**TECHNICAL REPORT**

**TITLED**

**GLOBAL SUPERSTORE ANALYSIS**

**BY**

**OBIWULU CHINYEAKA DAVID**

**Date: 10TH JULY 2024**

**Table of Contents**

* **Introduction**
* **Methodology**
* **Tools and Database**
* **Data Variables**
* **Questions and Exploratory Data Analysis**
* **Recommendations**
* **Conclusion**

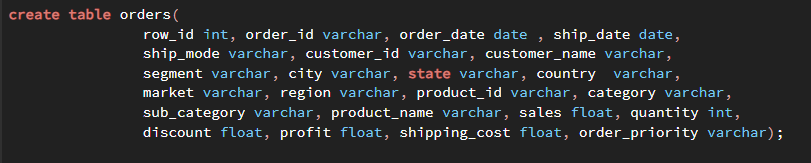
**Introduction**

Global Superstore is a global online retailer based in New York, boasting a broad product catalogue and aiming to be a one-stop-shop for its customers. The Global superstore’s clientele, hailing from 147 different countries, can browse through an endless offering with more than 10,000 products. This large selection comprises three main categories and subcategories as: office supplies (e.g., appliances, binders, arts, envelops, fasteners, labels, papers, storages, supplies, and staples,), furniture (e.g., chairs, bookcases, furnishing, and tables), and technology (e.g., accessories, copiers, machines, and smartphones)**.**

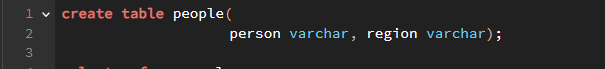
**Methodology**

This study employed a descriptive approach to analyze the global terrorism dataset because it particularly considered already generated dataset to solve prevailing business challenges. The methodology consisted of the following steps:

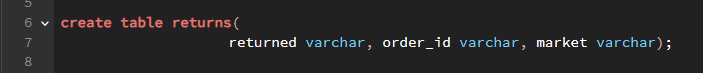
* **Data Collection:** The dataset was delivered by Vephla Group Limited and it contained three combined worksheets namely: Orders, People and Returns. These three-dataset contained:
  + The Order dataset and it consists of fifty-one thousand two hundred and ninety-one (51291) rows (incidents) and twenty-four (24) columns, spanning the year 2011–2014.
  + The People dataset contains only two (2) columns, and fourteen (14) rows.
  + The Returns dataset consists of three (3), and one thousand one hundred and seventy-four (1174) rows.
* **Data Cleaning:** The dataset was cleaned with Excel and was inspected for missing values, duplicates, word cases, strange characters such as apostrophes and quotation marks, spaces, removal of unwanted columns, and inconsistencies.
  + The data was cleaned using a Power Query.
  + Duplicates were removed by clicking on data in the task bar of the excel file and then click on remove duplicate in the ribbon of data, I made sure all the named columns were checked and then I clicked okay to proceed.
  + The data was crosschecked for inconsistency in word format and converted all the words to sentence case.
  + The trim command was used to remove all the extra spaces in the dataset to ensure proper analysis.
  + Find and replace was used to find strange characters and was replaced with blanks to ensure clean datasets.
  + The postal code column was removed at the end of the cleaning because it is not related to the analysis and there are several other alternatives like states and cities that could stand in if needed.
  + After cleaning, the data size was to 59291 fifty-nine thousand two hundred and ninety-one rows and twenty-three (23) columns.
  + The Data was formatted, separated and standardized for analysis.
* **Table creation and Data Importation:** After cleaning of the datasets, PostgreSQL was used to create table for each of the dataset before they were finally imported.
  + **Table Creation for Orders Dataset:**



* **Table Creation for Peoples Dataset:**



* **Table Creation for Returns Dataset:**



**Database Variables for Orders Dataset**

1. **Row ID:** This is the unique number of each row.

2. **Order ID:** This column represents the unique number given during each order.

3. **Order Date:** The date of the order.

4. **Ship Date:** The date when the order was shipped.

5. **Ship Mode:** The mode of carriages.

6. **Customer ID:** The unique identity of each customer.

7. **Customer Name:** The full names of the customers.

8. **Segment**: This column describes the part or department of the customer.

9. **City:** The city where the customers are located.

10. **State:** The state of the customers.

11. **Country:** Country name of the customers.

12. **Market:** Name of the market where the products are sold.

13. **Product ID:** Unique number of each product.

14. **Category:** This column describes the category where each product belongs.

15. **Sub-Category:** This column tells the sub-category of each product.

16. **Product Name:** Tells the name of the products.

17. **Sales:** Shows the amount each product was sold.

18. **Quantity:** Shows the quantity of each product that was sold.

19. **Discount:** Identifies the amount of discount that was given in each product sold.

20. **Profit:** Shows the profit made in the sales of each product.

21. **Shipping Cost:**  The cost of shipping and deliveries.

22. **Order Priority:** Shows the level of urgency of the order.

**Database Variables for Peoples Dataset**

1. **Persons:** Identifies the name of the sales person.

2. **Region:** It highlights the region of the agents.

**Database Variables for Returns Dataset**

**1. Returned:** This column indicates whether the product was returned or not.

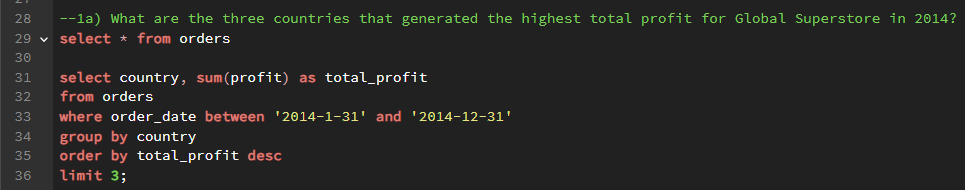
**2. Order ID:** This is the unique number of each order.

**3. Market:** This is the name of the market where the products are sold.

**Questions and Exploratory Data Analysis**

This segment shows the SQL codes and the charts that was used to provide insights to the answered questions.

* **What are the three countries that generated the highest total profit for the Global Superstore in 2014:** This question seeks to know the best three countries in terms of generated profit in the year 2014. I selected all the countries and summed up their profits in 2014 then limited the result by 3 from descending order just to find out who performed best from top to bottom.

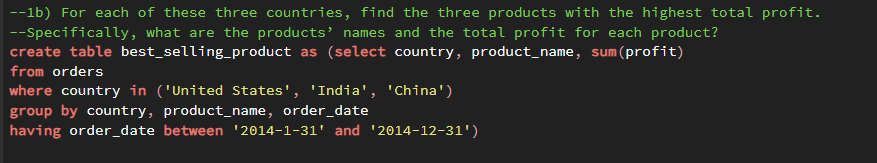


**The resulting visual of the above query is represented below:**

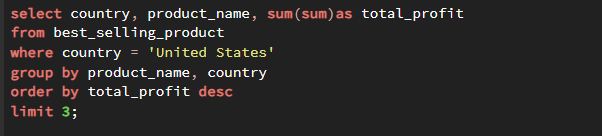
1a. The above visual showcased the insight that the United States is the country with the most profit of eighty-six thousand six hundred and seven US dollars while India and China came second and third with forty-six thousand five hundred and eighteen US dollars and forty-three thousand six hundred and twenty US dollars respectively.

* **For each of these three countries (USA, INDIA, and CHINA), find the three products with the highest total profit. Specifically show the product names, and the total profit of each profit:**

To answer these questions correctly, I have to firstly break down the complexity of the question by writing a query that will separate all three countries and the total generated profits in the stated year (2014). The successful query was then kept in a table named ‘best-selling product.’

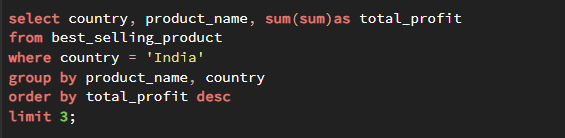


* From the best-selling product table, the following queries was written to answer the above questions and the visuals are represented below for insights:



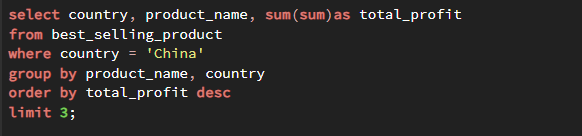
1bi. The above insight shows that in the USA, Canon image CLASS 2200 Advanced Copier machine is the best-selling product in 2014 and it made a total profit of fifteen thousand six hundred and eighty US dollars in a calendar year. Also, Hewlett Packard LaserJet 3310 Copier and GBC DocuBind TL300 Electric Binding System made the second and third best-selling products with three thousand six hundred and twenty-four US dollars, and one thousand nine hundred and eleven US dollars respectively.

* This is a query written to show the best three products sold in India as it relates to their total sum of profit made in a calendar year. The next image following the query is the result of the SQL query.



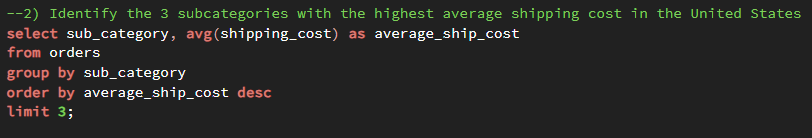
1bii. The below image shows that in India, Sauder Classic Bookcase, Traditional and Cisco Smart Phone, with Caller ID and Hamilton Beach Refrigerator, Red are the best-selling products in 2014. Furthermore, Sauder made the highest profit of two thousand four hundred and twenty US dollars while Cisco and Hamilton Beach Refrigerator made second and third with one thousand six hundred and nine US dollars and one thousand four hundred and forty US dollars respectively.

* The below PostgreSQL code called out China and the sum of profit made for the best three sold products in 2014 and the next image is the result of the query.



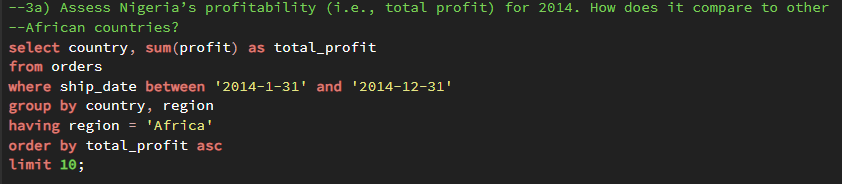
1biii. The insight shown below reveals that the best three selling products in China in the year 2014 are: Bush Classic Bookcase, Mobile and HP Copy Machine, Colour and Sauder Classic Bookcase, Metal. Sauder Classic is the Top one best-selling with one thousand four hundred and sixty-three US dollar, while HP Copy and Bush Classic are the second and third respectively.

* **Identify the 3 subcategories with highest average shipping cost in the United States:** This query was written to pull out best three product subcategories in the United States. The aggregate function of average was used thereby the need to group became inevitable. The alias function was also used in place of AS to rename ‘avg(shipping\_cost)’ to Average ship cost. Further visual showcases the derived insight and result of the written query.



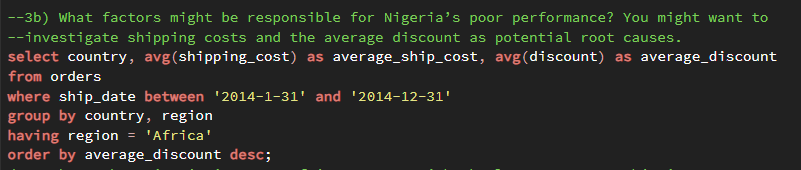
2. The below image is a resulting insight derived from the above query and it explains that the three products with the highest average shipping cost in the USA are Bookcases, Copiers, and Tables. Tables top with chart with ninety-two US dollars and seventy-five cents while Copiers and Bookcases and second and third with seventy-one US dollars and seventy-five cents, and sixty-four US dollars and forty-nine cents respectively.

**3a. Assess Nigeria’s profitability (i.e total profit) for 2014. How does it compare to other African countries?** The below image carries a PostgreSQL query that out list African countries and the sum of profits they made in the business year of 2014. The limit 10 and the order by ascending statement was used to reduce the countries to 10 from the lowest profit to the highest profit. The next image is an insight that revealed the answer to the query.



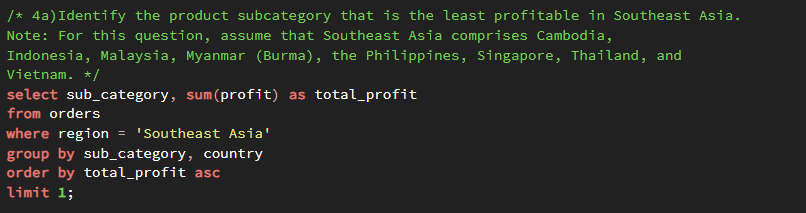
This segment is represented by the scattered plot chart and it reveals that Nigeria made a loss and is the highest sum of loss of $(22,328.03) in deficit for the business year of 2014.

**3b. What factors might be responsible for Nigeria’s poor performance? You might want to investigate shipping cost and the average discount as potential root causes:**



The above image shows that Nigeria is the top 3 African country with the highest average discount which would only lead to reduction in average profit and thereby lead to the poor performance at the end of a business year. It was also observed that Nigeria has is the least 6 average shipping cost, this implies that the shipping process to other countries is relatively cheap when it should be relatively high so they can make more cumulative profit from it. This could be due to poor exchange rate and inflation in Nigeria market today.

**4a. Identify the product subcategory that is the least profitable in Southeast Asia.** This image has shown queries written to ascertain the product that has the lowest sum of profit in Southeast Asia. I went further to limit it to 1 because I seek to establish only the least product. This analysis was presented with a card in the dashboard/ visualization.



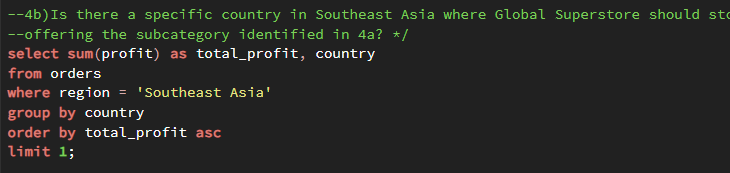
The below card shows that Table is the product with the least profit of $(10,680.28)

Least profitable PRODUCT in ASIA?

**TABLE: $ (10,680.28)**

**4b. Is there a specific country in the Southeast Asia where Global Superstore should stop offering the product subcategory in 4a (Tables)?**

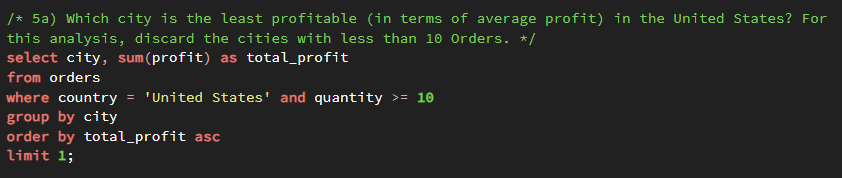
The below query answered the question of which country should stop selling tables and the answer is Indonesia because it made the highest loss in Southeast Asia. Tables made a loss of $(10,680.28) which is more than every of country in Southeast Asia and the query was represented with a card just below the query.



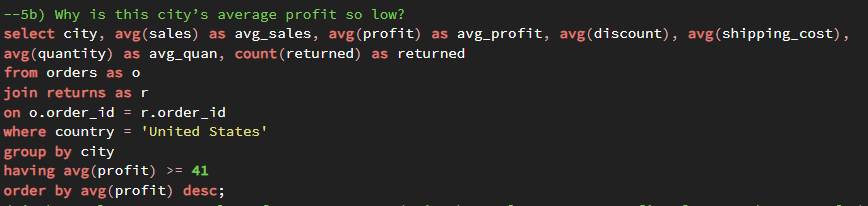
Least profitable COUNTRY in ASIA?

**INDONESIA: $ (10,680.28)**

**5a. Which city is the least profitable (in terms of average profit) in the United States?** The image below is a query that shows the least profitable city in the United States. The sum aggregate function was used therefore the need to group with city became necessary. The limit function was used to reduce the number of the cities to the least one by using ascending order. It was confirmed that Concord city has the highest average loss of $(1,862.3), which is the least profit of all cities.

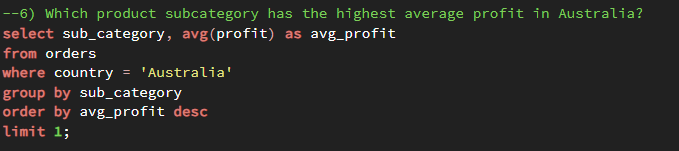


**5b. Why is this city’s (Concord) average profit so low?**

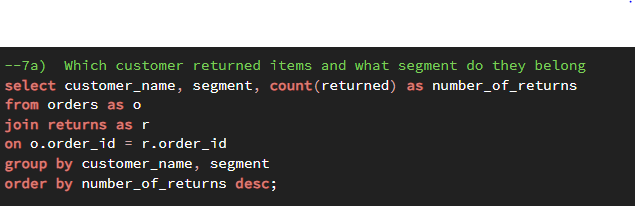


The above image is a query statement that shows the reason why Concord city has a low average profit in USA. It shows that Concord has a low average sale of $225.98. Also, Concord City has a low average profit of 41.742and this is as a result of the high average shipping cost, and a high average discount charges amongst other economical, and political factors.

**6. Which product subcategory has the highest average profit in Australia?** This query shows that appliances have the most average profit of $139 in Australia and the query looked at subcategory, along with average profit. I went ahead to filter further for only Australia as a country.

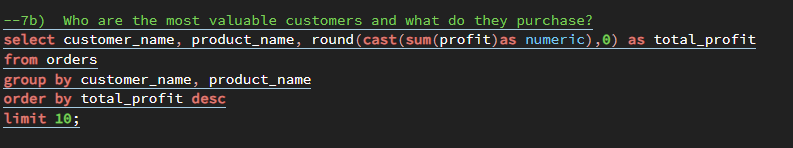


**7a. Which customer returned items and what segment do they belong?** This question seeks to make know the customers that made the most returns and what they returned. I selected customer name and the segment to identify the customers and the segments where the goods where returned. I used the join syntax to combine the returns table and orders table that way I was able to add count of returned to the select option. The linking column is order id since it is common in both datasets, I group by customer name and segment and ordered by number of returns from the highest to the lowest and the result is the chart below.



This visual shows that Ted Butterfield returned items 31 times in the consumer segment which is the highest number of returns. Gray Hwang returned 20 times and in the consumer segment while Nat Gilpin and Steven Ward returned 19 and 18 in the corporate segment respectively. Barry Franz returned 15 items in home office segment while Tom Boeckenhauer returned 15 items in the consumer segment which are the least.

**7b. Who are the most valuable customers and what did they purchase?** I took a twist to this final question by selecting customer name. product name and rounded up the sum of profit to 0 thereby making the figures whole numbers. I grouped by the none aggregated selected column and ordered by total profit in descending order. I went ahead to limit by 10 which means I want the top 10 valuable product customers and the products they have purchases more.



The below image is an insight that answered the above question and above query. It is clear that Canon image Class 2200 Advanced copier made the highest profit of $8,400 which means it is the most valuable product and Tamara Chand is the most valuable customer. Raymond Buch is the second most valuable customer with $6,720 profit made from his purchase of Canon ImageClass 2200 Advanced copier. Cynthia Arntzen is the tenth valuable customer and he purchased Apple Smart Phone with Caller ID with $2,818 profit made from it.

**Recommendation**

* I recommend that low profitable countries especially Nigeria and other African countries should take a critical look into the operation schedules of the high profitable countries such as the USA, India, and China. Also, South Africa being the most profitable in Africa continent should be considered as a case study for other African countries.
* Less profitable products should be advertised more and more awareness should be created to ensure it gain the market in the coming business yet. Additionally, scheme such as enviable discounts for the purchase of these low profit product in bulk should be enacted in other to boast sales.
* Considering that Nigeria had a high average shipping cost which has led to a very low profit in the previous business year, the government could come up with strategies and macro-economical agreement that would enable factories to be built in Nigeria or seek avenue that would scrub the shipping cost to the barest minimum. Youths should also be empowered and home-made products should be patronized to encourage skill development and affordable and profitable commodities.
* The challenge of exchange rates in Nigeria as well as other African countries with has led to either a loss or little yearly profit should also be looked into.
* Items in the consumer segment seem to have more returns than any other segment and this could be as a result of several reasons. I recommend that quality control should inspect products in the consumers segments to ensure it meets consumption standard. Also, reasons as to why these products are returned should be properly documented; this will help to improve the general customer service and satisfaction and at the long run, create loyal customers and increase profit at the end of every business year.

**Conclusion**

Nigeria has low average profit in 2014 but can greatly improve in the next business year if the shipping cost can be reduced and irrelevant discounts be cut down. Products that are highly profitable should be push forward for more profits while measures such as advertisement and product awareness should be considered in other to improve the profitability of the low profit products.

This analysis has contributed to my understanding of business analytics and reporting, highlighting the need for sustainable business to combat the challenges that lead to losses in business and ensure continuous profitability and customer satisfaction. By leveraging data-driven insights, it is clear that effective strategies can lead to proper decisions making that will guarantee the increase of profit in businesses nationwide.