



Instacart Market Basket Analysis Report

💡 Project Overview

This project analyzes the *Instacart Online Grocery Basket dataset* to identify *buying patterns and frequently purchased product combinations*.

Using the *Apriori Algorithm*, we explored customer purchase behavior, discovered strong associations, and visualized them through an *interactive Power BI dashboard*.

🎯 Objectives

- Find items that are most frequently bought together
- Identify strong product associations using Support, Confidence, and Lift
- Create clear Power BI dashboards for interactive insights
- Provide business recommendations for cross-selling and promotions

📊 Dataset Overview

Source: [Kaggle – Instacart Online Grocery Basket Analysis](#)

Files Used:

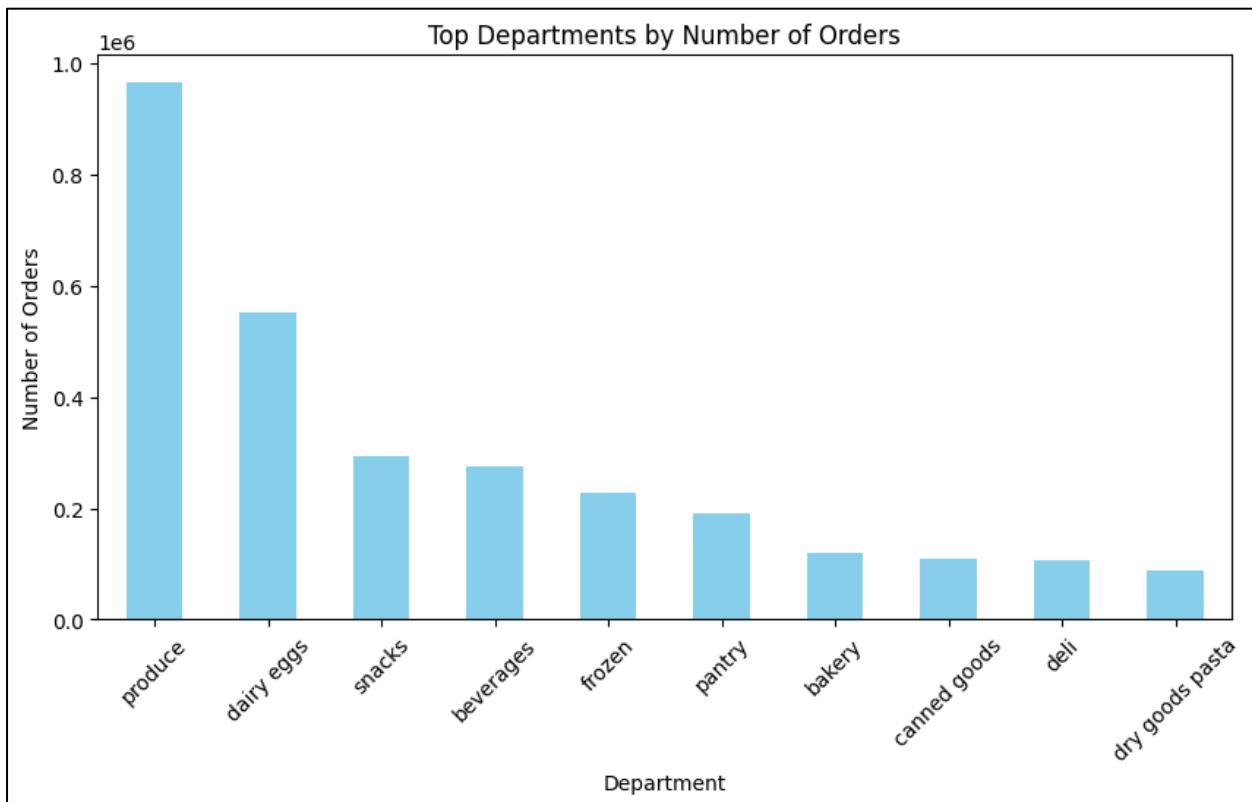
- orders.csv – Contains customer order information
- products.csv – Product details
- order_products_prior.csv – Product list for previous orders
- aisles.csv and departments.csv – Product categorization

📝 Data Cleaning & Preparation

- Merged multiple CSV files to create a single unified dataset.
- Removed missing and duplicate values.
- Standardized column names for easy analysis.
- Prepared a basket dataset with unique orders and their purchased items.



Exploratory Data Analysis (EDA)

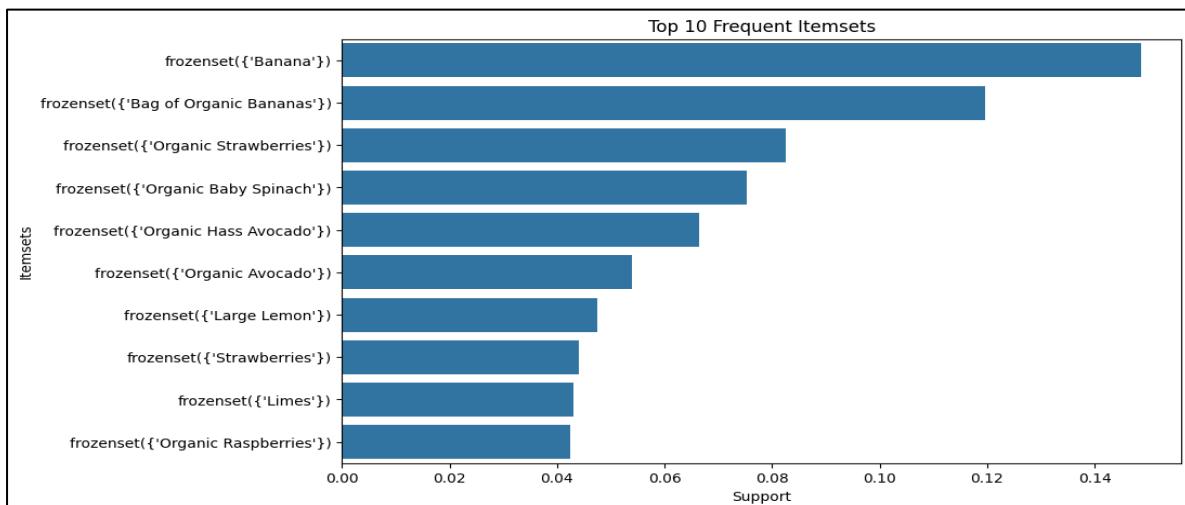


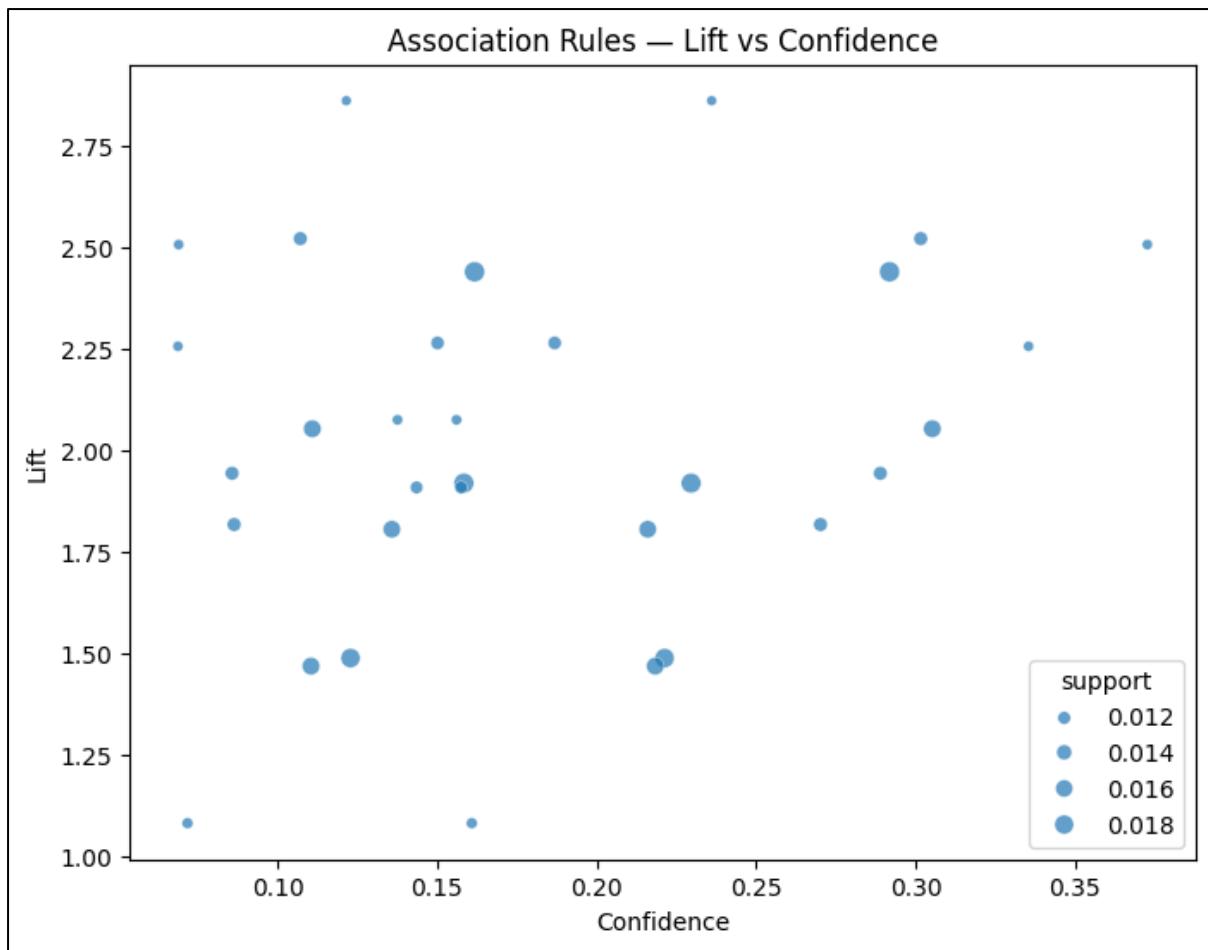
Insights:

- The Produce department leads with nearly 1 million orders, followed by Dairy & Eggs, Snacks, and Beverages.
- Essential and perishable goods dominate shopping baskets.
- Indicates that users rely heavily on Instacart for fresh and daily groceries.



Market Basket Analysis (Apriori Algorithm)





Metrics:

- Support → How often items appear together
- Confidence → Probability that item B is purchased when A is bought
- Lift (>1) → Strength of relationship

Findings:

- Top frequent products: Banana, Bag of Organic Bananas, Organic Strawberries, Organic Baby Spinach
- High lift values between Organic Raspberries → Organic Strawberries and Bag of Organic Bananas → Organic Raspberries
- Scatter plot (Lift vs Confidence) shows strong product associations among organic fruit categories.

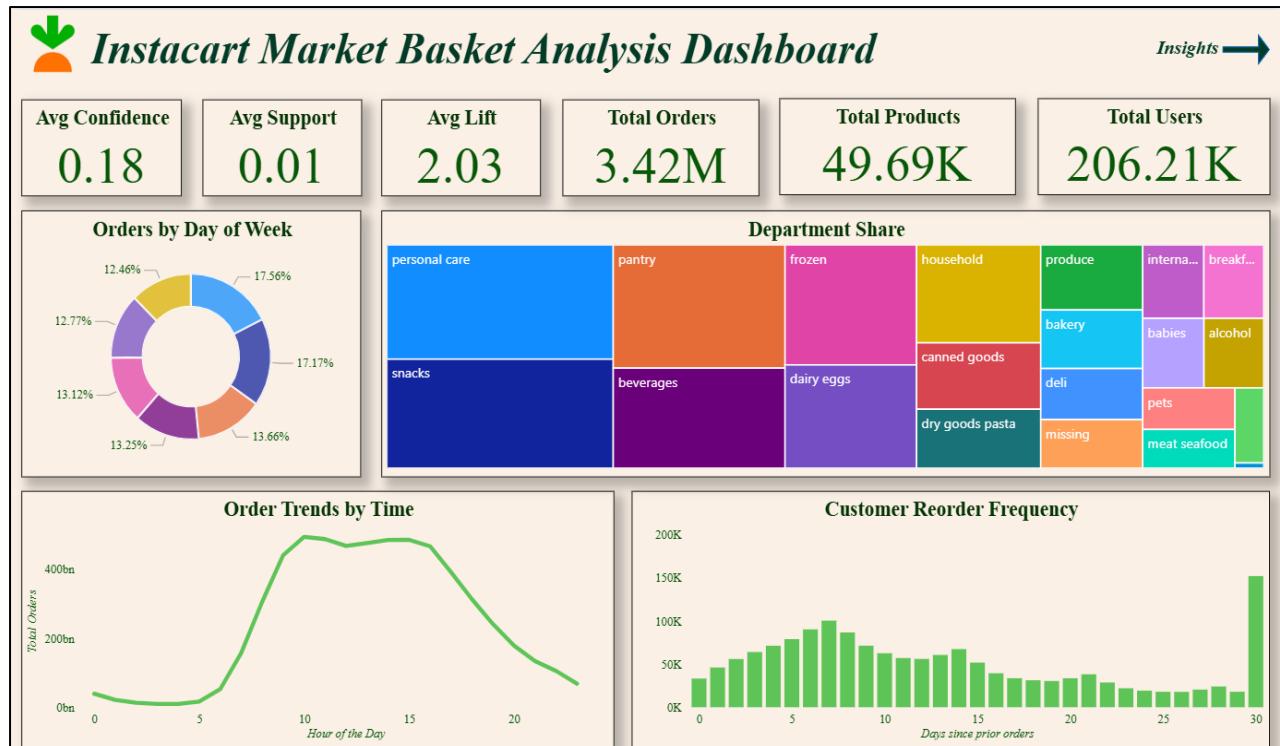
Conclusion:

Organic fruits and vegetables show high association and strong buying dependency.

Dashboard & Visualization

Created an interactive *Power BI Dashboard* with two pages:

1 Overview Dashboard

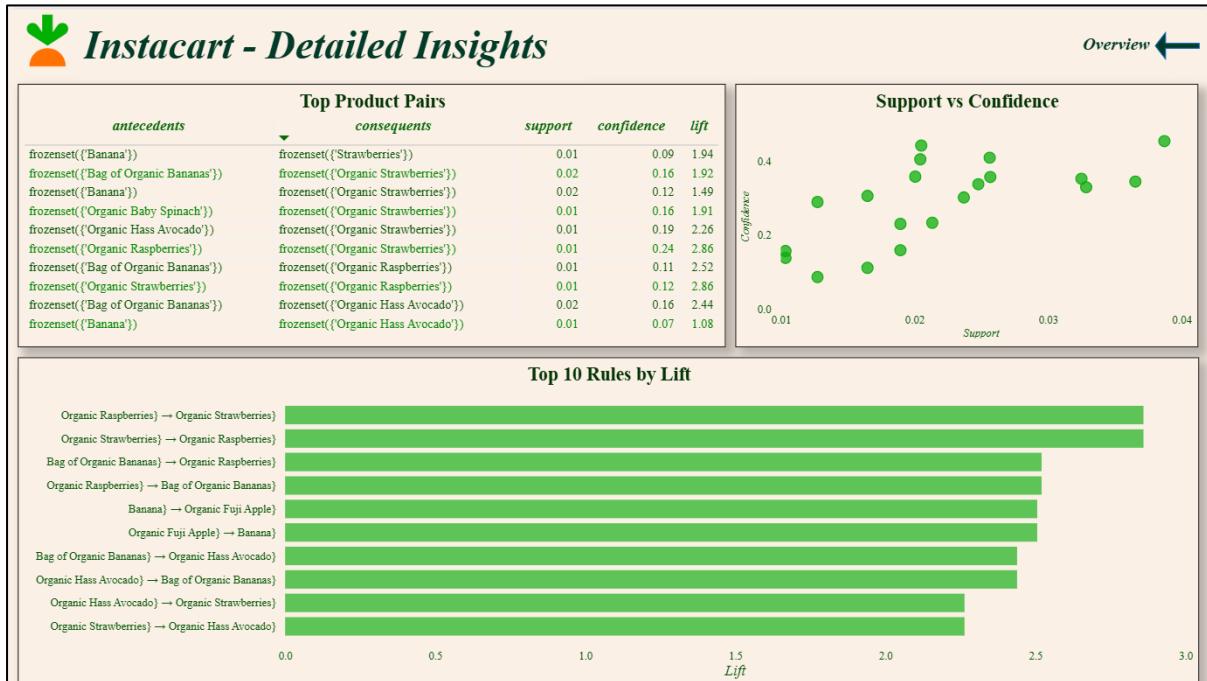


Highlights:

- Avg Confidence: 0.18 | Avg Support: 0.01 | Avg Lift: 2.03
- Total Orders: 3.42M | Total Products: 49.69K | Total Users: 206.21K
- Peak shopping hours: Between 9 AM and 12 PM
- Departments like Pantry, Frozen, and Personal Care dominate share
- Reorders mostly happen every 7–10 days

Insight: Customers prefer shopping during morning hours and frequently reorder essentials weekly.

2 Insights Dashboard



Highlights:

- Top Product Pairs: Organic Raspberries → Organic Strawberries, Bag of Organic Bananas → Organic Raspberries
- Scatter (Support vs Confidence): Shows strong reliability of frequent rules
- Top 10 Rules by Lift: Focused on organic fruits and daily-use combinations

Insight: Organic products form the backbone of Instacart sales — a clear opportunity for combo promotions.



Key Insights

- Produce & Dairy are the most purchased departments
- Organic items (fruits and greens) dominate frequent baskets
- Customers mostly shop between 9 AM – 12 PM
- High lift values show strong cross-selling potential



Business Impact

- Helps Instacart design *combo offers* (like “Banana + Strawberries” discounts)
- Assists in *store layout optimization* — placing associated items close together
- Supports *recommendation systems* in-app (“You may also like...” feature)

- Enables *data-driven marketing campaigns* focused on organic and daily essentials

Conclusion

The analysis of Instacart's dataset reveals strong purchasing patterns centered around organic and essential grocery items.

Using the *Apriori algorithm* and *Power BI dashboards*, we identified valuable insights into customer behavior that can guide marketing, inventory, and sales strategies.

Author

Krupa Gajra

Data Analyst | Python | Power BI

 Year: 2025

 [LinkedIn](#)