Centennial College

Capstone

“Analytic Plan: Instacart**”**



Prepared by

Victor Lara

Table of Contents

1. Business Opportunity Brief……………………………………………2
   1. Supporting Insights………………………………………....2
   2. Project Gains……………………………….………………..3
2. Analytics Objective…………………………………………………4
   1. Assumptions…………………………………………………5
   2. Success measures/metrics………………………………...5
   3. Methodology and Approach……………………………….5
3. Dependencies and Risks…………………………………………..6

4.0 References………………………………………………………….7

## 1.0 Business Opportunity Brief

Instacart is the North American leader in online grocery delivery. Instacart shoppers offer same-day delivery and pickup services to bring groceries and other essentials to busy people and families across the U.S. and Canada. Instacart has partnered with almost 600 national, regional and local retailers to deliver from nearly 40,000 stores across 5,500 cities in North America. Instacart’s delivery service is available to 85% of U.S. households and 70% of Canadian households.

Online grocery purchases have jumped to 10% of the $1 trillion industry, more than triple what they were before the pandemic. However, that growth represents a big risk because Instacart’s customers might return to picking out their own products once the pandemic passes. Therefore, Instacart has to define and implement new strategies in order to mitigate these risks.

In addition, each purchase of Instacart is becoming more valuable. For example, Instacart was grossing more than $3 per order by mid-2020, up from a loss of more than $2 per order at the start of 2019. Since the pandemic started, Instacart has delivered three consecutive quarters of positive cash flow, earnings before interest, taxes, depreciation and amortization. Moreover, Instacart’s customers increased from 130,000 in 2019 to 500,000 in 2020. However, consumers expectations and preferences have risen as well. As a result, promotional prices, personalized advertisement, seasonal campaigns, etc. are strategies that every online grocery deliver should consider as part of their marketing campaign.

In order to achieve Instacart’s marketing goals, the company is building an advertising platform that attract targeted audience base on consumers behaviour such as previous orders, time between purchases, purchase time (day and hour) and purchase frequency.

**The specific ask:**

Predict which previously purchase products will be in a user’s next order using “The Instacart Online Grocery Shopping Dataset 2017”. This dataset contains a sample of over 3 million grocery orders from more than 200,000 Instacart users.

For each user, Instacart provides between 4 and 100 of their orders, with the sequence of products purchased in each order. They also provide the week and hour of day the order was placed.

## 1.1 Supporting Insights

Instacart competitors include Amazon Fresh, Walmart Grocery, DoorDash, Dumpling and Postmates. Two of the biggest and most widely available of this list are Amazon Fresh and Walmart. For this analysis we will focus on Instacart main competitors in The USA and Canada.

Amazon Fresh is essentially a digital grocery store. It used to be an add-on service but now comes free for Amazon Prime members. The store offers fresh produce, meat, dairy, seafood, packaged foods, Whole Foods 365 products and household goods, such as cleaning supplies. Instacart and Amazon Fresh operate differently, Amazon put together orders in a warehouse, while Instacart uses personal shoppers who go to local stores near customers location. In addition, customers can even shop for electronics, clothing and toys on Amazon Fresh.  The interface is the same as the rest of Amazon, so using it is easy and familiar. For example, customers just add items to the cart and then check them out.  Amazon Fresh offers free two-hour delivery in most cases if the customer meets an order minimum which vary by city. For a fee in some areas, customers can get their delivery within the hour. Pickup is available in some cities and is free.

Walmart Grocery is available most major cities. Previously, the service allowed customers to order from Walmart for same-day pickup. Now, users can have that order delivered to them, also same-day. Walmart Grocery has existed on a separate app from the main Walmart app, but the company plans to merge them soon. Compared with Instacart, Walmart has physical presence in North America while Instacart not. However, Instacart uses personal shoppers to pick up items from the local stores. Customers can even order from Costco and Sam’s Club without needing a membership. Some areas even allow customers to order from liquor stores, pet stores and pharmacies. As customers are shopping, Instacart indicates substitutes if any items are unavailable. Moreover, customers receive live updates and can track the shopper’s progress on GPS.

As of April 2021, Walmart lead the online grocery market, capturing around 47 percent of grocery delivery and pickup sales. Instacart followed closely behind, accounting for a similarly significant market share at 45 percent. Amazon and the other players capture less than 15 percent of the market.

## 1.2 Project Gains

Once our predictive analysis is completed we will be able to provide valuable information for decision making. For example, problems that are time consuming such as understand your target audience, simplify and improve customer experience, connect with the customers, identify patterns and preferences and advertise globally will be solved in a more efficient way. As a result, we will deliver relevant content, generate more sales, reach new customers and maintain the existing ones, improve the customer engagement, increase the number of items per order, increase average order value, increase the traffic, reduce workload and generate detailed reports.

The following table describes the objectives of this analysis, the goals and indicators that we will study in order to measure the success of the project.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  | **Instacart** | | |
|  |  |  |  |  |
| ***Objectives*** |  | **Sales Growth** | **Brand Awareness** | **Customer Loyalty** |
| ***Goals*** |  | Increase Revenue | Acquire New Customers | Increase Engagement |
| ***KPIs*** |  | Number of items per order,  Number of transactions. | Number of new visitors,  Number of returning visitors. | Number of returning customers,  Online order frequency. |
|  |  |

## Analytics Objective

As we describe in the previous section, the three objectives of this analysis are to increase sales, brand awareness and customer loyalty. The following hypothesis were considered base on a common set of criteria and the outcome of each goal:

We believe that we could increase the revenue if we:

* Designed specific advertising for a specific audience
* Made recommendations base on the customer behavior
* Provided substitute products
* Offered promotional prices and seasonal discounts

We believe that we could acquire new customers if we:

* Offered a new customer discount
* Gave existing users a reward for a referral
* Gave a reward to returning visitors
* Partnered with more grocery stores
* Had a positive feedback from the existing customers

We believe that we could increase the engagement if we:

* Developed a recommendation engine that convert shoppers into customers
* Offered free delivery shipping to loyal customers
* Send mails about new products to the existing customers
* Send email to clients who haven´t been active for a while with discount coupons

## 2.1 Other related questions and Assumptions:

The following assumptions could affect the analysis:

* We assume that the sample is not bias and has been selected randomly.
* Order\_dow column doesn’t indicate the name of the day, we assume that the numbers are related to the days of the week.
* The dataset doesn’t include the target; therefore, we cannot compare our results.
* We assume that because is a relational dataset every table can be linked with its pair.
* We assume that all the columns are relevant for this analysis.
* We assume that an accurate predictive model can be built with the dataset.

## 2.2 Success measures/metrics

In order to define the success of our project we identified the following indicators and metrics:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Instacart** | | | | | | |
| **Objectives** | **Sales Growth** | | **Brand Awareness** | | **Customer Loyalty** | |
| **Goals** | Increase Revenue | | Acquire & keep customers | | Increase Engagement | |
| **KPIs** | Number of items per order | Number of transactions | Number of new visitors | Number of returning visitors | Purchase Frequency | Number of referrals |
| **Targets** | 20% / order | 20 % / day | 500,000 / month | 250,000 /month | 20 % / month | 20 % /month |

We classified the targets by KPI’s and inticated the term that is being estimated.

## 2.3 Methodology and Approach

**Type of Analysis:** *Neural Networks, Gradient Boosting Machine.*

*The initial approach will be merging the different data frames into a single one. After that, I will try to classify the customers. Finally, I will be able to predict which products will be in the customer’s next orders.*

**Methodology:**

*We will start by exploring all the features we will be able to answer questions related to Number of Purchases, Hour and Day Ordered, Most Ordered Products, Orders per day and week, Best Selling department, Best Selling Aisles, and so on. The second step will be to find all possible clusters among the different customers. We can try to reduce the features using Principal Component Analysis. Finally, we will use the best predictive model. Fir this analysis we can use Neural Network, lightGBM and/or XGBOOST.*

**Output:** *The output will be a set of insights, rules and strategic recommendations that will help us to build an accurate recommendation system that helps the company to make better marketing decisions.*

## 3.0 Dependencies and Risks

|  |  |  |
| --- | --- | --- |
| **Risk** | **Likelihood (based on historical data)** | **Impact** |
| *Amount of data being collected* | *Medium* | *The outcome of our analysis will not be representative to the entire population. Therefore, the analysis will be irrelevant.* |
| *Collecting pre-covid data* | *High* | *The outcome will consider a pre and not a post-covid environment. The analysis will not be meaningful.* |
| *Improper Exploration Techniques* | *Low* | *We will not be able to showcase the relevant features of our data.* |
| *Selecting the best Predictive Model* | *Medium* | *If our model is not the one with the highest accuracy, our prediction will not cause a positive effect on the sales.* |
| *Consider all tables of the dataset* | *Low* | *Some of the tables may not be relevant for the prediction.* |

4.0 References

Sorvino, C. (2021, January 28). *Instacart Survived Covid Chaos — But Can It Keep Delivering After The Pandemic?*Forbes. https://www.forbes.com/sites/chloesorvino/2021/01/27/instacart-survived-covid-chaos---but-can-it-keep-delivering-after-the-pandemic/?sh=3049cb13bfa1

Curry, D. (2021, November 11). *Instacart Revenue and Usage Statistics (2021)*. Business of Apps. https://www.businessofapps.com/data/instacart-statistics/

Curry, D. (2021, November 11). *Instacart Revenue and Usage Statistics (2021)*. Business of Apps. https://www.businessofapps.com/data/instacart-statistics/