

# Relatório Técnico

INF2102 – Projeto Final de Programação

Professora: Clarisse Sieckenius de Souza

Aluno: Rodrigo Brito de Freitas Lima

Orientadora: Simone Diniz Junqueira Barbosa

## Sumário

1.	Introdução.....	4
2.	Descrição e objetivos gerais do Software .....	5
2.1	Análise de Logs .....	5
2.2	Questionários.....	5
2.3	Entrevistas.....	6
2.4	Sugestões .....	6
3.	Objeto de estudo .....	7
4.	Requisitos.....	8
4.1	Requisitos Funcionais .....	8
4.2	Requisitos Não Funcionais.....	8
5.	Modelo de dados e outras dimensões relevantes.....	9
5.1	Modelo de Dados.....	9
5.1.1	acesso_cidades.csv.....	9
5.1.2	acesso_estado.csv .....	9
5.1.3	acesso_paginas.csv.....	10
5.1.4	acessos.csv .....	10
5.1.5	referencia.csv .....	10
5.1.6	telas_acessadas.csv .....	10
5.1.7	Titulos_data.csv.....	11
5.1.8	Titulos_data_hora.csv .....	11
5.1.9	jogo.csv.....	11
5.1.10	Engajamento.csv.....	11
5.1.11	Como_acessa_data.csv.....	12
5.1.12	Onde_acessa_data.csv .....	12
5.1.13	Referencia_atualizada.csv .....	12
5.1.14	Arquivo de retorno do questionário.....	12
5.2	Outras dimensões relevantes (Dados Técnicos e Melhorias).....	12
5.2.1	Dados Técnicos.....	12
5.2.2	Melhorias .....	13
6.	Descrição Funcional do Sistema .....	13
6.1	Coleta de Dados .....	14
6.2	Processamento dos dados.....	15
6.2.1	Pré-processamento.....	16
6.2.2	Processamento .....	16
6.2.3	Sugestões de Melhoria .....	17

<b>7. Manual do usuário.....</b>	<b>17</b>
<b>7.1 Premissas .....</b>	<b>17</b>
<b>7.2 Google Analytics .....</b>	<b>18</b>
<b>7.3 Processamento .....</b>	<b>27</b>
<b>7.3.1 Gera_referencia.ipynb .....</b>	<b>28</b>
<b>7.3.2 Analise_Titulos.ipynb .....</b>	<b>29</b>
<b>7.3.3 Analise_Titulos-Hora.ipynb .....</b>	<b>30</b>
<b>7.3.4 Analise_Paginas.ipynb .....</b>	<b>30</b>
<b>7.3.5 Analise de Acessos.ipynb .....</b>	<b>31</b>
<b>7.3.6 Analise_consulta_zerada.ipynb.....</b>	<b>31</b>
<b>7.3.7 Analise_tecnologia.ipynb .....</b>	<b>31</b>
<b>7.3.8 Identificando_links.ipynb .....</b>	<b>32</b>
<b>7.3.9 Analise_questionario.ipynb .....</b>	<b>32</b>
<b>8. Cenários de Sucesso .....</b>	<b>33</b>
<b>8.1 Primeiro cenário .....</b>	<b>33</b>
<b>8.2 Segundo cenário .....</b>	<b>34</b>
<b>9. Cenários de Falha.....</b>	<b>35</b>
<b>9.1 Primeiro cenário .....</b>	<b>35</b>
<b>9.2 Segundo cenário .....</b>	<b>36</b>

# 1. Introdução

Sistemas legados ainda têm muita importância para corporações e usuários. Processos de modernização têm sido bastante explorados e pesquisas são realizadas, grande parte das quais com o foco voltado para a reescrita do código com a troca da tecnologia legada para uma tecnologia mais moderna.

Esses projetos são sempre cercados de muita expectativa; porém, muitas vezes temos cenários caóticos de desenvolvimento. Vários fatores contribuem para esses problemas, incluindo a falta de conhecimento dos novos desenvolvedores nas linguagens legadas e no negócio. A maioria das ferramentas criadas para extração de regras de negócio não consegue recuperar todas. Isso traz muita frustração para o usuário, que espera que o novo e moderno sistema resolva seus problemas e, na maioria das vezes, recebe um produto com visual moderno mas que apresenta vários problemas de funcionalidade.

Nesse contexto, propomos uma metodologia que se baseia em análise de logs, formulação de questionários e realização de entrevistas a fim de buscar um entendimento sobre o uso do sistema, preferência dos usuários, dificuldades de entendimento e dificuldades de navegação, visando a gerar sugestões de redesign do sistema legado e tornar as interfaces estáticas em interfaces inteligentes, sem alteração das regras de negócio implementadas.

## 2. Descrição e objetivos gerais do software

Nosso objetivo maior é buscar ideias de redesign para os sistemas legados, transformando os sistemas atuais em sistemas inteligentes e mantendo as suas regras de negócio. Para isso, sugerimos três etapas de análise: a análise de logs, a criação de questionários online para os usuários e a realização de entrevistas com usuários-chave do sistema legado.

As etapas sugeridas estão encadeadas e os resultados de cada etapa norteiam a etapa seguinte.

### 2.1 Análise de Logs

A análise de logs é a primeira etapa da metodologia. Ela deve ser realizada durante todo o ciclo de vida do projeto de modernização do sistema legado. Seus resultados serão utilizados para a criação dos questionários, para o roteiro da entrevista e serão a peça central para a definição de recomendações de melhoria.

Para a realização desta análise, precisamos de um programa coletor de dados da utilização do sistema. Existem algumas opções no mercado e pode ser criada uma solução caseira. No estudo de caso conduzido, utilizamos o Google Analytics.<sup>1</sup>

A coleta de dados começou no dia 03/08/2023 e, tendo em vista o volume de acessos ao site em questão, foi necessário ao menos um mês de coleta de dados para iniciarmos uma análise consistente que nos possibilitasse gerar os primeiros insumos para criarmos os questionários, criarmos o roteiro das entrevistas e esboçarmos as primeiras recomendações para o redesign.

É necessário definir uma frequência para a recuperação dos logs. Essa frequência deve ser escolhida principalmente tendo como base o volume de uso do sistema. No estudo de caso realizado, definimos uma frequência de recuperação dos logs de uma semana. Nessa definição, é importante verificar se a ferramenta de coleta de dados tem alguma limitação na extração dos dados. Em nosso caso, o Google Analytics tem uma limitação de 5000 linhas para exportação dos arquivos, e isso foi levado em consideração no momento da definição da frequência de coleta.

### 2.2 Questionários

Após o prazo inicial de análise de logs, obtivemos dados suficientes para a criação de questionários a serem aplicados aos usuários de forma mais genérica. A intenção

---

<sup>1</sup> <https://analytics.google.com/>

com os questionários é identificar preferências no uso do sistema, identificar dificuldades que possam levar à não utilização de determinadas áreas do sistema, identificar dificuldades que afetem diretamente a experiência do usuário e medir a aceitação sobre ideias de redesign já identificadas na análise de logs.

O tempo e o formato de aplicação do questionário dependem da natureza do sistema, dos usuários envolvidos e do prazo de entrega do projeto. Para o objeto deste estudo, o questionário ficou disponível por um mês e foram coletadas 42 respostas.

Informações que caracterizam o perfil dos usuários devem ser obtidas no questionário; elas serão úteis no momento de analisar as respostas. Também vale perguntar no questionário, sempre que necessário, sobre a disponibilidade de participação de entrevista. Dado que alguns sistemas legados podem ter usuários-chave conhecidos, essa pergunta pode ser desnecessária; já para o caso de sites abertos ao público, essa informação se torna muito útil para o posterior recrutamento para as entrevistas.

## 2.3 Entrevistas

Considerando as análises de log e as informações de uso obtidas com os questionários, definimos um roteiro de entrevistas. No caso de redesign de sistema legado, as entrevistas devem ser realizadas com usuários-chave do sistema, pessoas que vão contribuir com informações relevantes. O roteiro deve focar principalmente nas “áreas de sombra” do sistema, ou seja, áreas do sistema com poucos acessos. Buscamos identificar os problemas encontrados no acesso a essas áreas, por que usuários evitam acessá-las, as maiores dificuldades no acesso ao sistema, as maiores oportunidades de melhoria e novas ideias, sempre visando melhorar a experiência do usuário no acesso das informações contidas no sistema.

## 2.4 Sugestões

Com todas as etapas finalizadas e com as informações das diferentes fontes de dados compiladas, o desenvolvedor/pesquisador começa a criar cenários de redesign que atinjam o objetivo de transformar a experiência do usuário saindo de um sistema antigo e estático para um sistema moderno e inteligente, utilizando novas ferramentas e técnicas de design, sistemas de recomendação, sistemas adaptados tanto a desktops como a dispositivos móveis, como *smartphones* e *tablets*.

### 3. Objeto de estudo

O sistema desenvolvido neste trabalho se destina à análise de dados coletados com o Google Analytics dos registros (*logs*) de uso de um sistema legado, com vistas ao seu redesign. Trata-se de um conjunto de programas desenvolvidos na forma de jupyter notebooks.

Para realizar a pesquisa, escolhemos como sistema legado o site de estatísticas de futebol do Fluminense Football Club.<sup>2</sup> O site foi desenvolvido em Python/Flask, Javascript, HTML e Bootstrap. O site é público e com acesso aberto, sem necessidade de login. Isso prejudica a análise de preferências individuais. O design é estático, sendo necessários múltiplos cliques para a consulta de uma única estatística. Algumas consultas contam com um grande número de opções de busca, e essas opções ficam dispostas dentro de um mesmo objeto (elemento do tipo *combobox*); e outras consultas podem ficar escondidas em áreas do site que, para o usuário final, podem não ter nenhuma relação com o seu objetivo. Algumas áreas do site são pouco acessadas e fornecem visualizações e informações muito interessantes; e outras trazem a informação de forma complicada e pouco didática, o que dificulta o entendimento e afasta os usuários. Existem links de acesso rápido em diversos pontos, porém nem sempre são bem sinalizados.

O site já passou por duas grandes reformulações. Inicialmente em 1995 era uma aplicação Cobol com arquivos VSAM como a base de dados. Em 2007, foi iniciado um trabalho para trazer estas informações para o mundo da Internet e o site foi lançado em 2008, desenvolvido em PHP com HTML e validações em Javascript.

Na segunda reformulação, ocorrida em 2020, o sistema passou para o Python/Flask, todas as consultas foram migradas e, com a facilidade de trabalhar com dataframes da linguagem Python, novas funcionalidades foram incluídas. A interface foi extensamente reformulada e o site agora é responsivo, embora testes mais abrangentes em diferentes tamanhos de tela não foram conduzidos.

Com a ajuda da metodologia de análise de logs para sistemas legados apoiada por este trabalho, faremos a terceira grande reformulação, com foco maior no design das telas, das interações e da navegação. As regras de negócio permanecerão sem alteração e serão mantidas no novo design, e as alterações serão codificadas em Python/Flask.

---

<sup>2</sup> <https://fluzao.xyz/>

O código do site de estatísticas de futebol do Fluminense Football Club), objeto do caso deste estudo, não tem código aberto e não estará presente no repositório relacionado ao trabalho. Já os dados levantados referentes ao uso, acesso e links do site serão compartilhados como parte importante deste trabalho.

## 4. Requisitos

### 4.1 Requisitos Funcionais

**RF1** – O sistema deve ter títulos de página bem definidos de forma que seja possível reunir as páginas do mesmo assunto, permitindo uma melhor validação dos resultados.

**RF2** – A ferramenta de coleta de logs deve ser capaz de gerar relatórios no formato CSV.

**RF3** – O sistema deve possibilitar a divisão das classes/consultas em grupos e subgrupos; isso facilita a análise de logs principalmente para análise de poucos acessos.

**RF4** – O sistema deve ser capaz de gerar a análise de logs para qualquer período desejado.

**RF5** – O sistema deve guardar todos os gráficos gerados com a indicação do período consultado para posterior análise de resultados e comparações.

**RF6** – O sistema deve realizar análises de títulos de páginas por data, títulos de páginas por data e hora, acessos totais, acessos por perfil, acessos por cidade, acessos por estado, páginas por código, engajamento, dispositivo que foi realizado o acesso, ferramenta usada para encontrar a página.

**RF7** – O sistema deve conseguir percorrer os HTMLs existentes para identificação de links através de suas âncoras (tag “a” da linguagem).

**RF8** – O sistema deve conseguir identificar os links existentes codificados através do comando “url\_for”.

**RF9** – O sistema deve ser capaz de identificar e listar todas as páginas com determinado número de acessos únicos.

### 4.2 Requisitos Não Funcionais



**RNF1** – Extensibilidade – O sistema deve permitir a criação de análises de novos parâmetros que sejam solicitados.

**RNF2** – Confiabilidade – O sistema deve ser confiável e atender aos requisitos.

**RNF3** – Usabilidade – O sistema deve ser de fácil entendimento e uso.

## 5. Modelo de dados e outras dimensões relevantes

A metodologia de análise de logs para sistemas legados não possui um banco de dados próprio: os arquivos de entrada de cada programa da análise de logs precisam seguir um formato predefinido. Abaixo listamos os arquivos de entrada e seus formatos, a origem de cada coluna e os campos que precisam ser formatados na etapa de pré-processamento, tal como explicado no Manual de utilização. Esses campos serão gerados automaticamente na análise de logs na próxima versão do sistema.

### 5.1 Modelo de Dados

Esta seção descreve as tabelas utilizadas pelo sistema.

#### 5.1.1 acesso\_cidades.csv – Arquivo com a informação de acesso por cidades

Colunas	Descrição	Origem
Cidade	Cidade de onde o usuário acessou o site	Analytics
Data	Data por extenso do acesso	Analytics
Usuários	Número de usuários recorrentes acessando o site	Analytics
Novos Usuários	Número de novos usuários acessando o site	Analytics
Total	Total de usuários – Usuários + Novos Usuários	Excel
DT	Data formatada	Excel

#### 5.1.2 acesso\_estado.csv – Arquivo com a informação de acesso por Estado

Colunas	Descrição	Origem
Regiao	Estado de onde o usuário acessou o site	Analytics
Data	Data por extenso do acesso	Analytics
Usuário	Número de usuários recorrentes acessando o site	Analytics
Primeira Visita	Número de novos usuários acessando o site	Analytics
Total	Total de usuários	Analytics
Usuarios	Total de usuários	Analytics

DT	Data formatada	Excel
----	----------------	-------

### 5.1.3 acesso\_paginas.csv – Arquivo com a informação de acesso por páginas do sistema

Colunas	Descrição	Origem
Pagina	Página acessada	Analytics
Visualizações	Número de visualizações	Analytics

### 5.1.4 acessos.csv – Arquivo com a informação de acesso gerais do sistema

Colunas	Descrição	Origem
Data	Data do acesso	Analytics
Novos usuarios	Número de novos usuários	Analytics
Recorrentes	Número de usuários recorrentes acessando o site	Analytics
Total	Total de usuários	Excel

### 5.1.5 referencia.csv – Arquivo com a ligação de consulta com a página x código de acesso interno do sistema

Colunas	Descrição	Origem
Perguntas	Número da pergunta a que se refere a pesquisa	Mapa
Consulta	Nome da consulta	Mapa
Grupo	Grupo da consulta	Mapa
Subgrupo	Subgrupo da consulta	Mapa
Código	Código para acesso da consulta	Mapa
Pagina	Início do nome da página – referente ao grupo	Mapa
Pagina2	Complemento do nome da página – subgrupo	Mapa
Visualizacao	Sempre zerado – será populado no primeiro processo	--
Usuario	Sempre zerado – será populado no primeiro processo	--

### 5.1.6 telas\_acessadas.csv – Arquivo com os dados das páginas/códigos acessados por data

Colunas	Descrição	Origem
Data	Data do acesso	Analytics
Visualizações	Número de visualizações	Analytics
Usuarios	Número de usuários recorrentes acessando o site	Analytics
Novos usuarios	Número de novos usuários	Analytics
Endereco	Página acessada	Excel

Página	Página acessada sem o código	Excel
Codigo	Codigo da pagina acessada	Excel

#### 5.1.7 Titulos\_data.csv – Arquivo com os dados de títulos das páginas por data

Colunas	Descrição	Origem
Titulo	Título da Página acessada	Analytics
Extenso	Data por extenso	Analytics
Sessões	Sessões que acessaram a página	Analytics
Usuarios	Número de usuários recorrentes acessando o site	Analytics
Novos usuarios	Número de novos usuários	Analytics
Data	Data do acesso formatada	Excel

#### 5.1.8 Titulos\_data\_hora.csv – Arquivo com os dados de títulos das páginas por data/hora

Colunas	Descrição	Origem
Titulo	Título da Página acessada	Analytics
Extenso	Data por extenso	Analytics
Sessões	Sessões que acessaram a página	Analytics
Usuarios	Número de usuários recorrentes acessando o site	Analytics
Novos usuarios	Número de novos usuários	Analytics
Data	Data do acesso formatada	Excel
Hora	Hora do acesso	Excel

#### 5.1.9 jogo.csv – Arquivo com os jogos cadastrados no site – Específico para o objeto de estudo

Colunas	Descrição	Origem
res_flu	Resultado do Fluminense	BD
res_adv	Resultado do Adversário	BD
Cod_jogo	Código do jogo no BD	BD
Adversário	Nome do Adversário	BD
Data	Data do Jogo	BD
Agrupado	Formatação do título da página	BD

#### 5.1.10 Engajamento.csv – Arquivo com tempo médio diário que o usuário passa acessando o sistema

Colunas	Descrição	Origem
Data	Data do acesso	Analytics
Segundos	Média de utilização do sistema no dia	Excel

#### 5.1.11 Como\_acessa\_data.csv - Arquivo com os dispositivos usados por data

Colunas	Descrição	Origem
Dispositivo	Dispositivo usado para entrar no site	Analytics
Extenso	Data por extenso fornecida pelo Analytics	Analytics
Usuário	Número de usuários que acessaram	Analytics
Sessão	Número de sessões abertas	Analytics
Data	Data formatada para geração de relatórios	Excel

#### 5.1.12 Onde\_acessa\_data.csv - Canal de acesso usado para iniciar a navegação

Colunas	Descrição	Origem
Data	Data do acesso	Analytics
Extenso	Data por extenso fornecida pelo Analytics	Analytics
Usuário	Número de usuários que acessaram	Analytics
Sessão	Número de sessões abertas	Analytics
Data	Data formatada para geração de relatórios	Excel

#### 5.1.13 Referencia\_atualizada.csv- Arquivo com a ligação de consulta com a página x código de acesso interno do sistema, atualizado com a quantidade de acesso

Colunas	Descrição	Origem
Perguntas	Número da pergunta a que se refere a pesquisa	Mapa
Consulta	Nome da consulta	Mapa
Grupo	Grupo da consulta	Mapa
Subgrupo	Subgrupo da consulta	Mapa
Código	Código para acesso da consulta	Mapa
Pagina	Início do nome da página – referente ao grupo	Mapa
Pagina2	Complemento do nome da página – subgrupo	Mapa
Visualizacao	Populado no processo gera_referencia	Processo
Usuario	Populado no processo gera_referencia	Processo

#### 5.1.14 Arquivo de saída do questionário

Este arquivo não tem uma estrutura definida, pois depende das perguntas que serão feitas e do sistema de formulários utilizado. Em geral, os sistemas de formulário online permitem exportar os dados em CSV.

## 5.2 Outras dimensões relevantes (Dados Técnicos e Melhorias)

### 5.2.1 Dados Técnicos

A linguagem de programação utilizada na construção do sistema foi o Python através do Anaconda, um gerenciador de ambiente e integração de vários softwares e linguagens para ciência de dados. Foram instaladas as bibliotecas Pandas, Plotly, Numpy. Não existe um front-end integrado, ficando a cargo do pesquisador/desenvolvedor executar o notebook que mais lhe convier, de acordo com a análise que desejar fazer.

### 5.2.2 Melhorias

Como melhorias futuras, pretendemos unificar os códigos, hoje distribuídos em jupyter notebooks, e criar uma área administrativa para a definição dos caminhos para armazenar os gráficos de saída, dos caminhos para os arquivos de entrada e das datas de corte da análise. Todos esses parâmetros fariam parte de um arquivo de configuração utilizado na execução de todos os programas da análise de logs.

Uma outra oportunidade de melhoria diz respeito à facilidade de escolha de qual programa executar e com qual configuração, e assim fazer execuções personalizadas, gerando os resultados de seu interesse.

Finalmente, antevemos também como melhoria a construção das rotinas de geração de datas automáticas, a partir da coluna Extenso, gerada pela Google Analytics, de forma a eliminar etapas de pré-processamento, conforme descrito no Manual.

## 6. Descrição Funcional do Sistema

O sistema busca apoiar a metodologia de análise, que é composta de três partes: coleta de dados do uso do sistema (Google Analytics), processamento dos dados nos programas desenvolvidos e análise dos resultados com formulação de sugestões de redesign (Figura 1).

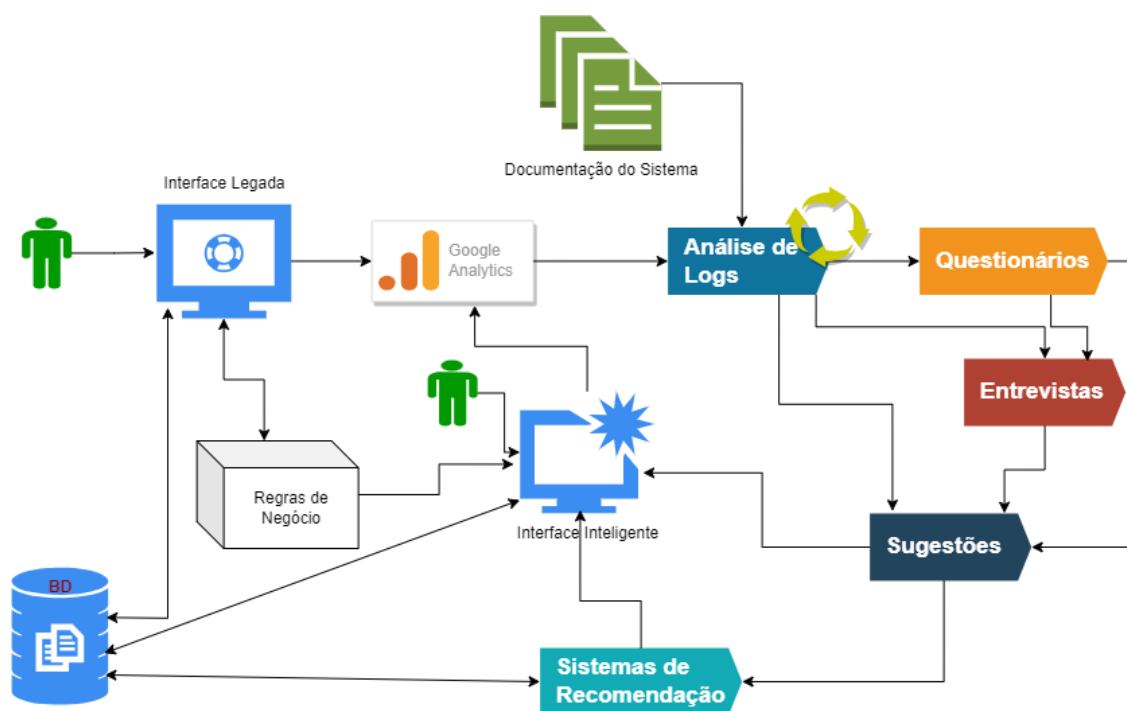


Figura 1 - Metodologia Proposta

## 6.1 Coleta de Dados

A coleta de dados será realizada na ferramenta de análise que melhor atender as necessidades do pesquisador/desenvolvedor. No caso do estudo sendo realizado, escolhemos o Google Analytics.

Uma vez escolhido o programa, a coleta de dados é iniciada. A metodologia prevê a busca de informações dos títulos das páginas por data e por data/hora, das páginas por data, acesso de usuários, acesso de usuários por cidade e por estado, engajamento, qual o dispositivo utilizado e como iniciou o acesso.

A metodologia é flexível, tanto para excluir qualquer uma dessas consultas, bem como para incluir novas que não tenham sido identificadas para o estudo em andamento. Um exemplo é que, para atender ao estudo, não temos a necessidade de buscar informações sobre a geração de receita ou número de cliques; todos esses dados estão disponíveis no Analytics e podem ser baixados e utilizados para a análise fornecendo informações para futuras sugestões de redesign.

A frequência da coleta dos dados vai depender do volume de utilização do sistema. O Google Analytics tem uma limitação de 5000 linhas de log por relatório; portanto, deve ser avaliada qual frequência melhor atende ao sistema legado. Para o estudo realizado, foi determinada uma frequência semanal para todos os relatórios.

A equipe de desenvolvedores deve determinar ainda o período inicial de análise de logs, após o qual devem existir insumos suficientes para a elaboração e aplicação do questionário.

No questionário, as perguntas devem ser direcionadas no sentido de entender com mais profundidade alguns aspectos de navegação identificados pelos logs do sistema. Não existe uma ferramenta específica para o questionário, ficando a cargo de cada desenvolvedor a escolha da ferramenta. Para o estudo sendo realizado, foi escolhido o Google Forms.

A aplicação do questionário também vai variar de sistema a sistema. Podemos enviar para pessoas específicas ou disponibilizar um link dentro do sistema, que permanece por um determinado período para quem quiser responder. O importante é termos uma quantidade de respostas que seja possível analisar e que traga *insights* que ajudem no redesign do sistema legado. Os dados gerados serão coletados e adicionados aos resultados da análise de logs e vão ajudar na criação dos roteiros de entrevistas.

No estudo do site, foi disponibilizado um questionário feito no Google Forms por um mês como um link dentro do sistema. Obtivemos 42 respostas, os resultados foram extraídos da plataforma Google e trabalhados em um notebook específico para a análise desses dados.

Por fim, tivemos a etapa de Entrevistas, na qual usuários-chave serão convidados a conversar sobre detalhes de navegação, entendimento de informações mostradas na tela e o seu funcionamento, propostas de melhoria e extensões, e demais assuntos que sejam importantes e pertinentes para o sistema legado. Não prescrevemos uma ferramenta específica para apoiar as entrevistas, ficando a cargo de cada desenvolvedor coletar esses dados da maneira que lhe convier. Esses dados também são trabalhados em um notebook específico.

No estudo do site, as entrevistas foram feitas com base em um roteiro criado com base na análise dos dados de log e do questionário. As entrevistas são gravadas e as respostas compiladas em um dataset a ser analisado.

## 6.2 Processamento dos dados

Após a coleta dos dados dos logs, as informações precisam ser processadas, transformadas em gráficos, tabelas e arquivos de saída, fornecendo insumo para as sugestões de redesign dos sistemas legados.

Este processo deve ser dividido em duas fases: uma manual de pré-processamento dos dados, e outra automática, fazendo uso dos programas criados para esta análise. Desenvolvemos sete programas de análise de logs; porém, a metodologia é flexível, e somente o que for necessário deve ser executado. A criação de novos programas de análise tende a ser facilitada: mantendo as nomenclaturas

de campos e seus tipos, podemos reutilizar o que já foi criado para novas análises, com pequenos ajustes.

### 6.2.1 Pré-processamento

O pré-processamento dos dados consiste em concatenar datasets de análise de logs recém captados da ferramenta de coleta de dados (no caso, Google Analytics) com um dataset principal, que tem todos os datasets de análise juntos. Caso exista uma coluna de data sem formatação, é necessário criar manualmente uma coluna de data formatada. Já está em curso uma melhoria para a geração automática desta data nos programas de análise, mas ainda sem data prevista de implementação.

Outro ponto a ser abordado no pré-processamento dos dados é em relação aos acentos: em determinadas instalações, os acentos podem se tornar um problema nos programas automáticos, seja na análise ou na exibição dos dados. Por isso criamos, para o estudo, uma macro no aplicativo Notepad++, que realiza ajustes de acentuação e transforma a vírgula (,) em ponto e vírgula (;), para facilitar acesso a esses datasets via o aplicativo Excel, para eventuais inspeções e ajustes manuais.

Para finalizar o pré-processamento, os arquivos coletados e o arquivo completo são abertos no Excel e o conteúdo do parcial é concatenado. Com isso, é gerado um arquivo atualizado para o uso nos programas de análise automática dos logs.

### 6.2.2 Processamento

Com a geração dos datasets completos com os dados do sistema legado, podemos executar os programas de análise de log. Como mencionado anteriormente, existe uma flexibilidade quanto ao que deve ser executado para cada sistema legado; vai depender da disponibilidade das informações, da qualidade das informações e da importância das informações para a análise.

No estudo sendo realizado, foram criados nove programas para análise de logs. Foi criado ainda um programa para recuperar *tags* dentro de páginas HTML; este programa não faz parte do conjunto de programas de análise de logs e é executado



no início do ciclo de vida da modernização, com a intenção de capturar e listar todos os links existentes no sistema.

Uma melhoria já identificada será a criação de um programa que identifique todas as *tags* TITLE das páginas HTML do sistema legado, para trazer um mapeamento dos títulos das páginas que contribuirá para uma análise de logs mais eficiente. Para o caso do estudo em andamento, o sistema legado já tinha os títulos todos bem formados e por isso a identificação dos títulos via programa não foi necessária; mas, para aplicação da metodologia em sites em que isso não ocorra, isso seria bastante útil.

Em geral, o processamento dos datasets é independente entre si. As únicas duas dependências são a análise de páginas acessadas e análise de páginas nunca acessadas ; estes dois programas dependem da geração da referência atualizada. Os demais programas de análise dos títulos por data e por data e hora, acessos ao sistema, informações de tecnologia e análise de questionário podem ser executados a qualquer momento.

Em cada programa é possível definir datas de início e fim da análise, permitindo a limitação de um período de tempo para a geração das informações. Os gráficos e datasets de saída são salvos com essas datas embutidas no nome, para permitir avaliar de forma incremental os dados coletados. Uma sugestão de melhoria no processo é criar uma área de configuração e disparo dos programas, onde o pesquisador/desenvolvedor possa escolher as datas na tela e executar o programa do seu interesse.

### 6.2.3 Sugestões de Melhoria

As sugestões de melhoria serão realizadas após análise do desenvolvedor dos dados coletados nas três fases do projeto. Deve ser gerado um relatório com recomendações e uma explicação, com base nos dados, sobre os motivos que levaram o desenvolvedor a julgar que a melhoria proposta trará ganhos para a experiência do usuário.

## 7. Manual do usuário

O site sendo analisado pode ser de código fechado. Os códigos referentes à metodologia de redesign de sistemas legados através da análise de logs estão no GIT [https://github.com/RodLima26/Analise\\_legado](https://github.com/RodLima26/Analise_legado).

### 7.1 Premissas

Para o melhor aproveitamento da metodologia de redesign de sistemas legados pela análise de logs, algumas premissas devem ser atendidas. O não cumprimento

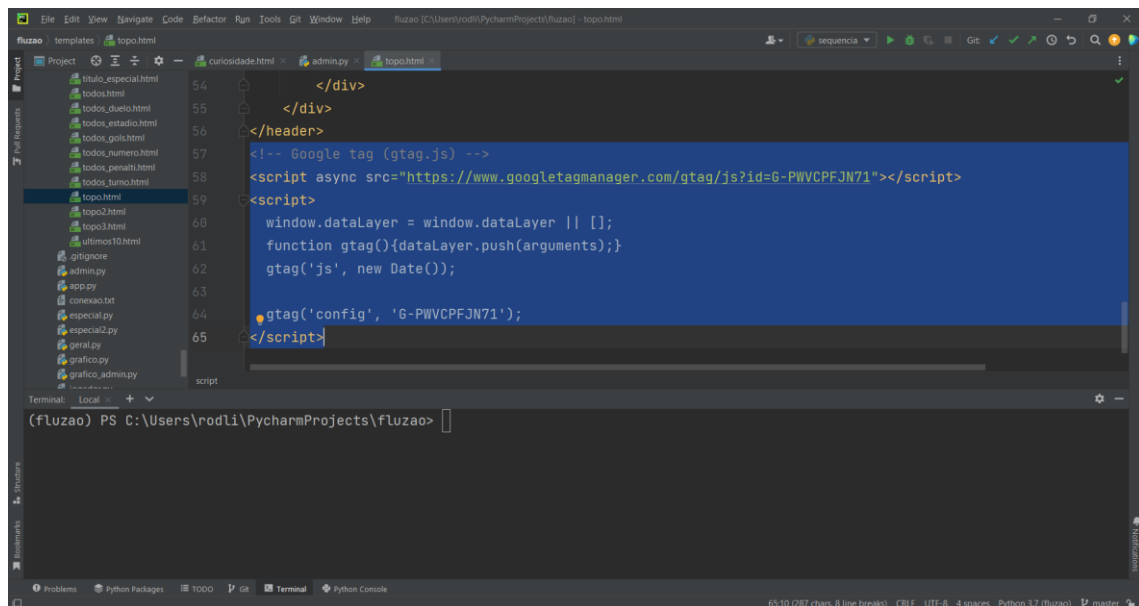
destas premissas não invalida o processo, porém o torna bem mais complexo e difícil de chegar a uma sugestão robusta de melhoria.

- 1) Utilização da ferramenta Google Analytics 4 para a extração de relatórios de navegação, páginas, títulos das páginas por data, títulos das páginas por data hora, acesso de usuários totais e acesso de usuários por cidade e estado, engajamento, dispositivo utilizado e qual a fonte para a entrada no sistema.
- 2) O sistema precisa estar com os títulos (tag TITLE das páginas HTML) bem formatados permitindo uma melhor análise da navegação.
- 3) Um mapa com as páginas e seus códigos de acesso pode ser criado, gerando o arquivo referencia.csv, que será utilizado para a identificação dos acessos a cada consulta/página do sistema. Essa informação gera um bônus de análise e não impede ou invalida as demais análises. Um pedaço desse mapa pode ser criado usando o programa de identificação de links que será detalhado neste relatório.
- 4) Dividir o sistema previamente em grupos e subgrupos ajuda na análise das páginas acessadas verificando os acessos em diferentes áreas de informação, identificando em quais pontos precisamos agir para que todo o sistema seja visualizado, trazendo mais informações para o usuário.

## 7.2 Google Analytics

O primeiro passo a ser realizado é identificar qual ferramenta de coleta de dados será utilizada. A ferramenta indicada pela metodologia é o Google Analytics, gratuito, de fácil implementação e de grande contribuição para a coleta de dados em todos os níveis de um site.

Para ter acesso à ferramenta, é necessário ter uma conta na Google, depois procurar pelo Google Platform e finalmente Google Analytics. Deve-se seguir o procedimento de cadastro e, ao final, gerar o código HTML sugerido para embutir nas páginas HTML do sistema legado, de forma a cobrir toda a navegação realizada pelo sistema. No objeto de estudo deste trabalho, o código foi replicado no arquivo topo.html, código HTML comum a todas as páginas do site, conforme Figura 2.



The screenshot shows the PyCharm IDE interface. The left sidebar displays a project tree with files like `curiosidade.html`, `admin.py`, and `topo.html`. The main editor window is open to `topo.html`, showing HTML code with a Google Analytics tag implementation. The code includes a comment `<!-- Google tag (gtag.js) -->`, a script tag `<script async src="https://www.googletagmanager.com/gtag/js?id=G-PWVCPFJN71"></script>`, and a `<script>` block containing `window.dataLayer = window.dataLayer || [];`, `function gtag(){dataLayer.push(arguments);}`, `gtag('js', new Date());`, and `gtag('config', 'G-PWVCPFJN71');`. The bottom panel shows a terminal window with the command prompt `(fluzao) PS C:\Users\rodli\PycharmProjects\fluzao>`.

```
54     </div>
55   </div>
56 </header>
57
58 <!-- Google tag (gtag.js) -->
59 <script async src="https://www.googletagmanager.com/gtag/js?id=G-PWVCPFJN71"></script>
60 <script>
61   window.dataLayer = window.dataLayer || [];
62   function gtag(){dataLayer.push(arguments);}
63   gtag('js', new Date());
64   gtag('config', 'G-PWVCPFJN71');
65 </script>
```

Terminal: Local +  
(fluzao) PS C:\Users\rodli\PycharmProjects\fluzao>

**Figura 2 - Código para captura de dados pelo Google Analytics**

Após a publicação do sistema com a tag do Google Analytics, devemos estabelecer a frequência de recuperação dos dados. Isso vai depender da quantidade de acessos que o sistema legado apresenta. Existe uma limitação de download de 5000 linhas em arquivos CSV pelo Google Analytics. Portanto, quanto maior o número de acessos, maior deverá ser a frequência de captura dos dados.

O Google Analytics já fornece um bom número de relatórios em sua versão inicial. Listamos a seguir os locais onde extrair os relatórios utilizados em nossa análise. É importante frisar que nem todos os relatórios listados vão fazer sentido para todos os sistemas, e novos relatórios podem ser extraídos ou criados sempre que necessário pelo desenvolvedor. Para isso, o Google Analytics fornece um passo-a-passo bem detalhado e fácil de entender. Todos os relatórios devem ser baixados no formato CSV, clicando no ícone de compartilhar escolhendo a opção de download do arquivo.

## 1) Acessos gerais - Página principal da opção de Relatórios

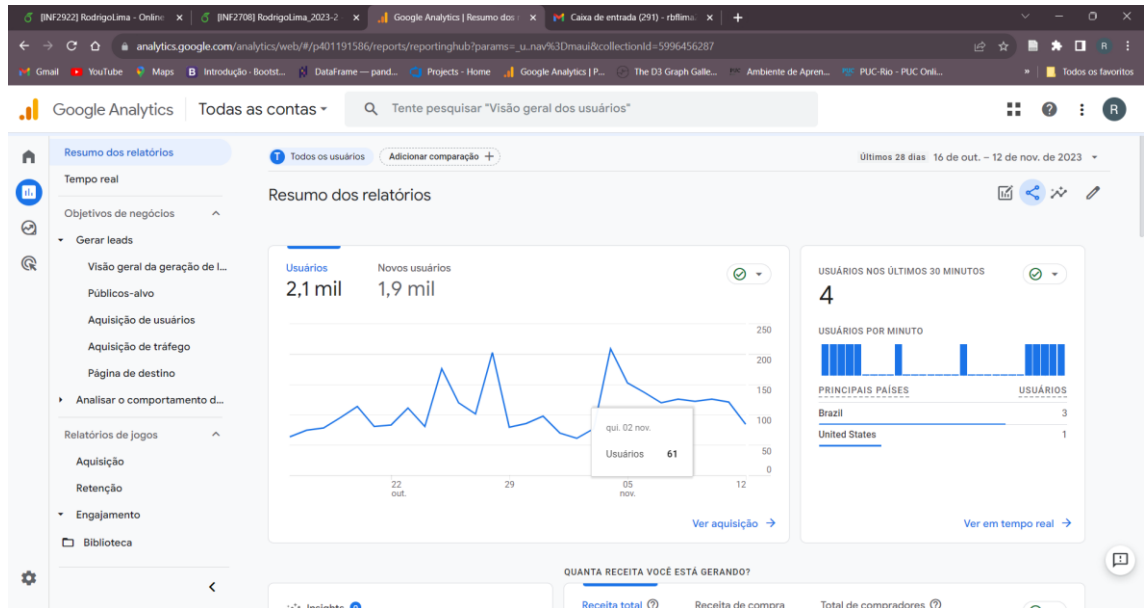


Figura 3 - Relatório Inicial do Google Analytics

## 2) Acessos por cidade – Objetivos de Negócios/Gerar Leads/Visão geral da geração de leads → encontrar o gráfico de acessos por cidade, entrar na página específica de acessos por cidade conforme especificado na Figura 3 e incluir a data conforme mostrado na figura 4

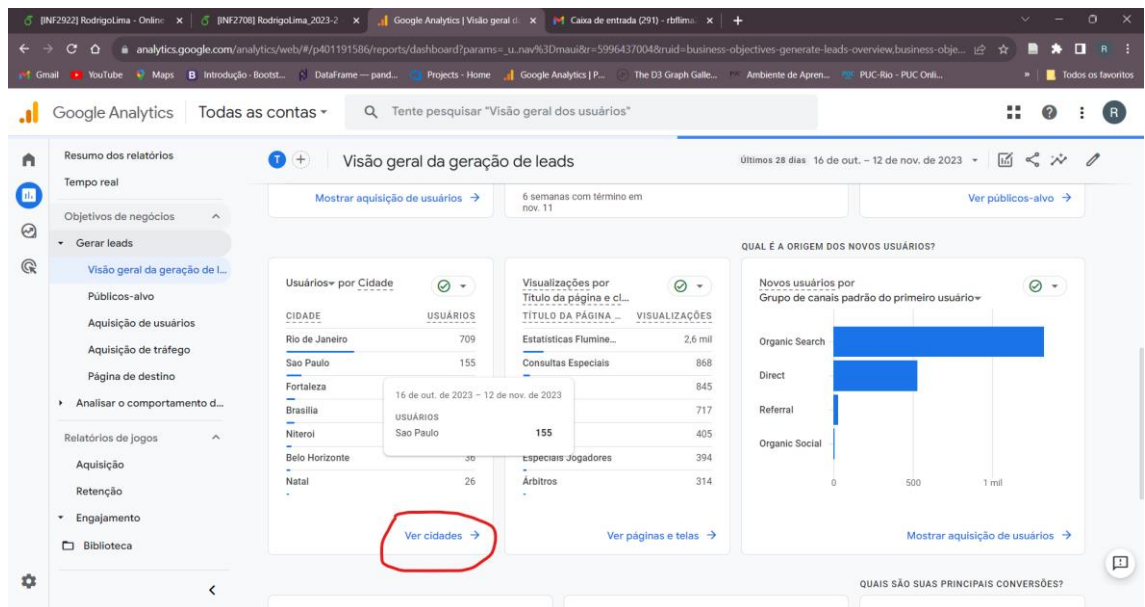


Figura 4 - Relatório de acessos por cidade

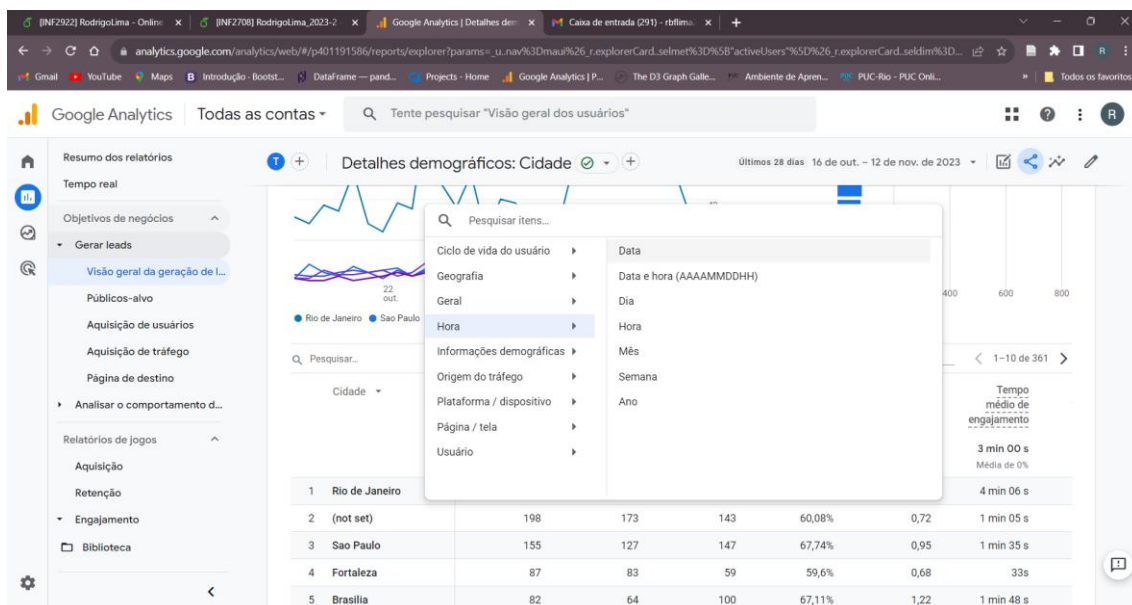


Figura 5 - Incluindo a coluna data

### 3) Engajamento – Relatório de jogos/Engajamento/Engajamento

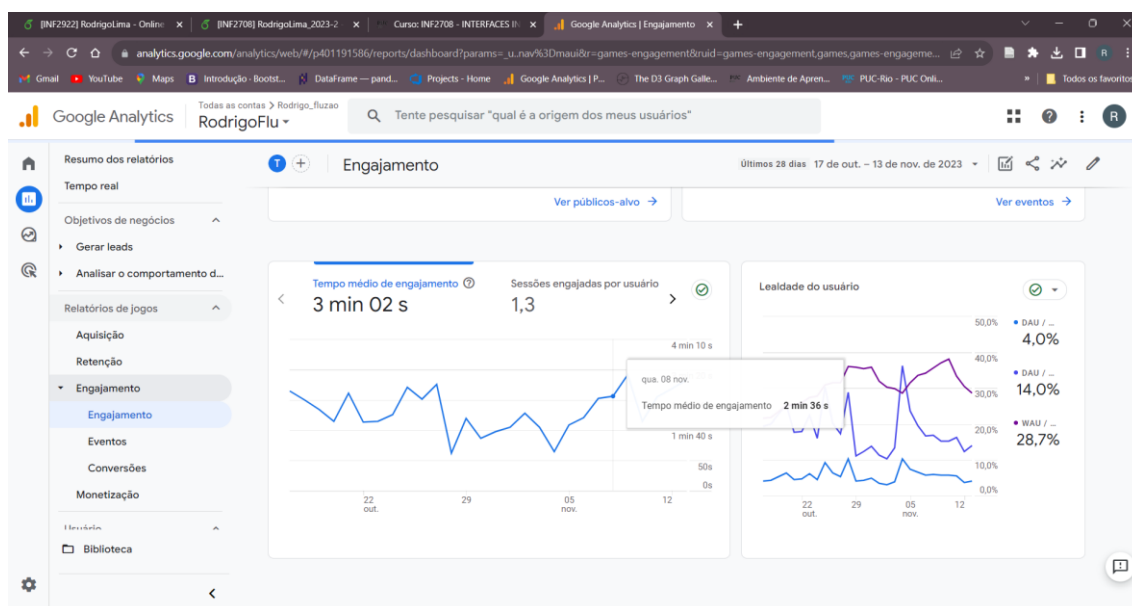


Figura 6 - Relatório de Engajamento

#### 4) Dispositivos de acesso - Relatório de jogos/Aquisição

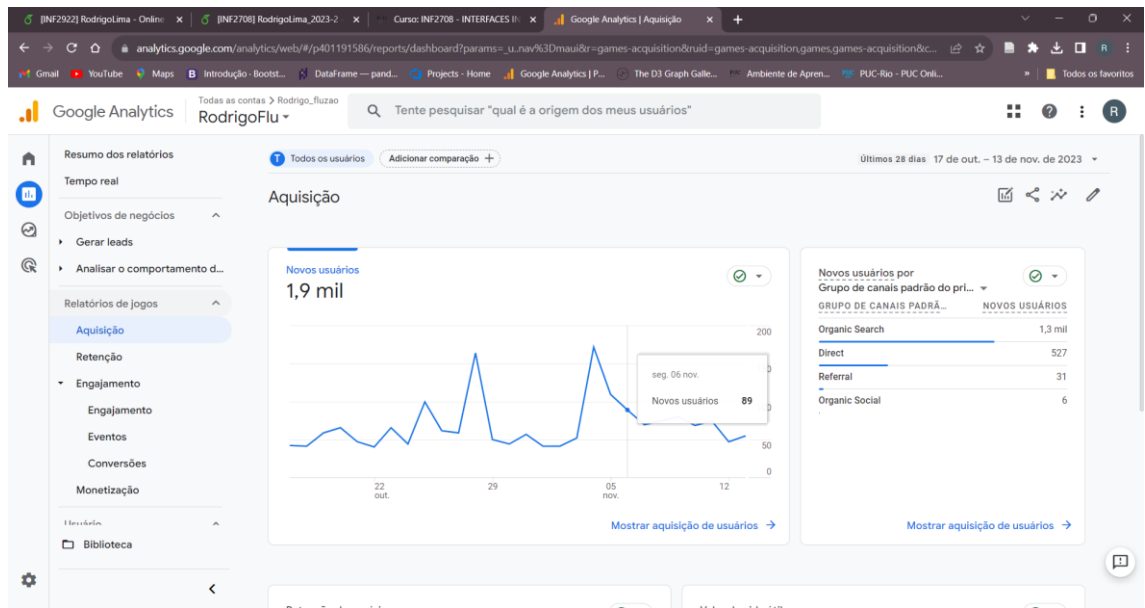


Figura 7 - Relatório de como iniciou o acesso ao sistema

#### 5) Como acessa o site – Tecnologia / Visão Geral, entrar na opção dispositivos de plataforma e adicionar a coluna data

The screenshot displays the Google Analytics 'Detalhes de tecnologia: Categoria da plataforma / dispositivo' (Technology Details: Platform/Device Category) report. The table shows data for various devices and platforms, including web/mobile, web/desktop, and web/tablet, with columns for Users, New Users, Sessions, and Engagement Rate.

	Categoria da pl. / dispositivo	Data	Usuários	Novos usuários	Sessões engajadas	Taxa de engajamento	Sessões engajadas por usuário
			2.162 100% do total	1.909 100% do total	2.916 100% do total	70,47% Média de 0%	1,35 Média de 0%
1	web / mobile	20231104	168	148	119	60,1%	0,71
2	web / mobile	20231028	160	143	95	51,63%	0,59
3	web / mobile	20231105	112	88	97	69,29%	0,87
4	web / mobile	20231025	106	63	102	66,67%	0,96
5	web / mobile	20231111	91	62	86	73,5%	0,95
6	web / mobile	20231106	86	67	73	72,28%	0,85
7	web / mobile	20231108	79	55	70	66,67%	0,89
8	web / desktop	20231025	69	37	91	88,35%	1,32
9	web / mobile	20231110	69	43	67	71,28%	0,97

Figura 8 - Relatório de dispositivos de acesso

## 6) Páginas e Consultas

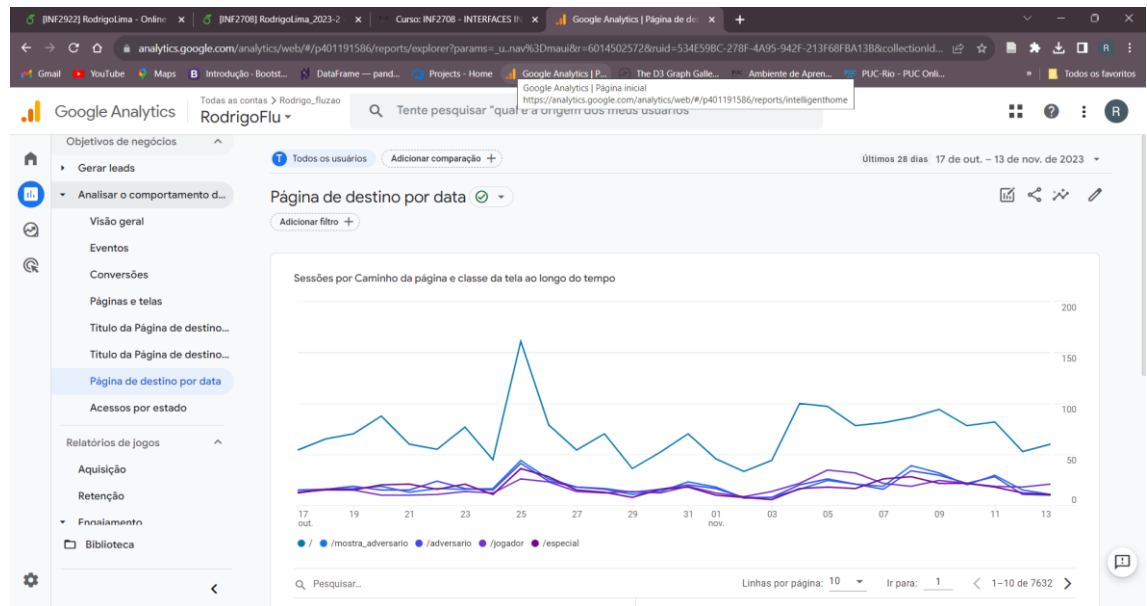


Figura 9 - Relatório de Páginas e consultas por data

## 7) Título das páginas por data

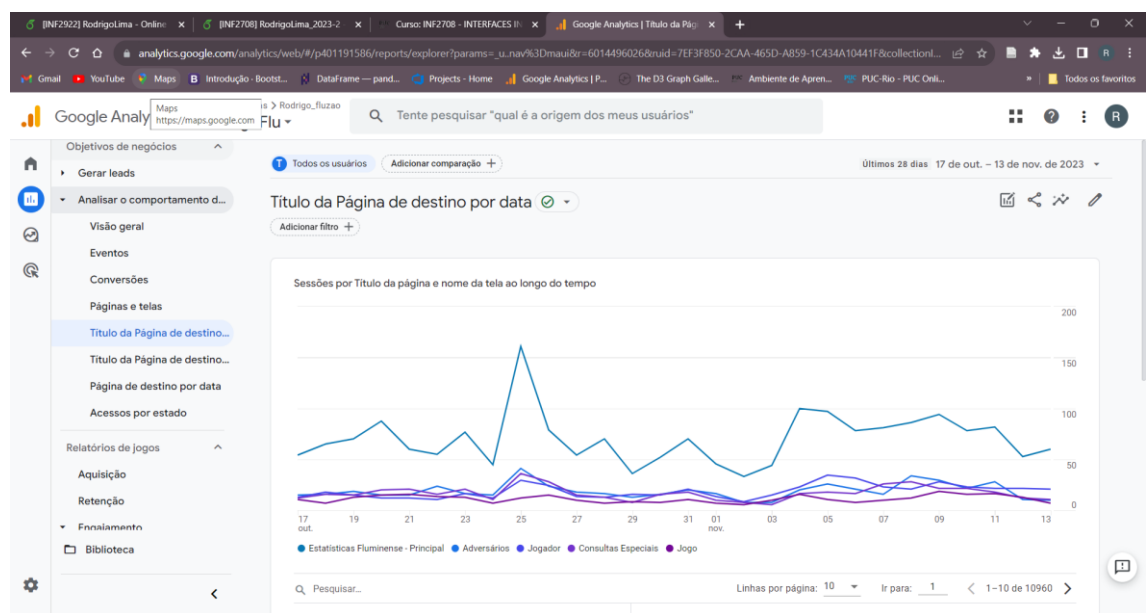


Figura 10 - Relatório de títulos acessados por data

8) Título das páginas por data e hora

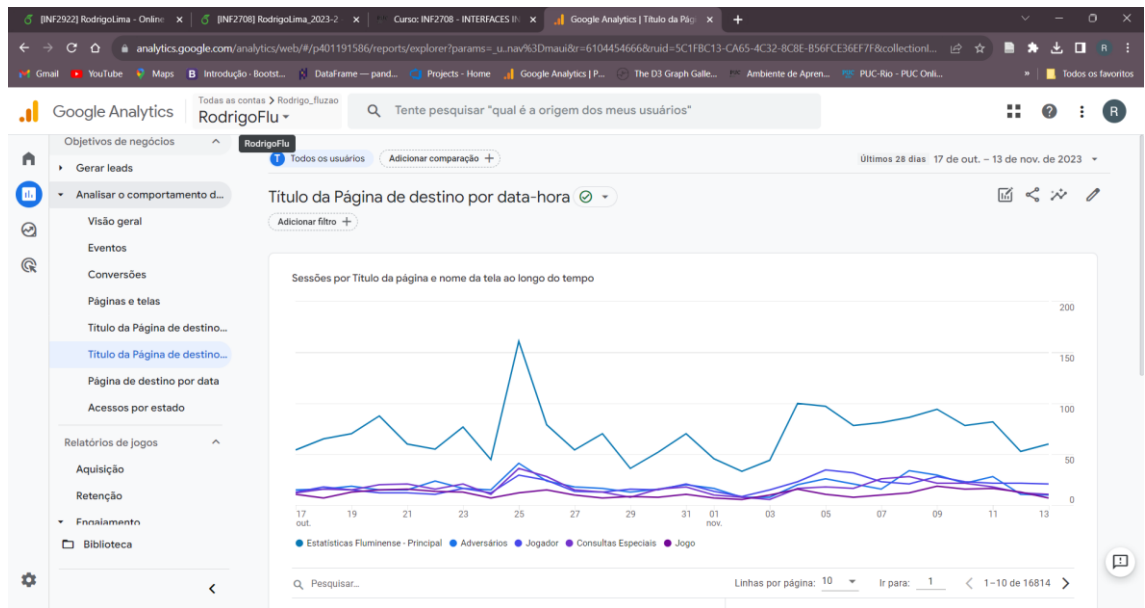


Figura 11 - Relatório de títulos acessados por data e hora

9) Acesso por estado

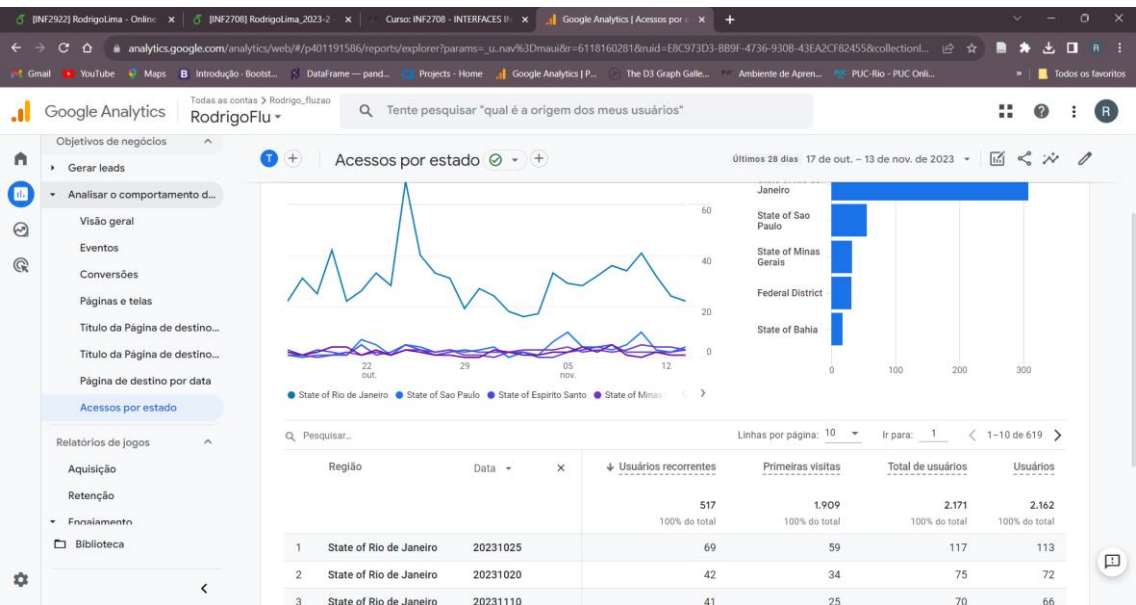


Figura 12- Relatório de acessos por estado



Os relatórios de Páginas e Títulos, itens 6, 7 e 8 da lista acima não são fornecidos inicialmente pelo Google Analytics e precisam ser definidos pelo desenvolvedor. Para criar um relatório dentro da ferramenta, devemos clicar em Biblioteca no menu lateral esquerdo e depois em Criar novo relatório, e seguir as instruções, incluindo as colunas que são importantes para a análise a ser realizada.

Os arquivos baixados do analytics deverão ter o nome formatado com um texto que indique a que os dados se referem e as datas de corte. Recomendamos que sejam guardados dentro de uma pasta logs do projeto principal, para futura referência ou auditoria.

Devido à formatação de acentos e ao caractere de divisão dos campos (,) é executada uma macro no Notepad++ nos arquivos CSV baixados que remove todos os acentos e troca o caractere de “,” para “;”, desta forma, é possível a abertura no Excel sem problema.

Por conta da limitação de 5000 linhas do Google Analytics, após a formatação, os arquivos devem ser abertos e concatenados com os seus respectivos arquivos “completos” (resultantes de concatenações de períodos anteriores) que vão estar armazenados na pasta Arquivos; a formatação do nome deve ser igual à dos logs sem a data de corte, o que facilita a correlação para a concatenação.

Os campos de data são exportados no formato extenso do Google Analytics, Ex: 20230920, desta forma ainda é necessário, dentro do Excel, a transformação para o formato de data em outra coluna. Isso ainda é realizado manualmente, mas já identificamos como oportunidade de melhoria fazer a transformação automática na execução das análises.

Ao final deste processo, são gerados os arquivos CSV que serão entrada para as análises.

- 1) acessos.csv
- 2) acesso\_cidades.csv
- 3) engajamento.csv
- 4) onde\_acessa\_data.csv
- 5) como\_acessa\_data.csv
- 6) telas\_acessadas.csv
- 7) titulos\_data.csv
- 8) titulos\_data\_hora.csv
- 9) acesso\_estado.csv

Além desses, o arquivo referencia.csv deve ser gerado como apoio aos programas que verificam a quantidade de consultas com nenhum ou pouco acesso. Segue abaixo um exemplo dos dados que esse arquivo deve conter.

Salvamento Automático Referencia.csv

Pesquisar

Rodrigo Lima

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibir Ajuda Acrobat

Comentários Compartilhamento

Colar

Calibri 11 A\*

Fonte

Alinhamento

Número

Estilos

Formato Condicional

Formatar como Tabela

Estilos de Célula

Inserir

Excluir

Formatar

Células

Classificar e Filtrar

Localizar e Selecionar

Suplementos

Criar PDF e compartilhar o link

Criar PDF e compartilhar via Outlook

Adobe Acrobat

POSSÍVEL PERDA DE DADOS Alguns recursos podem ser perdidos se você salvar esta pasta de trabalho no formato (.csv) delimitada por vírgulas. Para preservar estes recursos, salve-a em um formato de arquivo do Excel.

Não mostrar novamente Salvar como...

B3 Especial

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
142	153	Adversario x Arbitro	Especiais	Adversario	100101000	consulta_especial	mostra_especial		0	0				
143	154	Adversario x Estádio	Especiais	Adversario	100102000	consulta_especial	mostra_especial		0	0				
144	155	Adversário em campeonato	Especiais	Adversario	100103000	consulta_especial	mostra_especial		0	0				
145	156	Adversário x Jogador	Especiais	Adversario	100104000	consulta_especial	mostra_especial		0	0				
146	157	Adversário x Técnico	Especiais	Adversario	100105000	consulta_especial	mostra_especial		0	0				
147	158	Adversário do mesmo país	Especiais	Adversario	200000000	consulta_espadv	mostra_espadv2		0	0				
148	159	Adversário do mesmo estado	Especiais	Adversario	201000000	consulta_espadv	mostra_espadv2		0	0				
149	160	Adversário do mesmo estado em campeonato	Especiais	Adversario	201103000	consulta_espadv	mostra_espadv2		0	0				
150	161	Adversários Internacionais	Especiais	Adversario	202000000	mostra_espadv	N/A		0	0				
151	162	Adversário em campeonato x Estádio	Especiais	Adversario	100103102	consulta_especial	mostra_especial		0	0				
152	163	Adversário por ano	Especiais	Adversario	106100000	consulta_espadv	mostra_espadv2		0	0				
153	164	Adversários em vantagem	Especiais	Adversario	203000000	consulta_espadv	mostra_adversario		0	0				
154	165	Principais Adversários	Especiais	Adversario	250000000	consulta_prinadv	mostra_prinadv		0	0				
155	166	Principais Adversários em campeonato	Especiais	Adversario	250103000	consulta_prinadv	mostra_prinadv		0	0				
156	167	Principais Adversários em estádio	Especiais	Adversario	250102000	consulta_prinadv	mostra_prinadv		0	0				
157	168	Todos os jogos contra adversário	Especiais	Adversario	100999000	consulta_especial	mostra_todos		0	0				
158	169	Lista de adversários por estado	Especiais	Adversario	201150000	consulta_espadv	mostra_espadv2		0	0				
159	170	Lista de adversários por país	Especiais	Adversario	200150000	consulta_espadv	mostra_espadv2		0	0				
160	171	Mais penalities contra o Fluminense	Especiais	Adversario	210000000	mostra_penalti	N/A		0	0				
161	172	Mais penalities a favor do Fluminense	Especiais	Adversario	211000000	mostra_penalti	N/A		0	0				
162	173	Mais penalities convertidos pelo Fluminense	Especiais	Adversario	212000000	mostra_penalti	N/A		0	0				
163	174	Mais penalities perdidos pelo Fluminense	Especiais	Adversario	213000000	mostra_penalti	N/A		0	0				
164	175	Disputa de penalty por adversário	Especiais	Adversario	100290000	consulta_disputa	mostra_disputa		0	0				
165	176	Cartões contra adversário	Especiais	Adversario	100289000	consulta_cartao	N/A		0	0				

Referencia

Pronto Acessibilidade: não disponível

Figura 13 - Exemplo do arquivo referencia.csv

Para auxiliar a análise dos logs relacionados aos jogos de futebol mais acessados da história, geramos também um arquivo jogo.csv. Segue abaixo um exemplo do arquivo, que é exclusivo para a análise do site Fluzao.xyz.

```
C:\Users\rodil\OneDrive\Documentos\Mestrado\Arquivos\jogo.csv - Notepad++
Arquivo  Editar  Localizar  Visualizar  Formatar  Linguagem  Configurações  Ferramentas  Macro  Executar  Plugins  Janela  ?
new 3  x  new 1  x  novo 1  x  jogo.csv  x  referencia.csv  x  jogo 2  x
5951  51:8702:River Plate;02/05/2023;Fluminense 5 X 1 River Plate -
5952  1:1:8703:Vasco;06/05/2023;Fluminense 1 X 1 Vasco -
5953  2:0:8704:Cruzeiro;10/05/2023;Fluminense 2 X 0 Cruzeiro -
5954  2:0:8705:Cuiaba;13/05/2023;Fluminense 2 X 0 Cuiaba -
5955  0:0:8706:Flamengo;16/05/2023;Fluminense 0 X 0 Flamengo -
5956  0:1:8707:Botafogo;20/05/2023;Fluminense 0 X 1 Botafogo -
5957  0:1:8708:The Strongest;25/05/2023;Fluminense 0 X 1 The Strongest -
5958  0:2:8709:Corinthians;28/05/2023;Fluminense 0 X 2 Corinthians -
5959  0:2:8710:Flamengo;01/06/2023;Fluminense 0 X 2 Flamengo -
5960  2:1:8711:Bragantino;04/06/2023;Fluminense 2 X 1 Bragantino -
5961  0:2:8712:River Plate;07/06/2023;Fluminense 0 X 2 River Plate -
5962  2:2:8713:Goias;11/06/2023;Fluminense 2 X 2 Goias -
5963  1:1:8714:Atlético MG;21/06/2023;Fluminense 1 X 1 Atlético MG -
5964  2:1:8715:Bahia;24/06/2023;Fluminense 2 X 1 Bahia -
5965  1:1:8716:Sporting Cristal;27/06/2023;Fluminense 1 X 1 Sporting Cristal -
5966  0:1:8717:São Paulo;01/07/2023;Fluminense 0 X 1 São Paulo -
5967  2:0:8718:Internacional;09/07/2023;Fluminense 2 X 0 Internacional -
5968  0:0:8719:Flamengo;16/07/2023;Fluminense 0 X 0 Flamengo -
5969  0:2:8720:Coritiba;24/07/2023;Fluminense 0 X 2 Coritiba -
5970  1:0:8721:Santos;29/07/2023;Fluminense 1 X 0 Santos -
5971  1:1:8722:Argentinos Juniors;01/08/2023;Fluminense 1 X 1 Argentinos Juniors -
5972  2:1:8725:Palmeiras;05/08/2023;Fluminense 2 X 1 Palmeiras -
5973  2:0:8726:Argentinos Juniors;08/08/2023;Fluminense 2 X 0 Argentinos Juniors -
5974  1:2:8727:Grêmio;13/08/2023;Fluminense 1 X 2 Grêmio -
5975  3:1:8728:América MG;19/08/2023;Fluminense 3 X 1 América MG -
5976  2:0:8729:Olimpia;24/08/2023;Fluminense 2 X 0 Olimpia -
5977  2:2:8730:Atletico PR;27/08/2023;Fluminense 2 X 2 Atletico PR -
5978  3:1:8731:Olimpia;31/08/2023;Fluminense 3 X 1 Olimpia -
5979  1:0:8732:Fortaleza;03/09/2023;Fluminense 1 X 0 Fortaleza -
5980  2:4:8733:Vasco;16/09/2023;Fluminense 2 X 4 Vasco -
5981  1:0:8734:Cruzeiro;20/09/2023;Fluminense 1 X 0 Cruzeiro -
5982  2:2:8735:Internacional;27/09/2023;Fluminense 2 X 2 Internacional -
5983  0:3:8736:Cuiaba;30/09/2023;Fluminense 0 X 3 Cuiaba -
5984  2:1:8737:Internacional;04/10/2023;Fluminense 2 X 1 Internacional -
5985  0:2:8738:Botafogo;08/10/2023;Fluminense 0 X 2 Botafogo -
5986  3:3:8739:Corinthians;19/10/2023;Fluminense 3 X 3 Corinthians -
5987  0:1:8740:Bragantino;22/10/2023;Fluminense 0 X 1 Bragantino -
5988  5:3:8741:Goias;25/10/2023;Fluminense 5 X 3 Goias -
5989  0:2:8742:Atlético MG;28/10/2023;Fluminense 0 X 2 Atlético MG -
5990  0:1:8743:Bahia;31/10/2023;Fluminense 0 X 1 Bahia -
5991  2:1:8744:Boca Juniors;04/11/2023;Fluminense 2 X 1 Boca Juniors -
5992  0:0:8745:Internacional;08/11/2023;Fluminense 0 X 0 Internacional -
5993  1:1:8746:Flamengo;11/11/2023;Fluminense 1 X 1 Flamengo -
Normal text file      length: 368.072  lines: 5.993      Ln: 5.993  Col: 55  Pos: 368.071      Windows (CR LF)  UTF-8      INS
```

Figura 14 - Exemplo do arquivo jogo.csv

## 7.3 Processamento

O processamento dos notebooks criados não tem uma ordem específica e deve ser executado sempre que o desenvolvedor quiser fazer uma análise. Dentro de cada programa temos os limitadores de data, que permitem fazer um recorte dos datasets entre as datas escolhidas e fazer uma análise de um período de tempo menor. Essas análises podem evidenciar o comportamento de acesso ao sistema reagindo a eventos externos e trazer valiosa contribuição para as sugestões de ajuste, criação de sistema de recomendação.

A única exceção à regra de ordem de execução se encontra nas pesquisas em que o arquivo de referência (referencia.csv) for criado. Nesse caso, o programa gera\_referencia.ipnyb deve ser executado antes do Analise\_Paginas.ipnyb e do Analise\_consulta\_zerada.ipnyb. O arquivo de referencia.csv pode já existir na instalação e ser necessário realizar somente uns ajustes de formato. Outra opção é gerar uma primeira versão executando o programa Identificando\_links.ipnyb; esse programa identifica todos os links chamados dentro das páginas HTML do site legado, através da tag a e do comando url\_for. Após sua execução será necessária a compilação dos dados para adequação ao formato do arquivo referencia.csv. Novas

adições de identificação de links e chamadas a consultas vão tornar o programa cada vez mais robusto e confiável.

A divisão das consultas ou páginas do sistema em grupos e subgrupos de correspondência auxilia a análise de logs a identificar pontos do sistema legado com menos acesso, pontos com muito acesso, e fornece dados que serão explorados nos questionários, entrevistas e, mais importante, na identificação de oportunidades e formulação de recomendações de melhoria. Ajuda também a focar análises em determinadas partes do sistema, trazendo maior entendimento do uso do sistema em todas as áreas.

No arquivo de páginas e consultas, a página inicial aparece com o caminho da consulta em branco. Isso pode ser contornado incluindo manualmente a palavra “inicial”. Caso seja deixado em branco, pode trazer algum prejuízo no relatório a ser gerado, em termos de documentação.

Todo o processamento é realizado utilizando o Jupyter Notebook, acessado dentro do ambiente Anaconda, as bibliotecas envolvidas no processo foram a numpy, pandas, plotly, matplotlib e warnings.

A seguir descrevemos cada programa, suas entradas, saídas e objetivo principal, bem como alguns detalhes sobre o objeto de estudo (site legado) específico.

### 7.3.1 Gera\_referencia.ipynb

**Entradas** – referencia.csv e telas\_acessadas.csv

**Saídas** – referencia\_atualizada.csv

**Objetivo** – Atualizar as colunas Visualizacao e Usuario com as quantidades de acesso, gerando o referencia\_atualizada.csv que será a base de todo o resto da análise de páginas acessadas.

**Detalhe específico do objeto de estudo** - O campo código da referencia.csv é único no sistema legado e portanto é utilizado para atualizar os acessos de todas as páginas que contenham código no relatório gerado pelo analytics(telas\_acessadas.csv), páginas de link e gráfico também são atualizadas com o código de acesso, as demais páginas são consideradas páginas de trânsito e não contém informações finais por isso não são consideradas importantes para o estudo e neste caso serão desconsideradas.

### 7.3.2 Analise\_Titulos.ipynb

**Entradas** – Titulos\_data.csv

**Saídas** – Gráficos de análise de navegação, os gráficos que vão ser gerados dependem da análise de título que está sendo realizada

**Objetivo** – Realizar a análise da navegação dos usuários por data, fazendo o corte por data de acesso e por palavras-chave que agrupem consultas/páginas de mesma natureza para que seja possível estabelecer uma conexão dos acessos com eventos externos, permitindo a criação de um sistema de recomendação por conteúdo. Podem ser criados filtros por palavra-chave tantos quanto necessário para realizar a melhor análise. Outro objetivo é identificar as páginas mais acessadas do sistema legado independente de data e agrupamentos, outra análise executada é sobre os maiores acessos por data mas sem agrupamento e por último removendo algumas páginas, e.g. página inicial, para identificar algum viés de acesso.

**Detalhe específico do objeto de estudo** – Esta sendo analisado um sistema legado de estatística de um clube de futebol, então é importante analisar qual o adversário que tem mais acesso no geral, qual adversário que tem mais acesso perto dos confrontos, quais jogadores tem a sua página mais acessada, se existe alguma conexão com os dados informados em transmissões esportivas com as consultas realizadas em seguida.

### 7.3.3 Analise\_Titulos-Hora.ipynb

**Entradas** – Titulos\_data\_hora.csv

**Saídas** – Gráficos de análise de navegação, os gráficos que vão ser gerados dependem da análise de título que está sendo realizada

**Objetivo** – Realizar a análise da navegação dos usuários por data e hora, fazendo o corte por data de acesso e por palavras-chave que agrupem consultas/páginas de mesma natureza para que seja possível estabelecer uma conexão dos acessos com eventos externos, permitindo a criação de um sistema de recomendação por conteúdo. Podem ser criados filtros por palavra-chave tantos quanto necessário para realizar a melhor análise. Outro objetivo é identificar as páginas mais acessadas do sistema legado independente de data e agrupamentos, outra análise executada é sobre os maiores acessos por data mas sem agrupamento e por último removendo algumas páginas, e.g. página inicial, para identificar algum viés de acesso.

**Detalhe específico do objeto de estudo** – Este estudo foi criado com a intenção de identificar se durante os jogos o sistema recebia mais ou menos acessos, foi apurado que o sistema recebia menos acessos durante os jogos, porém seus picos são na noite anterior e logo após a disputa.

### 7.3.4 Analise\_Paginas.ipynb

**Entradas** – referencia\_atualizada.csv; jogo.csv; telas\_acessadas.csv

**Saídas** – Gráficos de análise de navegação, os gráficos que vão ser gerados dependem da análise de páginas, grupos/subgrupos e jogos que está sendo realizada

**Objetivo** – Realizar a análise da navegação nas páginas dos usuários por data, fazendo o corte por data de acesso, identificar páginas com pouco ou nenhum acesso, os grupos e subgrupos com maior quantidade relativa de consultas sem acesso, grupos e subgrupos com mais acessos totais.

**Detalhe específico do objeto de estudo** – Para o estudo em questão é importante identificar os jogos mais acessados, os títulos dos jogos nesse caso podem se repetir pelo fato de não ter a data como parte integrante, por isso usamos o código do jogo o que só é coletado na análise de páginas.

### 7.3.5 Analise de Acessos.ipynb

**Entradas** – acessos.csv; acesso\_cidades.csv; acesso\_estado.csv; Engajamento.csv

**Saídas** – Gráficos de quantidade de acessos e engajamento.

**Objetivo** – Análise dos acessos gerais, por cidade e por estado, buscando correlações entre os acessos e eventos externos, além do tempo de engajamento atual, um dos objetivos do estudo é aumentar este tempo com informações dispostas de forma mais inteligente pelo novo sistema.

**Detalhe específico do objeto de estudo** – Não existe nenhuma análise especial voltada diretamente para o objeto de estudo neste programa.

### 7.3.6 Analise\_consulta\_zerada.ipynb

**Entradas** – referencia\_atualizada.csv;

**Saídas** – consultas\_menos\_acesso.csv; Gráficos de consultas zeradas ou com pouco acesso, grupos percentualmente com menos acesso.

**Objetivo** – Realizar a análise das consultas com pouco ou nenhum acesso, essa informação é base para a geração dos questionários e entrevistas.

**Detalhe específico do objeto de estudo** – Não temos nada específico neste programa.

### 7.3.7 Analise\_tecnologia.ipynb

**Entradas** – como\_acessa\_data.csv; onde\_acessa\_data.csv

**Saídas** – Gráficos de Dispositivos e ferramentas de busca utilizadas para iniciar acesso ao sistema legado.

**Objetivo** – Realizar a análise dos dispositivos e ferramentas de busca, importante avaliar para identificar nível de responsividade exigido e necessidade de maior exposição em redes sociais e outros meios para aumentar o acesso e engajamento.

**Detalhe específico do objeto de estudo** – Não temos nada específico neste programa.

### 7.3.8 Identificando\_links.ipynb

**Entradas** – biblioteca HTML do sistema legado

**Saídas** – Lista de links acessados pela tag a e lista de links acessados via url\_for.

**Objetivo** – Gerar lista inicial de referência caso não exista uma documentação robusta com essa informação, identificar os links existentes no sistema para ajudar na recomendação de criação de novos links caso seja identificado que são insuficientes.

**Detalhe específico do objeto de estudo** – Não temos nada específico neste programa.

### 7.3.9 Analise\_questionario.ipynb

**Entradas** – Arquivo com o compilado das respostas do questionário.

**Saídas** – Gráficos gerados a partir da análise das respostas e correlações identificadas.

**Objetivo** – Gerar gráficos com as distribuições das respostas, analisar correlações entre respostas, identificar pontos comuns de melhoria apontados pelos usuários, gerar insumo para o roteiro de entrevista.

**Detalhe específico do objeto de estudo** – Não temos nada específico neste programa.



## 8. Cenários de Sucesso

### 8.1 Primeiro cenário

#### **Situação do Sistema Legado**

- 1) O sistema legado possui documentação robusta, e cada página HTML possui um título único e que indica claramente seu conteúdo
- 2) Páginas sendo acessadas com códigos distintos, funções/classes separadas para cada grupo relacionado e tudo isso documentado de forma robusta e completa, criação do arquivo referencia.csv é uma tarefa simples diante de todo o cenário
- 3) Definição clara entre grupos e subgrupos.
- 4) Não existe a necessidade de se executar o programa de identificação de links para a geração do arquivo referencia.csv, ficando somente a necessidade de se executar o mesmo para análise de eficiência do uso de links.
- 5) Google Analytics ou análogo instalado e gerando todos os relatórios apresentados no Manual com sucesso.

#### **Desenvolvimento da Análise**

Análise transcorre sem problemas, respeitando as frequências de recuperação e análise de logs, tempo de exposição do questionário e entrevistas com usuários chave realizadas com sucesso.

#### **Resultados**

Sugestões realizadas, implementadas e avaliadas com sucesso, sistema legado modernizado e transformado em uma interface inteligente sem perda de nenhuma regra de negócio, trazendo conforto e melhorando a experiência do usuário no uso do sistema, aumentando produtividade, engajamento, exposição da marca, aumentando geração de receita.

## 8.2 Segundo cenário

### Situação do Sistema Legado

- 1) O sistema legado possui documentação robusta com títulos criados em todas as páginas e que tenham significado, porém alguns títulos podem ser duplicados e precisam ser explicados de uma outra forma.
- 2) Páginas sendo acessadas com códigos distintos, funções/classes separadas para cada grupo relacionado, porém sem documentação robusta, criação do arquivo referencia.csv não é possível diante do cenário encontrado.
- 3) Existe alguma definição entre grupos e subgrupos.
- 4) Com a execução o programa de identificação de links é possível fazer a geração do arquivo referencia.csv com uma boa acurácia sendo possível utiliza-lo para a análise de páginas.
- 5) Google Analytics ou análogo instalado e gerando todos os relatórios apresentados no Manual com sucesso.

### Desenvolvimento da Análise

Precisa ser gasto um tempo maior para a criação do arquivo de referencia.csv e de tempos em tempos é necessária uma revisão para a identificação de novas páginas adicionadas, identificação dos títulos duplicados e criação de lógica dentro dos notebooks para tratar o problema afastando qualquer ruído na análise dos logs, respeitando as frequências de recuperação e análise de logs, tempo de exposição do questionário e entrevistas com usuários chave realizadas com sucesso.

### Resultados

Com mínima interferência no resultado final as sugestões realizadas, implementadas e avaliadas com sucesso, sistema legado modernizado e transformado em uma interface inteligente sem perda de nenhuma regra de negócio, trazendo conforto e melhorando a experiência do usuário no uso do sistema, aumentando produtividade, engajamento, exposição da marca, aumentando geração de receita.

## 9. Cenários de Falha

### 9.1 Primeiro cenário

#### **Situação do Sistema Legado**

- 1) O sistema legado não possui documentação robusta algumas páginas não possuem título e outras que possuem tem títulos genéricos que não conseguem produzir uma distinção sobre o assunto que está sendo abordado na página/consulta.
- 2) Páginas sendo acessadas com códigos duplicados, funções/classes misturadas, impossível dividir os acessos em grupos, criação do arquivo referencia.csv não é possível diante deste cenário.
- 3) Indefinição de grupos e subgrupos de consultas/páginas
- 4) Mesmo com a execução do programa de identificação de links ainda não é possível a criação do arquivo referencia.csv.
- 5) Google Analytics ou análogo instalado e gerando alguns dos relatórios apresentados no Manual com sucesso, devido a falha de definição dos títulos.

#### **Desenvolvimento da Análise**

A definição clara dos títulos e a geração mesmo que incompleta de um arquivo de referência são as duas premissas para que a metodologia traga algum ganho e gere alguma sugestão, com o cenário acima a análise não é realizada.

#### **Resultados**

Análise não realizada, nenhuma sugestão é possível de ser realizada, o melhor em um caso como esse é partir para uma modernização tradicional onde alguma ferramenta extraia regras de negócio do sistema legado e o design é feito completamente do início, em um caso como esse não vale a pena investir em ajustar o sistema legado nem criar uma documentação.

## 9.2 Segundo cenário

### Situação do Sistema Legado

- 1) O sistema legado possui muitos títulos em duplicidade ou genéricos porém todas as páginas isso se aplica a todas as páginas/consultas tornando a identificação da navegação muito confusa porém possível.
- 2) Páginas sendo acessadas com códigos duplicados, funções/classes misturadas, criação do arquivo referencia.csv não é possível diante do cenário encontrado.
- 3) Existe alguma definição entre grupos e subgrupos.
- 4) Com a execução o programa de identificação de links é possível fazer a geração do arquivo referencia.csv com uma péssima acurácia tornando ele muito pouco eficiente na análise de páginas acessadas.
- 5) Google Analytics ou análogo instalado e gerando alguns dos relatórios apresentados no Manual com sucesso, devido a falha de definição dos títulos.

### Desenvolvimento da Análise

Alguma análise dos logs é feita porém os resultados são incapazes de trazer qualquer tipo de informação que possa ser utilizada para que seja feita uma sugestão de modernização com segurança.

### Resultados

A aplicação da metodologia não traz os resultados esperados e o projeto de modernização não pode ser concluído, a solução é fazer um projeto de modernização tradicional onde alguma ferramenta extraia regras de negócio do sistema legado e o design é feito completamente do início, em um caso como esse não vale a pena investir em ajustar o sistema legado nem criar uma documentação.