

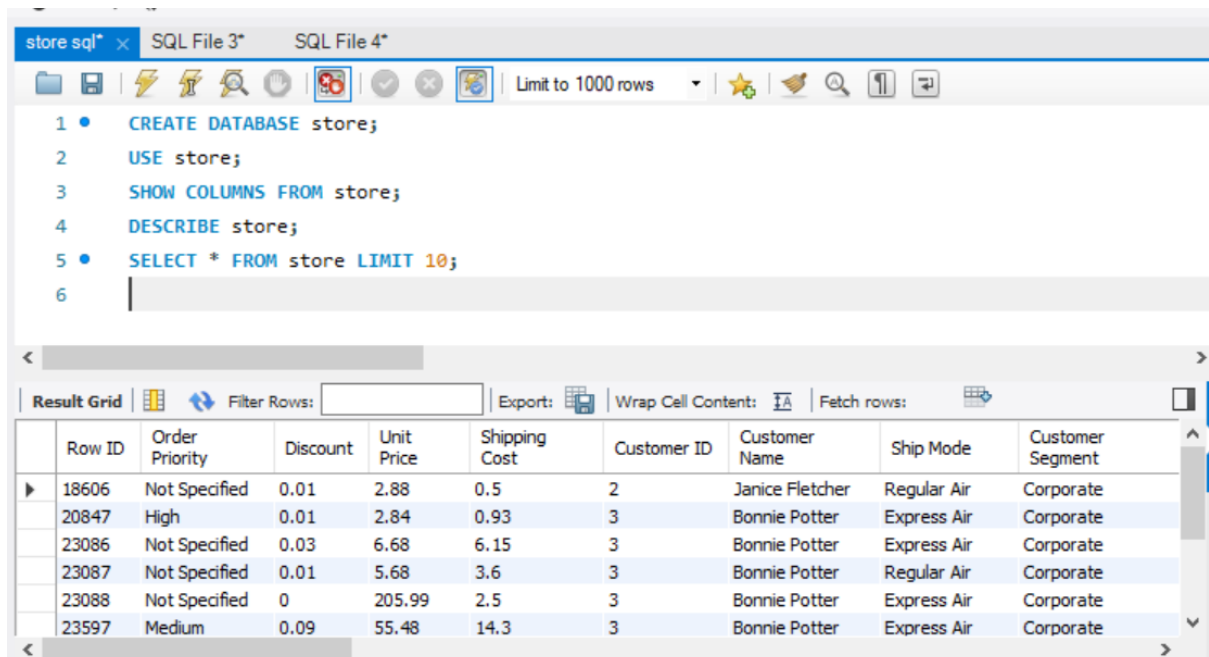
STORE ANALYSIS with SQL

1. Introduction

This report provides a comprehensive analysis of sales data from a retail store using SQL (MySQL). The dataset includes order details such as sales, profit, shipping information, customer segments, and product categories. The objective is to derive key business insights that can guide decision-making in areas like revenue, product performance, and shipping efficiency.

2. Initial Setup

We begin by creating and selecting the database, followed by examining the structure and previewing initial records.



The screenshot shows a SQL IDE with a script editor and a result grid. The script editor contains the following SQL commands:

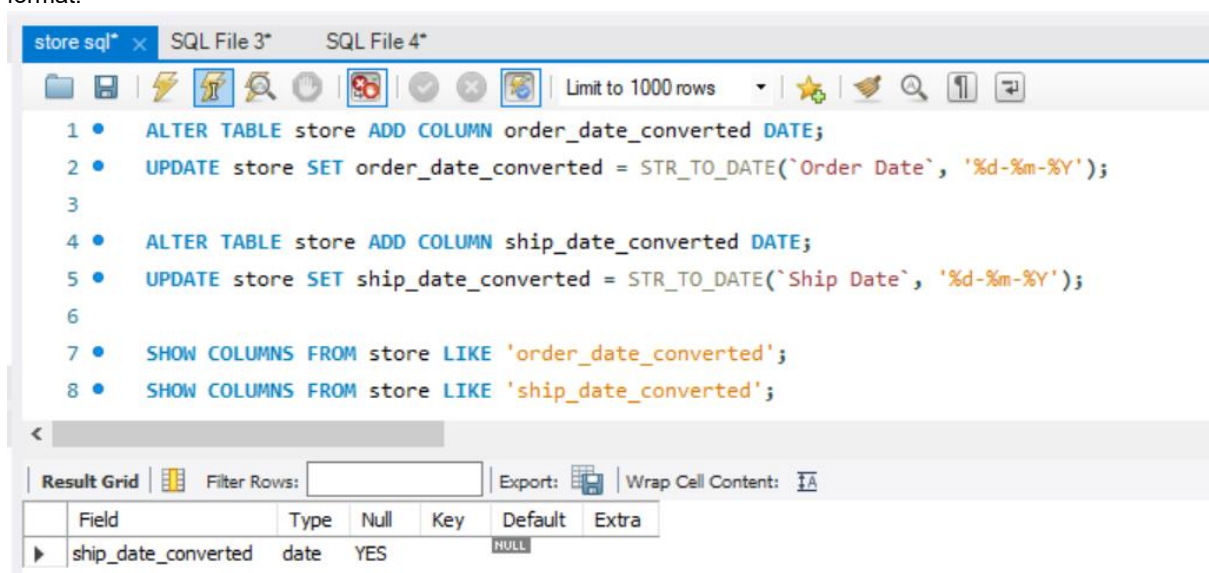
```
1 • CREATE DATABASE store;
2 • USE store;
3 • SHOW COLUMNS FROM store;
4 • DESCRIBE store;
5 • SELECT * FROM store LIMIT 10;
6
```

The result grid displays the following data:

Row ID	Order Priority	Discount	Unit Price	Shipping Cost	Customer ID	Customer Name	Ship Mode	Customer Segment
18606	Not Specified	0.01	2.88	0.5	2	Janice Fletcher	Regular Air	Corporate
20847	High	0.01	2.84	0.93	3	Bonnie Potter	Express Air	Corporate
23086	Not Specified	0.03	6.68	6.15	3	Bonnie Potter	Express Air	Corporate
23087	Not Specified	0.01	5.68	3.6	3	Bonnie Potter	Regular Air	Corporate
23088	Not Specified	0	205.99	2.5	3	Bonnie Potter	Express Air	Corporate
23597	Medium	0.09	55.48	14.3	3	Bonnie Potter	Express Air	Corporate

3. Data Cleaning and Preparation

To analyse dates effectively, we convert the Order Date and Ship Date columns from string format to SQL date format.

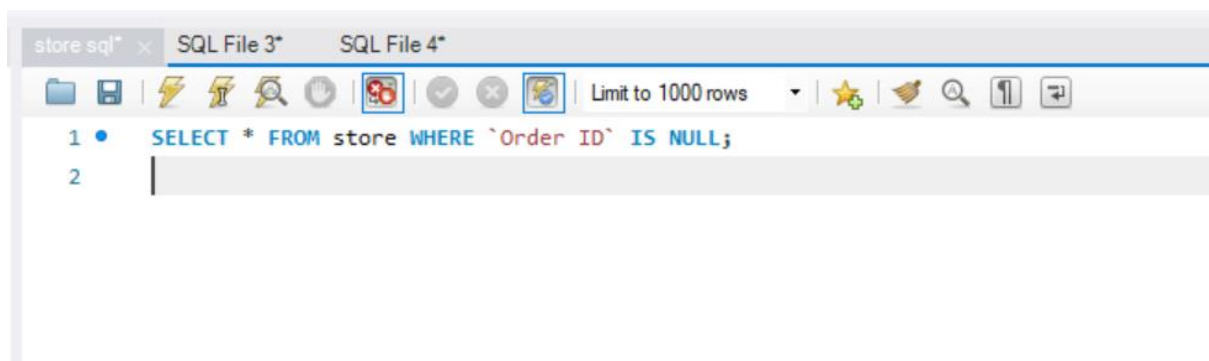


The screenshot shows a SQL IDE with a script editor and a result grid. The script editor contains the following SQL commands:

```
1 • ALTER TABLE store ADD COLUMN order_date_converted DATE;
2 • UPDATE store SET order_date_converted = STR_TO_DATE('Order Date', '%d-%m-%Y');
3
4 • ALTER TABLE store ADD COLUMN ship_date_converted DATE;
5 • UPDATE store SET ship_date_converted = STR_TO_DATE('Ship Date', '%d-%m-%Y');
6
7 • SHOW COLUMNS FROM store LIKE 'order_date_converted';
8 • SHOW COLUMNS FROM store LIKE 'ship_date_converted';
```

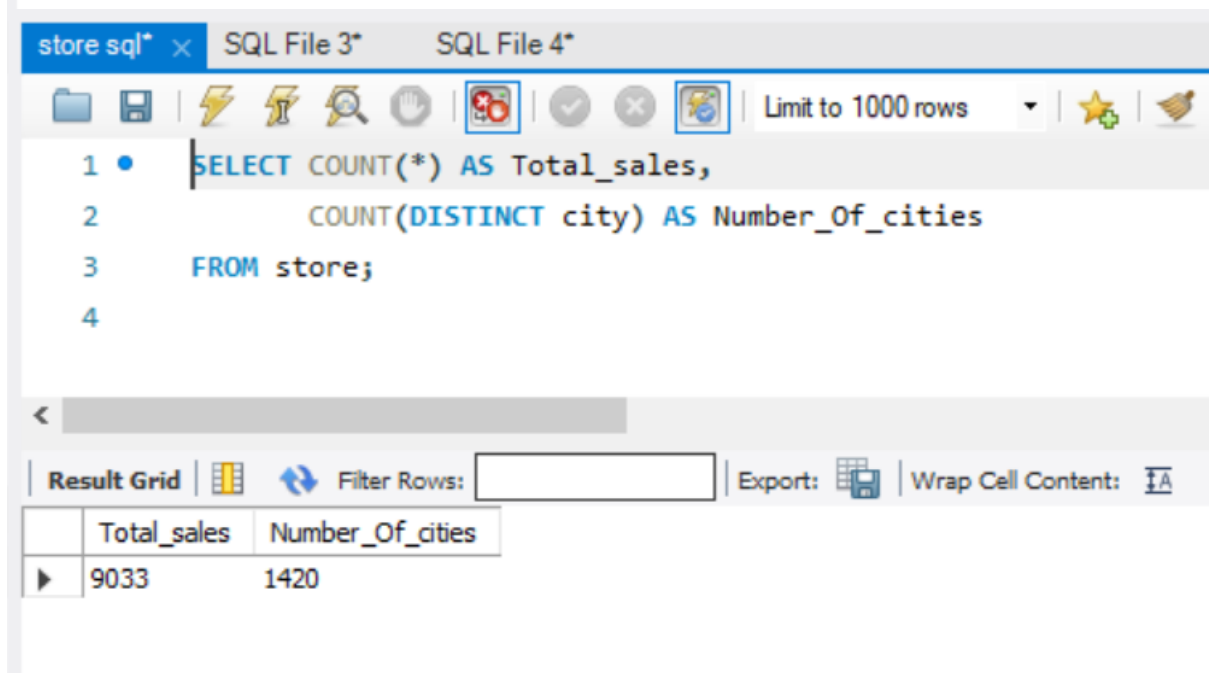
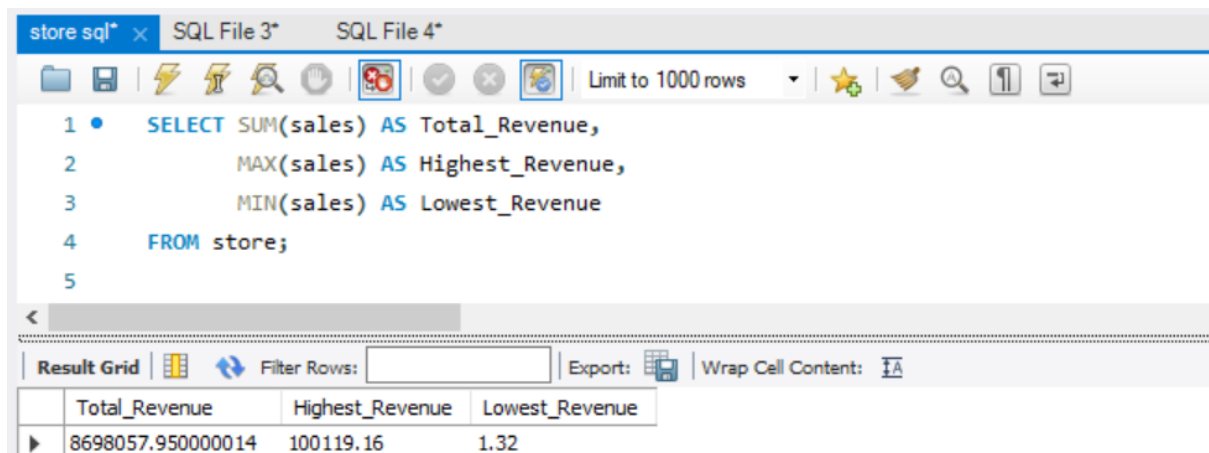
The result grid displays the following data:

Field	Type	Null	Key	Default	Extra
ship_date_converted	date	YES		NULL	



4. Revenue and Sales Overview

We analyse total, maximum, and minimum revenue, as well as the total number of sales and number of cities involved.



5. Yearly Sales and Profit Growth

We calculate year-over-year growth in both sales and profit using common table expressions (CTEs).

```
1 with cte as(select
2     YEAR(order_date_converted) as year,
3     Round(sum(sales), 2) as total_sales,
4     Round(sum(profit), 2) as total_profit
5     from store group by YEAR(order_date_converted)),
6 yoy_growth as (select
7     year, total_sales, total_profit,
8     Lag(total_sales) over ( order by year ) as previous_total_sales,
9     Lag(total_profit) over ( order by year ) as previous_total_profit
10    from cte),
11 yoy_growth_percentage as ( select
12     year, total_sales, total_profit, previous_total_sales, previous_total_profit,
13     Round(((total_sales / previous_total_sales) -1) * 100, 2) as sales_growth_percentage,
14     Round(((total_profit / previous_total_profit) -1) * 100, 2) as profit_growth_percentage
15    from yoy_growth )
16 select
17     year, total_sales, total_profit, sales_growth_percentage, profit_growth_percentage
18 from yoy_growth_percentage;
```

Result Grid

	year	total_sales	total_profit	sales_growth_percentage	profit_growth_percentage
▶	2010	1864104.43	201679.48	NULL	NULL
	2011	1913396.26	285359.03	2.64	41.49
	2012	2118545.49	333336.65	10.72	16.81
	2013	2802011.77	429820.83	32.26	28.94

There is a 32.26% YoY growth between 2012 to 2013.

The Store total sales grew from \$1,864,104.43 in 2010 to \$2,802,011.77 in 2017, an increase of 50.31%.

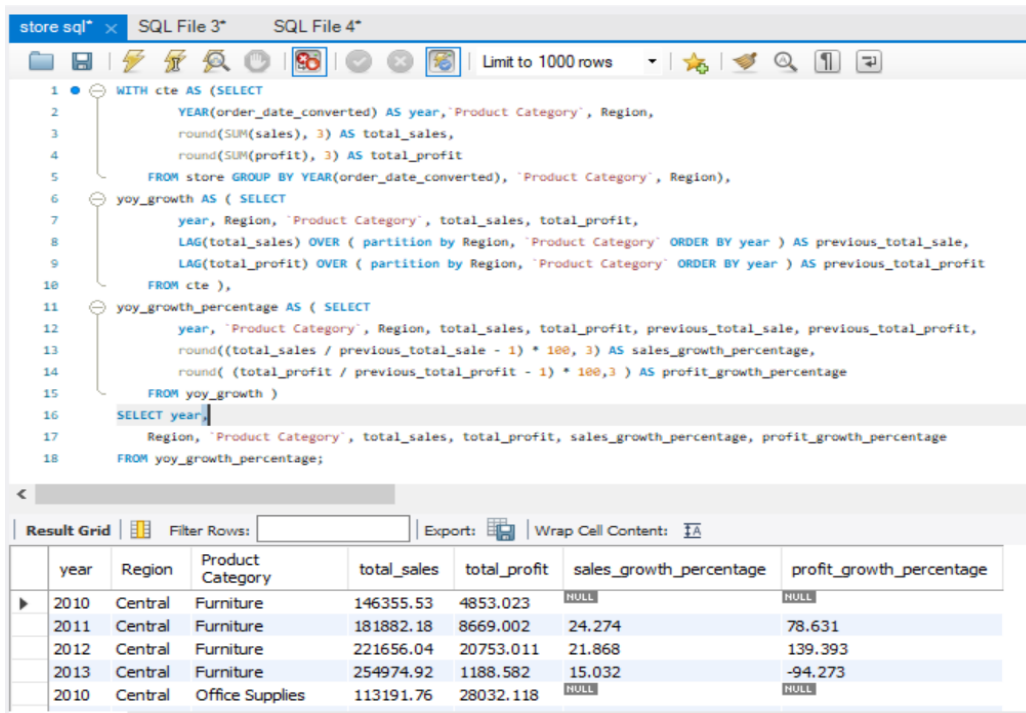
The total profit grew from \$201,679.48 to \$ 429,820.83 in the same period, an increase of 113.12%.

The sales growth percentage was highest in 2013 (32.26%) and lowest in 2011 (2.64%).

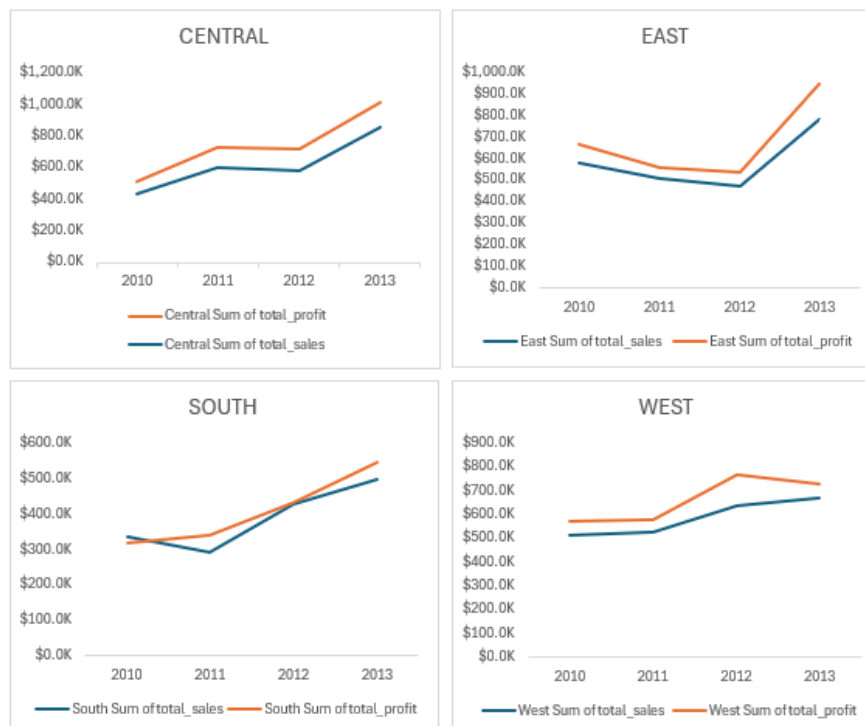
The profit growth percentage was highest in 2011 (41.49%) and lowest in 2012 (16.81%).

6. Category & Region-wise YoY Growth

This section analyses how each product category has performed in different regions on a yearly basis.



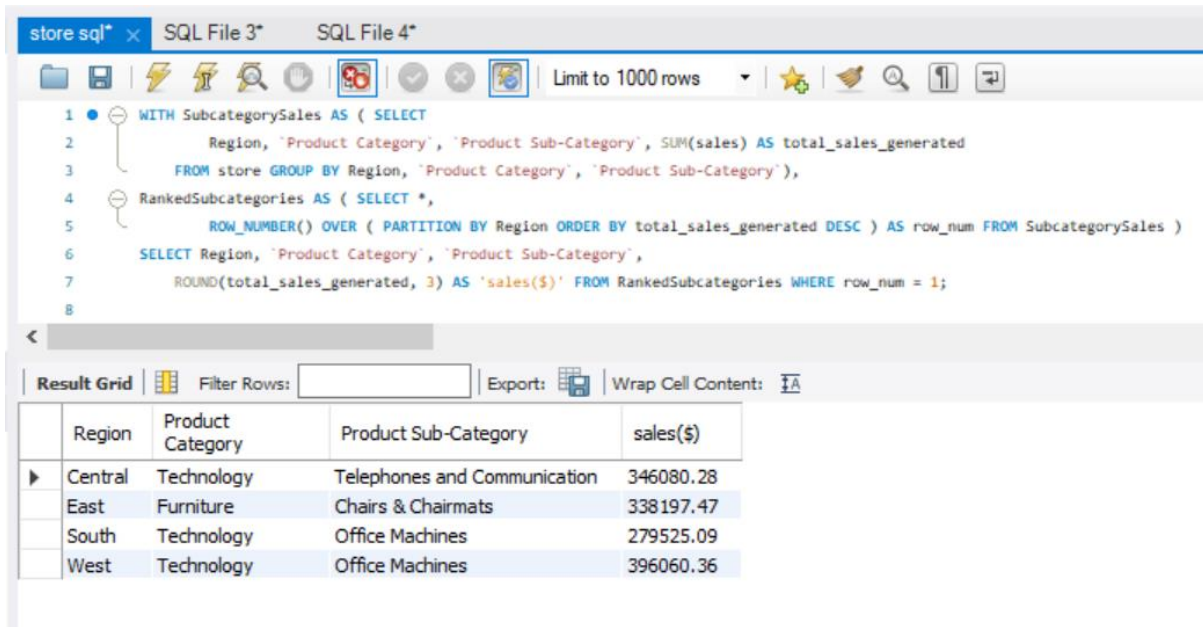
The results will have 48 rows, corresponding to the 4 regions * 3 categories * 4 years. This will allow us to analyse the year-over-year (YoY) growth for each region and category.



The Store experienced significant sales and profit growth from 2010 to 2013, with the most pronounced increase occurring in the Central region. The East region also saw strong growth, while the west region experienced more modest gains. In contrast, the South region had the least growth, although sales increased. Overall, each region demonstrated growth from 2010 to 2013. However, the South and West regions experienced steeper sales growth from 2010 to 2013, accompanied by negative profit growth. In contrast, the Central and East regions maintained consistent growth in both sales and profit from 2010 to 2013.

7. Top Revenue Generating Categories/Subcategories by Region

We determine which product sub-category generates the highest revenue in each region.



The screenshot shows a SQL query in SQL Studio. The query uses a Common Table Expression (CTE) named 'SubcategorySales' to calculate the total sales generated for each region, product category, and product sub-category. It then uses another CTE named 'RankedSubcategories' to rank these sub-categories by total sales within each region. Finally, it selects the top-ranked sub-category for each region, displaying the region, product category, product sub-category, and the rounded total sales.

```
1 WITH SubcategorySales AS ( SELECT
2     Region, 'Product Category', 'Product Sub-Category', SUM(sales) AS total_sales_generated
3 FROM store GROUP BY Region, 'Product Category', 'Product Sub-Category'),
4 RankedSubcategories AS ( SELECT *,
5     ROW_NUMBER() OVER ( PARTITION BY Region ORDER BY total_sales_generated DESC ) AS row_num FROM SubcategorySales )
6 SELECT Region, 'Product Category', 'Product Sub-Category',
7     ROUND(total_sales_generated, 3) AS 'sales($)' FROM RankedSubcategories WHERE row_num = 1;
```

The result grid shows the following data:

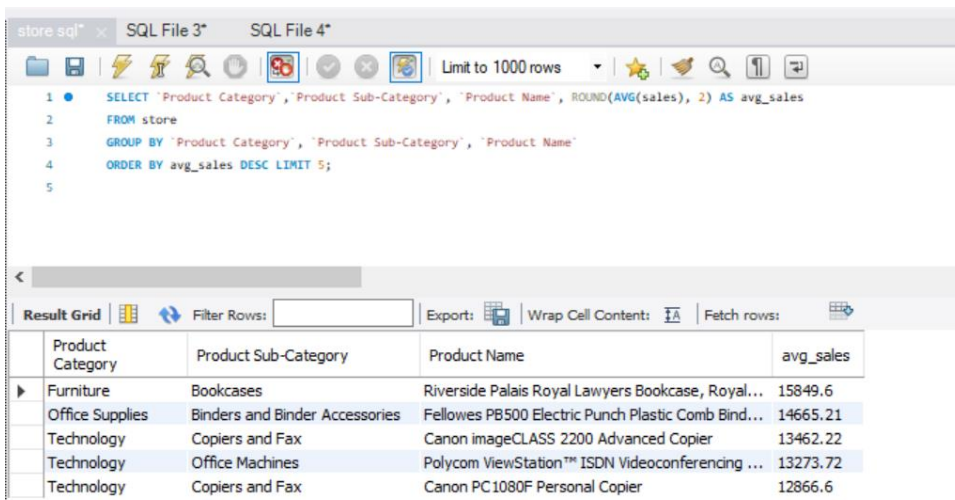
	Region	Product Category	Product Sub-Category	sales(\$)
▶	Central	Technology	Telephones and Communication	346080.28
	East	Furniture	Chairs & Chairmats	338197.47
	South	Technology	Office Machines	279525.09
	West	Technology	Office Machines	396060.36

The West region led in Office Machines sales from the Technology category, totalling \$396,060.36. The Central region followed closely, with the highest Telephones and communication sales in the Technology category, amounting to \$346,080.28. The East region outperformed, with the highest chairs and chair mats sales, reaching \$338,197.47. The South region also saw strong office machine sales, totalling \$279,525.09.

In summary, the West and south regions excelled in office machine sales, while the East and central regions dominated in chairs & chair mats and Telephones and communication sales within their respective categories.

8. Top Performing and Underperforming Products

Top 5 Products with Highest Average Sales



The screenshot shows a SQL query in SQL Studio. The query selects the top 5 products with the highest average sales, grouped by product category, product sub-category, and product name. The average sales are calculated using the AVG function and rounded to two decimal places.

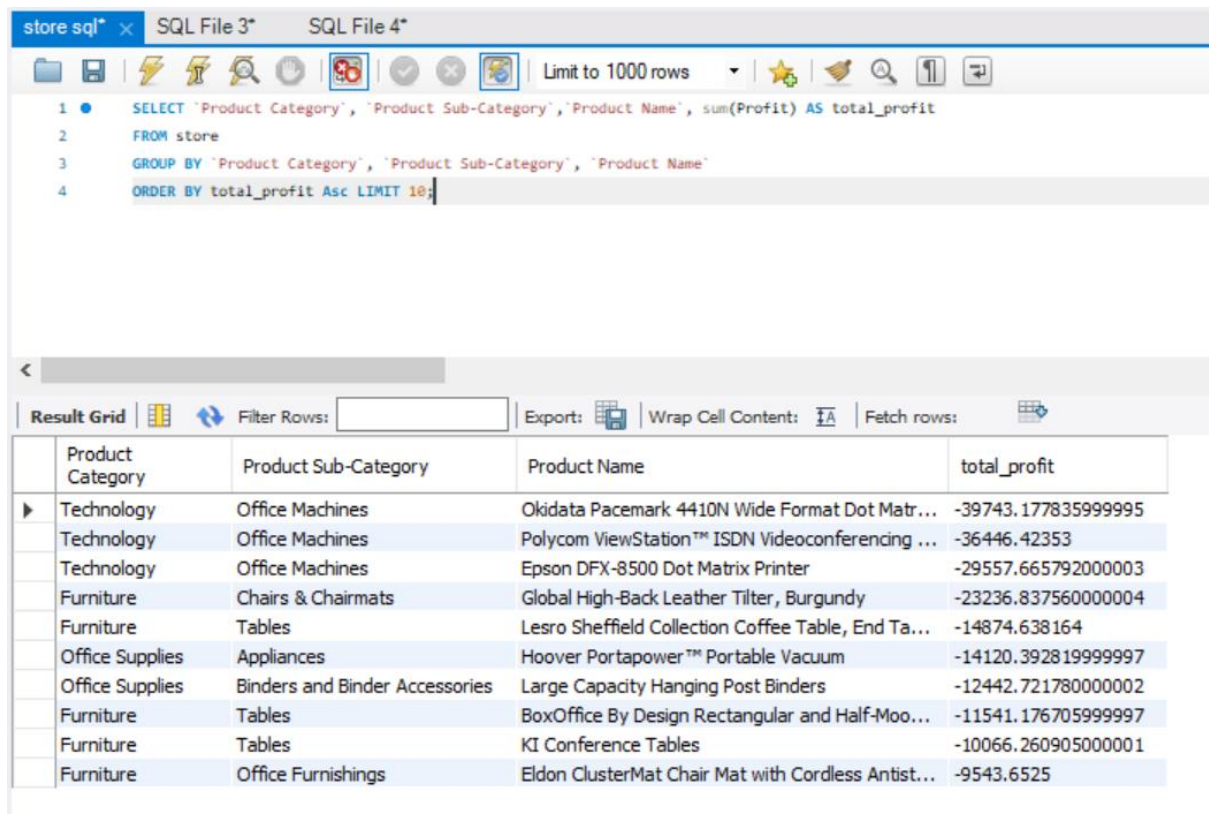
```
1 SELECT 'Product Category', 'Product Sub-Category', 'Product Name', ROUND(AVG(sales), 2) AS avg_sales
2 FROM store
3 GROUP BY 'Product Category', 'Product Sub-Category', 'Product Name'
4 ORDER BY avg_sales DESC LIMIT 5;
```

The result grid shows the following data:

	Product Category	Product Sub-Category	Product Name	avg_sales
▶	Furniture	Bookcases	Riverside Palais Royal Lawyers Bookcase, Royal...	15849.6
	Office Supplies	Binders and Binder Accessories	Fellowes PB500 Electric Punch Plastic Comb Bind...	14665.21
	Technology	Copiers and Fax	Canon imageCLASS 2200 Advanced Copier	13462.22
	Technology	Office Machines	Polycom ViewStation™ ISDN Videoconferencing ...	13273.72
	Technology	Copiers and Fax	Canon PC1080F Personal Copier	12866.6

The query finds the top 5 products with the highest average sales. It does this by grouping the data by product Category, product Name, and Product Sub-category, and then calculating the average sales for each product within each sub-category. The query then orders the results by the average sales in descending order.

Top 10 Loss-Making Products



store sql* x SQL File 3* SQL File 4*

Limit to 1000 rows

```
1 SELECT 'Product Category', 'Product Sub-Category', 'Product Name', sum(Profit) AS total_profit
2 FROM store
3 GROUP BY 'Product Category', 'Product Sub-Category', 'Product Name'
4 ORDER BY total_profit Asc LIMIT 10;
```

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

	Product Category	Product Sub-Category	Product Name	total_profit
▶	Technology	Office Machines	Okidata Pacemark 4410N Wide Format Dot Matr...	-39743.177835999995
	Technology	Office Machines	Polycom ViewStation™ ISDN Videoconferencing ...	-36446.42353
	Technology	Office Machines	Epson DFX-8500 Dot Matrix Printer	-29557.665792000003
	Furniture	Chairs & Chairmats	Global High-Back Leather Tilter, Burgundy	-23236.837560000004
	Furniture	Tables	Lesro Sheffield Collection Coffee Table, End Ta...	-14874.638164
	Office Supplies	Appliances	Hoover Portapower™ Portable Vacuum	-14120.392819999997
	Office Supplies	Binders and Binder Accessories	Large Capacity Hanging Post Binders	-12442.721780000002
	Furniture	Tables	BoxOffice By Design Rectangular and Half-Moo...	-11541.176705999997
	Furniture	Tables	KI Conference Tables	-10066.260905000001
	Furniture	Office Furnishings	Eldon ClusterMat Chair Mat with Cordless Antist...	-9543.6525

The query finds the top 10 products with the lowest total profit. It does this by grouping the data by product name, subcategory, and category, and then ordering the results by total profit in ascending order. This means that the query will return the products with the lowest total profit first.

9. Customer Segment Analysis

We identify which customer segment places the highest number of orders in each state.

The screenshot shows a SQL query in a window titled 'SQL File 4*'. The query uses Common Table Expressions (CTEs) to calculate the total orders for each customer segment by state and then identify the segment with the highest number of orders in each state.

```
1 WITH cte AS ( SELECT
2     'State or Province', 'Customer Segment', COUNT('Order ID') AS total_orders
3 FROM store
4 GROUP BY 'State or Province', 'Customer Segment' ),
5 cte2 AS ( SELECT
6     'State or Province', MAX(total_orders) AS max_orders
7 FROM cte
8 GROUP BY 'State or Province')
9 SELECT a.'State or Province', a.'Customer Segment', a.total_orders
10 FROM cte AS a
11 JOIN cte2 AS b ON a.'State or Province' = b.'State or Province' AND a.total_orders = b.max_orders
12 ORDER BY 'State or Province';
```

The result grid below the query shows the following data:

	State or Province	Customer Segment	total_orders
▶	Alabama	Home Office	55
	Arizona	Home Office	51
	Arkansas	Corporate	57
	California	Corporate	368
	Colorado	Home Office	67
	Connecticut	Corporate	30
	Delaware	Consumer	13
	District of Columbia	Corporate	23
	Florida	Consumer	160
	Georgia	Corporate	on

10. Top 15 States by Number of Orders

This section provides insights into the top-performing states in terms of order count, including average order value and average profit per order.

The screenshot shows a SQL query in a window titled 'SQL File 3*'. The query selects the top 15 states based on the number of orders, along with their region, average order value, and average profit per order.

```
1 select
2     'State or Province', Region, count(*) as no_of_orders, round(Avg(sales), 2) as avg_order_value,
3     round(avg(profit), 2) as profit_per_order
4 from store
5 group by 'State or Province', Region order by count(*) desc limit 15
```

The result grid below the query shows the following data:

	State or Province	Region	no_of_orders	avg_order_value	profit_per_order
▶	California	West	980	1156.15	82.15
	Texas	Central	618	862.97	170.35
	Illinois	Central	557	1153.72	207.47
	New York	East	553	1451.84	194.19
	Florida	South	495	983.13	36.25
	Ohio	East	382	746.08	176.72
	Washington	West	320	1567.14	129.22
	Michigan	Central	317	1015.04	154.7

11. Shipping Analysis

Average Time to Ship an Order

We calculate the average number of days taken to ship an order after placement.

```
1 SELECT
2     ROUND(AVG(DATEDIFF(ship_date_converted, order_date_converted)), 2) AS avg_shipping_days
3 FROM store;
4
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: [IA](#)

avg_shipping_days
2.04

Order Distribution by Shipment Type

store sql* x SQL File 3* SQL File 4*

Limit to 1000 rows

```
1 SELECT `Ship Mode`, COUNT(*) AS total_orders,
2     ROUND((COUNT(*) * 100.00 / (SELECT COUNT(*) from store)), 2) AS percentage
3 FROM store
4 GROUP BY `Ship Mode`
5 ORDER BY percentage DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: [IA](#)

Ship Mode	total_orders	percentage
Regular Air	6715	74.34
Delivery Truck	1266	14.02
Express Air	1052	11.65

CONCLUSION

The SQL-based analysis has provided valuable insights into various aspects of the business, including revenue generation, product performance, regional trends, and customer behavior. These findings can assist in making data-driven strategic decisions to optimize product offerings, marketing efforts, and logistics operations.