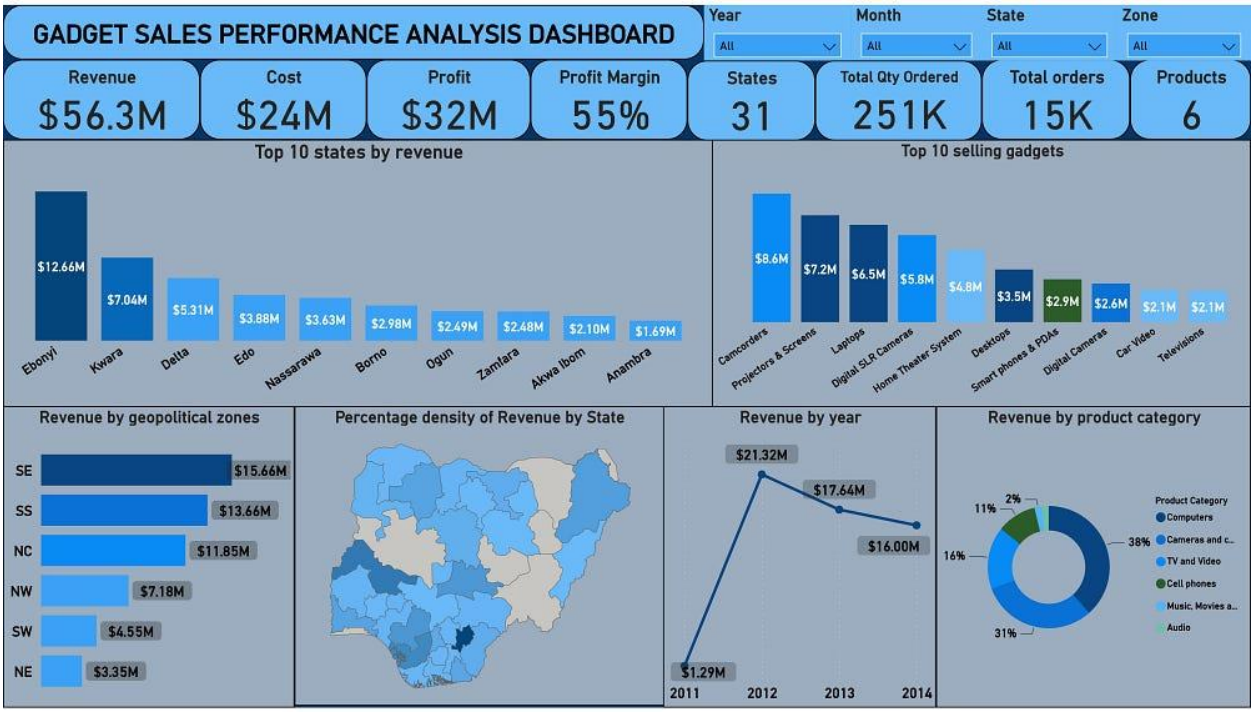
Data analysis: wrangling, cleaning, analysis and visualization of gadget stores' sales performance data using power query and power BI



Gadget sales analysis dashboard

Background

The term "**data cleaning**" simply means making data more accurate for analysis. It entails identifying data errors and then correcting them by changing, updating, or removing data. The opposite of clean data is dirty data. Dirty data is data that is incomplete, incorrect, or irrelevant to the problem you are trying to solve. Various kinds of dirty data may include duplicate data (*which may skew or inflate the analysis*), outdated data, incomplete data (*missing several important fields*), inaccurate data, or even inconsistent data (*which may contain different formats in the same field, especially the* ***date format***). To be considered clean, data must be **complete**, **consistent** (*all columns must have the correct data type and format*), and **accurate**.

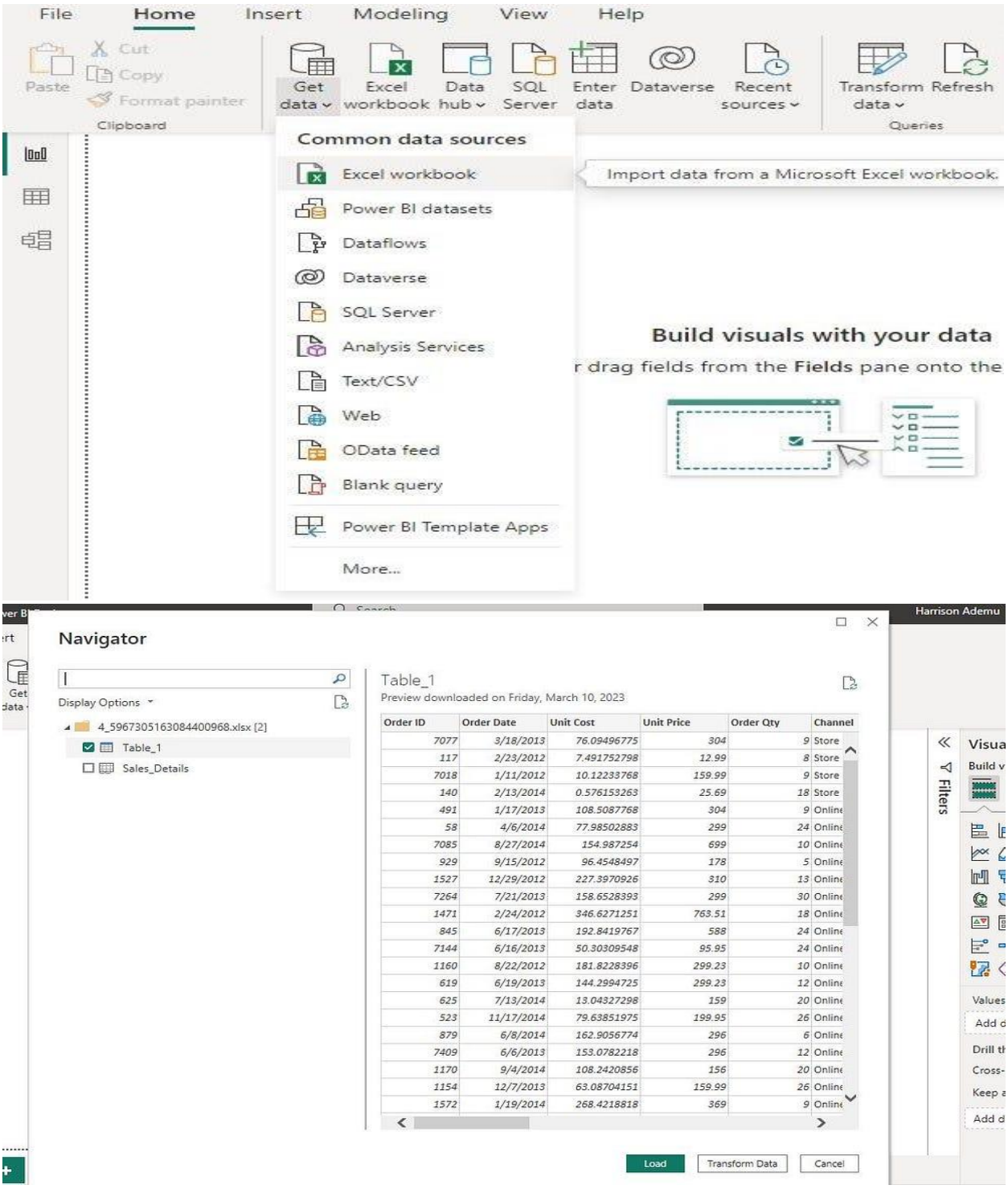
Data wrangling, on the other hand, is simply transforming data from one "raw" data form into another format with the intent of making it more ***accessible and easier to analyze.*** During this process, the simple **ETL** ***(****Extract, Transform, Load****)*** method was used to extract the data from a CSV format (comma-separated value), Transform it from the Navigator field into the power query in order to clean the data and then finally Load it into power BI to visualize.

About the dataset

The dataset contains details about gadget stores under the same company located in 31 states in Nigeria including the Federal Capital Territory, and the sales performance of each individual store from 2011–2014 in terms of the gadgets (products and their subcategories) sold during that time. The dataset initially had 15000 rows and 12 columns, but during the analysis, a few more columns were added to contain calculations that were relevant to the analysis. The dataset can be gotten here.

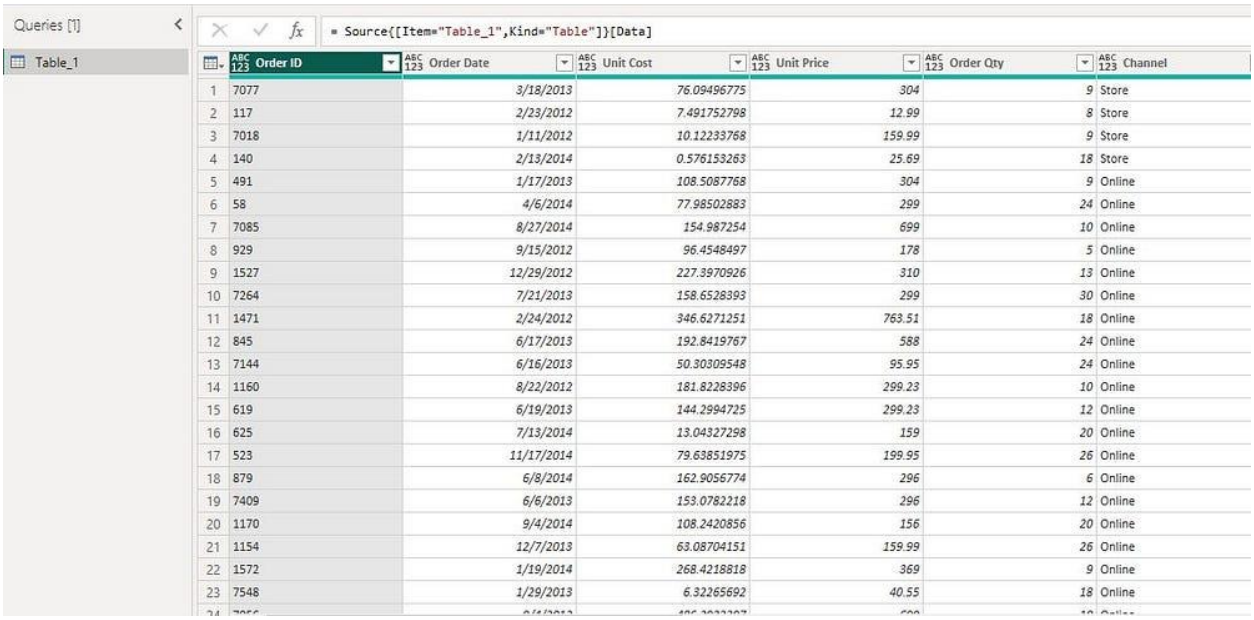
Data cleaning steps:

The dataset was imported into the navigator field and transformed to the power query where the cleaning was done.



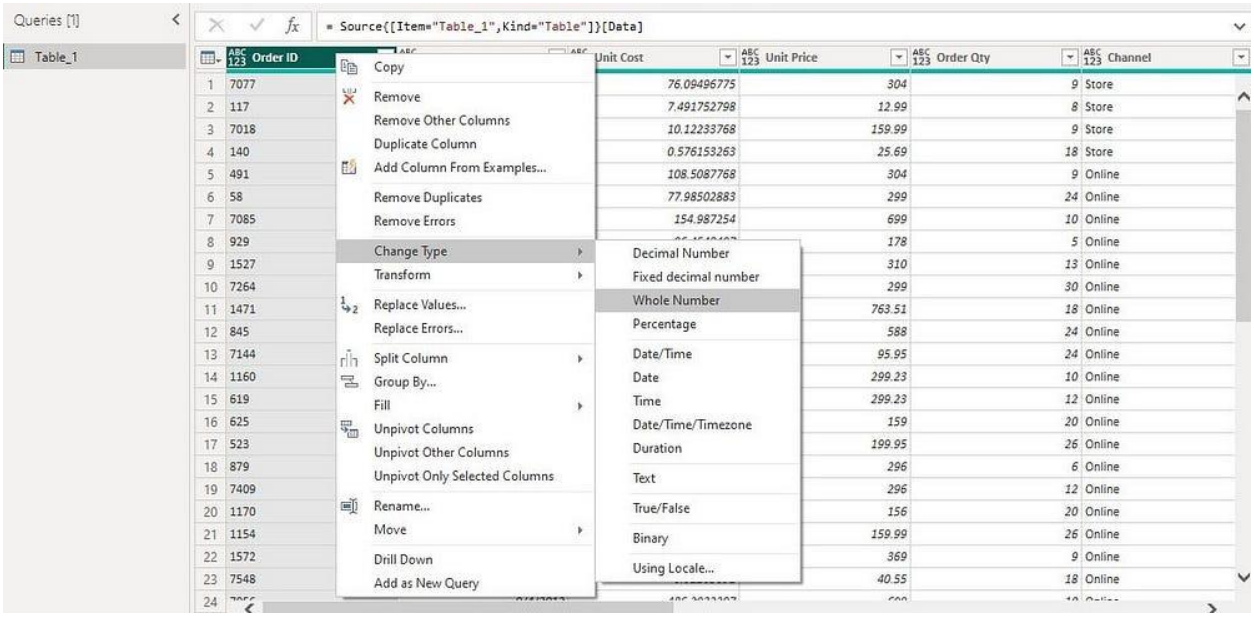
Upon transforming into the power query editor, it was observed that the dataset was dirty and needed to be cleaned before loading into power BI. *Don't worry, stick with me, and I'll show you every step taken to transform this dirty data into a more appealing dataset.*

Notice how the data type of every column is a string and a numerical (*ABC123*); each column needed to be changed into the right data type.

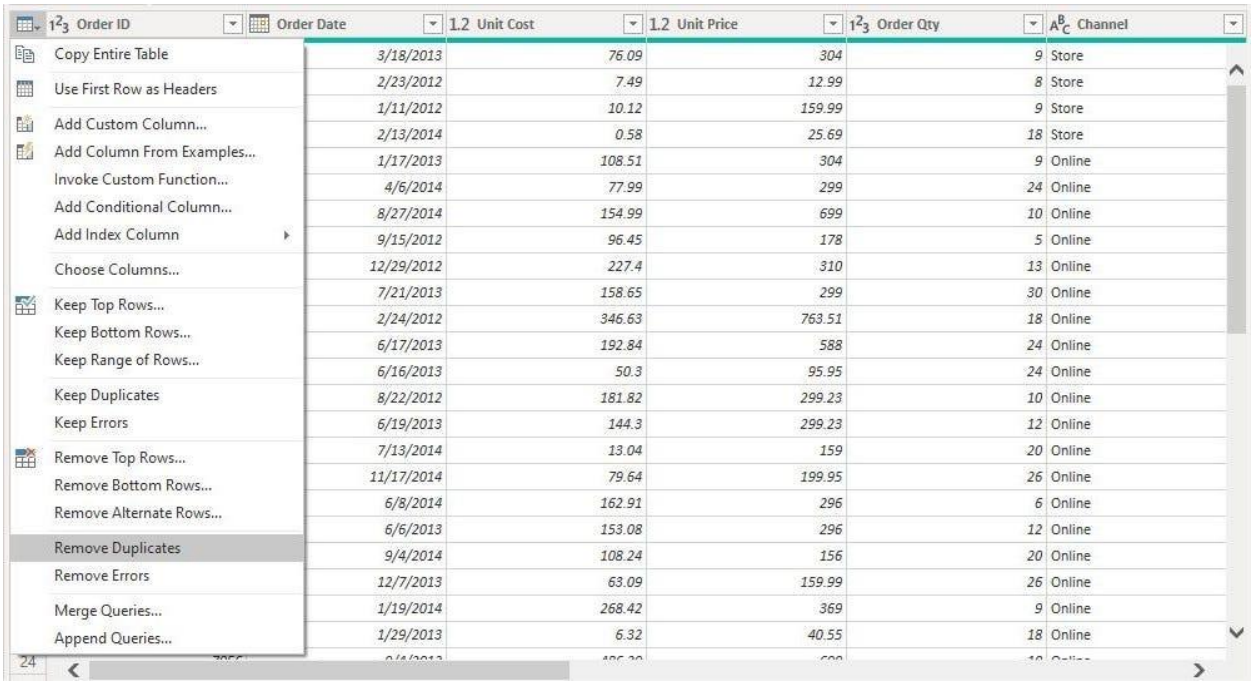


	Order ID	Order Date	Unit Cost	Unit Price	Order Qty	Channel
1	7077	3/18/2013	76.09496775	304	9	Store
2	117	2/23/2012	7.491752798	12.99	8	Store
3	7018	1/11/2012	10.12233768	159.99	9	Store
4	140	2/13/2014	0.576153263	25.69	18	Store
5	491	1/17/2013	108.5087768	304	9	Online
6	58	4/6/2014	77.98502883	299	24	Online
7	7085	8/27/2014	154.987254	699	10	Online
8	929	9/15/2012	96.4548497	178	5	Online
9	1527	12/29/2012	227.3970926	310	13	Online
10	7264	7/21/2013	158.6528393	299	30	Online
11	1471	2/24/2012	346.6271251	763.51	18	Online
12	845	6/17/2013	192.8419767	588	24	Online
13	7144	6/16/2013	50.30309548	95.95	24	Online
14	1160	8/22/2012	181.8228396	299.23	10	Online
15	619	6/19/2013	144.2994725	299.23	12	Online
16	625	7/13/2014	13.04327298	159	20	Online
17	523	11/17/2014	79.63851975	199.95	26	Online
18	879	6/8/2014	162.9056774	296	6	Online
19	7409	6/6/2013	153.0782218	296	12	Online
20	1170	9/4/2014	108.2420856	156	20	Online
21	1154	12/7/2013	63.08704151	159.99	26	Online
22	1572	1/19/2014	268.4218818	369	9	Online
23	7548	1/29/2013	6.32265692	40.55	18	Online

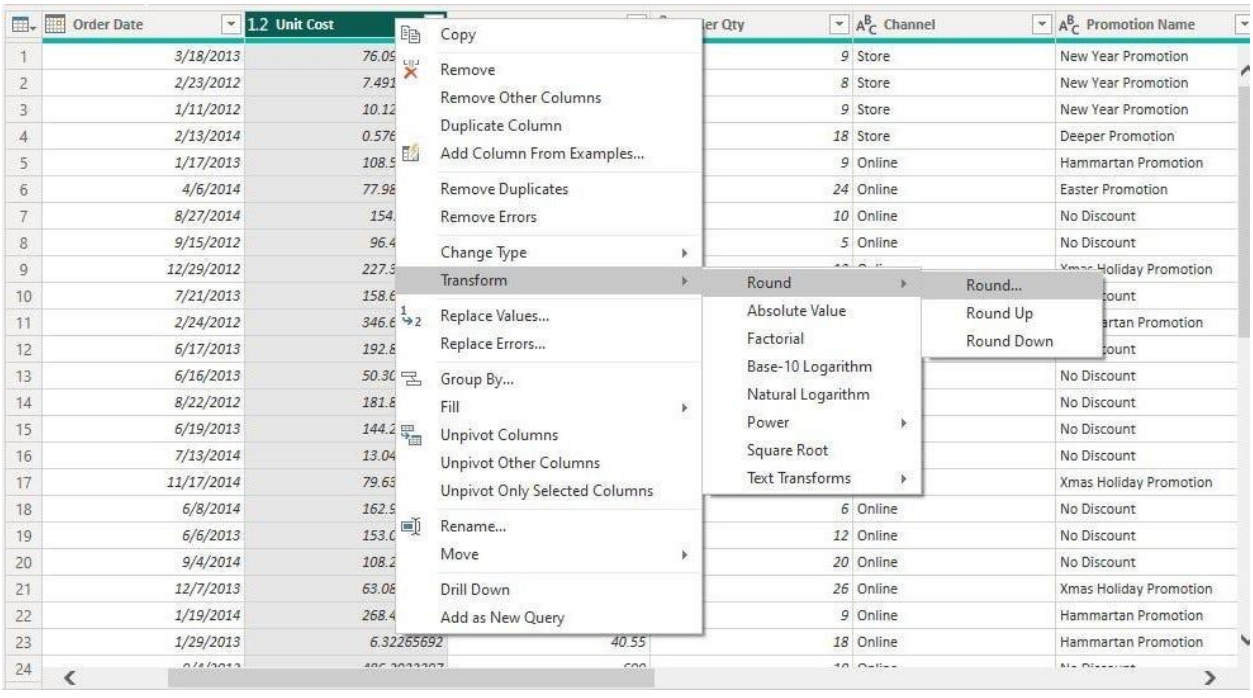
Each column's data type was changed to the right data type to make it more consistent.



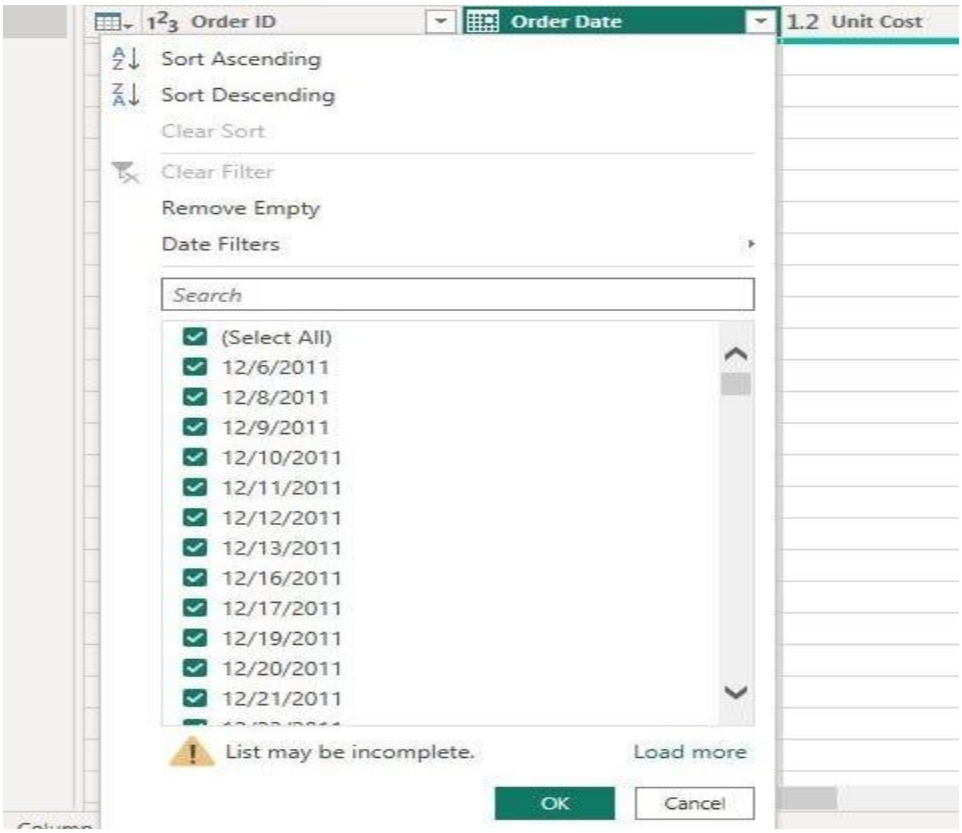
The dataset was then checked for **duplicates**, which were removed to avoid a skewed or inflated analysis.



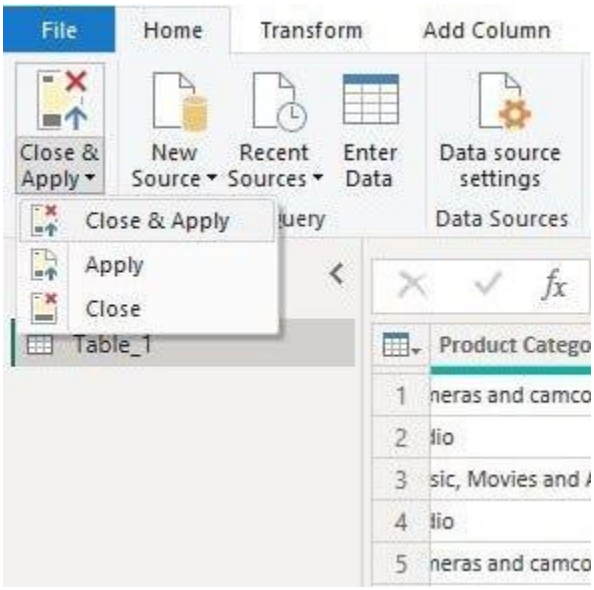
Next, the unit cost had a few more decimals than needed, so for accurate calculations and better visualization, the unit cost was rounded to two decimal places.



The *order date* column was checked by clicking the dropdown icon to filter any inconsistencies in the date format. After investigating, it was found that the date formats were all in unison with the format: '*mm/dd/yyyy'*.



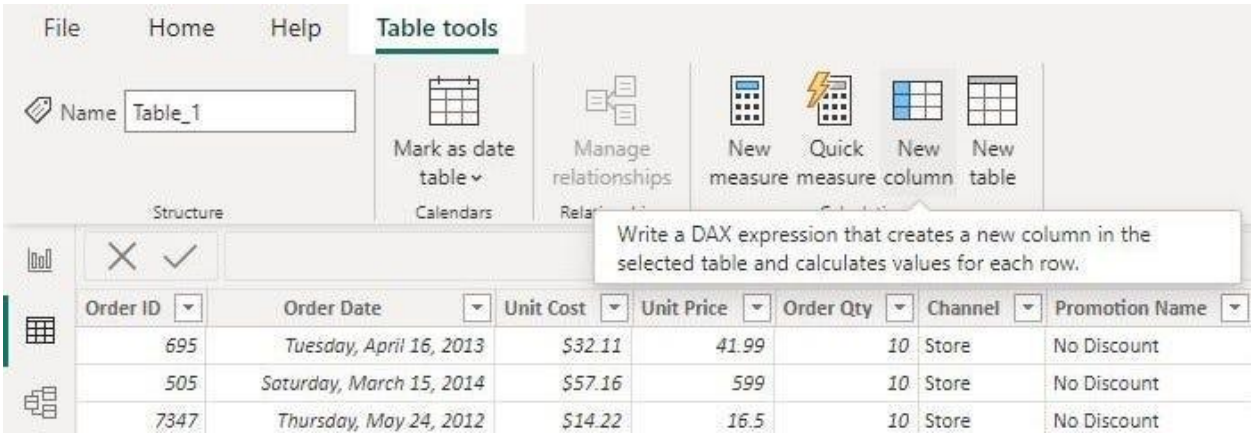
Finally, once I was satisfied that the data was clean and ready for visualization, I loaded the dataset into power BI by applying all of the cleaning steps to it by clicking 'close and apply'.



Once inside the power BI platform, I visualized a few insights by comparing different columns from my dataset in order to build a dashboard, analyzing the sales performance of different stores in different states in Nigeria.

In order to provide a more accurate analysis of the sales performances of these stores, I created three new columns and performed some mathematical calculations to show the revenue, gross profit, and gross profit margin generated from different stores in various states.

I created these columns by clicking on the data field, and then clicking ‘New column’. Calculating the revenue, gross profit, and gross profit margin was possible using DAX (*data analysis expressions*) to input the formulas.

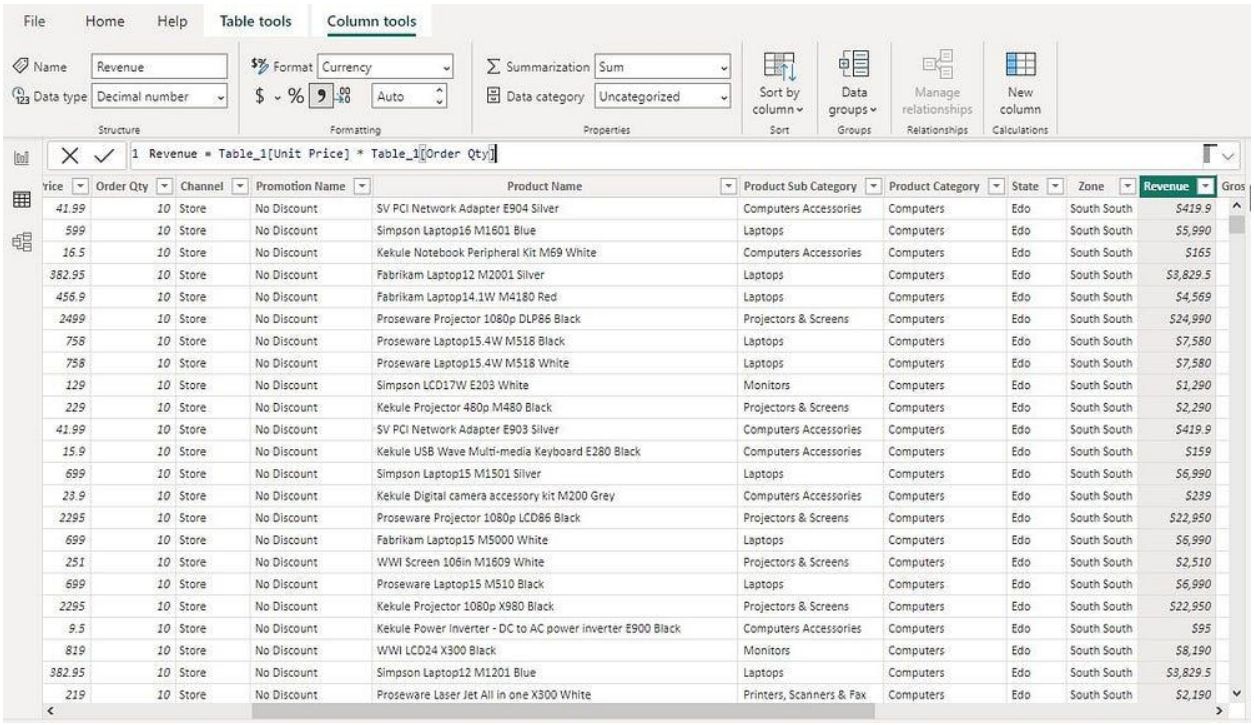


Next, click ‘New column’ to add a new column for the revenue calculations.

For revenue, input the formula:

Revenue = Table_1[Unit Price] * Table_1[Order Qty]

A new column was created automatically, calculating the revenue by multiplying the *unit price* by the *order quantity* for each row and filling up the new column. The format for the revenue column was changed to *currency.*

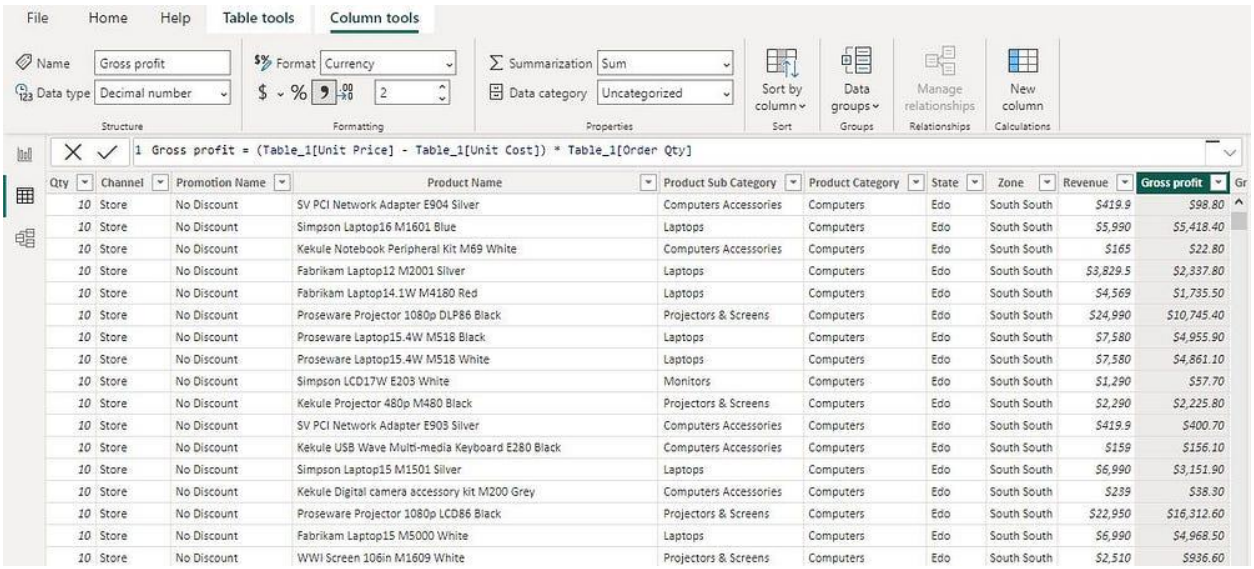


rice	Order Qty	Channel	Promotion Name	Product Name	Product Sub Category	Product Category	State	Zone	Revenue	Gros
41.99	10	Store	No Discount	SV PCI Network Adapter E904 Silver	Computers Accessories	Computers	Edo	South South	$419.9	
599	10	Store	No Discount	Simpson Laptop16 M1601 Blue	Laptops	Computers	Edo	South South	$5,990	
16.5	10	Store	No Discount	Kekule Notebook Peripheral Kit M69 White	Computers Accessories	Computers	Edo	South South	$165	
382.95	10	Store	No Discount	Fabrikam Laptop12 M2001 Silver	Laptops	Computers	Edo	South South	$3,829.5	
456.9	10	Store	No Discount	Fabrikam Laptop14.1W M4180 Red	Laptops	Computers	Edo	South South	$4,569	
2499	10	Store	No Discount	Proseware Projector 1080p DLP86 Black	Projectors & Screens	Computers	Edo	South South	$24,990	
758	10	Store	No Discount	Proseware Laptop15.4W M518 Black	Laptops	Computers	Edo	South South	$7,580	
758	10	Store	No Discount	Proseware Laptop15.4W M518 White	Laptops	Computers	Edo	South South	$7,580	
129	10	Store	No Discount	Simpson LCD17W E203 White	Monitors	Computers	Edo	South South	$1,290	
229	10	Store	No Discount	Kekule Projector 480p M480 Black	Projectors & Screens	Computers	Edo	South South	$2,290	
41.99	10	Store	No Discount	SV PCI Network Adapter E903 Silver	Computers Accessories	Computers	Edo	South South	$419.9	
15.9	10	Store	No Discount	Kekule USB Wave Multi-media Keyboard E280 Black	Computers Accessories	Computers	Edo	South South	$159	
699	10	Store	No Discount	Simpson Laptop15 M1501 Silver	Laptops	Computers	Edo	South South	$6,990	
23.9	10	Store	No Discount	Kekule Digital camera accessory kit M200 Grey	Computers Accessories	Computers	Edo	South South	$239	
2295	10	Store	No Discount	Proseware Projector 1080p LCD86 Black	Projectors & Screens	Computers	Edo	South South	$22,950	
699	10	Store	No Discount	Fabrikam Laptop15 M5000 White	Laptops	Computers	Edo	South South	$6,990	
251	10	Store	No Discount	WWI Screen 106in M1609 White	Projectors & Screens	Computers	Edo	South South	$2,510	
699	10	Store	No Discount	Proseware Laptop15 M510 Black	Laptops	Computers	Edo	South South	$6,990	
2295	10	Store	No Discount	Kekule Projector 1080p X980 Black	Projectors & Screens	Computers	Edo	South South	$22,950	
9.5	10	Store	No Discount	Kekule Power Inverter - DC to AC power inverter E900 Black	Computers Accessories	Computers	Edo	South South	$95	
819	10	Store	No Discount	WWI LCD24 X300 Black	Monitors	Computers	Edo	South South	$8,190	
382.95	10	Store	No Discount	Simpson Laptop12 M1201 Blue	Laptops	Computers	Edo	South South	$3,829.5	
219	10	Store	No Discount	Proseware Laser Jet All in one X300 White	Printers, Scanners & Fax	Computers	Edo	South South	$2,190	

This step was repeated for the gross profit column, by using the formula:

Gross profit = (Table_1[Unit Price]—Table_1[Unit Cost]) * Table_1[Order Qty]

The format for the gross profit was changed to *currency.*

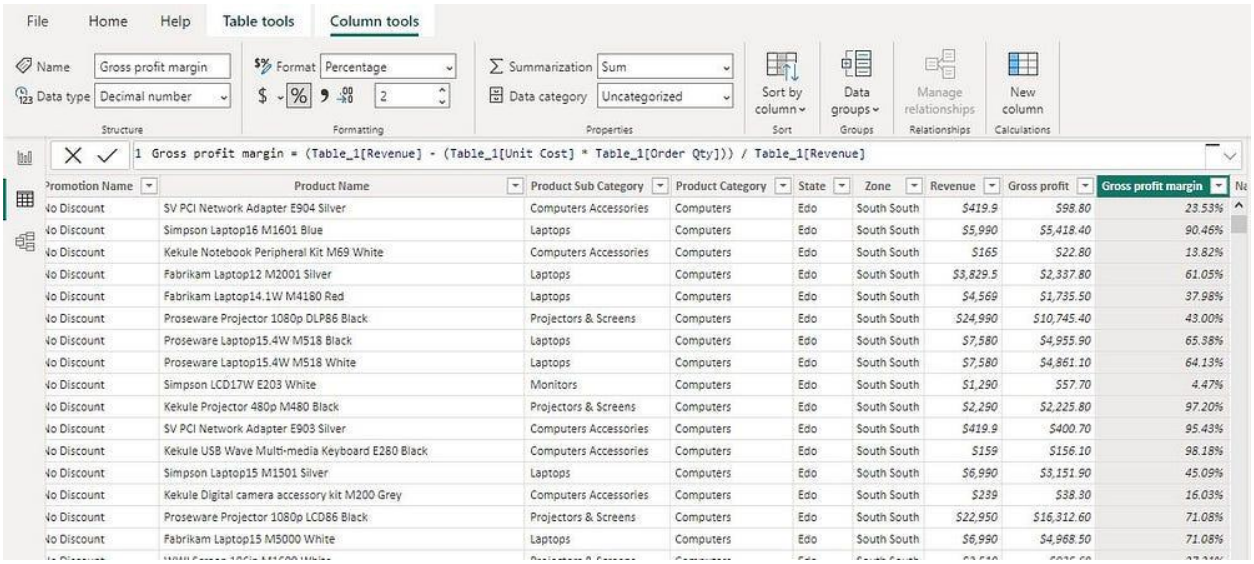


Qty	Channel	Promotion Name	Product Name	Product Sub Category	Product Category	State	Zone	Revenue	Gross profit	Gr
10	Store	No Discount	SV PCI Network Adapter E904 Silver	Computers Accessories	Computers	Edo	South South	$419.9	$98.80	
10	Store	No Discount	Simpson Laptop16 M1601 Blue	Laptops	Computers	Edo	South South	$5,990	$5,418.40	
10	Store	No Discount	Kekule Notebook Peripheral Kit M69 White	Computers Accessories	Computers	Edo	South South	$165	$22.80	
10	Store	No Discount	Fabrikam Laptop12 M2001 Silver	Laptops	Computers	Edo	South South	$3,829.5	$2,337.80	
10	Store	No Discount	Fabrikam Laptop14.1W M4180 Red	Laptops	Computers	Edo	South South	$4,569	$1,735.50	
10	Store	No Discount	Proseware Projector 1080p DLP86 Black	Projectors & Screens	Computers	Edo	South South	$24,990	$10,745.40	
10	Store	No Discount	Proseware Laptop15.4W M518 Black	Laptops	Computers	Edo	South South	$7,580	$4,955.90	
10	Store	No Discount	Proseware Laptop15.4W M518 White	Laptops	Computers	Edo	South South	$7,580	$4,861.10	
10	Store	No Discount	Simpson LCD17W E203 White	Monitors	Computers	Edo	South South	$1,290	$57.70	
10	Store	No Discount	Kekule Projector 480p M480 Black	Projectors & Screens	Computers	Edo	South South	$2,290	$2,225.80	
10	Store	No Discount	SV PCI Network Adapter E903 Silver	Computers Accessories	Computers	Edo	South South	$419.9	$400.70	
10	Store	No Discount	Kekule USB Wave Multi-media Keyboard E280 Black	Computers Accessories	Computers	Edo	South South	$159	$156.10	
10	Store	No Discount	Simpson Laptop15 M1501 Silver	Laptops	Computers	Edo	South South	$6,990	$3,151.90	
10	Store	No Discount	Kekule Digital camera accessory kit M200 Grey	Computers Accessories	Computers	Edo	South South	$239	$38.30	
10	Store	No Discount	Proseware Projector 1080p LCD86 Black	Projectors & Screens	Computers	Edo	South South	$22,950	$16,312.60	
10	Store	No Discount	Fabrikam Laptop15 M5000 White	Laptops	Computers	Edo	South South	$6,990	$4,968.50	
10	Store	No Discount	WWI Screen 106in M1609 White	Projectors & Screens	Computers	Edo	South South	$2,510	$936.60	

The same step was repeated for the gross profit margin, by using the formula:

Gross profit margin = (Table_1[Revenue] - (Table_1[Unit Cost] * Table_1[Order Qty])) / Table_1[Revenue]

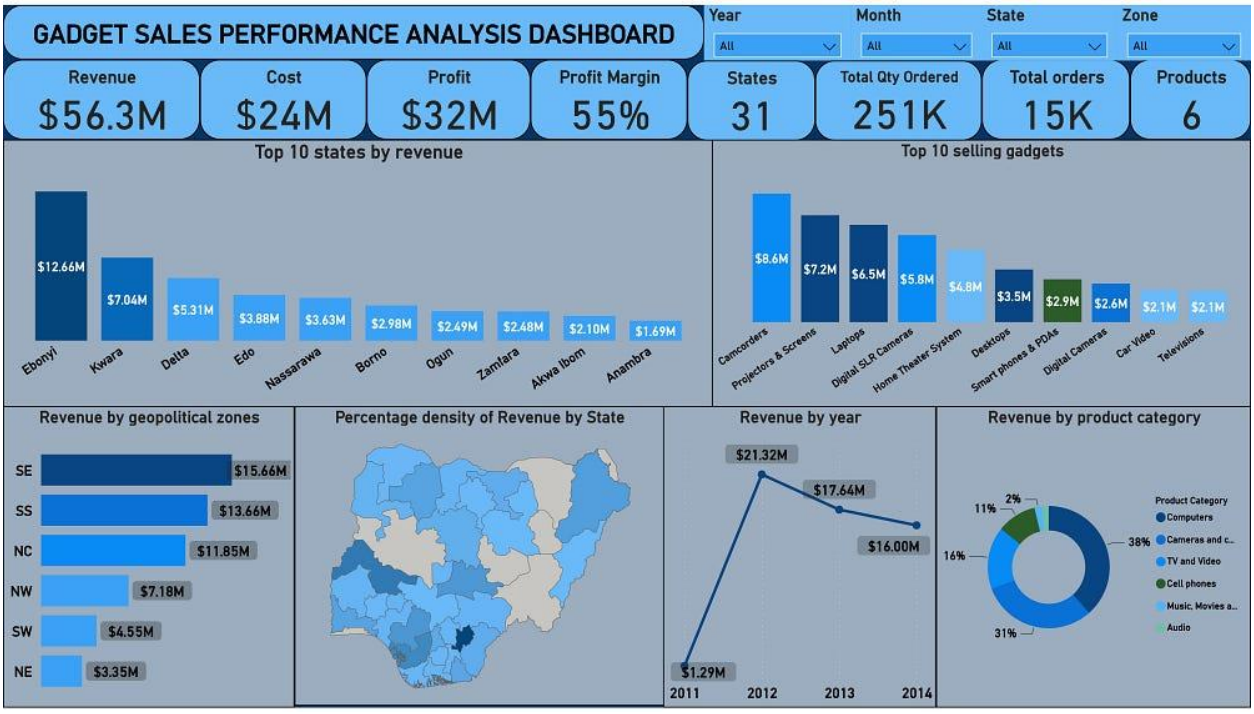
The format of the gross profit margin column was changed to *percentage*.



Promotion Name	Product Name	Product Sub Category	Product Category	State	Zone	Revenue	Gross profit	Gross profit margin
No Discount	SV PCI Network Adapter E904 Silver	Computers Accessories	Computers	Edo	South South	$419.9	$98.80	23.53%
No Discount	Simpson Laptop16 M1601 Blue	Laptops	Computers	Edo	South South	$5,990	$5,418.40	90.46%
No Discount	Kekule Notebook Peripheral Kit M69 White	Computers Accessories	Computers	Edo	South South	$165	$22.80	13.82%
No Discount	Fabrikam Laptop12 M2001 Silver	Laptops	Computers	Edo	South South	$3,829.5	$2,337.80	61.05%
No Discount	Fabrikam Laptop14.1W M4180 Red	Laptops	Computers	Edo	South South	$4,569	$1,735.50	37.98%
No Discount	Proseware Projector 1080p DLP86 Black	Projectors & Screens	Computers	Edo	South South	$24,990	$10,745.40	43.00%
No Discount	Proseware Laptop15.4W M518 Black	Laptops	Computers	Edo	South South	$7,580	$4,955.90	65.38%
No Discount	Proseware Laptop15.4W M518 White	Laptops	Computers	Edo	South South	$7,580	$4,861.10	64.13%
No Discount	Simpson LCD17W E203 White	Monitors	Computers	Edo	South South	$1,290	$57.70	4.47%
No Discount	Kekule Projector 480p M480 Black	Projectors & Screens	Computers	Edo	South South	$2,290	$2,225.80	97.20%
No Discount	SV PCI Network Adapter E903 Silver	Computers Accessories	Computers	Edo	South South	$419.9	$400.70	95.43%
No Discount	Kekule USB Wave Multi-media Keyboard E280 Black	Computers Accessories	Computers	Edo	South South	$159	$156.10	98.18%
No Discount	Simpson Laptop15 M1501 Silver	Laptops	Computers	Edo	South South	$6,990	$3,151.90	45.09%
No Discount	Kekule Digital camera accessory kit M200 Grey	Computers Accessories	Computers	Edo	South South	$239	$38.30	16.03%
No Discount	Proseware Projector 1080p LCD86 Black	Projectors & Screens	Computers	Edo	South South	$22,950	$16,312.60	71.08%
No Discount	Fabrikam Laptop15 M5000 White	Laptops	Computers	Edo	South South	$6,990	$4,968.50	71.08%

Data visualization:

After cleaning and analyzing the data, it was finally time to create some visualizations from the insights discovered in the dataset, in order to make the gathered data easier for stakeholders to comprehend.



Stick around a little more and I'll show you how I built this amazing dashboard.

Data visualization Steps:

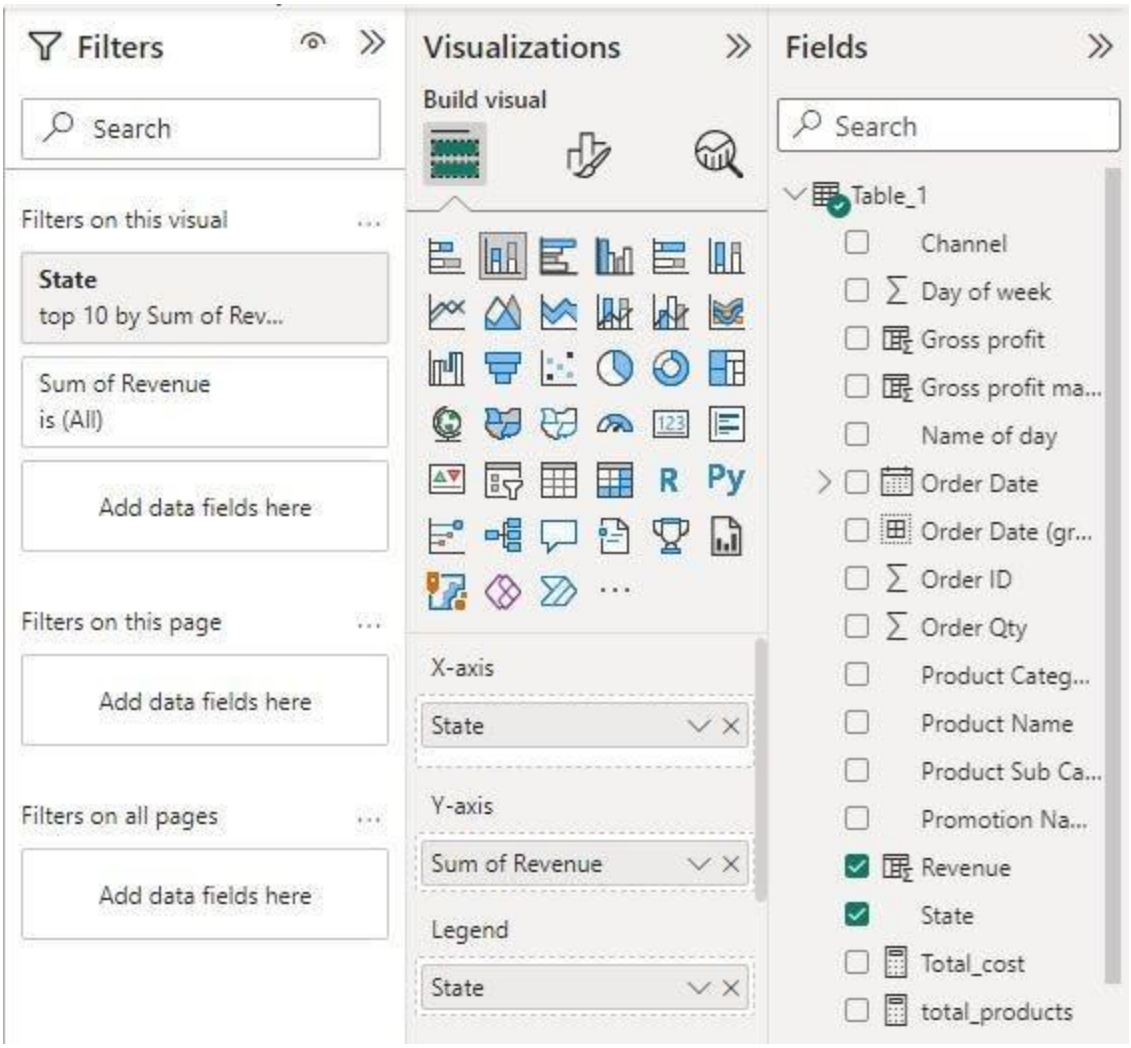
Visualizing data is important because it passes a comprehendible message across to stakeholders in the form of pictures; it helps them gain meaningful insight from vast amounts of data, "*a picture truly is worth more than a thousand words!*".

It is important to give some quality time studying the dataset to be able to generate meaningful insights from it.

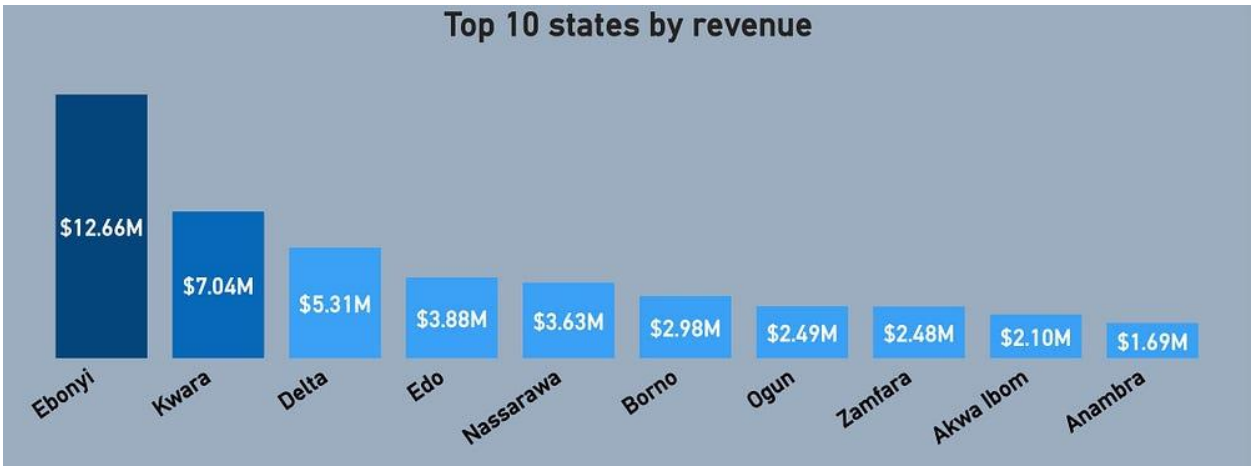
Here are some of the insights i got from the dataset:

1. **The revenue generated by stores in each state, i also narrowed it down to the top 10 states with the highest revenue:**

I created this visual by comparing each state (x-axis) by the total revenue generated by each state (y-axis), by simply checking each box in the fields pane, the intellisense of power BI allows it sometimes to automatically figure out what axis each field should belong to. On the visualization pane, I selected the '*stacked column chart'* to be used to visualize my comparism. Knowing what chart to use and when to use them is an important skillset every data analyst should have. Next, on the filters pane i filtered the *state* to only show the top 10 states with the highest revenue because there were 31 states in total and showing all of them was not relevant to my analysis dashboard.

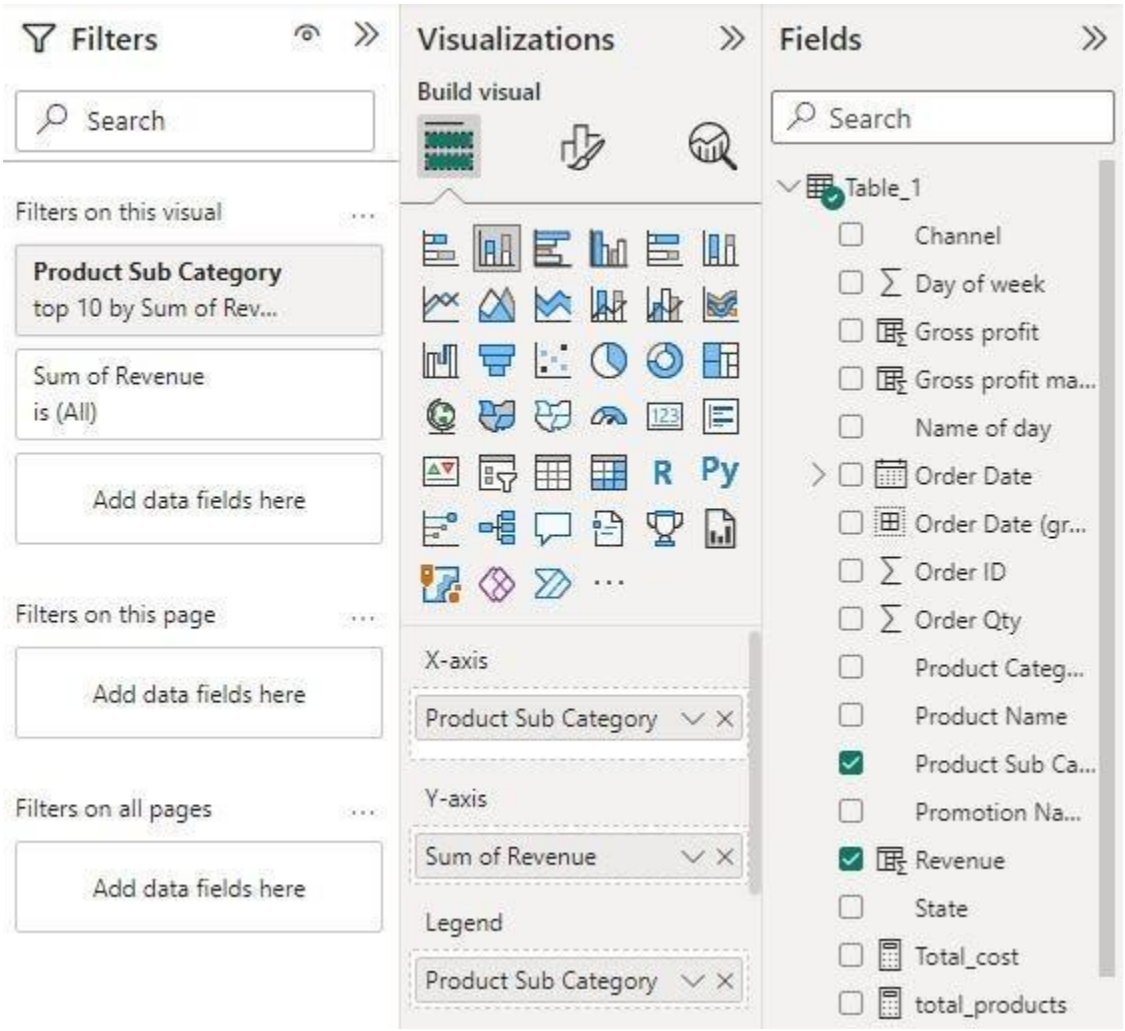


The result of my visualization:

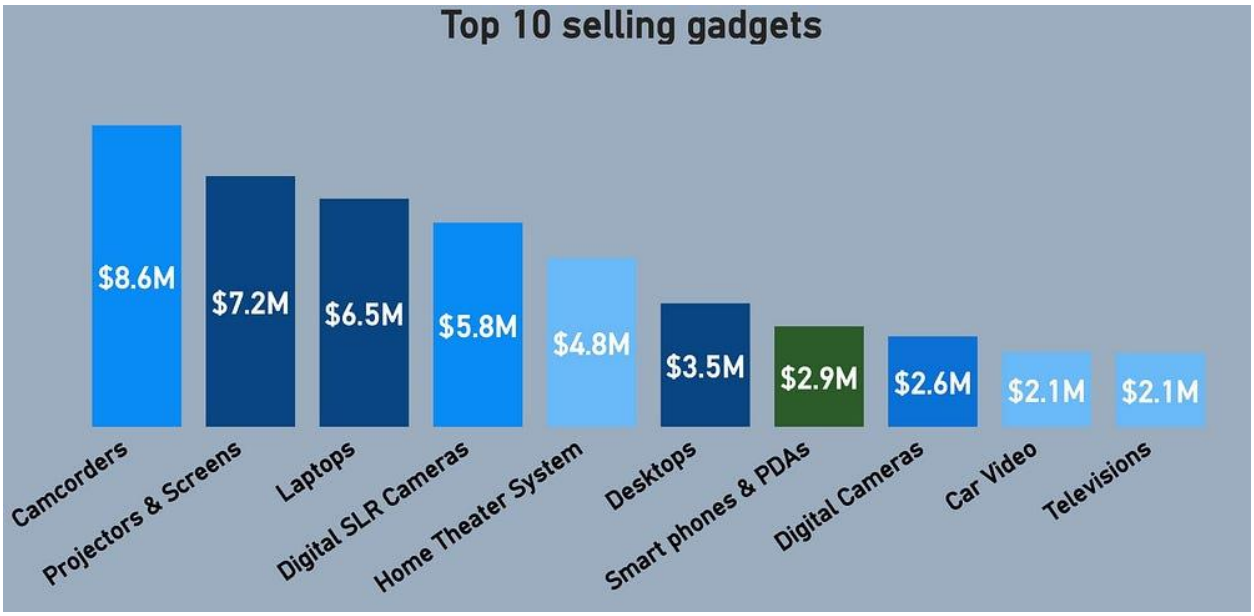


2. **The revenue generated by the top 10 highest selling gadgets (*product subcategory*):**

Similarly, the my first visual, I created this by comparing the product subcategory (x-axis) by the total revenue generated by each of them (y-axis). I also used a '*stacked column chart'* and i filtered the product subcategory to show the top 10 selling product subcategories by revenue generated.

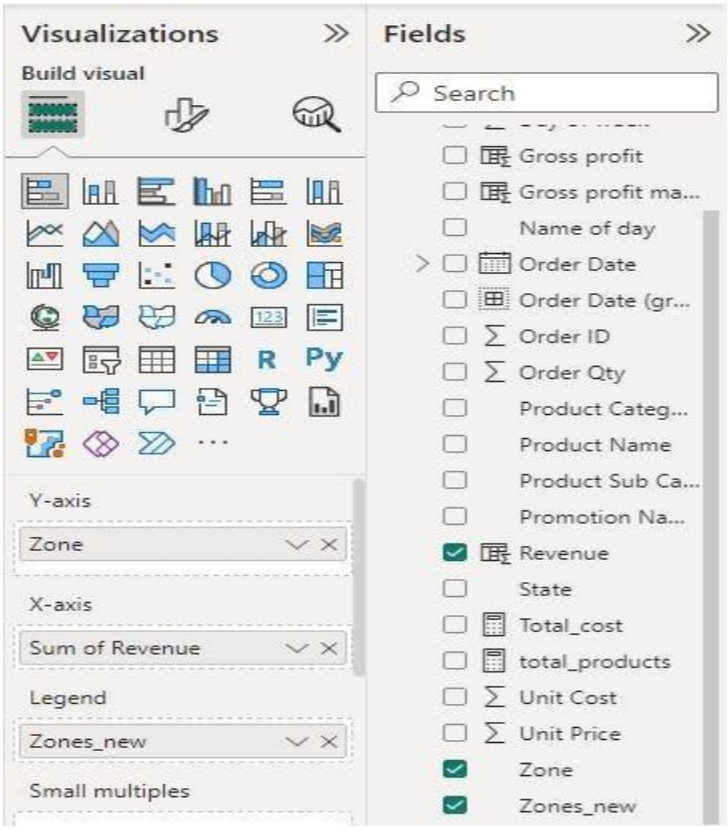


The result of my visualization:

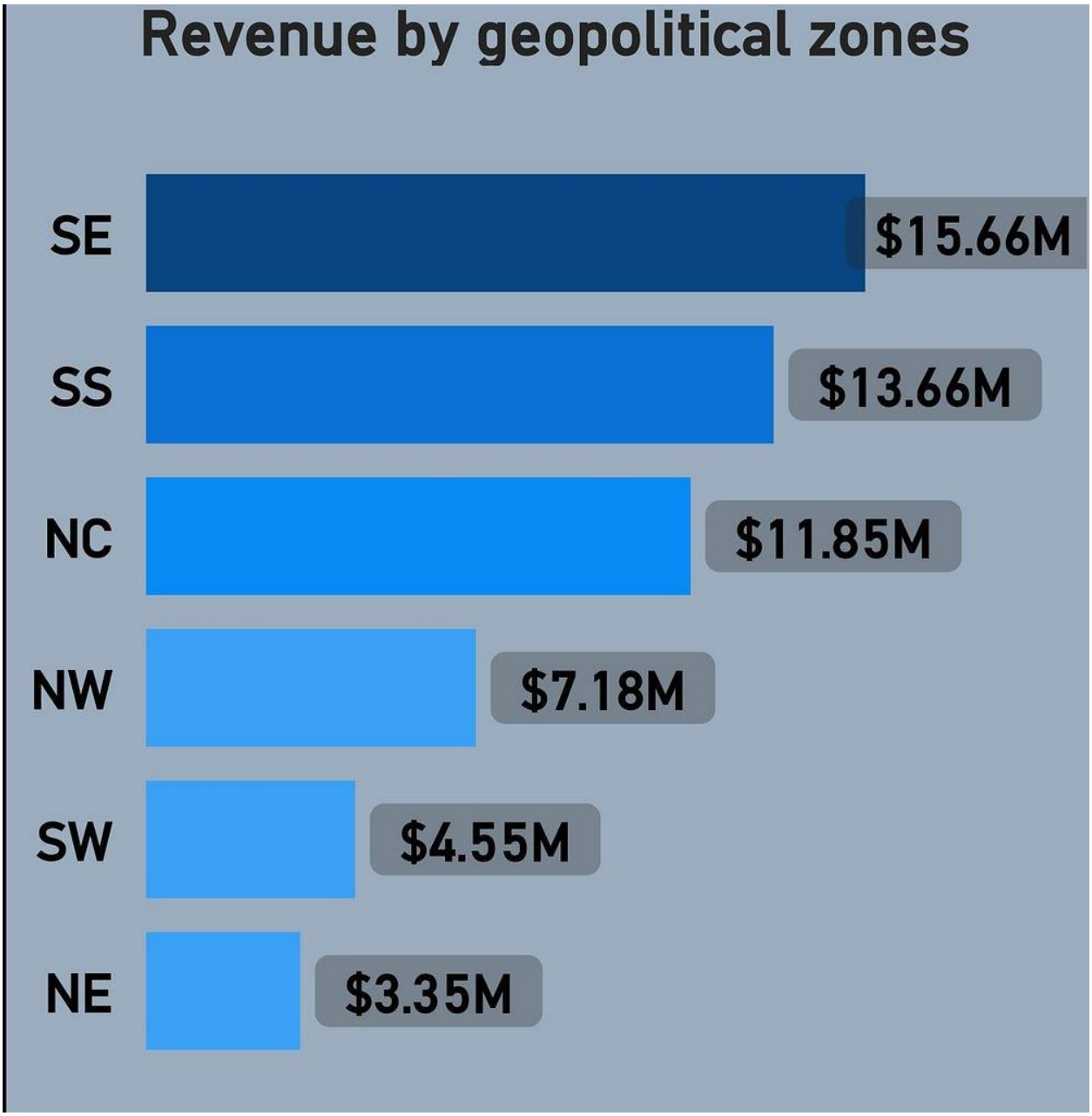


3. **The revenue generated by each geopolitical zones in Nigeria:**

I created this visual by comparing the geopolitical zones (y-axis) by the revenue generated by each (x-axis). A '*stacked bar chart'* was used for the visualization.



The result of my visualization:

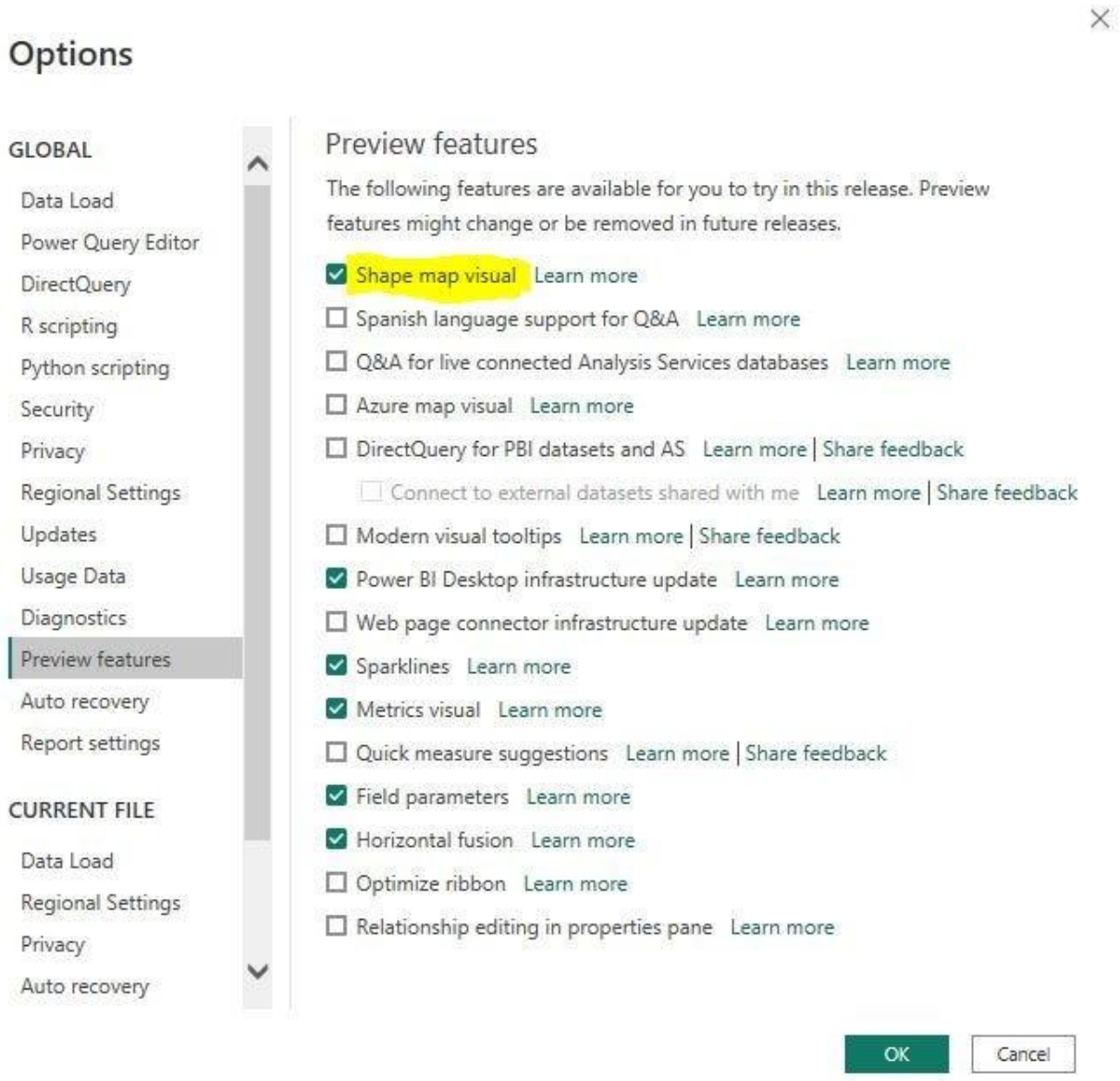


4. **The percentage density of revenue by states:**

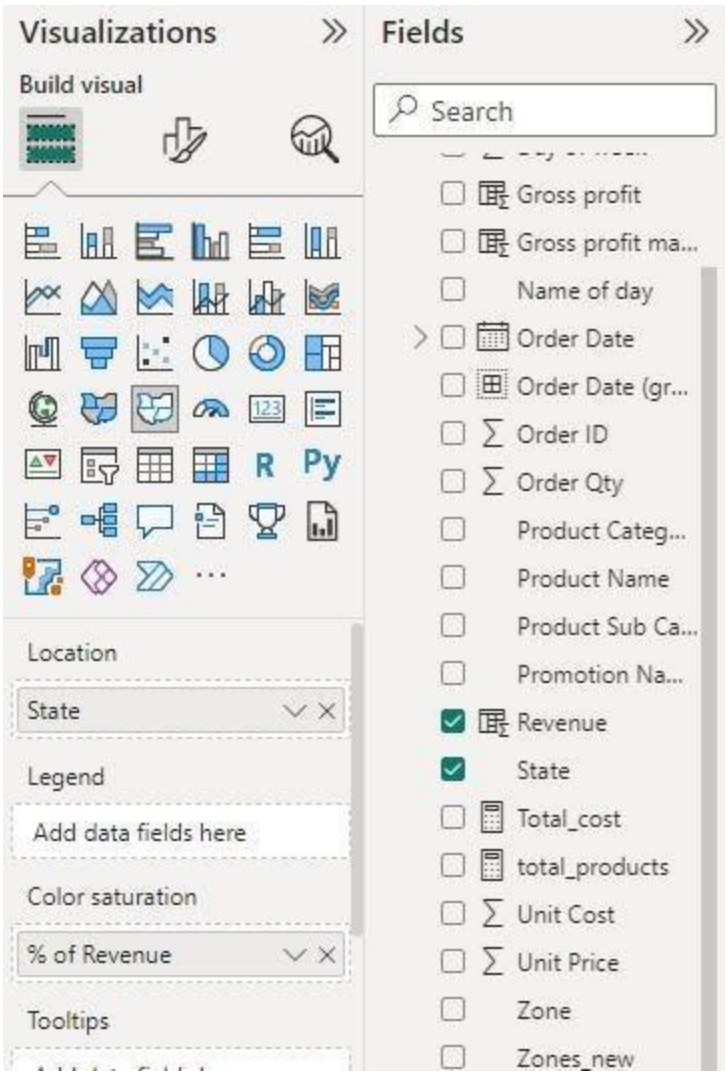
The percentage density shape map shows the concentration by states in terms of revenue generated, '*the higher the revenue the darker the gradient, the lower the revenue the lighter the gradient'.*

Shape maps are used to specifically narrow down the map of a country, region or state, rather than the entire continent.

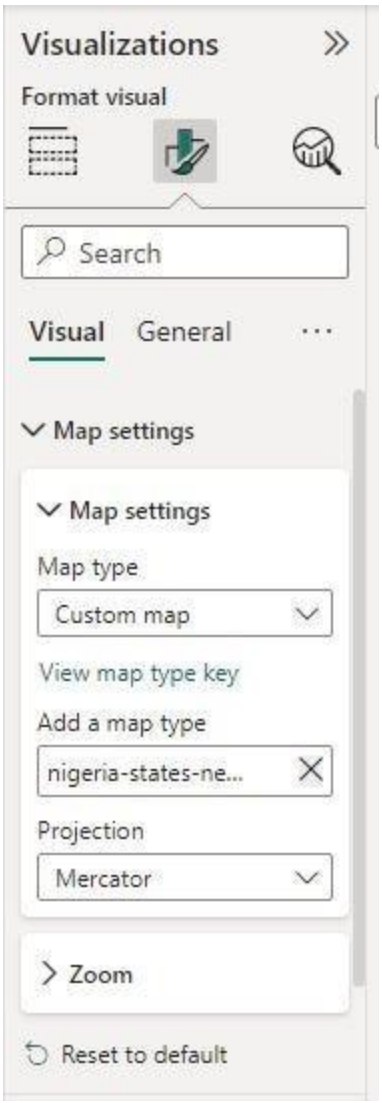
Creating a shape map takes a few extra steps because the shape map are not automatically enabled in your power BI features. In order to add the shape map, follow the steps: Sign-in to the power BI service, click file >> option and settings >> options >> preview features >> **enable** shape map visual >> **click OK**.



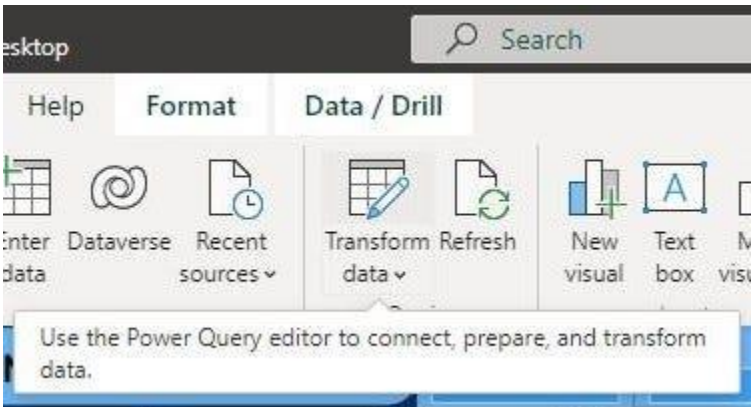
After the shape map was added to my visualization pane, I inserted the shape map into my dashboard and I used the state as **Location** and the revenue generated as the **saturation** to show the concentration of the states in terms of revenue *(in percentage)*.



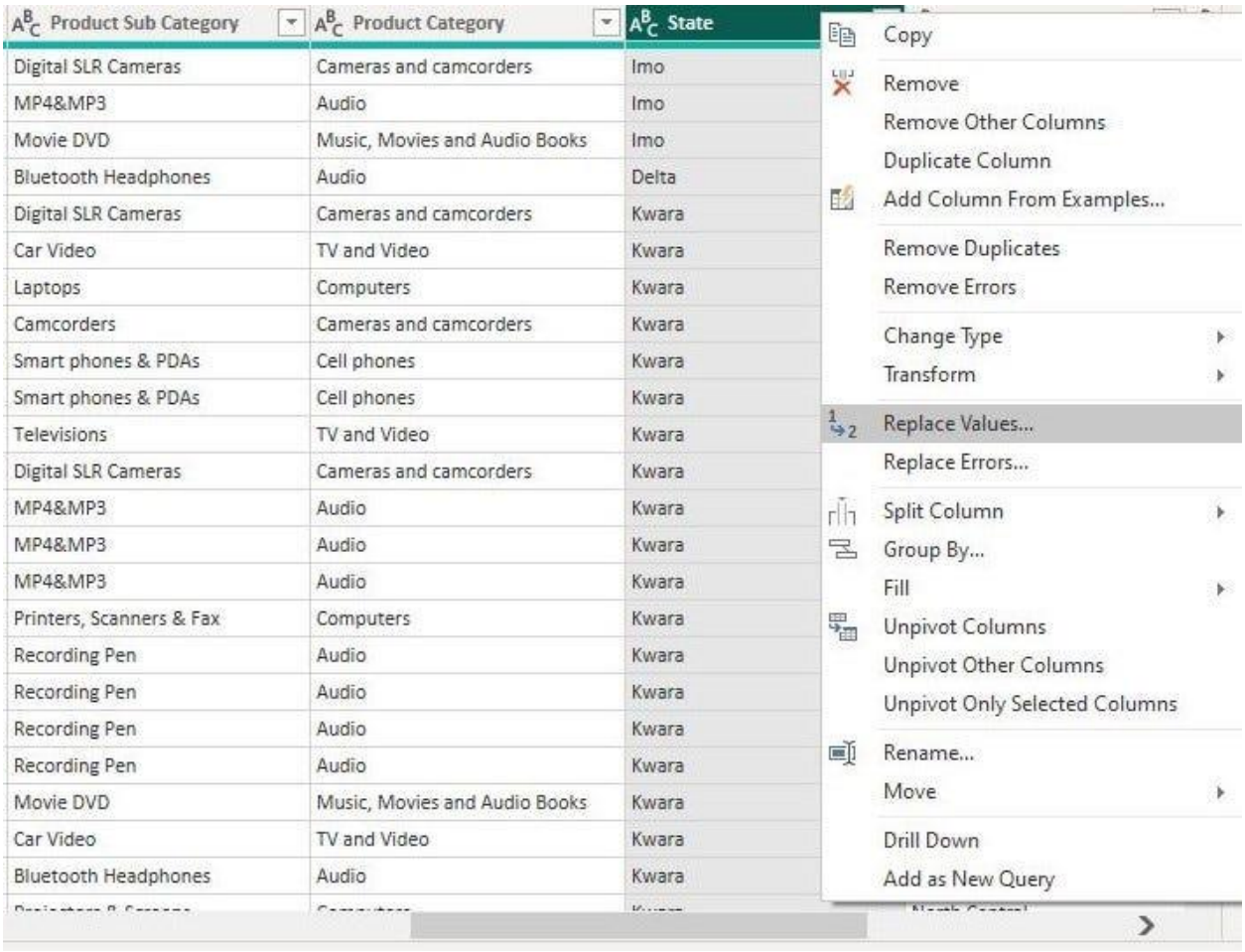
I noticed that after inserting the state as Location, power BI still gave me a map of America instead of a map of Nigeria, this was normal as power BI uses the map of America as the default map. In order to rectify this, I clicked on format visual >> Map settings >> clicked the '*map type*' dropdown and selected '*Custom map*' >> then I clicked *'add map type*' by browsing through my file directory >> selected a shape map (.JSON) file (which is a Nigerian states custom shape map in JSON format). Download the custom Nigerian states shape map file here. As soon as i added the custom shape map, the Nigerian map was inserted with all 36 states of Nigeria and the Federal Capital Territory.



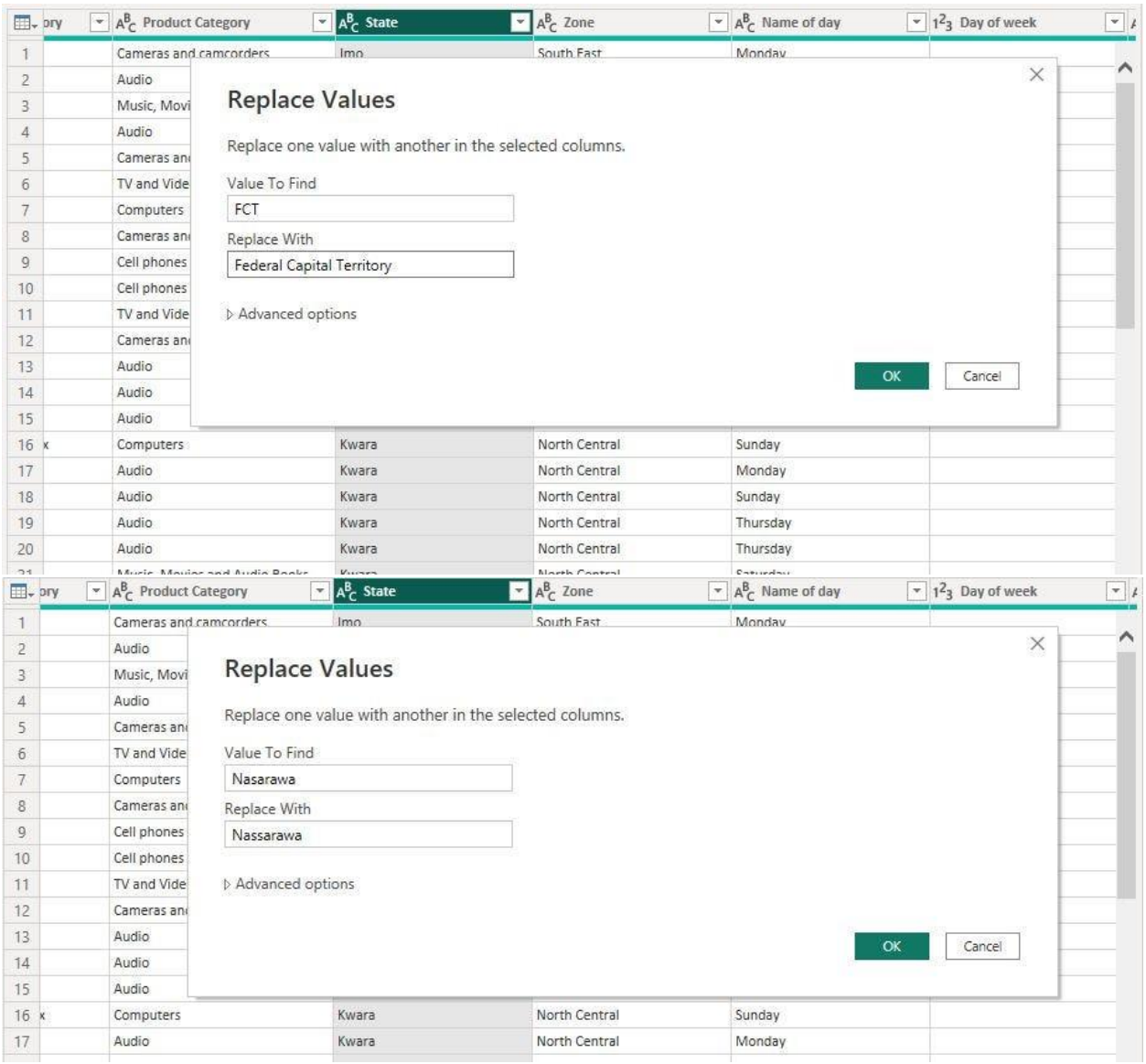
But there was a problem with my map, the FCT and Nasarawa state were both blank on the map even though they were a part of the 31 states in which the gadgets store were located. So I made some investigations and I found that 'FCT' was named as 'Federal Capital Territory' and Nasarawa was named 'Nassarawa' on the custom shape map file. In order to rectify this, I transformed my dataset to return back to power query to further clean the data.



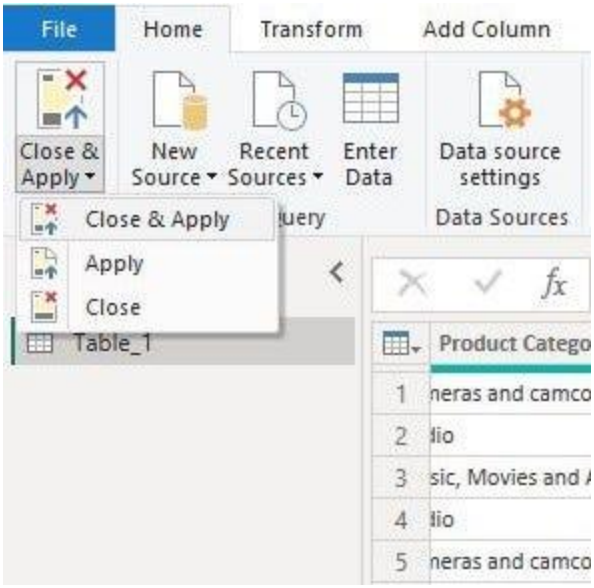
In the power query, I used ‘replaced values’ to find and replace the names the states.



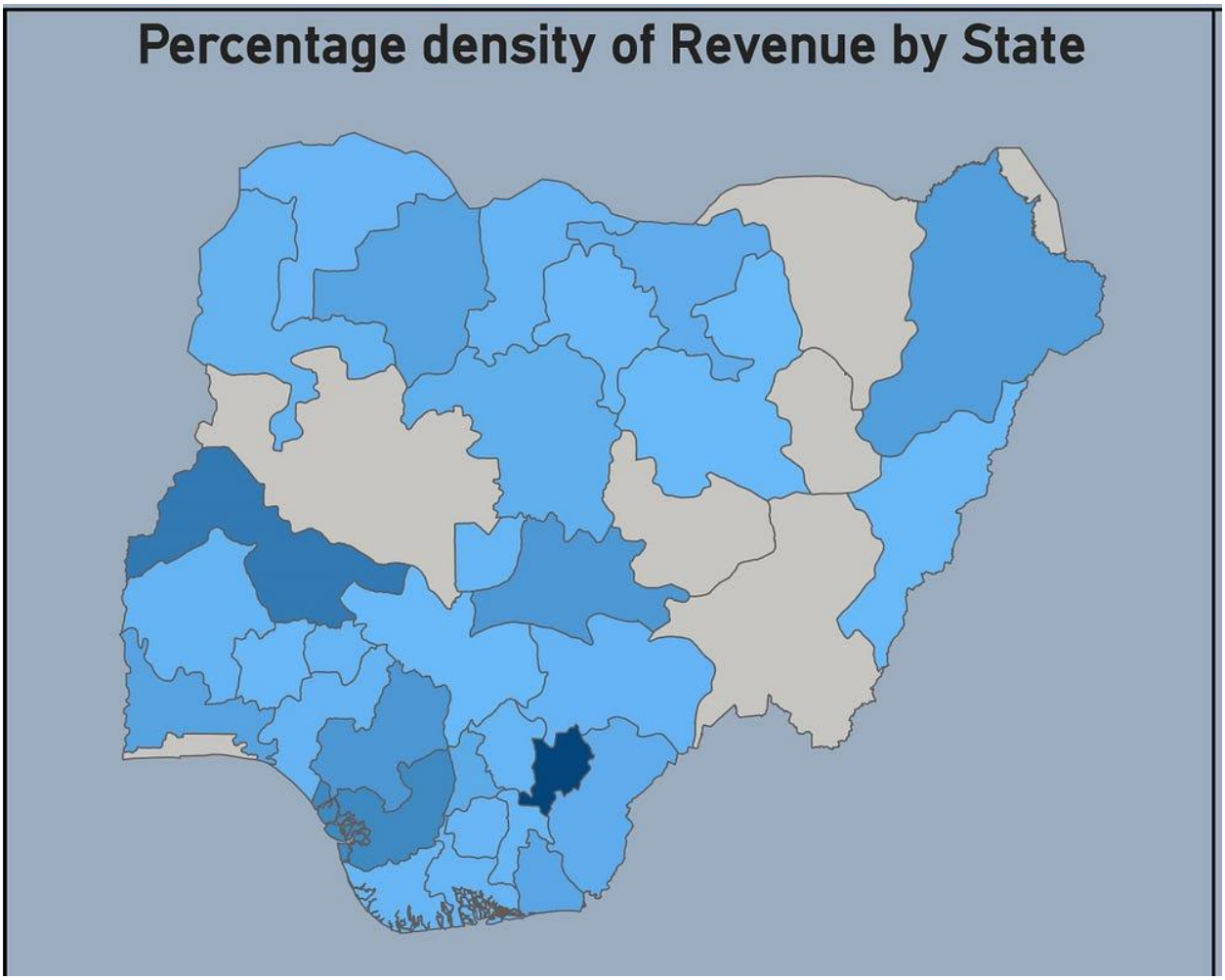
Then i replaced the names of the states with the corresponding names in the shape map file.



Finally, I clicked on ‘close and apply’ in order to return back to power BI. The changes were effected and all 31 Nigerian states were inserted into the custom shape map.

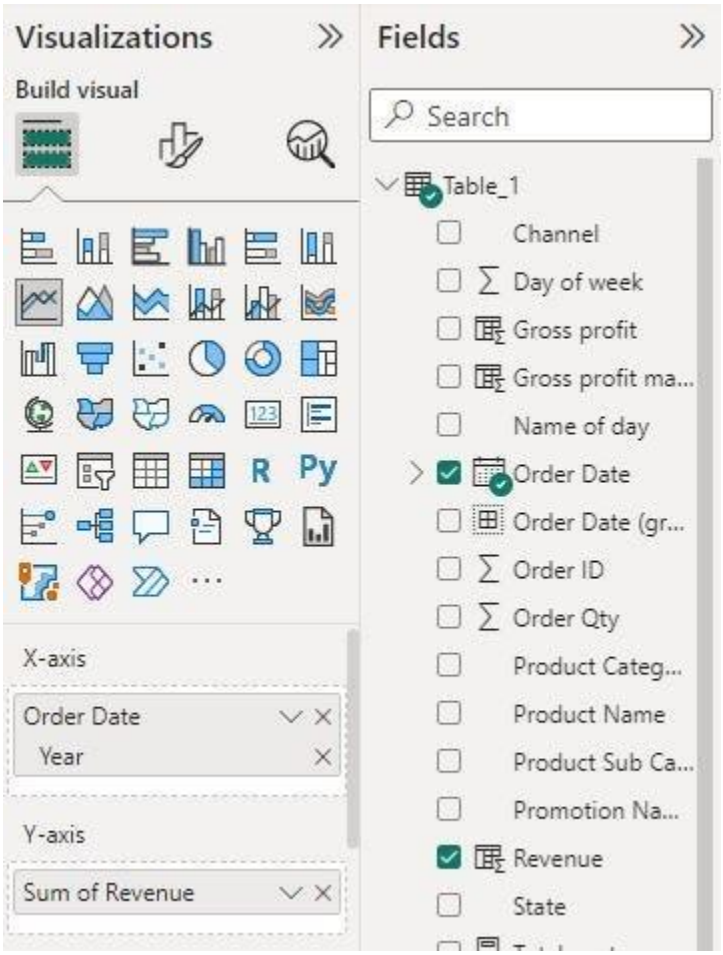


The result of my shape map visualization:

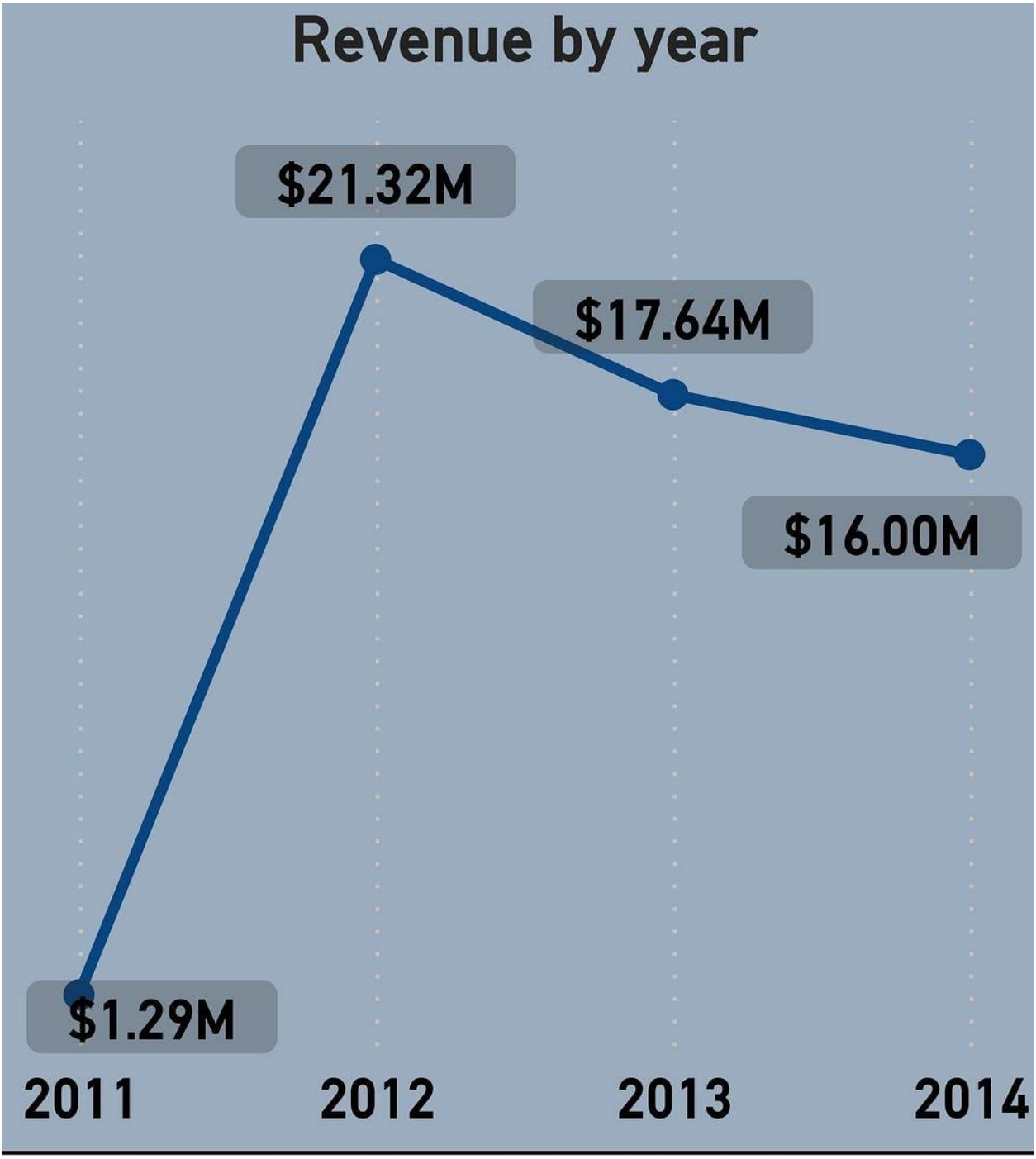


5. **The revenue generated in each year from 2011–2014:**

I created this visual by comparing the years (x-axis) by the revenue generated in each year (y-axis). A line chart was used to show this comparism, because line charts are the perfect visual chart to show changes over time.

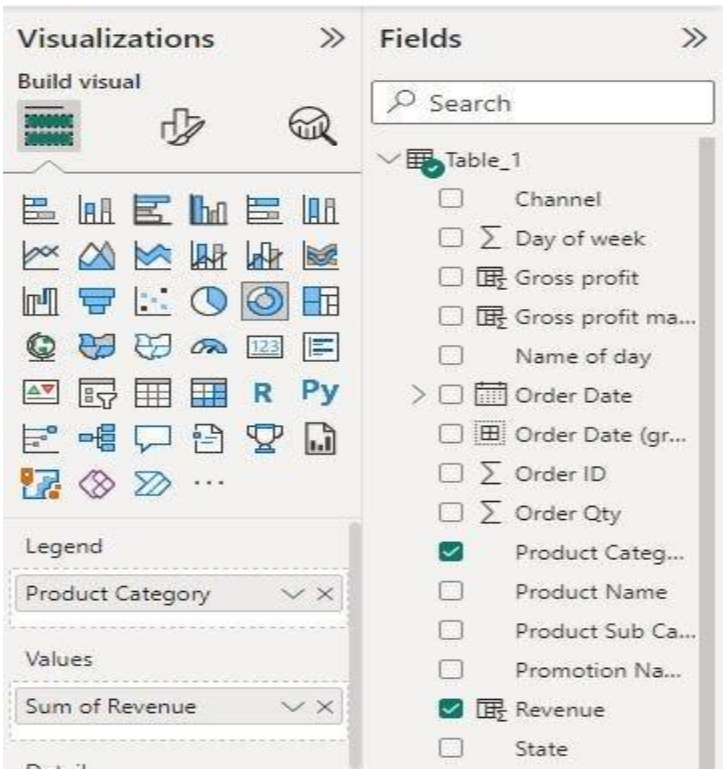


The result of the visualization:

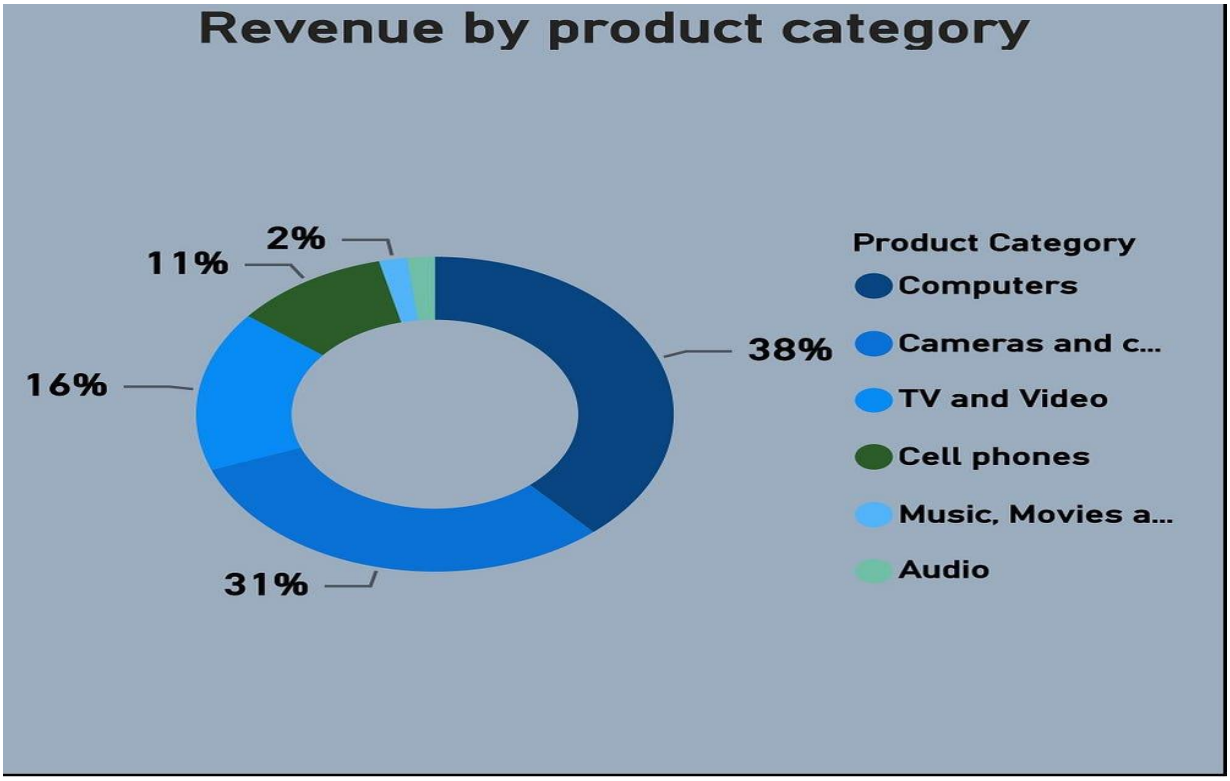


6. **The revenue generated by each product category:**

Finally, I created this visual using a doughnut chart to compare the product categories by the revenue generated by each product category. The doughnut was used because doughnut charts are perfect for percentage comparisms between multiple items, the doughnut and the pie chart perform the same function, but I chose to use the doughnut chart this time simply because of preference.



The result of my visualization:



SLICERS:

slicers enable a user to sort and filter a packed report and view only the information they want based on certain criteria. I added 4 slicers to filter my analysis based on the year, month, state, and geopolitical zone.

Slicers can be added from the visualization pane

CARDS:

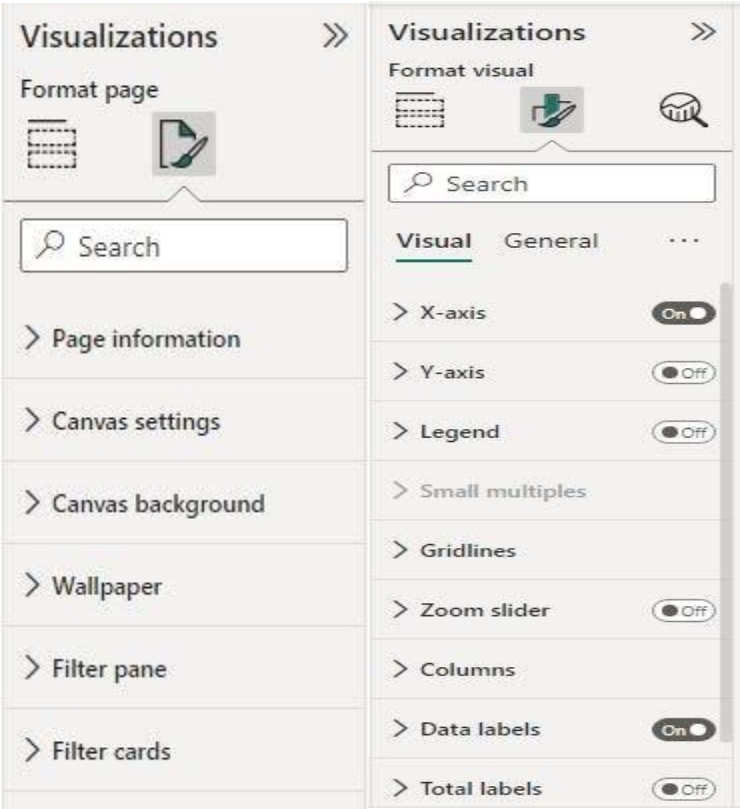
Cards are used to display KPIs *(key performance indicators)*, which are aggregated information of a single numeric measure value. Cards are useful because they key values that interest stakeholders upon first glance.

A few of these cards added shows the revenue and profit, these are key values that catch the attentions of stakeholders, wouldn't you agree?

Cards can be added from the visualization pane

USING FORMAT VISUAL:

A format visual pane contains settings to format the entire dashboard area and each individual charts.



Findings:

1. The gadget store has a total of 6 products, which are further divided into 22 sub-product categories for sale.
2. The gadget store is located in 31 states in Nigeria.
3. Ebonyi state in the south east region of Nigeria sold the most products thereby generating the highest revenue from 2011 to 2014.
4. Computers are the highest selling products with over 38% of the products ordered in the last 4 years, and the camcorders are the highest selling gadgets in the *cameras and camcorders* products sub-category.
5. The highest revenue was generated in 2012 having profit worth $12 million, with Ebonyi state selling over $4.7 million worth of products.
6. The total orders all over the country was 15,000 orders and having 251,000 quantities with those orders.
7. There was a decline in revenue generated from 2012 to 2013 and a further decline from 2013 to 2014. The decline was as a result of the reduction in total orders.

Recommendation:

1. With the progressive sales performance of the gadget stores, more stores should be created in the remaining 6 states in Nigeria, especially in Lagos state which is a cosmopolitan city.

CONCLUSION

Cleaning the gadget sales dataset was a bit challenging, but having a clean data is always encouraged in order to obtain accurate analysis and the best visualization possible.