Amrita Vishwa Vidyapeetham

Amrita School of Computing, Chennai

Computer Science and Engineering-Cyber Security

20CYS215 ML in Cybersecurity

**AWS Honeypot Analysis**

V. Jayaraj, E. Kuralamudhan, Navaneethan

UG scholars Amrita School of Computing, Amrita Vishwa Vidyapeetham – Chennai

**1 INTRODUCTION**

* 1. **ABSTRACT**

In recent years, cloud computing has become a popular platform for hosting web applications and services. However, as the adoption of cloud computing has increased, so too have the number of cyber-attacks. To detect and prevent these attacks, organizations often use honeypots, which are decoy systems designed to attract and monitor malicious activity. This research paper focuses on analyzing honeypot data collected from an Amazon Web Services (AWS) instance.

This research paper focuses on analyzing honeypot attack data to improve the understanding of the attack landscape and develop effective countermeasures. Our research also covers the analysis of honeypot data using various techniques, including machine learning and statistical analysis. Our findings indicate that analyzing honeypot attack data can provide valuable insights into the methods and motivations of attackers, enabling organizations to develop more effective security measures.

**1.2 HONEYPOT**

The primary purpose of a honeypot system is to gather intelligence on the attacker’s tactics, techniques, and procedures (TTPs) and use this information to improve the overall security posture of the organization. By monitoring the attacker’s activities in the honeypot environment, security analysts can gain insights into the types of attacks that are being launched and develop effective countermeasures to prevent them from happening in the future.

The function of a honeypot is to represent itself on the internet as a potential target for attackers and to gather information and notify defenders of any attempts to access the honeypot by unauthorized users.

* 1. **WHY ANALYZING IS ESSENTIAL?**

**1.3.1 Early Detection of Attacks:**

Honeypots can be used as a decoy to attract attackers, which can help detect attacks at an early stage. Analyzing honeypots can provide insights into the methods used by attackers, their motivations, and their capabilities, enabling organizations to take proactive measures to prevent real attacks.

**1.3.2 Understanding the Attack Landscape:**

By analyzing the data collected from honeypots, organizations can gain a better understanding of the types of attacks that are being launched against their systems, the frequency of these attacks, and the tools and techniques used by attackers. This information can be used to develop more effective security measures and strategies to protect against these attacks.

**1.3.3 Identifying Vulnerabilities:**

Honeypots can be used to simulate various systems and applications, allowing organizations to identify vulnerabilities in their IT infrastructure that can be exploited by attackers. Analyzing the data collected from honeypots can provide insights into the security weaknesses of the organization and help prioritize remediation efforts.

**1.3.4 Improving Incident Response:**

Honeypots can be used to test incident response procedures and provide security teams with opportunities to practice responding to different types of attacks. Analyzing honeypot data can help identify gaps in incident response procedures and enable organizations to improve their response capabilities.

Overall, analyzing honeypots is an essential component of an organization’s security strategy, as it can provide valuable insights into the current threat landscape, vulnerabilities in their IT infrastructure, and help improve incident response capabilities.

**Keywords:**

Cloud Computing, machine learning, statistical analysis, unauthorized users, proactive measures, incident response

1. **LITERATURE REVIEW & PROBLEM DESCRIPTION:**
   1. **RELATED SURVEY STUDIES**
      1. **“An Overview of Honeypot Systems” by Neha Titarmare**

Here the author first gives a brief introduction about honeypot and why is it necessary in the contemporary world, she describes how it acts as a trap mechanism and also gives the characteristics of Honeypots.

The author also mentions the various advantages and disadvantages of Honeypots and classifies them based on its interaction as ‘High’, ‘Medium’ & ‘Low’. They are also classified based on its purpose as ‘Production System’ & ‘Research System’.

The author ends by talking about the concept of ‘Honeynet’ which in a nutshell is a collection of Honeypots and ‘Honeywalls’ which provides data control.

|  |  |
| --- | --- |
| **Low-Interaction Honeypot** | **High-Interaction Honeypot** |
| Emulates OSs and services | No Emulation; real OS & services provided |
| Simple and easy to install and deploy. | Can be complex to install and deploy |
| Minimal risk as the emulated services control what attackers can and cannot do. | Increased risk as attackers are provided with real operating systems to interact with. |
| Captures limited information which is mainly transactional data and very limited interaction. | Can capture far more information including new tools, communication and attacker keystrokes |

**Table .** Comparison of Low And High Interaction Honeypots

* + 1. **Honeypots and honeynets: issues of privacy**

Here the author starts by explaining about the concepts of Honeypot and Honeynets and goes on to talk w.r.t the legal framework of the European Union about what data are legally allowed to be collected and retained.

The author gives his view on privacy and personal data protection w.r.t to the legal framework in EU law by describing the basic concepts of personal data protection and data retention. He also talks about IP addresses and other privacy issues such as network monitoring and concludes the paper.

* + 1. **Network Defence Strategy by Jianchao Hong**

In this paper, the author introduces us to the Core Technology of Honeypots such as Internet spoofing technology, Data analysis and Data collection.

He says about the design of the defense system, modules and its Analog testing system such as overflow attack, DoS and DDoS attacks. The author ends by raising a question on how to protect the Honeypot itself from an intruder who tries to corrupt the system log files present in it.

* + 1. **“Honeypot Data Analysis” from the book Honeypots for windows by Roger A. Grimes**

This chapter from the book covered the structured approach to honeypot analysis. It reviewed all the different ways to examine honeypot data, including analyzing network traffic, changes to the file system, and changes to the OS.

It starts by answering a question of why is it necessary to analyze data and goes on to brief about forensic investigations of the attack such as “Was the attack automated or manual?”, “How did the initial compromise happen?” and “What did the Hacker/Malware do after initial compromise?”.

* + 1. **Systematic Review of Graphical Visual Methods in Honeypot Attack Data Analysis**

This paper reviews the visualization practices and methods commonly used in the discovery and communications of attack patterns based on Honeypot network traffic data.

It explains about Honeypot types and its log files. Then it briefs about the log management system used in honeypots. It then describes the importance of visualization for easy understanding. The key process involved are Data Abstraction, Visual Analysis & Abstraction, Knowledge Extraction.

It goes on to explain about Graphical Perception and its various methods to implement, Common visualization channels and methods. It then reviews various related survey studies and analysis approach.

* + 1. **Threat Prediction using Honeypot and Machine Learning**

This research paper again starts off with introducing the concepts of Honeypots in general then it discusses about open-source honeypots such as D – Shield (low interaction honeypot developed by Internet Strom Center) and T – Pot.

Then it briefs about cloud services and the need for security in cloud platforms. It also explains about how Honeypot systems are deployed in the cloud and by the end it analyses the results by creating various graphs which displays the ‘OS Distribution’, ‘Attacks based on countries’, ‘Username and Password Cloud’.

Atlast it uses Machine Learning for analyzing the attacks and prediction of attacks, Notable algorithms used are Decision Tree, Logistic Regression and Random Forest.

# A Comparative Analysis of Honeypots on Different Cloud Platforms

This journey by Christopher Kelly starts with an abstract explaining about various Cloud platforms such as Amazon Web Services (AWS), Google Cloud Platform (GCP) and Microsoft Azure.

It also specifies an incident where a misconfigured AWS instance allowed the exposure of hundreds of thousands of customer details, including bank account and social security numbers.

It then briefs about how honeypots acts as a deployable resource and lure attackers to collect their data. Then some related works are discussed and then it briefs about a methodology which works based on the combination if IaaS and containers and then it graphically plots the results.

* + 1. **Honeypot Technologies for Malware Detection and Analysis**

This research paper offers a brief summary of the latest developments in honeypot technologies used for malware detection and analysis.This includes not only honeypot software, but also methodologies to analyze captured honeypot data.

It then describes about the primary objective of antimalware systems as to defend computers against attacks launched by malicious users. It then explains one of the most used tools for network protection, firewalls. Firewalls help protect inside devices from outside attacks.

Some related works are then discussed and usage, advantages, disadvantages of honeypots are then explained and then concludes by drawing some conclusions from Honeypots and acknowledging the works of his fellow mates who helped him in this research.

**2.1.9 Visualizing data using Matplotlib and Seaborn by Arnav Oberoi, Rahul Chauhan**

Here the author introduces about visualization as a graphical

representation of data through the use of pictorial design.

The usual data visualization methods, such as scatter plots, bar charts,

histograms, line charts, and pie charts.

It starts with exploring Matplotlib library which is a graphics package for

data visualization in python which has arisen as a key component in the Python

Data Science Stack and it is well integrated with Numpy and pandas.

It then compares the libraries Matplotlib and Seaborn on the basis of univariate

plots and multivariate plots for data science. It also then gives examples of

various types of Graphical representation using both Matplotlib and Seaborn.

**2.1.10 Making Scientific Publication Plot using Python (Medium)**

This blog from medium first explains the disadvantages of using MATLAB and how python replaces it. It introduces us to various plot parameters, colors, Figures and Axes and gives a general idea of how to plot a scientific graph.

It also describes about spines in a plot, Tick parameters, Adjusting and plotting Range/Ticks, Axis labels, Secondary Axis Ticks, Legends. It ends by saying how to save our plot using “plt.savefig” and explaining about its parameters such as dpi, format(png, jpg, pdf), transparent, bbox\_inches and plot.show() to print the plot.

**2.2 PROBLEM DESCRIPTION**

**2.2.1 Problem Environment**

With the increasing frequency and complexity of cyber-attacks, organizations need to implement proactive measures to protect their systems and data. Honeypots are decoy systems designed to detect and prevent attacks by luring attackers away from real systems. Analyzing the data from honeypots can provide valuable insights into the types of attacks being launched and how attackers are targeting systems.

**2.2.2 Objectives**

The objective of this research is to analyze the data collected from honeypots to improve the understanding of cyber-attacks and develop effective countermeasures. This research aims to identify the types of attacks being launched, the tactics and techniques used by attackers, and the motivations behind the attacks. The research also aims to identify trends and patterns in honeypot data that can be used to develop more effective security measures.

**2.2.3 Assumptions**

This research assumes that the honeypots are properly configured and managed to attract and detect a range of attacks. It also assumes that the data collected from the honeypots is accurate and representative of the types of attacks being launched against real systems.

**2.2.4 Definition**

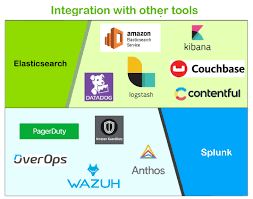
Honeypots are decoy systems designed to attract and monitor malicious activity. They are typically used to detect and prevent cyber-attacks by luring attackers away from real systems. Analyzing the data from honeypots involves examining the types of attacks being launched, the tactics and techniques used by attackers, and the motivations behind the attacks. The data can be used to develop more effective security measures and improve the understanding of the threat landscape.

1. **METHODOLOGY**

**3.1 EXISTING PRACTICE**

The existing practice for analyzing data from honeypots involves using various techniques and tools to extract insights and trends from the collected data. These techniques include statistical analysis, machine learning, and visualization.

One common approach is to use open-source tools such as Elastic Stack, Splunk, or Snort to collect, store, and analyze honeypot data. These tools enable security teams to perform real-time analysis of honeypot data and identify potential threats.



Visualization is also an important aspect of analyzing honeypot data. Tools such as Kibana or Grafana can be used to create dashboards and visual representations of the data, making it easier to identify patterns and trends.

Overall, the existing practice for analyzing honeypot data involves using a combination of techniques and tools to extract insights and improve the understanding of the threat landscape.

**3.2 PROPOSED METHOD**

One proposed method for analyzing data from honeypots is to use machine learning algorithms in combination with data visualization tools such as Matplotlib and Seaborn to detect patterns and anomalies in honeypot data and can also be used to identify new or previously unknown attack types and develop predictive models to identify future attacks.

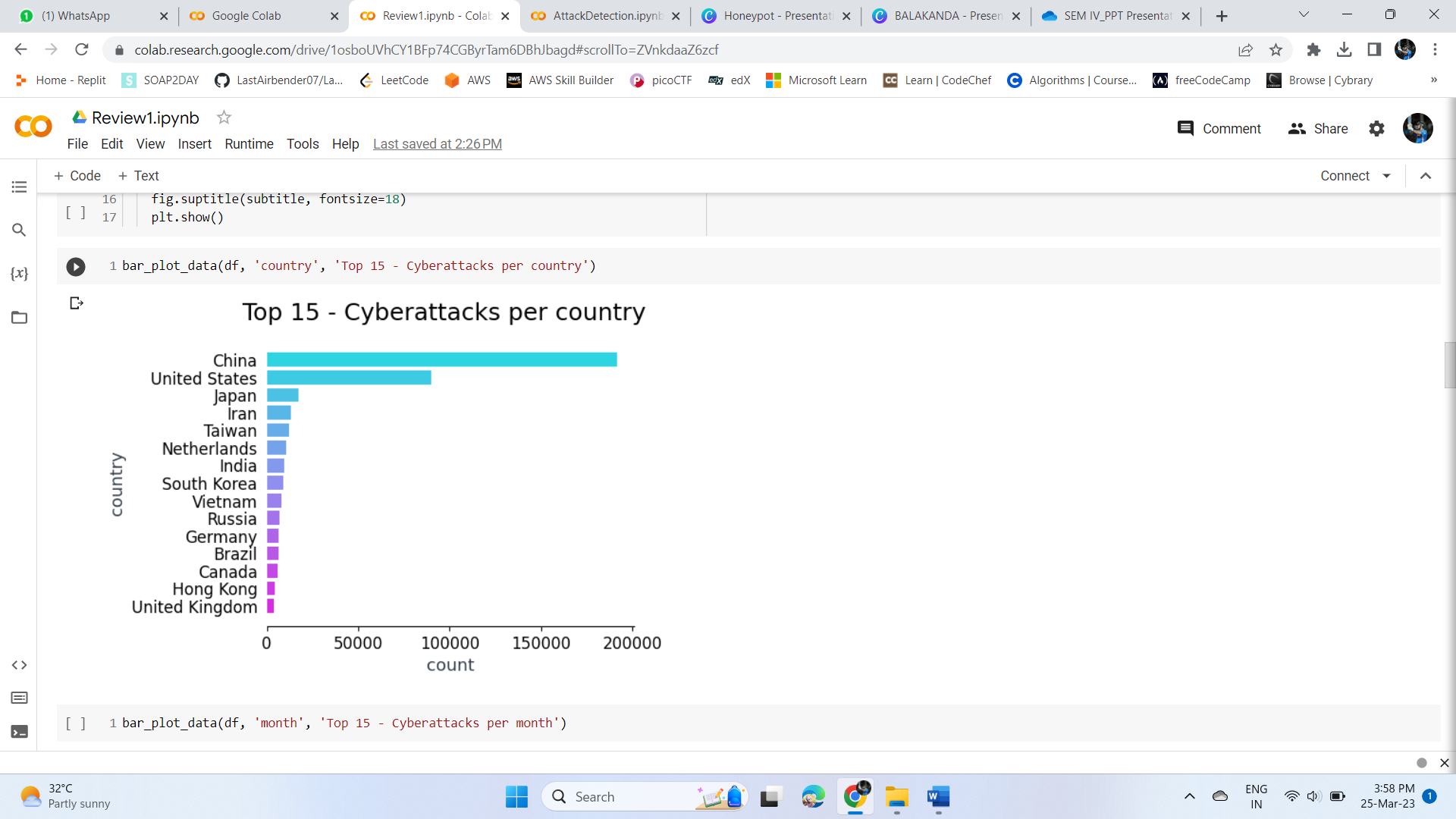
The first step would be to preprocess the data collected from the honeypot and convert it into a suitable format for analysis. This may involve cleaning the data, removing outliers and noise, and converting it into a structured format.

Next, machine learning algorithms such as clustering or anomaly detection can be applied to the preprocessed data to identify patterns and anomalies in the data. This may include identifying attack types, attack sources, or attack frequencies.

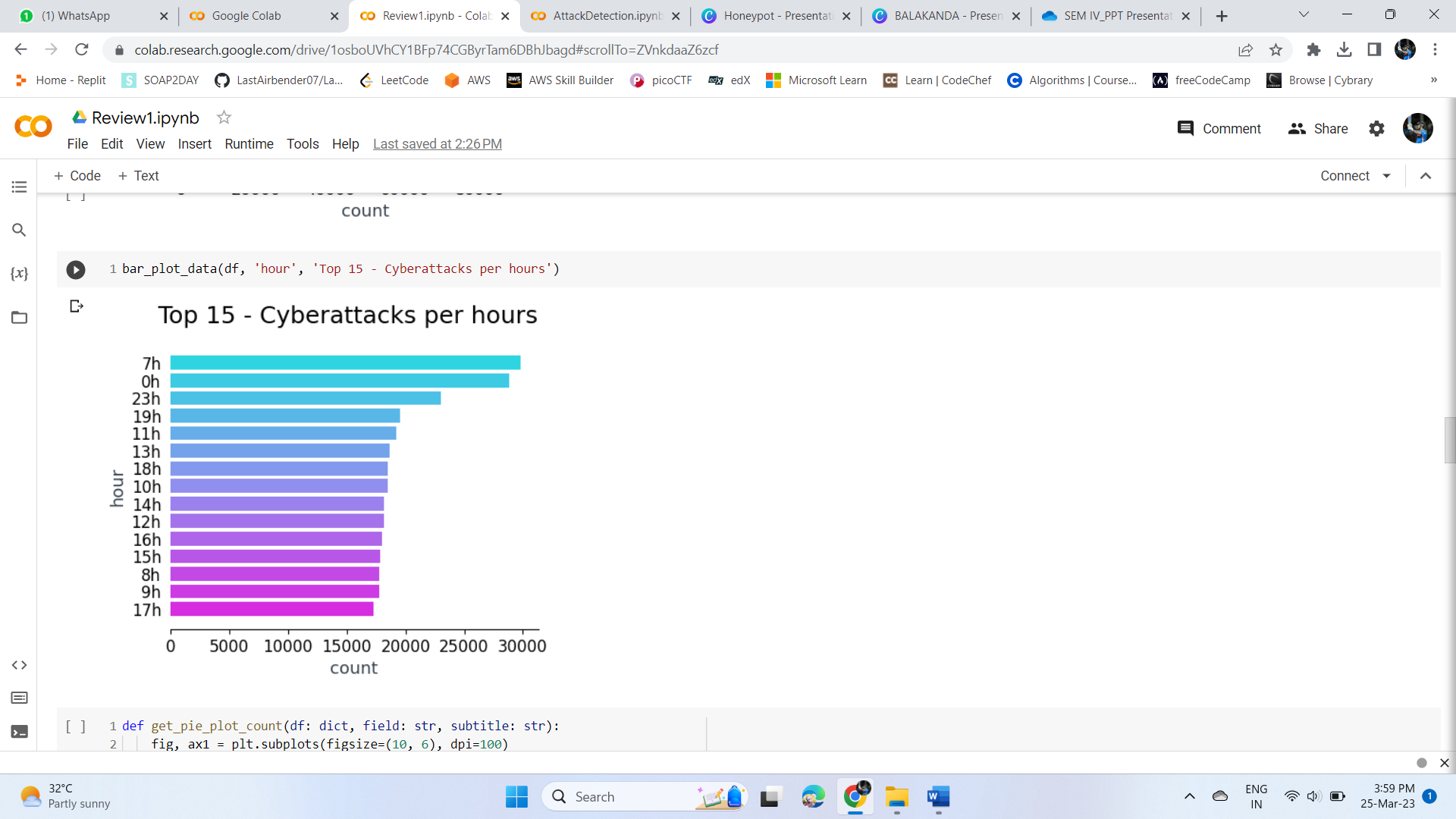
Once the machine learning analysis is complete, the results can be visualized using Matplotlib and Seaborn. These tools enable the creation of various types of visualizations such as bar charts, pie charts, scatter plots, and heatmaps. The visualizations can be used to identify trends and patterns in the data, making it easier to understand the types of attacks being launched and the methods used by attackers.

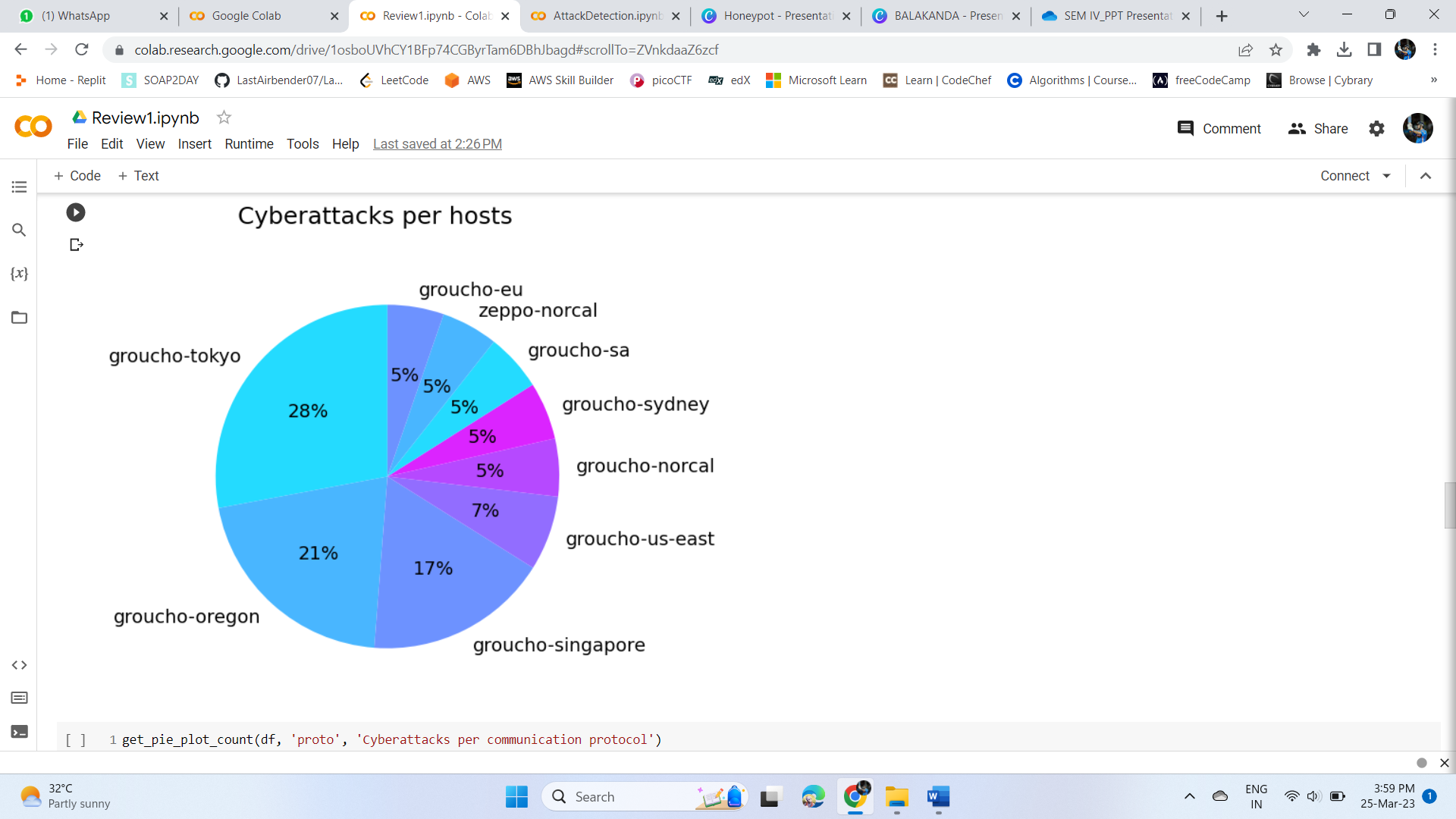
Overall, the proposed method involves using machine learning algorithms to identify patterns and anomalies in the data and using data visualization tools to create visualizations of the results. This approach can provide valuable insights into the threat landscape and enable organizations to develop more effective security measures.

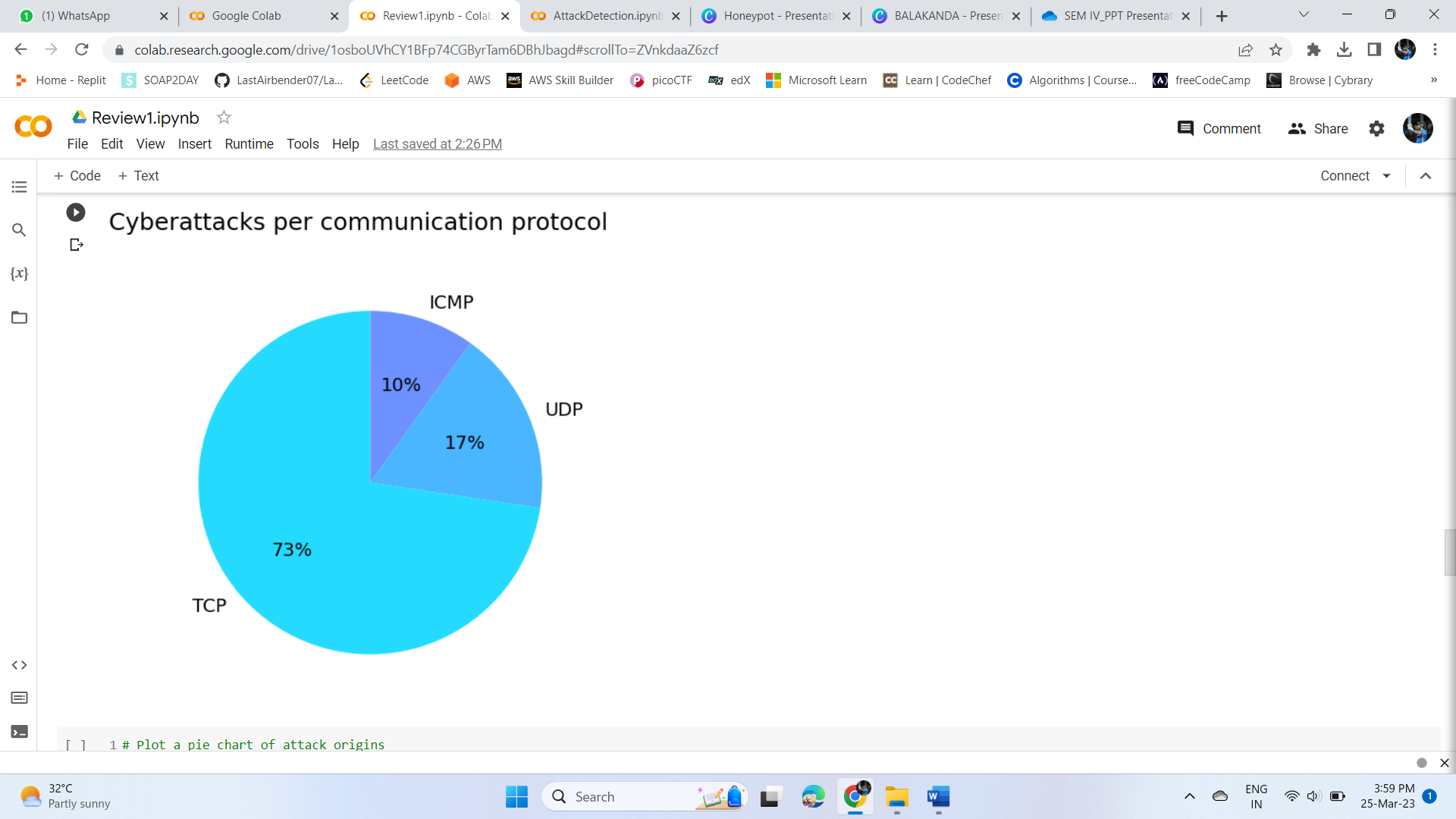
1. **RESULTS**



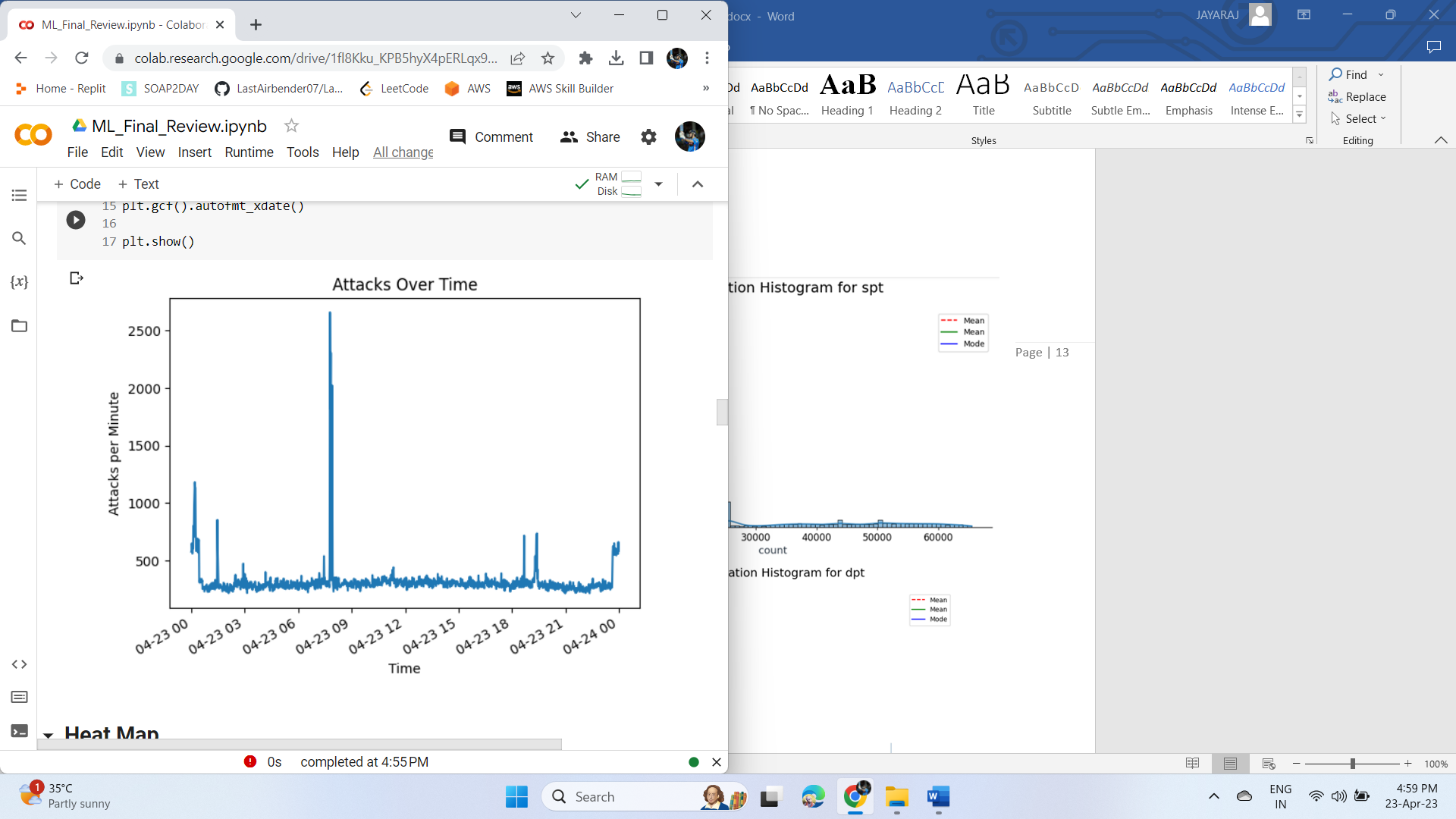


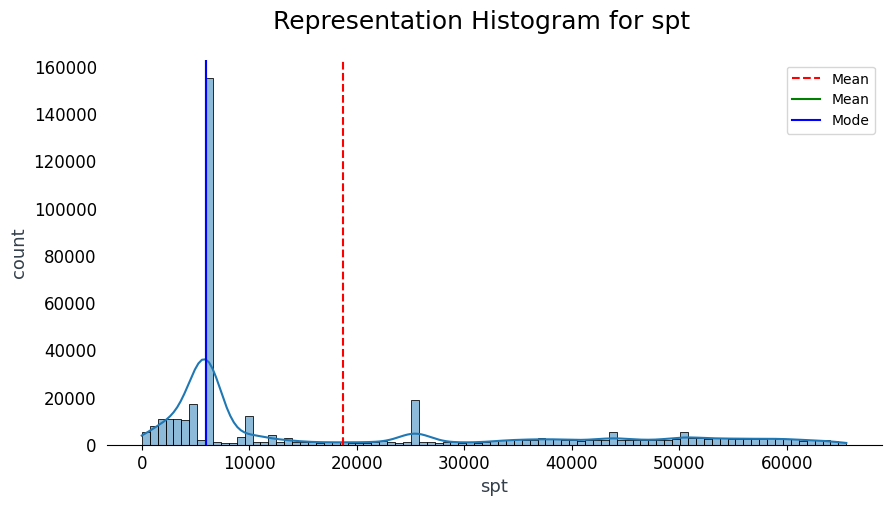


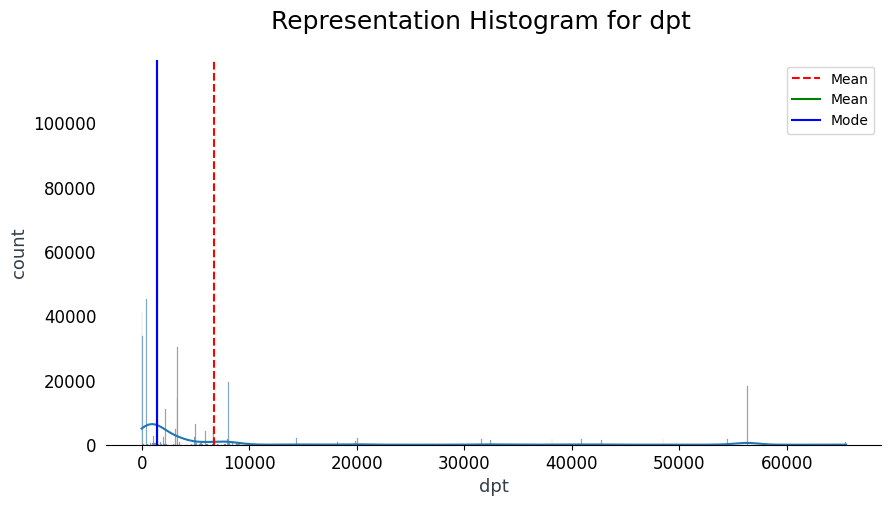


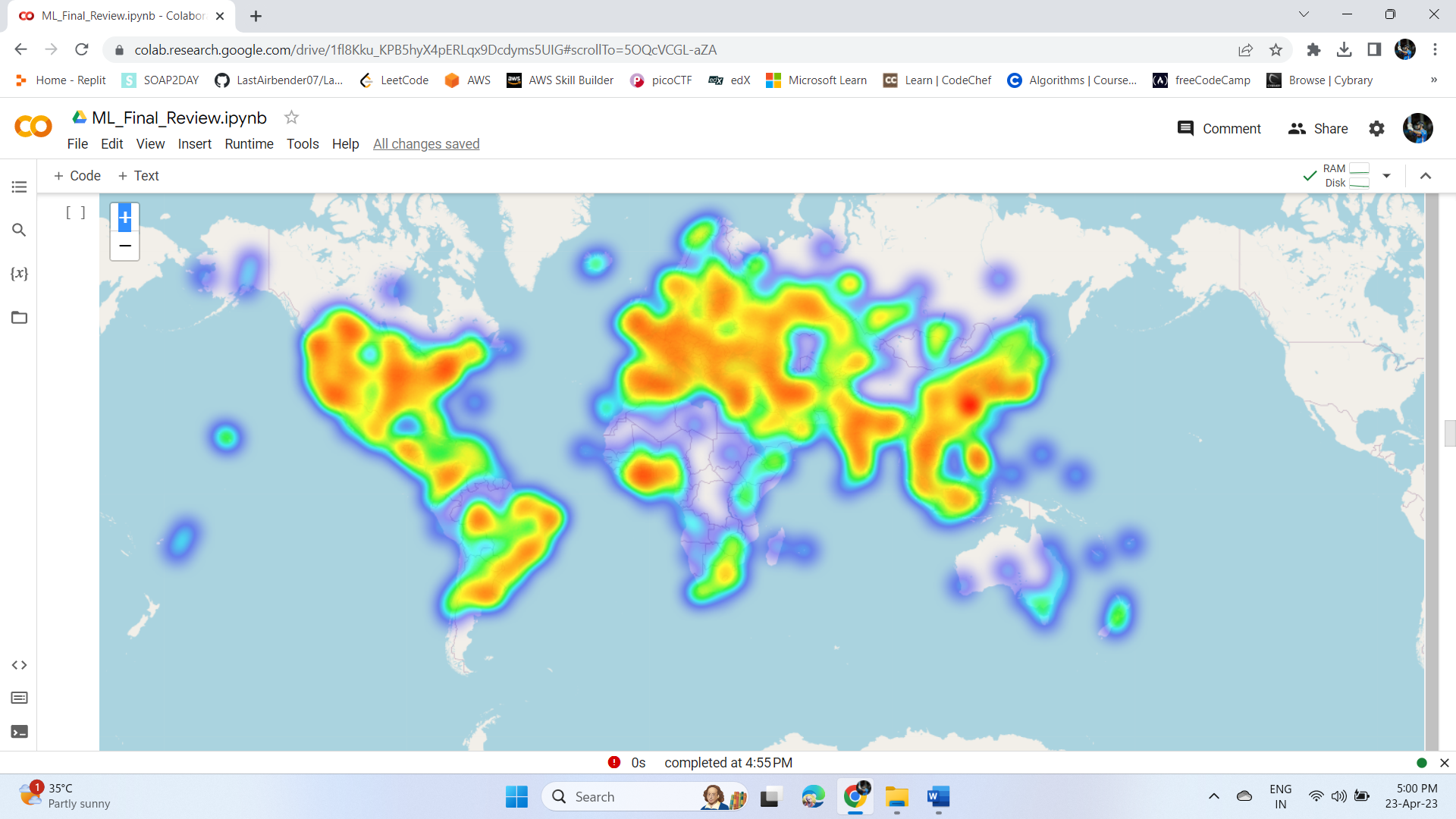


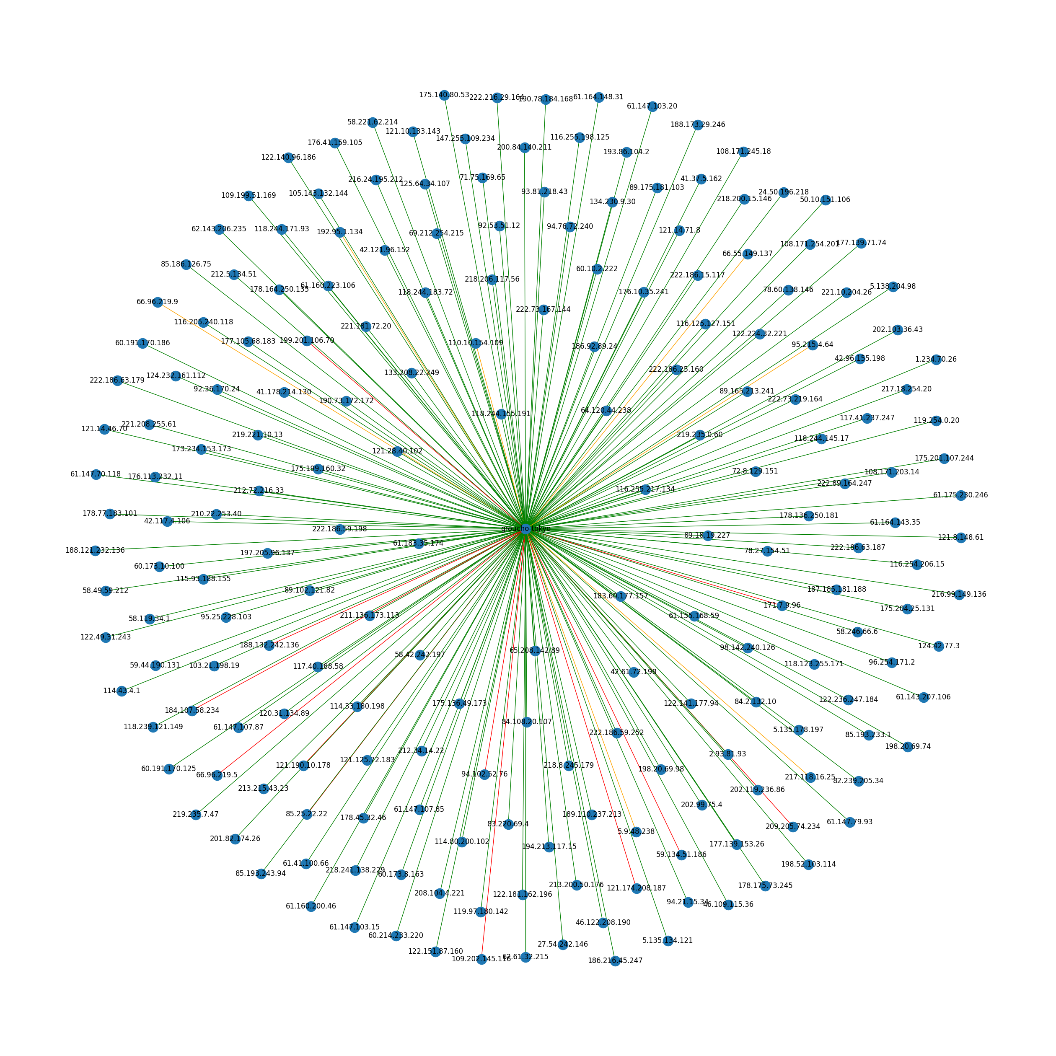


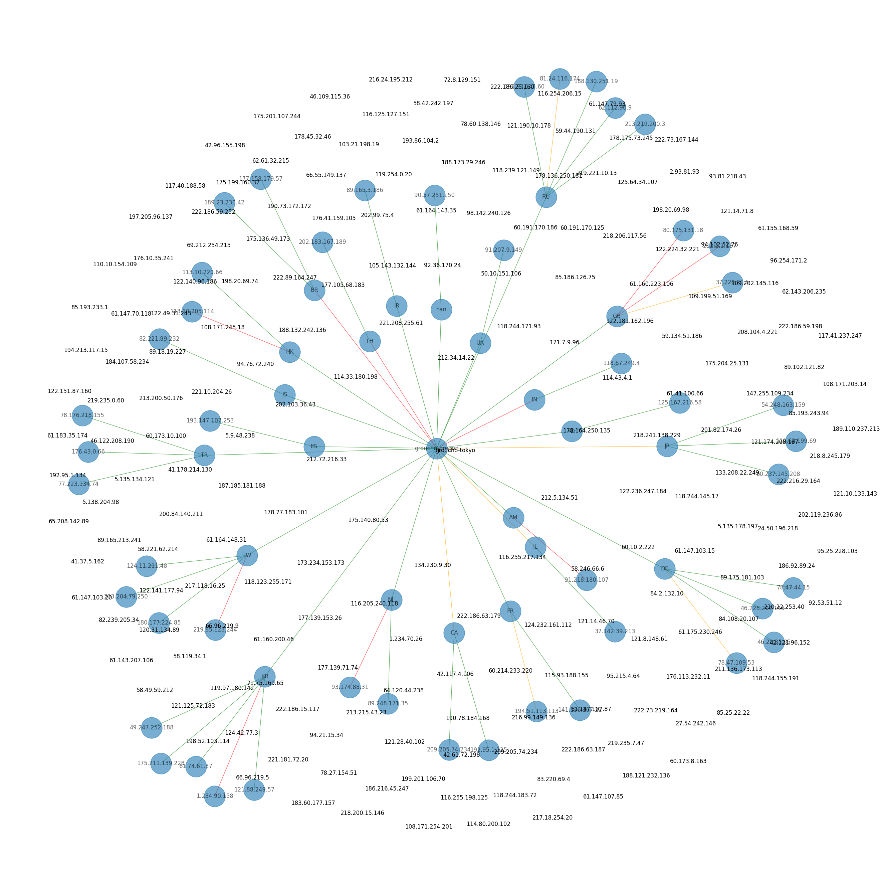


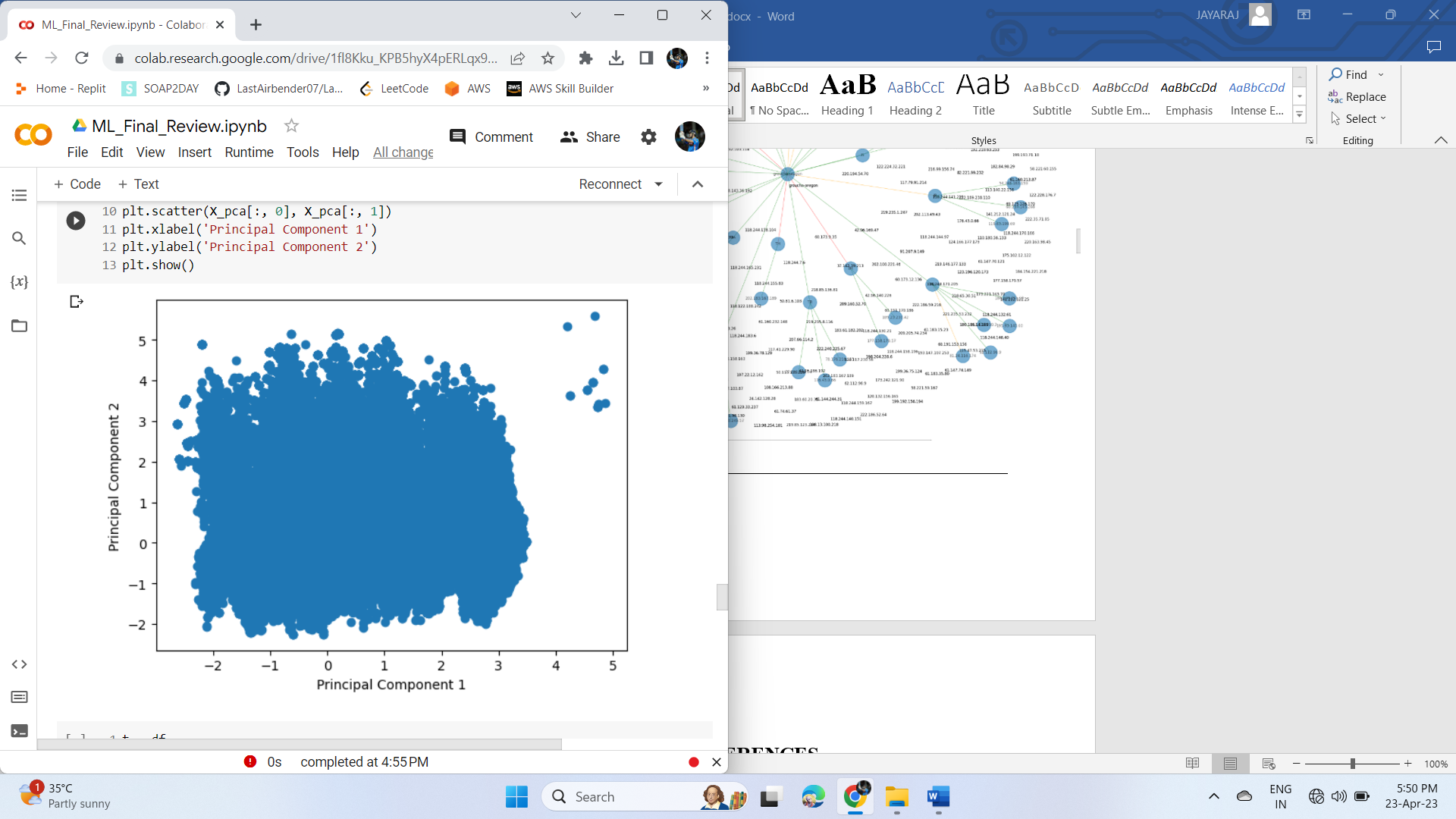
****

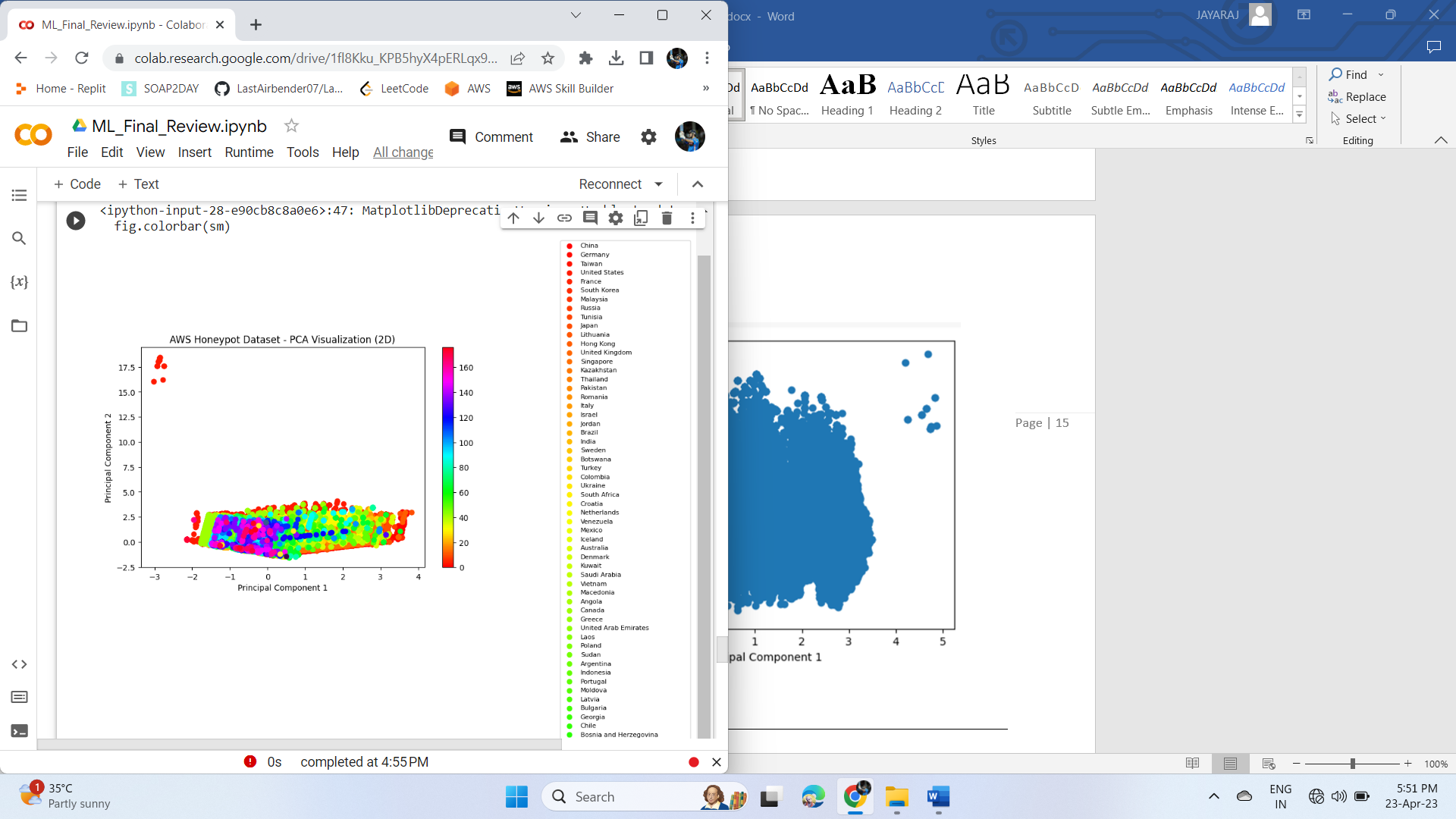
****

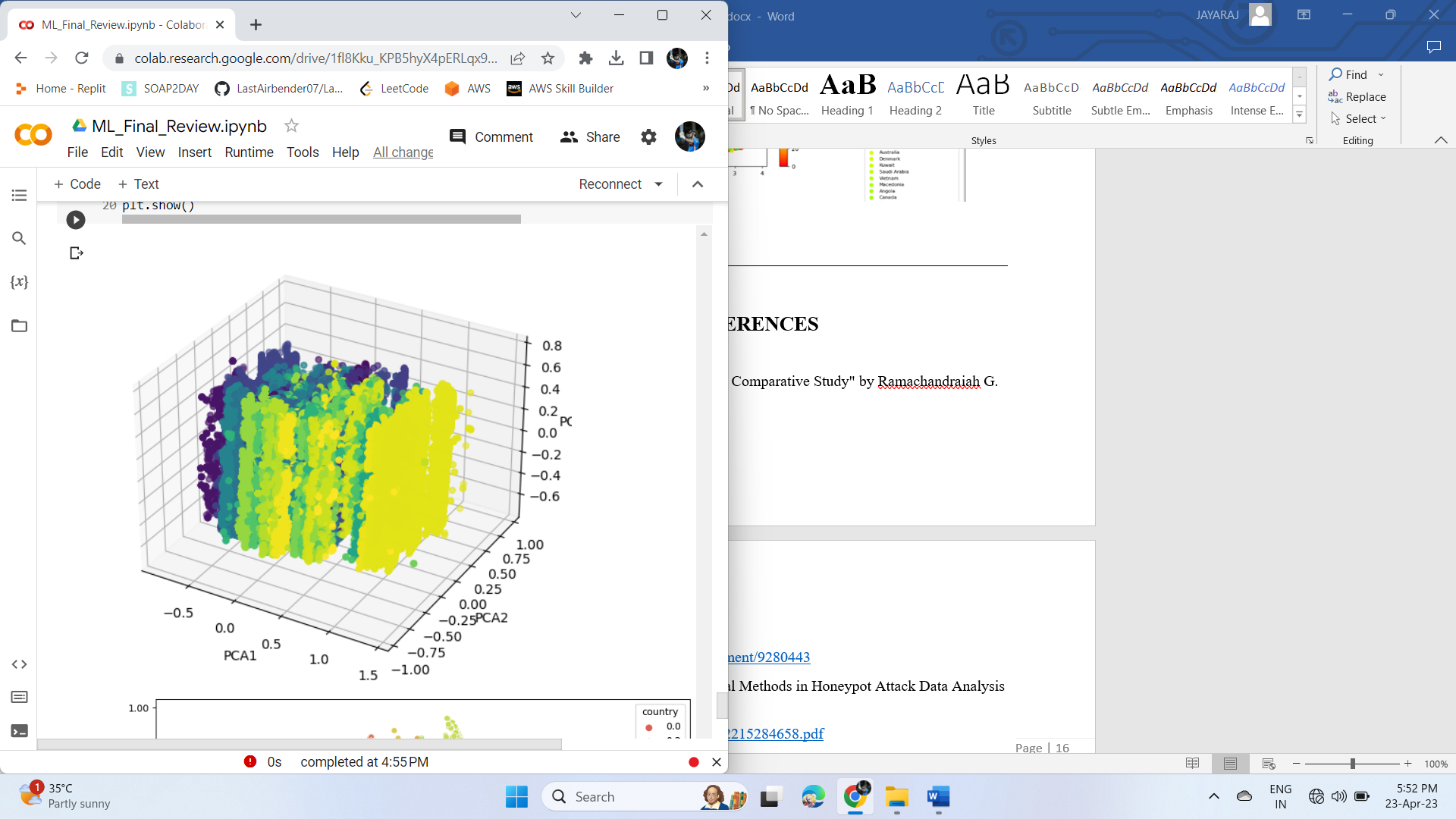


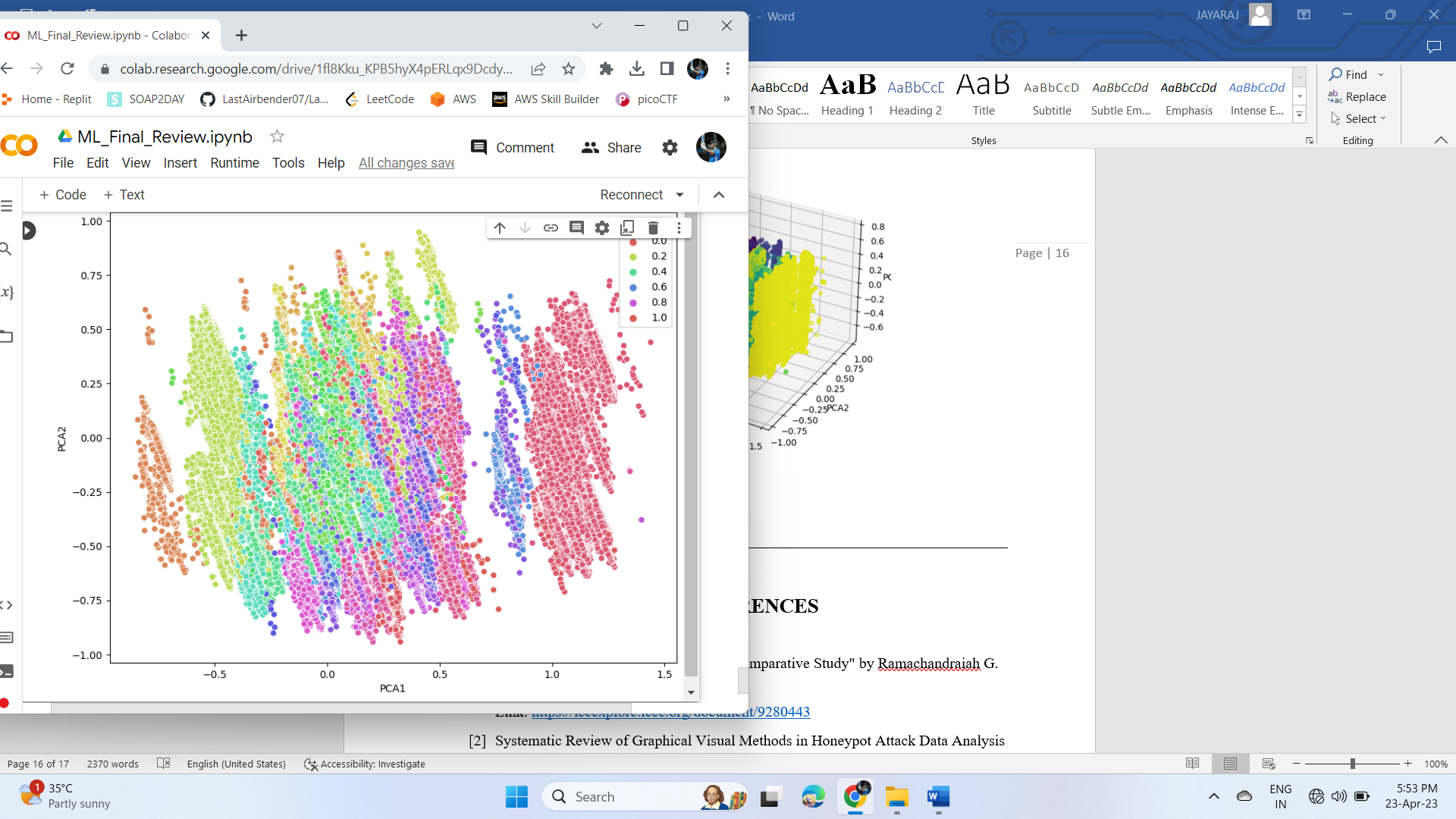












1. **CONCLUSION**

This Dataset has 4,51,581 data points collected from 9:53pm on 3 March 2013 to 5:55am on 8 September 2013.The deployment of a Honeypot can be beneficial in determining current attack strategies and probes being used by hackers due to the information collected.

Using the AWS Honeypot, attacks were identified on a number of ports and the most popular port attacked was spt 6000 to dpt 1433, as a compromise here can have the biggest reward for the attacker. The time of day at the source was interesting as the majority of attacks occurred between 20:00 pm and 0:30 am.

Most used hosts were “groucho-tokyo” and “groucho-oregon” and the most repeated IP address were “175.146.199.252” and “200.186.189.218” and a majority of the IPs were traced to be from China (42%) and US (20%).

However, it is important to note that this dataset is a honeypot dataset, which means that the attacks recorded in the dataset may not be representative of real-world attacks. Therefore, any conclusions drawn from this dataset should be considered with caution and validated using other sources of data.

1. **REFERENCES**

[1] Honeypot Analysis: An Overview and Comparative Study" by Ramachandraiah G. and Venkatesha

Link: <https://ieeexplore.ieee.org/document/9280443>

[2] Systematic Review of Graphical Visual Methods in Honeypot Attack Data Analysis by Gbenga Ikuomenisan

<https://www.scirp.org/pdf/jis_2022082215284658.pdf>

# [3] Honeypots and honeynets: issues of privacy

# <https://jis-eurasipjournals.springeropen.com/articles/10.1186/s13635-017-0057->

# [4] “An Overview of Honeypot Systems” by Neha Titarmare

# <https://www.researchgate.net/publication/332113726_An_Overview_of_Honeypot_Systems>

[5] Honeypot Analysis: Tools and Techniques" by Muhammad Munwar Iqbal, Syed Asad Ali Shah, and Muhammad Arif.

[6] Honeypot Technologies for Malware Detection and Analysis by Dragoș Draghicescu, PhD

<https://web.archive.org/web/20210921210754/https://revista.unap.ro/index.php/XXI_FCSM/article/download/1145/1116>

[7] Hatice Beyza, “Password Attack Analysis Over Honeypot” MatDer 1995

<https://dergipark.org.tr/en/download/article-file/1879578>

[8] “Honeypot Data Analysis” from the book Honeypots for windows by Roger A. Grimes

<https://link.springer.com/chapter/10.1007/978-1-4302-0007-9_11>

[9] Network Defence Strategy by Jianchao Hong

<https://iopscience.iop.org/article/10.1088/1757-899X/322/5/052033/pdf>

[10] Threat Prediction using Honeypot and Machine Learning - Prof. Suvarna Aranjo

<https://web.archive.org/web/20220809040120/https://www.ijraset.com/best-journal/threat-prediction-using-honeypot-and-machine-learning>