Marketing and Retail Analytics project report

By: Ambrish Verma

Date: 6-April-2024

Project 1 – RFM Analytics

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.

Part A – Automobiles Sales Data

Data Summary:

- There are a total of 2747 records and 20 attributes in the data report.
- There are no missing value in the records for any of the attributes.
- The information of the data attributes:

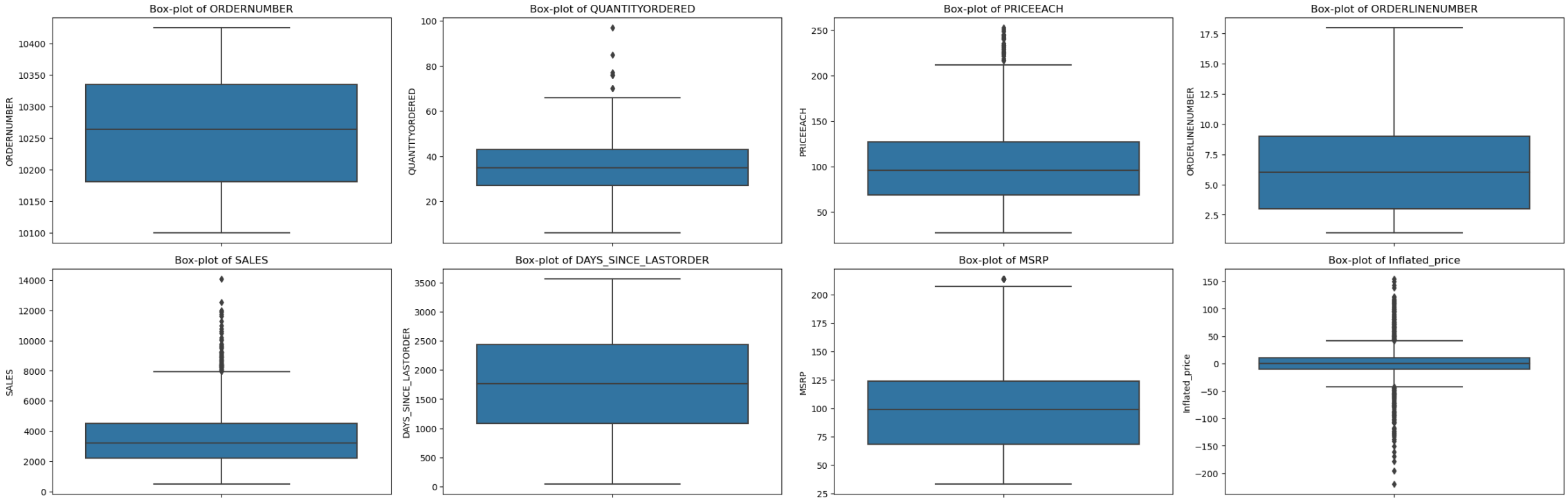
Column	Count	Non-null	Data type
ORDERNUMBER	2747	non-null	int64
QUANTITYORDERED	2747	non-null	int64
PRICEEACH	2747	non-null	float64
ORDERLINENUMBER	2747	non-null	int64
SALES	2747	non-null	float64
ORDERDATE	2747	non-null	datetime64[ns]
DAYS_SINCE_LASTORDER	2747	non-null	int64
STATUS	2747	non-null	object
PRODUCTLINE	2747	non-null	object
MSRP	2747	non-null	int64
PRODUCTCODE	2747	non-null	object
CUSTOMERNAME	2747	non-null	object
PHONE	2747	non-null	object
ADDRESSLINE1	2747	non-null	object
CITY	2747	non-null	object
POSTALCODE	2747	non-null	object
COUNTRY	2747	non-null	object
CONTACTLASTNAME	2747	non-null	object
CONTACTFIRSTNAME	2747	non-null	object
DEALSIZE	2747	non-null	object

Data Summary - Continued

- First order date: 2018-01-06
- Report generation date: 2020-06-01
- No anomaly in customer details. All the data in the attributes: Phone, Address attributes are linked to the customer's name.
- Total 89 different customers exist.
- 118 orders have status of: ‘Cancelled’, ‘On-hold’, and ‘Disputed’. These are from a total of 10 customers.
- It is assumed that if the attribute: ‘PRICEEACH’ is higher than or equal to ‘MSRP’, the sale is profitable, else it is loss-making.

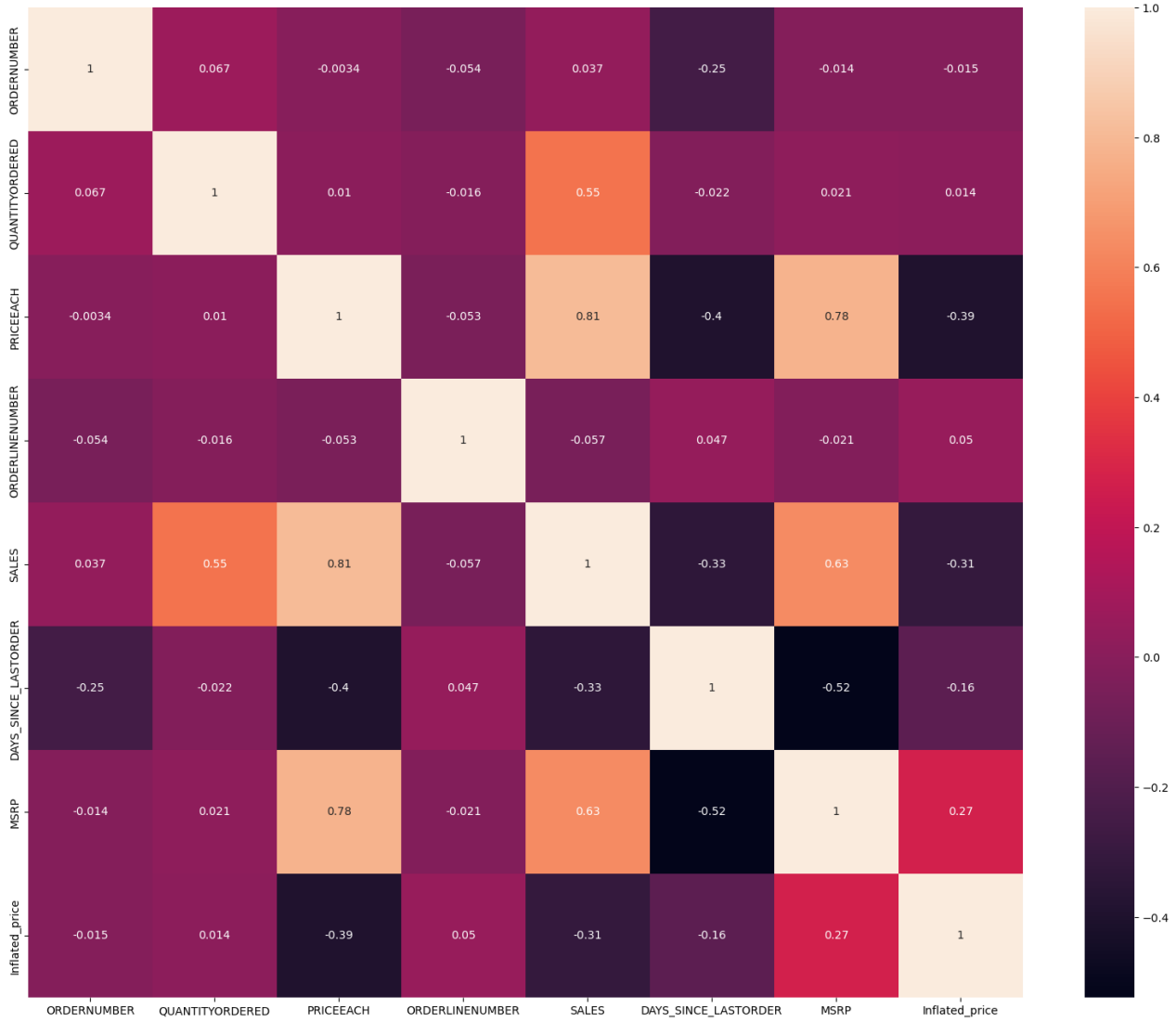
Exploratory Data Analysis – Univariate Analysis of Numeric Attributes

- The inflated price is a derived column by subtracting PRICEEACH from MSRP
- Most orders have number of each item within the range 3-9
- The difference between MSRP and PRICEEACH have significant outliers. Hence, the MSRP is not strictly followed.



Bivariate Analysis – Numeric Columns

- DAYS_SINCE_LAST_ORDER has negatively correlated With SALES. This is in line with the assumption that customer who has ordered long back is not contributing to sales.
- MSRP and PRICEEACH are highly positively correlated with SALES.
- QUANTITYORDERED is positively correlated with SALES. Hence, more the quantity more the SALES.

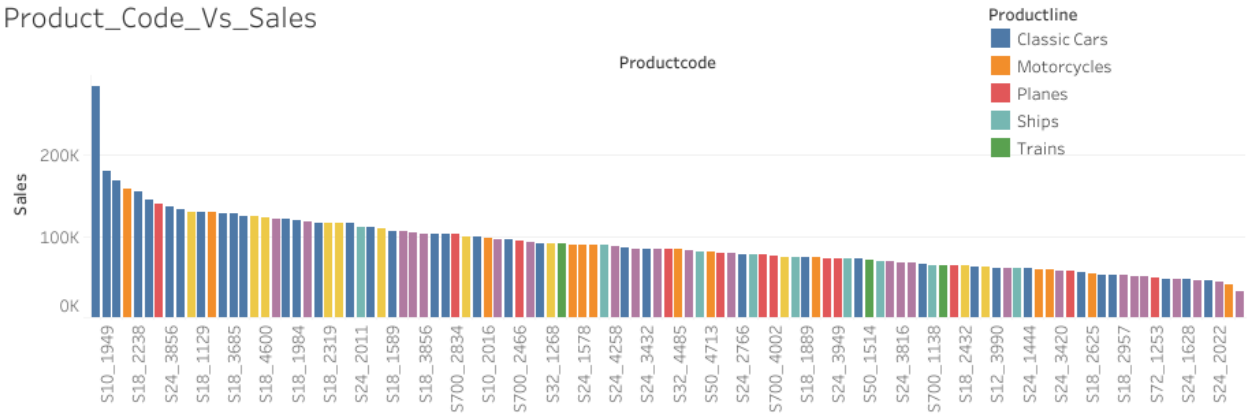


Automobile_Sales

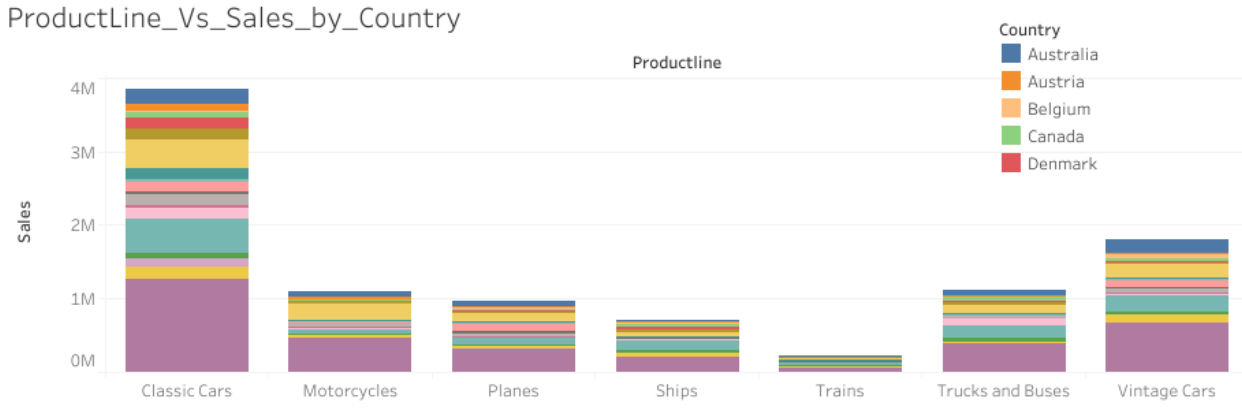
The Product code and product lines are distributed with respect to the sales made by the company. Of all the sales, the sale by Vlassic Cars are the highest, and that of trains are the lowest. Some product codes if planes are given the highest discounts and they make the most sales as well.

There is a growing trend in overall sales figure. The discounts offered mostly range between 0 and 50. The increasing sales trend is mostly refl..

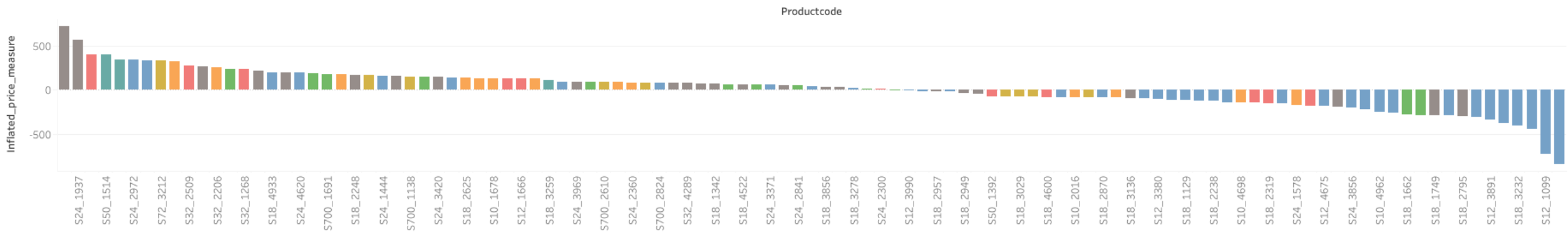
Product_Code_Vs_Sales



ProductLine_Vs_Sales_by_Country



ProductLine_Vs_InflatedPrice

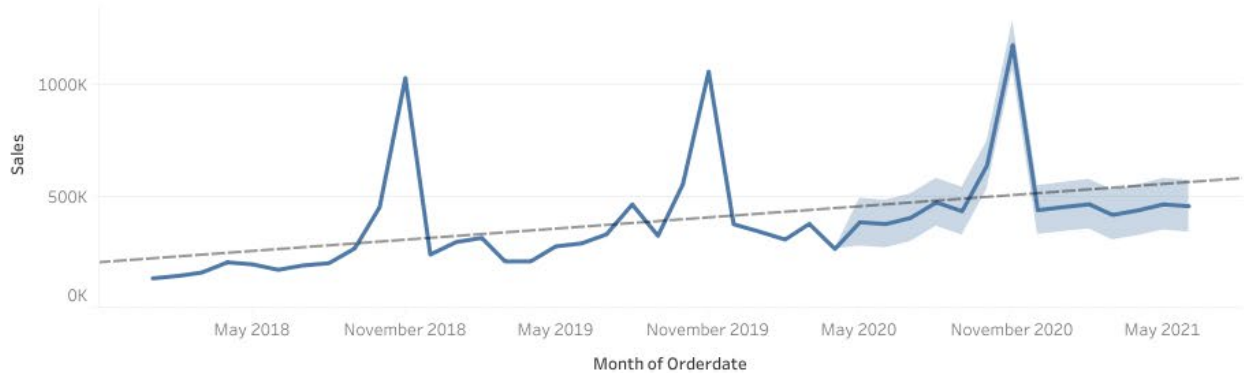


Automobile_Sales

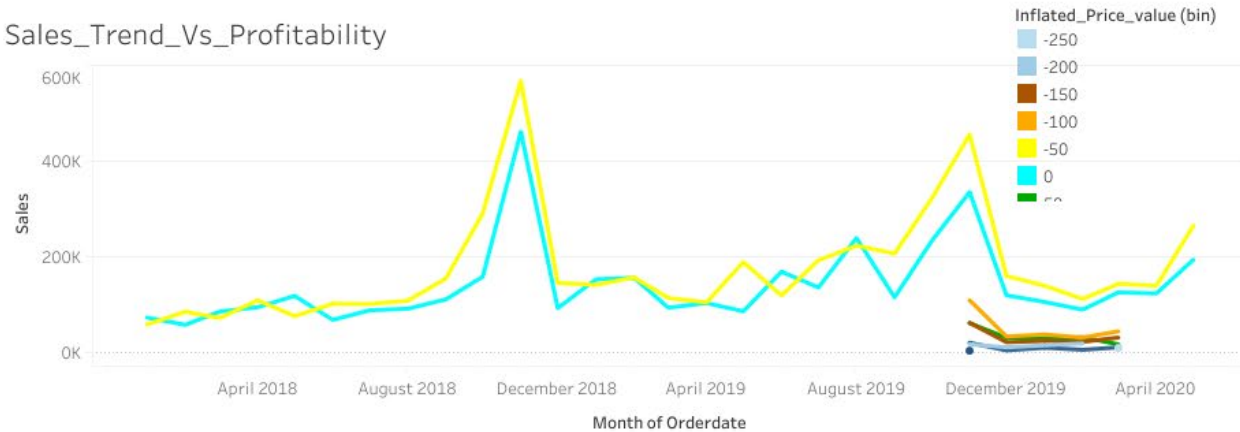
The Product code and product lines are distributed with respect to the sales made by the company. Of all the sales, the sale by Vlassic Cars are ..

There is a growing trend in overall sales figure. The discounts offered mostly range between 0 and 50. The increasing sales trend is mostly reflected in the prodct lines: Classic Cars and Vintage Cars.

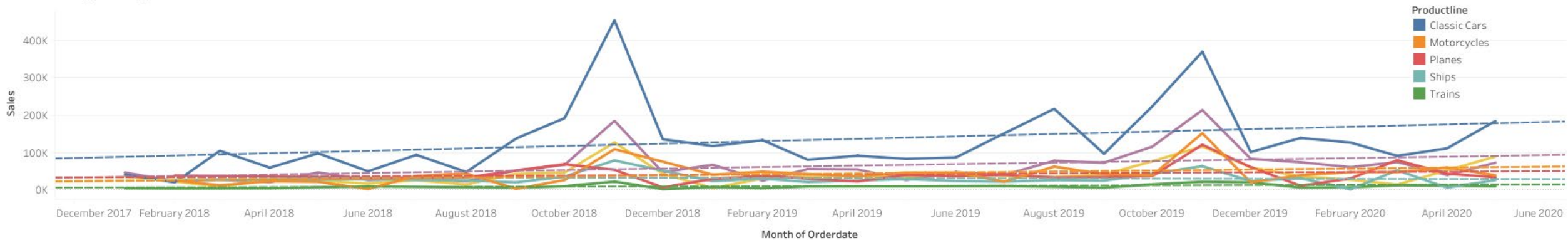
SalesTrend_Forecast



Sales_Trend_Vs_Profitability



OrderDate_Sales_ProductLine



EDA – Multivariate Analysis

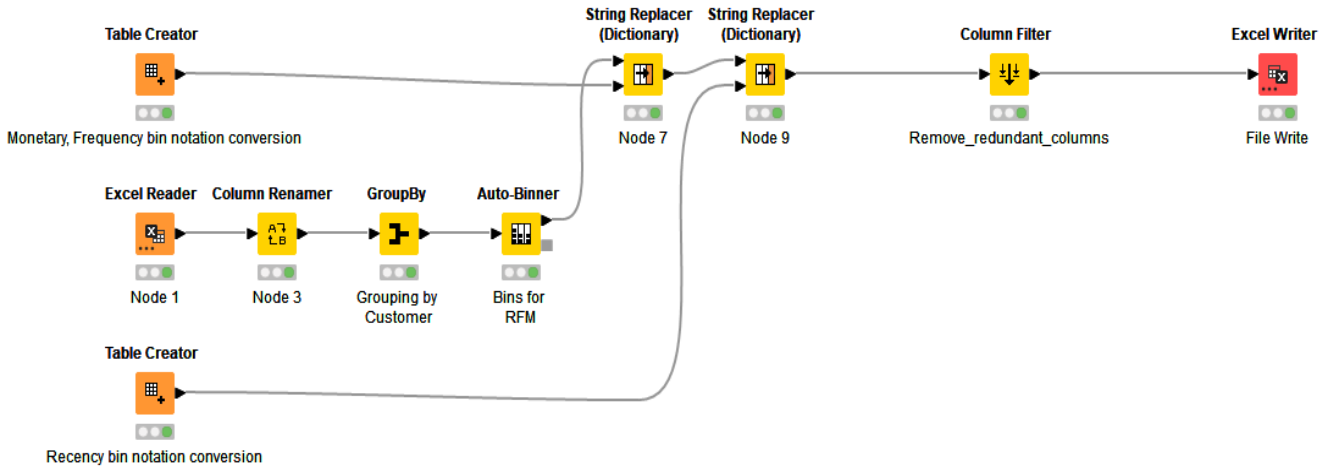
- There is a growing trend in overall sales figure.
- The discounts offered mostly range between 0 and 50.
- The increasing sales trend is mostly reflected in the product lines: Classic Cars and Vintage Cars
- The Product code and product lines are distributed with respect to the sales made by the company.
- Of all the sales, the sale by Classic Cars are the highest, and that of trains are the lowest.
- Some product codes if planes are given the highest discounts and they make the most sales as well.

RFM Analysis

- Recency Frequency Monetary Analysis is the analysis of customer transactional behaviour to understand how important the customer is for the seller.
- Recency tells when was the last time the customer transacted. Frequency is about the quantity of order and Monetary tells us about the total sale amount the customer gave the seller.
- In the current problem, the assumptions made are:
  - **DAYS_SINCE_LASTORDER** denotes the recency
  - Total Count of **ORDERNUMBER** per customer is for frequency
  - **SALES** column is the monetary field that can be used for RFM analysis
  - The above columns were then validated using the other available attributes and were found to be correct.
- For better readability, the following changes were made to the column names:
  - SALES → Monetary
  - DAYS_SINCE_LAST_ORDER → Recency
- The data was grouped on Customer Name and the bins were then created for Recency, Frequency and Monetary.
- The bins were converted in the following manner:

Quantile	Bin	F, M Values	R Value
0.0 - 0.25	Bin 1	1	4
0.25 - 0.5	Bin 2	2	3
0.5 - 0.75	Bin 3	3	2
0.75 - 1	Bin 4	4	1

Knime Workflow:



- The workflow steps are:
  - The dataset is read using the node: Excel Reader. For better readability, the Columns are renamed as mentioned in previous slide.
  - The data is grouped using groupby node to group them based on customer name
  - The grouped data are then converted into Bins. Ex: Bin 1, Bin2 etc. based on the quantile values.
  - The bins are then converted again to the codes: 1,2,3,4 with 1 being the lowest/poorest and 4 being the highest.
- The final file has the RFM bins along with the customer names to identify which is our best customer and which are the ones who can be dropped.

Automobile_Sales

There is a growing trend in overall sales figure. The discounts offer..

The following graph shows the RFM metric. RFM metric of 111 is considered the worst and 444 is the best.

Top left customers are the ones who are on the verge of churning and bot..

RFM_Metric

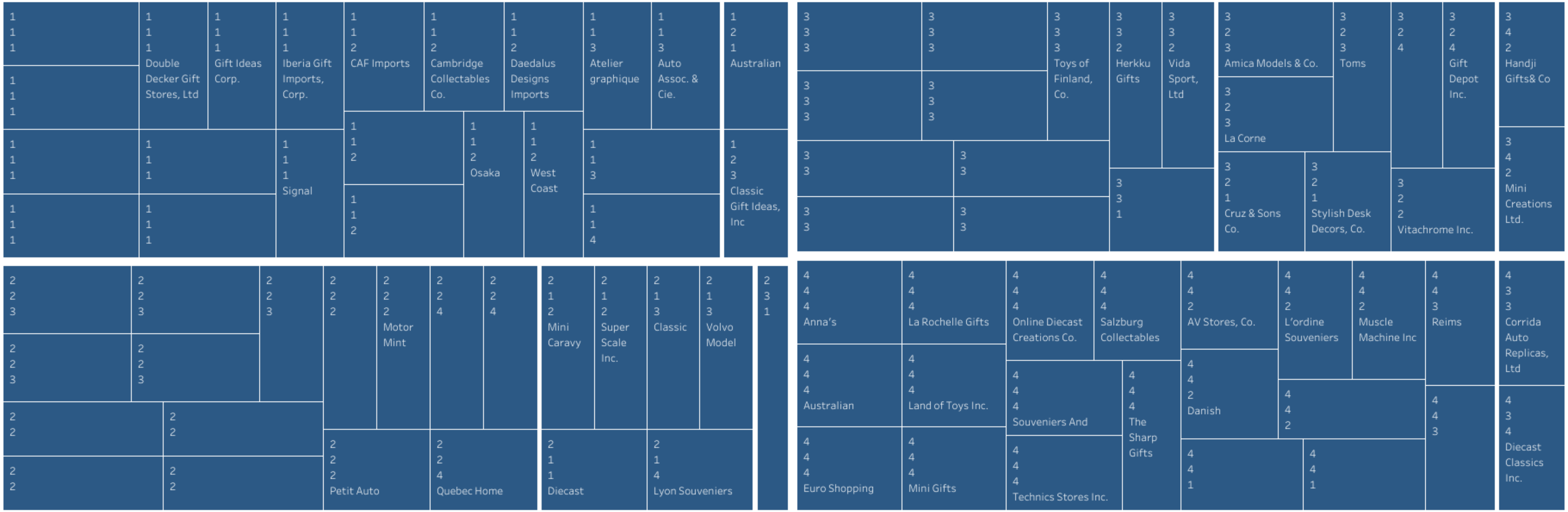
Monetary (Binned)	Recency (Binned)	ORDERNUMBER (Binned) 1	2	3	4
1	1	10	1		
	2	7			
	3	3	1		
	4	1			
2	1	1	4	1	
	2	2	3		
	3	2	5		
	4	1	3		
3	1		2	1	
	2		1	2	2
	3		3	5	
	4		2	4	
4	1				2
	2				5
	3			1	2
	4			1	11

Automobile_Sales

The following graph shows the RFM metric. RFM metric of 111 is considered the worst and 444 is the best.

Top left customers are the ones who are on the verge of churning and bottom right are the ones who are the best customers.

RFM_Metric



RFM Analysis – outcome 1

- Based on the bins, the best customers were identified as those who had the score of 444.
- The lost customer were identified as those who had a score of 111

The best customers:

- Anna's Decorations, Ltd
- Australian Collectors, Co.
- Euro Shopping Channel
- La Rochelle Gifts
- Land of Toys Inc.
- Mini Gifts Distributors Ltd.
- Online Diecast Creations Co.
- Salzburg Collectables
- Souveniers And Things Co.
- Technics Stores Inc.
- The Sharp Gifts Warehouse

The lost Customers:

- Alpha Cognac
- Auto-Moto Classics Inc.
- Bavarian Collectables Imports, Co.
- Clover Collections, Co.
- Double Decker Gift Stores, Ltd
- Gift Ideas Corp.
- Iberia Gift Imports, Corp.
- Mini Auto Werke
- Royale Belge
- Signal Collectibles Ltd.

RFM Analysis – outcome 2

- Based on the bins, the Loyal customers were identified as those who had the score of 4 for Frequency(ORDERNUMBER)
- The customer who were identified to be on the verge of churning were from the group having RFM score of as those who had a score of 333, 342, 334. (Sum of RFM scores 9 or 10)

The Loyal customers

- Dragon Souveniers, Ltd.
- Rovelli Gifts
- Handji Gifts& Co
- Mini Creations Ltd.
- AV Stores, Co.
- Danish Wholesale Imports
- L'ordine Souveniers
- Muscle Machine Inc
- Saveley & Henriot, Co.
- Reims Collectables
- Scandinavian Gift Ideas
- Anna's Decorations, Ltd
- Australian Collectors, Co.
- Euro Shopping Channel
- La Rochelle Gifts
- Land of Toys Inc.
- Mini Gifts Distributors Ltd.
- Online Diecast Creations Co.
- Salzburg Collectables
- Souveniers And Things Co.
- Technics Stores Inc.
- The Sharp Gifts Warehouse

Customers on the verge of churning:

- Heintze Collectables
- Marta's replicas Co.
- Souminen Souveniers
- Tokyo Collectables, Ltd
- Toys of Finland, Co.
- Auto Canal Petit
- Baane Mini Imports
- Oulu Toy Supplies, Inc.
- UK Collectables, Ltd
- Handji Gifts& Co
- Mini Creations Ltd.

Project 2 – 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.

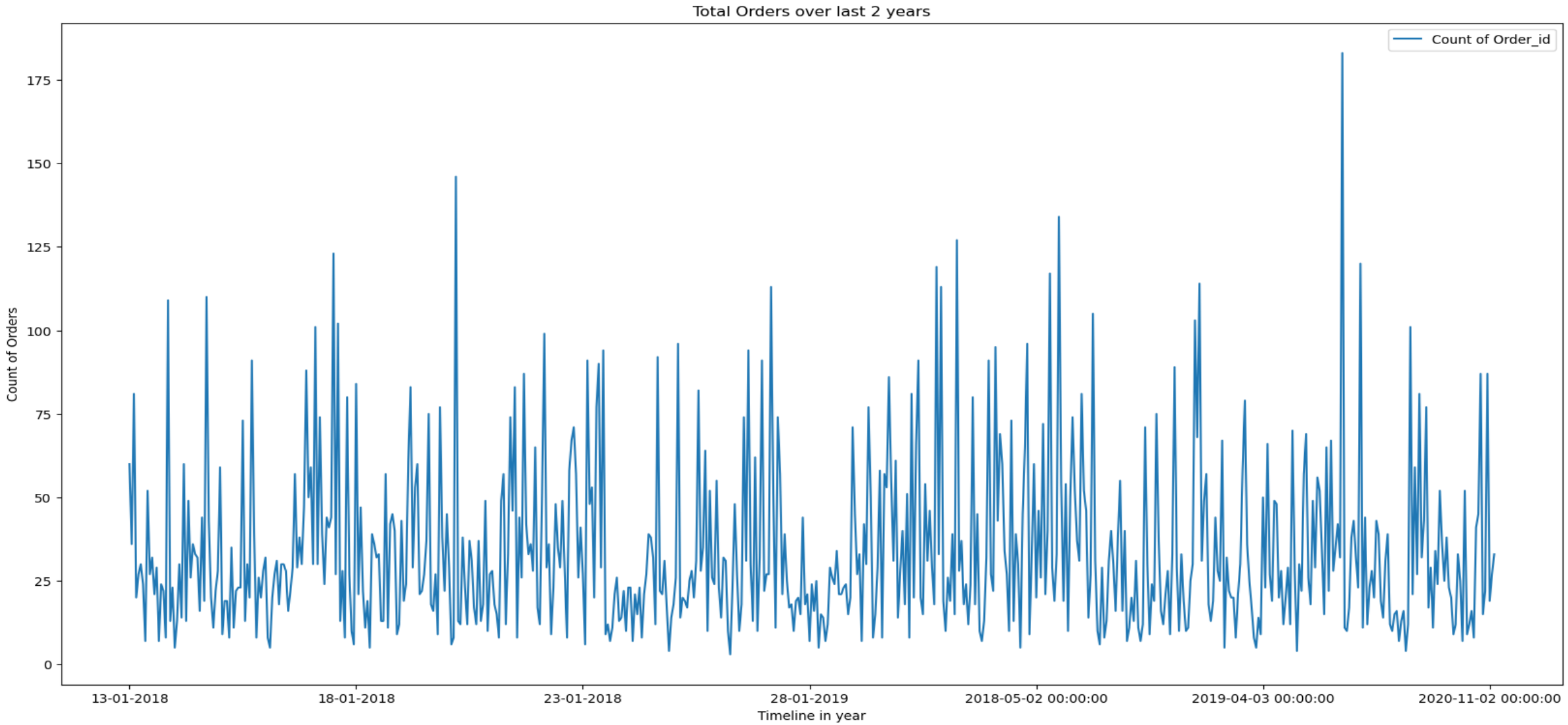
Exploratory data Analysis

- Data Summary:
  - There are a total of 20641 records and 2 attributes in the dataset provided.
  - The attributes and data-types provided are:

#	Column	Non-Null	Count	Dtype
0	Order_id	20641	Non-Null	Int64
1	Product	20641	Non-Null	object

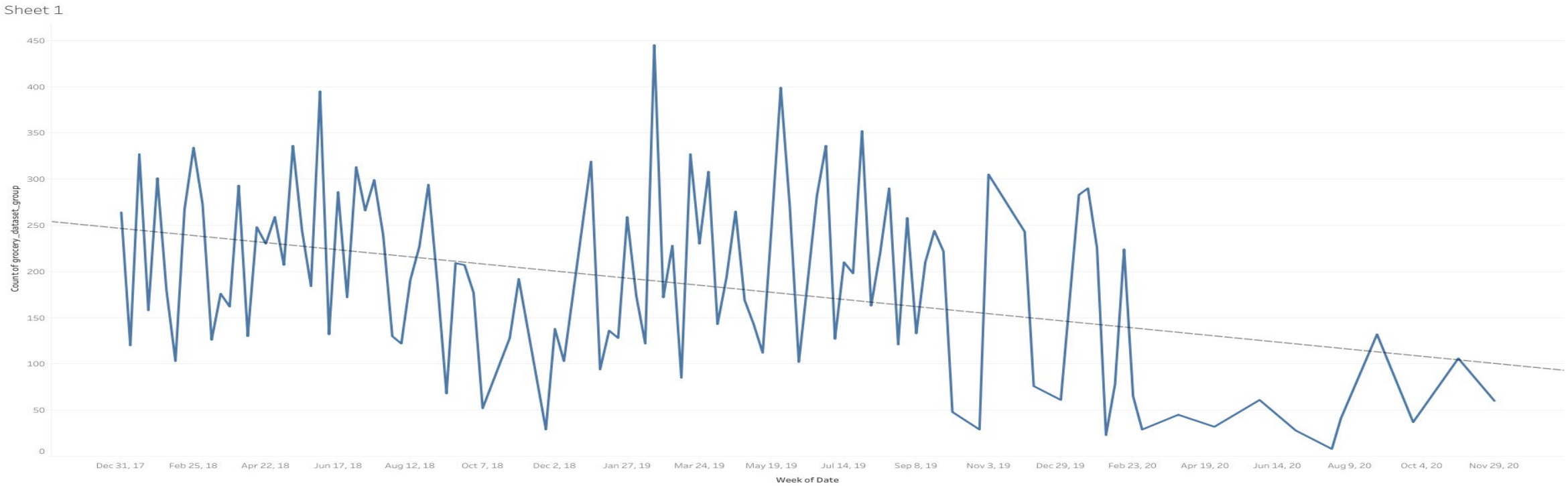
  - A Total of 37 products of different categories are sold at the store.
  - The dataset does not have a continuous data available for a thorough time-series analysis.

Order trend – day wise

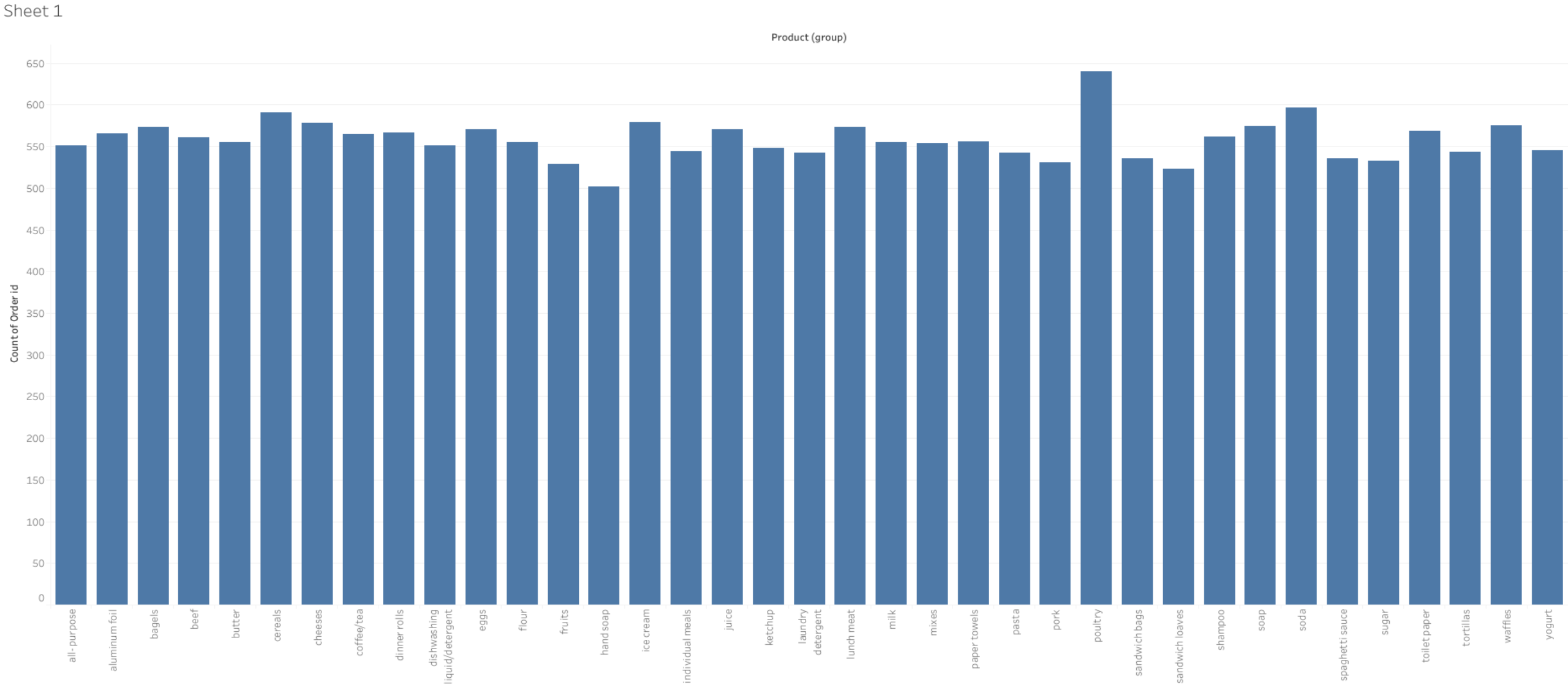


Order trend - monthly

- Downward trend of the number of orders can be observed.

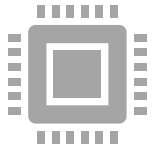


Count plot – Product group



Market Basket Analysis – Grocery Order data

For the problem, the MBA is done using KNIME tool.



The steps for MBA using KNIME are explained below:

The dataset was read using the CSV reader node.

This dataset was then grouped using the order_id as an input for the next node: cell_splitter.

In the cell_splitter, lists of products was generated.

The node 'Association Rule Learner was used to create the associations.

Minimum support value = 0.03 has been configured. Support value is a measure of how often an itemset appears in the transactions relative to the total transactions.

Minimum confidence of 0.6 has been configured. Confidence measures the likelihood that if item A is purchased, item B will also be purchased.

Knime workflow

MBA recommendation table (snippet).

Table "default" - Rows: 1270 | Spec - Columns: 6 | Properties | Flow Variables

Row ID	Support	Confide...	Lift	Consequent	implies	Items
rule17	0.031	0.795	2.194	paper towels	<---	[eggs,ice cream,pasta,lunch meat]
rule52	0.032	0.783	2.158	paper towels	<---	[eggs,ice cream,pasta,cereals]
rule18	0.031	0.729	2.066	flour	<---	[dishwashing liquid/detergent,cheeses,waffles,soda]
rule94	0.032	0.74	2.041	paper towels	<---	[eggs,dinner rolls,ice cream,pasta]
rule56	0.032	0.72	1.986	paper towels	<---	[eggs,poultry,ice cream,pasta]
rule14	0.031	0.778	1.951	ice cream	<---	[paper towels,eggs,pasta,lunch meat]
rule19	0.031	0.761	1.947	soda	<---	[dishwashing liquid/detergent,cheeses,flour,waffles]
rule136	0.033	0.717	1.931	pasta	<---	[paper towels,dishwashing liquid/detergent,eggs,ice cr
rule706	0.04	0.697	1.922	paper towels	<---	[all- purpose,individual meals,toilet paper]
rule29	0.031	0.714	1.914	spaghetti sauce	<---	[dinner rolls,poultry,laundry detergent,juice]
rule135	0.033	0.745	1.911	eggs	<---	[paper towels,dishwashing liquid/detergent,ice cream,p
rule137	0.033	0.691	1.905	paper towels	<---	[dishwashing liquid/detergent,eggs,ice cream,pasta]
rule55	0.032	0.706	1.901	pasta	<---	[paper towels,eggs,poultry,ice cream]
rule54	0.032	0.72	1.847	eggs	<---	[paper towels,poultry,ice cream,pasta]
rule93	0.032	0.685	1.845	pasta	<---	[paper towels,eggs,dinner rolls,ice cream]
rule259	0.036	0.641	1.833	sandwich loaves	<---	[all- purpose,flour,individual meals]
rule92	0.032	0.712	1.825	eggs	<---	[paper towels,dinner rolls,ice cream,pasta]
rule714	0.04	0.676	1.822	pasta	<---	[hand soap,soda,aluminum foil]
rule719	0.04	0.676	1.822	ketchup	<---	[butter,aluminum foil,soap]
rule420	0.038	0.632	1.81	sandwich loaves	<---	[paper towels,flour,individual meals]
rule749	0.041	0.671	1.808	ketchup	<---	[pork,sandwich bags,soap]
rule142	0.034	0.629	1.8	sandwich loaves	<---	[yogurt,hand soap,soap]
rule26	0.031	0.7	1.796	eggs	<---	[dishwashing liquid/detergent,ice cream,pasta,soda]
rule1116	0.046	0.65	1.793	paper towels	<---	[ice cream,pasta,lunch meat]
rule1269	0.055	0.649	1.791	paper towels	<---	[eggs,ice cream,pasta]
rule399	0.038	0.662	1.786	fruits	<---	[all- purpose,beef,lunch meat]
rule555	0.039	0.688	1.784	bagels	<---	[sandwich loaves,fruits,juice]
rule1015	0.044	0.676	1.781	soap	<---	[sandwich loaves,all- purpose,ketchup]
rule50	0.032	0.692	1.776	eggs	<---	[paper towels,ice cream,pasta,cereals]
rule116	0.033	0.667	1.774	mixes	<---	[all- purpose,hand soap,tortillas]
rule389	0.038	0.672	1.767	milk	<---	[sandwich loaves,pork,soda]
rule419	0.038	0.623	1.766	flour	<---	[pasta,mixes,coffee/tea]
rule51	0.032	0.655	1.762	pasta	<---	[paper towels,eggs,ice cream,cereals]
rule448	0.038	0.614	1.758	sandwich loaves	<---	[cheeses,hand soap,ketchup]
rule21	0.031	0.686	1.757	cheeses	<---	[dishwashing liquid/detergent,flour,waffles,soda]
rule669	0.04	0.652	1.756	ketchup	<---	[tortillas,coffee/tea,juice]
rule993	0.044	0.617	1.749	flour	<---	[yogurt,pasta,coffee/tea]
rule904	0.042	0.649	1.747	pasta	<---	[dinner rolls,hand soap,individual meals]
rule1161	0.047	0.654	1.745	beef	<---	[poultry,fruits,sugar]
rule2	0.031	0.648	1.745	ketchup	<---	[pasta,pork,soap]
rule27	0.031	0.648	1.745	pasta	<---	[dishwashing liquid/detergent,eggs,ice cream,soda]
rule552	0.039	0.647	1.742	ketchup	<---	[sandwich loaves,milk,coffee/tea]
rule1069	0.045	0.614	1.741	flour	<---	[cheeses,soda,ketchup]
rule1102	0.045	0.607	1.738	sandwich loaves	<---	[bagels,cereals,juice]
rule1186	0.047	0.659	1.736	soap	<---	[paper towels,all- purpose,ketchup]
rule987	0.043	0.645	1.736	ketchup	<---	[toilet paper,mixes,soap]
rule22	0.031	0.673	1.734	dishwashing liquid/detergent	<---	[cheeses,flour,waffles,soda]
rule133	0.033	0.691	1.733	ice cream	<---	[paper towels,dishwashing liquid/detergent,eggs,pasta
rule700	0.04	0.657	1.733	soap	<---	[all- purpose,flour,ketchup]
rule1202	0.047	0.628	1.732	paper towels	<---	[eggs,dinner rolls,ice cream]
rule1146	0.046	0.658	1.731	milk	<---	[paper towels,dishwashing liquid/detergent,spaghetti s
rule1268	0.055	0.643	1.731	pasta	<---	[paper towels,eggs,ice cream]
rule317	0.037	0.656	1.73	soap	<---	[spaghetti sauce,all- purpose,sandwich bags]
rule489	0.039	0.638	1.729	shampoo	<---	[milk,beef,lunch meat]
rule15	0.031	0.673	1.727	eggs	<---	[paper towels,ice cream,pasta,lunch meat]
rule1252	0.051	0.674	1.726	cheeses	<---	[bagels,cereals,sandwich bags]
rule1036	0.044	0.625	1.724	paper towels	<---	[ice cream,cereals,juice]
rule1228	0.048	0.625	1.724	paper towels	<---	[eggs,ice cream,cereals]
rule849	0.042	0.64	1.723	ketchup	<---	[milk,coffee/tea,pork]

Explanations of the recommendation table:

- The Column Items is the list of the products that are ordered.
- Market Basket Analysis recommends the items in the field ‘Consequent’ .
- Lift signifies the chances of the customer buying the consequent item if s/he has bought the items in the ‘Item’ bucket.
- Ex: For the first row, the chances of the customer buying the item: ‘paper towels’ are higher by 2.194 times if s/he has bought the items as mentioned in the column: ‘Item’.
- The support for this row is 0.031. This means that for the items in this bucket(e.g. egg, the number of transactions containing egg /total number of transactions = 0.031)
- Confidence for the first row is 0.795 indicates that there is 79.5% probability that if the customer has bought egg, ice cream, pasta, lunch meat, then s/he will purchase paper towel as well.

Suggestions

- Paper towels, eggs, ice cream, soda, pasta are among the most prominent recommendations in the analysis. Hence, these items should be placed very near to the billing counter.
- Also, for the above, giving a discount on the above group of items also will entice the customers to add the additional items for purchase/ Ex: Egg and paper towel can be used for a combo offer.
- Sandwich loaves are the item of choice for customers wo are buying the individual meals. Hence, giving a discount on sandwich loaves when one buys individual meals is a good recommendation.