

12/15/23



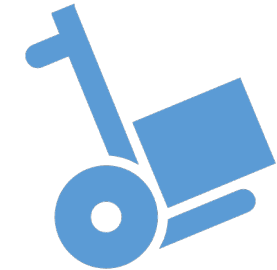


INDEX

- Problem statement
- The solution & its Potential impact
- Overview of the dataset and preprocessing procedures
- Important findings from EDA
- Baseline models and evaluation metrics
- Next steps for advanced modeling and producing your work

Problem statement

The grocery and food delivery industry has witnessed rapid growth in recent years, driven by the adoption of online shopping during and after the COVID-19 pandemic.



Predicting Future Purchase Behavior:

Customer habits are constantly evolving, making it challenging to rely solely on past purchase data for predicting future behavior.

Demand Forecasting and Inventory Management:

Industry growth, changing customer loyalty, and evolving habits can introduce unpredictability into demand forecasting and inventory management.

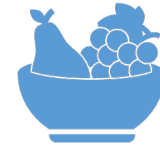
Maximizing Order Value:

While customers may be receptive to product suggestions, their purchasing patterns can also be unpredictable.

The solution

Efficient Cross-Selling
process with
recommendation
engine.

Potential impact of solution



This project's insights will benefit business owners, investors, and stakeholders within the grocery and food delivery ecosystem.



- Enhanced sales forecasting and resource allocation.



Advantages for the business:



- Optimized inventory management, reducing costs related to stockouts or overstock.



- Virtual store layout decisions, and promotional activities.



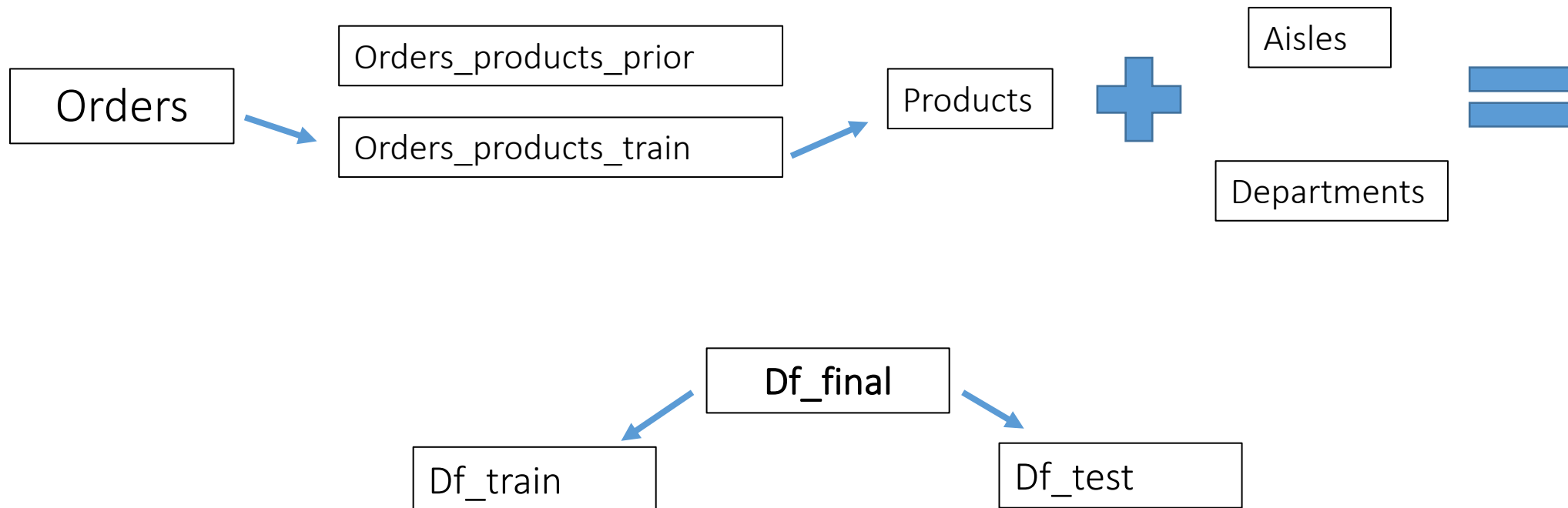
- Increased revenue through effective cross-selling, improving customer loyalty and satisfaction.

Work with dataset:

- Dataset size blocker:
 - Filtered **Orders** test dataset for test orders only (eval_set column).
 - Merged filtered **Orders** dataset with order products items dataset.
 - Combined **Orders** test dataset with products information.
 - Dropped columns that wont be used in the analysis 'order_dow', 'order_hour_of_day'
- Using test dataset with a fewer number of rows (train).



Overview of the dataset and preprocessing procedures:





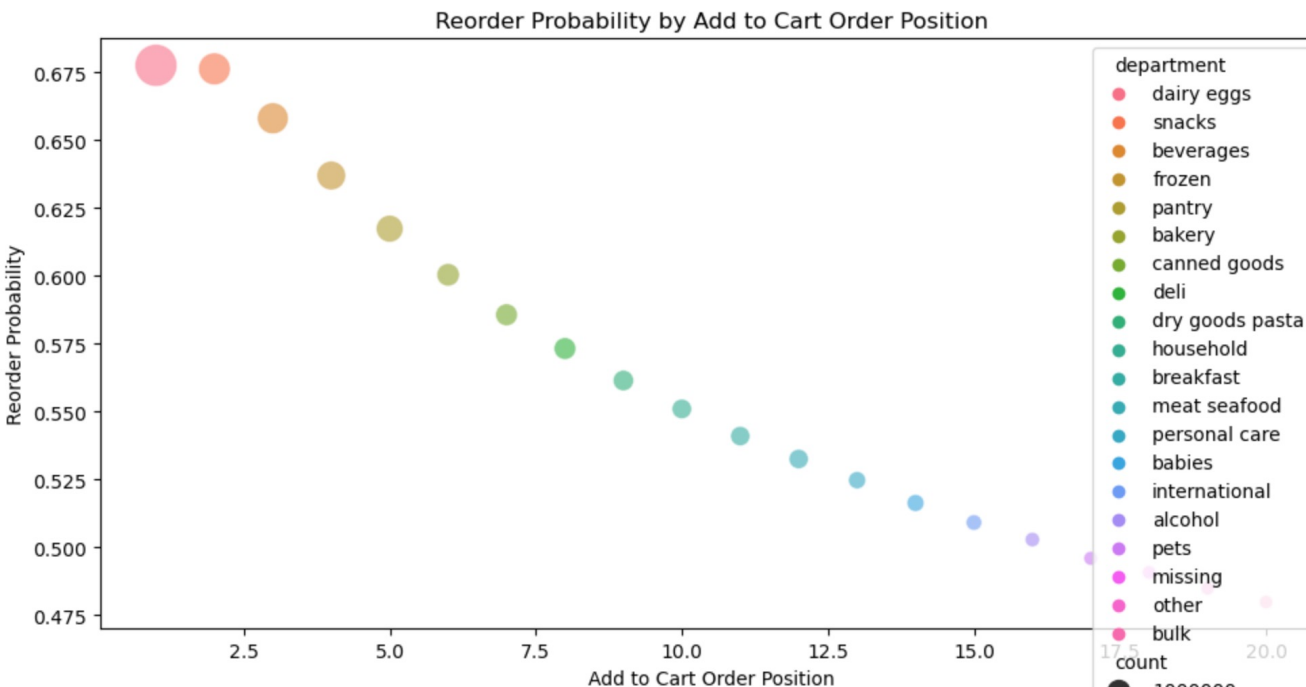
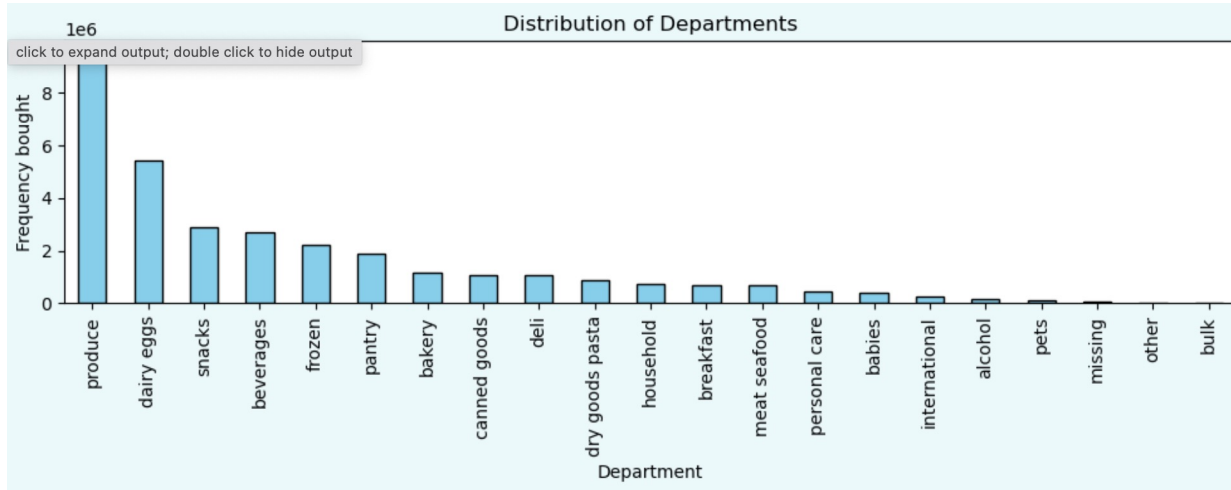
Features:

- User ID
- Order number
- Order of item added to the cart
- Items inside the order
- Product Name
- Aisle & Department Number
- Items frequently ordered together

Target Variable:

- Reordered item
- Items frequently ordered together
- Days since last order
- Product Basket items

Important EDA findings



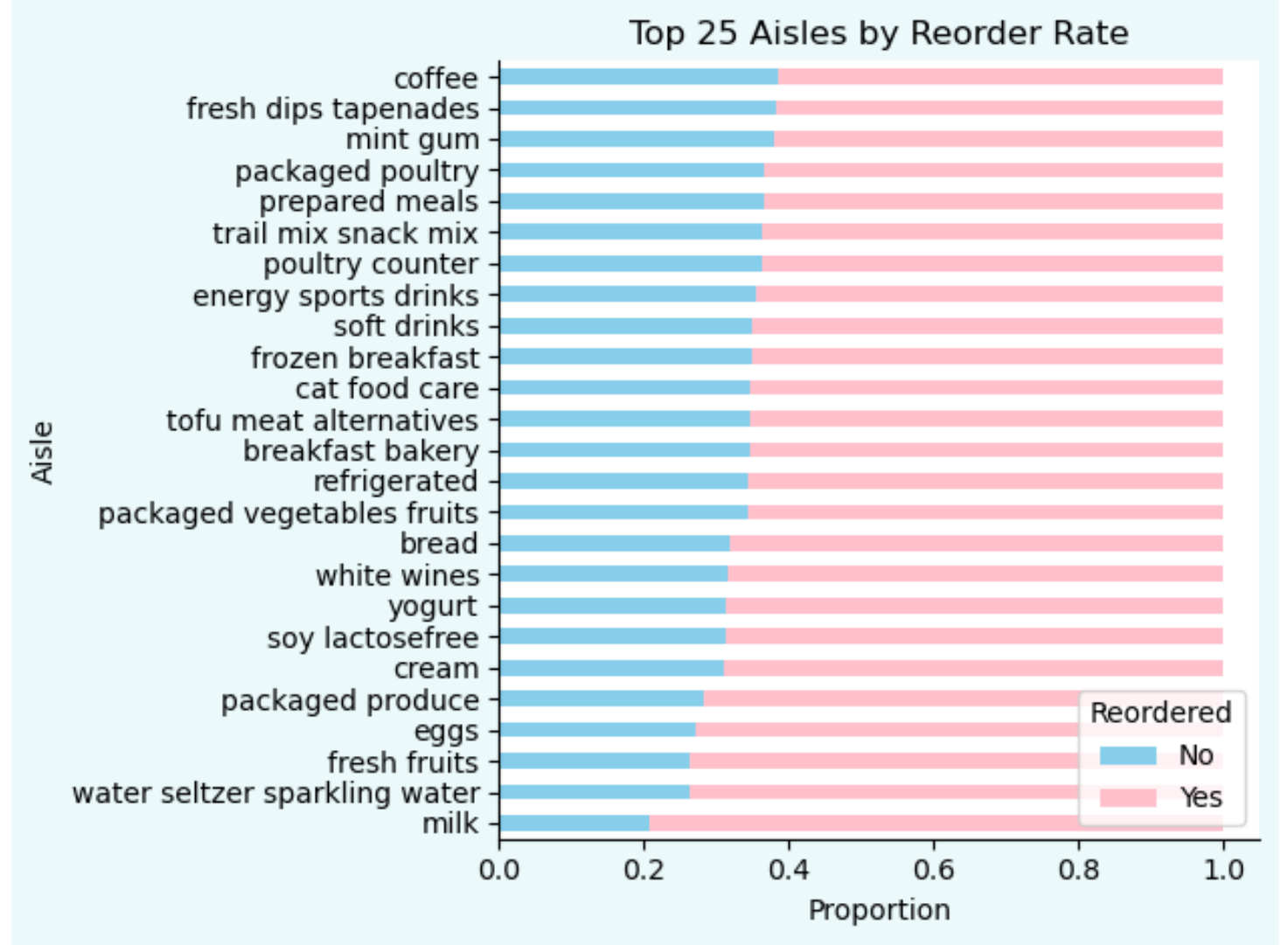
Top-Selling Products:

- Fresh fruits and vegetables, particularly bananas and organic bananas.
- A strong emphasis on organic options.
- A variety of berries and avocados, reflecting a preference for 'superfoods'.

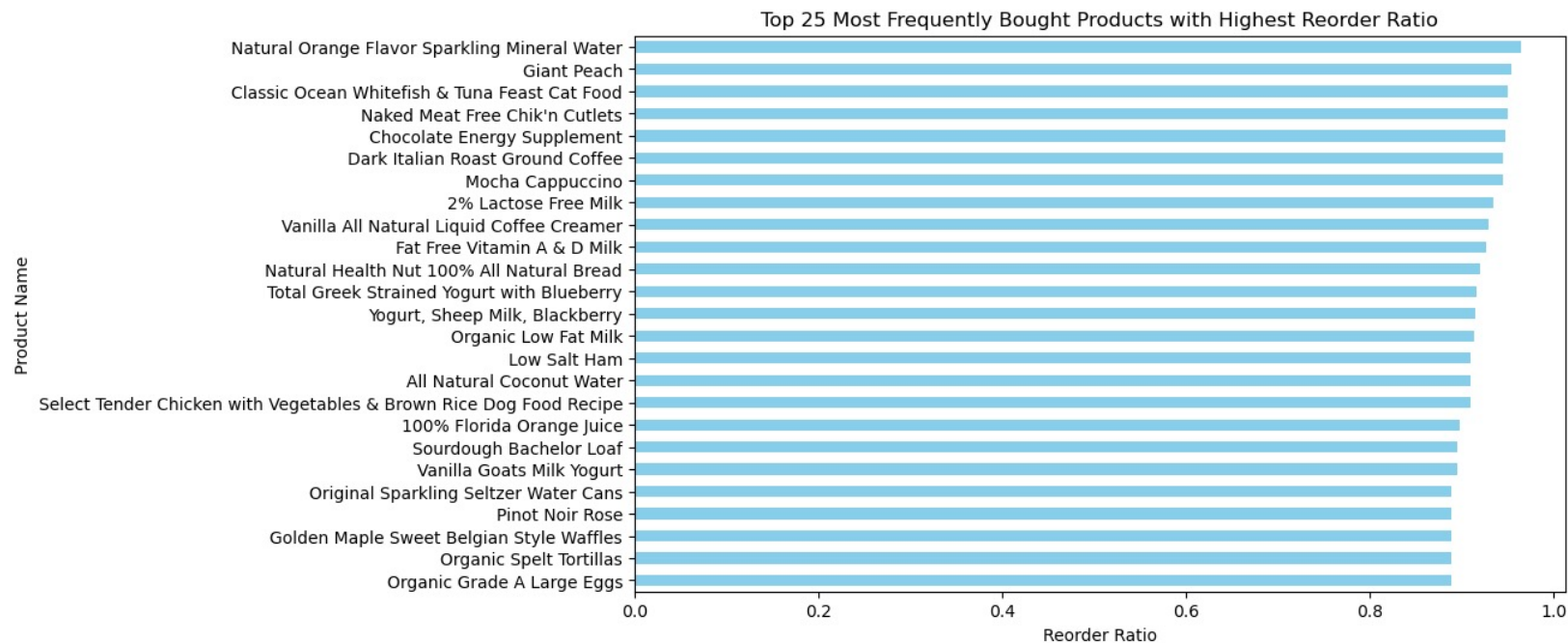
Correlation between when an item is added to the cart and its reorder probability, along with the volume of purchases per department.

Important EDA findings

- Fresh produce aisles (vegetables and fruits) are top choices.
- Dairy products like yogurt, cheese, and milk are popular.
- Snack-related aisles (chips, pretzels) and bread cater to various dietary needs.
- Lactose-free options indicate consideration of dietary restrictions.



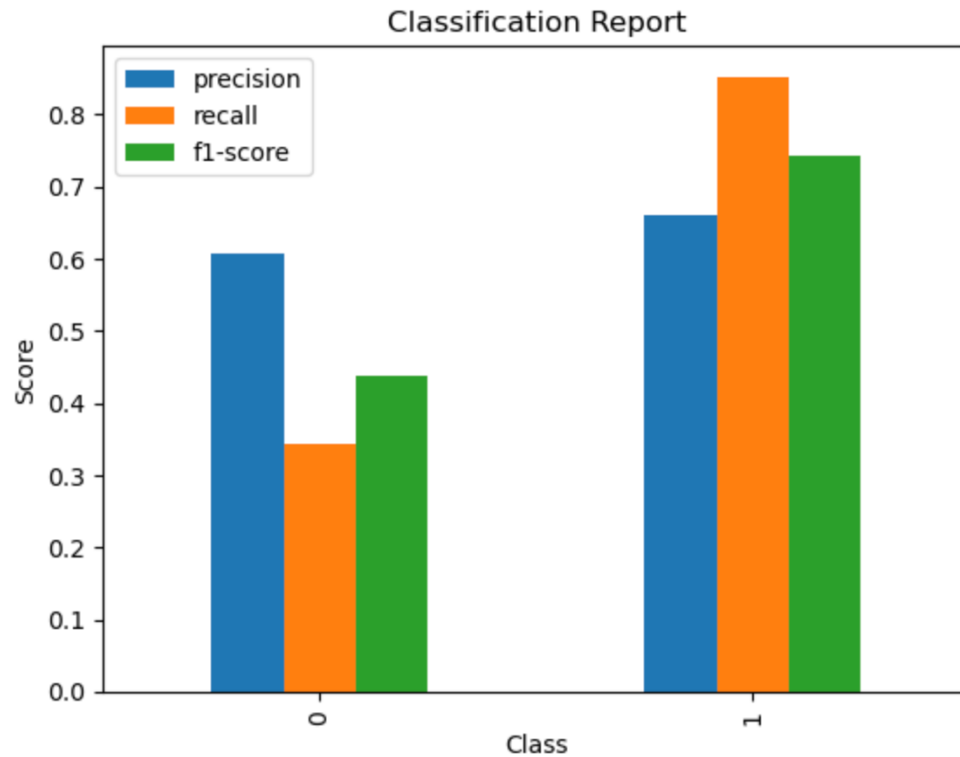
Important EDA findings



Popular and Staple Products:

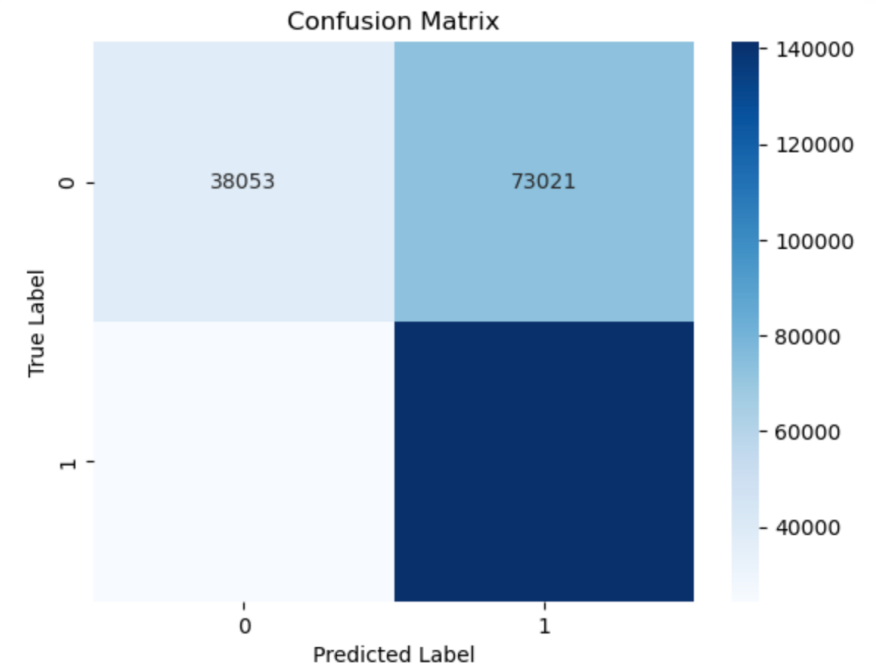
- Assortment includes beverages (e.g., Sparkling Mineral Water, Peach juice, Orange Juice), dairy, dairy substitutes, and groceries.
- These items are likely staples in customers' diets, indicating regular purchases.
- Reasons could include product quality, limited substitutes, or customer habits.

Baseline models and evaluation metrics:



The accuracy is moderately good, but there is definitely room for improvement, especially in terms of precision and recall for the not reordered class.

The model is better at identifying products that will be reordered than those that will not. This could be because there are more reordered instances in the dataset (evidence of class imbalance), or the features are more predictive for the reordered class. The relatively high number of False Positives (products predicted as reordered but actually not) suggests that the model might be erring on the side of predicting reorder.



Next steps

- Clustering and classification (Bag of words)
- Regression analysis
- Machine Learning: KNN and Decision Tree
- Train Machine Learning model



Thank
You