

INST 754

Final Project

**Cyber Risk Insight: Analyzing
IT Vulnerabilities and
Predicting Threat Periods**

**Pranav Adiraju
Srikanth Parvathala**

Cyber Security Overview

- CVE - Common Vulnerabilities and Exposures
- CVEs are a fundamental part of many cybersecurity practices and are used in numerous security products and services.
- They are essential for understanding the landscape of cybersecurity threats and for ensuring that systems are protected against known vulnerabilities.

Dataset

- The Common Vulnerabilities and Exposures (CVE) are publicly available datasets that provide references to all the publicly-known information security vulnerabilities and exposures.
- It is maintained by The National Cybersecurity FFRDC operated by the MITRE corporation.
- The analysis will utilize four datasets: CVE, Products, Vendor_Product, and Vendors. These datasets provide detailed information on known software vulnerabilities, including their severity levels, which are measured using the Common Vulnerability Scoring System (CVSS).

ETL Process

- We applied transformations to each dataset using Power BI Query Editor. This involved assigning headers, renaming columns, and modifying data types.
- Additionally, we merged the Products and Vendors datasets using 'cve_id' as the key column and subsequently integrated this merged dataset with the CVE dataset, again using 'cve_id' as the joining column.
- After completing the Extract, Transform, Load (ETL) process, we conducted data cleaning by removing null values from columns. This process yielded a refined dataset containing 241,979 rows and 15 columns.

ETL Process Steps

The screenshot displays the Power Query Editor interface. The main area shows a table with columns: `cve_id`, `mod_date`, `pub_date`, `cvss`, `cwe_id`, and `cwe_name`. The table contains 17 rows of CVE data. The formula bar at the top shows the query: `Table.RenameColumns(#"Changed column type", {"", "cve_id"})`. The right-hand pane, titled 'Query settings', shows the 'Applied steps' list with 'Renamed columns' selected.

Power Query Editor

Home Transform Add column View Help

Close & load Get data Enter data Options Manage parameters Refresh Properties Advanced editor Manage Choose columns Remove columns Keep rows Remove rows Filter rows Sort Split column Group by Data type: Text Use first row as headers Replace values Combine Export template Share

Queries [4]

- cve
- products
- vendors
- merged_cve_dataset

	cve_id	mod_date	pub_date	cvss	cwe_id	cwe_name
1	CVE-2019-165...	11/21/2019, 3:15:00 PM	11/21/2019, 3:15:00 PM	6.8	352	Cross-Site Request Forgery (CSRF)
2	CVE-2019-165...	11/21/2019, 3:15:00 PM	11/21/2019, 3:15:00 PM	4	732	Incorrect Permission Assignment for Critical Resource
3	CVE-2019-165...	11/21/2019, 3:15:00 PM	11/21/2019, 3:15:00 PM	4.3	639	Authorization Bypass Through User-Controlled Key
4	CVE-2013-2092	11/20/2019, 9:22:00 P...	11/20/2019, 9:15:00 PM	4.3	79	Improper Neutralization of Input During Web Page Generation ('Cross-site Scripting')
5	CVE-2013-2091	11/20/2019, 8:15:00 PM	11/20/2019, 8:15:00 PM	7.5	89	Improper Neutralization of Special Elements used in an SQL Command ('SQL Injection')
6	CVE-2013-1817	11/20/2019, 8:15:00 PM	11/20/2019, 8:15:00 PM	5	200	Information Exposure
7	CVE-2013-1816	11/20/2019, 8:15:00 PM	11/20/2019, 8:15:00 PM	5	20	Improper Input Validation
8	CVE-2012-1257	11/20/2019, 8:15:00 PM	11/20/2019, 8:15:00 PM	2.1	319	Cleartext Transmission of Sensitive Information
9	CVE-2011-4455	11/20/2019, 8:10:00 PM	11/20/2019, 7:15:00 PM	4.3	79	Improper Neutralization of Input During Web Page Generation ('Cross-site Scripting')
10	CVE-2011-4454	11/20/2019, 8:10:00 PM	11/20/2019, 7:15:00 PM	4.3	79	Improper Neutralization of Input During Web Page Generation ('Cross-site Scripting')
11	CVE-2010-4659	11/20/2019, 5:48:00 ...	11/20/2019, 5:15:00 PM	4.3	79	Improper Neutralization of Input During Web Page Generation ('Cross-site Scripting')
12	CVE-2010-4660	11/20/2019, 4:15:00 PM	11/20/2019, 4:15:00 PM	7.5	20	Improper Input Validation
13	CVE-2013-0195	11/20/2019, 3:15:00 PM	11/20/2019, 3:15:00 PM	4.3	79	Improper Neutralization of Input During Web Page Generation ('Cross-site Scripting')
14	CVE-2013-0194	11/20/2019, 3:15:00 PM	11/20/2019, 3:15:00 PM	4.3	79	Improper Neutralization of Input During Web Page Generation ('Cross-site Scripting')
15	CVE-2013-0193	11/20/2019, 3:15:00 PM	11/20/2019, 3:15:00 PM	4.3	79	Improper Neutralization of Input During Web Page Generation ('Cross-site Scripting')
16	CVE-2012-6136	11/20/2019, 3:15:00 PM	11/20/2019, 3:15:00 PM	4.9	276	Incorrect Default Permissions
17	CVE-2011-1028	11/20/2019, 3:15:00 PM	11/20/2019, 3:15:00 PM	7.5	20	Improper Input Validation

Query settings

- Properties
- Name: cve
- Applied steps
 - Source
 - Promoted headers
 - Changed column type
 - Renamed columns

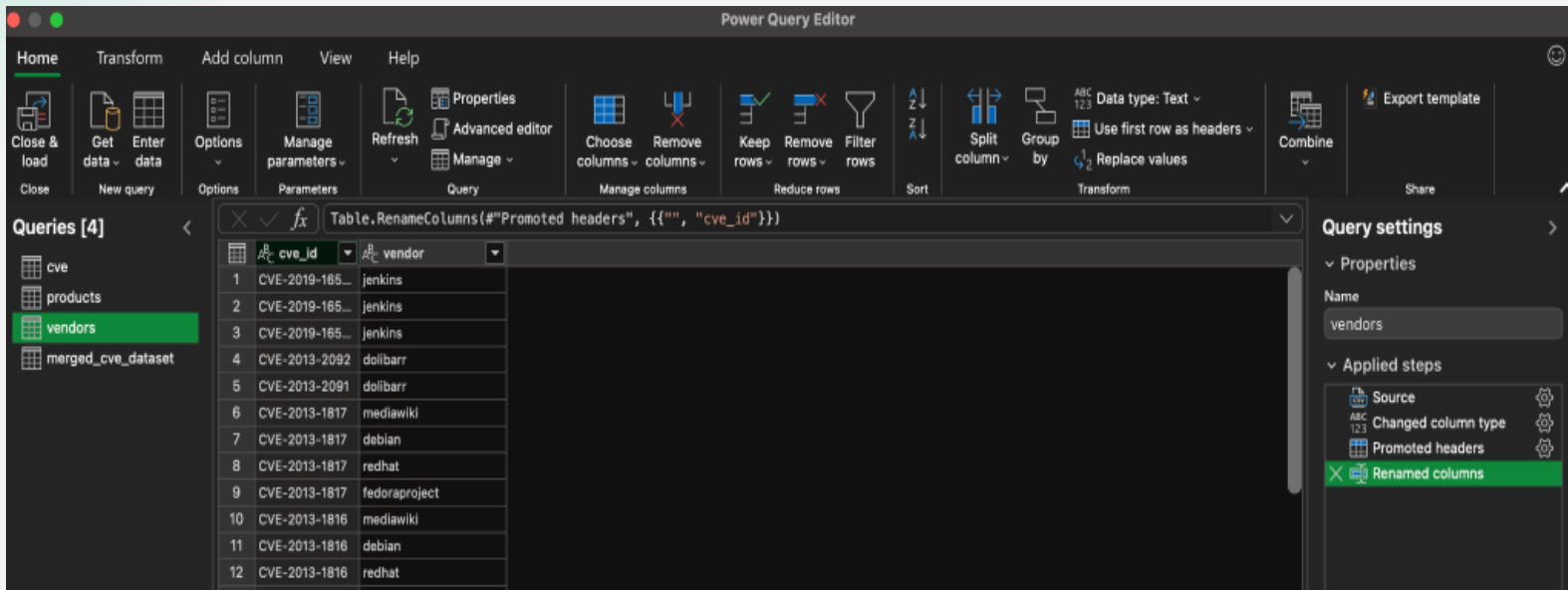
ETL Process Steps

The screenshot displays the Power Query Editor interface, which is used for data transformation and loading. The interface is divided into several sections:

- Home Tab:** Contains icons for 'Close & load', 'Get data', 'Enter data', 'Options', 'Manage parameters', 'Refresh', 'Advanced editor', 'Manage', 'Choose columns', 'Remove columns', 'Keep rows', 'Remove rows', 'Filter rows', 'Sort', 'Split column', 'Group by', 'Data type: Text', 'Use first row as headers', 'Replace values', 'Combine', and 'Export template'.
- Queries [4]:** A list of queries on the left side, including 'cve', 'products' (highlighted), 'vendors', and 'merged_cve_dataset'.
- Table View:** The main area showing a table with columns 'cve_id' and 'vulnerable_product'. The table contains 9 rows of data, including CVE-2019-165... and CVE-2013-2092.
- Query settings:** A panel on the right showing the 'Properties' section with the name 'products' and the 'Applied steps' section with 'Source' and 'Promoted headers' (highlighted).

The formula bar at the top of the table view shows the query formula: `Table.PromoteHeaders(Source, [PromoteAllScalars = true])`.

ETL Process Steps



The screenshot displays the Power Query Editor interface. The top ribbon includes tabs for Home, Transform, Add column, View, and Help. The 'Transform' tab is active, showing various data manipulation options like 'Close & load', 'Get data', 'Enter data', 'Options', 'Manage parameters', 'Refresh', 'Advanced editor', 'Manage', 'Choose columns', 'Remove columns', 'Keep rows', 'Remove rows', 'Filter rows', 'Sort', 'Split column', 'Group by', 'Data type: Text', 'Use first row as headers', 'Replace values', 'Combine', and 'Export template'.

On the left, the 'Queries [4]' pane lists four queries: 'cve', 'products', 'vendors' (highlighted), and 'merged_cve_dataset'.

The main workspace shows a table with the following data:

	cve_id	vendor
1	CVE-2019-165...	jenkins
2	CVE-2019-165...	jenkins
3	CVE-2019-165...	jenkins
4	CVE-2013-2092	dolibarr
5	CVE-2013-2091	dolibarr
6	CVE-2013-1817	mediawiki
7	CVE-2013-1817	debian
8	CVE-2013-1817	redhat
9	CVE-2013-1817	fedoraproject
10	CVE-2013-1816	mediawiki
11	CVE-2013-1816	debian
12	CVE-2013-1816	redhat

The formula bar at the top of the workspace shows the M code: `Table.RenameColumns(#"Promoted headers", {{{"", "cve_id"}}})`.

On the right, the 'Query settings' pane is open, showing the 'Properties' section with the name 'vendors' and the 'Applied steps' list. The 'Applied steps' list includes 'Source', 'Changed column type', 'Promoted headers', and 'Renamed columns' (highlighted with a green background).

ETL Process Steps

The screenshot displays the Power Query Editor interface. The top ribbon includes tabs for Home, Transform, Add column, View, and Help. The Home tab is active, showing various options like Close & load, Get data, Enter data, Options, Manage parameters, Refresh, Advanced editor, Manage, Choose columns, Remove columns, Keep rows, Remove rows, Filter rows, Sort, Split column, Group by, Data type, Use first row as headers, Replace values, Combine, and Export template.

The main area shows a table with the following columns: cve_id, vulnerable_produ..., vendor, modified_date, published_date, 1.2 cv..., 1.2 cwe_co..., and cwe_name. The table contains 13 rows of data, including CVE-2019-2211, CVE-2019-2212, CVE-2019-2213, CVE-2019-2214, CVE-2019-187..., CVE-2019-186..., CVE-2019-169..., CVE-2019-186..., CVE-2019-186..., CVE-2019-186..., CVE-2019-188..., CVE-2012-4384, and CVE-2012-4384.

The right sidebar shows the Query settings panel, which includes the Properties section (Name: merged_cve_dataset) and the Applied steps section. The Applied steps list includes Source, Expanded vendors, Merged queries, Expanded cve, Filtered rows, and Renamed columns (highlighted in green).

Queries [4]

- cve
- products
- vendors
- merged_cve_dataset

Table: RenameColumns("#Filtered rows", {"mod_date", "modified_date"}, {"pub_date", "published_date"})

	cve_id	vulnerable_produ...	vendor	modified_date	published_date	1.2 cv...	1.2 cwe_co...	cwe_name
1	CVE-2019-2211	android	google	11/14/2019, 9:36:00 ...	11/13/2019, 6:15:00 PM	7.8	89	Improper Neutralization of S
2	CVE-2019-2212	android	google	11/14/2019, 9:30:00 ...	11/13/2019, 6:15:00 PM	4.9	200	Information Exposure
3	CVE-2019-2213	android	google	11/14/2019, 9:24:00 ...	11/13/2019, 6:15:00 PM	6.9	416	Use After Free
4	CVE-2019-2214	android	google	11/14/2019, 9:19:00 ...	11/13/2019, 6:15:00 PM	7.2	269	Improper Privilege Managem
5	CVE-2019-187...	parallels_plesk_panel	parallels	11/14/2019, 9:14:00 ...	11/13/2019, 8:15:00 PM	4.3	79	Improper Neutralization of Ir
6	CVE-2019-186...	ng_firewall	untangle	11/14/2019, 8:57:00 ...	11/14/2019, 3:15:00 PM	6.5	89	Improper Neutralization of S
7	CVE-2019-169...	web_chat	enghouse	11/14/2019, 8:45:00 ...	11/13/2019, 7:15:00 PM	4.3	79	Improper Neutralization of Ir
8	CVE-2019-186...	ng_firewall	untangle	11/14/2019, 8:37:00 ...	11/14/2019, 3:15:00 PM	9	74	Neutralization of Special Ele
9	CVE-2019-186...	ng_firewall	untangle	11/14/2019, 8:23:00 ...	11/14/2019, 3:15:00 PM	3.5	79	Improper Neutralization of Ir
10	CVE-2019-186...	ng_firewall	untangle	11/14/2019, 8:19:00 ...	11/14/2019, 3:15:00 PM	3.5	79	Improper Neutralization of Ir
11	CVE-2019-188...	lavalite	lavalite	11/14/2019, 7:57:00 ...	11/13/2019, 8:15:00 PM	4.3	79	Improper Neutralization of Ir
12	CVE-2012-4384	letodms	trillexnet	11/14/2019, 7:23:00 ...	11/13/2019, 4:15:00 PM	4.3	79	Improper Neutralization of Ir
13	CVE-2012-4384	letodms	debian	11/14/2019, 7:23:00 ...	11/13/2019, 4:15:00 PM	4.3	79	Improper Neutralization of Ir

Query settings

Properties

Name: merged_cve_dataset

Applied steps

- Source
- Expanded vendors
- Merged queries
- Expanded cve
- Filtered rows
- Renamed columns

Descriptive Statistical Analysis

```
> summary(data)
```

cve_id	vulnerable_product	vendor	modified_date	published_date	cvss
Length:241979	Length:241979	Length:241979	Length:241979	Length:241979	Min. : 1.200
Class :character	Class :character	Class :character	Class :character	Class :character	1st Qu.: 4.300
Mode :character	Mode :character	Mode :character	Mode :character	Mode :character	Median : 6.400
					Mean : 6.194
					3rd Qu.: 7.500
					Max. :10.000

cwe_code	cwe_name	summary	access_authentication	access_complexity
Min. : 1	Length:241979	Length:241979	MULTIPLE: 31	HIGH : 5620
1st Qu.: 94	Class :character	Class :character	NONE :218079	LOW :132077
Median : 189	Mode :character	Mode :character	SINGLE : 23869	MEDIUM:104282
Mean : 216				
3rd Qu.: 287				
Max. :1188				

access_vector	impact_availability	impact_confidentiality	impact_integrity
ADJACENT_NETWORK: 6614	COMPLETE:68134	COMPLETE: 60398	COMPLETE: 57242
LOCAL : 34590	NONE :76415	NONE : 76256	NONE : 78110
NETWORK :200775	PARTIAL :97430	PARTIAL :105325	PARTIAL :106627

Data Visualization

Here are the research questions that our dataset will be answering:

1. Time Series Analysis of High/Severe Vulnerabilities over time.
2. Vulnerabilities by Access Complexity, Impact, CVSS scores.
3. Identifying Vendors/Products based on Vulnerability.

Data Visualization

LIVE DEMO

Thank you!!