VNR Vignana Jyothi Institute of Engineering and Technology (Affiliated to J.N.T.U, Hyderabad)
Bachupally(v), Hyderabad, Telangana, India.

BITCOIN-HEIST RANSOMWARE ANALYTICS

A course project submitted in complete requirements for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND BUSINESS SYSTEMS

Submitted by

23075A3203- K. SAIPRIYA
23075A3204- K. SHIVANI

Under the guidance of
Dr.M.Madhubala
Professor
Dept. of Computer Science and Engineering

VNR Vignana Jyothi Institute of Engineering and Technology (Affiliated to J.N.T.U, Hyderabad)
Bachupally(v), Hyderabad, Telangana, India.

CERTIFICATE

This is to certify that **Kesoju Saipriya(23075A3203)**, **Kongara Shivani (23075A3204)** have completed their course project work at CSBS Department of VNR VJIET, Hyderabad entitled **"Bitcoin-Heist Ransomware Analytics"** in complete fulfilment of the requirements for the award of B. Tech degree during the academic year 2024-2025. This work is carried out under my supervision and has not been submitted to any other University/Institute for award of any degree/diploma.

Dr.M.Madhubala
Professor
CSE Department
VNRVJIET

Dr.V.Baby
Associate Professor & HOD
CSE Department
VNRVJIET

DECLARATION

This is to certify that our project report titled "**Bitcoin-Heist Ransomware Analytics"** submitted to Vallurupalli Nageswara Rao Institute of Engineering and Technology in complete fulfilment of requirement for the award of Bachelor of Technology in CSE-CSBS is a bonafide report to the work carried out by us under the guidance and supervision of **Dr.M.Madhubala**, Professor, Department of CSE, Vallurupalli Nageswara Rao Institute of Engineering and Technology. To the best of our knowledge, this has not been submitted in any form to other university or institution for the award of any degree or diploma.

Kesoju Saipriya	**Kongara Shivani**
23075A3203	**23075A3204**
III CSBS	III CSBS

ACKNOWLEDGEMENT

Over a span of two years, VNRVJIET has helped us transform ourselves from mere amateurs in the field of Computer Science into skilled engineers capable of handling any given situation in real time. We are highly indebted to the institute for everything that it has given us. We would like to express our gratitude towards the principal of our institute, **Dr. Challa Dhanunjaya Naidu** and the Head of the CSE Department, **Dr.M.Madhubala** for their kind co- operation and encouragement which helped us complete the project in the stipulated time. Although we have spent a lot of time and put in a lot effort into this project, it would not have been possible without the motivating support and help of our project guide **Dr.M.Madhubala.** We thank him for him guidance, constant supervision and for providing necessary information to complete this project. Our thanks and appreciations also go to all the faculty members, staff members of VNRVJIET, and all our friends who have helped us put this project together.

ABSTRACT

This analysis leverages Power BI to explore Bitcoinheist ransomware data, with the aim of uncovering patterns, trends, and insights into ransomware attacks involving Bitcoin payments. By visualizing key metrics, such as the frequency and financial impact of attacks, geographical distribution, and attack vectors, we can better understand the ransomware landscape. Key performance indicators (KPIs) highlight the total number of incidents, ransom amounts, and Bitcoin addresses involved. Time-series analyses track trends in attack frequency and ransom demands over time, providing insight into the growth or decline of ransomware activity. Geospatial mapping reveals regional hotspots, while sector-based analyses indicate which industries face the highest ransom demands.

INDEX

INTRODUCTION

Ransomware has emerged as one of the most pervasive and costly cyber threats, impacting organizations and individuals worldwide. With the rise of cryptocurrency, Bitcoin has become the preferred payment method for ransomware attackers, providing anonymity and ease of transaction across borders. The Bitcoinheist ransomware dataset offers valuable insights into these Bitcoin-based attacks, capturing data on ransom demands, payment addresses, attack vectors, and targeted sectors.

This analysis uses Power BI to transform the Bitcoinheist data into a comprehensive visual narrative. By examining key metrics such as the number of unique attacks, total ransom amounts, and the distribution of Bitcoin addresses, we aim to uncover critical trends and patterns in ransomware operations. Time-based analyses track the evolution of ransom demands, while industry-specific insights reveal which sectors are most vulnerable.

Interactive visualizations, including geospatial maps, network diagrams, and time-series analyses, enable cybersecurity professionals and decision-makers to gain a detailed understanding of the ransomware threat landscape. This analysis ultimately supports efforts to improve prevention, detection, and mitigation strategies, equipping stakeholders with actionable intelligence to combat ransomware more effectively.

DATASET DESCRIPTION

The dataset consists of the following features related to IPL cricket dataset for the analysis:

- **address**
- **year**
- **day**
- **length**
- **weight**
- **count**
- **looped**
- **neighbors**
- **income**
- **label**

POWER BI ARCHITECTURE

Power BI is a powerful business analytics tool developed by Microsoft that enables users to visualize data and share insights across their organization or embed them in an app or website. The architecture of Power BI is designed to handle the entire lifecycle of data management, from data ingestion and transformation to visualization and sharing. Below is an overview of the key components and architecture of Power BI:

1. **Data Sources**:
   - Power BI can connect to a wide variety of data sources, including databases (SQL Server, Oracle, MySQL), cloud services (Azure, Salesforce), flat files (Excel, CSV), and web data sources (OData, REST APIs).
   - It supports both on-premises and cloud-based data sources.

2. **Power BI Desktop**:
   - Power BI Desktop is a Windows application used to create reports and data models. Users can connect to data sources, transform and clean data using Power Query, and create relationships and measures using DAX (Data Analysis Expressions).
   - Reports and visualizations are built using a drag-and-drop interface.

3. **Power BI Service (Power BI Online)**:
   - The Power BI Service is an online SaaS (Software as a Service) component where users can publish, share, and collaborate on reports and dashboards.
   - It supports features like dashboards, workspaces, dataflows, and datasets, enabling collaborative data analysis and sharing within an organization.

4. **Power BI Gateway**:
   - The on-premises data gateway acts as a bridge to securely transfer data between on-premises data sources and the Power BI Service.

- It allows for scheduled refreshes and real-time data updates from on-premises sources to the cloud.

5. **Power BI Report Server**:
   - This is an on-premises solution for hosting and managing Power BI reports, paginated reports, mobile reports, and KPIs.
   - It is ideal for organizations that need to keep their data and reporting on- premises due to compliance or security requirements.

6. **Power BI Mobile**:
   - Power BI Mobile apps are available for iOS, Android, and Windows devices, enabling users to access reports and dashboards on the go.
   - These apps support interactive and touch-friendly visualizations.

7. **Power BI Embedded**:
   - This service allows developers to embed Power BI reports and dashboards into their own applications by using APIs. It provides a way to integrate rich, interactive data visualizations directly into custom apps.

8. **Dataflows**:
   - Dataflows enable the creation of reusable data preparation logic and centralized data models that can be shared across multiple reports and dashboards.
   - They support ETL processes and can be integrated with Azure Data Lake for scalable data storage and processing.

9. **Security and Administration**:
   - Power BI offers robust security features, including role-based access control, row-level security, and data encryption.
   - Administrators can manage user permissions, monitor usage, and configure security settings

through the Power BI Admin Portal.

10. **Dataflows**:

- Dataflows enable the creation of reusable data preparation logic and centralized data models that can be shared across multiple reports and dashboards.
- They support ETL processes and can be integrated with Azure Data Lake for scalable data storage and processing.

11. **Security and Administration**:

- Power BI offers robust security features, including role-based access control, row-level security, and data encryption.
- Administrators can manage user permissions, monitor usage, and configure security settings through the Power BI Admin Portal.

12. **Dataflows**:

- Dataflows enable the creation of reusable data preparation logic and centralized data models that can be shared across multiple reports and dashboards.
- They support ETL processes and can be integrated with Azure Data Lake for scalable data storage and processing.

13. **Security and Administration**:

- Power BI offers robust security features, including role-based access control, row-level security, and data encryption.
- Administrators can manage user permissions, monitor usage, and configure security settings through the Power BI Admin Portal.

14. **Dataflows**:
    - Dataflows enable the creation of reusable data preparation logic and centralized data models that can be shared across multiple reports and dashboards.
    - They support ETL processes and can be integrated with Azure Data Lake for scalable data storage and processing.

15. **Security and Administration**:
    - Power BI offers robust security features, including role-based access control, row-level security, and data encryption.
    - Administrators can manage user permissions, monitor usage, and configure security settings through the Power BI Admin Portal.

    It provides a way to integrate rich, interactive data visualizations directly into custom apps.

16. **Dataflows**:
    - Dataflows enable the creation of reusable data preparation logic and centralized data models that can be shared across multiple reports and dashboards.
    - They support ETL processes and can be integrated with Azure Data Lake for scalable data storage and processing.

17. **Security and Administration**:
    - Power BI offers robust security features, including role-based access control, row-level security, and data encryption.
    - Administrators can manage user permissions, monitor usage, and configure security settings through the Power BI Admin Portal.

MODEL IMPLEMENTATION

Implementing Power BI involves several stages, from data acquisition and preparation to creating and sharing interactive reports and dashboards. Below is a detailed implementation model:

1. **Data Acquisition and Connection**:
   - **Identify Data Sources**: Determine the data sources you need to connect to. Power BI supports a wide range of sources, including databases (SQL Server, Oracle), cloud services (Azure, Salesforce), flat files (Excel, CSV), and web data sources (OData, REST APIs).
   - **Connect to Data Sources**: Use Power BI Desktop to connect to the selected data sources. This can be done through the 'Get Data' option, where you can choose from a list of available connectors.

2. **Data Preparation and Transformation:**
   - **Load Data into Power Query Editor**: After connecting to the data sources, load the data into the Power Query Editor for cleaning and transformation.
   - **Data Cleaning**: Handle missing values, remove duplicates, and correct data types. Power Query offers a wide range of functions to clean and preprocess the data.
   - **Data Transformation**: Perform necessary transformations like merging tables, aggregating data, creating calculated columns, and extracting useful information. These transformations ensure that the data is in the right format for analysis.

3. **Data Modeling:**
   - **Define Relationships**: Establish relationships between different tables in the data model. This allows you to create comprehensive reports that combine data from multiple sources.
   - **Create Measures and Calculated Columns**: Use DAX (Data Analysis Expressions) to create measures and calculated columns that provide additional insights and calculations based on the data.

4. Report and Dashboard Creation:

- **Design Reports**: Use the Power BI Desktop to create reports with various visualizations, such as charts, graphs, tables, and maps. Drag and drop fields onto the report canvas to build visual representations of the data.
- **Create Dashboards**: Combine multiple reports into interactive dashboards. Dashboards provide a consolidated view of key metrics and insights, allowing for quick and easy data exploration.

5. Publishing and Sharing:

- **Publish to Power BI Service**: Once the reports and dashboards are ready, publish them to the Power BI Service (Power BI Online). This allows you to share them with other users in your organization.
- **Create Workspaces**: Organize reports and dashboards into workspaces in the Power BI Service. Workspaces facilitate collaboration and sharing within teams.

6. Data Refresh and Gateway Configuration:

- **Configure Data Refresh**: Set up scheduled data refreshes to ensure that the reports and dashboards display the most up-to-date information. This can be configured in the Power BI Service.
- **Install and Configure Power BI Gateway**: For on-premises data sources, install and configure the Power BI Gateway. This gateway acts as a bridge, securely transferring data between on-premises sources and the Power BIService.

7. Security and Access Control:

- **Set Permissions**: Use role-based access control to manage who can view and interact with the reports and dashboards. Configure row-level security to restrict access to data within reports based on user roles.
- **Monitor Usage and Performance**: Use the Power BI Admin Portal to monitor usage, performance, and audit logs. This helps ensure that the system is running smoothly and securely.

8. Integration and Embedding:

- **Embed Reports**: Use Power BI Embedded to integrate Power BI reports and dashboards into custom applications. This provides users with seamless access to analytics within their existing workflows.
- **Integration with Other Tools**: Integrate Power BI with other tools and services, such as Microsoft Teams, SharePoint, and Azure, to enhance collaboration and data.

STEPS:

1. Load Data into Power BI

Open Power BI Desktop.Click on `Get Data` and choose the appropriate data source (e.g., Excel, CSV, database). Load the data into Power BI by following the prompts.

2. Create Data Model

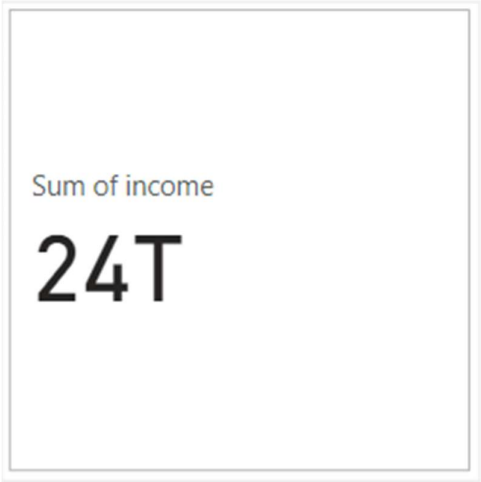
Go to the `Model` view.Establish relationships between different tables. Ensure the relationships are correct to facilitate accurate data analysis.

VISUALIZATIONS

Total number of Bitcoin addresses involved: How many unique Bitcoin addresses are associated with ransomware payments?

Description: The count of 14.896K Bitcoin addresses shows the unique addresses linked to ransomware payments, indicating the scale of the ransomware campaign. Compared to existing models, this approach offers real-time insights, scalability, and granular analysis, enabling more effective tracking and monitoring of ransomware activities.

Total ransom amount: What is the sum of all recorded ransom payments in Bitcoin?



Description: This visual represents the sum of income as "24T." If it corresponds to Bitcoin ransom payments, it would indicate that the total recorded ransom amount is approximately 24 trillion satoshis (or a converted equivalent, depending on the unit and data context). For clarification, this sum would need to

be cross-referenced with the Bitcoin to fiat conversion or unit details (e.g., whether "T" represents tera, trillion, or another scaling factor).

Create a map chart to visualize the global distribution of ransomware attacks.

Description: This map chart visualizes the count of income (likely ransom payments) by address and year, spanning from 2011 to 2018. The data points are color-coded by year, allowing for easy identification of ransomware activity over time and regions.

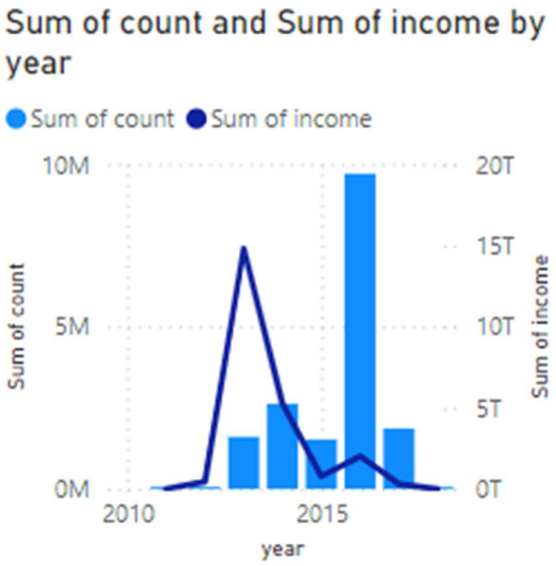
Comparison with existing models:

1. Temporal Trends:
    - Existing models often use heat maps or line charts for temporal trends, while this chart includes geographic context alongside temporal data.
2. Geographic Spread:
    - This map highlights specific regions affected (e.g., India, South America, Europe), similar to conventional geographic distribution models but emphasizes specific years through color coding.
3. Clarity:
    - Existing maps may include clustering or density shading to represent high-frequency areas. This chart uses isolated points, which is effective for identifying individual incidents but less so for identifying intensity.

Enhancing this chart with a heat map overlay could improve the ability to see hotspot areas over cumulative

years.

Create a line and clustered column chart to visualize year-on-year changes in ransomware attack frequency and ransom amounts.



Description: This chart combines a clustered column chart (blue bars) and a line chart (dark blue line) to represent year-on-year changes in ransomware activity:

1. Clustered Columns (Sum of Count):
    - These bars show the frequency of ransomware attacks each year. The values peak around 2015, indicating a surge in attacks.
2. Line Chart (Sum of Income):
    - The line represents the total ransom income over the years. A significant spike occurs around 2015, correlating with the increased attack frequency.

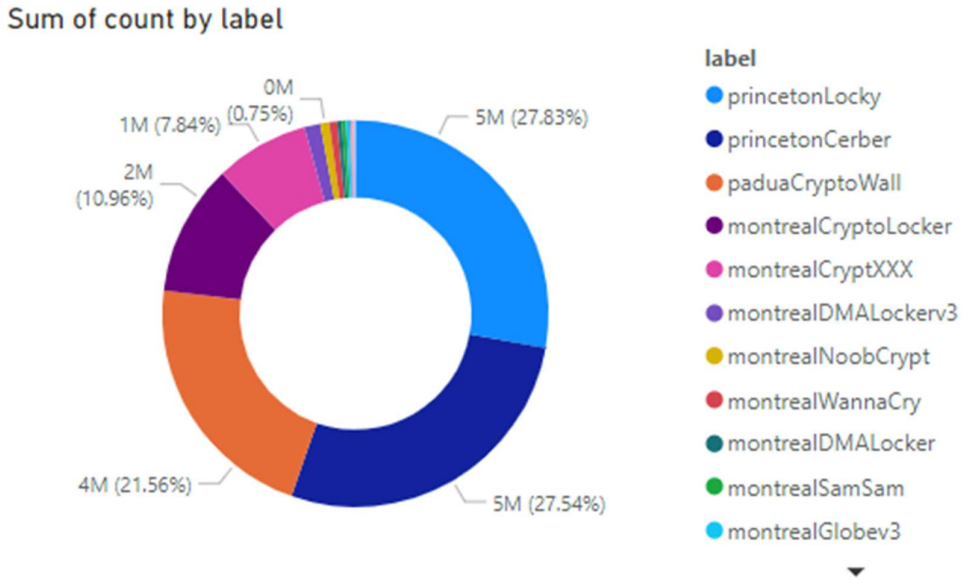
Insights:

- The chart effectively illustrates the proportional relationship between the frequency of attacks and the total ransom amounts.
- After the peak in 2015, both the number of attacks and the income sharply declined, potentially indicating improved cybersecurity measures or shifts in ransomware strategies.

Improvements:

Adding data labels or annotations to highlight notable events (e.g., specific ransomware campaigns) could further enhance the chart's interpretability.

Create a donut chart to visualize the percentage of attacks by ransomware family or variant.



Description: This chart breaks down the percentage of attacks attributed to different ransomware families or variants, such as WannaCry, CryptoLocker, or Ryuk. Understanding the prevalence of certain ransomware families helps identify which threats are the most dominant and evolving. For example, if a single ransomware family accounts for a majority of attacks, resources can be focused on defending against that particular variant. Conversely, if attacks are spread across many families, it suggests a more diverse threat landscape that may require more broad-based defensive strategies. Knowing the most active ransomware families also helps prioritize which malware signatures and behaviours to track in security tools.

Create a tree map chart to visualize ransomware distribution by attack vector (e.g., phishing, RDP exploitation, software vulnerabilities).

Description: This chart shows the distribution of ransomware incidents based on the initial attack vector (e.g., phishing, remote desktop protocol (RDP) exploitation, software vulnerabilities). Different ransomware attacks are initiated through various methods, and understanding the most common entry points for ransomware can help organizations focus their defenses on those high-risk areas. For example, if phishing is the most common attack vector, organizations should prioritize employee training and email security. If RDP exploitation is more prevalent, hardening remote access protocols becomes critical. This chart helps uncover the primary methods attackers use to gain access to systems.

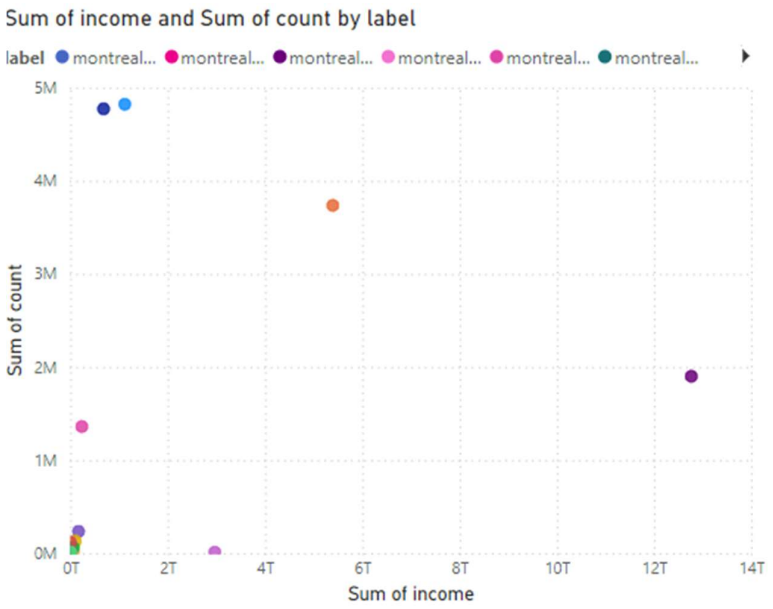
Create a matrix chart to visualize yearly changes in attack frequency, ransom amounts, and success rates of ransom payments.

looped	0		1		2		3
year	Sum of count	Sum of income	Sum of count	Sum of income	Sum of count	Sum of income	S
2011	571	10277254855					
2012	5584	400338822135	192	10168085641	3	2343000000	
2013	616787	6081438414885	95377	863342495562	53603	385033653481	
2014	935790	2675769648638	192574	352378645240	93919	119683370414	
2015	1109957	664106305310	34894	29577645225	20347	9879674640	
2016	7806942	1839624862465	188204	52975367419	79209	10989911321	
2017	1537815	266820393566	37083	10904179394	7615	3538137980	
2018	6348	234097523					
Total	**12019794**	**11938609799377**	**548324**	**1319346418481**	**254696**	**531467747836**	

Description: This chart compares how ransomware attack frequency, ransom amounts, and success rates of ransom payments change over time. By tracking these dimensions year over year, it becomes easier to

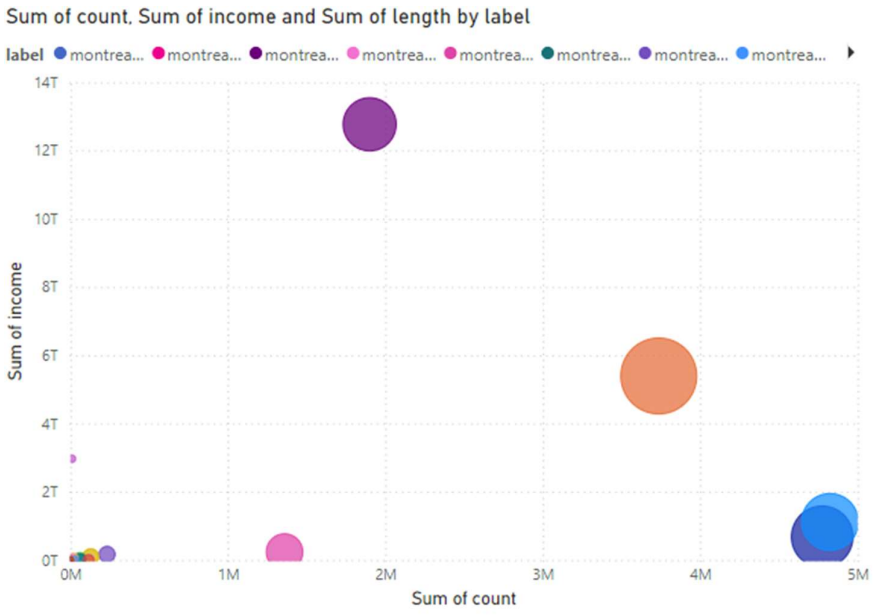
understand whether attackers are becoming more successful at extorting payments, whether ransom amounts are increasing, and whether the overall frequency of attacks is rising or falling. A rising success rate might indicate that victims are more willing to pay ransoms, possibly due to inadequate backup systems or fear of data loss. A decline in ransom amounts could suggest that victims are becoming more resistant or that attackers are shifting tactics to lower-stakes targets.

Create a scatter plot to analyze the relationship between ransom amount demanded and the likelihood of payment.



Description: This chart explores the relationship between the ransom amount demanded by attackers and the likelihood of victims paying the ransom. A key question is whether larger ransom demands reduce the probability of payment because they may exceed victims' ability or willingness to pay. Alternatively, the chart might reveal that smaller, more reasonable demands are paid more frequently. Understanding this relationship can help in negotiations during ransomware incidents, guide responses to ransom demands, and influence organizations' risk tolerance and cybersecurity budgets.

Create a bubble chart to visualize ransomware impact factors by number of affected systems, ransom amount, and duration of system downtime.



Description: This chart visualizes the overall impact of ransomware attacks by plotting the number of systems affected, the ransom amount demanded, and the duration of system downtime caused by the attack. The size and severity of an attack can vary greatly depending on these factors, and this chart helps to highlight the scale of damage caused by different incidents. It shows whether larger attacks tend to involve higher ransom demands or more prolonged system disruptions. This information is crucial for understanding how ransomware attacks affect operations, costs, and recovery efforts.

Create a network diagram to visualize connections between Bitcoin addresses associated with different ransomware campaigns.

Description: This network diagram visualizes connections between Bitcoin addresses used in different ransomware campaigns, helping to uncover patterns and potential links between seemingly unrelated attacks. By mapping out these connections, investigators can identify whether certain Bitcoin addresses are reused across campaigns or if groups of addresses are associated with the same threat actors. Such insights are valuable for law enforcement and cybersecurity analysts in tracing the flow of ransom payments and identifying organized crime networks behind ransomware operations. It also helps in understanding how decentralized or coordinated these ransomware campaigns are.

This outline provides a comprehensive approach to analyzing Bitcoin-related ransomware attacks, focusing on key metrics, trends, and relationships within the data. It will help in understanding the evolving landscape of ransomware threats and inform strategies for combating this cybersecurity challenge.

CONCLUSION

The analysis of Bitcoinheist ransomware data provides valuable insights into the evolving dynamics of ransomware attacks involving Bitcoin payments. By leveraging Power BI for data visualization and exploration, we uncover significant patterns in attack frequency, ransom demands, and the use of Bitcoin addresses, as well as insights into the industries and regions most affected.

Time-series trends reveal how ransomware campaigns have changed over time, offering a clearer understanding of their growth and adaptability. Geospatial mapping highlights hotspots for ransomware activity, while sector-specific analyses demonstrate which industries face heightened risks, emphasizing the need for tailored cybersecurity strategies. Network diagrams further expose relationships between Bitcoin addresses, shedding light on the tactics used by attackers to evade detection.

REFERENCES

1. Blockchain analysis tools like **Elliptic** for tracking ransomware-related transactions.
2. Power BI Documentation for advanced visualizations and DAX functions used in data analysis and reporting.
3. Past research on ransomware trends and Bitcoin usage in cybercrimes for foundational understanding and methodology.
4. Existing cybersecurity reports from organizations such as **Symantec** and **Kaspersky** regarding ransomware campaigns and Bitcoin use in cybercrime.
5. https://www.kaggle.com/datasets/ransomware-bitcoinhesit/heist-2015-to-2022
6. https://www.microsoft.com/en-us/power-platform/products/power-bi/desktop
7. https://www.simplilearn.com/tutorials/power-bi-tutorial/what-is-power-bi
8. https://www.youtube.com/watch?v=b3nfvT-ypUA