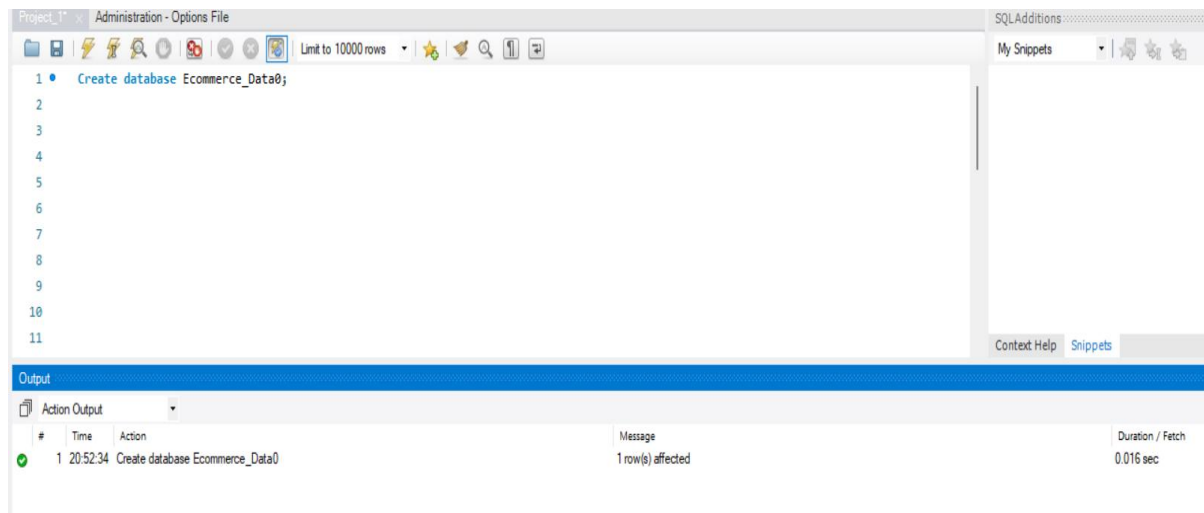# Project 5 E-commerce Return Rate Reduction Analysis
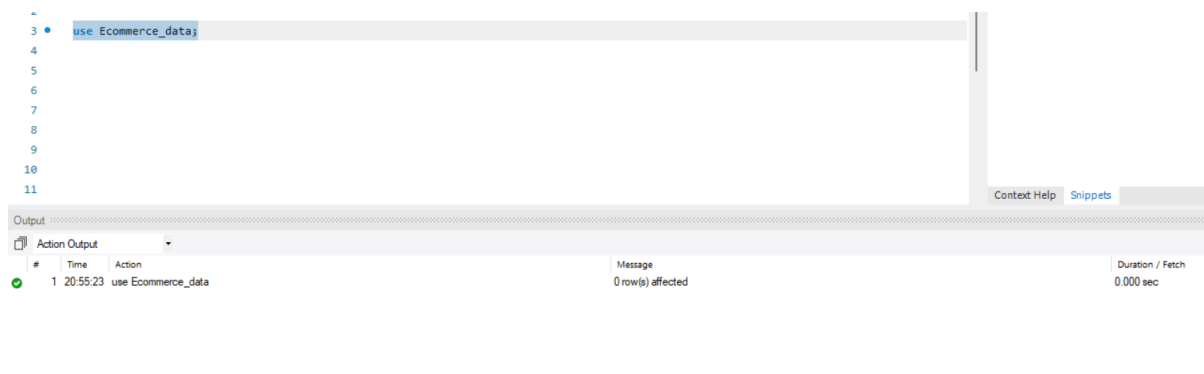
Data Cleaning by Using SQL

Step 1: Firstly, I Create a database by using command

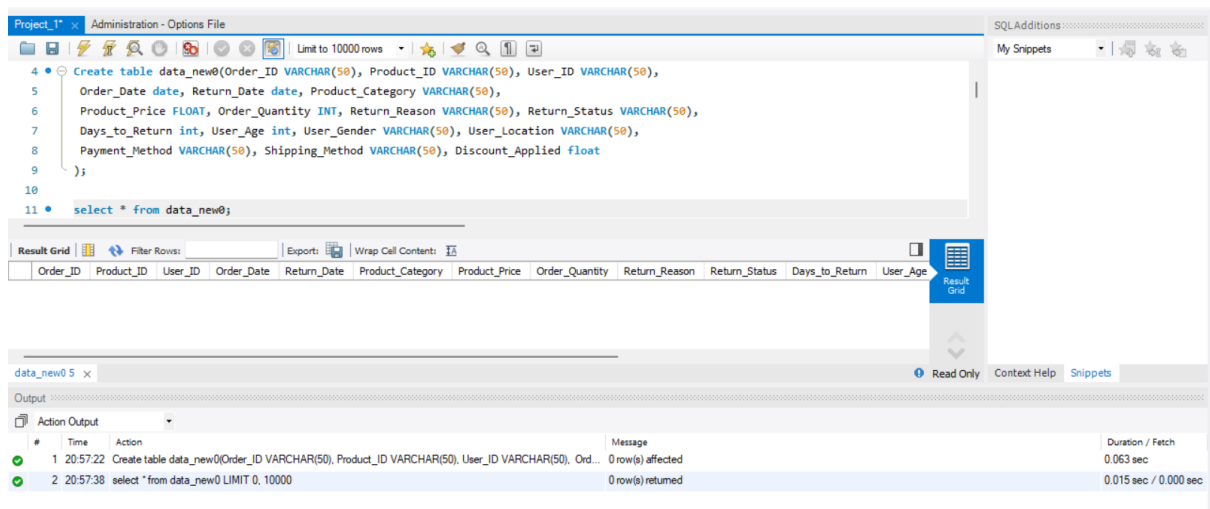Create database Ecommerce_Data;

Step 2: populate database for using purpose by using "use command"
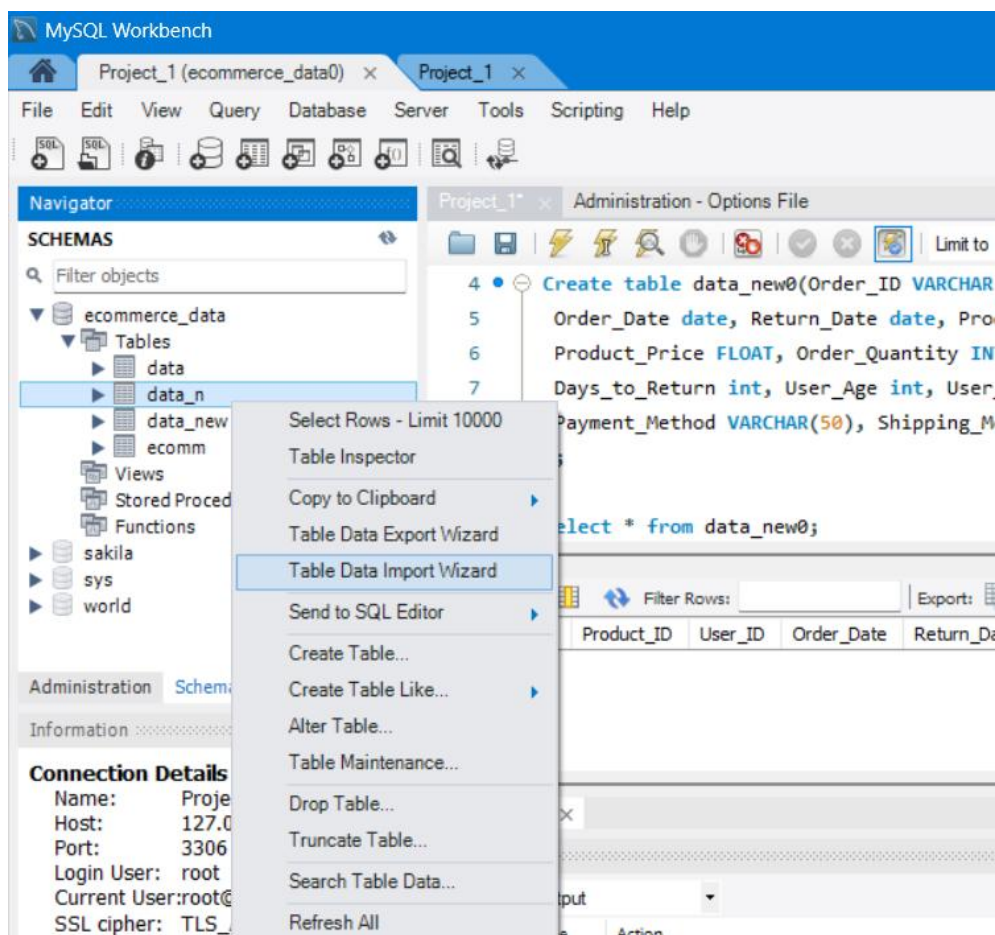
use Ecommerce_data;

step 3: Create table by using "Create command"

Create table data_new(Order_ID VARCHAR(50), Product_ID VARCHAR(50), User_ID VARCHAR(50),

Order_Date date, Return_Date date, Product_Category VARCHAR(50),

Product_Price FLOAT, Order_Quantity INT, Return_Reason VARCHAR(50), Return_Status VARCHAR(50),

Days_to_Return int, User_Age int, User_Gender VARCHAR(50), User_Location VARCHAR(50),

Payment_Method VARCHAR(50), Shipping_Method VARCHAR(50), Discount_Applied float

);

select * from data_new;

Step 4: Import the dataset by using "Imort data" tool in the sql

Step 5: Run the select query to fetch the data

select * from data_new;

Step 6: Then we use multiple select query to count the number of null that are present in the column

-- Data imported By CSV file

SELECT COUNT(*) FROM Data_new WHERE Order_ID IS NULL;

SELECT COUNT(*) FROM Data_new WHERE product_id IS NULL;

SELECT COUNT(*) FROM Data_new WHERE user_id IS NULL;

SELECT COUNT(*) FROM Data_new WHERE order_date IS NULL;

SELECT COUNT(*) FROM Data_new WHERE return_date IS NULL;

SELECT COUNT(*) FROM Data_new WHERE product_category IS NULL;

SELECT COUNT(*) FROM Data_new WHERE product_price IS NULL;

SELECT COUNT(*) FROM Data_new WHERE order_quantity IS NULL;

SELECT COUNT(*) FROM Data_new WHERE return_reason IS NULL;

SELECT COUNT(*) FROM Data_new WHERE return_status IS NULL;

SELECT COUNT(*) FROM Data_new WHERE days_to_return IS NULL;

SELECT COUNT(*) FROM Data_new WHERE user_age IS NULL;

SELECT COUNT(*) FROM Data_new WHERE user_gender IS NULL;

SELECT COUNT(*) FROM Data_new WHERE user_location IS NULL;

SELECT COUNT(*) FROM Data_new WHERE payment_method IS NULL;

SELECT COUNT(*) FROM Data_new WHERE shipping_method IS NULL;

SELECT COUNT(*) FROM Data_new WHERE discount_method IS NULL;

```
13      -- Data imported By CSV file
14
15 •    SELECT COUNT(*) FROM Data_new WHERE Order_ID IS NULL;
16 •    SELECT COUNT(*) FROM Data_new WHERE product_id IS NULL;
17 •    SELECT COUNT(*) FROM Data_new WHERE user_id IS NULL;
18 •    SELECT COUNT(*) FROM Data_new WHERE order_date IS NULL;
19 •    SELECT COUNT(*) FROM Data_new WHERE return_date IS NULL;
20 •    SELECT COUNT(*) FROM Data_new WHERE product_category IS NULL;
21 •    SELECT COUNT(*) FROM Data_new WHERE product_price IS NULL;
22 •    SELECT COUNT(*) FROM Data_new WHERE order_quantity IS NULL;
23 •    SELECT COUNT(*) FROM Data_new WHERE return_reason IS NULL;
24 •    SELECT COUNT(*) FROM Data_new WHERE return_status IS NULL;
25 •    SELECT COUNT(*) FROM Data_new WHERE days_to_return IS NULL;
26 •    SELECT COUNT(*) FROM Data_new WHERE user_age IS NULL;
27 •    SELECT COUNT(*) FROM Data_new WHERE user_gender IS NULL;
28 •    SELECT COUNT(*) FROM Data_new WHERE user_location IS NULL;
29 •    SELECT COUNT(*) FROM Data_new WHERE payment_method IS NULL;
30 •    SELECT COUNT(*) FROM Data_new WHERE shipping_method IS NULL;
```

Context Help   Snippets

Output

Action Output

#	Time	Action	Message	Duration / Fetch
✓	11 21:05:56	SELECT COUNT(*) FROM Data_new WHERE user_age IS NULL LIMIT 0, 10000	1 row(s) returned	0.000 sec / 0.000 sec
✓	12 21:05:59	SELECT COUNT(*) FROM Data_new WHERE user_gender IS NULL LIMIT 0, 10000	1 row(s) returned	0.015 sec / 0.000 sec
✓	13 21:06:01	SELECT COUNT(*) FROM Data_new WHERE user_location IS NULL LIMIT 0, 10000	1 row(s) returned	0.015 sec / 0.000 sec
✓	14 21:06:04	SELECT COUNT(*) FROM Data_new WHERE payment_method IS NULL LIMIT 0, 10000	1 row(s) returned	0.000 sec / 0.000 sec
✓	15 21:06:07	SELECT COUNT(*) FROM Data_new WHERE shipping_method IS NULL LIMIT 0, 10000	1 row(s) returned	0.000 sec / 0.000 sec
✗	16 21:06:10	SELECT COUNT(*) FROM Data_new WHERE discount_method IS NULL LIMIT 0, 10000	Error Code: 1054. Unknown column 'discount_method' in 'where clause'	0.016 sec

Step 8: Delete null values if it present
-- Delete null cell row

Delete FROM Data_new WHERE Order_ID IS NULL;

Delete FROM Data_new WHERE product_id IS NULL;

Delete FROM Data_new WHERE user_id IS NULL;

Delete FROM Data_new WHERE order_date IS NULL;

Delete FROM Data_new WHERE return_date IS NULL;

Delete FROM Data_new WHERE product_category IS NULL;

Delete FROM Data_new WHERE product_price IS NULL;

Delete FROM Data_new WHERE order_quantity IS NULL;

Delete FROM Data_new WHERE return_reason IS NULL;

Delete FROM Data_new WHERE return_status IS NULL;

Delete FROM Data_new WHERE days_to_return IS NULL;

Delete FROM Data_new WHERE user_age IS NULL;

Delete FROM Data_new WHERE user_gender IS NULL;

Delete FROM Data_new WHERE user_location IS NULL;

Delete FROM Data_new WHERE payment_method IS NULL;

Delete FROM Data_new WHERE shipping_method IS NULL;

Delete FROM Data_new WHERE discount_method IS NULL;

```
36        -- Delete null cell row
37
38  •    Delete FROM Data_new WHERE Order_ID IS NULL;
39  •    Delete FROM Data_new WHERE product_id IS NULL;
40  •    Delete FROM Data_new WHERE user_id IS NULL;
41  •    Delete FROM Data_new WHERE order_date IS NULL;
42  •    Delete FROM Data_new WHERE return_date IS NULL;
43  •    Delete FROM Data_new WHERE product_category IS NULL;
44  •    Delete FROM Data_new WHERE product_price IS NULL;
45  •    Delete FROM Data_new WHERE order_quantity IS NULL;
46  •    Delete FROM Data_new WHERE return_reason IS NULL;
47  •    Delete FROM Data_new WHERE return_status IS NULL;
48  •    Delete FROM Data_new WHERE days_to_return IS NULL;
49  •    Delete FROM Data_new WHERE user_age IS NULL;
50  •    Delete FROM Data_new WHERE user_gender IS NULL;
51  •    Delete FROM Data_new WHERE user_location IS NULL;
52  •    Delete FROM Data_new WHERE payment_method IS NULL;
53  •    Delete FROM Data_new WHERE shipping_method IS NULL;
54  •    Delete FROM Data_new WHERE discount_method IS NULL;
```

Output — Action Output

#	Time	Action	Message	Duration / Fetch
3	21:08:30	Delete FROM Data_new WHERE user_id IS NULL	0 row(s) affected	0.000 sec
4	21:08:33	Delete FROM Data_new WHERE order_date IS NULL	0 row(s) affected	0.000 sec
5	21:08:41	Delete FROM Data_new WHERE return_date IS NULL	0 row(s) affected	0.000 sec
6	21:08:43	Delete FROM Data_new WHERE product_category IS NULL	0 row(s) affected	0.000 sec
7	21:08:47	Delete FROM Data_new WHERE product_price IS NULL	0 row(s) affected	0.000 sec
8	21:08:51	Delete FROM Data_new WHERE order_quantity IS NULL	0 row(s) affected	0.000 sec

Step 9: -- Check for duplicates based on order_id

SELECT order_id, COUNT(*)

FROM data_new

GROUP BY order_id

HAVING COUNT(*) > 1;

```
56        -- Check for duplicates based on order_id
57
58  •    SELECT order_id, COUNT(*)
59        FROM data_new
60        GROUP BY order_id
61        HAVING COUNT(*) > 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: IA

order_id	COUNT(*)

Result 36 ×   Read Only   Context Help   Snippets

Output — Action Output

#	Time	Action	Message	Duration / Fetch
1	21:09:55	SELECT order_id, COUNT(*) FROM data_new GROUP BY order_id HAVING COUNT(*) > 1 LIMIT 0, 10000	0 row(s) returned	0.000 sec / 0.000 sec

Step 10 : -- Delete duplicates based on order_id

DELETE FROM data_new

WHERE id NOT IN (

 SELECT MIN(id)

 FROM data_new

 GROUP BY order_id);

```
63      -- Delete duplicates based on order_id
64
65 •    DELETE FROM data_new
66 ⊖   WHERE id NOT IN (
67         SELECT MIN(id)
68         FROM data_new
69         GROUP BY order_id
70      );
71
72      -- Negative price check
73
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

order_id	COUNT(*)

Result 36 ×

Output

Action Output

#	Time	Action	Message
✓ 1	21:09:55	SELECT order_id, COUNT(*) FROM data_new GROUP BY order_id HAVING COUNT(*) > 1 LIMIT 0, 10000	0 row(s) returned

Step 11: -- Negative price check

SELECT * FROM data_new WHERE product_price < 0;

```
72      -- Negative price check
73
74 •    SELECT * FROM data_new WHERE product_price < 0;
75
76
77
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Order_ID	Product_ID	User_ID	Order_Date	Return_Date	Product_Category	Product_Price	Order_Quantity	Return_Reason	Return_Status	Days_to_Return	User_Age

data_new 38 ×

Output

Action Output

#	Time	Action	Message
✓ 1	21:11:30	SELECT * FROM data_new WHERE product_price < 0 LIMIT 0, 10000	0 row(s) returned

Step 12: delete data for < 0

```
76 •    DELETE FROM data_new WHERE product_price < 0 OR order_quantity <= 0;
77
78
79
```

Output

Action Output

#	Time	Action	Message
✓ 1	21:12:15	DELETE FROM data_new WHERE product_price < 0 OR order_quantity <= 0	0 row(s) affected