**SQL CASE STUDY**

**Data Availability:**

The data set we will be using for our session comprises of 3 tables : Customer, Transactions & Product category information.



mysql> create database alabs;

Query OK, 1 row affected (0.00 sec)

mysql> use alabs;

Database changed

mysql> CREATE TABLE IF NOT EXISTS Customer(

-> customer\_Id bigint(10) PRIMARY KEY,

-> DOB varchar(10),

-> Gender varchar(5),

-> city\_code varchar(10)

-> );

Query OK, 0 rows affected (0.07 sec)

mysql> CREATE TABLE IF NOT EXISTS Transactions(

-> transaction\_Id bigint(20) NOT NULL,

-> cust\_id bigint(10),

-> tran\_date varchar(10),

-> prod\_subcat\_code varchar(10),

-> prod\_cat\_code varchar(10),

-> qty int,

-> rate double,

-> tax double,

-> total\_amt double,

-> store\_type varchar(10)

-> );

Query OK, 0 rows affected (0.03 sec)

mysql> CREATE TABLE IF NOT EXISTS prod\_cat\_info(

-> prod\_cat\_code varchar(10),

-> prod\_cat varchar(30),

-> prod\_subcat\_cod varchar(10),

-> prod\_subcat varchar(30)

-> );

Query OK, 0 rows affected (0.01 sec)

mysql> show tables;

+-----------------+

| Tables\_in\_alabs |

+-----------------+

| Customer |

| Transactions |

| prod\_cat\_info |

+-----------------+

3 rows in set (0.00 sec)

mysql> describe Customer;

+-------------+-------------+------+-----+---------+-------+

| Field | Type | Null | Key | Default | Extra |

+-------------+-------------+------+-----+---------+-------+

| customer\_Id | bigint(10) | NO | PRI | NULL | |

| DOB | varchar(10) | YES | | NULL | |

| Gender | varchar(5) | YES | | NULL | |

| city\_code | varchar(10) | YES | | NULL | |

+-------------+-------------+------+-----+---------+-------+

4 rows in set (0.00 sec)

mysql> describe Transactions;

+------------------+-------------+------+-----+---------+-------+

| Field | Type | Null | Key | Default | Extra |

+------------------+-------------+------+-----+---------+-------+

| transaction\_Id | bigint(20) | NO | | NULL | |

| cust\_id | bigint(10) | YES | | NULL | |

| tran\_date | varchar(10) | YES | | NULL | |

| prod\_subcat\_code | varchar(10) | YES | | NULL | |

| prod\_cat\_code | varchar(10) | YES | | NULL | |

| qty | int(11) | YES | | NULL | |

| rate | double | YES | | NULL | |

| tax | double | YES | | NULL | |

| total\_amt | double | YES | | NULL | |

| store\_type | varchar(10) | YES | | NULL | |

+------------------+-------------+------+-----+---------+-------+

10 rows in set (0.00 sec)

mysql> describe prod\_cat\_info;

+-----------------+-------------+------+-----+---------+-------+

| Field | Type | Null | Key | Default | Extra |

+-----------------+-------------+------+-----+---------+-------+

| prod\_cat\_code | varchar(10) | YES | | NULL | |

| prod\_cat | varchar(30) | YES | | NULL | |

| prod\_subcat\_cod | varchar(10) | YES | | NULL | |

| prod\_subcat | varchar(30) | YES | | NULL | |

+-----------------+-------------+------+-----+---------+-------+

4 rows in set (0.00 sec)

mysql> LOAD DATA INFILE '/home/cloudera/localdata/Customer.csv'INTO TABLE Customer FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;

Query OK, 5647 rows affected (0.09 sec)

Records: 5647 Deleted: 0 Skipped: 0 Warnings: 0

mysql> select \* from Customer limit 10;

+-------------+------------+--------+-----------+

| customer\_Id | DOB | Gender | city\_code |

+-------------+------------+--------+-----------+

|6783 | 01-05-1974 | M | 4

|66784 | 13-12-1991 | F | 10

|6785 | 29-06-1985 | F | 3

|6788 | 20-03-1972 | F | 2

|6794 | 28-02-1971 | F | 9

|6799 | 16-03-1970 | F | 5

|6803 | 23-06-1988 | F | 1

|66804 | 14-07-1979 | M | 10

|6805 | 07-01-1976 | F | 8

|66806 | 08-12-1991 | F | 10

+-------------+------------+--------+-----------+

10 rows in set (0.00 sec)

mysql> LOAD DATA INFILE '/home/cloudera/localdata/Transactions.csv'INTO TABLE Transactions FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;

Query OK, 23057 rows affected, 4579 warnings (0.26 sec)

Records: 23057 Deleted: 0 Skipped: 0 Warnings: 4579

mysql> select \* from Transactions limit 10;

+----------------+---------+------------+------------------+---------------+------+-------+---------+-----------+------------+

| transaction\_Id | cust\_id | tran\_date | prod\_subcat\_code | prod\_cat\_code | qty | rate | tax | total\_amt | store\_type |

+----------------+---------+------------+------------------+---------------+------+-------+---------+-----------+------------+

|80712190438 | 270351 | 28-02-2014 | 1 | 1 | -5 | -772 | 405.3 | -4265.3 | e-shop

|29258453508 | 270384 | 27-02-2014 | 5 | 3 | -5 | -1497 | 785.925 | -8270.925 | e-shop

| 51750724947 | 273420 | 24-02-2014 | 6 | 5 | -2 | -791 | 166.11 | -1748.11 | TeleShop

|93274880719 | 271509 | 24-02-2014 | 11 | 6 | -3 | -1363 | 429.345 | -4518.345 | e-Shop

| 51750724947 | 273420 | 23-02-2014 | 6 | 5 | -2 | -791 | 166.11 | -1748.11 | TeleShop

| 97439039119 | 272357 | 23-02-2014 | 8 | 3 | -2 | -824 | 173.04 | -1821.04 | TeleShop

|45649838090 | 273667 | 22-02-2014 | 11 | 6 | -1 | -1450 | 152.25 | -1602.25 | e-shop

| 22643667930 | 271489 | 22-02-2014 | 12 | 6 | -1 | -1225 | 128.625 | -1353.625 | TeleShop

|92372943 | 275108 | 22-02-2014 | 3 | 1 | -3 | -908 | 286.02 | -3010.02 | MBR

|50076728598 | 269014 | 21-02-2014 | 8 | 3 | -4 | -581 | 244.02 | -2568.02 | e-shop

+----------------+---------+------------+------------------+---------------+------+-------+---------+-----------+------------+

10 rows in set (0.00 sec)

mysql> LOAD DATA INFILE '/home/cloudera/localdata/prod\_cat\_info.csv'INTO TABLE prod\_cat\_info FIELDS TERMINATED BY ',' LINES TERMINATED BY '\n' IGNORE 1 LINES;

Query OK, 23 rows affected (0.04 sec)

Records: 23 Deleted: 0 Skipped: 0 Warnings: 0

mysql> select \* from prod\_cat\_info limit 10;

+---------------+-------------+-----------------+----------------------+

| prod\_cat\_code | prod\_cat | prod\_subcat\_cod | prod\_subcat |

+---------------+-------------+-----------------+----------------------+

| Clothing | 4 | Mens

|| Clothing | 1 | Women

| Clothing | 3 | Kids

| Footwear | 1 | Mens

|| Footwear | 5 | Women

| Footwear | 10 | Kids

| | Electronics | 4 | Mobiles

| | Electronics | 5 | Computers

|3 | Electronics | 8 | Personal Appliances

| | Electronics | 9 | Cameras

+---------------+-------------+-----------------+----------------------+

10 rows in set (0.00 sec)

**QUESTIONS:**

1.What is the total number of rows in all 3 tables in the database?

mysql> select ((select count(\*) from Transactions ) + (select count(\*) from Customer )+ (select count(\*) from prod\_cat\_info)) as Total\_Rows;

+------------+

| Total\_Rows |

+------------+

| 28727 |

+------------+

1 row in set (0.02 sec)

2.What is the time period of the transaction data we have available?

mysql> SELECT DATEDIFF(STR\_TO\_DATE(max(tran\_date), '%d-%m-%Y'),STR\_TO\_DATE(min(tran\_date), '%d-%m-%Y')) AS TIME\_PERIOD\_IN\_DAYS from Transactions;

+---------------------+

| TIME\_PERIOD\_IN\_DAYS |

+---------------------+

| 730 |

+---------------------+

1 row in set (0.02 sec)

3.What is the total number of transactions that have a return?

mysql> select count(case when total\_amt<0 or Qty<0 then 1 else 0 end) as total\_return\_transactions FROM Transactions;

+---------------------------+

| total\_return\_transactions |

+---------------------------+

| 23057 |

+---------------------------+

1 row in set (0.02 sec)

4.What is the count of Male & Female customers in the database?

mysql> select count(customer\_Id), Gender from Customer group by Gender;

+--------------------+--------+

| count(customer\_Id) | Gender |

+--------------------+--------+

| 2754 | F |

| 2893 | M |

+--------------------+--------+

2 rows in set (0.01 sec)

5.Which city do we have the maximum number of customers from and how many?

mysql> select count(customer\_Id) as total\_customers, city\_code from Customer group by city\_code order by total\_customers desc limit 1;

+-----------------+-----------+

| total\_customers | city\_code |

+-----------------+-----------+

| 595 | 3

+-----------------+-----------+

1 row in set (0.01 sec)

6.Which product category does the sub-category “DIY” belong to?

mysql> select prod\_cat from prod\_cat\_info where prod\_subcat like '%DIY%';

+----------+

| prod\_cat |

+----------+

| Books |

+----------+

1 row in set (0.00 sec)

7.How many sub-categories are there under the Books category?

mysql> select count(distinct prod\_subcat) from prod\_cat\_info where prod\_cat='Books';

+-----------------------------+

| count(distinct prod\_subcat) |

+-----------------------------+

| 6 |

+-----------------------------+

1 row in set (0.00 sec)

8.What is the maximum quantity of products ever ordered?

mysql> select max(qty) from Transactions;

+----------+

| max(qty) |

+----------+

| 5 |

+----------+

1 row in set (0.21 sec)

9.What is the maximum quantity of products ever returned?

mysql> select min(qty) from Transactions;

+----------+

| min(qty) |

+----------+

| -5 |

+----------+

1 row in set (0.02 sec)

10.Which store-type sells the maximum products, by value and by quantity?

By Quantity:

mysql> select sum(qty) as Quantity, store\_type as Top\_Store from Transactions group by store\_type order by sum(qty) desc limit 1;

+----------+-----------+

| Quantity | Top\_Store |

+----------+-----------+

| 22765 | e-shop

+----------+-----------+

1 row in set (0.05 sec)

By Value:

mysql> select sum(total\_amt) as Quantity, store\_type as Top\_Store from Transactions group by store\_type order by sum(total\_amt) desc limit 1;

+-------------+-----------+

| Quantity | Top\_Store |

+-------------+-----------+

|825151.97 | e-shop

+-------------+-----------+

1 row in set (0.05 sec)

11.What is the highest value of return recorded in the transaction data?

mysql> select min(total\_amt) from Transactions where total\_amt<0;

+----------------+

| min(total\_amt) |

+----------------+

| -8270.925 |

+----------------+

1 row in set (0.01 sec)

12.How many customers have >10 transactions with us, excluding returns?

mysql> select count(a.cust\_id) as high\_value\_customers from (select count(transaction\_id), cust\_id from Transactions where qty>0 group by cust\_id having count(transaction\_id)>10) as a;

+----------------------+

| high\_value\_customers |

+----------------------+

| 6 |

+----------------------+

1 row in set (0.02 sec)

13.What is the combined total revenue earned from the “Electronics” & “Clothing” categories, from “Flagship stores”?

mysql> select sum(total\_amt) as revenue from Transactions where store\_type like '%Flagship%' and prod\_cat\_code in (select distinct prod\_cat\_code from prod\_cat\_info where prod\_cat in ('Electronics','Clothing'));

+------------------+

| revenue |

+------------------+

| 3409559.27000001 |

+------------------+

1 row in set (0.08 sec)

14.What is the total revenue generated from “Male” customers under the “Electronics” category, by prod sub-cat?

mysql> select sum(total\_amt) as revenue, prod\_subcat\_code from Transactions where cust\_id in (select customer\_Id from Customer where Gender='M') and prod\_cat\_code in (select distinct prod\_cat\_code from prod\_cat\_info where prod\_cat='Electronics') group by prod\_subcat\_code order by revenue desc;

+-------------+------------------+

| revenue | prod\_subcat\_code |

+-------------+------------------+

| 1192413.235 | 4 |

| 1172702.245 | 9 |

| 1138983.17 | 10 |

| 1107593.435 | 8 |

| 1091417.34 | 5 |

+-------------+------------------+

5 rows in set (0.22 sec)

15. What is percentage of sales and returns by product sub category and find the top 5 sub categories in terms of sales.

Sales % by product sub category:

mysql> select @total\_sales:=sum(total\_amt) from Transactions where total\_amt>0;

+------------------------------+

| @total\_sales:=sum(total\_amt) |

+------------------------------+

| 54458559.22 |

+------------------------------+

1 row in set (0.02 sec)

mysql> select (sum(total\_amt)/@total\_sales)\*100 as Sales\_percent, prod\_subcat\_code from Transactions where total\_amt>0 group by prod\_subcat\_code order by Sales\_percent desc;

+------------------+------------------+

| Sales\_percent | prod\_subcat\_code |

+------------------+------------------+

| 17.5606372110702 | 4 |

| 13.0413517116915 | 3 |

| 12.9760239092128 | 10 |

| 12.7972023403083 | 1 |

| 8.88567683263803 | 11 |

| 8.66324658157198 | 12 |

| 4.57761099578352 | 7 |

| 4.36242045148987 | 2 |

| 4.31430724325357 | 9 |

| 4.29402875965399 | 8 |

| 4.27668024706879 | 5 |

| 4.25081371625755 | 6 |

+------------------+------------------+

12 rows in set (0.03 sec)

Returns % by product sub category:

mysql> select @total\_returns:=sum(total\_amt) from Transactions where total\_amt<0;

+--------------------------------+

| @total\_returns:=sum(total\_amt) |

+--------------------------------+

| -5874586.64 |

+--------------------------------+

1 row in set (0.01 sec)

mysql> select (sum(total\_amt)/@total\_returns)\*100 as Returns\_percent, prod\_subcat\_code from Transactions where total\_amt<0 group by prod\_subcat\_code order by Returns\_percent desc;

+------------------+------------------+

| Returns\_percent | prod\_subcat\_code |

+------------------+------------------+

| 18.9197775624261 | 4 |

| 13.7071963415625 | 3 |

| 13.1533219671776 | 1 |

| 12.010850264692 | 10 |

| 8.85738158080856 | 11 |

| 8.6837668122297 | 12 |

| 4.71148724091335 | 2 |

| 4.43692009281525 | 7 |

| 4.06670493840908 | 5 |

| 3.91899131136144 | 6 |

| 3.85471810830251 | 8 |

| 3.67888377930196 | 9 |

+------------------+------------------+

12 rows in set (0.01 sec)

Top 5 sub categories in terms of sales:

mysql> select prod\_subcat\_code, (sum(total\_amt)/@total\_sales)\*100 as Sales\_percent from Transactions where total\_amt>0 group by prod\_subcat\_code order by Sales\_percent desc limit 5;

+------------------+------------------+

| prod\_subcat\_code | Sales\_percent |

+------------------+------------------+

| 4 | 17.5606372110702 |

| 3 | 13.0413517116915 |

| 10 | 12.9760239092128 |

| 1 | 12.7972023403083 |

| 11 | 8.88567683263803 |

+------------------+------------------+

5 rows in set (0.04 sec)