INSTACART ANALYSIS

SHRADHA ATMARAM YEWALE – 014725503

VINEET BATTHINA – 015408887

DHRUV DINESH SONI - 015942927

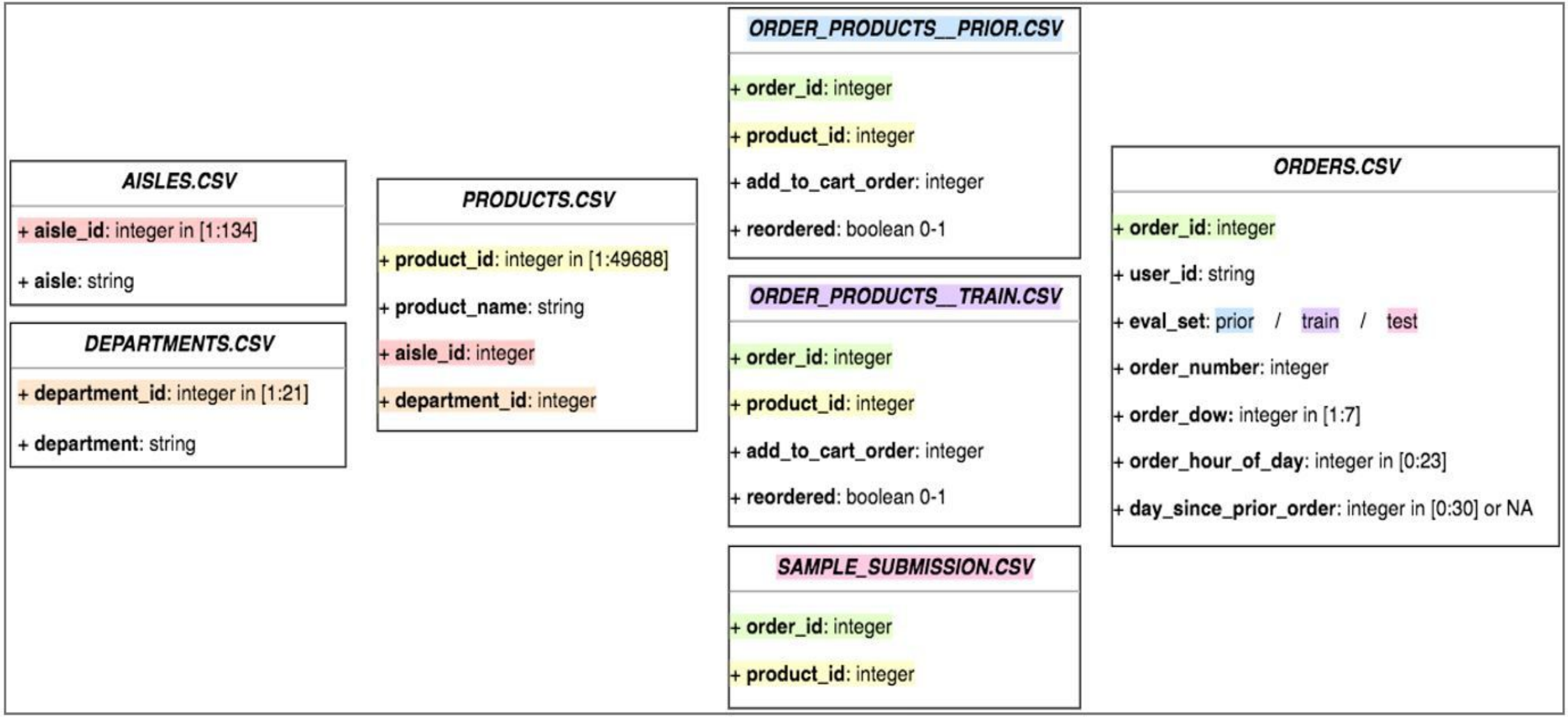
INTRODUCTION

- Instacart provides grocery delivery services
- It is a platform where a customer picks, packs, and delivers the orders within the customer's designated time frame
- Currently they use data to develop models for recommendations

OBJECTIVE

- Predict future behavior of order
- Predict which previously purchased products will be in a user’s next order
- Understand the buying behavior of customer and organise the store and give recommendations accordingly for promotions

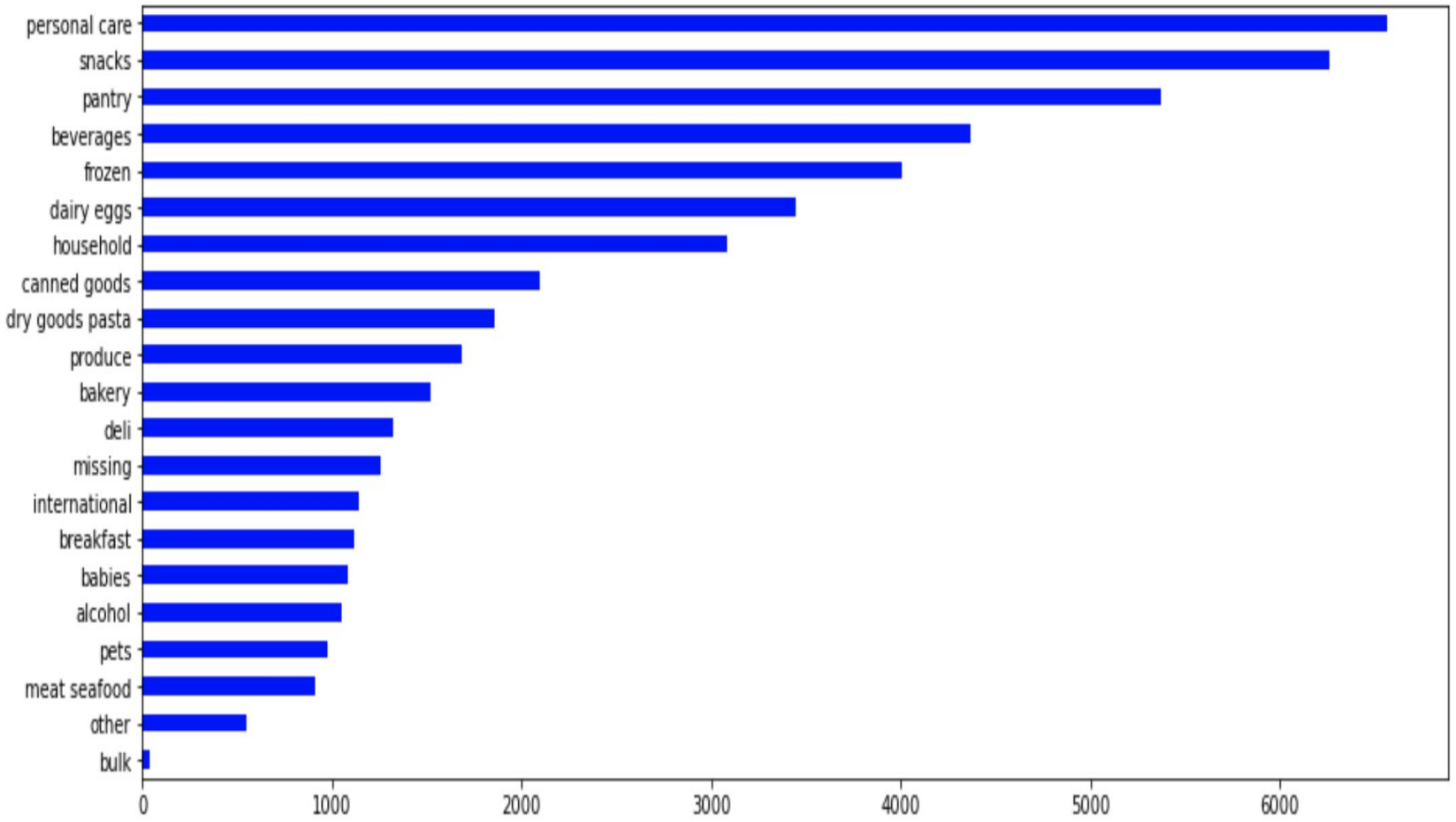
DATASET


AISLES.CSV
+ aisle_id: integer in [1:134]
+ aisle: string
DEPARTMENTS.CSV
+ department_id: integer in [1:21]
+ department: string
PRODUCTS.CSV
+ product_id: integer in [1:49688]
+ product_name: string
+ aisle_id: integer
+ department_id: integer
ORDER_PRODUCTS__PRIOR.CSV
+ order_id: integer
+ product_id: integer
+ add_to_cart_order: integer
+ reordered: boolean 0-1
ORDER_PRODUCTS__TRAIN.CSV
+ order_id: integer
+ product_id: integer
+ add_to_cart_order: integer
+ reordered: boolean 0-1
SAMPLE_SUBMISSION.CSV
+ order_id: integer
+ product_id: integer
ORDERS.CSV
+ order_id: integer
+ user_id: string
+ eval_set: prior / train / test
+ order_number: integer
+ order_dow: integer in [1:7]
+ order_hour_of_day: integer in [0:23]
+ day_since_prior_order: integer in [0:30] or NA

EXPLORATORY DATA ANALYSIS

- Customer orders peek monthly and weekly
- Most ordered products are fruits like bananas, strawberries and organic products
- The fresh food and fresh vegetables aisles are the most frequently visited
- The number of products for personal care, snacks and pantry are the highest

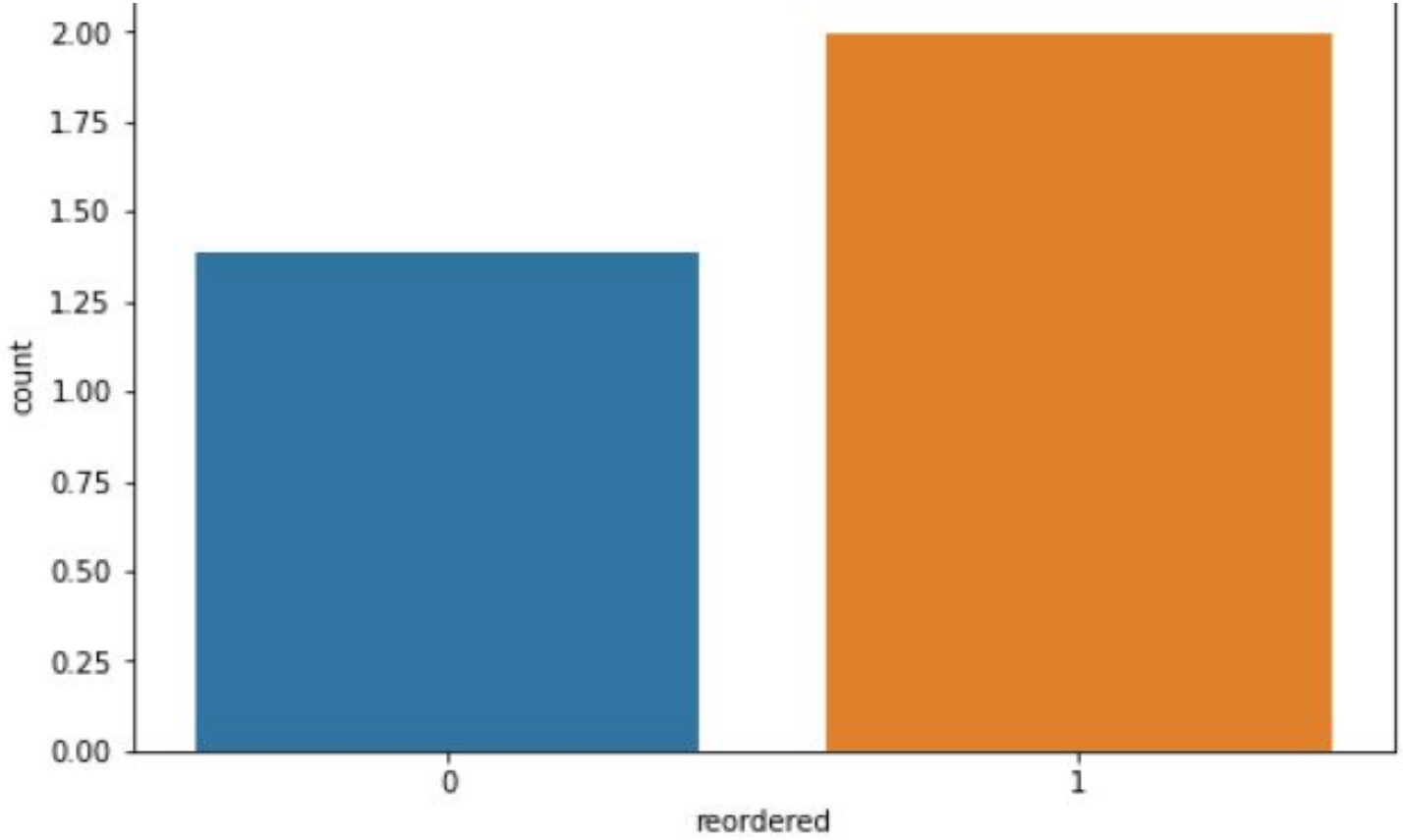
VISUALIZATION



Product Frequency plot according to department

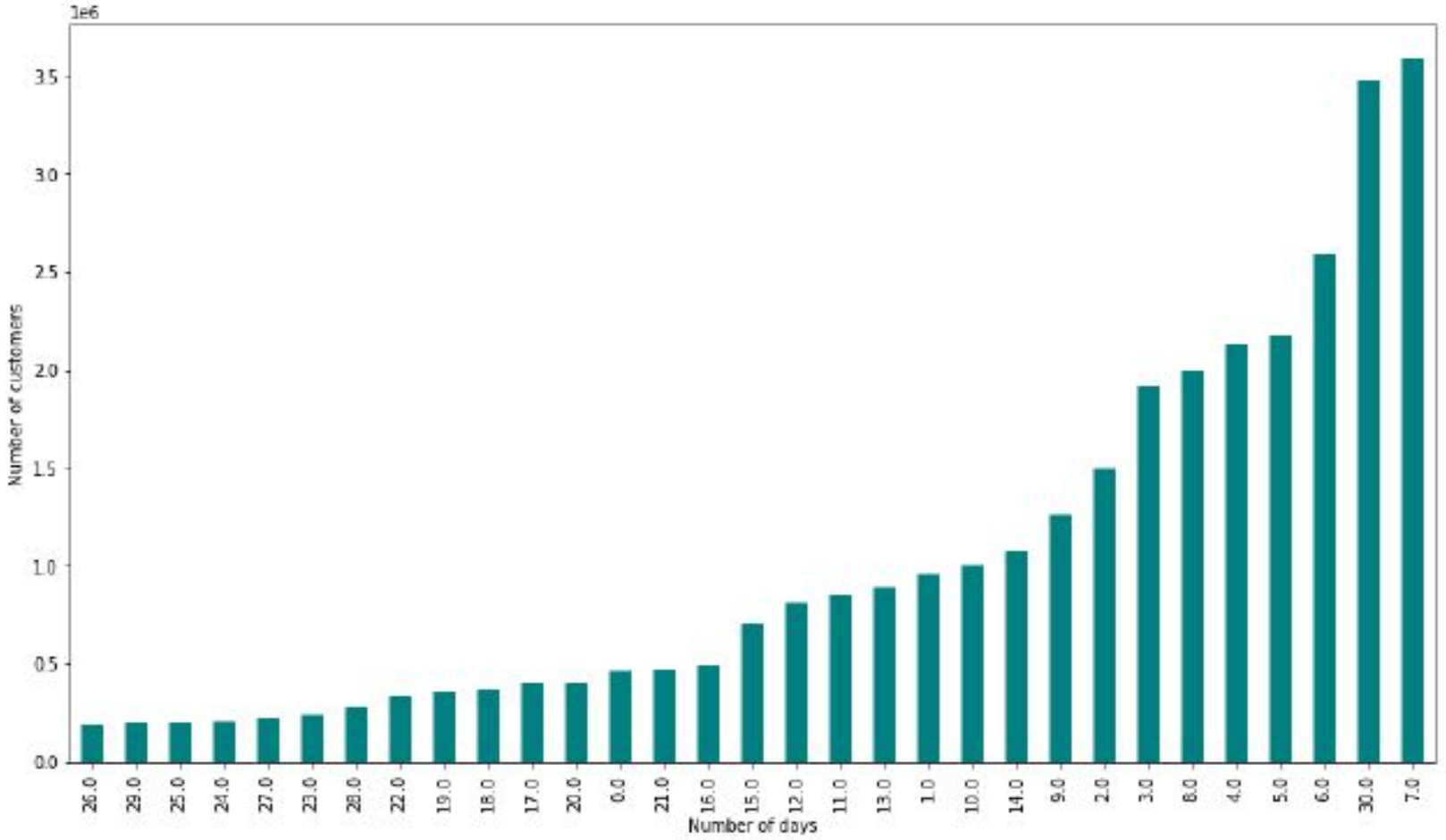
VISUALIZATION

VISUALIZATION



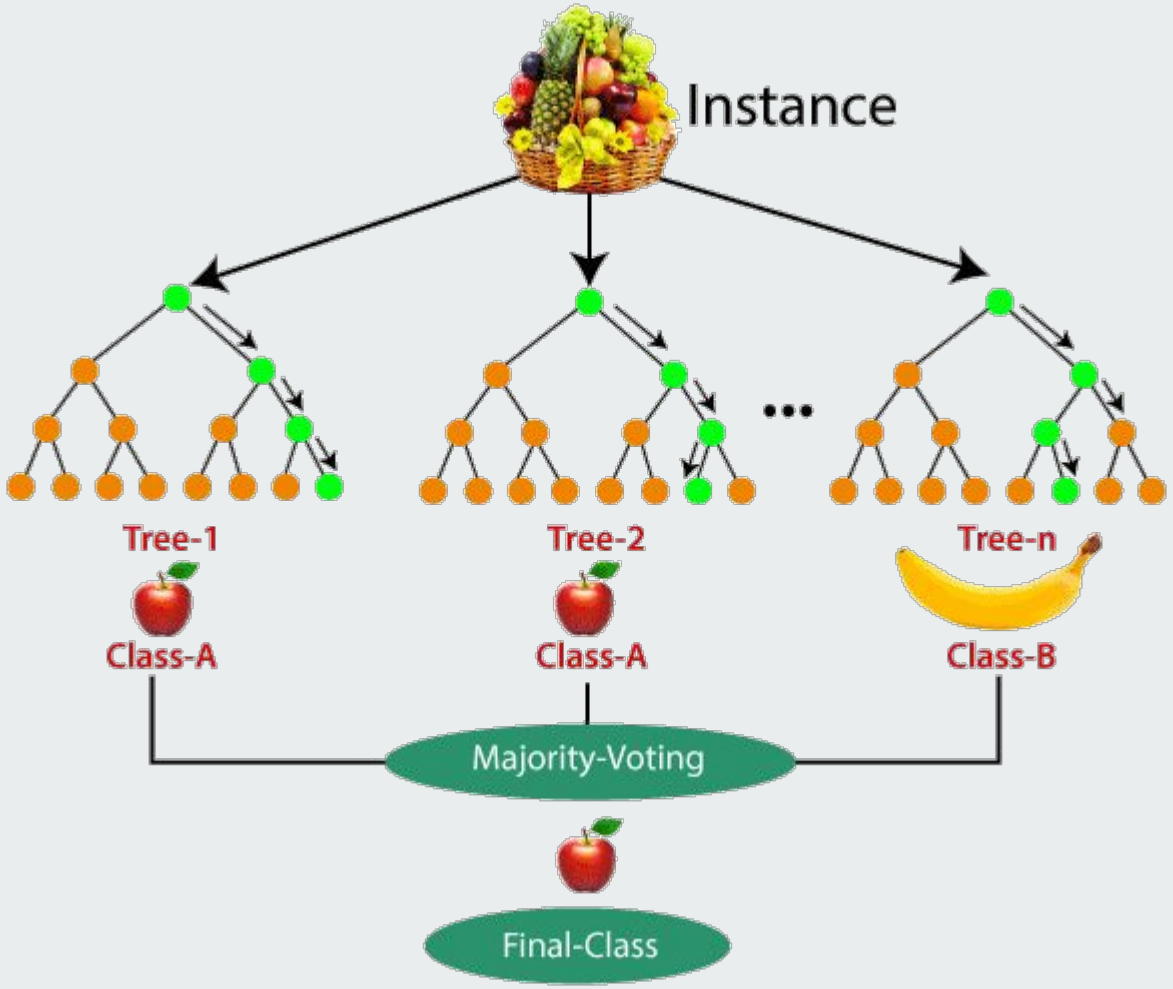
Number of products reordered

VISUALIZATION



Number of days after customer reorders

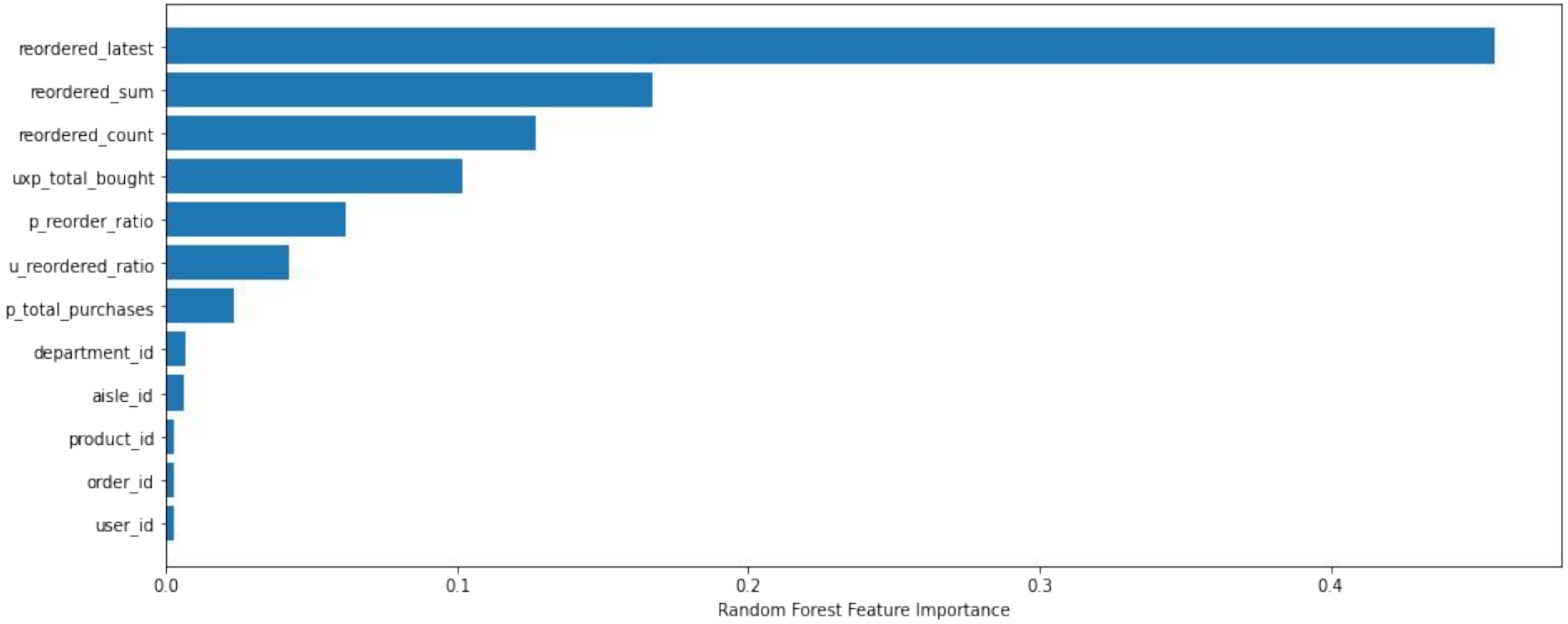
CLASSIFICATION



Random Forest

- Predicting which previously purchased product will be in the customer's next purchase.
- Random Forest model gave 90.45% accuracy
- Implemented using parameters max features as "log2", max depth as 11 and n estimators as 24
- 8 million records
- 33% data used to test the model

TOP FEATURES



ASSOCIATION RULE MINING

- Association rule mining is a technique to identify underlying relations between different items
- The main idea behind Apriori algorithm: All non-empty subsets of a frequent itemset must also be frequent
- Support: Fraction of transactions that contain an itemset
- Confidence: Measures how often items in Y appear in transactions that contain X
- Shortcomings of Apriori algorithm: size of itemset from candidate generation, scanning the itemset database over and over again
- FP Growth

EXAMPLES

A	B
Organic Raspberries	Organic Strawberries
Organic Fuji Apple	Banana
Organic Raspberries	Bag of Organic Bananas
Organic Hass Avocado	Bag of Organic Bananas
Cucumber Kirby	Banana
Organic Avocado	Banana
Organic Strawberries	Bag of Organic Bananas
Strawberries	Banana
Large Lemon	Banana
Organic Baby Spinach	Bag of Organic Bananas
Limes	Banana
Organic Baby Spinach	Banana
Organic Strawberries	Banana

A	B
Non Fat Raspberry Yogurt	Icelandic Style Skyr Blueberry Non-fat Yogurt
Icelandic Style Skyr Blueberry Non-fat Yogurt	Non Fat Raspberry Yogurt
Icelandic Style Skyr Blueberry Non-fat Yogurt	Vanilla Skyr Nonfat Yogurt
Vanilla Skyr Nonfat Yogurt	Icelandic Style Skyr Blueberry Non-fat Yogurt
Total 2% Lowfat Greek Strained Yogurt With Blueberry	Total 2% with Strawberry Lowfat Greek Strained Yogurt
Total 2% with Strawberry Lowfat Greek Strained Yogurt	Total 2% Lowfat Greek Strained Yogurt With Blueberry
Total 2% with Strawberry Lowfat Greek Strained Yogurt	Total 2% Lowfat Greek Strained Yogurt with Peach
Total 2% Lowfat Greek Strained Yogurt with Peach	Total 2% with Strawberry Lowfat Greek Strained Yogurt
Yellow Bell Pepper	Orange Bell Pepper
Frozen Organic Wild Blueberries	Organic Whole Strawberries
Organic Whole Strawberries	Frozen Organic Wild Blueberries
Sparkling Lemon Water	Lime Sparkling Water
Peach Pear Flavored Sparkling Water	Sparkling Water Grapefruit
Sparkling Lemon Water	Sparkling Water Grapefruit
Lime Sparkling Water	Sparkling Water Grapefruit

FUTURE WORK

- Extend our model for analysis of all the users, and also find association between all the products
- More features about the user and the product can be considered, such as the size of the user's cart, the average day of the order, and so on.
- We can use clustering techniques for customer segmentation. This can be used for personalized recommendation.

CONCLUSION

- Random forest classifier - 90.45% accuracy.
- Association Rule Mining results
- Targeted advertisements to customers.
- Physical stores- organise items effectively.

REFERENCES

[1] Annie, L. C. M., & Kumar, A. D. (2012). Market basket analysis for a supermarket based on frequent itemset mining. International Journal of Computer Science Issues (IJCSI), 9(5), 257.
[2] P. Pravallika and K. Narendra, “Analysis on Medical Data sets using Apriori Algorithm Based on Association Rules,” IJSRSET, vol. 4, no. 1, pp. 717–722, 2018
[3] R. Gangurde, D. B. Kumar, and D. S. D. Gore, “Building Prediction Model using Market Basket Analysis,” Int. J. Innov. Res. Comput. Commun. Eng., vol. 5, no. 2, pp. 1302–1309, 2017
[4] M. Kaur and S. Kang, “Market Basket Analysis: Identify the Changing Trends of Market Data Using Association Rule Mining,” Procedia Comput. Sci., vol. 85, no. Cms, pp. 78–85, 2016
[5] R. Kanapaka, “Association Rule Mining using Apriori algorithm For food dataset,” Int. J. Comput. Appl., vol. 112, no. 4, p. 8887, 2015
[6] D. Hunyadi, “Performance comparison of Apriori and FP-Growth algorithms in generating association rules,” Proc. Eur. Comput. Conf., pp. 376–381, 2011.

THANK YOU