Open in app ↗   Get unlimited access

Search Medium

Michel Nguegang
Mar 1 · 20 min read · ▶ Listen

🔖 Save   🐦   f   in   🔗   ⋯

Superstore Sales Analysis Demonstration

Identifying the patterns in our products, regions, categories and customer segments for efficiency and profit optimization with SQL

With the constant advancement of technology and innovation, having clear cut databases and possessing departments to breathe life into interpreting them will set you in the top realm of efficiency. Execution is important but we need great inputs to yield even better outputs. There is a necessity to catch up with the growing demand and fierce competition in the market . That is why we need leverage insights to provide more value to the customer as value is the key to everything.

Our analysis in question will be carried with Excel, SQL and finally Tableau. Excel will serve as first repository for our data, SQL will give meaning to our data and Tableau will give a clear face to our data. The following sales performance analysis will follow the 6 steps of Data Analysis which are: Ask, Prepare, Process, Analyze, Share and Act.

Step 1: Ask

In this step, we will define the business problem given to us which was interpreted as "What are the best products, regions, categories and customer segments for the Superstore to target or avoid in order to increase profitability?"

Business 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 ?

Deliverables:

- A clear summary of the business objectives.

- A full documentation of all the data cleaning, manipulation and analysis.

- A dashboard with visualizations and main outcomes.

- Recommendations based on our insights and analysis.

👏 1 | 💬 | ⋯

Step 2: Prepare

In this phase, we will identify and assess the features of our 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.

Moving on to the data processing, we will use Excel for cleaning

Step 3: Process

We will process and clean our data with the help of Excel as the file is already a CSV file so a look through of our data with Excel can be ideal to:

- Observe our data

- Check for missing data with the help of conditional formating

- Remove duplicate rows

- Correctly format columns for easy SQL analysis

While exploring our dataset, we can perform and notice the following;

- Our data looks correct and consistent. Everything looks well structured for further analysis it just needs a little editing.

- With the use of conditional formating, which is a technique to highlight certain values of interest, we set our new formatting rule to be 'Format only cells that contain'. Then we added on it to be "Format only cells with': Blanks'. Our color of choice was yellow. Then we zoomed out to have a birds eye view of our dataset. We can confirm that our dataset contains no missing values. Another method would be to go to the 'Data' section of Excel and hit filters and filter rows for blanks.

- With the command 'remove duplicates', there was no instance where the data was duplicated with all the exact parameters for a customer in all columns. So all the rows had some variety to it hence returning no duplicate data for our data set.

- Finally, made sure Order Date and Ship Date were well formatted by formating it to dates (which they already were) and formatted the Sales and Profits columns from numbers to currencies since we our talking about financials. The discount column was not formated to currency as in this context, the discount is more of a percentage value. So discount will later be formatted from number to percentage by just multiplying it by 100.

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.

Step 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 SQL and check the first 5 rows to make sure it imported well.

```
SELECT *
FROM superstore
LIMIT 5;
```

rowid integer	orderid character varying (20)	orderdate date	shipdate date	shipmode character varying (20)	customerid character varying (20)	customername character varying (100)
1	1	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute
2	2	CA-2016-152156	2016-11-08	2016-11-11	Second Class	CG-12520	Claire Gute
3	3	CA-2016-138688	2016-06-12	2016-06-16	Second Class	DV-13045	Darrin Van Huff
4	4	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell
5	5	US-2015-108966	2015-10-11	2015-10-18	Standard Class	SO-20335	Sean O'Donnell

The first 5 rows and 7 columns of our dataset

Okay, let's perform an exploratory data analysis with our input on the superstore dataset. A list of tasks will be answered followed by the query input and query result. At the end of our analysis, we will transition to a dashboard reflecting the answers to the most important components that will solve the business problem posed upon us.

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

The years were grouped by order date, so we can observe data for the year 2014, 2015, 2016 and 2017.

```
SELECT DATE_TRUNC('year', orderdate) AS year,
SUM(sales) AS total_sales,
SUM(profit) AS total_profit
FROM superstore
GROUP BY year
ORDER BY year ASC;
```

This query produced the following result:

	year timestamp with time zone 🔒	total_sales 🔒 numeric	total_profit 🔒 numeric
1	2014-01-01 00:00:00-06	484247.56	49544.06
2	2015-01-01 00:00:00-06	470532.46	61618.69
3	2016-01-01 00:00:00-06	609205.86	81795.27
4	2017-01-01 00:00:00-06	733215.19	93439.77

Total sales and Total profits for each year

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?**

This is done to see the periods where our company has been the most impactful. So that in the future, we can tailor our operations where we see fit like maximizing our resources like advertisement, customer service and our overall presence during those times of the year. This is solved with the code below;

```
SELECT
  date_part('year', orderdate) AS year,
  CASE
    WHEN date_part('month', orderdate) IN (1,2,3) THEN 'Q1'
    WHEN date_part('month', orderdate) IN (4,5,6) THEN 'Q2'
    WHEN date_part('month', orderdate) IN (7,8,9) THEN 'Q3'
    ELSE 'Q4'
  END AS quarter,
  SUM(sales) AS total_sales,
  SUM(profit) AS total_profit
FROM superstore
GROUP BY year, quarter
ORDER BY year, quarter;
```

The code above produces the following:

	year double precision 🔒	quarter 🔒 text	total_sales 🔒 numeric	total_profit 🔒 numeric
1	2014	Q1	74447.86	3811.20
2	2014	Q2	86538.77	11204.16
3	2014	Q3	143633.18	12804.73
4	2014	Q4	179627.75	21723.97
5	2015	Q1	68851.74	9264.94
6	2015	Q2	89124.28	12190.92
7	2015	Q3	130259.51	16853.70
8	2015	Q4	182296.93	23309.13
9	2016	Q1	93237.20	11441.39
10	2016	Q2	136082.33	16390.30
11	2016	Q3	143787.43	15823.63
12	2016	Q4	236098.90	38139.95
13	2017	Q1	123144.84	23506.21
14	2017	Q2	133764.33	15499.42
15	2017	Q3	196251.94	26985.26
16	2017	Q4	280054.08	27448.88

The total sales and total profits for each year per quarter

Now this table will aid us in knowing what quarters were the most profitable to us from 2014–2017. This can help to pave the way for investment and marketing strategies. Computing the table above through Excel gives us the following :

Quarters (2014 - 2017)	Total Sales	Total Profit
Q4	$878,077.66	$110,621.93
Q3	$613,932.06	$72,476.32
Q2	$445,509.71	$55,284.80
Q1	$359,681.64	$48,023.74

Most performing quarters from 2014–2017

The data above shows that the period of October, November and December are our best selling 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. Let's get into the regions.

3. What region generates the highest sales and profits ?

This is solved with the code below:

```
SELECT region, SUM(sales) AS total_sales, SUM(profit) AS total_profits
FROM superstore
GROUP BY region
ORDER BY total_profits DESC;
```

	region character varying (20)	total_sales numeric	total_profits numeric
1	West	725457.93	108418.79
2	East	678781.36	91522.84
3	South	391721.90	46749.71
4	Central	501239.88	39706.45

Total Profits and Sales by Region

We can observe above that the West region is the one with the most sales and brings us in the highest profits. The East region is pretty good looking for our company too. Those 2 regions are definitely areas of interest if we want to maximize our profits and expand our business. Concerning the South region, we do not gain a lot of revenue but still the profits are there. It is the Central region that is quite alarming as we generate way more revenue than the South region but do not make at least the same profits over there. The Central region should be on our watchlist as we could start to think on how we could maybe put our resources in the other regions instead. Let's observe each regions profit margins for further analysis with the following code:

```
SELECT region, ROUND((SUM(profit) / SUM(sales)) * 100, 2) as profit_margin
FROM superstore
GROUP BY region
ORDER BY profit_margin DESC
```

	region character varying (20)	profit_margin numeric
1	West	14.94
2	East	13.48
3	South	11.93
4	Central	7.92

Profit margins by region

Profit margins are a measure of a company's profitability and are expressed as the percentage of revenue that the company keeps as profit. So we can see that the West and East are really good. The South region despite almost selling less than half of the West region in revenue has a good profit margin of 11.93% which is great. However the Central region is still not convincing. Let's move on and try to pinpoint the data in each region.

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

States

Firstly, Let's discover what states are the top 10 highest and lowest and then we will move on to the cities. For the states, it can be found with the following code:

```
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;
```

This produces the following results:

	state character varying (20)	total_sales numeric	total_profits numeric	profit_margin numeric
1	California	457687.68	76381.60	16.69
2	New York	310876.20	74038.64	23.82
3	Washington	138641.29	33402.70	24.09
4	Michigan	76269.61	24463.15	32.07
5	Virginia	70636.72	18598.00	26.33
6	Indiana	53555.36	18382.97	34.33
7	Georgia	49095.84	16250.08	33.10
8	Kentucky	36591.75	11199.70	30.61
9	Minnesota	29863.15	10823.22	36.24
10	Delaware	27451.07	9977.37	36.35

Top 10 State's total sales and profits with their profit margins

The decision was to include profit margins to see this under a different lens. The data shows the top 10 most profitable states. Besides we can see the total sales and profit margins. Profit margins are important and it allows us to mostly think long-term as an investor to see potential big markets. 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:

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

	state character varying (20)	total_sales numeric	total_profits numeric
1	Texas	170187.98	-25729.29
2	Ohio	78258.21	-16971.37
3	Pennsylvania	116512.02	-15560.04
4	Illinois	80166.16	-12607.89
5	North Carolina	55603.09	-7490.81
6	Colorado	32108.12	-6527.86
7	Tennessee	30661.92	-5341.66
8	Arizona	35282.02	-3427.87
9	Florida	89473.73	-3399.25
10	Oregon	17431.14	-1190.48

Bottom 10 State's total sales and profits

Our least profitable markets are listed above. 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.

Cities

The top cities are found with the code below:

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

	city character varying (50)	total_sales numeric	total_profits numeric	profit_margin numeric
1	New York City	256368.12	62037.08	24.20
2	Los Angeles	175851.33	30440.94	17.31
3	Seattle	119540.74	29156.13	24.39
4	San Francisco	112669.09	17507.39	15.54
5	Detroit	42446.94	13181.79	31.05
6	Lafayette	25036.20	10018.38	40.02
7	Jackson	24963.85	7581.67	30.37
8	Atlanta	17197.84	6993.69	40.67
9	Minneapolis	16870.54	6824.61	40.45
10	San Diego	47521.05	6377.24	13.42

Top 10 Cities' total sales and profits with their profit margins

The top 3 cities that we should focus on are New York City, Los Angeles and Seattle.

The bottom 10 cities are:

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

	city character varying (50) 🔒	total_sales 🔒 numeric	total_profits 🔒 numeric
1	Philadelphia	109077.09	-13837.83
2	Houston	64504.71	-10153.48
3	San Antonio	21843.54	-7299.06
4	Lancaster	9891.48	-7239.08
5	Chicago	48539.59	-6654.55
6	Burlington	21668.08	-3622.88
7	Dallas	20131.90	-2846.55
8	Phoenix	11000.27	-2790.85
9	Aurora	11656.47	-2691.76
10	Jacksonville	44713.18	-2323.80

Bottom 10 Cities' total sales and profits with their profit margins

The bottom 3 are Philadelphia, Houston and San Antonio. We have 2 cities from Texas in our top 3 so it is clear that we have start redesigning some strategies and how we operate in those cities.

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

First, let's observe the correlation between discount and average sales to understand how impactful one is to the other.
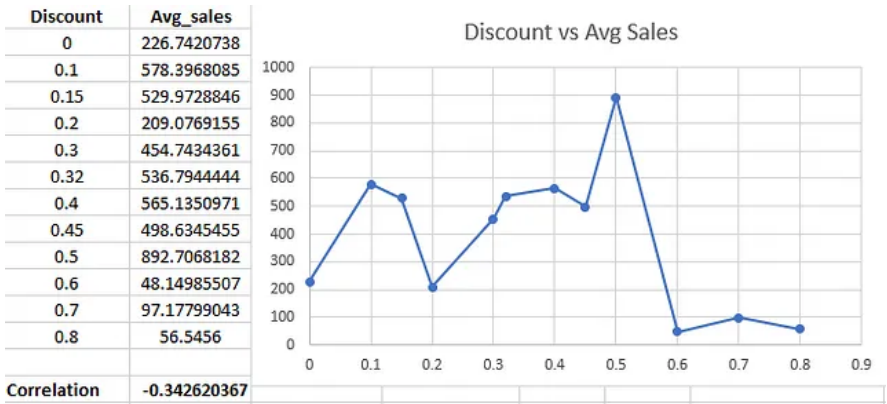
```
SELECT Discount, AVG(Sales) AS Avg_Sales
FROM superstore
GROUP BY Discount
ORDER BY Discount;
```

This produces the following:

	discount numeric (4,2) 🔒	avg_sales numeric 🔒
1	0.00	226.7420737807419758
2	0.10	578.3968085106382979
3	0.15	529.9728846153846154
4	0.20	209.0769155045118950
5	0.30	454.7434361233480176
6	0.32	536.7944444444444444
7	0.40	565.1350970873786408
8	0.45	498.6345454545454545
9	0.50	892.7068181818181818
10	0.60	48.1498550724637681
11	0.70	97.1779904306220096
12	0.80	56.5456000000000000

Discount vs Avg Sales

Seems that for each discount point, the average sales seem to vary a lot. Let's check the correlation with a graph in Excel

Discount	Avg_sales
0	226.7420738
0.1	578.3968085
0.15	529.9728846
0.2	209.0769155
0.3	454.7434361
0.32	536.7944444
0.4	565.1350971
0.45	498.6345455
0.5	892.7068182
0.6	48.14985507
0.7	97.17799043
0.8	56.5456
Correlation	-0.342620367

Discount vs Avg Sales correlation graph

They almost have no linear relationship. This noted by the correlation coefficient of -0.3 and the shape of the graph. However we can at least observe that at a 50% discount, (0.5 * 100 to convert it to percentage) our average sales are the highest it can be. Maybe it is a psychology technique or it's just the right product category that is discounted.

Let's observe the total discount per product category:

```
SELECT category, SUM(discount) AS total_discount
FROM superstore
GROUP BY category
ORDER BY total_discount DESC;
```

	category character varying (50) 🔒	total_discount numeric 🔒
1	Office Supplies	947.80
2	Furniture	368.89
3	Technology	244.40

Most discounted Categories

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.

```
SELECT category, subcategory, SUM(discount) AS total_discount
FROM superstore
GROUP BY category, subcategory
ORDER BY total_discount DESC;
```

	category character varying (50) 🔒	subcategory character varying (50) 🔒	total_discount numeric 🔒
1	Office Supplies	Binders	567.00
2	Technology	Phones	137.40
3	Furniture	Furnishings	132.40
4	Furniture	Chairs	105.00
5	Office Supplies	Paper	102.60
6	Furniture	Tables	83.35
7	Office Supplies	Appliances	77.60
8	Office Supplies	Storage	63.20
9	Technology	Accessories	60.80
10	Office Supplies	Art	59.60
11	Furniture	Bookcases	48.14
12	Technology	Machines	35.20
13	Office Supplies	Labels	25.00
14	Office Supplies	Envelopes	20.40
15	Office Supplies	Fasteners	17.80
16	Office Supplies	Supplies	14.60
17	Technology	Copiers	11.00

Most discounted subcategories (product type)

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:

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

	category character varying (50) 🔒	total_sales numeric 🔒	total_profit numeric 🔒	profit_margin numeric 🔒
1	Technology	836154.10	145455.66	17.40
2	Office Supplies	719046.99	122490.88	17.04
3	Furniture	741999.98	18451.25	2.49

Categories with their total sales, total profits and profit margins

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. Furnitures are 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, SUM(sales) AS total_sales, SUM(profit) AS total_profit
FROM superstore
GROUP BY region, category
ORDER BY total_profit DESC;
```

	region character varying (20)	category character varying (50)	total_sales numeric	total_profit numeric
1	West	Office Supplies	220853.20	52609.89
2	East	Technology	264974.04	47462.12
3	West	Technology	251991.86	44303.98
4	East	Office Supplies	205516.15	41014.52
5	Central	Technology	170416.29	33697.60
6	South	Technology	148771.91	19991.96
7	South	Office Supplies	125651.31	19986.52
8	West	Furniture	252612.87	11504.92
9	Central	Office Supplies	167026.33	8879.95
10	South	Furniture	117298.68	6771.23
11	East	Furniture	208291.17	3046.20
12	Central	Furniture	163797.26	-2871.10

Highest total sales and profits per Category in each region

These our are best categories in terms of total profits in each region. The West is in our top 3 two times with Office Supplies and Technology and the East with Technology. Among the total profits, the only one that fails to break even is the Central Region with Furniture where we operate at a loss when selling it there.

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

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

	state character varying (20)	category character varying (50)	total_sales numeric	total_profit numeric
1	New York	Technology	127483.48	42186.89
2	California	Office Supplies	142351.89	37748.46
3	California	Technology	159271.12	29470.23
4	New York	Office Supplies	90020.03	25994.03
5	Washington	Technology	50536.71	15019.42
6	Michigan	Office Supplies	37723.76	15005.31
7	Washington	Office Supplies	40084.42	11189.48
8	Indiana	Technology	26323.25	11000.88
9	Georgia	Office Supplies	26715.81	9800.91
10	California	Furniture	156064.67	9162.91
11	Minnesota	Office Supplies	19406.54	7780.51
12	Virginia	Technology	24145.16	7407.75
13	Washington	Furniture	48020.16	7193.80
14	Delaware	Technology	14562.22	6239.04
15	Virginia	Office Supplies	21169.61	5985.91

Top Highest total sales and profits per Category in each state

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) :

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

	state character varying (20)	category character varying (50)	total_sales numeric	total_profit numeric
1	Texas	Office Supplies	44490.43	-18584.66
2	Ohio	Technology	35676.04	-12649.93
3	Texas	Furniture	60593.34	-10436.17
4	Illinois	Furniture	28274.57	-9076.32
5	Illinois	Office Supplies	19907.92	-8354.17
6	Pennsylvania	Furniture	39354.94	-7196.70
7	Pennsylvania	Office Supplies	34941.78	-5172.11
8	Ohio	Furniture	24199.13	-4206.38
9	North Carolina	Technology	26083.10	-3583.28
10	North Carolina	Furniture	15155.47	-3486.44
11	Colorado	Technology	10966.31	-3471.58
12	Tennessee	Office Supplies	12347.88	-3199.11
13	Pennsylvania	Technology	42215.30	-3191.23
14	Arizona	Furniture	13525.33	-2744.91
15	Colorado	Furniture	13243.05	-2683.15

Top Lowest total sales and profits per Category in each state

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 observe the total sales and total profits of each subcategory with their profit margins:

```sql
SELECT subcategory, SUM(sales) AS total_sales, SUM(profit) AS total_profit, ROUND(SUM(profit)/SUM(sales)*100, 2) AS profit_marg
FROM superstore
GROUP BY subcategory
ORDER BY total_profit DESC;
```

	subcategory character varying (50)	total_sales numeric	total_profit numeric	profit_margin numeric
1	Copiers	149528.01	55617.90	37.20
2	Phones	330007.10	44516.25	13.49
3	Accessories	167380.31	41936.78	25.05
4	Paper	78479.24	34053.34	43.39
5	Binders	203412.77	30221.64	14.86
6	Chairs	328449.13	26590.15	8.10
7	Storage	223843.59	21279.05	9.51
8	Appliances	107532.14	18138.07	16.87
9	Furnishings	91705.12	13059.25	14.24
10	Envelopes	16476.38	6964.10	42.27
11	Art	27118.80	6527.96	24.07
12	Labels	12486.30	5546.18	44.42
13	Machines	189238.68	3384.73	1.79
14	Fasteners	3024.25	949.53	31.40
15	Supplies	46673.52	-1188.99	-2.55
16	Bookcases	114880.05	-3472.56	-3.02
17	Tables	206965.68	-17725.59	-8.56

Subcategories with their total sales, total profits and profit margins

Out of our 17 subcategories nationwide, our biggest profits comes from Copiers, Phones, Accessories and Paper. The profits and profit margins on Copiers and Papers especially are interesting for the long run. Our losses came from Tables, Bookcases and Supplies where we are uncapable of breaking even. Those 3 should be further reviewed as the sales are there, (except Supplies) but we cannot generate profits from them.

Now let's see the highest total sales and total profits per subcategory in each region:

```sql
SELECT region, subcategory, SUM(sales) AS total_sales, SUM(profit) AS total_profit
FROM superstore
GROUP BY region, subcategory
ORDER BY total_profit DESC;
```

	region character varying (20)	subcategory character varying (50)	total_sales numeric	total_profit numeric
1	West	Copiers	49749.23	19327.25
2	East	Copiers	53219.46	17022.88
3	West	Accessories	61114.10	16484.62
4	West	Binders	55961.11	16096.78
5	Central	Copiers	37259.57	15608.86
6	Central	Phones	72403.26	12323.14
7	East	Phones	100615.02	12314.70
8	West	Paper	26663.73	12119.10
9	East	Binders	53498.06	11267.92
10	East	Accessories	45033.38	11195.89
11	South	Phones	58304.43	10767.35
12	East	Chairs	96260.65	9357.77
13	West	Phones	98684.39	9111.06
14	East	Paper	20172.62	9015.29
15	West	Storage	70532.84	8645.49

Top 15 Subcategories with the highest total sales and total profits in each region

These above display the best subcategories per region.

Now let's see the least performing ones:

```
SELECT region, subcategory, SUM(sales) AS total_sales, SUM(profit) AS total_profit
FROM superstore
GROUP BY region, subcategory
ORDER BY total_profit ASC;
```

	region character varying (20)	subcategory character varying (50)	total_sales numeric	total_profit numeric
1	East	Tables	39139.82	-11025.39
2	South	Tables	43916.19	-4623.06
3	Central	Furnishings	15254.35	-3906.18
4	Central	Tables	39155.07	-3559.68
5	Central	Appliances	23581.98	-2638.61
6	Central	Bookcases	24157.16	-1997.92
7	West	Bookcases	36004.18	-1646.50
8	Central	Machines	26797.38	-1486.07
9	South	Machines	53890.98	-1438.90
10	East	Bookcases	43819.34	-1167.65
11	East	Supplies	10760.10	-1155.11
12	Central	Binders	56923.26	-1043.71
13	Central	Supplies	9467.37	-661.88
14	West	Machines	42444.14	-618.95
15	South	Supplies	8318.93	1.89

Top 15 Subcategories with the lowest total sales and total profits in each region

We are unable to break-even with 14 subcategories. Tables and Furnishings are our biggest losses in profits in the East, South and Central region.

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

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

	state character varying (20)	subcategory character varying (50)	total_sales numeric	total_profit numeric
1	New York	Machines	43183.50	17320.09
2	New York	Phones	47502.62	13399.26
3	New York	Binders	32456.90	11096.01
4	California	Accessories	37255.01	11095.68
5	Michigan	Binders	22821.97	11079.48
6	California	Binders	28473.11	10002.16
7	Washington	Copiers	20249.83	9442.43
8	Indiana	Copiers	18499.93	8849.97
9	California	Paper	16757.95	7977.42
10	California	Copiers	24559.51	7889.85
11	California	Storage	45112.29	7204.27
12	New York	Chairs	46634.20	7111.03
13	California	Appliances	24175.92	6992.27
14	California	Phones	67964.57	6309.30
15	Washington	Binders	18156.14	6061.81

Top 15 Highest total sales and profits per Subcategory in each state

Machines, Phones and Binders perform very well in New York. Followed by Accessories and Binders in California and Michigan respectively.

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

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

	state character varying (20)	subcategory character varying (50)	total_sales numeric	total_profit numeric
1	Texas	Binders	9042.64	-14705.12
2	Ohio	Machines	8978.25	-11770.95
3	Illinois	Binders	4538.56	-7204.34
4	Texas	Appliances	2407.75	-6147.24
5	North Carolina	Machines	12620.66	-5384.81
6	Pennsylvania	Binders	6266.12	-4571.03
7	New York	Tables	13779.03	-4535.65
8	Colorado	Machines	3313.69	-4384.26
9	Illinois	Tables	6550.72	-4309.76
10	North Carolina	Tables	9681.72	-3684.25
11	Tennessee	Binders	5148.05	-3635.93
12	Pennsylvania	Phones	19702.41	-3606.96
13	Texas	Furnishings	3766.71	-3312.67
14	Pennsylvania	Bookcases	5230.76	-2896.77
15	Ohio	Phones	14634.98	-2778.87

Top 15 Highest total sales and profits per Subcategory in each state

Binders are our biggest losses in Texas and Illnois. Machines are not profitable in Ohio at all. We should observe and rethink our strategies in those areas.

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

Let's verify this information:

```
SELECT productname, SUM(sales) AS total_sales, SUM(profit) AS total_profit
FROM superstore
GROUP BY productname
ORDER BY total_profit DESC;
```

	productname character varying (150)	total_sales numeric	total_profit numeric
1	Canon imageCLASS 2200 Advanced Copier	61599.83	25199.94
2	Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind	27453.38	7753.06
3	Hewlett Packard LaserJet 3310 Copier	18839.68	6983.89
4	Canon PC1060 Personal Laser Copier	11619.83	4570.94
5	HP Designjet T520 Inkjet Large Format Printer - 24" Color	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 Headset System	9367.29	3696.28
9	Ibico EPK-21 Electric Binding System	15875.92	3345.29
10	Zebra ZM400 Thermal Label Printer	6965.70	3343.53
11	Honeywell Enviracaire Portable HEPA Air Cleaner for 17' x 22' Room	11304.44	3247.02
12	Hewlett Packard 610 Color Digital Copier / Printer	8899.82	3124.94
13	Plantronics CS510 - Over-the-Head monaural Wireless Headset System	10822.36	3085.04
14	Canon Imageclass D680 Copier / Fax	8959.87	2799.97
15	Fellowes PB300 Plastic Comb Binding Machine	8070.20	2518.06

Top 15 most profitable products

These Copiers, Machines and Printers are definetly the main foundations of our profits. The Canon imageClass 2200 Advanced Copier, Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind and the Hewlett Packard LaserJet 3310 Copier are our top 3. We should keep up the stock with these. Let's verify our less proftable ones:

```sql
SELECT productname, SUM(sales) AS total_sales, SUM(profit) AS total_profit
FROM superstore
GROUP BY productname
ORDER BY total_profit ASC;
```

	productname character varying (150)	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	-2876.11
5	Bush Advantage Collection Racetrack Conference Table	9544.72	-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	16656.21	-1299.19
9	Balt Solid Wood Round Tables	6518.76	-1201.06
10	BoxOffice By Design Rectangular and Half-Moon Meeting Room Tables	1706.26	-1148.44
11	Riverside Furniture Oval Coffee Table, Oval End Table, End Table with Drawer	4446.18	-1147.41
12	Epson TM-T88V Direct Thermal Printer - Monochrome - Desktop	1212.71	-1057.23
13	Hon 2090 "Pillow Soft" Series Mid Back Swivel/Tilt Chairs	5282.42	-989.04
14	O'Sullivan 4-Shelf Bookcase in Odessa Pine	2740.20	-975.12
15	Bretford "Just In Time" Height-Adjustable Multi-Task Work Tables	5634.90	-964.20

Top 15 less profitable products

The Cubify CubeX 3D Printer Double Head Print, Lexmark MX611dhe Monochrome Laser Printer and the Cubify CubeX 3D Printer Triple Head Print are the products that operate the most at a loss. We should take this into account if we are thinking about modifying our stock.

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

This can be verified with the help of the following query:

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

	segment character varying (20)	total_sales numeric	total_profit numeric
1	Consumer	1161401.34	134119.33
2	Corporate	706146.44	91979.45
3	Home Office	429653.29	60299.01

Goods Segment ordered by total profits

The consumer segment brings in the most profit followed by Corporate and then Home office. Let's move on.

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

This can be solved with the following;

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

	total_customers bigint 🔒
1	793

Total number of customers

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

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

	region character varying (20) 🔒	total_customers bigint 🔒
1	West	686
2	East	674
3	Central	629
4	South	512

Total customers per region

We surely had customers moving around regions which explains why they all do not add up to 793. Since there could be double counting. The West is the area where we have the biggest market of all. Statewise, here are the numbers:

	state character varying (20) 🔒	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
11	Michigan	106
12	Arizona	100
13	Tennessee	84
14	Georgia	83
15	Colorado	75

Top 15 states with the most customers

We have the most customers in California, New York and Texas. The areas where we have the least that passed by there are:

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

	state character varying (20)	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
11	New Hampshire	17
12	Iowa	17
13	South Carolina	19
14	Louisiana	21
15	New Mexico	22

Top 15 states with the least customers

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

Let's say we want to build a loyalty and rewards program in the future. What customers spent the most with us? That is generated the most sales. It is always important to cater for our best customers and see how we can provide more value to them as it its cheaper to keep a current customer than it is to acquire a new one. We will also check the total profits just fo further analysis. We can find out what we are looking for with the following query:

```
SELECT customerid,
SUM(sales) AS total_sales,
SUM(profit) AS total_profit
FROM superstore
GROUP BY customerid
ORDER BY total_sales DESC
LIMIT 15;
```

	customerid character varying (20)	total_sales numeric	total_profit numeric
1	SM-20320	25043.07	-1980.75
2	TC-20980	19052.22	8981.32
3	RB-19360	15117.35	6976.09
4	TA-21385	14595.62	4703.80
5	AB-10105	14473.57	5444.81
6	KL-16645	14175.23	806.84
7	SC-20095	14142.34	5757.42
8	HL-15040	12873.30	5622.43
9	SE-20110	12209.44	2650.67
10	CC-12370	12129.08	2177.05
11	TS-21370	11891.75	2371.73
12	GT-14710	11820.12	2163.42
13	BM-11140	11789.64	-1659.97
14	SV-20365	11470.94	1199.45
15	CJ-12010	11164.97	858.75

Top 15 customers that generated the most sales compared to total profits.

The display of the customer names are on file but showing the unique Customer id is a form of pseudonymization for security reasons. What is actually interesting to see is that customer ID 'SM-20320' is the customer who spent the most with us but is not bringing us profit. We still have to reward his/her loyalty. It is customer ID 'TC-20980' who is second in the list but brings us the most profit. So we really have to thank our top customers and keep them on deck.

12. Average shipping time per class and in total

Finally, the average shipping time regardless of the shipping mode that is chosen is found with the following function:

```
SELECT ROUND(AVG(shipdate - orderdate),1) AS avg_shipping_time
FROM superstore
```

	avg_shipping_time numeric
1	4.0

Average shipping time

The shipping time in each shipping mode is:

```
SELECT shipmode,ROUND(AVG(shipdate - orderdate),1) AS avg_shipping_time
FROM superstore
GROUP BY shipmode
ORDER BY avg_shipping_time
```

shipmode character varying (20) 🔒	avg_shipping_time numeric 🔒
1	Same Day	0.0
2	First Class	2.2
3	Second Class	3.2
4	Standard Class	5.0

Average shipping time by shipping mode

Here we have all the information we need to transition to our Tableau dashboard. The data visualization is our following step.

Step 5: Share

Onto sharing our observations, below you will find a picture of the interactive dashboard that represents the main KPIs and information on the collected Superstore data which was realized and demonstrated with Tableau.

The link to the interactive dashboard will be found under the screenshot below. Here are our findings:
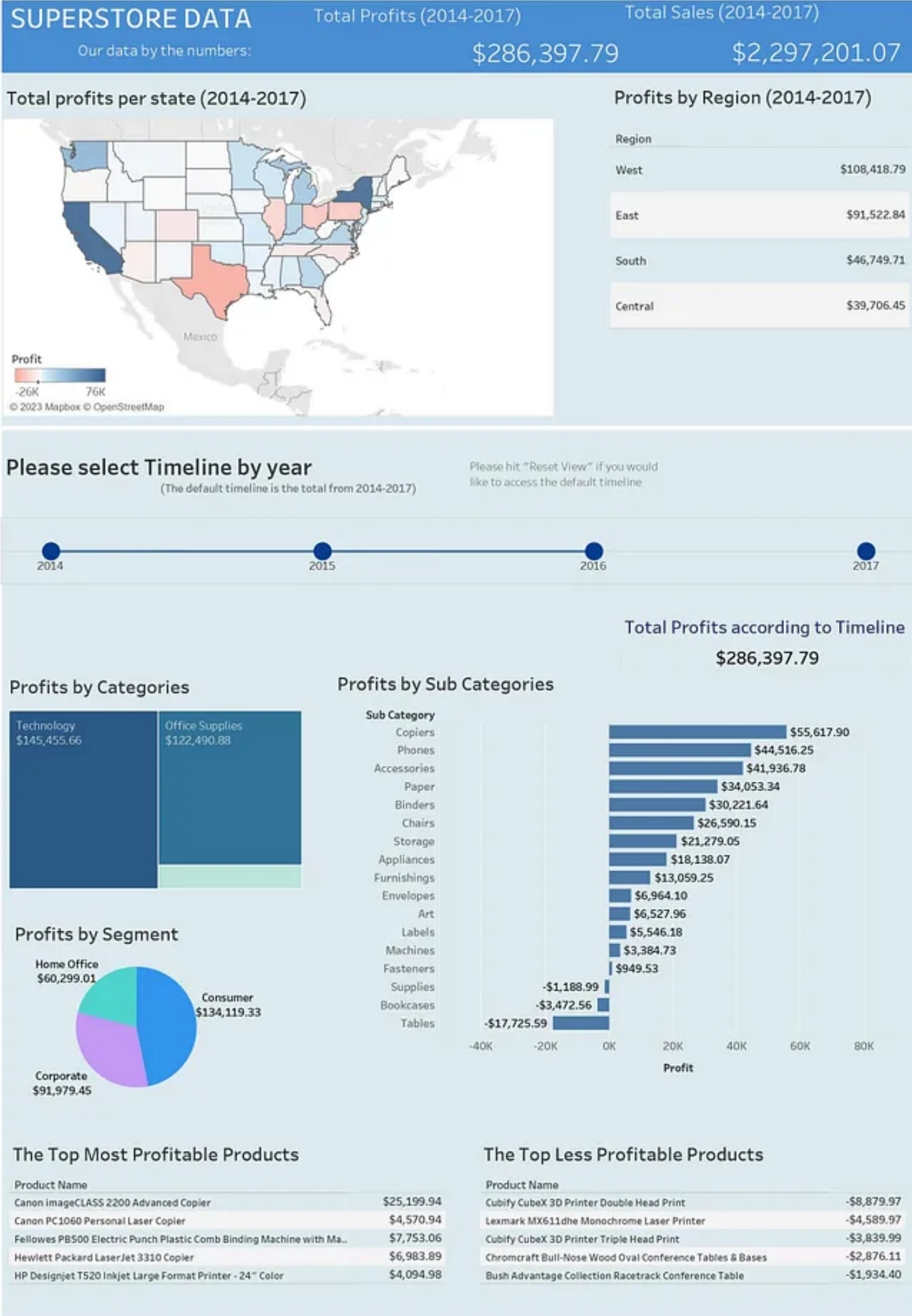
Superstore Interactive Dashboard Screenshot

This is an in depth analysis of the Superstore dataset. You will find 2 interactive dashboards focused on the yearly sales aspect and another one that is more quarterly sales oriented with both dashboards displaying data collected from 2014–2017.

If there is mainly an interest in the yearly (timeline) data on profits, categories, sub categories, segments and products, then below you will find the direct link to the dashboard:

https://10az.online.tableau.com/t/micheltableau/views/SalesDataAnalysis-YearlyFocus2/SalesDataAnalysis-YearlyFocusDashboard?:origin=card_share_link&:embed=n

If it is instead the quarterly (timeline) data that interests you, then the results are in this dashboard instead:

https://10az.online.tableau.com/t/micheltableau/views/SalesDataAnalysis-QuarterlyFocus/SalesDataAnalysis-QuartelyFocusDashboard?:origin=card_share_link&:embed=n

The 2 dashboards and its contents are in the main source file below:

https://10az.online.tableau.com/#/site/micheltableau/projects/338892?:origin=card_share_link

Finally, let's establish suggestions for an effective plan of action to solve our business objectives.

Step 6: Act

With everything that was covered, here are our conclusions and future recommendations for the success of our Superstore:

- Our profits got progressively better. Our sales too even with a short halt in 2015. We should keep the pace up on that aspect.

- Our most profitable quarter all year round was Q4. To maximize even more profits, we must make sure to have enough stock and push our marketing and customer service to make the most out of the October — December festive period.

- The most performing regions are the West then the East, South and Central regions in that order. The Central region brings in at least $100,000 more in sales than the South region but still makes less profits than it. There is work to be done in the Central region if we really want to keep that market. However, I believe it is better to take some of the resources in our Central region to instead our West region stores as we are more profitable there and could really establish ourselves as kingpin in that region.

- California, New York and Washington are our most profitable markets and most present ones especially in terms of sales as states. We have to focus more on them. Our least profitable markets are Texas, Ohio, and Pennsylvania. Which I believe that we should decrease our presence there or even put a halt at our store locations there as sales in Texas and Pennsylvania are in the $100,000s but are unable to convert to profits.

- New York City, Los Angeles and Seattle are our most profitable cities and we list them as being a top priority because it is easier to rule a city than ruling a state. If we gain the city, gaining the state will be less challenging. Philadelphia, Houston and San Antonio are the cities where we lose the most money. We have 2 cities from Texas in our top 3 cities so it is clear that we have start rethinking about really wanting to carry business there, the better option would be to stop.

- Out of the 3 categories, Technology and Office Supplies are the best in terms of profits. Plus they seem like a good investment because of their profit margins. Furnitures are still making profits but do not convert well in overall. With low profits and low profit margins, we should start to see what more we can bring to the furniture department. The sales are there but they do not translate smoothly.

- Still under categories but regionally, Office supplies in the West brings the most profit so we must increase the cap of those materials over there. Same thing with the East and office supplies and both the East and West with Technology. However, furniture in the Central region is the only category that doesn't convert to profits so it would be better to take some of these resources to the West region which is the biggest gainer in terms of Furniture.

- Statewise, Technology and Office supplies brings us the most profit in the state of New York and California. We have to increase the availability of these goods in these states for better profits. However, Office supplies in Texas, Technology in Ohio and Furniture in Texas and Illinois are our biggest losses so we have to drastically reduce these type of products in those areas.

- Out of our 17 subcategories nationwide, our biggest profits comes from Copiers, Phones, Accessories and Paper. The profits and profit margins on Copiers and Papers especially are interesting for the long run. We should immediately push these products as we have a great market share with these items. Our losses came from Tables, Bookcases and Supplies where we are uncapable of breaking even. We must spend less time and money with them. Especially with tables because compared to our only 3 losses, Tables lost us $17725 which is huge compared to our other losses of $3472 and $1188 which came from Bookcases and Supplies respectively.

- For subcategories regionally, Copiers in the West and East with Accessories and Binders in the West are products that we always have to have in stock and promote for more profit. While Tables in the East, South and Central with furnishings in the Central region are the top products where we lose money so we should direct our attention elsewhere.

- In what concerns subcategories by state, Machines, Phones and Binders perform very well in New York. Followed by Accessories and Binders in California and Michigan respectively so there is a need to accentuate our business there with those products. For our biggest losses, Binders in Texas and Illinois with machines in Ohio are not profitable at all. We have to decrease stock in those places.

- For the particular products, The Canon imageClass 2200 Advanced Copier, Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind and the Hewlett Packard LaserJet 3310 Copier are our top 3 in profits. We must always keep up the stock with these. For our losses, The Cubify CubeX 3D Printer Double Head Print, Lexmark MX611dhe Monochrome Laser Printer and the Cubify CubeX 3D Printer Triple Head Print are the products that operate the most at a loss. We should certainly discontinue those products.

- Out of the 3 segments, The consumer segment brings in the most profit followed by Corporate and then Home office. We must give more importance to the consumer segment even if all the 3 are profitable.

- Finally, for our clientele, we have 793 customers total, and we have the most customers in California, New York and Texas. The case of Texas is pretty ironic since it is also the state that losses us the most money. So we must take a critical decision about Texas

first as we absolutely can't break through now. California and New York are pretty obvious, we have to be outstanding and be the best of what there is to offer in our respective niche.

Thank you for your time and commitment to the lecture!

Superstore   Sql   Excel   Tableau   Data Analysis