PGPDSBA JULY 2021

MRA Project Milestone 2

By.
S.VISHAL
PGPDSBA
JULY 2021

Contents of the presentation:

- Problem Statement
- Synopsis–
- EDA – Inferential analysis
  - Sales Data
  - Descriptive
  - Summary
  - Trend - Yearly, Quarterly, Monthly, and Weekly
- Market Basket Analysis – Overview
- Market Basket Analysis on Grocery data
  - Association Rules
  - KNIME workflow
  - Output from Association Rule
  - Recommendations

Problem Statement:

A Grocery Store shared the transactional data with you.

Your job is to identify the most popular combos that can be suggested to the Grocery Store chain after a thorough analysis of the most commonly occurring sets of items in the customer orders.

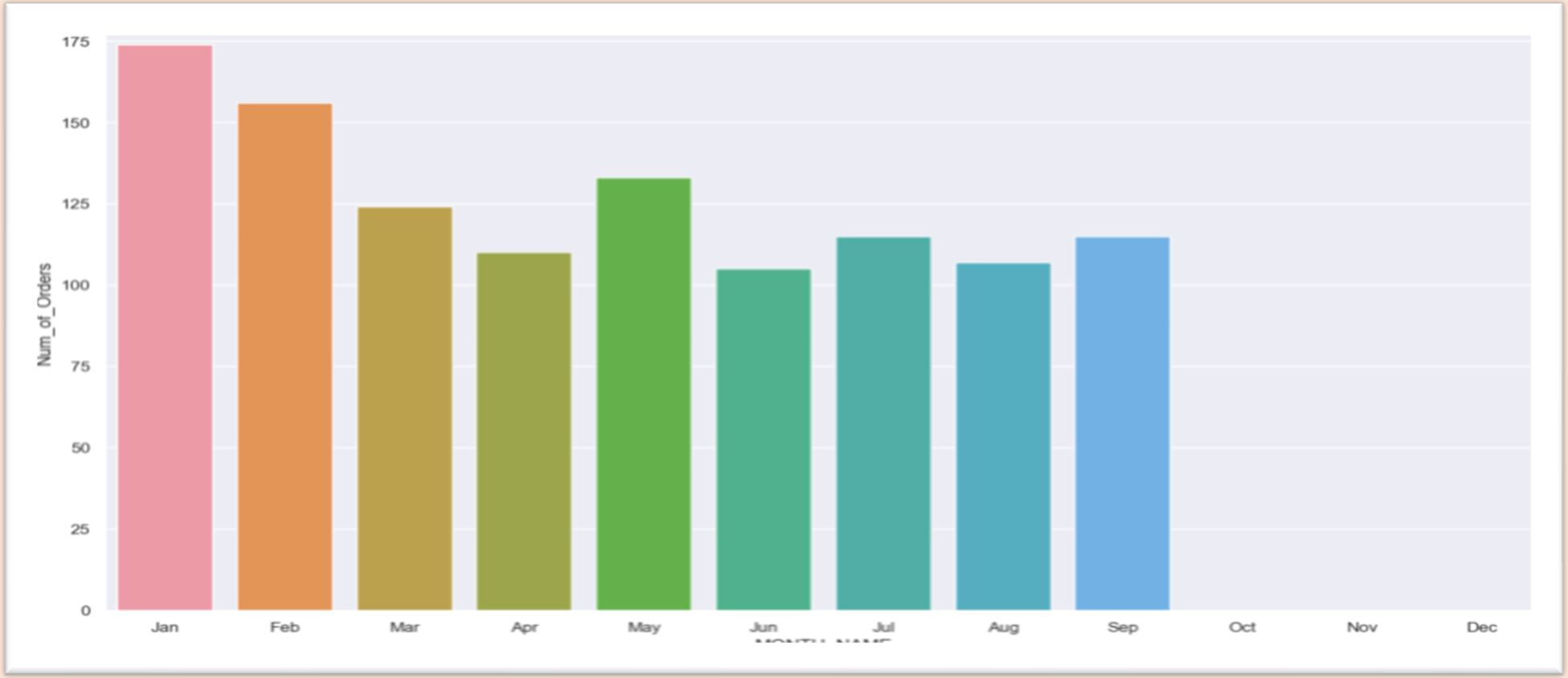
The Store doesn’t have any combo offers. Can you suggest the best combos & offers?

Synopsis:

- Total No. of Sales records = 20641
- 3 No. of Variables
- Zero missing entries
- Zero duplicate entries
- Date in the dataset is from 1st Jan 2018 to 26th Feb 2020
- 603 days of data
- 37 range of products
- 1139 number of invoices
- Max. transaction on 8th Feb 2019 with 183 transactions
- Avg. number of daily orders = 2
- Avg. number of products per order = 18

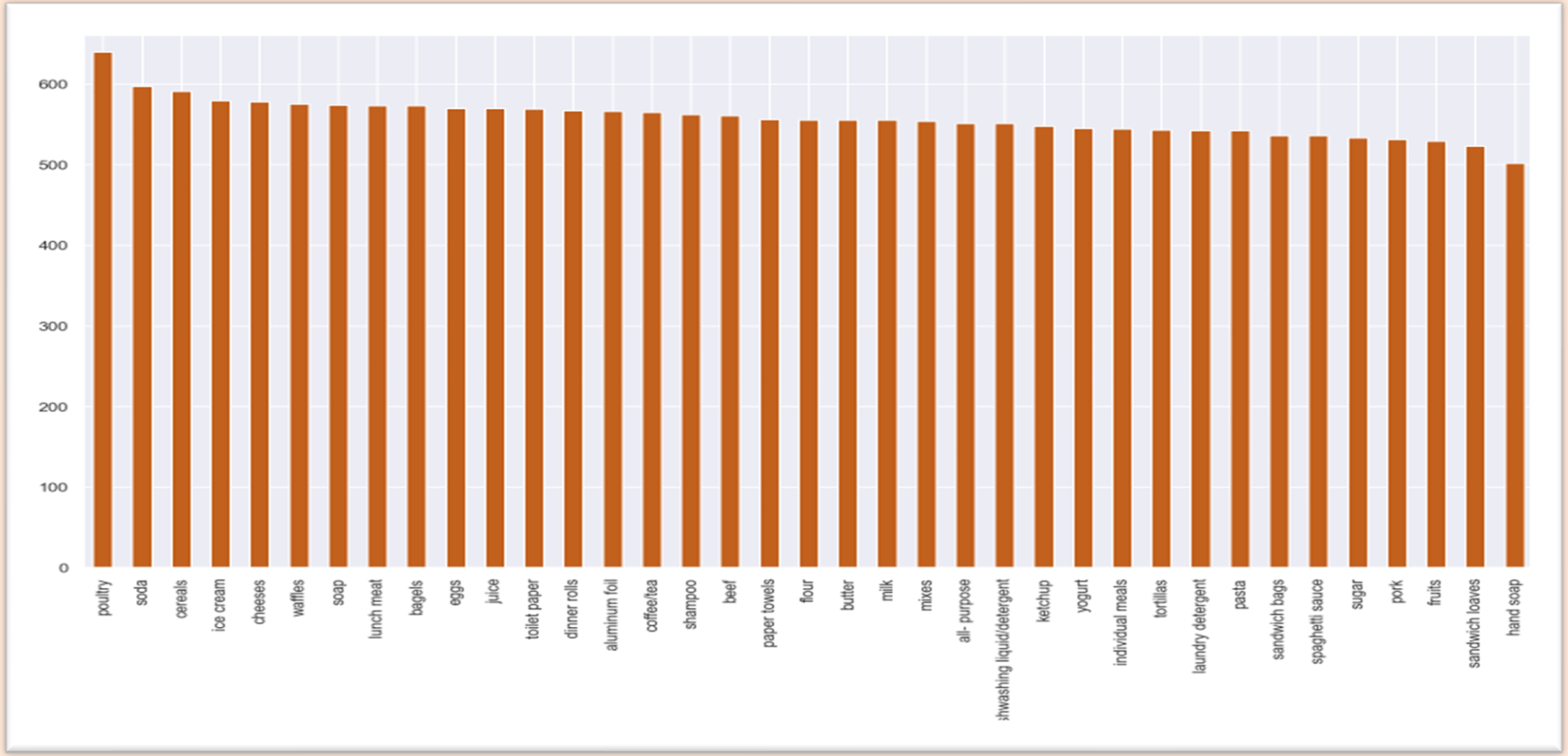
Synopsis contd...

- Store has not submitted the data for Oct, Nov, and Dec months
  Or
- Stores remains closed during 3 months period – Oct, Nov, and Dec
- Maximum transactions are observed in the month of Jan followed by Feb month



Synopsis contd...

- All products are ordered with a similar amount in frequencies
- Poultry is ordered with the highest of 640 transactions
- Hand soap is the product with the least order



Synopsis contd...

- Tools used in this project
  - Python (Jupiter Notebook) – for basic EDA and Time series
  - KNIME – for Market Basket Analysis
  - Excel – for summary pivot and views

- Sales Data:

Date	Order_id	Product
01-01-2018	1	yogurt
01-01-2018	1	pork
01-01-2018	1	sandwich bags
01-01-2018	1	lunch meat
01-01-2018	1	all- purpose
01-01-2018	1	flour
01-01-2018	1	soda
01-01-2018	1	butter
01-01-2018	1	beef

- Summary:

```
<class 'pandas.core.frame.DataFrame'>
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)
```

- Data Description

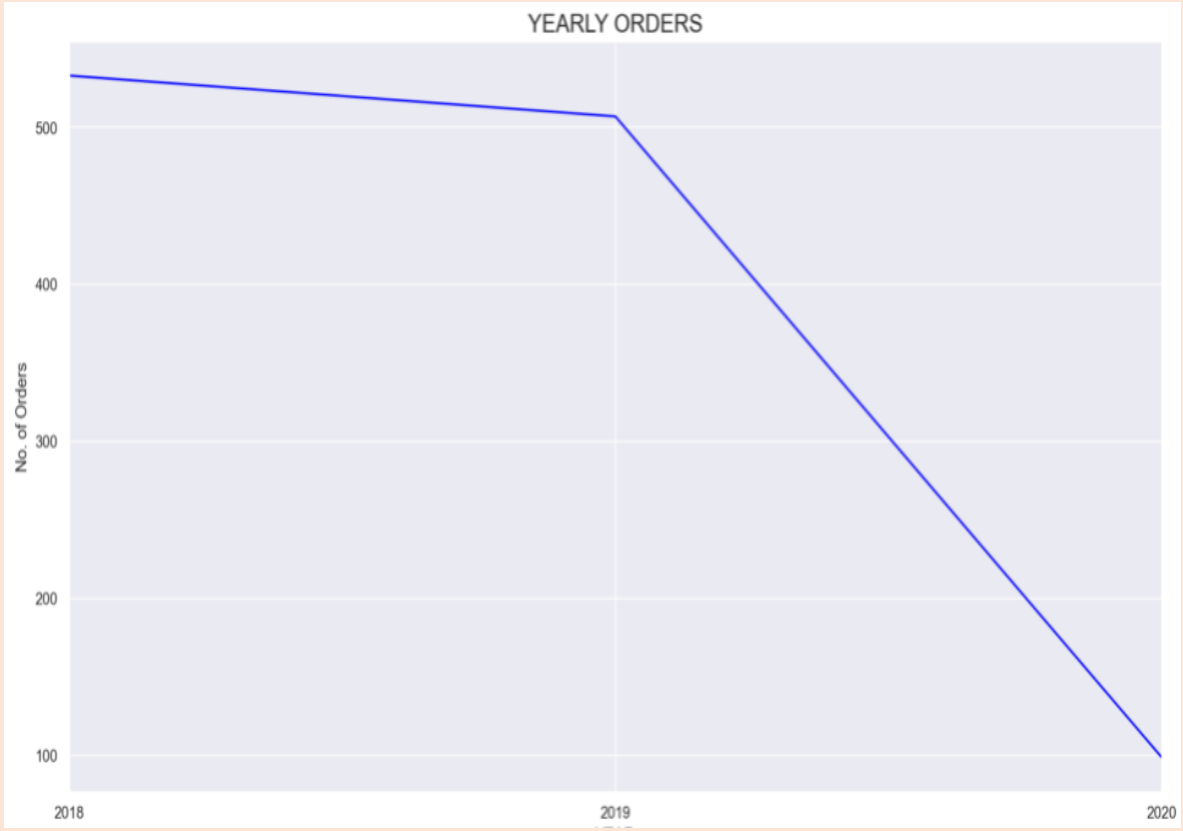
	count	mean	std	min	25%	50%	75%	max
Order_id	20641.0	575.986289	328.557078	1.0	292.0	581.0	862.0	1139.0

	count	unique	top	freq
Product	20641	37	poultry	640

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20641 entries, 0 to 20640
Data columns (total 3 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Date      20641 non-null  datetime64[ns]
 1   Order_id  20641 non-null  int64
 2   Product   20641 non-null  object
dtypes: datetime64[ns](1), int64(1), object(1)
memory usage: 483.9+ KB
```

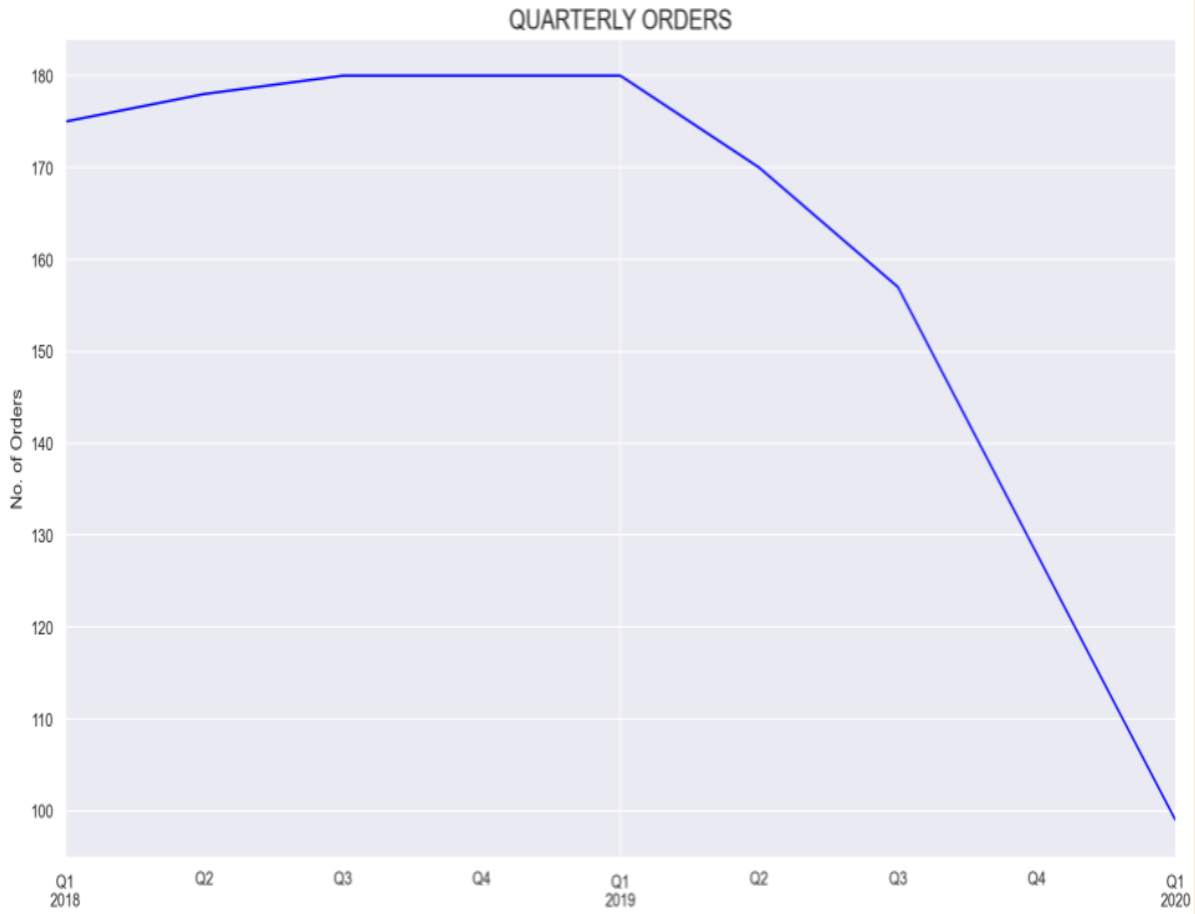
Yearly Trend:

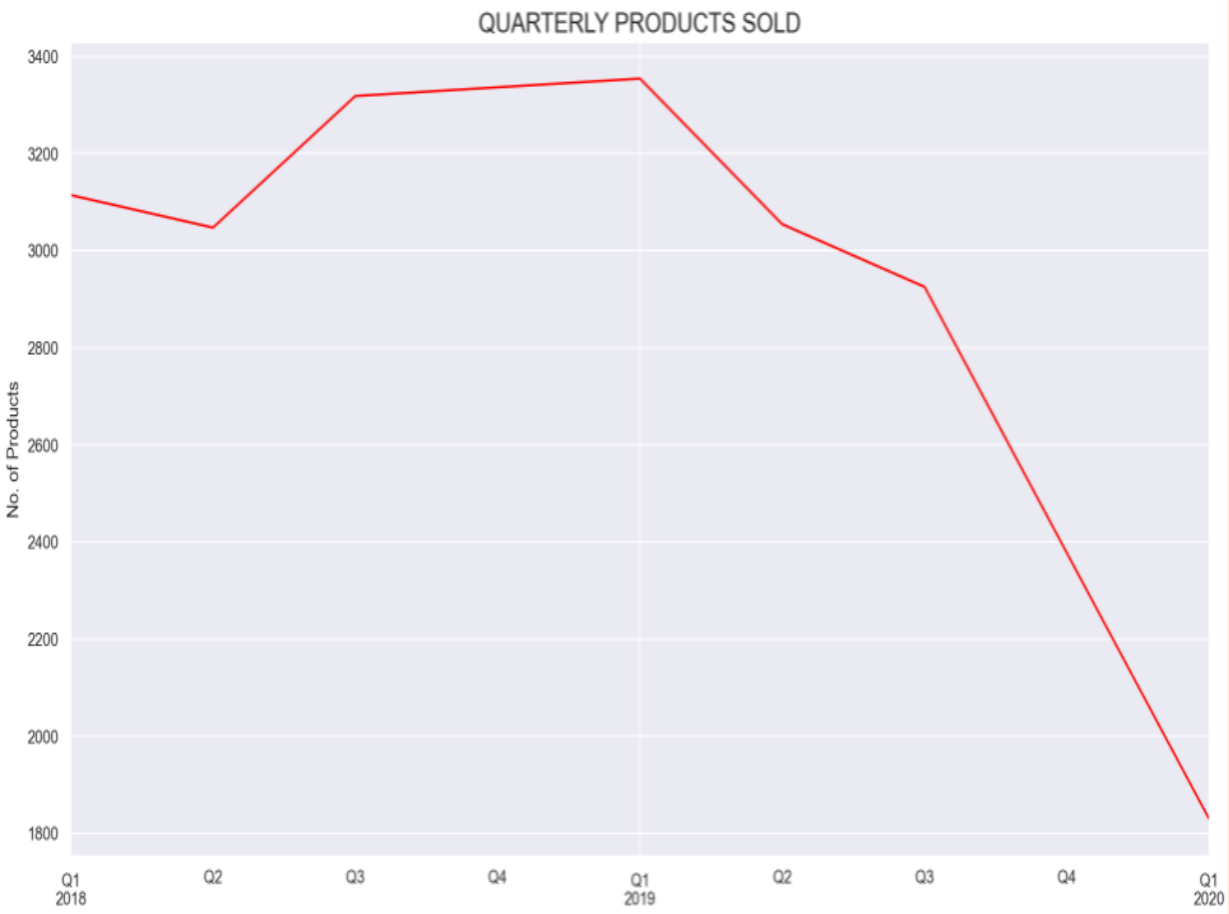
- No. of transactions and no. of products sold is gradually decreased from 2018 to 2019
- Please note: the data is available only for first two months for the year 2020



Quarterly Trend:

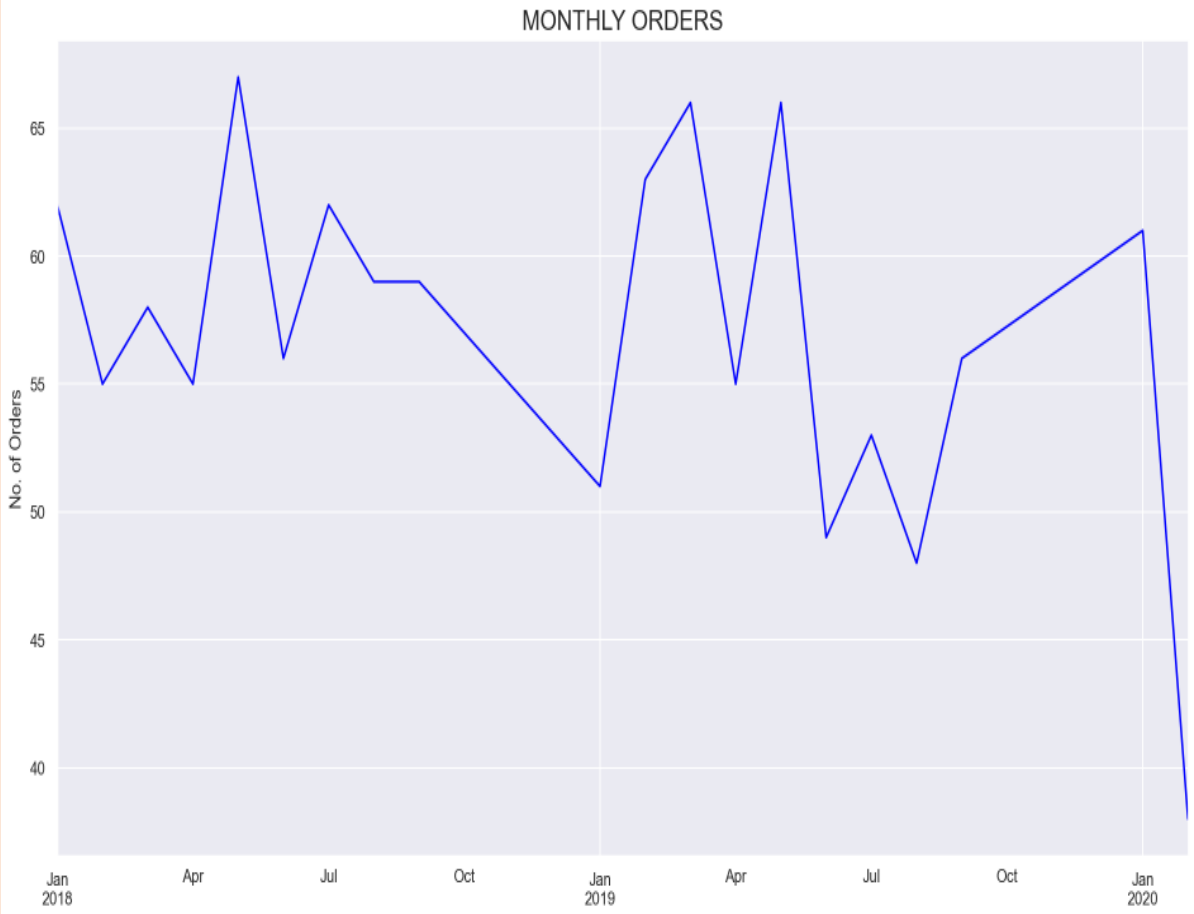
- No. of orders are increased from Q1 to Q3 in 2018 but decreased from Q1 to Q4 in 2019
- Products sold is increased from Q2 to Q3 in 2018 but decreased from Q2 to Q3 in 2019



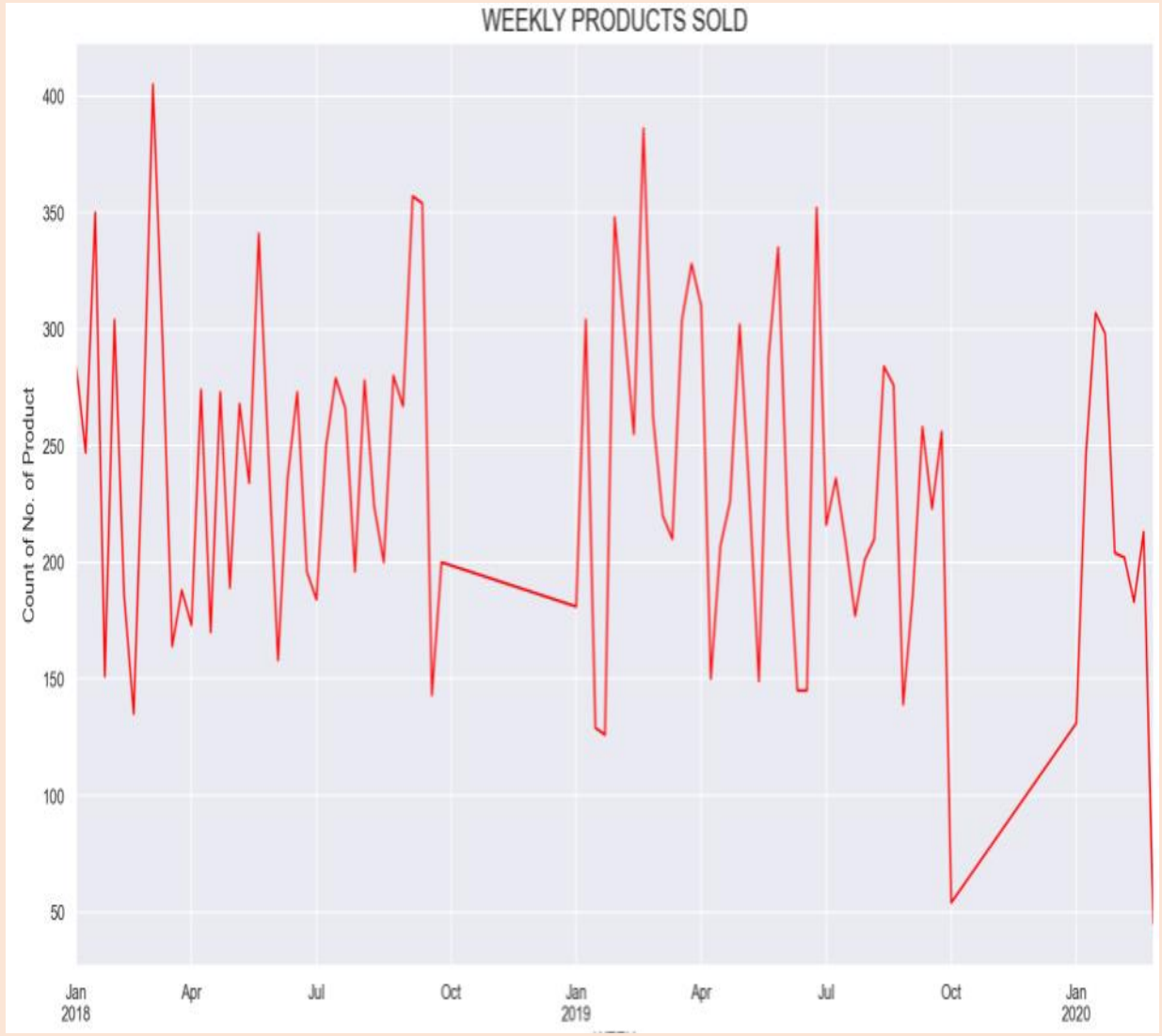


Monthly Trend:

- No exact pattern was found however we can notice that high orders places in May 2018 and 2019

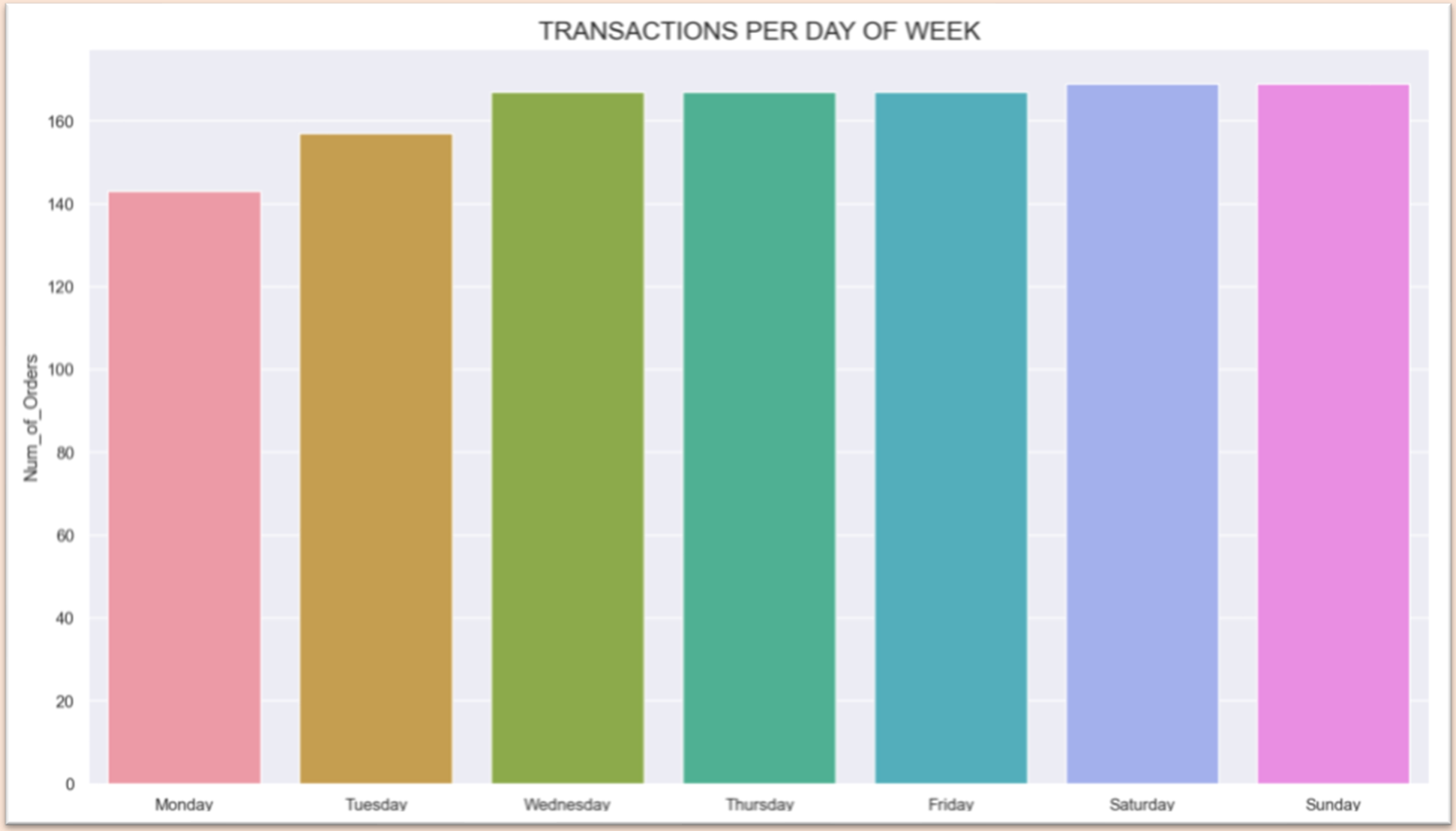


Weekly Trend:



Transactions per day of the week:

- It is seen that Monday and Tuesday has the lowest transactions



Market Basket Analysis:

- MBA is a strategy adopted by the retailers to gauge customer buying pattern
- This investigates a group of item customer ends up buying together
- This strategy finds the relationship between the items in a customer’s shopping cart based on the frequencies
- This level of strategy helps in understanding the right items grouped which can be recommended based on the shopping basket acts as a recommendation system
- Association rule algorithm is used as a data mining technique in MBA
- This tries to associate different items in a shopping cart with the other products purchased along with the basket products
- It acts as a recommendation system to the customers also gives us the results in analyzing the associations

Support, Confidence, and Lift:

- SUPPORT –
  - Support of A – is the fraction of transactions of A out of the total transactions
  - If item A is bought 100 times out of the total 1000 transactions of the store, then Support of A = 100/1000 = 0.1 (10%)
  - Similarly, if items A and B are together bought 50 times, then Support of A and B = 50/1000 = 0.05 (5%)

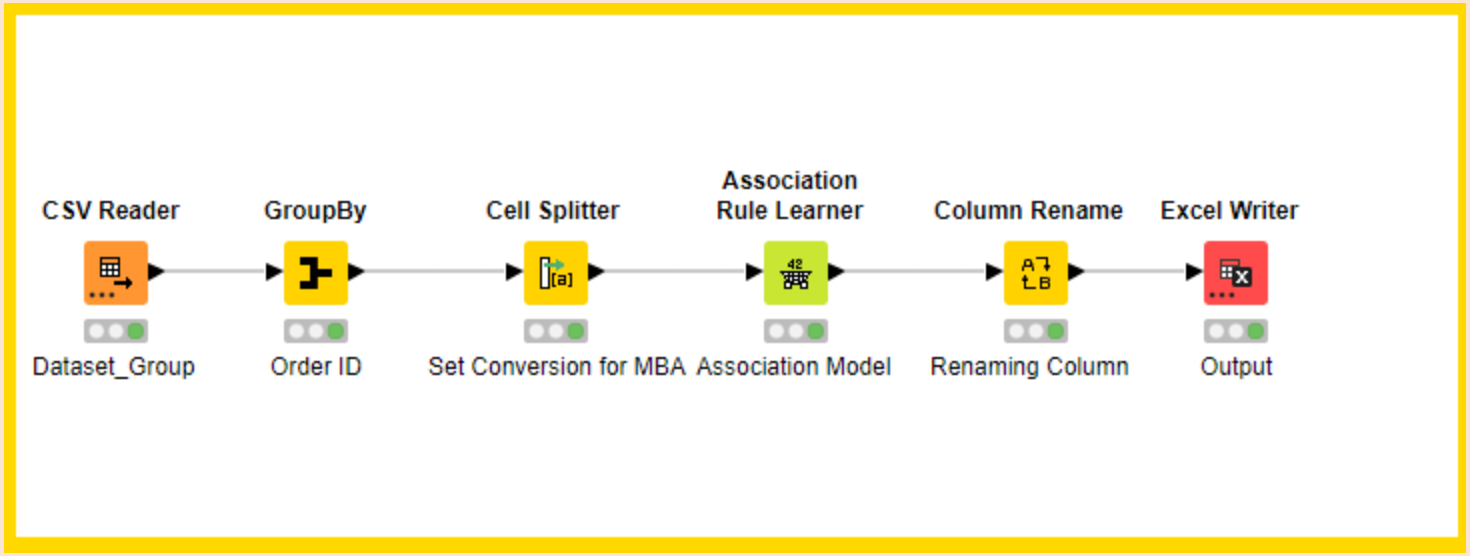
- CONFIDENCE –
  - Confidence (A è B) – is the likelihood of a customer buying item A, will also buy item B
  - This is the Probability of B given that A has been bought

- LIFT –
  - Lift is the most important metric to consider when choosing an association rule
  - Given A is bought, then – Lift is the % increase in the chance of buying

MBA on Grocery store dataset:

- We perform MBA using KNIME on the given Grocery Store Dataset
- We choose multiple thresholds for Min Support and Min Confidence to filter out less frequent and less appropriate rules
- Finally chosen thresholds for –
  - Support = 0.03 (3%)
  - Confidence = 0.55 (55%)
- The above values indicate –
  - We want to create rules with only those items which appear in at least 3% of transactions
  - We want a minimum Confidence of 55% in the rule i.e. Suggested item should be bought at least 55 out of every 100 times Basket items are bought
- We get an output of 10550 rules
- All rules - Average Lift, Confidence, and Support is 1.48, 0.57, and 0.04 respectively

KNIME Workflow– Market Basket Analysis:



Output head– Final Output of the rules table:

Table "default" - Rows: 10550 | Spec - Columns: 6 | Properties | Flow Variables

Row ID	D Support	D Confidence	D Lift	S Suggested Item	S implies	S Basket Items
rule234	0.031	0.795	2.194	paper towels	<---	[eggs, ice cream, pasta, lunch meat]
rule557	0.032	0.783	2.158	paper towels	<---	[eggs, ice cream, pasta, cereals]
rule235	0.031	0.729	2.066	flour	<---	[dishwashing liquid/detergent, cheeses, waffles, soda]
rule1076	0.032	0.74	2.041	paper towels	<---	[eggs, dinner rolls, ice cream, pasta]
rule562	0.032	0.72	1.986	paper towels	<---	[eggs, poultry, ice cream, pasta]
rule231	0.031	0.778	1.951	ice cream	<---	[paper towels, eggs, pasta, lunch meat]
rule236	0.031	0.761	1.947	soda	<---	[dishwashing liquid/detergent, cheeses, flour, waffles]
rule1718	0.033	0.717	1.931	pasta	<---	[paper towels, dishwashing liquid/detergent, eggs, ice cream]
rule7402	0.04	0.697	1.922	paper towels	<---	[all- purpose, individual meals, toilet paper]
rule247	0.031	0.714	1.914	spaghetti sauce	<---	[dinner rolls, poultry, laundry detergent, juice]
rule1717	0.033	0.745	1.911	eggs	<---	[paper towels, dishwashing liquid/detergent, ice cream, pasta]
rule1719	0.033	0.691	1.905	paper towels	<---	[dishwashing liquid/detergent, eggs, ice cream, pasta]
rule561	0.032	0.706	1.901	pasta	<---	[paper towels, eggs, poultry, ice cream]
rule560	0.032	0.72	1.847	eggs	<---	[paper towels, poultry, ice cream, pasta]
rule1075	0.032	0.685	1.845	pasta	<---	[paper towels, eggs, dinner rolls, ice cream]
rule3309	0.036	0.641	1.833	sandwich loaves	<---	[all- purpose, flour, individual meals]
rule1074	0.032	0.712	1.825	eggs	<---	[paper towels, dinner rolls, ice cream, pasta]
rule7450	0.04	0.676	1.822	pasta	<---	[hand soap, soda, aluminum foil]
rule7488	0.04	0.676	1.822	ketchup	<---	[butter, aluminum foil, soap]
rule5092	0.038	0.632	1.81	sandwich loaves	<---	[paper towels, flour, individual meals]
rule7833	0.041	0.671	1.808	ketchup	<---	[pork, sandwich bags, soap]
rule1758	0.034	0.629	1.8	sandwich loaves	<---	[yogurt, hand soap, soap]
rule243	0.031	0.7	1.796	eggs	<---	[dishwashing liquid/detergent, ice cream, pasta, soda]
rule10009	0.046	0.65	1.793	paper towels	<---	[ice cream, pasta, lunch meat]
rule10515	0.055	0.649	1.791	paper towels	<---	[eggs, ice cream, pasta]
rule4892	0.038	0.662	1.786	fruits	<---	[all- purpose, beef, lunch meat]
rule6411	0.039	0.688	1.784	bagels	<---	[sandwich loaves, fruits, juice]
rule9513	0.044	0.676	1.781	soap	<---	[sandwich loaves, all- purpose, ketchup]
rule555	0.032	0.692	1.776	eggs	<---	[paper towels, ice cream, pasta, cereals]
rule1378	0.033	0.667	1.774	mixes	<---	[all- purpose, hand soap, tortillas]
rule4792	0.038	0.672	1.767	milk	<---	[sandwich loaves, pork, soda]
rule5074	0.038	0.623	1.766	flour	<---	[pasta, mixes, coffee/tea]
rule556	0.032	0.655	1.762	pasta	<---	[paper towels, eggs, ice cream, cereals]
rule5536	0.038	0.614	1.758	sandwich loaves	<---	[cheeses, hand soap, ketchup]
rule238	0.031	0.686	1.757	cheeses	<---	[dishwashing liquid/detergent, flour, waffles, soda]
rule7092	0.04	0.652	1.756	ketchup	<---	[tortillas, coffee/tea, juice]
rule9384	0.044	0.617	1.749	flour	<---	[yogurt, pasta, coffee/tea]
rule8803	0.042	0.649	1.747	pasta	<---	[dinner rolls, hand soap, individual meals]

Recommendations:

Suggested Item	Count
poultry	1330
cheeses	598
lunch meat	575
soda	549
eggs	499
yogurt	487
dinner rolls	427
ice cream	411
waffles	404
juice	393

- The table shows the top 10 items suggested by the rules
- Most of these items are stored in a refrigerator
- It is recommended to have Refrigerators lined up on a sidewall which is accessible from all aisles
- Poultry, Eggs, Ice Cream, Meat buyers tend to buy Paper Towels more often
- It is recommended to keep Paper Towels for use and on a shelf to sell near Refrigerators
- Looking at the cart mix of products, it seems customers very often come to pick up ingredients for pasta
- It is recommended to make a Pasta Bag – containing eggs, cheese, pasta, and spaghetti sauce

- Beef and/or Pork buyers are seen to have a high likelihood of also buying cleaning products such as soap, hand soap, shampoo, and dishwashing liquid
- It is recommended to have a small shelf of choicest cleaning agents near the deep freezers containing these meats
- Poultry is the most sold item – most customers seem to come primarily for this
- Maximize this by cross-selling other items with this like – Buy 2 Poultry get 50% off on Paper Towels or Buy 3 Poultry get 10/- store credit

THANK YOU
S.VISHAL