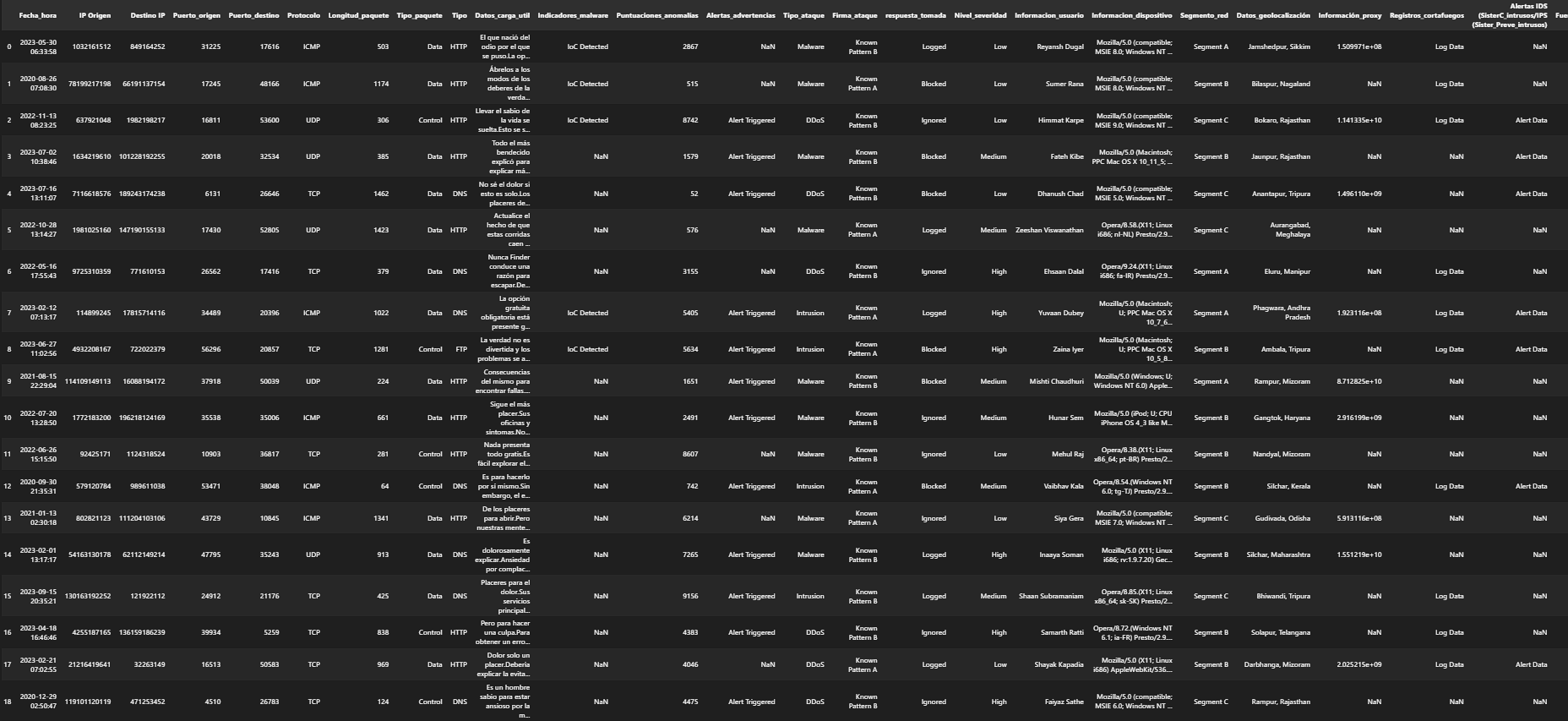
**ESTUDIO PREVIO DEL DATASET**

**ESTRUCTURA DEL DATASET**

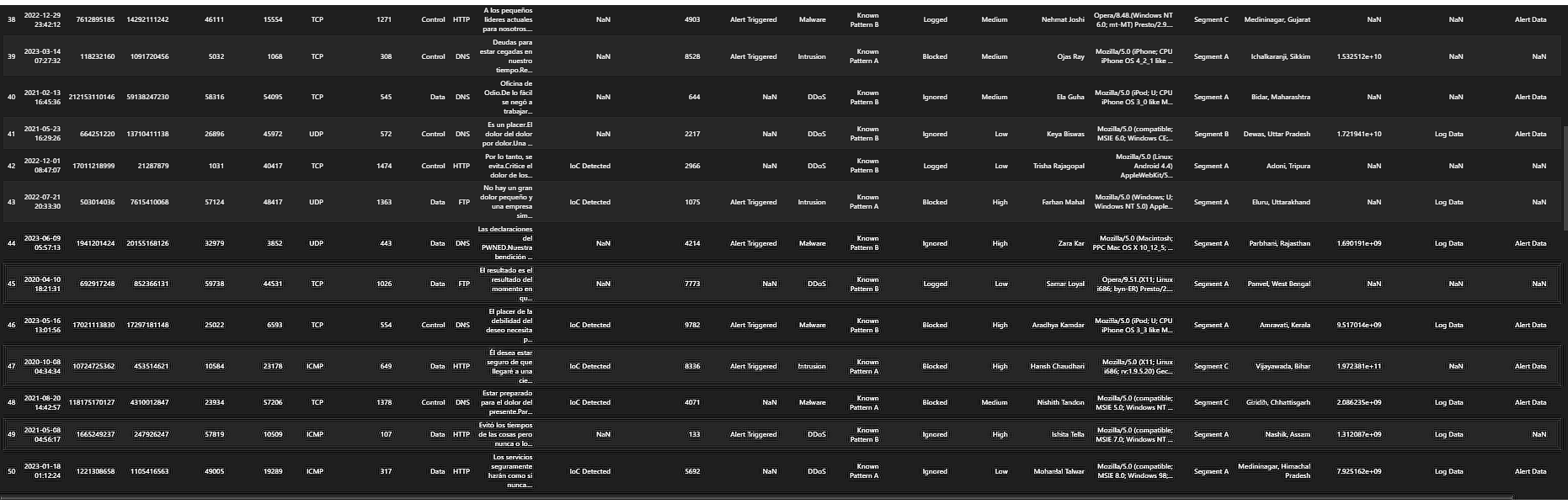
El Dataset ha sido extraído de le web: <https://www.kaggle.com/datasets/teamincribo/cyber-security-attacks> y trata se compone de 26 columnas con 40000 registros cada una de ellas. Las he traducido al castellano y he modificado alguna columna. Hay NaN pero son datos que necesaria saber del que hizo el Dataset ya que la media en un puerto de salida no tiene mucho sentido.

Em general, Trata sobre ataques Malware, de intrusión y DDoS (denegación de servicios), aportando muchos datos técnicos como IPs, puertos, geolocalizaciones, etc.,

**PRIMERAS 50 FILAS**

****

****

****

**DESCRIPCION DE LAS COLUMNAS**

****

**PREGUNTAS INICIALES**

1.- Distribución de Ataques:

* ¿Cuál es la distribución de los diferentes tipos de ataques en el conjunto de datos?



* ¿Cuántos incidentes de cada tipo de ataque se registran?

La columna 'Tipo\_ataque' contiene 3 tipos de ataque: Malware, DDoS y Desconocido. Hay un total de 40.000 incidentes registrados en el marco de datos, con 16.000 incidentes de malware, 17.000 incidentes de DDoS y 7.000 incidentes de desconocidos.

Finalidad: Obtener una comprensión general de la incidencia de diferentes tipos de ataques.

2.- Análisis Temporal:

* ¿Cómo varían los ataques a lo largo del tiempo en el conjunto de datos?

Para analizar la variación de los ataques a lo largo del tiempo en el conjunto de datos, podemos agrupar los datos por la columna Fecha\_hora y luego contar el número de ocurrencias de cada tipo de ataque (Tipo\_ataque), carga útil de datos (Datos\_carga\_util), análisis de resultados (Resultado\_analisis), etiqueta (Etiqueta\_x), indicadores de malware (Indicadores\_malware) y nivel de gravedad (Nivel\_severidad). Luego podemos trazar los resultados para visualizar la variación de los ataques a lo largo del tiempo.

* ¿Hay patrones estacionales o tendencias en los ataques a lo largo de los años?

Finalidad: Identificar posibles patrones temporales en los ataques.

3.- Origen de los Ataques:

* ¿Cuál es la distribución geográfica de los ataques (por país, región)?

Mapas de distribución por calor de ciudades y ips

* ¿Existen patrones en los países de origen de los ataques?

Si, todos están en la India

Finalidad: Entender la geografía de los ataques y posiblemente identificar ubicaciones de mayor riesgo.

4.- Correlaciones y Relaciones:

* ¿Hay correlaciones entre diferentes variables en el conjunto de datos?

Sí, existen correlaciones entre diferentes variables en el conjunto de datos, siendo es probable que las columnas 'Puntuaciones\_anomalias' y 'Tipo\_ataque' estén correlacionadas, ya que las puntuaciones de anomalías más altas generalmente están asociadas con actividad maliciosa.

* ¿Existen relaciones entre el tipo de ataque y otras características?

Finalidad: Identificar posibles factores que estén relacionados entre sí y podrían tener implicaciones en la ciberseguridad

**HIPOTESIS A ESTABLECER Y CONSECUENCIAS PRACTICAS**

1.- Preguntas:

* ¿Los ataques se dirigen a infraestructuras criticas u otras instituciones sensibles?
* ¿Los ataques se realizan durante el horario comercial?
* ¿Los ataques se dirigen a los sistemas ubicados en Europa?

2.- Consecuencias prácticas:

* Si se puede identificar que los ataques se dirigen a los sistemas de información financiera, se puede tomar medidas para reforzar la seguridad de estos sistemas.
* Si se puede identificar que los ataques se realizan durante el horario comercial, se pueden tomar medidas para aumentar la seguridad de la red durante esos intervalos.
* Si se puede identificar que los ataques se dirigen a sistemas infraestructuras criticas ubicados en Europa, se pueden tomar medidas para aumentar la seguridad em estas infraestructuras