MSBX 5420

Basket Analysis

Rafael Cintron,
Pranathi Manthri,
Murali Prateek Manthri

Business Problem

Customers often purchase items that are ingredients for the same meal or are commonly paired together, yet these items may be located far apart in the store

Project Idea

Analyze transaction data to identify supplementary, in goal of placing them closer together, enhancing convenience and increasing the likelihood of higher basket sizes

Goal

Create a model that can detect associations between products, categories, and aisles

The Data

- Sourced from Kaggle
- Original Size: 713mb
- Instacart Dataset
  - Same-Day Grocery Delivery/Pickup Platform
  - User orders on app/web and order is fulfilled by an Instacart Shopper

Metadata

aisles.csv

- Aisle Name
- Aisle ID
- Product Organization

products.csv

- Product name
- Product ID
- Aisle
- Departments

orders.csv

- Order ID
- Order number
- Weekday of the Order
- Hour of the Order
- User ID
- Days since the Prior Order

departments.csv

- Department Name
- Department ID

order_products__prior.csv

- Order
- Products
- Reordered Items

order_products__train.csv

- Order
- Products
- Reordered Items

Data Preparation

Merge Sheets

- **Union**
  - order_products__prior
  - order_products__train
- **Inner Join on**
  - aisles
  - departments
  - products
  - orders
- **Split (back to test and train)**
  - order_products__prior
  - order_products__train

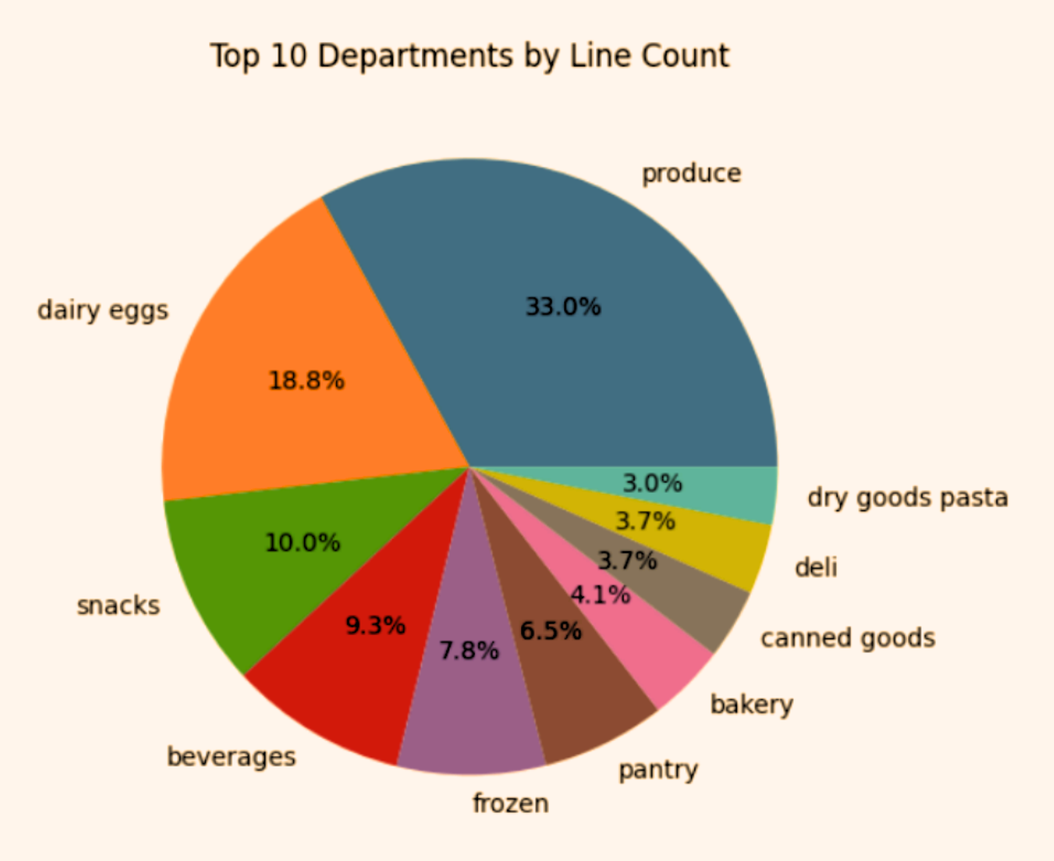
Prep/ Clean

- Remove Null Values

New Sub-Datasets

- Top Products
- Top Departments
- Reorder Rate by Department
- Orders by Day of the Week
- Orders by Hour
- Top Aisle-Department Combinations
- Average Basket Size

Exploratory Data Analysis



30% Produce

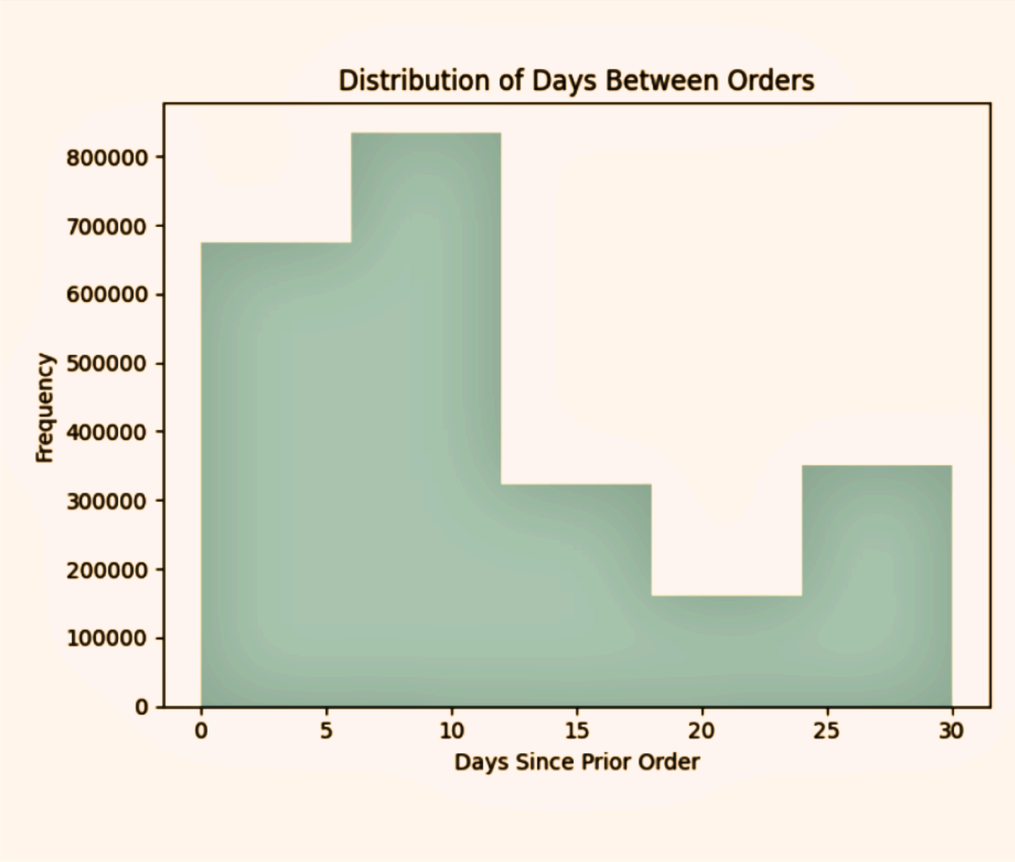
19% Dairy & Eggs

10% Snacks

< 10% Everything Else

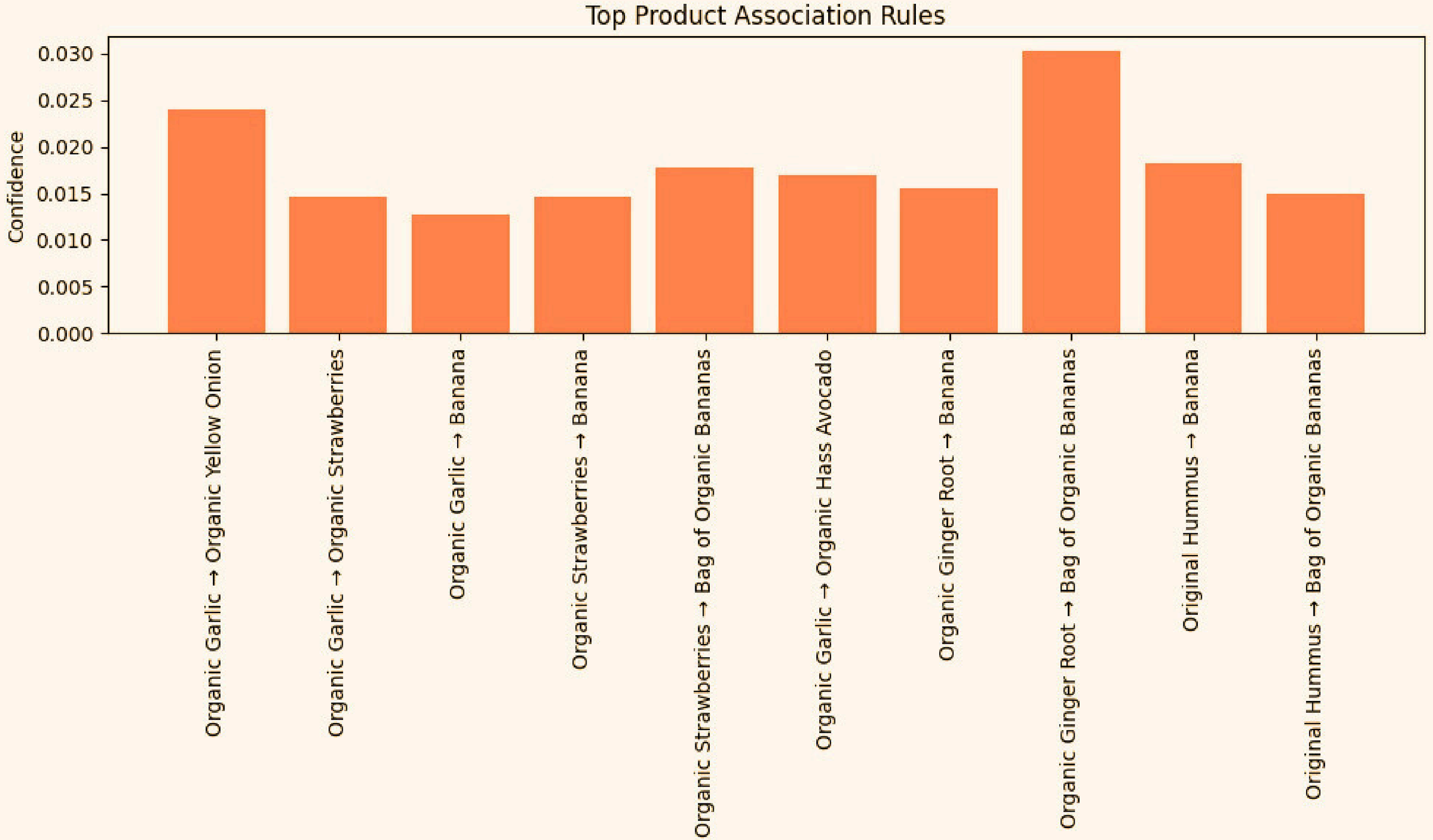
Exploratory Data Analysis

- Largest Reorder Bucket: **7–10 days**
- Reorder Bucket of **25–30 days** is also higher than expected
- Business Insights
  - You can target two distinct campaigns:
    - Weekly Push for Staple Products
    - Monthly Reminder for Bulk Products



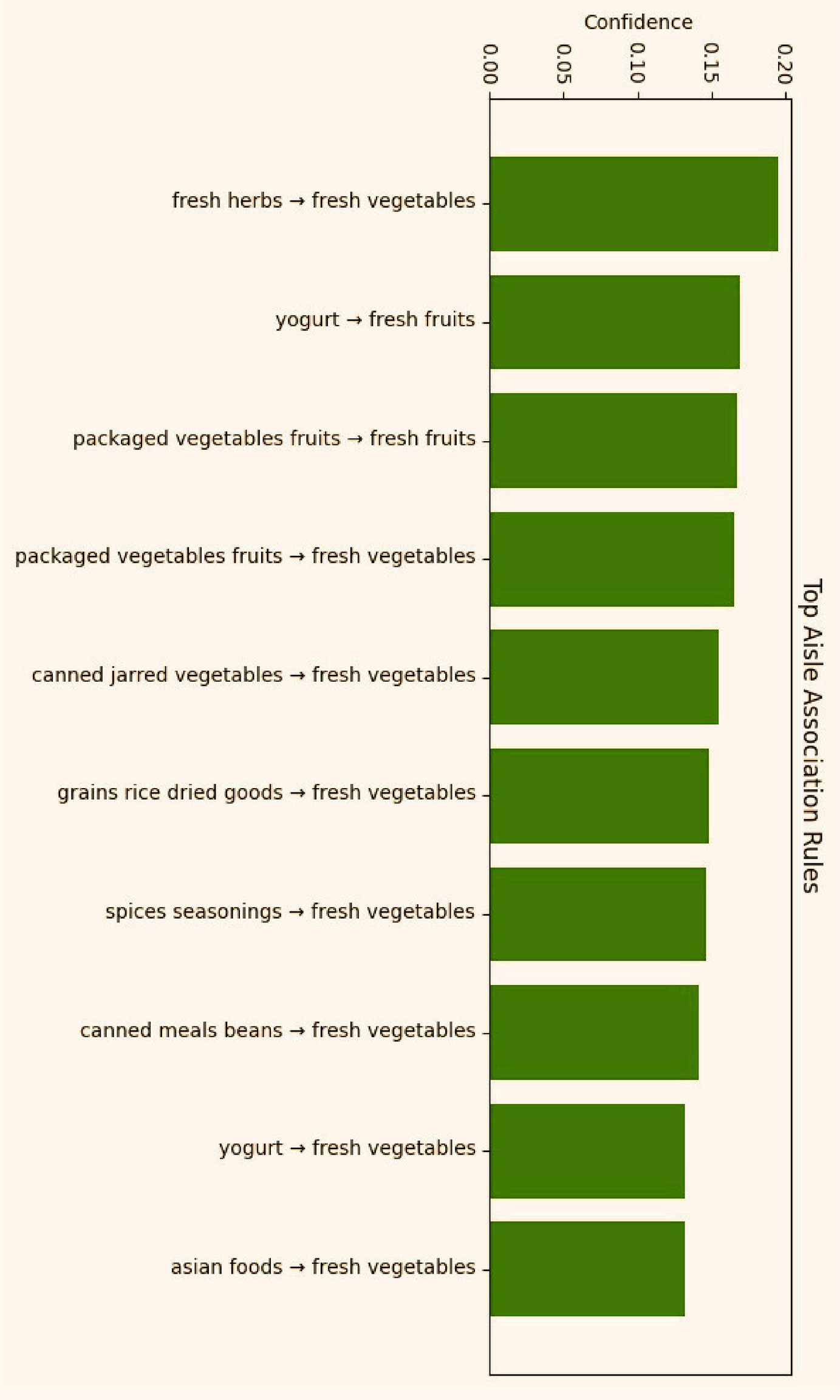
Product-Based Association Rule

- Goal: Find which **Individual Products** are co-purchased within Customer Orders frequently
- Algorithm: FP-Growth
- Process
  1. Group Orders by **Products**
  2. Find association likelihood between **Products**
  3. Evaluation: Support, Confidence, and Lift


Top Product Association Rules
Confidence
0.030
0.025
0.020
0.015
0.010
0.005
0.000
Organic Garlic → Organic Yellow Onion
Organic Garlic → Organic Strawberries
Organic Garlic → Banana
Organic Strawberries → Banana
Organic Strawberries → Bag of Organic Bananas
Organic Garlic → Organic Hass Avocado
Organic Ginger Root → Banana
Organic Ginger Root → Bag of Organic Bananas
Original Hummus → Banana
Original Hummus → Bag of Organic Bananas

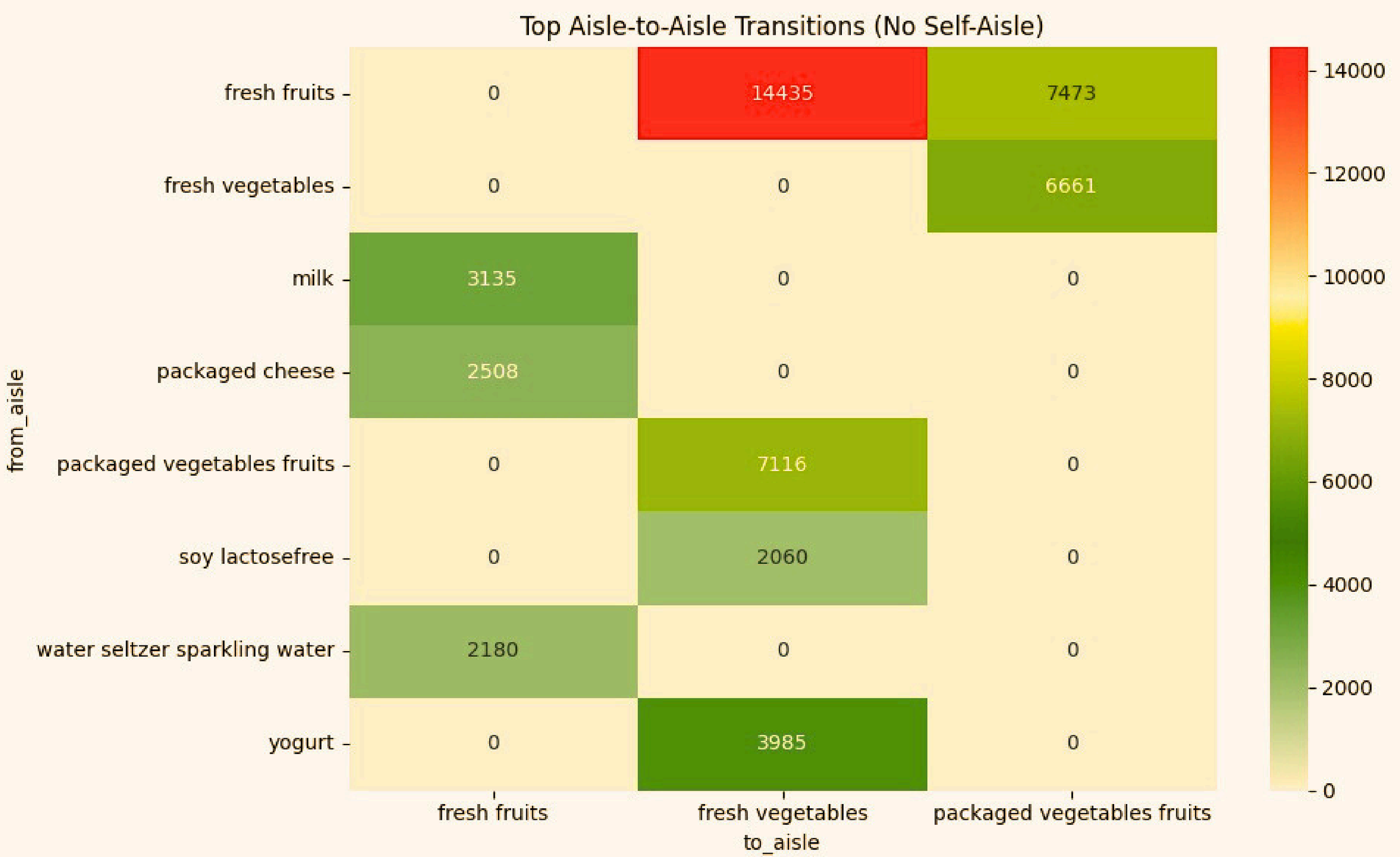
Aisle-Based Association Rule

- Goal: Find what **Aisles** have a high frequency of co-purchased Products
- Algorithm: FP-Growth
- Process
  1. Group Orders by **Aisles**
  2. Find association likelihood between **Aisles**
  3. Evaluation: Support, Confidence, and Lift


Confidence
0.00
0.05
0.10
0.15
0.20
fresh herbs → fresh vegetables
yogurt → fresh fruits
packaged vegetables fruits → fresh fruits
packaged vegetables fruits → fresh vegetables
canned jarred vegetables → fresh vegetables
grains rice dried goods → fresh vegetables
spices seasonings → fresh vegetables
canned meals beans → fresh vegetables
yogurt → fresh vegetables
asian foods → fresh vegetables
Top Aisle Association Rules

Cart Path Transition

- Goal: Find **Sequences of Aisle Purchases** based on how Products are added to the cart
- Process
  1. Create add_to_cart_order field
  2. Find **Order Transitions between Aisles** for each **Customer**
  3. Aggregate Transitions across the dataset
  4. Evaluation: **Frequency**


Top Aisle-to-Aisle Transitions (No Self-Aisle)
from_aisle
fresh fruits
fresh vegetables
milk
packaged cheese
packaged vegetables fruits
soy lactosefree
water seltzer sparkling water
yogurt
0
14435
7473
0
0
6661
3135
0
0
2508
0
0
0
7116
0
0
2060
0
2180
0
0
0
3985
0
fresh fruits
fresh vegetables
packaged vegetables fruits
to_aisle
14000
12000
10000
8000
6000
4000
2000
0

Model Comparisons

Product-Based Association Rule

- Co-purchase relationships at the item level
- Product recommendations and bundle creation
- Identifies what items are frequently bought together
- Enhances personalization and basket size

Aisle-Based Association Rule

- Associations between aisles
- Store layout design and cross-merchandising
- Identifies what types of products are commonly paired
- Improves department adjacencies and impulse sales

Cart Path Transition

- Sequence of item additions and customer movement
- Layout simulation and path optimization
- Analyzes how customers shop
- Enhances navigation efficiency and shopping experience

Recommendations

Prioritize Fruits and Vegetables

Organic Fruits, Vegetables, and Onion Family Members should be in the same section as Inorganic

- High confidence with several Product Associations

Freezers for Frozen Fruit should be next to Fresh Fruit

- The Fresh and Frozen, Fruit and Vegetables are sold at the highest rates
- High confidence with several Category Associations
- High Frequency of Aisle Transitions

Thank you

Questions?