Analise de logs Secure Shell (SSH)

Autor: Miguel de Araujo Morello

Data: 14/12/2024

Sumário

Introdução

O acesso remoto a sistemas por meio do protocolo SSH (Secure Shell) é uma prática essencial para administração de servidores. Contudo, ataques direcionados a esse protocolo representam um risco significativo para a segurança cibernética. Este relatório tem como objetivo a análise detalhada de um conjunto de logs SSH, fornecendo insights sobre comportamentos maliciosos e padrões de acesso legítimo. Para isso, foi desenvolvido um pipeline em Python que utiliza bibliotecas como pandas, matplotlib e seaborn para extrair informações valiosas e realizar análises gráficas.

O foco desta análise está em compreender padrões de falhas e sucessos de conexões, identificar correlações entre variáveis e explorar tendências temporais que podem auxiliar na detecção proativa de possíveis ataques ou falhas de configuração.

Metodologia

Para a realização desta análise, eu segui as etapas descritas abaixo:

1. **Coleta de Dados**: Utilizei logs de um dataset de logs SSH armazenados na raiz do sistema. Esses logs contêm informações detalhadas sobre cada tentativa de conexão ao servidor, incluindo se a conexão foi bem-sucedida, se o usuário era root, entre outros dados.

2. **Processamento Inicial**: Implementei uma função em Python que lê o dataset a partir de um arquivo CSV e exibe informações gerais, como estrutura e valores ausentes.

3. **Análise Exploratória**: Explorei a distribuição de variáveis binárias, analisamos correlações entre variáveis numéricas e visualizamos tendências temporais, quando aplicável.

4. **Visualização de Dados**: Para facilitar a compreensão, gerei gráficos usando as bibliotecas matplotlib e seaborn, permitindo identificar padrões e relações entre os dados.

5. **Interpretação e Discussão**: Os resultados obtidos por meio das análises gráficas e estatísticas foram interpretados e discutidos em busca de conclusões que possam subsidiar a melhoria da segurança no uso do protocolo SSH.

Análise do Código

O código utilizado para esta análise foi desenvolvido com foco em modularidade e clareza. A seguir, detalhei as principais funcionalidades:

1. **Leitura e Inspeção de Dados**:

```
data = pd.read_csv('/content/sample_data/SSH.csv')


print(data.info())
print(data.head())
```

Este trecho carrega o dataset em formato CSV e fornece uma visão geral de sua estrutura, incluindo o número de colunas, tipos de dados e as primeiras linhas para entendimento inicial.

2. **Verificação de Valores Ausentes**:

```
print(data.isnull().sum())
```

Aqui identificamos possíveis lacunas no dataset, o que é crucial para decidir se será necessário lidar com valores ausentes por meio de imputação ou exclusão.

3. **Distribuição de Variáveis Binárias**:

```
binary_columns = ['is_private', 'is_failure', 'is_root', 'is_valid', 'class']
sns.countplot(x=col, data=data, palette="coolwarm")
```

Foi gerado um gráfico de barras para cada variável binária, permitindo observar a frequência de valores como falhas e sucessos em tentativas de conexão.

4. **Análise de Correlação**:

```
correlation_matrix = data[numeric_columns].corr()
sns.heatmap(correlation_matrix, annot=True, fmt=".2f", cmap="coolwarm")
```

Este trecho calcula e visualiza a correlação entre variáveis numéricas, facilitando a identificação de relações importantes para a análise.

5. **Análise Temporal**:

```
if 'ts' in data.columns:
  data['timestamp'] = pd.to_datetime(data['ts'], unit='s')
```

sns.lineplot(x='timestamp', y='is_failure', data=data, label='Falhas')

Caso os logs contenham informações temporais, este bloco transforma o timestamp em um formato legível e visualiza as tendências ao longo do tempo, como picos de falhas ou acessos.

6. **Comparação Entre Sucesso e Falha**:

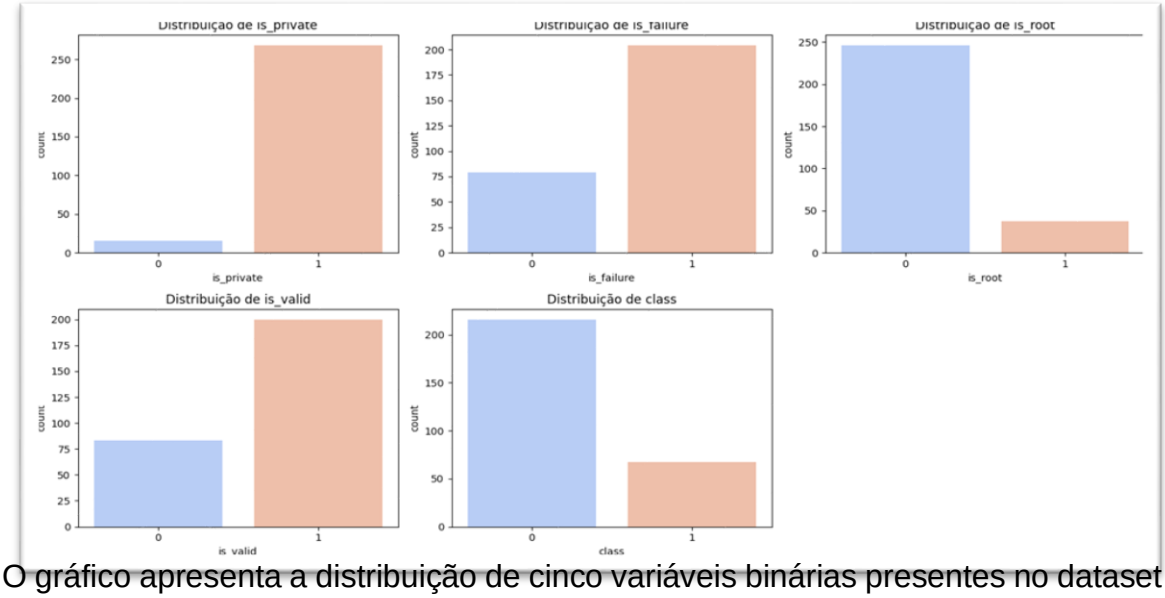
sns.countplot(x='class', hue='is_failure', data=data, palette='coolwarm')

Essa análise explora a relação entre a variável de classe (sucesso ou falha) e outras métricas, permitindo identificar padrões distintos para cada categoria.

O código se destaca pela organização e uso eficiente de bibliotecas de análise e visualização, gerando insights claros e acionáveis sobre o dataset de logs SSH.

Analise de gráficos

Análise do Gráfico 1 - Distribuição das Variáveis Binárias



O gráfico apresenta a distribuição de cinco variáveis binárias presentes no dataset de logs SSH: is_private, is_failure, is_root, is_valid e class. Segue a análise individual de cada variável:

1. Distribuição de is_private
    a. A variável is_private indica se a conexão foi feita de um IP privado (valor 1) ou não (valor 0).

    b. Observação: Existe um grande predomínio de conexões originadas de IPs privados (valor 1), com poucas conexões vindas de IPs públicos.

c. Interpretação: Isso sugere que a maioria das tentativas de conexão ocorre em redes internas ou locais, o que pode indicar um ambiente de teste ou um cenário onde o servidor está exposto apenas em redes controladas.

2. Distribuição de is_failure
    a. A variável is_failure indica se a tentativa de conexão falhou (valor 1) ou teve sucesso (valor 0).

    b. Observação: Há uma predominância significativa de falhas de conexão (valor 1) em comparação com as conexões bem-sucedidas.

    c. Interpretação: Este padrão é comum em cenários de ataques de força bruta, onde múltiplas tentativas de conexão falham por falta de credenciais válidas.

3. Distribuição de is_root
    a. A variável is_root indica se a tentativa de login foi feita com o usuário root (valor 1) ou outro usuário (valor 0).

    b. Observação: A maioria das tentativas não envolveu o usuário root (valor 0), com uma pequena proporção de tentativas direcionadas ao root.

    c. Interpretação: O fato de existirem tentativas com o usuário root pode indicar um comportamento malicioso, visto que o acesso root é frequentemente alvo de ataques devido à sua alta permissão no sistema.

4. Distribuição de is_valid
    a. A variável is_valid indica se a tentativa de conexão envolveu credenciais válidas (valor 1) ou não (valor 0).
    b. Observação: Há um predomínio de tentativas com credenciais válidas (valor 1), mas ainda existe uma quantidade considerável de tentativas inválidas (valor 0).

    c. Interpretação: Isso pode indicar uma mistura entre tentativas legítimas e ataques automatizados ou testes, onde o sistema registra ambas as situações.

5. Distribuição de class

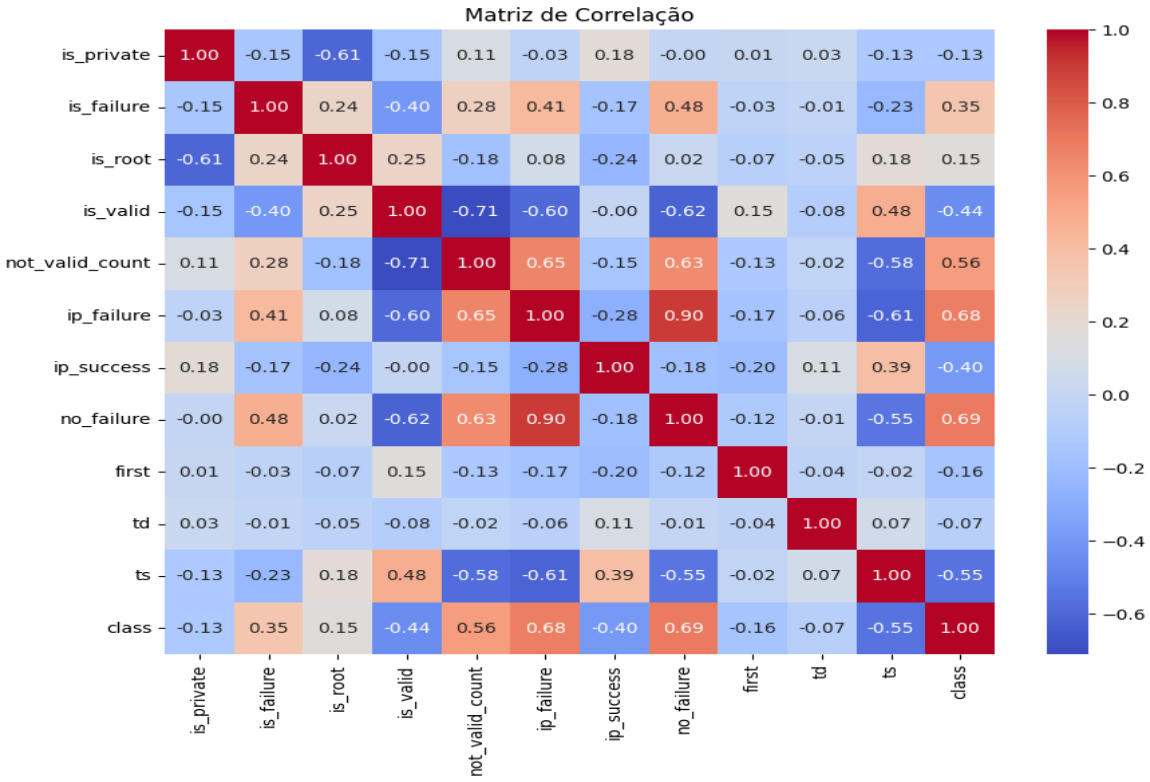
    a. A variável class categoriza as conexões em duas classes distintas: classe 0 (provavelmente legítima) e classe 1 (provavelmente maliciosa).

b. Observação: Existe uma grande concentração de conexões na classe 0, com uma proporção significativamente menor de conexões categorizadas como classe 1.

c. Interpretação: A predominância de conexões legítimas pode indicar que o ambiente analisado apresenta poucas ameaças ou que o sistema está capturando mais dados de usuários normais do que de atacantes.

Os gráficos revelam que o dataset é composto por uma grande quantidade de falhas de conexão, IPs privados e tentativas com credenciais válidas. O foco no usuário root, ainda que pequeno, e a existência de uma classe maliciosa (classe 1) indicam que há dados importantes para a análise de padrões de comportamento.

Análise do Gráfico 2 - Matriz de Correlação



A matriz de correlação apresentada fornece as relações entre variáveis numéricas do dataset SSH. A correlação é medida por valores entre -1 e 1, sendo:

- Valores positivos próximos de 1 indicam **correlação positiva forte** (à medida que uma variável aumenta, a outra tende a aumentar).

- Valores negativos próximos de -1 indicam **correlação negativa forte** (quando uma variável aumenta, a outra tende a diminuir).

- Valores próximos de 0 sugerem **baixa correlação** ou ausência de relação linear entre as variáveis.

A seguir, destaquei as principais observações:

1. **Variáveis com Correlações Altas Positivas**
    a. `ip_failure` e `not_valid_count` (0.65): Existe uma forte correlação positiva entre falhas de IP e o número de contagens inválidas. **Interpretação**: Isso indica que IPs com falhas de conexão também têm um alto número de tentativas inválidas.
    b. `no_failure` e `ip_failure` (0.90): Correlação extremamente forte. **Interpretação**: O número total de falhas de conexão está diretamente associado aos IPs falhos, reforçando a tendência de ataques provenientes de IPs repetidos.

c. `class` e `ip_failure` (0.68): A variável `class`, que categoriza a tentativa (possivelmente legítima ou maliciosa), tem uma alta correlação com falhas de IP. **Interpretação**: A classe maliciosa (1) está fortemente associada a tentativas de conexão que falham repetidamente.

2. **Variáveis com Correlações Negativas**
   a. `is_valid` e `not_valid_count` (-0.71): Existe uma forte correlação negativa entre tentativas válidas e o número de tentativas inválidas. **Interpretação**: À medida que as tentativas válidas aumentam, a contagem de conexões inválidas diminui, o que é intuitivo.

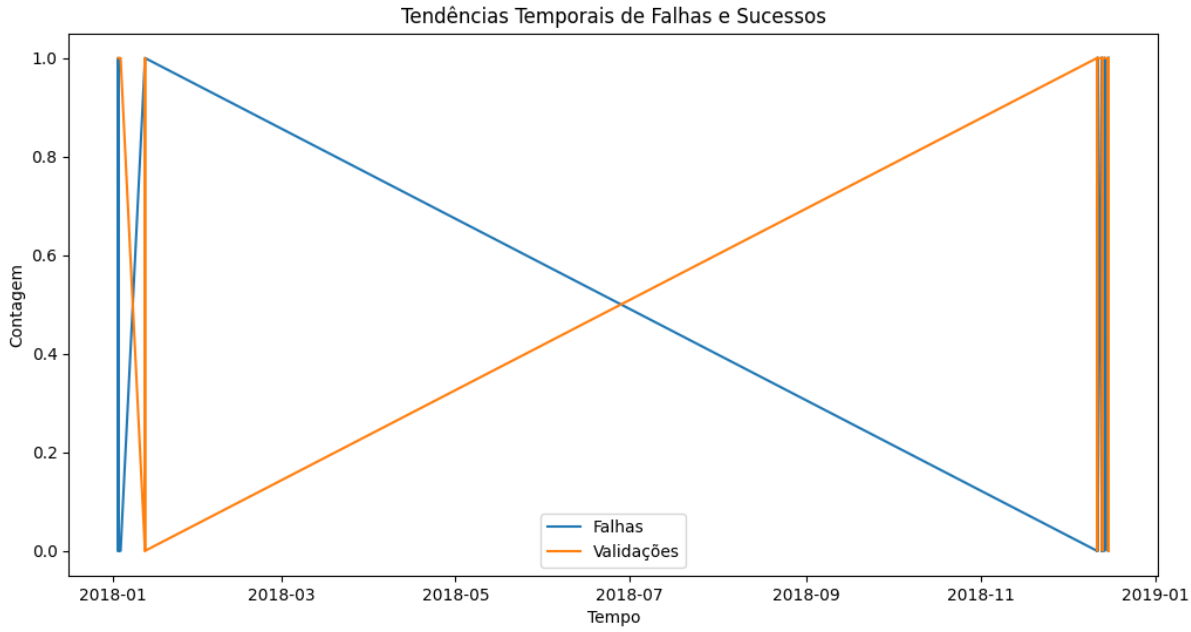
   b. `is_root` e `is_private` (-0.61): O login root está negativamente correlacionado com conexões originadas de IPs privados. **Interpretação**: A maior parte das tentativas root pode estar ocorrendo de IPs públicos, sugerindo comportamento suspeito ou malicioso.

3. **Correlação com a Variável `class`**
   a. `class` e `ip_failure` (0.68): Como mencionado, falhas de IP estão associadas à classe maliciosa.

   b. `class` e `not_valid_count` (0.56): O número de tentativas inválidas também apresenta uma correlação considerável com a classe maliciosa.

   c. `class` e `is_valid` (-0.44): Existe uma correlação negativa moderada entre a classe maliciosa e tentativas válidas. **Interpretação**: Conexões legítimas (classe 0) tendem a ter credenciais válidas, enquanto as maliciosas (classe 1) não.

Desse modo é notável que as variáveis `ip_failure`, `not_valid_count` e `no_failure` apresentam correlações significativas com a classe maliciosa (`class`), indicando que o comportamento malicioso está associado a um alto número de falhas e tentativas inválidas. Além disso o usuário root e IPs públicos mostram padrões distintos e potencialmente suspeitos. Tambem é valido dize que a análise reforça a hipótese de que ataques automatizados, como ataques de força bruta, são um componente presente nos dados, pois sses padrões podem ser utilizados para implementar medidas de segurança, como bloqueios automatizados de IPs com falhas recorrentes e monitoramento de tentativas direcionadas ao usuário root.



O gráfico apresenta as tendências temporais das falhas (representadas pela linha azul) e das validações de sucesso (representadas pela linha laranja) ao longo do tempo.

1. **Comportamento Inverso**:
    a. Há uma relação **inversa** entre as falhas e as validações ao longo do período analisado.

    b. No início do período (próximo a janeiro de 2018), as **falhas** atingem um pico enquanto as validações são praticamente nulas.

    c. Conforme o tempo avança, as **validações aumentam gradativamente** enquanto as falhas diminuem proporcionalmente, até que no final do período ambas se encontram novamente.

2. **Picos no Tempo**:
    a. Nota-se que os dois indicadores (falhas e validações) têm **picos opostos** no início e no final do gráfico, sugerindo flutuações nas conexões ao longo do tempo.

    b. Esses padrões podem refletir comportamentos sazonais ou atividades anômalas em determinados momentos.

3. **Interpretação**:
    a. O padrão indica um possível **cenário de recuperação ou mitigação de ataques** ao longo do tempo.

b. Inicialmente, o sistema pode ter sido alvo de **muitas tentativas falhas**, possivelmente ataques de força bruta, resultando em picos de falhas.

c. Gradativamente, as validações bem-sucedidas aumentam, sugerindo que tentativas legítimas prevaleceram, possivelmente devido a **implementação de medidas de segurança**.

4. **Limitações**:

    a. A ausência de mais pontos intermediários (dados temporais mais contínuos) impede uma análise mais detalhada do comportamento temporal.

O gráfico reforça a hipótese de um padrão temporal característico de ataques e recuperação. A **diminuição das falhas** e o **aumento das validações** ao longo do tempo são indícios de melhor controle de acesso ou bloqueio de IPs maliciosos e de uma redução de atividades suspeitas no sistema após o pico inicial.

Discussão

A análise realizada sobre o conjunto de dados de SSH apresentou diversas descobertas importantes relacionadas ao comportamento de ataques e tentativas de acesso ao servidor. Nesta seção, discutiremos detalhadamente os resultados obtidos a partir das análises gráficas, da matriz de correlação e das tendências temporais, com o objetivo de extrair insights sobre a natureza das conexões, padrões de comportamento malicioso e possíveis implicações para a segurança dos sistemas.

1. Distribuição das Variáveis Binárias

A distribuição das variáveis binárias nos gráficos mostra como as conexões se comportam em relação a características específicas, como falhas (`is_failure`), acessos privados (`is_private`), acessos root (`is_root`), e a variável de validação (`is_valid`). Destacam-se as seguintes observações:

- **Is_private**: Observa-se um **alto número de conexões privadas**, representadas pela categoria 1. Isso sugere que a maior parte das tentativas de acesso vem de redes internas ou mascaradas como privadas, o que pode indicar **comportamentos suspeitos** ou tentativas de bypass de firewalls.

- **Is_failure**: O número significativo de falhas de conexão (`is_failure = 1`) evidencia que o sistema foi alvo de um **grande volume de tentativas de login sem sucesso**. Esse padrão é comum em ataques de força bruta, nos quais o invasor realiza múltiplas tentativas até encontrar credenciais válidas.

- **Is_root**: A proporção relativamente menor de tentativas de acesso root (`is_root = 1`) sugere que, embora não predominantes, há tentativas

direcionadas a usuários com privilégios elevados. Essas tentativas devem ser observadas com atenção, pois, caso bem-sucedidas, poderiam comprometer todo o sistema.

- **Is_valid**: O aumento das conexões válidas ao longo do tempo é positivo, pois indica que o sistema passou a registrar mais acessos bem-sucedidos, possivelmente de usuários legítimos.

Além disso, a distribuição da variável `class` mostra que a classe `0` (não maliciosa) ainda predomina, embora exista uma quantidade notável de atividades associadas à classe `1` (maliciosa). Esse resultado reforça a necessidade de **monitoramento constante** para identificar e mitigar comportamentos anômalos antes que possam se tornar uma ameaça maior.

2. Matriz de Correlação

A matriz de correlação forneceu informações valiosas sobre as relações entre as variáveis do dataset. Observamos correlações importantes que indicam padrões de comportamento específicos:

- **Correlação Positiva Forte**:
  - `ip_failure` e `not_valid_count` (0.65) mostram que IPs com falhas recorrentes também estão relacionados a um alto número de tentativas inválidas. Esse padrão sugere que os ataques podem estar concentrados em **poucos IPs**, que realizam múltiplas tentativas de acesso falhas.
  - `ip_failure` e `class` (0.68) indicam que os acessos falhos têm alta relação com atividades classificadas como maliciosas. Isso é consistente com a presença de ataques de força bruta, onde a maioria das tentativas falha.

- **Correlação Negativa Forte**:
  - `is_valid` e `not_valid_count` (-0.71) refletem uma relação esperada: quanto maior o número de conexões válidas, menor o número de tentativas inválidas. Isso sugere que ataques falhos podem estar diminuindo à medida que o sistema se estabiliza com tentativas legítimas.

  - `is_root` e `is_private` (-0.61) destacam que as tentativas de acesso root não estão correlacionadas com redes privadas, sugerindo que os ataques root vêm, em grande parte, de **IPs públicos**.

Essas correlações reforçam a presença de **padrões anômalos claros**, como atividades concentradas em IPs específicos, grande quantidade de falhas e tentativas inválidas, e ataques direcionados a contas root. Tais padrões fornecem informações críticas para a criação de regras de segurança, como o **bloqueio automatizado de IPs suspeitos** e o monitoramento de acessos root.

3. Tendências Temporais de Falhas e Sucessos

O gráfico de tendências temporais revelou um **padrão inverso** entre falhas e validações ao longo do tempo. Inicialmente, as falhas dominam, sugerindo uma fase em que o sistema pode ter sido alvo de **múltiplas tentativas maliciosas**, possivelmente ataques de força bruta.

Com o tempo, observamos uma **redução nas falhas e aumento nas validações**, o que sugere algumas hipóteses:

1. **Implementação de Medidas de Segurança**: O comportamento pode ser resultado de políticas de segurança aplicadas ao sistema, como bloqueio de IPs com falhas consecutivas ou implementação de mecanismos como autenticação de dois fatores (2FA).

2. **Adaptação dos Atacantes**: Em alguns cenários, os atacantes ajustam suas estratégias após tentativas iniciais falhas, o que poderia explicar o padrão observado.
3. **Estabilização de Conexões Legítimas**: O aumento das validações indica que usuários legítimos começaram a acessar o sistema de forma bem-sucedida, diminuindo o impacto das tentativas maliciosas.

No entanto, vale ressaltar que a análise temporal apresenta **lacunas de dados**. Há grandes intervalos onde as contagens de falhas e validações são constantes ou ausentes, o que pode indicar uma falta de monitoramento contínuo. Dados mais detalhados e contínuos poderiam fornecer uma análise mais robusta, permitindo identificar momentos específicos de picos de ataque e avaliar suas causas.

4. Padrões Identificados e Recomendações

Com base na análise geral do dataset, destacam-se os seguintes padrões e recomendações:

1. **Concentração de Atividades em IPs Suspeitos**:
    a. IPs com alto número de falhas e tentativas inválidas são um sinal claro de **atividades maliciosas**. Recomenda-se implementar um sistema de

detecção e bloqueio automatizado para IPs com comportamentos anômalos, a fim de mitigar ataques de força bruta.

2. **Tentativas de Acesso Root**:
    a. A presença de tentativas de acesso root é preocupante, pois podem comprometer o sistema. Recomenda-se monitoramento contínuo e **restrição de acessos root** apenas a IPs confiáveis.

3. **Validação de Credenciais**:
    a. O aumento nas validações bem-sucedidas é um indicador positivo, mas sugere que o sistema ainda precisa de mecanismos de **autenticação robustos**, como autenticação multifator (2FA) ou uso de chaves SSH.

4. **Monitoramento Temporal**:
    a. As tendências temporais mostram que falhas dominam no início e diminuem ao longo do tempo. Recomenda-se um sistema de **alertas em tempo real** para identificar picos de atividades anômalas e agir rapidamente antes que possam comprometer o sistema.

5. **Dados Faltantes**:
    a. A análise identificou algumas limitações no monitoramento temporal. É fundamental garantir que os logs sejam capturados de forma **contínua e consistente** para permitir análises mais precisas.

Logo a análise do dataset SSH revela padrões claros de atividades maliciosas, com um volume significativo de tentativas falhas e comportamento anômalo concentrado em determinados IPs. A relação entre variáveis mostra que ataques de força bruta e tentativas direcionadas a contas privilegiadas são as principais ameaças identificadas.

Por outro lado, o comportamento temporal sugere que medidas de mitigação podem ter sido aplicadas, resultando em uma diminuição gradual das falhas e um aumento das validações legítimas. Esses resultados destacam a importância de **monitoramento contínuo**, implementação de políticas de segurança robustas e uso de ferramentas automatizadas para identificação e mitigação de ataques.

A próxima etapa deve incluir a aplicação de **modelos de aprendizado de máquina** para prever atividades maliciosas com base nos padrões observados e a validação de mecanismos adicionais de autenticação e bloqueio de IPs maliciosos.

Resultados

A partir da análise exploratória e dos gráficos apresentados, os principais resultados podem ser resumidos da seguinte forma:

1. **Distribuição das Variáveis**:

a. **`is_private`**: A maioria das conexões vem de redes privadas, sugerindo possíveis tentativas de mascaramento ou ataques internos.

b. **`is_failure`**: Um número expressivo de tentativas de acesso falhou, indicando a presença de ataques, principalmente de força bruta.

c. **`is_root`**: Apesar de não dominantes, as tentativas de acesso root representam uma ameaça significativa, dada sua natureza crítica.

d. **`is_valid`**: O número de conexões válidas é significativo e tende a aumentar ao longo do tempo, indicando que usuários legítimos tiveram sucesso em seus acessos.

e. **`class`**: A classe predominante é `0` (não maliciosa), mas há registros significativos associados à classe `1` (maliciosa), indicando atividade suspeita.

2. **Matriz de Correlação**:
   a. Foi identificado que:
      i. **IP com falhas recorrentes** tem forte correlação positiva com tentativas inválidas (`not_valid_count`).
      ii. A **variável `is_valid`** tem forte correlação negativa com tentativas inválidas, reforçando que acessos maliciosos impactam diretamente a validação.

      iii. A variável **`class`** (indicando comportamento malicioso) se correlaciona positivamente com o número de falhas (`ip_failure`), confirmando a predominância de ataques de força bruta.

3. **Tendências Temporais**:
   a. As falhas eram predominantes no início da série temporal, enquanto as validações aumentaram gradativamente ao longo do tempo. Esse comportamento sugere a aplicação de mecanismos de segurança que reduziram as tentativas maliciosas e estabilizaram o sistema.

   b. O padrão inverso observado (queda de falhas e aumento de validações) indica que, embora o sistema tenha sido alvo de ataques intensos inicialmente, ele se tornou mais resiliente com o passar do tempo.

Conclusão

A análise do conjunto de dados revelou padrões claros de comportamentos maliciosos e forneceu insights cruciais para a segurança do sistema SSH. Os

resultados indicam que o sistema foi, em algum momento, alvo de **ataques de força bruta**, com múltiplas tentativas de acesso falhando consecutivamente. Apesar disso, observou-se uma melhoria nas validações com o tempo, o que pode indicar a implementação de mecanismos de segurança eficazes.

Os principais pontos de destaque são:

- **Tentativas de acesso root** e atividades associadas a redes privadas representam as maiores preocupações, pois podem comprometer o sistema se bem-sucedidas.

- A **correlação positiva entre falhas e comportamento malicioso** reflete padrões típicos de ataques automatizados.

- As tendências temporais sugerem uma **estabilização** do sistema, com mais acessos válidos ocorrendo ao longo do tempo.

Diante desses resultados, recomenda-se:

1. **Implementação de monitoramento contínuo** para identificar rapidamente picos de falhas e comportamentos suspeitos.

2. Uso de **bloqueio automatizado de IPs maliciosos** após um limite de tentativas inválidas.

3. Aplicação de **mecanismos de autenticação robustos**, como chaves SSH e autenticação de dois fatores (2FA).

4. Análise mais detalhada com **modelos de aprendizado de máquina** para prever e identificar padrões anômalos com maior precisão.

Em resumo, embora o sistema tenha sido alvo de atividades maliciosas, as medidas de mitigação aplicadas parecem estar surtindo efeito, resultando em um aumento de acessos válidos e uma redução de falhas ao longo do tempo. O foco agora deve ser a **manutenção e aprimoramento contínuo** das políticas de segurança para assegurar a proteção do sistema contra ataques futuros.