975	
May	1,155	1,182	
June	930	897	
July	1,054	943	
August	1,159	961	
September	1,105	1,021	

Weekly Data

We have data for week 1 to week 40 for 2018 and 2019.

For 2020, we have data for 9 weeks.

We have highest sales in week 8 of 2019 and lowest in week 40 of 2018.

Weekwise Order Details

Week of Da..	2018	Date 2019	2020
Week 1	233	151	106
Week 2	294	298	244
Week 3	327	136	310
Week 4	158	128	290
Week 5	286	345	213
Week 6	188	317	138
Week 7	150	195	239
Week 8	267	445	224
Week 9	389	219	65
Week 10	309	247	
Week 11	175	197	
Week 12	176	327	
Week 13	162	230	
Week 14	273	364	
Week 15	176	182	
Week 16	248	194	
Week 17	230	265	
Week 18	256	280	
Week 19	221	230	
Week 20	336	141	
Week 21	245	251	
Week 22	163	387	
Week 23	240	209	
Week 24	201	116	
Week 25	286	191	
Week 26	172	282	
Week 27	243	269	
Week 28	219	221	
Week 29	299	210	
Week 30	239	198	
Week 31	176	173	
Week 32	311	232	
Week 33	221	271	
Week 34	227	290	

Week 34	227	290
Week 35	302	121
Week 36	371	207
Week 37	279	253
Week 38	209	210
Week 39	207	244
Week 40	15	107

Sales Trend

We can see that there is an slight rise in trend then there is sudden downfall in sales.
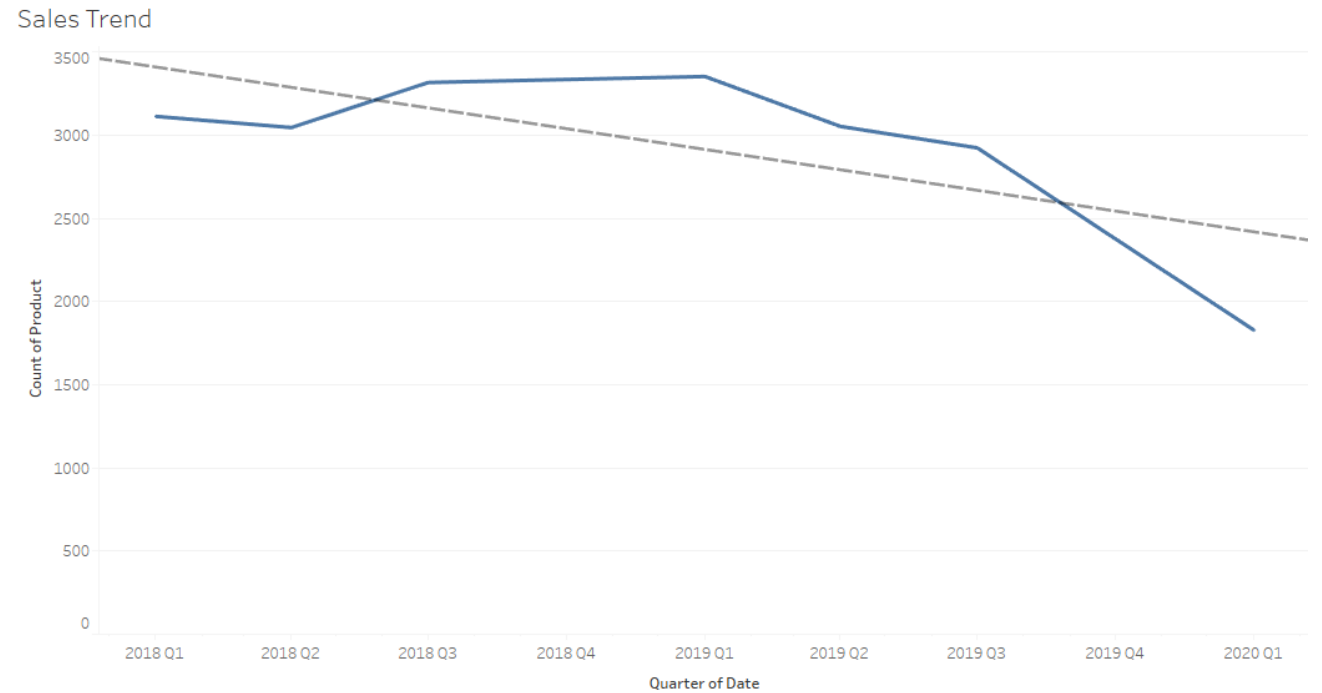
Sales Seasonality

We can see that there is downward fall from January to February in 2018 and rise for both years 2018 and 2019 in may-June.

There is again downfall in July and rise in august month

Product wise monthly sales

We have listed all products and number of times product has been in each month of 2018, 2019 and 2020.

Flour has been least ordered product - 11 times in February 2020.

Cheese has been most ordered product – 48 times in May 2018.

MonthWise Product Details - Table

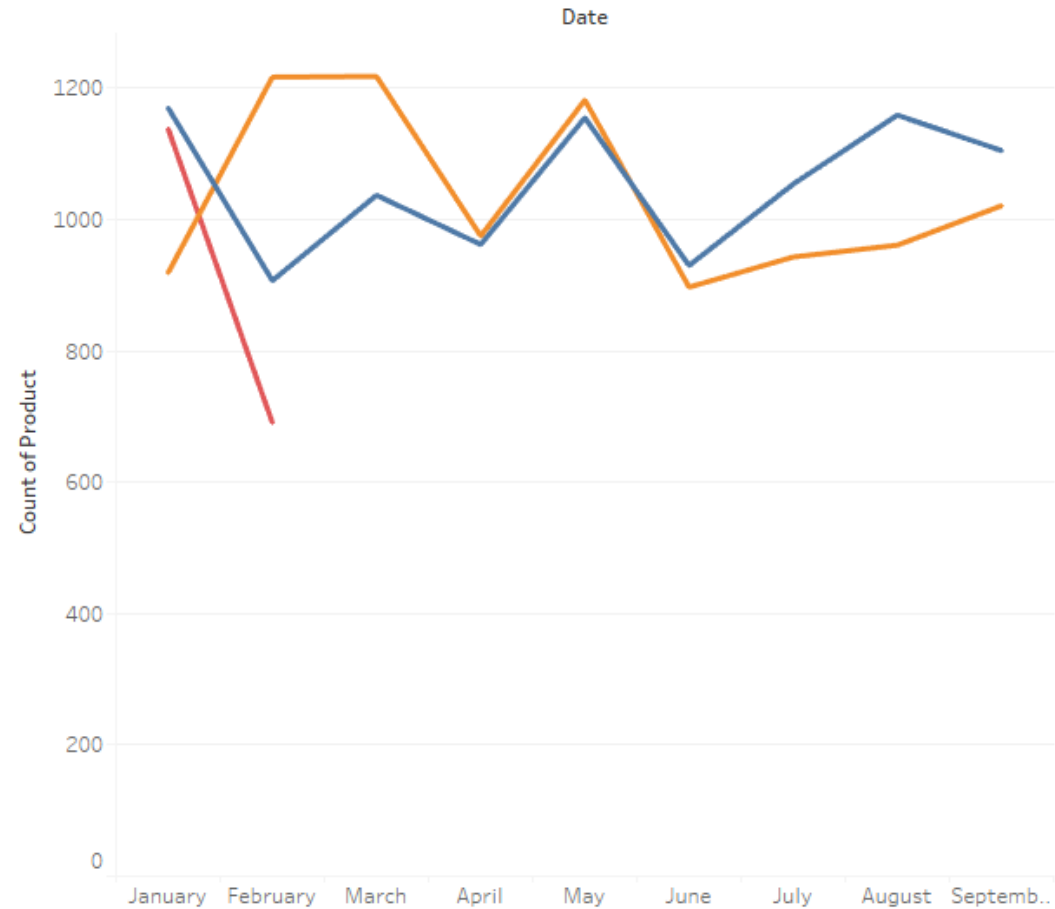| Product	2018									2019									2020	
	January	February	March	April	May	June	July	August	Septem..	January	February	March	April	May	June	July	August	Septem..	January	February
all- purpose	35	24	31	29	28	19	21	29	32	23	30	27	26	34	27	27	24	29	33	23
aluminum foil	35	21	28	26	29	30	33	30	34	22	35	31	36	24	26	22	29	33	28	14
bagels	27	32	31	33	29	26	19	36	26	35	32	42	27	27	30	21	28	24	32	16
beef	31	32	32	23	38	36	23	27	26	26	29	35	20	26	22	22	30	25	30	28
butter	28	21	28	25	38	25	22	31	35	24	26	45	18	45	23	20	18	32	34	17
cereals	33	27	32	21	48	30	27	42	31	28	38	25	31	24	19	25	32	27	34	17
cheeses	35	29	31	23	26	29	28	27	31	29	31	36	27	18	36	28	31	29	27	27
coffee/tea	33	24	31	22	30	31	33	30	31	30	23	29	28	29	33	35	22	27	31	13
dinner rolls	29	17	31	32	30	20	32	32	21	24	38	29	35	30	21	33	21	25	41	26
dishwashing liqui..	27	32	31	22	23	12	28	31	25	29	40	38	30	37	29	21	20	27	34	15
eggs	26	22	37	29	33	20	27	34	30	28	30	41	30	37	27	22	17	33	32	15
flour	30	21	31	28	29	26	44	34	37	19	33	26	27	33	23	25	15	23	40	11
fruits	25	29	29	21	32	24	28	42	26	21	30	26	27	38	26	16	30	17	26	16
hand soap	35	28	26	26	31	23	28	29	28	25	33	24	18	25	17	22	17	25	22	20
ice cream	33	24	27	31	24	29	30	37	29	25	38	32	30	23	24	30	30	25	37	21
individual meals	30	28	28	24	27	32	19	28	41	17	26	33	34	24	15	27	30	32	38	11
juice	36	20	36	29	27	17	23	40	34	31	28	38	31	43	19	25	31	20	24	18
ketchup	28	34	24	29	35	14	36	36	29	28	27	35	14	35	28	16	15	31	35	19
laundry detergent	34	13	23	18	33	24	25	37	35	28	34	37	23	28	33	18	32	27	20	20
lunch meat	32	30	20	28	31	25	42	31	30	16	38	31	24	26	26	37	28	30	30	18
milk	26	23	20	32	33	22	34	27	26	32	42	33	22	34	26	23	29	29	26	16
mixes	31	26	26	25	29	24	36	31	38	23	47	33	17	23	25	21	24	19	36	20
paper towels	26	21	27	29	34	22	24	23	29	30	38	33	25	41	24	29	18	29	36	18
pasta	30	22	28	35	29	33	32	26	30	22	31	22	31	36	18	28	21	23	31	14
pork	30	28	30	18	21	26	29	26	31	23	25	30	26	27	18	25	24	31	39	24
poultry	36	27	30	34	35	26	29	31	39	23	37	40	31	31	29	34	29	34	35	30
sandwich bags	24	25	30	20	31	30	29	27	31	19	34	36	26	36	21	19	26	25	24	23
sandwich loaves	38	21	21	24	33	26	24	22	15	30	24	30	24	28	19	34	29	26	31	24
shampoo	47	28	24	28	28	30	33	30	27	23	31	40	16	23	25	28	30	29	27	15
soap	35	16	30	33	28	22	27	43	26	23	32	40	22	40	19	28	26	35	27	22
soda	40	18	20	24	36	24	32	32	30	19	45	36	31	36	24	20	37	39	38	16
spaghetti sauce	31	21	32	36	28	25	27	27	27	19	30	41	24	33	22	25	30	21	23	14
sugar	29	28	30	18	26	36	22	23	33	25	39	24	22	30	17	37	28	33	21	12
toilet paper	39	28	28	23	34	23	26	30	29	21	26	35	31	44	31	23	21	26	26	25
tortillas	30	22	19	20	33	27	35	26	27	20	27	34	27	43	25	21	29	27	35	16
waffles	35	25	28	20	39	23	21	45	30	33	41	27	28	32	23	32	28	26	23	16
yogurt	21	20	27	24	37	19	26	27	26	26	29	24	36	39	27	24	32	28	32	21

Products Ordered Together

We have stacked products in rows and columns.

Numbers represent how many times 2 products have occurred at same time in orders.

Product Ordered Together

Product (co..	all-purpose	aluminum foil	bagels	beef	butter	cereals	cheeses	coffee/tea	dinner rolls	dishwashing..	eggs	flour	fruits	hand soap	ice cream	individual m..	juice
all-purpose	427	179	171	165	168	172	169	167	177	183	182	162	171	155	179	159	
aluminum f..	179	438	192	182	168	189	193	181	189	181	179	176	180	158	201	185	
bagels	171	192	439	178	173	185	191	173	177	191	191	155	191	150	182	188	
beef	165	182	178	427	146	192	182	161	176	179	186	154	189	164	184	168	
butter	168	168	173	146	419	171	182	163	173	170	181	155	161	150	175	168	
cereals	172	189	185	192	171	451	199	193	195	193	199	167	184	158	189	182	
cheeses	169	193	191	182	182	199	445	191	191	186	193	182	186	158	204	183	
coffee/tea	167	181	173	161	163	193	191	432	181	175	181	173	164	159	188	174	
dinner rolls	177	189	177	176	173	195	191	181	443	188	198	176	168	175	184	170	
dishwashin..	183	181	191	179	170	193	186	175	188	442	195	173	191	163	188	181	
eggs	182	179	191	186	181	199	193	181	198	195	444	170	173	164	190	175	
flour	162	176	155	154	155	167	182	173	176	173	170	402	157	151	165	162	
fruits	171	180	191	189	161	184	186	164	168	191	173	157	422	157	188	165	
hand soap	155	158	150	164	150	158	158	159	175	163	164	151	157	394	163	158	
ice cream	179	201	182	184	175	189	204	188	184	188	190	165	188	163	454	185	
individual ..	159	185	188	168	168	182	183	174	170	181	175	162	165	158	185	428	
juice	175	196	185	177	168	178	171	178	179	174	188	157	155	157	190	177	
ketchup	176	180	180	177	163	169	183	187	180	177	189	169	168	158	178	162	
laundry det..	185	183	174	161	180	193	183	166	188	186	186	166	169	163	181	173	
lunch meat	182	183	180	180	182	192	183	190	185	198	181	178	186	163	198	202	
milk	173	183	184	183	159	198	178	173	178	194	192	166	167	169	188	165	
mixes	173	196	187	165	174	198	186	169	185	205	181	171	170	173	175	170	
paper towels	178	182	175	177	164	187	172	175	177	181	189	158	171	152	194	176	
pasta	182	179	173	167	160	184	182	177	197	184	193	169	164	169	191	167	
pork	175	167	177	171	155	171	171	169	163	174	172	165	146	148	160	161	
poultry	200	200	191	194	189	206	206	199	222	213	201	179	191	176	204	194	
sandwich b..	163	172	176	178	175	180	193	174	174	175	166	156	175	158	178	164	
sandwich lo..	162	171	165	162	161	170	177	166	170	150	153	161	159	158	180	182	
shampoo	173	181	180	175	166	181	180	174	193	166	172	163	160	157	171	173	
soap	186	179	166	182	169	185	185	181	178	182	190	172	169	154	190	182	
soda	170	184	179	177	176	178	192	181	195	190	213	180	171	158	192	189	
spaghetti s..	170	178	190	180	171	190	188	165	196	167	186	151	177	165	181	180	
sugar	166	162	172	177	157	178	174	162	187	173	172	164	166	157	179	169	

Summary

- We have orders with as high as 34 products and low as 3 products.

- Customers have ordered various items in single order.

- We have data for 2018,2019 and 2020. For 2018 and 2019, we have data from January to September. For 2020, we have data for January and February only.

- While comparing monthly data of all years, Highest sales are in march 2019 and Lowest sales in February 2020.

- We can see that there is downward fall from January to February in 2018 and rise for both years 2018 and 2019 in may-June. There is again downfall in July and rise in august month.

- We have data for week 1 to week 40 for 2018 and 2019.For 2020, we have data for 9 weeks. We have highest sales in week 8 of 2019 and lowest in week 40 of 2018.

- While analyzing transactions to find products sold. Poultry has been ordered most number of times and Hand soap has been ordered least number of times.

- Flour has been least ordered product - 11 times in February 2020 and Cheese has been most ordered product – 48 times in May 2018.

KNIME Workflow
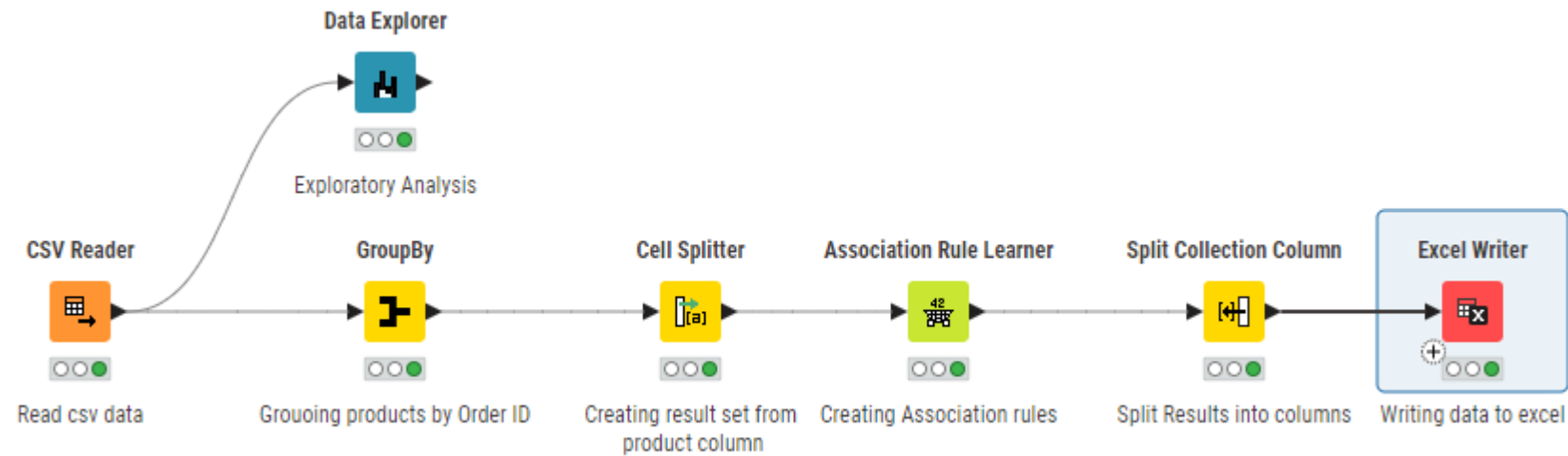
Minimum Support is 0.09. Item set length is 10. Minimum confidence is 0.5.

Association Rules

- It means relationship/patterns between different products based on order data.

- We try to identify items that have been ordered together.

- For example, People who ordered milk have also ordered bread.

- This activity helps business in multiple ways like devising promotional strategies including products in association rules or implementing product placement plan in store in order to increase product sales.

- It also helps to identify frauds and customer behavior and targeting customers with right products.

Association Rules

Various measures to find important rules:

- Support means number of times items appeared together in dataset.
- Confidence means support for 2 items compared to support of item in itemset.
- We specify minimum support and confidence and choose the rules that have support and confidence greater than threshold.
- Lift is number of times confidence of our rule meets threshold set by us.

Association Rules

RowID	Support	Confidence	Lift	Consequent	implies	Split Value 1	Split Value 2
rule0	0.090430202	0.52284264	1.33824217	soda	<---	dinner rolls	pasta
rule1	0.090430202	0.533678756	1.3721447	dinner rolls	<---	pasta	soda
rule2	0.090430202	0.528205128	1.42228284	pasta	<---	dinner rolls	soda
rule3	0.090430202	0.52284264	1.33824217	soda	<---	ice cream	waffles
rule4	0.090430202	0.50990099	1.27924499	ice cream	<---	waffles	soda
rule5	0.090430202	0.536458333	1.36085978	waffles	<---	ice cream	soda
rule6	0.090430202	0.515	1.3705257	mixes	<---	poultry	aluminum foil
rule7	0.090430202	0.525510204	1.24699192	poultry	<---	mixes	aluminum foil
rule8	0.090430202	0.504901961	1.31297565	aluminum foil	<---	ice cream	cheeses
rule9	0.090430202	0.533678756	1.33889891	ice cream	<---	cheeses	aluminum foil
rule10	0.090430202	0.512437811	1.31161049	cheeses	<---	ice cream	aluminum foil
rule11	0.090430202	0.504901961	1.31297565	aluminum foil	<---	poultry	ice cream
rule12	0.090430202	0.515	1.29203744	ice cream	<---	poultry	aluminum foil
rule13	0.090430202	0.512437811	1.21597222	poultry	<---	ice cream	aluminum foil
rule14	0.090430202	0.556756757	1.32113739	poultry	<---	dinner rolls	mixes
rule15	0.090430202	0.5	1.28846154	dishwashing liquid/deterge	<---	poultry	cereals
rule16	0.090430202	0.533678756	1.26637522	poultry	<---	dishwashing liquid/deterge	cereals
rule17	0.091308165	0.562162162	1.33396396	poultry	<---	dinner rolls	lunch meat
rule18	0.091308165	0.517412935	1.32434457	soda	<---	eggs	poultry
rule19	0.091308165	0.527918782	1.35427814	eggs	<---	poultry	soda
rule20	0.091308165	0.517412935	1.33032355	dinner rolls	<---	eggs	poultry
rule21	0.091308165	0.525252525	1.24638047	poultry	<---	eggs	dinner rolls
rule22	0.091308165	0.538860104	1.3854665	dinner rolls	<---	eggs	pasta
rule23	0.091308165	0.527918782	1.35427814	eggs	<---	dinner rolls	pasta
rule24	0.091308165	0.525252525	1.41433245	pasta	<---	eggs	dinner rolls
rule25	0.091308165	0.504854369	1.26658398	ice cream	<---	poultry	cheeses
rule26	0.091308165	0.509803922	1.30486891	cheeses	<---	poultry	ice cream
rule27	0.091308165	0.509803922	1.20972222	poultry	<---	ice cream	cheeses
rule28	0.092186128	0.552631579	1.41448847	soda	<---	eggs	soap
rule29	0.092186128	0.527638191	1.35355833	eggs	<---	soda	soap
rule30	0.092186128	0.509708738	1.31051524	dinner rolls	<---	poultry	cereals
rule31	0.092186128	0.538461538	1.27772436	poultry	<---	dinner rolls	cereals
rule32	0.093064091	0.540816327	1.4063694	yogurt	<---	juice	aluminum foil
rule33	0.093064091	0.527363184	1.37138508	aluminum foil	<---	yogurt	juice
rule34	0.093064091	0.524752475	1.39322394	juice	<---	yogurt	aluminum foil
rule35	0.094820018	0.545454545	1.39611849	soda	<---	eggs	dinner rolls
rule36	0.094820018	0.507042254	1.30365943	dinner rolls	<---	eggs	soda
rule37	0.094820018	0.553846154	1.42079002	eggs	<---	dinner rolls	soda
rule38	0.094820018	0.507042254	1.34934843	mixes	<---	dishwashing liquid/deterge	poultry
rule39	0.094820018	0.52173913	1.34448161	dishwashing liquid/deterge	<---	poultry	mixes
rule40	0.094820018	0.526829268	1.25012195	poultry	<---	dishwashing liquid/deterge	mixes
rule41	0.095697981	0.542288557	1.41019787	aluminum foil	<---	poultry	juice
rule42	0.095697981	0.556122449	1.31963223	poultry	<---	juice	aluminum foil
rule43	0.095697981	0.545	1.44698135	juice	<---	poultry	aluminum foil
rule44	0.099209833	0.579487179	1.48992302	dinner rolls	<---	spaghetti sauce	poultry
rule45	0.099209833	0.509009009	1.36414414	spaghetti sauce	<---	dinner rolls	poultry
rule46	0.099209833	0.576530612	1.3680591	poultry	<---	dinner rolls	spaghetti sauce
rule47	0.194907814	0.501128668	1.18913657	poultry	<---	dinner rolls	

Association Rules

- Rule 0: Support is 9% meaning that customers bought soda, dinner rolls and pasta . Confidence is 52% meaning that customers who bought dinner rolls and pasta also bought soda. Lift is 33% that means there is 33% chance that customer who bought dinner rolls and pasta will buy soda.

- Rule 10: Support is 9% meaning that customers bought cheeses, icecream and aluminum foil. Confidence is 51% meaning that customers who bought icecream and aluminum foil also bought cheeses. Lift is 31% that means there is 31% chance that customer who bought icecream and aluminum foil will buy cheeses.

Discount Offers and Combo Offers

Combo Offers

- Buy dinner rolls and poultry combo offers.
- Buy dinner rolls and pasta and get soda free.
- Pack and eat offer: Buy icecream and cheeses and get aluminium foil
- Dinner offer: Buy poultry and eggs and dinner rolls are on us.

Discount Offers

- 10% off on icecream if you buy poultry
- Buy eggs and soda to get dinner rolls at 15% off
- Buy dinner rolls and pasta at 20% off
- Dishwashing liquid at 10 % off on purchasing cereals.