Amazon Fresh Analytics

This presentation demonstrates the use of SQL for data modeling, definition, manipulation, constraints, aggregations, joins, subqueries, normalization, and real-world analysis on the Amazon Fresh dataset.

The project is structured into tasks, covering:

- Data Modeling and Basic Queries
- DDL, DML, and Constraints
- Aggregations, Joins, and Subqueries
- Normalization and Real-World Insights

Approach:

01	02	03
Task description	SQL query used	Insight gained from execution

Objective:

To showcase a complete end-to-end SQL workflow, from database design and cleaning to advanced queries, and provide actionable business insights for Amazon Fresh.

Task 1 – ER Diagram

Task Description:

Create an ER diagram for the Amazon Fresh database to understand the relationships between tables (Customers, Products, Orders, Suppliers, Order_Details, Reviews).

Insight:

The ER diagram shows the following relationships:
https://drive.google.com/file/d/1EBbyRp9fBNOiN2cYHDiSW3RR7DFBqBcN/view?usp=sharing

- Customers place multiple Orders.
- Each Order contains multiple Order_Details linked to Products.
- Products are supplied by Suppliers.
- Customers provide Reviews for Products.

Primary Keys and Foreign Keys

Customers

PRIMARY KEY (CustomerID)

Orders

PRIMARY KEY (OrderID)

FOREIGN KEY (CustomerID) REFERENCES Customers(CustomerID)

Order_Details

PRIMARY KEY (OrderDetailID)

FOREIGN KEY (OrderID) REFERENCES Orders(OrderID)

FOREIGN KEY (ProductID) REFERENCES Products(ProductID)

Products

PRIMARY KEY (ProductID)

FOREIGN KEY (SupplierID) REFERENCES Suppliers(SupplierID)

Suppliers

PRIMARY KEY (SupplierID)

Reviews

PRIMARY KEY (ReviewID)

FOREIGN KEY (ProductID) REFERENCES Products(ProductID)

FOREIGN KEY (CustomerID) REFERENCES Customers(CustomerID)

Insight:

- Each table has a unique primary key for record identification.
- Relationships are maintained via foreign keys:
  - Orders link to Customers.
  - Order_Details link to Orders and Products.
  - Products link to Suppliers.
  - Reviews link to both Customers and Products.
- This ensures consistency across entities and enables meaningful joins for analysis.

Task 3 — Querying Customers and Products

Task Description

- Retrieve all customers from a specific city.
- Fetch all products under the "Fruits" category.

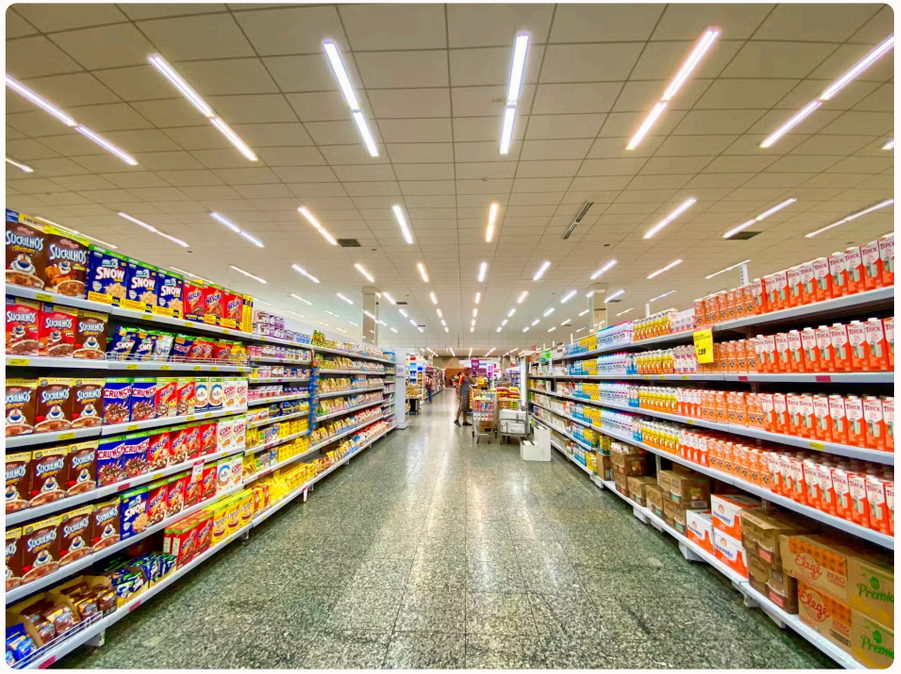
Query

```
-- Customers from a specific city
SELECT * FROM Customers WHERE City = 'East Ronald';

-- Products under Fruits category
SELECT * FROM Products WHERE Category = 'Fruits';
```

Insight

- Found 3 customers located in East Ronald (India) with varied ages and Prime membership status.
- Identified 89 fruit products in the catalog across multiple sub-categories (e.g., Blue Fruit, Popular Fruit, However Fruit).
- Fruits appear to be a major category in the product portfolio, suggesting strong customer choice coverage.



Recreate Customers Table with Constraints

Task Description

Write DDL statements to recreate the Customers table with the following constraints:

- CustomerID as the Primary Key.
- Age cannot be null and must be greater than 18 (CHECK constraint).
- Add a UNIQUE constraint for Name.

Query

```
CREATE TABLE Customers (
    CustomerID VARCHAR(36) PRIMARY KEY,
    Name VARCHAR(100) UNIQUE,
    Age INT NOT NULL CHECK (Age > 18),
    Gender VARCHAR(10),
    City VARCHAR(100),
    State VARCHAR(100),
    Country VARCHAR(100),
    SignupDate DATE,
    PrimeMember VARCHAR(3) DEFAULT 'No'
);
```

Insight

1

Every customer has a unique ID and unique name.

2

Invalid ages (≤18 or NULL) are not allowed, improving data quality.

3

Default value "No" ensures Prime membership field consistency.

Successfully applied data integrity constraints ensuring better data quality and consistency.

Insert 3 New Products

Task Description

Insert 3 new rows into the Products table using INSERT statements.

```
INSERT INTO Products (ProductID, ProductName, Category, SubCategory, PricePerUnit, StockQuantity, SupplierID)
VALUES
('P101', 'Organic Apple', 'Fruits', 'Fresh Fruits', 120, 500, 'S001'),
('P102', 'Brown Bread', 'Bakery', 'Breads', 45, 300, 'S002'),
('P103', 'Almond Milk', 'Dairy', 'Milk Substitutes', 180, 200, 'S003');
```

Insight

Organic Apple

Fruit category with 500 units in stock

Brown Bread

Bakery category with 300 units in stock

Almond Milk

Dairy category with 200 units in stock

Catalog diversity improved across Fruits, Bakery, and Dairy categories, helping expand the product range for customers.

Update Product Stock Quantity

Task Description

Update the stock quantity of a product where ProductID matches a specific ID.

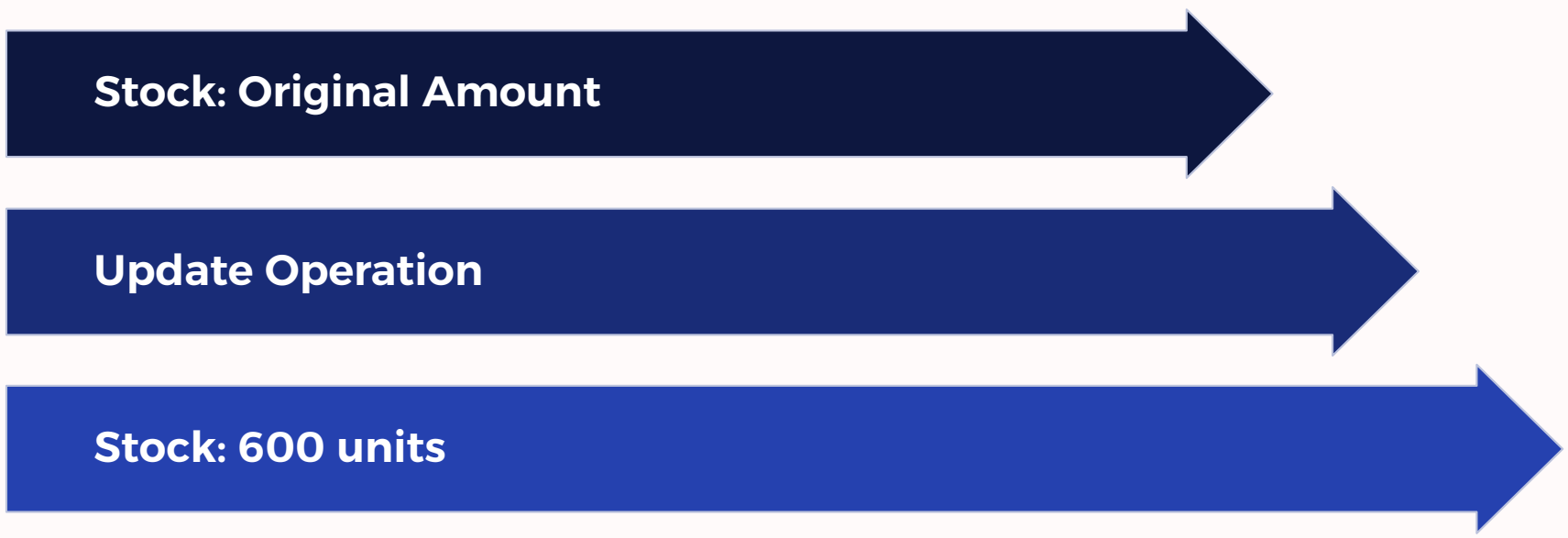
Query

```
UPDATE Products
SET StockQuantity = 600
WHERE ProductID = '0006853b-74cb-44a2-91ed-699aa31c55b5';
```

Insight

The stock for the selected product was successfully updated to 600 units.

This confirms that inventory levels can be dynamically adjusted, ensuring accurate tracking of product availability.



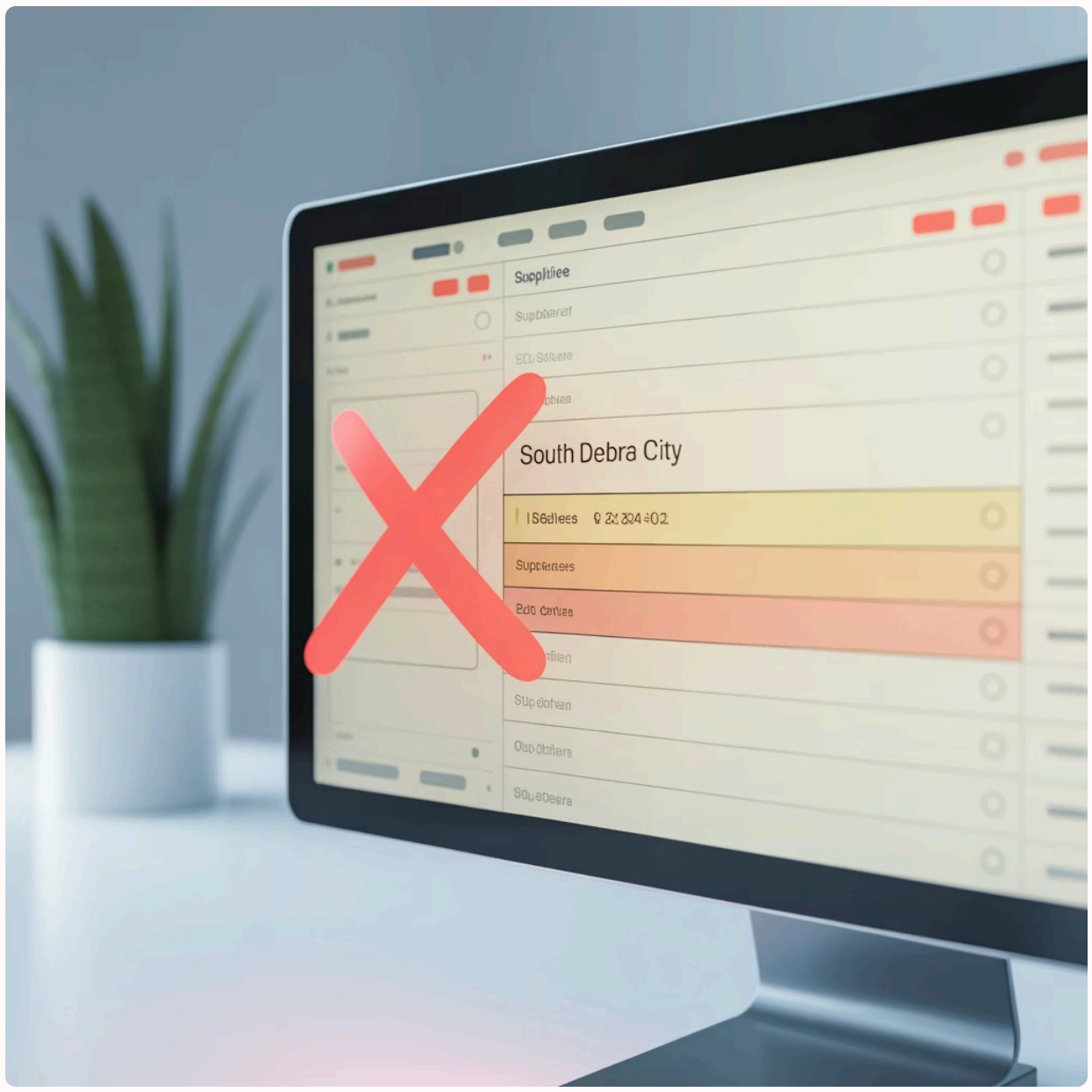
Delete Supplier by City

Task Description

Delete a supplier from the Suppliers table where their City matches a specific value.

Query

```
DELETE FROM Suppliers
WHERE City = 'South Debra';
```



Insight

- The query executed successfully, and 1 supplier was deleted from the dataset.
- This confirms that there was at least one supplier located in South Debra, and the record was removed.
- Targeted cleanup queries like this help maintain accurate supplier data by removing entities from irrelevant or outdated locations.

SQL Constraints and Operators

Description

- Ensure data quality by enforcing constraints.
- Add a CHECK constraint for review ratings to be between 1 and 5.
- Add a DEFAULT constraint for the PrimeMember column in Customers.

Query

```
-- Ensure Ratings are between 1 and 5
ALTER TABLE Reviews
ADD CONSTRAINT chk_rating
CHECK (Rating BETWEEN 1 AND 5);

-- Default value for PrimeMember
ALTER TABLE Customers
MODIFY PrimeMember VARCHAR(3) DEFAULT 'No';
```

Rating Constraint

Ensures only valid review ratings (1–5) can be inserted, improving reliability of review data.

Default Constraint

Standardizes entries, so new customers without explicit membership info are marked as "No" by default.

Validation Result

After applying constraints, existing rows were validated successfully meaning no violations were found in the current dataset.

Clauses and Aggregations

Task Description

Use SQL clauses and aggregations to answer key business questions:

1. Find orders placed after 2024-01-01.
2. List products with average ratings greater than 4.
3. Rank products by total sales (units sold).

Recent Orders

9,000+ orders placed post-Jan 2024, confirming strong ongoing customer activity.

Example: Orders in early 2025 show significant order amounts (₹201–₹900+).

Highly Rated Products

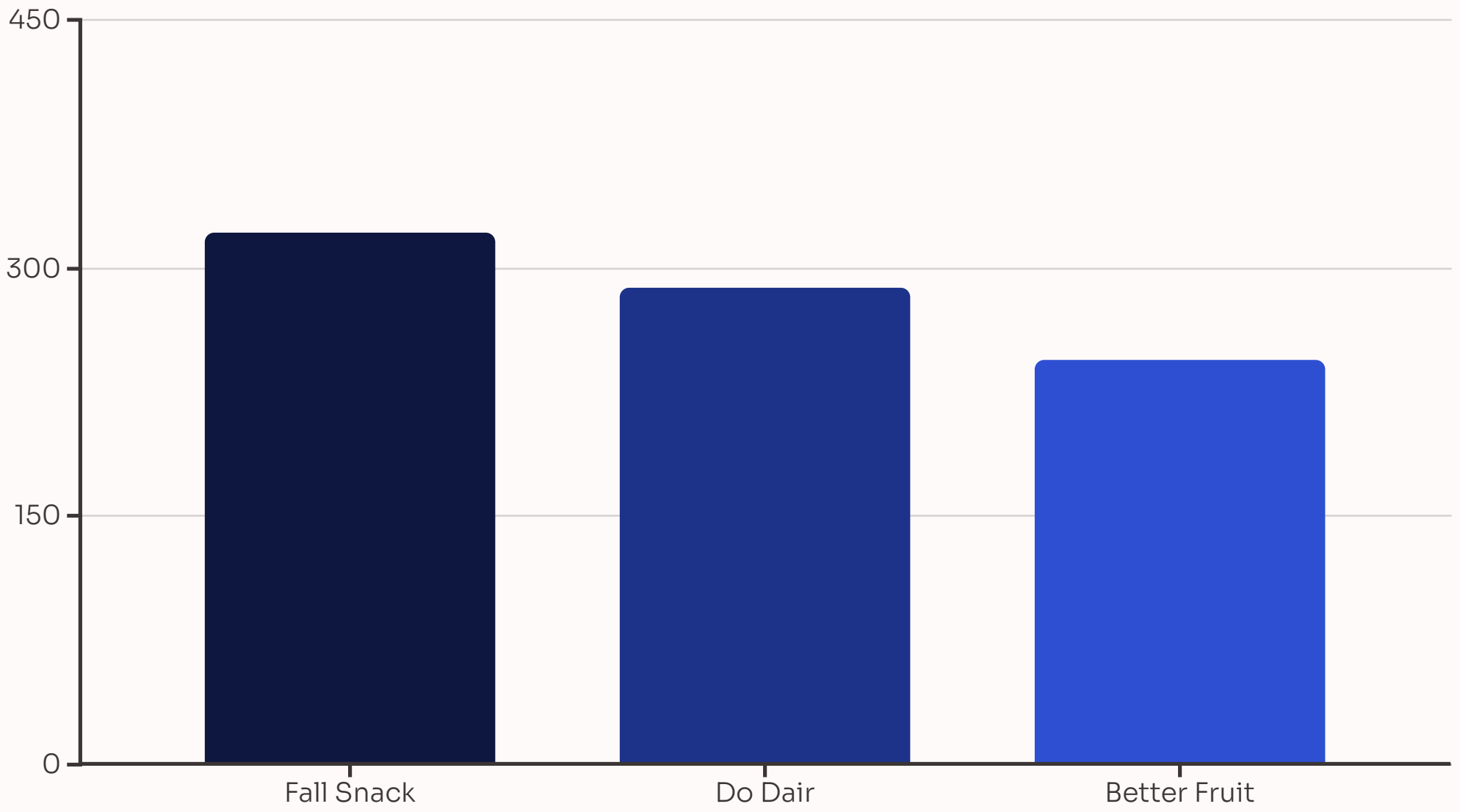
Multiple products (e.g., Four Fruit, Dark Fruit, Society Fruit) achieved perfect 5.0 average ratings.

Indicates strong customer satisfaction in specific product lines.

Top-Selling Products

Fall Snack (321 units), Do Dair (288 units), and Better Fruit (244 units) are leading.

These products dominate sales volume, highlighting them as key revenue drivers.



Identifying High-Value Customers

Task Description:

Amazon Fresh wants to identify top customers based on their total spending.

We calculate each customer's total spending, rank them, and highlight customers who spent more than ₹5,000.

₹83,122

Highest Spender

Michael Garcia across 25 orders

923

High-Value Customers

Customers who spent more than ₹5,000

₹80,000+

Top Tier Spending

Several customers including Kristopher Douglas, Stacey Webb, Sean Allen, and Allison Carson

This insight helps in targeting premium customers for loyalty programs or exclusive offers.

Complex Aggregations and Joins

Total Revenue per Order

Highest-grossing order: ₹36,638.

Multiple orders > ₹25,000, indicating bulk or premium customers.

```
SELECT o.OrderID,
      SUM(od.Quantity *
od.UnitPrice - od.Discount) AS
total_revenue
FROM Orders o
JOIN Order_Details od ON
o.OrderID = od.OrderID
GROUP BY o.OrderID
ORDER BY total_revenue DESC
LIMIT 10;
```

Customers with Most Orders (2025)

Michael Garcia: most active with 15 orders.

Other frequent buyers: Madeline Martin, Yvette Murray, Ralph Evans, Hannah Day (12 orders each).

Strong repeat-customer loyalty observed.

Supplier with Most Products in Stock

Top supplier: Unknown Supplier with 124,843 units.

Acts as a key inventory backbone, critical for supply stability.

Normalization (Products Table to 3NF)

Description

To eliminate redundancy and improve data integrity, the Products table was normalized to Third Normal Form (3NF) by splitting categories and subcategories into separate tables and linking them via foreign keys.

Original Products Table

Contains redundant category and subcategory data

Normalized Structure

Categories Table → SubCategories Table → Products Table

Benefits

Reduced duplication, improved consistency, better scalability

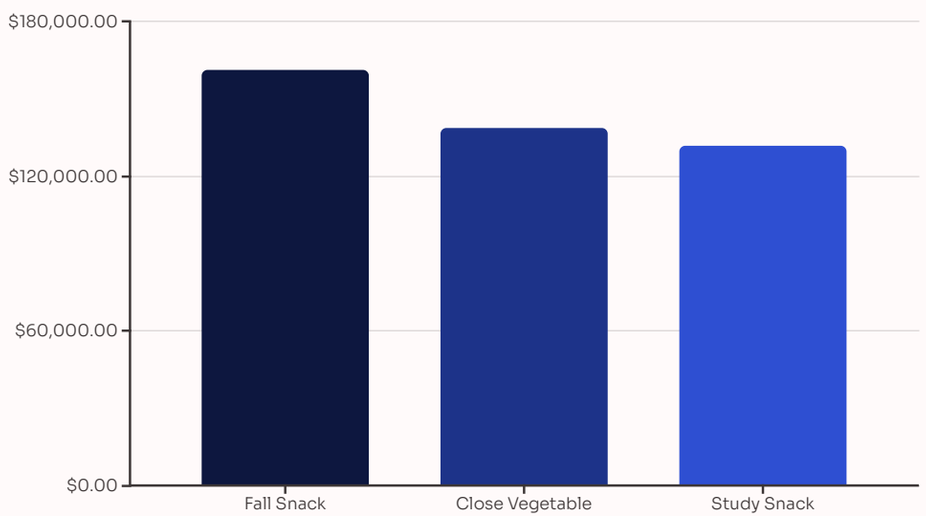
Example: "Organic Apple" → SubCategory: Fresh Fruits → Category: Fruits.

This structure improves scalability, as new categories or subcategories can be added without altering the Products table.

Subqueries and Nested Queries

Top 3 Products Based on Sales Revenue

```
SELECT p.ProductName,
      SUM(od.Quantity * od.UnitPrice - od.Discount) AS
total_revenue
FROM Order_Details od
JOIN Products p ON od.ProductID = p.ProductID
GROUP BY p.ProductName
ORDER BY total_revenue DESC
LIMIT 3;
```



Customers Who Haven't Placed Any Orders

```
SELECT c.CustomerID, c.Name
FROM Customers c
WHERE c.CustomerID NOT IN (
    SELECT DISTINCT o.CustomerID
    FROM Orders o
);
```

Found 11 customers who have never placed an order despite being registered.

Indicates potential customer churn risk or an opportunity for re-engagement campaigns (e.g., discounts, onboarding emails).

Advanced Insights

Description:

This task focuses on deriving deeper business insights through advanced SQL queries, including prime member concentration, most frequently ordered categories, and average order value per customer.

Prime Member Concentration

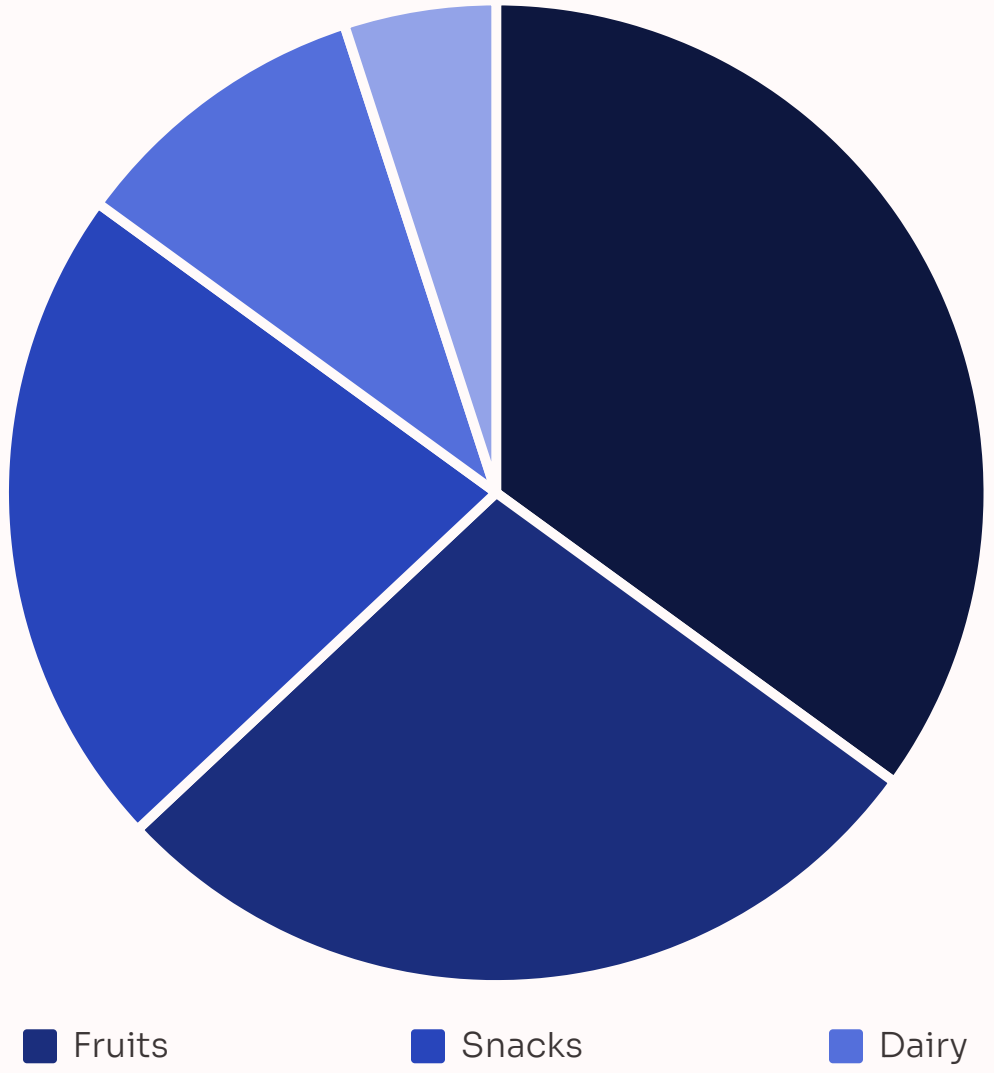
Prime members are concentrated in a few key cities (e.g., New David, Josephport, East Kristine), suggesting strong local loyalty.

Top Categories

The most frequently ordered categories are Meat, Fruits, and Snacks, showing high demand in essentials and daily consumables.

Customer Order Value

Top customers maintain a higher-than-average order value, highlighting opportunities for personalized promotions to sustain loyalty.



These insights can drive targeted marketing campaigns, inventory planning, and customer retention strategies.

Thank You

We extend our sincere gratitude for taking the time to review this comprehensive analysis of Amazon Fresh's operational and customer data. This document has aimed to provide a deep dive into our database structure, query capabilities, and the critical insights derived from leveraging advanced SQL techniques. From understanding customer behavior to optimizing product stock and identifying high-value segments, the presented analytics are designed to inform strategic decisions and enhance overall business performance.

We appreciate your attention and look forward to discussing how these data-driven insights can be put into action.