US Superstore Data Analysis and Visualization Project.

The US Superstore dataset was obtained via Kaggle, and it provides some insights into the online orders of the business. The online market has been growing taking up some of the market share over the past two decades. Customers are able to purchase items such as books, toys, clothes, and more without having to leave their own homes. The data contents purchases collected from customers starting 2014 and ending in 2018. There are 9994 total rows and 21 columns in the dataset. Columns include information such as Order ID, Order Date, Ship Mode, Customer ID, Customer Name, etc. The data was downloaded in a CSV format and uploaded to Big Query for analysis. My goal for this project was to generate some insights using SQL and visualize the results using Tableau. Some of the questions I would be answering are:

* Which Product Category sold the most and is profitable?
* Which Product Sub-Category sold the most and is profitable?
* Which Customer Segment provides the most profit?
* Which Customer Segment has made the most shipments?
* Which cities have made the most sales?
* The bottom 10 cities with the least sales.
* Top 10 customers that have made the most sales in 2017.
* Which Region had the highest profits?
* Which 10 states had less than 100 quantities sold in 2017?

Before beginning my analysis, I ran this query to ensure the dataset would have no issues in Big Query.

Text

Description automatically generated with medium confidence

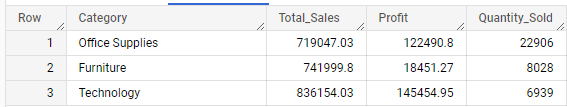
1. **Which Product Category sold the most and is profitable?**

I first began by running this query to generate the output I wanted to receive.

Text

Description automatically generated

Using this, I would receive the Product Category, Total Sales, Profit, and the Quantity Sold while grouping them by the Category. The Sales and Profit were summed up and round to 2 decimal places while the Quantity sold was just summed up. The output came out to be:



We can see that Technology made the most from sales and had the highest profit while having the least amounts of units sold. Technology made $836,154.03, Furniture made $741,999.80, and Office Supplies made the least in sales at $719,047.03. Technology such as computers and smartphones tend to be expensive purchase someone can make. That is the reason why they are so profitable while being able to sell the least number of units.

Chart, waterfall chart

Description automatically generated

1. **Which Product Sub-Category sold the most and is profitable?**

The query used for this is similar to the one for finding the most profitable Product Category, one of the changes made is that Category was switched out to find the Sub-Category instead. They were ordered descending so the highest results would appear first in the table.

Text

Description automatically generated

Table

Description automatically generated

Phones were the best performing in terms of sales since it had $330,007.05 in sales, $44,515.73 in profit, and a quantity sold of 3,289. Furthermore, items included in the technology category would be Machines, Accessories, and Copiers. While not making as much in sales, they also tended to have comparable profits when compared to phones. Furniture items tended to perform poorly in profits. Customers are probably looking towards dedicated furniture stores and would rather see the item in person instead of making the purchase online. Tables and Bookcases made the superstore lose profits, so it would not be advisable to push those items online.

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

1. **Which customer segment provides the most profits?**

Now, I would like to examine which customer segment is the most profitable. The query I used involved the Segment and Profit columns. The Profits were summed up, rounded to 2 decimal places, and grouped by the segment.

Text

Description automatically generated

Chart

Description automatically generated with low confidence

The consumer market ended up being the most profitable for online sales at $134,199.21. Corporate and Home Offices are not too far behind either.

Chart, bar chart

Description automatically generated

1. **Which Customer Segment made the most shipments?**

The query I used for this question is fairly similar to the one used for finding the highest profit made. I wanted to know who made the most shipments to see if it would follow the segment that made the highest profit. Instead of summing the shipment method, I used a COUNT function to get a count of the total amount of shipments. These were then grouped by the segment to get a more accurate result.

Text

Description automatically generated

Graphical user interface, text, application

Description automatically generated

The results fall in line with what we saw with profits. Consumers made the most of shipments at 5,191 while Corporate made 3,020 and Home Offices made 1,783.

1. **Which cities made the most sales?**

Here, I used the city, Sales, and Quantity columns to generate the results in SQL. The sales were summed up and rounded to 2 decimal places while the quantity sold was just summed. They were both grouped by the city column. Then, they were ordered by Sales in descending order so the highest results would appear first.

Text

Description automatically generated

Table

Description automatically generated

It is not surprising to see locations such as New York City, Los Angeles, and Seattle take the top results in terms of sales and quantity sold. New York City was especially high having $256,368.16 in sales and selling 3,417 items. The higher population an area has, we can expect to see higher sales from them.

Chart, bar chart

Description automatically generated

1. **The bottom 10 cities with the least number of sales.**

The query used here is the same as the last one, except I removed the descending so it would display the lowest results first.

Text

Description automatically generated

Table

Description automatically generated

The three cities with the least number of online sales would be Abilene, Texas, Elyria, Ohio, and Jupiter, Florida. They had $1.39, $1.82, and $2.06 in total sales respectively.

1. **Top 10 Customers that made the most sales in 2017.**

Here, I wanted to work with the Customer ID, Customer Name, Sales, and Quantity columns. As usually, sales were summed up and rounded to 2 decimal places while quantity was just summed up. A specific condition I wanted to look for was sales that were made only in 2017 so I added a WHERE clause specifying the range of dates I wanted. They were grouped by Customer ID and Customer Name, ordered by Sales descending, and I only wanted 10 names, so I added a limit of 10.

Text

Description automatically generated

Graphical user interface, application, table

Description automatically generated

Here, we can see that the customer that made the highest number of sales would be Raymond Buch at $14,203.28. Based on what we saw, we can infer that Raymond tended to make technology purchases in 2017. I mention this because he has a lot of sales while not buying as many items. The same could be said for Tom Ashbrook and Hunter Lopez as well.

1. **Which region had the highest profits?**

I wanted to know what the most profitable region was, so I would be using the Region and Profit columns for this query. Profit was summed up and rounded to 2 decimal places like before and they would be grouped by the Region. The regions were ordered descending so I would see the highest amounts first.

Text

Description automatically generated

A picture containing graphical user interface

Description automatically generated

The West region of the US ended up having the highest profits at $108,418.45. The East Region is not too far behind at $91,522.78. I believe this is because many major cities tend to be located in these areas like New York City, Los Angeles, Seattle, and many others. Furthermore, the South and Central may be more rural areas since they do not have many profits for online orders.

Chart

Description automatically generated

1. **Which 10 States had less than 100 quantities sold in 2017?**

What I wanted to know next was 10 states that were not able to sell over 100 units in 2017. I would be working with the State and Quantity columns for this question. Quantity would be summed up as well. I wanted the results to only be from 2017 so I specified a range of dates in the WHERE clause and grouped it by the state. I added a HAVING condition specifying that I wanted states that made less than 100 quantities sold and ordered it by that as well. I added a limit displaying 10 results as well.

Text

Description automatically generated

Graphical user interface, application, table

Description automatically generated

These are 10 states that sold less than 100 items in 2017. New Mexico comes close to making it to 100 but falls short at just 98. It ranges all the way to Louisiana which were only able to make 50 online sales. Perhaps online shopping isn’t as popular in these states when compared to the others.

There were some other observations I wanted to make as well using Tableau. I wanted to see what the popular shipping was when making online purchases.

Chart, bar chart

Description automatically generated

Standard Class shipping is the popular option when making online purchases. Standard likely does not incur an extra cost which may be why it is used. Online shoppers would not want to pay extra just to receive their item a few days quicker.

Chart, line chart

Description automatically generated

Here, I wanted to see how monthly profits for the product categories would change throughout the year. The red line is for Technology, Orange for Office Supplies, and Blue for furniture. Technology sales consistently make the highest profits from online orders only getting beaten out 4 of the 12 months. Office Supplies make the most profits during the months of January, June, August, and September. My theory is that August and September are usually regarded as the back-to-school season, so many teachers, students, and parents are going to be buying supplies to prepare for the academic year. For schools, there is usually a winter break around the months of December and January so they have time to go out and resupply on anything they might be missing. Purchases also tend to increase around the months of November and December. This time is the holiday season so many people would be looking to make purchases on Black Friday, or finish their holiday shopping.

**Recommendations**

Based on everything I had seen, here are some recommendations I have for the Superstore:

* I would recommend for them to focus mainly on Technology and Office Supplies for their online market. Hardly anyone is making any furniture purchases and is even causing them to lose profits. When it comes to furniture, people would prefer to view the item in person to see if it will match their current decorations and fit in their home. They should consider removing furniture items that cause them to lose profits, or lower prices when viewing the item online. Ordering phones and binders online would be okay since they know exactly what they are going to get.
* They should try to market their online store towards consumers and corporations located in the East and West regions of the US. These two regions are where the majority of their profits come from. In fact, 70% of all online purchases come from these two alone. These are where major cities tend to be located such as New York City and Los Angeles. The Central and Southern regions may be composed of rural areas where shipping may not be readily available.
* US Superstore should also hold promotions depending on the time of year. For example, they should hold sales on office supplies during the months of August and September since that is what makes the most profit at the time. Both Technology and Office Supplies should be heavily promoted during November and December since that is where many do their holiday shopping.