**Objectives of the project:**

1. **Identify Intrusions and Patterns**

* Investigate the relationship between login attempts, failed logins, and attack detection to identify potential attack patterns or anomalies.
* Analyze unusual time access and protocol types to identify suspicious behaviors associated with attacks.

1. **Session Behavior Analysis**

* Explore the duration of sessions based on encryption type and protocol type to find correlations between session length and attack likelihood.

1. **Protocol and Browser Analysis**

* Determine how different protocols and browsers impact the frequency of attacks and session behavior.

1. **Predictive Modelling of Attacks**

* Build a predictive model (using SQL for basic queries and aggregates) that could help in early detection of intrusions based on past data patterns.

1. **Detection of Multiple Failed Login Attempts**

* Analyze failed login attempts with a specific threshold to detect potential brute force or credential stuffing attacks.

**SQL Queries and Outputs:**

**1. Session Duration vs. Attack Detection**

**Objective:** Determine if session duration correlates with attack detection. Suspiciously short or long sessions might have higher attack probabilities.

* What is the average session duration for sessions with and without attack detection, and how does it vary across different protocol types?

**Query:**

SELECT

protocol\_type,

AVG(session\_duration) AS avg\_session\_duration,

COUNT(CASE WHEN attack\_detected = 'Yes' THEN 1 END) AS attack\_count,

COUNT(CASE WHEN attack\_detected = 'No' THEN 1 END) AS normal\_count

FROM cybersecurity\_intrusion\_data

GROUP BY protocol\_type;

**Output Example:**

|  |  |  |  |
| --- | --- | --- | --- |
| Protocol\_type | avg\_session\_duration | Attack\_count | Normal\_count |
| TCP | 278.6241204611592 | 3263 | 2682 |
| UDP | 285.0403640455737 | 1182 | 1002 |
| ICMP | 270.4286439684935 | 269 | 189 |

**2. Failed Logins and Attack Detection Correlation**

**Objective:** Find if there's a strong relationship between failed login attempts and attack detection.

* How many failed login attempts occurred before an attack was detected, and is there a pattern?

**Query:**

SELECT

session\_id,

SUM(failed\_logins) AS total\_failed\_logins,

attack\_detected

FROM cybersecurity\_intrusion\_data

GROUP BY session\_id, attack\_detected

HAVING SUM(failed\_logins) > 3; -- Threshold of failed logins (for example, 3 or more failed logins)

**Output Example:**

|  |  |  |
| --- | --- | --- |
| Session\_id | Total\_failed\_logins | Attack\_detected |
| SID\_yesyesyes39 | 4 | no |
| SID\_yesyesyes68 | 4 | no |
| SID\_yesyesno58 | 4 | no |
| SID\_yesyes259 | 4 | no |
| SID\_yesyes264 | 4 | no |
| SID\_yesyes273 | 4 | no |
| SID\_yesyes3yes9 | 4 | no |
| SID\_yesyes352 | 4 | no |
| SID\_yesyes364 | 4 | no |
| SID\_yesyes382 | 4 | no |
| SID\_yesyes4yes… | 4 | no |
| SID\_yesyes424 | 4 | no |
| SID\_yesyes49no | 4 | no |
| SID\_yesyes494 | 4 | no |
| SID\_yesyes5yes6 | 4 | no |
| SID\_yesno7yes3 | 4 | no |

**3. Protocol and Browser Usage with Attack Detection**

**Objective:** Examine how attack frequency varies with different protocols and browsers.

* Which protocols and browsers are most commonly associated with attack detection?

**Query:**

SELECT

protocol\_type,

browser\_type,

COUNT(session\_id) AS total\_sessions,

SUM(CASE WHEN attack\_detected = 'Yes' THEN 1 ELSE 0 END) AS attack\_count

FROM cybersecurity\_intrusion\_data

GROUP BY protocol\_type, browser\_type

ORDER BY attack\_count DESC;

**Output Example:**

|  |  |  |  |
| --- | --- | --- | --- |
| Protocol\_type | Browser\_type | Total\_sessions | Attack\_count |
| TCP | Chrome | 3173 | 1812 |
| TCP | FireFox | 1233 | 694 |
| UDP | Chrome | 1173 | 657 |
| TCP | Edge | 929 | 506 |
| UDP | FireFox | 450 | 251 |
| UDP | Edge | 323 | 199 |
| TCP | Safari | 311 | 182 |
| ICMP | Chrome | 264 | 161 |
| TCP | Unknown | 299 | 69 |
| UDP | Safari | 104 | 55 |

**4. Predictive Analysis of Attack Detection Based on Failed Logins**

**Objective:** Build a predictive query to detect potential attacks based on failed login attempts.

* Can we predict if an attack is more likely based on failed logins within a session?

**Query:**

SELECT

session\_id,

failed\_logins,

CASE

WHEN failed\_logins > 5 THEN 'High Risk'

WHEN failed\_logins BETWEEN 3 AND 5 THEN 'Moderate Risk'

ELSE 'Low Risk'

END AS risk\_level,

attack\_detected

FROM cybersecurity\_intrusion\_data

WHERE failed\_logins > 2; -- Focus on sessions with failed login attempts

**Output Example:**

|  |  |  |  |
| --- | --- | --- | --- |
| Session\_id | Failed\_logins | Risk\_level | Attack\_detected |
| SID\_yesyesyes96 | 3 | Moderate risk | no |
| SID\_yesyesyesno5 | 3 | Moderate risk | no |
| SID\_yesyesyesno6 | 3 | Moderate risk | no |
| SID\_yesyesyesno7 | 3 | Moderate risk | no |
| SID\_yesyesyesnono | 3 | Moderate risk | no |