

A Data-Driven Exploration of Sales and Profit Optimization in the Superstore Database

Objective

In today's competitive business landscape, data-driven decision-making has become paramount for organizations seeking to thrive and outperform their competitors. The [Superstore Dataset](#) has a ton of information about sales and profit, and we're going to dig into it using SQL to find out all sorts of useful things.

Our main goal is to make the store's sales and profit better by finding patterns, spotting trends, and getting useful insights from the data. With SQL, we can look closely at everything about the Superstore – from how sales vary in different regions to which products make the most money, who the best customers are, product profitability, customer segmentation, and shipping efficiency.

Through preparation and analysis of data, we aim to explore avenues, for the Superstore's expansion streamline its operations, and ultimately boost profitability. Leveraging SQL as our tool we embark on a journey of exploration, learning, and strategic decision-making to propel the Superstore towards success, in the realm of retail.

1. Sales and Profit Analysis

Analyzing sales and profit trends over time provides insights into the Superstore's financial performance and growth trajectory. So we need to perform data analysis on the superstore dataset.

1. To obtain the first 5 rows from the database

```
SELECT top 5 *  
FROM [SalesAnalysis].[dbo].[superstore]
```

2. Total Sales year-wise

```
SELECT YEAR(Order_Date) AS Year,  
SUM(sales) AS Total_Sales  
FROM [SalesAnalysis].[dbo].[superstore]  
GROUP BY YEAR(Order_Date)  
ORDER BY Year;
```

One might encounter the same issue with the data type mismatch for the profit column if it's stored as nvarchar. This query will sum up the profits yearwise, converting the profit column to a numerical data type using CAST(), similar to the previous example with the sales column.

Note: Using nvarchar datatype for storing profit values in a database isn't a standard practice, as nvarchar is typically used for storing Unicode character data. Storing numerical data such as profit as nvarchar can lead to issues with calculations, sorting, and querying. so we change the data type of profit. So in the next step, we change the datatype of Profit.

--Note

```
ALTER TABLE [SalesAnalysis].[dbo].[superstore]  
ALTER COLUMN profit DECIMAL(10,4);
```

--3 Total Profit yearwise

```
SELECT YEAR(Order_Date) AS Year,  
SUM(profit) AS Total_Profit  
FROM [SalesAnalysis].[dbo].[superstore]  
GROUP BY YEAR(Order_Date)  
ORDER BY Year;
```

The above data shows the profit over the years has increased with each year although there is a fall in total sales in 2015 after 2014.

1. Regional Analysis

--4. Total Sales and Profit Region-wise

```
SELECT region,  
ROUND(SUM(sales),2) AS Total_Sales, SUM(profit) AS Total_Profit  
FROM [SalesAnalysis].[dbo].[superstore]  
GROUP BY region  
ORDER BY Total_Profit DESC;
```

Most of the sales and profit is brought by the WEST region. Next is the east region which is also beneficial for the company. So these two regions are of our interest. While the central region does not make much profit. So to deal with more observation we need to check on the profit margin of this region.

5. Profit Margin

```
SELECT region,  
ROUND((SUM(profit)/sum(sales)), * 100,2) AS Profit_Margin
```

```
FROM [SalesAnalysis].[dbo].[superstore]
GROUP BY region
ORDER BY Profit_Margin DESC;
```

Profit margin measures the profitability and revenue of the company. Just as the sales and profit of east and West the profit is good. Although the south region has fewer sales but the profit margin is relatively good. The central region does not provide much profit altogether. So next we expand to states and respective cities.

4. States and Cities Analysis

Analyzing sales and profit metrics at the state and city levels provides insights into localized market dynamics and opportunities.

--Top 10 states by sales and profit

```
SELECT TOP 10 states,
SUM(Sales) AS Total_Sales,
SUM(Profit) AS Total_Profits,
ROUND((SUM(Profit)/SUM(Sales))*100,2) AS profit_margin
FROM [SalesAnalysis].[dbo].[superstore]
GROUP BY state
ORDER BY Total_Profits DESC;
```

For states and cities, we take into consideration the sales, profit, and profit margin. In terms of sales and profit California, New York and Washington give a good result.

The least profitable states are Texas, Ohio, and Pennsylvania which have more Sales than Washington (one of the top profitable states) but these states are in loss.

--top 10 cities by sales and profit

```
SELECT TOP 10 CITY,
SUM(Sales) AS Total_Sales,
SUM(Profit) AS Total_Profits,
ROUND((SUM(Profit)/SUM(Sales))*100,2)AS profit_margin
FROM [SalesAnalysis].[dbo].[superstore]
GROUP BY city
ORDER BY Total_Profits DESC;
```

The top 3 cities are New York City, Los Angeles and Seattle.

--Note

```
ALTER TABLE [SalesAnalysis].[dbo].[superstore]
ALTER COLUMN discount decimal(10,4);
```

The discount also has some impact on sales of a product so we need to find the relationship between them.

```
SELECT Discount, AVG(Sales) AS Avg_Sales
FROM [SalesAnalysis].[dbo].[superstore]
GROUP BY Discount
ORDER BY Discount;
```

For each discount, the sales vary. There is no linear relationship between discounts and sales. But we observe that if the discount is 50 % then the average sales is highest. So now we move towards identifying discount and product relationships.

```
SELECT category, SUM(discount) AS total_discount
FROM [SalesAnalysis].[dbo].[superstore]
GROUP BY category
ORDER BY total_discount DESC;
```

So the most discounted item is Office Supplies

– Most discounted subcategory

```
SELECT category, Sub_Category, SUM(discount) AS total_discount
FROM [SalesAnalysis].[dbo].[superstore]
GROUP BY category, Sub_Category
ORDER BY total_discount DESC;
```

The top three discounted products are Binders, Phones, and Furnishing. But we notice that office supplies-binders have a huge margin concerning other sub-categories. Focusing on binders we can check its sales and profit. So for category, we check it with respective states to explore the category.

–Sales , profit, and margin wrt category

```
SELECT category, SUM(sales) AS total_sales, SUM(profit) AS total_profit,
ROUND(SUM(profit)/SUM(sales)*100, 2) AS profit_margin
FROM [SalesAnalysis].[dbo].[superstore]
GROUP BY category
ORDER BY total_profit DESC;
```

Here we find out which category has the highest sales, profit, and margin. So technology has the highest sales, profit, and profit margin. We also notice that although furniture has good sales but it doesn't make a good profit margin. So next we check total sales and profit in each region.

--Highest total sales and profits per Category in each region

```
SELECT region, category, SUM(sales) AS Total_Sales, SUM(profit) AS Total_Profit
FROM [SalesAnalysis].[dbo].[superstore]
GROUP BY region, category
ORDER BY total_profit DESC;
```

We notice that the West is the top region in the category of office supplies and technology. Now we check state-wise.

```
Select DISTINCT category
FROM [SalesAnalysis].[dbo].[superstore]
```

--to obtain the highest sales and profit per category per state

```
SELECT state, category, SUM(sales) AS Total_Sales, SUM(profit) AS Total_Profit
FROM [SalesAnalysis].[dbo].[superstore]
GROUP BY state, category
ORDER BY total_profit DESC;
```

Here we have state and category which shows that the top 3 categories are technology in New York and California, and office supplies in California.

--total sales and total profits of each subcategory with their profit margins

```
SELECT sub_category, SUM(sales) AS total_sales, SUM(profit) AS total_profit,
ROUND(SUM(profit)/SUM(sales)*100, 2) AS profit_margin
FROM [SalesAnalysis].[dbo].[superstore]
GROUP BY sub_category
ORDER BY total_profit DESC;
```

--so copier has the highest sales margin in the subcategory

--To obtain which subcategory generate highest sales and profit regionwise

```
SELECT region, category, sub_category, SUM(sales) AS total_sales, SUM(profit) AS
total_profit
FROM [SalesAnalysis].[dbo].[superstore]
GROUP BY region, category, sub_category
ORDER BY total_profit DESC;
```

--highest total sales and total profits per subcategory statewide

```
SELECT state, sub_category, SUM(sales) AS total_sales, SUM(profit) AS
total_profit, ROUND(SUM(profit)/SUM(sales)*100, 2) AS profit_margin
FROM [SalesAnalysis].[dbo].[superstore]
```

```
GROUP BY state, sub_category  
ORDER BY total_profit DESC;
```

So machines, phones, and binders perform well in New York.

4. Product Analysis

Analyzing product categories and their performance helps identify high-profit opportunities and areas for improvement.

--least profitable and highest profitable product

```
SELECT product_name, SUM(sales) AS total_sales, SUM(profit) AS total_profit  
FROM [SalesAnalysis].[dbo].[superstore]  
GROUP BY product_name  
ORDER BY total_profit DESC;
```

These Copiers, Machines, and printers are the most profitable. So we also check the less profitable products.

```
SELECT product_name, SUM(sales) AS total_sales, SUM(profit) AS total_profit  
FROM [SalesAnalysis].[dbo].[superstore]  
GROUP BY product_name  
ORDER BY total_profit ASC;
```

Which segment makes the most profit and sales?

--to obtain a segment that makes maximum sales and profit

```
SELECT segment, SUM(sales) AS total_sales, SUM(profit) AS total_profit  
FROM [SalesAnalysis].[dbo].[superstore]  
GROUP BY segment  
ORDER BY total_profit DESC;
```

5. Total Number of Customer

--total customer

```
SELECT COUNT(DISTINCT customer_id) AS total_customers  
FROM [SalesAnalysis].[dbo].[superstore]
```

--total customer yearwise

```
SELECT YEAR(Order_Date) AS Year,
COUNT(DISTINCT customer_id) AS Total_Customers
FROM [SalesAnalysis].[dbo].[superstore]
GROUP BY YEAR(Order_Date)
ORDER BY Year;
```

--total customer regionwise :

```
SELECT region, COUNT(DISTINCT customer_id) AS total_customers
FROM [SalesAnalysis].[dbo].[superstore]
GROUP BY region
ORDER BY total_customers DESC;
```

Although we have 793 customers the region does not add up to 793. It shows double counting. Also, observe that West has the highest number of customers. So next we explore statewise.

--Total customer statewise :

```
SELECT state, COUNT(DISTINCT customer_id) AS total_customers
FROM [SalesAnalysis].[dbo].[superstore]
GROUP BY state
ORDER BY total_customers DESC;
```

So most customers are in California, NY and Texas while least is from Wyoming, north Dakota, and West Virginia.

6. Customer segmentation

Segmenting customers based on their spending behavior allows for targeted marketing and loyalty programs.

```
SELECT TOP 15 customer_id,
SUM(sales) AS total_sales,
SUM(profit) AS total_profit
FROM [SalesAnalysis].[dbo].[superstore]
GROUP BY customer_id
ORDER BY total_sales DESC;
```

Customer ID 'SM-20320' stands out as the top spender among our customers, despite not generating the highest profit. Yet in customer segmentation, he can hold the reward points. On the other hand, Customer ID 'TC-20980' ranks second in spending but contributes significantly

to our profits. Therefore, it's crucial to express our gratitude to our top customers and prioritize maintaining strong relationships with them.

There are many other ways to build loyalty and reward programs. As we use basic SQL we just find out the customer segmentation concerning the sales and profit made. While exploring other aspects we have different methods for Customer Segmentation:

There are various segmentation methods such as:

Demographic segmentation

Behavioural segmentation

Psychographic segmentation

Rfm analysis

Clustering algorithm

These methods can be dealt with in other section.

7. Shipping Analysis

Analyzing shipping times and modes provides insights into operational efficiency and customer satisfaction levels.

–Average shipping time

```
SELECT AVG(DATEDIFF(day, order_date, ship_date)) AS avg_shipping_time  
FROM [SalesAnalysis].[dbo].[superstore];
```

-- Shipping time in each shipping mode is

```
SELECT ship_mode,  
ROUND(AVG(DATEDIFF(day, order_date, ship_date)), 1) AS avg_shipping_time  
FROM [SalesAnalysis].[dbo].[superstore]  
GROUP BY ship_mode  
ORDER BY avg_shipping_time;
```

Conclusion

The Superstore Dataset analysis offers valuable insights into sales and profit optimization strategies. By leveraging these insights, the Superstore can refine its product offerings, target specific regions and customer segments, and enhance operational efficiency to drive sustainable growth and profitability.

Here are our conclusions and future recommendations based on the analysis of our Superstore data:

1. Overall Performance: Our profits have been improving steadily over time, along with our sales, even though we had a brief dip in 2015. It's important to keep up this positive momentum.

2. Regional Performance: The West region performs the best, followed by the East, South, and Central regions. Despite having higher sales than the South region, the Central region's profits are lower. We need to consider shifting resources from the Central region to strengthen our presence in the more profitable West region.

3. State and City Performance:

California, New York, and Washington are our most profitable states, while Texas, Ohio, and Pennsylvania are the least profitable. We should consider reducing our presence in the less profitable states and focus more on our top-performing markets.

New York City, Los Angeles, and Seattle are our most profitable cities, while Philadelphia, Houston, and San Antonio are the least profitable. We should prioritize our efforts in profitable cities and reconsider our presence in less profitable ones.

4. Product Category Analysis:

Technology and Office Supplies are our most profitable categories, whereas Furniture shows lower profitability. We need to explore ways to improve the profitability of our Furniture department.

Office supplies perform exceptionally well in the West region, indicating a need to increase stock in that category. Conversely, Furniture in the Central region is not profitable, suggesting a reallocation of resources to more profitable regions.

5. Subcategory Analysis: Copiers, Phones, Accessories, and Paper are our most profitable subcategories, while Tables, Bookcases, and Supplies show losses. We should focus on promoting profitable subcategories and reducing stock to unprofitable ones.

6. Customer Segment Analysis: The Consumer segment brings in the most profit, followed by Corporate and Home Office segments. We should prioritize catering to the Consumer segment to maximize profitability.

7. Customer Distribution: We have a total of 793 customers, with the highest concentration in California, New York, and Texas. Despite having many customers in Texas, it's our least profitable state. We need to make strategic decisions regarding our presence in Texas and focus on delivering exceptional service in California and New York.

By implementing these recommendations, we can strengthen our position in profitable markets, optimize our product offerings, and enhance overall profitability for the Superstore.