



instacart

data analysis

tanner martz

Problem Statement

What makes a product popular on Instacart?

Data Needed

1. Intrinsic Product attributes

- a. Nutritional value
- b. Product category

2. Customer behavior patterns

- a. Add to cart order
- b. Reorder rate

The Data LifeCycle

1. Data Collection
2. Data Processing (ETL)
3. Exploratory Analysis & Data Visualization (EDA)
4. Analysis, Hypothesis Testing, & ML (Model)
5. Insight & Policy Decision

1 Collect

2 ETL

3 EDA

4 ML Model

5 Insights



Data Collection

1. Kaggle Dataset [here](#).
 - a. 3 million+ grocery orders from 200,000+ users
 - b. Customer behavior patterns
 - c. Product categorization
2. USDA Nutrition Profiles Database [here](#).
 - a. Foundational nutritional information

1 Collect

2 ETL

3 EDA

4 ML Model

5 Insights



Data Processing (ETL)

1. Kaggle Dataset

- a. Data imported from CSV files

2. USDA Nutrition Profiles Database

- a. API call and save data to CSV file
 - i. Calories, Protein, Fat, Carbohydrate, Water

1

Collect

2

ETL

3

EDA

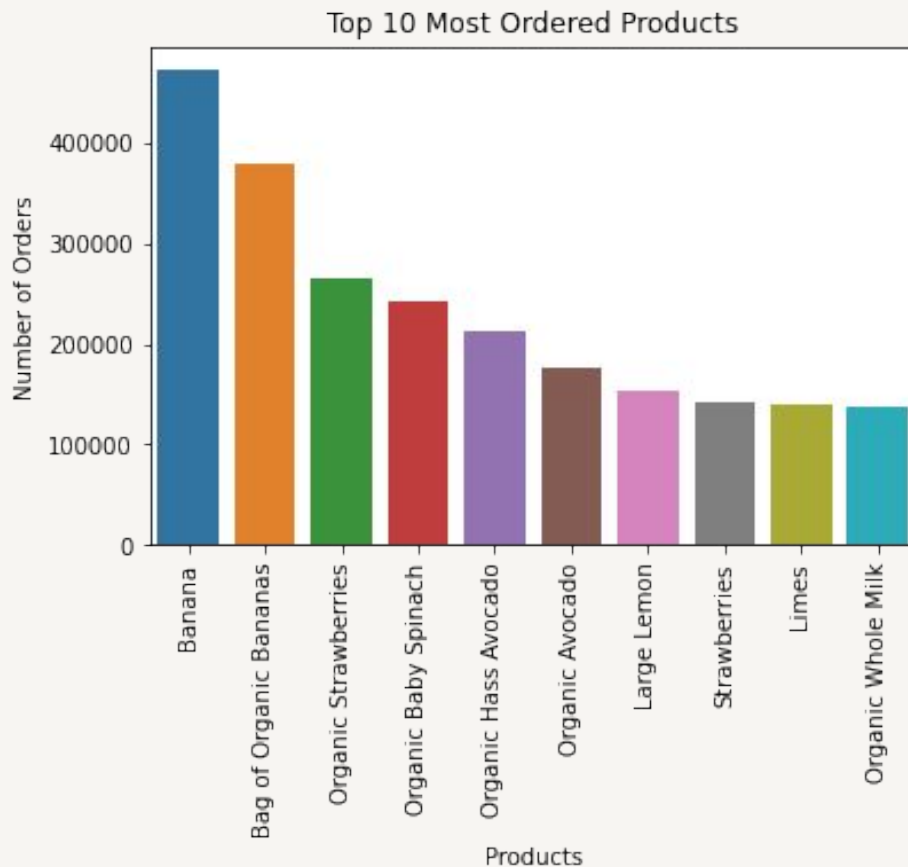


4

ML Model

5

Insights



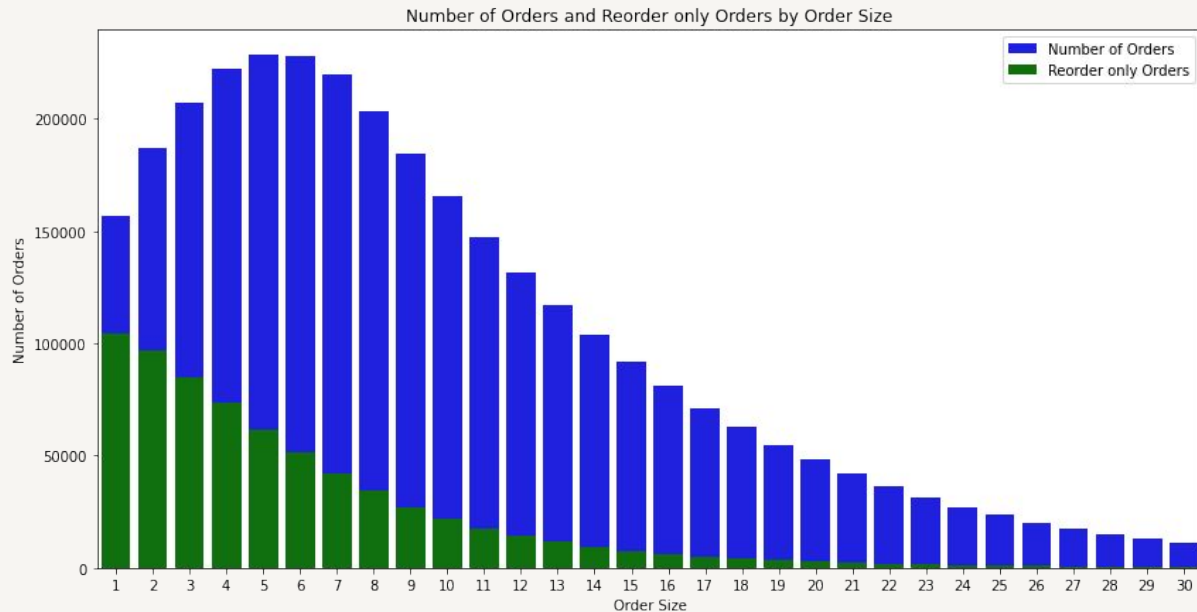
1 Collect

2 ETL

3 EDA
..

4 ML Model

5 Insights



1 Collect

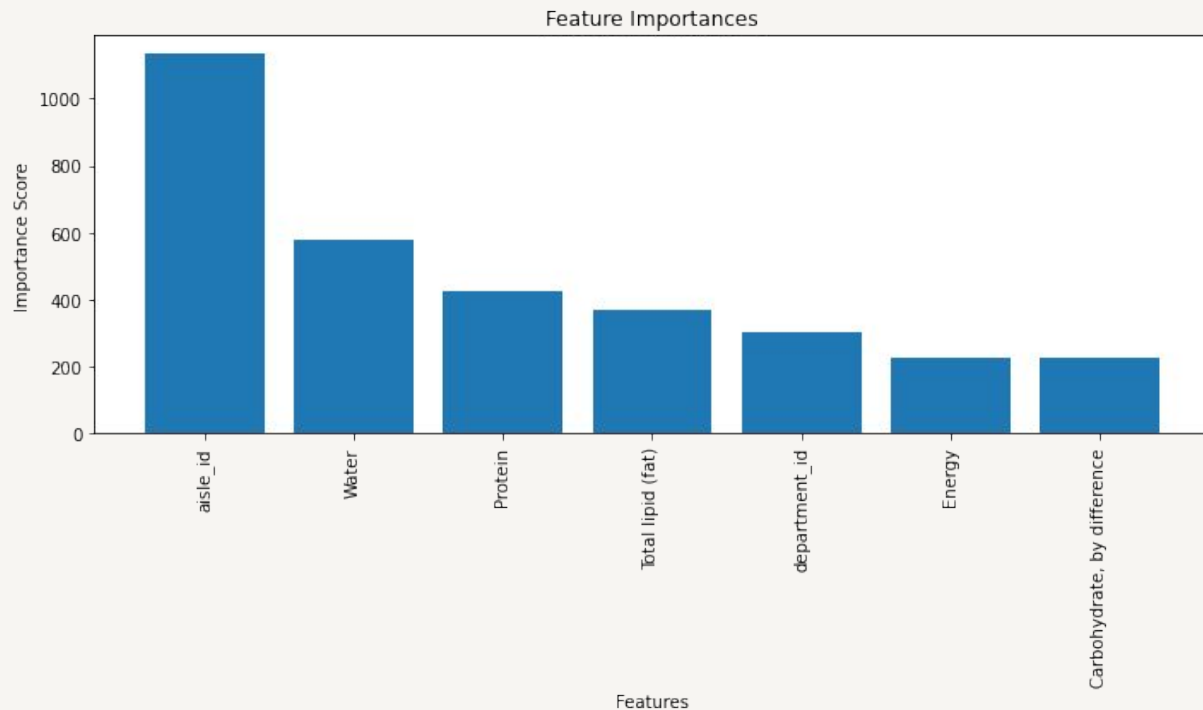
2 ETL

3 EDA

4 ML Model



5 Insights



1 Collect

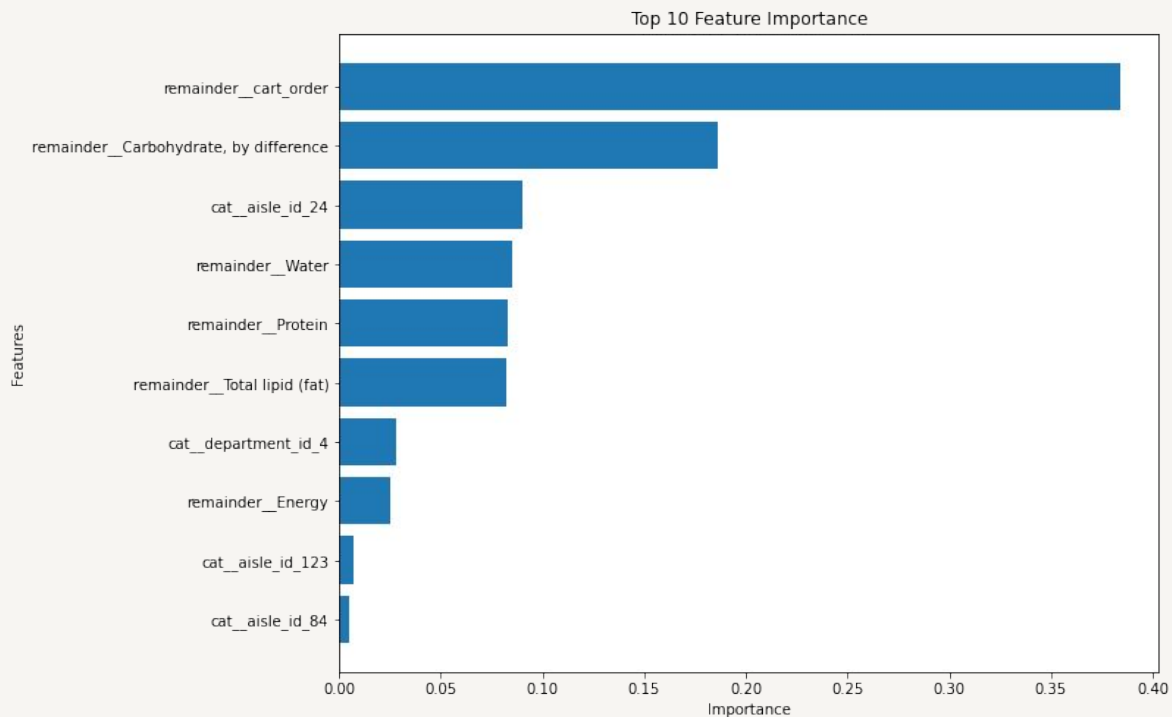
2 ETL

3 EDA

4 ML Model



5 Insights



1 Collect

2 ETL

3 EDA

4 ML Model
...

5 Insights

(Persuaded into buying) Top 10 Products with Largest Rank Difference:

	product_name	rank_file1	rank_file2	rank_difference
83	Organic Lacinato (Dinosaur) Kale	99	69	30
76	Organic Ginger Root	84	56	28
62	Organic Red Onion	64	37	27
74	Green Bell Pepper	79	52	27
73	Red Peppers	78	51	27
77	Small Hass Avocado	86	60	26
65	Organic Italian Parsley Bunch	68	43	25
57	Organic Cilantro	59	34	25
75	Boneless Skinless Chicken Breasts	80	58	22
36	Organic Grape Tomatoes	38	19	19

(Desuaded from buying) Top 10 Products with Smallest Rank Difference:

	product_name	rank_file1	rank_file2	rank_difference
16	Soda	17	90	-73
53	Organic Reduced Fat Milk	55	93	-38
55	Whole Milk	57	92	-35
22	Hass Avocados	23	57	-34
15	Spring Water	16	49	-33
51	100% Raw Coconut Water	53	83	-30
56	Granny Smith Apples	58	88	-30
67	Sparkling Lemon Water	71	100	-29
37	Organic Reduced Fat 2% Milk	39	64	-25
58	Organic Bartlett Pear	60	84	-24

1 Collect

2 ETL

3 EDA

4 ML Model

5 Insights

Insight & Policy Decision

Perishability is the #1 predictor of most ordered instacart products

Customers are subject to other behavioral factors that may nudge them to purchase “unessential” products