* Data file Global Cybersecurity Threats from 2015 to 2024 as csv file.
* Data source Kaggle.com
* https://www.kaggle.com/datasets/atharvasoundankar/global-cybersecurity-threats-2015-2024
* Data consist of 10 different countries record of getting attacked from different source targeting multiple industries.
* The data also contain financial loss in Million highlighting Number of affected users during that event.
* The data also has Security vulnerability Type, Defense Mechanism used by countries and resolution time for each event.

Four Problem Statement and Hypotheses:

1st Problem Statement:

* What are the trends in the frequency and financial impact of different types of Attack type in India, China, and the USA from 2015 to 2024?

Hypothesis 1: The USA has experienced a higher overall financial impact from cyberattacks compared to India and China from 2015 to 2024.

2nd Problem Statement:

* Is there a similarity in incident resolution time caused by security vulnerability type and Defense mechanism used in India, China and the USA?

Hypothesis 2: Incidents involving “Social Engineering” as a vulnerability tend to have shorter resolution times across all three countries compared to those involving “Unpatched software”.

3rd Problem Statement:

* Which Industries are the most frequently targeted and suffer the highest financial losses from cybersecurity threats in India, China and the USA?

Hypothesis 3: The “Retail” industry is among the top three most frequently targeted industries in all three countries. There is also a strong relation between Incident Frequency and Financial Loss.

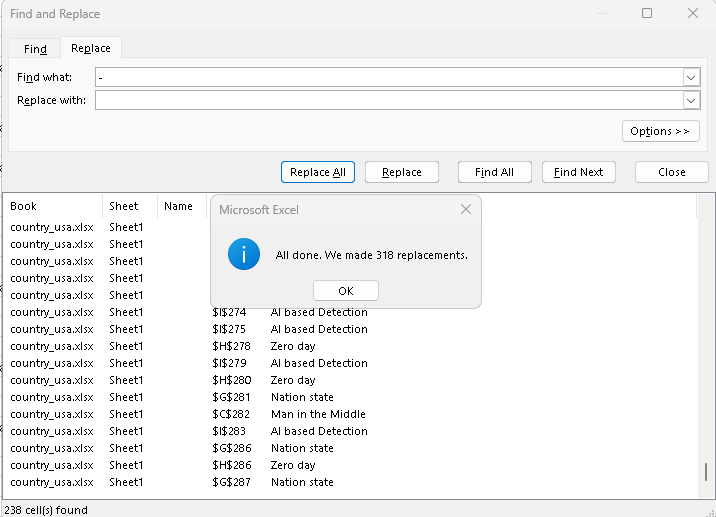
4th Problem Statement:

* Is there any correlation between the severity of the financial impact and the number of affected users in cybersecurity incidents within India, China and the USA?

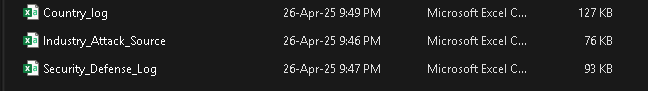
Hypothesis 4: Across India, China, and the USA there is a positive correlation between the financial impact and the affected users due to cyber-attack. In other word incidents with higher financial losses tend to impact a larger number of users.

Analysis Steps:

* I only want to analyze three country India, China and USA so I filtered the country column and picked only three countries to populate data with their information, I also created
* The unique incident key for each individual events and I made the column in the original dataset before separating data files.
* I also double checked the columns and change their type according their values. I had to double check column names, strings with any other character types and containing digits and assigning them as Numbers or floats.
* I selected the column and using *Right-click >> Format Cell >> Number* and assigned decimal places depending on if they had floating points. For floating points, I used two decimal places.
* To make sure the values are readable in SQL developer I had to replace “ - “ with space on columns Attack Type, Attack Source, Defense Mechanism Used and Security Vulnerability Type using Find & Select >> Replace.

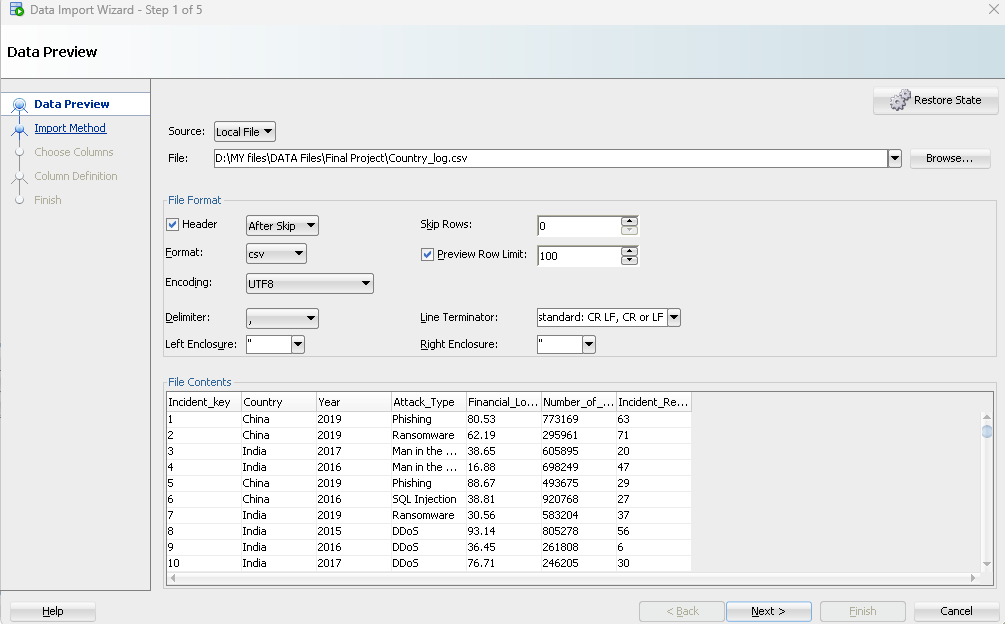


* After that I renamed the columns with space to “\_” to make it acceptable in SQL developer.
* After that I filtered the columns and removed columns that I don’t needed. And I created a 3 CSV file named country\_log, Industry\_Attack\_Source, and Security\_Defense\_log.

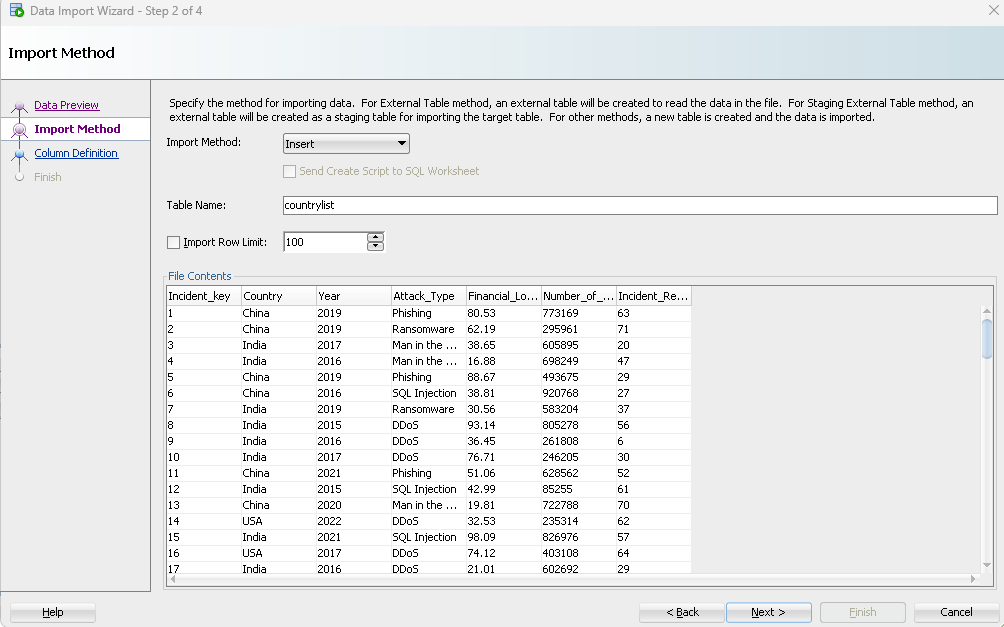


Now I am ready to import the files in my SQL developer to start my analysis.

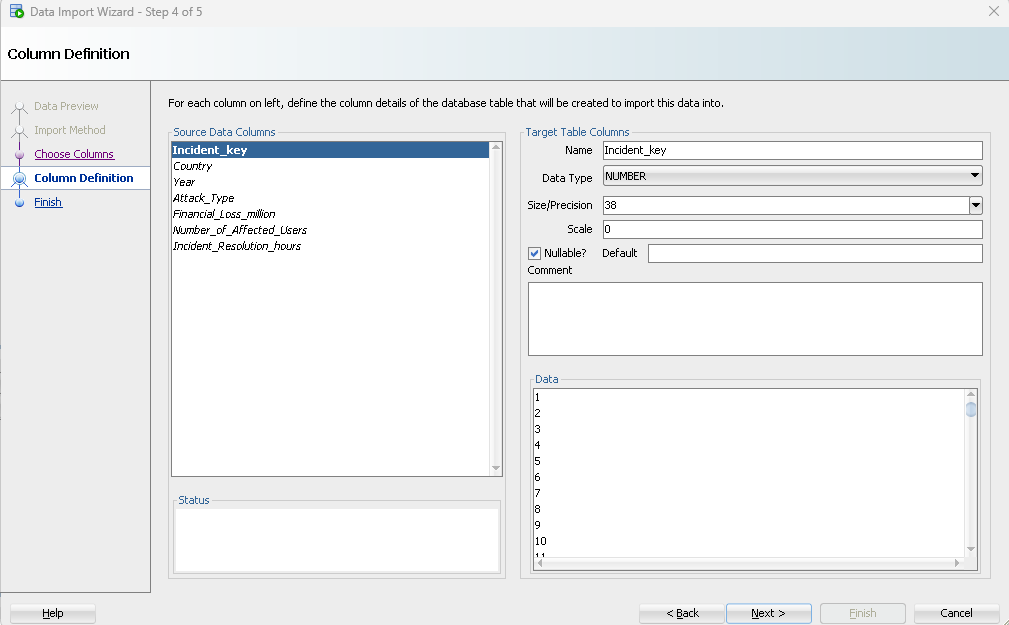
* I logged into my SQL developer database and imported the three CSV files.
* Importing steps: *Right click on the database Table and Left click on the Import Data.*
* After that Data Import Wizard pops up and we have to find the location of the data using Browse.
* Navigate to the drive and folder where the CSV files are located and select the Country\_log.csv file.



* Include the header and make any changes, in my case default and click Next. Next import method box appears and we choose “Insert give a table name and depending on Import Row limit import. In my case I want to import all the rows so leave it unchecked.



* Next, I have options to include and exclude columns. In my case my analysis needs all the columns so I don’t make any changes and click Next to move on to next step.
* Now I am in Column Definition and there are no issues with the columns as we formatted previously.

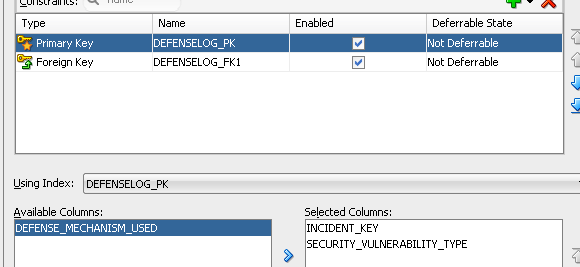
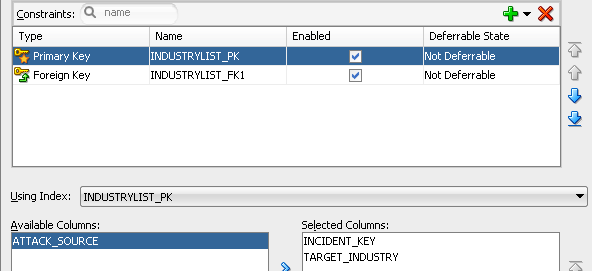


* Now we hit next to Finish and Task successful message pops up.
* I repeat the same process to import other CSV files as well.

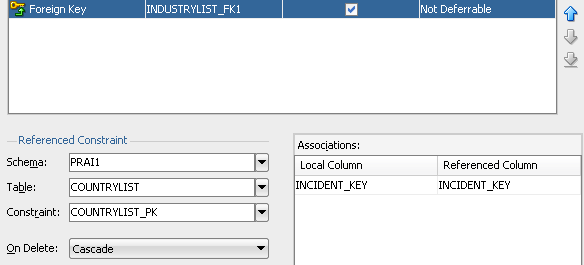
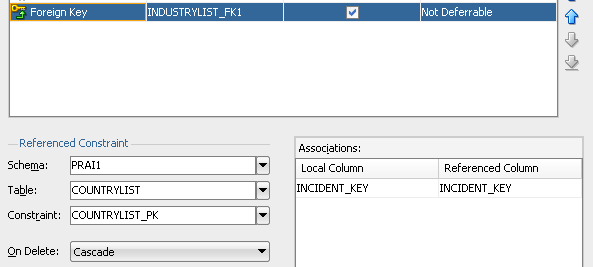


Now I have to assign Primary Key and Foreign Key.

* I select the Data and go to pencil icon and Constraints and using add sign drop down option I choose Primary Key for the COUNTRYLIST TABLE and I choose INCIDENT\_KEY and save it.
* After that for DEFENSELOG and INDUSTRYLIST I set INCIDENT\_KEY as primary key along with SECUTIRY\_VULNERABILITY\_TYPE and TARGET\_INDUSTRY.

* Then I created a foreign key for both in table I choose COUNTRYLIST and its constraint with cascade on Delete.

Now with all that setup I can finally start working on my Problem statement and Hypothesis. The data was already sorted between 2014 till 2024 so we don’t need to worry about the year unless we want to analyze for specific year.

1st Problem Statement:

* What are the trends in the frequency and financial impact of different types of Attack type in India, China, and the USA from 2015 to 2024?

To pull this data I need to use countrylist table and count all the attack type for each country. This will give a frequency for attack for each attack type per country.

Using SQL query:

SELECT

year, country, attack\_type,

COUNT(\*) AS "Incident Frequency"

FROM

countrylist

WHERE

country IN ('India', 'China', 'USA')

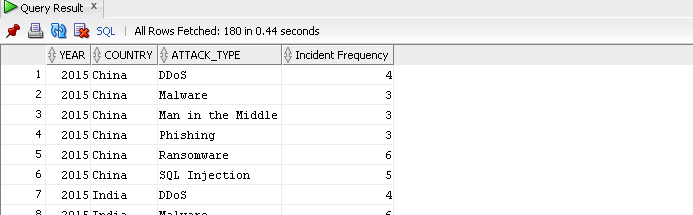
AND year BETWEEN 2015 AND 2024

GROUP BY

year, country, attack\_type

ORDER BY

year, country, attack\_type;



Hypothesis 1: The USA has experienced a higher overall financial impact from cyberattacks compared to India and China from 2015 to 2024.

Now for this analysis I am going to run another SQL query and combine my both data in Excel for visualization and further analysis.

Using SQL query:

SELECT

year, country, attack\_type,

SUM(financial\_loss\_million) AS "Total Financial Impact"

FROM

countrylist

WHERE

country IN ('India', 'China', 'USA')

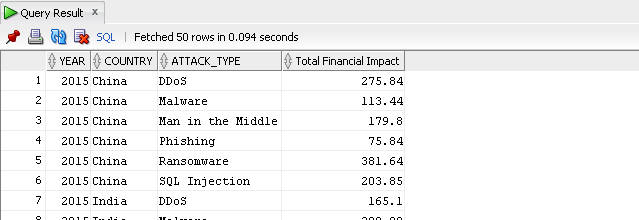
AND YEAR BETWEEN 2015 AND 2024

GROUP BY

year, country, attack\_type

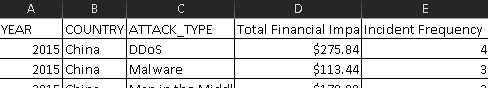
ORDER BY

year, country, attack\_type;

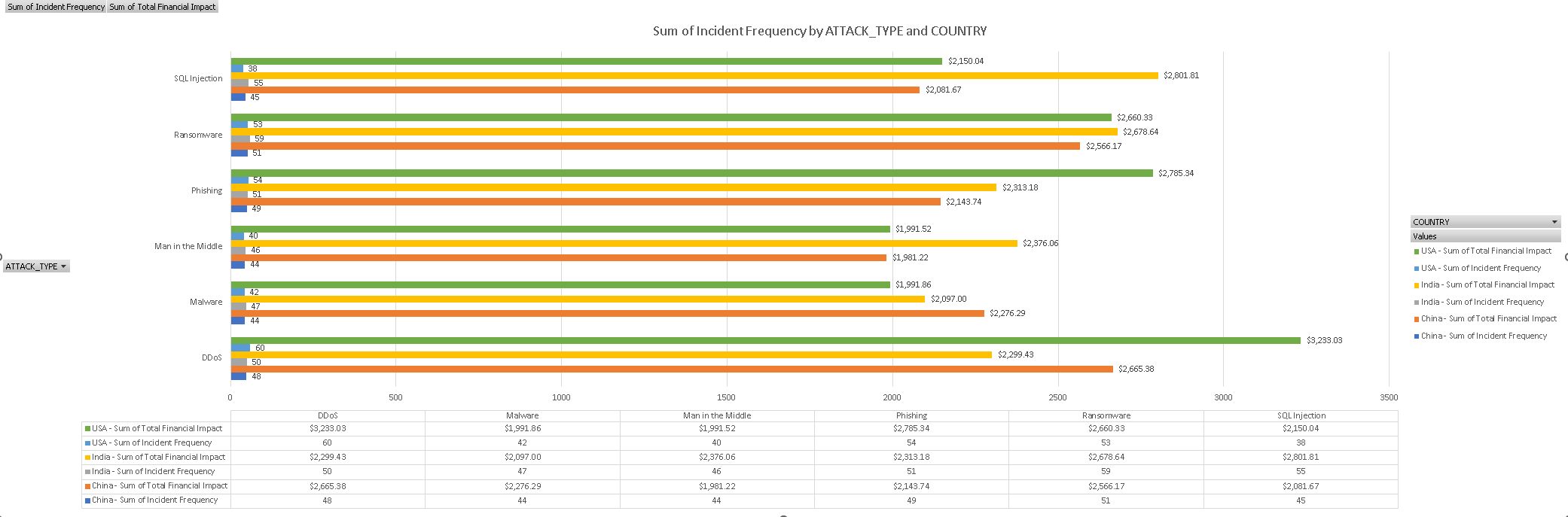


Now that we have gather all the data for our problem one, we will take this data in our excel file and create a visualization to analyze the relationship between our columns and verify our hypothesis.

I have organized those data as:



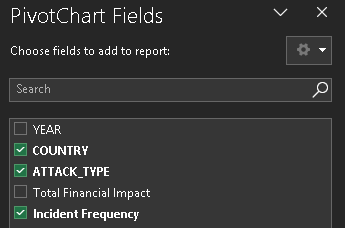
Now we will select all the columns that we need in this case we can select all the columns or also ignore the year column if we don’t want analysis specifically for year and we will insert a clustered bar chart.

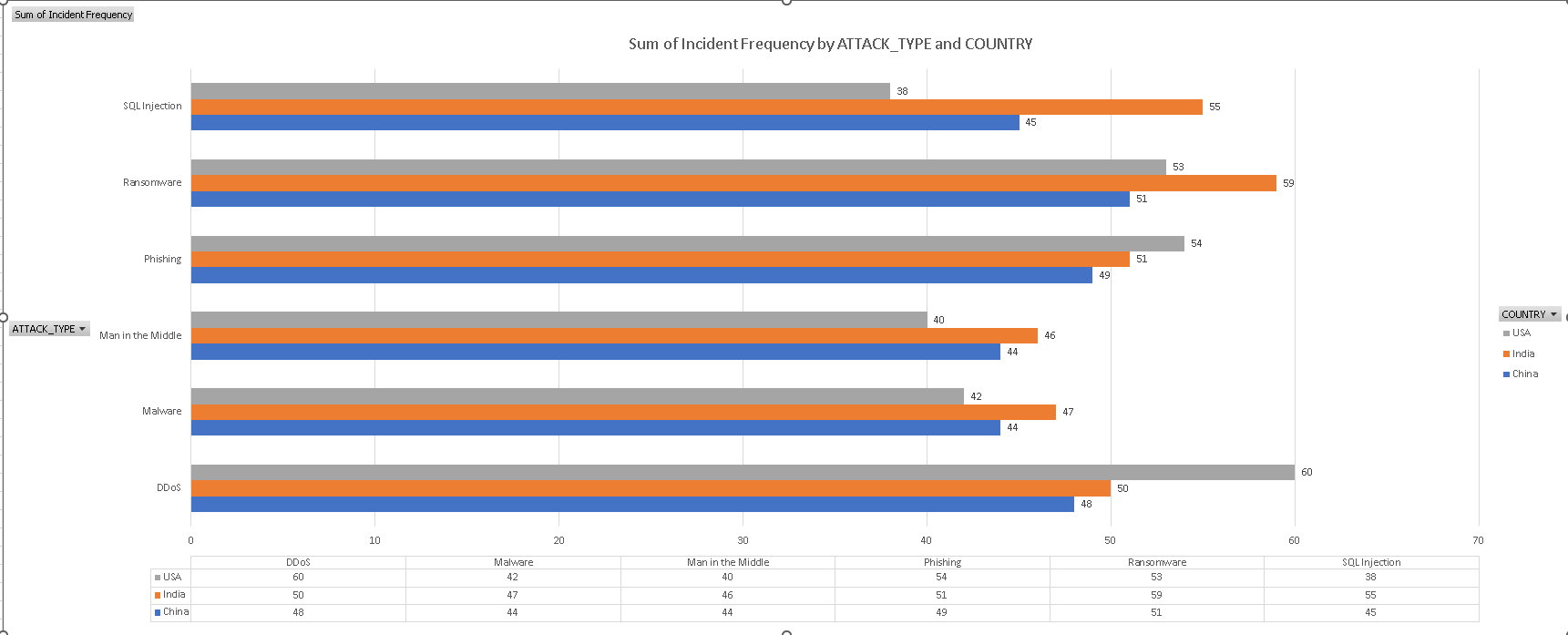


Now first of all we will answer our problem statement:

* What are the trends in the frequency and financial impact of different types of Attack type in India, China, and the USA from 2015 to 2024?

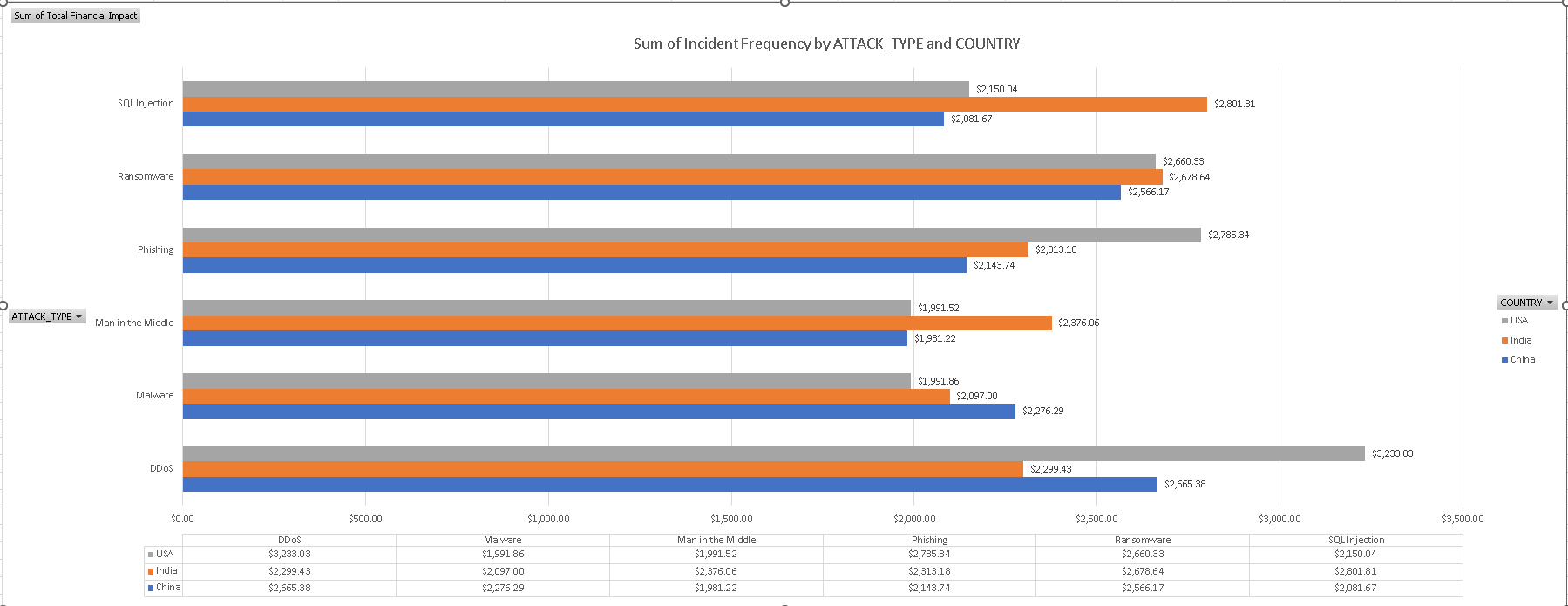
First of all, using our PivotChart Fields, we will only select our columns country, Attack\_type and Incident Frequency to answer the trend in the frequency.



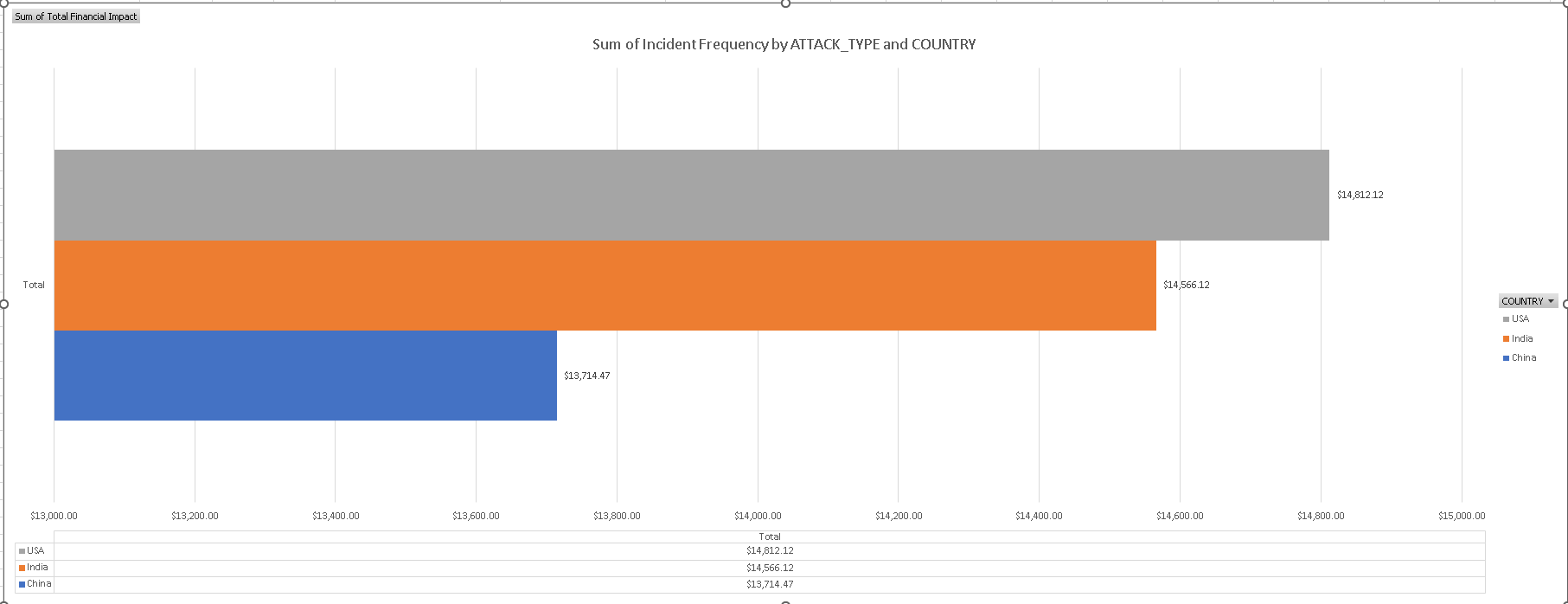


We can see a visual trend of the attack types for each country. Which shows that DDos attacks has the highest attack frequency in the USA and Ransomware in China and India.

Next, we will select our columns Total Financial Impact and uncheck the Incident Frequency to answer the financial impact of different types of attack. From the below visualization we can see that DDos attack caused the highest financial impact in the USA of $3,233.03 million and China of $2,665.38 million, where in India SQL injection attack caused the highest financial impact of $2,801.81 million.



Now to answer our Hypothesis if USA experienced the highest overall financial impact from the cyberattacks, we need to sum our total financial loss in million for each country and analyze the result. We can get that report by unchecking our attack type column and we can see that report on the image below. The USA has experienced the highest overall financial impact from cyberattacks, totaling $14,812.12 millions compared to India whose total is $14,566.12 million and China $13,714.47 million. Proving our Hypothesis.



We can also get those total financial loss per country using SQL query as well:

SELECT

country,

SUM(financial\_loss\_million) AS "Total Financial Impact"

FROM

countrylist

WHERE

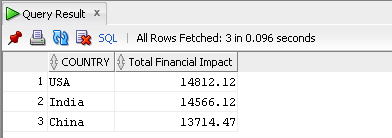
year BETWEEN 2015 AND 2024

GROUP BY

country

ORDER BY

"Total Financial Impact" DESC;



2nd Problem Statement:

* Is there a similarity in incident resolution time caused by security vulnerability type and Defense mechanism used in India, China and the USA?

Our second problem question is little bit complex than our first one and it is not straight forward. To find any similarity in incident resolution time caused by security vulnerability type and defense mechanism used in India, China and the USA we have to visualize and analyze their average resolution time. To get this data we will go back to our SQL developer and write query that will give us country its security vulnerability type, defense mechanism used and their average resolution time.

We will use SQL query:

SELECT

cl.country,

dl.security\_vulnerability\_type,

dl.defense\_mechanism\_used,

ROUND(AVG(cl.incident\_resolution\_hours), 2) AS "AVG resolution time"

FROM

countrylist cl

JOIN

defenselog dl ON cl.incident\_key = dl.incident\_key

WHERE

cl.country in ('India', 'China', 'USA')

GROUP BY

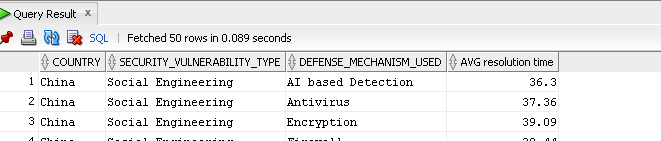
cl.country, dl.security\_vulnerability\_type, dl.defense\_mechanism\_used

ORDER BY

cl.country, dl.security\_vulnerability\_type, dl.defense\_mechanism\_used;

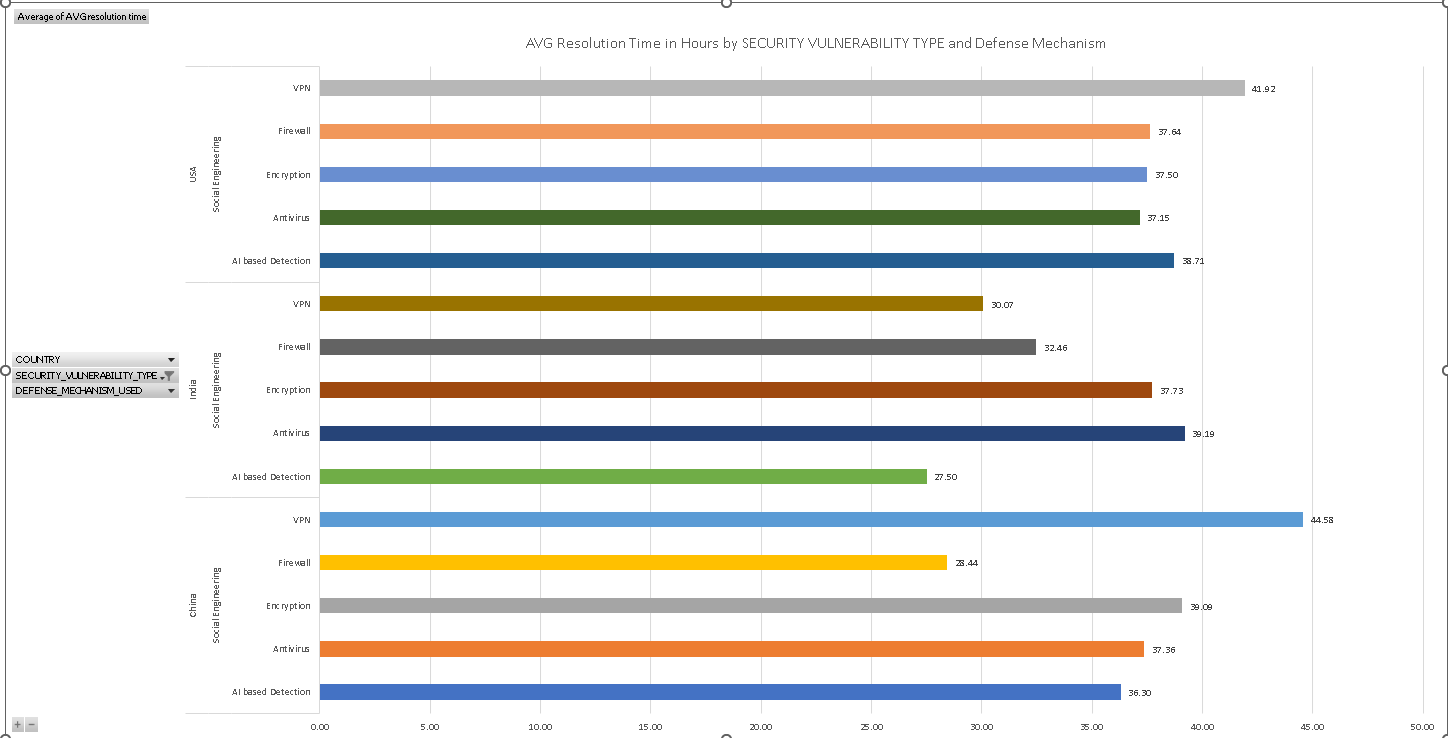
In our query we had to join countrylist table with defenselog table with their incident key column as for Inner Join that will give us data needed to answer our problem statement one.

Example:



Now we will take this data and open it in our excel to create a visualization. We will use clustered bar chart again and by default it will show a very large chart with all the values. But we can simplify using each security vulnerability type defense mechanism used and its average resolution time to answer our question.

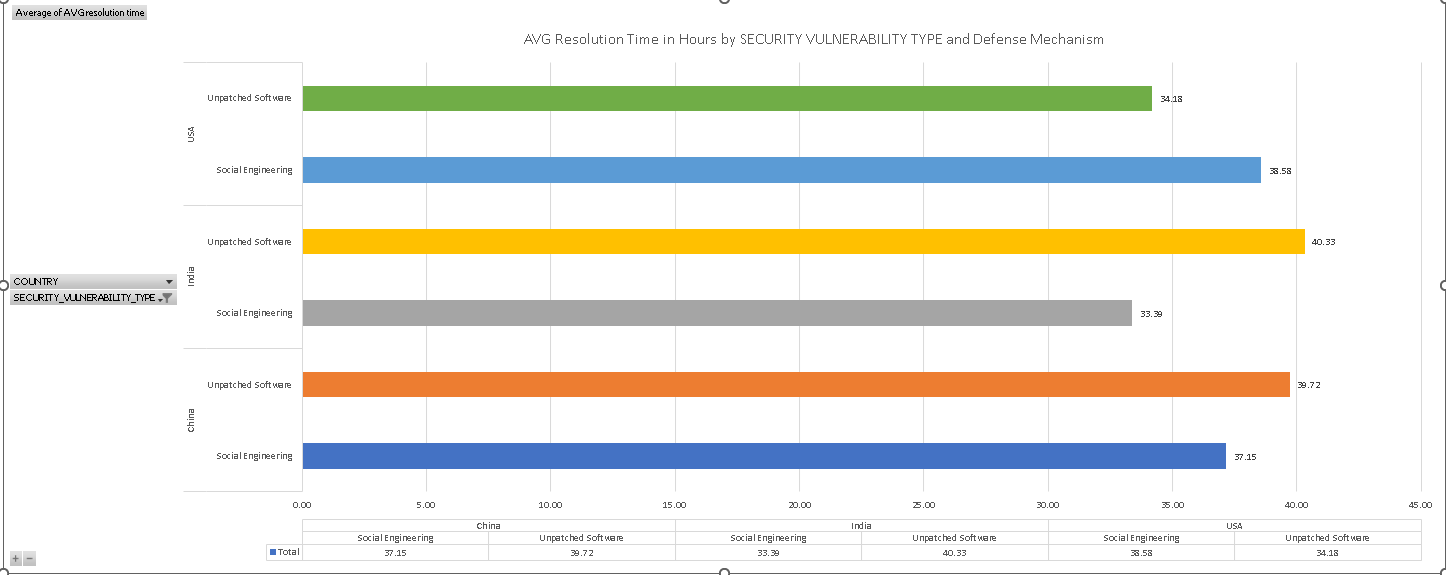
For example, I will only check social engineering as security vulnerability type and sort our data to check any similarities.



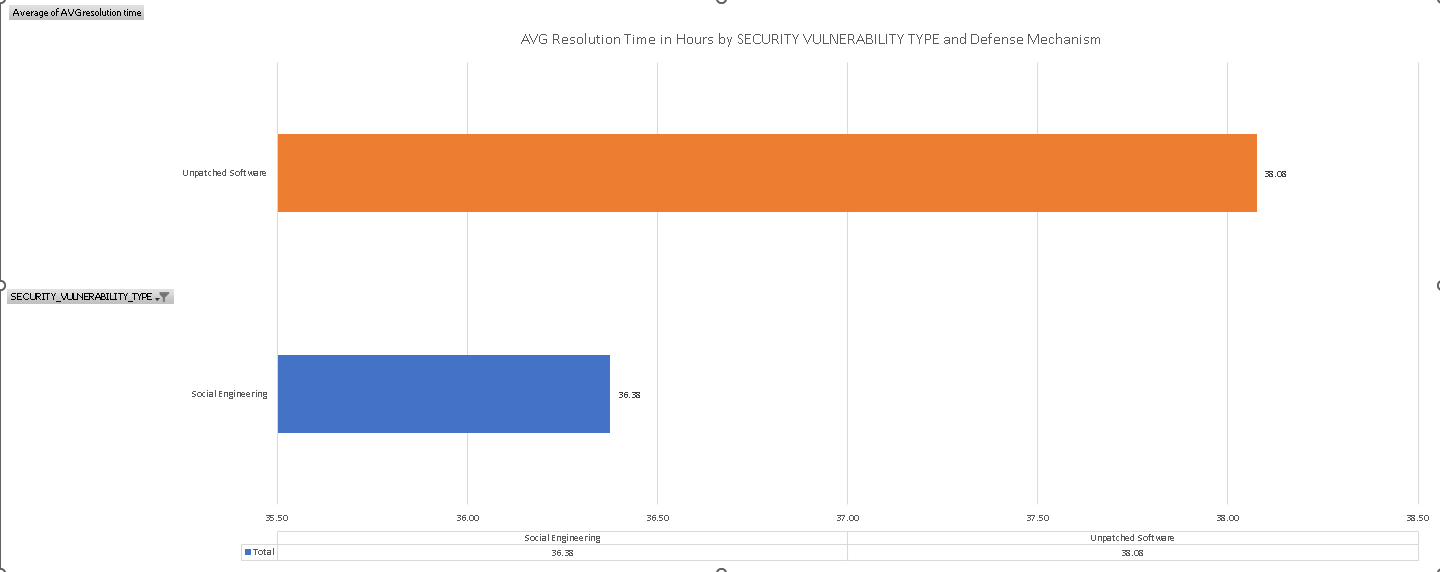
As we can see from the chart there are no similarities in the in the incident resolution time for this scenario. We can repeat and check the same for others as well and find there is no consistent similarities in incident resolution time across three countries.

Hypothesis 2: Incidents involving “Social Engineering” as a vulnerability tend to have shorter resolution times across all three countries compared to those involving “Unpatched software”.

Now we will check if our hypothesis is true. We will only check social engineering and unpatched software in security vulnerability type and uncheck the defense mechanism used column from our PivotChart fields to analyze our data.



We can see that we cannot prove our hypothesis using this relationship because in the USA social Engineering average time is higher than unpatched software. But if we combine all of the countries data than we see social engineering has the shortest average time than that of unpatched software. Which partially proves our hypothesis.



3rd Problem Statement:

* Which Industries are the most frequently targeted and suffer the highest financial losses from cybersecurity threats in India, China and the USA?

This is one of the interesting questions I came up with and here we need to sort our data using country and its industrylist their incident frequency and total financial loss in millions.

First, we will filter our data using SQL query:

SELECT

cl.country, il.target\_industry,

COUNT(\*) AS "Incident Frequency",

ROUND(SUM(cl.financial\_loss\_million), 2) AS "Total Financial Loss"

FROM

countrylist cl

JOIN

industrylist il ON cl.incident\_key = il.incident\_key

WHERE

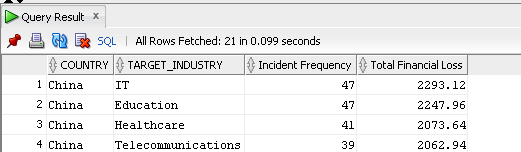
cl.country IN ('India', 'China', 'USA')

GROUP BY

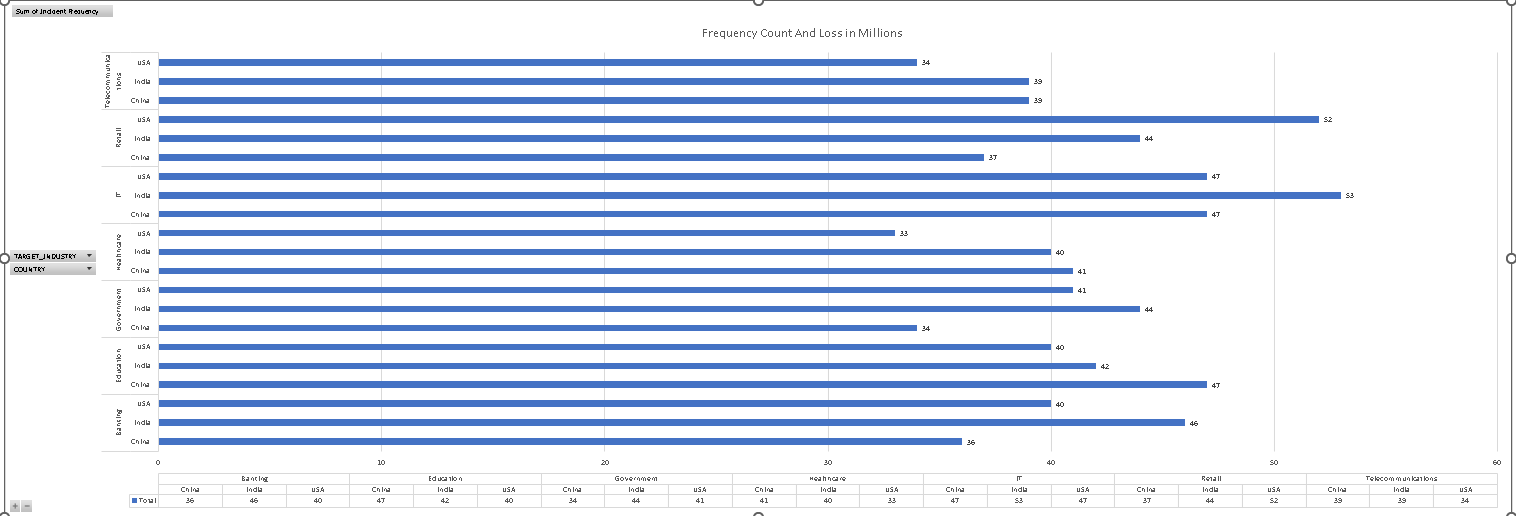
cl.country, il.target\_industry

ORDER BY

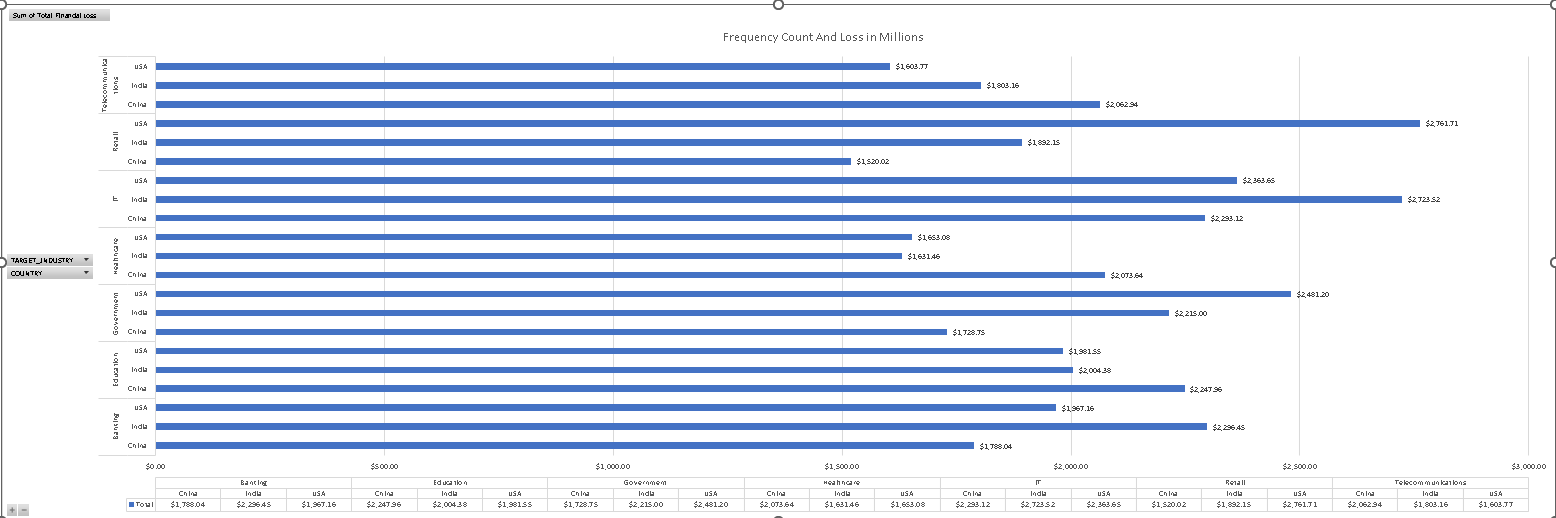
cl.country, "Total Financial Loss" DESC;



Now let’s take this data into our excel and create a visualization to answer our question. We will use the same clustered chart for this question and we can see that across India, China, and the USA, the IT industry faces the highest incident frequency, while retail is top target in the USA.

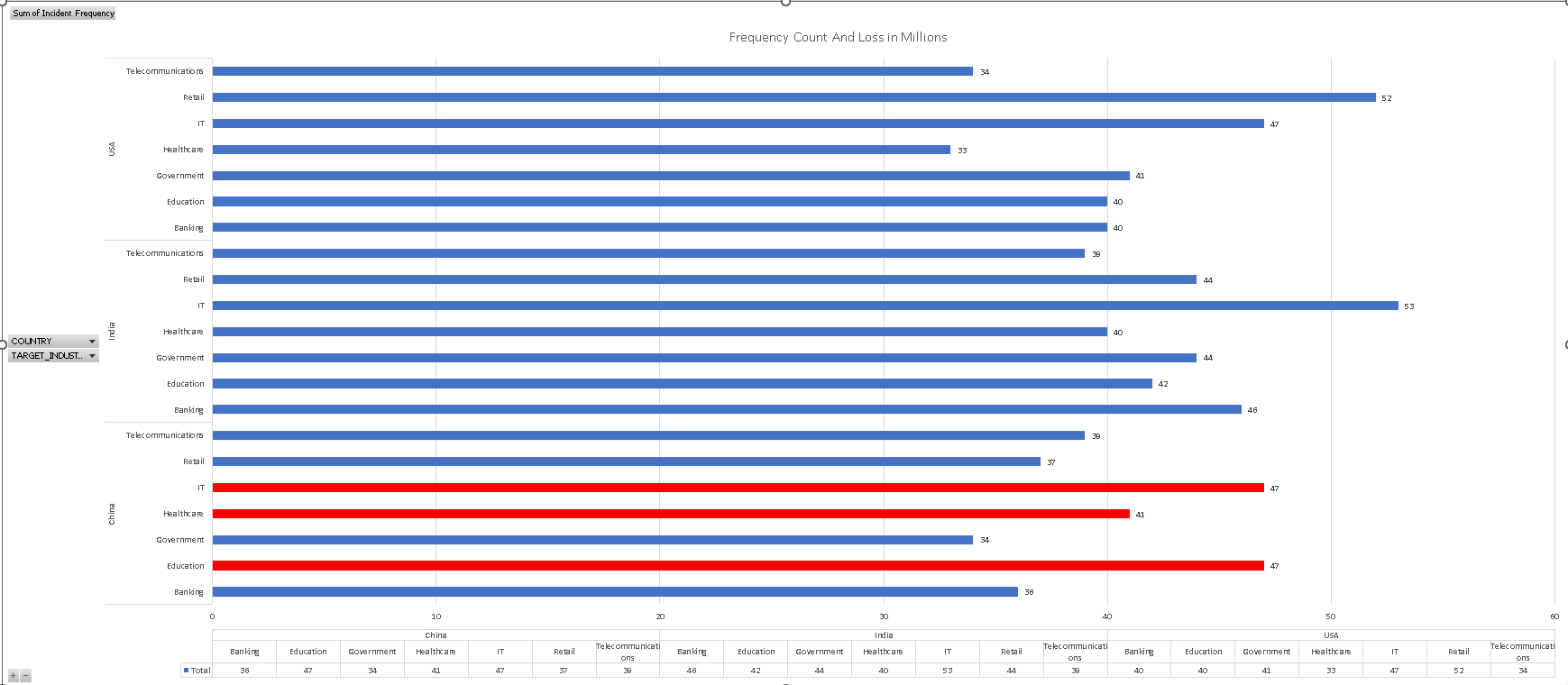


In term of financial loss in millions the IT industry and the Retail industry experience the significant impact.

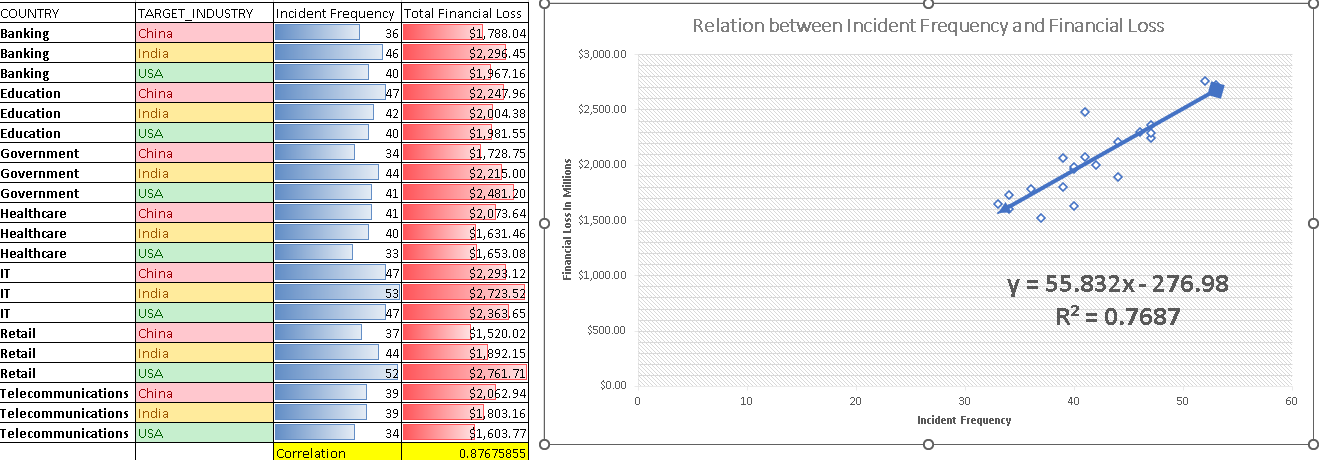


Hypothesis 3: The “Retail” industry is among the top three most frequently targeted industries in all three countries. There is also a strong relation between Incident Frequency and Financial Loss.

Now looking at the data above we cannot prove our hypothesis that retail is among the top three most frequently targeted industries in all countries. Also, if we see the chart below which shows the frequency of incident the retail does not fall in top three industries for all countries. In China the top three industries with highest incident frequency are Education, Health and IT. Which directly falsify our hypothesis.



Next, we also have to check if there is a strong relation between Incident Frequency and Financial Loss. We can analyze this relationship using Scatter plot which will help us to visualize the relationship.



If we analyze the scatter plot with the trend line, we can see a strong positive correlation between incident frequency and financial loss. As the number of incidents rises, the financial loss tends to increase significantly proving our hypothesis. We can also reevaluate this relationship using its correlation coefficient which is 0.88 rounded close to value one.

4th Problem Statement:

* Is there any correlation between the severity of the financial impact and the number of affected users in cybersecurity incidents within India, China and the USA?

Our final question tends to find if there is any kind of relationship between the financial impact and the number of affected users due to cybersecurity incidents within each country. We can easily answer this question by analyzing its correlation coefficient.

Using our SQL developer, we will run a query that will show us correlation coefficient for each country between affected users and financial loss:

We will use SQL query:

SELECT

country,

ROUND(CORR(financial\_loss\_million, number\_of\_affected\_users), 3)

AS "Correlation\_coefficient"

FROM

countrylist

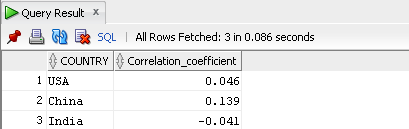
WHERE

country IN ('India', 'China', 'USA')

GROUP BY

country;

This will give us:



And before jumping into conclusion I want to take the actual data and visualize this relationship using excel.

So, I will sort the data using SQL query:

SELECT

country, financial\_loss\_million AS "Loss In Million",

number\_of\_affected\_users As "Affected User Count"

FROM

countrylist

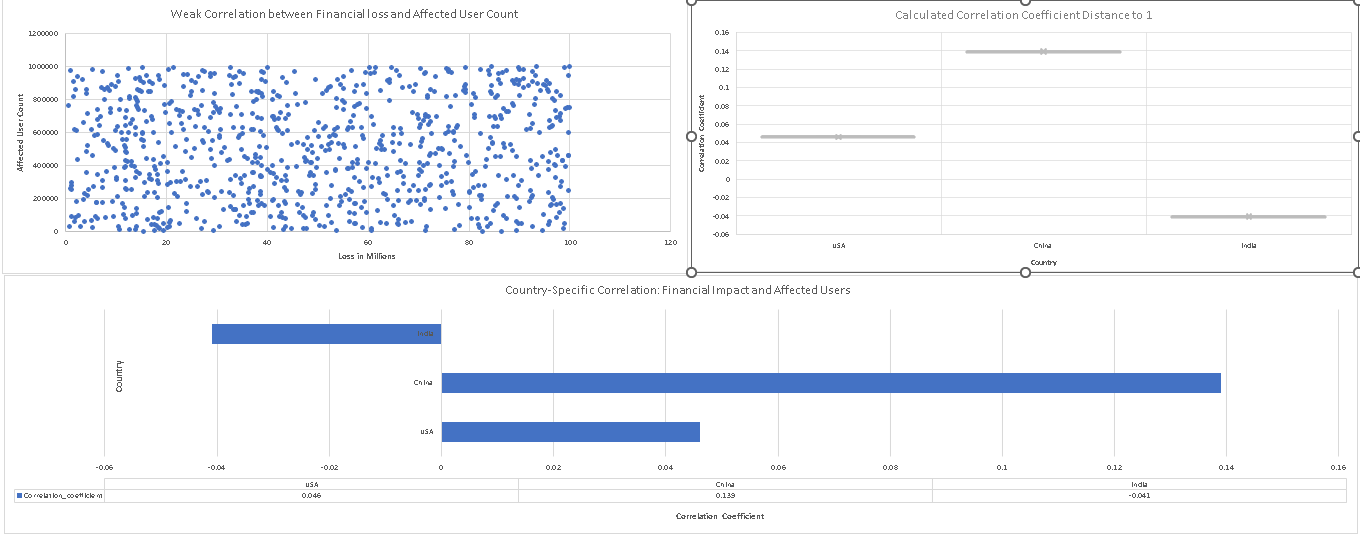
WHERE

country IN ('India', 'China', 'USA')

ORDER BY

country, "Loss In Million" DESC;

And take this data into excel and create a scatter plot, box and whisker, and clustered chart that show me the relationship between affected users and financial loss.



Hypothesis 4: Across India, China, and the USA there is a positive correlation between the financial impact and the affected users due to cyber-attack. In other word incidents with higher financial losses tend to impact a larger number of users.

Looking at the above charts we can conclude that the data does not strongly support our hypothesis of a positive correlation between financial impact and the numbers of affected users across India, China and the USA. The correlation is weak and positive in the USA and China, and slightly negative in India, indicating a minimal linear relationship overall.

Summary:

Based on our analysis the USA had the higher overall financial impact from cybersecurity. There are no consistent similarities across countries in term of incident resolution times. The retail industry is not in the list of top three frequently targeted industry because in China the top three are Education, Health and IT industries. But there is a strong positive correlation between incident frequency and financial loss with a correlation coefficient of 0.88. Lastly, there is no positive correlation between financial impact and affected users across the three countries.