Sales Analysis

SQL PROJECT ON SUPERSTORE SALES ANALYSIS

Introduction

My Self Roni Sarkar, an enthusiastic data analyst with a passion for uncovering insights through data. Today I successfully completed a first project on SQL, analyzing superstore sales data. This achievement showcases a strong foundation in data analysis and a dedication to continuous learning.
.

BRIEF SUMMARY

With Michel Nguegang's expert guidance, I successfully completed an SQL project on Superstore sales analysis. His support in understanding the dataset and crafting precise SQL queries was invaluable. The project involved analyzing sales trends, identifying key performance metrics, and generating insights on product performance and customer behavior. Michel's mentorship ensured a thorough and accurate analysis, enhancing my SQL skills and deepening my understanding of data analytics.

PROJECT QUESTION

1. What are total sales and total profits of each year.
2. What region generates the highest sales and profits.
3. What state and city brings in the highest sales and profits.
4. The relationship between discount and sales and the total discount per category.
5. What category generates the highest sales and profits in each region and state.
6. What subcategory generates the highest sales and profits in each region and state.
7. What are the names of the products that are the most and least profitable to us?
8. What segment makes the most of our profits and sales.
9. How many customers do we have (unique customer IDs) in total and how much per region and state?
10. Customer rewards program.

FEATURES OF SUPERSTORE DATASET:

- The data is publicly available through Kaggle under https://www.kaggle.com/datasets/vivek468/superstore-dataset-final.

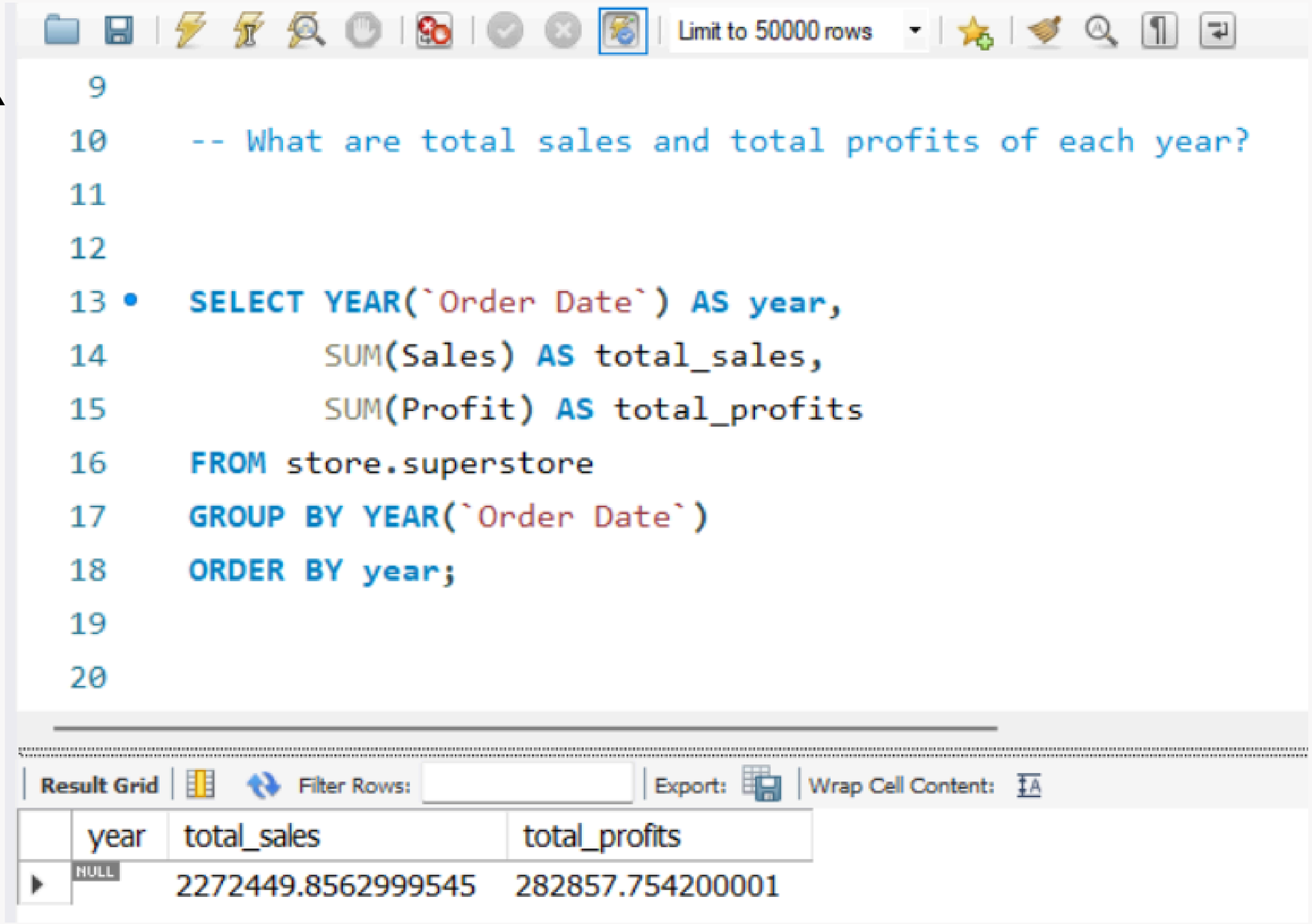
- It comes with 9995 rows with 9994 being pure data and the other one row being the column headers. It contains data recorded between the 3rd of January 2014 (the first order date) to the 5th of January 2018 (the last shipping date). (The last order date is the 30th of December 2017, so we will instead use the order dates range to represent our 4 years of business)

- It contains the data of 793 customers.

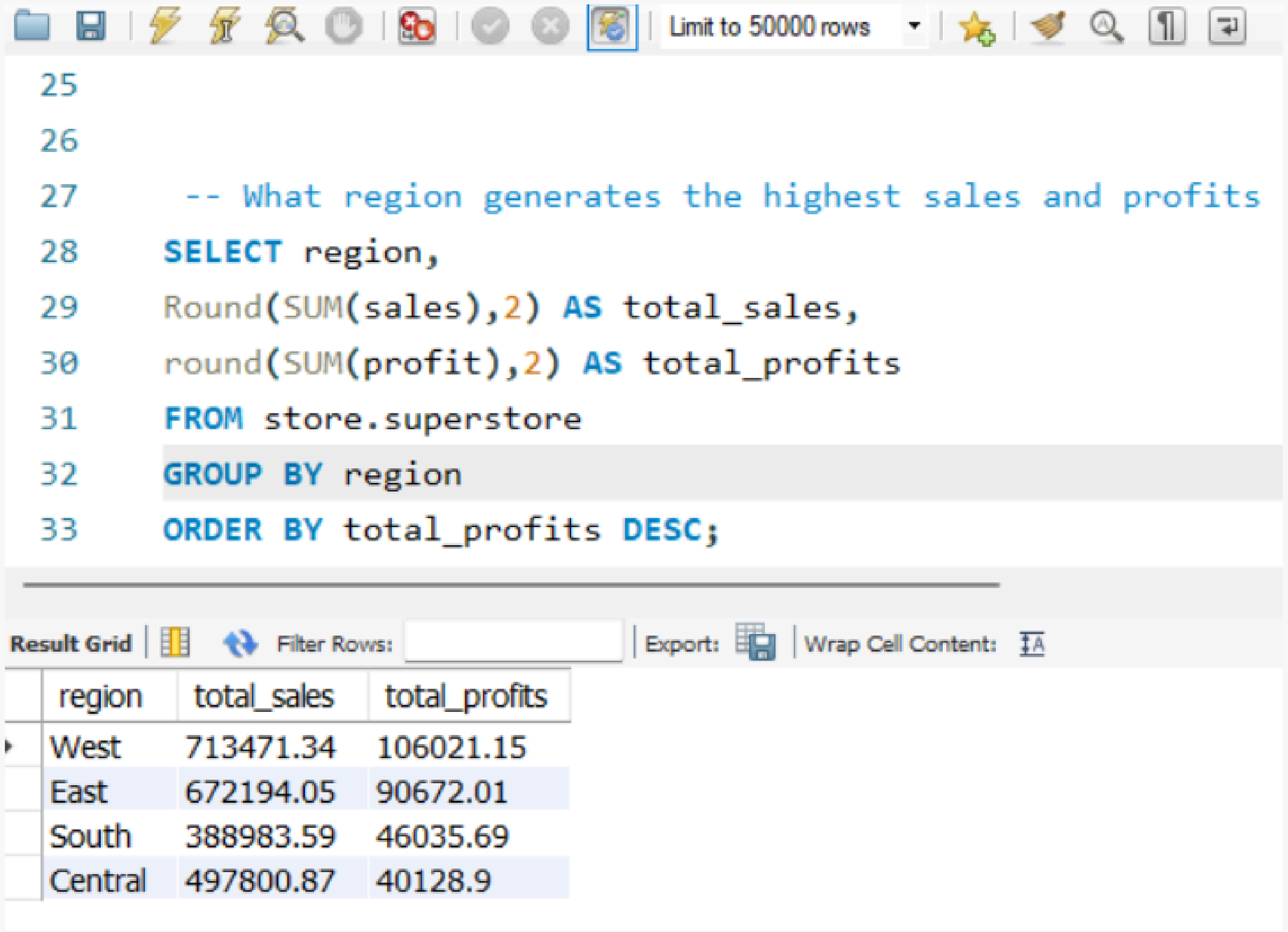
- The data contains the 21 columns namely; Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID , Customer Name, Segment, Postal Code, City, State, Country, Region, Product ID, Category, Sub-Category, Product Name, Sales, Quantity, Discount and Profit.

- The only limitations of our dataset that I could mention is that the most recent data point was almost 6 years ago. So our data is not current. However, our data is quite reliable, original, comprehensive and is cited.

-- What are total sales and total profits of each year?



What region generates the highest sales and profits



```
25
26
27      -- What region generates the highest sales and profits
28     SELECT region,
29     Round(SUM(sales),2) AS total_sales,
30     round(SUM(profit),2) AS total_profits
31     FROM store.superstore
32     GROUP BY region
33     ORDER BY total_profits DESC;
```

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

region	total_sales	total_profits
West	713471.34	106021.15
East	672194.05	90672.01
South	388983.59	46035.69
Central	497800.87	40128.9

Top 10 state and city brings in the highest sales and profits

```
40
41  -- top 10 state and city brings in the highest sales and profits
42
43  SELECT state,city, round(SUM(sales),2) AS total_sales, round(SUM(profit),2) AS total_profits,
44  ROUND((SUM(profit) / SUM(sales)) * 100, 2) as profit_margin
45  FROM store.superstore
46  GROUP BY state,city
47  ORDER BY total_profits DESC
48  limit 10;
```

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

state	city	total_sales	total_profits	profit_margin
New York	New York City	255248.97	61624.06	24.14
California	Los Angeles	173168.87	29806.92	17.21
Washington	Seattle	117772.58	28869	24.51
California	San Francisco	110917.04	17176.67	15.49
Michigan	Detroit	42302.9	13117.05	31.01
Indiana	Lafayette	19596.08	8959.26	45.72
Delaware	Newark	20289.21	8008.34	39.47
Georgia	Atlanta	17197.84	6993.66	40.67
Minnesota	Minneapolis	16870.54	6824.58	40.45
California	San Diego	47115.06	6309.3	13.39

The relationship between discount and sales and the total discount per category

```
49
50    -- The relationship between discount and sales and the total discount per category
51 •  SELECT Discount, round(AVG(Sales),2) AS Avg_Sales
52    FROM store.superstore
53    GROUP BY Discount
54    ORDER BY Discount;
55
```

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

Discount	Avg_Sales
0	230.36
0.1	578.4
0.15	529.97
0.2	213.22
0.3	454.74
0.32	536.79
0.4	567.54
0.45	498.63
0.5	892.71
0.6	47.25
0.7	102.36
0.8	58.63

What category generates the highest sales and profits in each region and state

```sql
-- What category generates the highest sales and profits in each region and state ?

 SELECT category, round(SUM(sales),2) AS total_sales,  round(SUM(profit),2) AS total_profit
ROUND((SUM(profit) / SUM(sales)) * 100, 2) as profit_margin
from store.superstore
group by category
order by total_profits desc;
```

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

category	total_sales	total_profits	profit_margin
Technology	835900.07	145387.1	17.39
Office Supplies	703502.93	120489.89	17.13
Furniture	733046.86	16980.77	2.32

What subcategory generates the highest sales and profits in each region and state

```sql
-- What subcategory generates the highest sales and profits in each region and state ?
SELECT `Sub-Category`,
       ROUND(SUM(sales), 2) AS total_sales,
       ROUND(SUM(profit), 2) AS total_profits,
       ROUND((SUM(profit) / SUM(sales)) * 100, 2) AS profit_margin
FROM store.superstore
GROUP BY `Sub-Category`
ORDER BY total_profits DESC;
```

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

Sub-Category	total_sales	total_profits	profit_margin
Copiers	149528.03	55617.82	37.2
Phones	329753.09	44447.88	13.48
Accessories	167380.32	41936.64	25.05
Paper	75356.12	32712.17	43.41
Binders	199905.72	29983.02	15
Chairs	328449.1	26590.17	8.1
Storage	216803.21	21527.91	9.93
Appliances	107532.16	18138.01	16.87
Furnishings	82752.23	11588.64	14
Art	27118.79	6527.79	24.07
Envelopes	15339.49	6460.87	42.12
Labels	12486.31	5546.25	44.42
Machines	189238.63	3384.76	1.79
Fasteners	3008.66	942.44	31.32
Supplies	45952.47	-1348.57	-2.93
Bookcases	114880	-3472.56	-3.02
Tables	206965.53	-17725.48	-8.56

What are the names of the products that are the least profitable to us?

```sql
-- What are the names of the products that are the most and least profitable to us?

SELECT `Product Name`, region ,
       round(SUM(sales),2) AS total_sales,
       round(SUM(profit),2) AS total_profit
FROM store.superstore
GROUP BY `Product Name`,region
ORDER BY total_profit ASC
limit 5;

```

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

Product Name	region	total_sales	total_profit
Cubify CubeX 3D Printer Double Hea...	East	6299.98	-9239.97
Cubify CubeX 3D Printer Triple Head ...	South	7999.98	-3839.99
Lexmark MX611dhe Monochrome La...	West	2549.98	-3399.98
GBC DocuBind P400 Electric Binding ...	Central	8710.34	-3048.62
Chromcraft Bull-Nose Wood Oval Co...	South	6611.76	-2865.1

What are the names of the products that are the most profitable to us?

```sql
-- What are the names of the products that are the most and least profitable to us?

SELECT `Product Name`, region ,
       round(SUM(sales),2) AS total_sales,
       round(SUM(profit),2) AS total_profit
FROM store.superstore
GROUP BY `Product Name`,region
ORDER BY total_profit DESC
limit 5;
```

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

Product Name	region	total_sales	total_profit
Canon imageCLASS 2200 Advanced ...	East	30099.91	10079.97
Canon imageCLASS 2200 Advanced ...	Central	17499.95	8399.98
Canon imageCLASS 2200 Advanced ...	West	13999.96	6719.98
Fellowes PB500 Electric Punch Plastic ...	South	7625.94	3812.97
GBC Ibimaster 500 Manual ProClick Bi...	Central	10653.72	3804.9

What segment makes the most of our profits and sales

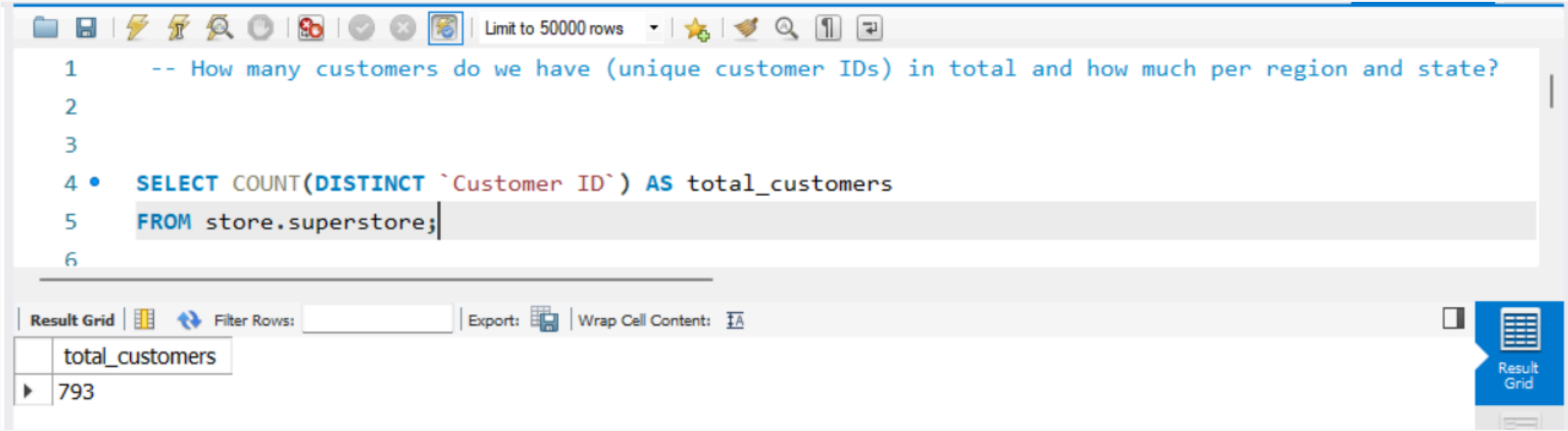
```sql
-- What segment makes the most of our profits and sales

SELECT segment, region ,
       round(SUM(sales),2) AS total_sales,
       round(SUM(profit),2) AS total_profit
FROM store.superstore
GROUP BY segment,region
ORDER BY total_profit DESC;
```

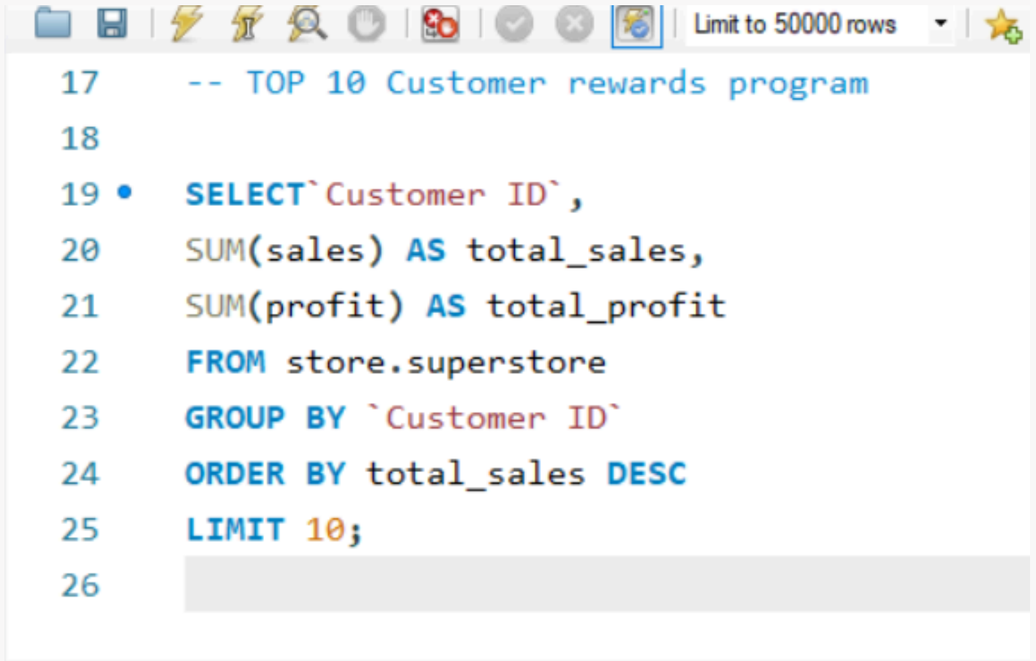
Result Grid | Filter Rows: | Export: | Wrap Cell Content:

segment	region	total_sales	total_profit
Consumer	West	357818.71	56307.68
Consumer	East	347820.22	40418.12
Corporate	West	221531.06	33496.18
Consumer	South	194634.69	26901.63
Home Office	East	126875.7	26611.18
Corporate	East	197498.14	23642.71
Corporate	Central	157157.68	18661.51
Home Office	West	134121.58	16217.29
Corporate	South	120417.63	14565.9
Home Office	Central	90750.61	12425.05
Consumer	Central	249892.58	9042.35
Home Office	South	73931.27	4568.16

How many customers do we have (unique customer IDs) in total and how much per region and state?



Top 10 Customer rewards program



```
17    -- TOP 10 Customer rewards program
18
19 •  SELECT`Customer ID`,
20    SUM(sales) AS total_sales,
21    SUM(profit) AS total_profit
22    FROM store.superstore
23    GROUP BY `Customer ID`
24    ORDER BY total_sales DESC
25    LIMIT 10;
26
```

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

Customer ID	total_sales	total_profit
SM-20320	25043.05	-1980.7392999999997
TC-20980	19017.847999999998	8964.482600000001
RB-19360	15117.339	6976.0959
TA-21385	14595.62	4703.788299999999
AB-10105	14355.610999999997	5438.9075
SC-20095	14142.333999999999	5757.411899999999
KL-16645	14071.917	768.8711999999999
HL-15040	12873.297999999999	5622.4292000000005
SE-20110	12209.438000000002	2650.6768999999995
CC-12370	12129.072	2177.0493

Region wise total customer

```
21    --  Region wise total customer
22 •  SELECT region, COUNT('Customer ID') AS total_customers
23    FROM store.superstore
24    GROUP BY region
25    ORDER BY total_customers DESC;
26
```

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

region	total_customers
West	3099
East	2756
Central	2260
South	1579

THANK YOU!