

# **Fall 2017- MIS584 - Business Intelligence**

## **Instacart: BI application**

**Team 10:**

**Kavin Chandrasekaran**

**Sishi Hao**

**Abdulaziz Alajaji**

**Fangling Zhang**

**Submitted to:**

**Dr. Nima Kordzadeh**

**Executive Summary:**

This report is to propose a business intelligence solution to Instacart and show how Instacart could use benefit from using a business intelligence comprehensive solution. Instacart is an on-demand same-day grocery delivery platform, using a sharing economy model, founded in 2012 in San Francisco, California. In five years, Instacart has grown into a national company in more than 26 metro areas and has many partner stores across the country. Although Instacart has used Segment as the platform to collect data and analyze customers' and shoppers' behavior, Instacart still has problems: inefficient resources arrangement, long delivery time and poor shopper performances, not good products recommendation. In order to solve problems, we provide detailed analysis about the current state and how to incorporate our proposal within the framework. The business solutions includes two kinds dashboards and recommendation system to help Instacart for its three main problems listed as the following:

- Regional managers can use a dashboard to monitor different Key Performance Indicators (KPI) in each city/state, and then can focus on high profitable customers, give them priority to reach in-store part-time shoppers or arrange full time shoppers serve these customers exclusively. Instacart can provide these customers delivery fee discount or products coupon from stores within a loyalty program, which can also benefit stores. The dashboard can also show most popular products which can be shared with stores so that they can arrange inventory and marketing.
- Shopper manager can use a dashboard to monitor delivery times to reduce

delivery time and eventually improves customer satisfaction. Shopper manager can figure out in a specific area, the slower delivery is distributed in which hours of a specific day of week with the order percentage of that hours in that day. If in a specific hour, the delivery is very slow and order percentage is large, manager can arrange more in-store shoppers or reward part-time shoppers to deliver in that hour. Peak hour and day of week (more order volume) can be shared with shoppers. If shoppers want to own more money, they can deliver more in peak time. If a customer want to get their products more quickly, they can avoid the peak time. What's more, fastest shopper should be rewarded and slowest shoppers should be trained more or punished.

- The proposed solution has a forecasting system that can help personalize advertisements, service and arrange delivery in advance. Using the dashboard, marketing managers can identify top users who who have highest order frequency. Managers can give these customers, who order frequently, a top priority to assign shoppers and get the order delivered faster and receive special loyalty discount. 'Products in the next order for top users' can help marketing manager to recommend products and send stores' coupon to customers. It also can help stores to arrange products in advance so to reduce delivery time. Marketing manager can provide 'Top 5 products in each department in next orders' to specific stores, and help them plan inventory and promotion.

## Introduction

Instacart is a same-day grocery delivery platform in major cities of USA. Customers select groceries through a smartphone app or website from various retailers

and delivered by a personal shopper. Instacart was founded in year 2012. As of March 2017, it services 36 markets, composed of 1,200 cities in 25 states in America.

Being considered as a leader in on-demand economy, Instacart have competitors include Shipt, Postmates and so on. It's also competing against delivery services provided by grocery chains such as Safeway delivery and dedicated grocery delivery services without physical stores such as Amazon Fresh.

Since Instacart wants to grow customers base and provide perfect customer experience, Instacart choose Segment as a single platform to collect data from customers and analyze them as you can see below:



**Figure: 1:** *Instacart work-flow chart.*

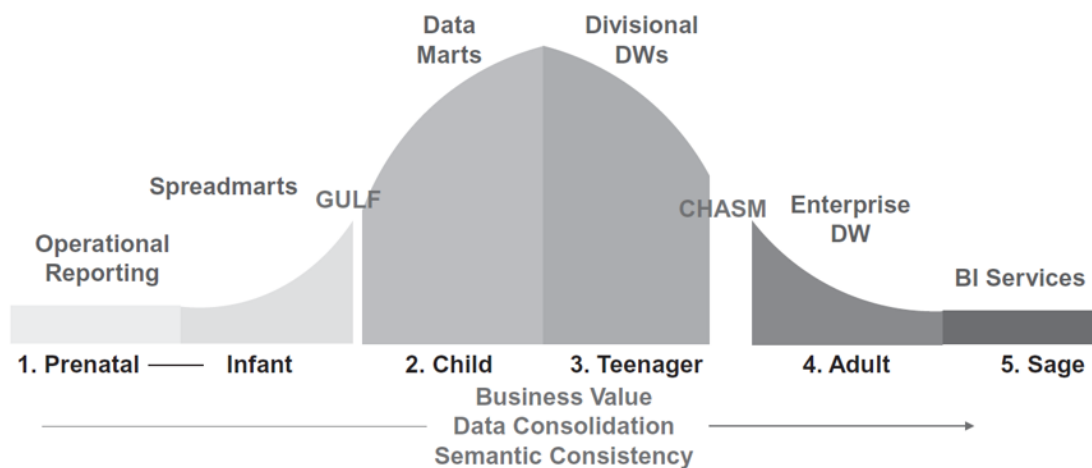
**How Instacart uses Segment:**

Segment is a customer data collection platform with multi-platform support. Because Instacart needs to track all data from different platforms in order to analyze them together, the company uses Segment Sources to collect customers' data from iOS, Android, web and other third-party cloud services like Zendesk, SendGrid, Facebook Ads. What's more, Instacart uses Segment Warehouses to schematize and load data into data warehouses, which is RedShift, and utilize Segment Integrations to push data out for analytics teams and marketing teams.

Segment plays an important role in helping Instacart gain a dynamic analytics environment when Instacart wants to grow customer base and to improve shoppers' experiences. For example, Instacart uses Segment to obtain customer behavior data from different platforms with email data, which would let Instacart measure if customers receive better services from Instacart or whether contacting through email would increase order numbers. Also, the analytics team in Instacart finds their shopper app in Android where should be improved using feedback data from Instacart app. By combining data with shoppers table in Redshift, Instacart could be able to identify behavioral differences between new and experienced shoppers, and help new shoppers become experienced shoppers faster.

Instacart's customers have gotten better experiences when they use Instacart than before, since Instacart learns about whether experiences are good or bad. Data team in Instacart would continue to use Segment to improve experiences, as it's the tool to get the data everywhere it needs to go and then back to warehouses so that Instacart could analyze it to keep improving business.

In Figure 2 below, we have included a business intelligence maturity model charts that is commonly used to evaluate the current state of implementing business intelligence framework within an organization. Until now, implementation of business intelligence in Instacart is in the chasm stage between the teenager stage and adult stage. Instacart is migrating separate business intelligence to an integrated enterprise-wide business intelligence.



**Figure: 2:** *Business Intelligence Maturity Model.*

## The Proposed Business Intelligence Solution

Our proposed solution includes dashboards which would facilitate the regional, shoppers and marketing managers to explore the key areas to focus on. Given these dashboards, the managers would be able to quickly gather insights about the current and past key performance indicators. For example, In a particular location within a timeframe, managers can see a trending line for customers acquisitions status and another trending line for the average delivery to aim for customer satisfaction eventually.

Also one of the task managers can do with the aid of our BI solution is to determine what is the preferred time to order for customers and what days of week has the most numbers of orders. Using our BI solution will let managers use dashboards which can be analysed with multiple filtering and drill-down techniques to help gather insights so managers can make informed decisions.

By using a combination of the data from the dashboards the regional managers could identify the timeframes and locations of slower deliveries and could improve strategies to find more shoppers for a specific location and specific time.

We could address three major problems that are faced by the managers.

### **Resource management:**

The regional managers are responsible for managing the local resources. There are two types of resources, one is the full-service shopper, who performs both the shopping and delivering part of fulfilling the order. The other resource is the in-store shoppers, who are present in a participating store and help with shopping by collecting the items for the full-service shopper to deliver or for the customer to pick-up. Apart from the shoppers there are in-store liaisons called Shift leads who supervise and train the other shoppers. This is an important role, as their knowledge and efficiency would directly impact the performance and delivery times of other shoppers.

### **Shopper performances:**

A key performance indicator (KPI) for Instacart's business is their promised shorter delivery time. Intuitively, we use the delivery time as the measure to evaluate the performance of the shoppers. We also consider the number of orders that the shopper has delivered and their average delivery times to identify more loyal shoppers

and the casual shoppers. We also use this data to identify the timeframes when they are delivering the fastest and slowest to improve and incentivize faster deliveries apart from the tip.

### **Product recommendation:**

In order to provide a highly customized and personalized experience to the customers, our proposal includes a predictive model to predict the customers' next order including when the order is going to be, and what products had the most likelihood to be included in their order. Once we have a reliable model, we would be able to provide personalized coupons and also better manage the inventory in the store. Though the inventory is managed by the store and not the Instacart, the shift lead would be able coordinate with the local store to plan for the future demands. We could also plan for the shopper demand, once we have the time-frame for the future orders. If there is high shopper demand during that time, we could suggest the potential locations for existing or new shoppers to meet the new demand.

### **Schema:**

Our proposed solution uses a data warehouse which collects data about all the orders, the customer information about the users who placed the orders, the product information and the information about the shopper who delivered that order. The `user_orders` table, contains all the collective information. It merges the orders from different tables and joins it with the products that are present in that order. This way we could query for summarized values on a granular level of the products instead of the



orders. We then join that data with the customer information like the location. Finally we have the information about the shopper who delivered the order. We have shopper information like the location and how their delivery time. Since we have all the data about the orders in one table, we can easily and efficiently query to make dashboards and ad-hoc reports, which is key to our solution. Our star schema for the proposed solution is below.

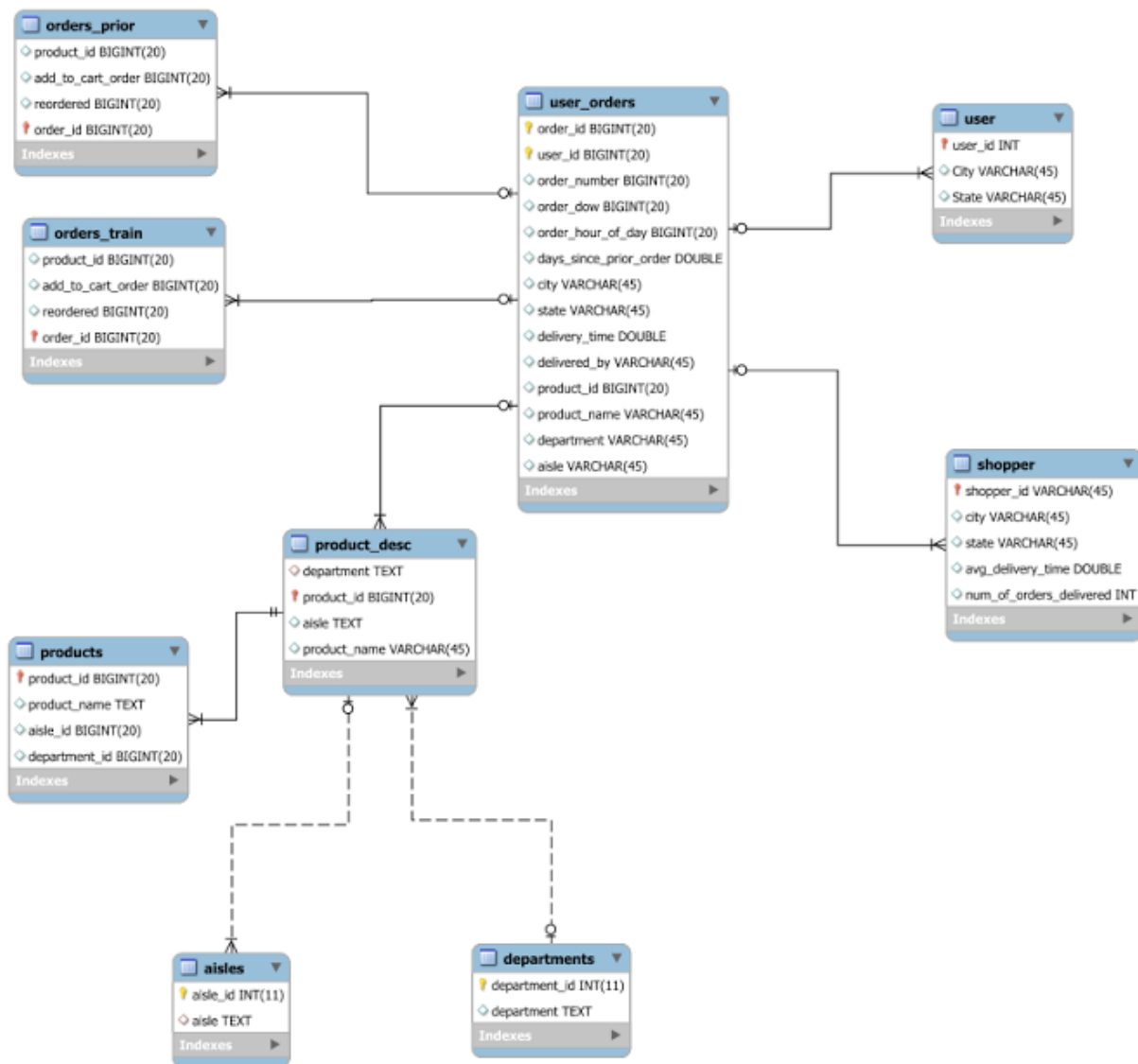
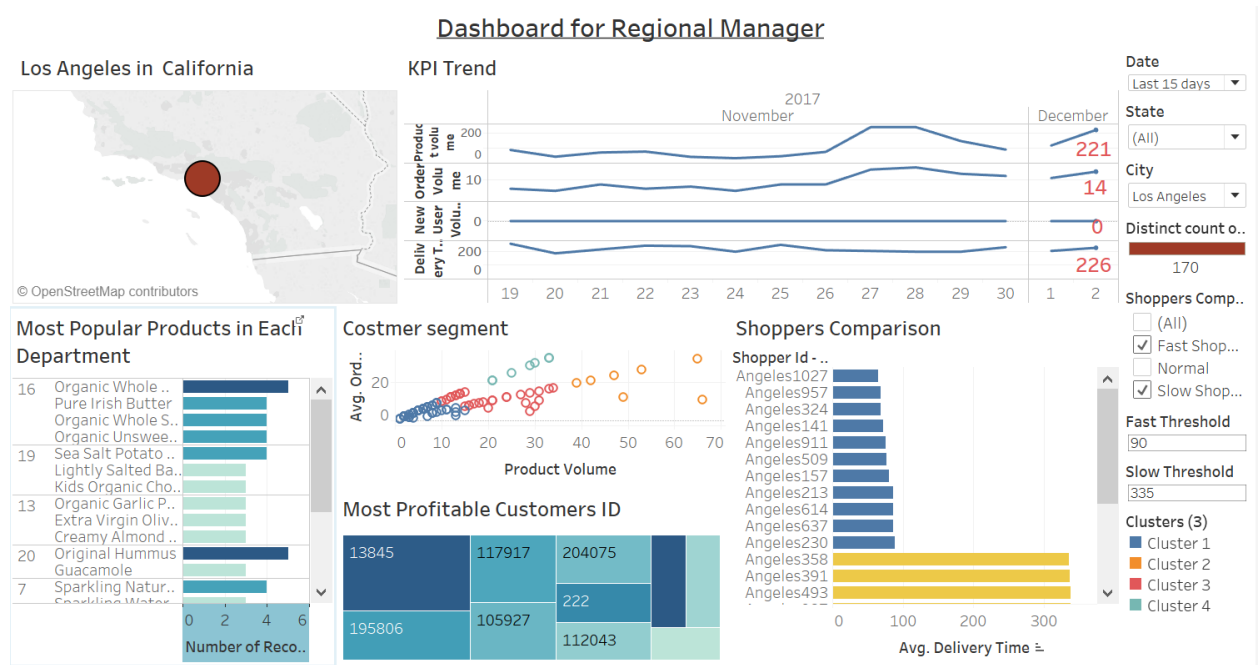


Figure: 3: InstaCart Star-schema for Data Warehouse.

## Three Use Cases/Prototypes

### 1. Tactical dashboard for Regional Manager



**Figure 4: Tactical Dashboard for Regional Manager.**

Our first dashboard is for the regional managers. The regional managers are for specific regions and they can focus on a specific city or state or a group of states. Once they identify their area of interest, they can also look for a specific time frame, like the past 15 days or a month or by quarters or years. We have used different KPIs in this dashboard and all of them are calculated using the selected timeframe range, which can

be a specific day, range of days, range of weeks, range of months or even a particular or range of years. The KPIs are listed below:

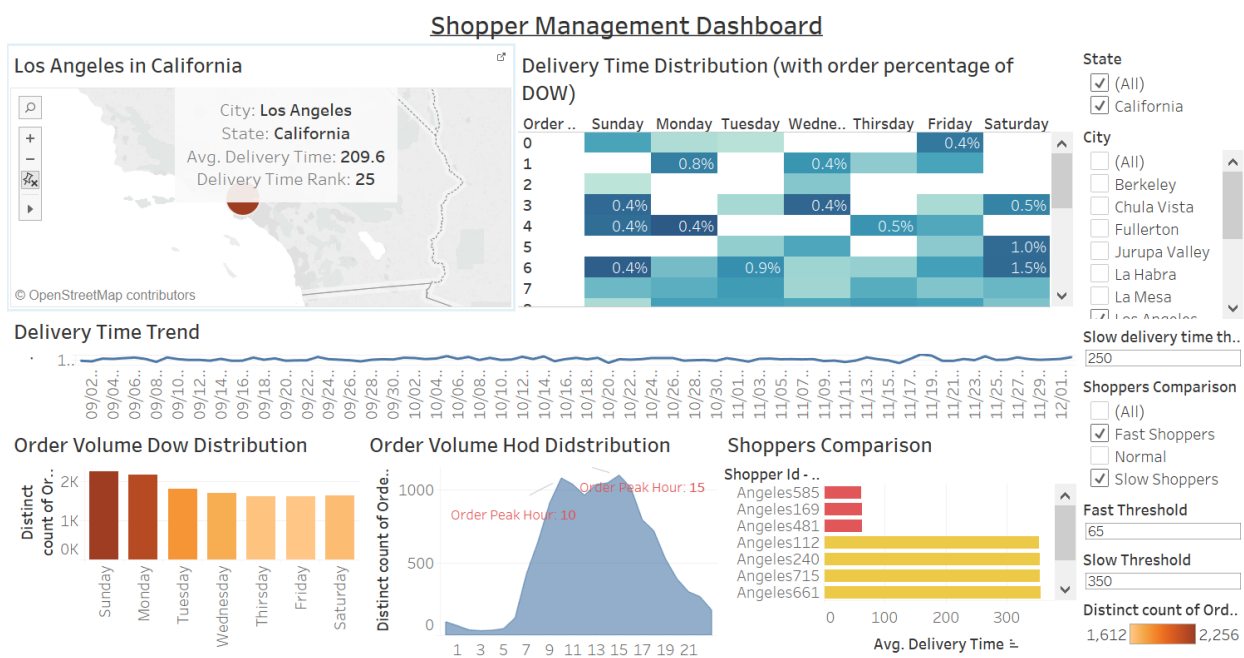
- 1) Product volume: The number of products purchased.
- 2) Order volume: The number of orders placed.
- 3) New User acquisition volume: In the given frame, how many new users have placed orders.
- 4) Average Delivery time: the average time a shopper takes to deliver an an order.

On Each KPI we have also have a current indicator which has a KPI value for today. For example, delivery time has an indicator value of 226, which means for today, the average delivery time is 226 minutes, and for that particular KPI we want it to be reduced as much as possible. From the dashboard, managers can see customer segments with most profitability in terms of how frequently they order and how many products they purchase. Because Instacart can have commissions on each purchase, the number of purchased products can be a proxy for how much the order can cost as we don't have the prices for each product as well in the given dataset. Not only managers can determine customer segments but they can also use the right below chart to see customers IDS of customers with most profitability. This would be helpful to make valuable customers get better service. Then, Instacart can get loyalty from these valuable customers.

Manager can assign a full time shopper or facilitate a direct communication between part-time shopper and customer to make sure the items are delivered accurately and quickly. The customer can also be provided with delivery fee discount or

coupon of products from stores. What's more, managers can use most popular products charts to make recommendations to the stores to ensure they have the product available in their inventory. In some stores, where we have Instacart's in-store shoppers, managers can direct them to arrange these items ahead of times because they are most likely to be included in each order for the customers.

## 2. Tactical dashboard for Shopper Manager



**Figure: 5: Shopper Management Dashboard.**

Our second dashboard is for Shopper Manager. Shoppers Managers are acting like HR Department in a typical organizational structure. They are monitoring the performance of shoppers and are in contact with shoppers directly. They are the ones who are responsible for finding out if some locations needs more resources or not. They can use this dashboard to first analyse the overall delivery times distributed on the different times of day and the different days of weeks. As seen in our previous

dashboard, this dashboard also has the capability to only choose a specific location but also can choose a specific time frame. The location filtering mechanism can be done either using the filtering parameters on the right side or by directly clicking on a specific location in the map.

'Delivery Time distribution' graph is very helpful to manager. The deeper the color is, the delivery time is longer. Manager should pay more attention to the deeper grids. The percentage number in the grids mean the order volume percentage in that one hour of that one day. The sum of one day's percentage number equals 100%. For easier interpretation, we just show the slow delivery hours' order percentage here. If the percentage is very small, managers can just ignore that hour. But if the number is large, managers should arrange more in store/part-time shopper in that hour.

Another component is the delivery trending line, which shows a clear indicator about the aggregated delivery times within a particular location. This can help managers to spot weird patterns and analyze the problem very efficiently. For instance, on a specific timeframe you may notice a significant high peak of average delivery time, and then managers can look at the number of assigned shoppers at the time to see if the problem happened because of a resource assignment issue or was it just because high traffic at that time.

Managers can look at the shopper comparison chart and see the performance of shoppers. This chart will indicate if there are shoppers who consistently have either fast or slow delivery times on average. Shopper manager can make an informed decision by judging their performance this way and can provide incentives to those who are meeting their delivery time SLA(service level agreement) and investigate more about the issues

for the other shoppers who have relatively slow delivery times. Moreover, one of the additional components we have in this dashboard are the two order volume charts on bottom left. Order volume indicates how many orders are placed given that day or more specifically given that hour of day. Using these charts, Shopper managers can assign resources accordingly and making sure they have shoppers on ground waiting for orders ahead of time to make proactive decisions ahead of time to reduce delivery times, and eventually gain customer satisfaction. To provide the user with easy-to-use filtering techniques we created two parameters on the right side based on user-preference, the manager will choose what time of deliveries is considered to be fast or slow. Based on the filters the charts will be generated and will show top slow/fast shoppers.

### 3. Forecasting

To personalize service for a customer, one important thing we do here is to predict the products in the next order for each customers, which is a recommendation system. With that, we can recommend and provide coupon of these specific products to each customer. Two ways to produce a list of recommendation are tried here, XGBoost and AutoEncoder.

XGBoost model is a kind of decision tree algorithm. Before setting up model, we create many dimensions like `user_reorder_ratio`, `product_reorder_times` based on the original data. Then I used these dimensions as predictor variables, re-order or not as response variables to train a XGBoost model. The training loss is 0.26 and accuracy is 91%.

Then we use the trained model to get a list of recommended products for each user.

The result is as follows:

```

[0]      train-rmse:0.465223
[10]     train-rmse:0.299993
[20]     train-rmse:0.270648
[30]     train-rmse:0.266006
[40]     train-rmse:0.265029
[50]     train-rmse:0.26465
[60]     train-rmse:0.264401
[70]     train-rmse:0.264196
[79]     train-rmse:0.263965

```

```

sum((val>0.50).astype(int)==y_val)/y_val.shape[0]
0.90981035623187811

```

**Figure 7:** *Training error and training accuracy for our predictive model.*

Then we use the trained model to get a list of recommended products for each user.

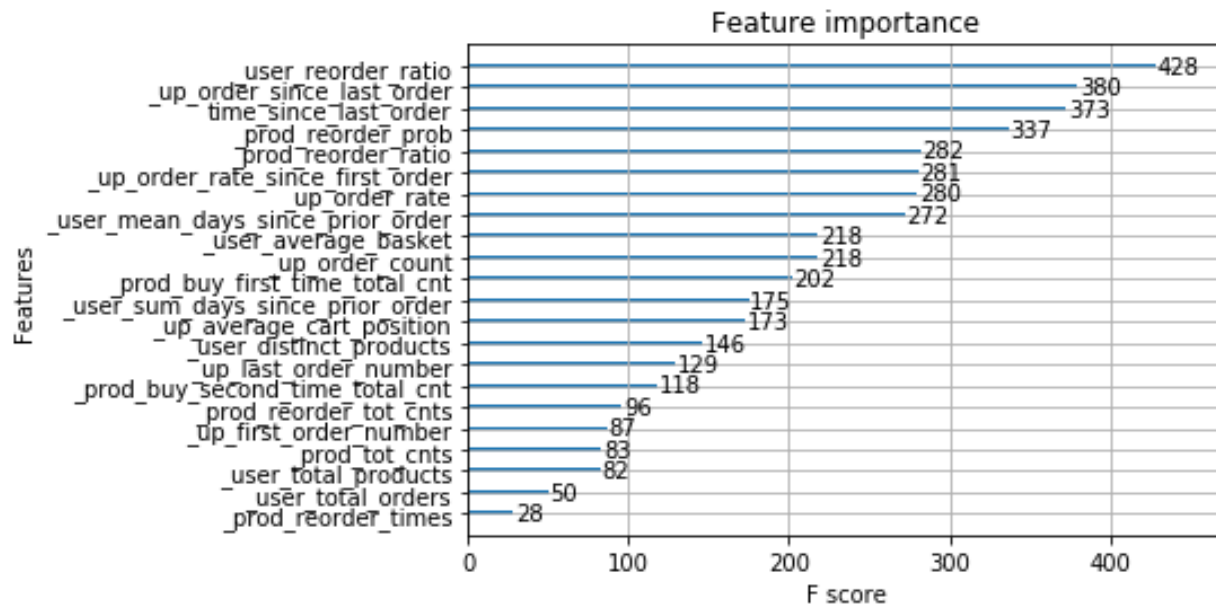
The result is as follows:

user_id	order_id	products
36855	17	None
35220	34	None
187107	137	None
115892	182	5479 39275 9337
35581	257	49235 24852
113359	313	46906 13198 21903 12779 45007
173814	353	None
55492	386	4920 47766 22124 40759 38281 39180 28985 45066 24852 15872 21479 42265
120775	414	20392
33565	418	None
78090	437	4589 27966
190409	452	40386 19370 5262 13166 36606 36313 27966 36735 19019 13176 47209 28666
119489	474	None
92343	492	46667 24852 44632 21137 43086
173962	497	31964 1831
152263	513	None
97952	517	21903 37646 24852

**Table 1:** List of recommended products for each user

According to this model, the most important dimensions to predict if a user will reorder a product or not are user\_reorder\_ratio, up\_order\_since\_last order (latest order number –

latest order number with include this specific product) and time\_since\_last\_order are the most important variables.



**Figure: 8:** Predictive model Feature Importance Plot.

The second method we used is Auto-Encoder, it is a supervised deep learning model. I used the original data to train a model directly and got 0.12 loss, which is less than XGBoost loss. The products recommended for each person is as follows:

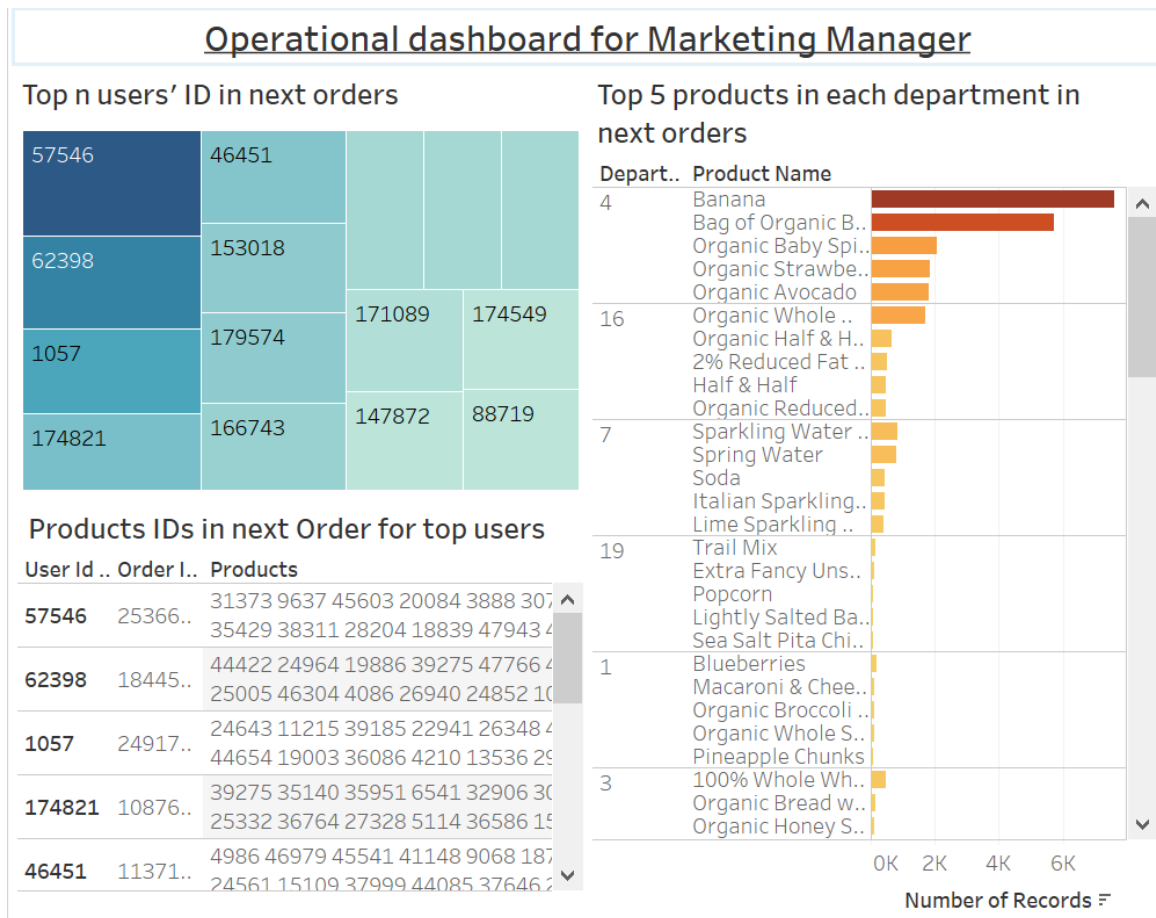
```
epoch: 72 loss: 0.133045987603
epoch: 73 loss: 0.139793299091
epoch: 74 loss: 0.131039042164
epoch: 75 loss: 0.137722507221
epoch: 76 loss: 0.129101792708
epoch: 77 loss: 0.135736489753
epoch: 78 loss: 0.12724634649
epoch: 79 loss: 0.133808062834
epoch: 80 loss: 0.125460800498
```



order_id	products
17	13107
34	2596 16083 21137 39180 39475 43504 44663 47766 47792
137	2326 5134 23794 24852 25890 38689 41787
182	5479 9337 11520 13629 32109 33000 39275 41149 47209 47672
257	4605 13870 21137 24838 24852 27104 27966 28476 29837 30233 37646 39475 45013 49235
313	12779 13198 14077 21903 25890 28535 30391 45007 46906 47626
353	35561 40688
386	260 2326 4920 6046 8174 10343 15872 17652 21479 21903 22124 24852 25513 27156 28985 30450 35221 36695 37935 38281 39180 40759 42265 45066 47766
414	4472 7539 14947 19006 20392 20564 21230 21376 21709 27705 27845 31215 31730 33320 36865 44292
418	1503 5262 13702 30489 38694 40268 41950 47766

**Figure: 8:***Predictive model 2 Auto Encoder: Training Loss & Output of predictions.*

With the recommendation system result, we designed an operational prediction dashboard for marketing manager. Top n users are acquired by adding the product volume in the next order. This graph in the dashboard can tell the marketing manager who will buy most products in near future. Then he can give these customers priority to reach shoppers and additional discount. 'Products in the next order for top users' can help marketing manager to decide which specific coupon or information to send. Marketing manager can provide 'Top 5 products in each department in next orders' to their stores, help stores on inventory and promotion.



**Figure: 9: Operational dashboard for Marketing Manager.**

## Implementation

Instacart is a technology driven company, but need to treasure data mining more in the future. It is facing very strong competitors such as Amazon. Instacart need to analyze and mine the data in its system more deeply, recommend products more accurately and personalize their service for customers, shoppers and stores. According to Kotter's 8-Step Change Model, Instacart should implement some steps.

- Create Urgency

Managers should communicate what's happening in the marketplace and with competitors. More and more players in the market, including not only startup, but also some large companies, are getting into this fresh delivery industry. This industry is getting more and more competitive. Instacart should tell their staff how strong competitors like amazon are doing and how they threaten instacart. These actions can urge company to pay more attention to use data and mine data and use them to get a competitive edge over others.

- Create vision to change

The values are central to change are: product delivery volume, order volume, new user volume and delivery time. Each region (city/state) should focus on these KPI and their trend. The strategies to improve these values include recognizing most important customers, personalizing service, train shoppers and communicate most popular products to customers and stores. Instacart should practice its "vision speech" often.

- Apply vision to all aspects of operations

Training: Instacart should train their staff on data analysis and data mining, at least how to use the results. Shoppers are part-time employees, but very important to the delivery time and customer satisfaction. Instacart usually gives simple training to shoppers. In the future, shopper managers should train them how to use shopper management dashboard and understand the order peak hour of each day and day of each week, shopper rank.

Marketing: When making the marketing plan, instacart can use the customer segmentation result and pay more attention to more profitable users, especially the top users in the next order. Instacart can give them more priority. These customer can get

shoppers faster than regular customers, discounted delivery fee. Instacart can even urge the stores to give extra coupon to these important customers.

Performance review: More importantly, instacart should set up an efficient system for shoppers performance review and rewards. For example, each month, reward money to fastest and most order delivery shoppers. Slowest shoppers should be trained more and order assignment could be adjusted based on the shoppers' delivery time to prefer shoppers who deliver faster, which could motivate the shoppers to deliver faster to get more orders.

- Create Short-Term Wins

Products forecasting and recommendation system is key technology to Instacart. Instacart should review the forecasting accuracy and recommendation effect periodically. Once the accuracy improved, instacart should reward the relative team and staffs. Every region can be ranked according to KPI each month, the best growth regions should be rewarded too.

## Summary

## and

## Conclusion

Instacart could address three major problems including managing local resources, reducing delivery time and providing personalized experience to customers.

First, we created tactical dashboard for regional managers. Regional managers are responsible for specific areas. By using this dashboard, they could focus on their own cities and identify their area of interests. They could see customer segments with most profitability, which would be helpful to make valuable customers receive better service. Second, we built another tactical dashboard for shopper manager. Shopper managers are responsible for monitoring the performance of shoppers, contacting with

shoppers directly and managing resources in different locations. They could use this dashboard to analyze delivery time on different times in different locations. Also this dashboard would give the information about which shopper are doing great and which are not. Using this tactical dashboard shopper managers can assign resources accordingly to gain customer satisfaction. When we designed dashboards, we found it difficult to collect real data from Instacart. To address it, we collected data from other online grocery stores and also randomly created some data to finish our dashboard.

To provide personalized service, we developed a recommendation system to predict and then recommend products to each customer. According our prediction model, the most important dimensions are the products the customers bought in previous orders and how long it has been since their last order. We generated an operational prediction dashboard to tell the marketing manager, who will buy the most products, in the near future, so that customers could receive discounts before they place their next order. Also, using this operational dashboard could help managers to decide which coupon or information flyer should be sent to customers and help stores update the inventory to meet the expected demand.