SUPERSTORE SALES ANALYSIS BREAKDOWN

With growing demands and cut-throat competition in the market, a Superstore Giant in the USA is seeking to understand what works best for them. They would like to understand which products, regions, categories, and customer segments they should target or avoid. The superstore chain deals in technological products, office supplies, and furniture.

This analysis will be carried out with Excel, and SQL and visualized using PowerBi. Excel will serve as the first repository for the data, and SQL will be used for manipulating, transforming, and giving meaning to the data. In contrast, PowerBi will give a visual representation of the data. The following sales performance analysis will follow the 6 steps of Data Analysis: Ask, Prepare, Process, Analyse, Share, and Act.

1. Ask

The business problem given to us was interpreted as "What are the best products, regions, categories, and customer segments for the Superstore to target or avoid increasing profitability?"

Business objectives:

- How can we optimize the 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 full documentation of all the data cleaning, manipulation, and analysis.
- A dashboard with visualizations and main outcomes.
- Recommendations based on the insights and analysis.

2. Prepare

The data is publicly available on my GitHub; Superstore Data

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 the 4 years of business.

It contains the data of 793 customers.

The data contains the 21 columns namely;

Row ID => Unique ID for each row.

Order ID => Unique Order ID for each Customer.

Order Date => Order Date of the product.

Ship Date => Shipping Date of the Product.

Ship Mode=> Shipping Mode specified by the Customer.

Customer ID => Unique ID to identify each Customer.

Customer Name => Name of the Customer.

Segment => The segment where the Customer belongs.

Country => Country of residence of the Customer.

City => City of residence of the Customer.

State => State of residence of the Customer.

Postal Code => Postal Code of every Customer.

Region => Region where the Customer belongs.

Product ID => Unique ID of the Product.

Category => Category of the product ordered.

Sub-Category => Sub-Category of the product ordered.

Product Name => Name of the Product

Sales => Sales of the Product.

Quantity => Quantity of the Product.

Discount => Discount provided.

Profit => Profit/Loss incurred.

The only limitation of the dataset given is that the most recent data point was almost 6 years ago, which means the data is not current. However, the data is quite reliable, original, comprehensive, and is cited.

3. Process

This involves the cleaning of the data with the help of Excel as the file is already a CSV file so a look-through of the data with Excel can be ideal. The following was done;

- Observe the data
- Check for missing data, spaces/blanks with the help of conditional formatting
- Remove duplicate rows.

- Correctly format columns for easy SQL analysis

While exploring the dataset, the following was noticed.

- The data looks correct and consistent. Everything looks well-structured for further analysis it just needs a little editing.
- With the use of COUNTBLANK, blanks were checked for and there were 0 blanks. Just for extra verification, conditional formatting was noticed, which is a technique to highlight certain values of interest, the new formatting rule was set to be 'Format only cells that contain', then "Format only cells with': Blanks'. The color of choice was yellow. Then we zoomed out to have a bird's eye view of the dataset. We can confirm that the 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 them hence returning no duplicate data for the data set.
- Finally, made sure the Order Date and Ship Date were well-formatted by formatting them to dates (which they already were) and formatted the Sales and Profits columns from numbers to currencies since we are talking about financials. The discount column was not formatted to currency as in this context, the discount is more of a percentage value. So, the discount will later be formatted from number to percentage by just multiplying it by 100.

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

4. Process

SQL Analysis is done using Microsoft SQL Server Management Studio, the following questions were answered using queries that can be accessed via; Superstore-Sales-SQL-Data-Analysis Project

- What are the total sales and total profits of each year?
- What are the total profits and total sales per quarter?
- What region generates the highest sales and profits?
- What state and city brings in the highest sales and profits?
- The relationship between discount and sales and the total discount per category
- What category generates the highest sales and profits in each region and state?
- What subcategory generates the highest sales and profits in each region and state?
- What are the names of the products that are the most and least profitable to us?

- What segment makes the most of the profits and sales?
- How many customers do we have (unique customer IDs) in total and how much per region and state?
- What customers spent the most with us?
- What is the average shipping time per class and in total?

5. Share

The 'share' part of this analysis involves the visualization of the data analysed using SQL, the visualization was done using Microsoft PowerBi. The visualization can be found here; Superstore-Sales PowerBi Report

6. Act

To act on the data analysed, there is a need to come up with inferences from the analysis, make conclusions, and give recommendations.

- The profits got progressively better. The sales too even with a short halt in 2015. We should keep the pace up on that aspect.
- The most profitable quarter all year round was Q4. To maximize even more profits, we must have enough stock and push the 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 profit 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 the Central region to instead of the West region stores as we are more profitable there and could really establish ourselves as a kingpin in that region.
- California, New York, and Washington are the most profitable markets and most present ones, especially in terms of sales as states. We have to focus more on them. The least profitable markets are Texas, Ohio, and Pennsylvania. Which I believe that we should decrease the presence there or even put a halt at the 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 the most profitable cities and we list them as being a top priority because it is easier to rule a city than rule 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 the top 3 cities so it is clear that we have to 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. Furniture is still making profits but does 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 bring 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.
- State-wise, Technology and Office supplies bring us the most profit in the states 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 the biggest losses so we have to drastically reduce these types of products in those areas.
- Out of the 17 subcategories nationwide, the biggest profits come from Copiers, Phones, Accessories, and Paper. The profits and profit margins on Copiers and Papers especially are interesting in the long run. We should immediately push these products as we have a great market share with these items. The losses came from Tables, Bookcases, and Supplies which we are incapable of breaking even. We must spend less time and money with them. Especially with tables because compared to the only 3 losses, Tables lost us $17725 which is huge compared to the 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 the 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 the business there with those products. For the 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 image Class 2200 Advanced Copier, Fellowes PB500 Electric Punch Plastic Comb Binding Machine with Manual Bind, and the Hewlett Packard LaserJet 3310 Copier are the top 3 in profits. We must always keep up the stock with these. For the 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 the Corporate and then the Home office. We must give more importance to the consumer segment even if all the 3 are profitable.

- Finally, for the 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 loses 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 the respective niche.