greatlearning
Learning for Life

Business report

Sales dataset

DATA SCIENCE AND BUSINESS ANALYTICS

greatlearning
Learning for Life

Content:

greatlearning
Learning for Life

1. Define the problem and conduct exploratory data analysis

1.1. Problem definition

Business Context:

In the highly competitive grocery retail industry, understanding customer buying patterns is crucial for enhancing sales, increasing customer satisfaction, and improving profitability. By identifying frequently purchased item combinations, grocery stores can craft effective marketing strategies, optimize inventory management, and tailor promotions to meet customer needs. Leveraging Point of Sale (POS) data can unlock valuable insights that drive customer-centric offerings, such as combo packs, discounts, and targeted promotions, which can increase basket size and improve customer retention. This analysis aligns with business goals by maximizing revenue, reducing operational costs, and boosting customer loyalty.

Objectives:

As a business analyst, the goal is to analyze the POS transactional data to identify frequently purchased item combinations. Using association rule mining or similar techniques, the aim is to uncover patterns that will help the store create targeted combo offers and discounts, ultimately driving revenue growth by increasing customer purchases and average basket size.

Data Description:

The dataset consists of transactional data from a grocery store, where each row represents a product purchased in a specific order. The columns in the dataset are as follows:

- Date: The date when the transaction took place.
- Order_id: A unique identifier for each customer order.
- Product: The individual item purchased in the transaction.

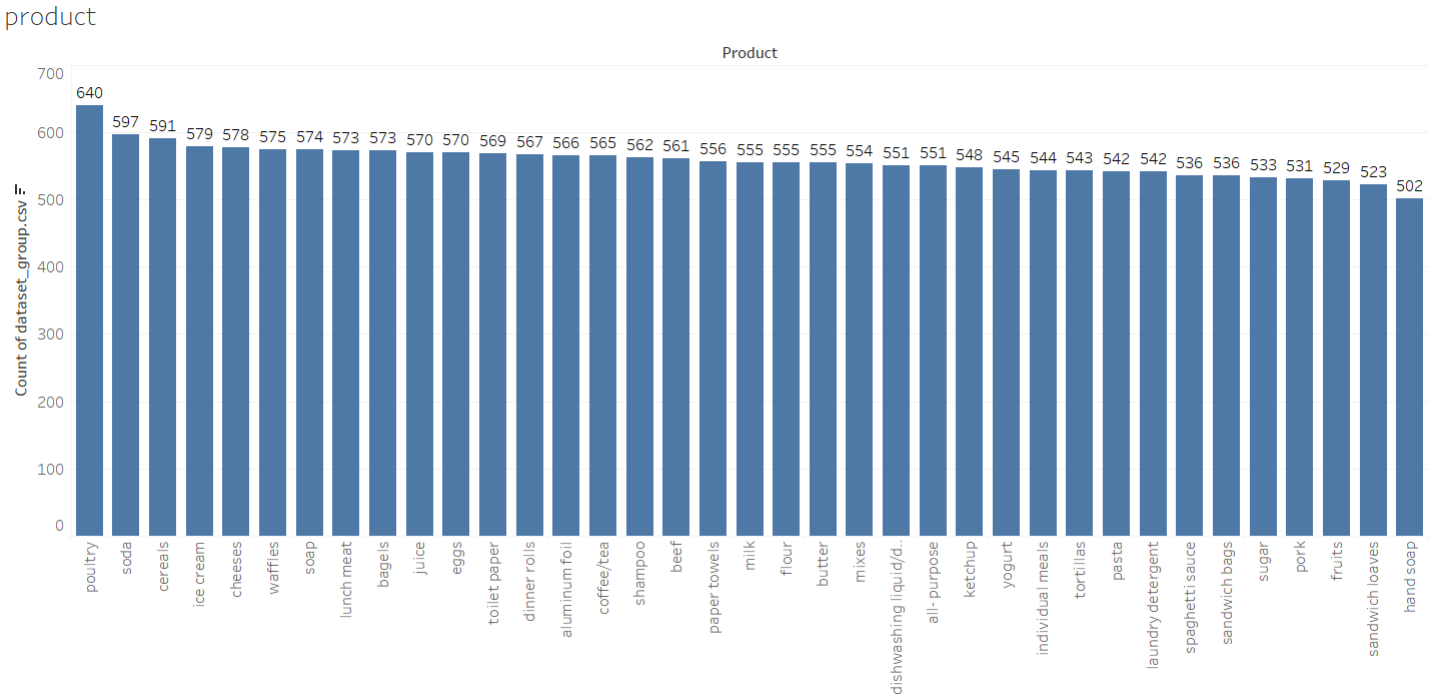
1.2. Statistical Summary:

	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

- **Order id** is the only item which is in integer datatype because of which statistical summary is showing only order id not date and product.

1.3. Univariate analysis.

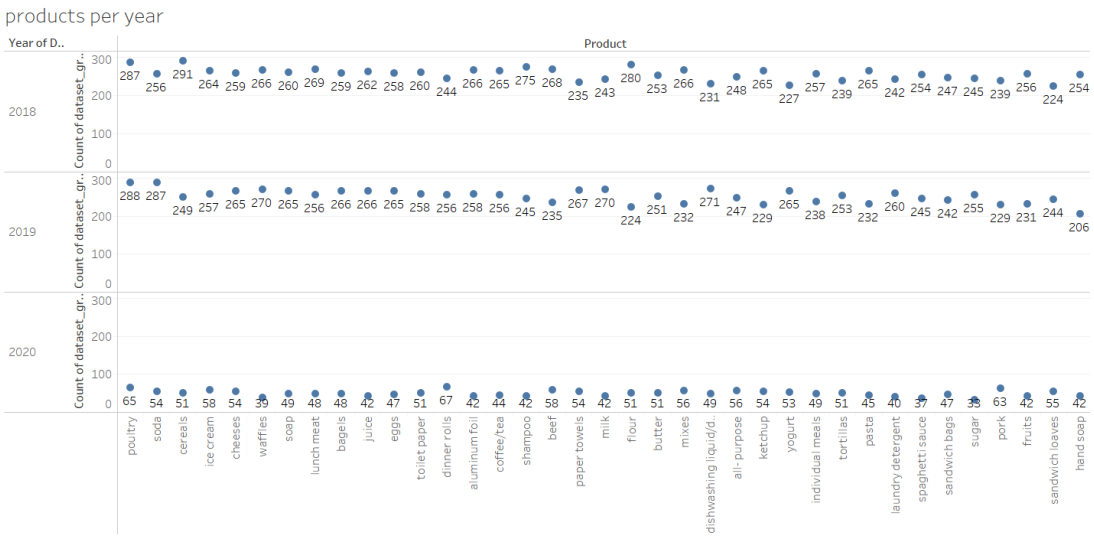
- **Product:**



- Poultry has the highest sales across all the products with the count of 640.
- Second highest consumer purchased product is soda.
- As per the above plot hand soap is the product which is purchased less but still it purchased by 502 times.

1.4. Bivariate analysis and time series analysis:

- **Products sold per year:**



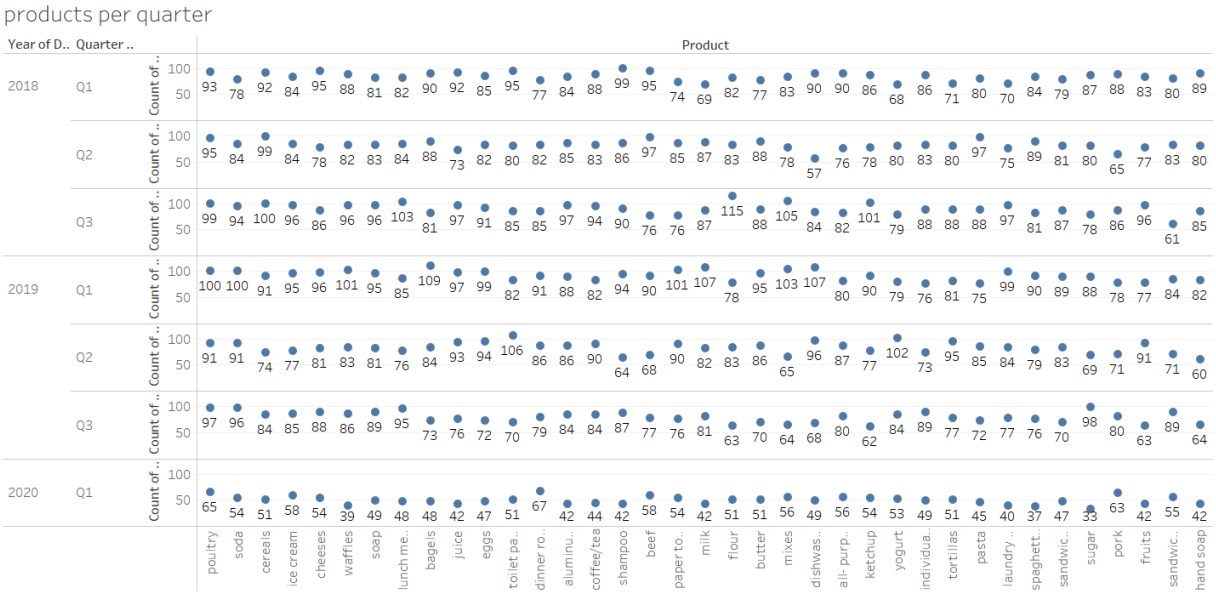
products per year(text)

Count of dataset_...
33 291

Product	Date 2018	2019	2020	Grand Total
poultry	287	288	65	640
soda	256	287	54	597
cereals	291	249	51	591
ice cream	264	257	58	579
cheeses	259	265	54	578
waffles	266	270	39	575
soap	260	265	49	574
lunch meat	269	256	48	573
bagels	259	266	48	573
juice	262	266	42	570
eggs	258	265	47	570
toilet paper	260	258	51	569
dinner rolls	244	256	67	567
aluminum foil	266	258	42	566
coffee/tea	265	256	44	565
shampoo	275	245	42	562
beef	268	235	58	561
paper towels	235	267	54	556
milk	243	270	42	555
flour	280	224	51	555
butter	253	251	51	555
mixes	266	232	56	554
dishwashing liquid/..	231	271	49	551
all- purpose	248	247	56	551
ketchup	265	229	54	548
yogurt	227	265	53	545
individual meals	257	238	49	544
tortillas	239	253	51	543
pasta	265	232	45	542
laundry detergent	242	260	40	542
spaghetti sauce	254	245	37	536
sandwich bags	247	242	47	536
sugar	245	255	33	533
pork	239	229	63	531
fruits	256	231	42	529
sandwich loaves	224	244	55	523
hand soap	254	206	42	502

- As said above these above graphs are also showing the same poultry and soda have the highest sales of 640 and 597.
- Sale has been increased year by year but the data of 2020 has a single quarter that's why it's showing the less sales.
- Most of the product purchases have increasing trend but some of them have decreasing trend.

- **Products sold per quarter:**



product per quarter(text)

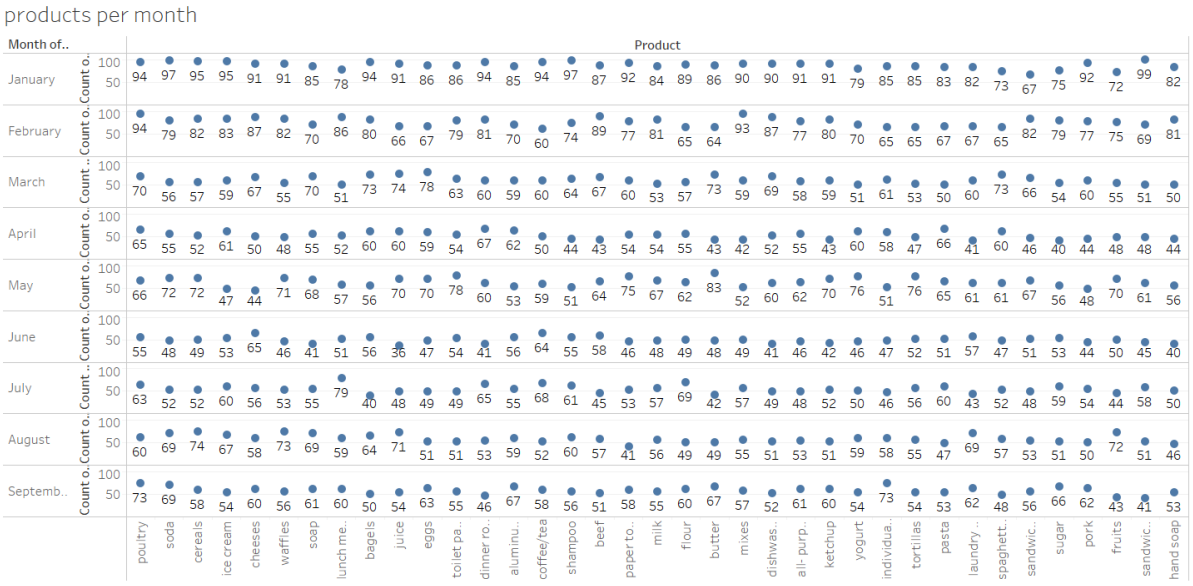
Count of dataset_..
136 258

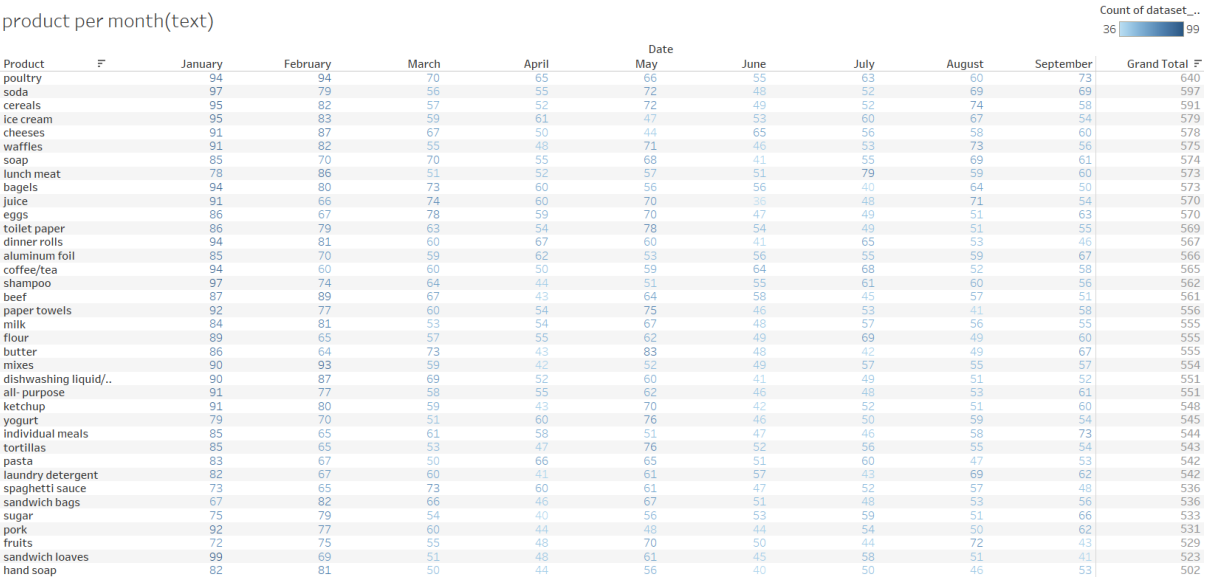
Product	Date Q1	Q2	Q3	Grand Total
poultry	258	186	196	640
soda	232	175	190	597
cereals	234	173	184	591
ice cream	237	161	181	579
cheeses	245	159	174	578
waffles	228	165	182	575
soap	225	164	185	574
lunch meat	215	160	198	573
bagels	247	172	154	573
juice	231	166	173	570
eggs	231	176	163	570
toilet paper	228	186	155	569
dinner rolls	235	168	164	567
aluminum foil	214	171	181	566
coffee/tea	214	173	178	565
shampoo	235	150	177	562
beef	243	165	153	561
paper towels	229	175	152	556
milk	218	169	168	555
flour	211	166	178	555
butter	223	174	158	555
mixes	242	143	169	554
dishwashing liquid/..	246	153	152	551
all- purpose	226	163	162	551
ketchup	230	155	163	548
yogurt	200	182	163	545
individual meals	211	156	177	544
tortillas	203	175	165	543
pasta	200	182	160	542
laundry detergent	209	159	174	542
spaghetti sauce	211	168	157	536
sandwich bags	215	164	157	536
sugar	208	149	176	533
pork	229	136	166	531
fruits	202	168	159	529
sandwich loaves	219	154	150	523
hand soap	213	140	149	502

- Same as told above poultry and soda has highest sale.
- Quarter 1 is showing the highest sales but it might be because we have only 1st quarter data in 2020.

greatlearning
Learning for Life

- **Products sold per month:**



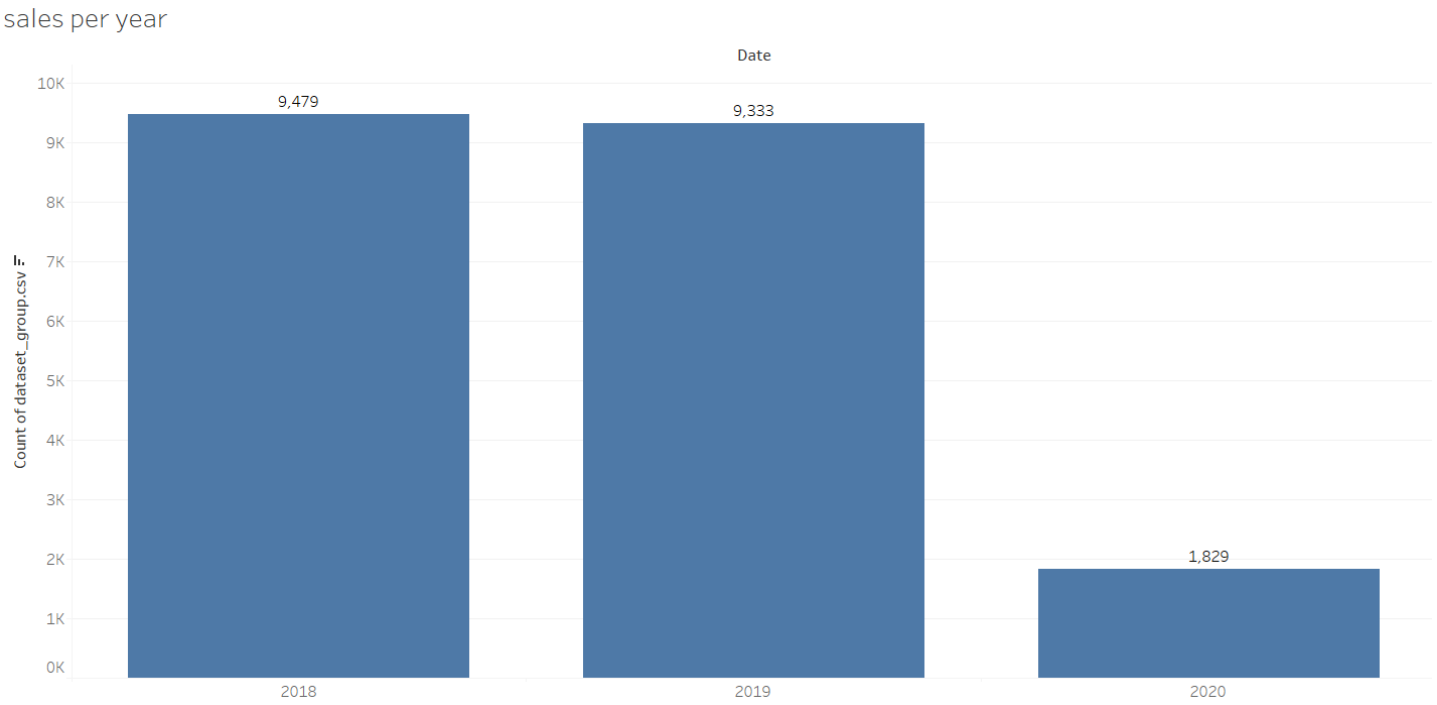


Product	January	February	March	April	May	June	July	August	September	Grand Total
poultry	94	94	70	65	66	55	63	60	73	640
soda	97	79	56	55	72	48	52	69	69	597
cereals	95	82	57	52	72	49	52	74	58	591
ice cream	95	83	59	61	47	53	60	67	54	579
cheeses	91	87	67	50	44	65	56	58	60	578
waffles	91	82	55	48	71	46	53	73	56	575
soap	85	70	70	55	68	41	55	69	61	574
lunch meat	78	86	51	52	57	51	79	59	60	573
bagels	94	80	73	60	56	56	40	64	50	573
juice	91	66	74	60	70	36	48	71	54	570
eggs	86	67	78	59	70	47	49	51	63	570
toilet paper	86	79	63	54	78	54	49	51	55	569
dinner rolls	94	81	60	67	60	41	65	53	46	567
aluminum foil	85	70	59	62	53	56	55	59	67	566
coffee/tea	94	60	60	50	58	64	68	52	58	565
shampoo	97	74	64	44	51	55	61	60	56	562
beef	87	89	67	43	64	58	45	57	51	561
paper towels	92	77	60	54	75	46	53	41	58	556
milk	84	81	53	54	67	48	57	56	55	555
flour	89	65	57	55	62	49	69	49	60	555
butter	86	64	73	43	83	48	42	49	67	555
mixes	90	93	59	42	52	49	57	55	57	554
dishwashing liquid/..	90	87	69	52	60	41	49	51	52	551
all- purpose	91	77	58	55	62	46	48	53	61	551
ketchup	91	80	59	43	70	42	52	51	60	548
yogurt	79	70	51	60	76	46	50	59	54	545
individual meals	85	65	61	58	51	47	46	58	73	544
tortillas	85	65	53	47	76	52	56	55	54	543
pasta	83	67	50	66	65	51	60	47	53	542
laundry detergent	82	67	60	41	61	57	43	69	62	542
spaghetti sauce	73	65	73	60	61	47	52	57	48	536
sandwich bags	67	82	66	46	67	51	48	53	56	536
sugar	75	79	54	40	56	53	59	51	66	533
pork	92	77	60	44	48	44	54	50	62	531
fruits	72	75	55	48	70	50	44	72	43	529
sandwich loaves	99	69	51	48	61	45	58	51	41	523
hand soap	82	81	50	44	56	40	50	46	53	502

- It has the highest sales in beginning months like January and February.
- Company can launch new products at this month to increase sales
- In remaining months company should do marketing to improve sales.

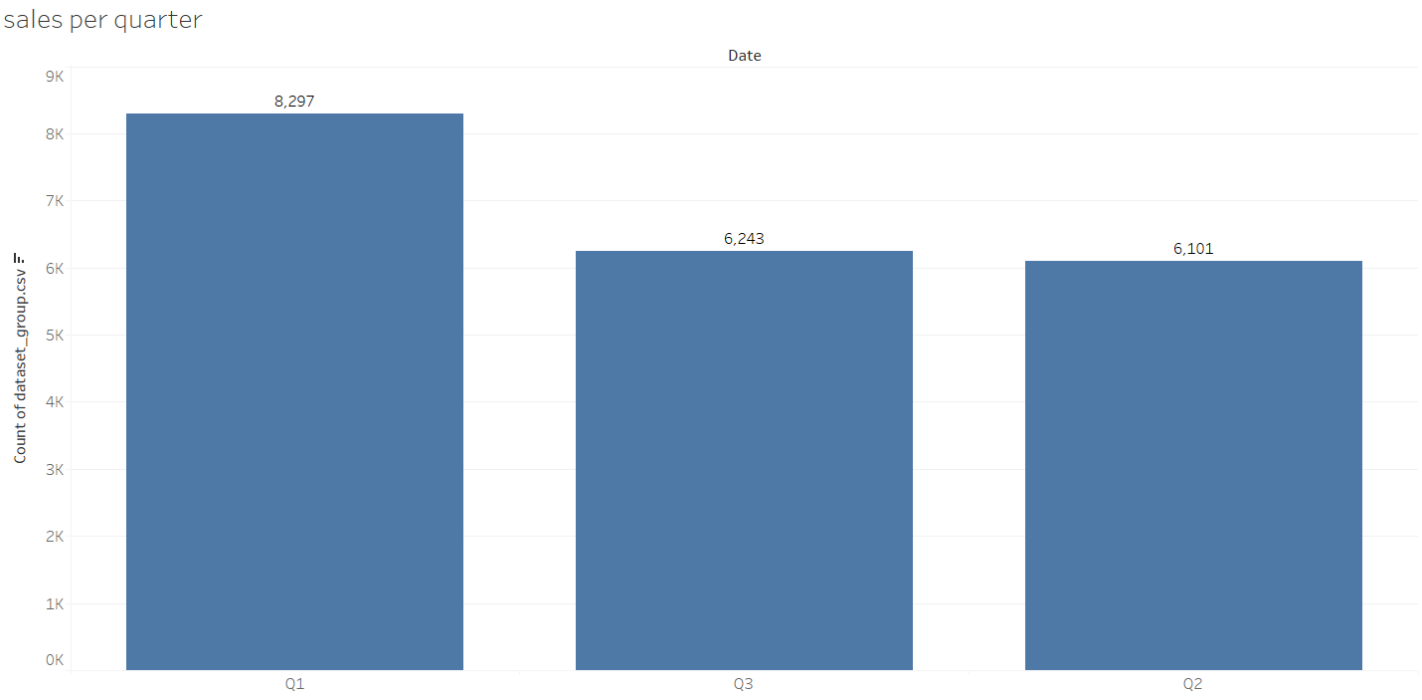
1.5. Multivariate analysis.

- **Sales per Year:**



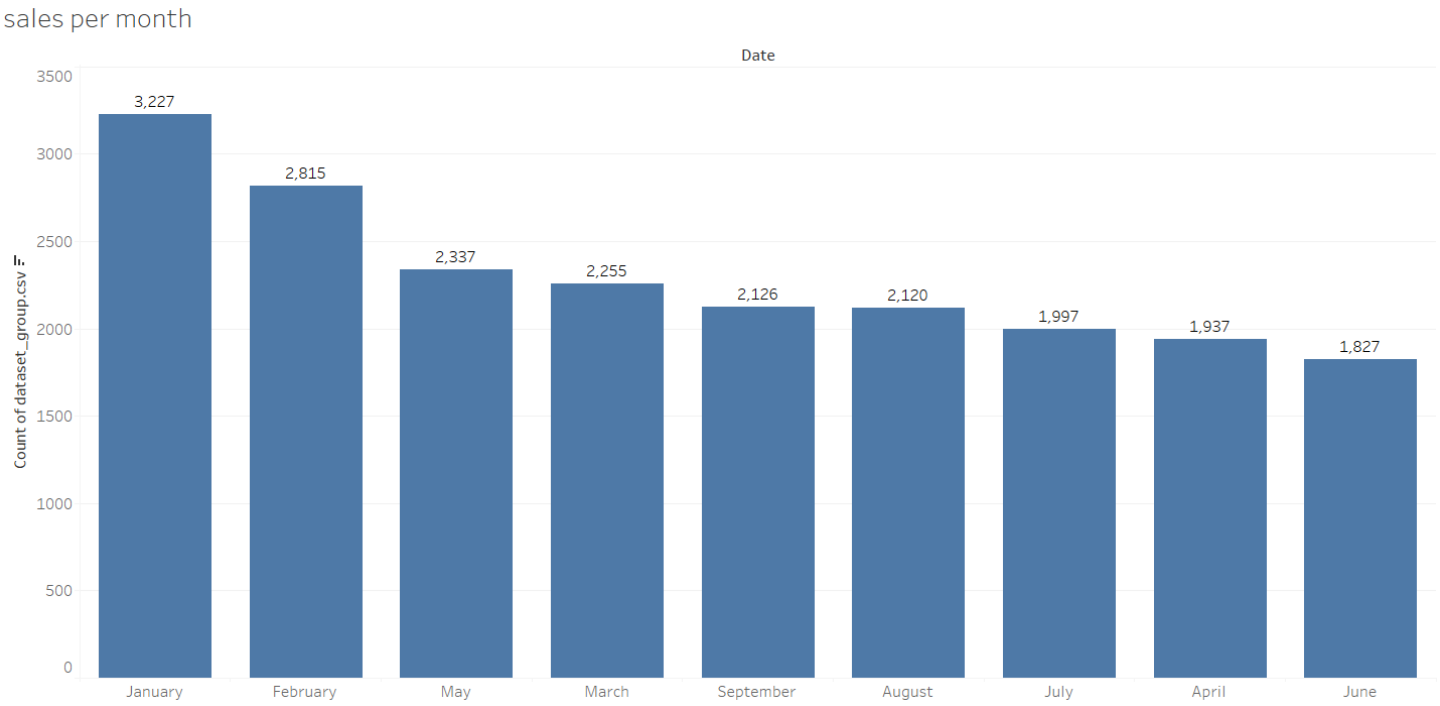
- Sales has been decreased from 2018 to 2019.
- Company should change the strategies and try to improve sales by doing marketing and all.
- Company must try to understand what all the things which are affecting the sales, that to be understood and take actions against those things.

- **Sales per quarter:**



- Quarter 1 has the highest sales as said above.
- Sales decreases from 1st quarter to 2nd quarter but later on increased in 3rd quarter which must be understood why is it happening.
- Try promotional activities in 2nd quarter.

- **Sales per month:**



- Highest sales in the month of January.
- Then in the month of February.
- And a very low sales in the month of June.

2. Market Basket Analysis

2.1. The process of using association rules in the context of this project

Association rules in this project were used to identify patterns in customer purchasing behavior. Using the **Association rules**, we detected which items are frequently purchased together. These patterns are valuable for:

- Designing promotional bundles
- Cross-selling products
- Planning store layout or e-commerce recommendations

Steps followed:

1. Transactions were grouped by Order ID.
2. Items within each order were treated as itemsets.
3. The Association rules was applied via KNIME to extract association rules.
4. Rules were filtered based on support and confidence thresholds to ensure relevance.

2.2. The threshold values of support and confidence

Metric	Value Used
Support	0.03 (5%)
Confidence	0.65 (65%)

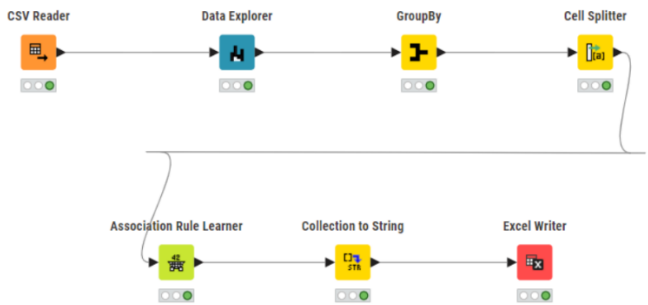
Why these thresholds?

- **Support = 0.03** ensures we capture rules appearing in at least **3%** of total transactions, focusing on common purchasing behavior.
- **Confidence = 0.65** helps ensure the rules have predictive strength, i.e., if a customer buys item A, there's at least a **65%** chance they will also buy item B.

2.3. KNIME workflow

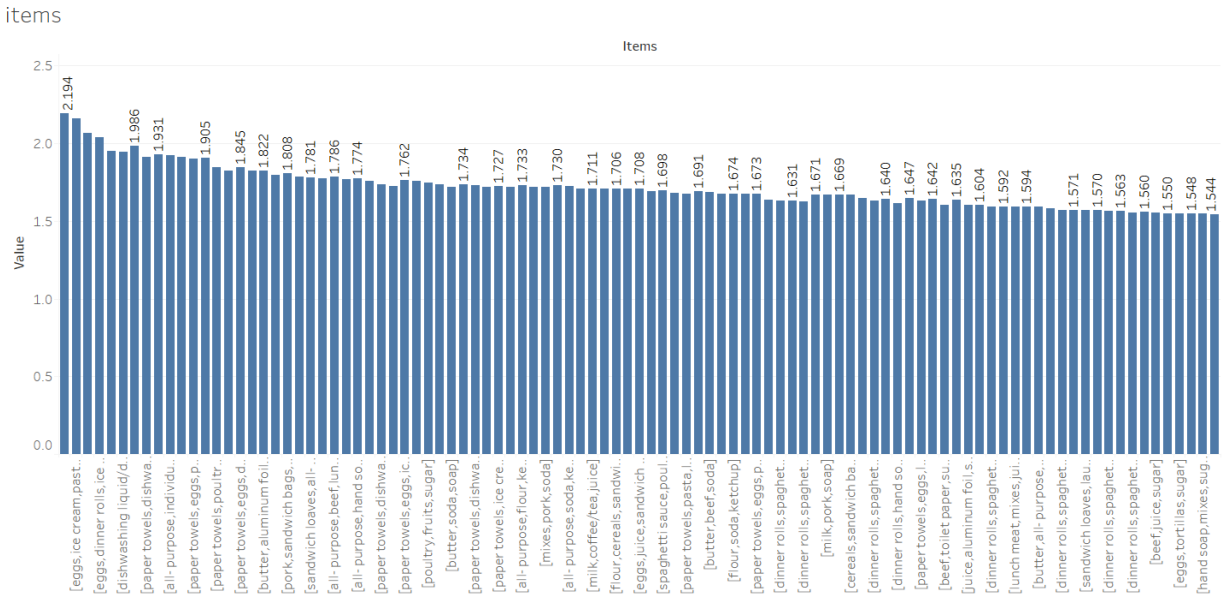
Below is a simplified explanation of your KNIME pipeline used for Market Basket Analysis:

1. **CSV Reader** – Loaded sales data (Dataset_group.csv)
2. **Data Explorer** – To explore the data
3. **GroupBy Node** – Grouped products by Order ID
4. **Cell Splitter** – Converted the item list into individual items
5. **Association Rule Learner** – Applied association rule algorithm with specified support and confidence



2.4. Arrange the associations obtained in a tabular manner

#	RowID	Support	Confidence	Lift	Consequent	implies	Items
		Number (doub..	Number (double)	Number (double)	String	String	Set
3	rule2	0.031	0.795	2.194	paper towels	<---	[eggs,ice cream,pasta,...]
16	rule15	0.032	0.783	2.158	paper towels	<---	[eggs,ice cream,pasta,...]
4	rule3	0.031	0.729	2.066	flour	<---	[dishwashing liquid/detergent,cheeses,waffles,...]
25	rule24	0.032	0.74	2.041	paper towels	<---	[eggs,dinner rolls,ice cream,...]
20	rule19	0.032	0.72	1.986	paper towels	<---	[eggs,poultry,ice cream,...]
1	rule0	0.031	0.778	1.951	ice cream	<---	[paper towels,eggs,pasta,...]
5	rule4	0.031	0.761	1.947	soda	<---	[dishwashing liquid/detergent,cheeses,flour,...]
29	rule28	0.033	0.717	1.931	pasta	<---	[paper towels,dishwashing liquid/detergent,eggs,...]
54	rule53	0.04	0.697	1.922	paper towels	<---	[all- purpose,individual meals,toilet paper]
11	rule10	0.031	0.714	1.914	spaghetti sauce	<---	[dinner rolls,poultry,laundry detergent,...]
28	rule27	0.033	0.745	1.911	eggs	<---	[paper towels,dishwashing liquid/detergent,ice cream,...]
30	rule29	0.033	0.691	1.905	paper towels	<---	[dishwashing liquid/detergent,eggs,ice cream,...]
19	rule18	0.032	0.706	1.901	pasta	<---	[paper towels,eggs,poultry,...]
18	rule17	0.032	0.72	1.847	eggs	<---	[paper towels,poultry,ice cream,...]
24	rule23	0.032	0.685	1.845	pasta	<---	[paper towels,eggs,dinner rolls,...]
23	rule22	0.032	0.712	1.825	eggs	<---	[paper towels,dinner rolls,ice cream,...]
55	rule54	0.04	0.676	1.822	pasta	<---	[hand soap,soda,aluminum foil]
56	rule55	0.04	0.676	1.822	ketchup	<---	[butter,aluminum foil,soap]
57	rule56	0.041	0.671	1.808	ketchup	<---	[pork,sandwich bags,soap]
9	rule8	0.031	0.7	1.796	eggs	<---	[dishwashing liquid/detergent,ice cream,pasta,...]
39	rule38	0.038	0.662	1.786	fruits	<---	[all- purpose,beef,lunch meat]
46	rule45	0.039	0.688	1.784	bagels	<---	[sandwich loaves,fruits,juice]



- **Eggs, ice cream, pasta, lunch meat –** These are the basket which is highly used by the customers.
- **Eggs, ice cream, pasta and cereals –** These are 2nd highly used basket.
- **Eggs, ice cream and pasta** are most commonly purchased consumables.

2.5. Explain the support, confidence and lift values calculated

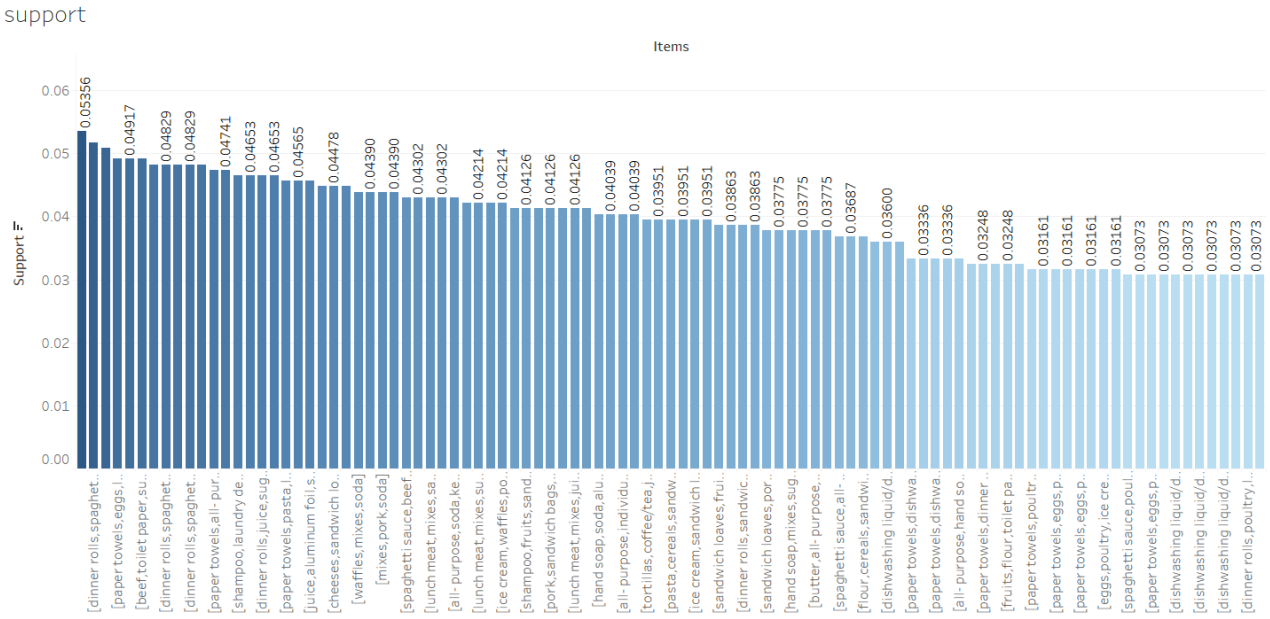
Support: Frequency of a rule in the dataset.
Example: A support of 0.03 means the combination appears in 3% of all transactions.

Confidence: Likelihood of the consequent being purchased when the antecedent is bought.
Example: A confidence of 0.68 means 68% of the time when a poultry is bought, a dinner rolls is also purchased.

Lift: Measures how much more likely the consequent is purchased with the antecedent than by chance.
Lift > 1 implies a strong relationship.

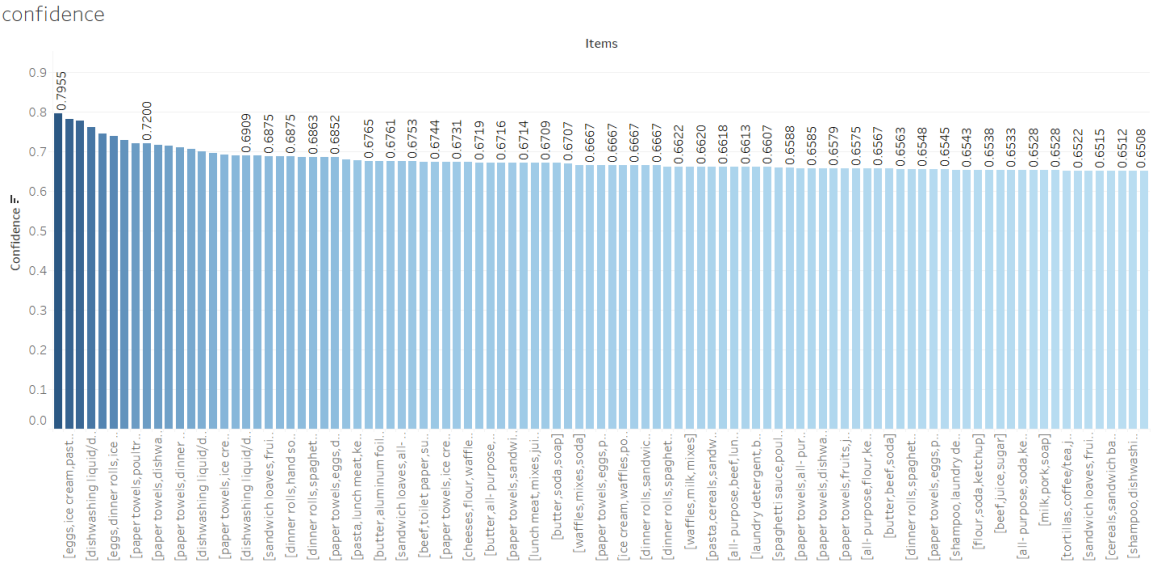
Example: A lift of 1.98 indicates the rule is 98% stronger than random co-occurrence.

Support:



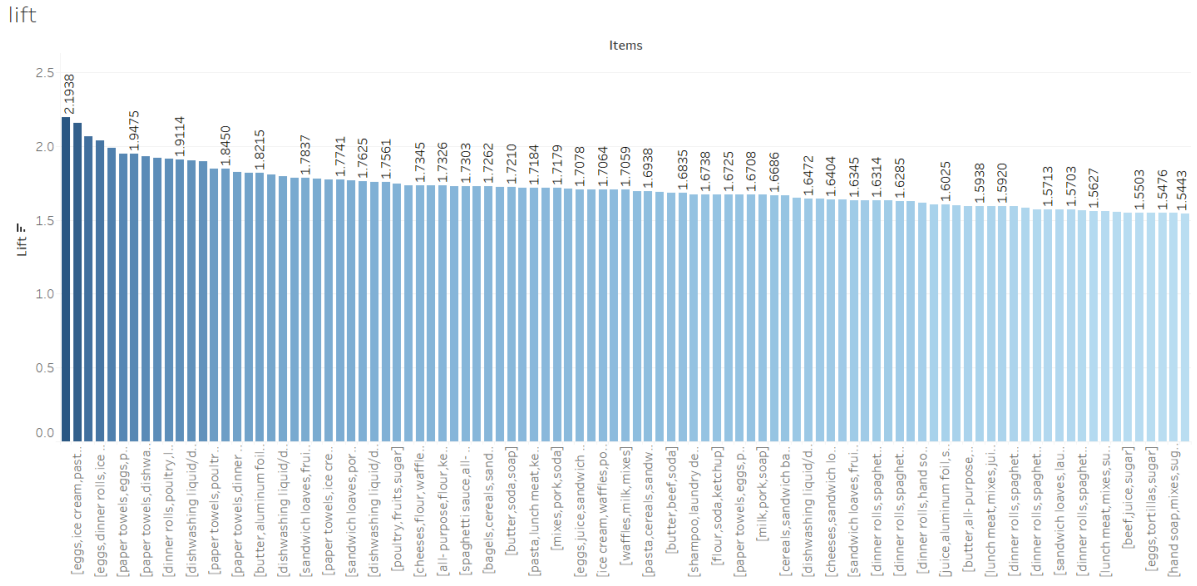
- After giving 5% support we get **dinner rolls, spaghetti sauce, laundry detergent** shows the highest support.
- **Dinner rolls, spaghetti sauce and ice cream** are having the 2nd highest support.

Confidence:



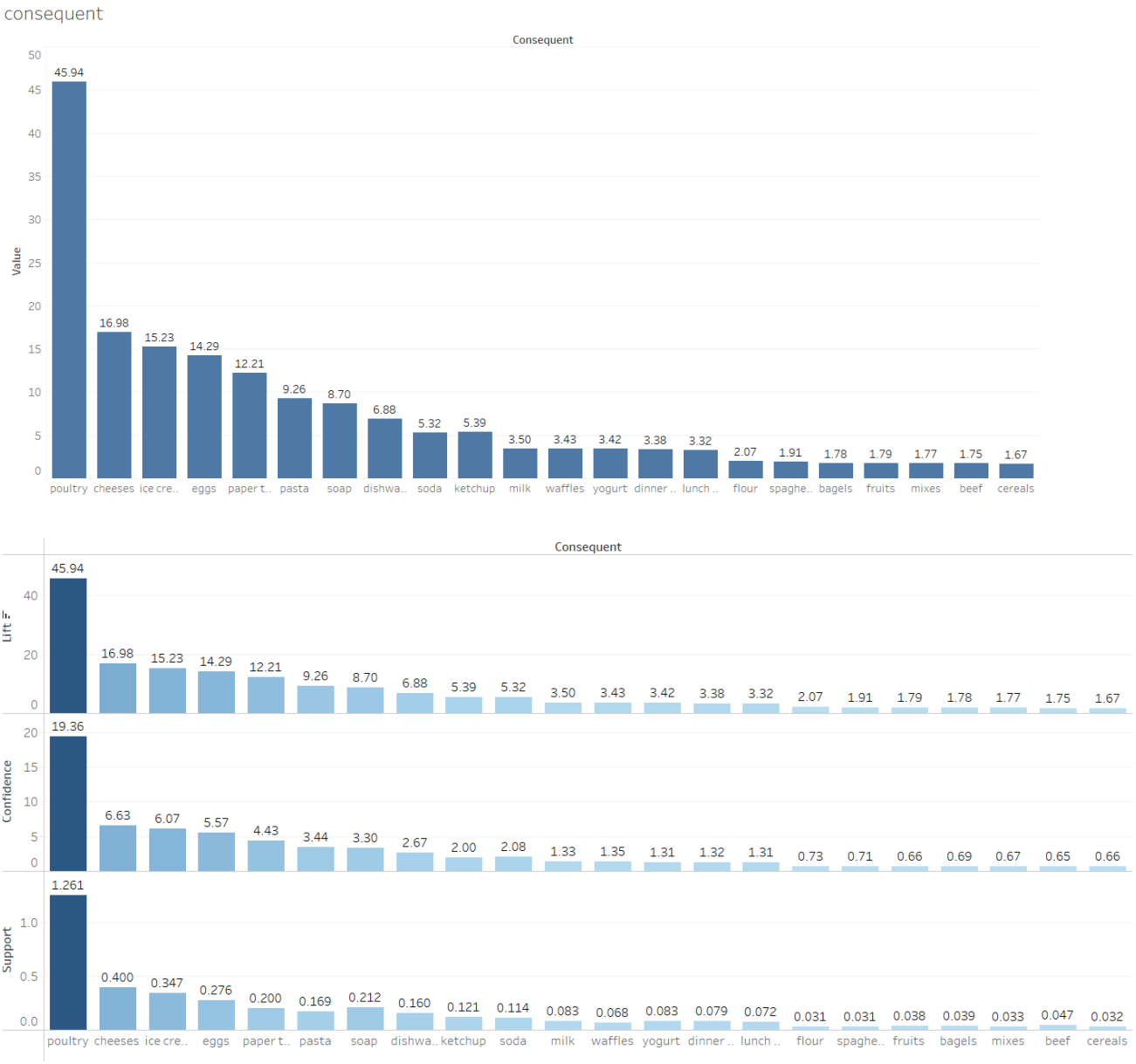
- After giving 65% of confidence **Eggs, ice cream, pasta, lunch meat** – are the bundle which got the highest confidence of 80%.
- **Same** as said above are the 2nd highest confidence but here in 3rd highest confidence **paper towel, eggs, pasta and lunch meat.**
- **Eggs, ice cream, pasta, lunch meat and paper towel** can place together to increase sales.

Lift:



- As the arrange shown above shows the highest lift across all the items.

Most purchased items:



- **Poultry** has the highest purchases
- **Cheeses** is 2nd highest sales
- **Poultry** should be placed between the items which makes a bundle because of poultry others things can be sold as well. Bundles such as poultry, dinner rolls, spaghetti sauce.
- **Cheese** can be placed with sandwich bags and cereals.

3. Inferences and Recommendations

3.1. Key insights from the analysis conducted

Based on the association rules generated using support = 0.03 and confidence = 0.65, the following insights were observed:

- Common Pairings: Certain products are frequently purchased together, such as:
    - Dinner rolls ↔ poultry
    - Sandwich bags + Cheese ↔ Cereals
    - Eggs ↔ Paper towels
- Product Affinity:
    - Items like Dinner rolls, Poultry, and Spaghetti Sauce often act as strong consequents, indicating they are commonly added to the cart after certain items.
- Lift > 1 in most strong rules, meaning the items are being bought together significantly more often than by random chance.
- Eggs and paper towels showed a high correlation in purchases, especially in mid-size bulk orders.
- Some items, such as eggs, paper towels or ice cream, indicate consumables are often purchased together, suggesting routine office restocking.

3.2. Business recommendations

Based on the analysis and insights gained, here are practical business recommendations:

Combo Offers

- Dinner rolls + Poultry
  → Bundle Price or 10% discount if bought together.

- Sandwich bags + Cheese + Cereals
  → "Office Essentials Pack" with 1 free item on purchase of 3.

- Paper towels + ice cream
  → Consumable bundle: Buy one, get 25% off on the second.

- Soda + All-purposes
  → Offer combo deals during month-end/quarter-end when businesses restock.

Cross-selling Recommendations

- Suggest Sandwich bags on the cart page if Cheese is added.

- Recommend Eggs when Paper towel is in the cart (or vice versa).
  "Buy Two, Get One Free"

- On high-frequency items like dinner rolls, Poultry, Spaghetti Sauce, run promotions like:
  - *"Buy Poultry, get 1 Dinner rolls free"*
  - *"Buy 3 Dinner rolls, get 1 Spaghetti Sauce free"*

Inventory Planning

- Increase stock of items like Paper towels, Eggs, and Ice cream that appear in multiple high-confidence rules.

- Prepare bundled kits to simplify logistics and improve upselling.

Targeted Marketing

- Use insights to:
  - Send targeted promotions to customers who bought Waffles, encouraging dishwashing liquid/ detergent add-ons.
  - Create email campaigns offering bundled discounts for ice cream or consumables.
