**Superstore Sales Analysis**

**Project objective:**

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 to increase profitability?”

**Business objectives:**

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

**Deliverables:**

* A clear summary of the business objectives.
* A complete documentation of all the data cleaning, manipulation, and analysis.
* A dashboard with visualizations and primary outcomes.
* Recommendations based on our insights and analysis.

**Data description:**

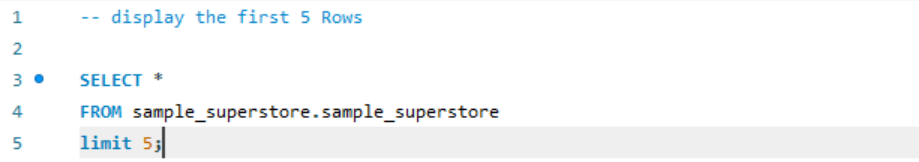
* We will identify and assess the features of our Superstore Dataset:
* It comes with 9995 rows, with 9994 being pure data and the other row being the column headers. It contains data recorded between the 3rd of January 2014 (the first order date) and the 5th of January 2018 (the last shipping date). (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 data on 793 customers.
* The data contains 21 columns.

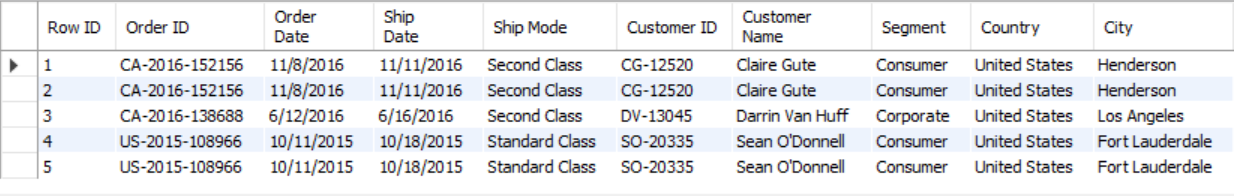
**Schemas:**

|  |
| --- |
| Column Name Type   * **Row ID**: Integer * **Order ID**: String * **Order Date**: Date (MM/DD/YYYY) * **Ship Date**: Date (MM/DD/YYYY) * **Ship Mode**: String * **Customer ID**: String * **Customer Name**: String * **Segment**: String * **Country**: String * **City**: String * **State**: String * **Postal Code**: String * **Region**: String * **Product ID**: String * **Category**: String * **Sub-Category**: String * **Product Name**: String * **Sales**: Float * **Quantity**: Integer * **Discount**: Float * **Profit**: Float |

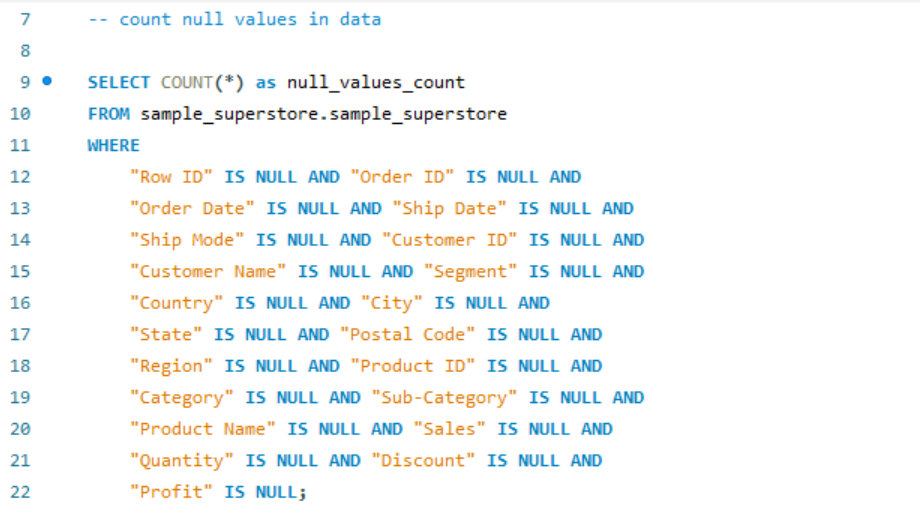
**Analyze:**

* **Let’s load our data into SQL and check the first 5 rows to ensure it is imported well.**

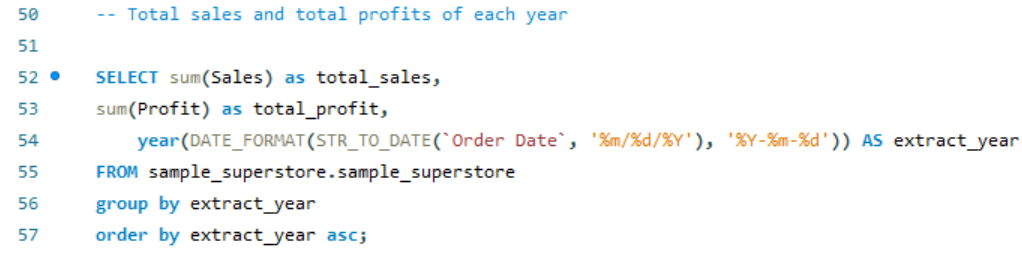


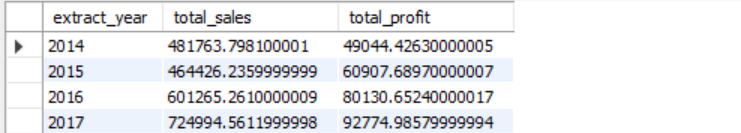


* **Count null values in the dataset?**

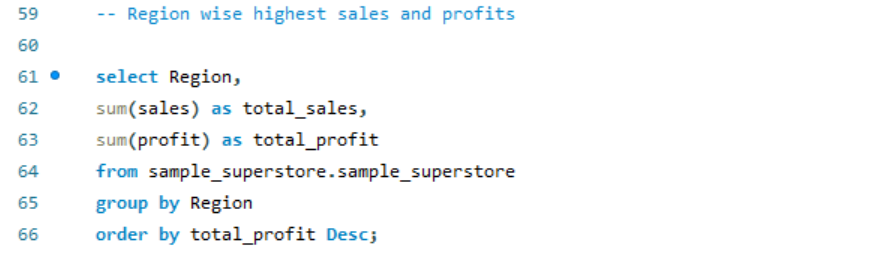


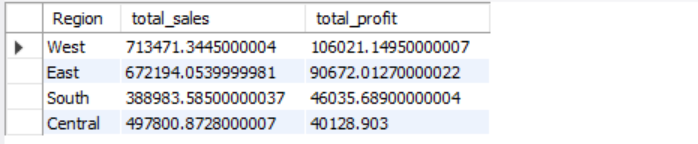
* **What are the total sales and total profits of each year?**
* The years were grouped by order date, so we can observe data for the years 2014, 2015, 2016, and 2017.



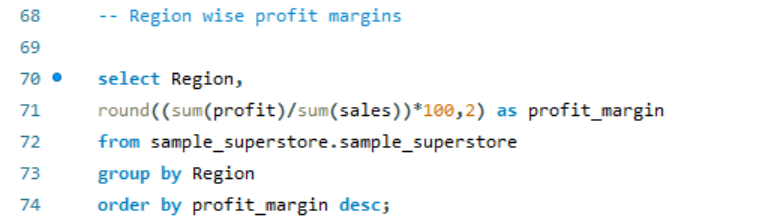
****

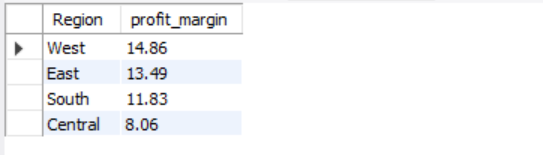
* The data above shows how the profits over the years have steadily increased, with each year being more profitable than the other despite having a fall in sales in 2015, our financial performance
* **What region generates the highest sales and profits?**

****

****

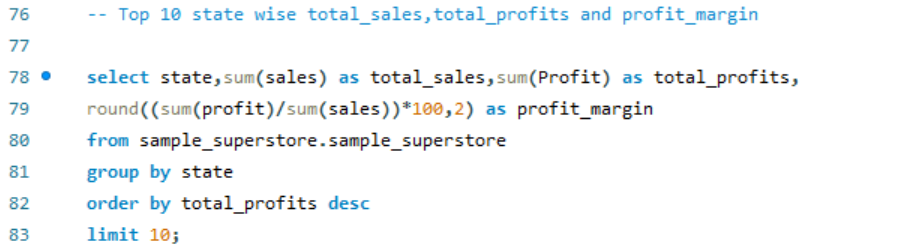
* We can observe above that the West region has the most sales and brings us the highest profits. The East region is pretty good looking for our company too. Those 2 regions are areas of interest if we want to maximize our profits and expand our business. We have gained a little revenue in the South region, but the profits remain. The Central region is alarming as we generate way more income than the South region but make different profits. The Central region should be on our watchlist as we could start to think about how we could put our resources in the other regions instead.
* **Let’s observe each regions profit margins for further analysis with the following code:**

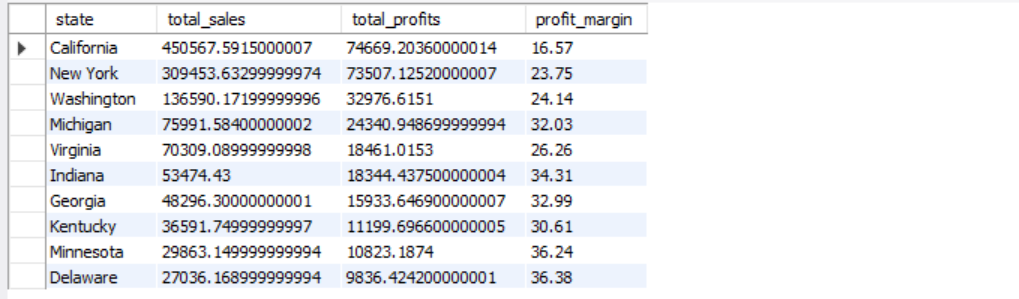


****

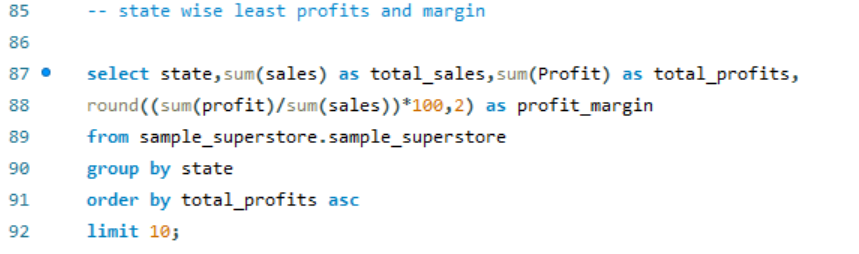
* Profit margins are a measure of a company’s profitability and are expressed as the percentage of revenue that the company keeps as profit. So, we can see that the West and East are good. The South region despite almost selling less than half of the West region in revenue has a good profit margin of 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.
* **What state and city brings in the highest sales and profits?**

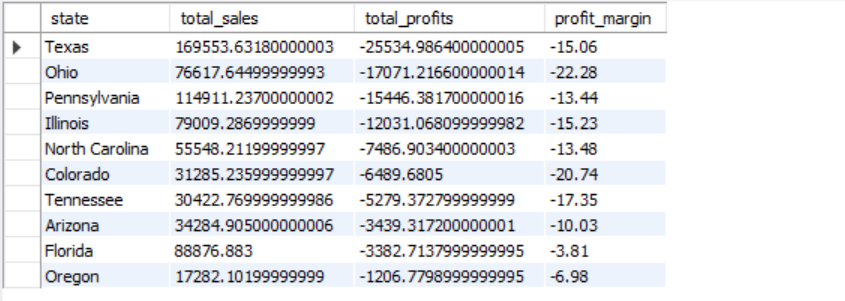
**First, let’s discover which states have the top 10 highest and lowest ratings and then move on to the cities. For the states, it can be found with the following code:**



****

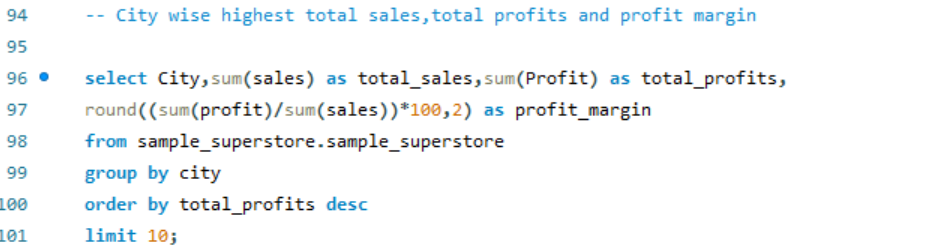
* The decision was to include profit margins to see this through a different lens. The data shows the top 10 most profitable states. We can also see the total sales and profit margins. Profit margins are significant, allowing us to think long-term as investors to see potential big markets. In terms of profits, California, New York, and Washington are our most profitable and present markets, especially in sales, which are so high that it would take so much for the profit margins to be higher. However, the profits are significant, 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.

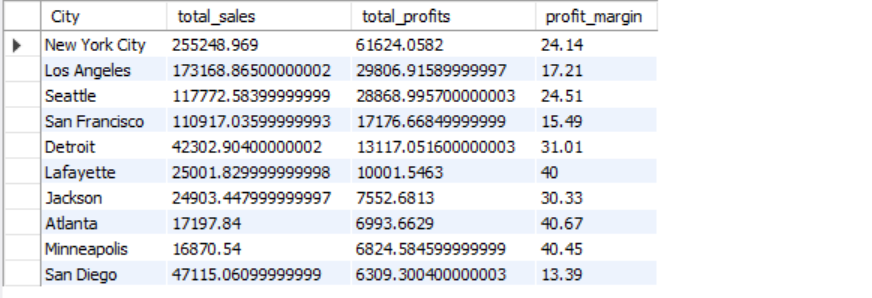




* Our least profitable markets are listed above. The top 3 are Texas, Ohio, and Pennsylvania. Texas and Pennsylvania are especially alarming as they have more than 100,000 in sales, with Texas having more sales than Washington (which made $33402.70 in profits) but a loss of $25729.29.

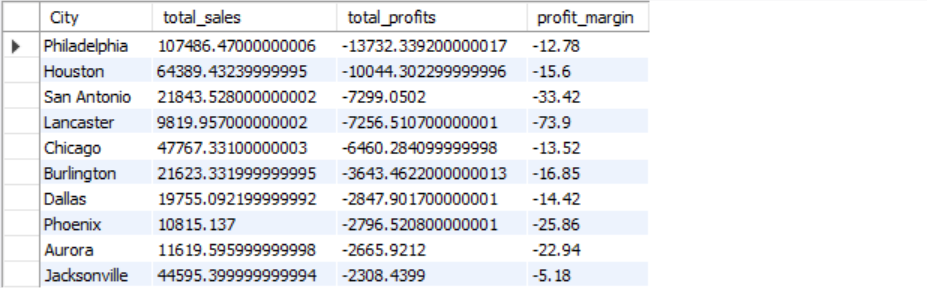
**The top cities are found with the code below:**





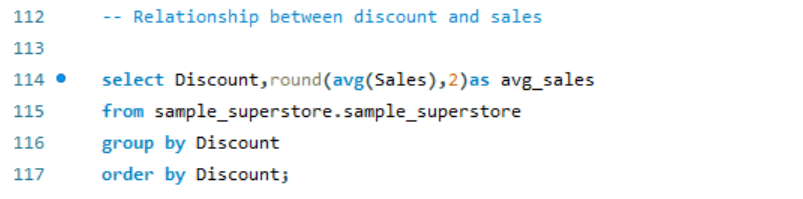
* The top 3 cities that we should focus on are New York City, Los Angeles, and Seattle.

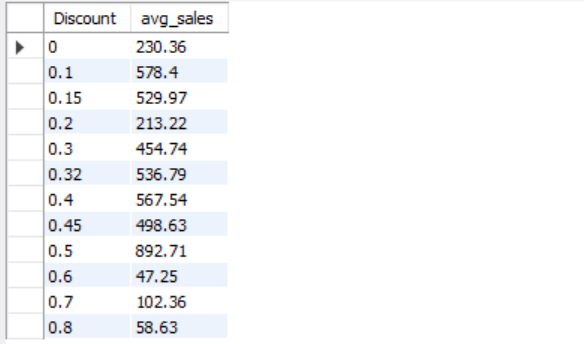




* The bottom 3 are Philadelphia, Houston, and San Antonio. We have 2 cities from Texas in our top 3, so we have started redesigning some strategies and how we operate in those cities.
* **The relationship between discount and sales and the total discount per category**

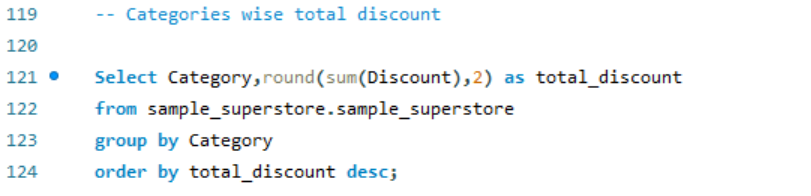
**First, let’s observe the correlation between discounts and average sales to understand how impactful one is to the other.**

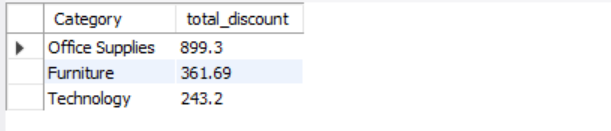




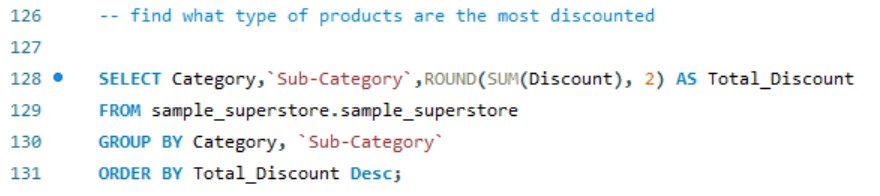
* For each discount point, the average sales vary a lot.

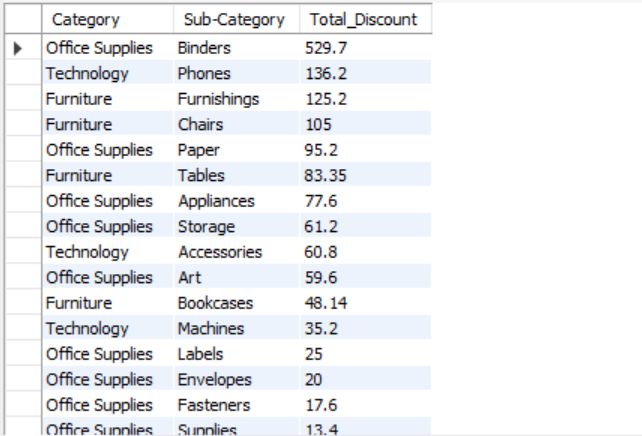
**Let’s observe the total discount per product category:**

****

****

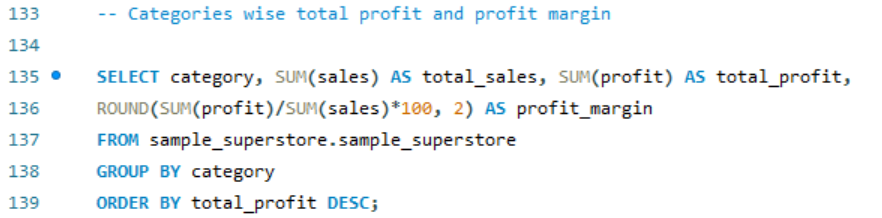
* Office supplies are the most discounted items, followed by furniture and technology. We will later dive into how much profit and sales each generates. Before that, let’s zoom in on the category section to see exactly what type of products are the most discounted.
* **find what type of products are the most discounted**

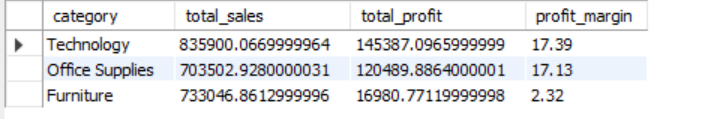


****

* Binders, Phones, and Furnishings are the most discounted items. But the gap between binders and the others is 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.
* **What category generates the highest sales and profits in each region and state?**

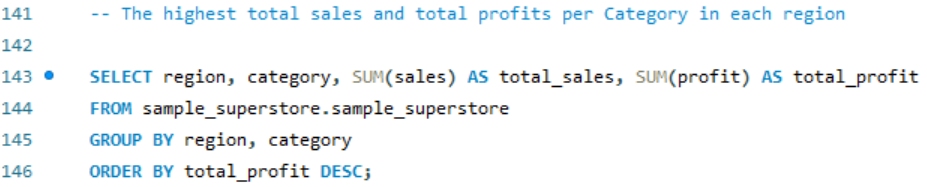
**First, let’s observe the total sales and total profits of each category with their profit margins:**

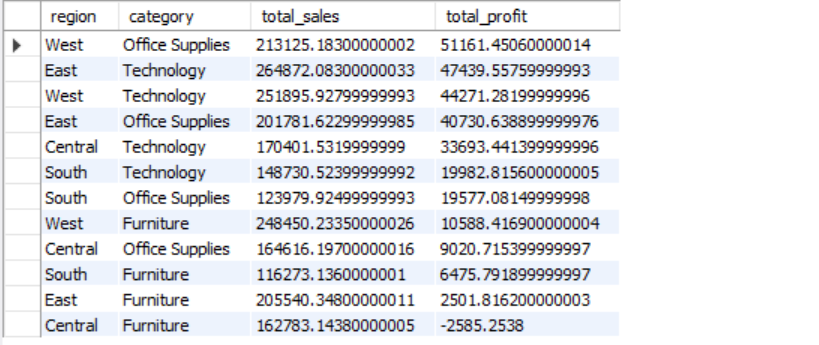
****

****

* Out of the 3, 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 does not convert well overall.

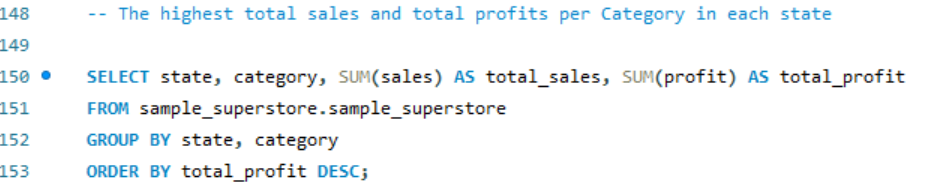
**Let’s observe the highest total sales and total profits per Category in each region:**

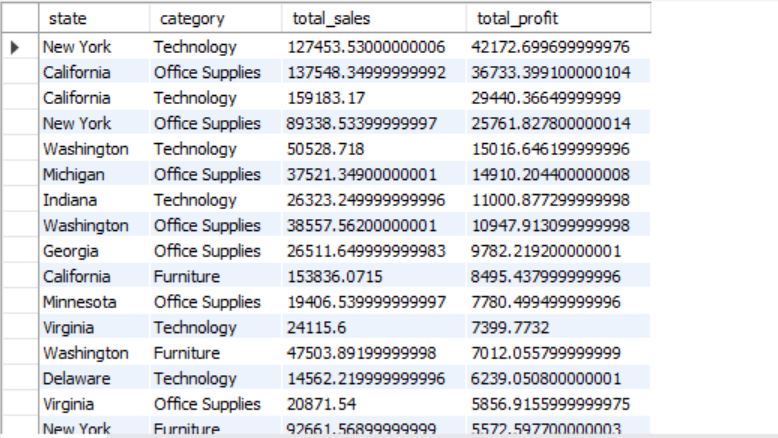


****

* These are the best categories in terms of total profits in each region. The West is in our top 3, with Office Supplies and Technology, twice, and the East has 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.

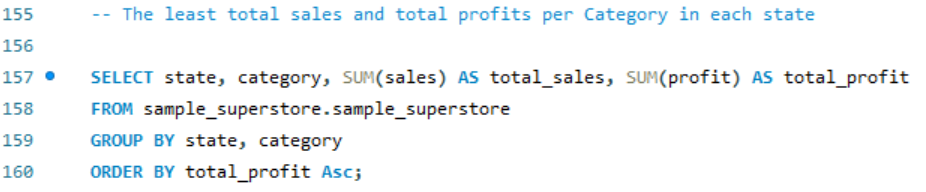
**Now let’s see the highest total sales and total profits per Category in each state:**

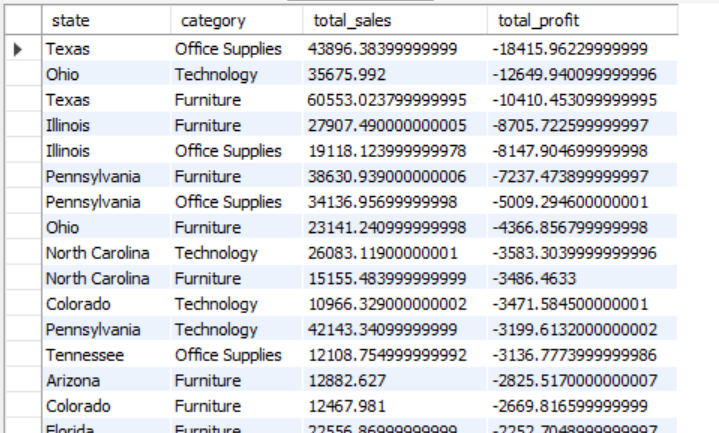
****

****

* The table above shows the highest-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.

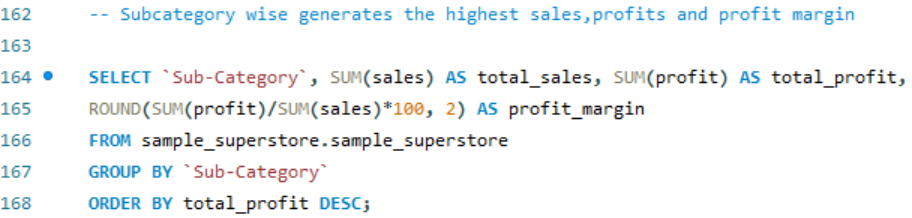
**Let’s check the least profitable ones by just changing our ‘ORDER BY’ clause too ascending (ASC) :**

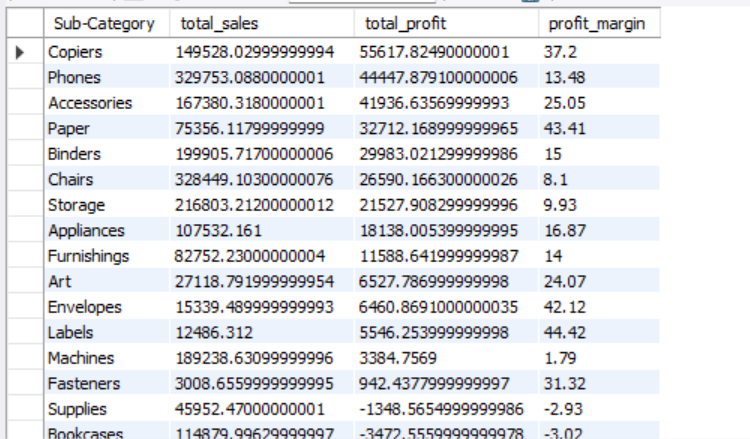




* Office Supplies in Texas, Technology in Ohio & Furniture in Texas, and Illinois are our biggest losses. Let’s move on to subcategories.
* **What subcategory generates the highest sales and profits in each region and state?**

**Let’s observe the total sales and total profits of each subcategory with their profit margins:**

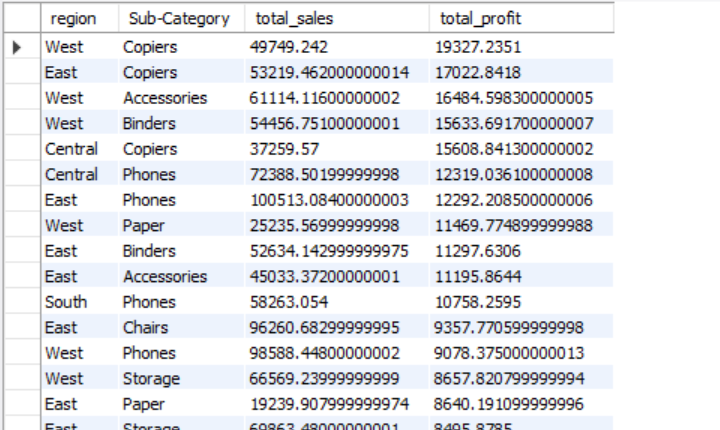
****

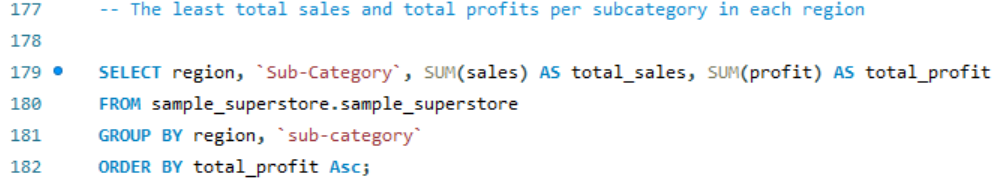


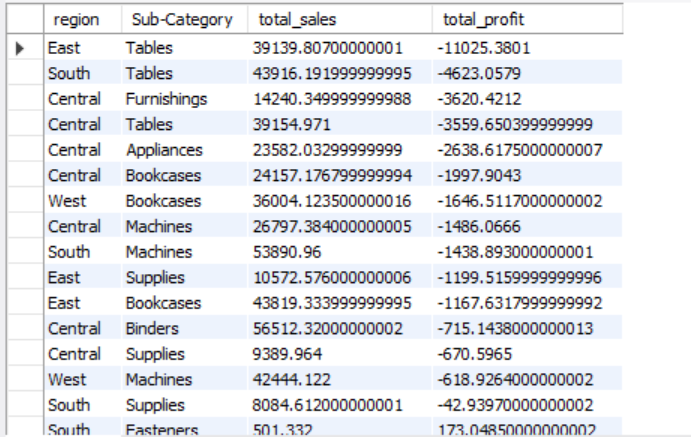
* Out of our 17 subcategories nationwide, our most significant profits come from Copiers, Phones, Accessories, and Paper. The profits and profit margins on Copiers and Papers especially are interesting in the long run. Our losses came from tables, bookcases, and supplies, which we were unable to break even with. Those 3 should be further reviewed as there are sales (except for supplies), but we cannot generate profits from them.

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

****

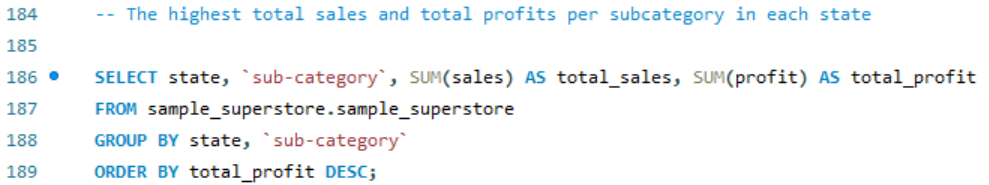
****

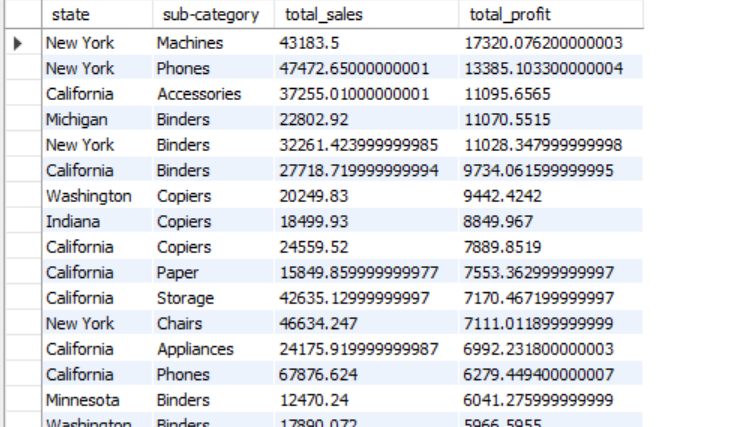
****

****

* We are unable to break even with 14 subcategories. Tables and Furnishings are our most significant losses in profits in the East, South and Central regions.

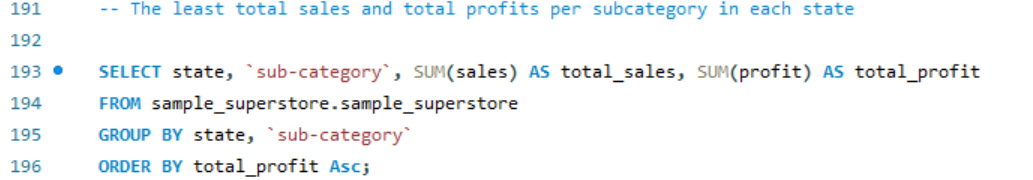
**Now let’s see the highest total sales and total profits per subcategory in each state:**

****

****

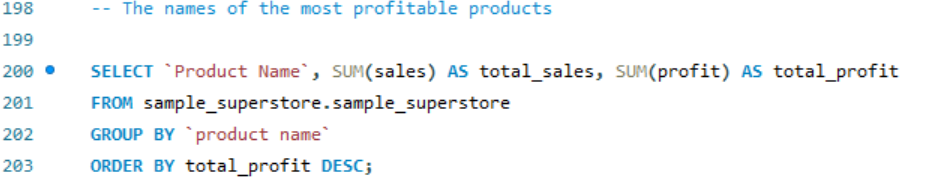
* Machines, Phones, and Binders perform very well in New York. Followed by Accessories and Binders in California and Michigan, respectively.

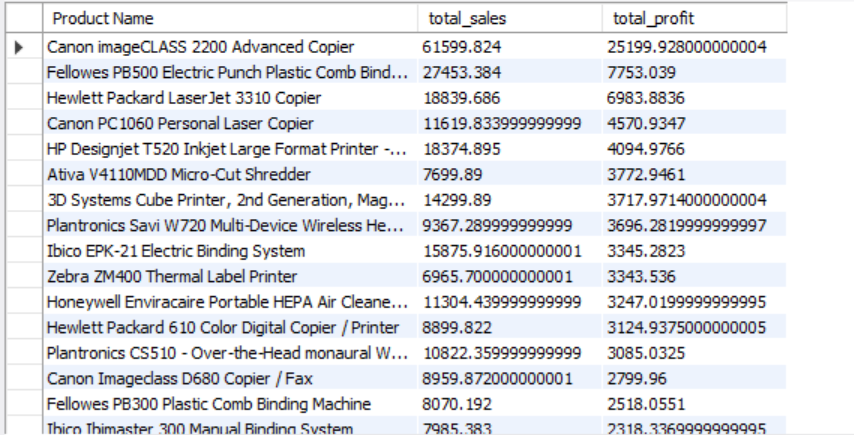
**Let’s see the lowest sales and profits. Still for biggest lost in profits.**

****



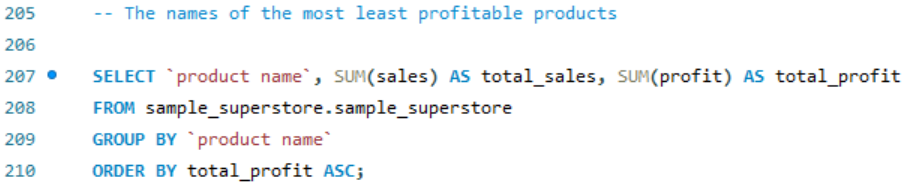
* Binders are our most significant losses in Texas and Illinois. Machines are not profitable in Ohio at all. We should observe and rethink our strategies in those areas.
* **What are the names of the most and least profitable products to us?**

****



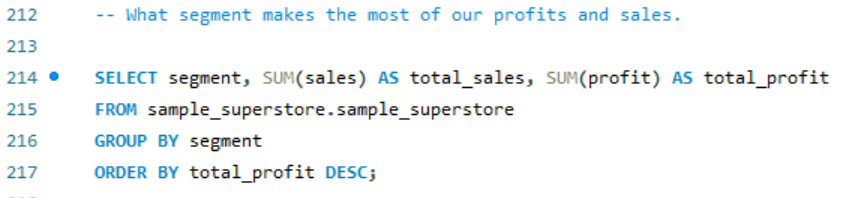
* These Copiers, Machines, and Printers are the main foundations of our profits. The Canon imageClass 2200 Advanced Copier, Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind, and the Hewlett Packard LaserJet 3310 Copier are our top 3.

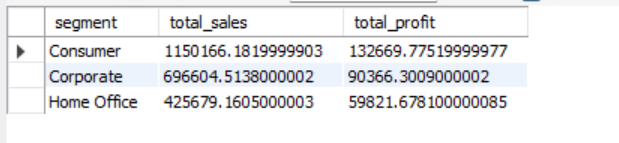
**Let’s verify our less profitable ones:**



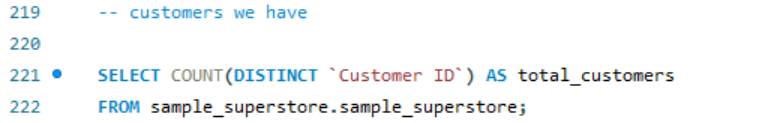


* The Cubify CubeX 3D Printer Double Head Print, Lexmark MX611dhe Monochrome Laser Printer, and the Cubify CubeX 3D Printer Triple Head Print are the products that operate the most at a loss.
* **What segment makes the most of our profits and sales**



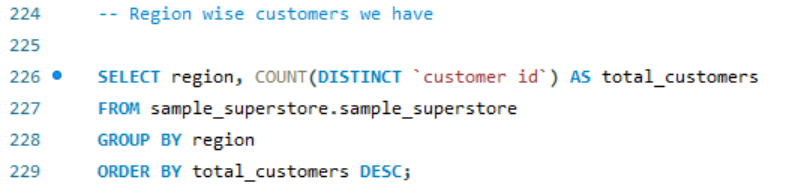
****

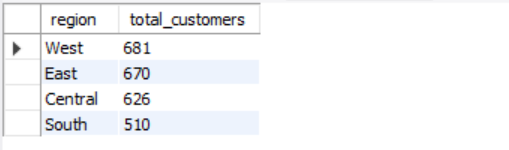
* The consumer segment brings in the most profit followed by Corporate and then Home office
* **How many customers do we have (unique customer IDs) in total and how much per region and state?**

****

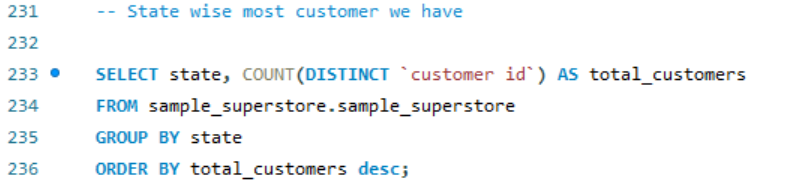
****

* We’ve had 793 customers between 2014 and 2017
* **Let’s check Region wise customers we have**

****

****

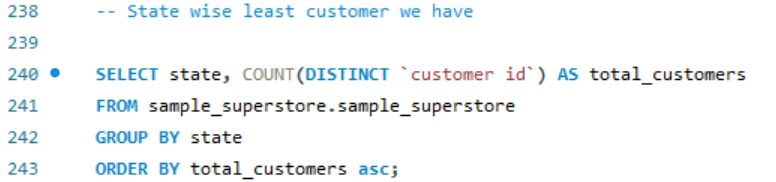
* We surely had customers moving around regions which explains why they all do not add up to 793. Since there could be double counting. The West is the area where we have the biggest market of all.
* **Let’s check State wise most customers we have**

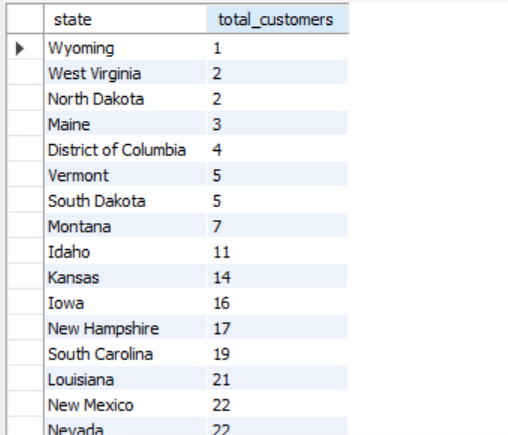




* We have the most customers in California, New York and Texas.

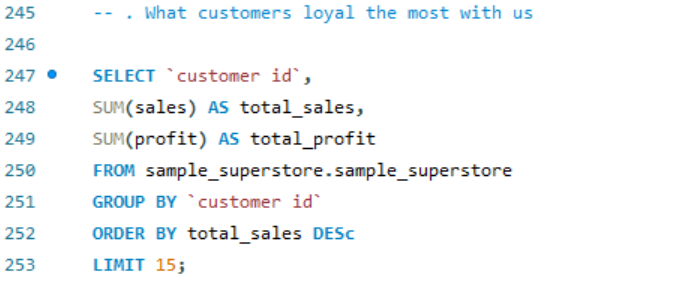
**Let’s check State wise most customers we have**

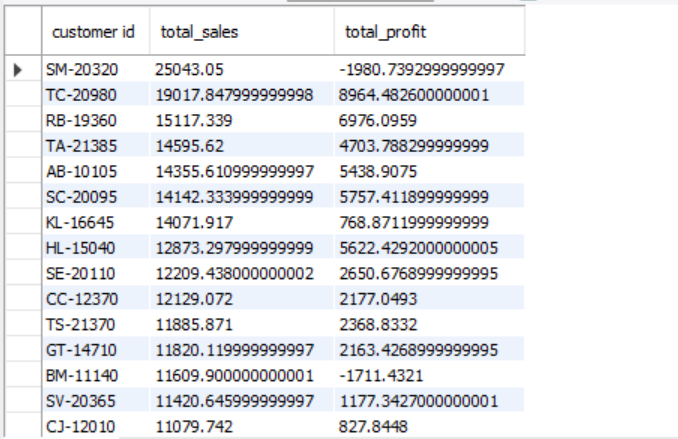




* Wyoming, North Dakota and West Virginia are the places where we had the least customers carry on business with us there.
* **Customer rewards program**

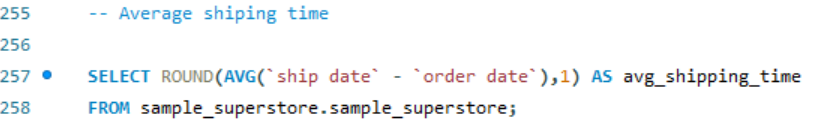
**Let’s say we want to build a loyalty and rewards program in the future. What customers spent the most with us? That generated the most sales. It is always important to cater to our best customers and see how we can provide more value to them, as it’s cheaper to keep a current customer than to acquire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





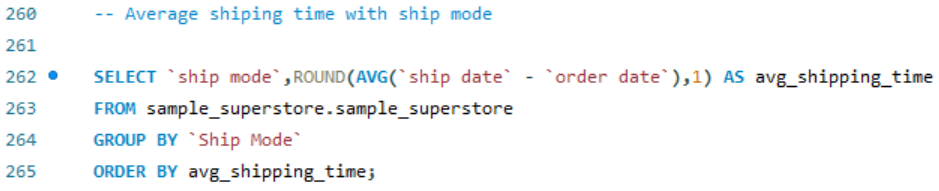
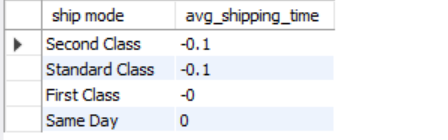
* The display of the customer names is on file but showing the unique customer ID is a form of pseudonymization for security reasons. What is interesting to see is that customer ID ‘SM-20320’ is the customer who spent the most with us but is not bringing us profit. We still must reward his/her loyalty.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





**The shipping time in each shipping mode is?**

**** ****

* Average shipping time -0.1 is second class and standard class.