Análisis Exploratorio en pruebas de detección de intrusiones

Este reporte está enfocado en el reconocimiento de intentos de penetración o ataques a una página web, la información se provee de cybersecurity.csv el cual compila distintas maneras de ver y comprender como los usuarios maliciosos intentan acceder o violar una página.

Preparación de los datos: es un reto rellenar los datos que no están completos así como dejar los datos lo más limpios posibles para entenderlos y hacerlos más legibles a la hora de trabajar con ellos.

Retos:

- **Las relaciones entre los datos**: hacer un gráfico con la cantidad de ataques que se han hecho comparándolo con 1 y 0 donde (1 es un ataque) y (0 es una sesión normal)

- **Datos nulos o vacíos**: verificar el estado de las columnas y determinar qué columnas tenían datos vacíos los cuales podrían perjudicar a la hora de trabajar con el dataset.

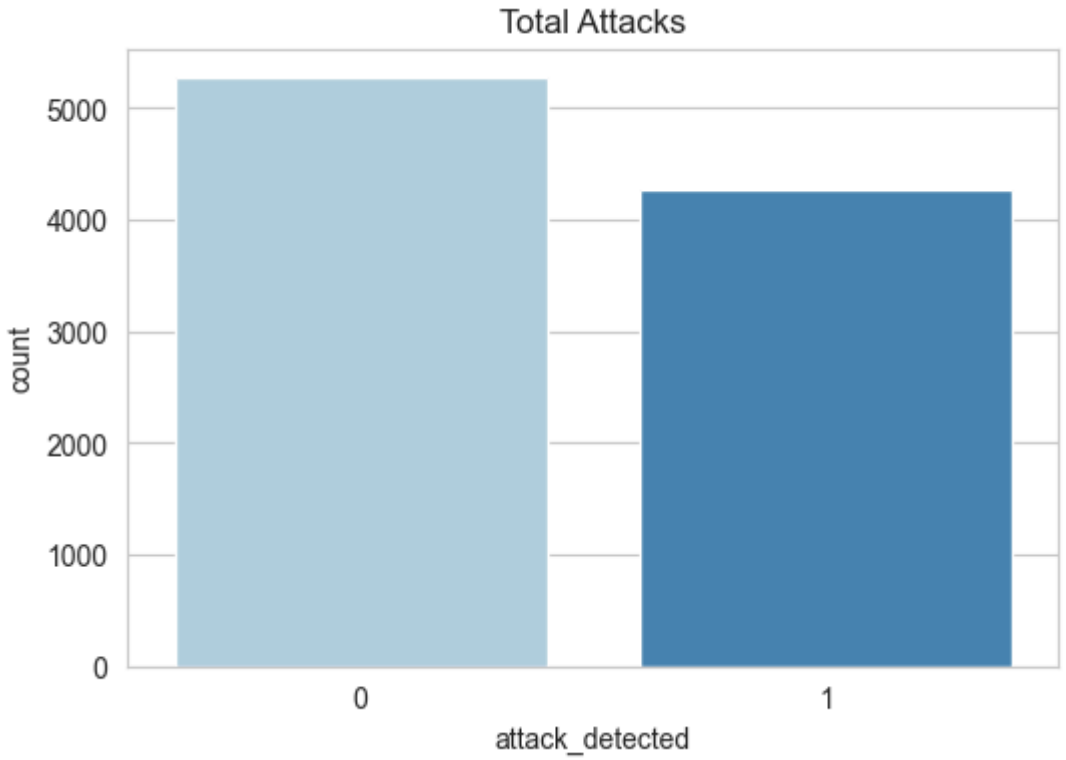
- **Comprender**: saber con qué datos estoy trabajando y para que lo estoy haciendo, lo cual se explica por medio de los gráficos y la descripción del dataset.

¿De qué maneras se aprovecha este dataset?: explica cada aspecto en detalle, incluyendo la estructura del conjunto de datos, la importancia de las características, los posibles enfoques de análisis y cómo se puede utilizar para el machine learning.

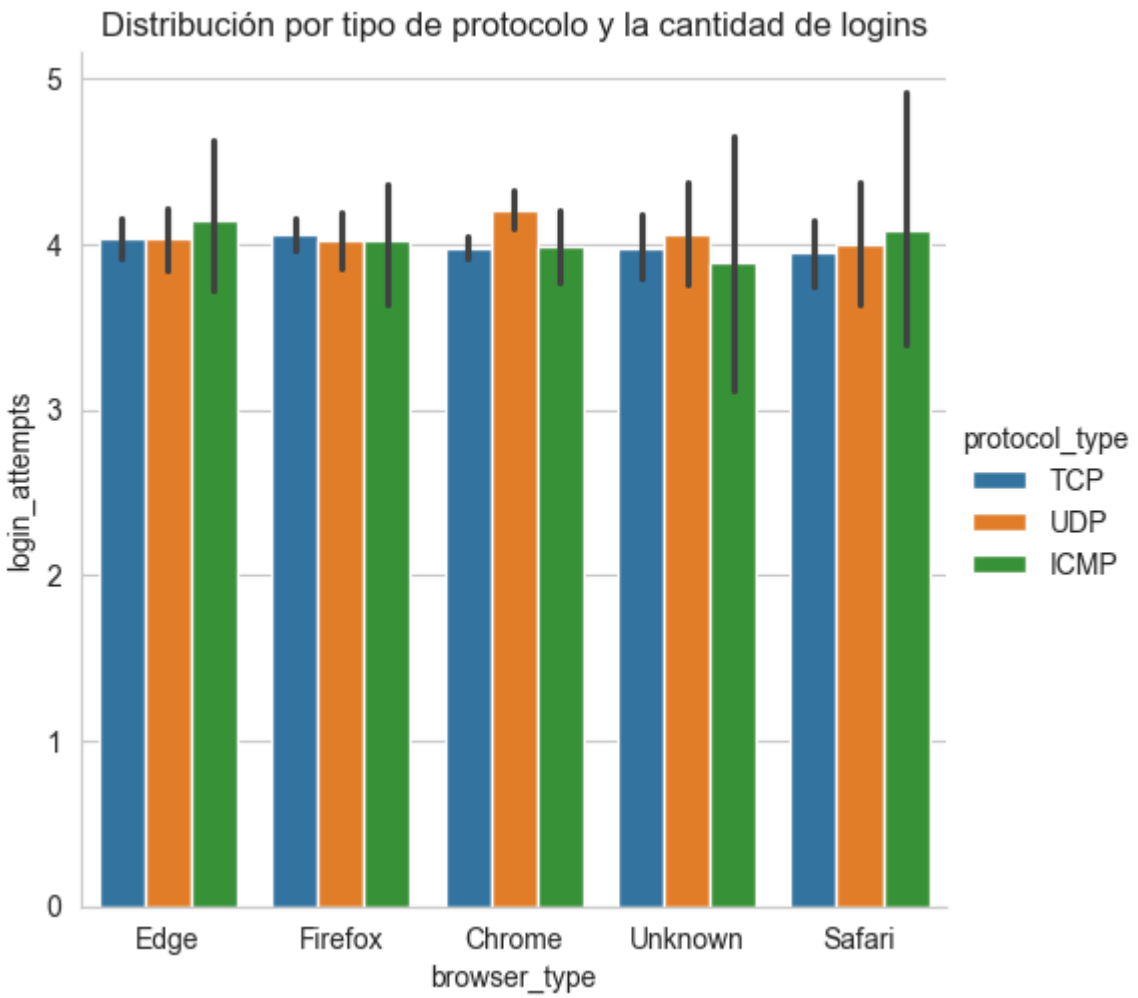
El dataset consta de características basadas en la red y basadas en el comportamiento del usuario. Cada característica proporciona información valiosa sobre posibles amenazas cibernéticas.

Ya que el dataset va bastante enfocado al machine learning, este gráfico ayuda a deducir y distinguir los distintos patrones de ataque, (0 es actividad

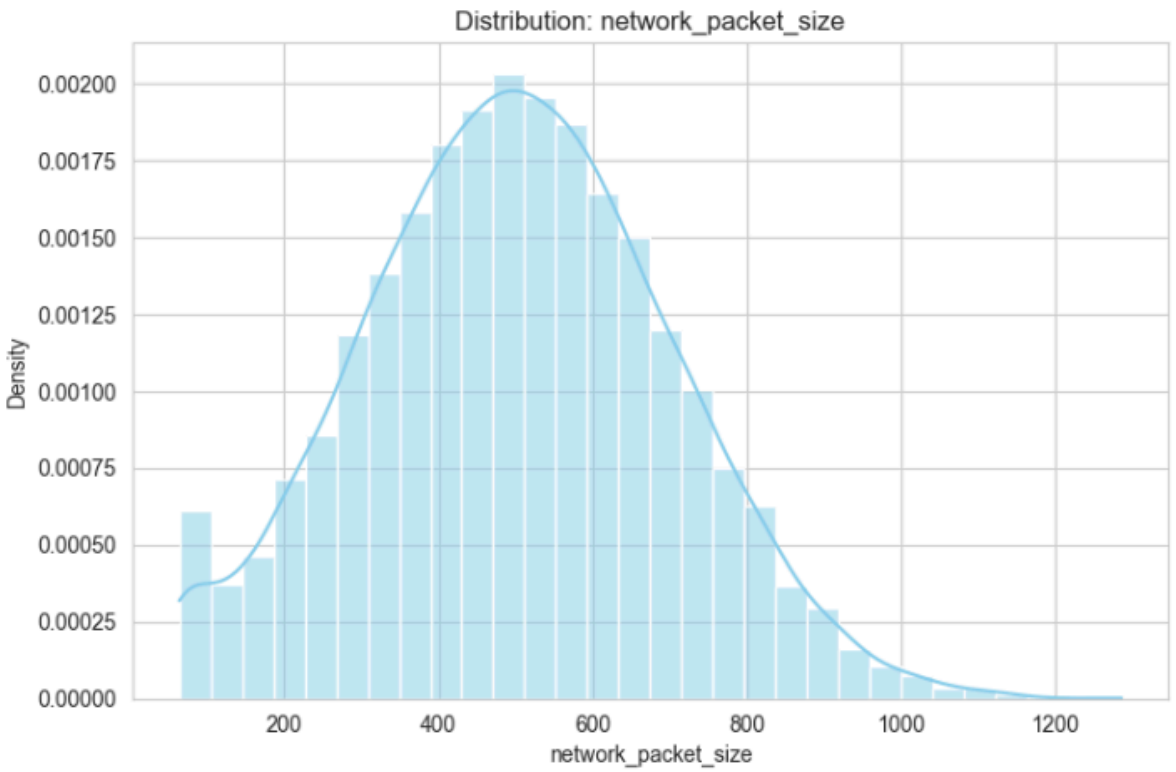
normal) (1 es que se ha detectado un ataque), esto ayudaría a un modelo a clasificar en datos binarios, cuando se presenta un ataque.



Cuantas veces intentaron abrirse paso y con qué protocolo ayuda a entender de qué manera los agentes maliciosos atacan y prediciendo cómo lo harán, el más (explotable) sería el ICMP el cual se vulnera con la denegación de servicios o mayormente conocido como (DoS).



Representación del tamaño de los paquetes en el rango de 64 a 1500 bytes un atacante anormalmente usa paquetes muy pequeños o muy grandes para reconocimiento o intentos de explotación, esto le ayudará al modelo a detectar y/o predecir cuando un usuario malicioso va a realizar un ataque.


Distribution: network_packet_size
Density
0.00200
0.00175
0.00150
0.00125
0.00100
0.00075
0.00050
0.00025
0.00000
200
400
600
800
1000
1200
network_packet_size