MMA831: Marketing Analytics

Final Project Report
Instacart

August 20, 2021

Team Kipling

Faye Ding	20250430
Matthew Gottlieb	20259474
Adaure Ibe	20254264
Angie Jung	20254266
Justin Liu	20254488
Mahendar Rawat	20254140

Queen's Smith School of Business

Contents

Executive Summary

The purpose of this project is to explore the value that Instacart can bring to its users and shareholders by implementing marketing strategies powered by Machine Learning. Being one of the fastest-growing e-grocery brands in North America, Instacart seeks to develop a strategy to achieve a sustainable competitive advantage over its main competitors, such as Amazon and Walmart. This report provides an overview of three modelling techniques (K-Means, Apriori, XGBoost) developed for customer segmentation, products association, and recommender system to address Instacart's objective of increasing sales.

Exploratory data analysis was conducted to gain high-level business insight into Instacart products and users. For example, our team noted that a high percentage of Instacart users are organic fruits and vegetable lovers and follow an urban lifestyle (weekend orders). Another unique insight was that most Instacart users were new or exploring customers. The loyal customers only represented 5% of the user base. This insight highlights that retaining this new customer segment is necessary for Instacart to build brand loyalty among users and ultimately improve sales.

Our first model for customer segmentation results in three clusters. We assessed several statistics for each cluster, such as the average number of total orders, average number of days between orders per user, and average number of items purchased for each order.
The second modelling technique, Apriori, focuses on which products will most likely be purchased together. We found some insightful product associations such as organic Hass avocado with a bag of organic bananas, organic strawberries with organic bananas and banana with organic avocado. The understanding of the product associations will help the business in exploring cross or up-sell opportunities.
Lastly, our *R*ecommender system model predicts the items a user is most likely to purchase in their next order. We introduced several key features to help our objective. These included reorder ratio of a user, reorder ratio of a product, the last five products purchased by a user and many more.
Based on our findings, our suggested recommendations and estimated economic impact include:
Strengthen negotiation power with grocery stores by using customer insights to earn additional commissions from local stores. The estimated economic impact is approximately $476,000 per day.
Issuing 15% and 10% coupons to targeted cluster 1 and 0 respectively. The estimated economic impact is additional sales of $246,557.25.
Promote bundled platter and gift basket based on association rules. The estimated benefit is an additional $13 per platter and $8 per basket.
Continuously develop a comprehensive recommendation system to detect hidden sales opportunities The estimated economic impact is an additional $1,400,000 per day.
With the suggested recommendations above, Instacart can improve sales and maintain robust and sustainable growth for the future.

Business Overview

About Instacart

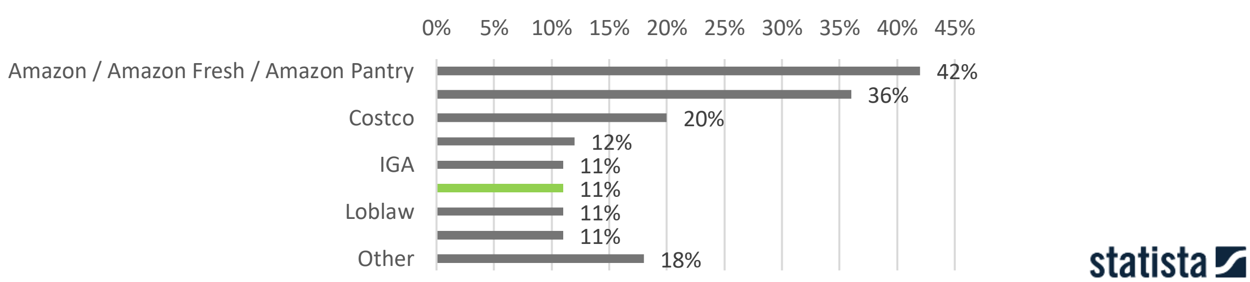
Instacart is an online grocery delivery startup. Without operating any grocery stores, the company implements sharing economy-based business model by connecting customers with personal shoppers and providing a platform for retailers to sell their products. Users can order products through Instacart's app or website, and shoppers who are independent contractors in Instacart, receive orders, shop, and deliver items to their doorsteps. The company partners with over 350 retailers across the United States and Canada and generate revenues mainly from membership, delivery, and placement fees.

Instacart's value proposition emphasizes customer experience and the freshness of the products. The company focuses on on-time delivery within an hour of their customers' orders and personalized shopping experience. In terms of customers, Instacart has a high share of 25 ~ 45-year-old urban users who tend to have less time for a proper meal and who are active social media users[1]. As a result, Instacart is one of the leading online grocery delivery brands in terms of market share and customer satisfaction in Canada and the United States.

Industry Analysis

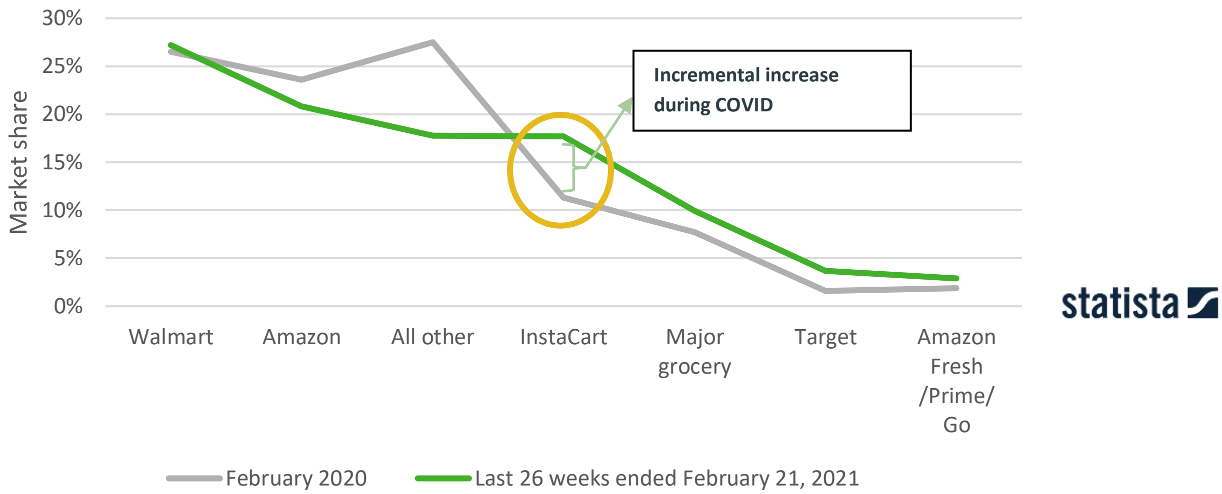
The online grocery market has experienced rapid growth in the past years, especially since the pandemic. Currently, Instacart is among the top e-grocery brands in North America, representing about 30% of the market share with its major competitors, including Walmart, Amazon Fresh and Costco. Furthermore, other on-demand food delivery services such as Uber or DoorDash have also expanded their businesses to e-grocery. The market expects increased competition as more consumers are showing patterns of shifting their grocery shopping online.

- **Survey - Recent grocery orders placed, Canada 2020**



[1] Statista Report 'Grocery delivery services: Instacart in Canada (2021)',

- **Online food grocery market share in the US, 2020 ~ 2021**



- **Porter's 5 Forces Analysis**

Threat of new entrants	• Moderate - entry barrier exists due to the high costs associated with online grocery shopping such as delivery and membership fees
Bargaining power of buyers	• High - buyers can always choose to go grocery shopping at the expense of their convenience, or switch to another service. Cost of switch is low.
Threat of substitute products	• High - Strong competitors offering/ launching similar services
Bargaining power of suppliers	• Moderate - due to the growing risk of COVID-19, opportunities exist for Instacart to improve bargaining power towards suppliers
Rivalry among existing competitors	• High - retail giants including Amazon Fresh, Walmart are main competitors and Uber and Door Dash also expanded their business to e-grocery

Key Marketing Challenges

To keep up with a growing customer base and continue to drive sales, Instacart needs to develop an ongoing strategy to differentiate its service to compete with some of the world's most prominent retail players, including Walmart and Amazon Fresh and similar on-demand delivery player, Uber. Due to the low switching costs by users and bargaining power towards retailers, Instacart must leverage customer insights effectively to maximize user convenience and satisfaction and increase their customer base and sales. Leveraging customer insights, key marketing challenges for Instacart include 1) efficiently marketing target customers based on different purchasing patterns and 2) attract and retain customers in a highly competitive market.

Business Problem

Instacart can drive sales and increase its user base by developing the following three strategies.

Customer Profiling

- Understand customer mix
- Identify different customer segments with their unique needs
- Recognize customer purchasing behaviours and trends

Targeting Customer

- Identify the most profitable customer segment

Recommendation System

- Categorize products with strong associations
- Products to recommend based on customer buying patterns

With Instacart's lean operations and strong retailer networks, understanding customer mix, targeting the profitable customer segment, and providing personalized search results for all users will ultimately improve overall customer experience, increase user base and unlock business values for partner retailers by helping them better understand its customers and generate higher sales.

Analytical Solutions

As mentioned in the previous section, Instacart is seeking to improve sales through observing customer purchasing behaviour, identifying customer segments, and improving product recommendations. To achieve this, we will be implementing various analytical solutions to solve Instacart's business problems. These analytical solutions involve using machine learning techniques to resolve the multiple challenges Instacart is currently facing.

Identifying Instacart's customers implementing targeted marketing can be done through segmenting customers via clusters. This entails segregating its customers based on their mutual similarities and differences, given their demographic and consumer behaviour trends. The main benefit of this analytical tool is that it will allow Instacart to improve its overall customer experience through personalized marketing. When Instacart understands its clients' preferences, it can offer services and products that are best suited to their unique needs. This could potentially translate into an improved relationship with its customers due to increased customer satisfaction and loyalty. Customer segmentation and targeted marketing also allow Instacart to reduce costs associated with mass marketing by only sending targeted ads to the customers who they predict will most likely be interested in a specific product or service. The third benefit of this analytical tool is that Instacart can focus on the most profitable cluster and make it its primary customer base. That way it can utilize its resources better by channelling them to cluster with the highest return on investment. All these added benefits will ultimately result in increased revenue for Instacart.

Technique: A few machine learning tools can be used to achieve this, namely K-Means, Hierarchical clustering and DBSCAN. We opted to use K-Means for this project because it is easier to use and had a faster processing time than the other algorithms.

We conducted the next analytical solution to solve Instacart’s recommendation problem for newly acquired customers through Cross & Up-Selling. With newly acquired customers, it’s a challenge to make effective product recommendations because there isn’t historical data about these clients our algorithm of choice can leverage. Nonetheless, Instacart can still make recommendations to these newly acquired customers using association rules. Its algorithm can learn patterns from the transactions of its previous clients on what goods customers tend to buy together. The algorithm can then use these patterns to predict what products are likely to be associated with the products its new customers purchase. The main benefit of this is that Instacart can recommend additional products to customers which they are likely to buy. They can recommend complementary products as well as similar products that are higher end. The effective implementation of this system can overall translate into increased sales due to customers making additional purchases from what they originally intended to buy.

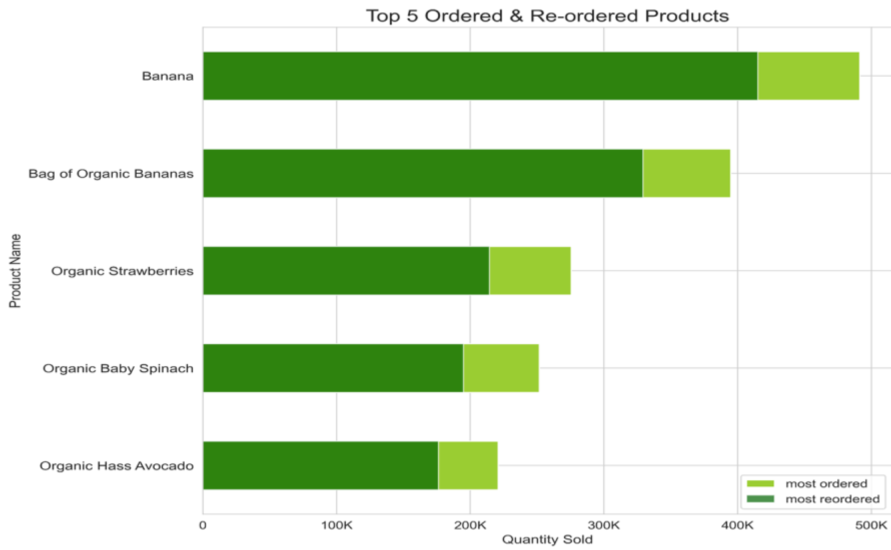
Technique: We explored various ML techniques, including Apriori, Eclat and FP-growth but opted to use the Apriori algorithm. It was simple to use and the most interpretable.

The final analytical solution is an overall effective recommender system. The focus of this tool is to cater to making product recommendations for already existing customers based on their previous buying patterns. The key benefit of this is developing personalized marketing strategies for these customers based on insights derived from historical data. The recommender system then uses these insights to make predictions on what products the consumer might likely pair with the existing products in their current basket. Instacart can leverage this through bundling associated products together, or even displaying a recommended product at the checkout screen to encourage buyers to purchase additional items. This will overall translate into increased sales for Instacart.

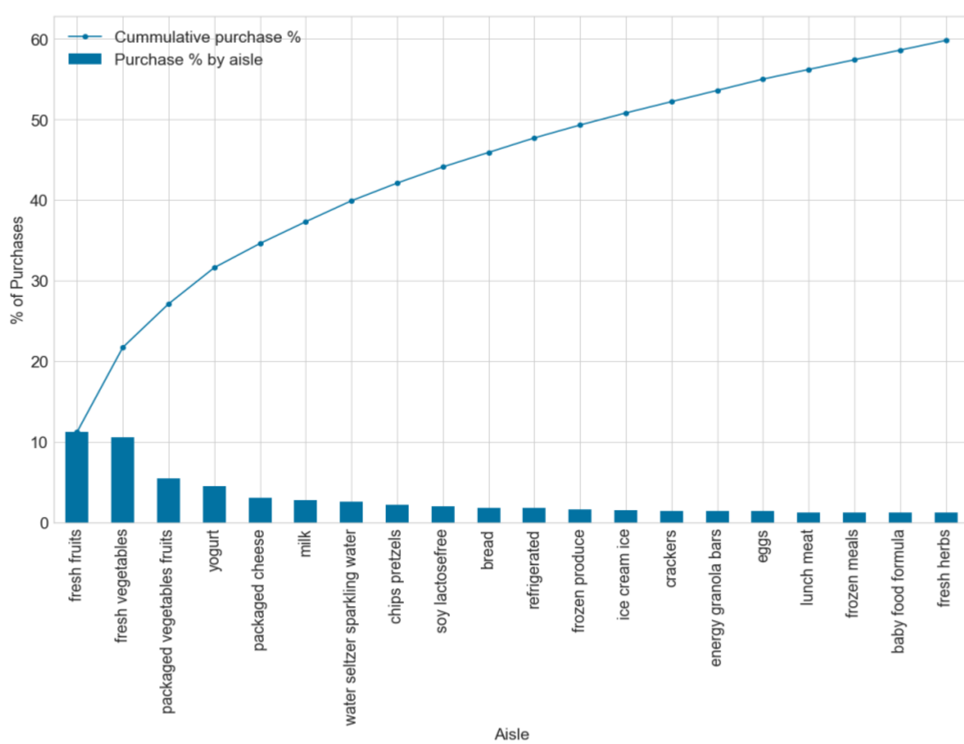
Technique: XGBoost Classifier

Understand the Data (EDA)

Products and Departments

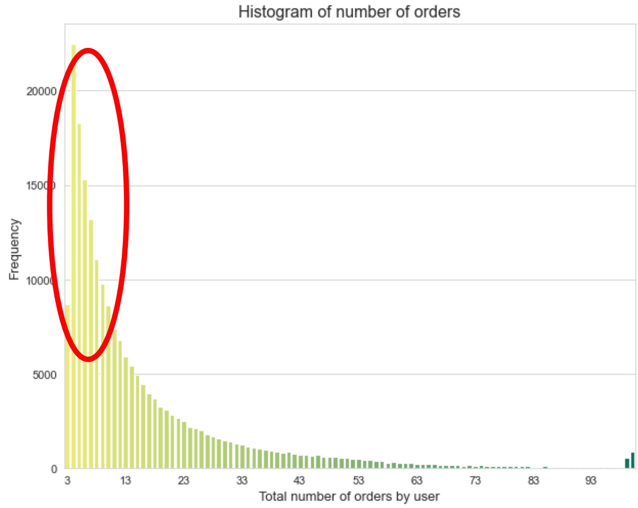


From the graph above, the most popular products sold (banana, strawberry, avocado, etc.) are associated with fruits or organic, healthy food. This presents the insights of the customers that they tend to have healthy eating behaviors; this may also imply that the majority of Instacart customers are millennials because they are the biggest cohort of organic shoppers.

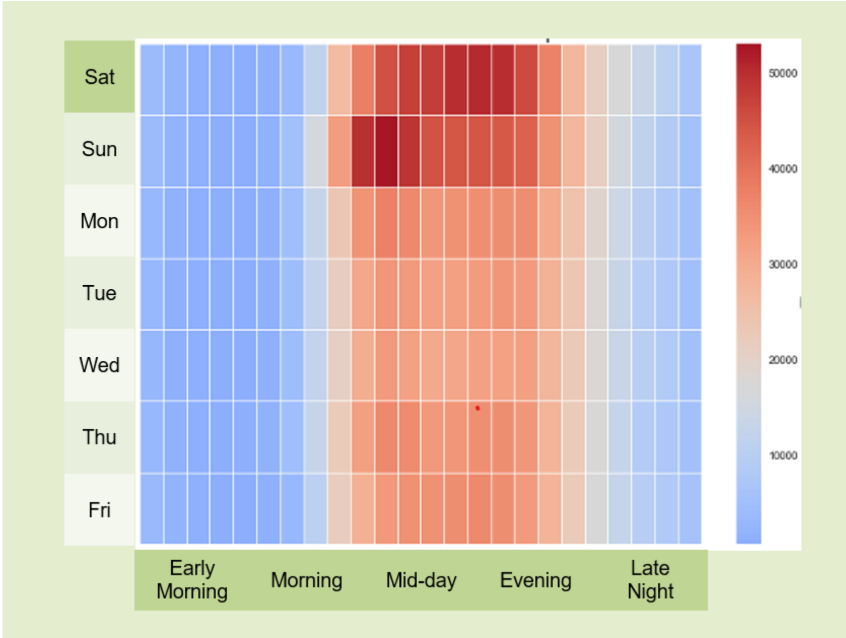


60% of all items purchased come from the 20 aisles above; and the most popular products are fresh fruits and vegetables. Most products fall under grocery category which resonates with Instacart's business model that they focus on providing groceries to customers. This also supports with the insights discovered previously.

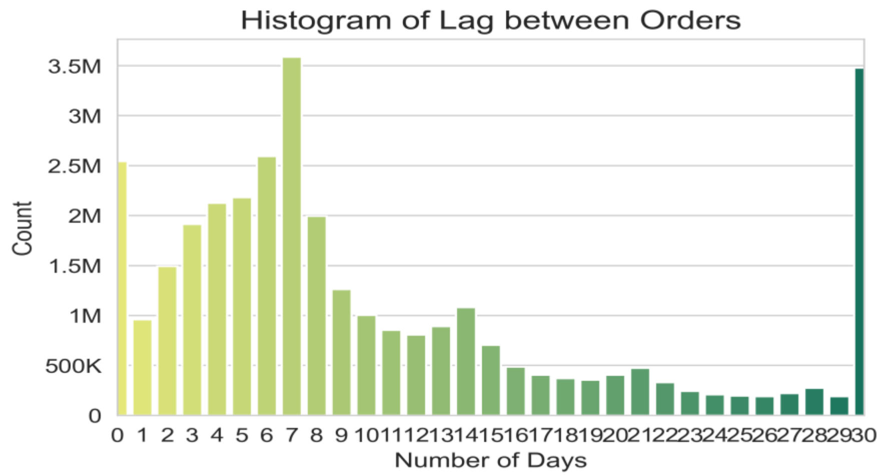
Order and Users



From the graph above, we can see that the histogram is skewed to the right; high frequency falls under the orders between 3 to 23. In other words, Instacart customers tend to put small orders when they shop.81% of customers are most likely to put orders under 25 and most of them are still new to the Instacart app. As the order increases, the count of customers decreases drastically which means loyal customers are still minimal, only count for 5% of the user base.



In addition to the user and order insights, we can find out that most customers shop during weekends using Instacart; and they shop mostly during mid-day time. This may imply that Instacart customers are busy working during weekdays, and they only have time shopping for groceries during the weekends.



From above, we can find out that the graph is skewed to the right and most shopping gap are between 3 to 9 days. The most counts for the number days of gap are 7 days which implies that majority of the users shop at Instacart on a weekly base. It clearly supports the insights discovered from the purchase heatmap.

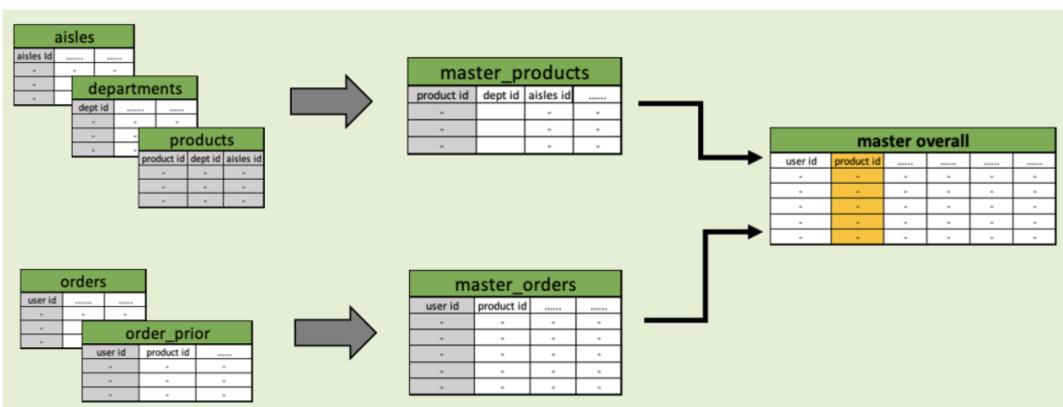
Conclusion of EDA Findings:

- Instacart customers mostly shop for healthy, organic fruits and vegetables
- Customers most likely are millennials
- Only 5% of users are loyal and most users are still new to Instacart
- Customers mostly shop during weekends because they work in weekdays

Data Preparation

When beginning to prep the data, it is first essential to understand the original sources and formatting of it. Our data was accessible directly through Kaggle and was robust in its offerings, with seven total zip files available to us. The data files included information about the products themselves, customer orders, the aisles that the products sit in, and the department they belong to. We opted to load the zip files directly into our notebooks and open them within our Python environment rather than opening them locally due to their size.

Once loaded into Python, the next step was to determine how to merge these datasets together to create a master file. We found a commonality among our datasets through various id codes and used these to build links; this included features such as product id, aisle id, department id, and order id. Once merged together, we created a master data frame that we could use for our EDA and modelling later on. It is also important to note that there was no missing data; however, we did notice that new users (approximately 6%) had NaN for historical purchases, which makes sense. Therefore, we replaced NaN with 0. Due to the nature of the data, there was also a high amount of cardinality within it. We expected this due to the nature of Instacart's business, as there can be thousands of individual products available to purchase in a given store. This was particularly evident within the "product name" and "product id" features.



Once our main data frame was created, we were able to build off of it during our EDA process. For example, we created a sub data frame to generate a heatmap using the features "order hour of the day", "order day of the week" and "user id". This enabled us to gain insights into the number of unique users by day of the week and hour of the day. In another example, we created a sub data frame by grouping together product name and order id. We then aggregated it by counting each unique instance a product was ordered. This allowed us to view the top selling products based on the total number of times it was ordered.

Overall, the biggest obstacle in the data preparation section of our analysis was the merging of our seven datasets. However, once merged, we were happy to observe a robust master dataset with relatively few missing values across all features that we could then feed into our models.

Modelling and Performance

Once we have a thorough understanding of the data, we performed an exploratory data analysis to gather some high-level business insights. Now, we proceed to apply the modelling techniques we discussed earlier to solve our business problems.

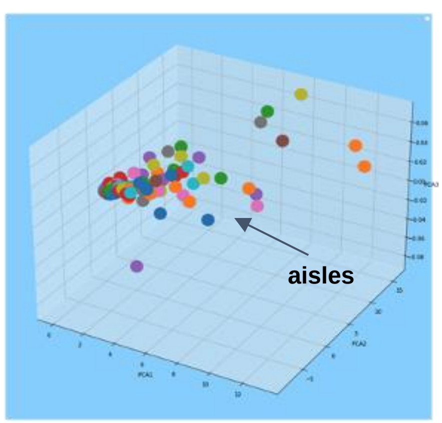
Customer Segmentations

The goal is to understand the market mix of Instacart to create customer groups that reflect similar purchasing behaviour. This exercise will help our team find the distinct characteristics of different customer groups and ultimately help the Marketing team address each group more effectively. For instance, Instacart could create targeted marketing campaigns to appeal to each customer group, which would be far more effective than one generic campaign for all.

Instacart's purchasing data was available to us at three levels, i.e., by product, aisle, and department. We can perform customer segmentation at any of these three levels. Still, to keep the data in a sizable volume, we decided to utilize aisles representing categories of products to segregate the customers using K-Means clustering. This unsupervised learning algorithm is one of the most popular techniques and is easy to implement.

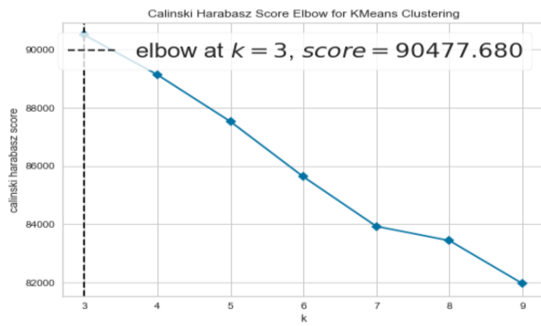
We performed Principal Component Analysis (PCA) to reduce the number of dimensions from 134 to 3 components for the following reasons.
a) to be able to visualize the data
b) K-Means does not work well with higher dimension data
c) to eliminate redundant features

We understand that by reducing the data, we are losing the information. Still, in our case, we can retain ~55% of the variance, which seems to be a reasonable trade-off for the reasons mentioned above. After PCA, to visualize the position of aisles relative to each other, we used aisles PCA coordinates to plot the aisles on a 3D graph.

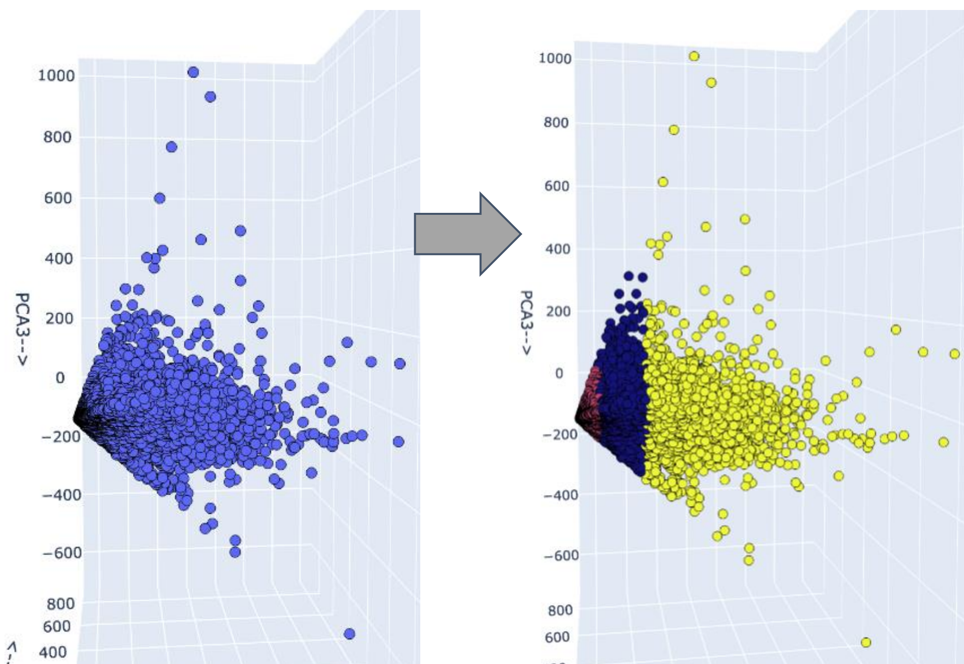


Note: Excluded Yogurt, Fresh fruits, Fresh vegetables, and Packaged vegetable fruits aisles in above graph to zoom in on the other aisles.

Once we have the PCA transformed data, the next step is to determine the appropriate number of clusters (K) we want the KMeans algorithm to categorize the data. We used Calinski Harabasz (CH) Score to find the optimal number of clusters. We observed the highest CH score at the three clusters.



Now, we apply KMeans algorithm using the result of our hyperparameter tuning exercise to sort the customers into their respective clusters. Below chart on the right visualizes the clusters in different colors, which seems to be distinct and well separated.



Customers in PCA space **Clusters formed after KNN**

In the end, we used Silhouette Score to measure how similar the points are to their own cluster. It essentially checks the goodness of our clustering technique. A score of 0.63 indicates well apart and distinct clusters.

```
score = silhouette_score(pca_aisle_trans, labels_top_n)
print("Silhouette Score: ", score)
```

```
Silhouette Score:  0.6396256805029651
```

Market Basket Analysis

This is a crucial technique used by businesses to discover insightful relations between two or more products. The idea is to search for combinations of items purchased together frequently in a basket (cart). The goal here is to leverage the knowledge of associations between products to up-sell and cross-sell products, to recommend products or to create new pricing models to improve sales.

To discover associations between Instacart's products, we used the Apriori association rule learning algorithm to mine frequent item sets.

```
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
```

We first started with aisles data to find associations between product categories, however, we found many redundant associations between aisles with low support and low lift. Some examples are shown in below table.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
448	(packaged cheese, packaged vegetables fruits, ...	(fresh fruits)	0.112038	0.807374	0.105832	0.944602	1.169968	0.015375	3.477130
435	(packaged vegetables fruits, milk, yogurt)	(fresh fruits)	0.121191	0.807374	0.114249	0.942719	1.167636	0.016403	3.362827
391	(packaged cheese, fresh vegetables, yogurt)	(fresh fruits)	0.126435	0.807374	0.118440	0.936766	1.160263	0.016360	3.046244
420	(fresh vegetables, packaged vegetables fruits,...	(fresh fruits)	0.192543	0.807374	0.180198	0.935886	1.159173	0.024744	3.004447

To overcome this issue, we made another attempt using product-level data to find meaningful associations between products by clusters. We observed that product associations have much higher support and lift compared to aisles. The is due to the large volume of data available for products, which means there is a higher chance of discovering relationships.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
2	(Organic Hass Avocado)	(Bag of Organic Bananas)	0.041087	0.096078	0.011279	0.274518	2.857232	0.007331	1.245961
3	(Bag of Organic Bananas)	(Organic Hass Avocado)	0.096078	0.041087	0.011279	0.117394	2.857232	0.007331	1.086457
4	(Organic Strawberries)	(Bag of Organic Bananas)	0.063429	0.096078	0.015147	0.238798	2.485443	0.009052	1.187492
5	(Bag of Organic Bananas)	(Organic Strawberries)	0.096078	0.063429	0.015147	0.157648	2.485443	0.009052	1.111853
9	(Banana)	(Organic Avocado)	0.123296	0.046537	0.012160	0.098629	2.119351	0.006423	1.057791

Recommender System

To improve the customer experience of Instacart users, we developed a Recommender System model to predict the items a user is most likely to purchase in their next order based on past purchasing behavior. The objective of building this model is to increase the order size of users to eventually improve Instacart's revenue. To build this model, we used user, product and order data provided to us in the competition to predict if a user will reorder a product or not. The dataset includes the user's historical purchase data and other details such as aisles, department, day of purchase, the hour of purchase, days since last purchase and reorder.

Selecting the data

We combined the segregated data into one master file to extract new features for the utility matrix by users and by products. From our understanding of the data, we knew that we would need to explore new features to be able to prepare a robust dataset.

Constructing data (Feature Engineering)

We extracted several features by combining the order, user, and product datasets. Some of these features proved to be very important for our model (see appendix for the list of important features).

Following are the new features we introduced:

1. Total number of orders by user (users_totals_u)
2. Average number of orders per user (number_of_orders_u)
3. Average number of days between orders (average_days_between_u)
4. Average number for reordered products by user (reorder_ratio_u)
5. Order day of the week with the most items (most_dow_u)
6. Order hour of the day with the most items (most_hour_u)
7. Total amount of different items bought (specific_items_p)
8. Total amount of purchases per product (total_purchased_p)
9. Ratio of repurchase for each product (reorder_ratio_p)
10. Mean position that the product is added to the cart (cart_position_p)
11. Number of times a user bought a specific product (times_bought_up)
12. Times a user purchased the product after purchasing it once (amt_bought)
13. Total orders used for calculating range, support (total_orders)
14. Finding when the user has bought a product the first time (first_order_num)
15. Number of times product bought in last five orders (last_five_up)
16. Ratio of the products bought in the last five orders (ratio_last_five_up)

We used target encoding to convert aisle_id and department_id into more meaningful values before splitting the training and validation data. Here is a visual of how the final dataset looked like.

user id	product id	New Features			Reorder
-	-	-	-	-	1
-	-	-	-	-	0
-	-	-	-	-	1
-	-	-	-	-	1
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-
-	-	-	-	-	-

Modelling – XGBoost Classifier

At this point, we have the final dataset ready to be used for modelling. We used a XGBoost Classifier model to fit the training data with tuned parameter values derived from a grid search.

```
XGBClassifier(base_score=0.5, booster='gbtree', colsample_bylevel=1,
              colsample_bynode=1, colsample_bytree=1, gamma=0, gpu_id=-1,
              importance_type='gain', interaction_constraints='',
              learning_rate=0.300000012, max_delta_step=0, max_depth=6,
              min_child_weight=1, missing=nan, monotone_constraints='()',
              n_estimators=100, n_jobs=8, num_boost_round=10,
              num_parallel_tree=1,
              parameters={'colsample_bytree': 0.4, 'eval_metric': 'logloss',
                          'max_depth': 5, 'scale_pos_weight': 0.6,
                          'subsample': 0.8},
              random_state=0, reg_alpha=0, reg_lambda=1, scale_pos_weight=1,
              subsample=1, tree_method='approx', validate_parameters=1,
              verbosity=None)
```

We then used a function to tune the threshold to classify reorder or no reorder to maximize the F1 score. The function suggested the best F1 score of 0.4337 at a probability threshold of 0.20.

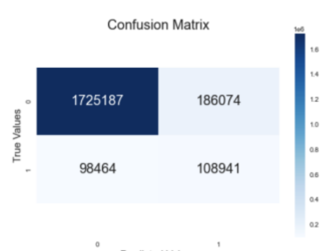
Model Evaluation and Performance

At last, when we used our model on the validation dataset to evaluate the performance of our model. Our made was able to correctly classify reorders and not reorders with 87% accuracy. The model has a low precision (0.37) and recall (0.53) for predicting reorders, which could be improved by further tuning the parameters with a broader array of values and taking measures to correct the data imbalance. Below are the classification report and confusion matrix on the test set.

```
F1 Score: 0.4336650611042554
              precision    recall  f1-score   support

         0.0       0.95      0.90      0.92   1911261
         1.0       0.37      0.53      0.43    207405

    accuracy                           0.87   2118666
   macro avg       0.66      0.71      0.68   2118666
weighted avg       0.89      0.87      0.88   2118666
```



Kaggle Performance

Name	Submitted	Wait time	Execution time	Score
xgbfinal_submission.csv	4 days ago	1 seconds	1 seconds	0.37337

Complete

Jump to your position on the leaderboard

Top score on the leaderboard is 0.409

Top-5 rows of prediction submitted on the Kaggle competition:

	order_id	products
0	2774568	17668 18599 21903 22035 39190 43961 47766
1	1528013	21903 38293
2	1376945	8309 13176 14947 27959 28465 33572 34658 35948...
3	1356845	7076 10863 11520 13176 14992 22959
4	2161313	196 10441 11266 12427 14715 27839 37710

Recommendations

Here are some recommendations for Instacart to transform the analytic results from observed customer purchasing behavior, identified customer segmentation, and enhanced product recommendation systems into action steps to increase business values and drive sales.

Negotiation power on partnership with grocery stores

To establish a sustainable partnership and strengthen the negotiation power on better product pricing, Instacart can leverage its data and findings on customer purchasing behaviour to help local grocery stores improve their inventory planning and sales forecasting. Most local grocery stores without loyalty programs will be more willing to partner up with Instacart, which puts Instacart in an advantageous position in terms of negotiation power.

Assuming Instacart currently adds $0.1 markup on all items sold on their website compared to local grocery stores. This is the only margin they make on products aside from delivery and loyalty fees. However, by providing data analytics on customer purchasing behavior to local grocery stores, in exchange, they can negotiate a 2% commission rate per order from partnered stores on top of the $0.1 markup per item.

Assumptions:

Markup	$0.1 per item
# Of Items per order	9
# Of Orders per day	280,000
Total markup per day without commission per day	**$0.1 x 9 x 280,000 = $252,000 per day**
Average order sales in $	$40
# Of Orders per day	280,000
2% commission	$40*2% = $0.8 per order
Total commission per day	**$0.8 x 280,000 = $224,000 per day**

As per calculation above, Instacart can benefit from additional 2% commission gain, which is $224,000 per day from the local stores based on the assumptions above.

Targeted offers based on cluster

From our analysis on customer segments, cluster 1 is our targeted consumer group because they buy fewer times (six times lower than the ideal cluster), not very often (# of days in order gap is 3 times more compared to the ideal cluster), and fewer items (Avg. basket size is almost half of Ideal cluster).

There is a two-step plan for customers in cluster 1 that Instacart can adapt to convert customers from cluster 1 to cluster 2. The first step is assuming we issue a 15% off discount to encourage customers to buy more than 3 times, shorten the gap from 17 days to 10 days between the days when they make a purchase, and bring up the number of sales items from 9 to 13 or more to convert customers in cluster 1 to cluster 0.

Discount	15%
Duration	12 months
Current Avg. No of orders	10
Expected Avg. No of orders	33
Current Avg. # of items per order (Avg. basket size)	9
Expected Avg. # items per order (Avg. basket size)	13
Current Avg. order gap	17
Expected Avg. order gap	10
Average item sales in $	$5
Total sales before issuing 15% discount to cluster 1	$5x9x10x21 (every 17 days represent 21 times in 12 months) = $9,450
Total sales after issuing 15% discount to cluster 1	$5x13x0.85x33x37 (every 10 days represent 37 times in 12 months)=$67,460.25
Total sales marginal gain on discounts issued to cluster 1	**$67,460.25- $9,450 = $58,010.25**

The second step is to convert customers in cluster 0 to cluster 2, which is the ideal cluster. If we issue a 10% discount to customers in cluster 0 to stimulate customers to buy more than 2 times, shorten the gap from 10 days to 6 days between the days, and bring up the number of sales items from 13 to 16 or more to convert customers in cluster 0 to cluster 2.

Discount	10%
Duration	12 months
Current Avg. No of orders	33
Expected Avg. No of orders	61
Current Avg. # of items per order (Avg. basket size)	13
Expected Avg. # items per order (Avg. basket size)	16
Current Avg. order gap	10
Expected Avg. order gap	6
Average item sales in $	$5
Total sales before issuing 10% discount to cluster 0	$5x13x33x37 (10 days represent 37 times in 12 months) = $79,365
Total sales after issuing 10% discount to cluster 0	$5x16x0.9x61x61(6 days represent 61 times in 12 months)=$267,912
Total sales marginal gain on discounts issued to cluster 0	**$267,912- $79,365= $188,547**

As per calculations above, Instacart is able to generate $58,010.25 + $188,547 = $246,557.25 more sales based on the assumptions for cluster 1 and 0 by issuing 15% and 10% coupon respectively to target customers.

Bundle discounts based on association rules

The association rules will stimulate cross and upsell based on antecedents and consequents coming from customer purchasing history.

Based on the Market Basket Analysis, when customers purchase organic avocados, they also buy organic bananas. This suggests that there is an upsell/cross-sell opportunity if we bundle these two products together. For example, we can introduce a veggie and fruit platter or a gift basket that has these two items together to local stores.

Assuming one bag of avocados can sell for $5 on its own, and a bag of organic bananas can sell for $6. Each bag of avocado and bananas can make 3 platters. Suppose we bundle and present them in a freshly cut platter. In that case, we can tag a bundled price of $8 per platter, and that's $8x3=$24 for one bag of organic avocados and bananas.

Assumptions:

One bag of avocados	$5
One bag of organic bananas	$6
Each bag of avocado and bananas	3 platters
Bundled platter with avocados and organic bananas	$8 per platter
Total sales of platters with one bag of avocado and bananas	$8x3=$24
Total marginal gain on bundled platters	**$24-$5-$6=$13 per platter**
Price per basket	$19
Total marginal gain on bundled basket	**$19-$5-$6=$8 per basket**

From calculations above, Instacart can earn additional $13 per platter and $8 per basket by bundling these two products together.

Comprehensive recommendation system

The recommendation system built in our model can effectively detect hidden sales opportunities with the existing customers' purchases. When a repeated customer who purchases organic strawberry will review the product recommending system and pick up organic banana before check out will certainly drive up sales.

Assumptions:

Additional sales generated by recommendations	$5 per item on average
Additional items picked from the recommendations	1 item
# Of Orders per day	280,000
Total margin gain on recommendation	**$5x280,000 = $1,400,000**

From the margin gain calculations above, by improving the recommendation system, Instacart can earn additional $1,400,000 per day based on the list of assumptions.

With all the recommendations above, Instacart can improve sales through high negotiation power, targeted marketing, bundled discounts, and comprehensive recommendation system to maintain a robust and sustainable growth for future.

Appendix