Malware Data Analysis

Thesis submitted in partial fulfillment of
the requirements of the degree of

Masters in Science with Specialization in

Cyber security

by

Stalin Chetty

Roll No: 02

Gr No: 3510246

Under the Supervision of

Prof. Vishal Badgujar

April 2023

Nagindas Khandwala College,
Malad (West)

(Affiliated to University of Mumbai)

MUMBAI, 400064
MAHARASHTRA
YEAR - 2023

CERTIFICATE

This is to certify that the dissertation entitled **"Malware Data Analysis"** is a bonafide work of "**STALIN CHETTY" (Roll No: 02 and G.R. No: 3510246)** submitted to the Nagindas Khandwala College (Autonomous), Mumbai in partial fulfillment of the requirement for the award of the degree of **"Masters in Science with Specialization in Cybersecurity"**.

Prof. Vishal Badgujar
Internal-Examiner

Prof.
External-Examiner

Supervisor's Certificate

This is to certify that the dissertation entitled **"MALWARE DATA ANALYSIS"** submitted by **STALIN CHETTY, Roll No: 02 and G.R. No: 3510246,** is a record of original work carried out by him under my supervision and guidance in partial fulfillment of the requirements of the degree of **Masters in Science with Specialization in Cybersecurity** at Nagindas Khandwala College(Autonomous),Mumbai 400064 . Neither this dissertation nor any part of it has been submitted earlier for any degree or diploma to any institute or university in India or abroad.

Prof. Vishal Badgujar
Internal Examiner

Declaration of Originality

I, **Stalin Chetty, Roll No: 02** and **G.R. No: 3510246**, hereby declare that this dissertation entitled "**Malware Data Analysis"** presents my original work carried out as a Master Student of Nagindas Khandwala College (Autonomous),Mumbai 400064. To the best of my knowledge, this dissertation contains no material previously published or written by another person, nor any material presented by me for the award of any degree or diploma of Nagindas Khandwala College (Autonomous), Mumbai or any other institution. Works of other authors cited in this dissertation have been duly acknowledged under the sections "Reference" or "Bibliography". I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission.

I am fully aware that in case of any non-compliance detected in future, the Academic Council of Nagindas Khandwala College (Autonomous), Mumbai may withdraw the degree awarded to me on the basis of the present dissertation.

Date:

Place: Stalin Chetty

Acknowledgement

I remain immensely obliged to **Prof. Vishal Badgujar**, for providing me with the idea of this topic, and for his invaluable support in garnering resources for me either by way of information or computers also his guidance and supervision which made this Project happen.

I would like to say that it has indeed been a fulfilling experience for working out this Project.

Abstract

The project on malware data analysis with the help of Kibana in the local host machine aims to analyze and visualize various types of malware data using the Kibana platform. The project involves the collection of malware data, it is a dataset which contains the data's of the malicious phishing attack that took place on the internet, such as url's of the affected websites, and the kind of attack. Then these data's are uploaded in into Elasticsearch, a search and analytics engine. Kibana is then used to create interactive dashboards and visualizations, which help in identifying and analyzing malware threats the kibana helps to visualize the data in a graphical format, then we can arrange the graphical formats in one display and can view the data formation .The results of the analysis can be used by security professionals to detect and mitigate malware threats in real-time. Overall, this project demonstrates the effectiveness of using Kibana as a tool for analyzing and visualizing malware data on a local host machine.

Table of Contents

CHAPTERS 1:INTRODUCTION	**1-3**
1.1 Introduction	1
1.2 Problem Statements	2
1.3 Aim	3
1.4 Objectives	
CHAPTER 2 : LITERATURE SURVEY	4
CHAPTER 3 : METHEDOLOGY	**5-14**
3.1 Requirements: Hardware and Software	5
3.2 Methedology	6-14
CHAPTER 4 : CONCLUSION	**15**
REFERENCES	**16**

CHAPTER 1: INTRODUCTION

1.1 INTRODUCTION

Nowadays, Internet becomes an essential part of the daily life of many people. Several services are available and growing daily on the internet. These services are being used by an increasing number of people. One example of an Internet business service is online banking or advertising. Similar to the real world, there are those online who wish to harm others by taking advantage of honest users whenever money is involved. So kind of attacks are called as phishing attack. Phishing attacks are the practice of sending fraudulent communications that appear to come from a reputable source. It is usually done through email. The goal is to steal sensitive data like credit card and login information, or to install malware on the victim's machine. Phishing is a common type of cyber attack that everyone should learn about in order to protect themselves. Phishing starts with a fraudulent email or other communication that is designed to particular victim. The message is made to look as though it comes from a trusted sender. If it fools the victim, he or she is trapped into providing confidential information, often on a scam website. Sometimes malware is also downloaded onto the target's computer. Which starts scanning the system and starts taking the sensitive information from the system. The Project is to identify the malware analysis in database manner with the help of Elastic tool kibana. It is a visualization tool that is used to show the graphical representation of the datasets used of phishing attacks in the internet.

As we have to perform the data visualization in the local host machine. We need to download the Elastic search package from the official website an extract the package. The next is to download the kibana package from the official website of the elastic search and extract the packages and make configuration changes in the configuration file. After extracting the package open the command prompt of the elastic search bin file and the kibana bin file. And write the command as for elastic search "elasticsearch.bat" and for Kibana "Kibana.bat". In the elastic cmd it provides username and password for the kibana local host which configure the elastic search directly in the kibana local host.

After the configuration of elastic search in kibana, it setups the dashboard of the elastic search. The next is to download the dataset of malicious phishing on the internet. And upload the dataset to the dashboard and setup the data for the visualization. After the data is uploaded import it to the dashboard view. Then create visualization and view the data in the graphical manner according to the need.

1.2 PROBLEM STATEMENT

The rising incidence and sophistication of malware threats in today's digital environment serve as the issue statement for the project on malware data analysis using Kibana on the local host machine. Malware, which includes viruses, spyware, and ransomware, can seriously harm computer systems, networks, and data, resulting in losses in money and harm to a person's or an organisation's reputation. Security professionals need tools that can help them analyse and visualise huge volumes of malware data in real-time in order to properly detect and mitigate these threats. Create interactive dashboards and visualisations that can help in the detection and mitigation of malware threats using the potent data visualisation platform Kibana. But there is little study on the efficiency

The Purpose of this project is to analysis the data of the malicious datasets of the phishing attack that have been taken place earlier on the internet in the local host website with the help of elastic and kibana . And to represent the data in the graphical format.

1.3 Aim

The goal of the project on malware data analysis using the Kibana platform on the local host system is to analysis and visualize different forms of malware data. The project entails the gathering of malware data, which is a dataset including information about malicious phishing attacks that have occurred online, including the urls of the websites that were targeted and the nature of the attacks. These data are then put into the search and analytics engine Elasticsearch. The creation of interactive dashboards and visualizations using Kibana then aids in the detection and analysis of malware risks. The kibana program aids in the graphical representation of data, allowing us to arrange the graphical representations in a single display and observe the table of contents.

1.3 OBJECTIVE

A project on malware risk analysis that utilizes online datasets to analyze phishing attack details can provide valuable insights into the nature and extent of these types of attacks. The following are some observations regarding such a project:

- The project highlights the importance of conducting a comprehensive risk assessment to identify potential risks that could lead to malware infections. This involves identifying critical assets, assessing the likelihood and impact of potential threats efforts based on the level of risk.

- Identification of key risk factors: The project can help identify key risk factors associated with phishing attacks, such as the type of phishing email, the use of social engineering techniques, and the susceptibility of certain user groups.

- Need for advanced analytical tools: To extract meaningful insights from the datasets, the project should utilize advanced analytical tools such as kibana. These tools can help identify patterns and trends in the data that may not be apparent through manual analysis.

CHAPTER 2: LITERATURE SURVEY

Sr no	Author	Title	Advantages
1	Kevin Martin	Using Kibana to Visualize Network Data.	It explains how to use Kibana to visualize network data and gain insights into potential security threats
2	Akash Mahajan	Elastic Stack for Security Analytics	An overview of using the Elastic Stack for security analytics and monitoring.
3	Hamad Al-Mohannadi	Cyber Threat Intelligence from Honeypot Data Using Elasticsearch	The new threat intelligence technique that evaluates honeypot log data to identify attacker behaviour and find attack trends
4	Guy Bruneau	Malware Analysis with elastic-agent and Microsoft Sandbox	Microsoft describes the "Windows Sandbox supports simple configuration files, which provide a minimal set of customization parameters for Sandbox
5	Vlad-Andrei Zamfir, Mihai Carabas, Costin Carabas, Nicolae Tapus	Systems Monitoring and Big Data Analysis Using the Elasticsearch System	The Elasticsearch and other tools like Kibana and Logstash are integrated in order to have an affordable solution to web based integration of search and data analytics.

CHAPTER 3: METHEDOLOGY

3.1 REQUIREMENTS

Hardware Used:-

1. Processor: A quad-core processor with a clock speed of at least 2.4 GHz or higher is recommended.

2. RAM: A minimum of 8 GB of RAM is recommended for running Kibana and Elasticsearch. However, for larger datasets, it is recommended to have at least 16 GB of RAM or higher.

3. Storage: A solid-state drive (SSD) with a minimum of 500 GB of storage capacity is recommended for storing the Elasticsearch indices and Kibana dashboards.

Software and Tools Used:-

1. Elasticsearch

2. Kibana.

3. Malware data sets: malicious_phishing 44.595 KB

4. Chrome Web browser

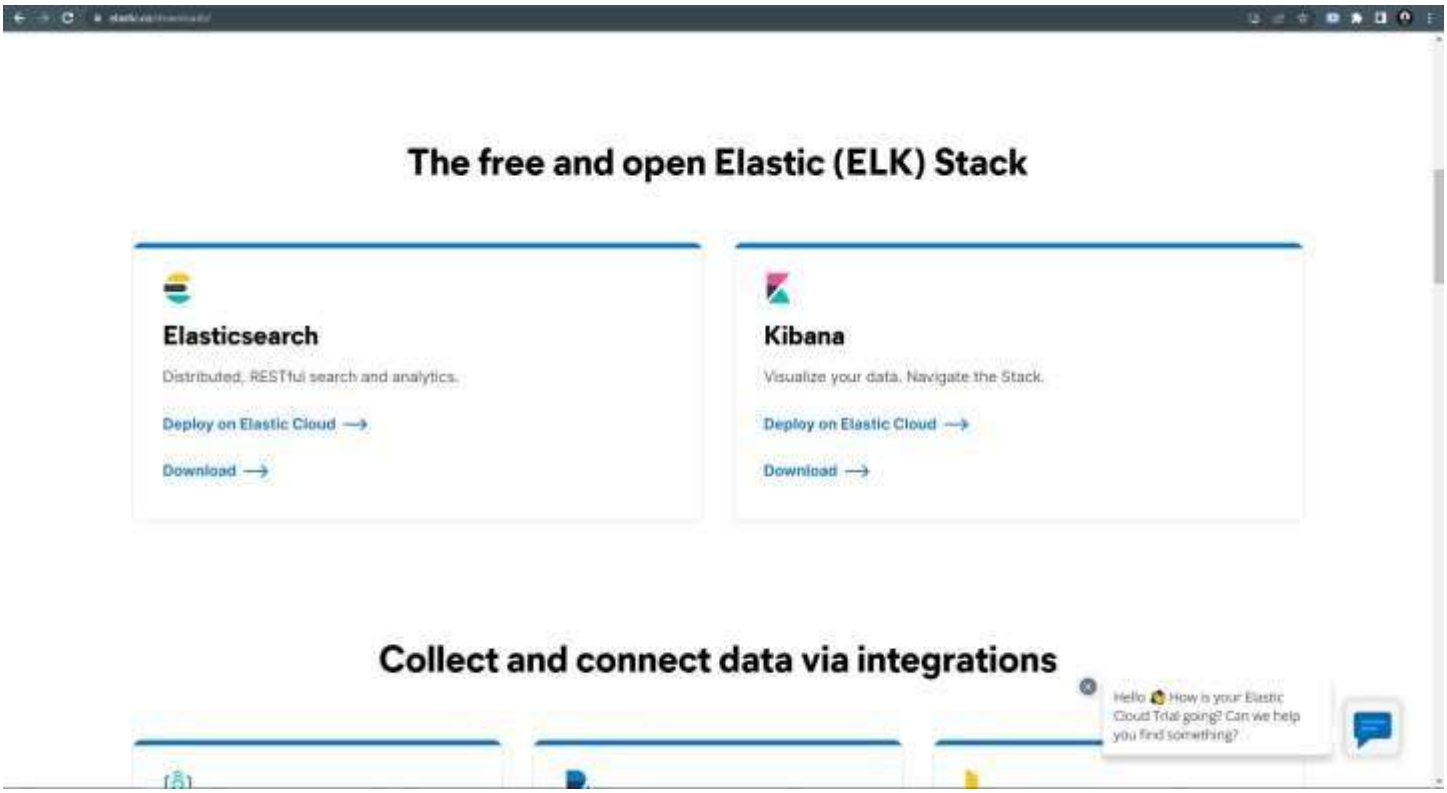
3.2 METHEDOLOGY

Here are the steps to set up Elasticsearch and Kibana on Windows to run it on the local host:

1. Download Elasticsearch and Kibana:

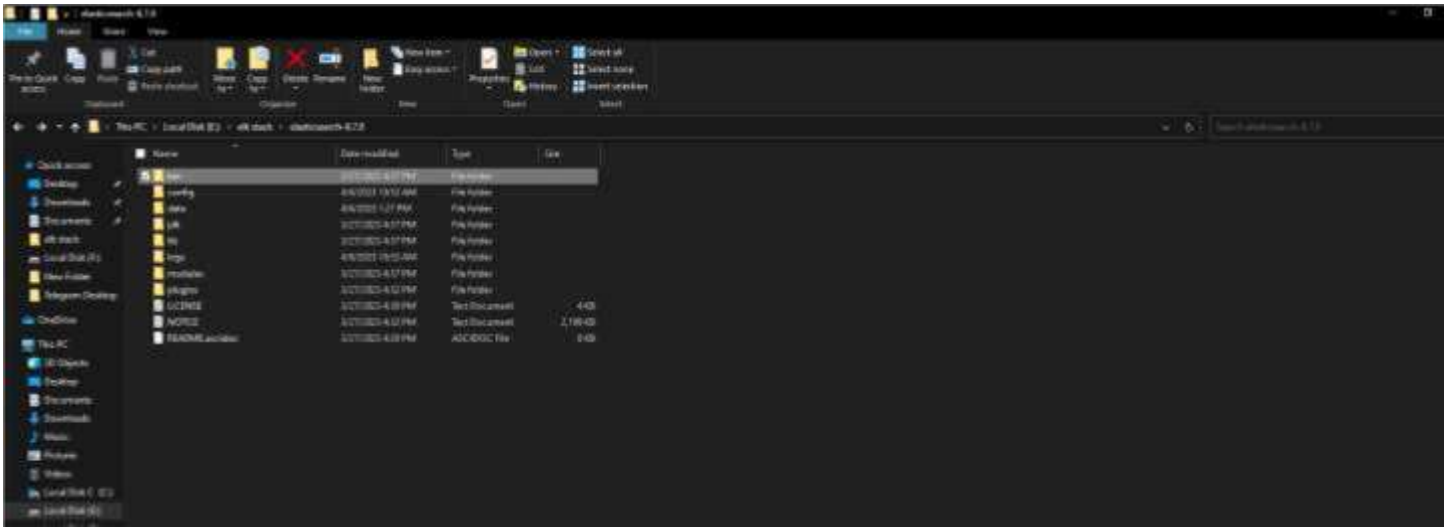
• Go to the Elasticsearch download page at https://www.elastic.co/downloads/elasticsearch and download the latest version of Elasticsearch.

• Go to the Kibana download page at https://www.elastic.co/downloads/kibana and download the latest version of Kibana.

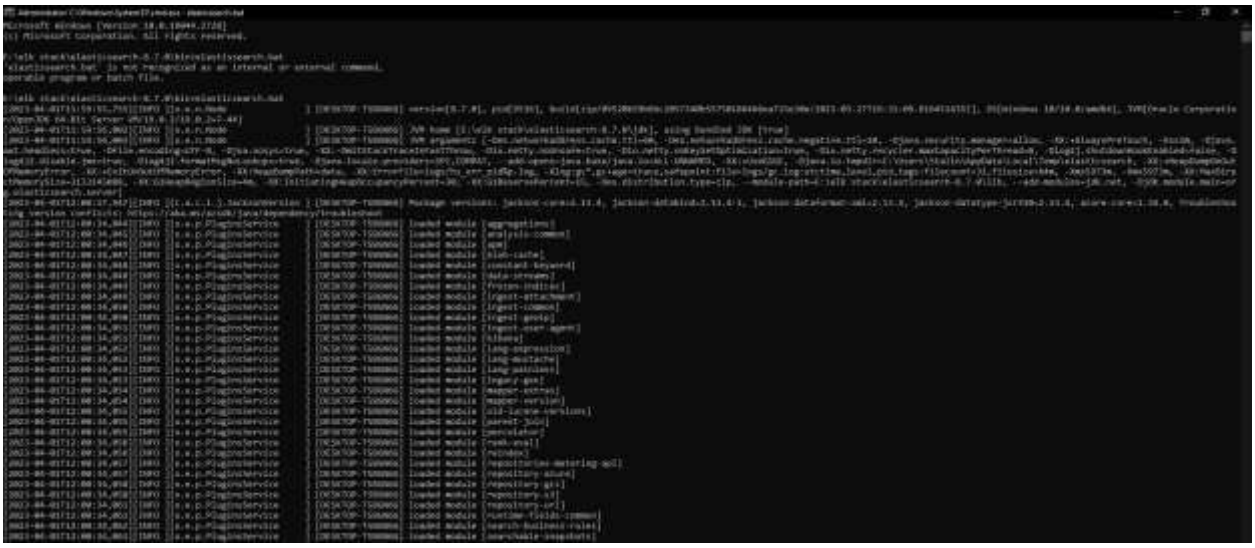


2. Install Elasticsearch:

• Extract the Elasticsearch archive that you downloaded in the previous step to a directory on your system, such as C:\elasticsearch.

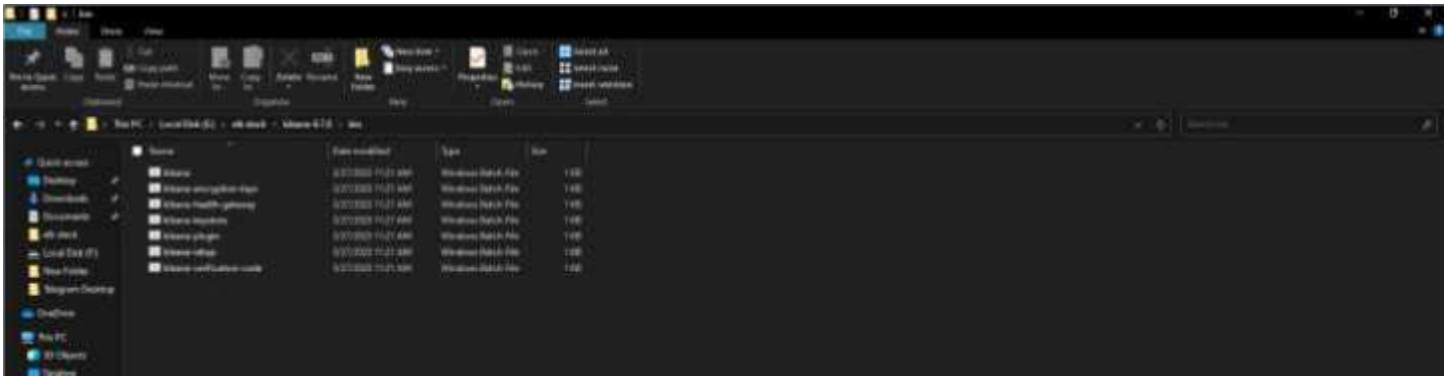


• Open a command prompt as an administrator, navigate to the Elasticsearch directory, and run the command bin\elasticsearch.bat. This will start Elasticsearch.



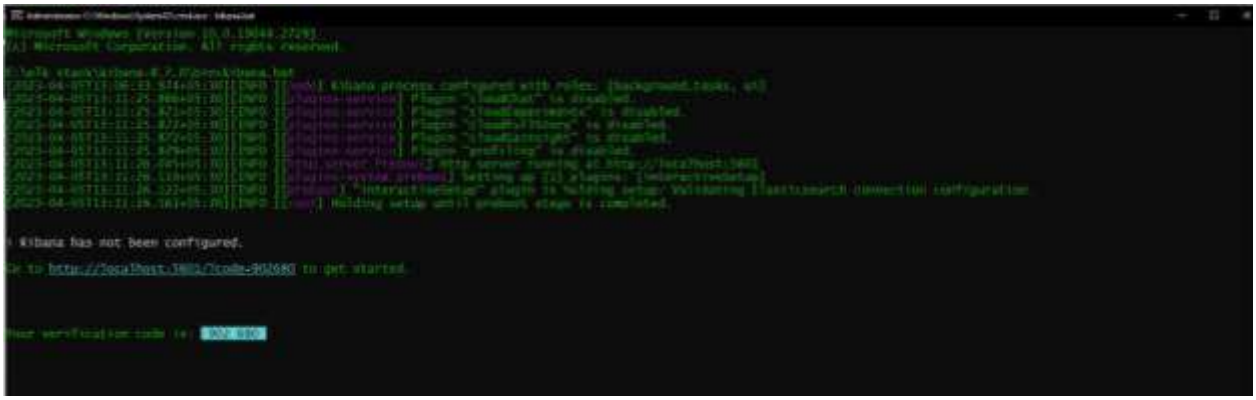
3. Install Kibana:

• Extract the Kibana archive that you downloaded in step 1 to a directory on your system, such as C:\kibana.



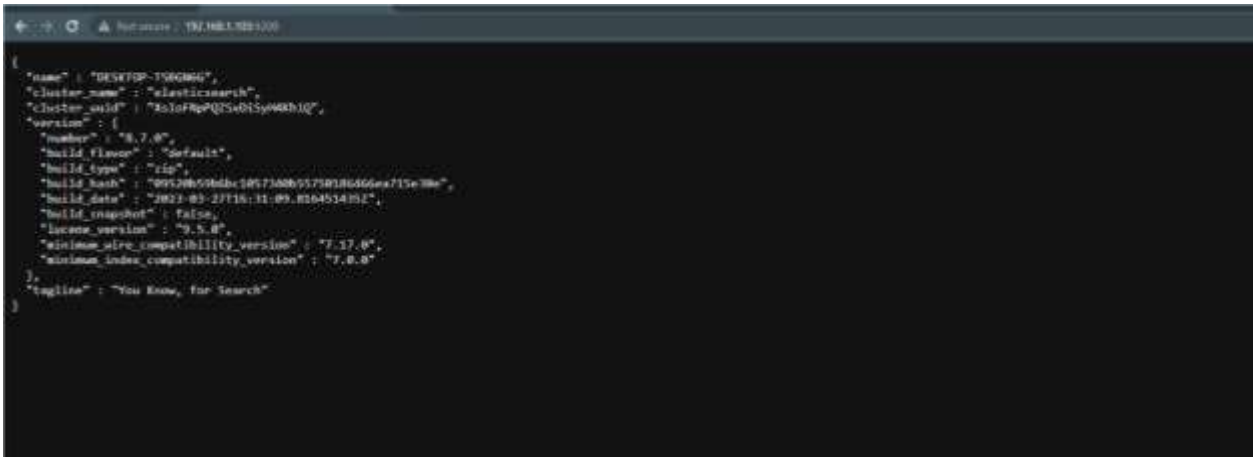
• Open the Kibana configuration file, located at C:\kibana\config\kibana.yml, and uncomment the line elasticsearch.hosts: ["http://localhost:9200"].

• Open a command prompt as an administrator, navigate to the Kibana directory, and run the command bin\kibana.bat. This will start Kibana.

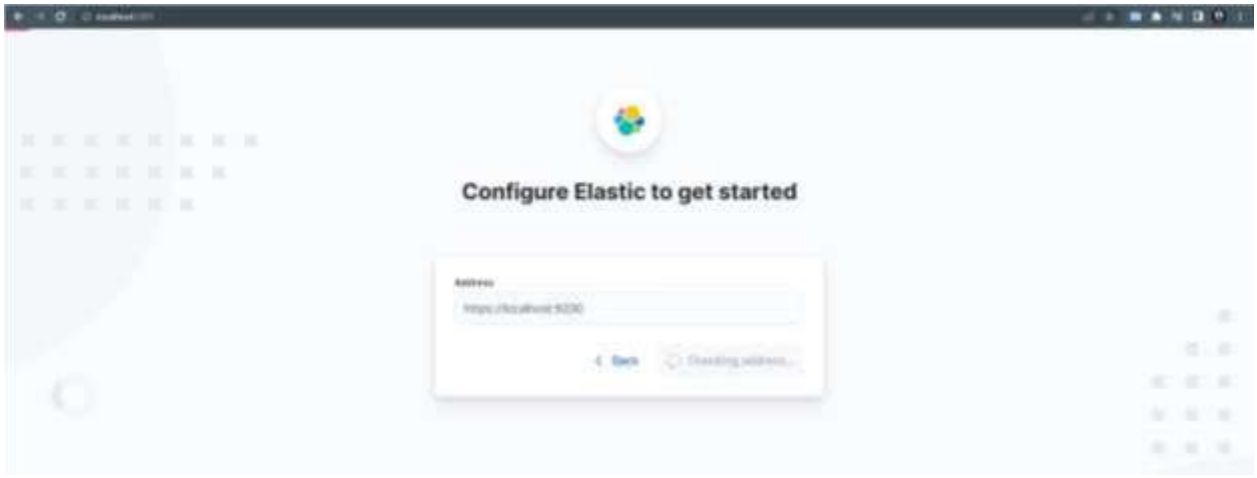


4. Verify that Elasticsearch and Kibana are running:

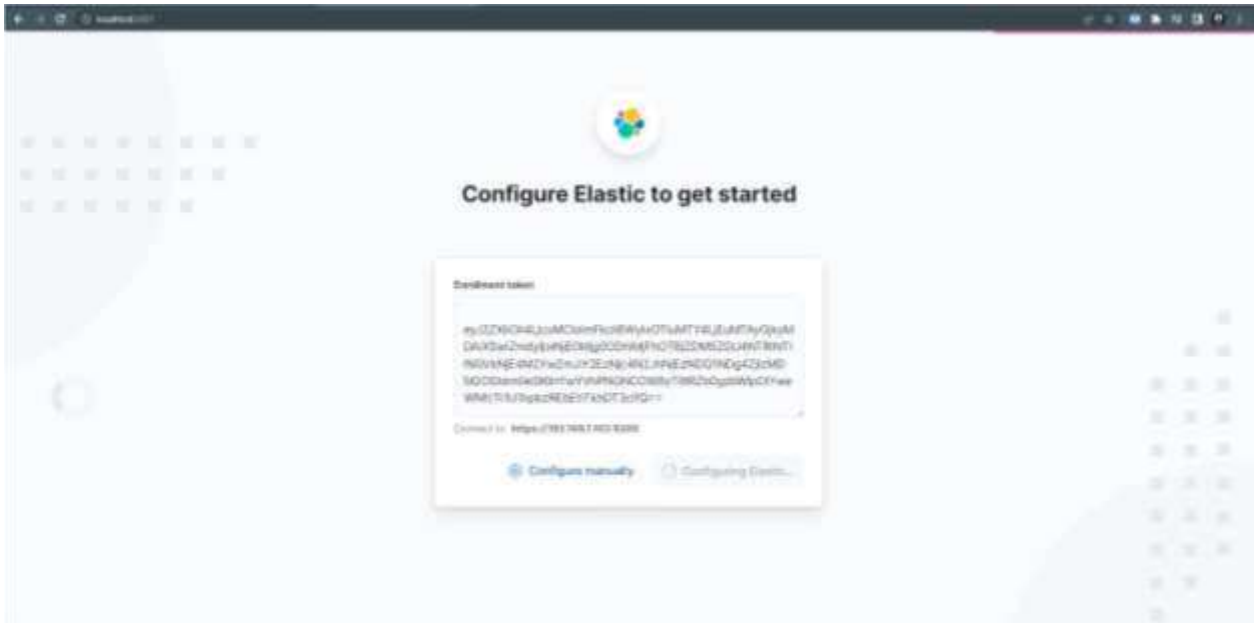
• Open a web browser and go to http://localhost:9200. You should see a response from Elasticsearch.



- Open another tab in the same web browser and go to http://localhost:5601. You should see the Kibana homepage.

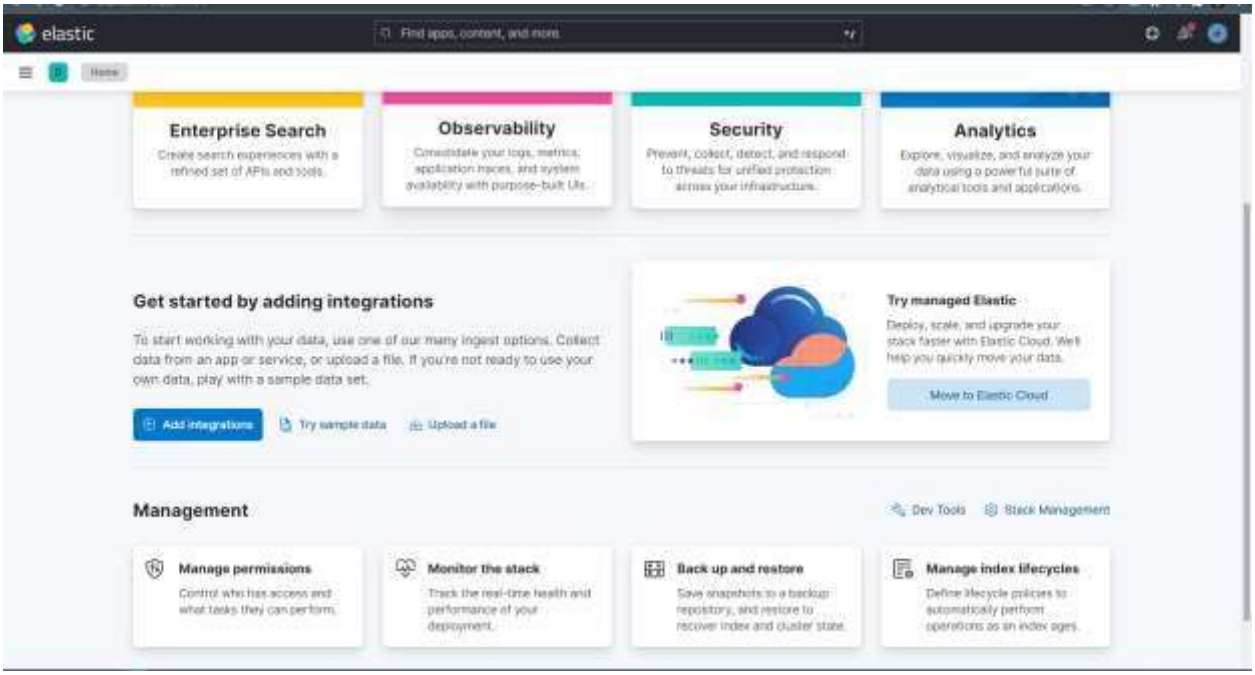


 Next is to provide the token at is available in the elastic search command prompt.



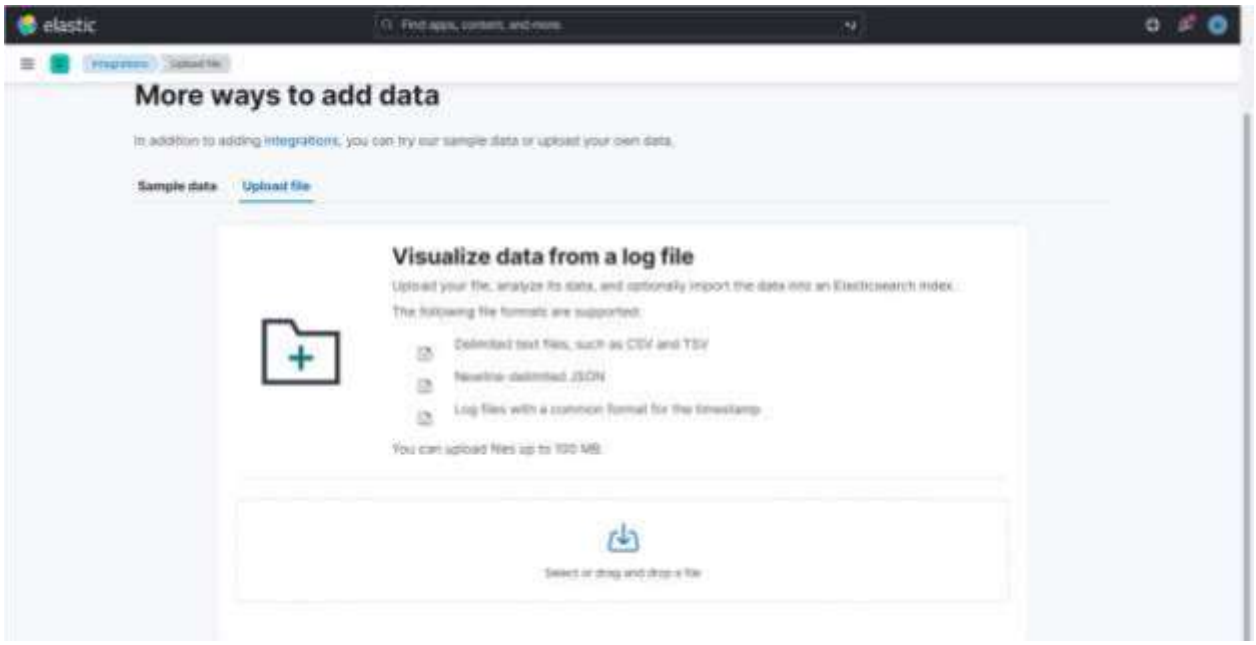
5. Start exploring data in Kibana:

- In Kibana, go to the "Management" section and create an index pattern for your data.

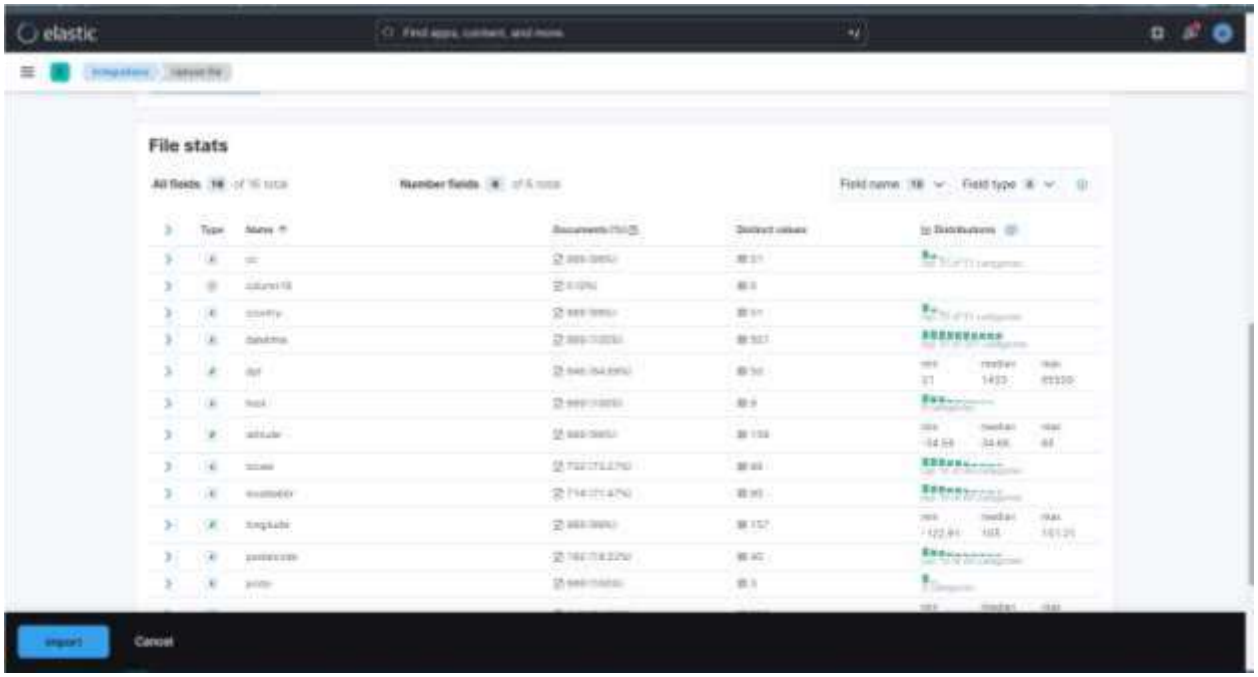


- Go to the "Discover" section and start exploring your data by running queries and visualizing the results.

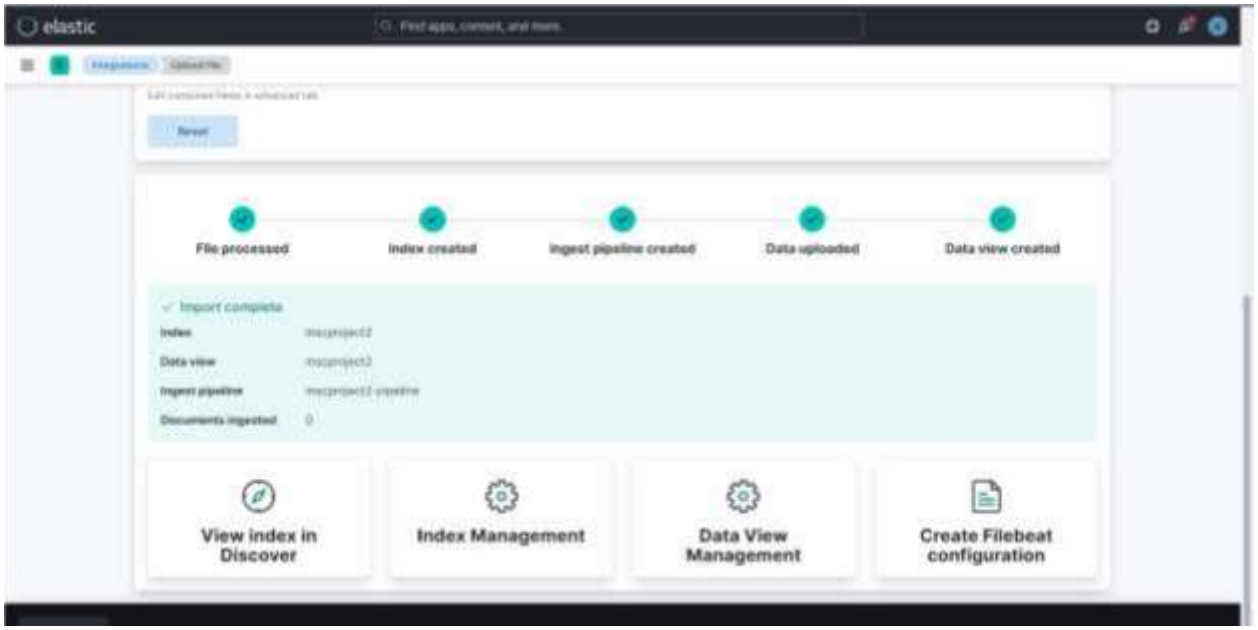
6. The further step is to upload the data in the local host website.



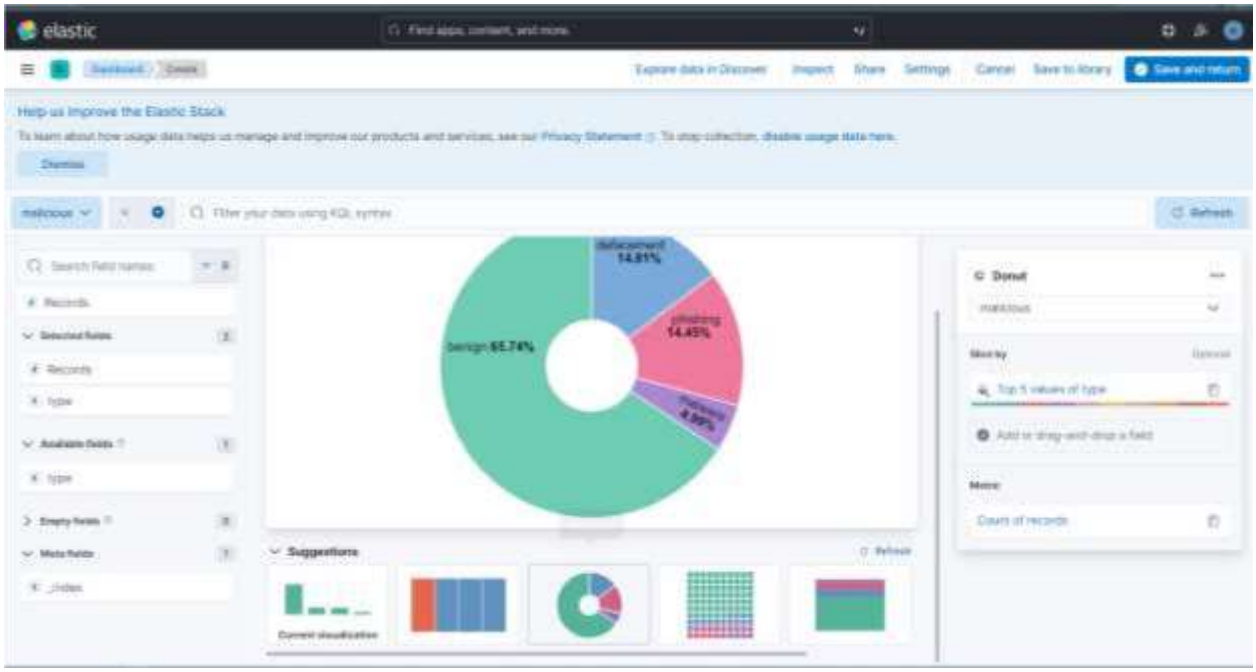
7. After uploading the dataset it starts scanning the file and provides with the output, then we have to import the file with a random file_name .

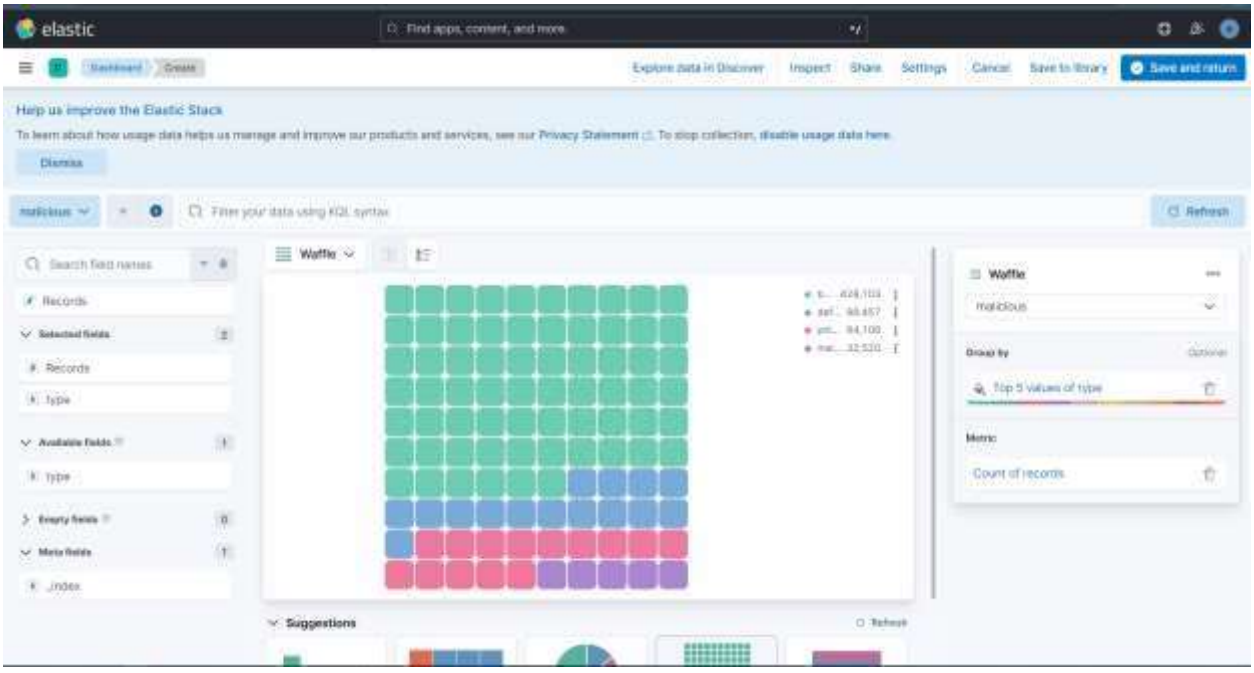


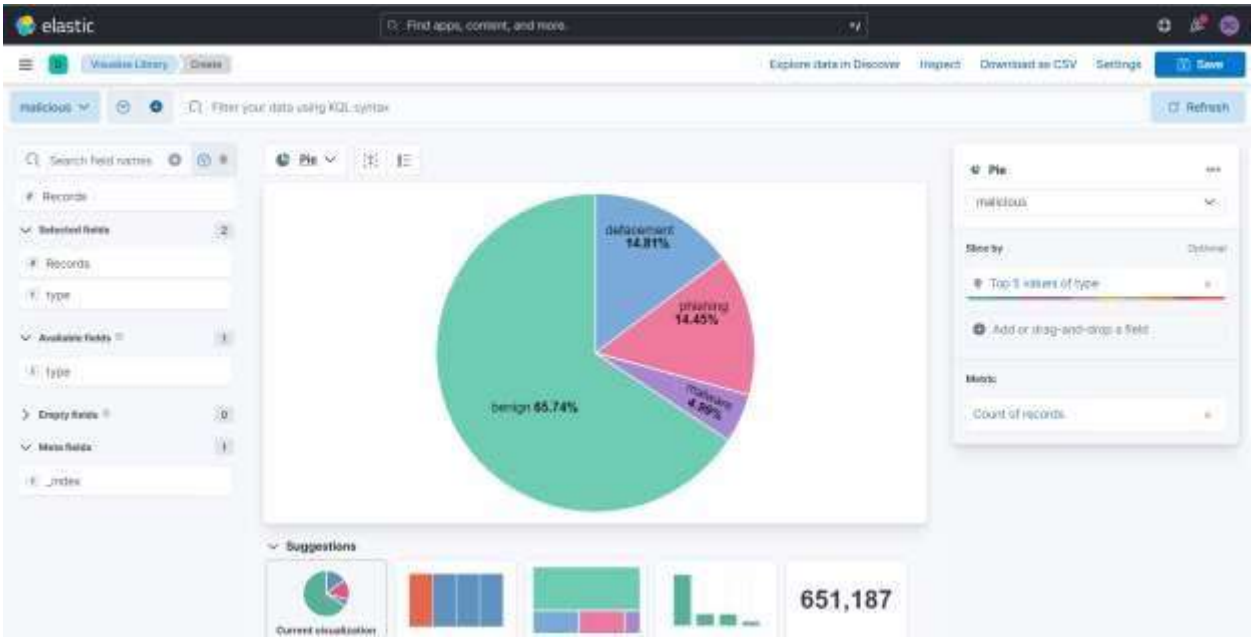
8. After importing the data in the dashboard it provides the output, and the permission to view the data.

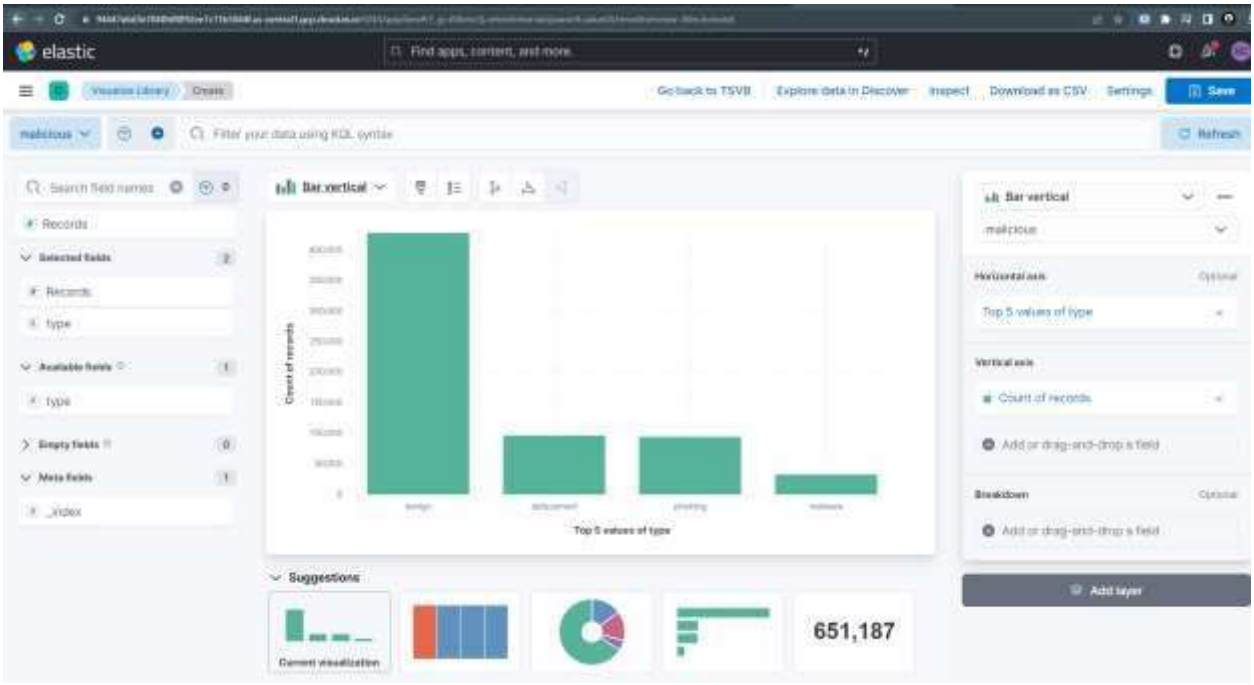


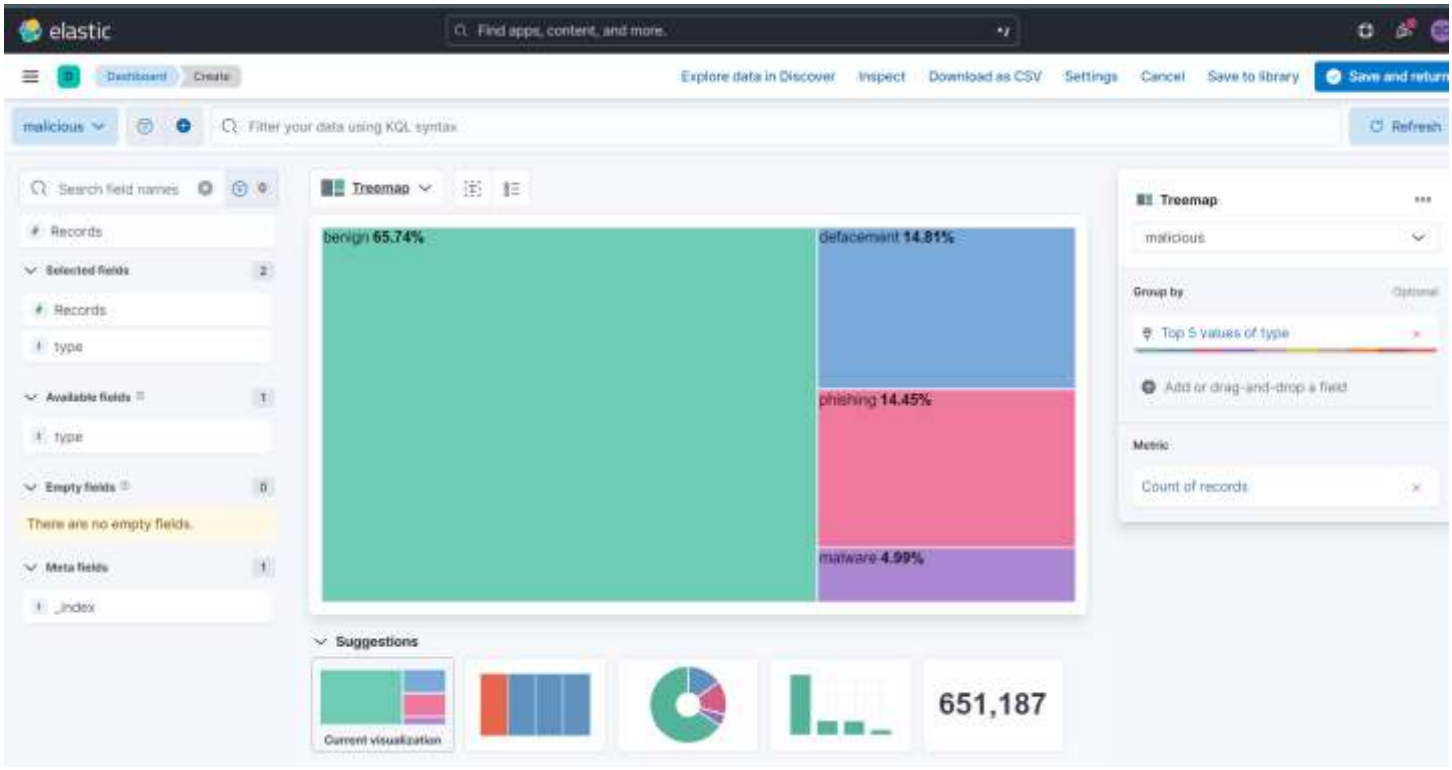
9. Further we can view the data in the graphical manner.

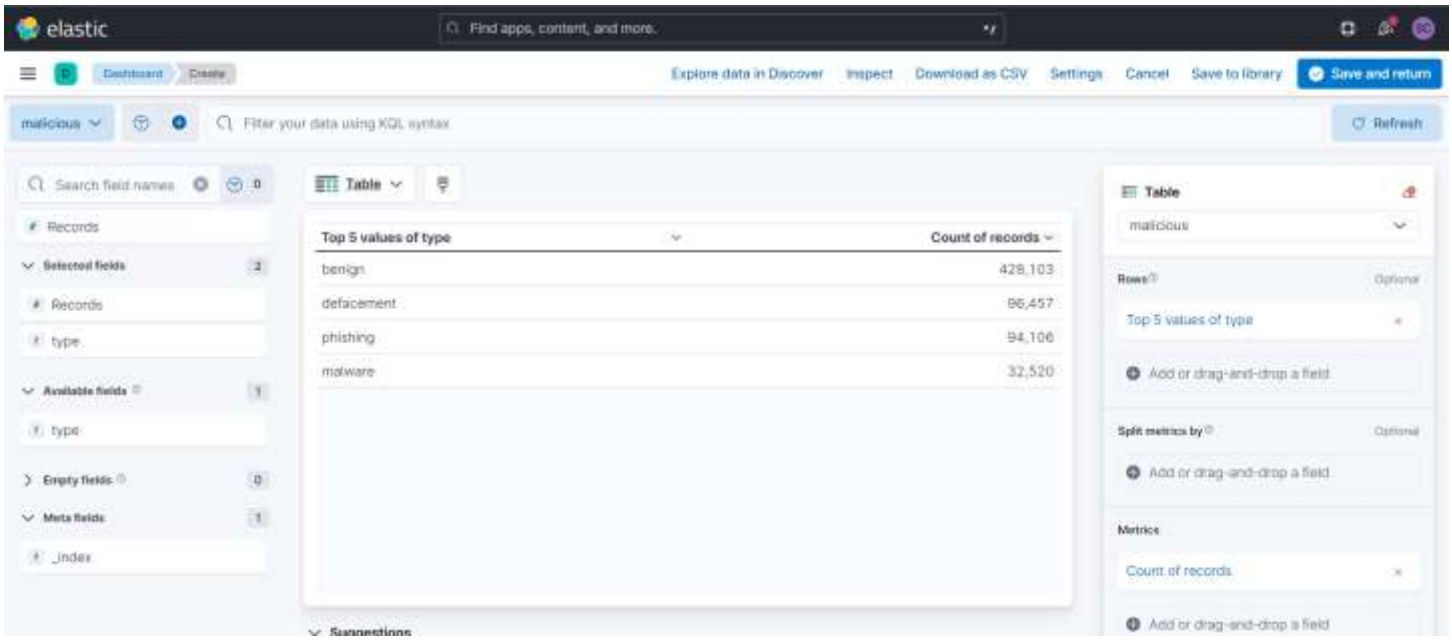


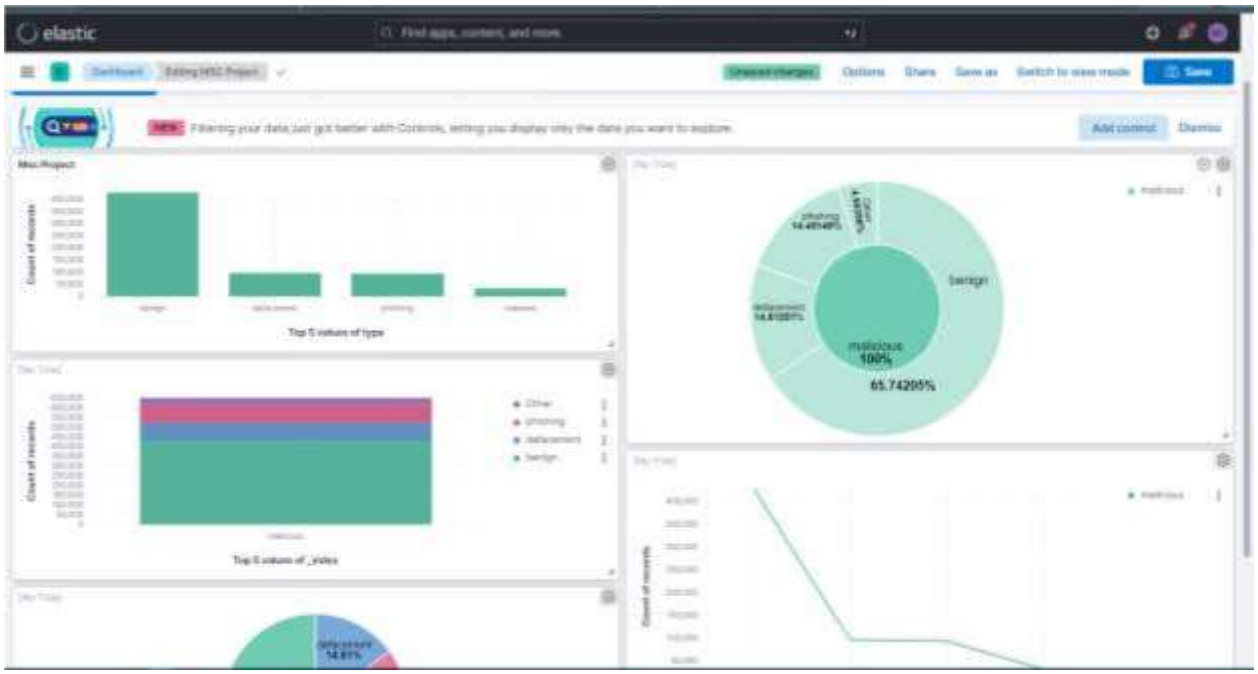



elastic
Find apps, content, and more.
Explore data in Discover
Inspect
Download as CSV
Settings
Save
Refresh
Pie
defacement 14.81%
phishing 14.45%
benign 65.74%
Suggestions
651,187

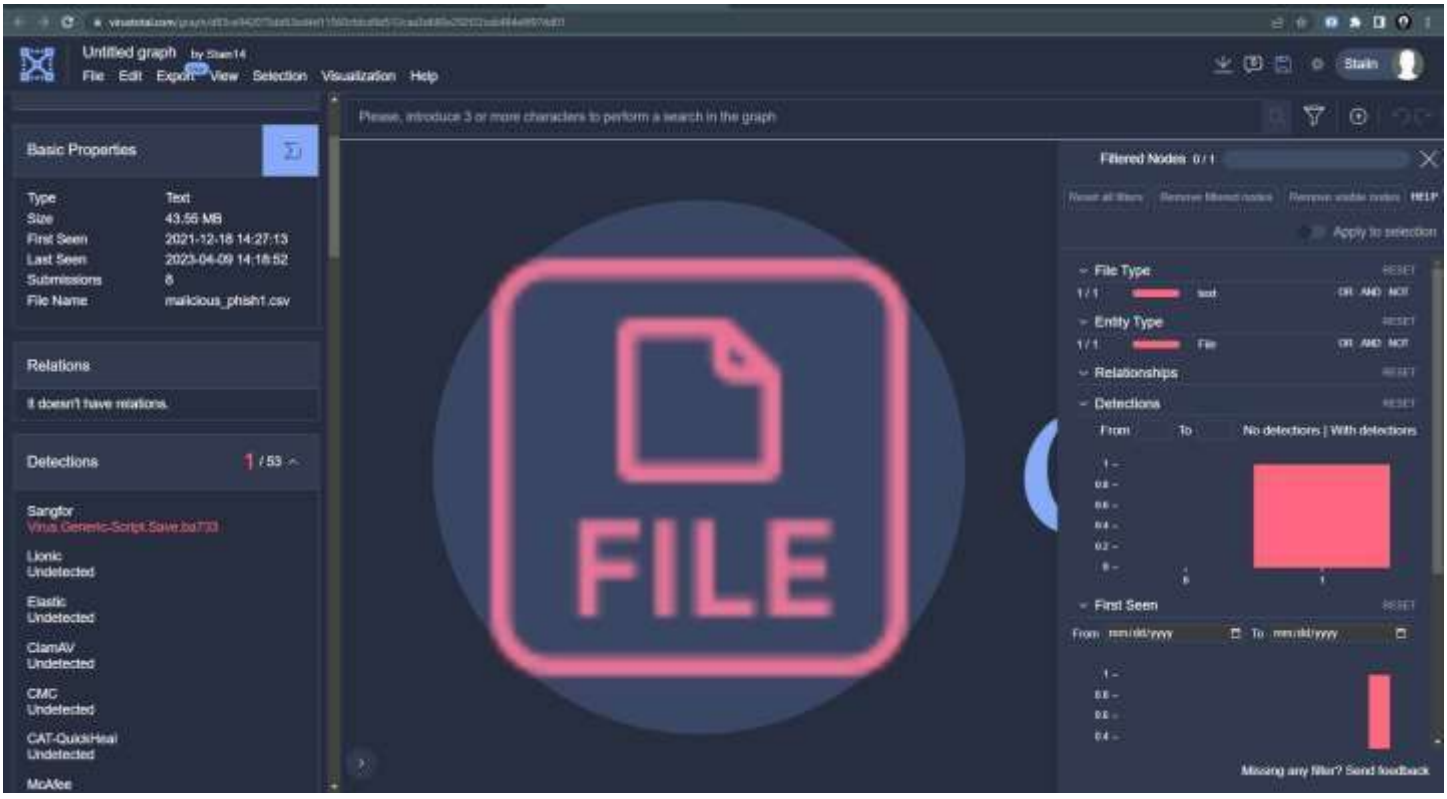

elastic
Find apps, content, and more.
Go back to TSVB
Explore data in Discover
Inspect
Download as CSV
Settings
Save
Refresh
Bar vertical
Count of records
Top 5 values of type
Suggestions
651,187


elastic
Find apps, content, and more.
Explore data in Discover
Inspect
Download as CSV
Settings
Cancel
Save to library
malicious
Filter your data using KQL syntax
Refresh
Search field names
Records
Selected fields
type
Available fields
Empty fields
There are no empty fields.
Meta fields
_index
Treemap
benign 65.74%
defacement 14.81%
phishing 14.45%
malware 4.99%
Suggestions
Current visualization
651,187
Group by
Top 5 values of type
Add or drag-and-drop a field
Metric
Count of records


elastic
Find apps, content, and more.
Explore data in Discover
Inspect
Download as CSV
Settings
Cancel
Save to library
Save and return
malicious
Filter your data using KQL syntax
Refresh
Search field names
Records
Selected fields
type
Available fields
Empty fields
Meta fields
_index
Table
Top 5 values of type
Count of records
benign 428,103
defacement 96,457
phishing 94,106
malware 32,520
Rows
Top 5 values of type
Add or drag-and-drop a field
Split metrics by
Metrics
Count of records



10. With the help of virustotal we finded the vulnerability in the malicious_phising. csv file



CHAPTER 4: CONCLUSION

In conclusion, setting up Elasticsearch and Kibana on a local machine is a straightforward process that enables users to visualize and analyze data in real-time.

For a project on malware risk analysis, Elasticsearch and Kibana can be used to store and visualize datasets related to malware and cybersecurity threats. By creating dashboards and visualizations, analysts can gain insights into potential risks and make informed decisions on how to mitigate them.

Additionally, Elasticsearch's powerful search capabilities allow analysts to quickly find specific data points and investigate potential incidents. Effective malware analysis requires collaboration among security professionals, adoption of best practices, and automation of analysis wherever possible.

Overall, using Elasticsearch and Kibana for malware risk analysis provides a powerful toolset for analyzing cybersecurity threats and enhancing overall network security.

REFERENCES

1. Elastic search official website : https://www.elastic.co/downloads/

2. Video reference taken to download and configure the Elasticsearch and Kibana https://www.youtube.com/watch?v=45vE3pgtpyY&ab_channel=SimplifyingTech

3. Reference taken to upload the data in the kibana and using the visualization dashboard . https://www.youtube.com/watch?v=RHkc4RQw7cg&t=434s&ab_channel=SimplifyingTech

4. https://www.onlinetutorialspoint.com/elasticsearch/how-to-install-elasticsearch-on-windows-10.html

5. Virus total website used to find the malware in the datasets. https://www.virustotal.com/gui/home/search