HUMBER INSTITUTE OF TECHNOLOGY

AND ADVANCED LEARNING

Capstone Course

BIA-5450-0LA

Instacart Market Basket analysis

Jeet Patel - N01510518

Submitted to : Ammar Al-Qaraghuli

Table of contents

1. Executive Summary:

The report focuses on a detailed analysis of the online delivery service Instacart's operational efficiency, consumer tastes and inventory management. The data consisting of more than 3 million grocery orders is sourced from Kaggle and includes information such as consumer orders, item details etcetera[1].

The purpose of this assignment is to understand the various challenges faced by Instacart and using machine learning algorithms to optimise performance. The report highlights issues such as order cancellation, delays in delivery time, inefficient demand, and supply. By using descriptive, prescriptive, diagnostic, and predictive analytics, insights are gathered on most popular products, reasons for order cancellations and delivery delays[5]. This information is useful to implement innovative solutions for customer retention.

To streamline processes, a new solution design that applies advanced machine learning techniques, like Gradient Boosting Machines and Random Forests, is utilised. The solution integrates seamlessly into Instacart's existing IT architecture, supporting real-time prediction capabilities and personalized product recommendations. Cloud-based scalable infrastructure ensures efficient handling of large volumes of data and model predictions.

This infrastructure is beneficial in providing a user-friendly virtual grocery store layout, real time information on products and electronic shelf-labels[4]. The implemented solutions if effective to bridge the gap between a virtual store and in-store shopping experience.

By periodic model training, refining the models, data augmentation and hypermeter tuning, developed solution can be improved to achieve higher accuracy, precision, and recall, leading to better inventory management and supply chain strategies for Instacart [12].

2. Introduction

This report analyzes different approaches, and algorithms implemented to provide a guide to improve existing inventory management practices and supply chain strategies for Instacart, specifically focusing on predicting item reorders [13]. This forecasting would be beneficial to ensure stock prioritization for the popular items. Moreover, predicting item reorders aids Instacart in optimizing operational efficiency. It prevents issues such as reducing stockout and minimizing overstocking. This leads to optimizing stocking strategies, faster delivery, customer retention and high loyalty. In addition, this process could help strengthen Instacart’s relations with their suppliers, by anticipating accurate demands and overall improvement in coordination [12].

As the online delivery service faced various challenges a combination of descriptive, prescriptive, diagnostic, and predictive analytics is vital to obtain data-driven insights and strategic solutions. This report aims at discussing the business problems, analytical questions, different methods to optimise the overall performance of the virtual grocery store.

3. Business Problem Overview:

Instacart, an important competitor in the online grocery delivery space, pursue to strengthen its operational capabilities and consumer experience. Pivotal challenges

comprise of precisely forecasting goods demand, improving pricing system, reducing order cancellations, and minimising delivery wait times[13]. To understand these challenges, the Instacart Market Basket Analysis focuses on using data-informed insights to enhance supply and demand management, collaboration, and customer retention.
The study focuses on using past performance data, customer taste and preferences, and item details to provide relevant predictions that improves the Instacart shopping experience. By leveraging this information, Instacart wishes to position itself as a customer-driven platform that not only meets but exceeds customer expectations.
Through amalgamation of analytics techniques, this analysis focus on answering important questions pertaining to product requirement, pricing strategies, order cancellations, delivery delays, and customer buying habits [14]. By unveiling these insights, Instacart can carry out strategic solutions that accelerate to enhanced operational efficiency and improved consumer loyalty. The subsequent sections discuss on the analytics questions and potential solutions that this analysis seeks to address.

4. Analytics Questions

4.1.Type: Descriptive Analytics.

Question: Which product is in high demand and has an important contribution towards revenue generation?[3]
Insights: Identify the top-selling product, assess their role in the contribution towards the overall revenue, and identify the customer preference trends [3].
Action: Allocate the resources based on the demand for a particular product, and optimize the inventory based on the demand [3].

4.2.Type: Prescriptive Analytics.

Question: How can Instacart enhance its price to increase its profitability and revenue?[3]
Insights: Analyze the pricing flexibility and know the willingness of customers to pay by providing optimal pricing strategies for different products [3].
Action: Adjust pricing policy so that they can improve their profitability and revenue, bring on new offers like points, or special discounts so that more customers are attracted to buy it [3].

4.3.Type: Diagnostic Analytics.

Question: What are the main reasons behind the cancellation of an order and how to control it?[3]

Insights: To identify the reason behind the cancellation of the order such as out-of-stock of the product, or delay in delivery. Based on the reason behind the cancellation provide an optimal solution [3].
Action: Improve inventory management, improve the delivery process, give some kind of concession to the customer like a discount on the next order, or cashback offer, and address the primary reason behind the cancellation of orders [3].

4.4.Type: Predictive Analytics.

Question: Based on the customer's buying habits can we predict the potential future buying of them?[3]
Insights: Analyse the customer purchase history, geographical location, browsing behavior, and the product inside the cart to generate a predictive model for the customer's future purchases [3].
Action: Optimize the recommendations system based on the customer purchase history, providing an attractive product line-up for the customers, and give points or some offers on the products that a customer buys frequently [3].

4.5.Type: Diagnostic Analytics

Question: What things lead to the delay in the delivery time?[3]
Insights: Based on the number of product orders, supply knows the exact reason behind the delay in the delivery of the order time [3].
Action: Based on the feedback and review given by the customer, implement the necessary steps to be taken like contacting with supplier, or asking for give a detailed description of the product and the address to be delivered [3].

5. Scope Statement

5.1 Product Scope Description:
We are trying to solve the stated business difficulties and requirements, the product scope entails creating and deploying several innovations and functionalities within Instacart's online grocery delivery platform. This entails raising customer happiness, streamlining processes, enhancing partner cooperation, optimizing costs and returns, and scouting out potential new markets [11][1].

5.2 Product Acceptance Criteria:

- Shows a notable improvement in metrics for customer satisfaction, such as a decline in customer complaints and an increase in positive feedback [11][1].
- By introducing improved inventory management systems, barcode scanning, and thorough shopper training, order fulfilment errors are reduced to a minimum [11][1].
- Offers a trustworthy and effective inventory management system that interfaces with affiliate stores to guarantee precise tracking of product availability, automatic stock replenishment, and real-time inventory visibility [11][1].

5.3. Project Deliverables:

- An upgraded online platform with features for better order accuracy, increased inventory control, and sophisticated data analytics [11][1].
- Accurate product information, price, and inventory data are provided by integration interfaces and APIs with partner grocery stores [11][1].
- Customer support systems, such as specialized support staff [11][1].

6. Data sources key data entities and flows

Data from Kaggle Data Sets [1] that includes a sample of over 3 million grocery orders from over 200000 Instacart users and close to 50,000 different products has been gathered by us. These orders are organized in a relational set of files that describe customers' orders over.

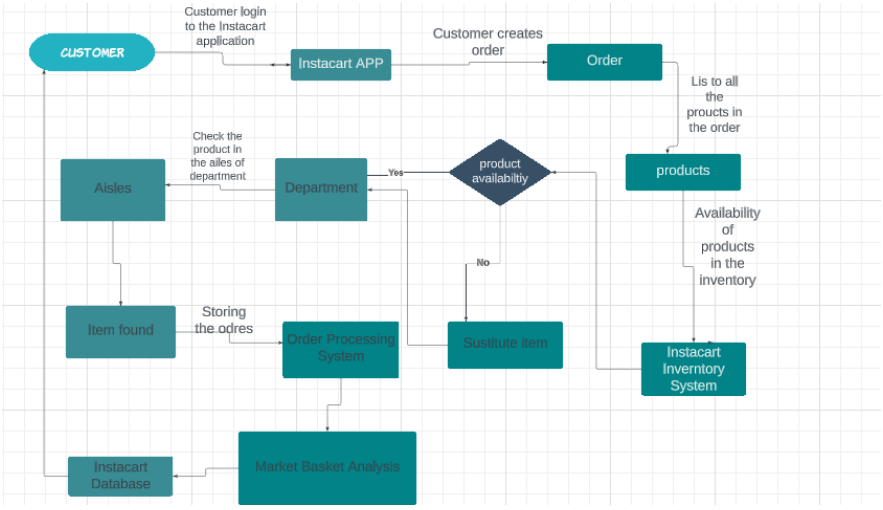
It includes information on when a product was bought, how quickly it was bought in relation to other products, and many other things. Additionally, every entity (customer, item, order, and aisle) has an associated unique identifier. This dataset is a severely skewed subset of the production data from Instacart and includes orders from numerous shops.
Each ID in the dataset is completely random, and none of the values can be traced back to another ID. Additionally, the product and order of orders are included in the information provided about the users. Only goods brought in by numerous customers from various stores are included; retail ID is not given. These databases allow for the discovery of numerous intriguing patterns.

The description of each file is as below:

A. aisles.csv: contains aisle_id, aisle
B. department.csv: contains department_id, department.
C. orders_product_train.csv: contains order_id, product_id, add_to_cart_order, reordered
D. orders_product_prior.csv: contains order_id, product_id, add_to_cart_order, reordered
E. orders.csv: contains order_id, user_id, eval_set, order_number, order_dow, order_hour_of_day, order_since_prior_order

F. products.csv: contains product_id, product_name, aisle_id, department_id g. sample_submission.csv: contains order_id, products
G. sample_submission.csv: contains order_id, product

6.1. Data Flow Diagram:



In the above data flow diagram, the entities are order, customer, department, aisles, products and systems Instacart Inventory system, Market Basket Analysis, Order Processing system and Instacart database:

1. Customer logs into the Instacart mobile app, selects products, and creates an order.
2. In the Order entity process, all the products listed in the order are stored and checked in the Instacart inventory system to verify their availability.
3. If the product is available, the system checks the department data entity and aisles data entity to locate the item.
4. Once the item is found, the order is stored in the order processing system for further processing, including market analysis to predict which products an Instacart consumer is likely to purchase again.
5. The data obtained from the market analysis is stored in the database.
6. The order is sent to the customer, along with predictions and recommendations for future purchases.[1]
7. On the other hand, if the product is not available, the system checks the department data entity and aisles data entity to recommend a substitute item for the order.
8. The substitute item is stored in the order processing system, and the same steps as mentioned above (steps 4-6) are followed for market analysis, storing data, and sending the order to the customer.

Data Files	*Data Entities*
Aisles.csv	• *aisle_id:* Unique identifier for each aisle. • *aisle:* Name or description of the aisle.

Order_Products_Train.csv	• *order_id:* Unique identifier for each order. • *product_id:* Unique identifier for each product. • *add_to_cart_order:* The sequence/order in which the product was added to the cart in that order. • *reordered:* Binary value indicating whether the product has been reordered in a subsequent order.
Order_Products_Prior.csv	• *order_id:* Unique identifier for each order. • *product_id:* Unique identifier for each product. • *add_to_cart_order:* The sequence/order in which the product was added to the cart in that particular order. • *reordered:* Binary value indicating whether the product has been reordered in a subsequent order.
Orders.csv:	• *order_id:* Unique identifier for each order. • *user_id:* Unique identifier for each user/customer. • *order_number:* The sequence/order number of the order for each user (1 for the first order, 2 for the second order, and so on). • *order_dow:* The day of the week the order was placed (0 - Saturday, 1 - Sunday, ..., 6 - Friday). • *order_hour_of_day:* The hour of the day the order was placed (0 - 23).
Departments.csv	• *department_id:* Unique identifier for each department. • *department:* Name or description of the department.
Sample_Submission.csv	• *order_id:* Unique identifier for each order. • *products:* Predicted products for the corresponding order.
Products.csv	• *product_id:* Unique identifier for each product. • *product_name:* Name or description of the product. • *aisle_id:* Unique identifier for the aisle to which the product belongs.

	• *department_id:* Unique identifier for the department to which the product belongs.[1]

7. Brief overview of data manipulation process and data output

7.1. Data Manipulation

To ensure data quality and enhance the analysis for the Instacart Market Analysis, we performed thorough data cleaning, including removing duplicates, merging, and consolidating relevant information from the seven data files, handling missing values, standardizing categorical variables, and addressing any other inconsistencies or errors present in the dataset [8]. There was a total of 7 data files which we loaded in Python so that we can perform data cleaning.

```
In [13]: #Open data base which was saved in Jupyter already

aisles = pd.read_csv('aisles.csv', encoding = "ISO-8859-1")
order_products_train = pd.read_csv('order_products__train.csv', encoding = "ISO-8859-1")
order_products_prior = pd.read_csv('order_products__prior.csv', encoding = "ISO-8859-1")
orders = pd.read_csv('orders.csv', encoding = "ISO-8859-1")
departments = pd.read_csv('departments.csv', encoding = "ISO-8859-1")
sample_submission = pd.read_csv('sample_submission.csv', encoding = "ISO-8859-1")
products = pd.read_csv('products.csv', encoding = "ISO-8859-1")
```

```
In [15]: all_orders = order_products_prior.merge(orders, on="order_id", how="left") \
             .merge(products, on="product_id", how="left") \
             .merge(departments, on="department_id", how="left") \
             .merge(aisles, on="aisle_id", how="left")
```

We have merged four files to make it better for analysis:

Out[16]:

	order_id	product_id	add_to_cart_order	reordered	user_id	eval_set	order_number	order_dow	order_hour_of_day	days_since_prior_order	product_na
0	2	33120	1	1	202279	prior	3	5	9	8.0	Organic Wh
1	2	28985	2	1	202279	prior	3	5	9	8.0	Mich Organic I
2	2	9327	3	0	202279	prior	3	5	9	8.0	Garlic Pov
3	2	45918	4	1	202279	prior	3	5	9	8.0	Coconut Bu
4	2	30035	5	0	202279	prior	3	5	9	8.0	Nat Sweet
...	...	...	...	...	...	...	...	...	...	...	
32434484	3421083	39678	6	1	25247	prior	24	2	6	21.0	Free & C Nat Dishwas Deter
32434485	3421083	11352	7	0	25247	prior	24	2	6	21.0	Organic Sandv Crac Peanut Bu
32434486	3421083	4600	8	0	25247	prior	24	2	6	21.0	All Nat French T St
32434487	3421083	24852	9	1	25247	prior	24	2	6	21.0	Bar
32434488	3421083	5020	10	1	25247	prior	24	2	6	21.0	Organic Sv & Salty Pe Pretzel Gra

32434489 rows × 15 columns

There were many missing values which we have replaced with “NA”. After replacing, here’s the count of the missing values now:

In [21]:

```
# Check for missing values
missing_values = all_orders.isnull().sum()

# Display the count of missing values
print(missing_values)
```

```
order_id                  0
product_id                0
add_to_cart_order         0
reordered                 0
user_id                   0
eval_set                  0
order_number              0
order_dow                 0
order_hour_of_day         0
days_since_prior_order    0
product_name              0
aisle_id                  0
department_id             0
department                0
aisle                     0
dtype: int64
```

To ensure accurate and comprehensive analysis for the Instacart Market Analysis, we

performed a series of data cleaning tasks on all the files. This included replacing column names with more descriptive and accurate labels and addressing any inconsistencies in the values present in the dataset to ensure data integrity and reliability for subsequent analysis [9].

7.2. DATA STORAGE:

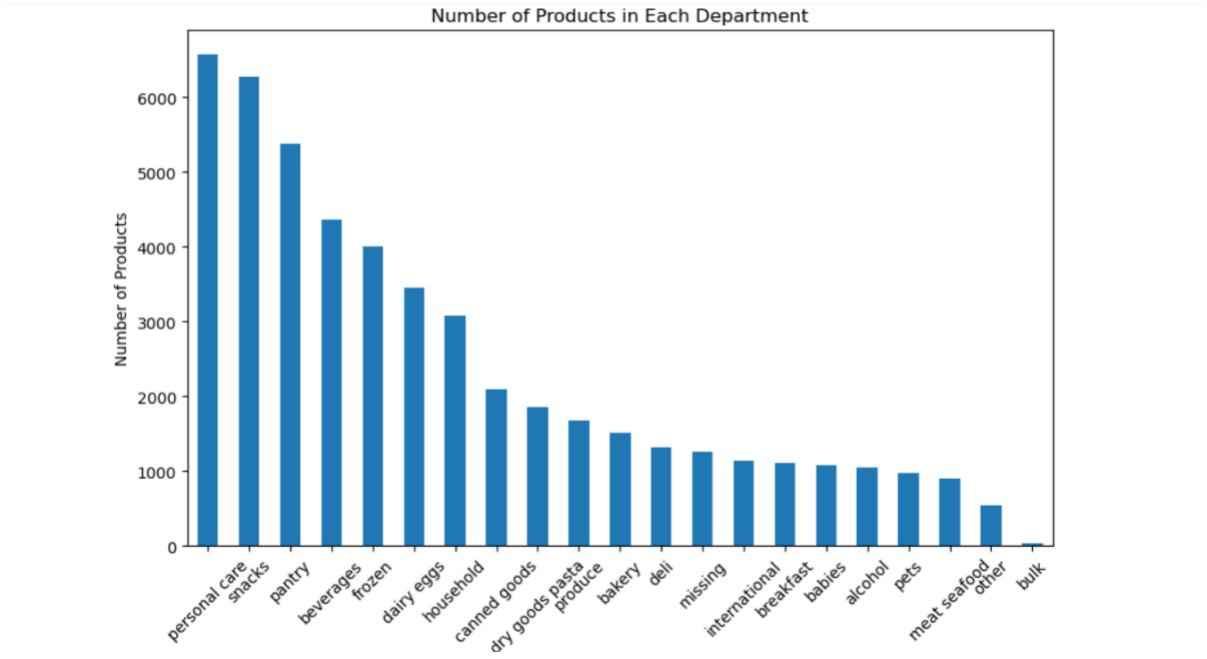
In the data preparation process for data analysis, it is crucial to understand where and how the data is stored during the analysis phase and the security rules for accessing both the source data and the output. This report aims to provide a concise overview of data storage options, security measures, and storage size requirements for different types of data [7].

For Instacart Analysis: The dataset for embedding training is extensive, consisting of 1.02 billion rows, which amounts to over 14 gigabytes (GB). It encompasses a total of 49.6 thousand unique items/products and 3.34 million unique orders.

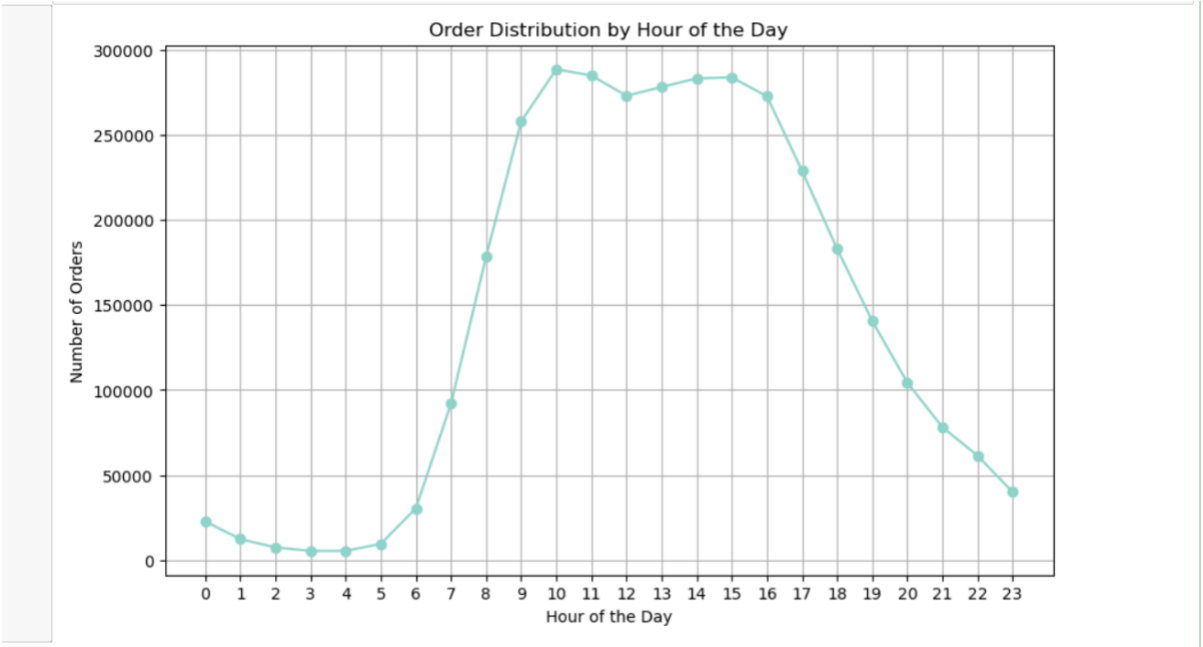
7.3. Data Output:

- We have done data loading into Python and added all the parameters that were required in it.
- Alongside we merge different data sets to get some meaningful insights and output based on the data we had.
- Here are some visualizations and the solutions that we tried to provide.

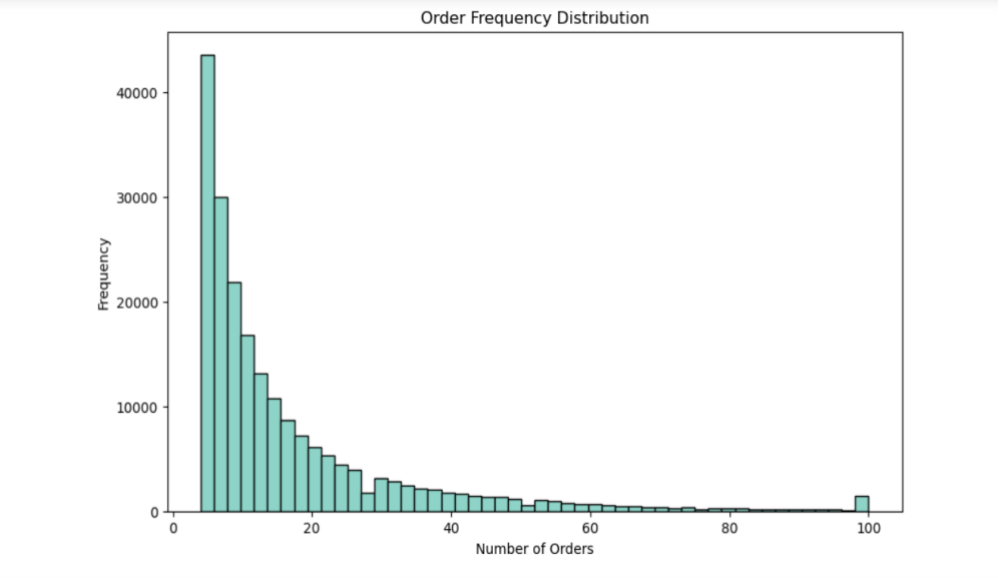
7.3.1. To know the different varieties of products that are in different departments we created a bar chart by executing the following code, we merged to dataset products and departments respectively and then we can see in the graph that the department is sorted based on the product count [7]. Basically, this chart helps us to understand and visualize the available products across different departments.



7.3.2. Next up, is a line chart that we have created to calculate the number of orders for each hour of the day and visualized the basic distribution of the order that has come during the day. We can clearly see the different ratio of orders that have arrived during the peak hour time and, a decline in the order during off hours [7].



7.3.3. To see the order pattern of a customer, we have created a histogram. Here we have calculated the occurrence of an order from different customers and made a frequency distribution of the order count [7]. Moreover, the frequency at the y-axis shows the customers placing an order, and the x-axis shows the no of orders [7]. This histogram helps us in understanding the diverse range of customers placing different orders based on their preferences.



8. New Solution Design and It's Fit into the existing IT architecture:

Our team has performed data visualization and logistic regression for Instacart market basket analysis to predict customer reordering behaviour. Building upon these existing efforts, this proposal outlines a new solution design that integrates advanced machine learning techniques and scalable infrastructure to further improve the accuracy and efficiency of product reordering predictions [6].

1. Data Collection and Preparation: The first step in the new solution design is to enhance data collection and preparation processes. This involves gathering historical order data, customer information, product details, and order timestamps. The data should be cleaned, pre-processed, and transformed into a format suitable for analysis and modelling [6].

2. Feature Engineering: To improve the predictive power of the model, additional features need to be engineered from the raw data. This can include customer-specific features like purchase frequency, average order size, and customer loyalty metrics, product-specific features like popularity, and frequency of appearance in orders, and temporal features like time of day or day of the week [6].
3. Advanced Machine Learning Models: While logistic regression provides valuable insights, it's essential to explore more sophisticated machine learning models to enhance prediction accuracy. Consider incorporating algorithms such as Gradient Boosting Machines (GBM), Random Forests, or Neural Networks. These models can capture more complex patterns in the data and lead to more accurate product reordering predictions [5].
4. Model Evaluation and Validation: A rigorous evaluation and validation process should be established to ensure the new ML models are effective and reliable. Techniques like cross-validation and performance metrics (e.g., precision, recall, F1-score) will measure the models' performance on a test dataset [5].
5. Real-time Prediction: To offer timely and personalized product recommendations, the new solution should support real-time prediction capabilities. As new orders come in, the model should quickly assess the likelihood of reordering for each product and recommend the most relevant ones to customers [5].
6. Scalable Infrastructure: With Instacart's vast customer base and extensive product catalog, the solution must be built on a scalable infrastructure to handle large volumes of data and model predictions efficiently. Consider leveraging cloud-based services like AWS, Google Cloud, or Microsoft Azure for elastic scaling and cost-effectiveness [5].
7. Model Deployment and Integration: The new machine learning models should be integrated into Instacart's existing IT architecture seamlessly. An API-based approach can be employed to facilitate easy communication between the front-end systems and the ML backend. This will enable the real-time product recommendations to be seamlessly integrated into the Instacart user interface [5].
8. Monitoring and Maintenance: Once the new solution is implemented, a robust monitoring and maintenance plan should be put in place. Regularly monitor model performance, data quality, and infrastructure health. Retrain the models periodically using updated data to ensure they remain accurate and relevant [5].

8.1. New Solution Accuracy Output:

```
Accuracy: 0.7055
Classification Report:
              precision    recall  f1-score   support

           0       0.68      0.47      0.56       787
           1       0.71      0.86      0.78      1213

    accuracy                           0.71      2000
   macro avg       0.70      0.66      0.67      2000
weighted avg       0.70      0.71      0.69      2000
```

8.2. Probability Prediction Output:

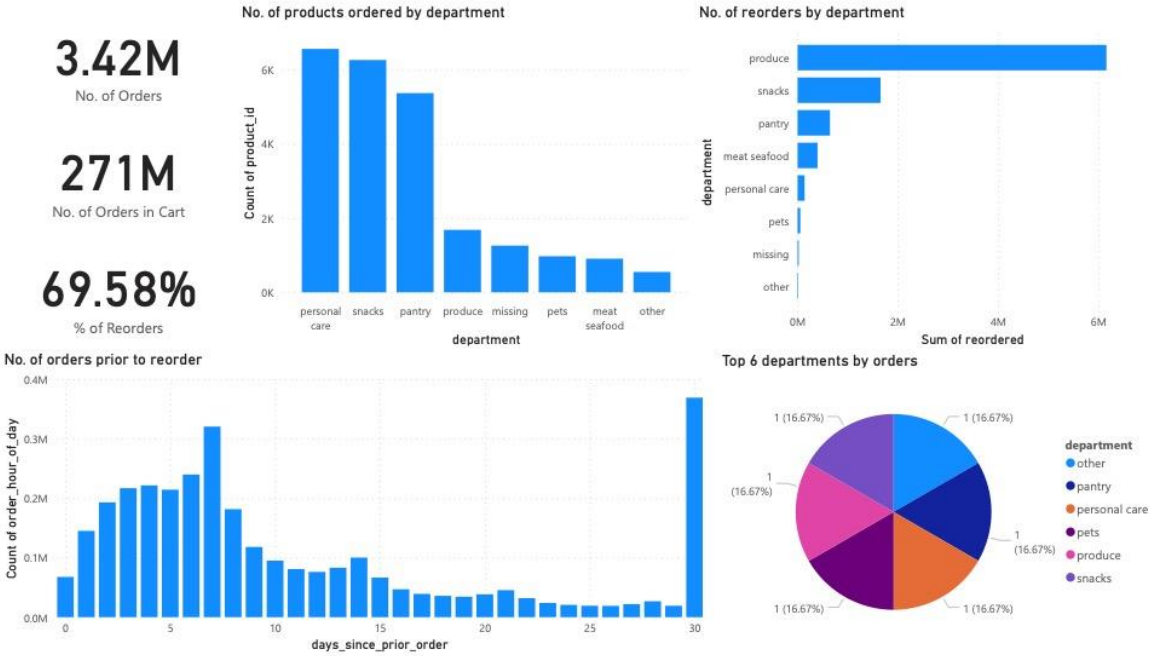
```
Top 10 Most Likely Items to be Reordered:
                                          product_name  reorder_probability
8910                     Organic Ginseng Mandarin Tonic             0.993102
3288                               Organic Raspberries              0.992857
8001                                 Almond Meal/Flour              0.992277
668                          Cozy Chamomile Herbal Tea              0.992055
8773                               Organic Gala Apples              0.991877
5454   Cheese Finely Shredded Mexican Four Cheese Blend             0.991559
8776                             Organic Tomato Cluster             0.990449
3295                                Super Greens Salad              0.989756
6005                         Organic Sweet Potato Fries             0.989301
6004  Dairy Free Coconut Milk Blueberry Yogurt Alter...             0.989075

Top 10 Least Likely Items to be Reordered:
                                          product_name  reorder_probability
3396                               Organic Baby Spinach             0.502118
8384  Hello Morning Blueberry, Banana & Quinoa Oatmeal              0.502076
4514                      Original Tofurky Deli Slices              0.502076
9766                      Organic Whole Grassmilk Milk              0.501480
8036                      Natural Cane Turbinado Sugar              0.500933
1803                            100% Whole Wheat Bread              0.500782
353                               Honey Roasted Turkey              0.500777
1022                                   Homestyle Ranch              0.500416
9767                         Organic Red Radish, Bunch              0.500171
6870                                         Firm Tofu              0.500018
```

By incorporating advanced machine learning models, scalable infrastructure, and real-time prediction capabilities, the enhanced Instacart Market Basket Analysis solution will empower the platform to offer more accurate and personalized product recommendations to its customers. This will not only enhance customer satisfaction but also drive business growth and improve the overall efficiency of the platform.

9. New Solution Implementation and Outcome Testing

9.1. Exploratory Data Analysis (EDA)



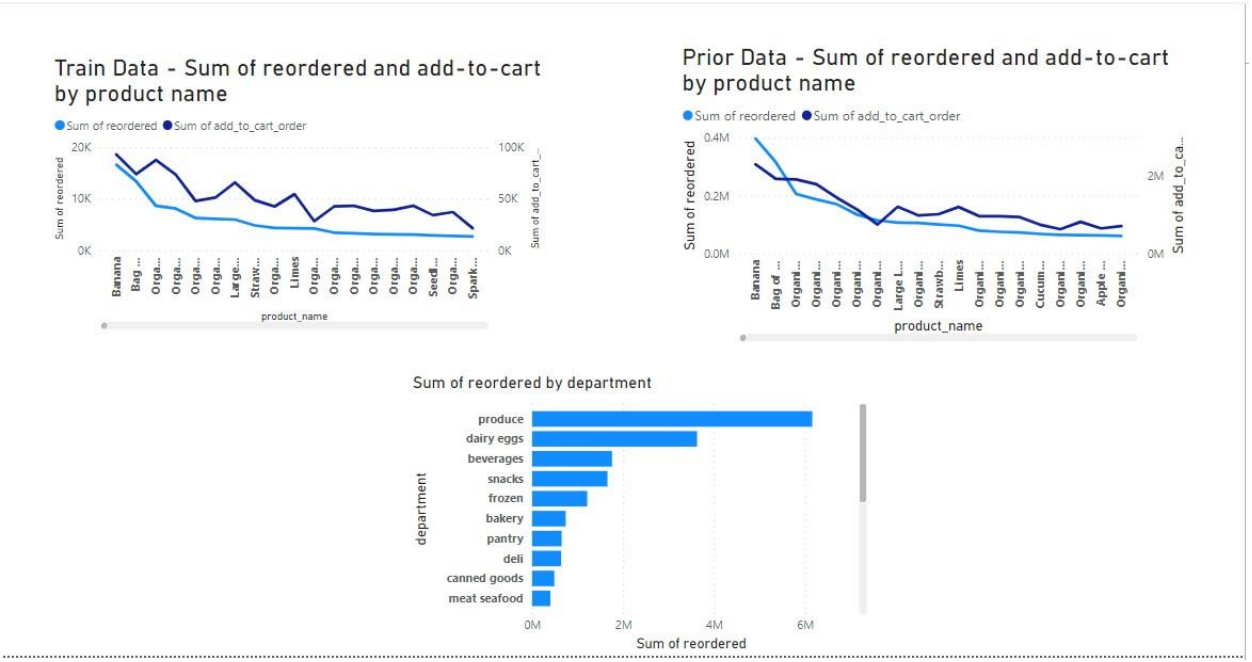
The above shown are a few visualizations that we did to find the buying pattern of a customer by using Instacart. We will go one by one with each visualization.

- Number of Orders: The total number of orders is 3.42 million, To get this figure firstly we selected Card Visualization on our dashboard after that, we went on to the orders table and, selected the column named order_number and that is how we have that figure on the dashboard.
- Number of Orders in Cart: The number of orders in a cart suggests the total number of orders in the cart of all the Instacart customers. To acquire the exact number, we have to select the Card Visualization on the dashboard, then go to order_products_prior table, and select the column named add_to_cart_order. Once we followed the steps, we have 271 million orders that are in the cart of all customers.
- Percent of Reorders: Reorder suggests the average percentage of all the customers ordering the same product again through using Instacart. To get this data we select Card Visualization on the dashboard, then go to order_products_prior table, and select the column named Count of order_id % difference from reordered. Resulting in we get 69.58% of reorders.
- No of reorders by department: This visualization shows the graph above all the departments that are available in Instacart which department was the one where the

most number of reorders were placed by the customers. To get this visualization we first select the stack bar chart on our dashboard, then put the department column from the department's table on the y-axis and the reordered column from the order_products_prior table on the x-axis. Based on the graph we can say that the Produce department has generated the maximum number of reorders compared to other departments.

- No of products ordered by department: The visualization is used in finding the department with the maximum orders based on the departments. In order to get the visuals we select a stacked column chart for our dashboard, then select the department column from the departments table on the x-axis, and count of product_id column from the products table on the y-axis. The visualization clary suggests that personal care is the department with the maximum number of orders followed by snacks.
- Top 6 departments by orders: This graph will be helpful in showing the top 6 departments out of all the departments present in Instacart. To get the visualization we used Pie Chart for our dashboard, after that we selected department and department_id columns from the department's table. The visual suggests that each department has the same amount of contribution which is 16.67%.
- No of orders prior to reorder: This graph shows the day on which the highest number of orders are usually placed on our platform. To get this visualization we use the Slack column chart on our dashboard. From the orders table, we get the day_since_prior_order column and order_hour_of_day columns. Based on the graph we get to see that people will go and order the maximum either at the end of the month or they will buy the most during the beginning of the first week of the month.

9.2. Visual Description of the Outcome Testing



The outcome testing (line graphs) shows the top 15 products that were added-to-cart, reordered and divided into train and prior data.

- Prior Data: The prior data shows the existing IT infrastructure, which shows the fluctuations between the orders in the cart and the percentage of all the customers ordering the same products again. (Left-Line chart)
- Train Data: The train data illustrates the implemented solutions and optimization, which reflects a balanced parallel movement of the sum of re-ordered and added to the cart lines across the product names. (Right-Line chart)

9.3. Implemented Solutions

A. Enhanced Department Ordering:

- The solution organizes products based on departments, streamlining the process of online grocery shopping to make it simpler for customers to find what they need fast. [10].

- Customers may now browse a virtual grocery store with a layout that resembles an actual one and has distinct departments for produce, personal care, beverages, pantry, snacks and many more. [10].

- By enhancing the shopping experience, the solution seeks to reduce customer frustration and improve customer fulfillment and loyalty. [10].

- The function is enhanced to adjust to specific client buying preferences over time, making subsequent transactions even more efficient and personalized. [10].

- The new solution also optimizes product search algorithms, allowing for more accurate and relevant product recommendations based on customer preferences and previous orders. [10].

B. Electronic Shelf Labels:

- This feature provides information like health attributes and whether a product is SNAP (Supplemental Nutrition Assistance Program) eligible. Customers can scan a QR code and get recipes that incorporate each product. [10].

- These Electronic Shelf Labels are integrated with Instacart's online platform, providing real-time updates on product prices, promotions, and availability. [10].
- Customers can access product information, nutritional facts, and customer reviews through the Electronic Shelf Labels, aiding them in making informed decisions while shopping in-store. [10].
- By implementing Electronic Shelf Labels, Instacart aims to bridge the gap between online and in-store shopping, making the process more convenient, informative, and enjoyable for its customers [10].

- Customers who want to shop in-store and have their items delivered or ready for pickup through the app can do so with ease through the app. [3]

9.4. Strategic Alignment of New Solutions

Business Problem	Business Requirement	Data Flow Optimization	New Solutions
Operational efficiency	Inventory management system	Order processing system	Enhanced department ordering
Partner collaboration	Partner integration	Inventory system	Electronic shelf label

10. Potential solution optimization

10.1. Optimization:

To optimize the solution, the following steps can be taken:

1.1. **Feature Engineering**: Consider identifying and incorporating additional relevant features that may improve the model's performance [2].

1.2. **Hyperparameter Tuning**: Experiment with different hyperparameter values for the chosen classification model to find the optimal combination that maximizes performance [2].

1.3. **Model Selection**: Try different binary classification algorithms (e.g., Logistic Regression, Random Forest, Gradient Boosting) to identify the most suitable model for predicting item reorders [2].

1.4. **Ensemble Methods**: Explore ensemble methods, such as stacking or bagging, to combine the predictions of multiple models for better accuracy [3].

1.5. **Data Augmentation**: Consider augmenting the dataset to balance class distributions, especially if there is a significant class imbalance.

1.6. **Cross-Validation**: Implement k-fold cross-validation to evaluate the model's performance on different subsets of the data and reduce overfitting [3].

1.7. **Iterative Refinement**: Continuously review the model's performance, gather feedback from stakeholders, and iteratively refine the solution based on the insights gained.

The implementation of different approaches, algorithms, and models for predicting item reorders using the Instacart dataset provided valuable insights into customer buying patterns and popular products. The binary classification model showed promising results with room for further optimization. By following the outlined client usage instructions and exploring optimization strategies, Instacart can enhance its inventory management practices, minimize stockouts, reduce overstocking, and optimize operational efficiency, leading to improved customer retention and loyalty. The implementation of this solution has the

potential to strengthen Instacart's relationship with suppliers by anticipating accurate demands and enhancing coordination in the supply chain [4].

References

[1] Kaggle. Instacart Market Basket Analysis. https://www.kaggle.com/competitions/instacart-market-basket-analysis/data

[2] Vincent, A.M., & Jidesh, P. (2023). An improved hyperparameter optimization framework for AutoML systems using evolutionary algorithms. Scientific Reports, 13(1), 16001. https://www.nature.com/articles/s41598-023-32027-3

[3] OpenAI. (2023, May 12). ChatGPT. https://chat.openai.com/chat

[4] Rocca, J. (2019, April 22). Ensemble methods: bagging, boosting, and stacking. Towards Data Science. https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205

[5] GitHub. Instacart-Market-Basket-Analysis. https://github.com/archd3sai/Instacart-Market-Basket-Analysis

[6] Gurudath, S. (2020). Market Basket Analysis & Recommendation System using Association Rules. ResearchGate. https://www.researchgate.net/publication/343484851_Market_Basket_Analysis_Recommendation_System_Using_Association_Rules

[7] OpenAI. (2023, May 12). ChatGPT. https://chat.openai.com/chat

[8] Kaggle. Instacart Market Basket Analysis. https://www.kaggle.com/competitions/instacart-market-basket-analysis/data

[9] Red Hat. (n.d.). Red Hat OpenShift: Cloud services. https://www.redhat.com/en/technologies/cloud-computing/openshift/cloud-services

[10] McIntosh, D. (2022, September 19). Introducing Connected Stores: Making Shopping Seamless. Instacart Platform Connected Stores. https://www.instacart.com/company/updates/introducing-connected-stores-making-shopping-seamless/Retrieved August 5th,2023

[11] Usmani, P. F. (2023, May 21). PM Study Circle. Project Scope Statement. https://pmstudycircle.com/project-scope-statement/

[12] Andre Ye (2020, March 30). How Instacart Uses Data Science to Tackle Complex Business Problems? https://towardsdatascience.com/how-instacart-uses-data-science-to-tackle-complex-business-problems-774a826b6ed5

[13] Elizabeth Mixon (2021, March 16). Instacart: Delivering Incredible Customer Experiences with Advanced Analytics and Machine Learning https://www.aidataanalytics.network/data-science-ai/articles/instacart-advanced-analytics-and-machine-learning

[14] Mark Fairhurst (June 13, 2022). Here's How Instacart Platform Is Doing Grocery Businesses More Harm Than Good https://www.mercatus.com/blog/how-instacart-platform-is-doing-grocery-businesses-more-harm-than-good/