Exploratory Data Analysis and Market Basket Analysis

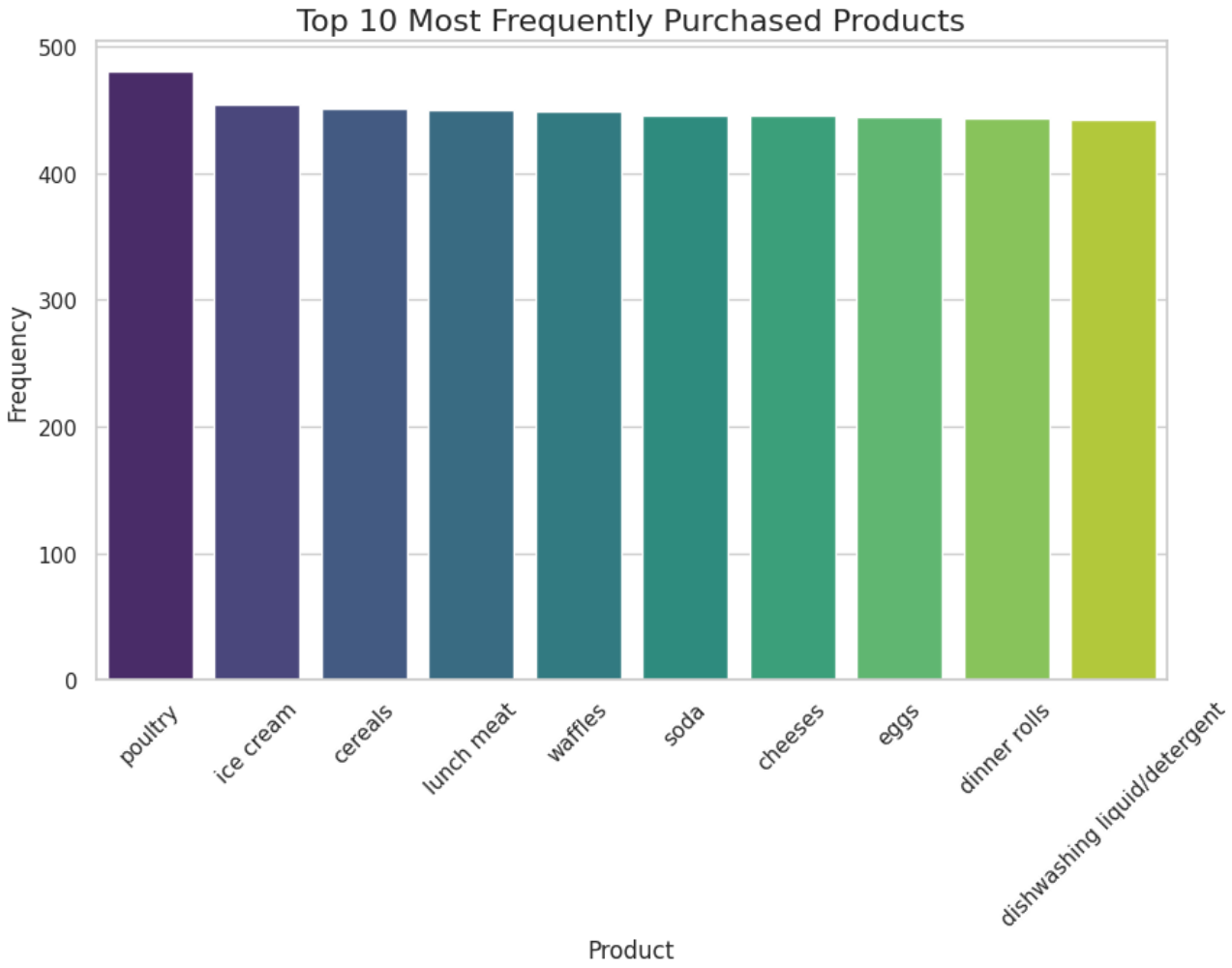
Problem Statement

Identify frequently purchased item combinations using POS data to optimize marketing strategies, inventory management, and promotions, driving revenue growth, cost efficiency, and customer loyalty in the competitive grocery retail industry.

Objective

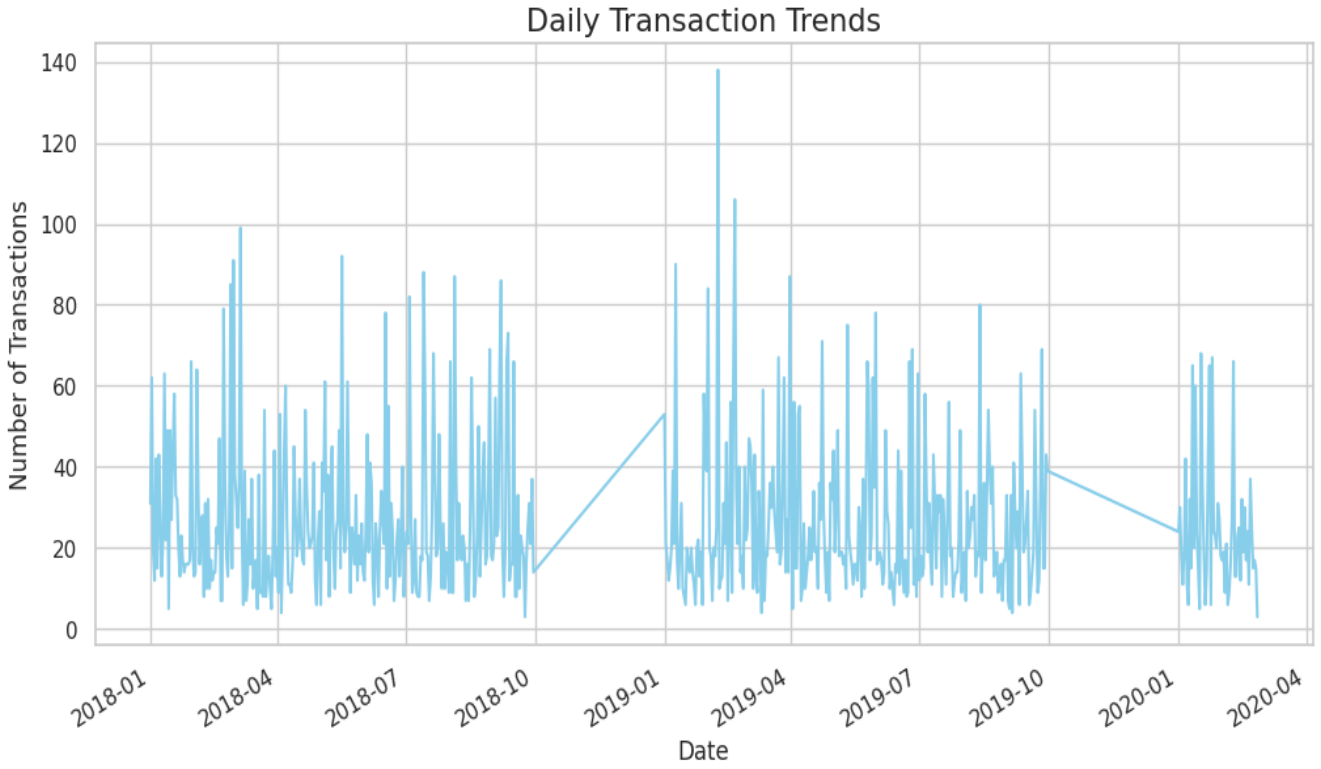
- Analyze POS transactional data to uncover customer buying patterns.
- Identify frequently purchased item combinations using association rule mining or similar techniques.
- Use insights to create targeted combo offers and discounts.Drive revenue growth by increasing customer purchases and average basket size.

Top 10 Most Frequently Purchased Products



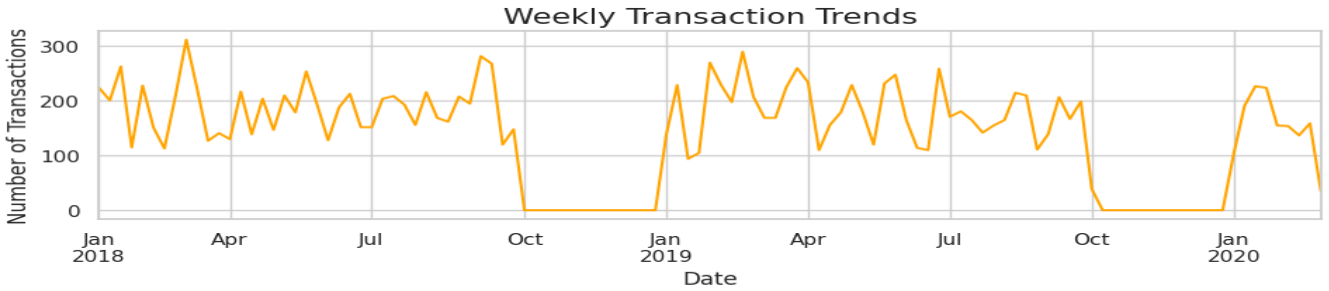
- Poultry is the most frequently purchased product, with over 500 transactions.
- Ice cream and cereals are the next most frequently purchased, each with over 350 transactions.
- Lunch meats, waffles, and soda all have between 275-300 transactions.
- Cheese, eggs, and dinner rolls have between 225-250 transactions.
- Dishwashing liquid/detergent is the least frequently purchased product in the top 10, with just under 175 transactions.

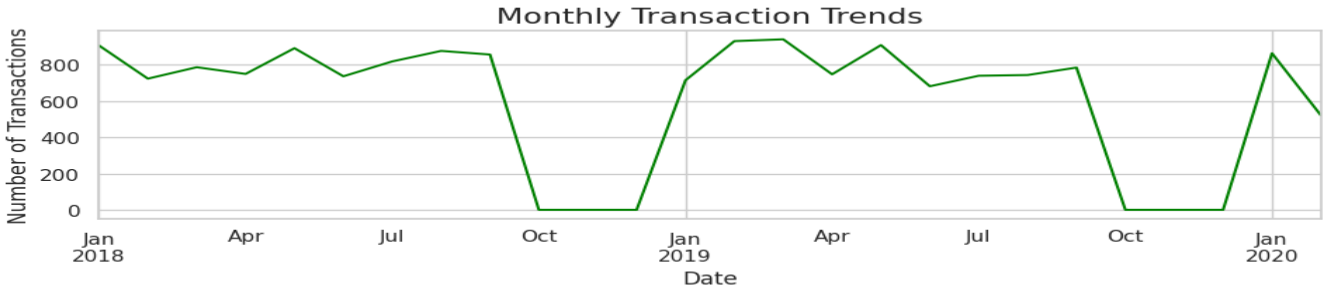
Daily Transaction Trends

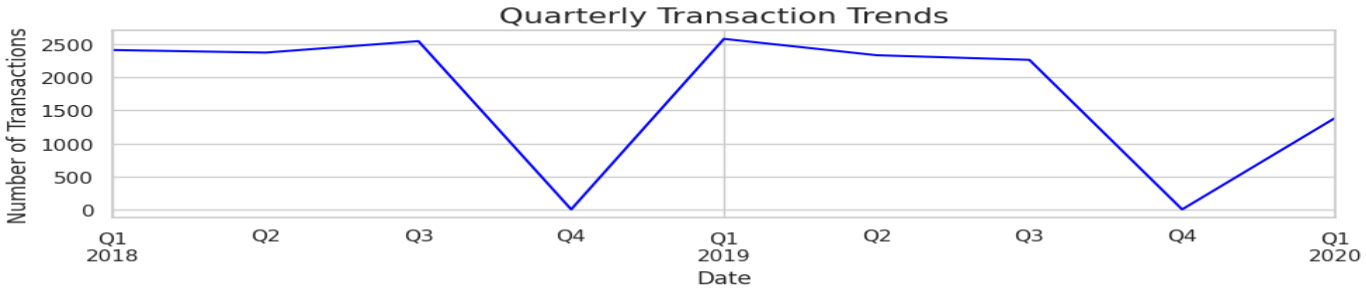


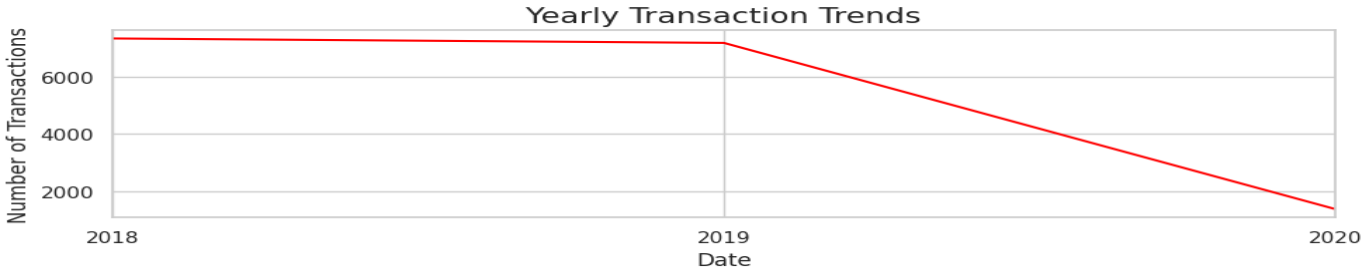
- Significant daily fluctuations: The number of transactions varies widely from day to day, with some peaks reaching over 100 transactions and valleys dipping below 20 transactions.
- Weekly patterns: There appears to be a weekly cyclical pattern, with higher transaction volumes on certain days of the week, likely corresponding to weekends or peak shopping days.
- Seasonal trends: Broader seasonal variations are also visible, with higher transaction counts in the middle of the year compared to the beginning and end.
- Anomalous spikes: A few extreme spikes in transaction volume stand out, potentially indicating special events, promotions, or other unusual circumstances.

Weekly, Monthly, Quarterly, and Yearly trends in Purchases









Weekly Trends:

- Weekly transaction volumes fluctuate significantly, ranging from around 50 to over 250 transactions per week.
- There is a clear weekly cyclical pattern, with higher transaction counts on weekends and lower volumes during weekdays.

Monthly Trends:

- Monthly transaction totals vary from around 400 to over 800, also exhibiting seasonal patterns.
- Transaction volumes tend to peak in the middle of the year and decline towards the end.

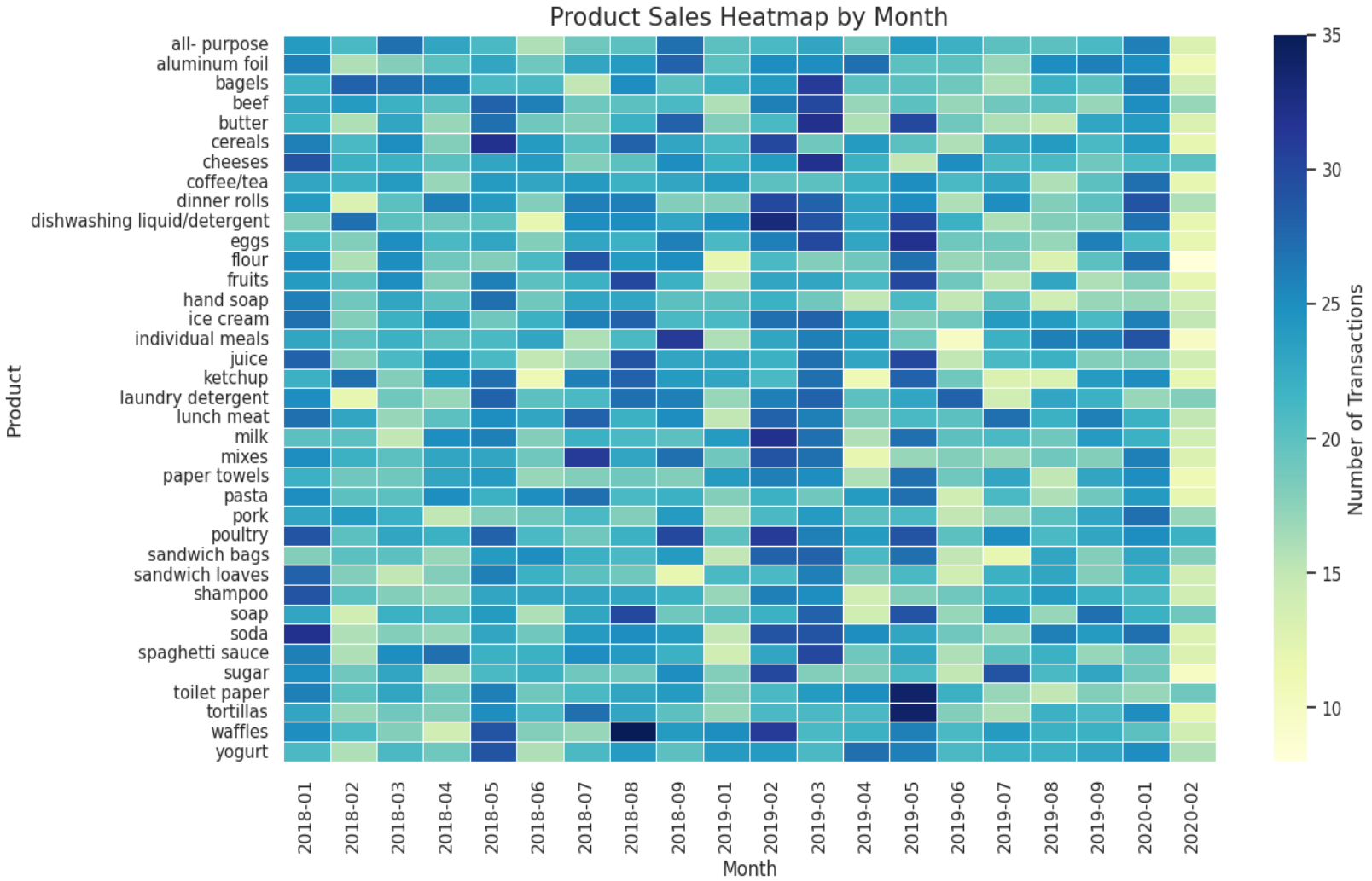
Quarterly Trends:

- Quarterly transaction totals range from around 1,500 to over 2,300.
- There is a clear quarterly seasonality, with Q4 being the busiest season.

Yearly Trends:

- It’s almost constant. The 2020 data is only available for the first two months.

Product Sales Heatmap



- Seasonal patterns in sales for many products
  - Higher sales of cereal, lunch meat, and dinner rolls in colder months
  - Higher sales of ice cream and fruits in warmer months
- Consistent year-round sales for poultry, cheese, and eggs
- High-volume product categories include poultry, ice cream, cereals, lunch meat, and waffles

Market Basket Analysis

Overview

Market Basket Analysis (MBA) identifies relationships between items purchased together. It uses transactional data to derive actionable insights.

Key Concepts

Association Rules:
E.g., If a customer buys *bread*, they are likely to buy *butter*.

Metrics Used:

- Support: Frequency of the itemset.
- Confidence: Likelihood of items being purchased together.
- Lift: Strength of the association (values > 1 indicate a positive correlation).

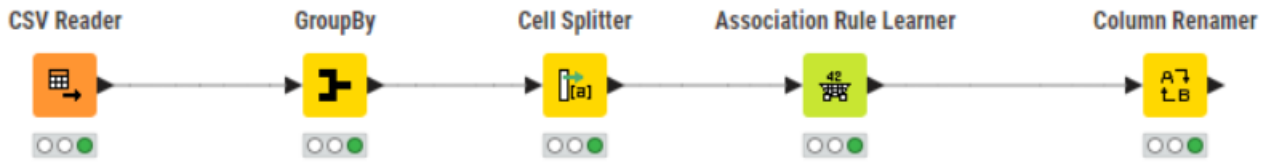
Benefits

- Increases revenue through targeted strategies.
- Enhances customer satisfaction with personalized offerings.
- Optimizes inventory management.

Association Rules

- Transformed the transaction data.
- Applied the association rule to identify frequent itemsets, using a minimum support threshold of 0.05 (5%).
- Generated association rules based on the frequent itemsets, with a minimum confidence threshold of 0.5 (50%).
- Evaluated the generated rules by calculating the support, confidence, and lift values.
- Sorted the rules by lift to surface the strongest product associations.

KNIME Workflow



Association Rule Output

RowID	Support *Number (double)*	Confidence *Number (double)*	Lift ↓ *Number (double)*	Rec_Item *String*	imp... *String*	Basket_Items *Set*
rule59	0.055	0.649	1.791	paper towels	<---	[eggs,ice cream,pasta]
rule58	0.055	0.643	1.731	pasta	<---	[paper towels,eggs,ice cream]
rule21	0.051	0.674	1.726	cheeses	<---	[bagels,cereals,sandwich bags]
rule3	0.05	0.64	1.7	juice	<---	[yogurt,toilet paper,aluminum foil]
rule18	0.051	0.63	1.678	mixes	<---	[yogurt,poultry,aluminum foil]
rule20	0.051	0.611	1.66	sandwich bags	<---	[cheeses,bagels,cereals]
rule52	0.054	0.642	1.651	dinner rolls	<---	[spaghetti sauce,poultry,laundry deterg
rule40	0.052	0.641	1.649	dinner rolls	<---	[spaghetti sauce,poultry,ice cream]
rule7	0.05	0.62	1.645	juice	<---	[yogurt,poultry,aluminum foil]
rule43	0.052	0.686	1.628	poultry	<---	[dinner rolls,spaghetti sauce,ice cream
rule49	0.052	0.634	1.627	eggs	<---	[paper towels,dinner rolls,pasta]
rule50	0.052	0.602	1.621	pasta	<---	[paper towels,eggs,dinner rolls]
rule24	0.051	0.63	1.621	dinner rolls	<---	[spaghetti sauce,poultry,cereals]
rule57	0.055	0.63	1.616	eggs	<---	[paper towels,ice cream,pasta]
rule11	0.05	0.613	1.616	coffee/tea	<---	[yogurt,cheeses,cereals]
rule44	0.052	0.628	1.614	dinner rolls	<---	[spaghetti sauce,poultry,juice]
rule35	0.052	0.628	1.61	eggs	<---	[dinner rolls,poultry,soda]
rule54	0.054	0.598	1.603	spaghetti sauce	<---	[dinner rolls,poultry,laundry detergent]
rule29	0.051	0.604	1.589	milk	<---	[poultry,laundry detergent,cereals]
rule42	0.052	0.59	1.581	spaghetti sauce	<---	[dinner rolls,poultry,ice cream]
rule45	0.052	0.584	1.566	spaghetti sauce	<---	[dinner rolls,poultry,juice]
rule56	0.055	0.624	1.565	ice cream	<---	[paper towels,eggs,pasta]
rule51	0.052	0.567	1.565	paper towels	<---	[eggs,dinner rolls,pasta]

Interpretation of Association Rule Output

Support:

- Support represents the proportion of transactions in the dataset that contain the items involved in the rule.
- For example, Rule 59 has a support of 0.055, meaning 5.5% of the total transactions include both the antecedent (e.g., "eggs, ice cream, pasta") and the consequent (e.g., "paper towels").

Confidence:

- Confidence measures the likelihood of the consequent item being purchased when the antecedent is present.
- Rule 59 has a confidence of 0.649, indicating that 64.9% of the time, customers who bought "eggs, ice cream, pasta" also bought "paper towels.“

Lift:

- Lift assesses the strength of the association rule relative to the random chance of buying the consequent.
- A lift greater than 1 implies a positive association between items.
- Rule 59 has a lift of 1.791, indicating that customers buying the antecedent items are 1.791 times more likely to buy "paper towels" compared to random purchase behavior.

Insights

- Rule 58 indicates that customers who purchase "paper towels, eggs, and ice cream" are likely to also buy "pasta" (Confidence: 0.643, Lift: 1.731).
- Pasta and Dinner Rolls: Items like pasta, dinner rolls, and poultry frequently appear together in rules (e.g., Rule 49, Rule 44, Rule 7). This suggests that customers often purchase these items for family meals or gatherings.
- Ice Cream and Dairy Products: Rules involving ice cream paired with products like yogurt, milk, and butter (e.g., Rule 56, Rule 29) highlight a preference for dessert and dairy combos.
- Rule 44 shows that "paper towels, eggs, and dinner rolls" often lead to purchasing "dinner rolls" again (Confidence: 0.628, Lift: 1.614).
- Staples for Quick Meals: Items like pasta, spaghetti sauce, and poultry are common ingredients for quick meals (e.g., Rule 40, Rule 43), suggesting these could be packaged as "easy meal" bundles.
- High lift values (e.g., >1.5) in several rules confirm strong associations among these items, suggesting opportunities for bundle promotions.

Recommendations

- Bundle Items with High Lift: Create product bundles combining items frequently purchased together (e.g., "eggs, ice cream, pasta, and paper towels"). Highlight the cost savings to encourage bulk purchases.
- Optimize Cross-Promotions: Use in-store or online recommendations to suggest items with high confidence. For example, suggest "paper towels" when a customer adds "eggs, ice cream, and pasta" to their cart.
- Stock and Shelf Placement: Place strongly associated items near each other in physical stores (e.g., dinner rolls near poultry and spaghetti sauce). This can increase convenience for customers and boost sales.
- Dynamic Discounts for Loyalty Members: Offer targeted discounts on items in high-confidence rules for loyalty program members. For example, a 10% discount on "paper towels" for customers who frequently buy "eggs and pasta."

Example Discounts

1. **Bundle Discount**:
    1. "Buy eggs, ice cream, and pasta, get 20% off paper towels."
2. **Threshold Offer**:
    1. "Spend $20 on dairy and bakery items, get 15% off dinner rolls."
3. **Buy One, Get One (BOGO)**:
    1. "Buy any two poultry items, get 50% off spaghetti sauce."
4. **Flash Sale**:
    1. "Weekend special: 10% off mixes when you purchase juice and sandwich bags."

Possible Combos with Lucrative Offers

1. **Breakfast Essentials Combo**:
   - **Items**: Eggs, Bread, Coffee/Tea, Juice
   - **Offer**: "Buy 3, get the 4th item at 50% off."
   - **Rationale**: These items often form a staple breakfast, encouraging customers to buy them together.
2. **Dinner Delight Combo**:
   - **Items**: Poultry, Spaghetti Sauce, Pasta, Dinner Rolls
   - **Offer**: "Get 20% off dinner rolls with the purchase of the other three items."
   - **Rationale**: These items are common ingredients for a hearty dinner, incentivizing higher spending.
3. **Snack & Dessert Combo**:
   - **Items**: Ice Cream, Mixes, Cheeses, Sandwich Bags
   - **Offer**: "Buy any 2 snacks and get 15% off sandwich bags."
   - **Rationale**: Combines snacks with a useful accessory for carrying or storing food.

Next Steps Based on Insights

1. **Promotional Campaigns**:
   Design and launch cross-promotions for frequently associated items with high lift values (e.g., offer discounts on "paper towels" when customers purchase "eggs and pasta").

1. **Upsell and Cross-Sell**:
   Use customer purchasing data to target customers with personalized discounts on frequently bought-together items, encouraging them to spend more on their next purchase.

1. **Custom Product Bundles**:
   Based on these combinations, create and test various product bundles, offering discounts for bundle purchases (e.g., pasta + spaghetti sauce + dinner rolls).