

Project Title	Amazon Fresh Analytics
Skills take away From This Project	Data Modeling and ER Diagrams SQL Basics: Queries, Clauses, and Aggregations Database Management: DDL, DML, and Constraints Advanced SQL: Joins, Subqueries, and Normalization ACID Transactions and TCL Concepts Business Use Case Analysis
Domain	E-commerce and Retail Analytics

## **Problem Statement:**

Amazon Fresh is an e-commerce platform specializing in groceries and daily essentials. As the platform expands, its management seeks to optimize operations, enhance customer satisfaction, and boost revenue. The challenge lies in effectively managing large volumes of data, including customer profiles, product inventories, supplier information, orders, and customer reviews.

The problem is to design a relational database for Amazon Fresh that organizes this data efficiently, allows for meaningful analysis, and supports critical business decisions. This project aims to solve key business problems, such as identifying top-performing products, analyzing customer purchasing patterns, and ensuring inventory management aligns with demand trends.

## **Business Use Cases:**

**Customer Insights:** Analyze purchasing patterns to identify top customers and cities for targeted marketing.

**Product Management:** Monitor product demand, stock levels, and performance of product categories.

**Supplier Analysis:** Identify key suppliers and optimize supply chain operations.

**Revenue Analysis:** Generate revenue trends and identify high-performing regions or categories.

**Customer Satisfaction:** Use review data to enhance product offerings and improve customer experience.

## **Approach:**

## Data Modeling and Basic Queries

- **Task 1:** Create an **ER diagram** for the Amazon Fresh database to understand the relationships between tables (e.g., Customers, Products, Orders).
- **Task 2:** Identify the **primary keys** and **foreign keys** for each table and describe their relationships.
- **Task 3:** Write a query to:
  - Retrieve all customers from a specific city.
  - Fetch all products under the "Fruits" category.

## Data Definition Language (DDL) and Constraints

- **Task 4:** Write **DDL statements** to recreate the Customers table with the following constraints:
  - CustomerID as the primary key.
  - Ensure Age cannot be null and must be greater than 18.
  - Add a unique constraint for Name.

## Data Manipulation Language (DML)

- **Task 5:** Insert 3 new rows into the Products table using INSERT statements.
- **Task 6:** Update the stock quantity of a product where ProductID matches a specific ID.
- **Task 7:** Delete a supplier from the Suppliers table where their city matches a specific value.

## SQL Constraints and Operators

- **Task 8:** Use SQL constraints to:
  - Add a CHECK constraint to ensure that ratings in the Reviews table are between 1 and 5.
  - Add a DEFAULT constraint for the PrimeMember column in the Customers table (default value: "No").

## Clauses and Aggregations

- **Task 9:** Write queries using:
  - **WHERE** clause to find orders placed after 2024-01-01.
  - **HAVING** clause to list products with average ratings greater than 4.
  - **GROUP BY** and **ORDER BY** clauses to rank products by total sales.

## Task 10: Identifying High-Value Customers

### Scenario:

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

1. Calculate each customer's total spending.
2. Rank customers based on their spending.
3. Identify customers who have spent more than ₹5,000.

## Complex Aggregations and Joins

- **Task 11:** Use SQL to:
  - Join the Orders and OrderDetails tables to calculate total revenue per order.
  - Identify customers who placed the most orders in a specific time period.
  - Find the supplier with the most products in stock.

## Normalization

- **Task 12:** Normalize the Products table to 3NF:
  - Separate product categories and subcategories into a new table.
  - Create foreign keys to maintain relationships.

## Subqueries and Nested Queries

- **Task 13:** Write a subquery to:
  - Identify the top 3 products based on sales revenue.
  - Find customers who haven't placed any orders yet.

## Real-World Analysis

- **Task 14:** Provide actionable insights:
  - Which cities have the highest concentration of Prime members?
  - What are the top 3 most frequently ordered categories?

## Results:

- Design an efficient relational database.
- Generate SQL queries to answer key business questions.
- Create actionable reports for stakeholders.
- Apply ACID principles to ensure data integrity.
- Normalize datasets for optimized database structure.

## Project Evaluation metrics:

- Accuracy and efficiency of SQL queries.
- Quality of the ER diagram and normalization.
- Completeness of business questions answered.
- Effective implementation of ACID transactions.
- Submission of well-structured and documented code.

## Technical Tags:

SQL, DDL, DML, TCL, ACID Transactions, Database Normalization, E-commerce Analytics, Business Insights

## Data Set:

**AMAZON-FRESH-ANALYSIS-DATASET**

The dataset consists of the following CSV files:

1. **Customers.csv:** Customer details including demographics and Prime membership.
2. **Products.csv:** Product information such as categories and stock levels.
3. **Suppliers.csv:** Supplier details and locations.
4. **Orders.csv:** Order details including customer and date.
5. **Order\_Details.csv:** Product-specific order information.
6. **Reviews.csv:** Customer ratings and reviews for products.

## Data Set Explanation:

### 1. Customers Dataset

- Details: Information about customers, like their names, cities, and Prime membership status.
- Data Types:
  - Unique customer ID: Number (INT)
  - Name: Text (VARCHAR)
  - City: Text (VARCHAR)
  - Age: Number (INT)
  - Prime membership: Yes/No (BOOLEAN)

### 2. Products Dataset

- Details: Information about products available, including price and stock.
- Data Types:
  - Unique product ID: Number (INT)
  - Product name: Text (VARCHAR)
  - Category: Text (VARCHAR)
  - Price: Decimal number (DECIMAL)
  - Stock quantity: Number (INT)

### 3. Suppliers Dataset

- Details: Information about suppliers, like their names and contact details.
- Data Types:
  - Unique supplier ID: Number (INT)
  - Supplier name: Text (VARCHAR)
  - City: Text (VARCHAR)
  - Contact number: Text (VARCHAR)

### 4. Orders Dataset

- Details: High-level details of orders placed by customers.
- Data Types:
  - Unique order ID: Number (INT)
  - Customer ID: Number (INT) (links to Customers)
  - Order date: Date (DATE)
  - Total order amount: Decimal number (DECIMAL)

## 5. Order Details Dataset

- Details: Specific products within each order.
- Data Types:
  - Unique order detail ID: Number (INT)
  - Order ID: Number (INT) (links to Orders)
  - Product ID: Number (INT) (links to Products)
  - Quantity ordered: Number (INT)
  - Subtotal: Decimal number (DECIMAL)

## 6. Reviews Dataset

- Details: Feedback from customers about products.
- Data Types:
  - Unique review ID: Number (INT)
  - Product ID: Number (INT) (links to Products)
  - Customer ID: Number (INT) (links to Customers)
  - Rating: Number (TINYINT) (1 to 5 stars)
  - Review text: Long text (TEXT)

## Project Deliverables:

1. **SQL Scripts:** Queries and commands used in the project.
2. **ER Diagram:** Visual representation of the database structure.
3. **Reports:** Output for business questions.
4. **Documentation:** Explanation of approach, assumptions, and findings.

Created By:	Verified By:	Approved By:
Gaurav Kumar	Pooranam Annamalai	Pooranam Annamalai