

MRA PROJECT — MILESTONE 2

Market Basket Analysis (Association Rules)



The diagram illustrates a data analysis process flow. It consists of six rectangular boxes arranged in a 2x3 grid. The top row contains three orange boxes, and the bottom row contains three light orange boxes. Each box is connected to a vertical line at the top and bottom, which then branches into two horizontal lines to form a rectangular frame around the text. The text in each box is in a bold, black, sans-serif font. The background is split horizontally into a light blue top half and a light green bottom half.

**PROBLEM
STATEMENT**

**DATA
SUMMARY**

**EXPLORATORY
DATA
ANALYSIS**

**SALES
TREND**

**MARKET BASKET
ANALYSIS**

SUGGESTIONS



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 menu items in the customer orders. The Store doesn't have any combo meals. Can you suggest the best combo meals?

DATA SUMMARY

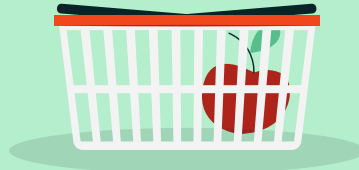


	Order_id	Date	Order_id	Product
count	20641.000000	2018-01-01	1	yogurt
mean	575.986289	2018-01-01	1	pork
std	328.557078	2018-01-01	1	sandwich bags
min	1.000000	2018-01-01	1	lunch meat
25%	292.000000	2018-01-01	1	all- purpose
50%	581.000000	HEAD OF THE TABLE		
75%	862.000000		Date	Product
max	1139.000000	count	20641	20641
DATA DESCRIPTION (NUMERIC)		unique	603	37
		top	2019-02-08	poultry
		freq	183	640
		DATA DESCRIPTION (CATEGORICAL)		

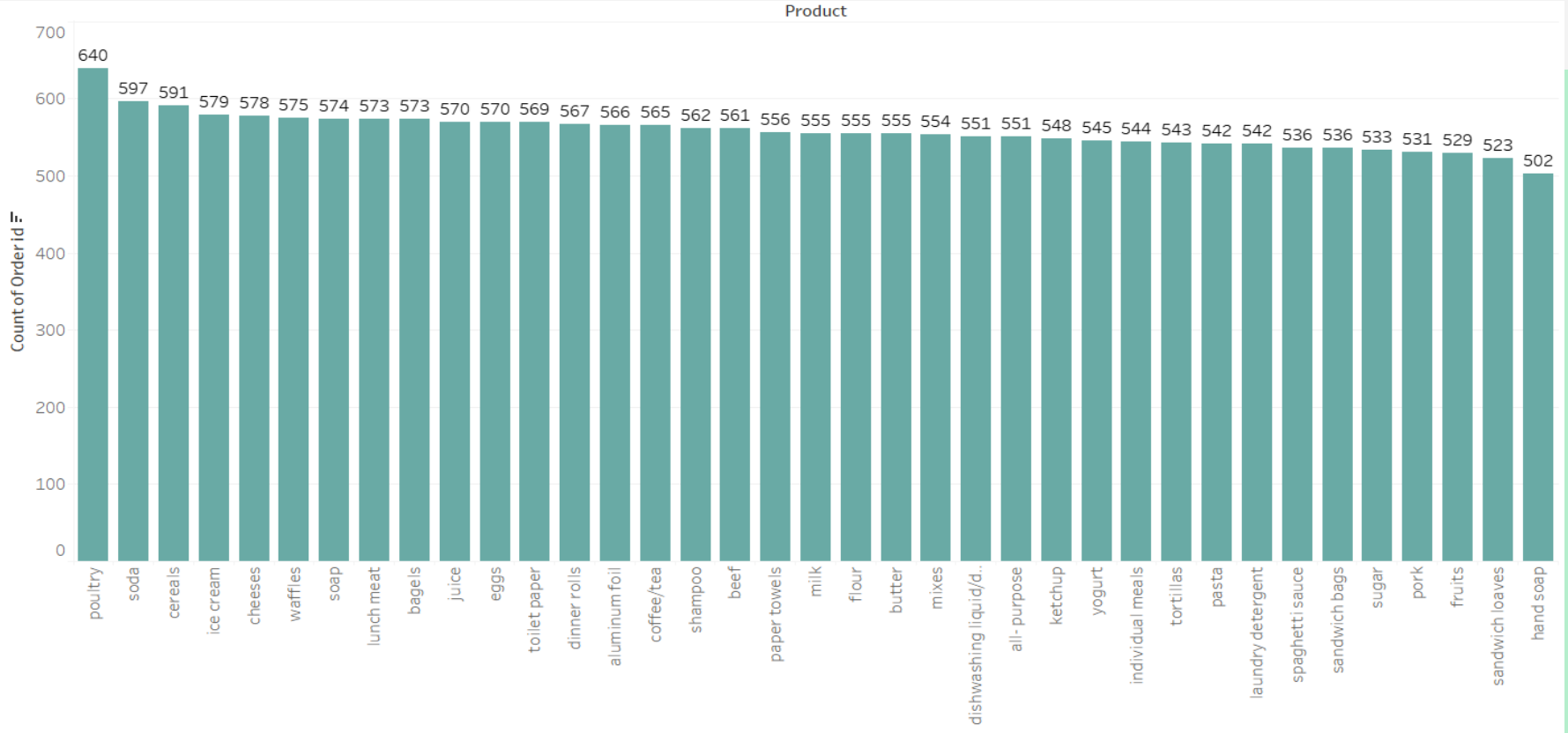
- Total number of duplicate rows: 4730.
- There are no missing values.
- The shape of the dataset is (20641, 3).

DATA INFERENCE

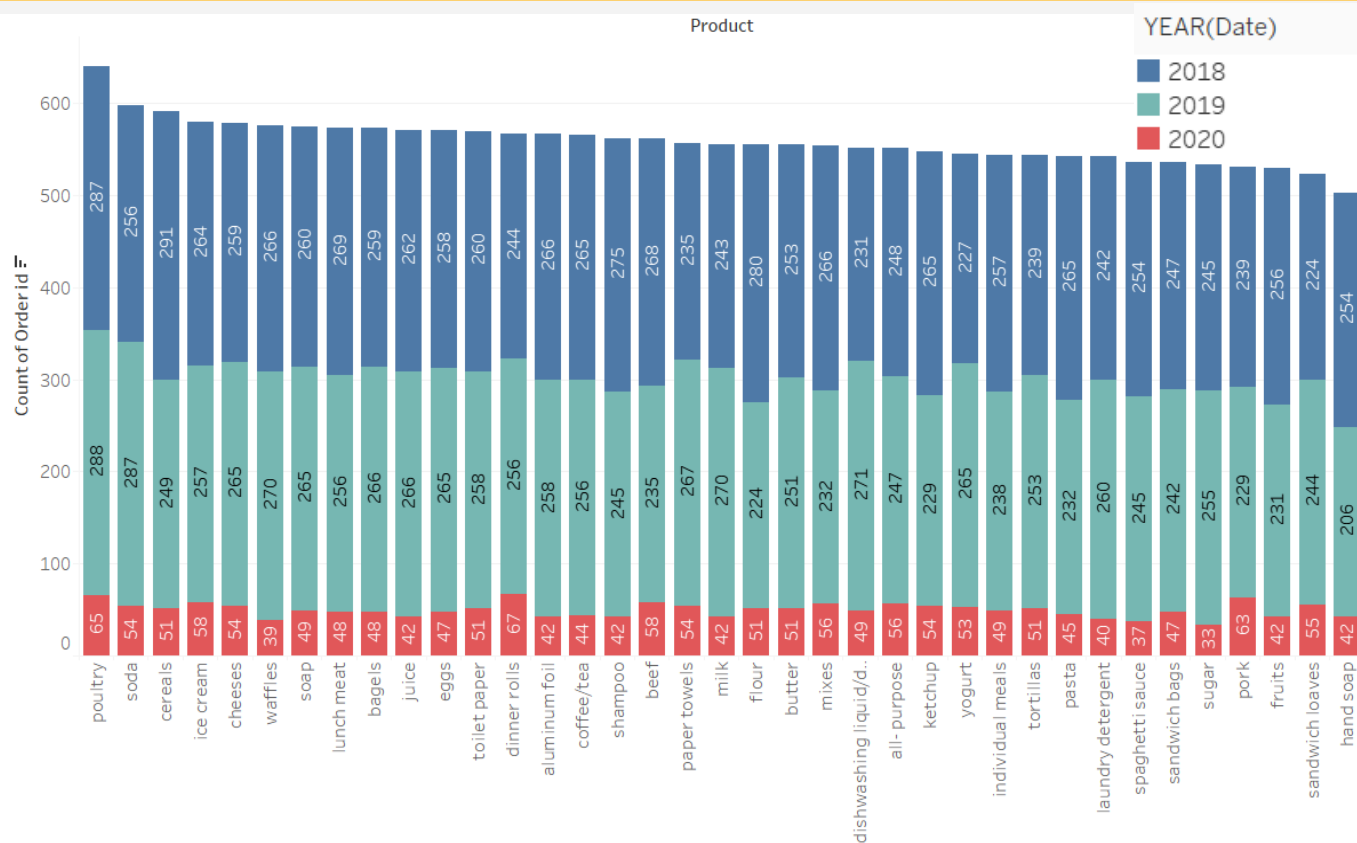
- There are no null data with 3 variables (1 ordinal variable, 1 date time and 1 object datatypes).
- There are 4730 duplicate rows assuming there was more than one quantity of a product bought in a single order.
- There are a total of 37 unique products in the Grocery Store, among which Poultry is the highest sold product.
- The Order_id is an identifier here and finding outliers using Boxplot or Histogram is meaningless.
- The highest sales for the day was reported on 2019-02-08.



EXPLORATORY DATA ANALYSIS- BIVARIATE



EXPLORATORY DATA ANALYSIS- BIVARIATE



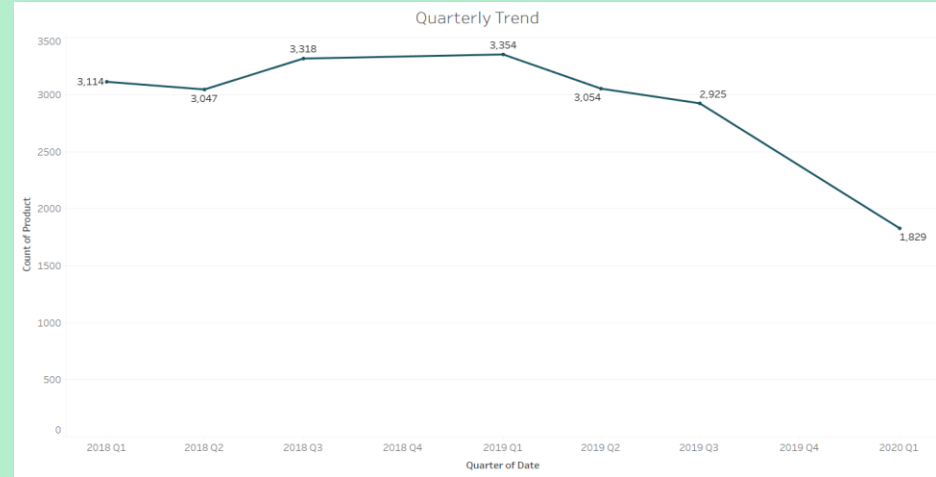
The highest order placed is for poultry followed by soda and then cereals.

In 2018, cereals was ordered the most.

For 2019 and 2020, poultry has the highest order quantity.

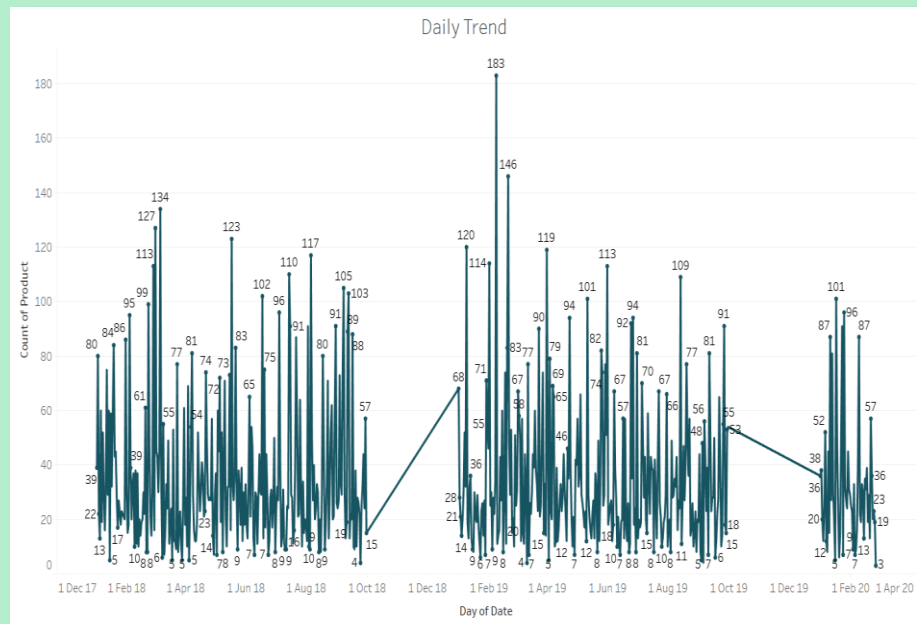
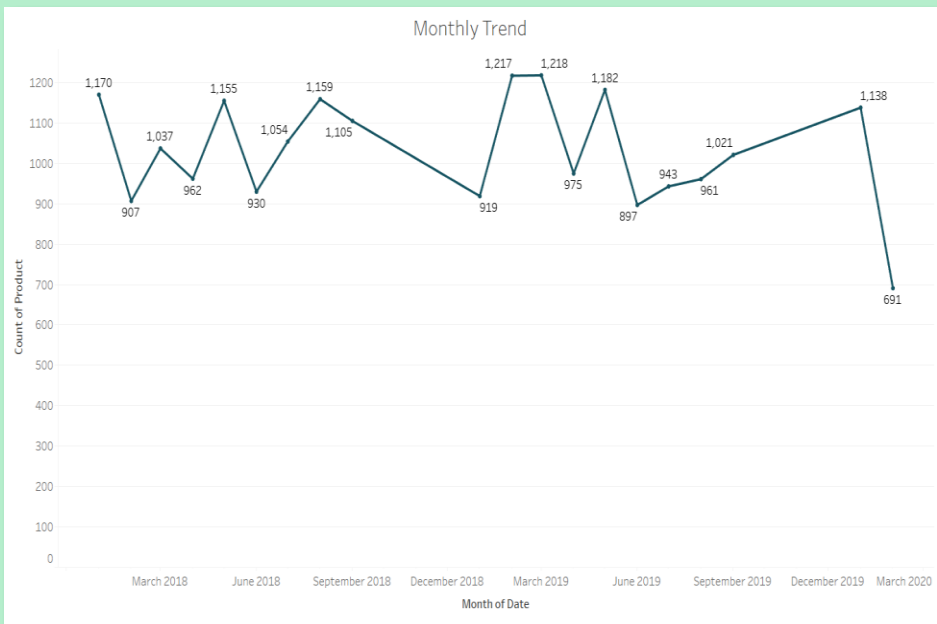


TRENDS ACROSS YEARS/ QUARTERS



- The sales decreased a bit from 2018 to 2019 but has a sharp decrease in the year 2020.
- From 3rd Quarter of 2018 to 1st Quarter of 2019, there was a good number of products sold but it started decreasing from 1st Quarter of 2019.

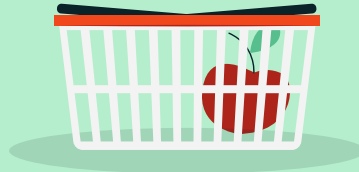
TRENDS ACROSS MONTHS/ DAYS



There is no monthly or daily trend to be seen.

EDA AND TRENDS – INFERENCE

- The orders most received are for Poultry and the least is for hand soap. Though there isn't much difference in the total. The orders of medium size are received mostly.
- The store receives order ID varies from 1 to 500. 34 is the highest item list on orders and 3 is the least.
- The store is facing a decline in order on 2020. Compared to the last 2 years. The sales decreased a bit from 2018 to 2019 but has a sharp decrease in the year 2020.
- From 3rd Quarter of 2018 to 1st Quarter of 2019, there was a good number of products sold but it started decreasing from 1st Quarter of 2019.

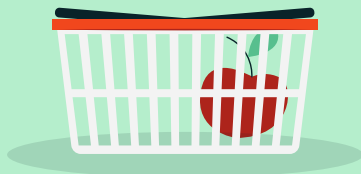


MARKET BASKET ANALYSIS (ASSOCIATION RULES)

Market basket analysis analyzes customer behavior by finding associations between the different items that customers place in their shopping baskets. This can help businesses develop marketing strategies by gaining an insight into which items are frequently purchased together by customers.

Association Rules: are widely used to analyze retail association between different objects in a set, find frequent patterns in a transaction database.

For instance, if customers are buying bread, how likely are they to also buy butter on the same trip to the supermarket. This information can be used to increase sales by placing these goods together.



ASSOCIATION RULES

BASED ON SUPPORT

$\text{support}(A) = \text{Number of transaction in which A appears} / \text{Total number of transactions}$

It is the ratio of transactions involving number of transaction of the item by the total number of transactions made.

BASED ON CONFIDENCE AND LIFT

Confidence: It is whether the product sales are popular on individual sales or through combined sales has been calculated.

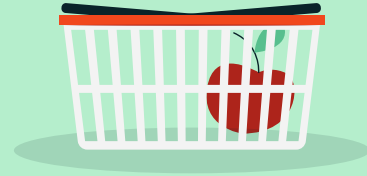
The lift of a rule is defined as:

$$\text{lift}(X \rightarrow Y) = \frac{\text{supp}(X \cup Y)}{(\text{supp}(X) * \text{supp}(Y))}$$

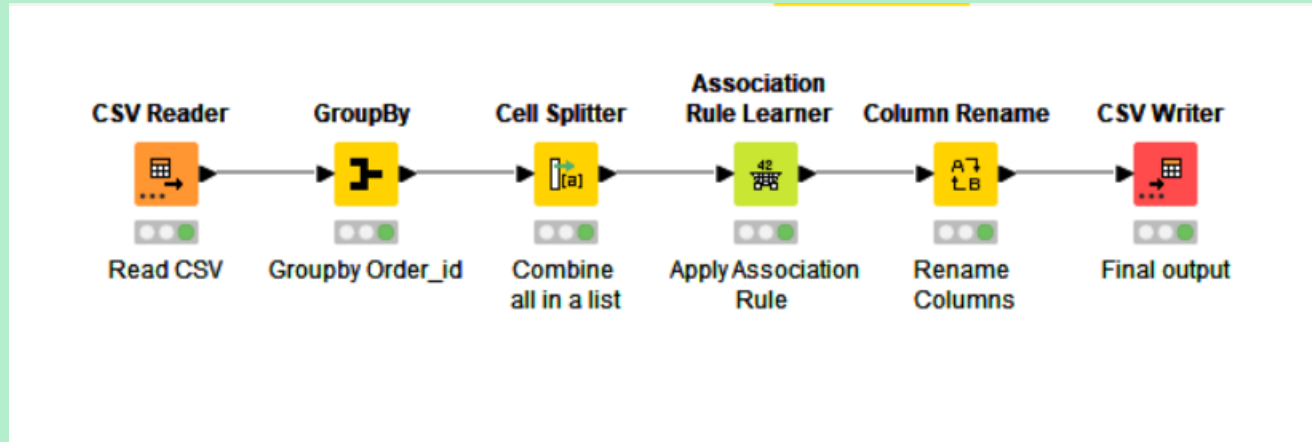
It calculated for knowing the ratio for the sales. (confidence percent/ support percent).

It is the ratio of the observed support divided by individual support of the items.

If the value of lift is greater than 1, it means that the itemset Y is likely to be bought with itemset X, while a value less than 1 implies that itemset Y is unlikely to be bought if the itemset X is bought.

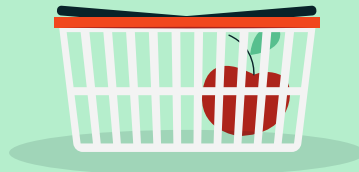


KNIME WORKFLOW IMAGE



THRESHOLD VALUES OF SUPPORT AND CONFIDENCE

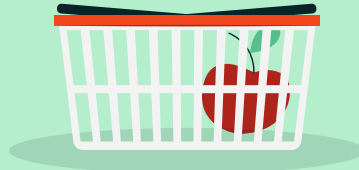
- The above KNIME Workflow has been used to determine the rule using Association Learner Rule.
- At Support 0.5 and Confidence 0.60, there are 24 rules and combination of products that could be sold together having a positive correlation i.e. lift is more than 1.
- A lower threshold would not make sense as the number of combinations are the product would be more and hard to implement.



ASSOCIATIONS IDENTIFIED

Support	Confidence	Lift	Recommended_item	Recommended_with	Items_list
0.0553117	0.649484536	1.79119343	paper towels	<---	[eggs, ice cream, pasta]
0.0553117	0.642857143	1.73100304	pasta	<---	[paper towels, eggs, ice cream]
0.0509219	0.674418605	1.726208518	cheeses	<---	[bagels, cereals, sandwich bags]
0.0500439	0.640449438	1.700400723	juice	<---	[yogurt, toilet paper, aluminum foil]
0.0509219	0.630434783	1.677722471	mixes	<---	[yogurt, poultry, aluminum foil]

The top 5 recommended items along with the products that should be placed together are presented above in the descending order of Lift.



ASSOCIATIONS IDENTIFIED

Support	Confidence	Lift	Recommended_item	Recommended_with	Items_list
0.0500439	0.640449438	1.700400723	Juice <---		[yogurt, toilet paper, aluminum foil]
0.0500439	0.619565217	1.644952873	Juice <---		[yogurt, poultry, aluminum foil]
0.0500439	0.612903226	1.615964755	coffee/tea <---		[yogurt, cheeses, cereals]
0.0500439	0.6	1.42375	Poultry <---		[dishwashing liquid/detergent, laundry detergent, mixes]
0.0509219	0.630434783	1.677722471	Mixes <---		[yogurt, poultry, aluminum foil]
0.0509219	0.610526316	1.659640749	sandwich bags <---		[cheeses, bagels, cereals]
0.0509219	0.674418605	1.726208518	Cheeses <---		[bagels, cereals, sandwich bags]
0.0509219	0.617021277	1.55828655	Cereals <---		[cheeses, bagels, sandwich bags]
0.0509219	0.630434783	1.620914712	dinner rolls <---		[spaghetti sauce, poultry, cereals]
0.0509219	0.637362637	1.512408425	Poultry <---		[dinner rolls, spaghetti sauce, cereals]
0.0509219	0.604166667	1.589251347	milk <---		[poultry, laundry detergent, cereals]
0.0517998	0.627659574	1.610144719	eggs <---		[dinner rolls, poultry, soda]
0.0517998	0.641304348	1.648861517	dinner rolls <---		[spaghetti sauce, poultry, ice cream]
0.0517998	0.686046512	1.627931202	poultry <---		[dinner rolls, spaghetti sauce, ice cream]
0.0517998	0.627659574	1.613779357	dinner rolls <---		[spaghetti sauce, poultry, juice]
0.0517998	0.602040816	1.428592687	poultry <---		[dinner rolls, spaghetti sauce, juice]
0.0517998	0.634408602	1.627458103	eggs <---		[paper towels, dinner rolls, pasta]
0.0517998	0.602040816	1.621098085	pasta <---		[paper towels, eggs, dinner rolls]
0.0535558	0.642105263	1.650920756	dinner rolls <---		[spaghetti sauce, poultry, laundry detergent]
0.0535558	0.655913978	1.556429211	poultry <---		[dinner rolls, spaghetti sauce, laundry detergent]
0.0553117	0.623762376	1.564901644	ice cream <---		[paper towels, eggs, pasta]
0.0553117	0.63	1.616148649	eggs <---		[paper towels, ice cream, pasta]
0.0553117	0.642857143	1.73100304	pasta <---		[paper towels, eggs, ice cream]
0.0553117	0.649484536	1.79119343	paper towels <---		[eggs, ice cream, pasta]

RECOMMENDATIONS

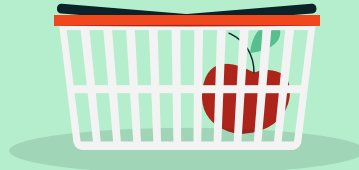
- At support: 0.5 and confidence: 0.60, our 15 rules lift values are above 1. i.e. there is a positive correlation within the itemset.
- The rules have most set to have poultry as recommend item. It also being the most sold item, creating combo with poultry will be highly beneficial.
- The introductory offer for new customer with point rewarding scheme for next purchase.
- The order drop during Q4 and year end, end of season sales will be profitable during this time.
- Rise is sales are seen at the year start, discounts or limited time period event price that would be helpful during end of the year or weekdays to boost sales.
- Sandwich bag placed between Cheeses and Bagels. Would increase Sandwich bag sale.



SUGGESTIONS

Possible Combos with Lucrative Offers:

- Giving buy one get one free for paper towel with eggs, ice cream and pasta.
- Cheese should be placed near bagel, cereals and sandwich bags.
- Juice should be placed at a distance from yogurt, toilet paper and aluminum foil since it would bound the consumer to explore more as Juice is the most bought item with the following products.
- 40% discount on spaghetti sauce when sold with dinner rolls and poultry combo.
- 20% off on hand soap with poultry purchase since hand soap is the lowest selling item and poultry being the highest.
- Loyalty points to redeem at the end of the month/year regular customer. Would import retention rate.





THANK YOU

