M2 Project Deliverable 1

Common software weaknesses in CISAs Known Exploitable Catalog

**Focus Area Background:**

I have been working as a cybersecurity engineer for a few years now, and I want my project to be cybersecurity related. My project will source data from the Cybersecurity and Infrastructure Security Agency (CISA), the National Institute of Standards and Technology (NIST), and the Massachusetts Institute of Technology Research and Engineering (MITRE). Cybersecurity specific terms such as Common Vulnerabilities and Exposures (CVE), Common Weakness Enumeration (CWE), and Common Vulnerability Scoring System (CVSS) will be frequently used. A brief description of each is as follows:

* CVEs are managed by NIST and provide a public method of sharing information about specific cybersecurity vulnerabilities.
* The CWEs database is managed by MITRE and is a category system for types of hardware and software weaknesses and vulnerabilities.
* CVSS is a way to evaluate and rank reported vulnerabilities in a standardized and repeatable way, as determined by the Forum of Incident Response and Security Teams (FIRST).

This project will investigate how CISAs catalog of known exploited vulnerabilities relate to information published in NIST and MITRE databases. This catalog is a collection of vulnerabilities frequently exploited by malicious cyber actors against federal enterprises. Vulnerabilities published in this catalog, by operational directive mandate, must be addressed by Federal Civilian Executive Branch agencies (FCEB) by prescribed due dates.

**Data Acquisition and Question Set:**

Focusing on CVEs published in CISAs known exploitable catalog narrows the scope of the project such that it can be completed during this course. The first requirement for my project was defining a reasonably scoped hypothesis, and CISAs known exploitable catalog provides a suitable foundation for this. By default, the catalog does not include any CWE or CVSS information for each CVE. Instead, I created a Python script to scrape and parse data from the NISTs CVE database and MITREs CWE database. This enriched the known exploitable CVEs with their respective CVSS scores and CWE IDs. The resulting dataset included the default columns provided by CISA, and "CWE", "Phishing", "base\_score\_CVSS", "Severity", and "CVSS\_Version". I also added a formula to compute the "timeToFix" for each vulnerability. See appendix A for a sample of the data since Blackboard does not process .csv files correctly.

With this more comprehensive data set, several questions can possibly be answered. Which CWEs are most frequently associated with the vulnerabilities in this dataset? What types of software weaknesses do these prevalent CWEs represent? What implications arise about commonly exploited software weaknesses from these findings? Are specific vendors or products more frequently associated with certain CWEs? All of these relate to my hypothesis; *Are certain software weaknesses, as defined by CWEs, more prevalent within CISAs known exploitable catalog than others?* Investigating this may lead to interesting outcomes regardless of the answer. The most common CWEs in the known exploitable dataset can highlight reoccurring software weaknesses. A recurring vulnerability category within the dataset suggests a pressing need for federal enterprises to prioritize countermeasures against those vulnerabilities, and for software developers to limit these weaknesses during development. If there are no standout CWEs within the dataset, then this provides an interesting answer, and suggests that FCEBs must have robust cyber systems that address a variety of vulnerabilities.

**Initial Impressions and Concerns:**

Following the successful creation and deployment of my python parsing scripts, I had an enriched form of the known exploitable catalog. While CISA is the sole publisher of this catalog, the CVE numbers and CWE IDs within can be cross-referenced with data from NIST and MITRE. Both these entities are highly regarded in the cybersecurity domain. However as I analyzed my enriched data set, certain concerns emerged:

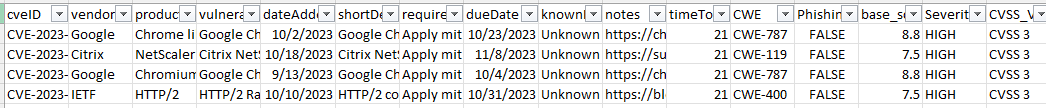
* The data is not free from missing values
* The presence of two CVSS versions poses challenges, especially when converting CVSSv2 base scores to CVSSv3.1, given the lack of an industry-standard conversion methodology.
* Some 'timeToFix' values appear negative, which might be a result of CISA setting due dates before the identification of certain CVEs.
* There is a high count of the 'none' category in the CWE column. I aim to discern the cause behind this and, if possible, map these CVEs to appropriate CWEs to enhance dataset completeness.

I will have to research and explain how I end up converting CVSSv2 scores to CVSSv3.1 socre for the next part of the project. Currently, I plan to on feeding the component values from CVSSv2 base scores into a CVSSv3.1 calculator. For now, I have simply added a "CVSS\_Version" column to the data set. Additionally, some of the 'timeToFix' values are negative because CISA prescribed their due dates for dates that were before the CVE was identified. I need to investigate this more. The most pressing issue with the dataset I have relates to the CWEs. My analysis relies on the CWE associated with each vulnerability, but the 'none' category in the CWE column has the highest count of CWE in the dataset. I need to determine the reason for this, and hopefully, I can appropriately categorize these CVEs to match CWEs. My initial assumption is that the CVEs were not fully understood at the time the CVE was published, and NIST never went back to link these CVEs to CWEs.

**Conclusion:**

My project will explore the relationship between CVEs in CISAs known exploitable catalog and MITREs CWE to better understand what type of vulnerabilities pose the most risk to FCEBs. As my project progresses, addressing the challenges described above and refining the dataset will be paramount. The insights gained from this analysis could potentially guide federal agencies in prioritizing countermeasures and assist software developers in preempting common software vulnerabilities.

**Appendix A:**

The following is a small sample of the enriched CISA known exploitable catalog since Blackboard does not display the attached .csv correctly.