**Analysis Report for Instacart Grocery’s Sales Data**

Elton Zhang

February 15, 2023

**Table of Contents**

Executive Summary…………………………………………………………………………………………………3

Background summary………………………………………………………………………………………………5

Methods and Data……………………………………………………………………………………………………6

Results……………………………………………………………………………………………………………………7

Discussion and further suggestions…………………………………………………………………………14

**Executive Summary**

This report analyzes the sales data of the online grocery delivery platform Instacart from two perspectives: customer preferences and store supplies. Advices for each aspect are made for future development of Instacart’s business and for improving the experience of users.

Key insights and business recommendations:

1. For customers, 70% of all purchases are ordered in 10am – 4pm, indicating the preference of more housewives or purchases from corporations, who prefer to purchase during the day. 40% of weekly sales are ordered in weekend, with two rush hours (Sunday daytime and Monday morning rush) present in the week. Grocery delivery store should hire more drivers that work especially during 10am – 4pm’s peak hours and during weekends and Mondays. Vouchers and sales activities should also be most heavily advertised on all of Instacart’s websites and mobile platforms during these peak hours, with more household-targeting promotions during the weekends and business-targeting promotions during Mondays.
2. 50% of customers have made less than 20 purchases in Instacart, while only 20% of customers purchase more than 25 times. It is advised for Instacart to have promotion activities for both new users and would-be loyal customers. A common but effective way to incentivize loyalty is to set up a progressive customer loyalty reward system that rewards promotion coupons when each customer achieve certain amount of purchases (e.g., a $50 off coupon when customer reach a total of 10 orders at Instacart). Membership bonuses starting from more than 20 orders can also be applied for the top 20% loyal customers. In combination with loyalty fostering, coupons of moderate value (e.g., $10 coupon for first purchase) should be distributed to new users under 5 purchases to motivate purchases of new users.
3. Most customers reorder at this platform in 1-8 days interval. More than 50% of customers order again within two weeks, more than 30 within a week, which indicates a relatively strong stickiness of users. We should send push notifications or email messages along with promotion coupons for the customers every 7 days or according to their normal purchase intervals. AI detection and suggestions of periodic purchase habits should also be adapted to better understand and incentivize regular purchasing.
4. For products, the most popular sales for Instacart overall are fruits, vegetables, and snacks; they are followed by frozen food, dried food, and instant food. Personal care and entertainment products have the least sales for Instacart, which makes much sense since its major business center is grocery delivery. It is advised to increase the purchase from their manufacturers for the most popular sales (i.e., fruits, vegetables, and snacks), and devise promotion events for them. Meanwhile, place them in more accessible and visible places both for the online platforms and offline retail stores.

**Background summary**

Instacart is an American grocery retail company that operates a grocery delivery and pick-up service in the United States and Canada. Its services are available from its website and mobile app. In contrast to amazon’s general online shopping and delivery, Instacart’s services specializes in grocery delivery: it connects grocery retailers with customers’ orders, delivering the orders in a smaller time frame than the average Amazon two-day delivery.

Instacart orders are fulfilled and delivered by a personal shopper, who picks, packs, and delivers the order within the customer's designated time frame—the fastest being within one hour or up to five days’ pre-order in advance. Retailers participating in Instacart's partnership program set the price of individual items on the Instacart marketplace, which are mostly the same prices as in-store. In addition, customers can pick up their pre-made orders from the store through a separate service. For stores that do not participate in Instacart's partnership program, customers can be charged a markup of about 15%-40% per order with individual items ranging from a negative markup to over 50%.

The usual delivery fee is $3.99 for orders of $35 or more and $7.99 under that amount. Regardless of the cost of the order, there is a 5% service fee with a minimum of $2 owed. Customers are also requested to leave a gratuity.

In this report, data about Instacart deliveries are sorted out and analyzed in search of key business insights for the corporation’s future development. Two major categories of analysis are carried out: one aiming at the customers and another analyzing products. Key characteristics of customers and orders are analyzed to better understand customers’ purchasing preferences, space-time trends, and loyalty of their purchases. Concerning products, the total sales count and total sales revenue are calculated and sorted by the departments and aisles of the products. Additionally, all-year-round, recurrent promotion events for popular departments and occasional promotion events can be held to further stimulate the sales of those products. Also, the retailers should increase purchase for those products from the manufacturers to smoothen the supply chain in response to supply shortages in peak periods of the week.

**Methods and Data**

Data charts used in the report are listed as follows:

1. orders.csv (415 thousand entries): listing the details of each order, including the day of week and the hour of day in which the order is made, the customer who made the purchase, and the time interval of the current order since the last order of the customer.
2. order\_products.csv (3.9 million entries): listing the products contained in each order and the chronological order with which they are added to cart.
3. products.csv (50 thousand entries): listing all products, the aisle and department they belong to.
4. aisles.csv (184 entries): listing names of all aisles.
5. departments.csv (23 entries): listing names of all departments.

Examination for the integrity of data is processed before data can be smoothly and reasonably analyzed. Firstly, search for missing data is conducted. In orders.csv, over 2 million missing entries are present in the column “day\_since\_prior\_order”; however, as those NaN values are present only for the first purchase every customer has made, they can be indicated as 0. In aisle.csv, the aisle\_id of pasta sauce is NaN, while there is no aisle with ID of 9; therefore it is confident for us to replace the aisle\_id of pasta sauce with 9.

In addition, the names of aisle\_id are corrupted with uninterpretable segmentation symbols. Using regular expressions to extract English words can clean the names as desired.

The department names are corrupted as well: there exists two departments called “illegal drugs” and “nuclear missiles” that appears to be irrelevant of Instacart’s service. Scrutiny upon their department id may reveal that "nuclear missiles" with ID of 18 is the same with ID of "babies"; "illegal drugs" with ID 6 is the same with ID of "international". Browsing through the department’s product names using regular expression to extract keywords find no words related to nuclear missiles nor illegal drugs. Therefore, it is safe to conclude that those two departments are redundant and should be directly deleted.

With clear data, safe analysis can be conducted. Python’s pandas module is used to read and manage databases, while multiple data visualization modules are used to visualize data and results.

**Results**

Distribution of purchase times over the week

Chart

Description automatically generated

*Figure 1: heat map of purchase density over an average week*

With the heat map, it becomes easier to analyze daily purchase distribution and purchase counts over the week. For all days, the only peak of purchases is at 10am – 4pm, where over 70% of all purchases that day are made. This suggests that more customers may not have a regular daytime job, since daytime workers have little time nor preference to order groceries during working hours. However, the peak of daytime orders indicate the preference of more housewives or purchases from corporations, who prefer to purchase during the day.

Viewing from a weekly perspective, the original data of 0 – 6 is ambiguous, since the “0” that represents day of week can be interpreted either as Monday or as Sunday. Nevertheless, the only peak of orders is near the weekends (with Saturday and Sunday holding nearly 40% of weekly sales), and there are less orders during the middle of weekdays. There are two rush hours present in the week—one on Sunday from 10am to 4pm (a weekend purchase for the coming week when many common households and businesses are free from work) and another on Monday from 9am to 11am (indicating a “Monday rush” of many corporations and workers).

In order to better incentivize purchases and increase revenue with the advantage of such peak hours, Grocery delivery store should hire more drivers that work especially during 10am – 4pm’s peak hours and during weekends and Mondays. Vouchers and sales activities should also be most heavily advertised on all of Instacart’s websites and mobile platforms during these peak hours, with more household-targeting promotions during the weekends and business-targeting promotions during Mondays.

Analysis of Customer Loyalty

Customer loyalty may be measured from two aspects: purchase capabilities (seen from total orders) and stickiness (seen from frequency of purchasing).

Chart, histogram

Description automatically generated

*Figure 2: plot of distribution of customer’s total purchases*

Chart, histogram

Description automatically generated

*Figure 3: cumulative plot of customer percentage for total purchases*

Grouping from the number of total purchases, we can see a clear divide between heavy users and the majority of common users obeying the power law: 50% of customers have made less than 20 purchases in Instacart, while only 20% of customers purchase more than 25 times. 100 counts is an outlier on this graph, probably indicating all customers with more than 100 purchases.

Ranking customers by number of orders made:

• customers who made over 87 orders are among the top 1% loyal customers

• customers who made over 50 orders are among the top 5% loyal customers

• customers who made over 35 orders are among the top 10% loyal customers

• customers who made over 23 orders are among the top 20% loyal customers

Concluding from such data, it is advised for Instacart to have promotion activities for both new users and would-be loyal customers. A common but effective way to incentivize loyalty is to set up a progressive customer loyalty reward system that rewards promotion coupons when each customer achieve certain amount of purchases (e.g., a $50 off coupon when customer reach a total of 10 orders at Instacart). Membership bonuses starting from more than 20 orders can also be applied for the top 20% loyal customers. In combination with loyalty fostering, coupons of moderate value (e.g., $10 coupon for first purchase) should be distributed to new users under 5 purchases to motivate purchases of new users.

Chart, histogram

Description automatically generated

*Figure 4: plot of customers’ days since prior order distribution*

In terms of order frequency, most customers reorder at this platform in 1-8 days interval. More than 50% of customers order again within two weeks, more than 30 within a week, which indicates a relatively strong stickiness of users. The exceptional peak at 30 days is similar to that of abovementioned peak at 100 purchases—a result of all intervals greater than 30 days is counted as 30 days.

In response to such stickiness status quo, it is advised to send push notifications or email messages along with promotion coupons for the customers every 7 days or according to their normal purchase intervals. AI detection and suggestions of periodic purchase habits should also be adapted to better understand and incentivize regular purchasing.

Most Purchased Items

Chart, histogram

Description automatically generated

*Figure 5: plot of distribution for departments’ order counts*

Grouping sales by departments, we may find that the top 6 departments stand out from the rest: produce, dairy eggs, snacks, beverages, frozen, and pantry. Among the rest, departments from “bakery” to “breakfast” have similar sales, while departments from “personal care” to “bulk” have similar sales.

It is therefore advised to design all-year-round, recurrent promotion events for produce and dairy eggs, snacks, beverages, frozen, and pantry departments, and occasionally design events for the rest of the departments

Chart

Description automatically generated with medium confidence

*Figure 6: plot of distribution for aisles’ order counts*

Power law also exists in the sales distribution of aisles—20% of aisles hold 80% of the sales, while the unpopular 80% contribute to only 20% of the sales.

The top 10 aisles and their sales are as follows:

|  |  |
| --- | --- |
| **Aisle name** | **Total orders** |
| fresh vegetables | 150609 |
| fresh fruits | 150473 |
| packaged vegetables fruits | 78493 |
| yogurt | 55240 |
| packaged cheese | 41699 |
| water seltzer sparkling water | 36617 |
| milk | 32644 |
| chips pretzels | 31269 |
| soy lactosefree | 26240 |
| bread | 23635 |

For the top 30 aisles, we should design occasional promotion events that further stimulate the sales of those products. In addition, suggest the retailers for an increase of the purchase for those products from the manufacturers to smoothen the supply chain in response to supply shortages in peak periods of the week.

Individually analyzing, the top 30 most-ordered products are as follows:

|  |  |
| --- | --- |
| **Product name** | **orders** |
| Banana | 92870 |
| Organic Strawberries | 87338 |
| Bag of Organic Bananas | 73739 |
| Organic Baby Spinach | 73237 |
| Large Lemon | 65595 |
| Limes | 54180 |
| Organic Hass Avocado | 50853 |
| Strawberries | 48324 |
| Organic Avocado | 47549 |
| Organic Zucchini | 43034 |
| Organic Cucumber | 42951 |
| Organic Blueberries | 42312 |
| Organic Raspberries | 42310 |
| Organic Garlic | 39006 |
| Organic Yellow Onion | 38030 |
| Organic Red Onion | 37729 |
| Organic Grape Tomatoes | 36868 |
| Asparagus | 36393 |
| Organic Cilantro | 33907 |
| Seedless Red Grapes | 33880 |
| Yellow Onions | 31666 |
| Organic Baby Carrots | 31192 |
| Organic Lemon | 29377 |
| Broccoli Crown | 28315 |
| Organic Whole Milk | 28087 |
| Honeycrisp Apple | 28028 |
| Red Peppers | 26702 |
| Organic Baby Arugula | 26094 |
| Fresh Cauliflower | 25759 |
| Organic Blackberries | 25388 |

Concluding from all the above, there is an agreement among the analyses: the most popular sales for Instacart overall are fruits, vegetables, and snacks; they are followed by frozen food, dried food, and instant food. Personal care and entertainment products have the least sales for Instacart, which makes much sense since its major business center is grocery delivery.

Therefore, it is advised to increase the purchase from their manufacturers for the most popular sales (i.e., fruits, vegetables, and snacks), and devise promotion events for them. Meanwhile, place them in more accessible and visible places both for the online platforms and offline retail stores.

**Discussion and further suggestions**

Further consultation and cooperation with data provider is needed:

* Find out which number among 0 – 6 corresponds to the days of weeks.
* Resolve the maximum value’s outlier issue for day since prior order and total orders of customers.
* Resolve and verify corrupted aisle name issue
* Resolve and verify the missing values

Further analyses could be conducted with more statistics for orders and products:

* Monthly trends and annual trends of orders can be analyzed through more detailed collection of order time
* Progressive trends of sales and orders can also be analyzed with the collection of more order data over the years.

A survey of customers may be conducted to collect subjective data and feedback:

* Collect feedback about purchasing preferences to check whether it is in agreement with the order data collected currently.
* Collect demographic data (e.g., gender, job, age, income, etc.) to further analyze their relationship with purchasing preferences.
* Ask for subjective feedback and advises from the customers to improve customer experience with the orders and the platform.