# Schema Design Documentation

# 1. Overview

This document describes the database schema design for the Retail Data Analysis. The schema is designed to support the project, ensuring scalability, normalization, and meeting all acceptance criteria.

# 2. Acceptance Criteria

* Identify key business entities from the dataset (e.g., Customer, Product, Transactions, etc.).
* Design schema that captures relationships between entities.
* ER diagram prepared with keys/relationships.
* Schema documented with table names, columns, data types, and purpose.

# 3. Entities and Relationships

## Key Entities

* **customer\_dim** – stores the customer details like name, email, phone etc.
* **product\_dim** – stores product details like name, brand, category etc and tell product sold.
* **date\_dim** – stores the date like day, month, weekdays etc.
* **payment\_dim** – stores the details regarding the payment like mode, status and id.
* **shipping\_dim** – stores shipping id and method of the customer product.
* **Transaction\_fact** – fact or central table for star schema and stores all the information and FK to other tables.

## Relationships

* Many Transaction → One Customer
* Many Transaction → One Product
* Many Transaction → One Payment
* Many Transaction→ One Shipping
* Many Transaction → One Date

# 4. Schema Design (Tables)

#### A. customer\_dim

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| customer\_id | INT (PK) | Unique ID for the customer |
| name | VARCHAR(100) | Full name |
| email | VARCHAR(100) | Email address |
| address | VARCHAR(250) | Full mailing address |
| phone | VARCHAR(20) | Phone number |
| city | VARCHAR(50) | City |
| state | VARCHAR(50) | State |
| zipcode | VARCHAR(20) | Zip/postal code |
| country | VARCHAR(50) | Country |
| age | INT | Age of the customer |
| gender | VARCHAR(20) | Gender (e.g., Male, Female, Other) |
| income | DECIMAL(12,2) | Annual income |
| customer\_segment | VARCHAR(50) | Segment type (e.g., Premium, Basic) |

#### 

#### B. product\_dim

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| product\_id | INT (PK) | Unique product ID |
| product\_name | VARCHAR(100) | Name of the product |
| product\_category | VARCHAR(100) | Category (e.g., Electronics) |
| product\_brand | VARCHAR(100) | Brand name |
| product\_type | VARCHAR(100) | Type (e.g., Smartphone, Laptop) |

#### C. date\_dim

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| date\_id | INT (PK) | Id for each date |
| year | INT | Year (e.g., 2025) |
| month | INT | Month (1–12) |
| day | INT | Day of the month (1–31) |
| weekday | VARCHAR(20) | Day name (e.g., Monday) |

#### D. payment\_dim

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| payment\_id | INT (PK) | Unique ID for the payment method |
| payment\_mode | VARCHAR(50) | e.g., Credit Card, PayPal, COD |
| payment\_status | VARCHAR(50) | e.g., Paid, Pending, Refunded |

#### E. shipping\_dim

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| shipping\_id | INT (PK) | Unique ID for the shipping method |
| shipping\_method | VARCHAR(50) | e.g., Standard, Same-Day |

#### F. transaction\_fact

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| transaction\_id | INT (PK) | Unique ID for the transaction |
| transaction\_date | TIMESTAMP | Timestamp of the transaction |
| customer\_id | INT (FK) | Link to customer\_dim |
| product\_id | INT (FK) | Link to product\_dim |
| date\_id | INT (FK) | Link to date\_dim |
| payment\_id | INT (FK) | Link to payment\_dim |
| shipping\_id | INT (FK) | Link to shipping\_dim |
| quantity | INT | Quantity of products bought |
| amount | DECIMAL(12,3) | Total amount paid |
| order\_status | VARCHAR(50) | e.g., Completed, Cancelled, Returned |
| ratings | INT | Rating given by the customer (e.g., 1–5) |
| feedback | TEXT | Optional textual feedback from the customer |

### 4. Relationships / Foreign Keys

| **Fact Table Column** | **References Table** | **Cardinality** |
| --- | --- | --- |
| customer\_id | customer\_dim | Many-to-One |
| product\_id | product\_dim | Many-to-One |
| date\_id | date\_dim | Many-to-One |
| payment\_id | payment\_dim | Many-to-One |
| shipping\_id | shipping\_dim | Many-to-One |

# 5. ER Diagram

