

Instacart Market Basket Analysis

Introduction

Instacart is an American company that operates a grocery delivery and pick-up service in the United States and Canada.

Instacart, a grocery-ordering and delivery app, aims to make it easy to fill your refrigerator and pantry with your personal favorites and staples when you need them. After selecting products through the Instacart app, personal shoppers review your order and do the in-store shopping and delivery for you. The company offers its services via a website and mobile app.

Business Problem

Market Basket Analysis is a data mining technique that is very much helpful to increase sales, helps us to give a better understanding of customer purchasing techniques. Using the purchase history done over a period we try to suggest the items to the customer that could be reordered.

The major benefits would be an increase in sales and customer satisfaction. Using the data we try to determine the products that are brought together with which retailers can optimize product placement, they can provide better deals which would encourage the customers to purchase the items which they had not thought of buying. This would thus help increase sales.

Market Basket analysis also allows companies to identify the important products and frequently purchased products which could potentially hurt their business if they are unavailable. For example, let's say we have predicted that so and so the product would be a part of these many users' next order. This would help the companies to be aware of the availability of the product and make sure it's available.

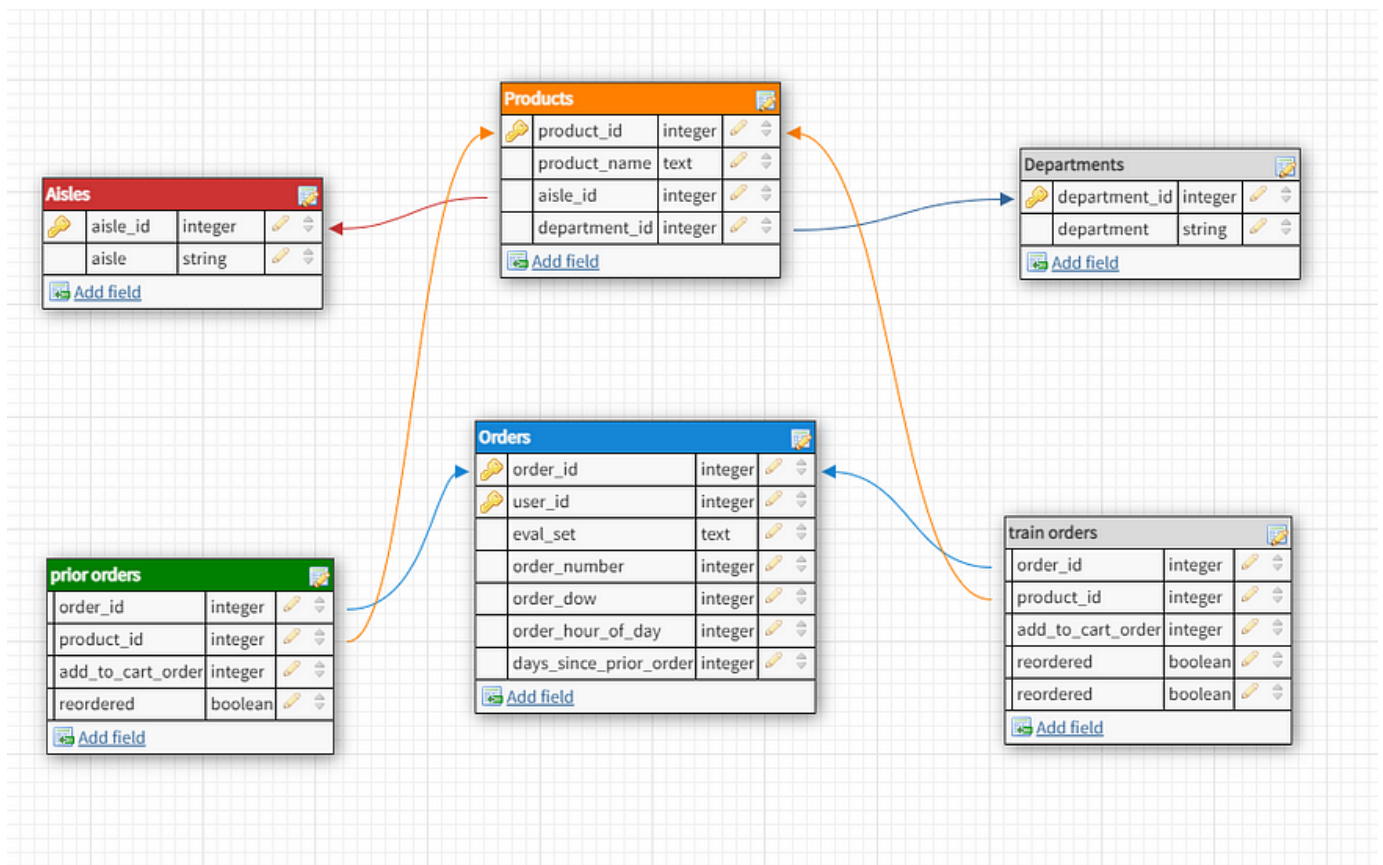
Dataset Description

The dataset is anonymized and contains a sample of over 3 million grocery orders from more than 200,000 Instacart users. For each user, we provide between 4 and 100 of their orders, with the sequence of products purchased in each order. We also provide the week and hour of day the order was placed, and a relative measure of time between orders. or more information, see the [blog post](#) accompanying its public release.

The dataset consists of several CSV files, including:

1. [orders.csv](#): Contains information about the orders placed by customers.
2. [products.csv](#): Contains information about the products available on Instacart.
3. [order_products__prior.csv](#): Contains information about products in prior orders.
4. [order_products__train.csv](#): Contains information about products in the training set of orders.
5. [aisles.csv](#): Contains information about product aisles (categories).
6. [departments.csv](#): Contains information about product departments.

These tables are related to each other as follows:



Data Dictionary

Refer to the provided data dictionary for a detailed description of each dataset and its columns.

Project Requirements

The goal of this project is to gain comprehensive insights into customer behavior, product performance, and purchasing patterns using transactional data. The project will be divided into several key stages, including data exploration and preprocessing, exploratory data analysis (EDA), and business questions and analysis. The following are the detailed requirements:

1. Data Exploration and Preprocessing

a. Load the Data:

- Load the CSV files into pandas DataFrames.
- Check for missing values and data types.

b. Data Cleaning:

- Handle missing values.
- Convert data types (e.g., datetime conversion).
- Merge DataFrames as needed (e.g., joining order_products with products).

c. Basic Descriptive Statistics:

- Calculate summary statistics for numerical features.
- Examine the distribution of categorical features.

2. Exploratory Data Analysis (EDA)

a. Customer Behavior:

- Average number of orders per user.
- Average time between orders for each user.
- Number of orders placed by each customer.
- Customer segments based on purchase frequency.

b. Product Analysis:

- Identify most popular products by frequency.
- Determine average order size (number of items per order).

c. Temporal Patterns:

- Analyze orders by day of the week and hour of the day.

d. Basket Analysis:

- Identify most frequently co-purchased items.
- Products often bought together on weekends vs. weekdays.

3. Business Questions and Analysis

a. Popular Products:

- Analyze sales distribution of top-selling products.
- Identify top 5 products commonly added to the cart first.
- Top 10 product pairs frequently purchased together.

b. Reorder Behavior:

- Analyze the reordered column to understand repeat purchase behavior.
- Products reordered the most.
- Reorder behavior based on day of the week and days since prior order.
- How the number of items in the cart impacts the likelihood of reordering.

c. Customer Segmentation:

- Segment customers based on the total amount they've spent on orders.

d. Department and Aisle Analysis:

- Best-selling department and aisle breakdown.
- What is the “produce” department? Break it down by aisle.
- Differences in purchasing behavior based on different departments or aisles.

By addressing these requirements, the project aims to provide a deep understanding of customer purchasing behavior, optimize product offerings, improve customer retention strategies, and enhance predictive capabilities for future planning.