Superstore Sales Analysis

In today's hyper-competitive landscape, data is the new frontier. It's the untapped goldmine that can propel businesses to unprecedented heights. To strike it rich, we must possess a robust, data-driven infrastructure. Think of it as building a high-performance engine: powerful databases are the core, analytical departments are the ignition, and insightful interpretation is the fuel.

We're in a race against time. The market is evolving at breakneck speed, demanding innovation and agility. To stay ahead, we must transform data into actionable intelligence. By harnessing the power of Excel, SQL, and Tableau, we'll embark on a data mining expedition. We'll extract, refine, and visualize our treasure trove of information, uncovering hidden patterns and opportunities.

Our journey will follow a proven roadmap: question, prepare, process, analyze, share, and act. This systematic approach will ensure we maximize the value of our data, ultimately delivering exceptional customer experiences.

1. Defining the Business Objective:

In this step we'll define certain business objectives which will reflect on our objective that is to identify optimal product, regional, and customer segments for Superstore to maximize profitability and answer the question "How can we optimize Superstore's product mix, geographic focus, and customer targeting to drive profit growth?"

Objectives:

- How can we optimize our profits? What are the emerging trends that we can we identify?
- How can we take these insights to build recommendations

For the scope of this project, we'll be only delivering a data exploration analysis using SQL, and processing our data and deriving insights from the same.

2. Assessing features from Data:

Our Superstore Dataset offers a snapshot into four years of business activity, spanning from January 2014 to December 2017. This meticulously recorded chronicle captures the transactions of 793 customers across the United States, represented by a robust collection of 9994 data points distributed across 21 key attributes.

From order details and shipping logistics to customer demographics and product specifics, the dataset provides a granular view of the business landscape. While its currency is limited to a period ending nearly six years ago, its reliability, comprehensiveness, and original nature make it a valuable resource for retrospective analysis and trend identification.

Essentially, this dataset serves as a historical record, offering insights into past performance and potential benchmarks for future operations.

3. Processing the Data:

We used Microsoft Excel to get one big overview of the data.

Our dataset appeared initially clean and organized, promising a smooth analytical journey. To ensure data integrity, we employed a dual-pronged approach.

Unmasking Missing Values: First, we cast a wide net using conditional formatting. By highlighting empty cells in yellow, potential missing values were brought into sharp focus. A subsequent manual scan through the filtered dataset confirmed the absence of these elusive gaps.

Duplicate Detection: To rule out redundant entries, we leveraged Excel's 'Remove Duplicates' function. The lack of any matches indicated data uniqueness across all customer parameters.

Formatting for Clarity: Finally, we refined the dataset's visual appeal and interpretability. Numeric columns representing monetary values were formatted as currency, while the discount column, inherently a ratio, was prepared for percentage conversion. Dates were verified to be in their correct format.

Through these steps, we've established a solid foundation for our analysis, with data that is accurate, complete, and ready for deeper exploration.

Now our dataset is ideal for analysis to discover relationships, trends and patterns that will give us a competitive edge and completely solve our business objectives

4. Analyze:

For the analysis part, we will string out the most important components of our data to answer our business objectives. Let's load our data into PostgreSQL and check the first 5 rows to make sure it imported well.

Query | Query History

```
CREATE TABLE PUBLIC.superstore(
    row_id int,
    order_id varchar(100),
    order_date date,
    ship_date date,
    ship_mode varchar(100),
    customer_id varchar(100),
    customer_name varchar(100),
    segment varchar(100),
    country varchar(100),
    city varchar(100),
    state varchar(100),
    postal_code int,
    region varchar(100),
    product_id varchar(100),
    category varchar(100),
    sub_category varchar(100),
    product_name varchar(255),
    sales decimal(12,5),
    quantity int,
    discount decimal(5,2),
    profit decimal(12,5)
);

COPY public.superstore FROM 'C:\Users\Public\superstore_dataset_csv.csv' DELIMITER ',' CSV HEADER;


select * from superstore limit 5;
```

Data Output | Messages | Notifications

	row_id integer	order_id character varying (100)	order_date date	ship_date date	ship_mode character varying (100)	customer_id character varying (100)	customer_name character varying (100)	segment character varying (100)
1	1	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer
2	2	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute	Consumer
3	3	CA-2016-138688	2016-06-12	2016-06-16	Second Class	DV-13045	Darrin Van Huff	Corporate
4	4	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer
5	5	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell	Consumer

So the data is loaded and set up in the SQL environment for further analysis.

Let's uncover certain insights first:

1. **What are total sales and total profits of each year?**

We selected sum of sales and profit grouped by year to find sum of sales and sum of profits for each year.

```sql
SELECT
    EXTRACT(year FROM order_date) AS year,
    ROUND(SUM(sales),2) AS TotalSales,
    ROUND(SUM(profit),2) AS TotalProfit
FROM superstore
GROUP BY year;
```

Data Output | Messages | Notifications

	year numeric	totalsales numeric	totalprofit numeric
1	2016	609205.60	81795.17
2	2017	733215.26	93439.27
3	2015	470532.51	61618.60
4	2014	484247.50	49543.97

The data above shows how the profits over the years have steadily increased with each year being more profitable than the other despite having a fall in sales in 2015, our financial performance

2. **What are the total profits and total sales per quarter?**

```sql
--2.    What are the total profits and total sales per quarter?

SELECT
    EXTRACT(YEAR FROM order_date) AS year,
    CASE
        WHEN EXTRACT(MONTH FROM order_date) IN (1,2,3) THEN 'Quarter-1'
        WHEN EXTRACT(MONTH FROM order_date) IN (4,5,6) THEN 'Quarter-2'
        WHEN EXTRACT(MONTH FROM order_date) IN (7,8,9) THEN 'Quarter-3'
        ELSE 'Quarter-4'
    END AS quarter,
    ROUND(SUM(sales),2) AS total_sales,
    ROUND(SUM(profit),2) AS total_profit
FROM superstore
GROUP BY year, quarter
ORDER BY year, quarter
```

Upon running the above query, we get the below result:

	year numeric	quarter text	total_sales numeric	total_profit numeric
1	2014	Quarter-1	74447.80	3811.23
2	2014	Quarter-2	86538.76	11204.07
3	2014	Quarter-3	143633.21	12804.72
4	2014	Quarter-4	179627.73	21723.95
5	2015	Quarter-1	68851.74	9264.94
6	2015	Quarter-2	89124.19	12190.92
7	2015	Quarter-3	130259.58	16853.62
8	2015	Quarter-4	182297.01	23309.12
9	2016	Quarter-1	93237.18	11441.37
10	2016	Quarter-2	136082.30	16390.34
11	2016	Quarter-3	143787.36	15823.60
12	2016	Quarter-4	236098.75	38139.86
13	2017	Quarter-1	123144.86	23506.20
14	2017	Quarter-2	133764.37	15499.21
15	2017	Quarter-3	196251.96	26985.13
16	2017	Quarter-4	280054.07	27448.73

If we extract the quarterly sum of sales and profits for all the years, we get:

	quarter text	total_sales numeric	total_profit numeric
1	Quarter-1	359681.58	48023.74
2	Quarter-2	445509.62	55284.54
3	Quarter-3	613932.11	72467.08
4	Quarter-4	878077.56	110621.66

The data above shows that the period of October, November and December are our bestselling months and our months where we bring in the most profit. Just by seeing this table, we can develop operation strategies pretty nicely as there is a clear buildup like a stock market rally from January to December then it dumps around the first 3 months.

3. What region generates the highest sales and profits?

We found out the total sales and profits for individual regions by grouping by regions as:

```
55  --3.    What region generates the highest sales and profits ?
56
57  SELECT
58      region,
59      ROUND(SUM(sales),2) AS total_sales,
60      ROUND(SUM(profit),2) AS total_profits
61  FROM superstore
62  GROUP BY region
63  ORDER BY total_profits DESC;
```

Data Output | Messages | Notifications

	region character varying (100)	total_sales numeric	total_profits numeric
1	West	725457.82	108418.45
2	East	678781.24	91522.78
3	South	391721.91	46749.43
4	Central	501239.89	39706.36

The West region emerges as a clear sales and profit leader. The East also contributes significantly to our bottom line. These two regions represent prime opportunities for growth and profit maximization.

While the South generates consistent profits, its revenue is comparatively lower. However, the Central region is a cause for concern. Despite higher revenue than the South, profit margins are unsatisfactory. A strategic reallocation of resources from the Central region to other areas could potentially enhance overall profitability.

We also found out the profit margins for different regions:

```sql
--3.    What region generates the highest sales and profits ?

SELECT
    region,
    ROUND(SUM(sales),2) AS total_sales,
    ROUND(SUM(profit),2) AS total_profits,
    ROUND(SUM(profit)/SUM(sales)*100,2) AS profit_margin_percentage
FROM superstore
GROUP BY region
ORDER BY total_profits DESC;
```

Data Output | Messages | Notifications

	region character varying (100)	total_sales numeric	total_profits numeric	profit_margin_percentage numeric
1	West	725457.82	108418.45	14.94
2	East	678781.24	91522.78	13.48
3	South	391721.91	46749.43	11.93
4	Central	501239.89	39706.36	7.92

Profit margin, a key indicator of profitability, reveals the percentage of revenue retained as profit. The West and East regions exhibit strong performance in this metric. Surprisingly, the South, despite generating roughly half the revenue of the West, boasts an impressive 11.93% profit margin. However, the Central region's performance remains underwhelming. To gain deeper insights, let's delve into the regional data.

4. What state and city bring in the highest sales and profits?

Let’s discover what states are the top 10 highest and lowest and then we will move on to the cities.

```sql
--4.    What state and city brings in the highest sales and profits ?

SELECT
    State,
    SUM(Sales) as Total_Sales,
    SUM(Profit) as Total_Profits,
    ROUND((SUM(profit) / SUM(sales)) * 100, 2) as profit_mar
FROM superstore
GROUP BY State
ORDER BY Total_Profits DESC
LIMIT 10;
```

Data Output | Messages | Notifications

	state character varying (100)	total_sales numeric	total_profits numeric	profit_mar numeric
1	California	457687.63150	76381.38710	16.69
2	New York	310876.27100	74038.54860	23.82
3	Washington	138641.27000	33402.65170	24.09
4	Michigan	76269.61400	24463.18760	32.07
5	Virginia	70636.72000	18597.95040	26.33
6	Indiana	53555.36000	18382.93630	34.33
7	Georgia	49095.84000	16250.04330	33.10
8	Kentucky	36591.75000	11199.69660	30.61
9	Minnesota	29863.15000	10823.18740	36.24
10	Delaware	27451.06900	9977.37480	36.35

In terms of profits, California, New York and Washington are our most profitable markets and most present ones especially in terms of sales. Which, are so high that it would take so much for the profit margins to be higher. However, the profits are great and the total sales show that we have the best part of our business share at those points so we need to boost our resources and customer service in those top states

Let's observe our bottom 10 states:

```
79 SELECT
80     State,
81     SUM(Sales) as Total_Sales,
82     SUM(Profit) as Total_Profits
83 FROM superstore
84 GROUP BY State
85 ORDER BY Total_Profits ASC
86 LIMIT 10;
```

Data Output | Messages | Notifications

	state character varying (100)	total_sales numeric	total_profits numeric
1	Texas	170188.04580	-25729.35630
2	Ohio	78258.13600	-16971.37660
3	Pennsylvania	116511.91400	-15559.96030
4	Illinois	80166.10100	-12607.88700
5	North Carolina	55603.16400	-7490.91220
6	Colorado	32108.11800	-6527.85790
7	Tennessee	30661.87300	-5341.69360
8	Arizona	35282.00100	-3427.92460
9	Florida	89473.70800	-3399.30170
10	Oregon	17431.15000	-1190.47050

The top 3 are Texas, Ohio and Pennsylvania. Texas and Pennsylvania are especially alarming as they have more than 100,000 in sales with Texas having more sales than Washington (which made $33402.70 in profits) but made a loss of $25729.29.

Now let's observe our top 10 cities and bottom 10 cities:

```
88 SELECT
89     City,
90     SUM(Sales) as Total_Sales,
91     SUM(Profit) as Total_Profits,
92     ROUND((SUM(profit) / SUM(sales)) * 100, 2) as profit_marg
93 FROM superstore
94 GROUP BY City
95 ORDER BY Total_Profits DESC
96 LIMIT 10
```

Data Output | Messages | Notifications

	city character varying (100)	total_sales numeric	total_profits numeric	profit_marg numeric
1	New York City	256368.16100	62036.98370	24.20
2	Los Angeles	175851.34100	30440.75790	17.31
3	Seattle	119540.74200	29156.09670	24.39
4	San Francisco	112669.09200	17507.38540	15.54
5	Detroit	42446.94400	13181.79080	31.05
6	Lafayette	25036.20000	10018.38760	40.02
7	Jackson	24963.85800	7581.68280	30.37
8	Atlanta	17197.84000	6993.66290	40.67
9	Minneapolis	16870.54000	6824.58460	40.45
10	San Diego	47521.02900	6377.19600	13.42

```sql
SELECT
    City,
    SUM(Sales) as Total_Sales,
    SUM(Profit) as Total_Profits
FROM superstore
GROUP BY City
ORDER BY Total_Profits ASC
LIMIT 10;
```

Data Output | Messages | Notifications

	city character varying (100)	total_sales numeric	total_profits numeric
1	Philadelphia	109077.01300	-13837.76740
2	Houston	64504.76040	-10153.54850
3	San Antonio	21843.52800	-7299.05020
4	Lancaster	9891.46400	-7239.06840
5	Chicago	48539.54100	-6654.56880
6	Burlington	21668.08200	-3622.87720
7	Dallas	20131.93220	-2846.52570
8	Phoenix	11000.25700	-2790.88320
9	Aurora	11656.47800	-2691.73860
10	Jacksonville	44713.18300	-2323.83500

Our analysis indicates that New York City, Los Angeles, and Seattle are our top three priority markets, while Philadelphia, Houston, and San Antonio rank lowest. The significant presence of two Texas cities in the top tier highlights a strategic opportunity. We must carefully evaluate our current approach in these markets and consider implementing revised strategies.

5. **The relationship between discount and sales and the total discount per category**

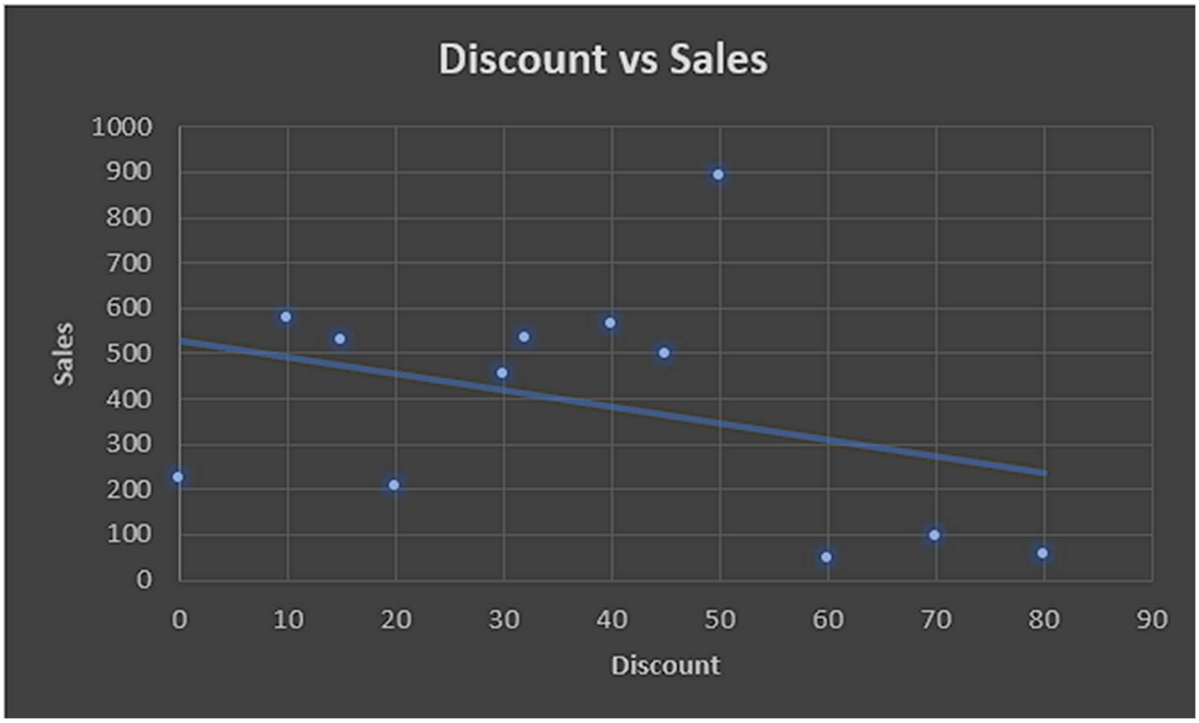
```sql
--5.The relationship between discount and sales and the total discount per category

SELECT
    Discount,
    AVG(Sales) AS Avg_Sales
FROM superstore
GROUP BY Discount
ORDER BY Discount;
```

Data Output | Notifications | Messages

	discount numeric (12,5)	avg_sales numeric
1	0.00000	226.7420737807419758
2	10.00000	578.3973510638297872
3	15.00000	529.9715673076923077
4	20.00000	209.0769395679518731
5	30.00000	454.7429735682819383
6	32.00000	536.7947703703703704
7	40.00000	565.1348737864077670
8	45.00000	498.6340000000000000
9	50.00000	892.7051515151515152
10	60.00000	48.1500000000000000
11	70.00000	97.1777081339712919
12	80.00000	56.5458533333333333

Upon plotting the above data in excel, we found that:



There appears to be a negligible linear correlation between the variables, as indicated by a correlation coefficient of -0.3 and the scatterplot's shape. Intriguingly, we observe a peak in average sales at a 50% discount. This phenomenon could be attributed to psychological factors influencing consumer behavior or the specific product category being discounted.

Let’s observe the total discount per product category.

```
116 SELECT
117     category,
118     ROUND(SUM(discount),2) AS total_discount
119 FROM superstore
120 GROUP BY category
121 ORDER BY total_discount DESC;
```

Data Output | Notifications | Messages

	category character varying (100)	total_discount numeric
1	Office Supplies	94780.00
2	Furniture	36889.00
3	Technology	24440.00

So, office supplies are the most discounted items followed by Furniture and Technology. We will later dive in into how much profit and sales each generate. Before that, let’s zoom in the category section to see exactly what type of products are the most discounted

```sql
SELECT
    category,
    sub_category,
    ROUND(SUM(discount),2) AS total_discount
FROM superstore
GROUP BY category, sub_category
ORDER BY total_discount DESC
```

Data Output | Notifications | Messages

	category character varying (100)	sub_category character varying (100)	total_discount numeric
1	Office Supplies	Binders	56700.00
2	Technology	Phones	13740.00
3	Furniture	Furnishings	13240.00
4	Furniture	Chairs	10500.00
5	Office Supplies	Paper	10260.00
6	Furniture	Tables	8335.00
7	Office Supplies	Appliances	7760.00
8	Office Supplies	Storage	6320.00
9	Technology	Accessories	6080.00
10	Office Supplies	Art	5960.00
11	Furniture	Bookcases	4814.00
12	Technology	Machines	3520.00
13	Office Supplies	Labels	2500.00
14	Office Supplies	Envelopes	2040.00
15	Office Supplies	Fasteners	1780.00
16	Office Supplies	Supplies	1460.00
17	Technology	Copiers	1100.00

Binders, Phones and Furnishings are the most discounted items. But the gap between binders and the others are drastic. We should check the sales and profits for the binders and other items on the list. But first let's move on to the categories per state.

6. **What category generates the highest sales and profits in each region and state?**

First, let's observe the total sales and total profits of each category with their profit margins:

```sql
--6.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 superstore
GROUP BY category
ORDER BY total_profit DESC
```

Data Output | Notifications | Messages

	category character varying (100)	total_sales numeric	total_profit numeric	profit_margin numeric
1	Technology	836154.03	145454.95	17.40
2	Office Supplies	719047.03	122490.80	17.04
3	Furniture	741999.80	18451.27	2.49

Out of the 3, it is clear that Technology and Office Supplies are the best in terms of profits. Plus, they seem like a good investment because of their profit margins. Furniture is still making profits but do not convert well in overall. Let’s observe the highest total sales and total profits per Category in each region

```sql
SELECT
    region,
    category,
    ROUND(SUM(sales),2) AS total_sales,
    ROUND(SUM(profit),2) AS total_profit
FROM superstore
GROUP BY region, category
ORDER BY total_profit DESC
```

Data Output | Notifications | Messages

	region character varying (100)	category character varying (100)	total_sales numeric	total_profit numeric
1	West	Office Supplies	220853.25	52609.85
2	East	Technology	264973.98	47462.04
3	West	Technology	251991.83	44303.65
4	East	Office Supplies	205516.06	41014.58
5	Central	Technology	170416.31	33697.43
6	South	Technology	148771.91	19991.83
7	South	Office Supplies	125651.31	19986.39
8	West	Furniture	252612.74	11504.95
9	Central	Office Supplies	167026.42	8879.98
10	South	Furniture	117298.68	6771.21
11	East	Furniture	208291.20	3046.17
12	Central	Furniture	163797.16	-2871.05

Office Supplies and Technology are the top-performing categories across regions, with the West excelling in both. While the East also thrives in Technology, the Central region faces challenges, particularly with Furniture, where it's the only category operating at a loss.

Now let’s see the highest total sales and total profits per Category in each state

```
    category,
    ROUND(SUM(sales),2) AS total_sales,
    ROUND(SUM(profit),2) AS total_profit
FROM superstore
GROUP BY region, category
ORDER BY total_profit DESC

SELECT
    state,
    category,
    ROUND(SUM(sales),2) AS total_sales,
    ROUND(SUM(profit),2) AS total_profit
FROM superstore
GROUP BY state, category
ORDER BY total_profit DESC;
```

	state character varying (100)	category character varying (100)	total_sales numeric	total_profit numeric
1	New York	Technology	127483.50	42186.79
2	California	Office Supplies	142351.95	37748.38
3	California	Technology	159271.08	29470.04
4	New York	Office Supplies	90020.04	25994.08
5	Washington	Technology	50536.71	15019.34
6	Michigan	Office Supplies	37723.76	15005.33
7	Washington	Office Supplies	40084.41	11189.52
8	Indiana	Technology	26323.25	11000.88
9	Georgia	Office Supplies	26715.81	9800.93
10	California	Furniture	156064.60	9162.97
11	Minnesota	Office Supplies	19406.54	7780.50
12	Virginia	Technology	24145.16	7407.75
13	Washington	Furniture	48020.15	7193.79
14	Delaware	Technology	14562.22	6239.05

The table above shows the most performing categories in each of our states. Technology in New York and Washington and Office Supplies in California. The 3 categories are all around good for our top 3 markets except the furniture category in Washington which is good but not as great as the others.

Let's check the least profitable ones by just changing our 'ORDER BY' clause too ascending (ASC).

```
SELECT
    state,
    category,
    ROUND(SUM(sales),2) AS total_sales,
    ROUND(SUM(profit),2) AS total_profit
FROM superstore
GROUP BY state, category
ORDER BY total_profit ASC;
```

	state character varying (100)	category character varying (100)	total_sales numeric	total_profit numeric
1	Texas	Office Supplies	44490.53	-18584.64
2	Ohio	Technology	35675.99	-12649.94
3	Texas	Furniture	60593.29	-10436.14
4	Illinois	Furniture	28274.52	-9076.29
5	Illinois	Office Supplies	19907.91	-8354.16
6	Pennsylvania	Furniture	39354.93	-7196.72
7	Pennsylvania	Office Supplies	34941.71	-5172.02
8	Ohio	Furniture	24199.15	-4206.32
9	North Carolina	Technology	26083.12	-3583.30
10	North Carolina	Furniture	15155.48	-3486.46

Office supplies in Texas, Technology in Ohio and Furniture in Texas and Illinois are our biggest losses. Let's move on to subcategories.

7. **What subcategory generates the highest sales and profits in each region and state?**

Let's see the highest total sales and total profits per subcategory in each state:

```
--7.What subcategory generates the highest sales and profits in each region and state ?

SELECT
    state,
    sub_category,
    ROUND(SUM(sales),2) AS total_sales,
    ROUND(SUM(profit),2) AS total_profit
FROM superstore
GROUP BY state, sub_category
ORDER BY total_profit DESC
```

	state character varying (100)	sub_category character varying (100)	total_sales numeric	total_profit numeric
1	New York	Machines	43183.50	17320.08
2	New York	Phones	47502.62	13399.19
3	New York	Binders	32456.91	11096.03
4	California	Accessories	37255.01	11095.66
5	Michigan	Binders	22821.97	11079.51
6	California	Binders	28473.17	10002.15
7	Washington	Copiers	20249.83	9442.42
8	Indiana	Copiers	18499.93	8849.97
9	California	Paper	16757.95	7977.54
10	California	Copiers	24559.52	7889.85

Let's see the lowest sales and profits. Still in order for biggest lost in profits

```
--7.What subcategory generates the highest sales and profits in each region and state ?

SELECT
    state,
    sub_category,
    ROUND(SUM(sales),2) AS total_sales,
    ROUND(SUM(profit),2) AS total_profit
FROM superstore
GROUP BY state, sub_category
ORDER BY total_profit ASC
```

	state character varying (100)	sub_category character varying (100)	total_sales numeric	total_profit numeric
1	Texas	Binders	9042.68	-14705.07
2	Ohio	Machines	8978.24	-11770.94
3	Illinois	Binders	4538.55	-7204.32
4	Texas	Appliances	2407.81	-6147.22
5	North Carolina	Machines	12620.66	-5384.81
6	Pennsylvania	Binders	6266.06	-4570.98
7	New York	Tables	13779.02	-4535.64
8	Colorado	Machines	3313.69	-4384.26
9	Illinois	Tables	6550.67	-4309.74
10	North Carolina	Tables	9681.73	-3684.25

New York demonstrates strong performance for Machines, Phones, and Binders. California and Michigan excel in Accessories and Binders, respectively. Texas and Illinois report significant losses in Binders. Ohio presents a challenge for Machine profitability. A comprehensive review of strategies in these regions is warranted.

8. **What are the names of the products that are the most and least profitable to us?**

Let’s find out the most profitable products from our data

```
--8.What are the names of the products that are the most and least profitable to us?


SELECT
    product_name,
    ROUND(SUM(sales),2) AS total_sales,
    ROUND(SUM(profit),2) AS total_profit
FROM superstore
GROUP BY product_name
ORDER BY total_profit DESC
```

	product_name character varying (255)	total_sales numeric	total_profit numeric
1	Canon imageCLASS 2200 Advanced Copier	61599.82	25199.93
2	Fellowes PB500 Electric Punch Plastic Comb Bindi...	27453.38	7753.04
3	Hewlett Packard LaserJet 3310 Copier	18839.69	6983.88
4	Canon PC1060 Personal Laser Copier	11619.83	4570.93
5	HP Designjet T520 Inkjet Large Format Printer - 24"...	18374.90	4094.98
6	Ativa V4110MDD Micro-Cut Shredder	7699.89	3772.95
7	3D Systems Cube Printer, 2nd Generation, Magenta	14299.89	3717.97
8	Plantronics Savi W720 Multi-Device Wireless Head...	9367.29	3696.28
9	Ibico EPK-21 Electric Binding System	15875.92	3345.28
10	Zebra ZM400 Thermal Label Printer	6965.70	3343.54

Also let’s find out the least profit-making products from our data:

```
--8.What are the names of the products that are the most and least profitable to us?


SELECT
    product_name,
    ROUND(SUM(sales),2) AS total_sales,
    ROUND(SUM(profit),2) AS total_profit
FROM superstore
GROUP BY product_name
ORDER BY total_profit ASC
```

	product_name character varying (255)	total_sales numeric	total_profit numeric
1	Cubify CubeX 3D Printer Double Head Print	11099.96	-8879.97
2	Lexmark MX611dhe Monochrome Laser Printer	16829.90	-4589.97
3	Cubify CubeX 3D Printer Triple Head Print	7999.98	-3839.99
4	Chromcraft Bull-Nose Wood Oval Conference Tables & Bases	9917.64	-2876.12
5	Bush Advantage Collection Racetrack Conference Table	9544.73	-1934.40
6	GBC DocuBind P400 Electric Binding System	17965.07	-1878.17
7	Cisco TelePresence System EX90 Videoconferencing Unit	22638.48	-1811.08
8	Martin Yale Chadless Opener Electric Letter Opener	16656.20	-1299.18
9	Balt Solid Wood Round Tables	6518.75	-1201.06
10	BoxOffice By Design Rectangular and Half-Moon Meeting Room ...	1706.25	-1148.44

Copiers, machines, and printers are our primary profit drivers. The Canon image Class 2200 Advanced Copier, Fellowes PB500 Electric Punch Plastic Comb Binding Machine, and Hewlett Packard LaserJet 3310 Copier consistently rank as our top three performers. Maintaining adequate stock levels for these products is crucial to our continued success.

The Cubify CubeX 3D Printers (Double and Triple Head) and the Lexmark MX611dhe Monochrome Laser Printer are our most unprofitable products. This should be considered when adjusting inventory levels.

9. What segment makes the most of our profits and sales?

The below query was used to find out required insights:

```sql
--9.What segment makes the most of our profits and sales ?


SELECT
    segment,
    ROUND(SUM(sales),2) AS total_sales,
    ROUND(SUM(profit),2) AS total_profit
FROM superstore
GROUP BY segment
ORDER BY total_profit DESC
```

	segment character varying (100)	total_sales numeric	total_profit numeric
1	Consumer	1161401.35	134119.21
2	Corporate	706146.37	91979.13
3	Home Office	429653.15	60298.68

The consumer segment brings in the most profit followed by Corporate and then Home office.

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

The above question can be solved by using the below query:

```sql
--10.How many customers do we have (unique customer IDs) in total and how much per region and state?


SELECT
COUNT(DISTINCT customer_id) AS total_customers
FROM superstore;
```

	total_customers bigint
1	793

We've had 793 customers between 2014 and 2017. Regionally, we had the following:

Evaluating region-wise, customer count:

```sql
SELECT
    region,
    COUNT(DISTINCT customer_id) AS total_customers
FROM superstore
GROUP BY region
ORDER BY total_customers DESC;
```

	region character varying (100)	total_customers bigint
1	West	686
2	East	674
3	Central	629
4	South	512

Customer mobility across regions likely explains why the total doesn't equal 793. This suggests potential overlap in the data. Despite this, the West remains our most significant market.

State-wise, here are the numbers:

The states with the greatest number of customer counts are:

```sql
SELECT
    state,
    COUNT(DISTINCT customer_id) AS total_customers
FROM superstore
GROUP BY state
ORDER BY total_customers DESC;
```

	state character varying (100)	total_customers bigint
1	California	577
2	New York	415
3	Texas	370
4	Pennsylvania	257
5	Illinois	237
6	Washington	224
7	Ohio	202
8	Florida	181
9	North Carolina	122
10	Virginia	107

Similarly, the states with least number of customer counts are:

```sql
SELECT
    state,
    COUNT(DISTINCT customer_id) AS total_customers
FROM superstore
GROUP BY state
ORDER BY total_customers ASC;
```

	state character varying (100)	total_customers bigint
1	Wyoming	1
2	North Dakota	2
3	West Virginia	2
4	Maine	3
5	District of Columbia	4
6	South Dakota	5
7	Vermont	5
8	Montana	8
9	Idaho	11
10	Kansas	14

Wyoming, North Dakota and West Virginia are the places where we had the least customers carry on business with us there.

11. Customer rewards program

To lay the groundwork for a future loyalty and rewards program, let's identify our highest-value customers – those who generate the most sales. Prioritizing these customers is crucial as retaining existing clients is generally more cost-effective than acquiring new ones. To gain a comprehensive understanding, we'll also calculate total profits. The following query will help us uncover these insights:

```sql
--11.Customer rewards program

SELECT
    customer_id,
    ROUND(SUM(sales),2) AS total_sales,
    ROUND(SUM(profit),2) AS total_profit
FROM superstore
GROUP BY customer_id
ORDER BY total_sales DESC
LIMIT 15
```

	customer_id character varying (100)	total_sales numeric	total_profit numeric
1	SM-20320	25043.05	-1980.74
2	TC-20980	19052.22	8981.32
3	RB-19360	15117.34	6976.10
4	TA-21385	14595.62	4703.79
5	AB-10105	14473.57	5444.81
6	KL-16645	14175.23	806.86
7	SC-20095	14142.33	5757.41
8	HL-15040	12873.30	5622.43
9	SE-20110	12209.44	2650.68
10	CC-12370	12129.07	2177.05
11	TS-21370	11891.75	2371.71
12	GT-14710	11820.12	2163.43
13	BM-11140	11789.63	-1659.96
14	SV-20365	11470.95	1199.42
15	CJ-12010	11164.97	858.74

For privacy reasons, customer names have been replaced with unique Customer IDs. While Customer ID ‘SM-20320’ is our top spender, they unfortunately aren’t contributing to our profit margins. Nonetheless, their loyalty warrants recognition. Conversely,

Customer ID ‘TC-20980’ is our second biggest spender and our most profitable customer. Retaining and rewarding both these high-value customers is crucial for our business

12. Average shipping time per class and in total

We can find the average shipping time in total by using the following query:

```
--12.Average shipping time per class and in total

SELECT
    ROUND(AVG(ship_date - order_date),1) AS avg_shipping_time
FROM superstore
```

	avg_shipping_time numeric
1	4.0

Also, we can find out the average shipping time for different classes as follows:

```
SELECT
    ship_mode,
    ROUND(AVG(ship_date - order_date),1) AS avg_shipping_time
FROM superstore
GROUP BY ship_mode
ORDER BY avg_shipping_time
```

	ship_mode character varying (100)	avg_shipping_time numeric
1	Same Day	0.0
2	First Class	2.2
3	Second Class	3.2
4	Standard Class	5.0

5. Proposed Structure:

Overall Performance

- **Positive Trend:** Profits and sales have shown consistent growth, with a minor setback in 2015.
- **Seasonal Impact:** Q4 is the most profitable quarter. Focus on inventory, marketing, and customer service for peak performance.

Regional Analysis

- **West Dominance:** The West region is the most profitable, followed by East, South, and Central.
- **Central Region:** Despite high sales, profit margins are lower than the South. Consider resource reallocation to the West.
- **Market Focus:** Prioritize California, New York, and Washington. Reduce presence or exit Texas, Ohio, and Pennsylvania.

Product Performance

- **Profitable Categories:** Technology and Office Supplies offer high profit margins. Furniture needs improvement.
- **Regional Strengths:** Office Supplies in West and East, Technology in West and East, Furniture in West.
- **Product Focus:** Increase Copiers, Phones, Accessories, and Paper. Discontinue Tables, Bookcases, and Supplies.
- **Top Performers:** Canon imageClass 2200, Fellowes PB500, Hewlett Packard LaserJet 3310.
- **Bottom Performers:** Cubify CubeX 3D Printers, Lexmark MX611dhe.

Customer Segmentation

- **Consumer Focus:** Prioritize the consumer segment.
- **Customer Concentration:** California, New York, and Texas have the highest customer base.
- **Texas Challenge:** Despite high customer count, Texas is unprofitable. Re-evaluate market presence.

Recommendations

- **Capitalize on Strengths:** Invest in high-performing regions, products, and customer segments.
- **Address Weaknesses:** Improve profit margins in the Central region, optimize furniture category, and exit unprofitable markets.
- **Data-Driven Decisions:** Continuously analyze performance metrics to inform strategic choices.