



Market Basket Analysis

Anel Akiyanova

Data Scientist

March, 2021



Problem Statement:

Perform Market Basket Analysis to identify products frequently purchased together

Practical application:

- Product placement
- Store shelf arrangement
- Cross-sales / -promotions
- Customer retention

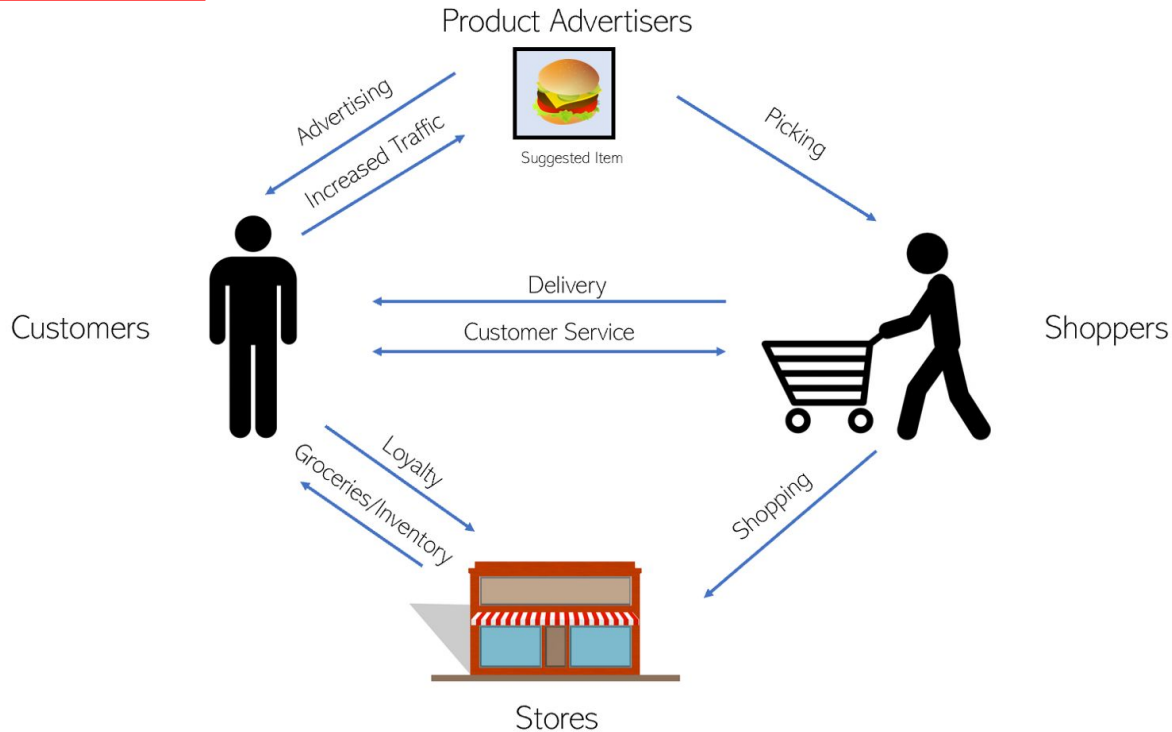


Table of Contents

- Background Information
- Data at a Glance
- Exploratory Data Analysis
- Modeling
- Conclusion and Recommendations
- Prediction App



Instacart Business Model



Source: Medium.com

- Delivery fees + annual membership
- **30% of Instacart purchases** comes from advertisement
- **90% returning** customers
- Express customers spend **\$500 / month**

Instacart Overview

- Surge in orders during coronavirus pandemic
- Expected FY2020 **Revenue \$35B**
- Valued at **\$17.7B (~0.5x Sales)**
- Uses in-house travel time estimator
- Employs Gradient Boosting Decision Tree to estimate fulfillment time



Data Cleaning and Organizing

Data

- 2017 Instacart Data
- 6 Datasets
- 3.4M unique orders
- 34M rows
- 50k products
- 206k user IDs

Cleaning

- Dropped rows with missing values
- Lowercase products
- Take out commas, quotes

Organizing

- Merging dataframes
- Splitting Test, Train, Prior orders
- Breakdown: time of day

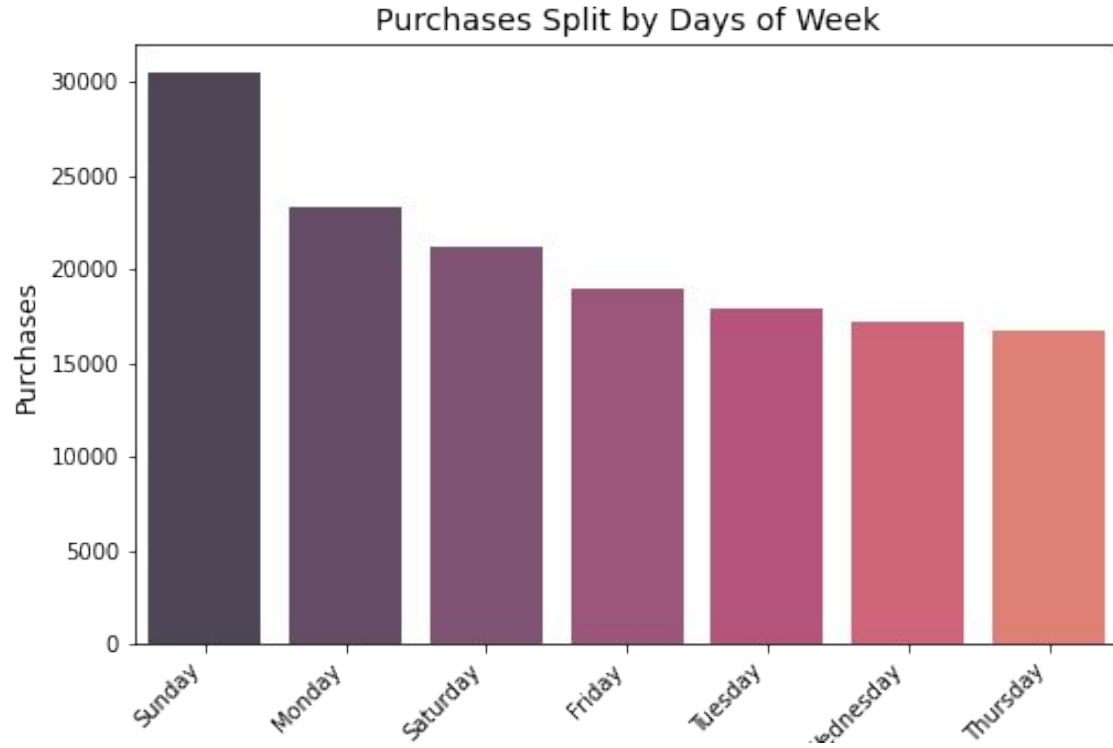
Orders: Days of Week

Busiest: Sunday

Slowest: Thursday

Comparison:

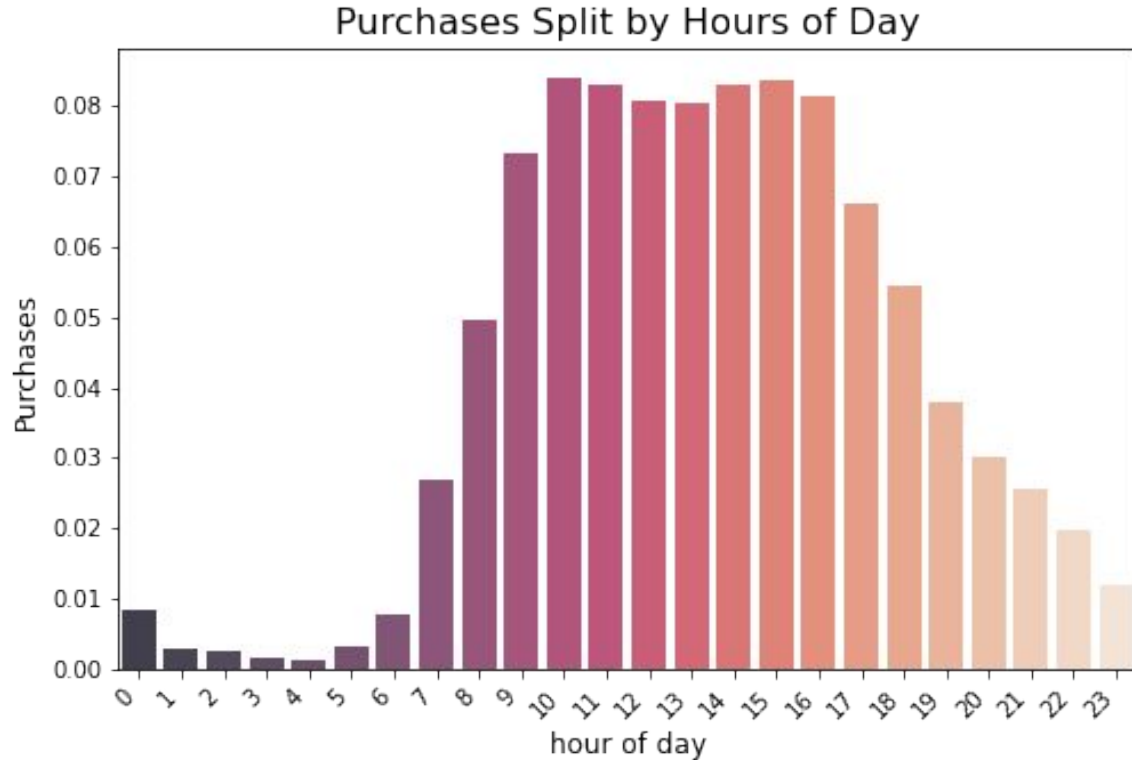
- 21% of all orders made on Sunday
- 12% of all orders made on Thursday



Orders: Hour of Day

Most Purchases are done between **11AM** and **3PM**

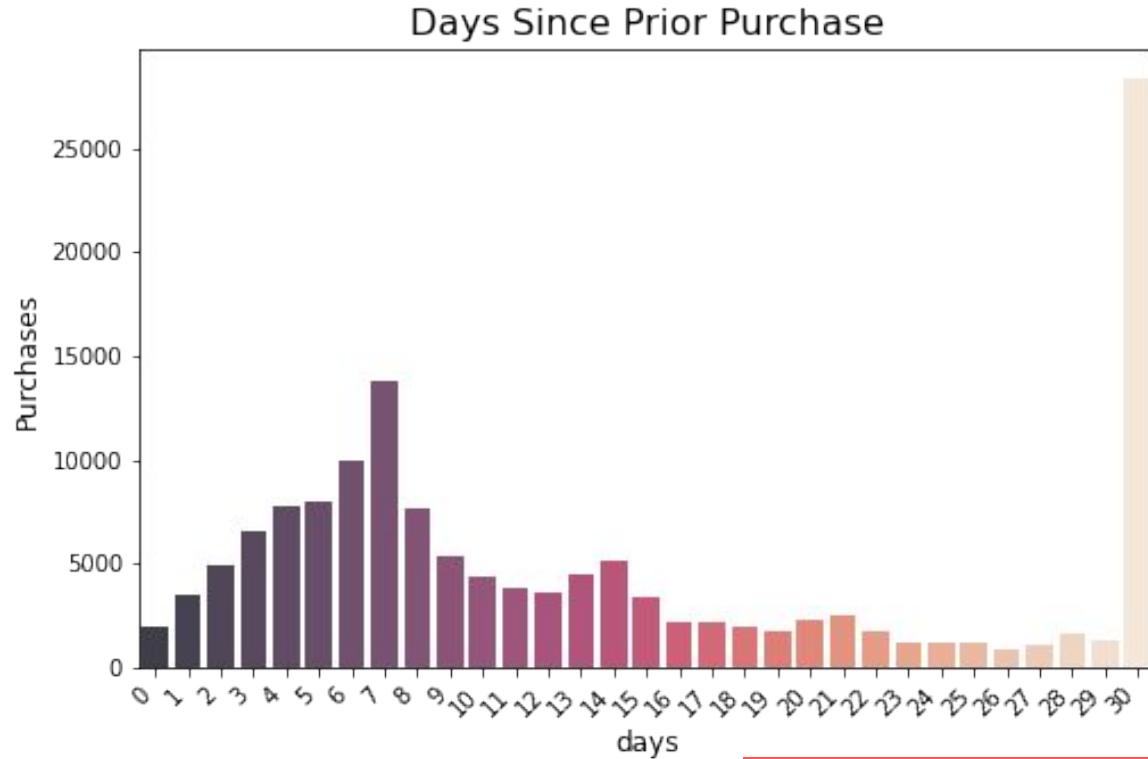
Slowest hours are at night between **1AM** and **5AM**



Orders: Reorders

Cyclical: Most customers reorder products in 7 and 30 days

Lowest repeating orders between **20 and 29 days** since initial order

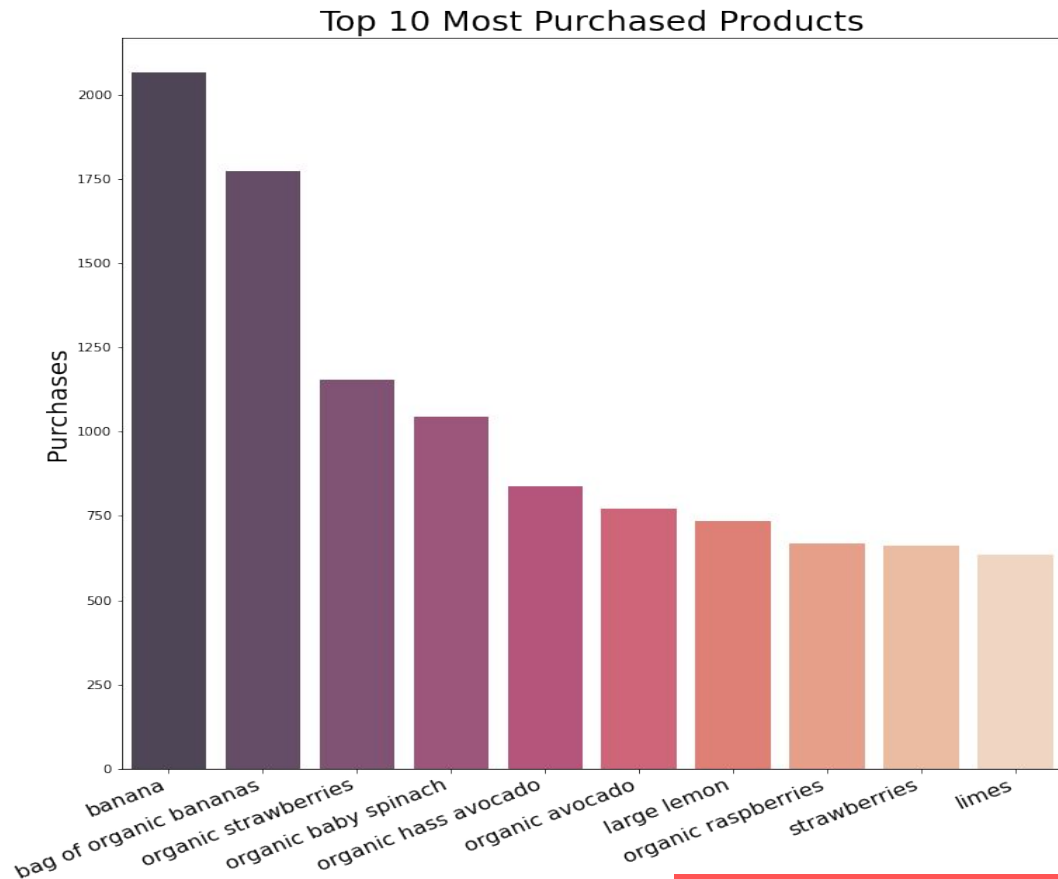


Orders: Products

Top product: Bananas

Top Departments:
Produce, Bakery, Dairy

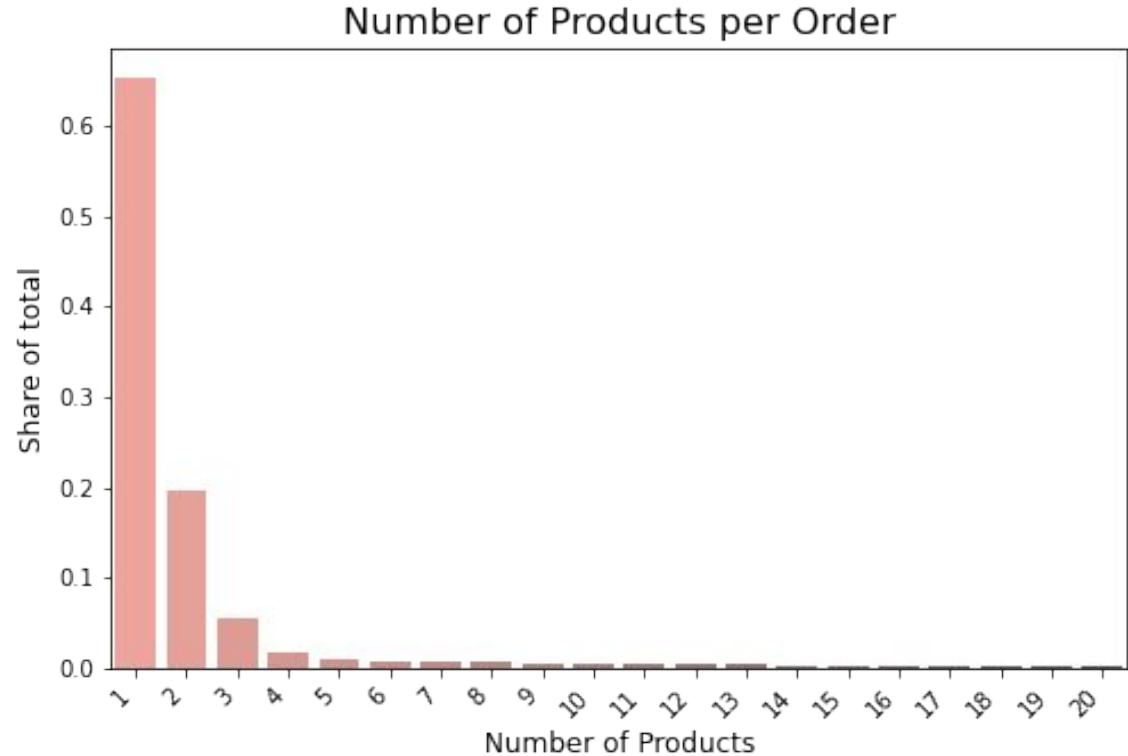
Products added to cart last: granola bars, cookies
cakes, pancake mixes



Orders: Number of Items

1-item orders comprise
over **60% of all orders**

Majority of orders have 1-3
items

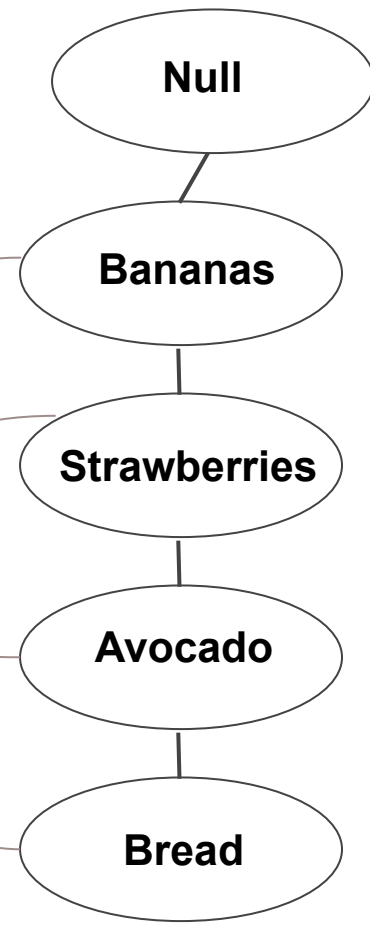


FP- Tree Example

Item	Frequency
Bananas	15
Strawberries	10
Avocado	7
Spinach	5
Bread	5

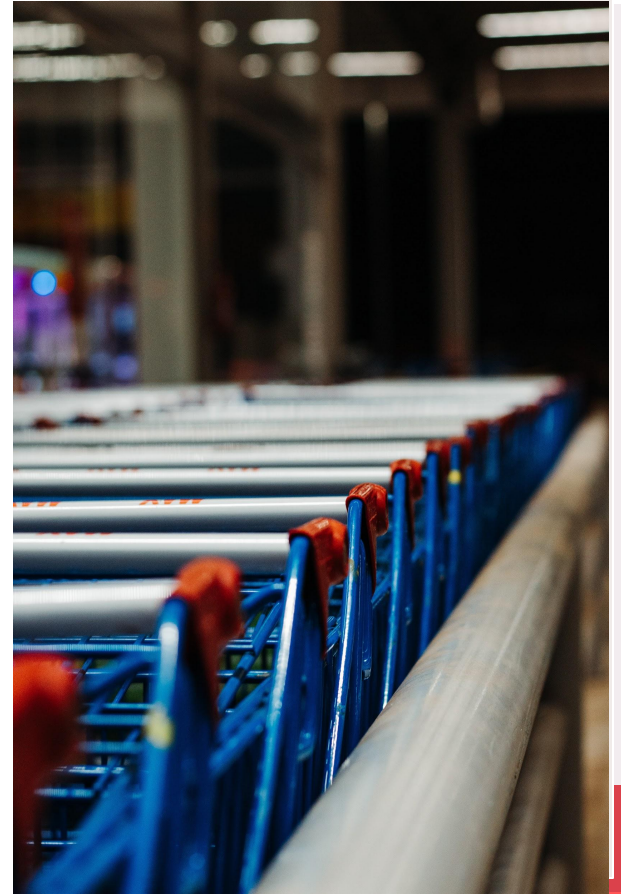


Order ID	Sorted Items
1	Bananas, Strawberries, Avocado, Bread
2	Strawberries, Spinach, Bread
3	Bananas, Strawberries, Spinach
4	Avocado, Bread
5	Bread



Modeling

- FP-Growth Algorithm
- Modeled on 5% of the data in Jupyter
- 50% of data in AWS Sagemaker
- Support Level: 1,000 transactions
- Confidence Level: 10%
- Frequent Antecedent: **Bananas**



Recommendations: Cross-selling, Cross-promotions

Offer:

- discounts on bananas to engage customers
- online recommendations with products purchased together by other customers
- Promotions on lower-priced item and recommend a pair

Customers also shopped for



Tanimura & Antle, Green Onions, 5.5 oz

\$1.10
(\$3.20/lb)

Add to Cart

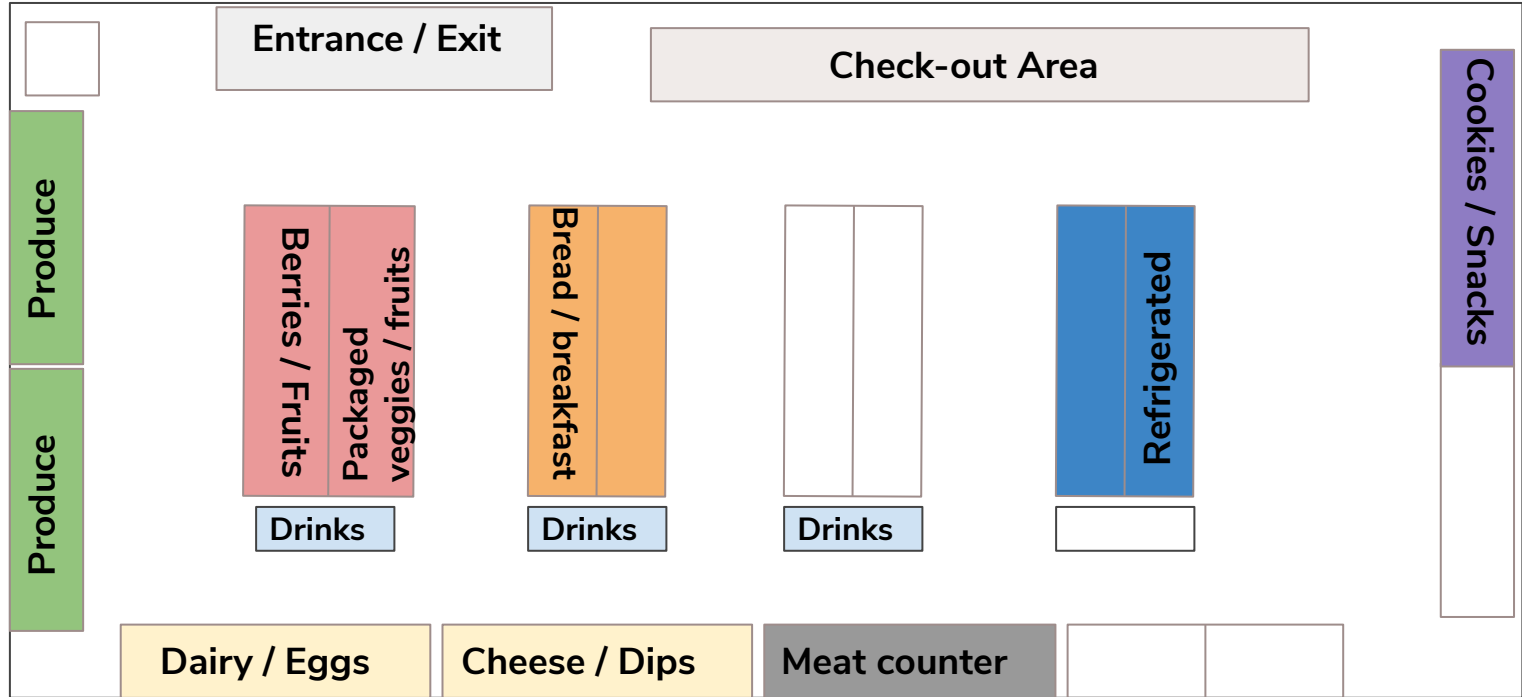


Fyffes, Organic Bananas, One Banded Bunch (5 ct)

\$1.49
(\$0.30/Count)

Add to Cart

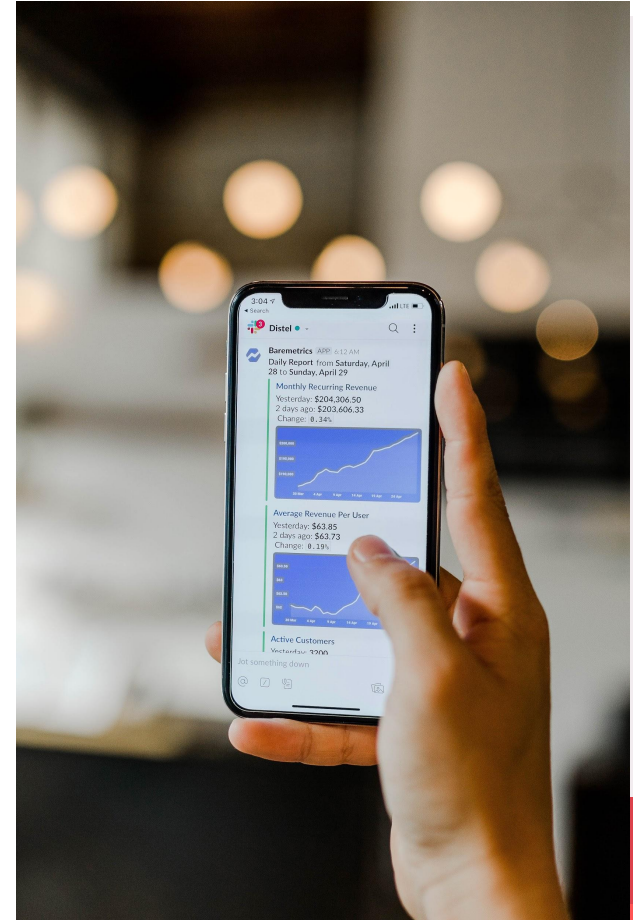
Recommendations: Store Layout



Association Rule App

Streamlit App predicting customer purchases:

- Takes current product as input
- Declares Frequent Pair + confidence level
- Interactive data dashboard





Thank you
for listening!

Full Project [link](#)

LinkedIn [profile](#)