

Instacart

Coding Comrades

Jared Hazlett, Arvind Bhamidipati, Joon Hee
Ariel, Akriti Suri, Regine Pablo



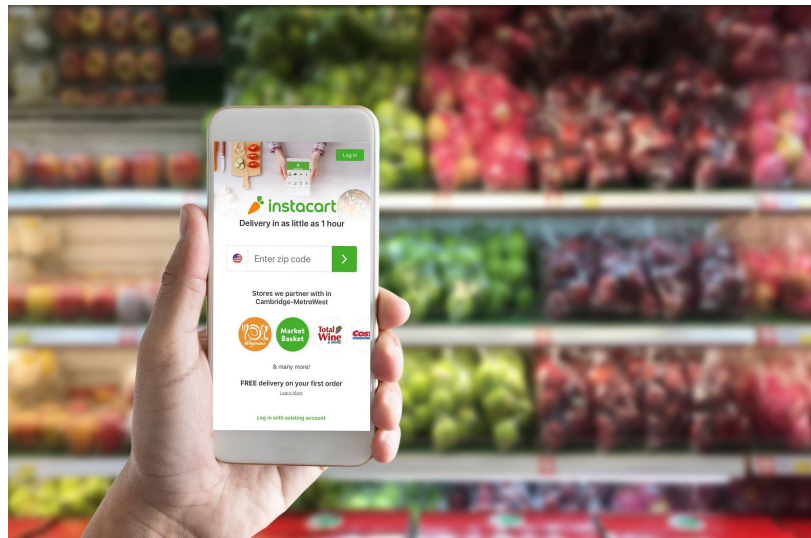
Background Overview

Why we used this Data:

- Data came from Instacart itself
- Instacart is challenging the Kaggle community to use this anonymized data on customer orders over time to predict which previously purchased products will be in a user's next order.

The Data itself

- Relational set of files describing customers' orders over time.
- Multiple raw CSV's with product names, aise, unique customer ID's



Project Question:



What we suggest Instacart customers to purchase?





Code Flow



- Clean Data
- Filter Data
- Machine Learning



- Load Data



- Website



- Statistics, Insights, and Visuals



Merging Data





Filtering Data

What we did

- Filter based on “*product name*”
Column to create a new column of
“*Organic or not*”
- Create a **conditional filter** to return
Vegetarian in a new column

Challenges:

- Limitations for vegan/non vegan
options

product_name	department_id	department	aisle_id	aisle	order_number	order_dow	order_hour_of_day	days_since_prior_order	add_to_cart_order	reordered	Organic vs Non Organic	Food Group	Vegan Friendly
Organic Italian Parsley Bunch	4	produce	16	fresh herbs	11	4	9	8	13	1	Organic	Vegetarian	Yes
Organic Cucumber	4	produce	83	fresh vegetables	11	4	9	8	11	1	Organic	Vegetarian	Yes
Organic Gala Apples	4	produce	24	fresh fruits	11	4	9	8	9	1	Organic	Vegetarian	Yes
Organic Bartlett Pear	4	produce	24	fresh fruits	11	4	9	8	8	1	Organic	Vegetarian	Yes
Organic Strawberries	4	produce	24	fresh fruits	11	4	9	8	18	1	Organic	Vegetarian	Yes
Organic Grapes				packaged									



nt	aisle_id	aisle	order_number	order_dow	order_hour_of_day	days_since_prior_order	add_to_cart_order	reordered	Organic vs Non Organic	Food Group	Vegan Friendly
28	16	fresh herbs	11	4	9	8	13	1	Organic	Vegetarian	Yes
28	83	fresh vegetables	11	4	9	8	11	1	Organic	Vegetarian	Yes
28	24	fresh fruits	11	4	9	8	9	1	Organic	Vegetarian	Yes
28	24	fresh fruits	11	4	9	8	8	1	Organic	Vegetarian	Yes
28	24	fresh fruits	11	4	9	8	18	1	Organic	Vegetarian	Yes



aisle	order_number	order_dow	order_hour_of_day	days_since_prior_order	add_to_cart_order	reordered	Organic vs Non Organic	Food Group	Vegan Friendly	Organic or Not	Veg Friendly
fresh herbs	11	4	9	8	13	1	Organic	Vegetarian	Yes	Organic	Vegetarian
fresh vegetables	11	4	9	8	11	1	Organic	Vegetarian	Yes	Organic	Vegetarian
fresh fruits	11	4	9	8	9	1	Organic	Vegetarian	Yes	Organic	Vegetarian
fresh fruits	11	4	9	8	8	1	Organic	Vegetarian	Yes	Organic	Vegetarian
fresh fruits	11	4	9	8	18	1	Organic	Vegetarian	Yes	Organic	Vegetarian
...
other



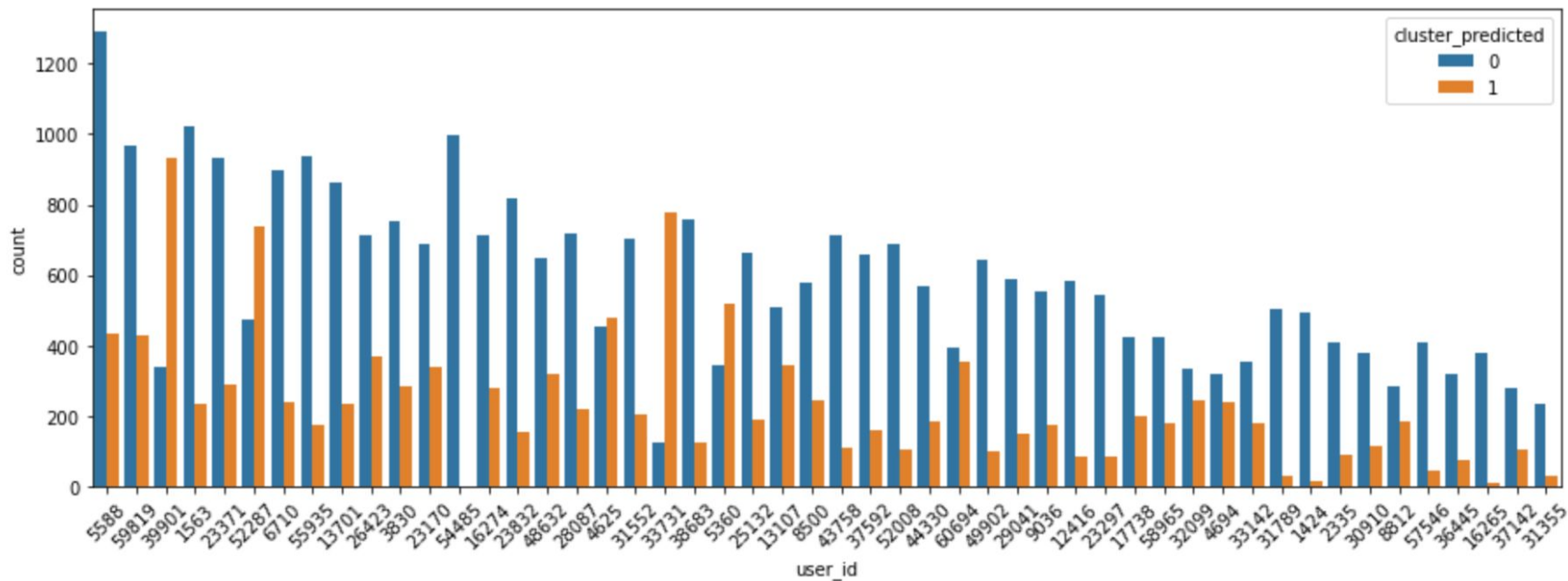
Machine Learning



- Using K modes clustering (Huang, Cao)
- Preprocessing the data
- Choosing K by comparing Cost against each K
- Predicting the clusters



Clustering with K=2





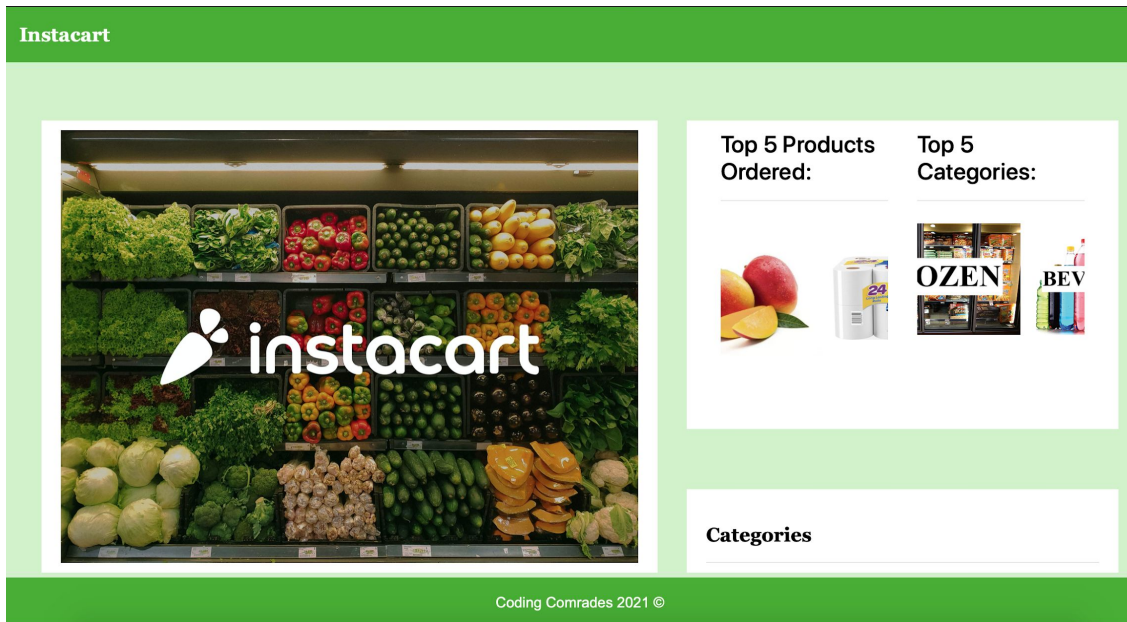
Website

What we did:

- Used *HTML* & *CSS*

Challenges:

- Product Table
 - Filtering Data



Website Demo





Statistics, Insights, Findings

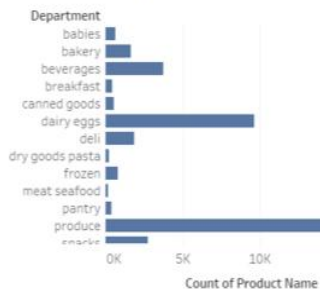


Food Group Data

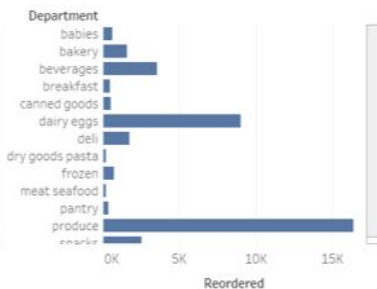
Food Group - No. times Product Purchased

Departm..	Product Name	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
babies	Baby Food Stage 2 B...	16	16	9	3	5	11	35
	Organic 4 Months B...	24	11	9	21	15	10	15
	Organic Apples, Carr...	14	10	8	8	5	3	15
	Organic Pears, Peas...	33	27	19	27	23	16	16
	Organic Strawberry ...	13	10	17	8	10	12	11
	Organic Sunny Days ...	21	8	9	4	13	9	14
	Stage 1 Apples Swe...	20	18	15	19	8	8	16
bakery	100% Whole Wheat ...	30	9	37	39	16	23	51
	Butter Bread	9	10	14	5	10	15	5
	Cinnamon Raisin Bre...	12	10	15	15	14	13	7
	Everything Bagels	11	5	39	44	42	35	32
	Gluten Free Plain Pa...	16	10	14	20	21	10	10

Products Purchased



Products Reordered



Time Dashboard

Type Of Day
☒ Weekday
☒ Weekend

DayOfWeek - Products (Graph)

DayOfWeek - Products (Table)

Dayof Week	
Sunday	4,473
Monday	5,251
Tuesday	7,256
Wednesday	5,802
Thursday	5,118
Friday	5,435
Saturday	7,895



TimeofDay

Hour Of ..

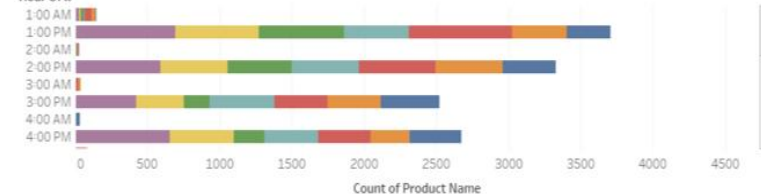


Tableau Demo



Conclusions



Appendix