

**Instacart: The Era of Fast and Convenient Retail?**

**Executive Summary**

May 7, 2023

Phillip Arash Mirian

BIA-678 Big data technologies

Stevens Institute of Technology – Spring 2023

Table of Contents

[1 Propose 3 decision problems that Instacart needs to consider to increase profitability.  Propose a Big Data technology to use for processing a very large amount of data(Spark,Hadoop)……………...…………….………………………………….………..……](#_Toc74830172)2

2 Popularity of Grocery Items: Please analyze the popularity of items. Complete an explanatory analysis of data using Tableau, Python, R. Craft and executive summary and generate Tableau twbx files………………………………………………………………………………4

3 Customer Segmentation: Perform clustering analysis for the customers. (Which groups of customers are similar?)………………….…………………………………………………………8

4 Product Recommendation: Which sets or products should be recommended to shoppers? (Hint: Association rules mining on Instacart products.)………………………….12

5 Address all your recommendations for Instacart in these areas: Data governance applications, Data science, and Artificial Intelligence. Summarize all of your findings…………………………………………………………………………………………………………………..17

Table of Figures

Figure 2.1 Frequency of Products Ordered and Reordered……………………………………………….5

Figure 2.2 Frequency of Products Ordered and Reordered by Department and Aisle..……..6

Figure 2.3 Proportion of Items at Checkout………………………………..……………………………….……7

Figure 2.4 Timing of Items Ordered…………………………………….…………………………………….……..8

Figure 3.1 Broad Customer Segmentation………………………………………………………………………..9 Figure 3.2 Principal Component Analysis (PCA)……………………………………….……………………..10 Figure 3.3 Feature Importance from PCA………………………………………………….…………………….10 Figure 3.4 Scree Plot for Clustering Analysis……………………………………………………………………11 Figure 3.5 Silhouette Scoring for Clustering Analysis………………………………………………………11 Figure 3.6 Silhouette Plots for Three PCA Kmeans Cluster Analysis…………………..…………...11 Figure 3.7 Silhouette Plots for Two PCA KMeans Cluster Analysis…………………………………..12 Figure 3.8 Snake Plot for Feature Segmentation Interpretation………………………………………12 Figure 5.1 Instacart Algorithmic Flow Chart………………………………………………………………….…17

References…………………………………………………………………………………………19

# Instacart: The era of fast and convenient retail?

## 1 Propose 3 decision problems that Instacart needs to consider to increase profitability.  Propose a Big Data technology to use for processing a very large amount of data (Spark, Hadoop).

Instacart is a same-day retail delivery service in the United States and Canada. This service allows customers to shop online for groceries and household goods and have them delivered to their doorstep. This company has exploded in the past couple of years dominating the online retail space competing with companies such as Shipt, Walmart, FreshDirect, Amazon Fresh and DoorDash to name a few. The company partners with local grocery stores and retailers offering a variety of products, and hires independent contractors called “shoppers” to pick up and deliver the orders. Customers will go to the website or app and place an order to be delivered within specified time windows including 1-hour, 2-hours or scheduled deliveries.

As we saw, Instacart competes with very well-known, innovative companies. In order stay competitive, Instacart will need to ensure they are able to differentiate themselves. If they are successfully able to do this, they will improve their profit margins. Some of the strategies Instacart has implemented to stay competitive include delivery fees, services fees, markup on products, advertising and promotions, subscription services, and selling data insights. Furthermore, they look to reduce non-fixed costs by optimizing their delivery systems to improve efficiency and speed thus increasing profits.

Let us look more closely at the at some of the decision Instacart needs to make to maintain these revenue streams. The first to consider are the four relationships of the marketplace model. That is, the relationships between the customer, shopper, product, and stores. An important component to this is how to balance supply and demand. If there are many customers who want certain products, there needs to be the retailers the customer prefers and sufficient shoppers to fulfill the demand. If demand cannot be met, customers will be turned away. On the other hand, if there is insufficient demand there will be idle shoppers. Thus, if this is not balanced there will be loss of profit.

When potential customers visit the website or app there are three possibilities that need to be observed when considering demand. The customer made a list and checked out. The customer made a list but did not checkout, possibly because the delivery times were not available. A potential customer is simply browsing without intent to purchase. All of these seem to indicate there will be a potential purchase, however patterns of total actual demand need to be interpreted. That is, actual purchases, and the purchasers that walked away. When looking at these combined, then a better picture is understood on how to account for demand.

If demand is well understood, the next reasonable decision Instacart needs to understand is how to fulfill their 1-hour delivery time slots. If this is not fulfilled, customer satisfaction goes down and they lose customer retention. Here, Instacart must predict how long it will take any individual shopper to perform their designated tasks. Things to consider in the time prediction are repeated trips to the same store, entering the store to shop, and leaving the car to deliver the groceries. Furthermore, there are shoppers who shop in stores only, those that enter the store and deliver, and those that deliver only. Yet another layer of complication to this are the multiple carts a delivery driver might be carrying. In other words, the optimal combination of routes will also need to be considered. Lastly, when trying to send shoppers to stores to purchase the items, they need to be sent to stores with the highest probability of having all the items. Here, there may be a tradeoff of distance for improved customer satisfaction as long as the delivery time can be met.

So far it was discussed how to ensure customer satisfaction. This makes sure customers are retained and will pay all those profiting fees discussed above. Another major revenue stream is from retailers. It might be wise for Instacart to make decisions on product recommendations featuring personalized products for individual customers. Meanwhile, recommending products showcasing new things to customers to both improve current recommendations as well as offering things they might not have known existed. Furthermore, having a good recommendation system, this may also help with fulfilling one of the delivery issues. Here if there are alternative products that can be well recommended to the customer, perhaps this will prevent a shopper having to travel as far for a store increasing efficiency. So, in this way customer satisfaction can be improved meanwhile featuring products form stores drawing customers to them in a potential meaningful way for the customer.

The three decisions stated above have complex algorithms with tremendous amounts of computation that are continuously updating. For example, the perpetual time series forecasting to predict demands, the gradient boosting decision trees to estimate fulfillment times needing to account for changing outliers, the greedy heuristics to perpetually map shoppers to stores, or the various recommender systems for product replacements and homepage featured products. This level of computation requires big data technologies including storing, governing, data preparation, feature extractions, machine learning training and then applying to improve the app.

A good big data technology is Amazon Web Services (AWS). There are several services that are offered that can be used to for the specific purposes of Instacart. Firstly, it might be good to consider Amazon Kinesis. This allows Instacart to ingest and process vast, real-time data from multiple sources including websites, apps and tangent helpful data including social media feeds and logs or blogs. The data gets ingested into Amazon S3 that is scalable and secure. It has low latency for accessing the data giving it the ability to be backed up for disaster recovery, content distribution, and data archiving. Also, S3 contains appropriate data governance enabling appropriate access controls, compliance, and lifecycle policies. Furthermore, S3 supports Amazon Athena. This can query the data directly from S3 entirely off premises whereby Amazon EMR enables Instacart to process their data using open-source frameworks such as Apache Spark. Spark is built on top of Hadoop’s distributed computing framework. The computing is designed to perform data processing and machine learning tasks on large datasets across multiple computers or nodes in parallel and supports multiple programming languages. This is a highly scalable and fault tolerant. It also supports Apache Hive to write SQL queries for analysts and data scientists to analyze large data sets more easily. Lastly, by using these AWS services Instacart can continue to customize the services they need depending on how their data infrastructure might change over time, and they can better customize what they want to pay for.

# Instacart: The era of fast and convenient retail?

## 2 Popularity of Grocery Items: Please analyze the popularity of items. Complete an explanatory analysis of data using  Tableau, Python, R. Craft and executive summary and generate Tableau twbx files.

Given the business decisions for Instacart mentioned in the previous section, we are now looking to understand its existing data. In this specific case it will allow us to better understand underlying patterns and insights, identify if there are biases or outliers, and appreciate if the business decisions are on the right track. We first look at Figure 2.1. This shows us the number of users and the number of orders made. We can see there are roughly 1 million people who have made 30 million orders. The most abundant items purchased are bananas, followed by organic bananas and then other fruits and vegetables primarily of the organic type. Notably, the top 13 items ordered are (in not exactly the same order) the same re-ordered items.

Delving deeper into the products we continue to see from figure 2.2, the highest proportion of sales come from produce, dairy and eggs, snacks, beverages, frozen then pantry making over 75% of the sales. Again, we see all the top departments of sales are the very same as those reordered. Breaking that down by aisle, these follow similar patterns demonstrating produce, then dairy and eggs, followed by subtle differences between pantry, frozen, beverages and bakery. These differences are not surprising given the small difference in percentages of sales between these departments. Again, we see reorders by aisle are very similar to orders.

Something very interesting to note is the number of products that exist within the aisles. We notice, still in Figure 2.2, the largest number of items that exist are among the snacks, frozen, personal care, and dairy products. More specifically, some of these items include candy chocolate, ice cream, vitamin supplements, pretzel chips and tea to name a few. This is a little

***Chart, bar chart, waterfall chart

Description automatically generated***

*Figure 2.1*

difficult to interpret because we think of these as being somewhat popular items, and in fact we do see from the middle line graph in Figure 2.3that there seem to be an inconsistent increase in the ratio of items reordered after being placed into the shopping cart after about 50 items. This may indicate that people are last-second placing items into the cart. These seem to be appropriate items that could categorize them as last-second items placed. However, given the amount of space these could take up, it may be wise to attend to this fact and make more space for items that are specifically high in demand. This may help avoid running out of items and address the goal of ensuring 1-hour deliveries.

Next, I want to bring your attention to purchasing patterns. The top bar graph in Figure 2.3 shows us approximately 75% of orders have 10 or less items. Further, the earlier the item is placed into the basket, the higher the chance is that it will be ordered as we stated above. This begins to tell us people may not be spending much time on the app or website and that they are likely purchasing specific items. This is very possible because we have already seen there are specific items that are clearly most purchased. In addition, in the bottom of Figure 2.3 we see those items placed into the cart first have a much higher proportion of orders that are reordered with the same trend of percentage of items in the cart at checkout just mentioned. That is, there remain to be small cart sizes upon reorders that are of the same products and if more items are added, there is less of a chance they will remain in a customer’s cart at checkout. This is something important to understand when it comes to marketing or recommending products to increase sales and improve profits. Instacart needs to identify what other products individual customers may be interested in and they need to do it in a way that

*Chart, pie chart

Description automatically generatedFigure 2.2*

will quickly catch the customers’ attention. Furthermore, we’ve seen there are many more products that exist that are not popular among purchases. If they are going to remain as part of the stock of items taking up retail space, then these could be the products recommended to the right people for up sale.

The last thing we can address from the data is when the orders are taking place. In Figure 2.4 we notice again that there is a larger proportion of sales that are reorders seen in the top left pie chart. The proportion of higher reorders are seen every hour of the day and day of the month. This makes sense, the longer the business is around, the more difficult it is to get new customers. Instacart would likely try to make efforts to realize an increase in the number of reorders to increase profits. The top right bar graph shows us that both orders and new orders follow the same ordering pattern by hour of the day. Orders are placed predominantly from 9am to 4pm. Not demonstrated here, but it was seen that this pattern persists for every day of the week. This pattern is a bit peculiar. Take pause on this for a moment and we can also see most orders and reorders take place during Sunday, Monday and Saturday respectively.

Chart, histogram

Description automatically generated

*Figure 2.3*

We go on to observe in the bottom bar graph of Figure 2.4 the most common orders occur every 7 and 30 days. We also see there are quite a high proportion of orders that occur 1 to 7 days and small increases every 14 and 21 days. This tells us Instacart has customers who perform orders mostly weekly and monthly, daily or every couple of days, as well as every 2 and 3 weeks.

In an effort to draw deeper insights into these patterns we may conclude that one of the reasons Instacart sees the most orders every 1 to 7 days is because approximately 45% of their customers are ordering from the produce, and dairy and egg departments. Further, the highest percentage of orders have 10 or less items. These are items that may be consumed regularly and have a relatively short shelf life. We can think of things like bananas or eggs as examples of this. Further, we see there is a higher likelihood of smaller cart sizes that would indicate a customer would likely need to make more frequent recurring orders. On the other hand, we see weekly, bi-weekly, tri-weekly and monthly shoppers who may be ordering other common products from departments such as pantry, beverages, frozen or bakery. These are items regularly consumed, but with a longer shelf life. Additionally, we see a kind of parabolic pattern to items remaining in the cart after about 50 items but with irregularity. This might mean there are specific items people know what they want and immediately place these into the basket. However, this population who make less frequent orders might need to plan what to eat for longer periods of time. It would be reasonable for there to be higher variability in what they decide to place in their cart. This could be the population to be more aggressive with advertisements and recommendations. Lastly, it is important to recognize when orders are taking place. As mentioned, we see they take place in the middle of the day mostly on weekends and Mondays. If most people are making quick, small basket purchases they can likely do this easily in the afternoon, maybe on lunch break to ensure deliveries will take place by the time they get home or to make food on weekend evenings. Here it is important to Chart

Description automatically generatedrecognize that there will need to be more shoppers available at these times.

*Figure 2.4*

Given this analysis it appears the business decisions may be on track. We now better understand purchasing behaviors to discern how much and when Instacart needs to have specific products. We perhaps understand the types of customers who would benefit more from advertisement and recommendations as they seem to have variability in their shopping behaviors. Lastly, we have a good idea when purchases are taking place to know better how much and when to mobilize an optimal number of shoppers.

# Instacart: The era of fast and convenient retail?

## 3 Customer Segmentation: Perform clustering analysis for the customers. (Which groups of customers are similar?)

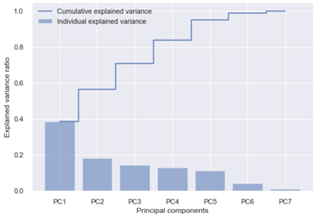
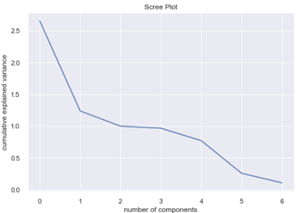
Here we want to understand who our customers are. Perhaps the best way to do this is through combining relevant features that might help explain customer behavior. These engineered features are shown in Figure 3.1. Along with this we see a rough idea of how each feature could be clustered. That is, primarily by quantity. In other words, customers can be segmented by a combination of feature, orders, or time customer. We could create high, medium and low segments and make recommendations for people within these segments.

*Figure 3.1*

However, this is a bit simplistic, nor does it tell us which features are the most important to offer the best ways to understand our customers.

*Chart

Description automatically generated*Alternatively, algorithmic techniques have been preferentially performed. Notably, the data to perform this was sampled from the total dataset due to computational limitations. Once the features were engineered, they were scaled and normalized for principal component analysis (PCA) and KMeans clustering purposes. Upon performing PCA, it appears that there are three features that explain most of the data. We see this when looking at Figure 3.2, the box plot shows us that it takes 4 principal components to explain approximately 82% of the data. However, the scree plot in Figure 3.2 tells us that the eigenvalue drops below 1 at 3 principal components. In this case it seems reasonable that we will have 3 principal components that explain the majority of our data. At this point we look to see the contribution each feature has to the PCA components. Those features having the highest influence on principal components were used for clustering analysis.



*Figure 3.2*

What is interesting is that businesses will often segment their customers based on features of recency, frequency and monetary value. If we look at which features contributed most to the principal components they are ‘total reorders’, ‘days since previous order’ and ‘total orders’ respectively, see Figure 3.3. It is not a far reach to understand these might otherwise be called *frequency, recency,* and *monetary value* respectively. Perhaps this is why businesses will often use this approach.

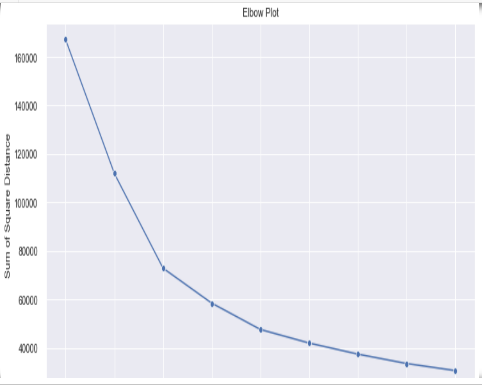
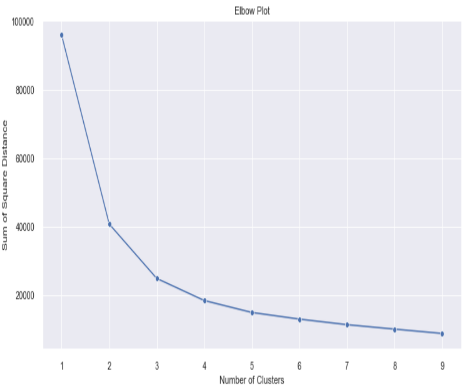
Table

Description automatically generated

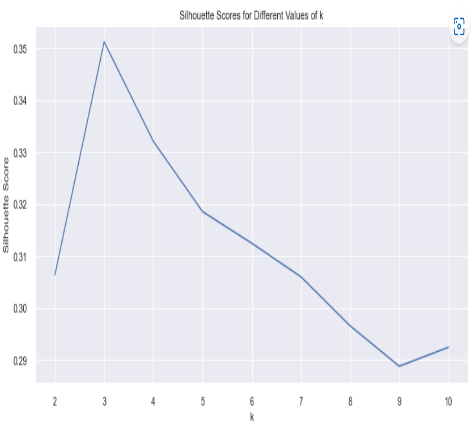
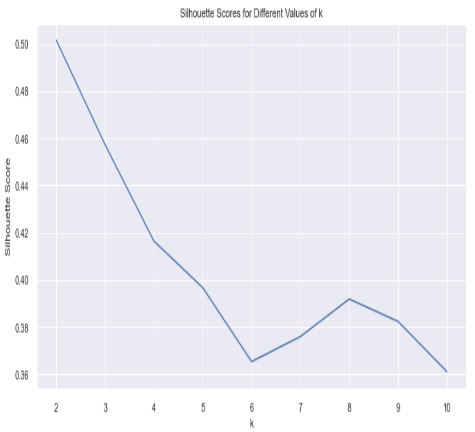
*Figure 3.3*

At this point we want to now consider how we can cluster our customers, that is, find out which groups have similarity in how often they shop, when was the last time they shopped and how much money they spend. To do this KMeans clustering was performed. To gain useful insights into this we need to determine the number of clusters that will capture the underlying pattern. This was first approached by using the elbow method. If we look at Figure 3.3 we see maybe there is an inflection point at 2 clusters when using 2 principal components and maybe another at 3 clusters when using 3 principal components. Notably 2, 3 and 4 principal components were assessed for clustering to ensure the patterns are optimally understood. Further, when we look at the silhouette score in Figure 3.4 we see again using 2 principal components that the data points better match their cluster when placed into 2 and 3 clusters when compared to 3 principal components seen in Figure 3.5.

In this case study, three principal components are used. The reason for this is because we need to identify enough characteristics on customers to identify who are the most valuable, where the needs are, how to properly market and offer promotions.



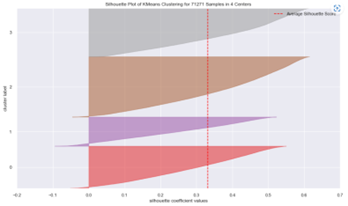
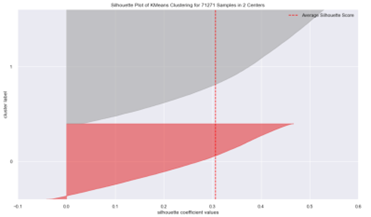
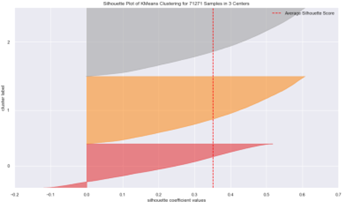
*Figure 3.4*



Additionally, the silhouette scores are relatively negligibly different so we can have good confidence the data points will continues to cluster well for a good segmentation. We can see this is in fact true when looking at Figure 3.6 and 3.7. Figure 3.6 shows us with 3 PCAs and 3 clusters there is the most homogenous clusters, though clearly the

*Figure 3.5*

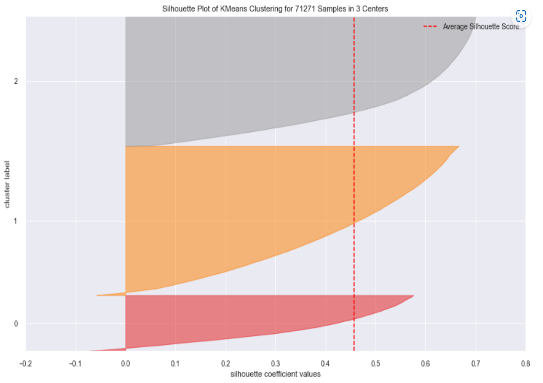
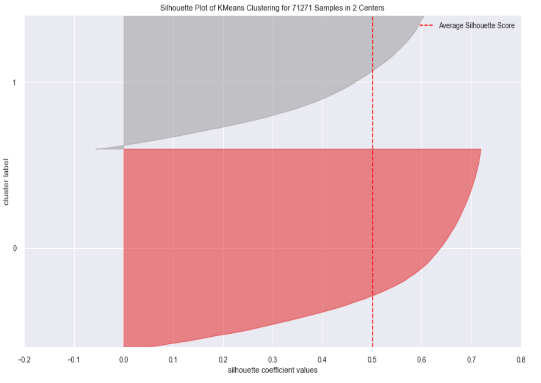
red cluster has fewer data points with some misclassification identified by the left bottom tail.



*Figure 3.6*

Further, it has the highest silhouette score of those with 3 PCAs. This contrasts to Figure 3.7 where despite the 2 PCAs having higher clustered silhouette scores, the clusters are not as homogenous and there are more clusters with misclassifications.

What we have now are three main features that have been determined to represent the majority of the data. Again, these are ‘total reorders’, ‘days since previous order’ and ‘total orders’. It has been determined that these features cluster best into 3 clusters. To best understand how to interpret this we look to Figure 3.8. This clearly demonstrates there are 3 types of customers. There are customers who have shopped recently and have spent a moderate amount of money. There are those who have not shopped so recently and have



*Figure 3.7*

spent a large amount of money. Lastly, there are those who have not shopped so recently and spent a small amount of money. Meanwhile, it appears all segments of customers have a nearly identical number of times they reorder. To note, the total number of orders gives insight into the amount of money spent as even if a customer has one item in their basket, if they have more total orders, they have likely spent more money cumulatively. This is especially since all customers have a nearly identical frequency of shopping. Here we have successfully segmented the customers and identified what underlying similarities they have within and among each other.

Chart, line chart

Description automatically generated

*Figure 3.8*

# Instacart: The era of fast and convenient retail?

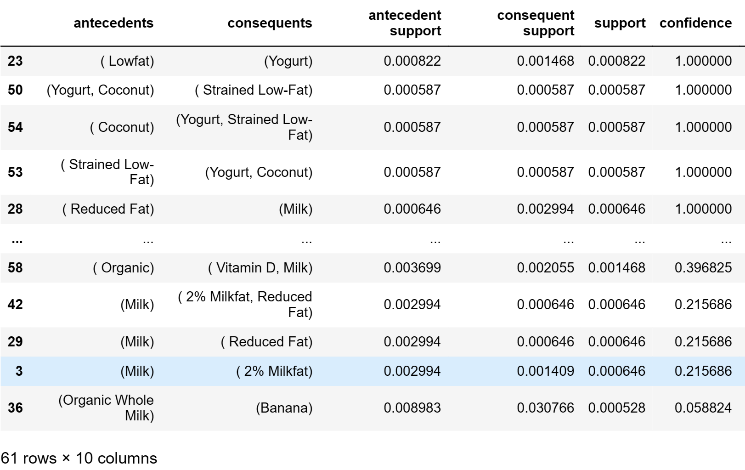
## 4 Product recommendation: which sets of products should be recommended to shoppers? (Hint: association rules mining on instacart products.)

Now that we understand who the customers are and a bit about their behaviors, let’s investigate what are some of their purchasing behaviors. To do this we are going to be performing a market basket analysis on our clustered customers. First the data is grouped by user id, order id and cluster with the products aggregated upon those groups to ultimately create shopping baskets. It is from these baskets that we determine how often one item is purchased with another, called an itemset. To determine frequent itemsets the apriori algorithm is used relying on its fundamental property of downward closure. It is important to note when invoking the association rule of the apriori algorithm, the minimum support was necessarily set to a one ten-thousandths to ensure an adequate number of itemsets were selected. Thus, the support indicates the fraction of the rows containing the itemset. Figure 4.1 provides an example of itemsets and their support for customers in cluster 0. It is important to notice there is an antecedent (the item used to make predictions about the other itemset) support, a consequent (the item predicted based on the occurrence of the antecedent) support, and a total support. The latter being the fraction of time these itemsets occur together.

*Figure 4.1*

Next, we want to be sure we know the probability of these items are occurring with one anotherTable

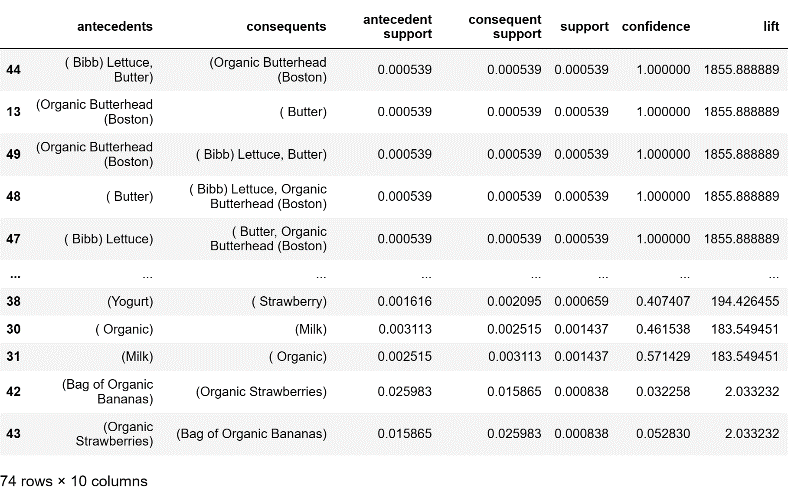
Description automatically generated. This is called confidence. In this case example, the confidence was set to 5%. This was set given the size of the dataset used and number of items clustered. We can see in Figure 4.2 by setting confidence to 5% this only offers 61 itemsets. It is important to appreciate that



*Figure 4.2*

the confidence obtained by one antecedent consequent pair will not necessarily have the same probability of the reverse.

The last item to pay attention to is the lift. This is a ratio of the confidence to consequent support. This provides the strength of association of items in the frequent itemsets; or if they would occur more than expected if they were statistically independent. If this ratio is greater than one, this in fact tells us there is a positive correlation and the itemset do occur more than expected. Meanwhile, this ratio less than one indicates a negative correlation, and a value of exactly one indicates there is no association between itemsets. We can see an example of this in Figure 4.3. for customers in segment 2. What is nice to see is that the lift value for



*Figure 4.3*

many itemsets are greater than one. This is important because we needed to set our confidence to a low threshold. Additionally, the support is also set to a very low threshold. This is to say, there are not many items purchased together, and those that are observed together generally have a low probability of being purchased together. By having a lift greater than one tells us there is at least a positive correlation between these items and they may in fact be purchased together.

Table

Description automatically generated Let us take a look at a specific example case. If we set the apriori rules as mentioned above to customers in cluster 0, we can see butter occurs with two different kinds of lettuce with 80% confidence and a lift greater than one, see Figure 4.4. A better example can be

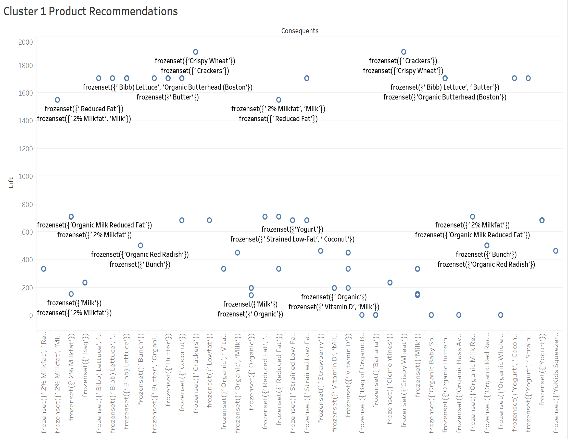
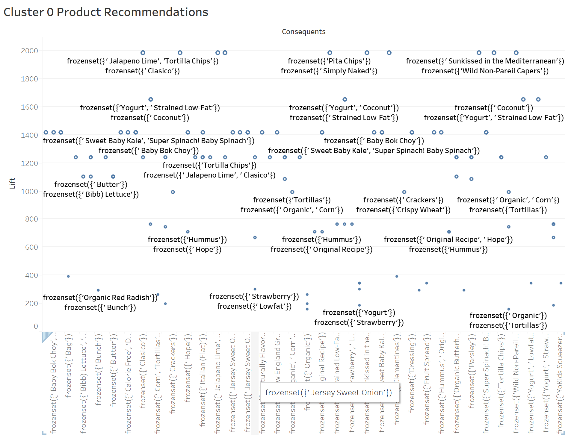
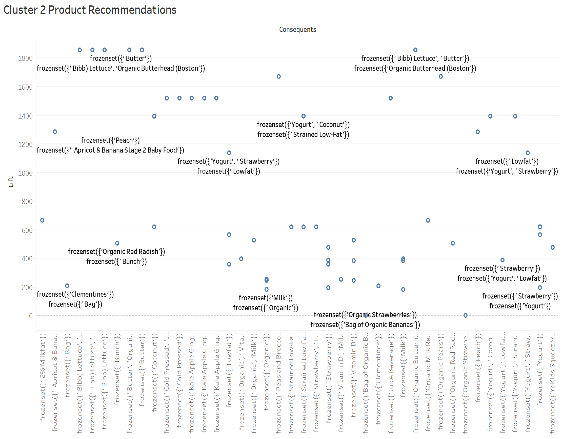
*Figure 4.4*

made by looking at Figure 4.5. We can see there are six recommendations to offer, all with a lift greater than one. If we pay close attention, it appears the data needs further preprocessing to better understand the items. Furthermore, we can look to figure 4.6 and get an idea of how the totality of these recommendations look. For a better appreciation, please observe Appendix B attached separately for complete customer recommendations.

Chart, bar chart

Description automatically generated

*Figure 4.5*



*Figure 4.6*

If we recall, we had three initial business problems. These include issues with customers adding to the basket, but not checking out, making it difficult to understand demand and how to mobilize resources. We had considered the necessity of cross-selling and upselling. Then we needed to make sure we could meet our delivery times. By understanding the customers as we have and their purchasing behaviors, we have created another ability to understand how to address these business goals. We have developed an understanding of customers and those with similar behaviors to recommend products or understand what products may be in higher demand or trending. This can lead to upselling and increased opportunities for product advertisement revenue streams. These behavioral patterns can help the business understand if there is overstock or items commonly purchased together to offer product bundles or promotions. We might be able to better understand why customer begin to place objects in their basket, but do not check out if these patterns help increase checkout behavior. Specifically, this might be understood through promotional strategies or offering new products based on trending behaviors thus having products they might have been looking for in the first place. These patterns will be lucrative in improving supply chain efficiencies. Lastly, the store might be organized based on these behaviors to improve shopper efficiency, helping meet the delivery times. Overall, the analyses, mining strategies and algorithms in this case study offers us insights into the needs and preferences of the customers to personalize their offerings in the efforts to provide an overall better customer experience.

# Instacart: The era of fast and convenient retail?

## 5 Address all of your recommendations for instacart in these areas: Data governance applications, data science, and artificial intelligence summarize all of your findings.

Instacart has grown dramatically over the past few years. As the pandemic has demonstrated, there may be shifting tides that will cause a business such as this the need to be able to have directed business goals. To do this, Instacart will need to ensure all their customers, shoppers, and brands’ data is properly managed. That is, Instacart will need to ensure they have good data strategy. In so doing they will be able to effectively use data to inform upon the current goals and help drive future goals.

Fundamentally a good data strategy relies on a solid data architecture. The foundation of such a strategy requires excellent data governance. Having a good data governance policy will do several things including increase the value of the company’s data, help solve issues with data, ensure correct regulation and compliance procedures, promote transparency, and help increase overall revenue to mention a few. It would be in Instacart’s interest then to have data stewards.

Of the many responsibilities of a data steward, a conglomerate of such includes upholding data management policies and maintaining the quality of data. That is, they would need to ensure the data is initially quarantined with proper hygiene rules applied, ensure that attributes of the data are defined, appropriate data access is maintained meanwhile making sure the data is not siloed such that it is accessible to those who need it quickly and easily, and that the data lifecycle is maintained. These will ensure appropriate analytics are being performed, offering proper insights meanwhile adequately protecting the data. Some platform tools Instacart could implement include Collibra, Informatica, Talen, IBM Inforshpere and Alation. This then offers Instacart a combination of options to ensure proper data governance. In other words, they might want to have data stewardship, and implement one such technology to their data architecture.

If these governance policies are well established, then there will be a great level of confidence from the insights that are drawn from various algorithmic techniques. However, Instacart requires very specialized methods to be applied to areas concerning shopper fulfillment times, shopper work availability, performance of new products launches, evaluation of product change to ranking, relevance and pricing of ads10. This requires a coordinated effort of data scientists with marketing, software engineers, and product manager to name a few. Data scientists also need to be working with key stakeholders to better understand the needs across the business to be able to deliver data-driven insights and recommendations. Examples of the aforementioned can be seen in Figure 5.1.

As mentioned above, Instacart requires specialized algorithmic techniques as their business requires unique business solutions. For example, taking the issues of fulfilling

Diagram

Description automatically generated

*Figure 5.1*

shoppers’ delivery times, they cannot simply use Google maps. This does not account for which store is closest with the highest probability of containing all the customer’s items. Further, the shopper will be picking up batches of orders that may contain frozen and fresh hot deli items. All the meanwhile needing to try and communicate with the customer if the customer ordered a low probability item, instead of a recommended item, that then became out of stock. Finally, and still only part of the story, delivery to the customer through a routing algorithm requires route efficiency keeping in mind some routes require the shopper to ambulate.11 All this within an hour delivery time precision to hundreds to thousands of customers at any moment throughout the day and week.

Again, there are four sides of the Instacart business. It would be wise for Instacart to have their data scientists organized based on these sides of the marketplace. That is the consumer, retailer, shopper, and advertiser.12 In this way Instacart can have a central function as a team of data scientists, meanwhile maintaining focused unique objectives for each business segment to optimally improve the quality of decision making.

It is within the scope of the data scientist who will are responsible for implementing the concepts of artificial intelligence. They implement machine learning algorithms that use their vast data to help Instacart improve business operation, drive revenue and ensure customer satisfaction. Some of the most relevant models used by Instacart are some iteration of product recommendations, predictive modeling including time series forecasting, fraud detection to ensure the credit card being used is not stolen, and natural language processing.

These components of artificial intelligence have allowed remarkable growth and opportunity for Instacart to become the growing empire it is today. However, as these technologies grow Instacart could capitalize on the strength of their data scientists to implement higher levels of prediction and automation. For example, since the recent advent of Chat GPT, Instacart has been an early adopter and realized the value to incorporate it into its business modeling. The way they adapted Chat GPT is to use it as a search engine that can respond to food-related questions. For example, this can be used to ask for recipe ideas, ingredients, or various customer specific healthy meal options.

The reason we can truly call this artificial intelligence is because it interacts with humans naturally, it augments decision making and it can be used at scale. In fact, Chat GPT is an open AI system able to seamlessly integrate with Instacart’s existing software through an application programming interface (API). This is taking natural language processing to an entirely different level. To expand on the simple integration, what they could look to do is have multiple API calls such that those ingredients the customer searched can automatically be placed into the shopping cart. Certain items can be promotional items that are automatically added to the shopping cart. This will be one way to help boost recommendation abilities and advertisement revenues. Further, Instacart now has a way to break into multiple markets should they choose. For example, Instacart can become more socially conscious and be a part of promoting healthy eating. This may provide opportunities to partner with other companies and increase advertising revenue streams.

It is important for Instacart to realize that this vastly increases the personalization of information of each of their users. Customers may even use this in a way beyond simply asking for recipes or meal suggestions. As can be seen with the chatbot example of Bus Uncle in Singapore, people may decide to have full conversations that become something personally identifying.12 What this implies is that Instacart will be sharing this information with outside sources such as Chat GPT that requires an added layer of privacy protection that in some ways may lie outside the hands of Instacart. Additionally, by integrating Chat GPT there may be unforeseen liabilities such as the possibility of recommending a product or recipe through some bias as a result of the user learning patterns. This may lead to inaccurate or deceiving results, and possibly be injurious as we see the increasing impact of food on human health.

Instacart is making an increasingly profound name for itself having increased revenue year by year with impact to begin to compete with Walmart and Amazon’s food market. They’re growing rapidly with newfound intelligence that do not have robust rules and ethical policies. It would be wise for Instacart to ensure trust by their users to stay in business. To do this they may consider implementing the 5 pillars of AI ethics. Specifically, this AI platform is fair to all customers. It has the ability to be explainable in its decision-making capacity. There is robustness in its security and data is used equitably. There is a transparency that this is in fact an AI making decisions, and not a human user. Lastly, all the customer’s data remains private, it remains in the hands of Instacart and is not inappropriately used. There are many amazing technological, mathematical, algorithmic, heuristic, and intelligent advances this company has been able to achieve. This offers them leverage to design some lofty business goals making them incredibly competitive in the marketplace. Thus, they are in a position to continue to help advance technologies. In this era of the rapid rise of AI without equally robust advances in ethics and privacy policies, Instacart needs to be one such entity to offer structure in these areas. The more people and communities can trust these applications, the more benefit people can reap, meanwhile offering continued growth for Instacart and its reach.

# Instacart: The era of fast and convenient retail?

## 6 References:

*Advertising on Instacart 101: Where do my ads show? - instacart corporate*. (n.d.). Retrieved March 27, 2023, from https://www.instacart.com/company/instacart-ads/wheredomyadsshow/

Grigoriy, G. (2021, March 21). *Data Science at Instacart: Making on-demand profitable*. Domino Data Lab. Retrieved March 26, 2023, from https://www.dominodatalab.com/blog/data-science-instacart

Hendelmann, V. (2023, January 13). *Instacart competitors: Ranking its 12 biggest rivals*. productmint. Retrieved March 26, 2023, from https://productmint.com/instacart-competitors/

Lu, D. (2022, September 10). *Evolving Doordash's substitution recommendations algorithm*. DoorDash Engineering Blog. Retrieved March 26, 2023, from https://doordash.engineering/2022/09/08/evolving-doordashs-substitution-recommendations-algorithm/

Sharda, R., Delen, D., & Turban, E. (2021). *Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*. Pearson.

Ye, A. (2020, March 30). *How Instacart uses data science to tackle complex business problems*. Medium. Retrieved March 26, 2023, from https://towardsdatascience.com/how-instacart-uses-data-science-to-tackle-complex-business-problems-774a826b6ed5

Zhang, Y. (2022, October 5). *Homepage recommendation with exploitation and exploration*. DoorDash Engineering Blog. Retrieved March 26, 2023, from https://doordash.engineering/2022/10/05/homepage-recommendation-with-exploitation-and-exploration/

Mixson, E. (2021, March 16). *Instacart: Delivering Incredible Customer Experiences with Advanced Analytics and Machine Learning*. AI, Data & Analytics Network. https://www.aidataanalytics.network/data-science-ai/articles/instacart-advanced-analytics-and machine-learning

‌Masters, K. (n.d.). *Brands Can Finally Control Their Product Data On Instacart*. Forbes. Retrieved May 4, 2023, from <https://www.forbes.com/sites/kirimasters/2021/05/06/brands-can-finally-control-their-product-data-on-instacart/?sh=698557e12d5d>(2022, December 21). *Getting to Know the Data Scientists* (Instacart, Ed.) [Review of *Getting to Know the Data Scientists*]. Instacart; Instacart. https://www.instacart.com/company/how-its-made/getting-to-know-the-data-scientists-at-instacart/?irgwc=1&utm\_source=instacart\_impact&utm\_medium=affiliate&utm\_term=partnertype-mediapartner\_adtype-TEXT\_LINK\_adname-Use%20your%20EBT%20SNAP%20benefits%20on%20select%20Instacart%20orders.%20Stay%20home%20and%20skip%20the%20lines%20with%20grocery%20delivery%20today.%20Terms%20apply.&utm\_campaign=partner-Bing%20Rebates%20by%20Microsoft&utm\_content=clickid-SM6WqHVtkxyIR9czljRB9S6SUkAVqlwhwWy1zs0\_campaignid-7412\_partnerid-2003851\_adid-1088185\_sharedid-EdgeBingFlow

‌

*The Algorithms That Make Instacart Roll*. (2021, March 2). IEEE Spectrum. https://spectrum.ieee.org/the-algorithms-that-make-instacart-roll

‌ (2022, July 25). *Building a Data-Driven company with Ananhita Tafvizi, Instacart’s Vice President and Head of Data Science* [Review of *Building a Data-Driven company with Ananhita Tafvizi, Instacart’s Vice President and Head of Data Science*]. Instacart; Instacart. <https://www.instacart.com/company/how-its-made/meet-anahita-tafvizi/?irgwc=1&utm_source=instacart_impact&utm_medium=affiliate&utm_term=partnertype-mediapartner_adtype-TEXT_LINK_adname-Use%20your%20EBT%20SNAP%20benefits%20on%20select%20Instacart%20orders.%20Stay%20home%20and%20skip%20the%20lines%20with%20grocery%20delivery%20today.%20Terms%20apply.&utm_campaign=partner-Bing%20Rebates%20by%20Microsoft&utm_content=clickid-SM6WqHVtkxyIR9czljRB9S6SUkAVqnx1wWy1zs0_campaignid-7412_partnerid-2003851_adid-1088185_sharedid-EdgeBingFlow>Svensson, L. (2023, March 23). *Introducing the Instacart Plugin for ChatGPT*. Instacart Corporate. https://www.instacart.com/company/updates/instacart-chatgpt/

‌

Loten, A. (n.d.). *Instacart Joins ChatGPT Frenzy, Adding Chatbot To Grocery Shopping App*. WSJ. <https://www.wsj.com/articles/instacart-joins-chatgpt-frenzy-adding-chatbot-to-grocery-shopping-app-bc8a2d3c>

‌ Nan, Wong, et al. “Bus Uncle Chatbot - Creating a Successful Digital Business (A).” *Singapore Management University*, 21 May 2021, pp. 1–6. *Harvard Business Review*.