**[TATA-Data-Visualisation-Empowering-Business-with-Effective-Insights](https://github.com/jagadishmali567/TATA-Data-Visualisation-Empowering-Business-with-Effective-Insights)**

**TASK 1** **(Framing the business scenario)**

**Concerns the CEO may have**

**1. Which region is producing the most profit, and which region is producing the least?**

• Given that it is dependent on sales, the primary source of money for the company, this question is significant to the CEO.

• The CEO must consider revenue analysis since senior executives are constantly thinking about how to boost profitability.

• Here, the CEO is curious about the viewing income by area to determine which locations are producing the most and which regions are producing the least.

• The CEO will be able to make decisions about how to increase revenue generation in the locations that are already producing the most using the data and research.

• The CEO will research the reasons why there aren't enough sales in the underperforming regions before attempting to modify the items and make them more appropriate for those areas.

**2. What is the trend for revenue on a monthly basis, and which months have seen the highest rise or fall in revenue?**

• The CEO will gain insight into the revenue's monthly pattern and how it changes over time.

• The CEO will then be able to examine the effects that internal corporate changes have had on sales.

• For instance, explain how the entrance of a new area or the debut of a new product increased sales for the online business throughout the month.

• The CEO might also examine whether there have been any internal delays that would have contributed to a possible decline.

• Senior management must do such analyses in order to plan forward and attempt to provide customers with the best possible experience.

**3. What quarters had the most revenue? Are sales impacted by the seasons?**

• Due of seasonality, there will always be months in the retail industry with higher demand.

• There will be instances where the data will undergo recurring, predictable modifications every year.

• Such seasonal months would need to be identified because the CEO would want to come up with a plan that would make the most of the months with higher demands.

**4. What percentage of overall income are the top customers responsible for? Are these customers essential to the company's success, or is the customers more diverse?**

• This research is essential since it will help the CEO determine what factors are most responsible for the overall income.

• It would be possible to determine which consumers contribute the most to sales by looking at the top customers of the retail establishment.

• The shop may then come up with a plan where the top customers can be targeted with additional goods they can purchase.

• Due to the fact that these customers are the store's top purchasers, this will guarantee more income for the business.

• A firm may benefit from having fewer consumers buy in larger quantities, but there may also be a negative aspect.

• Because these consumers make up the majority of the store's income and have the ability to bargain for lower pricing, retailers would have less negotiating leverage with them.

• The CEO must be informed about the customer diversity so that he may make advance plans.

• The strategy would be to broaden the customer base and focus on additional consumers who would boost sales in situations when the company is heavily dependent on a small number of customers.

**Concerns the CMO may have**

**5. How many customers make the same purchases again and over again? Do they place similar orders or do they place distinct orders?**

• The CMO's interest in observing the patterns in customer orders is evident from this query.

• The CMO is curious in the percentage of overall consumers that place repeat orders with them.

• The CMO will better understand the proportion of repeat consumers with the aid of this data.

• The study will also be performed to determine what they are purchasing second time.

• This will provide the CMO trends on which items and related products are in demand, allowing them to create a marketing plan to better target these customers.

**6. How long does it take for returning customers to place their next purchase after receiving the first one?**

• The CMO will be able to determine order frequency with the use of this study.

• This would entail figuring out how long it takes for customers to place new orders at the shop.

• It is assumed that customers who have just made a purchase will be thinking about the product and will likely buy or use it again in the future.

• The CMO may develop a strategy to persuade recent consumers to return and spend more after the data from the analysis is collected.

• It is possible to make an attempt to remind consumers who haven't made purchases from the shop in a while that it has been a while since their previous transaction.

**7. Which consumers have placed many orders and how much profit is generated by them?**

• How much a consumer spends at the store to buy goods determines the revenue for the business.

• Because of this, analysis must be performed to ascertain how much money is coming from the store's frequent consumers.

• The CMO can come up with a plan to increase recurring business from consumers who spend more on their purchases.

• A consumer should be encouraged to return to the store if they made a significant purchase there the first time, it is also crucial to remark.

• A marketing plan will guarantee that the shop will continue to earn more money from the high-paying consumers in the future.

**8. Who are the consumers who have returned the most frequently? What percentage of the revenue do they contribute?**

• Assessing which customers are returning the most frequently and how much they are contributing to the business is also crucial.

• Customers could require the same things on a weekly or monthly basis, but they don't have much financial worth.

• As a result, these consumers' contribution to revenue will be minimal.

• On the other hand, some customers could place orders twice a year and generate a lot of income.

• Because many customers only make purchases during specific months, management must make sure there are enough supplies on hand to fulfil their demands.

• More discounts must be offered to customers with big order quantities but low income so they may purchase in bulk and generate more sales.

**TASK 2 (DATA CLEANUP PROCESS)**

**1.REMOVING DUPLICATES**

-On spreadsheet, I found the duplicates of all columns and removed them.

Then I used **RStudio** to do the rest of the data cleaning as well as analysis, as its too large for spreadsheet to process and perform calculations. I also used RStudio due to the efficiency and accessibility of the program, its ability to handle huge amount of data and being able to create the visualizations.

### **2.Installing packages and library**

install.packages("tidyverse")

install.packages(“dplyr”)

install.packages("ggplot2")

install.packages("lubridate")

install.packages("janitor")

install.packages("ggpubr")

install.packages("skimr")

install.packages("here")

install.packages("ggrepel")

install.packages(“readr”)

library(tidyverse)

library(dplyr)

library(ggplot2)

library(lubridate)

library(janitor)

library(ggpubr)

library(skimr)

library(here)

library(ggrepel)

library(readr)

### **3.Importing and preparing the data**

online\_retail <- read.csv("C:/Users/Computer/Downloads/Online Retail Data Set.csv")

-I checked for duplicated rows in every table, just in case.

sum(duplicated(`online\_retail`))

it’s 0, therefore no duplicates found.

-Then I counted the number of rows with quantities less than 1. Then proceeded to exclude those rows.

online\_retail %>% filter(Quantity<1) %>% count()

online\_retail<-online\_retail %>% filter(Quantity>=1)

-Then I counted the number of rows with Unit Price<0. Then proceeded to exclude those rows.

online\_retail %>% filter(UnitPrice<0) %>% count()

online\_retail<-online\_retail %>% filter(UnitPrice>=0)

-I removed rows with NA values in every rows for accuracy and consistency.

online\_retail<-online\_retail %>% drop\_na()

-I cleaned and formatted the column names as well as rows to make sure of its accuracy and consistency.

clean\_names(online\_retail)

online\_retail<-rename\_with(online\_retail,tolower)

online\_retail <- online\_retail %>%

mutate(across(where(is.character), toupper))

I renamed the column unitprice for better understanding.

online\_retail <- online\_retail %>%

rename(unitprice\_dollars = unitprice)

I formatted the invoicedate column from “chr” to “datetime” format.

online\_retail<-online\_retail %>% mutate(invoicedate=ymd\_hms(invoicedate))

-I extracted the month from the invoicedate column.

online\_retail<-online\_retail %>% mutate(month=month(invoicedate,label=TRUE,abbr=TRUE))

-I extracted the year from the invoicedate column.

online\_retail<-online\_retail %>% mutate(year=year(invoicedate))

-Then I downloaded the xlsx file format of the cleaned data from RStudio into desktop to work in tableau for data vizualisation.

install.packages("openxlsx")

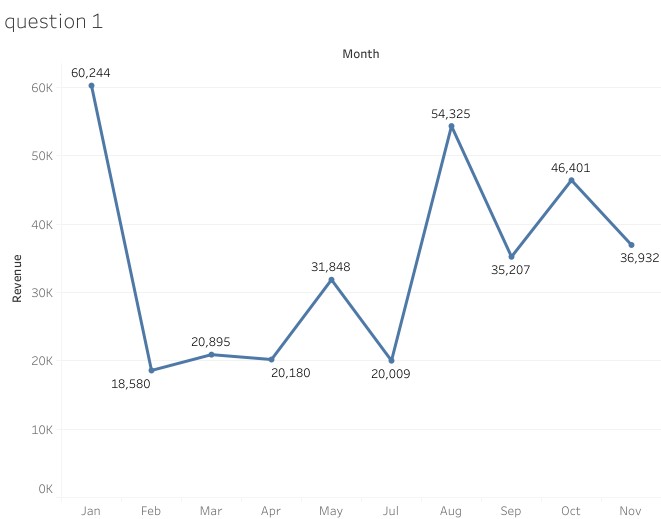
library(openxlsx)

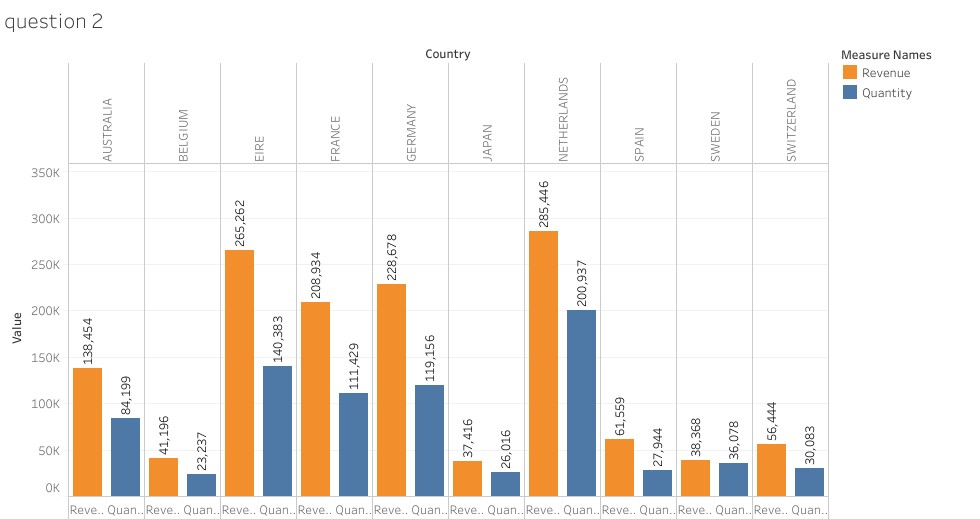
openxlsx::write.xlsx(online\_retail, file = "online\_retail.xlsx")

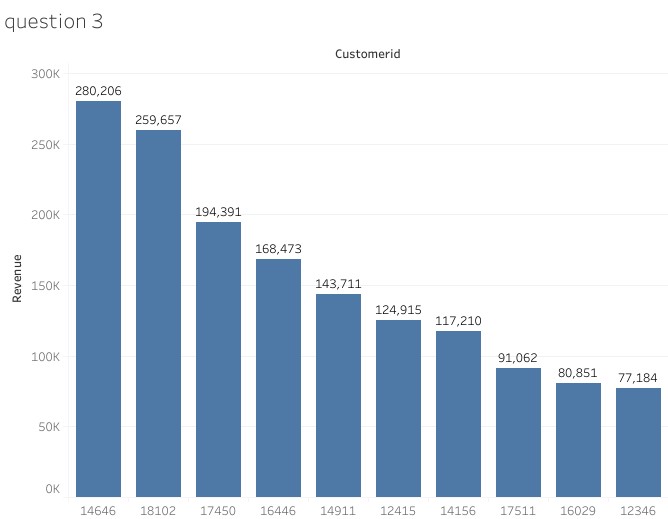
getwd()

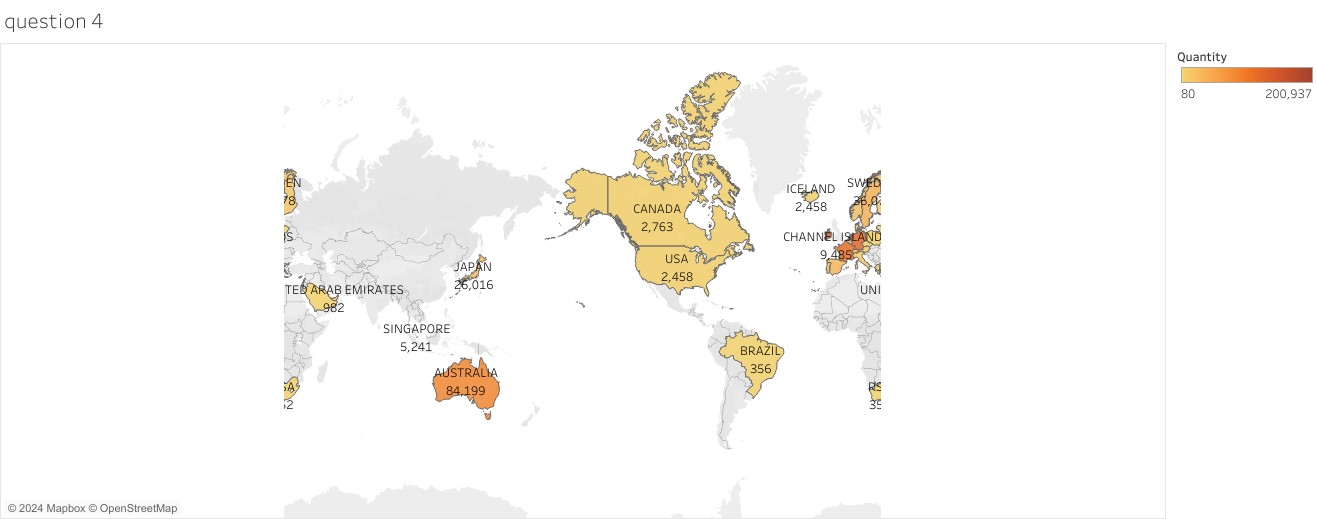
writexl::write\_xlsx(online\_retail, "online\_retail.xlsx")

**TASK 3 (Visualizations)**









**TASK 4 (COMMUNICATING INSIGHTS AND ANALYSIS)**

Good afternoon everyone,

I'm Tanupriya Das, and I appreciate the opportunity to share insights about our company.

Your leading questions were very helpful in understanding the insights you're looking for. I'm confident that the analysis will be compelling and useful for your upcoming business decisions.

I would like to assure you that I have provided the most accurate and up-to-date analysis.

And we will soon be discussing why that is.

* So, here we will be navigating through the three parts of today’s presentation ie.
* **Purpose statement**
* **The Cleanup process**
* **And Analysis with visualizations**

The Purpose statement here are the following,

(Refer to the slide points)

Now, moving on to the clean up process,

Firstly, on spreadsheet, I found the duplicates of all columns and removed them.

Then I used **RStudio** to do the rest of the data cleaning as well as analysis, as its too large for spreadsheet to process and perform calculations. I also used RStudio due to the efficiency and accessibility of the program, its ability to handle huge amount of data and being able to create the visualizations.

Then I installed packages and loaded them in order to work with the data set such as tidyverse, tidyr, dplyr etc to name a few.

Then I imported the data from my desktop, and prepared them by removing all the rows of quantity less than 1, and then removed all rows with unit price less than 0. I also removed all rows with NA values using the drop\_na().

Then I cleaned and formatted the columns and rows in the table for better consistency. Then I modified and transformed the dataset with mutate () and filter () for the analysis stage.

Now, moving on to the next stage ie. The analysis with visualizations stage.

1)Regarding your first query, the CEO has asked for a revenue trend to determine whether retail sales are seasonal. According to data, there are couple of months of the year 2011, that have significant development. January brought the highest revenue in 2011, totaling approximately 60 K $. But in February we saw a steep decline in revenue about 18.5 K $(about 3 times less than the previous month).

So, to summarize, the months of January, May, August and October saw a sharp rise in revenue. Whereas in Feb, Jul, September and Nov saw a sharp decline. Suggesting that, there are fluctuations throughout the year, with certain months experiencing sharp rises and others experiencing steep declines. Since there aren’t sufficient data for the months of June and Dec. No inferences can be made about them.

* The pattern observed here shows volatility rather than a clear seasonal trend. While there are months with higher and lower revenues, the lack of a consistent pattern suggests that factors other than seasonality might be influencing sales.
* Further analysis might be needed to identify underlying causes.

2)Next up, The top 10 countries with the most potential for growth are represented in the second graph.

Since the UK already has a large demand and I understand you are more interested in nations where demand may be boosted, the UK is not included in these statistics.

According to the data, sales of units and income are quite high in nations like the Netherlands, Ireland, Germany, and France. To guarantee that steps are taken to further seize these markets, I would propose concentrating on these nations.

3) For the 3rd query, the top 10 consumers who have made the most purchases from the business have been the subject of the third study. According to the statistics, there are not many differences between the top 10 consumer purchases. The fact that the highest revenue-producing consumer only spent 7.3% more than the second highest demonstrates that the company does not rely solely on a small number of consumers to generate income. This demonstrates that the state of business is positive.

4) The map chart concludes by comparing the places that have produced the greatest revenue to those that have not. Apart from the UK, it is clear that nations like the Netherlands, Ireland, Germany, France, and Australia generate large profits.

and the company should invest more in these nations to boost product demand. The map also reveals that the majority of sales occur only in the European zone, with only a small number in the American region and Japan. Along with Russia, there is no market for the items in Africa or Asia. Sales revenues and profitability might increase with the implementation of a fresh strategy focused on these areas.

I very much appreciate your time. Once you've reviewed the material, please let me know if you have any questions about the analysis or if you'd like to see any additional information. I would be glad to provide it for you.

Thank you.