1ª Análise de Logs

# Introdução

Para a primeira análise de logs extraídos do site [www.fluzao.xyz](http://www.fluzao.xyz) buscamos os logs de acesso diário, cidades de onde partiram os acessos, páginas de destino e títulos das páginas visitadas.

Para aproveitar as informações do log da Página de Destino foi feito um pré-tratamento dos dados dentro de uma planilha Excel, para isso foi utilizada a aba Menu Código da planilha Mapa do Site.xlsx já existente que lista todas as consultas e qual o caminho dentro do sistema que ela percorre, desta forma é possível isolar o caminho e o seu código e confrontar com a planilha referência criada, gerando uma planilha atualizada com as visualizações por página de acesso associada a consulta do site.

Nesta associação descrita acima alguns caminhos intermediários foram ignorados por não trazer nenhuma grande contribuição para a análise.

Foram utilizadas as bibliotecas Pandas, Numpy, Seaborn, Matplotlib e Plotly, para a visualização dos dados a biblioteca plotly se mostrou mais adequada, porém para essa primeira análise alguns gráficos foram gerados nas duas bibliotecas para comparação e a partir da próxima continuamos somente com o que for escolhido.

As análises foram realizadas em três notebooks e ainda precisam de algum ajuste nos comentários e agrupamento dos códigos que serão realizados posteriormente.

# Pré-processamento dos dados

Depois de criar a planilha de telas acessadas, separando os códigos da parte textual da url, o arquivo telas\_acessadas.csv é carregado e confrontado com o arquivo referencia.csv, a intenção é combinar o código da página, identificador único no sistema, contra os códigos do arquivo referencia.csv fazendo a ligação dos acessos com as consultas e assim sumarizando e contando a quantidade de visualizações de cada consulta existente no site, a saída deste processo é o arquivo referencia\_atualizada.csv que será utilizado na análise de páginas acessadas.

Alguns ajustes foram realizados nas páginas que não continham acesso via código, um exemplo são as páginas chamadas via link de outras páginas, nestas páginas o código corresponde a ids de banco de dados, então foi criado um grupo para os acessos via LINK.

Para fins de análise foram criados Grupos e Subgrupos para as consultas existentes no site.

Grupos criados são: **Inicial, Especial, Especiais, Adversário, Árbitro, Estádio, Campeonato, Jogador, Técnico, Jogo, Geral, Curiosidade, Contato, Créditos, Camisa, Geral e Link**

Subgrupos criados são: **Inicial, Especial, Adversário, Arbitro, Campeonato, Jogador, Técnico, Jogo, Geral, Curiosidades, Contato, Créditos, Camisa, Sequência, Frequência, Gráfico, Artilheiros, Assistência, Presença, Gols, Publico, Especiais, Frequência, Resultado, Sequência, Presidente, Brasileiro, Tecadv, Numeração, Estádio**

Algumas páginas de passagem foram excluídas da análise por não trazer nenhum benefício para o estudo.

# Análise por Título

Cada página do site tem o seu título informando a consulta que foi feita, neste título sempre temos os nomes das partes envolvidas ou um texto que indica a consulta realizada, por exemplo ao consultar o confronte entre Fluminense e Palmeiras o título fica Fluminense x Palmeiras, desta forma facilita a pesquisa e identificação de acessos com a mesma natureza.

## Pesquisa por palavra-chave

A primeira análise realizada nos logs consiste em identificar palavras-chave, um exemplo é o nome do adversário, portanto escolhemos um ou mais adversários e criamos uma lista, com essa lista filtramos no arquivo do relatório de títulos acessados por data por todas as linhas que tenham a palavra-chave, depois é gerado um gráfico de linha com os acessos diários de cada palavra-chave.

No gráfico abaixo foram escolhidos os últimos adversários do Fluminense do dia 02/08 até 22/08(Palmeiras, Argentino Juniors, Grêmio e América-MG), além de um próximo adversário (Olimpia) e um possível próximo adversário (Bolívar).

Gráfico, Gráfico de linhas

Descrição gerada automaticamente

A conclusão desta análise é que nos dias próximo ao jogo contra adversário encontramos um pico de visualizações das informações sobre o confronto, entre os dias 02 e 06 o Palmeiras foi o mais acessado, o jogo ocorreu no dia 05/08, já entre os dias 06 e 08/08 o Argentino Juniors passou a ser o mais acessado, a única “anomalia” da sequência foi o pico de acessos ao Olimpia, no dia 10/08 após vencer a sua eliminatória ele passou a ser o próximo adversário na principal competição do ano, então ele foi dominante nos dias 10 e 11/08, logo depois o dominante passou a ser o Grêmio com o jogo em 13/08, sendo dominante nos dias 12 e 13/08, com a semana inteira sem jogos tivemos 4 dias sem um grande pico e com o jogo contra o América-MG no dia 19/08, observamos um pico de acesso a partir do 18/08 até o dia 20/08.

Para o usuário final é importante ter disponível os dados do próximo adversário e em algum momento o futuro adversário de competições importantes pode ser bem importante, ainda não foi verificada a força dos adversários mais tradicionais do estado, essa análise será realizada e os resultados analisados.

O importante desse código gerado é que ele se adequa a qualquer parâmetro que queira ser estudo de qualquer natureza e de qualquer log de site que esteja com os títulos bem estruturados.

## Páginas mais acessadas

Esta análise traz quais as páginas são mais acessadas, o retorno não foi muito representativo, uma vez que a página inicial e algumas páginas de passagem ficam com muito mais acessos.

Gráfico, Gráfico de cascata

Descrição gerada automaticamente

## Confrontos mais acessados

Esta análise visa identificar quais são os confrontos mais acessados, difere da palavra-chave porque neste caso foram filtradas somente as páginas com título formado por “Fluminense x “.

Gráfico, Gráfico de barras

Descrição gerada automaticamente

A conclusão desta análise é que os confrontos atuais acabam sendo os mais visitados, outro ponto importante é notar a força dos clássicos regionais, mesmo que no período não tenha ocorrido jogo contra eles os três figuram entre os mais acessados, em uma proporção um pouco menor mas com a mesma lógica temos os clássicos regionais, Internacional, São Paulo, Corinthians e Santos, são importante adversários porém sem jogos no período, os demais confrontos necessitam de mais tempo de estudo para verificar se existe algum tipo de padrão ou se é só uma curiosidade, neste caso estariam os confrontos contra Guarani-Paraguai, San Lorenzo-Argentina e PSG-França.

## Análise de técnicos e ano

A análise de técnicos e uma outra análise com uma filtro em uma lista de anos (2023, 2022, 2021), jogadores, campeonatos e outras consultas tem o ano no título para fazer a diferença das buscas, indica que as informações atuais, de 2023, são as mais acessadas, um exemplo é que o técnico atual tem 8 vezes mais acessos que o segundo colocado.

A busca por informações atuais também segue um padrão, os mais acessados são os mais em evidência e por consequência os mais citados pela mídia, neste sentido a página mais acessada com o ano de 2023 é a do German Cano que é o artilheiro do Brasil e que vem sendo sempre citado em toda e qualquer transmissão, da mesma forma no mês de agosto o jogador John Kennedy assumiu um protagonismo nos jogos e muito festejado pela mídia, ele também foi o segundo mais acessado dentro do período pesquisado.

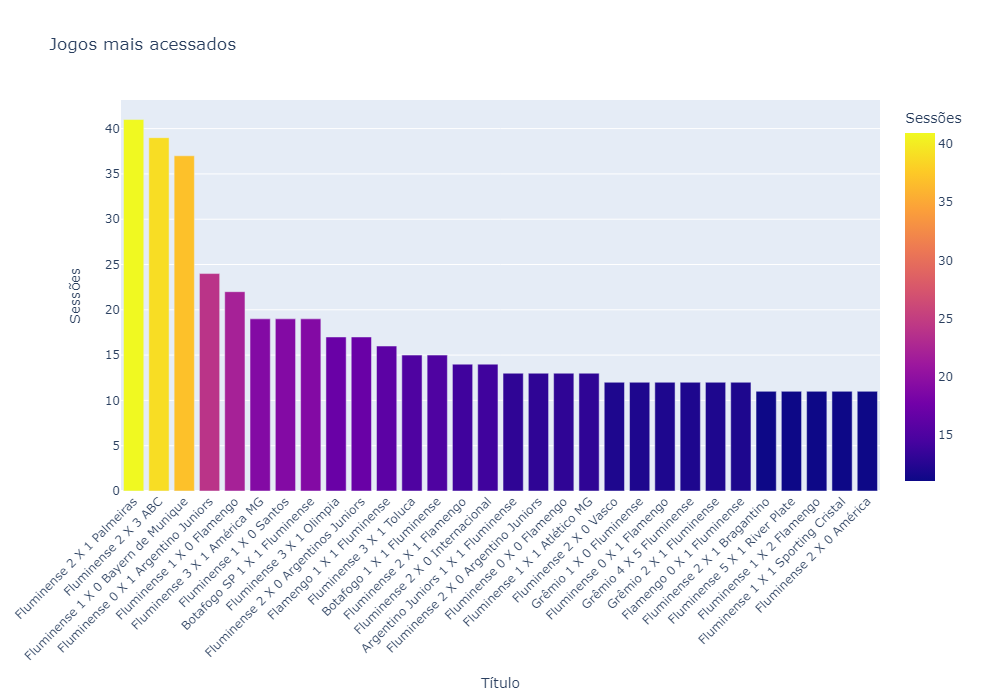
Gráfico, Gráfico de barras, Histograma

Descrição gerada automaticamenteGráfico, Gráfico de cascata

Descrição gerada automaticamente

## Jogos mais acessados via título da Página

Foi realizada uma análise para identificar os jogos mais acessados pelo título da página, foi identificado um problema porque o título dos jogos não contém a data do jogo e por isso jogos com o mesmo placar em datas diferentes são contabilizados como o mesmo jogo, em um primeiro momento de análise não traz muito desvio, mas com mais tempo de análise isso tende a gerar insights errados, utilizando a abordagem de análise por páginas acessadas conseguimos identificar o jogo com correção.



# Análise por página

Outra abordagem escolhida foi a abordagem por página, nesta abordagem é possível fazer a divisão por grupos e subgrupos além de trazer uma análise mais correta sobre os jogos mais acessados.

## Análise de grupos mais acessados

Uma imagem contendo Interface gráfica do usuário

Descrição gerada automaticamente

Conclusão:

Os usuários costumam utilizar os links presentes na página, pela análise de subgrupos chegamos à conclusão de que se todos os links do site ficarem facilmente identificáveis podemos facilitar a vida do usuário e aumentar o tempo de engajamento dentro do site, a maior parte dos acessos via link é feito nos jogos enquanto outros links sequer são acessados.

## Consultas por acesso

Existem muitas consultas que ainda não foram acessadas, é necessário realizar uma pesquisa mais profunda para entender o motivo pela qual as consultas não são acessadas ou são pouco acessadas, as duas principais razões seriam desconhecimento da consulta, pouco entendimento do que a consulta tem a oferecer e acesso prévio e não entendimento do resultado da consulta.

Gráfico

Descrição gerada automaticamente

## Consultas por grupo/subgrupo

Uma imagem contendo Gráfico

Descrição gerada automaticamente

Conclusão:

Corroborando com a análise de grupo grande parte dos maiores acessos vem dos mais diferentes links, desta forma é importante investir na melhoria dos links, criação de novos links.

A procura pela seção de gráficos também ficou bem abaixo do esperado e se for dado mais destaque aos gráficos tornando-os mais acessíveis e fáceis de manusear vai trazer uma melhor experiência aumentando o número de acessos.

## Grupos/Subgrupos Zerados de acesso

Quais os grupos/subgrupos com consultas que não tiveram acesso.

Gráfico, Gráfico de barras

Descrição gerada automaticamenteUma imagem contendo Aplicativo

Descrição gerada automaticamente

Conclusão:

A seção de Especiais-Jogador é a que apresenta mais consultas e mais consultas sem acesso, a distribuição das opções de consulta deve ser customizada para permitir uma melhor experiência e trazer facilidade na busca pelas informações, outro ponto importante é com relação aos gráficos e sua baixa procura de acessos.

## Jogos mais acessados

Conforme explicado no item de jogos mais acessados usando a página de destino traz uma estatística mais correta, porém traz alguns enigmas que não consigo responder por enquanto, o jogo mais acessado do período é um jogo de 1998 com acessos distribuídos por vários dias, o segundo maior podemos atribuir a curiosidade porque foi um jogo contra um gigante europeu realizado em 1976, depois deles os jogos mais acessados acabam seguindo a regra da proximidade com os confrontos.

Gráfico, Histograma

Descrição gerada automaticamente

Gráfico, Gráfico de barras

Descrição gerada automaticamente

# Acessos

Os acessos se mantêm bem constantes em uma média de 100 acessos diários, a maioria dos acessos são do Brasil mais precisamente do Rio de Janeiro, porém temos acessos de vários outros países e cidades bem incomuns.

Gráfico, Gráfico de linhas

Descrição gerada automaticamente

Gráfico, Gráfico de cascata

Descrição gerada automaticamente