



BC2407 ANALYTICS II: ADV PRED TECH

Semester 2 AY 2019/2020

COVER PAGE

Matriculation Name:

Lee Mun Yee (U1810810A)

Ong Qing Quan (U1810077E)

Tan Jing Yu, Denis (U1810101D)

Teo Kiat Wee, Jared (U1810545C)

Jovan Teo Xi Ze (U1810708L)

Tutor Name: Teoh Teik Toe

Seminar Group: 2

Team Number: 4

Assignment: Analytics 2 Group Report

Date: 1 April 2020

Contents

Executive Summary	1
Instacart Overview	2
Instacart Business Model	2
Online Grocery Sector Opportunities	3
Business Problem.....	5
Proposed Approach & Desired Outcomes	7
Analytics Solution.....	12
Conclusion	19
Appendices.....	20
Appendix A: Walmart's Balance Sheet for YA 2019, 2018 and 2017	20
Appendix B: Confidence Level for each pair of items in Instacart	20
Appendix C: Product Recommendations on product page	21
Appendix D: Product Recommendation on Cart Page	21
Appendix E: Code and Result for confidence level of all item pairs.....	22
Appendix F: Code and Result for confidence level of specific item pairs	22
Appendix H: Data Dictionary	23
References	25

Executive Summary

The retail grocery industry in the United States faces a precarious economic environment. Due primarily to competition from warehouse clubs, supercenters, and e-commerce, retail grocery sales have underperformed the U.S. retail sector and the overall U.S. economy, and employment growth in the industry has been stagnant. Yet, a large proportion of consumers maintain a strong preference for shopping at retail grocery stores, and total grocery industry sales and employment still exceed sales and employment at warehouse clubs/supercenters and e-commerce retailers. To compete in this setting, many retail grocers are turning to third-party online grocery delivery services offering online shopping and same-day grocery delivery, the largest of which is Instacart.

Our team came up with a business problem in which after solving, can help the online grocery stores in managing their business to gain an edge over the market. The specific business problem is to drive higher sales volume and customer retention.

The solution involves using association ruling to find out which products were frequently bought together. The solution is successful in solving both the business and analytics problems, while providing unique and insightful methods for the online grocery stores to gain an edge over their competitors and at the same time, incorporating analytical tools and technology to improve their efficiency and productivity, providing value for all stakeholders.

Instacart Overview

Instacart Inc. is a US based company which operates a crowdsourcing platform that provides same-day grocery delivery from multiple stores to customers (Acosta, 2019). Founded in 2012, Instacart quickly gained traction and has expanded its customer base to a reported figure of 500,000 to generate a revenue of \$2 billion in 2017 (Jungleworks, 2020). Instacart operates on an asset-light business model with neither warehousing nor physical stores with most of its expenses incurred on technology set-up, employee salaries and commission to shoppers. As a crowdsourcing platform with zero inventory, Instacart relies on delivery fees, memberships and markup pricing of 15% or more as its primary source of income (Stanley, 2016).

Instacart Business Model

Instacart operates as a four-sided marketplace, as shown in figure 1, with multiple stakeholders to coordinate with. The first two sides are the consumer and the personal shopper. The consumer uses the application to shop for groceries and have them delivered. The personal shopper goes to the store, purchase the groceries ordered by the consumer and delivers them straight to the doorstep. The next two sides are the retailers and the products. Instacart has established partnerships with 160 different retailers. Several of these retailers are household names such as Costco, Wholefoods and Target. These retailers already possess a huge and loyal customer base and they are allowing Instacart to access both their customer base and inventory. The last side refers to the products and they are closely interlinked with the retailers. The retailers own these products in their stores and warehouses. Instacart can provide value on the advertising side through its data science domain. By providing more relevant product offerings to the consumers through the application, increased sales conversion rate can be achieved.



Figure 1: Four-sided Marketplace

Online Grocery Sector Opportunities

The online grocery sector is set to disrupt the bricks-and-mortar grocery scene, but the extent and nature of disruption depends from market to market. In some markets, companies provide online-only home delivery options with no physical stores at all. In other markets, companies provide a mix of click-and-collect and delivery models. However, it is expected that bricks-and-mortar stores are going to face significant challenges moving forward with all these more convenient options surfacing from online groceries.

The online grocery scene has experienced rapid growth in recent years with a US report stating a 50% growth in online grocery sales 2018. However, the number of shoppers purchasing groceries online remains low with only 11% of people reported to have purchased groceries online for pickup or home-delivery at least once a month in a Gallup survey and 81% stating to have never purchased groceries online (Digital Commerce 360, 2019). Although the sector remains small, online grocery

sales has doubled in sales from 2016 to 2018 and is projected to grow by 17% to 19% each year over the next few years (Lipsman, 2019).

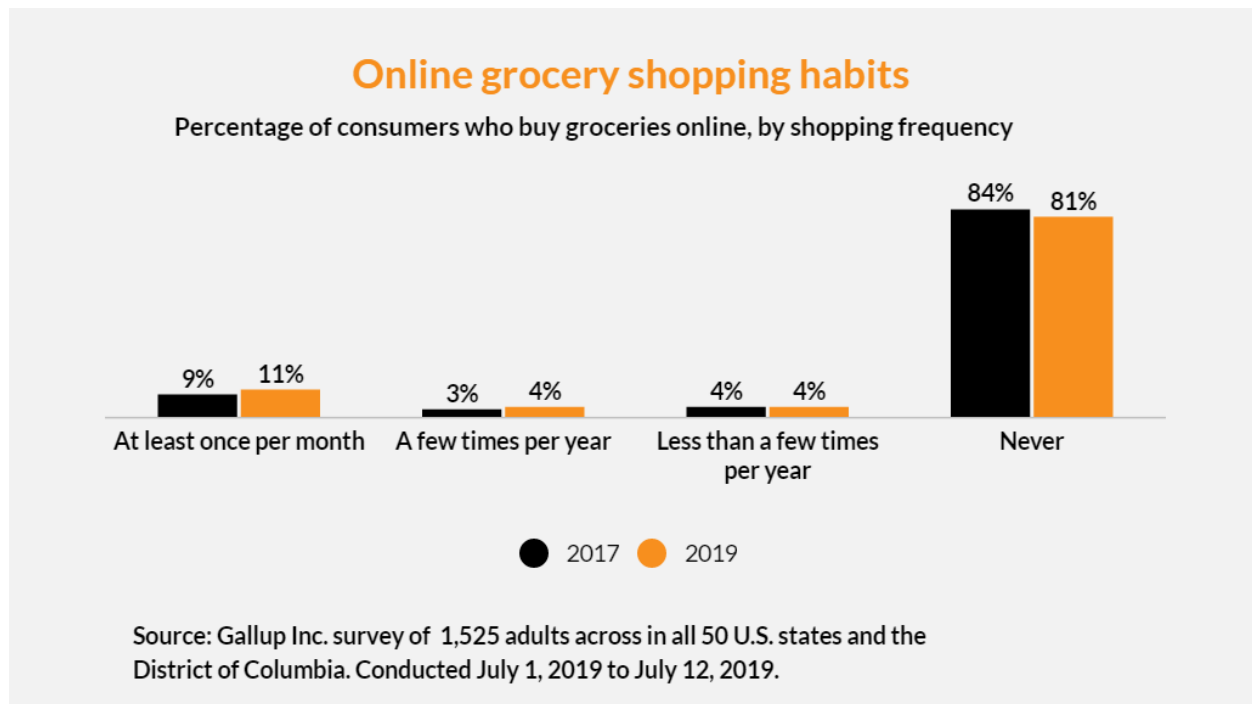


Figure 2: Online Grocery Shopping Habits

With the rapid growth of the online grocery scene, major players such as Amazon and Walmart have been ramping up their efforts in their pickup and delivery services in a bid to latch on this rising trend. In August 2019, research found that Walmart's online grocery operations had 62% more customers than Instacart (Stine, 2019), which holds a 14.5% market share in the online grocery industry (Pace, 2019).

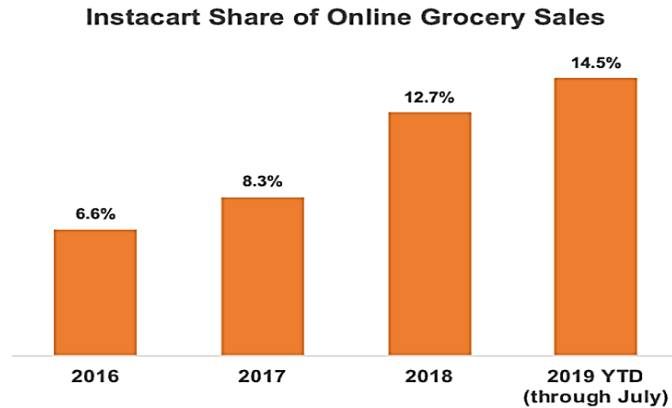


Figure 3: Instacart Share of Online Grocery Sales

In this small and competitive sector, maximizing sales and customer retention are crucial to success for Instacart, a company that solely relies on online grocery sales for revenue.

Plenty of opportunities exist within the online grocery sector. In the technology era, consumers are no strangers to online shopping using websites or mobile applications. In fact, many consumers are willing to pay a premium for convenience especially in developed countries. This is one of the main reasons why online shopping businesses are gaining traction globally. Compared to other products, groceries are demanded in almost every household as they are daily necessities. With the increasingly busier lifestyles of the people, an online grocery store seems to be a perfect business idea for today as well as tomorrow (Treder, 2019).

Business Problem

Instacart relies heavily on sales volume in maintaining profitability given its low profit margin of an estimate of 2% per order (Gaffney, n.d.). While Instacart has no inventory or warehousing, the high cost of technology development, salaries and commission to shoppers are reasons for the low margin. Furthermore, the lack of inventory and need to attract shoppers has resulted in the need for a high markup of 15% to remain profitable in each transaction.

The competitive landscape and low profit margins stress the importance of Instacart's capability in driving higher sales volume as well as a viable customer retention strategy. Therefore, our project seeks to address the 2 key issues of driving higher sales volume and customer retention through analytics.

Being a service-based company, customer satisfaction lies at the heart of Instacart's business. In a digital landscape, the need to better understand user needs and behaviors in designing the best user experience has become increasingly important in driving sales and establishing brand loyalty. Instacart currently uses analytics for various reasons such as their order fulfilment algorithm to drive operational efficiency and their tailored compensation system. However, in terms of customer retention and maximizing sales volume there is lack of a clear strategy which poses a significant risk to its business. Given that only 11% of shoppers purchase groceries online, a low market share of 14.5% and a low margin of 2%, it is crucial for Instacart to not only the retail grocery industry in the United States faces a precarious economic environment. Due primarily to competition from warehouse clubs, supercenters, and e-commerce, retail grocery sales have underperformed the U.S. retail sector and the overall U.S. economy, and employment growth in the industry has been stagnant. Yet, a large proportion of consumers maintain a strong preference for shopping at retail grocery stores, and total grocery industry sales and employment still exceed sales and employment at warehouse clubs/supercenters and e-commerce retailers. To compete in this setting, many retail grocers are turning to third-party online grocery delivery services offering online shopping and same-day grocery delivery, the largest of which is Instacart.

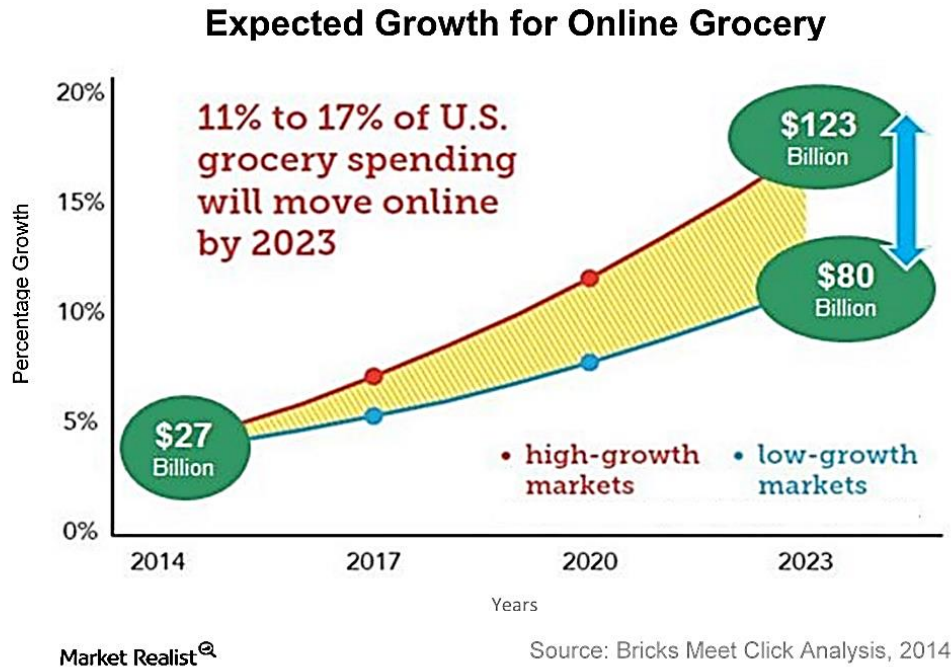


Figure 4: Expected Growth of Online Grocery Sales

Proposed Approach & Desired Outcomes

With the massive amount of data on Instacart's existing system and mobile applications, Instacart could predict what a customer is likely to purchase based on his/her past transactions and activities on Instacart's e-commerce platform. We propose to explore the associative relationship between Instacart's products based on past transaction records and make recommendations to customers accordingly.

Through analytics, Instacart could find out what products are commonly purchased together. For example, the model might tell us that a customer has a 60% chance of purchasing banana if he has bought a carton of milk. To achieve this objective, the associative rule mining is the perfect analytics tool as it is best working with categorical data (transactions data) and could lead to some hidden but useful discoveries.

When such associative rules are discovered, Instacart's interactive system could recommend products based on customers' past transactions and their activities on the e-commerce platform such as in-app search, views, items saved or liked and items in cart. In addition, different weightage could be given to various activities when making the recommendations. For example, the system would recommend 6 products based on a customer's past transactions but only 4 products based on what he or she likes or puts into cart.

Furthermore, bundle sales promotion campaign could be launched for products with high associative degree to encourage sales. A discount on sets of items that fit the customers' needs and wants will be a great incentive to buy from Instacart instead of its competitors. On the other hand, an intelligent recommendation system is also crucial for better customer shopping experience, sites like Netflix, Amazon and Spotify are constantly adjusting their recommendations to customers based on previous interaction. A great recommendation can lead to better customer engagement and hence customer retention (Tricon Infotech, 2019).

After the implementation of the program, we expect to see

- (a) an increase in sales revenue of 3% or more
- (b) an increase of customer retention rate of 10% or more

where the customer retention rate is calculated by the formula:

$$\text{Customer Retention Rate} = ((\text{No. of Customers at End of Period} - \text{No. of Customers Acquired During Period}) / \text{No. of Customers at Start of Period}) \times 100\%$$

Data exploration

We begin our analysis with preliminary data exploration to derive understanding with regards to the following questions to further optimize the implementation of our analytics solution:

1. Which product accounts for the most sales?
2. What time should our promotions be pushed out?
3. Which product are most frequently reordered?
4. Which set of items are most frequently bought together?

As can be seen from figure 5, fruits and vegetables are most frequently ordered products on Instacart. In fact, organic fruits and vegetables are seemingly popular amongst customers. Armed with this knowledge, association rules and further data exploration could be performed to determine which organic products can be packaged and marketed together with these organic fruits and vegetables.

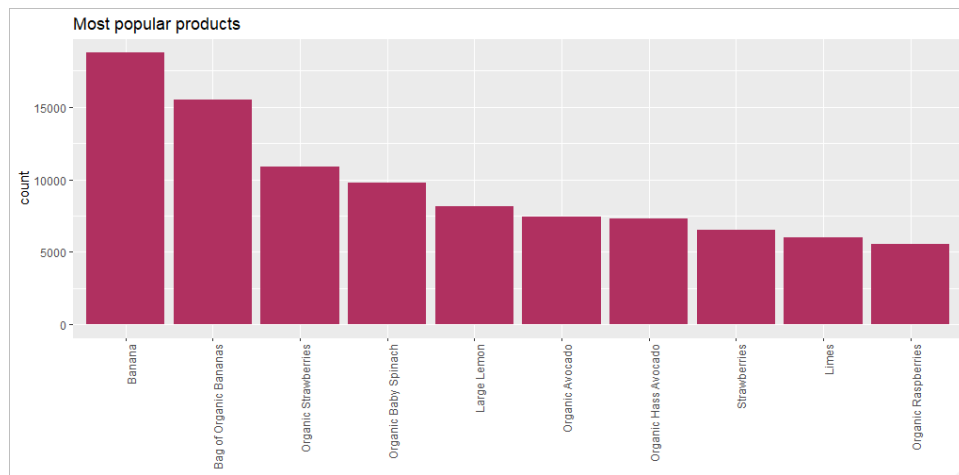


Figure 5: Graph of Most Popular Products

From figure 6, we can see that majority of Instacart's sales are aggregated between 10 a.m. to 5 p.m. This insight can be used to time our promotion and push notifications on promotions during

high volume period each day. This would maximize the conversion of marketing efforts through the application to boost sales.

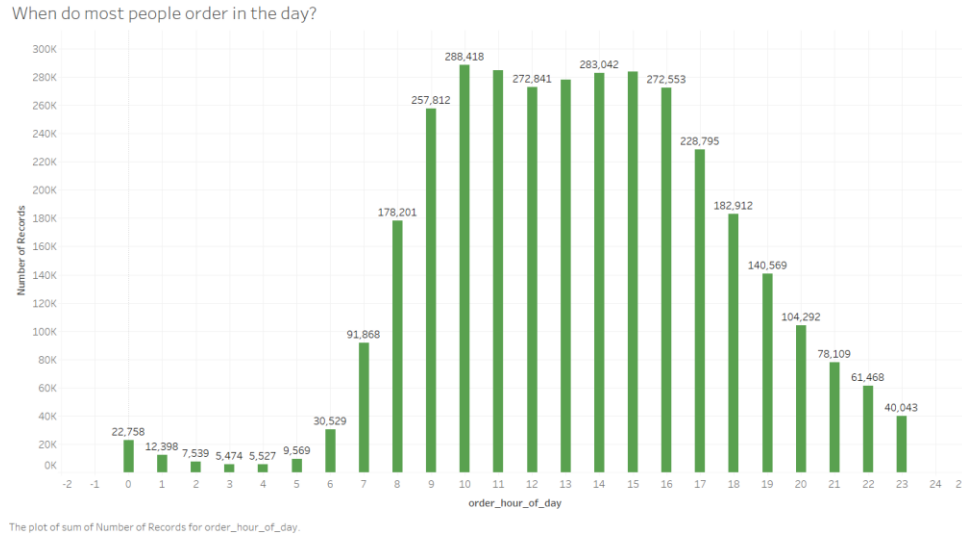


Figure 6: Graph of Products Sales for Different Hour of the Day

Another trend that Instacart can focus on would be the products that have a high chance of being reordered by customers. Instacart can use this information and push these items as “Popular items other customers buy”. By using this information, Instacart can also easily narrow down the products to recommend to existing customers as well as new ones. On the supplier side, Instacart can also make sure that on the days the customers usually reorder the products, for example, the 2% Lactose Free Milk, there would be enough stock for them. Instacart can also easily use the top reordered products as a baseline to form their association rule in order to easily bundle the products that customers usually reorder together to improve the overall customer experience on the app. This will ensure that customers will continue using the app to purchase their healthy milk which will in turn help to promote customer retention.

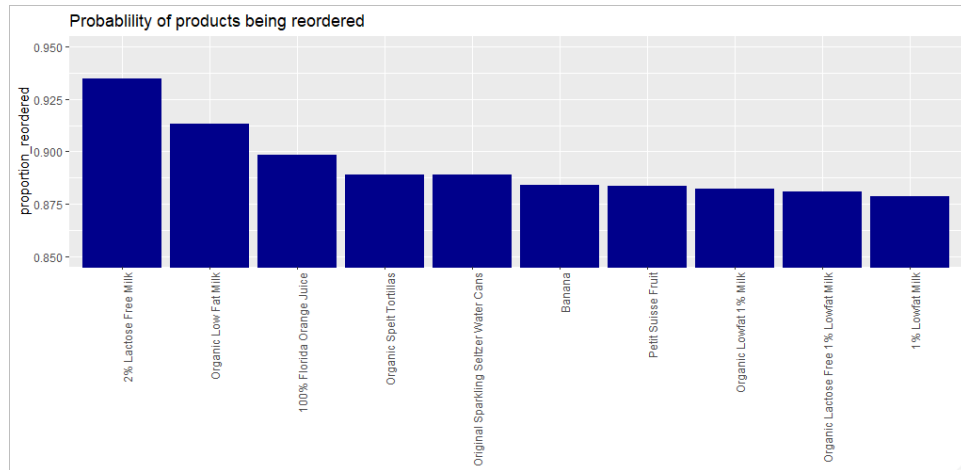


Figure 7: Graph of Probability of Products Being Reordered

From figure 8, we can sort the popularity of item pairs in descending order. Hence, Instacart can choose to focus on item pairs with high occurrence in our analytical model below or bundle the items together to be sold as a package. These actions may improve the overall customer experience as customers will not have the hassle of finding all the different items when they can just buy the recommended packages. Sales can also be improved as bundling of goods has a high correlation with impulse purchases (Hemmis, 2017). An increase in the number of impulse purchases will inadvertently increase the sales revenue.

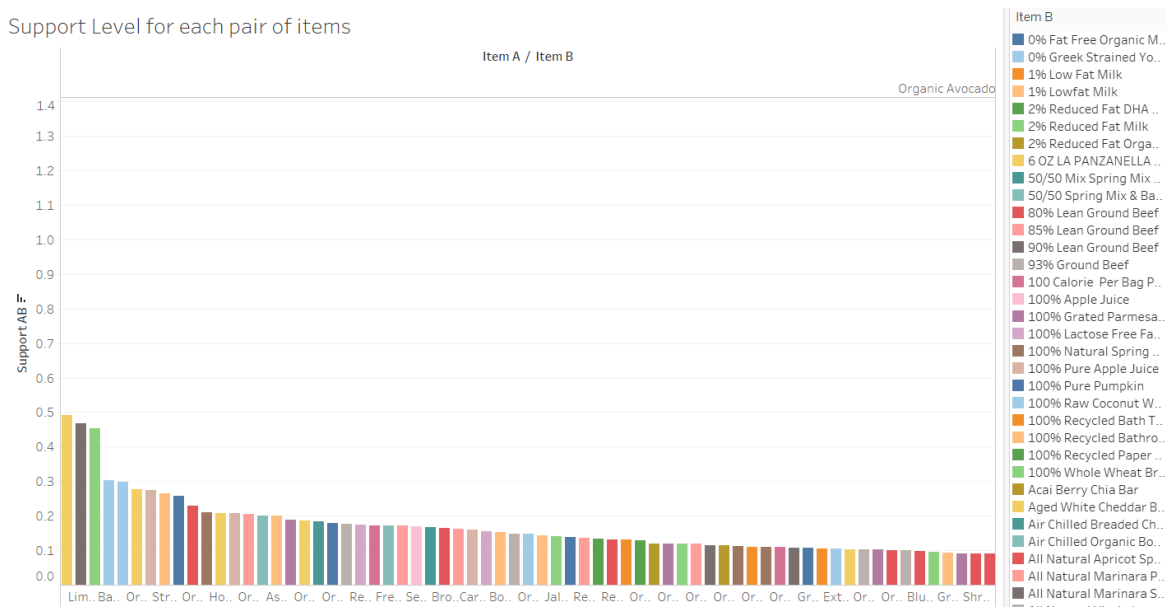


Figure 8: Graph of Support Level for Each Pair of Items

Analytics Solution

Our project focuses on the use of association rules to understand consumer behavior that defines in-app user experience and promotions in order to maximize items purchased per transaction and customer retention. The analytics solution is premised on using insight derived through association rules to determine the in-app item arrangement and recommendations by using items added to cart to predict the next item they are likely to purchase as well as the creation of promotion bundles.

Association Ruling will be used to resolve our business problem and discover sales correlations. Given the enormous data size from Instacart, Python's built-in Apriori package could not generate note-worthy Association Rule from our source data. We have also decided not to generate sample data from our dataset and derive our rules from the generated sample as there are many unique products within our dataset with low occurrence rate hence, certain abnormalities and interesting relationships among our products will be eradicated due to the randomness of sampling. To portray the utmost accurate association rule from our given real dataset from Instacart, we would be using

a series of python generators and leveraging on the Apriori algorithm to conduct Association Rule Mining. Our association rule is only comprised of 2 items, A and B due to the low support level for a basket of 3 or more items as Instacart has an extremely wide range of varied items.

Via the use of Association Rule Mining with a min support of A & B at 2%, we would generate the table shown below (sort by descending Confidence AtoB).

	itemA	itemB	freqAB	supportAB	freqA	supportA	freqB	supportB	confidenceAtoB	confidenceBtoA	lift
0	Oh My Yog! Pacific Coast Strawberry Trilayer Y...	Oh My Yog! Organic Wild Quebec Blueberry Cream...	860	0.028907	2856	0.095998	2271	0.076335	0.301120	0.378688	3.944745
2	Unsweetened Blackberry Water	Raspberry Essence Water	660	0.022184	3108	0.104468	2025	0.068066	0.212355	0.325926	3.119850
3	Organic Fiber & Protein Pear Blueberry & Spina...	Fiber & Protein Organic Pears, Raspberries, Bu...	606	0.020369	2782	0.093511	2167	0.072839	0.217829	0.279649	2.990560
4	Oh My Yog! Organic Wild Quebec Blueberry Cream...	Oh My Yog! Pacific Coast Strawberry Trilayer Y...	642	0.021579	2271	0.076335	2856	0.095998	0.282695	0.224790	2.944798
1	Oh My Yog! Pacific Coast Strawberry Trilayer Y...	Oh My Yog! Madagascar Vanilla Trilayer Yogyurt	640	0.021512	2856	0.095998	2567	0.086284	0.224090	0.249318	2.597119
5	Unsweetened Watermelon Water	Unsweet Peach Water	599	0.020134	4107	0.138048	1777	0.059730	0.145849	0.337085	2.441803

Figure 9: Sample Association Rules for Similar Items

Item A	Item added to the cart
Item B	Suggested item to be considered by customers.
Support AB	% simultaneous occurrence for Item A and B
Confidence AtoB	The confidence of the rule that A results in B
Lift	Ratio of the confidence of the rule and the expected confidence of the rule.

Table 1: Data Dictionary for Association Rules

In our analysis, we found that the lift for most of the association rules generated are more than 1 which indicates that the purchase of item A increases the likelihood of them purchasing item B. We further deduce that items that are similar in names have a much higher confidence association

rule than items that are dissimilar in names. This implies customers tend to purchase multiple products of the same brand in a single transaction. Instacart should therefore group products of similar nature and optimize the in-app arrangement of items based on the association rules. However, these association rules are rather self-explanatory since products that are similar in nature are naturally likely to be received in a similar fashion. We further sought for rules associating products of different types and brands for a more comprehensive product recommendation system.

	itemA	itemB	freqAB	supportAB	freqA	supportA	freqB	supportB	confidenceAtoB	confidenceBtoA	lift
45	Chocolate Peanut Butter	Coconut Chia Bar	970	0.032604	4649	0.156266	4646	0.156165	0.208647	0.208782	1.336069

Figure 10: Sample Association Rule for Dissimilar Item

The above association rule is an example of dissimilar products with a high confidence level and lift. These are association rules which are exceptionally insightful for companies. Association between similar products are intuitive to management but unique association between dissimilar products are often overlooked. These association rules could be used to pilot new promotions and recommendations for cart optimization.

An average customer would first search for a single item to purchase before looking for their next item. Using the association rules, we could recreate the users' experience through layered recommendation each time a product is added to cart. This allows maximization of sales value per transaction.

Analytical Solution Performance Measures

Once our solution is implemented Instacart will have to analyze the efficacy of our solution. This can be done via the use of Quantile Regression. Data will be collected before the implementation of our solution and after the implementation. The y-outcomes (ConfidenceAtoB, Support for item A and B and Lift) will be regressed against time to determine if our solution has led to an improvement in the aforementioned y-outcomes over time. The y-outcomes will be stratified into various quantiles (25th percentile, 50th percentile and 75th percentile) to give a complete analysis of how our model will affect the y-outcomes (e.g. Support) for different quantile range. Quantile Regression is an appropriate analytical tool to determine the performance of our model as it can hypothesized that product recommendation algorithms may not be effective on items that have very low purchase rate (low support) since these items are rather unique and customers may specifically want those products only.

A cost-performance matrix can be done on our model to determine the usability and sustainability of it. Walmart released an application on Aug 2018 to provide increase accessibility and convenience to their customers. In Walmart application, product recommendation is one of their key features to generate more sales and provide a better user experience to their customers. By referencing from Walmart's balance sheet for YA 2018 and 2019 (Appendix A), there is an increase in gross revenue of 3% after the implementation of their application. By benchmarking our model's success against Walmart's success and relevant industry standards (Crisp, 2018), we can hypothesis that our model should result in revenue improvement of around 2-5% to be classified as successful. However, the implementation of our model and database management of

company's data can result in high costs due to the high computational capacity and intricate complexities behind such an implementation (Hazel, 2020).

Hence, a cost-performance matrix and analysis must be done to ascertain the positive outlook our model can bring to Instacart. We expected the first few months of the implementation to be cost-ineffective and may result in higher operational cost whilst emanating poor improvement scores and revenue growth. This deplorable start to our model implementation can be attributed to the relative low confidence level and support of our association rules due to the wide myriad of distinct products in Instacart.

Model Improvements and Future Advancements

Fortunately, Association Rule is a rule-based machine learning which is used to discover relations via inputted data and threshold adjustments (min support/confidence level). Overtime, by receiving more inputted data, our model can improve on its confidence level by adopting a semi-supervised machine learning which stems from previous collected data (Berka & Rauch, 2010). Minimum support and confidence level can also be changed to provide higher quality association rules albeit they will be fewer in numbers. Grouping of items into item sets can also be considered to reduce the variations in our data.

The current Apriori algorithm can be computationally intractable if Instacart were to keep expanding on their database. Hence, a Gibbs-sampling-induced stochastic search procedure can be developed to randomly sample association rules via a newly developed variable: "Importance" and perform rule mining from the reduced transaction dataset generated by the sample (Qian, Rao,

Sun, & Wu, 2016). Via this new approach, more impactful association rules can be generated at a lower computational cost.

Analytical Solution Target

Ultimately, Instacart should aim to use our model to solve the underlying business problem aforementioned. Via the use of this model, Instacart can increase sales through successful item recommendations and larger item baskets. In addition, customer in-app experience and retention rate will improve as well since customers can find the items that they are looking for more seamlessly and quickly when such items are being recommended to them automatically when they add an item of similar nature to their cart. Given the decreased attention span among adults in the 21st century and the multitude of Instacart's competitors, having a good UI and cart suggestions which can significantly improve success of Instacart (Eaton, 2018). We aim to have an increased in sales and retention rate of 3% and 5% respectively, which can translate to an increase in profitability by over 75% since attracting new customers to e-grocery commerce can cost Instacart 5-25 times more than keeping an existing customer (Kulbyte, 2020).

Analytical Solution Demonstration

Instacart can further analyze the outputs of our model (Association Rule) via the use of analytical software like Tableau. From Appendix B, Instacart can view the graphical representation of the confidence level among the different item pairs. From the graph, Instacart can choose to implement chosen association rules (top few rules with highest confidence) into their software to recommend

specific items (Item B) in the product page to our customers when they add an item (Item A) into their cart or when they are viewing the item on its product page (Appendix C).

Instacart can also store the outputs of our model into a relational database management system (MySQL). From Appendix F, Instacart can view the associated rules and their confidence level of a specific item. For example, if Instacart wants to promote the sale of milk, they can look at the association rules resulted from Cereals (Item A) to see if milk (Item B) and cereals have association rules of high confidence. In addition, Instacart can also effectuate a ‘last-minute recommendations’ for their customers before they check out at the checkout page (Appendix D). These recommendations will be based on the top 3-5 association rules (ranked by confidence) that can be derived from the items (Item A) added into the customer’s cart (Appendix G).

Through the use of MySQL and Tableau, Instacart can analyze the relevance and accuracy of our model generated association rules and through multiple data collection processes, test phases and further improvements made to our model, Instacart will have a comprehensive vista of their customer’s buying pattern.

Monitoring of Analytical Solution

In assessing the achievement of our business objectives of maximizing sales value per transaction and customer retention, apart from the overall increase in revenue, we would also need to measure the changes in the average sales value per transaction as well as the sales attributed to customers in comparison to that of prior years. However, these alone would not give insights to the effectiveness of the analytics solution. In order to assess that, we would need to measure the conversion rate of recommendations and promotions that were based on association rules.

Conclusion

In conclusion, we believe that our model can resolve the aforementioned business problem and help increase sales and customer retention rate for Instacart. Instacart should aspire to further develop their data collection and expand on our model to achieve better results. For example, our model, with the right variables, can generate a specific and time-based association rule (e.g. Buying beer at night -> Buying potato chips). To further distant themselves from other competitors in the e-grocery market, Instacart should utilize the advantages of data science and develop a customer-friendly interface to promote both sales and retention rate.

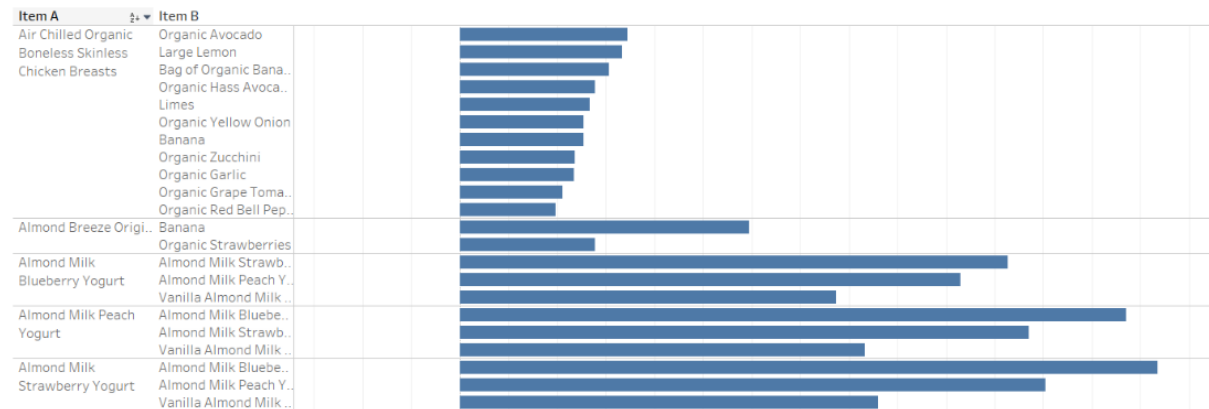
Appendices

Appendix A: Walmart's Balance Sheet for YA 2019, 2018 and 2017

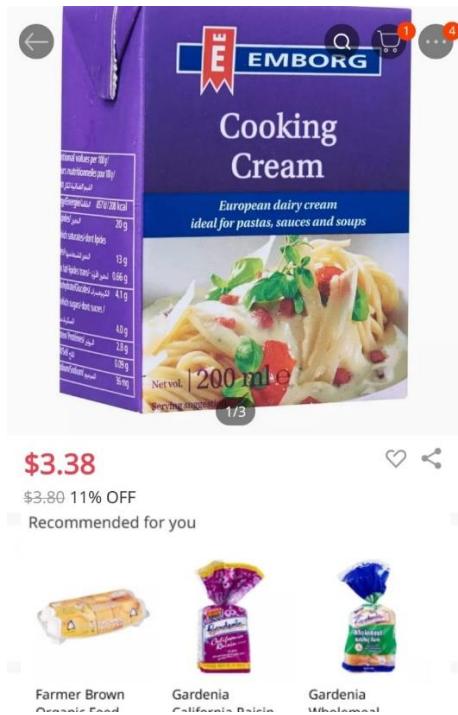
Breakdown	TTM	1/30/2020	1/30/2019	1/30/2018	1/30/2017
Total Revenue	523,964,000	523,964,000	514,405,000	500,343,000	485,873,000
Cost of Revenue	394,605,000	394,605,000	385,301,000	373,396,000	361,256,000
Gross Profit	129,359,000	129,359,000	129,104,000	126,947,000	124,617,000
✓ Operating Expenses					
Selling General and Administrati...	108,791,000	108,791,000	107,147,000	106,510,000	101,853,000
Total Operating Expenses	108,791,000	108,791,000	107,147,000	106,510,000	101,853,000
Operating Income or Loss	20,568,000	20,568,000	21,957,000	20,437,000	22,764,000
Interest Expense	2,599,000	2,599,000	2,346,000	2,330,000	2,367,000
Total Other Income/Expenses Net	1,958,000	1,958,000	-8,368,000	-3,136,000	-
Income Before Tax	20,116,000	20,116,000	11,460,000	15,123,000	20,497,000
Income Tax Expense	4,915,000	4,915,000	4,281,000	4,600,000	6,204,000
Income from Continuing Operations	15,201,000	15,201,000	7,179,000	10,523,000	14,293,000
Net Income	14,881,000	14,881,000	6,670,000	9,862,000	13,643,000

Appendix B: Confidence Level for each pair of items in Instacart

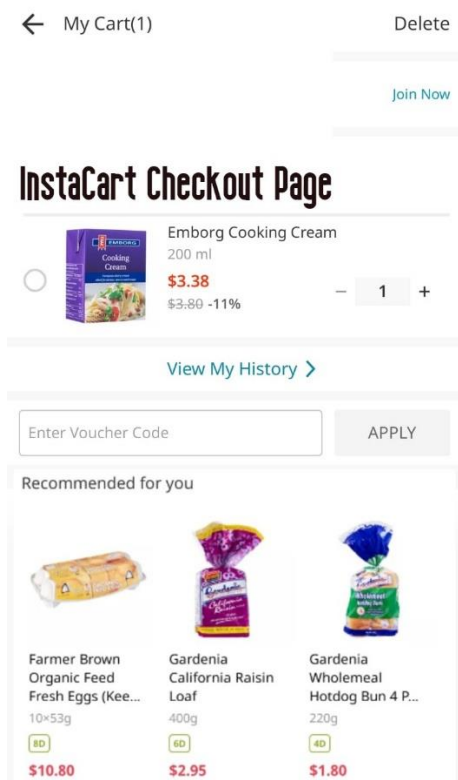
Confidence Level for Each Pair of Items



Appendix C: Product Recommendations on product page



Appendix D: Product Recommendation on Cart Page



Appendix E: Code and Result for confidence level of all item pairs

```
select itemA, itemB, supportAB, confidenceAtoB from project1.resultba2 where supportAB > 0.02
and confidenceAtoB > 0.2
order by confidenceAtoB desc;
```

itemA	itemB	supportAB	confidenceAtoB
Oh My Yog! Pacific Co...	Oh My Yog! Organic Wild Q...	0.02890697	0.301120448
Almond Milk Strawberr...	Almond Milk Blueberry Yogurt	0.05512492	0.287114846
Oh My Yog! Organic ...	Oh My Yog! Pacific Coast St...	0.021579389	0.282694848
Almond Milk Peach Yo...	Almond Milk Blueberry Yogurt	0.043326843	0.274080374
Greek Whole Milk Blen...	Organic Greek Whole Milk Bl...	0.023159189	0.260294673
Yotoddler Organic Pea...	Organic Whole Milk Strawbe...	0.051528355	0.250490196
Almond Milk Strawberr...	Almond Milk Peach Yogurt	0.046251152	0.240896359

Appendix F: Code and Result for confidence level of specific item pairs

```
select itemA, itemB, supportAB, confidenceAtoB from project1.resultba2 where supportAB > 0
and confidenceAtoB > 0 and itemA= "Coconut Chia Bar"
order by confidenceAtoB desc;
```

itemA	itemB	supportAB	confidenceAtoB
Coconut Chia Bar	Chocolate Peanut Butter	0.030251481	0.193715024
Coconut Chia Bar	Acai Berry Chia Bar	0.029243098	0.187257856

Appendix G: Code and Result for confidence level of items pairs derived from Cart Items

```
create view CartItems1( itemA, itemB, supportAB, confidenceAtoB)
as select itemA, itemB, supportAB, confidenceAtoB from project1.resultba2 where supportAB > 0
and confidenceAtoB > 0 and itemA= "Coconut Chia Bar" or itemA = "Chocolate Sea Salt" or
itemA= "Peanut Butter Bar"
order by confidenceAtoB desc;
```

```
select* from CartItems1 where itemB != "Coconut Chia Bar" and
itemB != "Chocolate Sea Salt" and itemB != "Chocolate Peanut Butter"
order by confidenceAtoB desc limit 3;
```

itemA	itemB	supportAB	confidenceAtoB
Chocolate Sea Salt	Coconut Chocolate Bar	0.028873358	0.198566805
Coconut Chia Bar	Acai Berry Chia Bar	0.029243098	0.187257856
Chocolate Sea Salt	Peanut Butter Egg White Protein Bar	0.021209649	0.145862228

Appendix H: Data Dictionary

Orders (3.4m rows, 206k users)		
Attribute	Description	Value
Order_id	order identifier	Nil
User_id	customer identifier	Nil
Eval_set	which evaluation set this order belongs in	Prior, Train, Test
Order_number	the order sequence number for this user	1 - nth
Order_dow	the day of the week the order was placed on	0 – 6
Order_hour_of_day	the hour of the day the order was placed on	0 - 23
Days_since_prior	days since the last order	0 - 30

Products (50k rows)		
Attribute	Description	Value
product_id	product identifier	Nil
Product_name	name of the product	Nil
Aisle_id	aisle product is in	Nil
Department_id	department product is in	Nil

Aisle (134 rows)		
Attribute	Description	Value
aisle_id	Aisle identifier	Nil
aisle	name of the aisle	Nil

Department (21 rows)		
Attribute	Description	Value
department_id	department identifier	Nil
department	name of the department	Nil

Order_products_SET(30m+ rows)		
Attribute	Description	Value
product_id	product identifier	Nil
Order_id	Order identifier	Nil
Add_to_cart_order	Order in which each product was added to cart	Nil
reordered	If product was ordered by this user in the past	1 if reordered, 0 if not.

where SET is one of the four following evaluation sets (eval_set in orders):

- "prior": orders prior to that users most recent order (~3.2m orders)
- "train": training data supplied to participants (~131k orders)

- "test": test data reserved for machine learning competitions (~75k orders)

References

3 Million Instacart Orders, Open Sourced. (n.d.). Retrieved April 1, 2020, from Instacart:

<https://www.instacart.com/datasets/grocery-shopping-2017>

Acosta, G. (2019, September 10). *How Instacart Has Outsmarted Amazon*. Retrieved from

Progressive Grocer: <https://progressivegrocer.com/how-instacart-has-outsmarted-amazon>

Berka, P., & Rauch, J. (2010). Machine Learning and Association Rules. *Czech Science Foundation*.

Crisp, A. (2018, December 14). *Global grocery markets forecast to create \$1.9 trillion opportunity by 2023*. Retrieved from IGD: <https://www.igd.com/articles/article-viewer/t/global-grocery-markets-forecast-to-create-19-trillion-opportunity-by-2023/i/20842>

Digital Commerce 360. (2019, August 27). *81% of US consumers say they never buy groceries online*. Retrieved from Digital Commerce 360:

<https://www.digitalcommerce360.com/2019/08/27/81-of-consumers-say-they-never-buy-groceries-online/>

Eaton, T. (2018, May 18). *UX Collection*. Retrieved from The power of good UI and how it enhances engagement: <https://uxdesign.cc/the-power-of-good-user-interface-and-how-it-enhances-engagement-the-new-currency-in-the-digital-43a59bcd9bda>

- Gaffney, P. (n.d.). *Toptal*. Retrieved April 4, 2019, from Is \$4.2 Billion Reasonable? How to Evaluate Instacart's Valuation: <https://www.toptal.com/finance/valuation/instacart-valuation>
- Hazel, T. (2020, March 13). *datanami*. Retrieved from Data is Cheap, Information is Expensive: <https://www.datanami.com/2020/03/13/data-is-cheap-information-is-expensive/>
- Hemmis, M. (2017, November 16). *Impulsive Buys: What is Impulse Item Shopping?* Retrieved from FastSpring: <https://fastspring.com/blog/4-proven-techniques-for-triggering-impulse-buys/>
- Jungleworks. (2020). *How Instacart Works: Comprehensive Business & Revenue Model*. Retrieved from Jungleworks: <https://jungleworks.com/how-instacart-works-makes-money-revenue-business-model/>
- Kulbytté, T. (2020, March 6). *5 Unique Ways to Increase Customer Retention (and Increase Profits!)*. Retrieved from SuperOffice: <https://www.superoffice.com/blog/customer-retention-tips-with-crm-software/>
- Lipsman, A. (2019, April 8). *eMarketer*. Retrieved from Grocery Ecommerce 2019 Online Food and Beverage Sales Reach Inflection Point: <https://www.emarketer.com/content/grocery-ecommerce-2019>
- Pace, M. (2019, October 22). *Digital Commerce 360*. Retrieved from Instacart finds online grocery success beyond Whole Foods: <https://www.digitalcommerce360.com/2019/10/22/instacart-finds-online-grocery-success-beyond-whole-foods/>

Qian, G., Rao, C. R., Sun, X., & Wu, Y. (2016, May 3). *Boosting association rule mining in large datasets via Gibbs sampling*. Retrieved from PNAS:

<https://www.pnas.org/content/113/18/4958>

Stanley, J. (2016, February 18). *Data Science at Instacart*. Retrieved from tech-at-instacart:

<https://tech.instacart.com/data-science-at-instacart-dabbd2d3f279>

Stine, L. (2019, August 14). *Walmart has more online customers than Instacart, report says*.

Retrieved from Grocery Dive: <https://www.grocerydive.com/news/walmart-has-more-online-customers-than-instacart-report-says/560854/>

Treder, T. (2019, February 12). *Online grocery business: Three ways to a key revenue stream*.

Retrieved from The Future of Customer Engagement and Experience: <https://www.the-future-of-commerce.com/2019/02/12/online-grocery-cpg/>

Tricon Infotech. (2019, April 11). *Improving Customer Engagement with Recommender Systems*.

Retrieved from Medium: <https://medium.com/@triconinfotech/improving-customer-engagement-with-recommender-systems-b423bdbb4e55>