Data Visualisation and Story Telling
Final Project

Encoding US Healthcare Data Breaches

Research question

"What are the trends, patterns, and impacts of healthcare data breaches in the United States?" This question is designed to guide an in-depth exploration of healthcare data breaches, focusing on three key aspects:

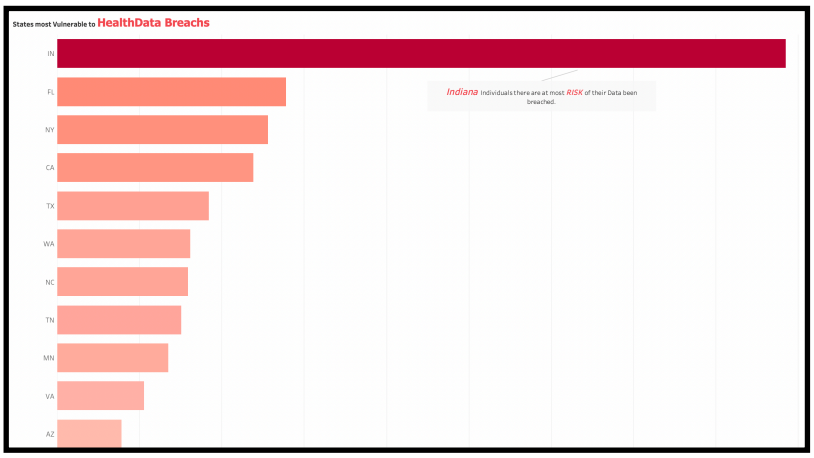
1. Trends: This refers to the general direction in which the data breaches are developing over time. Are healthcare data breaches becoming more frequent? Understanding these trends can help predict future breaches and inform proactive prevention strategies.
2. Patterns: This refers to recurring characteristics or factors in the data breaches. For example, are certain types of healthcare entities more prone to breaches? Are certain types of data more likely to be compromised? Identifying these patterns can help target resources and interventions where they are most needed.
3. Impacts: This refers to the consequences of the data breaches. How many individuals are typically affected by each breach? What types of data are typically compromised? What is the financial cost of these breaches? Understanding these impacts can help quantify the severity of the problem and prioritise efforts to mitigate these impact

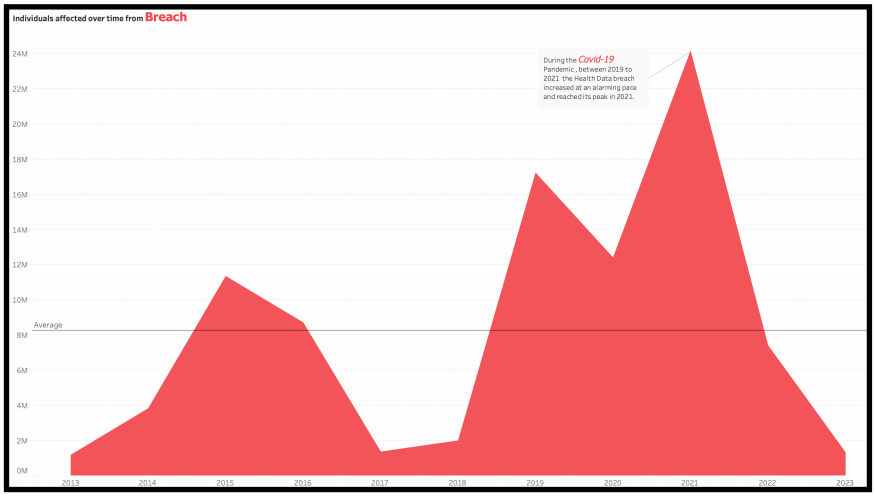
Audience

The primary audience for this project includes healthcare professionals, cybersecurity experts, policy makers, and researchers interested in healthcare data security. This question matters to them because understanding the patterns and impacts of healthcare data breaches can inform strategies to prevent future breaches and mitigate their effects. Professionals need to be aware of the risks associated with data breaches to ensure the privacy and security of patient information. The visualization will include functionalities such as filters that will allow the audience to interact with and explore the data in a meaningful way. This will enable the audience to focus on specific aspects of the data that are most relevant to their interests and responsibilities.

Descriptions of Healthcare Data Infographic Visuals

Used **bar charts** to plot the States vs the individuals affected by breaches over the last decade, this graph will help understand which state is at most risk of the health data been breached , so that the cybersecurity experts, policy makers can focus specifically on states where attention must be given to prevent future breaches.



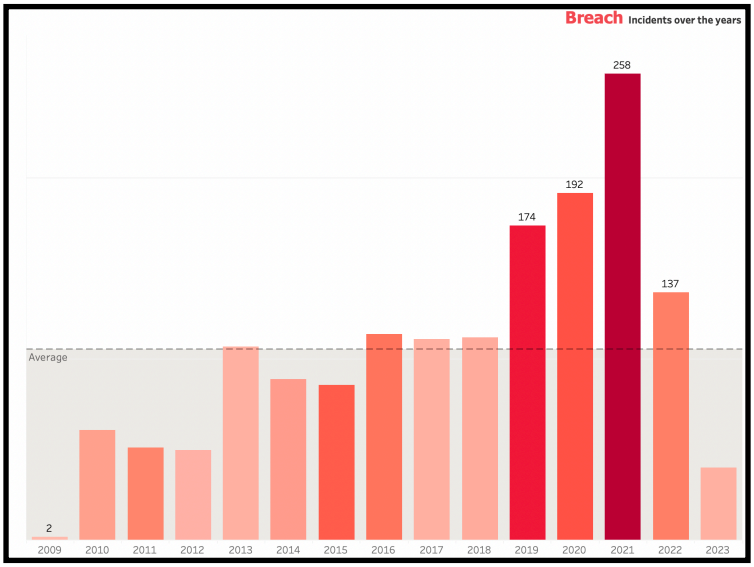


Utilised **area chart** to plot the sum of individuals affected over the period of 10 years, during the period of Covid-19 Pandemic ,there was an upsurge in the cases of Health data breaches.During that time healthcare system of US collapsed and maintaining the privacy of the data became a daunting task. Policy makers must focus on How to make effective strategies on how to handle health crisis in future so that it don't effect individuals privacy.

Prepared a **table** which highlights at which point of contact most of the Breach Happened. Attention must be given by the Cybersecurity experts and teams in order to make the Network servers breach proofs and the Data centres must be enhanced with high level security .

where **Breach** Occur the Most

Network Server	26,21,17,174
Email	3,34,61,866
Other	1,39,47,765
Desktop Computer	76,43,173
Electronic Medical Record	75,91,380
Laptop	69,40,205
Other Portable Electronic ..	32,98,031
Paper/Films	57,58,574



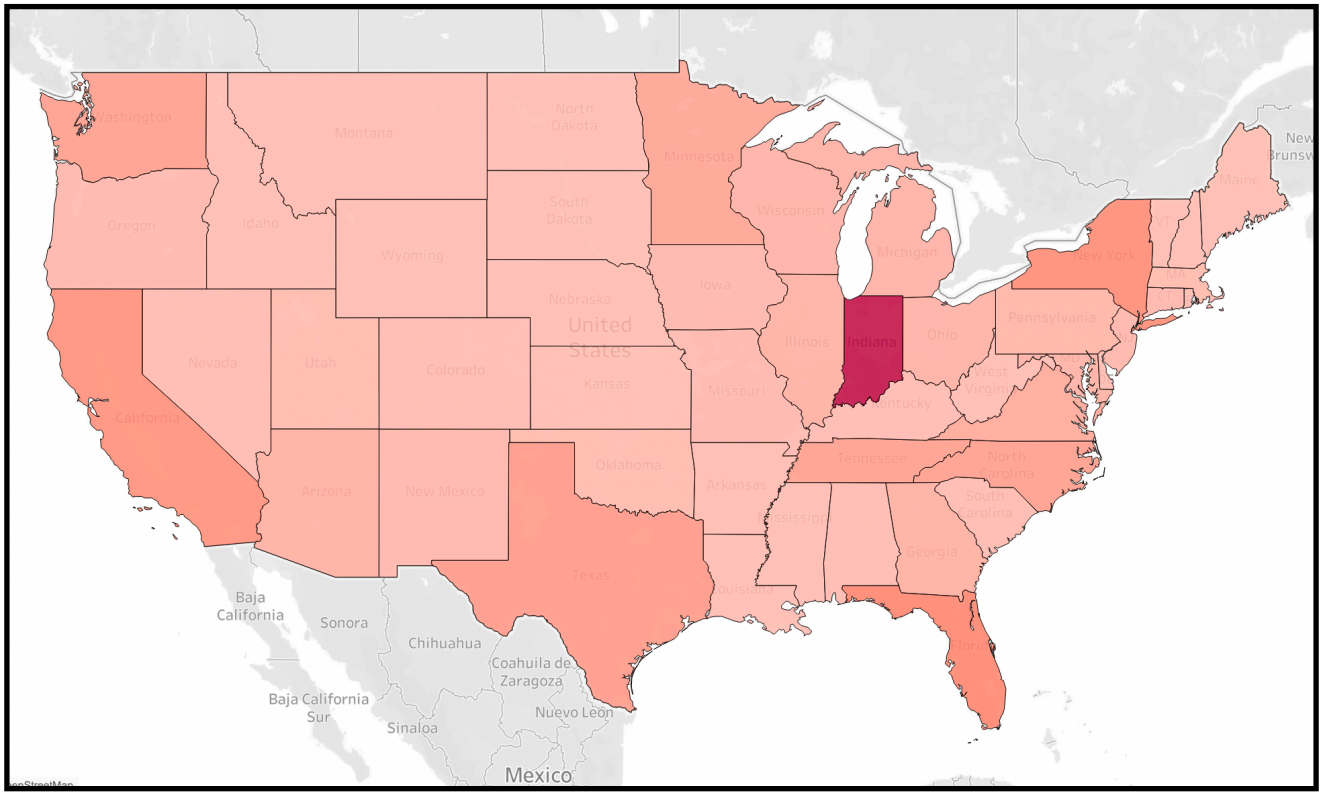
Plotted **bar chart** to highlight the number of breach incidents occurring over the period of time. As can be seen from the graphs , the period of Covid between 2019-2021 was prime time for Data breach. Also used the averages line in chart to highlight the average incidents occurred over the decade.

Prepared the **table** utilising the name of the organisations which are most affected by the breaches over the period of time categorising them by the states . One incident that happened in state of *Indiana , Anthem Inc.* Was the biggest Breach of health data by scale ever happened in the US till date.

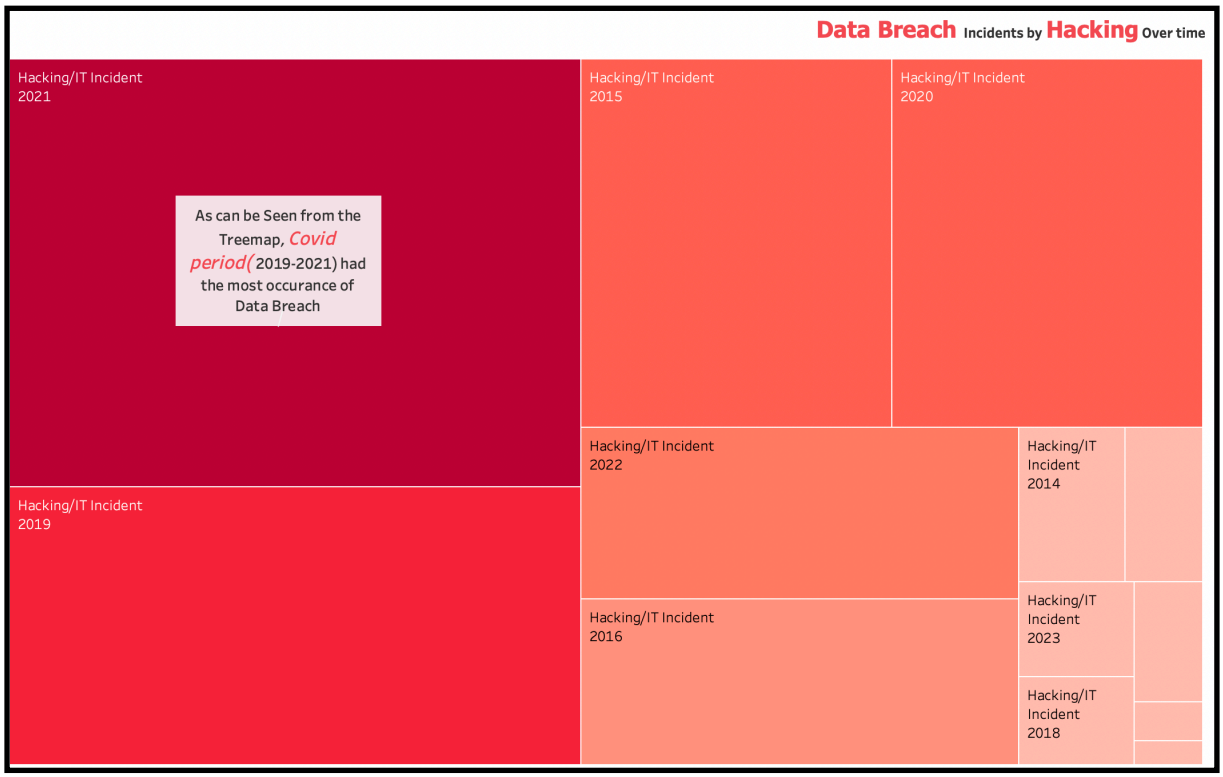
Entities By Breach which affected most of individuals

Name of Covered Entity	State	
Anthem Inc.	IN	7,88,02,023
Optum360, LLC	MN	1,15,00,000
Premera Blue Cross	WA	1,10,00,000
Laboratory Corporation of..	NC	1,02,51,784
Excellus Health Plan, Inc.	NY	93,58,891
Community Health Syste..	TN	61,21,158
Science Applications Inter..	VA	49,00,000
University of California, L..	CA	45,01,242
Community Health Syste..	TN	45,00,000
20/20 Eye Care Network, I..	FL	41,42,440

Used **geographical map** to show the geographical distribution of healthcare data breaches across the United States. This would help identify the certain regions which are more prone to breaches than others compared.



Utilised **vertical bar charts** to visualise the number of individuals affected by the breach and what were the reasons behind the breach. As can be seen from the bars, *Hacking/IT Incidents* is the entity mostly responsible behind the breaches, followed by theft and unauthorised access, which all can be protected if the correct policies to protect the individuals are in practise. Cybersecurity teams of the Entities must be more proactive in engaging with the stakeholders

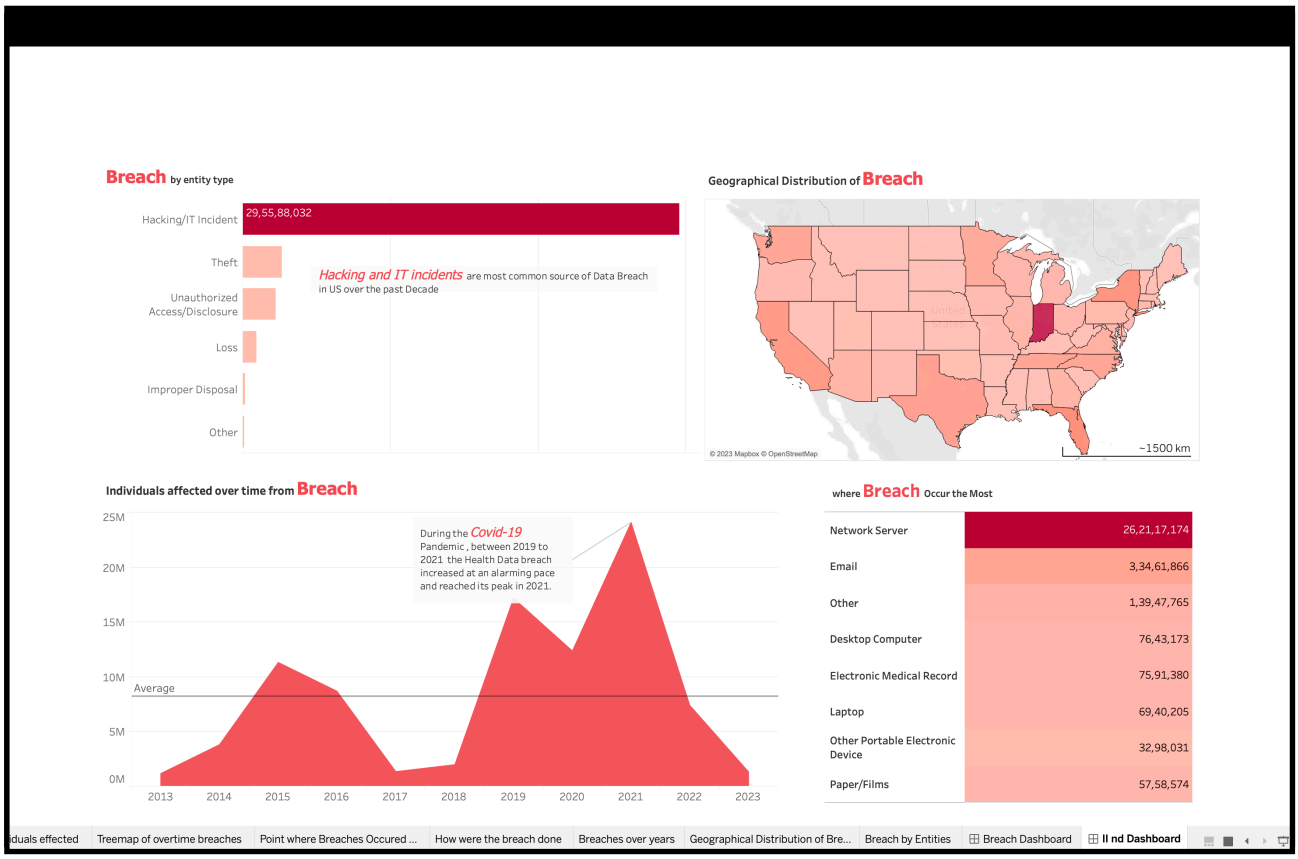


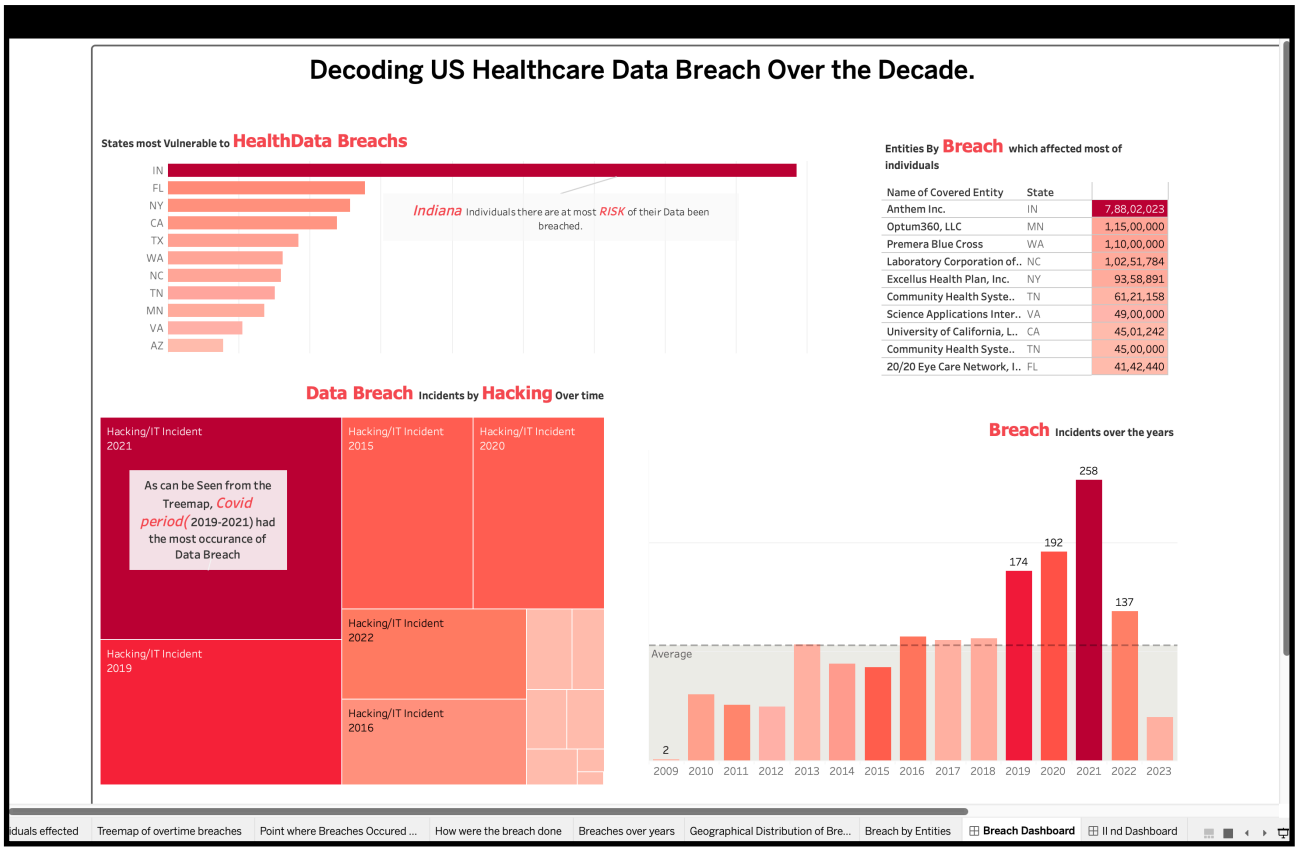
about the importance of the individuals privacy and need to protect it

Created the **tree-map,** specifically highlighting the *Hacking /IT Incidents* over the years as they were the leading cause behind the breach, again as I mentioned the rate of Hacking during the time of *Covid-19* was rampant and a lot of data was been Hacked.

My Final Dashboards

Below are the final Dashboard.I'm attaching the links of the Tableau Worksheet: https://public.tableau.com/app/profile/hamza.nadeem2614/viz/HealthcareDataBreachesinUS/Story1





Next Steps

The future direction of this project could involve several more complex steps:

Immediate Next Steps:

1. Expand the Dataset: The current dataset only includes major healthcare data breaches in the U.S. The project could be expanded to include minor breaches as well, or to include data breaches in other sectors or countries.
2. Incorporate Additional Variables: The current visualizations focus on the type, location, and impact of the breaches. Additional variables could be incorporated into the analysis, such as the response time to the breach, the measures taken to mitigate the breach, or the penalties imposed as a result of the breach.

More Complex Next Steps:

1. Predictive Analysis: The project could be developed into a predictive model that forecasts future healthcare data breaches based on historical trends and patterns. This could help healthcare entities and policymakers to anticipate and prevent future breaches.
2. Impact Assessment: A more in-depth analysis could be conducted to assess the impacts of data breaches on individuals and healthcare entities. This could involve linking the breach data with other datasets, such as patient health outcomes or healthcare entity financial data.

Improvements, Developments, or Alterations in Scope:

1. User Interface Improvements: The user interface of the visualizations could be improved to make them more intuitive and user-friendly. This could involve adding more interactive features, improving the visual design, or providing more detailed explanations and tooltips.
2. Real-Time Data Updates: The project could be developed to include real-time data updates, so that the visualizations always reflect the most recent data. This would require setting up a data pipeline to automatically update the dataset and refresh the visualizations.
3. Scope Alterations: The scope of the project could be altered to focus on a specific aspect of healthcare data breaches, such as breaches involving electronic health records, or breaches in a specific type of healthcare entity (e.g., hospitals, pharmacies, insurance companies). This would allow for a more detailed analysis of these specific areas.

Some potential future directions for healthcare data breach can be:

1. Integration of Machine Learning and AI: Incorporating machine learning and artificial intelligence techniques into data visualization tools can help identify patterns, trends, and anomalies more effectively. This could lead to the development of predictive models and early warning systems for healthcare data breaches.
2. Real-time Data Visualization: Developing real-time data visualization tools that can process and display data as it is generated can help healthcare professionals and organizations respond more quickly to emerging threats and trends.