

# 

**□** Author: ◊

SUHAS A. BARAGUNDI

**ឱ** Course Name: **◊** 

**DATA ANALYSIS** 



MAIMUNISA. MAM



### What you mean by cyber security:-

Cybersecurity is the practice of protecting computer systems, networks, and data from unauthorized access, damage, or disruption by cyber threats, encompassing technologies, processes, and policies to mitigate risks.

There are 6 types of cybersecurity measures discussed in this article – network security, application security, information security, cloud security, IoT security, and identity and access management

Bob Thomas is often called the "father of cybersecurity". In 1971, he created the first computer virus, called the "Creeper virus",

#### Abstract:-

This project analyzes cybersecurity login patterns using Power BI and Python. The goal is to detect unusual login behaviors and predict failed login attempts based on access duration. The dataset includes login records with timestamps, locations, and statuses. Power BI was used for visualization, and Python was employed for exploratory data analysis (EDA) and linear regression. Key findings indicate that failed logins often originate from specific regions and peak during certain hours. The regression model suggests a correlation between access duration and failed login attempts, enabling predictive insights for cybersecurity monitoring.

### **Introduction:-**

Cybersecurity threats are a growing concern for organizations. Unauthorized access attempts and login failures can indicate security breaches or bruteforce attacks. Analyzing login patterns helps identify risks and enhance security measures. This project leverages Power BI for visual insights and Python for statistical analysis, aiming to improve security monitoring.

# Importance of Cybersecurity Analysis:-

- Rising cyber threats demand proactive monitoring of login patterns.
- Analyzing login trends helps in preventing unauthorized access.
- Organizations can leverage predictive models to enhance security.

### Dataset Description:-

- Dataset Name: Cybersecurity Login Data
- **Total Records:-** 100,000+
- Key Columns:
  - User\_ID: Unique identifier for users
  - Timestamp: Date and time of login attempt
  - Login\_Status: Success or failed login
  - IP Address: Source of login attempt
  - Location: Geographical region of login
  - Access Duration: Duration of login session

### **Data Cleaning Process:-**

- Checked for missing values and handled inconsistencies.
- Converted timestamp into separate date and hour columns.
- Extracted failed login attempts for deeper analysis.
- Applied grouping and aggregation for better insights.

#### Power BI Dashboard Analysis:-

Power BI was used to create an interactive dashboard to analyze login trends.

### Visualizations in Power BI:-

Failed Logins by Location  $\rightarrow$  A geo-map highlights high-risk regions.

**Login Trends Over Time**\_→ A line chart showing peak login failure hours.

**Bar Chart for Failed Logins**  $\rightarrow$  Displays login failure rates per location.

Key Performance Indicators (KPIs) $\rightarrow$ Total login attempts, success vs. failure rates.

### Power BI Insights:-

- Geographic analysis helps detect cyber attack sources.
- Time-based trends reveal login vulnerabilities during specific hours.
- KPIs provide quick security monitoring statistics.

#### Power BI Dashboard

# **Cyber Security Analysis**

14**M** 

Sum of Access\_Duration

144K

Sum of Failed\_Login\_Count

Location

Beijing, China

London, UK

Mumbai, Indi

US

Login\_Status

Success

Success

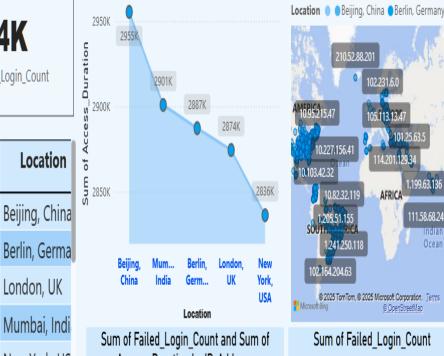
Success

10.00K

Count of User ID

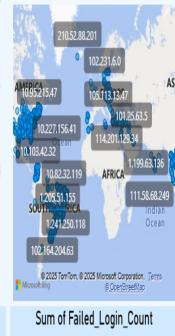
#### IP\_Address and its Location

	ai, I
ss New Yo	rk,
<u>.</u>	-1
ddress Count of U	ser_l
91.48	
53.12	
33.62	
12.79	
27.172	
9.188	
31	
1	000
	Count of U.  21.48  33.12  33.62  32.79  27.172  2.188



Sum of Access\_Duration by Location

Access Duration by IP Address



IP Address and Location

by Login\_Status



115K 29K Failed Success Login\_Status

# Exploratory Data Analysis (EDA) in Python:-

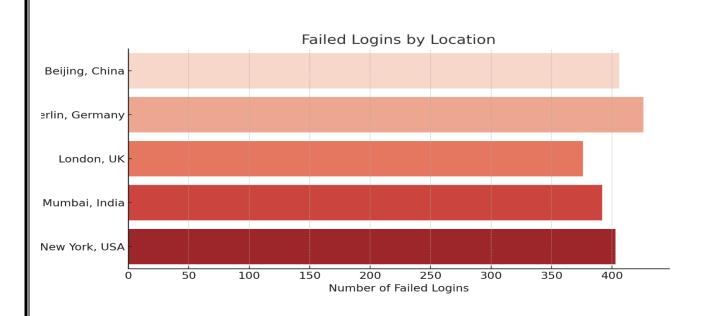
EDA was performed using Python libraries

- Pandas (Data Processing)
- > Seaborn & Matplotlib (Visualization)
- Scikit-learn (Linear Regression)

to find trends in login behavior.

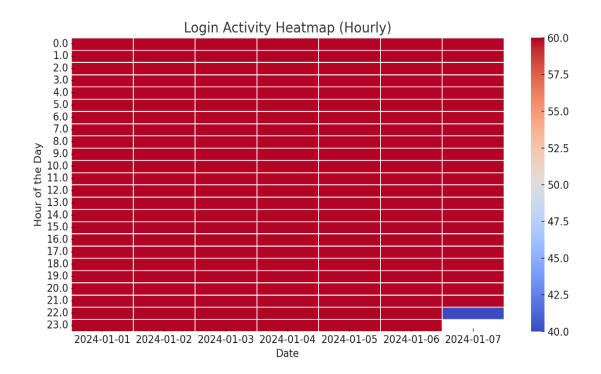
## Failed Logins by Location:-

Most failed logins originate from specific regions like New York, Beijing, and Berlin.



### Login Activity Heatmap:-

Peak failed login attempts occur during midnight to early morning hours.



### **User Login Trends:-**

Certain users have repeated failed attempts, indicating possible brute-force attacks.

### **Linear Regression Analysis:-**

### **Objective:-**

To predict the number of failed logins based on access duration.

#### **Model Used:-**

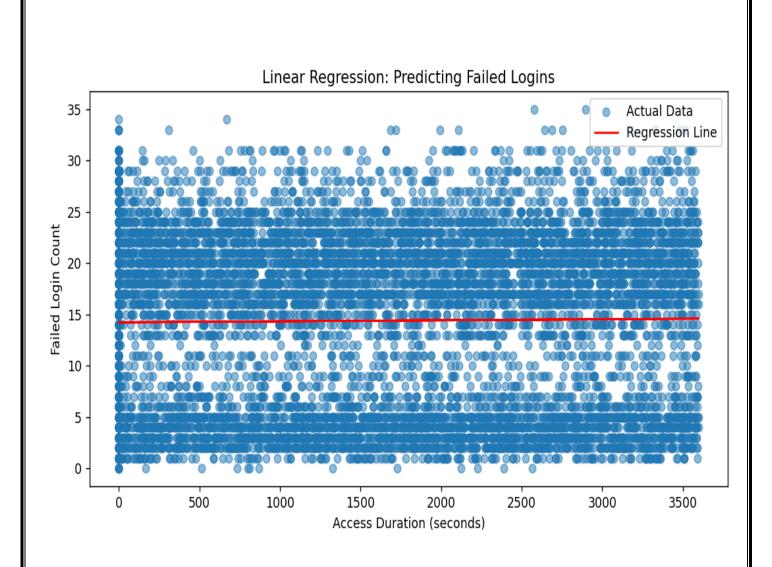
Linear Regression (Scikit-learn)Train-test split: 80% training, 20% testing

#### **Results:-**

Strong correlation found between access duration and failed login attempts.

Regression line equation:

Failed Logins =  $(m \times Access Duration) + b$ 



### Interpretation:-

Higher access duration often correlates with an increased number of failed login attempts. The model can be used to flag risky login sessions.

### Key Findings & Insights:-

### **Geographical Trends:-**

Certain locations show significantly higher failed login attempts, potentially indicating cyber attack hotspots

#### **Time-Based Trends:-**

Most failed logins occur during late-night hours.

#### **Predictive Model Performance:-**

Linear regression provides valuable insights into login failure patterns and can help enhance security monitoring.

#### Future Work:-

- Implement advanced machine learning models for anomaly detection.
- Use real-time alert systems in Power BI for immediate security monitoring.
- Expand the dataset for deeper analysis on network-based attacks.

# <u>References:-</u>

- Power BI documentation
- Pandas & Seaborn libraries for data visualization
- Scikit-learn for machine learning analysis
- Cybersecurity research articles on login pattern detection

#### **Conclusion:-**

- Power BI provided clear visual insights into login patterns.
- Python helped identify data-driven trends in failed login attempts.
- Linear regression demonstrated predictive potential for cybersecurity monitoring.