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

kaggle 🥕 instacart
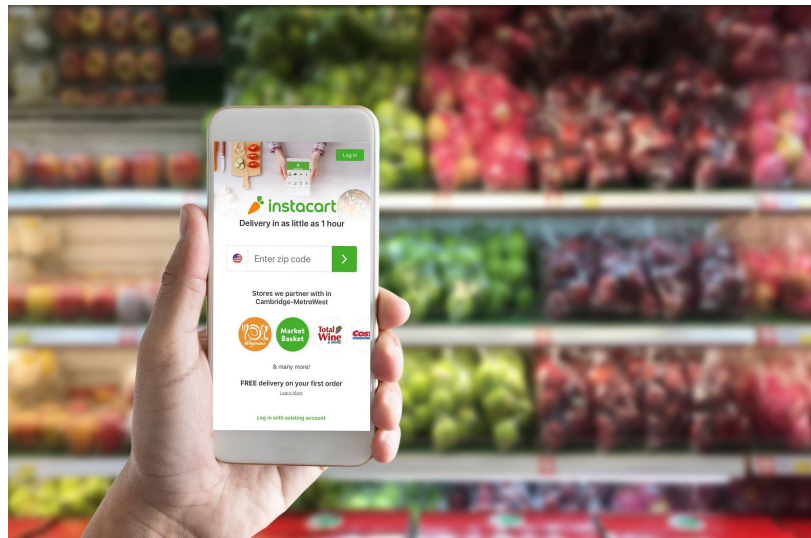
Project Question:

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	Fo Gro
Organic Italian Parsley Bunch	4	produce	16	fresh herbs	11	4	9	8	13	1	Organic	Vegeta
Organic Cucumber	4	produce	83	fresh vegetables	11	4	9	8	11	1	Organic	Vegeta
Organic Gala Apples	4	produce	24	fresh fruits	11	4	9	8	9	1	Organic	Vegeta
Organic Bartlett Pear	4	produce	24	fresh fruits	11	4	9	8	8	1	Organic	Vegeta
Organic Strawberries	4	produce	24	fresh fruits	11	4	9	8	18	1	Organic	Vegeta
Organic Grape				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
ce	16	fresh herbs	11	4	9	8	13	1	Organic	Vegitarian	Yes
ce	83	fresh vegetables	11	4	9	8	11	1	Organic	Vegitarian	Yes
ce	24	fresh fruits	11	4	9	8	9	1	Organic	Vegitarian	Yes
ce	24	fresh fruits	11	4	9	8	8	1	Organic	Vegitarian	Yes
ce	24	fresh fruits	11	4	9	8	18	1	Organic	Vegitarian	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	Vegitarian	Yes	Organic	Vegitarian
fresh vegetables	11	4	9	8	11	1	Organic	Vegitarian	Yes	Organic	Vegitarian
fresh fruits	11	4	9	8	9	1	Organic	Vegitarian	Yes	Organic	Vegitarian
fresh fruits	11	4	9	8	8	1	Organic	Vegitarian	Yes	Organic	Vegitarian
fresh fruits	11	4	9	8	18	1	Organic	Vegitarian	Yes	Organic	Vegitarian
...	...	...	...	...	...	...	...	...	...	...	...
other										Not	

pandas

Machine Learning

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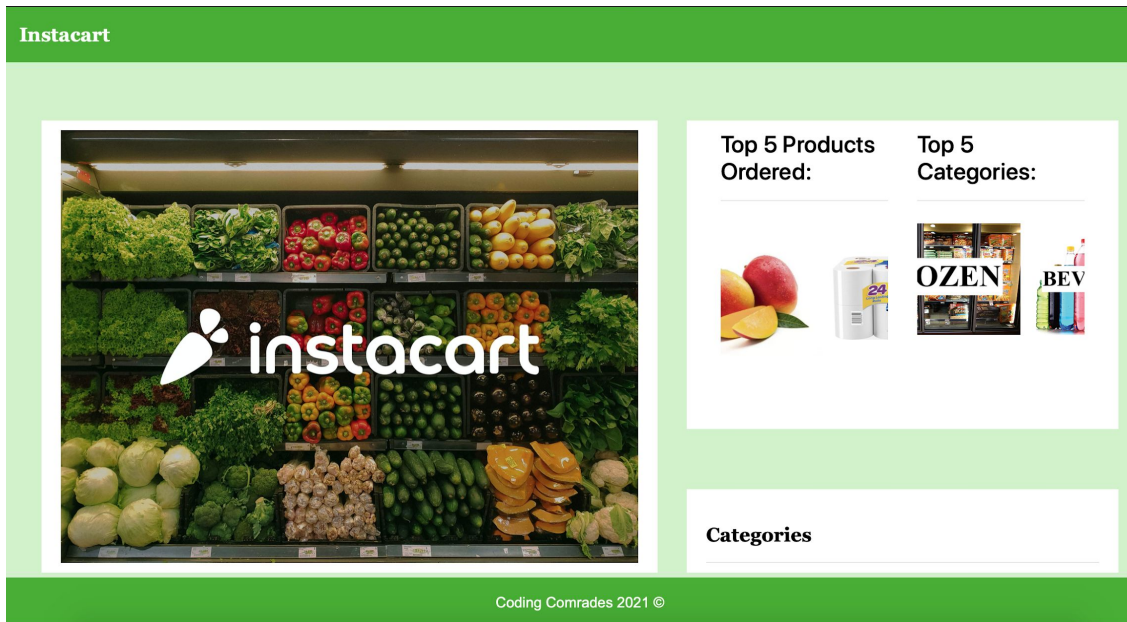- Predicting the clusters

Clustering with K=2
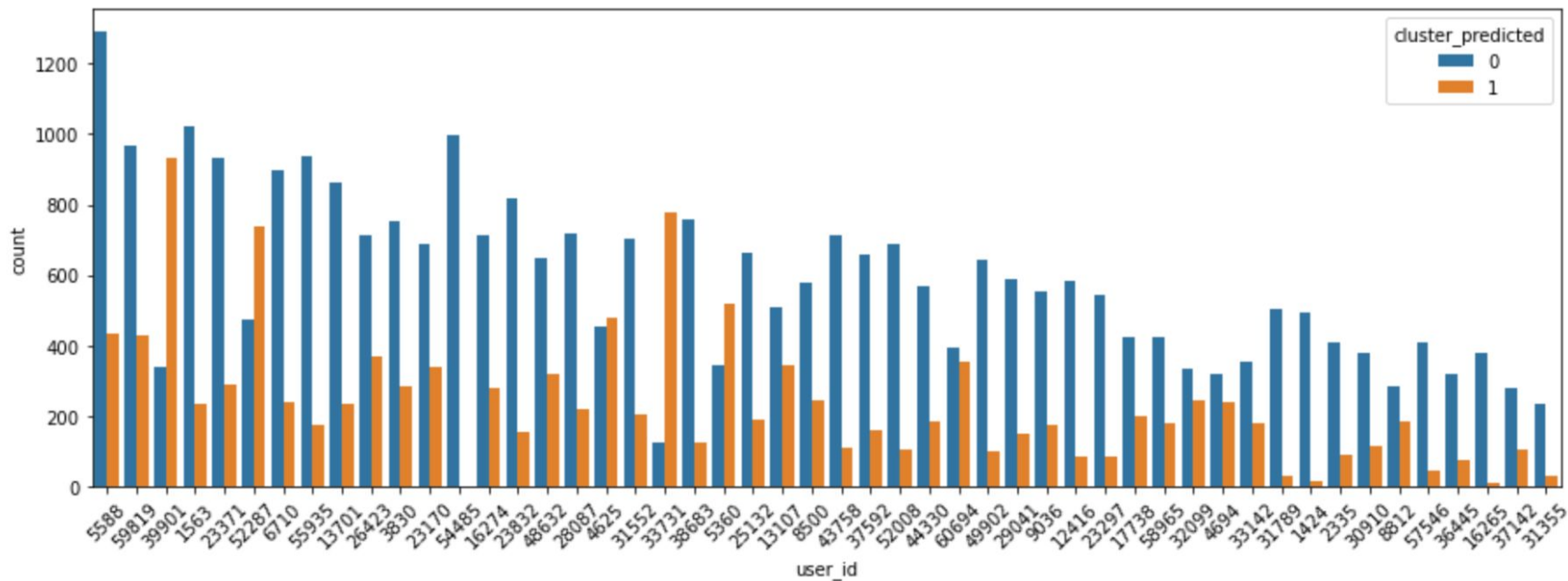
Website

What we did:

- Used *HTML & CSS*

Challenges:

- Product Table
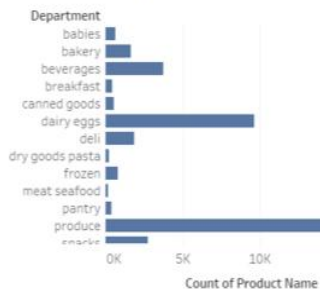 - Filtering Data
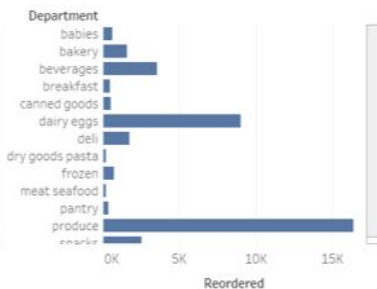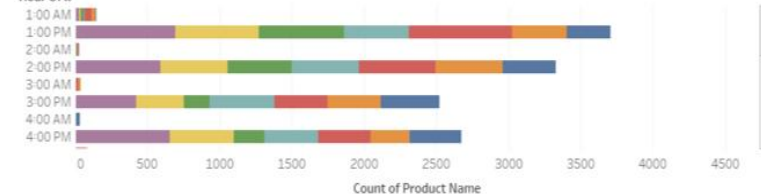
Website Demo

Statistics, Insights, Findings

Tableau Demo

Conclusions

Appendix