*Team 13*

*Instacart | Eller College of Management*

*INSTACART*

*A Take on Retail Analytics*

A close up of a logo

Description automatically generated

Table of Contents

[The business problem 1](#_Toc40051665)

[Problem Statement 1](#_Toc40051666)

[Client: Instacart 2](#_Toc40051667)

[Deeper Dig into Instacart’s business requirements 2](#_Toc40051668)

[Solution 2](#_Toc40051669)

[Methodology 3](#_Toc40051670)

[Dataset description 3](#_Toc40051671)

[Data Preprocessing 4](#_Toc40051672)

[Relative purchase frequency Analysis 5](#_Toc40051673)

[Market basket Analysis with Apriori Algorithm 6](#_Toc40051674)

[Visual Analytics with Tableau 9](#_Toc40051675)

[Recommendations based on Analysis 12](#_Toc40051676)

[Conclusion 13](#_Toc40051677)

Instacart: A take on Retail Analytics

# The business problem

Recent trends in grocery shopping show increasing dependence on online grocery platforms for even the basic daily supplies ranging from milk and bread to toilet paper. This has given a rise to numerous online grocery delivery platforms like Instacart, Postmates, Cornershop, Buymie, etc. Within the last few years, many big retailers like Target, Walmart, etc. have also come up with their own delivery mobile apps

However, the major focus of these online grocery platforms or delivery apps has been the increase of consumer base and sales at these platforms. The majority of advancements focus on consumers instead of retailers. With this project, our primary aim is to explore analytics such that consumer experience and basket size can be enhanced by enabling retailers through the study of consumer behavior and purchase patterns. These patterns can determine how marketing campaigns like discounts, product bundling, product aisle arrangements or product inventory management can be done.

# Problem Statement

This project focuses on how to engage retailers as well as consumers in the online delivery platforms by guiding retailers with the marketing campaigns and inventory management. This project enables retailers to focus on 3 perspectives listed below.

* Enhancing Customer Experience
* Increasing basket size of consumer basket
* Managing store inventory

According to a Forbes report on what happens to Old and Expired Supermarket Foods, we came across a common business scenario where Patricia Quillen, owner of ‘Country Discount Grocery’ found that 50% of her stock is outdated and 15% is close to its expiry. Small grocery store dealers who do not have extensive predictive data analytics resources like Walmart, etc. often face such issues throughout the year.

This project addresses the following important questions related to this scenario:

* Can we increase sales of this 15% of inventory by user friendly discounts?
* Can we estimate inventory requirements?
* Can we increase visibility of items low in demand?

# Client: Instacart

Instacart is a grocery pick-up and delivery service company operating in US and Canada. It doesn’t own any stores or warehouses and is dependent on close to 300 grocery stores and food retailers in the US to deliver products to its users. On the other hand, these stores have been able to increase their revenue through online sales via Instacart. Instacart has around 16 million monthly visitors on its app and website.

# Deeper Dig into Instacart’s business requirements

Instacart acts as a bridge between grocery stores and their customers. Today, it has developed its own loyal customers(users) who order grocery from the stores through its platform. Thus, Instacart requires to continuously enhance the experience that its consumers and retailers get on its app. On the other hand, Instacart also helps the grocery stores enhance their revenue. This helps Instacart grow its own revenues via delivery fee and also from the ‘mark-up’ prices of the grocery items in the store. Thus, profits increase as the users purchase more grocery items. Additionally, it is in the Instacart’s interests to help the stores in any way possible to establish a goodwill.

# Solution

Our solution proposes how a collaborative relationship between online grocery delivery platforms like Instacart and grocery dealers can benefit both mutually and solve multitude of retail issues. Instacart can utilize its sales data to study consumer behavior patterns and share these insights with its retailers to engage them more. On the similar grounds, retailers can utilize these insights as well as Instacart as a platform to run their marketing campaigns.

The proposed solution can be divided into two parts as follows.

* Analysis of frequent bought together products (by Apriori algorithm) : In this analysis, 3M+ Instacart orders are analyzed to conduct market basket analysis and define about 259 rules of products which are bought together.
* Analysis of basic consumer purchase patterns (Visual Analytics): In this analysis, purchase patterns on different hours of the day or weekday are analyzed along with product demand patterns.

## Based on above analysis, we identified a few recommendations that can be utilized by retailers. Methodology used can be extended and utilized by Instacart on wider dataset to find more interesting patterns and market basket analysis rules with greater support and confidence levels.

# Methodology

Methodology can be divided into 3 parts.

* Dataset description
* Data preprocessing
* Relative purchase frequency analysis
* Market basket Analysis with Apriori Algorithm
* Visual Analytics with Tableau

## Dataset description

We used the users’ dataset published by Instacart here:[*https://tech.instacart.com/3-million-instacart-orders-open-sourced-d40d29ead6f2*](https://tech.instacart.com/3-million-instacart-orders-open-sourced-d40d29ead6f2) . It is a relational set of files describing customers' (users) orders over time. The dataset is anonymized and contains a sample of over 3 million grocery orders from more than 200,000 Instacart users. For each user, Instacart provided between 4 and 100 of their orders, with the sequence of products purchased in each order. Instacart also provided the week and hour of day the order was placed, and a relative measure of time between orders.

Following are the files in the dataset with their attributes:

1. orders
   1. order\_id: order identifier
   2. user\_id: customer identifier
   3. eval\_set: which evaluation set this order belongs in
      1. "prior": orders prior to that users most recent order (~3.2m orders)
      2. "train": training data supplied to participants (~131k orders)
      3. "test": test data reserved for machine learning competitions (~75k orders
   4. order\_number: the order sequence number for this user (1 = first, n = nth)
   5. order\_dow: the day of the week the order was placed on
   6. order\_hour\_of\_day: the hour of the day the order was placed on
   7. days\_since\_prior: days since the last order, capped at 30 (with NAs for order\_number = 1)
2. products
   1. product\_id: product identifier
   2. product\_name: name of the product
   3. aisle\_id: foreign key
   4. department\_id: foreign key
3. aisles
   1. aisle\_id: aisle identifier
   2. aisle: the name of the aisle
4. departments
   1. department\_id: department identifier
   2. department: the name of the department
5. order\_products
   1. order\_id: foreign key
   2. product\_id: foreign key
   3. add\_to\_cart\_order: order in which each product was added to cart
   4. reordered: 1 if this product has been ordered by this user in the past, 0 otherwise

However, considering the time constraint we had, we decided to use only selected files and attributes for our analysis and solution. Following are those files:

1. orders
   1. order\_id
2. products
   1. product\_id
   2. product\_name
3. order\_products
   1. order\_id: foreign key
   2. product\_id: foreign key

## Data Preprocessing

We performed following data preprocessing activities to make the data suitable for use.

1. Merge relevant files (Orders and products)

Our dataset was distributed in different files. We only require data from files ‘orders’ and ‘products’. Thus, first we merged the files orders and products using file ‘order\_products’. We converted this data into a single dataframe.

1. Remove missing values
2. Remove noise: Irrelevant attributes

As mentioned above, we only required attributes: ‘order\_id’, ‘product\_id’, and ‘product\_name’.

1. Reduce dataset size

Reducing the number of data points was especially important when we used ‘Python’ Programming language for our analysis. ‘Python’ wasn’t able to group all three million data points and thus we initially reduced it to 1,00,000 data points at random.

1. Dataset binding : Our dataset had a list of orders which were reordered. In order to make them part of analysis, order with reordered value True were extracted and added again to transaction data.

## Relative purchase frequency Analysis

In this analysis we tried to visualize the relative purchase frequency of products high in demand.

A screenshot of a cell phone

Description automatically generated

It can be seen that banana is one of the most frequently bought product as compared to Asparagus or seedless red grapes. Such analysis can be a basis to place most bought products together in same aisle to reduce in-store movement. At the same time, to increase visibility of low selling products by placing them near high selling products.

## Market basket Analysis with Apriori Algorithm

We performed market basket analysis using Apriori algorithm to generate rules of items frequently bought together. The analysis required creation of transaction file that required merging of products belonging to same transaction id in comma separated values. With the analysis of 13,210 transaction records, the results of the algorithm gave a total of 259 rules altogether for minimum confidence level of 0.8 and minimum support level of 0.001. These 259 rules composition is as follows :

2 products (lhs + rhs) - 134 rules

3 products (lhs + rhs) – 113 rules

4 products (lhs + rhs) – 12 rules

The maximum support achieved is 0.0078 and confidence level of 1 and lift value of 930.

A screenshot of a cell phone

Description automatically generated

The above figure illustrates the summary of rules.

**Two-key Plot of Rules**

A screenshot of a cell phone

Description automatically generated

With the two-key plot, we can observe support and confidence level of different rules based on number of products in the rule. Colors signify the number of products in that rule. For example, order 2 in blue denotes rules containing only two products (lhs + rhs) but order 4 in red denotes rules containing 4 products (lhs + rhs). This plot also signifies how we have a smaller number of rules with 4 products as compared to 2 or 3 product rules.

An example of rules depicting Confidence and support levels with lift value can be visualized as

A close up of text on a white surface

Description automatically generated

According to the above figure, people who buys butter, Organic Butterhead (Boston) are also likely to buy lettuce. Similarly, top 20 rules can be visualized as follows in the parallel coordinates plot.

**A close up of text on a white background

Description automatically generated**

According to parallel coordinates plot we can identify rules like consumers who buys Mango and Happy Baby Spinach also buy Pear Baby Food.

We believe such rules and basket analysis is critical in providing future recommendations and creating product bundles to help improve the customer experience while increasing sales.

## Visual Analytics with Tableau

In this analysis, we established how purchase patterns change with the week or hour of the day. The analysis can be explained by 4 dashboards as follows.

A screenshot of a cell phone

Description automatically generated

In the above dashboard, it can be observed that produce like bananas, fresh vegetables, fruits, etc. are majorly bought on Sunday and Monday and reordered every 7 days. Similarly, by clicking on various departments, we can find weekly purchase patterns of each department and their reorder frequency.

A screenshot of a cell phone

Description automatically generated

In the above dashboard, it is apparent that alcohol is purchased on Fridays. However, other products like Deli and produce are purchased mostly on Sunday and Monday. It can be also observed that alcohols are purchased between 2 pm and 6 pm.

A screenshot of a cell phone

Description automatically generated

In the product demand analysis, it can be observed that organic unsweetened Almond milk is high in demand as compared to Coconut and Almond Milk blend in the dairy and eggs department.

A screenshot of a cell phone

Description automatically generated

In the above dashboard, it can be observed that average time for reordering baby food is between 9 to 10 days.

# Recommendations based on Analysis

Based on our both market basket analysis and visual analysis, a number of challenges faced by retailers and Instacart can be solved as follows.

1. **Discounts based on day of the week (weekdays – T/W/Th/F)** – As it has been observed that sales of departments vary with the day of the week and Tuesday, Wednesday and Thursdays are some of the less busy days of the week. This analysis can be utilized to offer discounts to all the consumers or certain segment of consumers to improve sales. For example, same kind of analysis is leveraged by Fry’s, Tucson to improve its sales on Tuesday by 10% discount on the total bill amount to University of Arizona students.
2. **Product bundling of high selling and low selling products in discount campaigns (e.g. String cheese with Yogurt cheese):** As observed certain products sell better than the other in same department. However, dairy products need to sold faster than other longer shelf life products. In that case, if customers are given discounts on certain items if bought together, it will be a great idea to upsell a few products. For example, Yogurt cheese which is a low selling product can be bought at 10% cheaper price if bought together with string cheese which is a high selling product.
3. **Recommending frequently bought together items (e.g. cold pressed juices to watermelon water buyer):** Various market basket analysis rules suggests the frequently bought together products. With the help of such analysis, it is possible to predict the next product user may buy if he/she places watermelon water in his/her cart. A real-time recommendation of cold-pressed juices will save the application user’s navigation time as well as basket size if user bought the product on suggestions.
4. **Employee management based on the hour of the day (10 am to 4 pm):** It can be observed that most of the daily purchases are done between 10 am to 4 pm. On Sunday, orders are placed after 2 pm. However, on Mondays these orders are placed before 10 am. Such analysis can be utilized to manage the hourly delivery partners of Instacart and retailers. With this analysis, one can estimate the number of partners at a particular hour of the day.
5. **Consumer targeted marketing campaigns** – With an estimate like facial care are reordered every 13-14 days, it is possible to provide consumers buying a facial care product with a discount coupon valid for next 14 days. At the same time, this can be utilized for other marketing campaigns like mail notifications/reminders to consumers for refill. For example, a grocery shopping web app ‘Grofers’ in India sends reminders to its consumers saying, ‘It’s time to buy your face cream’.
6. **Product placing to reduce in-store movement or improve visibility** - With relative frequency analysis, high selling products can be placed together and in a spacious aisle as compared to other products. At the same time, seedless red grapes if placed together with organic banana or Asparagus is placed with Zucchini, their visibility can be improved to enhance sales.

# Conclusion

To sum up, with the market basket analysis and analyzing customer purchase patterns for various products and also with time and day of the week, it is possible to help retailers. This can help retailers in managing product placement in stores, upselling products, running discount campaigns, employee management and also improving customer experience.

Appendix A: Market Basket Analysis

Complete market basket analysis and visualization of all the rules has been uploaded at Rpubs. Rules are demonstrated by means of dynamic charts depicting support, confidence and lift values of rules. The link to the analysis can be given as follows.

<https://rpubs.com/pooja18/612681>

Appendix B: Visual Analytics

The Tableau dashboards has been uploaded to the Tableau Public. The dashboard is dynamic. Please click on various departments, day of the week to observe customer purchase patterns. The dashboard can be accessed at the following link.

<https://public.tableau.com/profile/pooja.tyagi#!/vizhome/DerivingsalesatretailstoreswithInstacartordersinsights/PurchasebehavioratInstacart>