**Instacart Market Basket Analysis - Exploratory Data Analysis (EDA)**

**Project Overview**

In the realm of e-commerce and online grocery shopping, understanding customer behavior, product preferences, and purchasing patterns is crucial for enhancing customer experience and driving business growth. This project aims to delve deep into the Instacart dataset, an extensive compilation of data portraying customer orders over time. My objective is to conduct a comprehensive Exploratory Data Analysis (EDA) to unearth insights and patterns within this rich data set.

The Instacart dataset provides a detailed snapshot of customer interactions through various interconnected tables such as orders, products, departments, aisles, order\_products\_prior, and order\_products\_train. With around 3 million records, it offers a vast playground for data exploration and analysis.

**Goals of the Analysis**

Understand Customer Purchasing Behaviors: Analyze how customers interact with the platform and what influences their purchasing decisions.

**Product Analysis:** Explore the popularity and reorder trends of various products and understand how they are influenced by factors like placement in aisles or departments.

**Order Patterns**: Investigate the dynamics of customer orders, including time-based patterns and frequency.

**Data Integration:** Integrate various tables to create a holistic view of the data, facilitating a deeper understanding of how different aspects of the data relate to each other.

Through this EDA, I aspire to provide valuable insights that could aid in making data-driven decisions for business strategy and customer engagement improvements. This analysis serves as a precursor to further predictive modeling, aiming to forecast customer behavior and product trends.

**Dataset Descriptions**

The Instacart dataset is a rich, detailed collection of information related to grocery shopping and customer behavior. It comprises several interlinked tables, each offering unique insights. Here's a breakdown of each dataset:

**1. Orders Dataset**

Description: This dataset forms the backbone of the analysis, containing records of customer orders.

Key Fields: It includes order\_id, user\_id, order\_number (indicating the order sequence for a given customer), order\_dow (day of week), and order\_hour\_of\_day, days\_since\_prior\_order.

Purpose in Analysis: Utilized to understand customer ordering patterns, frequency, and timing preferences.

**2. Products Dataset**

Description: A comprehensive list of products available on the Instacart platform.

Key Fields: Contains product\_id, product\_name, along with aisle\_id and department\_id linking to respective datasets.

Purpose in Analysis: Central to analyzing product popularity, trends, and helping in identifying product-based preferences of customers.

**3. Departments Dataset**

Description: Categorizes products into different departments.

Key Fields: Includes department\_id and department.

Purpose in Analysis: Used to analyze department-wise sales trends and customer preferences, contributing to understanding how product categorization impacts purchasing behavior.

**4. Aisles Dataset**

Description: Details the aisles in which products are placed.

Key Fields: Comprises aisle\_id and aisle.

Purpose in Analysis: Aids in exploring the influence of product placement within store layout on purchasing patterns.

**5. Order\_Products\_Prior Dataset**

Description: Provides detailed historical data of products purchased in prior orders.

Key Fields: Includes order\_id, product\_id, and additional details like add\_to\_cart\_order and reordered status.

Purpose in Analysis: Key to understanding customer’s repurchase habits and the sequence of product addition to carts in past orders.

**6. Order\_Products\_Train Dataset**

Description: Like the Order\_Products\_Prior dataset but typically used for training predictive models.

Key Fields: Contains fields like order\_id, product\_id, with order sequence and reorder information.

Purpose in Analysis: Utilized as a training dataset for predictive modeling and validation of EDA insights.

**BASIC ANALYSIS**

After acquainting ourselves with the datasets, I embarked on an initial exploration to grasp the fundamental aspects of the Instacart customer orders and behavior. This basic analysis lays the groundwork for more intricate explorations later. Here’s a summary of the key questions addressed, and the insights derived:

**Total Number of Orders**

* **Objective:** To determine the scale of transactions in the dataset.
* **Approach:** Counted the total number of unique orders\_id in the orders dataset.
* **Insight:** Totally 3421083 orders were placed. This figure provides a sense of the dataset's volume, crucial for understanding the extent of customer interactions.

**Unique Users**

* **Objective:** To identify the size of the customer base.
* **Approach:** Calculated the number of unique user\_id in the orders dataset.
* **Insight:** There are totally 206209 users. This number gives an idea of the platform's user diversity and reach.

**Reorder Frequency**

* **Objective:** To understand how often customers reorder products.
* **Approach:** Analyzed the days since reorder in the order dataset to calculate the frequency of reorders.
* **Insight:** Reorder frequency helps in understanding customer loyalty and product stickiness.

**Average Number of Orders per Customer**

* **Objective:** To assess how engaged customers are with the platform.
* **Approach:** Computed the average number of orders per unique user\_id in the orders dataset.
* **Insight:** On average each user placed 17 orders. This metric is indicative of user engagement and platform dependency.

**ORDER AND RE-ORDER ANALYSIS**

**Peak Order Times**

* **Objective:** To find out the most popular times for placing orders.
* **Approach:** Aggregated orders based on order\_hour\_of\_day and visualized the distribution using a bar chart.
* **Insight:** The resulting bar chart elucidates the peak hours when most orders are placed, which is crucial for managing logistics, inventory, and customer support.

**Day of the Week Analysis**

Understanding Weekday Trends in Orders

* **Objective:** To uncover patterns in order frequency across different days of the week.
* **Approach:** The dataset originally represented days of the week as numerical values (0 to 6). To make the analysis more intuitive and readable, these numbers were converted to their corresponding weekday names (Sunday, Monday, etc.).
* **Visualization:** A bar chart or similar graphical representation was used to visualize the distribution of orders across different days.

**Insight:**

* This transformation provided a clearer understanding of customer ordering behavior in relation to weekends.
* It helped in identifying any specific days that experienced higher or lower order volumes, which can be crucial for inventory management, staffing, and marketing strategies.

**Average Reorder Rate Analysis**

**Objective**

* The objective of this analysis is to determine the average reorder rate across all orders on the Instacart platform. The reorder rate is a key indicator of customer satisfaction and product appeal.

**Approach**

* The approach taken was to calculate the proportion of reordered items within the total number of ordered items.

**Insight**

* The analysis revealed the average reorder rate of 59%, providing a clear indication of how frequently customers repurchase the same items. This rate is a vital metric for understanding customer buying behavior, product satisfaction, and loyalty.

**Analysis of Frequently Reordered Products**

**Objective**

* The primary objective of this analysis is to identify which products are frequently reordered by customers on the Instacart platform. Understanding the most reordered products can provide insights into customer preferences and product popularity.

**Approach**

* To achieve this, the following steps were undertaken:
* **Data Merging:** The products dataset was combined with the orders dataset to correlate product names with reorder information.
* **Data Filtering:** From this merged dataset, only the instances where reordered == 1 were extracted, focusing specifically on products that have been reordered.
* **Aggregation and Sorting:** The data was then aggregated to count the number of reorders for each product, and the products were sorted based on their reorder count.
* **Visualization:** A visualization, such as a bar chart, was created to display the top 10 most frequently reordered products. This provides a clear and immediate understanding of which products are most popular in terms of reorders.

**Insight**

* The resulting visualization of the top 10 reordered products offers valuable insights into customer purchasing behavior and product performance on the Instacart platform. Products that appear frequently in this list are likely to be essential items that customers prefer to buy regularly. This information can be crucial for inventory management, marketing campaigns, and understanding consumer trends.

**USER ANALYSIS: IDENTIFYING REGULAR AND ONE-TIME SHOPPERS**

**Objective**

* The aim of this part of the analysis was to categorize users into different segments based on their shopping frequency. This included identifying regular users, frequent users (those with more than 75 orders), and one-time shoppers.

**Approach**

* **Regular Users Identification:** Analyzed user ordering patterns to determine which customers consistently place orders over time. This involved looking at the frequency and regularity of their orders.
* **Frequent Users Categorization:** Set a threshold of more than 75 orders to classify users as frequent shoppers. This involved counting the total number of orders per user and identifying those who surpassed this threshold.
* **One-Time Shoppers:** Identified users who have placed only a single order. This was done by filtering users with only one order in their history.

**Insight**

* **Regular Users:** Understanding regular users helps in recognizing a loyal customer base and in tailoring services to enhance their shopping experience.
* **Frequent Users:** The identification of frequent users sheds light on a segment of customers who are highly engaged with the platform. This segment can be targeted for loyalty programs and can provide valuable feedback for service improvement.
* **One-Time Shoppers:** Identifying one-time shoppers is crucial for understanding potential barriers to customer retention. Analyzing this group can reveal insights into why they did not return for subsequent purchases, which is valuable for improving customer acquisition and retention strategies.

This user segmentation analysis is vital for personalized marketing, optimizing the customer experience, and strategic planning. Each segment – regular, frequent, and one-time shoppers – offers unique insights and opportunities for targeted business strategies.

**PRODUCT ANALYSIS: UNDERSTANDING ORDERING PATTERNS AND PREFERENCES**

**Most Ordered Products**

**Objective**

* To identify the most popular products on the Instacart platform, providing insights into customer preferences and product demand.

**Approach**

* Utilized the previously merged table of products and orders.
* Aggregated the data to count the total orders for each product.
* Sorted the products based on their total order count to identify the most ordered ones.
* Created a visualization, such as a bar chart, to display the top 10 most ordered products.

**Insight**

* This analysis highlights the products with the highest demand, which is crucial for inventory management and marketing strategies. Understanding the most popular products helps in stocking them adequately and leveraging their popularity for promotional activities.

**Ranking Departments by Orders**

**Objective**

* To rank the various departments based on the number of orders placed, giving an overview of department-wise customer preferences.

**Approach**

* Combined the departments dataset with all orders dataset to link department names with orders.
* Counted and aggregated the number of orders for each department.
* Created a visualization, like a bar chart, to rank the departments based on the total number of orders.

**Insight**

* The department ranking provides a clear view of which departments are most frequented by customers. This information can guide store layout optimizations, targeted promotions, and stock replenishment priorities.

**Time-Based Product Ordering Analysis**

**Objective**

* To explore if there are specific times or days when certain products are ordered more frequently.

**Approach**

* Analyzed the merged dataset of orders and products to examine ordering patterns across different times of the day and days of the week.
* Utilized a heatmap to visualize these patterns, with one axis representing the time or day and the other representing specific products or product categories.

**Insight**

* The heatmap analysis reveals time-based trends in product ordering, such as certain products being preferred at specific times of the day or on days. This can inform targeted marketing campaigns (like time-specific promotions) and help in anticipating demand fluctuations for better inventory management.

**CONCLUSION**

**Overview of the Project**

This exploratory data analysis (EDA) project on the Instacart dataset provided deep insights into customer purchasing patterns, product popularity, and user behavior. By dissecting various aspects of the data – from order frequencies and reorder trends to user segmentation and product demand – the analysis painted a comprehensive picture of the shopping dynamics on the Instacart platform.

**Key Findings**

* User Behavior: Segmentation of users into regular, frequent, and one-time shoppers highlighted diverse customer engagement levels.
* Product Popularity: Identification of the most ordered and frequently reordered products revealed customer preferences and essential items.
* Order Patterns: Analysis of order frequencies across different days and times, along with department-wise order rankings, offered valuable insights into customer shopping habits.

**Skills and Methodologies Employed**

* Data Wrangling and Preprocessing: The project began with extensive data cleaning and preparation, crucial for handling large datasets and ensuring accurate analysis.
* Data Merging and Integration: Combining multiple datasets was a key aspect, enabling a holistic view of the interconnected data.
* Exploratory Data Analysis: Employed a variety of EDA techniques, including statistical summaries, aggregation, and trend analysis.
* Data Visualization: Made extensive use of visualizations such as bar charts and heatmaps, which were instrumental in revealing patterns and conveying insights clearly.
* Programming and Tools: The project was primarily executed using Python, leveraging libraries such as Pandas for data manipulation and Matplotlib/Seaborn for visualization.
* Analytical Thinking: Critical to the project was the ability to interpret data, draw meaningful conclusions, and understand the implications of the findings in a real-world business context.

**Impact of the Project**

* The insights derived from this project have the potential to significantly influence business strategies for Instacart. By understanding customer preferences and behavior, Instacart can optimize its inventory, tailor its marketing strategies, and enhance the overall customer experience. Additionally, the findings can inform predictive modeling efforts, aiming to forecast future trends and customer needs.

**Future Scope**

* Predictive Modeling: The EDA sets the stage for advanced predictive analytics and machine learning models to forecast customer behavior.
* Personalized Marketing Strategies: Utilizing the insights for creating targeted marketing campaigns and personalized shopping experiences.
* Continuous Improvement: Regular updates to the analysis as new data becomes available to keep abreast of evolving customer trends.