



Instacart - USDA - Data Analysis

Aruna Rayapeddi
Data Scientist



Idea



Know what you eat before you buy any food product

- Know the Nutrition values (USDA)
- Know the Product details (Instacart)
- Instacart Shopping cart



What is Instacart?

- On-demand, Online, same-day grocery delivery and pick up platform in the U.S. and Canada.
- Groceries delivered in as fast as 1 hour.





What is USDA Food Composition Database ?

- Major source of food composition data in the United States
- Foundation for most food composition databases in the public and private sectors
- Used in developing the Dietary Guidelines for Americans and Dietary Reference Intakes for a wide variety of nutrients



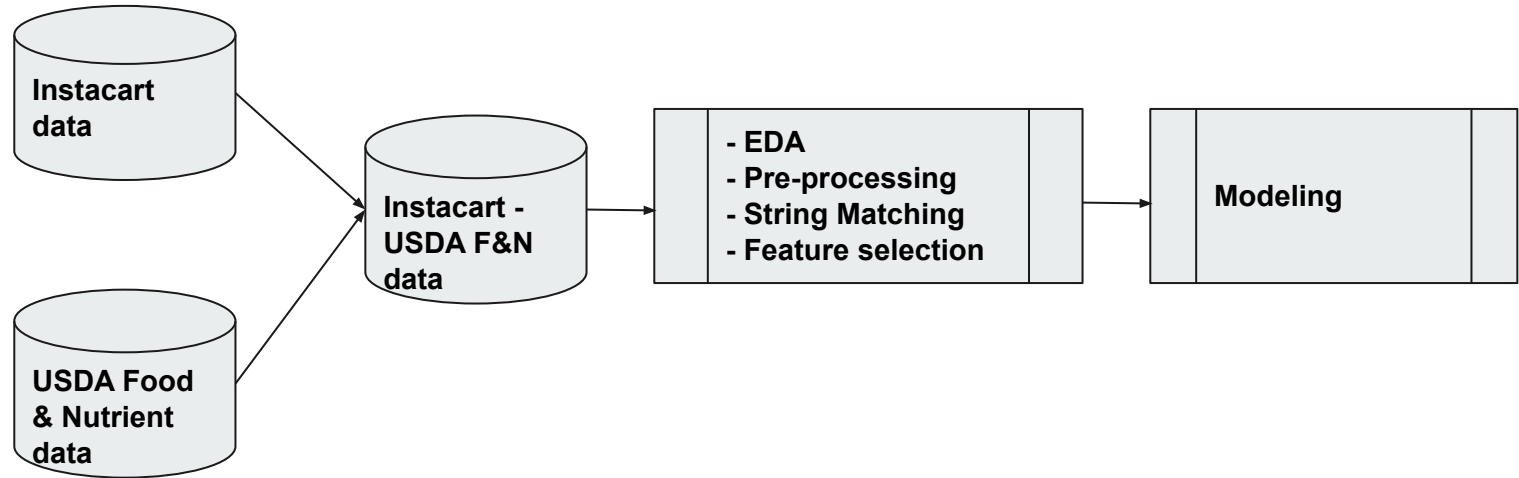


Objective

- Mapping Instacart's food items with USDA food and nutrients data, analyze and get insights on user's previous Instacart's food orders and predict whether or not the user reorders the product.
- The prediction also provides additional nutrient information on each of the food products.



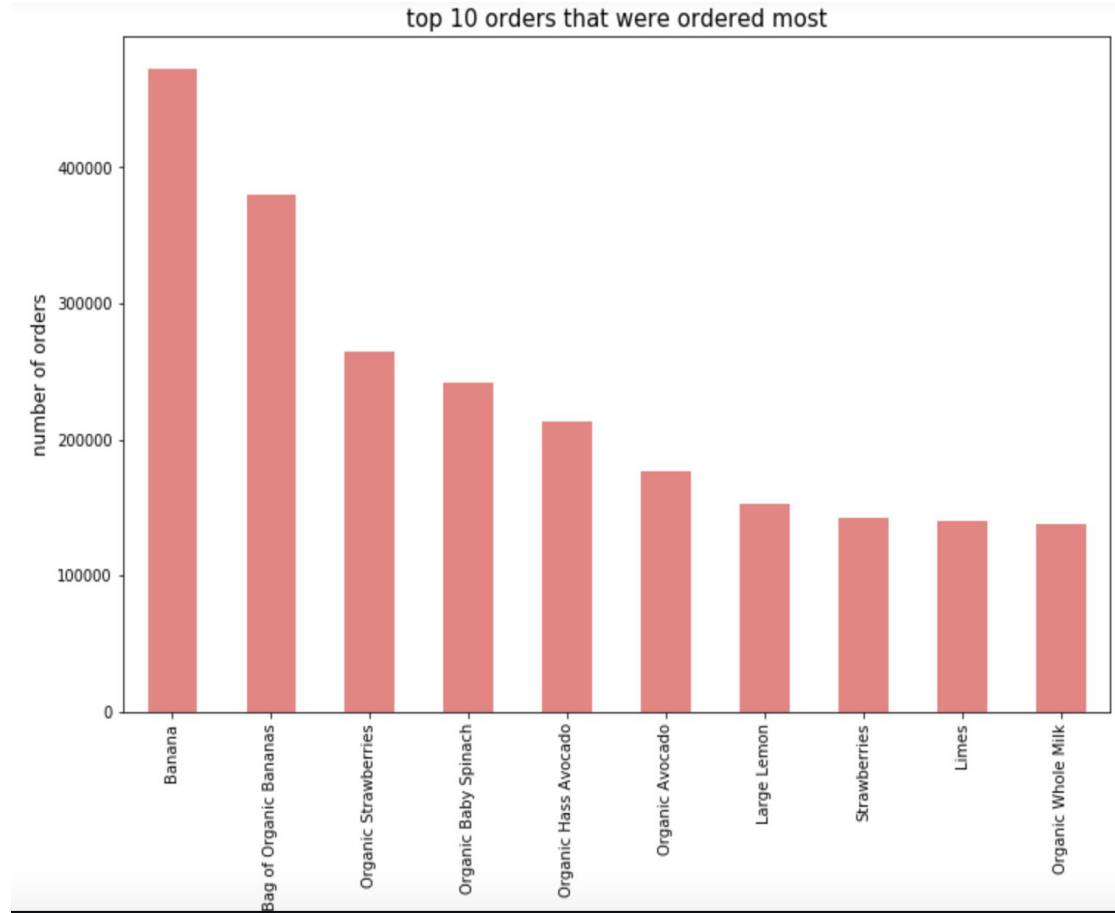
Process



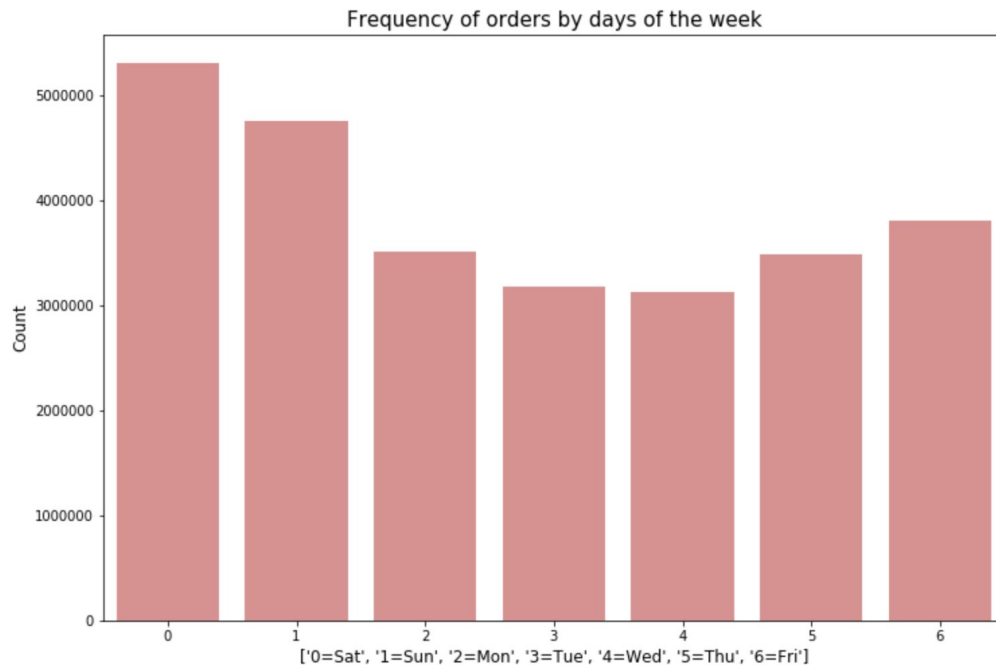
Top 10 products



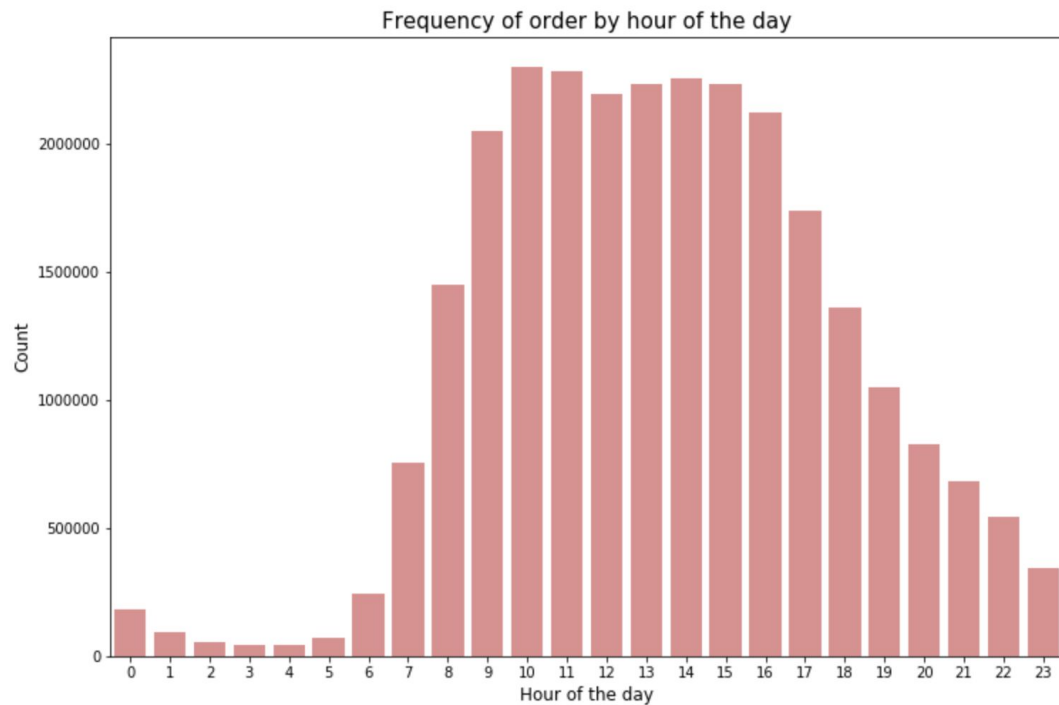
Bananas are the top choice in Orders



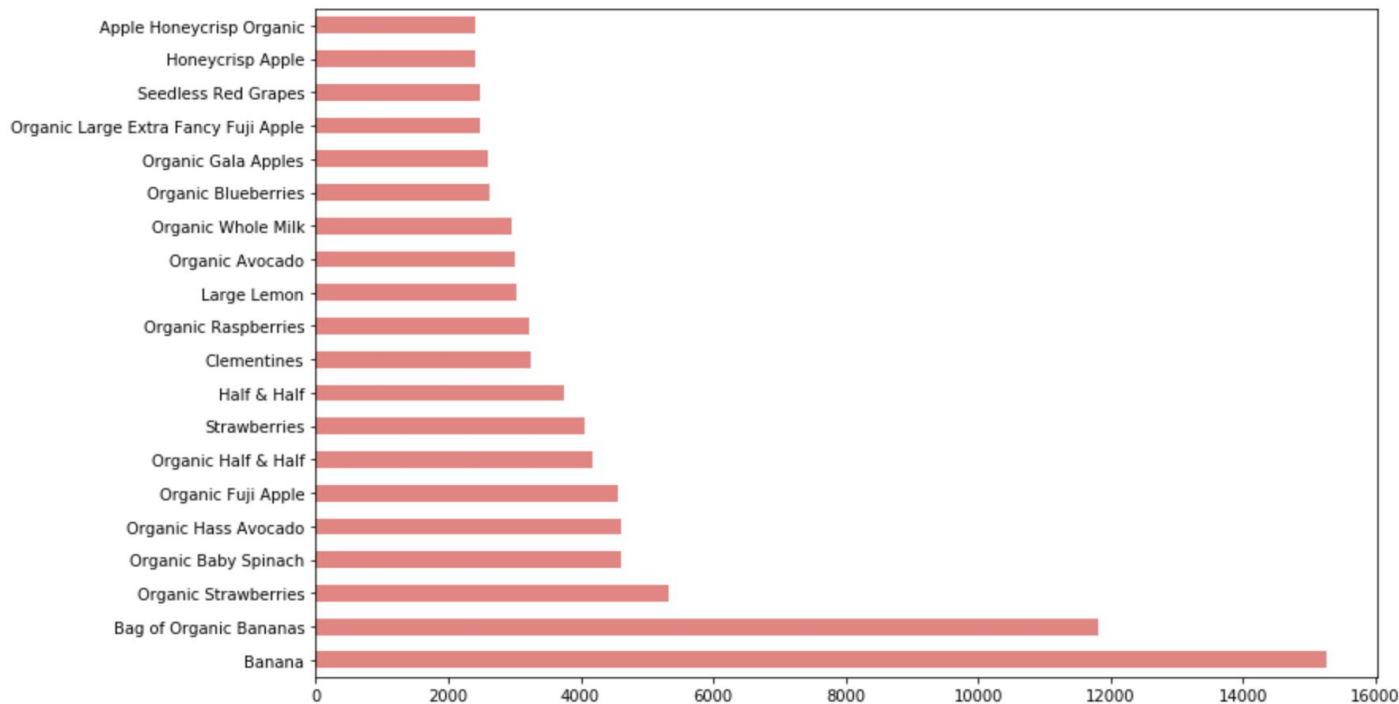
Frequency of Orders by days of the week



Frequency of Order by hour of the day



Top 20 products vegetarian users reordered



String Match between Instacart Product and USDA food

	NDB_No	Long_Desc	product_id	product_name
0	1119	Yogurt, vanilla, low fat.	30442	Low Fat Vanilla Yogurt
1	1295	Yogurt, vanilla, non-fat	30442	Low Fat Vanilla Yogurt
2	1179	Sour cream, light	26172	Light Sour Cream
3	1001	Butter, salted	42736	Unsalted Butter
4	19335	Sugars, granulated	26152	Granulated Sugar
5	11658	Spinach souffle	29116	Spinach Souffle
6	19296	Honey	49588	Honey
7	1146	Cheese, parmesan, shredded	25919	Parmesan Shredded Cheese
8	1117	Yogurt, plain, low fat	7054	Low Fat Plain Yogurt
9	18349	Rolls, french	517	French Rolls

Validate results by navigating to USDA Search

NDB_No	Long_Desc	product_name	Fiber, total dietary (g)
19165	Cocoa, dry powder, unsweetened	Unsweetened Cocoa Powder	37.0
11615	Chives, freeze-dried	Freeze-Dried Chives	26.2
		Freeze Dried Chives	26.2
2049	Thyme, fresh	Fresh Thyme	14.0
2075	Seasoning mix, dry, taco, original	Original Taco Seasoning Mix	13.3

Cocoa, dry powder, unsweetened

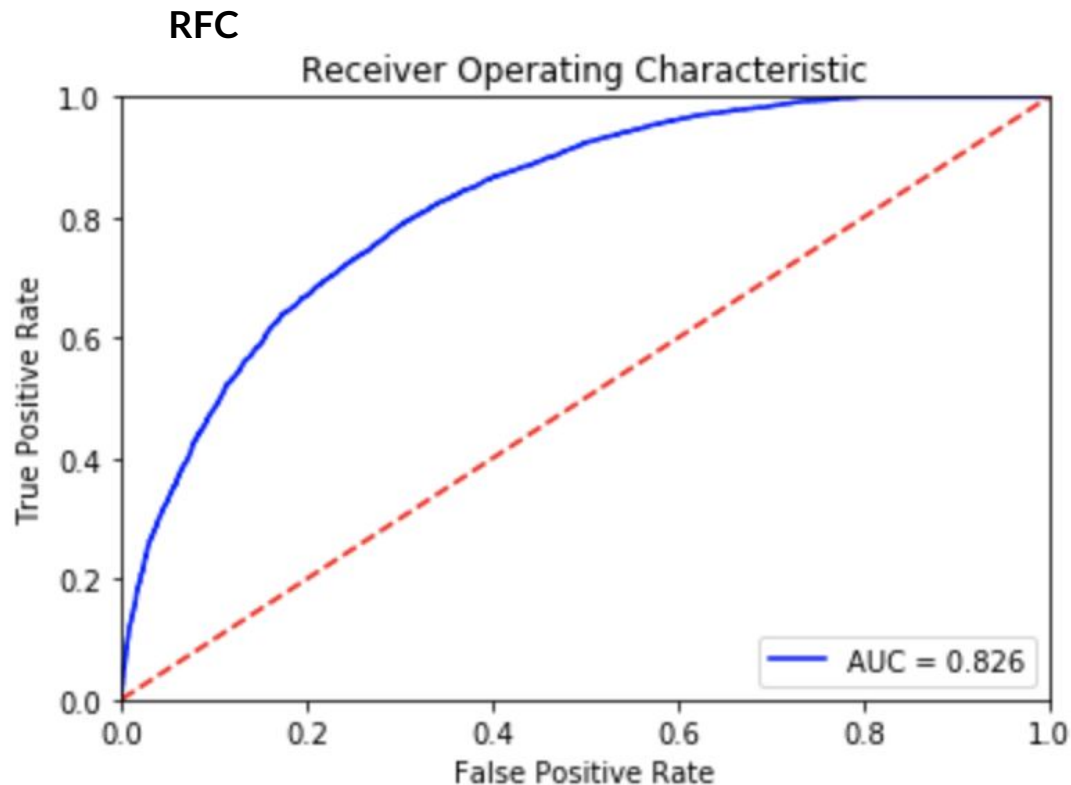
Data Type: SR Legacy **Food Category:** Sweets **FDC ID:** 169593 **NDB Number:** 19165
Published: 4/1/2019

Nutrients		Measures	
Portion:		100g	⬆⬇⬆
Fiber, total dietary		37	g

<https://ndb.nal.usda.gov/ndb/search/list>

Predictive Analysis

- Logistic Regression:
 - **accuracy: 0.697**
 - **F1 score: 0.667**
- Random Forest Classifier:
 - **accuracy: 0.713**
 - **F1 score: 0.703**
- Adaboost Classifier:
 - **accuracy: 0.730**
 - **F1 score: 0.737**
- XGBoost:
 - **accuracy: 0.815**
 - **F1 score: 0.819**





Takeaways

- Focus on clustering analysis and recommender systems.
- Focus on feature engineering.
- Correlated features for feature selection.