Personalized Recommendation System For instacart

Presented by Wisdom Chen, Boya Zeng, Jessy Hu, Yuqing Wu

Wisdom Chen

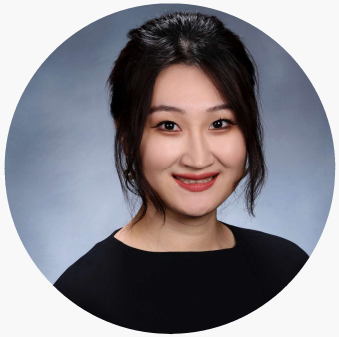
Data Scientist

Boya Zeng

Machine Learning Scientist

Jessy Hu

Business Intelligence Specialist



Yuqing Wu

Big Data Analyst

Our Mission

Translate complex data into actionable strategies, driving success and competitive advantage for our clients

OUR GOAL: LEVERAGING DATA ANALYTICS TO DRIVE PERSONALIZED MARKETING STRATEGIES IN ONLINE GROCERY SHOPPING

Situation

In the highly competitive online grocery delivery sector, businesses like Instacart strive to **understand customer behaviors, preferences, and segmentation to enhance service personalization and operational efficiency.**

Task

We aimed to dissect the extensive Instacart dataset available on Kaggle to **uncover actionable insights on customer purchasing behaviors.**

Action

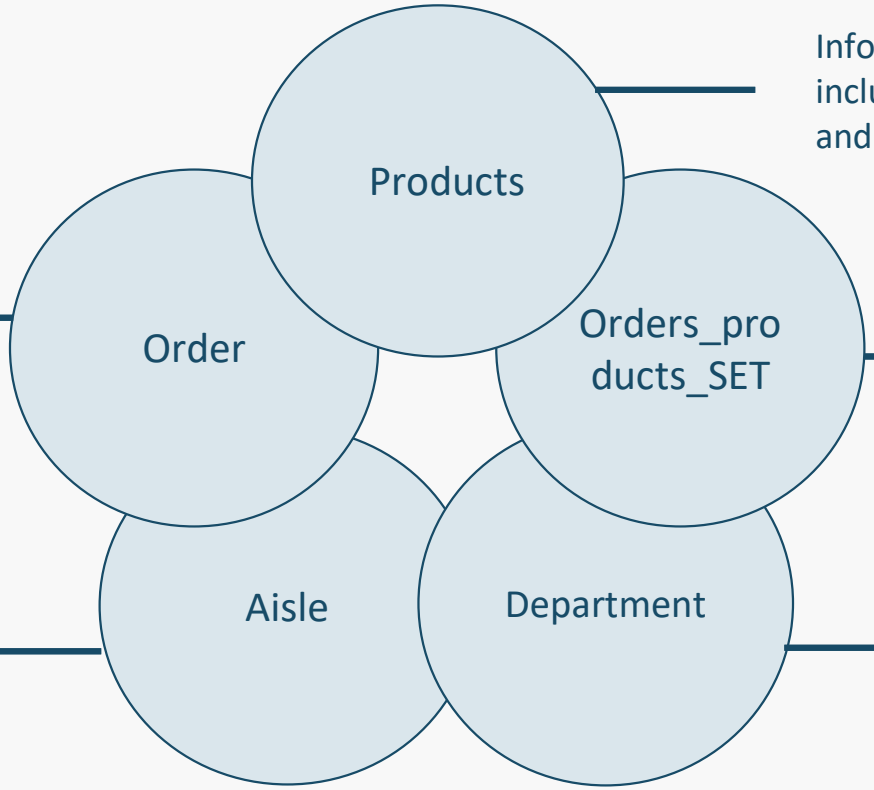
We will analyze **purchasing behaviors**, **segment customers**, perform **market basket analysis**, develop a **recommendation system**, and create a recommendation **dashboard**

Result

The recommendation system and dashboard facilitated data-driven decisions, setting the stage for a personalized marketing strategy to enhance customer engagement and boost sales.

DATA OVERVIEW: JOINED DATASET ENSURES QUALITY WITH OVER 3 MILLION RECORDS ACROSS 29 FIELDS CAPTURING COMPREHENSIVE INSTACART USER ACTIVITY

- ✓ Accuracy
- ✓ Completeness
- ✓ Reliability
- ✓ Relevance
- ✓ Timeliness



Information about each product, including its name, associated aisle and department

Information about each order associated User ID and order Time

Products in each order and whether the product was reordered

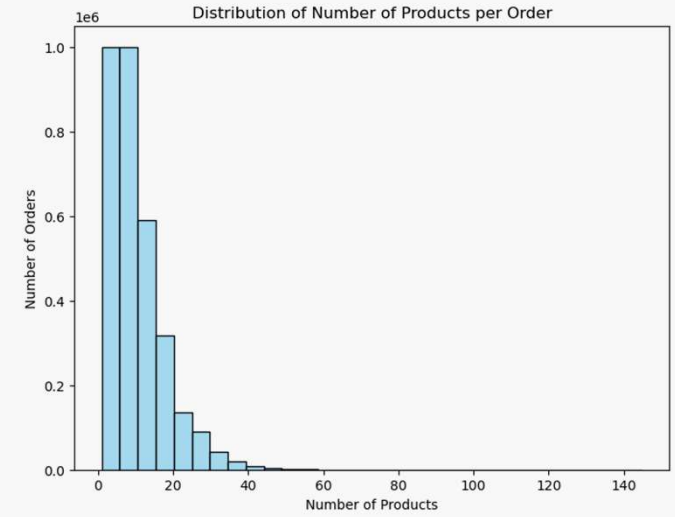
Each aisle's ID and name

Each department's ID and name

Joined data would have 3 Million + records and 29 fields in total，Including information of 200K Instacart user

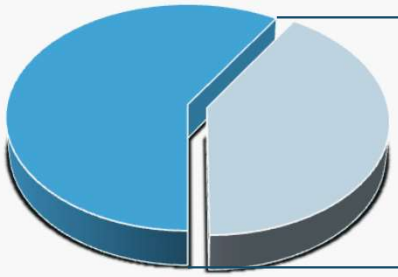
DATA EXPLORATION REVEALS 59% REORDER RATE; CUSTOMERS COMMONLY ORDER 20 ITEMS, PREFERRING FOOD AND PERSONAL CARE PRODUCTS

Product Portfolio



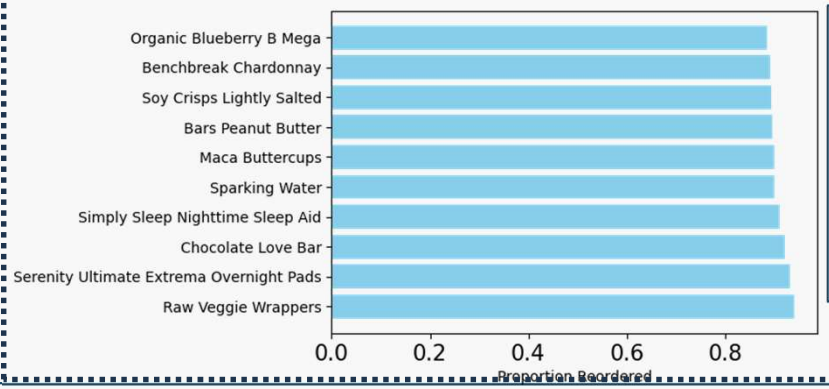
Customer typically purchase **20 products** within a single order

Reorder Distribution



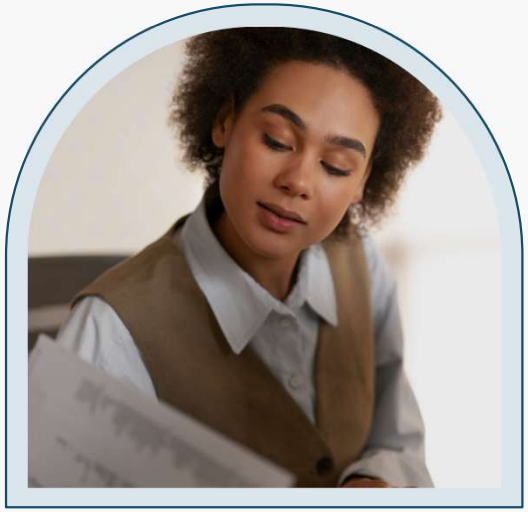
41% of the customer did not reorder since the initial purchase

Reordered Did Nor Reordered



Food and Personal Care are most frequently reordered by user

DATA EXPLORATION OF CUSTOMER'S PURCHASE BEHAVIOR



Name Anna Wilson

Age 22 years old

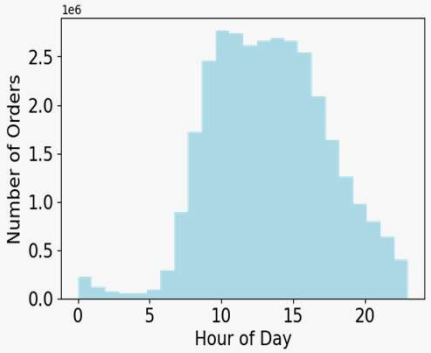
Profile Loyal Customer

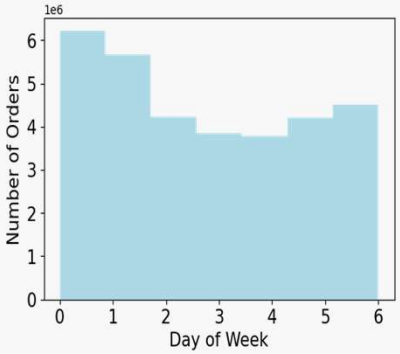
Trait 1 peak activity around midday

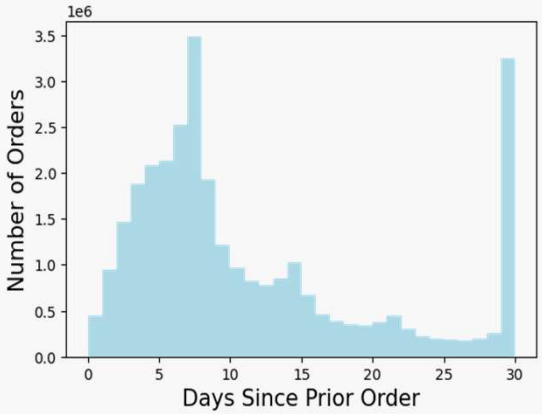
Trait 2 stocking up early in the week

Trait 3 daily or monthly replenishment.

Trait 4 routine and bulk purchasing behaviors







Customers ordering peaks midday and early in the week, with frequent same-day reorders or monthly patterns, while individual habits, like Anna's, vary from regular small purchases to intermittent bulk orders.

IDENTIFY DISTINCT **CUSTOMER SEGMENTS USING K-MEANS CLUSTERING** BASED ON THEIR SHOPPING PATTERNS TO TAILOR FURTHUR ANALYSIS

Data Preparation

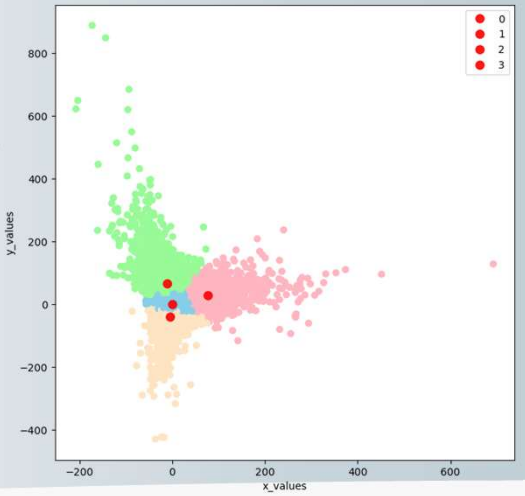
Aggregated purchase data by customer and aisle using cross tab

Aisle / User ID	Baby food formula	Yogurt	Tea	...
1	0	3	0	...
2	1	0	0	...
...	...	...	...	...
134	12	0	0	...

E.g., User #134 purchased 12 times of baby food formula

Segment Customer with K-Means

- Performed **PCA** for dimension reduction
- Utilized **k-means clustering** to segment customers into groups with similar aisle preferences.
- Determined optimal cluster number through **Elbow Method** and **Silhouette Score analysis**

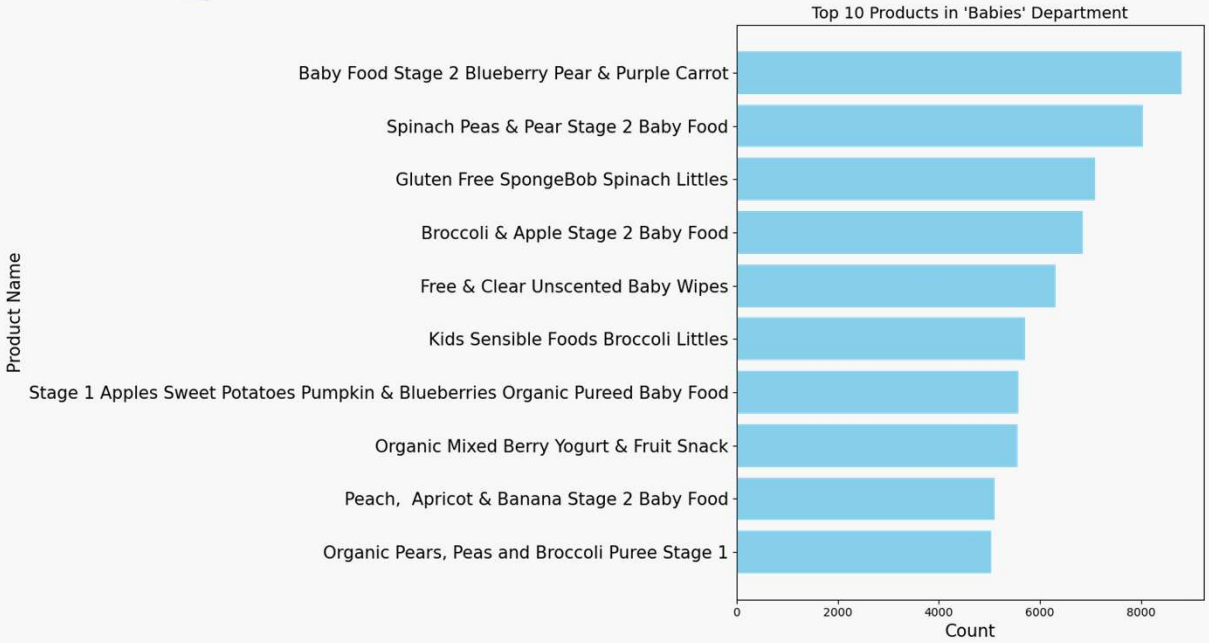


Distinguish Segment of Interest

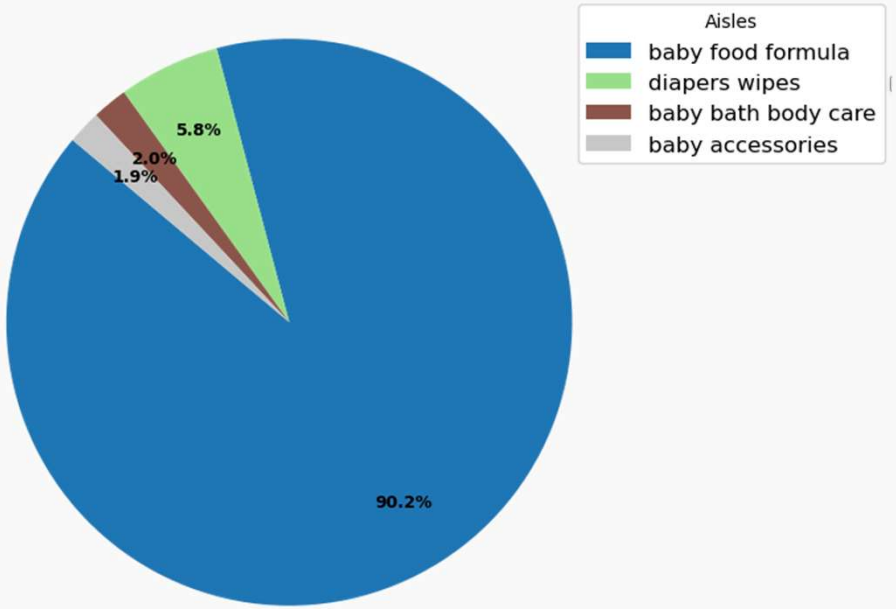
- 3 out of 4 clusters primarily purchase fresh fruits/vegetables
- **while 1 distinct cluster mainly buys baby food formula**.

Cluster 1	Cluster 2	Cluster 3	Cluster 4
Fresh veggie	Fresh fruit	Fresh fruit	**Baby food formula**
Fresh fruit	Fresh veggie	Yogurt	Fresh fruit
...	...	...	...

TAKE A CLOSER LOOK INTO BABIES DEPARTMENT



Top10 Products in Department of Babies



Distribution of Products by Aisle

Next Step：Identify patterns of product purchases and discover associations between different products.

PERFORM **MARKET BASKET ANALYSIS USING APRIORI ALGORITHM** TO IDENTIGY ASSOCIATION BETWEEN ITEMS

Association Rule Mining with Apriori

- Transform order data of cluster 4 to list of transactions with one-hot encoding

	Coffee Filters	Organic Milk	Greek Yogurt	...
0	True	True	False	...
1	False	True	False	...
2	False	False	False	...
...	...	...	...	...

- Set minimum threshold: 0.01 for support, 1 for lift

	Antecedents	Consequents	Support	Confidence	Lift
0	Whole Wheat Bread	Organic Bananas	0.01	0.30	1.51
1	Organic Bananas	Apple Honeycrisp Organic	0.02	0.08	1.62
...	...	...	...	...	...
193	Organic Hass Avocado	Organic Bananas	0.05	0.39	1.97

Item Sets Example

Support	0.01
Confidence	0.50
Lift	2.52

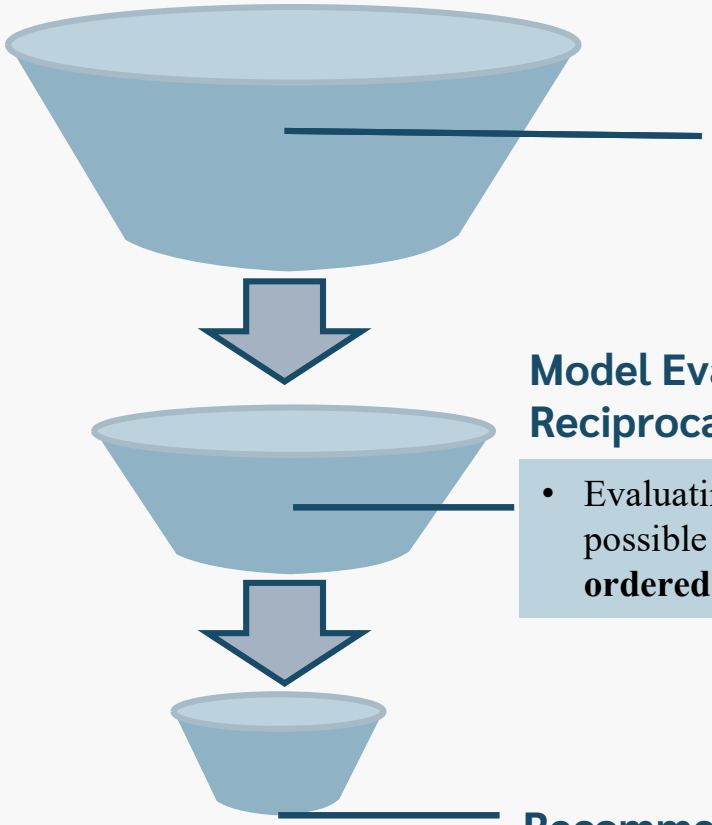
- High confidence
- Indicates these two items are often bought together

Support	0.01
Confidence	0.29
Lift	15.89

- High lift
- Indicates if A is sold the ratio of sale of B will increase more than 15 times.

BUILD **HYBRID RECOMMENDATION SYSTEM** AND EVALUATE IT USING MMR

SHOWCASE OUR RECOMMENDATION DASHBOARD



Hybrid Recommendation Systems

- **Collaborative Filtering:** Historical of reordered product
- **Content-Based:** Recommend items that are similar in content which are the users' choice

Model Evaluation -- Mean Reciprocal Rank (MRR)

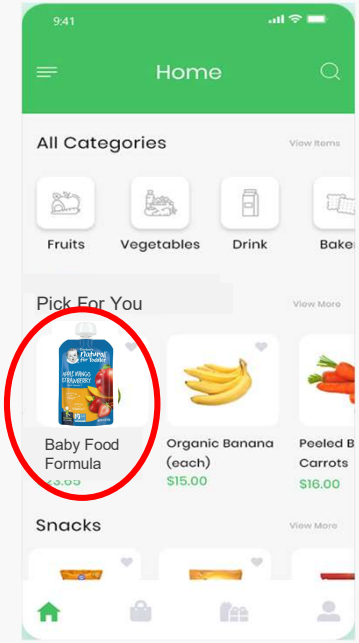
- Evaluating any process that produces a list of possible responses to a sample of queries, **ordered by probability of correctness**

Recommendation Dashboard

MRR (Apple)	
Items	**Scores**
Pink Lady Apple	0.78
Apple Cinnamon Fig Bar	0.62
Organics Vitamin C Apple Juice	0.51
Apple Apple Applesauce On The Go Pouches	0.49
Braeburn Apple	0.47

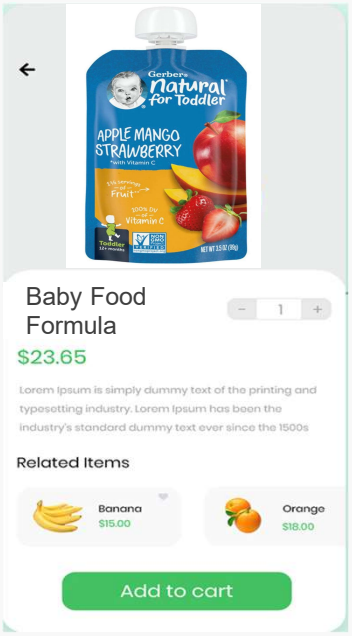
TAILORED **MARKETING STRATEGY** BASED ON OUR MODELING ENABLING POTENTIAL **BUSINESS VALUE**

Homepage Promotion



- ✓ Engage visitors from the moment they land on the homepage by using Recommendation System

Add-to-Cart Recommendation



- ✓ Maximize the opportunity for additional sales using Market Basket Analysis

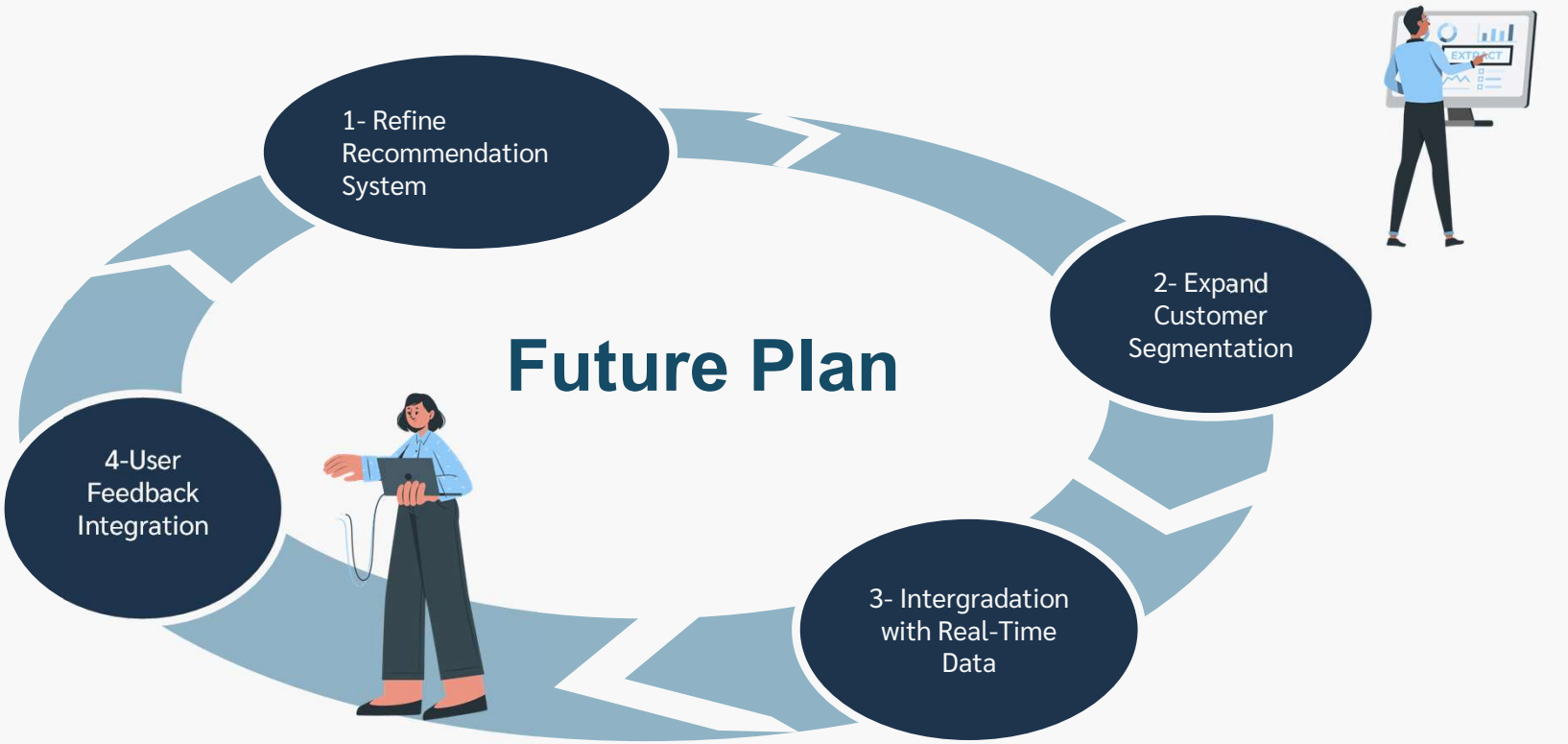
Business Value

- Enhanced Customer Segmentation
- Personalized Shopping Experience
- Efficient Marketing and Promotions

YC0

OUR **FUTURE PLAN** STAYS TRUE TO FLEXIBLE AGILE PPRINCIPLE WITH 4 STEPS

ALLOWS CONTINUOUS FEEDBACK AND ITERATIVE DEVELOPMENT



Slide 12

YC0 Business values + Marketing Strategy = 1 Slide

Wisdom Chen, 2024-03-02T20:19:56.119

Thank You

Any questions?

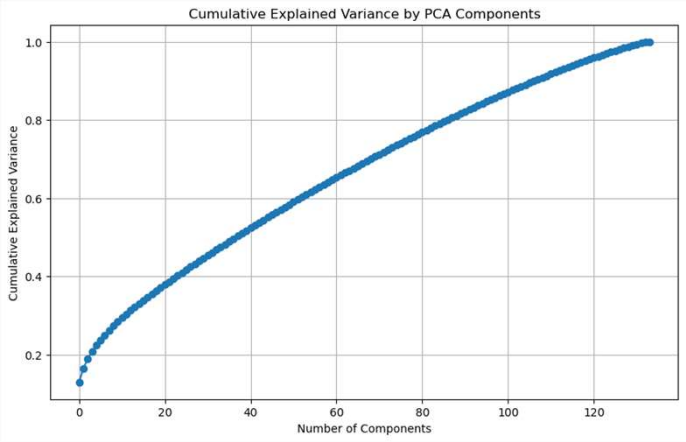
Team Contribution
EDA : Wisdom Chen, Boya Zeng, Jessy Hu
Customer Segment: Jessy Hu, Yuqing Wu
Recommendation System: Wisdom Chen
Market Basket Analysis: Yuqing Wu, Boya Zeng
Recommendation System Video Demo: Boya Zeng
Slides: All

Appendix

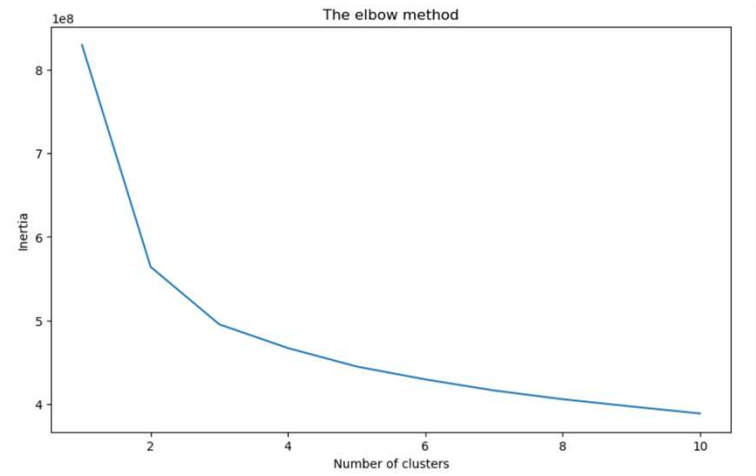
Dashboard Demo Video

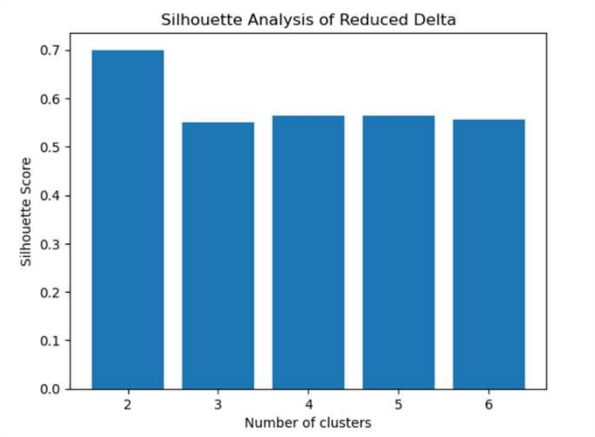
Back

PCA & Elbow Method &Silhouette Score Analysis



Selecting the number of components=78 to explain at least 75% of the variance





We will choose k=4 here. although k=2 seems to be the optimal, but there will be less information. k=5 will have the second largest silhouette score, but after we plot it out, we see one of the cluster is too small to compare. therefore, we ultimately chose k=4