* 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.

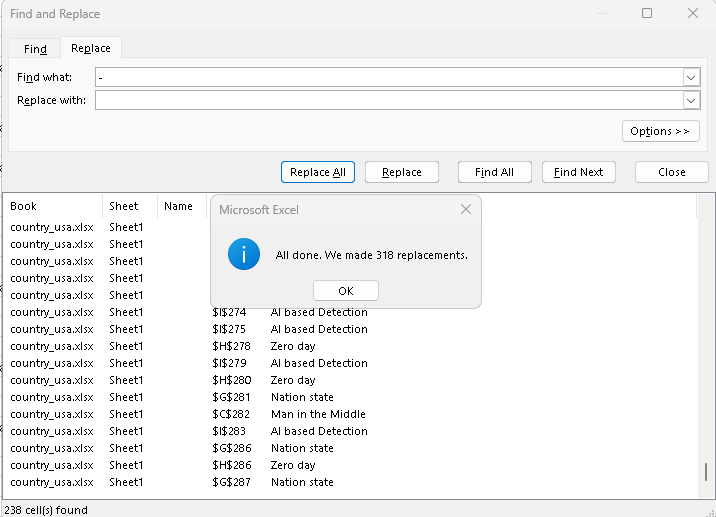
4th Problem Statement:

* 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?

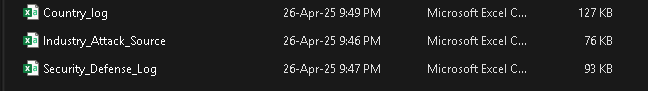
Hypothesis 4: Across India, China, and the USA there is a positive correlation between the financial impact and the affected users due to cyber-attack. In other word incidents with higher financial losses tend to impact a larger number of users.

Analysis Steps:

* I sorted the data according country information as I mentioned before 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.

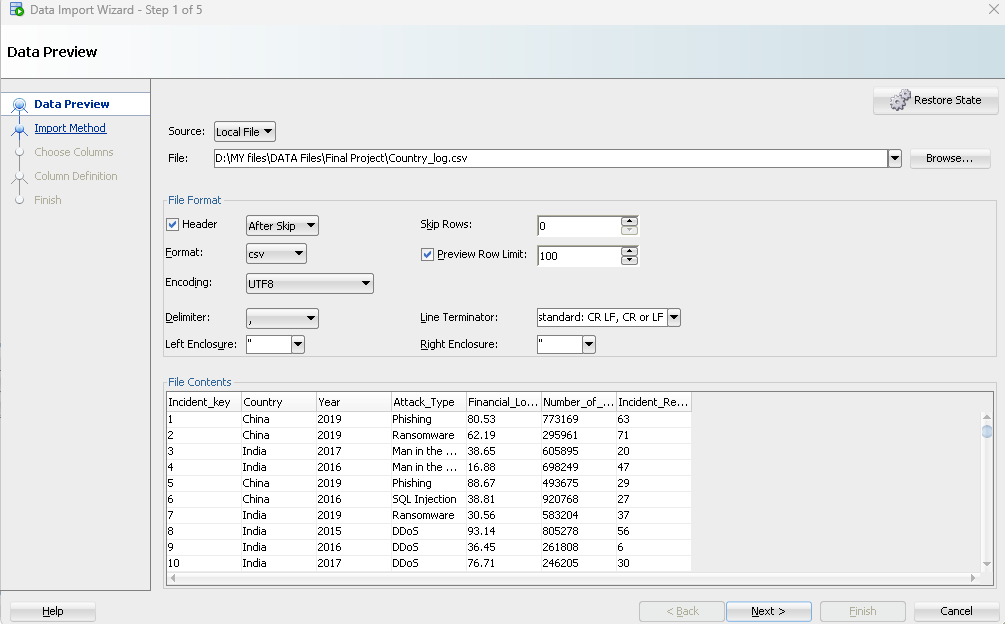


* After that I renamed the columns with space to “\_” to make it acceptable in SQL developer.
* After that I filtered the columns and removed columns that I don’t needed. And I created a 3 CSV file named country\_log, Industry\_Attack\_Source, and Security\_Defense\_log.

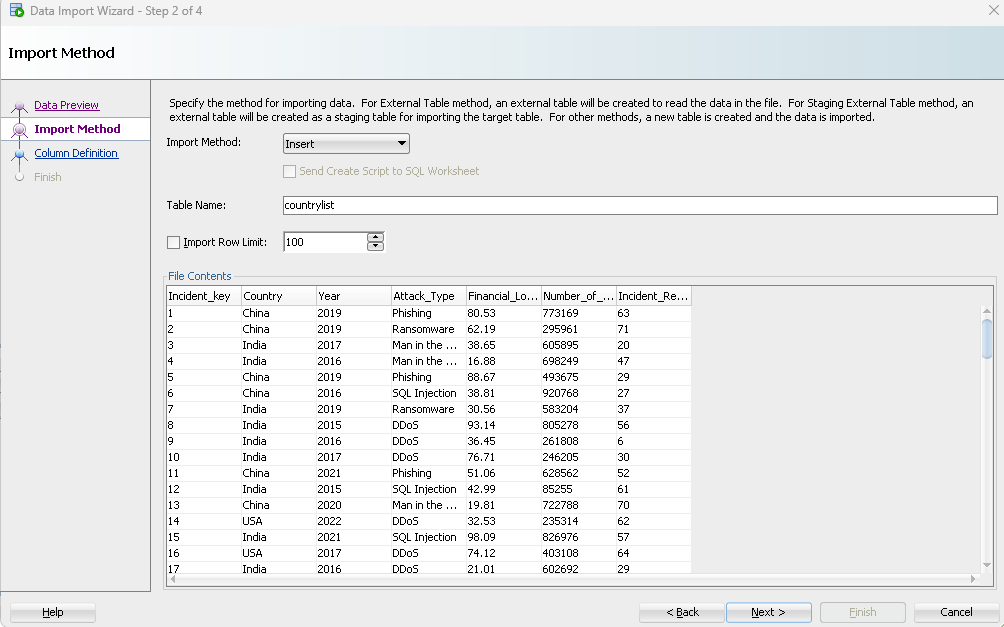


Now I was ready to import the files in my SQL developer to start my analysis.

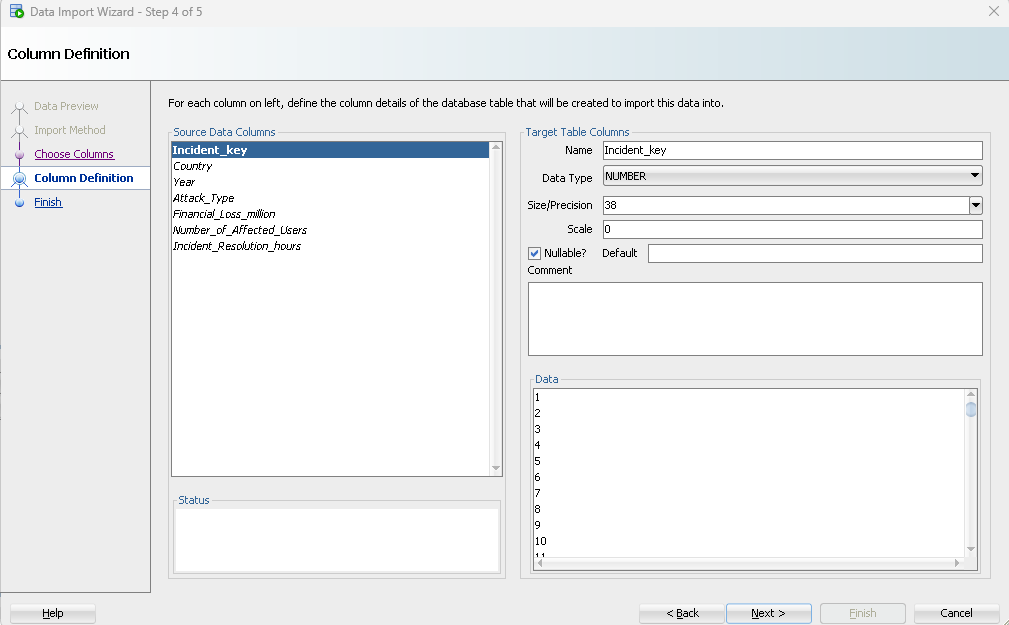
* I logged into my SQL developer database and imported the three CSV files.
* Importing steps: *Right click on the database Table and Left click on the Import Data.*
* After that Data Import Wizard pops up and we have to find the location of the data using Browse.
* Navigate to the drive and folder where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

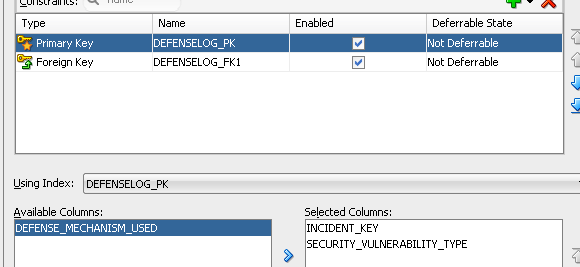
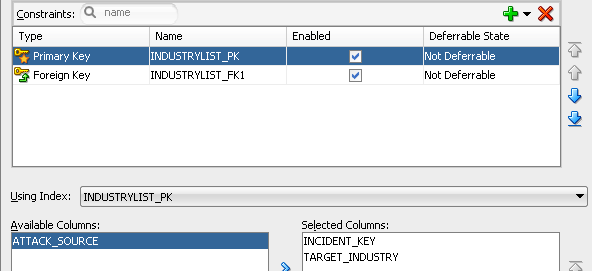


* Now we hit next to Finish and Task successful message pops up.
* I repeat the same process to import other CSV files as well.

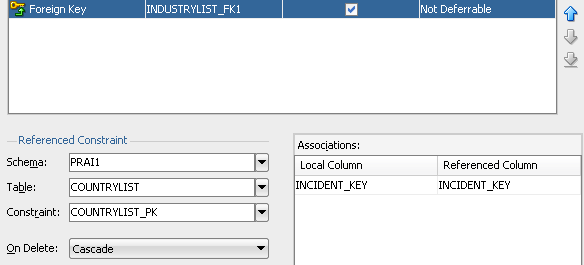
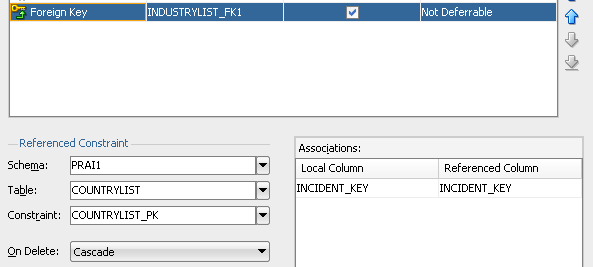


Now I have to assign Primary Key and Foreign Key.

* 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.