Optimize Operations:

The Impact of Basket Analysis on Grocery Retail Profitability

CLEAR DEMAND

Table of Contents

Unlock Growth

Retailers want to

- improve customer retention
- reduce churn
- acquire new customers by beating competition
- boost brand awareness
- promote customer advocacy

A basket analysis gives you insights into customer purchasing behavior.

To achieve these goals, retailers turn to a powerful tool:

basket analysis.

In this whitepaper we will focus on the best practices for conducting a comprehensive differential market basket analysis for grocery products.

Market Basket Analysis

By analyzing the items purchased by customers and the frequency with which they're bought together, retailers gain valuable insights into customer preferences and buying patterns.

A market basket analysis helps you optimize product selection, pricing, placement, and promotional strategies to better serve your customers.

This analysis is a key input to becoming a shopper destination that offers the best product selection, with the best price-value, unmatched convenience for a shopper experience that's imperative to retain existing customers and acquire new ones.

Retailers use this data mining technique to increase sales by better understanding customer purchasing patterns.

It's typically conducted using data sources like POS data, online purchase data, panel data, loyalty data, and/or receipt data which records the items purchased during a customer's shopping trip.

Predictive Market Basket Analysis

Starting with your own historical purchase data, you can use machine learning and statistical techniques to identify and unravel hidden correlations like product combinations that occur frequently in consumer purchase transactions.

Predictive models can effectively use these correlations and patterns to forecast things like products a user will buy again, products a user will try for the first time, or products a shopper will add to their cart during their next shopping trip (in store or online).

The goal of a predictive market basket analysis is to simplify the customer's shopping experience and improve cross-sell / upsell opportunities.

This analysis is a key input for retailers that offer personalized experiences for faster checkout and online product recommendations. In physical stores, it drives strategic decisions like product placement that enables a fast, frictionless purchase while maximizing the customer basket size.

Differential Market Basket Analysis

Using purchase data from your internal sources and external providers, you can perform a competitive analysis that helps unravel consumer purchase patterns across the market.

A differential market basket analysis helps retailers understand patterns like why consumers prefer to purchase a product from the same retail platform even when the price is the same across other retailers.

Consumer decisions are based on several choices:
» Available assortment
» Pricing
» Offered promotions
» Convenience (e.g., delivery time and pickup options)
» Experience
» Purchase history between stores, seasons, and time periods

The goal of a differential market basket analysis is to understand how changes in external factors impact purchase behavior, allowing retailers to make informed decisions about product assortment and pricing.

Step-by-Step

Implement your comprehensive differential market basket analysis with these 4 steps

Customer Transaction Data

Reliable Product Matching

ML and Statistical Models

Recommendations and Analysis

Step 1:
Get high-quality customer transaction level data.
Data quality is determined by the accuracy of captured information (correct price & purchase time), the depth of the capture (item & promotion details), and the coverage (stores, zip codes, and online shopping destinations).

Step 2:
Enable accurate product matching.
Accurately identifying items as the same product, related (product variation), or similar (within the same category) across retailer data sources is imperative for a comprehensive comparative analysis.

Step 3:
Use machine learning and statistical methods to create item-level relationships.
Then produce a single reconciled item-level dataset across disparate retail sources which you can analyze across dimensions like price bands, national versus private brands, location, or customer segments. It's recommended that you extend the product matching process to include item data from online competitor catalog and in-store audits.

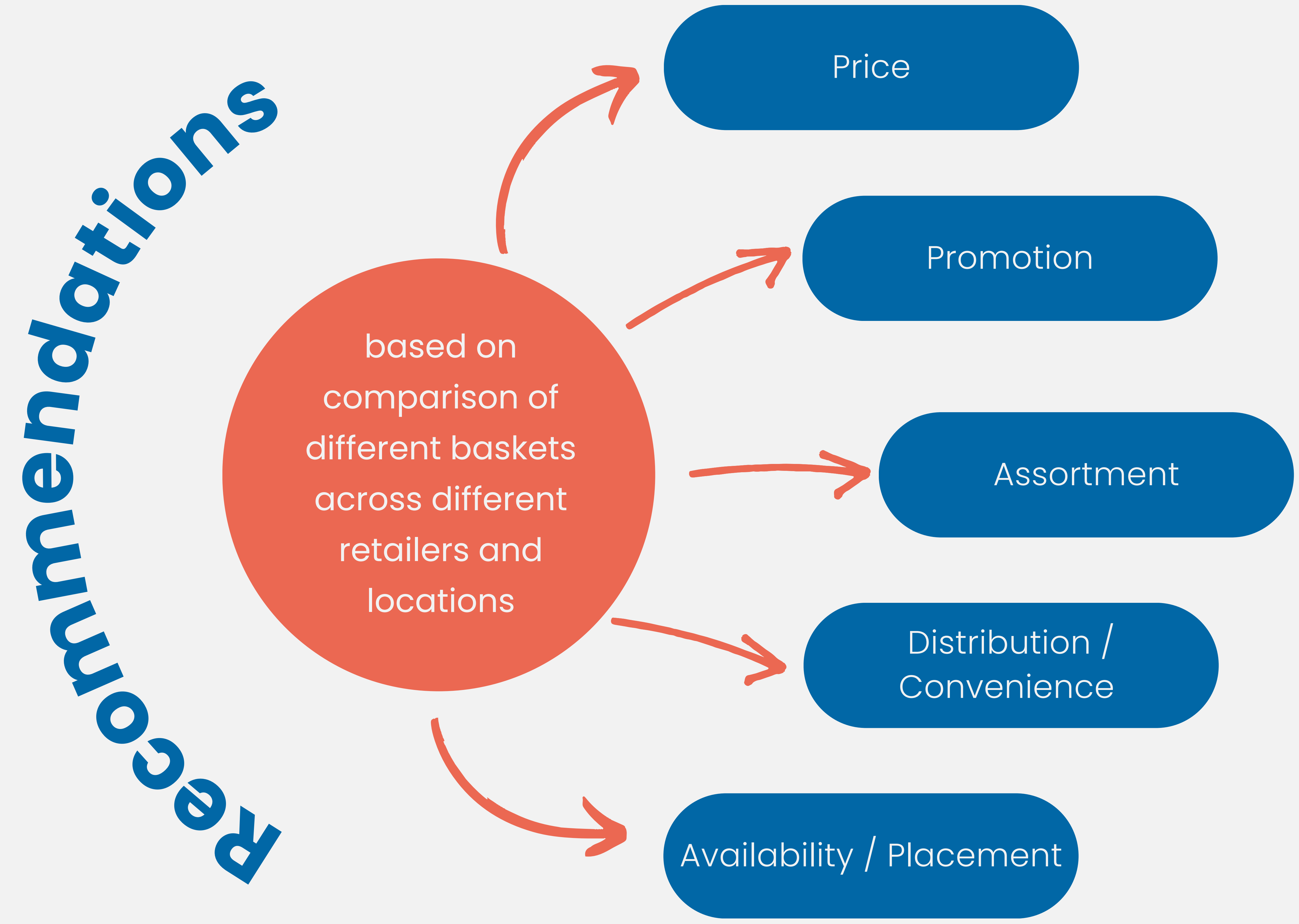
Additional information from these sources (online product availability, search rankings, reviews, ratings, sponsored ads, offline planogram, and product placement) provide valuable inputs for performing causal analytics on top of the transaction-level data.

Step 4:
Produce actionable insights from the reconciled item-level dataset.
Programmatically construct a market basket using popular items across categories at different retailers at a location level and compare them for end-to-end shopping experience in terms of available items, price and promotions offered, and additional convenience fees (like shipping or pick up charges if any).

Create many different baskets picking up alternate items from within a category (for example private label items versus national brands or normal versus USDA certified produce) for analysis and comparison.

Alternatively, past customer transaction-level data can be used for deducing the most popular mix of market baskets across locations and then use these as baskets as templates for ongoing benchmarking and comparison across competitors and locations.

Basket-level data can be further blended with additional external data sources like cross shopping reference data and census data as needed. At a regional level, this will help narrow down to the basket analysis and recommendations based on demographic information.


Recommendations
based on comparison of different baskets across different retailers and locations
Price
Promotion
Assortment
Distribution / Convenience
Availability / Placement

Every round of market basket analysis will help you identify the **key value items** (KVIs) that you should focus on. Your goal is to provide the best offers compared to your competitors **across channels** and **across locations**. KVIs will shift and adjust based on seasonality affects and market realities (like supply chain issues).

MATT SHACHTER
HEAD OF REVENUE GROWTH
CLEAR DEMAND

Operational Considerations

Conduct your basket analysis regularly and make sure you consider these 4 inputs

Frequency	Scope	Data Partner	Implementation

Frequency:
Depending upon resource and budget availability, your market basket analysis should be conducted at least once a quarter and ideally once a month. In parallel, invest in building a competitive intelligence program that can help lead the market on KVIs on the desired frequency (intraday, daily, weekly).

Scope:
Do a region-specific market basket analysis but know the competitors that matter to you. Start with a zip code and then aggregate upwards. Cover all digital channels and relevant physical locations in your analysis.

Data Partner:
Competitor data (transaction information and catalog information) is imperative for a comprehensive and accurate analysis. You'll most likely work with different data providers - each bringing a different dataset. For example, loyalty data may come from providers like GFK; consumer panel data may come from IRI or Nielsen IQ Slice; and receipt and web based may come from Bungee Tech. You should measure the data from every provider against depth, coverage, and accuracy metrics.

Implementation Partner:
Whether you are conducting the market basket analysis inhouse or working with a partner, work with a team that has demonstrated experience in blending and reconciling large datasets from different sources and building large scalable data and analytics services on top of it.

Daily Market Basket Analysis

At Clear Demand, we use the power of machine learning to produce the most differentiated market basket analytics for your retail business.

We bring the following strengths to any partnership:

- data
- product matching
- basket analysis
- daily & continuous measurement

DATA

To begin, we use ML to accurately capture more than a billion data points for millions of products, intraday, at a zip code level from hundreds of retailer destinations (websites, mobile apps and APIs).

We also use ML to extract item level data from consumer receipts provided by our customers or partners.

We invest heavily in ensuring the highest data quality using machine learning and manually audit of all data we collect.

The net result is highest level of coverage, depth, and accuracy of competitive intelligence data needed for market basket analysis.

PRODUCT MATCHING

We use proprietary machine learning models to perform the most comprehensive product matching across different data sources to find product relationships.

Our ML models are battle tested against poor data quality of catalogs and sparse receipt information.

Whether its finding exact products or similar private label equivalents, our product matching data allows accurate construction and changes to market baskets that are used for analysis across locations and competitors.

BASKET ANALYSIS

We partner with your business and analytics team to discover data patterns that are most valuable to your business.

We provide end-to-end data engineering support - from setting up automated pipelines for ingesting data from internal and external data sources - to hands-on data science involvement where our ML scientists and engineers conduct and optimize experiments to generate recommendations needed by your business teams.

Our goal is to setup a repeatable, cost-effective process to perform market basket analysis on a regular basis.

CONTINUOUS IMPROVEMENT

We believe in daily and continuous tracking of KVIs and baskets against the recommendations generated from a market basket analysis.

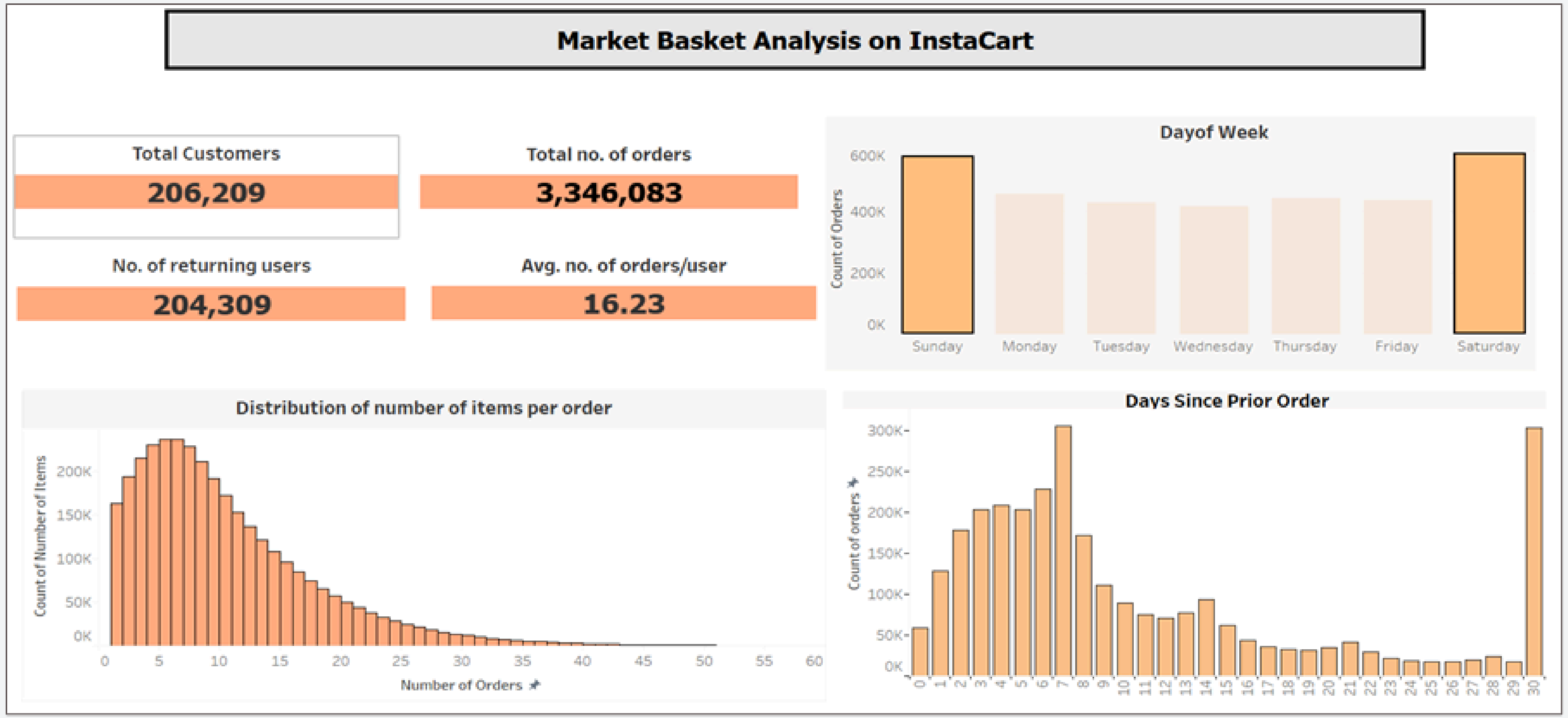
We can do this across key competitors and locations online and by continuously ingesting item level receipt data as it is made available.

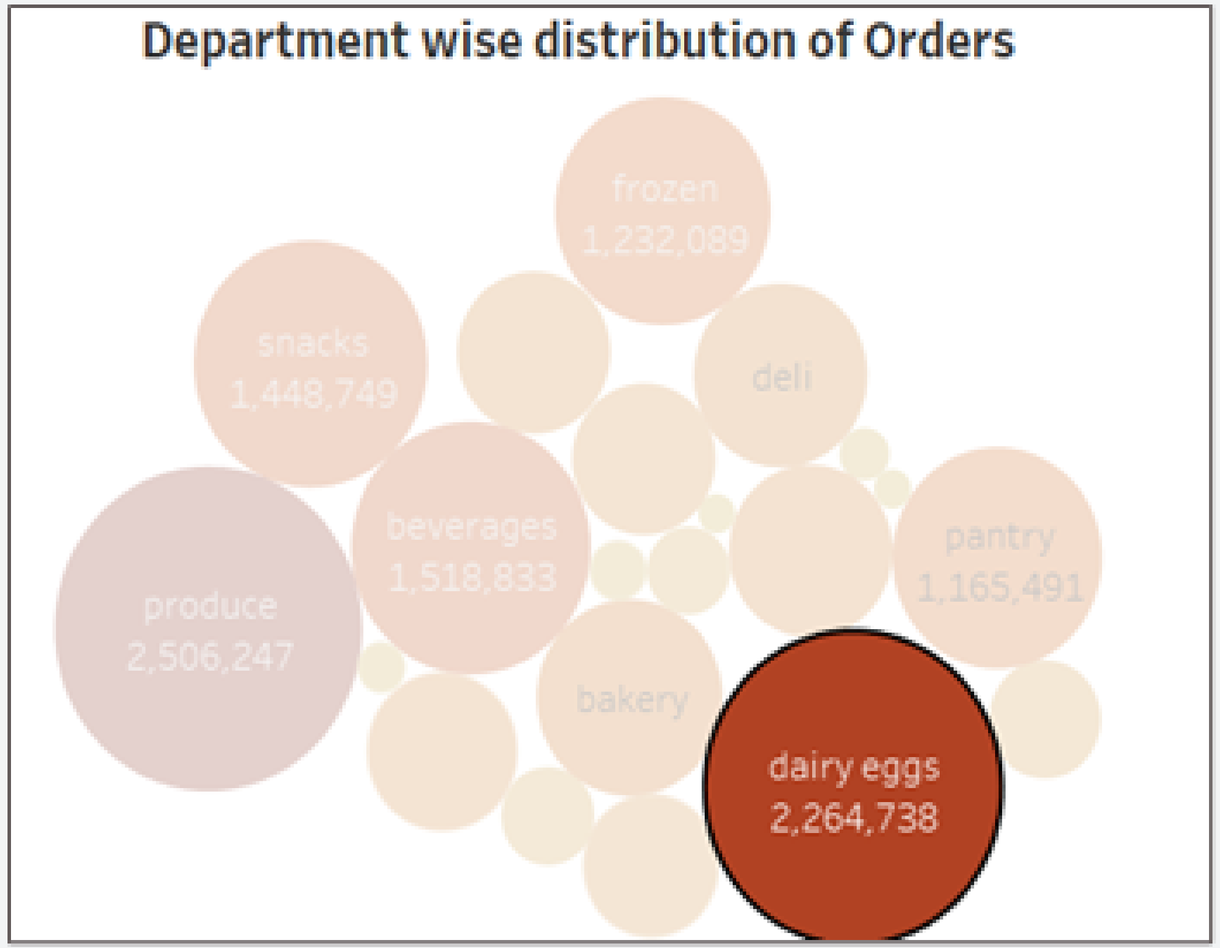
APPENDIX A

Example of Predictive
Market Basket Analysis

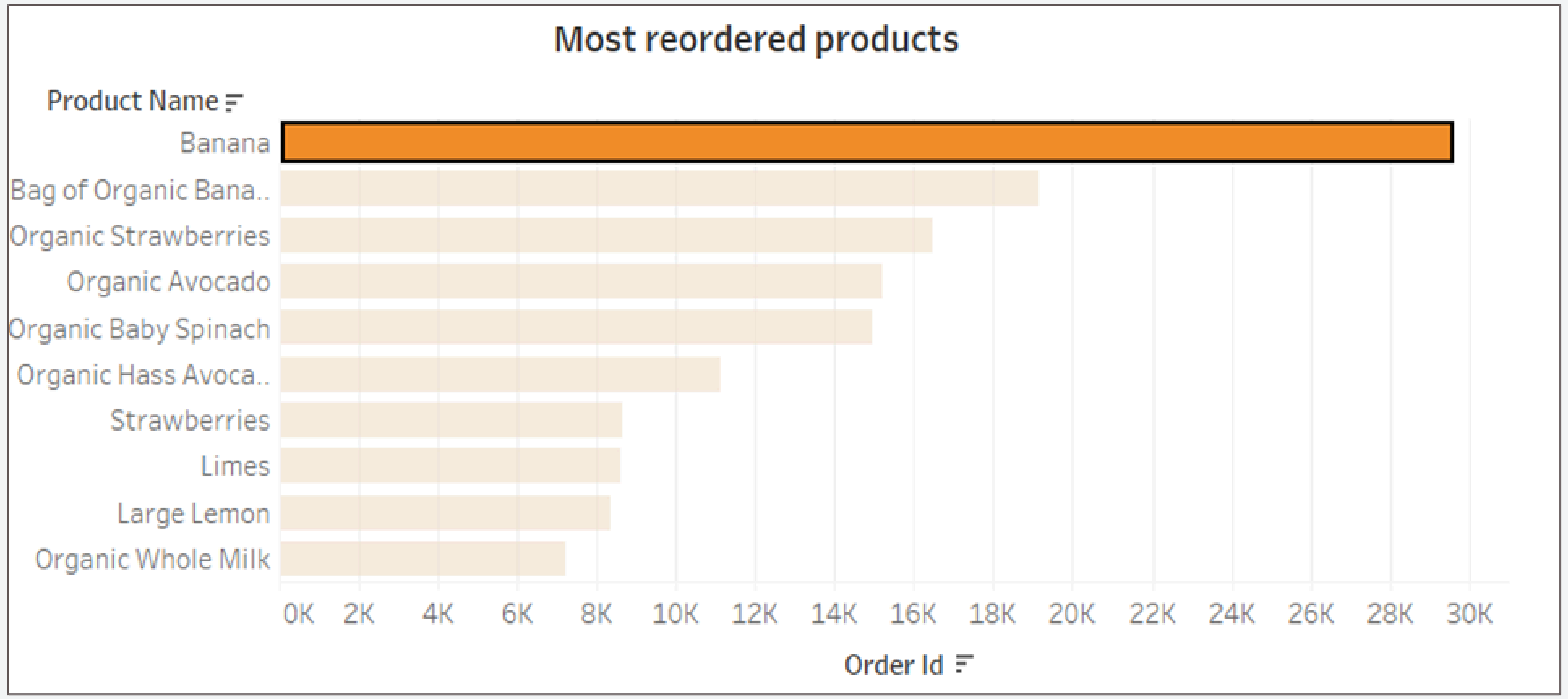
Instacart published a dataset of over 3 million grocery orders from more than 200K Instacart users on Kaggle. This dataset has information on order details, the flow of products in order, and dates of when the order was placed. It includes information on departments and aisles. We used this dataset to produce an example for **predictive market basket analysis** that can be produced to improve your consumer buying experience.

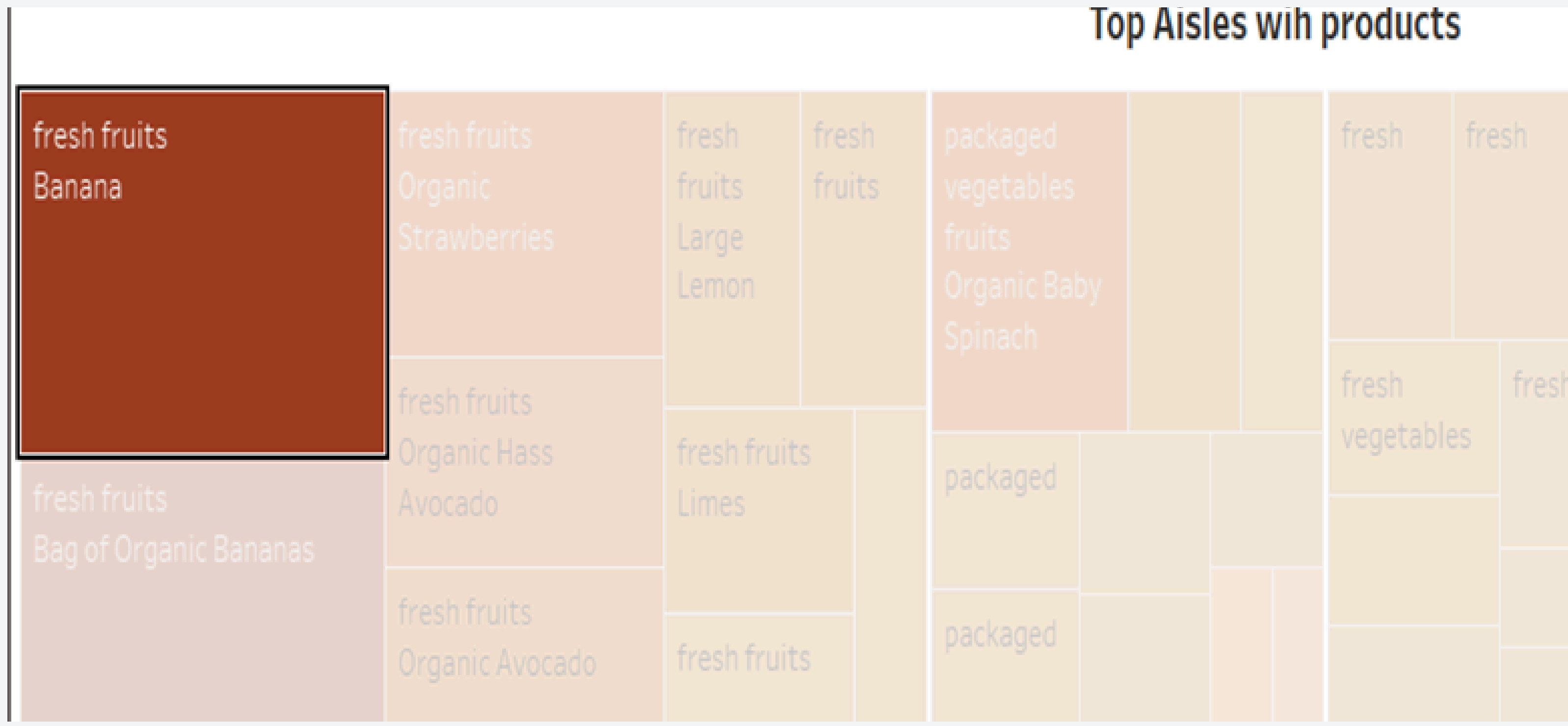
A few key findings:

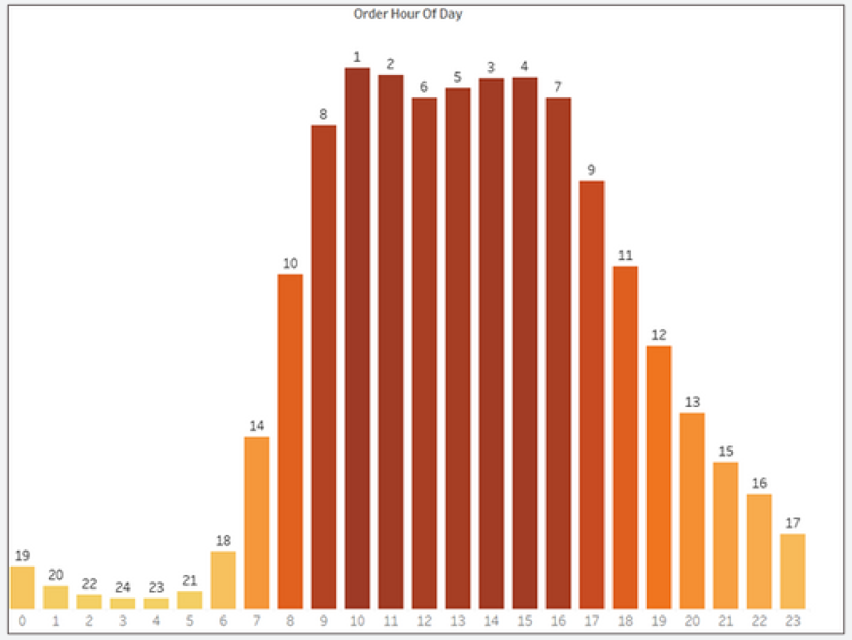


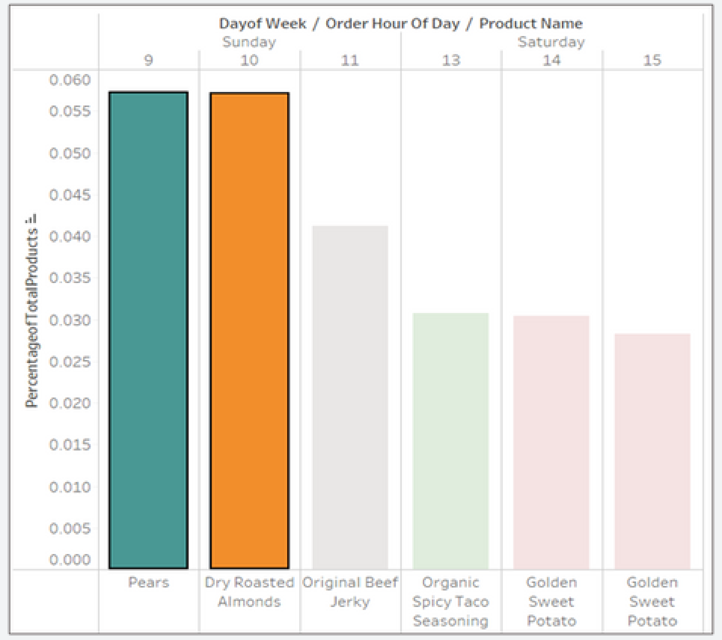


- Saturday & Sunday are the days when customers place most orders
- Customers tend to place orders toward the end of the month
- Most customers have 5 items in their order
- Dairy & eggs are a part of most orders









- Bananas were the most reordered product
- Fresh fruits are the top aisle & bananas are the top selling product
- Most orders are placed between 8am - 6pm
- Pears and dry/roasted almonds are sold the most during the weekend’s busiest hours

APPENDIX B

Example of Differential Market Basket Analytics Using Customer Transaction Level Data

To demonstrate the high-level process of performing a **differential market basket analysis** we will assume that we have we have obtained the following data from customer receipts and for simplification we have listed the same products across the retailers.

The data below represents 12 transactions across 6 retailers. Each transaction lists the items that were purchased along with the retailer where the purchase was made.

Transaction ID	Retailer	Butter	Milk	Bread	Eggs	Apples	Cabbage	Potatoes	Canned Peaches	Ground Beef
T1	Safeway	1	1	1	1	0	0	1	0	1
T2	Safeway	1	1	1	1	1	0	0	0	0
T3	Kroger	1	1	1	1	1	0	1	1	1
T4	Kroger	1	1	0	1	0	1	1	1	0
T5	Giantfoods	1	1	1	1	0	0	1	1	1
T6	Giantfoods	1	1	0	1	1	0	1	0	1
T7	Stop and Shop	0	1	1	1	1	1	1	0	0
T8	Stop and Shop	1	1	1	1	0	0	0	0	1
T9	HEB	1	1	1	1	1	0	1	0	0
T10	HEB	1	1	0	1	0	1	1	1	1
T11	Fred Meyer	1	1	1	1	1	0	1	0	0
T12	Fred Meyer	1	1	1	1	0	1	0	0	1

A dataset like this is prepared for analysis by converting it into a binary format, where a 1 indicates that the item was purchased and a 0 indicates that it was not. We can also remove any irrelevant columns, such as the transaction ID and retailer columns, which are not required for the analysis. The prepared data is shown below:

Butter	Milk	Bread	Eggs	Apples	Cabbage	Potatoes	Canned Peaches	Ground Beef
1	1	1	1	0	0	1	0	1
1	1	1	1	1	0	0	0	0
1	1	1	1	1	0	1	1	1
1	1	0	1	0	1	1	1	0
1	1	1	1	0	0	1	1	1
1	1	0	1	1	0	1	0	1
0	1	1	1	1	1	1	0	0
1	1	1	1	0	0	0	0	1
1	1	1	1	1	0	1	0	0
1	1	0	1	0	1	1	1	1
1	1	1	1	1	0	1	0	0
1	1	1	1	0	1	0	0	1

As a next step, the machine learning models or statistical technique uses association rules like Apriori, AIS, SETM or FP (frequent pattern) algorithms.

For this example, we use the Apriori function from the mlxtend library in Python to apply the Apriori algorithm to the prepared data. We set the minimum support level to 0.25, which means that an item set must appear in at least 25% of the transactions to be considered significant. The minimum support level can be adjusted based on the size of the dataset and the desired level of significance. The code for applying the Apriori algorithm is shown below:

from mlxtend.frequent_patterns import apriori frequent_itemsets = apriori(data, min_support=0.25, use_colnames=True)

The output of the Apriori function is a DataFrame that lists the frequent item sets along with their support levels. The support level indicates the percentage of transactions that contain the itemset. The resulting DataFrame is shown below:

Support	Itemsets
0.25	('Butter', 'Bread')
0.25	('Butter', 'Eggs')
0.25	('Butter', 'Milk')
0.25	('Bread', 'Eggs')
0.25	('Bread', 'Milk')

Support	Itemsets
0.25	('Cabbage', 'Potatoes')
0.25	('Ground Beef', 'Milk')
0.25	('Milk', 'Eggs')
0.25	('Potatoes', 'Eggs')
0.25	('Potatoes', 'Milk')

We can now generate association rules to understand the relationships between the item sets above.

Association rules are generated by setting a minimum confidence level, which is the percentage of transactions containing the antecedent (the items on the left-hand side of the rule) that also contain the consequent (the items on the right-hand side of the rule).

For example, if the confidence level is set to 0.5, a rule would only be generated if at least 50% of the transactions containing the antecedent also contain the consequent.

We can use the association_rules function from the mlxtend library to generate the association rules. The code for generating the association rules is shown below:

from mlxtend.frequent_patterns import association_rulesrules = association_rules(frequent_itemsets, metric="confidence", min_threshold=0.5

The output of the association_rules function is a DataFrame that lists the association rules along with their support, confidence, and lift values.

The support value for a rule is the percentage of transactions that contain both the antecedent and consequent, while the confidence value is the percentage of transactions containing the antecedent that also contain the consequent.

The lift value indicates the degree of dependence between the antecedent and consequent.

A lift value greater than 1 indicates a positive association, while a lift value less than 1 indicates a negative association. The resulting DataFrame of association rules is shown here:

Antecedents	Consequents	Support	Confidence	Lift
('Bread',)	('Butter',)	0.25	0.5	1.333
('Butter',)	('Bread',)	0.25	0.667	1.333
('Eggs',)	('Butter',)	0.25	0.5	1.333
('Butter',)	('Eggs',)	0.25	0.667	1.333
('Milk',)	('Butter',)	0.25	0.5	1.333
('Butter',)	('Milk',)	0.25	0.667	1.333
('Eggs',)	('Milk',)	0.25	0.5	1.25
('Milk',)	('Eggs',)	0.25	0.5	1.33
('Potatoes',)	('Cabbage',)	0.25	0.5	2.0
('Cabbage',)	('Potatoes',)	0.25	1.0	2.0

APPENDIX C

Example of Differential
Market Basket Analytics
Using Web Data

We'll demonstrate how a **simple basket analysis** can be performed across retailers by comparing the prices of established key value items across several retailers. Let's assume the KVIs are Butter, Milk, Bread, Eggs, Cabbage, Potatoes, and Canned Peaches. Let's construct and compare a basket across Safeway, Kroger, Giant Foods, Stop and Shop, HEB, and Target stores.

We will further assume that price, promotion, and availability data for these products from the retailers for a given location are all available **via web-based data collection** from retailer websites or mobile apps and as shown in the next slide.

When analyzed, it's easy to deduce from the data that…

- HEB offers the lowest prices across all items to its consumers
- HEB has all the key value items in stock

- Stop and Shop is the only other retailer with all items in stock but is priced higher across all items compared to HEB and other retailers. Just gaining price parity with the price leader in the market may significantly bump traffic and conversion given all other retailers above have some item out of stock.

- Giant Foods has opportunity to normalize the prices across items to improve price-value perception.

- All other retailers have opportunity to improve pricing and fix availability issues which will play a major role in customer loss to competitors.

	Safeway		Kroger		Giant Foods		Stop And Shop		HEB		Target	
	Price	In Stock	Price	In Stock	Price	In Stock	Price	In Stock	Price	In Stock	Price	In Stock
Butter	$3.99	Y	$3.89	Y	$4.29	Y	$4.19	Y	$3.79	Y	$3.99	Y
Milk	$2.49	Y	$2.39	Y	$2.69	Y	$2.59	Y	$2.29	Y	$2.49	N
Bread	$2.99	N	$2.89	Y	$3.29	Y	$3.19	Y	$2.69	Y	$2.99	Y
Eggs	$1.99	Y	$1.89	Y	$2.29	Y	$2.19	Y	$1.79	Y	$1.99	Y
Apples	$0.99/lb.	Y	$0.89/lb.	N	$1.19/lb.	Y	$1.09/lb.	Y	$0.79/lb.	Y	$0.99/lb.	Y
Cabbage	$0.79/lb.	Y	$0.69/lb.	Y	$0.99/lb.	Y	$0.89/lb.	Y	$0.59/lb.	Y	$0.79/lb.	Y
Potatoes	$0.79/lb.	Y	$0.69/lb.	Y	$0.99/lb.	N	$0.89/lb.	Y	$0.59/lb.	Y	$0.79/lb.	Y
Canned Peaches	$1.69	Y	$1.59	Y	$1.99	Y	$1.89	Y	$1.49	Y	$1.69	N
Ground Beef	$3.99/lb.	Y	$3.89/lb.	Y	$4.29/lb.	Y	$4.19/lb.	Y	$3.79/lb	Y	$3.99/lb.	Y

About

CLEAR DEMAND

Who We Are

We’re revolutionizing retail operations through efficient automation, improving our customers ability to compete and win in the market.

Mission

Clear Demand empowers profitable retail growth through data, machine learning, and automation. Take control of your product, category, and organizational profitability with our end-to-end retail pricing, promotion, and assortment solutions.