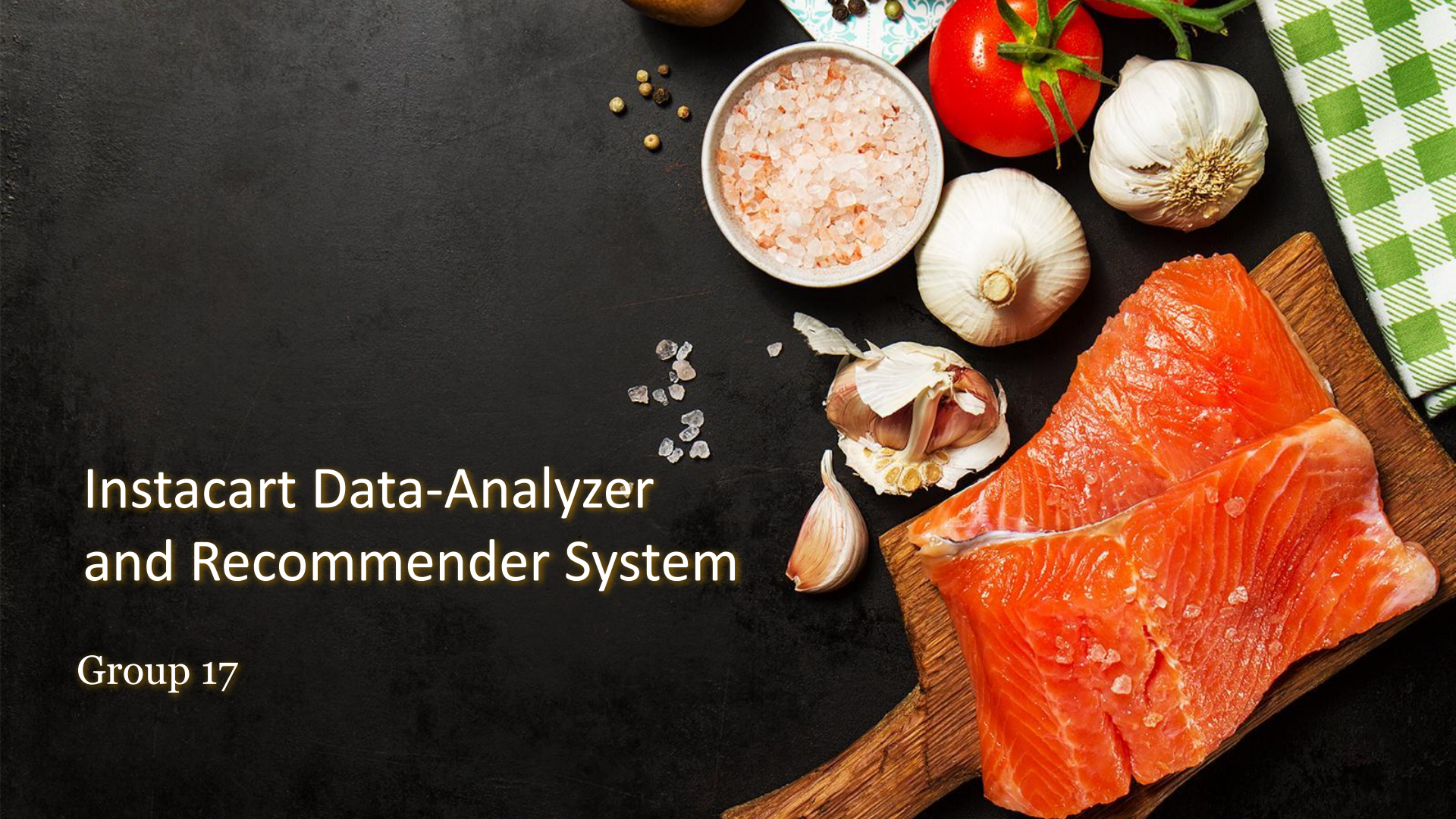


Instacart Data-Analyzer and Recommender System

Group 17

Overview

Dataset

Data Exploration

Recommender System

Dataset

Data Exploration

Recommender System

Instacart Market Basket Dataset

Orders
User: Who placed the order
Time: When the order was placed

Products
Aisle and department of the product

Order-Product
Which products were ordered
How many products were reordered
Order of placement in cart

Dataset

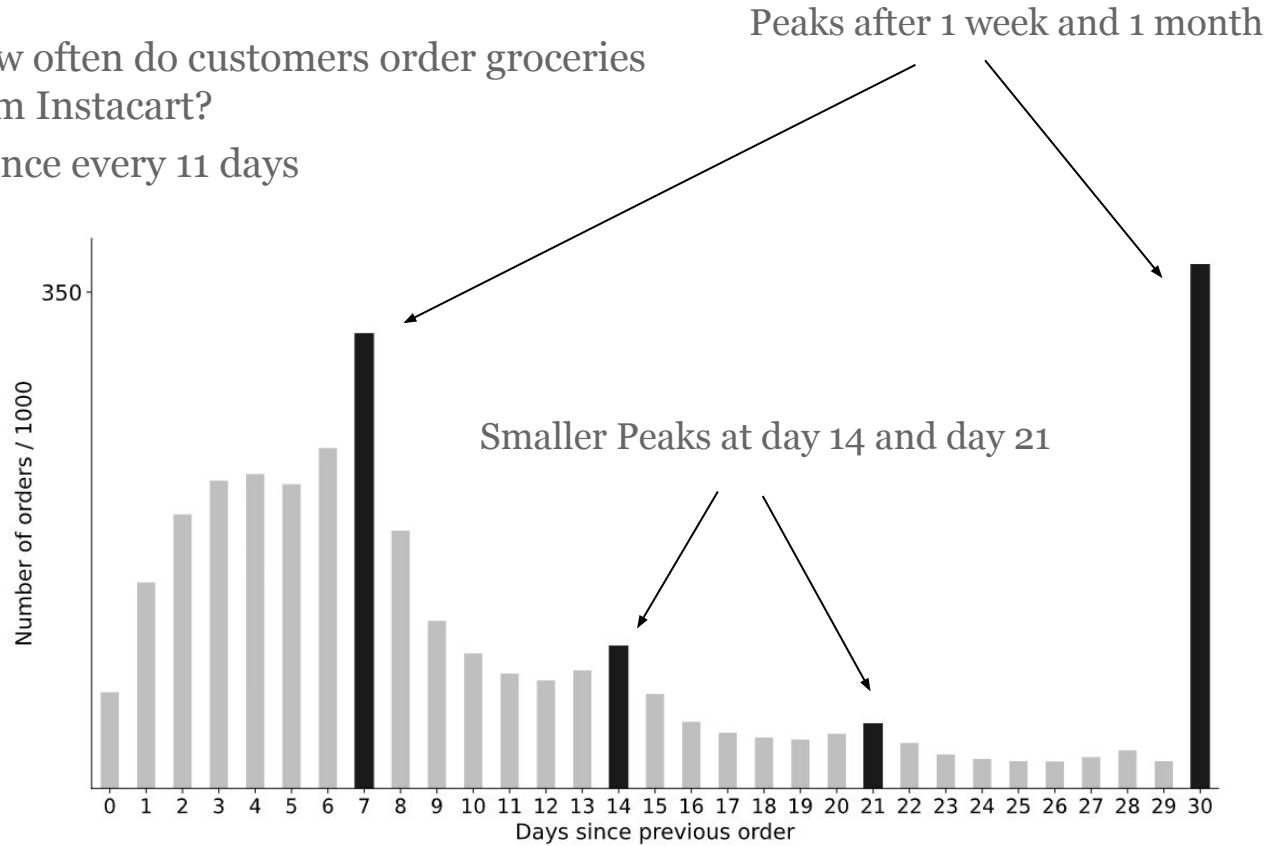
Data Exploration

Recommender System

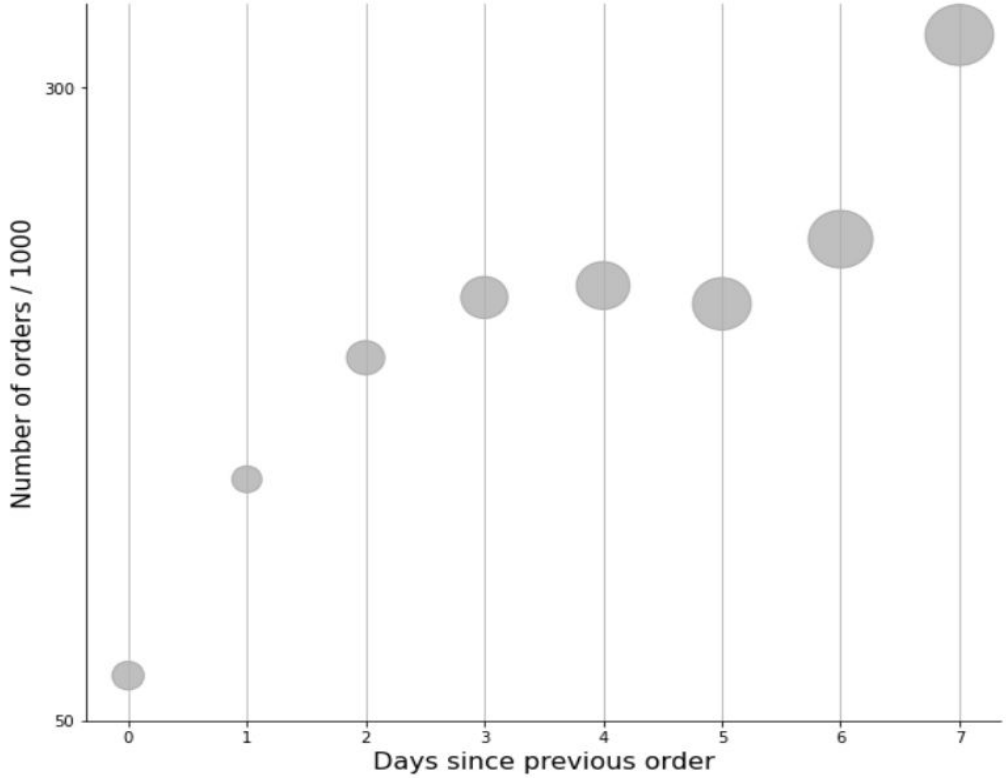
Order Frequency

How often do customers order groceries from Instacart?

≅ once every 11 days



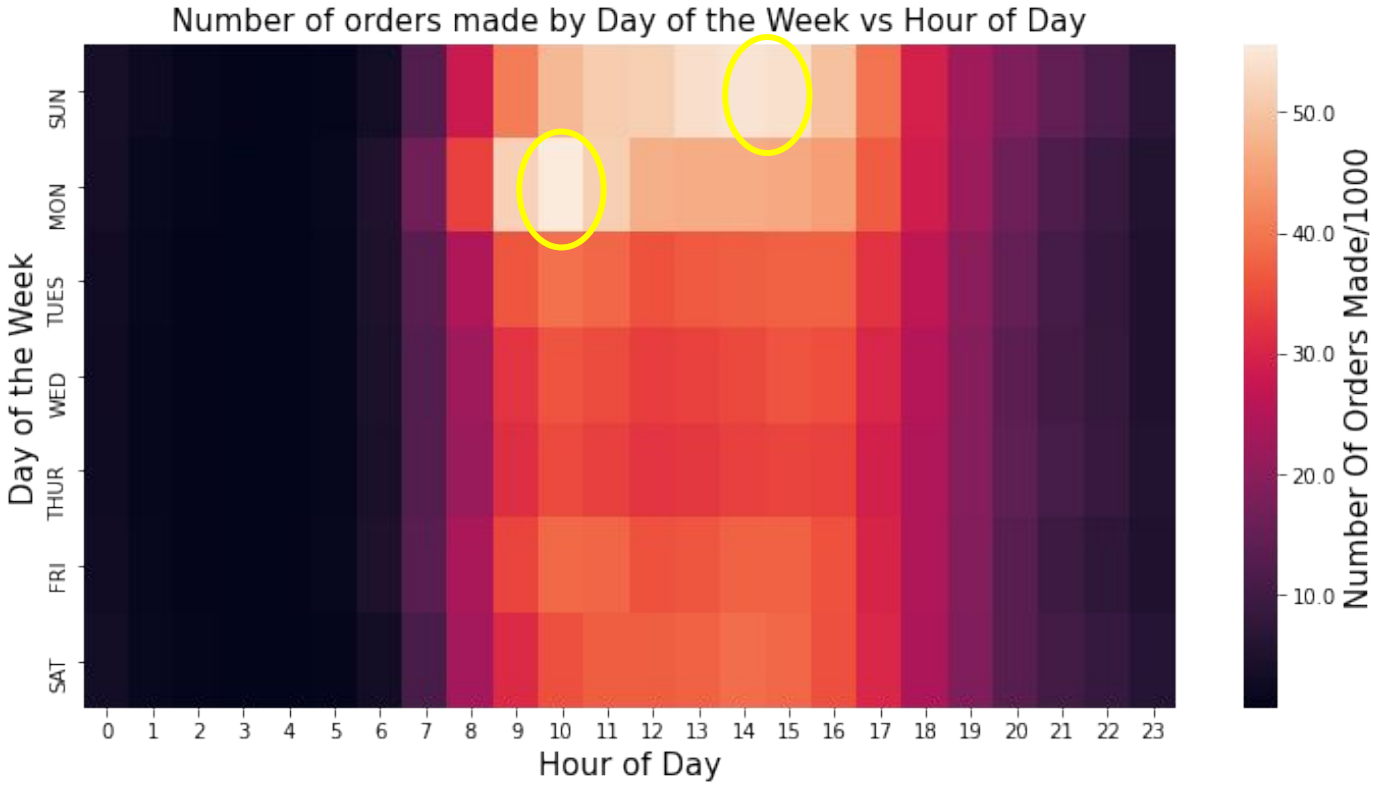
Order Frequency-Order Size Relation



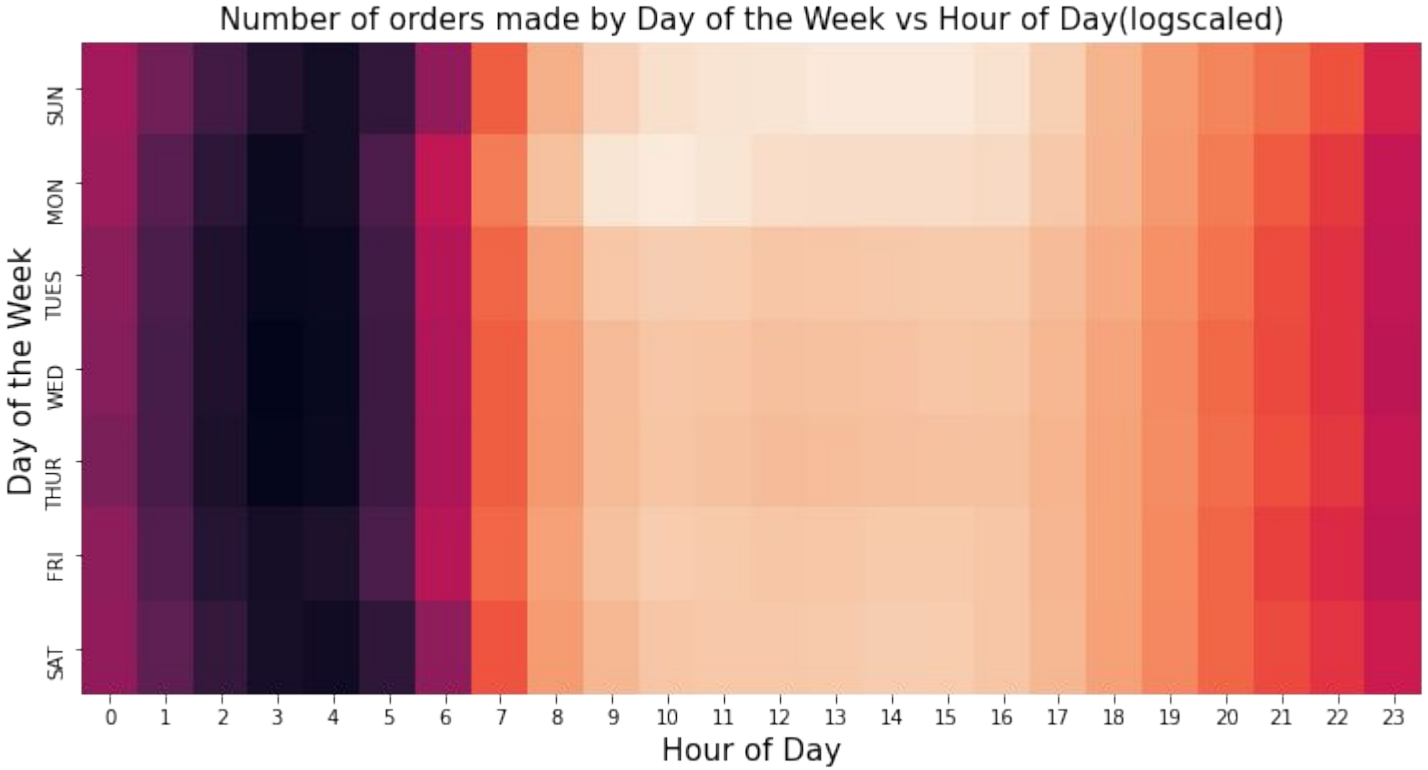
Bubble size indicates mean number of items in order.

Mean order size increases upto day 7 and then plateaus at about 10 items.

Order Times


Number of orders made by Day of the Week vs Hour of Day
Day of the Week
SUN
MON
TUES
WED
THUR
FRI
SAT
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
Hour of Day
Number Of Orders Made/1000
50.0
40.0
30.0
20.0
10.0

Order Times - Logarithmic plot



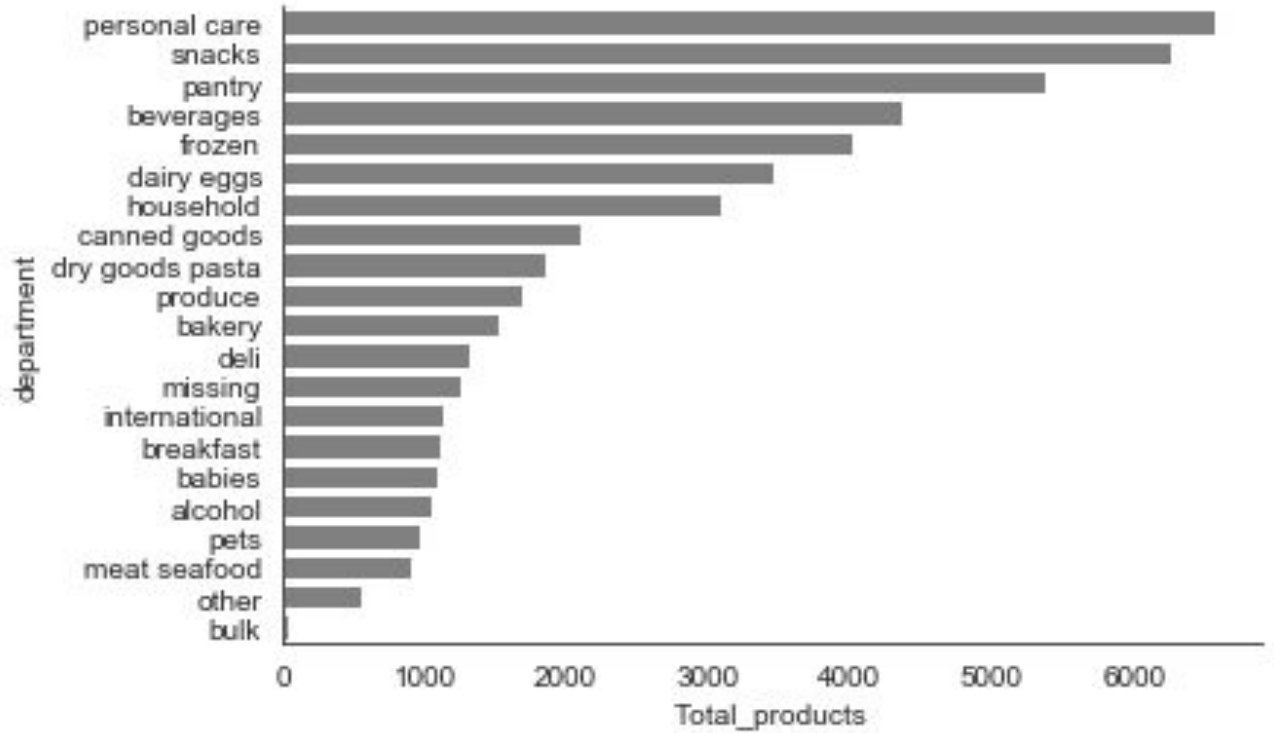
Most frequently purchased products

What products do consumers buy most often ?

Produce dominates other departments , perhaps because it is perishable.

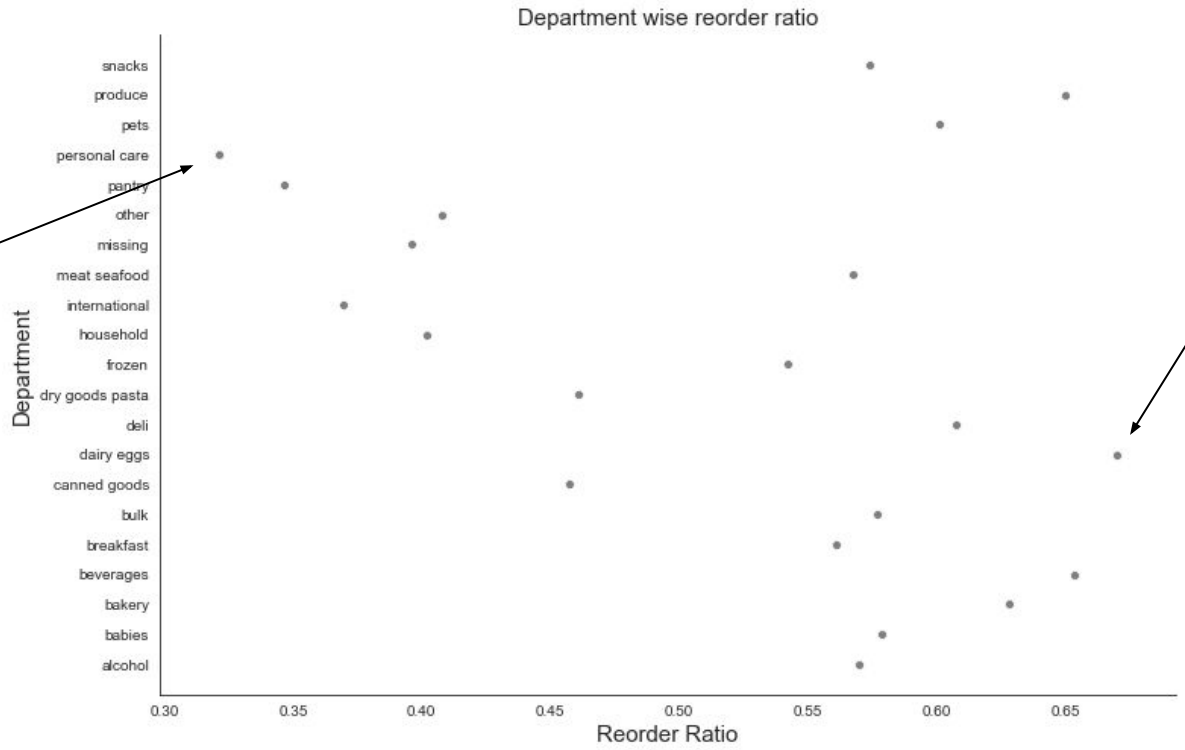
People buy organic food, despite it being more expensive.

Departments vs number of products



What products are reordered most often?

Personal Care has lowest reorder ratio



Dairy eggs have the highest reorder ratio

Dataset

Data Exploration

Recommender System

Recommender System

Method: User-based Collaborative Filtering.

- Purchases of all users are split into the prior set and the current set.
- Training: Prior set is used to build **frequency utility matrix**:
  - users as rows
  - products as columns
  - purchase frequencies as entries.
- Testing: Random 10000 users in the current set.

TF-IDF in this problem…

TF-IDF: Term Frequency–Inverse Document Frequency

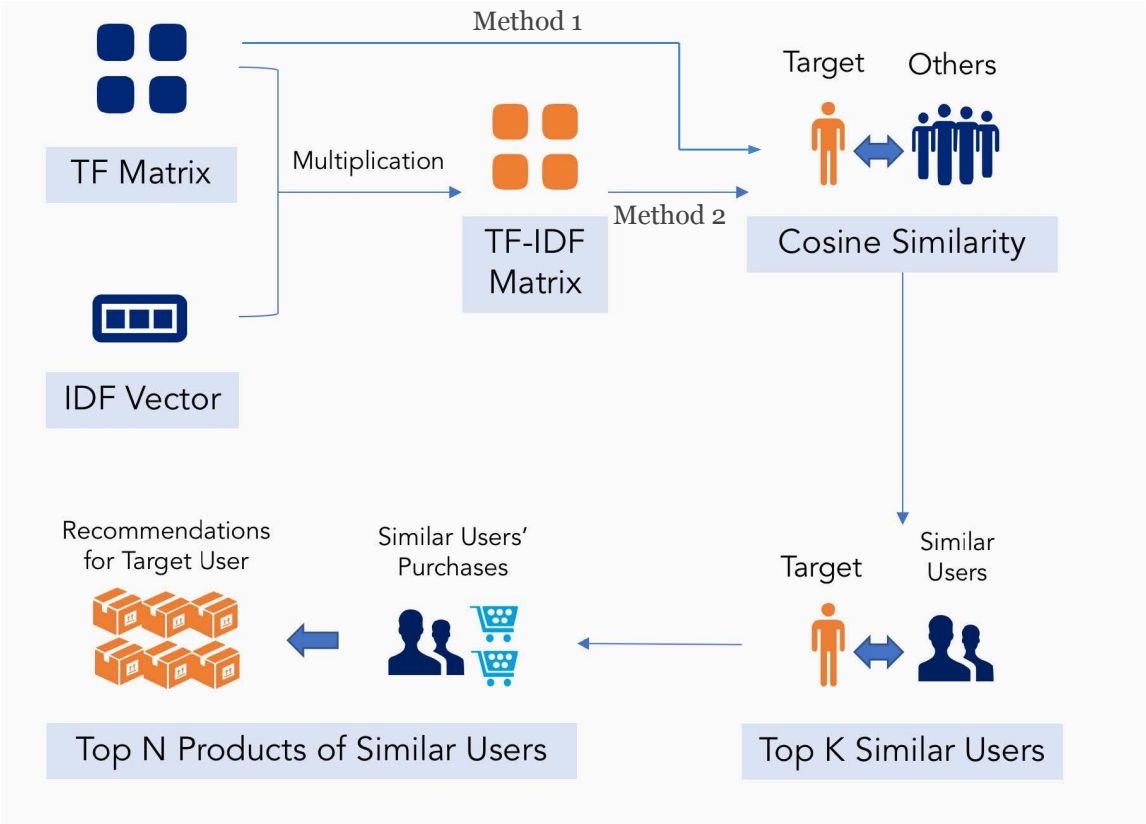
ANALOGY

Word	TERM		PRODUCT
	DOCUMENT		PURCHASE HISTORY
	QUERY		PURCHASE HISTORY of a TARGET USER
	SIMILAR DOCUMENTS Related to QUERY		PURCHASE HISTORIES OF USERS Similar to TARGET USER

How does TF-IDF work?

Frequency Matrix

Target User ID



Recommendation

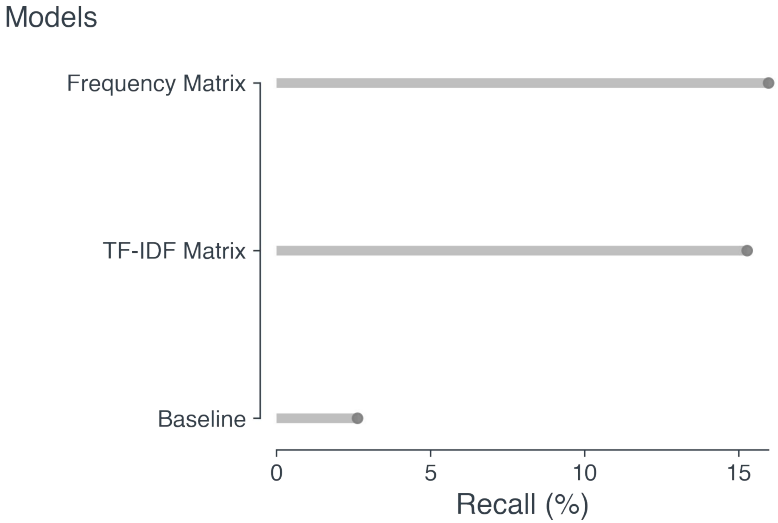
Example of recommendation

	Product	Department	Aisle
0	Original Beef Jerky	snacks	popcorn jerky
1	Soda	beverages	soft drinks
2	Pistachios	snacks	nuts seeds dried fruit
3	Bag of Organic Bananas	produce	fresh fruits
4	Clementines	produce	packaged produce
5	Crunchy Oats 'n Honey Granola Bars	snacks	energy granola bars
6	Zero Calorie Cola	beverages	soft drinks
7	Trail Mix	snacks	trail mix snack mix
8	Strawberries	produce	fresh fruits
9	Fuji Apples	produce	fresh fruits

For user with user_id = 1, the recommendation given by our system.
(K = 20, N =10)

Recommender Test Result

- **Baseline model** - generates recommendation based on the global popularity trending. (recall: **2.62%**) (N = 10)
- **User-based CF model** - generates recommendation based on user similarities over frequency/TF-IDF matrix. (recall: **15.97%/15.28%**)(K =20, N = 10)



Conclusion

- Customers seems to use Instacart at regular intervals
- Some time-intervals are much busier than others - specifically Sunday evenings and Monday mornings
- Perishable items like produce, dairy, eggs, are ordered very frequently
- Our user based CF model gives better recommendations than the baseline

References

- Dataset: https://www.kaggle.com/c/instacart-market-basket-analysis/overview
- Slide Template: https://slidesgo.com
- Kaggle Instacart EDA : https://www.kaggle.com/jungeun0121/instacart-eda
- Recommender System: https://github.com/nakulcr7/recommender-system-instacart

Any Questions?