Proyecto de Análisis de Datos: Exploración de Patrones en Eventos de Ciberseguridad

1. Introducción

La ciberseguridad es un campo crítico para las organizaciones, y la detección de patrones en eventos de seguridad es clave para prevenir ataques y reforzar defensas. Este proyecto se centra en analizar un conjunto de datos sobre ataques de ciberseguridad con el fin de identificar tendencias, comportamientos maliciosos recurrentes y vulnerabilidades potenciales en la red.

2. Objetivo General

Analizar patrones temporales, geográficos, y técnicos en los ciberataques registrados para identificar vulnerabilidades críticas y generar estrategias de seguridad informática más efectivas.

3. Objetivos Específicos

- **Identificar patrones temporales en los ciberataques** para determinar si existen tendencias específicas relacionadas con horas del día, días de la semana o períodos determinados.
- **Evaluar la asociación entre protocolos inseguros y ataques de alta severidad**, identificando cuáles representan mayores riesgos en el tráfico analizado.
- **Determinar los puertos de destino más vulnerables**, analizando el volumen de ataques dirigidos hacia ellos y los tipos de tráfico asociados.
- **Analizar patrones geográficos en las fuentes de ciberataques**, identificando regiones y ubicaciones que generan la mayor cantidad de amenazas según datos de geolocalización.
- **Examinar la relación entre las firmas de ataque detectadas y las acciones tomadas**, evaluando la probabilidad de respuestas específicas como bloqueos, alertas o registros.

4. Planteamiento del Problema

Los datos de eventos de ciberseguridad pueden ser complejos y contener valores ruidosos o inconsistentes, dificultando la detección de patrones claros. Este proyecto busca abordar estas limitaciones limpiando y analizando los datos para extraer información útil, ayudando a las organizaciones a priorizar medidas de defensa y mitigación.

5. Preguntas de Investigación

- **Análisis temporal de ciberataques:** ¿Hay patrones de tiempo específicos (diurnos, nocturnos, días de la semana) en los ataques según las marcas de tiempo?
- **Evaluación de protocolos inseguros:** ¿Qué protocolos están más frecuentemente asociados con ataques clasificados como de alto nivel de gravedad?}
- **Identificación de puertos vulnerables:** ¿Cuáles son los puertos de destino más comúnmente atacados y qué tipo de tráfico se dirige a ellos?
- **Análisis de patrones geográficos:** ¿Qué regiones geográficas generan la mayor cantidad de ataques según las direcciones IP de origen y los datos de geolocalización?
- **Relación entre firmas de ataque y acciones tomadas:** ¿Qué firmas de ataque tienen una mayor probabilidad de desencadenar acciones específicas (como bloqueos, alertas, etc.)?

6. Posible Solución

La solución consiste en usar Python y bibliotecas como pandas y numpy, para:

- Limpiar y organizar los datos.
- Visualizar patrones y relaciones entre variables clave.
- Generar estadísticas descriptivas y visualizaciones para responder las preguntas planteadas.
- Detectar correlaciones significativas entre eventos, protocolos, segmentos de red y geolocalización.

7. Consideraciones Futuras

Con una base de datos procesada, será posible implementar modelos predictivos para detectar comportamientos maliciosos en tiempo real. Esto puede incluir algoritmos de aprendizaje automático para clasificar tipos de ataques o identificar eventos con altas probabilidades de ser críticos.

8. Conclusión

El análisis de ciberataques revela que la mayoría ocurren en horarios nocturnos, con picos leves los domingos, lo que resalta la necesidad de vigilancia en horarios críticos. Los protocolos TCP, ICMP y UDP, junto con los puertos 34117 y 7508, son los más atacados, especialmente en servicios HTTP, DNS y FTP, lo que indica intentos de explotar servicios clave. La mayoría de los ataques se originan en regiones específicas de India, vinculadas a su creciente infraestructura tecnológica. Finalmente, el sistema responde de forma consistente ante diferentes firmas de ataque, mostrando un enfoque equilibrado en la gestión de amenazas.

9. Entregables

1. **Código en Google Colab:**
    - Limpieza y análisis de los datos.
    - Visualizaciones de patrones detectados.
2. **Informe detallado:**
    - Documento que explique los pasos tomados y los hallazgos clave.
3. **Presentación en PowerPoint:**
    - Resumen de los resultados y conclusiones para una audiencia técnica o directiva.