Instacart Recommender System

Aaron Xiao, Kiran Sidhu, Paul Flemming, Rozi Hagos, Tara Singh,

Agenda

instacart

- Overview
  - Business Objectives & Problem Formulation
- Methodology
  - Preliminary Analysis, Data Manipulation, Descriptive Analysi
  - Model Development & Model Evaluation
  - Insights and Results Summary
- Recommendations
- Performance Management
- Cost Benefit Analysis
- Conclusion

Business Problem | Analytical Objectives | Data Preparation | Model Development | Solution Deployment | Performance Management

Overview

instacart

Overview of the problem

Who are Instacart's customers?

What are their needs?

How will the customers use the solution?

What will success look like?

Business Objectives

- Increase Profit by maximizing the chance of finding the right products in a customer's cart and minimize the time to complete the transaction.
- Increase user engagement, satisfaction, and loyalty over time.

Objectives will be met by building a recommender system for Instacart that will recommend products that the consumer will be more inclined to try and purchase based on their previous order history.

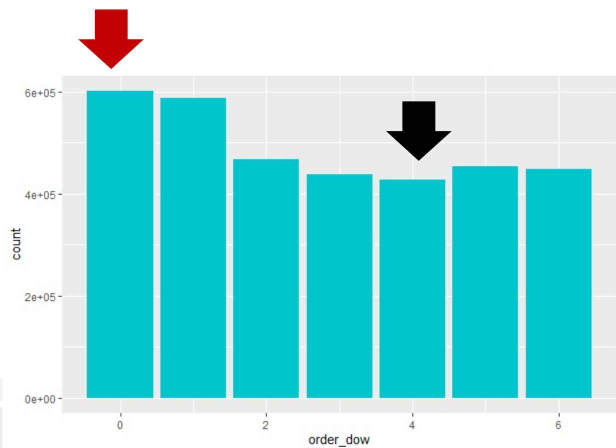
Problem Formulation

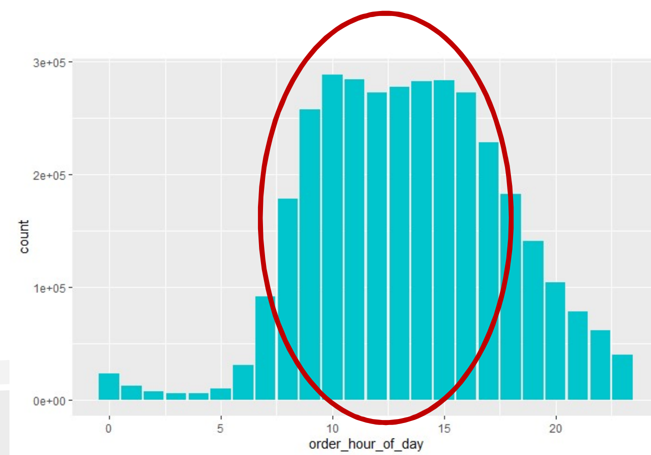
- The online grocery shopping market is expected to grow 500% over the next decade.
- In 2017, the customer average spend was $100 – 120 per order
- Even with quick expansion, Instacart is not yet profitable

Value Proposition:

Methodology

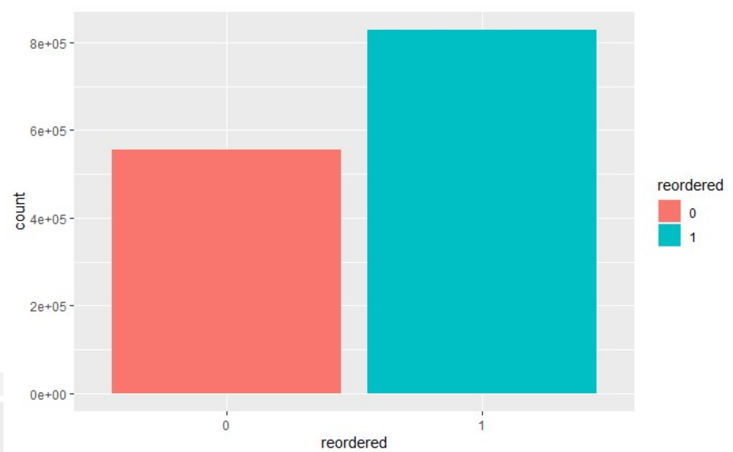
- The dataset contains samples of over 3 million grocery orders from more than 200,000 Instacart customers. The dataset contains 4 to 100 orders from each customer, with the add-to-cart sequence of products purchased in each order
- The week, hour of day the order was placed, and a relative measure of time between orders are also included in the datasets

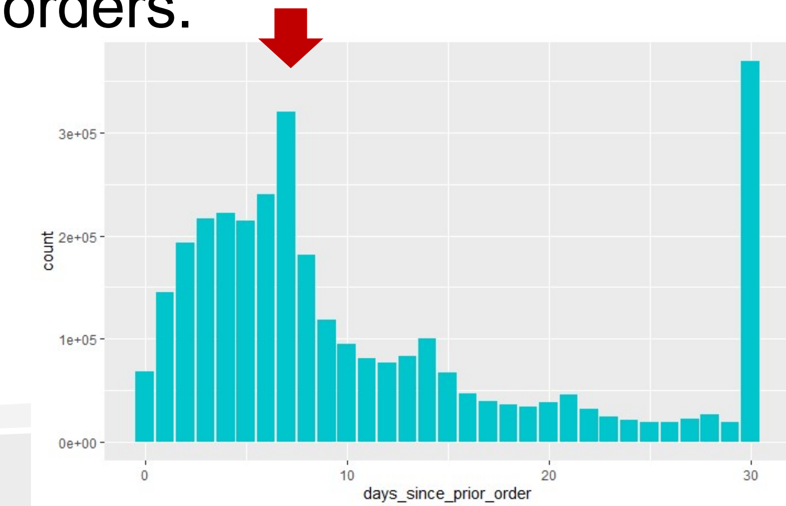




Preliminary Analysis

- Approximately 60% of the orders were from customers who have previously shopped on Instacart.
- We discovered that customers tended to reorder at around the seventh day after their initial order and they normally had at least 3 prior orders.

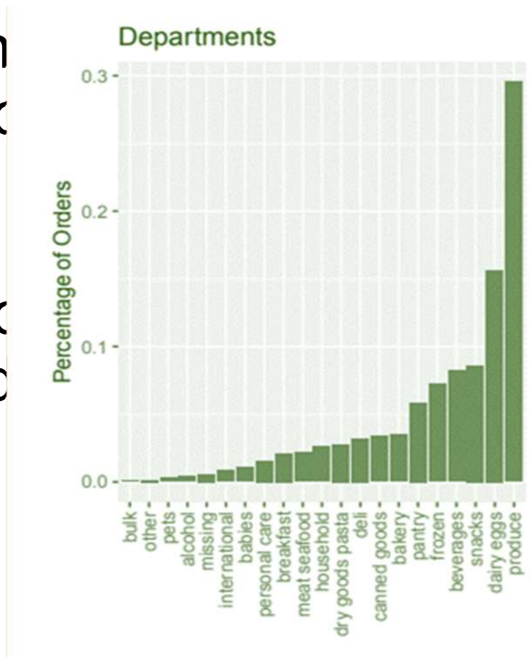


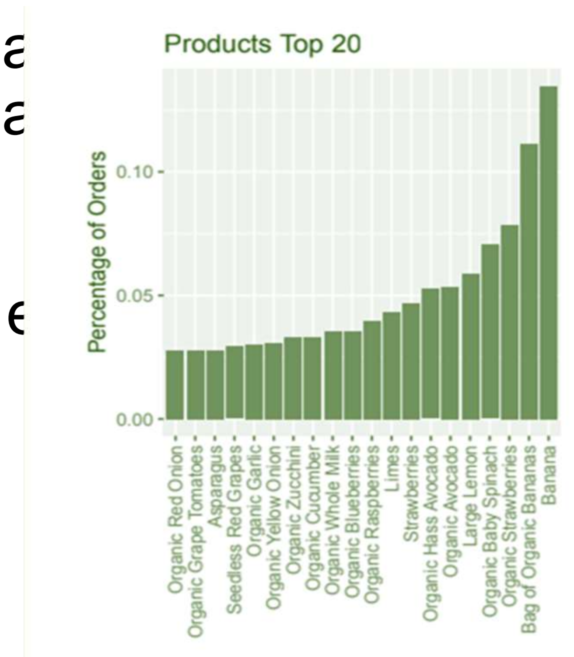


Descriptive Analysis

- Fresh produce was th a
  30% of the orders, follo a

  - 20 most popular pro e
    c



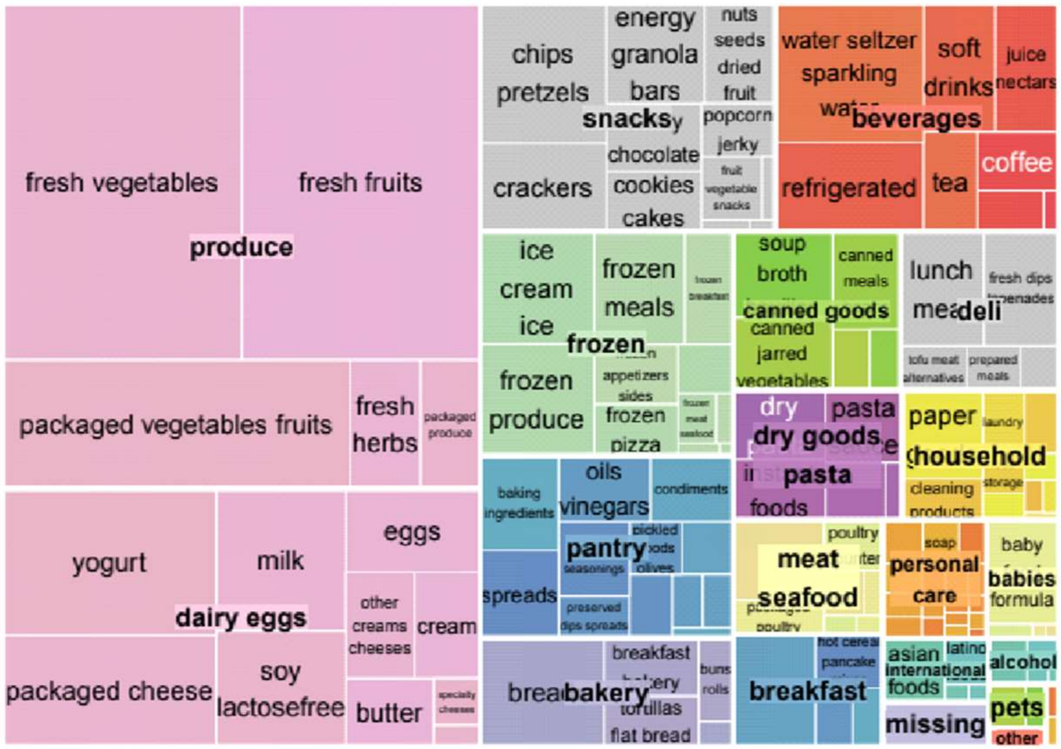


Descriptive Analysis

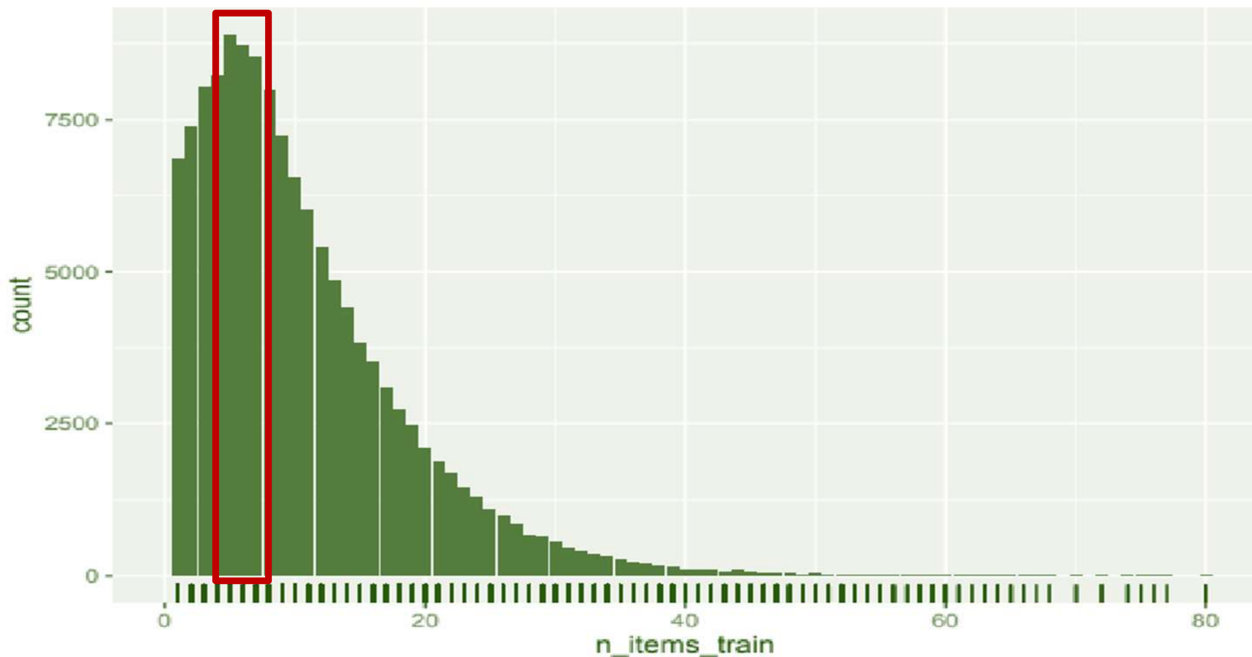
- Top 10

Descriptive Analysis

instacart



Data Manipulation



Model Development

- Use R recommenderlab package
- Total 5 algorithms have been used:
  - Association Rules
  - Random
  - Popular
  - Item Based Collaborative Filtering (IBCF)
  - User Based Collaborative Filtering (UBCF)
- There are 1.3M+ records cover total 49,000+ products.
- Consider the computing capability, the dataset is filtered scientifically to obtain a reasonable processing time.

	Item_1	Item_2	Item_3	...	Item_n-2	Item_n-1	Item_n
user_1	0	1	0		0	0	1
user_2	0	0	0		1	0	0
user_3	1	0	0		0	1	0
user_4	0	0	1		0	0	0
:							
:							
:							
user_n-3	0	1	0		1	0	0
user_n-2	1	0	0		0	0	1
user_n-1	1	0	0		1	0	0
user_n	0	0	1		0	1	0

47897 x 27846 rating matrix of class 'binaryRatingMatrix' with 350857 ratings.

Business Problem | Analytical Objectives | Data Preparation | **Model Development** | Solution Deployment | Performance Management

Model Development

instacart

```
scheme <- ratings_matrix %>%
  evaluationScheme(
        method = "split",
     k      = 2,
     train  = 0.8,
     given  = -1)
```

- Split the data by selecting train = 0.8 for a 80/20 train/test split.
- Set method = “split” and k = 2 for 2 runs. The results can then be averaged to produce a single evaluation set.
- Selecting given = -1 means that for the test users ‘all but 1’ randomly selected item is withheld for evaluation.

include the **random items** algorithm for benchmarking purposes.

```
algorithms <- list(
  "association rules" = list(name  = "AR", param = list(supp = 0.01, conf = 0.01)),
  "random items"      = list(name  = "RANDOM",  param = NULL),
  "popular items"     = list(name  = "POPULAR", param = NULL),
  "item-based CF"    = list(name  = "IBCF", param = list(k = 2)),
  "user-based CF"    = list(name  = "UBCF", param = list(method = "Cosine", nn = 20))
)
```

```
results <- recommenderlab::evaluate(
        scheme,
        algorithms,
        type  = "topNList",
        n  = c(3, 5, 10, 15)
)
```

- Select *type* = *topNList* to evaluate a Top N List of product recommendations
- Specify how many recommendations to calculate with the parameter *n* = *c(3, 5, 10, 15)*.

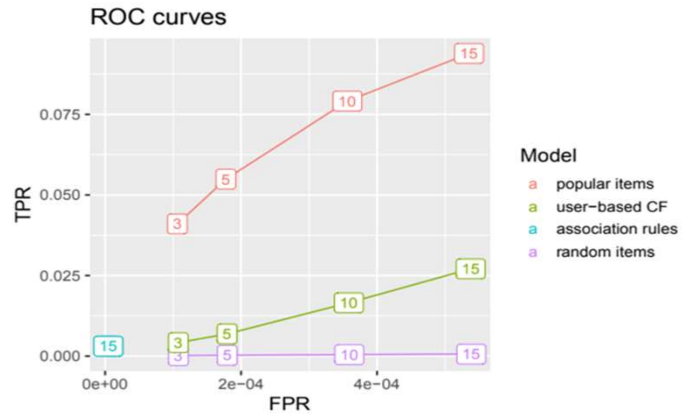
Model Evaluation

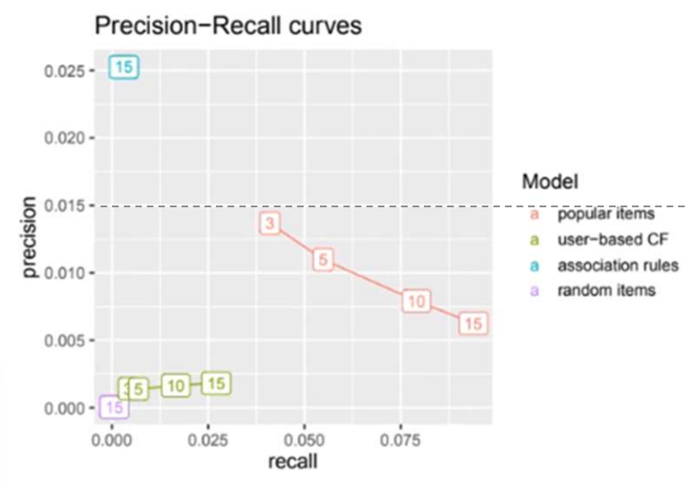
Model Performance Ranking:

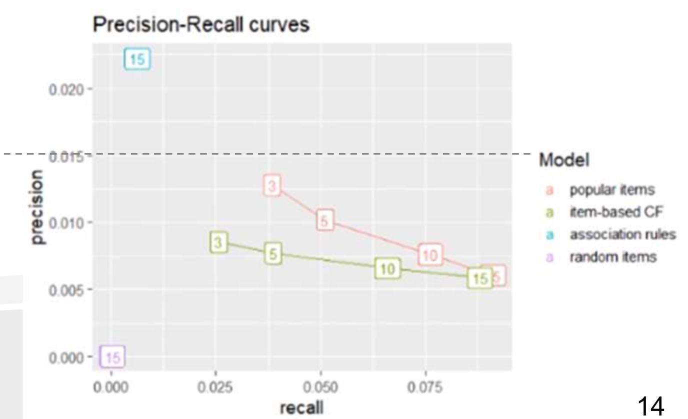
1. **popular items model**
2. **item-based CF model**
3. **user-based CF model**
4. **association rules model**
5. **random items model**

The item-based CF model is close to the popular items model. Both of them show significant higher TPRs for any given level of FPRs than other models.

This means that these 2 models are producing higher number of relevant recommendations (true positives) for the same level of non-relevant recommendations (false positives).







Model Evaluation

instacart

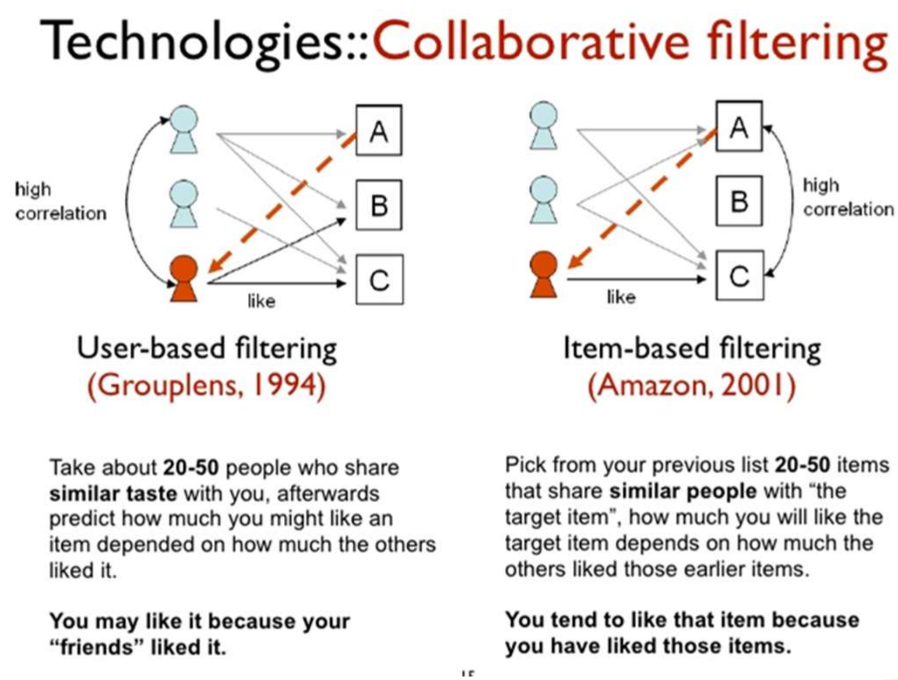
We randomly selected 5 items as a historical order from a customer, use it as the input to check our model to see if it provides something that sounds reasonable.

```
##     product_id                        product_name
##          <int>                              <fctr>
## 1:       39821 S.O.S Reusable Steel Wool Soap Pads
## 2:       31562                        Sweet Onions
## 3:        6844        Organic Unsulphured Molasses
## 4:       30960          Regular Pork Sausage Tube
## 5:       13914          Cheez-It Baked Snack Crackers
```

```
##     product_id                        product_name
##          <int>                              <fctr>
## 1:        1215 Kidz All Natural Baked Chicken Nuggets
## 2:        1940            Organic 2% Reduced Fat Milk
## 3:        4037                  Blackberry Preserves
## 4:        4658                Imported Mineral Water
## 5:        4913                  Table Water Crackers
```

Insights and Results Summary

- Popular items model seems to have the best results, but only be useful to suggest items for new customers.
- IBCF model would be our suggested recommender model base on the current data and processing capability. It looks for similar items based on the items users have already purchased.
- UBCF model would be another good choice if the similarity between users could be identified based on more historical orders. It recommends items by finding similar users to the targeted user.
- Ideally, the best model is a hybrid one that combines IBCF and UBCF. So the system can provide effective proposal of new items to a customer based on his own purchasing habit and the others with similar taste.
- Cold start for new users and new products



Recommendations

instacart

- Focus on personalization and improving user experience
- Focus on the strengths of buying online
- Customize the online experience
- Tailor the experience to individual needs
- Promote sale items, niche products or less frequently purchased items

Performance Management

- Model in motion
- Dataset constantly increasing
- Recommender model needs updating weekly for best results
- Regular updating will provide continuously refreshed and aligned recommendations for customers

Cost Benefit Analysis

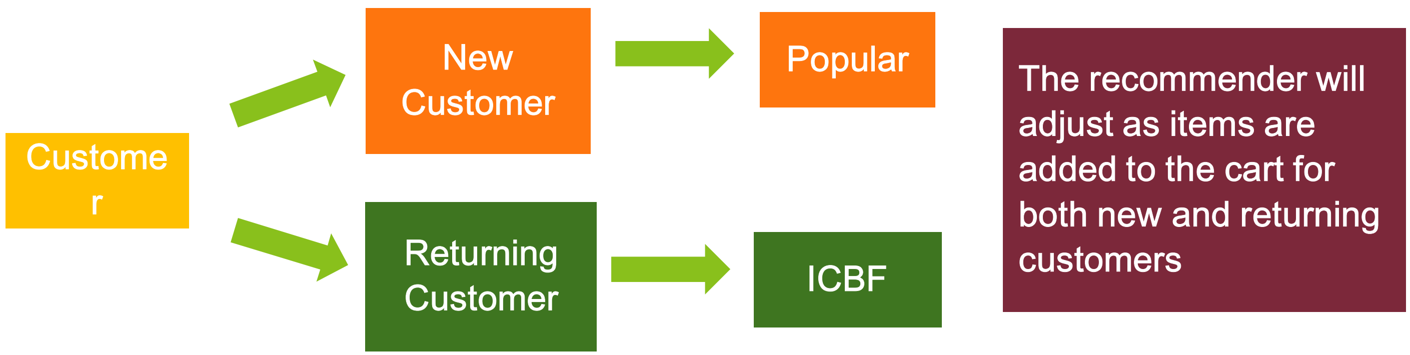
- Generally data most expensive part of data analytics
- Minimal cost for data acquisition, customers produce data
- Capital costs include storage array, model development, testing and evaluation
- Working costs data maintenance, processing, and data governance
- Working costs contribute to increased sales, greater revenue and profits as well as increased customer loyalty.
- As their customer base increases, there would also be an incremental increase in the above noted benefits
- Customer average spend of $100 – 120 per order.

Conclusion

- The business objective was to build a recommender system for Instacart
- The goal is to increase user engagement, satisfaction, and loyalty over time
- Popular items model was the best performing model
  - Fills in the gap for new customers
- IBCF tailored suggestions to customers cart contents
  - Creates a more personalized customer experience

Conclusion

instacart



- This strategy aligns with the business objectives and goals for future business growth and expansion instacart


Thank You