**HDSC August’22 Premiere Project by Team Tableau**

**Cyber Security Risk (2022 CISA Vulnerability)**

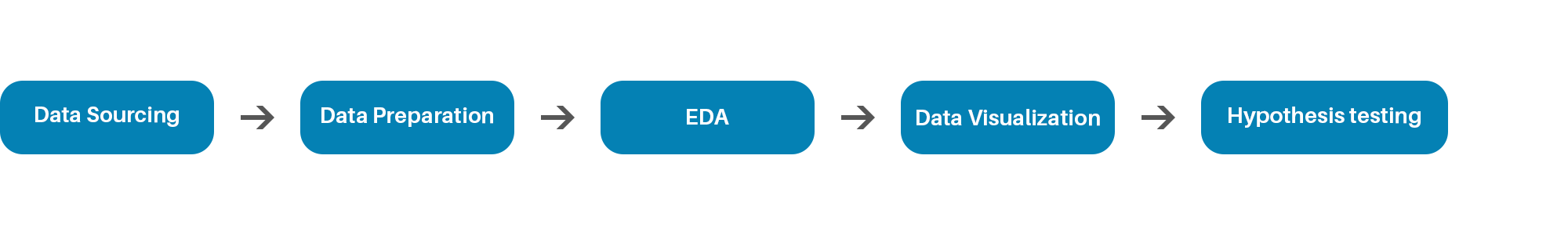
There has been an increase in cyber attacks recently, prompting efforts to mitigate cyber security risks and protect vital data. Without essential information about recent vulnerabilities or a proper understanding of available information on cyber security risk, it can be challenging to safeguard against these risks. The dataset obtained from the CISA Known Exploited Vulnerabilities catalog for 2022 offers valuable insights into the latest cyber security risks in the United States, including specific vulnerability names, targeted vendors and products, severity and complexity of vulnerabilities, and other pertinent details. This information, if properly organized and presented, will greatly aid in averting cyber security risks.

**Aim and Objectives**

The objective of this project is to prepare and analyze the available data to gain valuable insights for data-driven decision making. The project aims to achieve the following objectives:

* Analyze Vulnerability Severity Over Time: Measure the effectiveness of cyber security efforts by analyzing the severity of different security vulnerabilities over time.
* Hypothesis Testing: Perform hypothesis tests to evaluate specific assumptions related to cyber security risks.
* Exploratory Data Analysis: Conduct exploratory data analysis to understand the distribution of severity levels, average CVSS scores of products, vendor projects, and product names associated with security vulnerabilities.
* Visualization Dashboard: Develop a visualization dashboard that presents various metrics derived from the dataset. This dashboard will help security operatives make informed decisions by providing clear visual representations of the data.
* Recommendation Generation: Utilize the results of the analysis and hypothesis testing to generate actionable recommendations for security operatives. These recommendations will aim to enhance the cyber security posture and assist in avoiding potential risks.

**FLOW PROCESS**



**Data Sourcing**

The dataset for this project was obtained from Kaggle via the link below

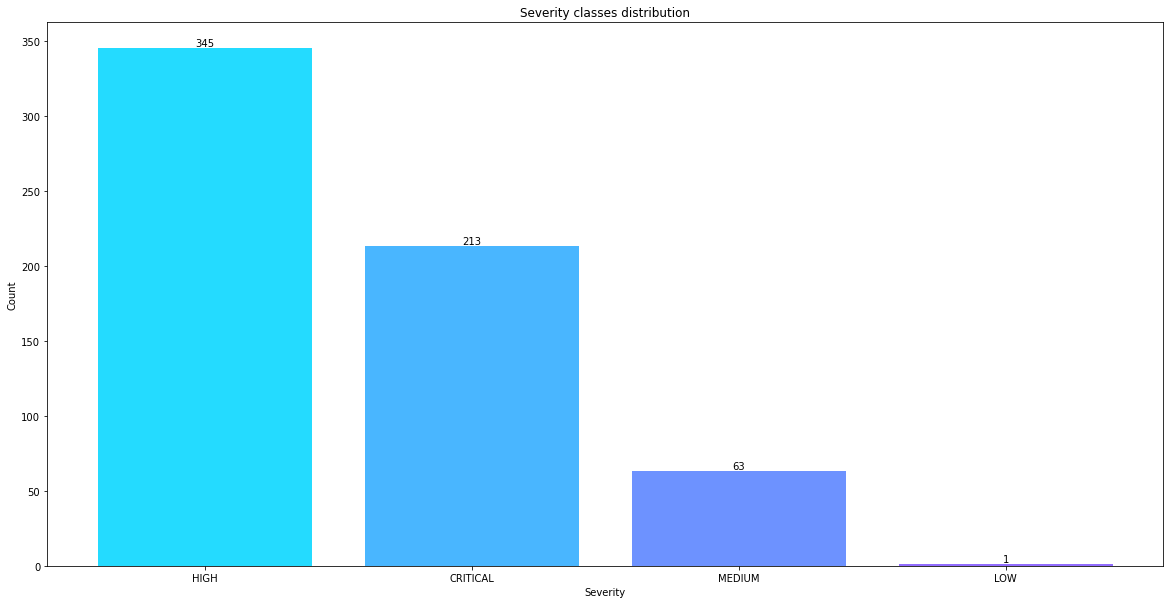
<https://www.kaggle.com/datasets/thedevastator/exploring-cybersecurity-risk-via-2022-cisa-vulne>

**Data Preparation:**

* Data Collection: In this stage, data on USA cyber security risks for 2022 were collected from Kaggle, a free data source platform. The collected data consisted of five different datasets.
* Data Discovery and Profiling: Each dataset underwent Pandas profiling. It was observed that all the datasets had the same number of columns and column headings. Additionally, it was found that most rows were identical across the five datasets. The profiling analysis also revealed that certain rows with a specific ID might have missing values in one dataset but not in another.
* Data Cleaning: The five datasets were merged into a single dataset, resulting in multiple duplicate rows. The combined dataset was sorted based on rows without missing values, and all duplicate rows were removed. It was also noted that the 'notes' column contained values in only one row, so it was eliminated from the dataset.
* Data Structuring: The data remained in the CSV (comma-separated values) format, which was the default format of the original data.
* Data Transformation: The data underwent further transformation to enhance its usability for analysis. The 'severity' and 'complexity' columns were converted into numerical data types, all the dates column in object data type were also converted to datetime data type and further modified to represent the days, months, and years.

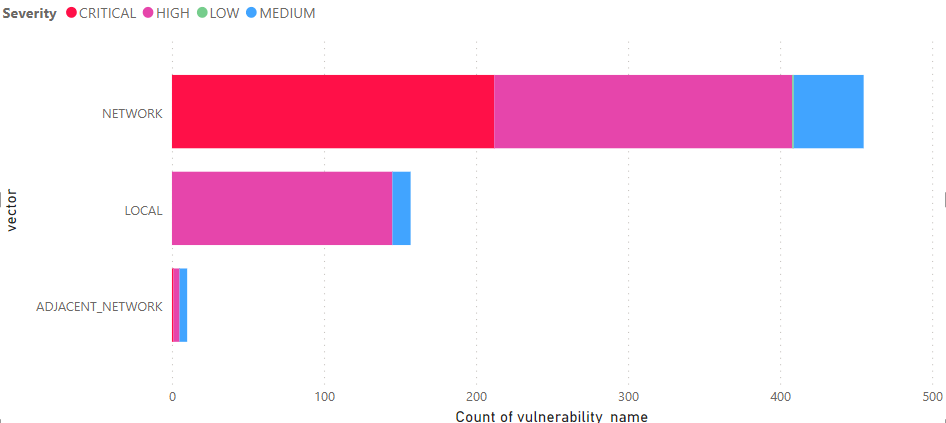
**Exploratory Data Analysis**The data was further explored to understand the distribution of the features associated with security vulnerabilities.

Firstly, a severity distribution bar chart was created to display the quantity of public exploits that are critical, high, medium, or low in terms of severity. It was clearly seen that most are highly severe.

****

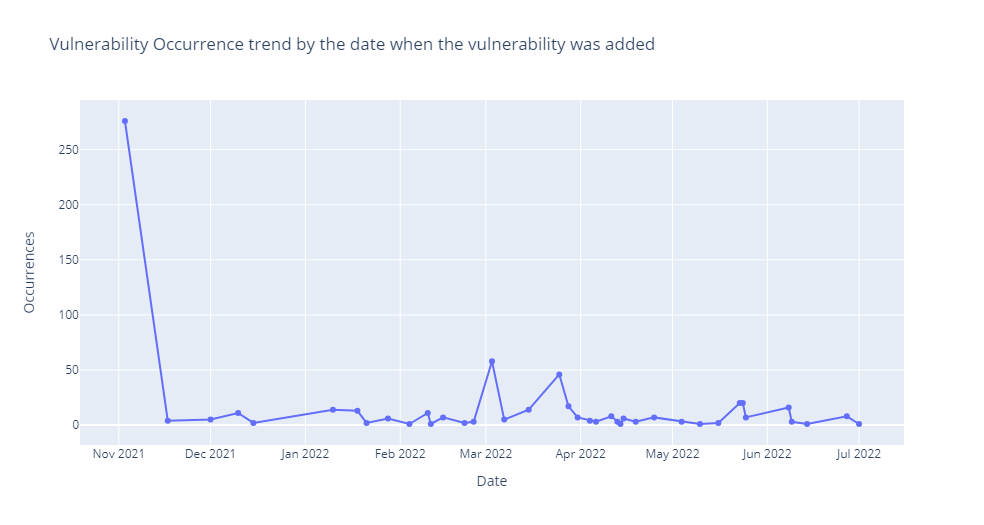
*Figure 1: Severity Distribution*

We proceed to compare the count of vulnerability names in relation to the unique vectors associated with the vulnerabilities in relation to the unique severities. It shows that the most frequent vector is “Network” and the most frequent severity associated to it is “Critical” closely followed by “High”.



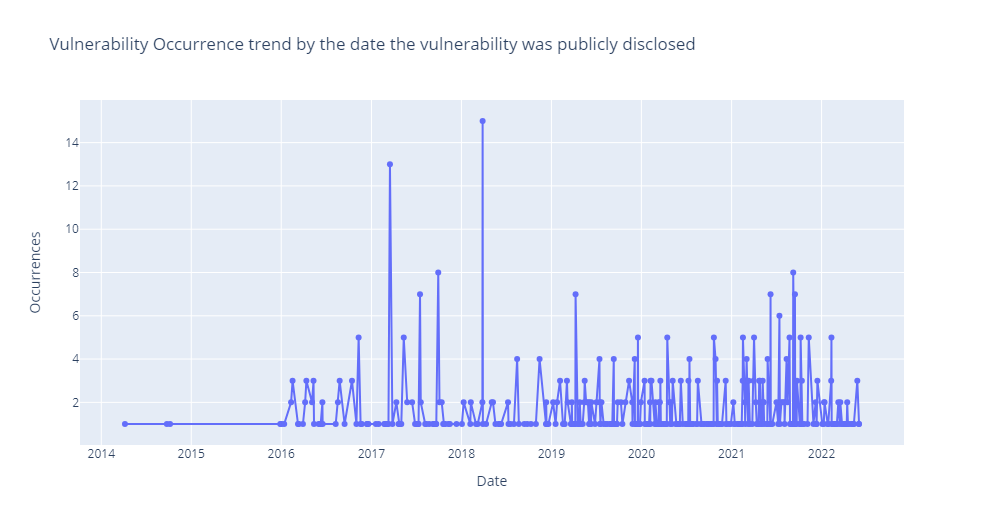
*Figure 2: Vulnerability name by Vector based on severity*

Then We decided to use plotly line graph to visualize the vulnerability occurrences in relation to the date added, Nov. 3rd, 2021 stood out followed by Mar. 3rd, 2022 and Mar. 25th, 2022.

**

*Figure 3: Vulnerability Occurrences over time by date added*

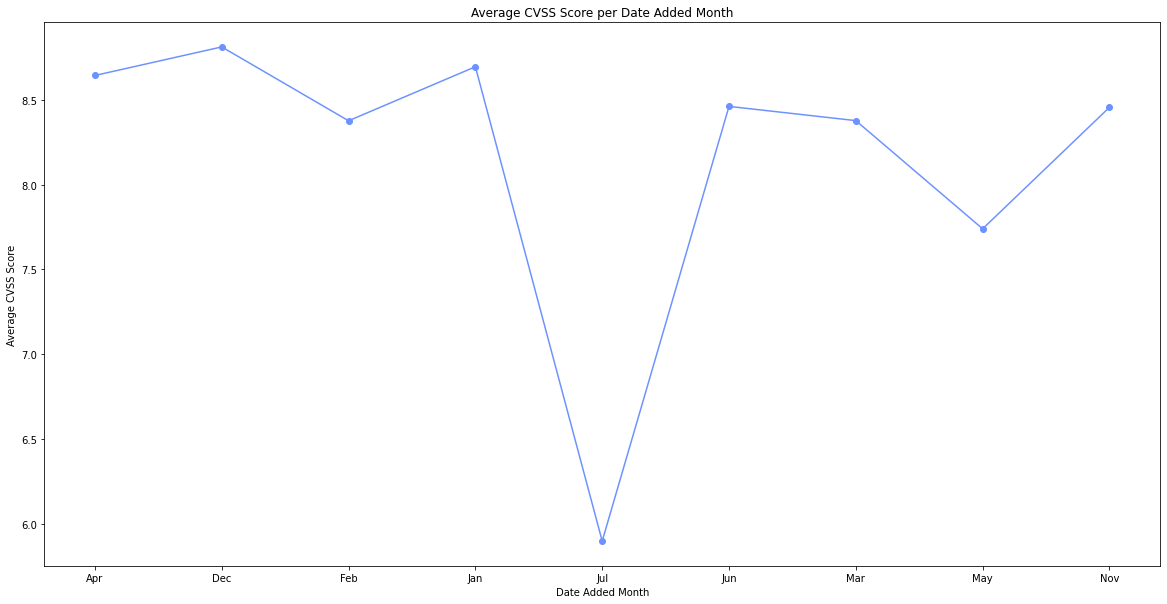
We utilized Plotly line graphs to visually represent the occurrence of vulnerabilities based on the dates they were publicly disclosed. March 2018 and March 2017 had the highest number of occurrences, respectively. These line graphs provide security operatives with the ability to investigate the events that took place during those peak periods and take preventive measures to avoid future incidents.



*Figure 4: Vulnerability occurrence by date the vulnerability was publicly disclosed*

In addition to the visualization of vulnerability occurrences, we also examined the average CVSS (Common Vulnerability Scoring System) scores based on the month the vulnerabilities were added. Our analysis revealed that December had the highest CVSS score, with June following closely behind. On the other hand, July recorded the lowest average CVSS score.

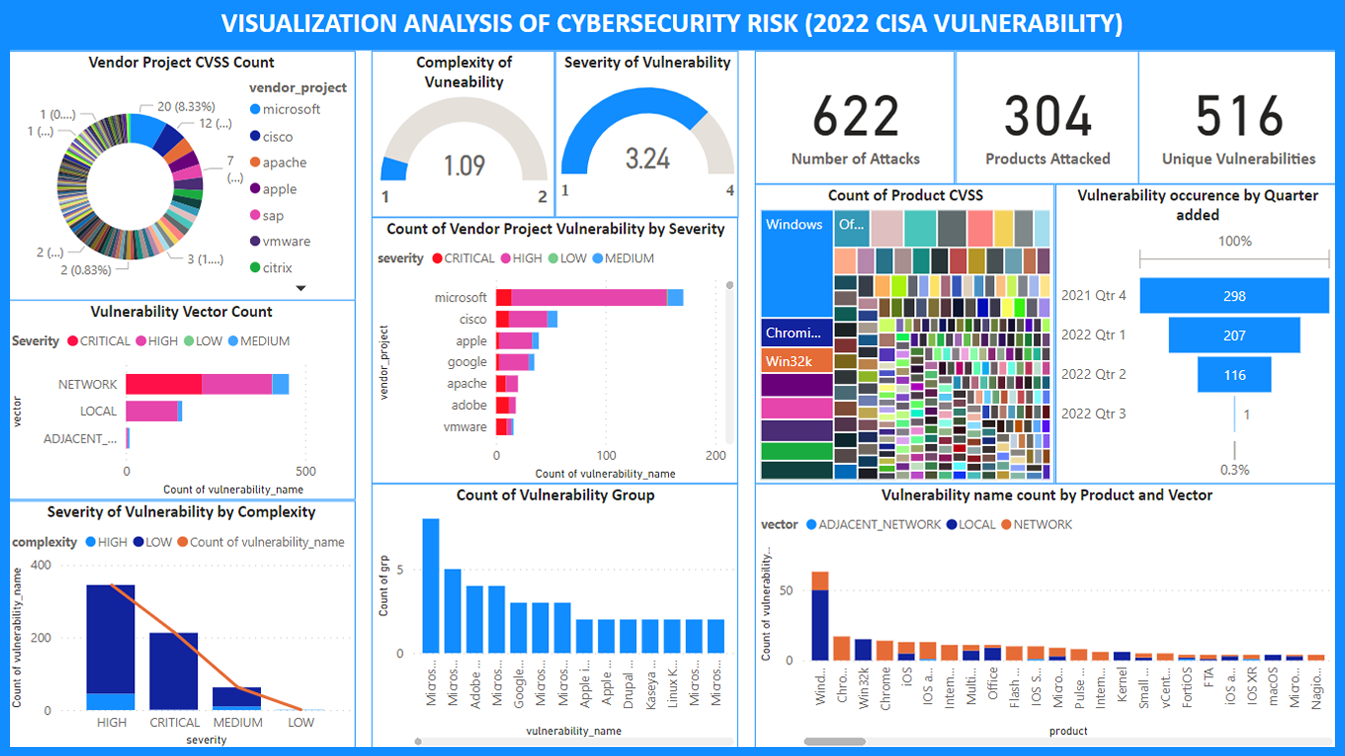
This information can greatly assist security operatives in several ways. First, it helps them identify the months during which vulnerabilities with higher severity levels tend to be added. This insight enables them to allocate resources and prioritize security measures accordingly, focusing more attention on the months with historically higher CVSS scores. By doing so, they can proactively address potential threats and mitigate the risk associated with vulnerabilities.

**

*Figure 5: Average CVSS Score per Date Added Month*

**Data Visualization Dashboard**

We utilized Power BI to create a dynamic visualization dashboard showcasing diverse data metrics and relationships.



*Figure 6: PowerBI Visualization Dashboard*

**Hypothesis Testing**

We conducted an in-depth statistical analysis to gain deeper insights into the data and provide meaningful recommendations. To achieve this, we tested eleven hypotheses, each focusing on different aspects of vulnerability severity and related factors. By testing the hypotheses, we aimed to gain a deeper understanding of the relationships and patterns within the data. The results of these tests will provide valuable insights for making informed recommendations and strategic decisions related to vulnerability management, prioritization, and mitigation efforts.

**Results**

Based on our analysis, we obtained the following results:

* Correlation: There is a strong positive correlation between the CVSS score and severity level of vulnerabilities. This implies that as the CVSS score increases, the severity level also tends to be higher.
* Date Added: Vulnerabilities discovered at an earlier date are more likely to have a higher severity level. This suggests that prompt identification and resolution of vulnerabilities are crucial in mitigating their potential impact.
* CWE Category: Certain CWE categories are found to be more likely to have a higher severity level. This highlights the importance of understanding the specific vulnerabilities associated with these categories and implementing appropriate preventive measures.
* Vendor Projects: Our analysis indicates that certain vendor projects are more likely to have a higher severity level. This finding underscores the significance of closely monitoring vulnerabilities related to these projects and prioritizing remediation efforts accordingly.
* Complexity Level: Vulnerabilities with a higher complexity level are more likely to have a higher severity level. This emphasizes the need for more comprehensive and intricate actions to address these complex vulnerabilities effectively.
* CVSS Score and Complexity: Interestingly, vulnerabilities with higher CVSS scores tend to require less complex actions to address them. On average, vulnerabilities with low CVSS scores have an average complexity level of 2.0. This suggests that while these vulnerabilities may have lower severity levels, they still demand attention and appropriate measures for mitigation.

**Conclusion and Recommendation**

While the presented results provide valuable insights, it's important to note that a comprehensive analysis and decision-making process benefit from a larger dataset. Expanding the data collection will enable deeper insights and a more accurate understanding of the relationships and trends within the vulnerabilities.

By acquiring more data, the company can enhance their decision-making process by:

* Validating and strengthening existing findings
* Discovering additional patterns and relationships
* Making more accurate predictions and forecasts
* Conducting deeper analysis on specific subgroups