Experiment 6 Output

Raw data was **cleaned and transformed** before being loaded into the data warehouse, ensuring consistency, accuracy, and completeness.

Creating Database and Using in MySQL

```
mysql> CREATE DATABASE RetailDataWarehouse;
Query OK, 1 row affected (0.03 sec)

mysql> show databases;
+--------------------+
| Database           |
+--------------------+
| classdb            |
| company            |
| employee           |
| hospital           |
| information_schema |
| movie              |
| mydatabase         |
| mysql              |
| performance_schema |
| retaildatawarehouse|
| root               |
| shopping           |
| student            |
| sys                |
| utd                |
+--------------------+
15 rows in set (0.00 sec)
```

Fig 1: Database in MySQL

Creating Tables

```
mysql> CREATE TABLE Raw_Sales (
    ->    ID INT,
    ->    Product VARCHAR(50),
    ->    Region VARCHAR(50),
    ->    Sales_Amount VARCHAR(20),   -- Stored as text, needs conversion
    ->    Date_Sold VARCHAR(20)       -- Inconsistent date format
    -> );
Query OK, 0 rows affected (0.02 sec)

mysql>
mysql> INSERT INTO Raw_Sales VALUES
    -> (1, 'Laptop', 'North', '50000', '2023-01-10'),
    -> (2, 'Phone', 'South', '32000.00', '10/02/2023'),
    -> (3, 'Laptop', 'East', NULL, '2023-03-15'), -- Missing Sales_Amount
    -> (4, 'Tablet', NULL, '25000', '2023-04-05'), -- Missing Region
    -> (5, 'Phone', 'West', 'invalid', '2023-05-20'); -- Invalid Sales_Amount
Query OK, 5 rows affected (0.00 sec)
Records: 5  Duplicates: 0  Warnings: 0
```

Fig 2: Creating Tables and Inserting Data

Extracting and Transforming Data

```
mysql> ALTER TABLE Raw_Sales ADD COLUMN Year INT;
Query OK, 0 rows affected (0.04 sec)
Records: 0  Duplicates: 0  Warnings: 0

mysql> UPDATE Raw_Sales SET Year = EXTRACT(YEAR FRO
Query OK, 5 rows affected (0.00 sec)
Rows matched: 5  Changed: 5  Warnings: 0

mysql> ALTER TABLE Raw_Sales ADD COLUMN Category VA
Query OK, 0 rows affected (0.02 sec)
Records: 0  Duplicates: 0  Warnings: 0

mysql> UPDATE Raw_Sales
    -> SET Category =
    ->     CASE
    ->         WHEN Product IN ('Laptop', 'Tablet')
    ->         WHEN Product = 'Phone' THEN 'Mobile'
    ->         ELSE 'Other'
    ->     END;
Query OK, 5 rows affected (0.00 sec)
Rows matched: 5  Changed: 5  Warnings: 0
```

Fig 3: Extraction and Transformation of Data

Loading data after cleaning

```
mysql> SELECT * FROM Raw_Sales;
+------+---------+---------+--------------+--
| ID   | Product | Region  | Sales_Amount | D
+------+---------+---------+--------------+--
|    1 | Laptop  | North   | 50000        | 2
|    2 | Phone   | South   | 32000.00     | 2
|    3 | Laptop  | East    | 0            | 2
|    4 | Tablet  | Unknown | 25000        | 2
|    5 | Phone   | West    | 0            | 2
+------+---------+---------+--------------+--
5 rows in set (0.00 sec)
```
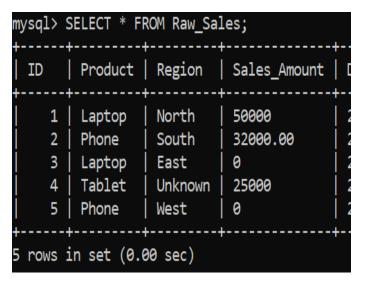
Fig 4: Cleaned data for warehouse

Madhurima Rawat DS 42