MRA Project ML 2

Priyadharshini K

PGP – DSBA Nov-C Batch, 2021

kpriyadharshinik@gmail.com

Tableau Public Link

https://public.tableau.com/views/MRA2_16330836865020/Sheet1?:language=en-US&:display_count=n&:origin=viz_share_link

Agenda [Table OF CONTENT]

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?

The project involves conducting a thorough analysis of Point of Sale (POS) Data for providing recommendations through which a grocery store can increase its revenue by popular combo offers & discounts for customers.

ABOUT DATA

➤ The dataset has a total of 3 variables – Numeric (Order id - int64(5)) , String (Product - object(12)), Date - datetime64(1)

➤ Nominal Variable : Date has 603 unique values while Product has 37 unique values

➤ Number Variable : Order Id

DATA PREVIEW

Numeric	Nominal	Data Preview	

Row ID	Date	Order_id	Product
Row0	01-01-2018	1	yogurt
Row1	01-01-2018	1	pork
Row2	01-01-2018	1	sandwich bags
Row3	01-01-2018	1	lunch meat
Row4	01-01-2018	1	all- purpose
Row5	01-01-2018	1	flour
Row6	01-01-2018	1	soda
Row7	01-01-2018	1	butter
Row8	01-01-2018	1	beef
Row9	01-01-2018	1	aluminum foil
Row10	01-01-2018	1	all- purpose

EDA – Univariate Analysis
Tool Used: KNIME

| Numeric | Nominal | Data Preview |

Search: [   ]

Column	Exclude Column	No. missings	Unique values	All nominal values	Frequency Bar Chart
Date	☐	0	603	08-02-2019, 20-02-2019, 06-03-2018, 01-03-2018, 17-05-2018, [...], 02-04-2019, 24-09-2018, 11-03-2019, 05-09-2019, 26-02-2020	
Product	☐	0	37	poultry, soda, cereals, ice cream, cheeses, [...], sugar, pork, fruits, sandwich loaves, hand soap	

Showing 1 to 2 of 2 entries

Poultry, Soda, Cereals have the maximum number of Orders in the descending fashion
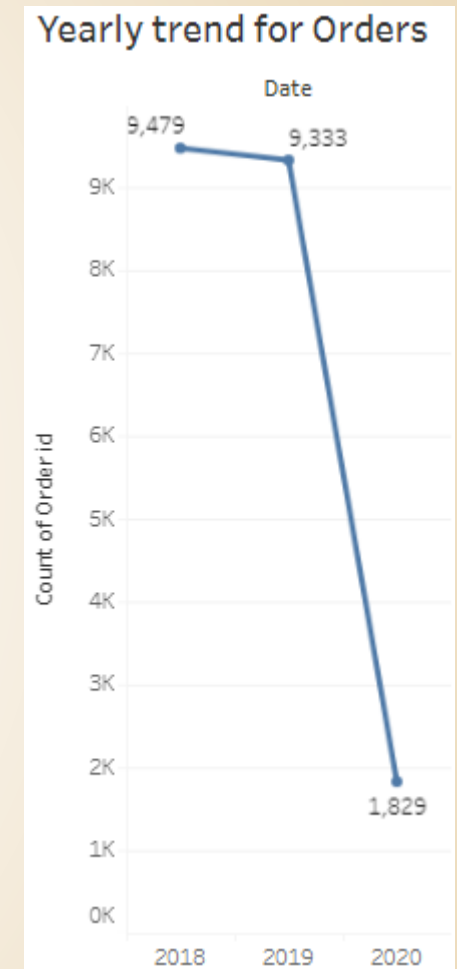
EDA – Univariate Analysis
Tool Used: Tableau
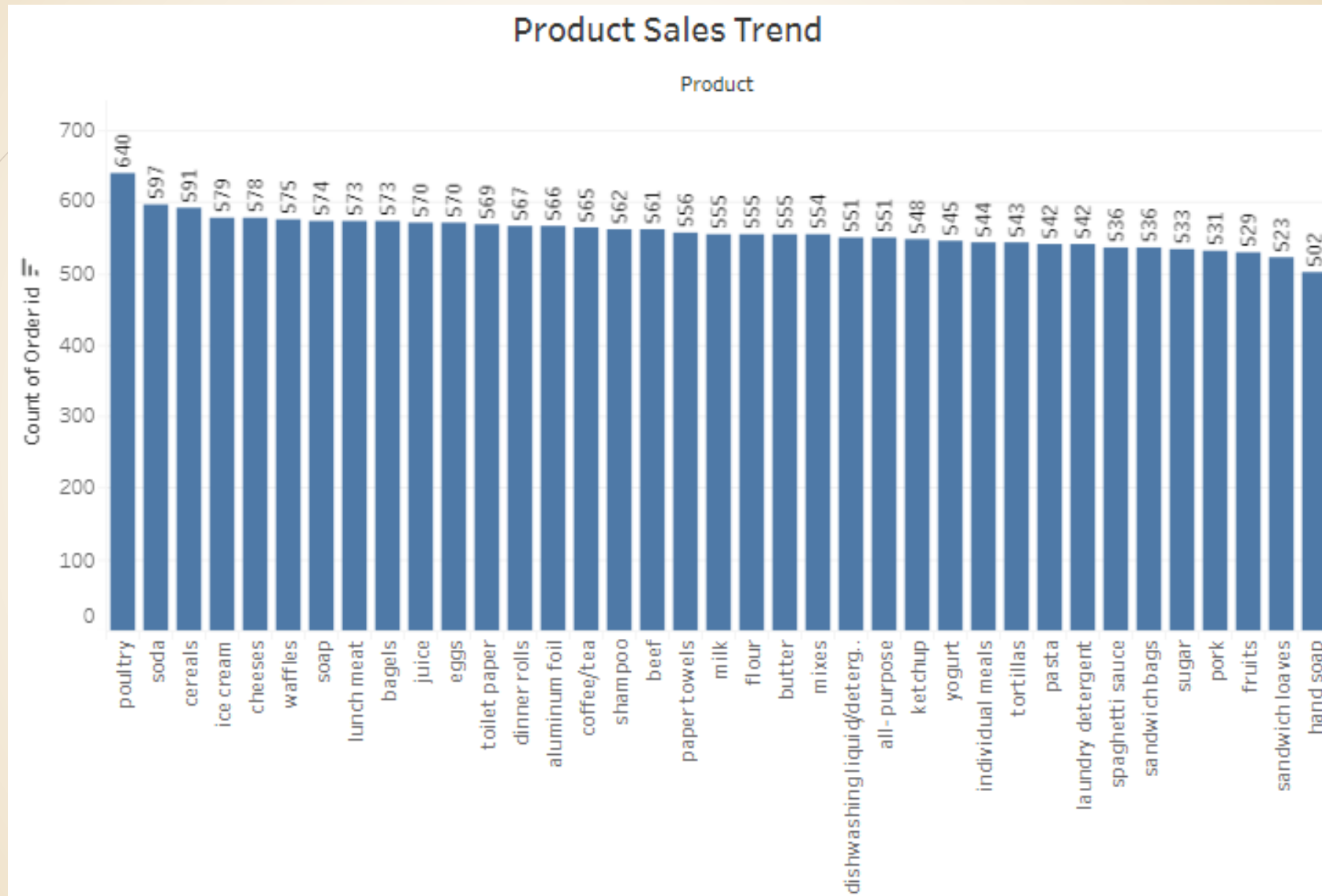
Poultry has the maximum number of Orders, followed by Soda and Cereals

EDA - Multivariate Analysis
Tool Used: Tableau
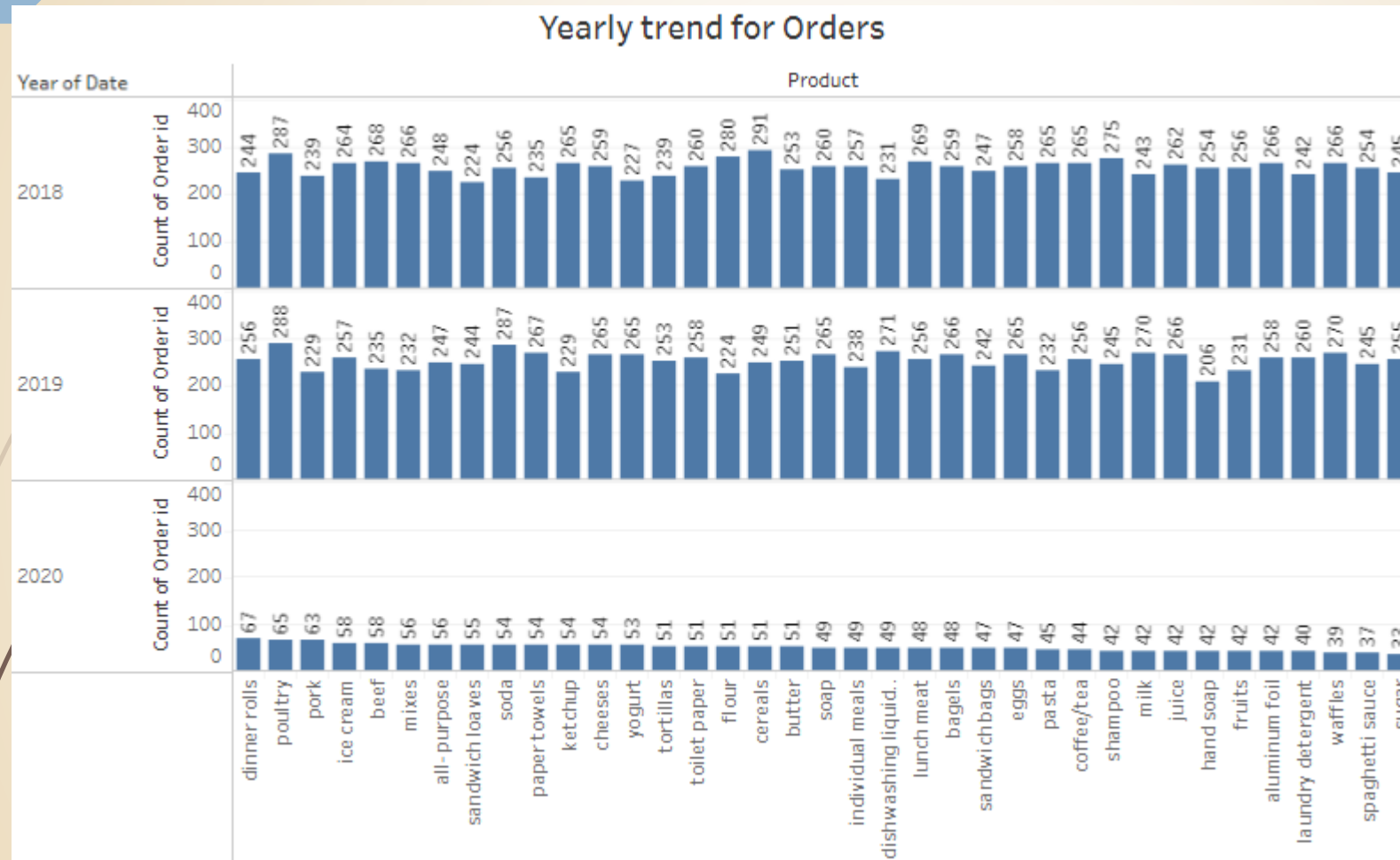
Data for Q2,Q3 is missing for 2020. The top product of Sales for 2018 differs from that of 2019

EDA - Multivariate Analysis
Tool Used: Tableau

Quarterly trend for Orders

Monthly trend for Orders
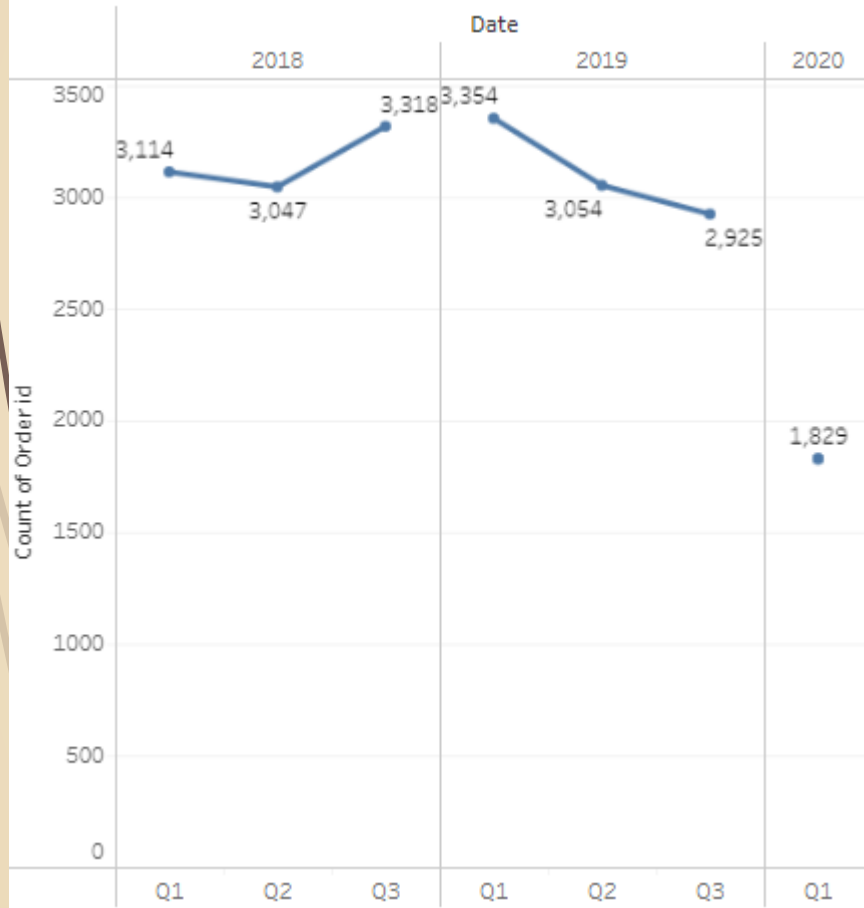
2018 – Q3 Sales is the highest, followed by Q1 and Q2
2019 – Q1 Sales is the highest, followed by Q2 and Q3

2018 – January records highest sales, while June records lowest
2019 – Q1 Sales is the highest, followed by Q2 and Q3

TREE CHART

WORD CLOUD

Poultry, Soda and Cereals form the highest order for all the years 2018,2019

2018

243 milk

266 waffles | 258 eggs | 280 flour | 268 beef

256 fruits | 244 dinner rolls | 266 mixes | 254 hand soap | 259 cheeses | 259 bagels | 291 cereals | 265 ketchup

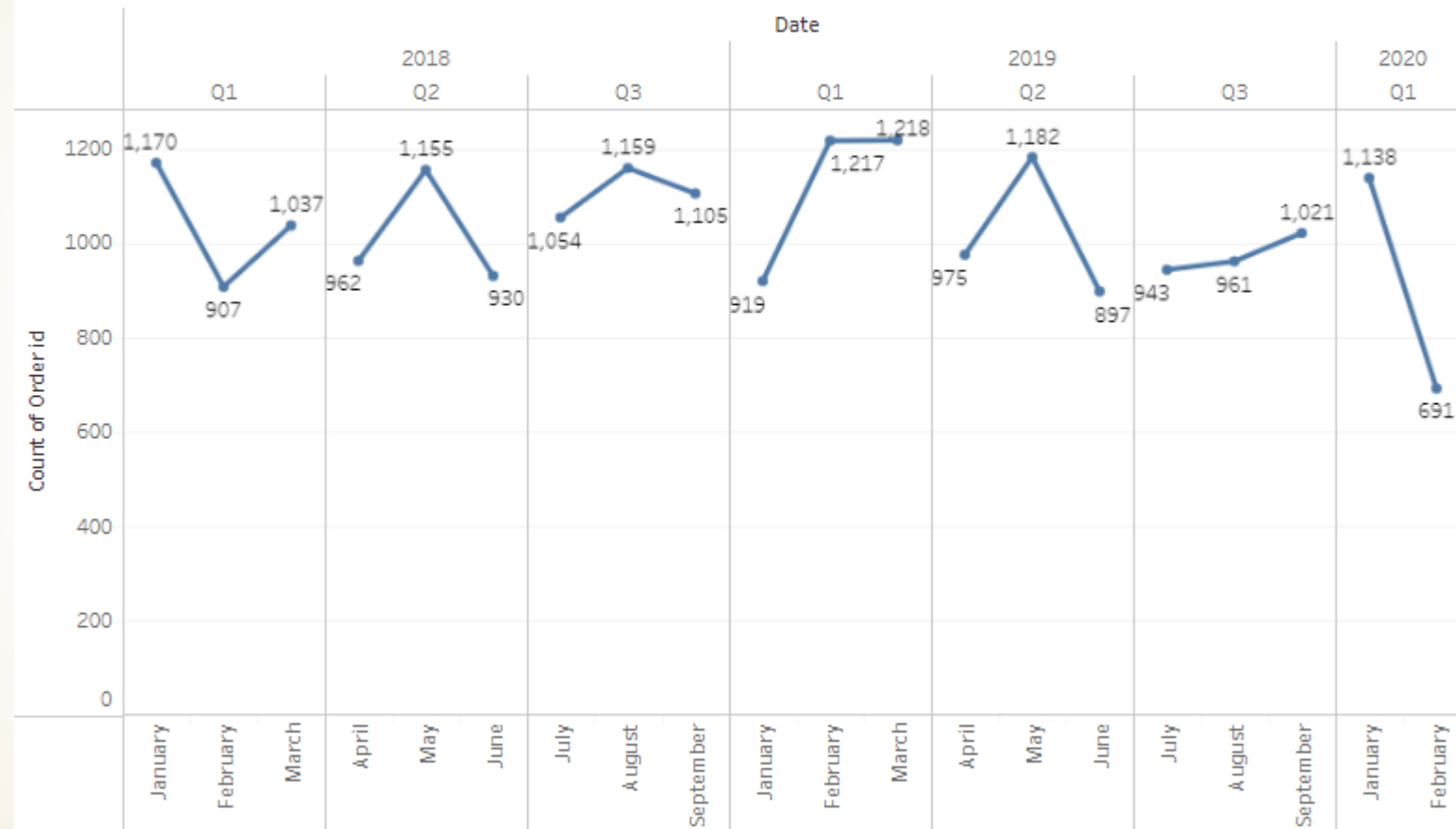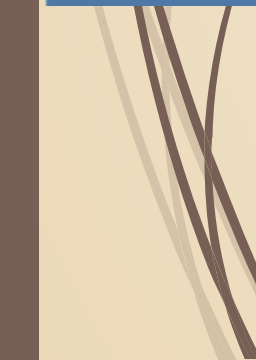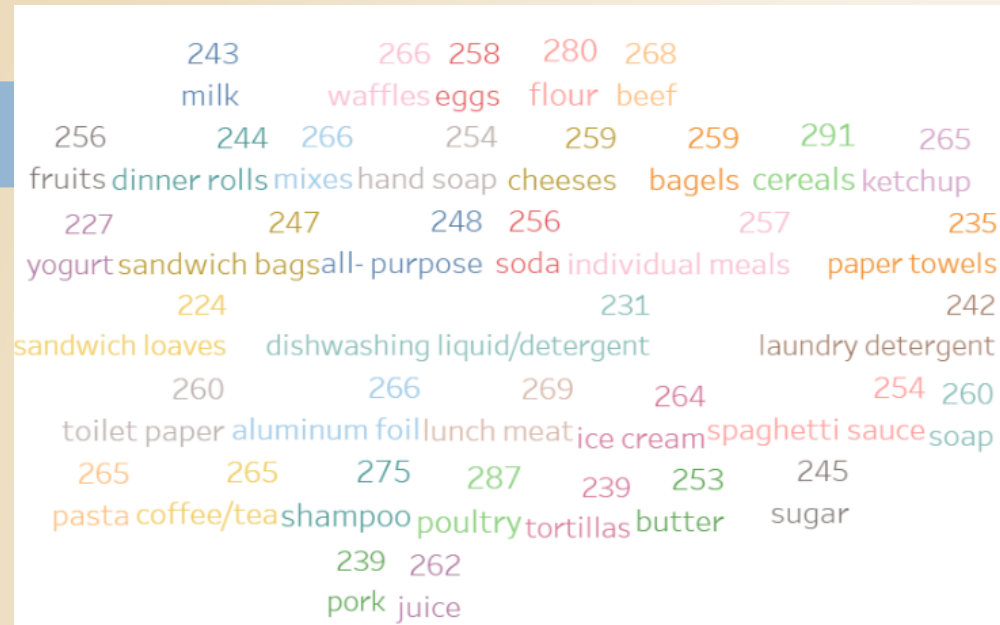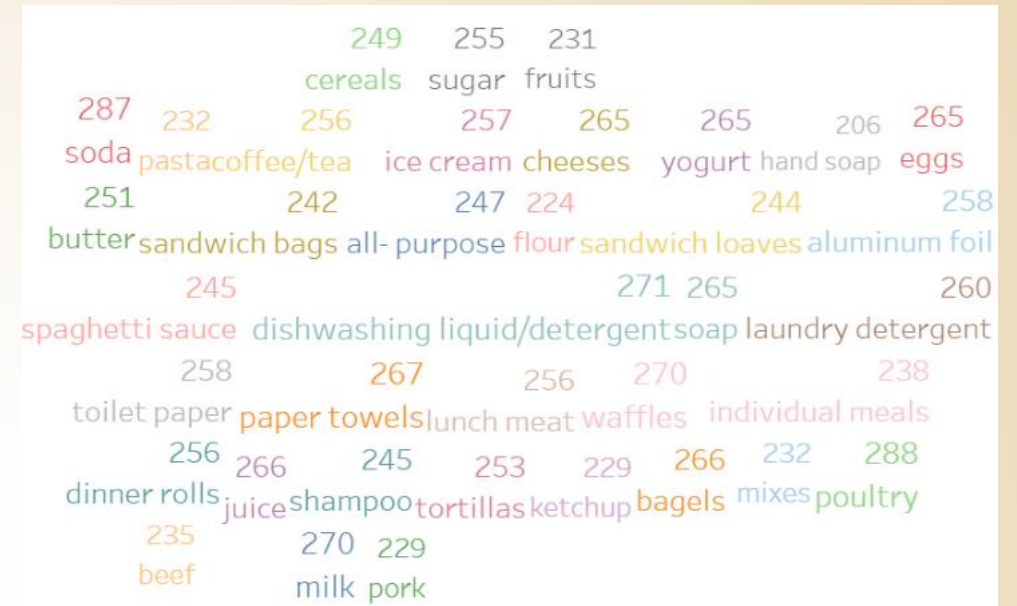227 yogurt | 247 sandwich bags | 248 all- purpose | 256 soda | 257 individual meals | 235 paper towels

224 sandwich loaves | 231 dishwashing liquid/detergent | 242 laundry detergent

260 toilet paper | 266 aluminum foil | 269 lunch meat | 264 ice cream | 254 spaghetti sauce | 260 soap

265 pasta | 265 coffee/tea | 275 shampoo | 287 poultry | 239 tortillas | 253 butter | 245 sugar

239 pork | 262 juice

2019

249 cereals | 255 sugar | 231 fruits

287 soda | 232 pasta | 256 coffee/tea | 257 ice cream | 265 cheeses | 265 yogurt | 206 hand soap | 265 eggs

251 butter | 242 sandwich bags | 247 all- purpose | 224 flour | 244 sandwich loaves | 258 aluminum foil

245 spaghetti sauce | 271 dishwashing liquid/detergent | 265 soap | 260 laundry detergent

258 toilet paper | 267 paper towels | 256 lunch meat | 270 waffles | 238 individual meals

256 dinner rolls | 266 juice | 245 shampoo | 253 tortillas | 229 ketchup | 266 bagels | 232 mixes | 288 poultry

235 beef | 270 milk | 229 pork

2020

53 yogurt | 33 sugar | 56 mixes

58 beef | 42 aluminum foil | 42 fruits | 65 poultry | 42 hand soap | 51 butter | 51 cereals | 54 cheeses

48 bagels | 47 sandwich bags | 37 spaghetti sauce | 47 eggs | individual meals | 56 all- purpose

40 laundry detergent | dishwashing liquid/detergent | 49 flour | 51 sandwich loaves | 55

42 milk | 51 toilet paper | 54 paper towels | 58 ice cream | 54 ketchup | 67 dinner rolls | 63 pork

45 pasta | 48 lunch meat | 51 tortillas | 44 coffee/tea | 42 shampoo | 54 soda | 49 soap
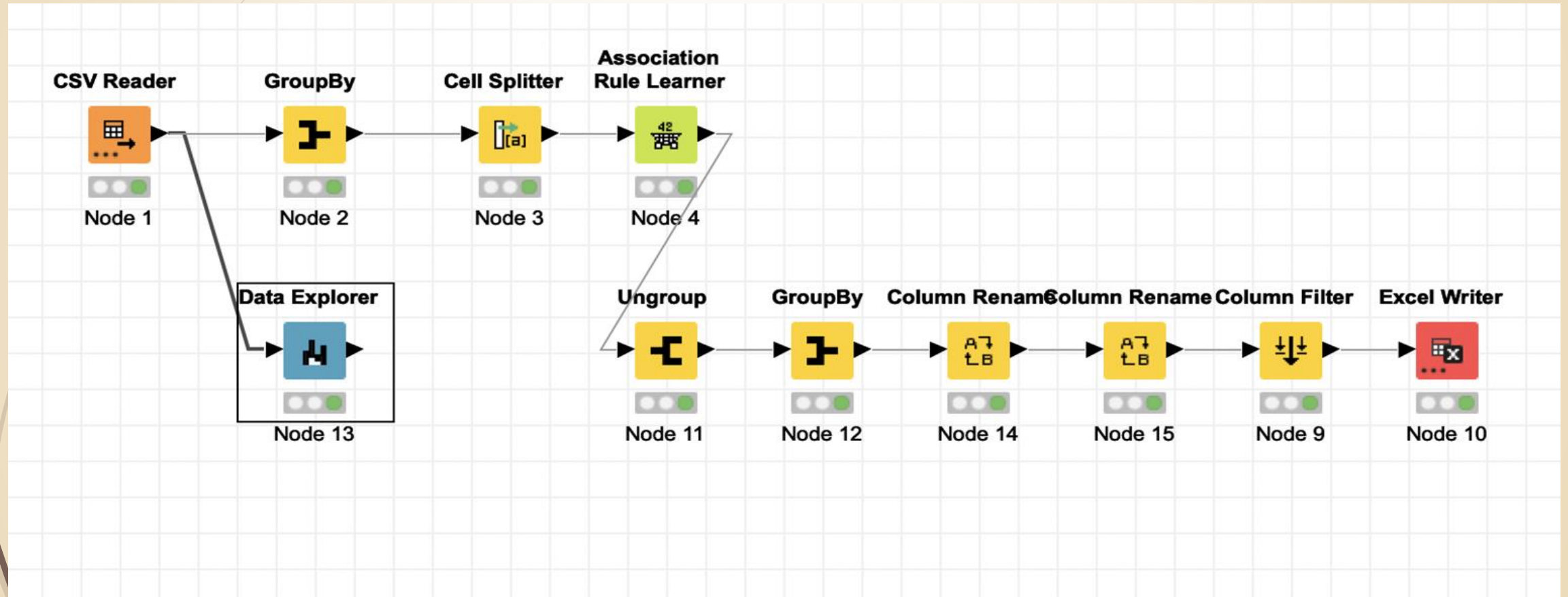
39 waffles | 42 juice

KNIME WORKFLOW

Association rules are "if-then" statements, that help to show the probability of relationships between data items, within large data sets in various types of databases. These rules are created by searching data for frequent if-then patterns and using the criteria support and confidence to identify the most important relationships.

Support is an indication of how frequently the items appear in the data. The probability of appearance of an item in the total number of items.

Confidence is an indication of how often the rule has been found to be true.
The *confidence* value of a rule, X→Y with respect to a set of transactions T, is the proportion of the transactions that contains 'X' which also contains 'Y'

Lift can be used to compare confidence with expected confidence, or how many times an if-then statement is expected to be found true.

An association rule has two parts: an antecedent (if) and a consequent (then). An antecedent is an item found within the data. A consequent is an item found in combination with the antecedent.

Support and Confidence threshold value determine the total number of rules/associations possible. This can be varied for multiple iterations to get the maximum probability.
For this dataset, **support is set at 0.02 while confidence is set at 0.75**. This means the occurrence of an item to be checked in the dataset is checked at 2% probability and 75% of the times the recommended product is found to be associated with the basket list of items.

ASSOCIATIONS IDENTIFIED USING MARKET BASKET ANALYSIS

(SAMPLE OUPUT OF THE WORKFLOW)

Support	Confidence	Lift	Recommended Product	implies	Basket Items
0.021071115	0.774193548	2.041218638	soap	<---	all- purpose, beef, lunch meat, sugar
0.023705004	0.771428571	2.033928571	soap	<---	all- purpose, flour, sandwich bags, ketchup
0.024582968	0.8	2.109259259	soap	<---	all- purpose, flour, soda, ketchup
0.021071115	0.8	2.109259259	soap	<---	all- purpose, flour, sugar, ketchup
0.020193152	0.793103448	2.135566968	ketchup	<---	all- purpose, hand soap, lunch meat, soap
0.021071115	0.774193548	2.041218638	soap	<---	all- purpose, individual meals, lunch meat, ketchup
0.020193152	0.766666667	2.114366425	paper towels	<---	all- purpose, individual meals, toilet paper, cereals
0.021071115	0.75	2.068401937	paper towels	<---	all- purpose, individual meals, toilet paper, coffee/tea
0.021071115	0.75	1.950342466	yogurt	<---	all- purpose, milk, tortillas, coffee/tea
0.026338894	0.75	2.01	spaghetti sauce	<---	all- purpose, soda, juice, soap
0.024582968	0.8	2.047640449	soda	<---	all- purpose, waffles, laundry detergent, juice
0.027216857	0.775	1.983651685	cheeses	<---	bagels, cereals, coffee/tea, sandwich bags
0.023705004	0.771428571	1.974510433	cheeses	<---	bagels, cereals, pork, sandwich bags
0.029850746	0.772727273	1.977834525	cheeses	<---	bagels, cereals, sandwich bags, aluminum foil
0.020193152	0.766666667	2.026063418	toilet paper	<---	bagels, cereals, sugar, ketchup

POSSIBLE ASSOCIATIONS

Row Labels	Average of Lift	Average of Confidence	Average of Support
beef	**2.199330015**	**0.824507389**	**0.020632133**
poultry, fruits, hand soap, sugar	2.207542599	0.827586207	0.021071115
shampoo, fruits, lunch meat, pork	2.191117431	0.821428571	0.020193152
cheeses	**2.193900482**	**0.857142857**	**0.026338894**
paper towels, cereals, sandwich bags, sugar	2.193900482	0.857142857	0.026338894
ice cream	**2.21933243**	**0.884615385**	**0.020193152**
paper towels, eggs, dinner rolls, pasta, lunch meat	2.21933243	0.884615385	0.020193152
ketchup	**2.223082533**	**0.825604839**	**0.022827041**
toilet paper, mixes, coffee/tea, soap	2.187795508	0.8125	0.022827041
tortillas, coffee/tea, juice, soap	2.258369557	0.838709677	0.022827041
mixes	**2.266960886**	**0.851851852**	**0.020193152**
yogurt, dishwashing liquid/detergent, all- purpose, hand soap	2.266960886	0.851851852	0.020193152
paper towels	**2.248930967**	**0.815459604**	**0.022827041**
dishwashing liquid/detergent, eggs, ice cream, pasta, lunch meat	2.187275612	0.793103448	0.020193152
eggs, dinner rolls, ice cream, pasta, lunch meat	2.349296027	0.851851852	0.020193152
eggs, dinner rolls, poultry, ice cream, pasta	2.265392598	0.821428571	0.020193152
eggs, ice cream, pasta, lunch meat	2.19375963	0.795454545	0.030728709
pasta	**2.243892829**	**0.833333333**	**0.021949078**
paper towels, dishwashing liquid/detergent, eggs, dinner rolls, ice c	2.243892829	0.833333333	0.021949078
sandwich loaves	**2.252749469**	**0.787176724**	**0.021071115**
yogurt, hand soap, aluminum foil, soap	2.235788317	0.78125	0.021949078
yogurt, hand soap, toilet paper, soap	2.269710622	0.793103448	0.020193152
soda	**2.180357886**	**0.851851852**	**0.020193152**
bagels, pasta, individual meals, pork	2.180357886	0.851851852	0.020193152
spaghetti sauce	**2.217931034**	**0.827586207**	**0.021071115**
waffles, laundry detergent, mixes, soap	2.217931034	0.827586207	0.021071115
sugar	**2.183440242**	**0.787878788**	**0.022827041**
poultry, flour, waffles, beef	2.183440242	0.787878788	0.022827041

ASSOCIATIONS GROUPED BY RECOMMENDED PRODUCT IN THE PRECEDING ORDER OF LIFT

Recommended Product	Row Labels	Lift	Confidence	Support
paper towels	eggs, dinner rolls, ice cream, pasta, lunch meat	2.349296027	0.851851852	0.020193152
paper towels	eggs, dinner rolls, poultry, ice cream, pasta	2.265392598	0.821428571	0.020193152
paper towels	eggs, ice cream, pasta, lunch meat	2.19375963	0.795454545	0.030728709
paper towels	dishwashing liquid/detergent, eggs, ice cream, pasta, lunch meat	2.187275612	0.793103448	0.020193152
sandwich loaves	yogurt, hand soap, toilet paper, soap	2.269710622	0.793103448	0.020193152
sandwich loaves	yogurt, hand soap, aluminum foil, soap	2.235788317	0.78125	0.021949078
mixes	yogurt, dishwashing liquid/detergent, all- purpose, hand soap	2.266960886	0.851851852	0.020193152
ketchup	tortillas, coffee/tea, juice, soap	2.258369557	0.838709677	0.022827041
pasta	paper towels, dishwashing liquid/detergent, eggs, dinner rolls, ice cream	2.243892829	0.833333333	0.021949078
ice cream	paper towels, eggs, dinner rolls, pasta, lunch meat	2.21933243	0.884615385	0.020193152
spaghetti sauce	waffles, laundry detergent, mixes, soap	2.217931034	0.827586207	0.021071115
beef	poultry, fruits, hand soap, sugar	2.207542599	0.827586207	0.021071115
beef	shampoo, fruits, lunch meat, pork	2.191117431	0.821428571	0.020193152
cheeses	paper towels, cereals, sandwich bags, sugar	2.193900482	0.857142857	0.026338894
ketchup	toilet paper, mixes, coffee/tea, soap	2.187795508	0.8125	0.022827041
sugar	poultry, flour, waffles, beef	2.183440242	0.787878788	0.022827041
soda	bagels, pasta, individual meals, pork	2.180357886	0.851851852	0.020193152

RECOMMENDATIONS:

- Paper Towels go with the basket list, eggs, dinner rolls, ice cream, pasta , lunch meat / poultry. Hence eggs, ice cream, meat and paper towels go together. Paper towels can be given as complimentary or at **offer price** during purchase of ice cream, meat, eggs.

- Sandwich loaves and Yogurt go together and hence can be given as, **"Buy Sandwich Loaves, Get Yogurt for Free".**

- Ketchup can be tied along with Tortilla as a **combo pack.**

- Similarly, spaghetti sauce & waffles make a good combo. Combo offers can be provided to these combinatory purchases.

- Beef is associated with poultry, lunch meat and pork. Hence Beef can be given as an **Add-on** product during purchase of lunch meat, pork / poultry at an add-on offer price.

- Cheeses and sandwich bags go together. Lucrative offers can be given to Cheeses during bulk/more than 5 purchase of sandwich bags as a **'Family Pack' Offer.**