

# 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.....  | 6  |
| 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.....                                       | 9  |
| 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.....   | 10 |
| 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.....                                     | 11 |
| 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 .....                          | 14 |
| 6.1    | Coleta de Dados .....   | 14 |
| 6.2    | Processamento dos dados.....                                  | 16 |
| 6.2.1  | Pré-processamento.....  | 16 |
| 6.2.2  | Processamento .....   | 17 |
| 6.2.3  | Sugestões de Melhoria .....                                   | 17 |

|   |           |
|---|-----------|
| <b>7. Manual do usuário.....</b>                | <b>18</b> |
| <b>7.1 Premissas .....</b>                      | <b>18</b> |
| <b>7.2 Google Analytics .....</b>               | <b>19</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 tem muita importância para corporações e usuários, processos de modernização têm sido bastante explorados, pesquisas são realizadas e grande parte 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, o principal é 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 conseguem recuperar 100%, tudo isso traz muito desconforto 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 que apresenta vários problemas de funcionalidade.

Tudo isso nos traz a proposta desta nova 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 gerar sugestões de redesign do sistema legado tornando as interfaces estáticas em interfaces inteligentes sem alteração de código das regras de negócio.

## 2. Descrição e objetivos gerais do Software

A nova metodologia tem como objetivo sugerir novas ideias de design para os sistemas legados transformando os sistemas atuais em sistemas inteligentes mantendo as regras de negócio, para isso foram sugeridas 3 etapas de análise, a análise de logs, a criação de formulários que serão expostos aos usuários e a realização de entrevistas com usuários chave do sistema.

As etapas sugeridas estão encadeadas e os outputs de cada etapa são dados norteadores da etapa seguinte.

### 2.1 Análise de Logs

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

Para a realização desta tarefa precisamos de um programa coletor de dados da utilização do sistema, existem algumas opções no mercado e pode existir ou ser criada uma solução caseira, no estudo de caso conduzido utilizamos o Google Analytics.

A coleta de dados começou no dia 03/08/2023 e é necessário ao menos 1 mês de coleta de dados para iniciarmos uma análise consistente que possibilite gerarmos os primeiros inputs para a criação dos questionários, criação do roteiro das entrevistas e esboçar as primeiras impressões sobre 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 sendo realizado foi definida uma frequência de recuperação dos logs de 1 semana, dado importante nesta definição é verificar se a ferramenta de coleta de dados tem alguma limitação na extração dos dados, o Google Analytics tem uma limitação de 5000 linhas no export dos arquivos e isso tem sido levado em consideração no momento da definição da frequência.

## 2.2 Questionários

Após o prazo inicial de análise de logs temos dados suficientes para a criação de questionários que serão aplicados aos usuários de forma mais genérica, a intenção é identificar preferências no uso do sistema, identificar dificuldades que possam levar a não utilização de determinadas áreas do sistema, identificar dificuldades que afetem diretamente a experiência do usuário, medir a aceitação sobre ideias de sugestões já identificadas na análise de logs.

O tempo e o formato de aplicação do questionário dependem da natureza do sistema, dos usuários envolvidos, do prazo de entrega do projeto, para o objeto deste estudo o questionário ficou disponível por 1 mês e foram coletadas 42 respostas.

Informações de perfil dos usuários devem ser contempladas no início do questionário, elas serão úteis no momento de analisar as respostas, a disponibilidade de participação na etapa de entrevista deve estar presente no questionário sempre que necessário, alguns sistemas legados podem ter usuários chave definidos o que torna essa pergunta desnecessária, já para o caso de sites abertos ao público essa informação é crucial.

## 2.3 Entrevistas

Tendo como input as análises de log e informações de uso vindas dos questionários podemos estabelecer um roteiro de entrevistas, essas dinâmicas devem ocorrer com usuários chave do sistema, pessoas que vão contribuir com informações relevantes, o roteiro deve focar principalmente nas áreas de sombra do sistema, quais os problemas encontrados no acesso destas áreas, porque usuários evitam acessar essas áreas, quais as maiores dificuldades no acesso ao sistema, quais as maiores chances de melhoria, novas ideias, sempre visando melhorar a experiência de acessar o sistema.

## 2.4 Sugestões

Com todas as etapas finalizadas e com as informações todas 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 a celular, tablets e desktops.

### 3. Objeto de estudo

Para realizar a pesquisa foi escolhido o site de estatísticas de futebol do Fluminense Football Club, ele está escrito em Python/Flask, Javascript, HTML, Bootstrap, o acesso ao site é aberto, sem necessidade de login, o que prejudica a verificação de preferências individuais, o design é estático sendo necessário 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(Combo), algumas consultas podem ficar escondidas em áreas do site que para o usuário final pode não ter nenhuma relação, algumas áreas do site são pouco acessadas e fornecem visualizações e informações muito interessantes, algumas consultas 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 podem ser melhor sinalizados.

O site já passou por 2 grandes reformulações, inicialmente 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 a linguagem de programação passou para o Python/Flask, todas as consultas foram migradas e com a facilidade de trabalhar com datasets da linguagem Python novas funcionalidades foram incluídas, o frontend sofreu um grande reformulação e o site é todo responsivo, porém testes mais completos em todos os tamanhos de tela não foram conduzidos.

Com a ajuda da metodologia de análise de logs para sistemas legados vai acontecer a terceira grande reformulação, com foco total no design de telas, as regras de negócio permanecerão sem alteração e serão linkadas ao novo design, as sugestões feitas serão codificadas em Python/Flask.

O código relacionado ao objeto de estudo é de uso particular e não vai estar presente no repositório relacionado ao trabalho, os dados levantados referentes ao uso, acesso, links 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 distinção 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 varrer os HTMLs existentes para identificação de links via a 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** – Escalabilidade – 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 pré-definido, abaixo estão listados os arquivos de entrada e seus formatos e a origem de cada coluna, alguns campos que precisam ser formatados na etapa de pré-processamento que está explicada no Manual de utilização, estes campos serão gerados automaticamente na análise de logs na próxima versão dos programas.

### 5.1 Modelo de Dados

#### 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 retorno do questionário

Esse arquivo não tem um formato definido, depende das perguntas que serão feitas

### 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 baixadas as bibliotecas Pandas, Plotly, Numpy, não existe um frontend de comunicação, ficando a cargo do pesquisador/desenvolvedor executar qual notebook mais lhe convém de forma manual.

### 5.2.2 Melhorias

Juntar todos os códigos, hoje feitos no formato de jupyter notebooks, e criar uma área administrativa para a definição dos caminhos de onde guardar os gráficos de saída, dos caminhos para os arquivos de entrada, datas de corte da análise, todos esses campos seriam variáveis globais para a execução em todos os programas da análise de logs.

Uma opção onde seja possível escolher qual programa executar e assim fazer execuções personalizadas, gerando os resultados de seu interesse.

Construção das rotinas de geração das datas automáticas a partir da coluna Extenso, gerada pela Google Analytics, desta forma será possível eliminar etapas do pré-processamento, conforme o que está descrito no Manual.

## 6. Descrição Funcional do Sistema

A metodologia é composta de 3 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.

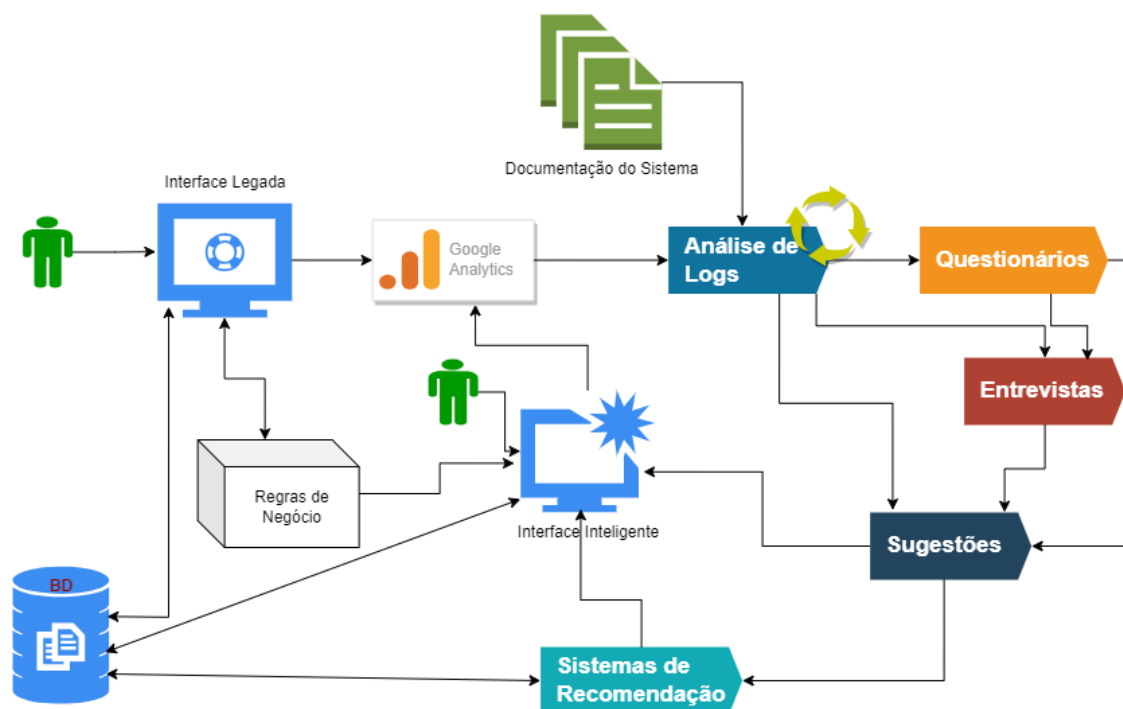


Figura 1 - Metodologia Proposta

### 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 foi escolhido o Google Analytics.

Uma vez escolhido o melhor 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, 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 e desta forma deve ser avaliada qual frequência melhor atende ao sistema legado, para o estudo sendo realizado foi determinada uma frequência semanal para todos os relatórios.

Deve ser determinado o período inicial de análise de logs, deve ser determinado pela equipe de desenvolvedores, após esse período 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 a melhor ferramenta, para o estudo sendo realizado foi escolhido o Google Forms.

A aplicação do questionário também vai variar de sistema a sistema, podemos fazer a execução da tarefa enviando nominalmente e coletando os resultados ou pode ser disponibilizado um link dentro do sistema que vai permanecer por um período para quem quiser responder, importante é termos uma quantidade de respostas que seja possível avaliar e trazer 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, foram 42 respostas, os resultados foram extraídos da plataforma Google e trabalhados em um notebook específico.

Por fim temos a etapa de Entrevistas onde usuários chave serão convidados a conversar sobre detalhes de navegação, entendimento de informações mostradas na tela e funcionamento da mesma, propostas de melhoria, o que deveria existir no sistema que não temos hoje e demais assuntos que sejam importantes e pertinentes para o sistema legado, não existe uma ferramenta específica ficando a cargo de cada desenvolvedor coletar esses dados da melhor forma que for possível, esses dados serão trabalhados em um notebook específico.

No estudo do site as entrevistas serão feitas com base em um roteiro criado tendo o questionário e a análise de logs como fontes primárias, não será utilizada nenhuma ferramenta específica, as entrevistas serão gravadas e as respostas compiladas em um dataset que será 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 2 fases, pré-processamento dos dados e outra automática fazendo uso dos programas criados para esta análise, existem 7 programas de análise de logs porém a metodologia é flexível, somente o que for necessário deve ser executado, a criação de novos programas de análise tende a ser facilitada, as nomenclaturas de campos, os seus tipos são iguais e 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(Google Analytics) com um dataset principal, que tem todos os datasets de análise juntos, caso exista a coluna de data sem formatação ainda é necessário fazer a criação manual de uma coluna de data formatada, já está em curso uma melhoria para a geração automática desta data nos programas de análise, 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 no estudo sendo realizado foi criada uma macro no aplicativo Notepad++ que realiza ajustes de acentuação e transforma a virgula (,) em ponto e vírgula (;), para facilitar acesso a esses datasets via o aplicativo Excel.

Para finalizar o pré-processamento os arquivos coletados e o arquivo completo são abertos no excel e o conteúdo do parcial é concatenado no completo, 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 já 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 9 programas para análise de logs, um único deles é utilizado para recuperação de tags dentro de programas HTML e não faz parte do ciclo de programas de análise de logs, este programa é executado no início do ciclo de vida da modernização com a intenção de listar todos os links existentes no sistema.

Uma melhoria já identificada e que será executada assim que possível será a criação de um programa que identifique todas as tags TITLE dentro dos HTMLs do sistema legado para trazer um mapeamento dos títulos das páginas que será extremamente importante para uma eficiente análise de logs, para o caso do estudo em andamento, o sistema legado já tinha os títulos todos ajustados e com sentido, desta forma a identificação dos títulos via programa não foi necessária mas para a metodologia e o seu sistema acoplado isso será bastante útil.

O processamento dos datasets é independente entre si, as únicas 2 dependências listadas são a análise de páginas consultadas e a análise de consultas zeradas, 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.

Dentro dos programas existem datas de início e fim da análise permitindo a limitação de um período de tempo para a geração das informações e os gráficos e datasets de saída são salvos com essas datas setadas fazendo parte do nome para permitir avaliar de forma incremental os dados coletados, uma sugestão de melhoria no processo ainda bem incipiente é criar um mecanismo 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 em cima dos dados coletados nas três fases do projeto, deve ser gerado um relatório com a sugestão e uma explicação com base nos dados com o motivo que levaram o desenvolvedor a entender que a melhoria proposta vai realmente trazer algum ganho na experiência do usuário.

## 7. Manual do usuário

### 7.1 Premissas

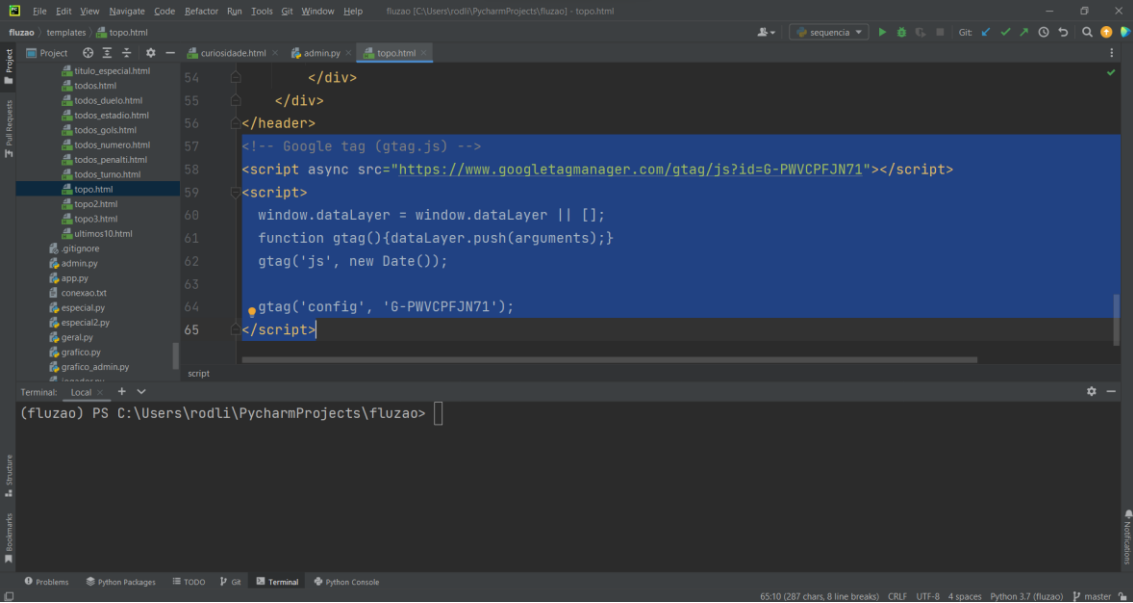
Para o melhor aproveitamento da metodologia de redesign de sistemas legados pela análise de logs algumas premissas devem ser atendidas, o não cumprimento destas premissas não invalida o processo, porém o torna bem mais complexo e difícil de chegar a uma sugestão robusta de melhoria.

- 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 do 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 invalida as demais análises. Um pedaço deste 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, isso ajuda na análise das páginas acessadas verificando os acessos em diferentes áreas de atuação, identificando quais pontos precisamos agir para que todo o sistema seja visualizado, trazendo mais informações para o usuário.
- 5) Para fins deste estudo os códigos referentes ao objeto sendo estudado são particulares e não farão parte deste manual, os códigos referentes a metodologia de redesign de sistemas legados através da análise de logs estão no GIT [https://github.com/RodLima26/Analise\\_legacy](https://github.com/RodLima26/Analise_legacy)

## 7.2 Google Analytics

Primeiro passo a ser realizado é identificar qual ferramenta de coleta de dados será utilizada, a ferramenta indicada pela metodologia é o Google Analytics, custo zero, fácil implementação e grande contribuição na coleta de dados em todos os níveis do sistema.

Para ter acesso a ferramenta é necessário ter uma conta na Google, depois procurar pelo Google Platform e depois Google Analytics, seguir o procedimento de cadastro e ao final pegar o código HTML sugerido e colocar em um ou mais HTMLs do sistema legado de forma a cobrir toda a navegação realizada pelo sistema, para efeito de comparação 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.



```
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>
```

**Figura 2 - Código para funcionamento do Google Analytics**

Após a publicação do sistema com a tag do 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 pela ferramenta, quanto maior o número de acessos menor deverá ser a frequência.

O Google Analytics já fornece um bom número de relatórios em sua versão inicial, segue abaixo a lista com os locais onde extrair os relatórios, 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 essa criação o Google analytics fornece um passo a passo bem detalhado e intuitivo, 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

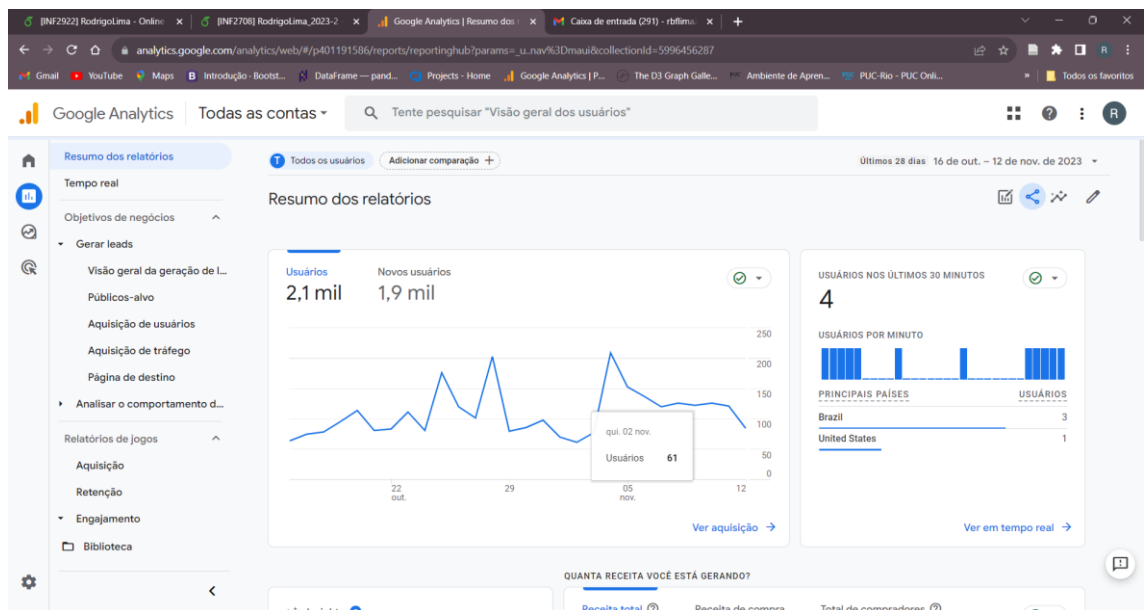


Figura 3 - Relatório Inicial do Google Analytics

## 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

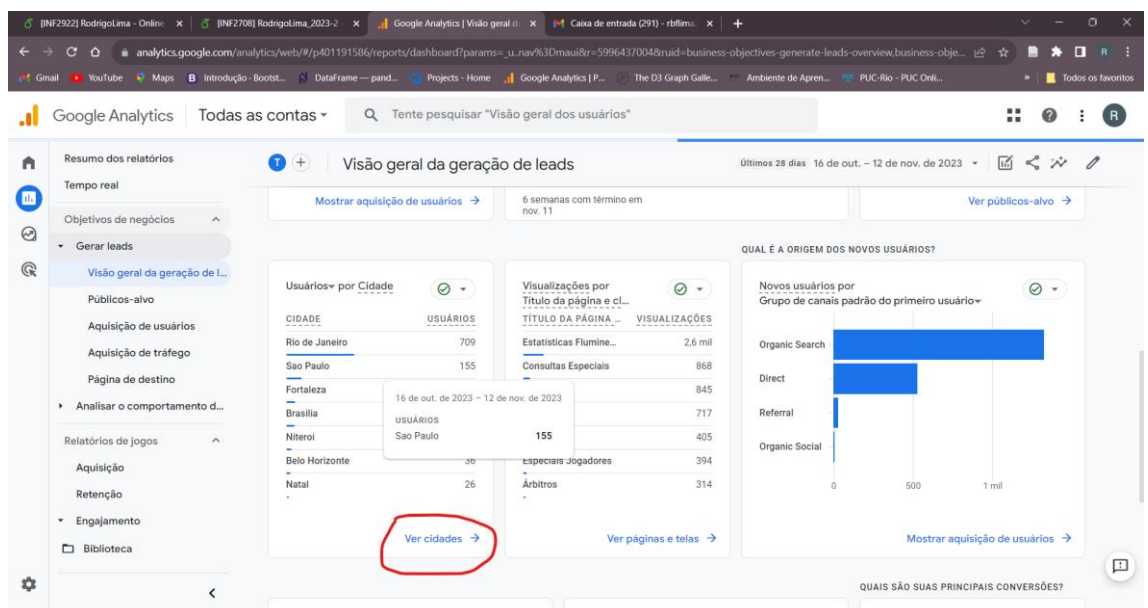


Figura 4 - Relatório de acessos por cidade

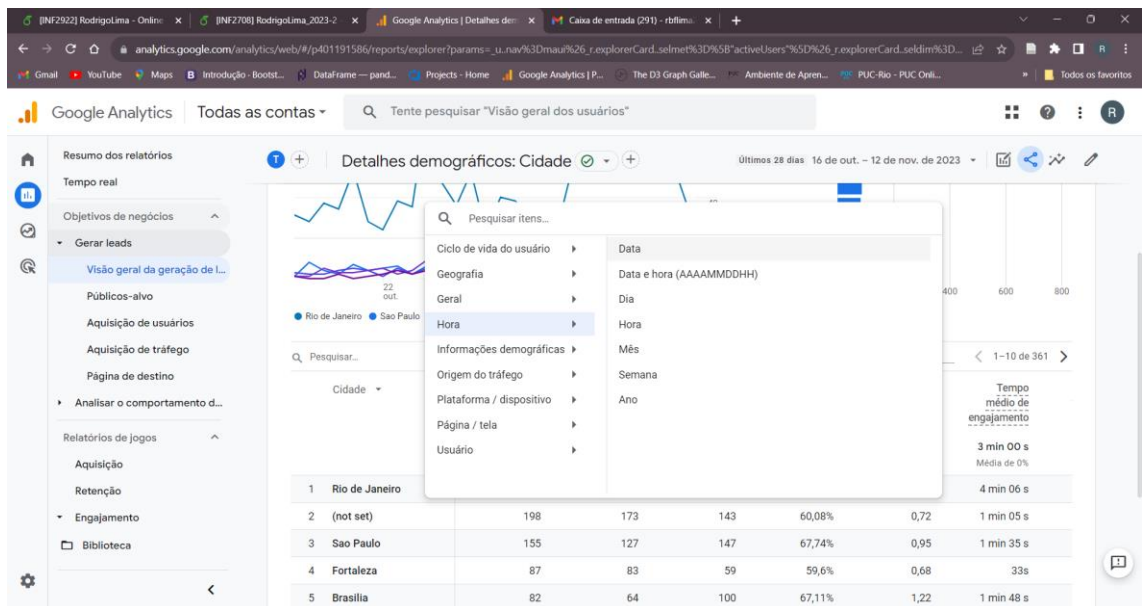


Figura 5 - Incluindo a coluna data

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

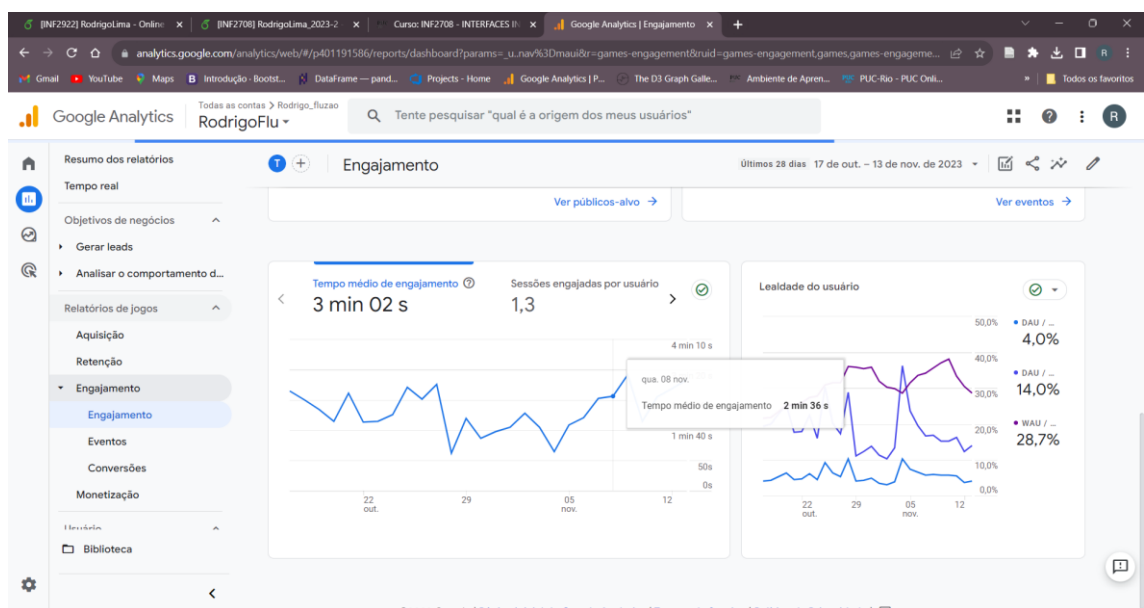


Figura 6 - Relatório de Engajamento

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

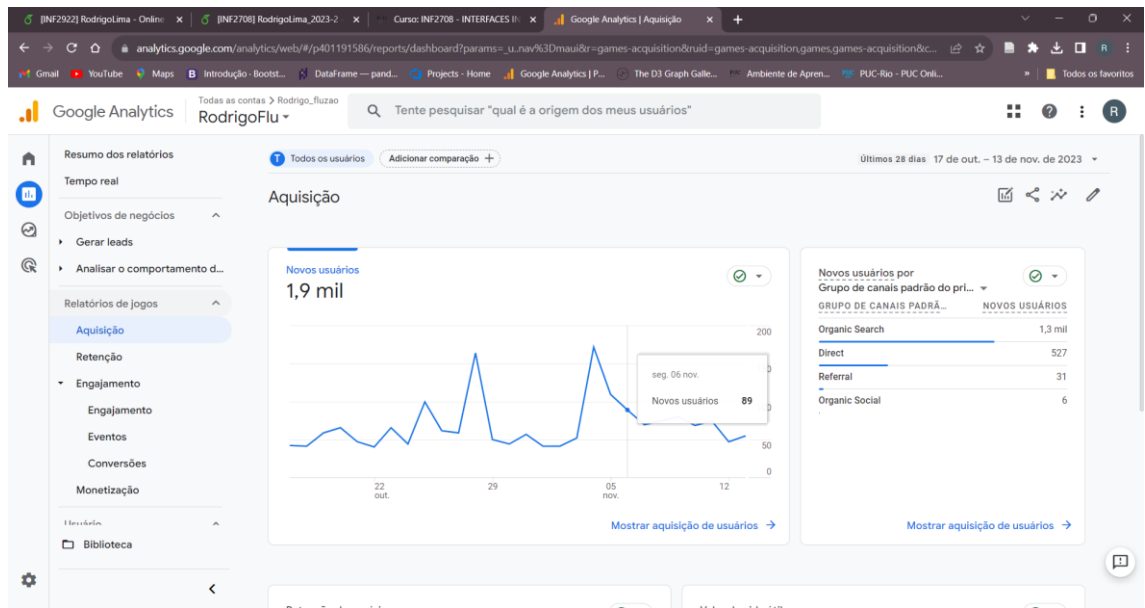


Figura 7 - Relatório de como iniciou o acesso ao sistema

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

The screenshot shows the Google Analytics 'Detalhes de tecnologia: Categoria da plataforma / dispositivo' (Technology Details: Platform / Device Category) report. The table displays user data for various platforms and devices over time.

|   | Categoria da pl. / dispositivo | Data     | Usuários               | Novos usuários         | Sessões engajadas      | Taxa de engajamento   | Sessões engajadas por usuário |
|---|--------------------------------|----------|------------------------|------------------------|------------------------|-----------------------|-------------------------------|
|   |                                |          | 2.162<br>100% do total | 1.909<br>100% do total | 2.916<br>100% do total | 70,47%<br>Média de 0% | 1,35<br>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

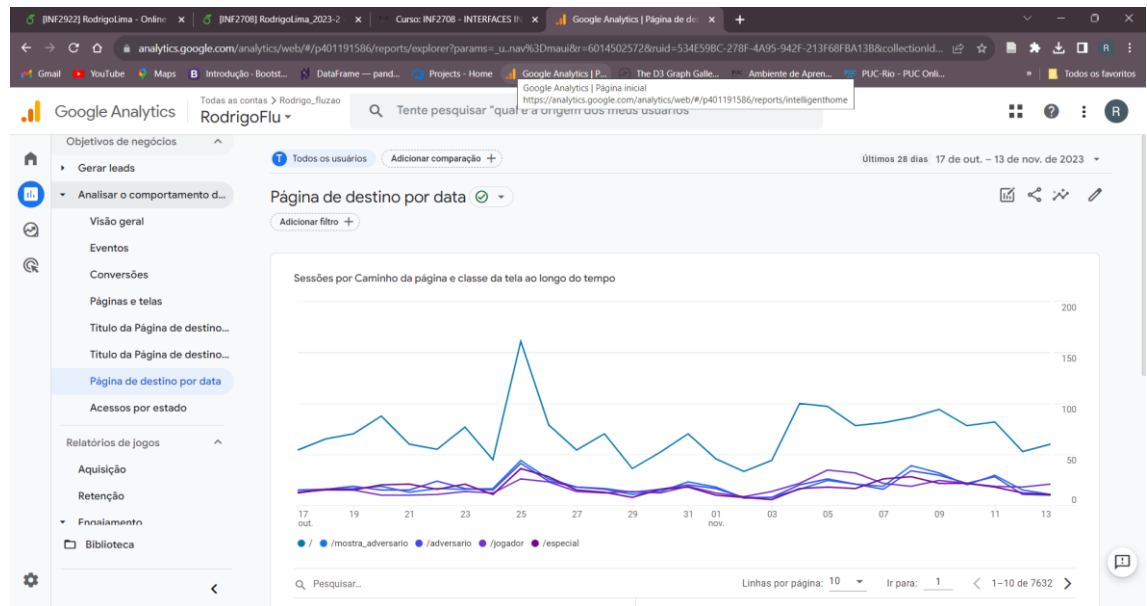


Figura 9 - Relatório de Páginas e consultas por data

## 7) Título das páginas por data

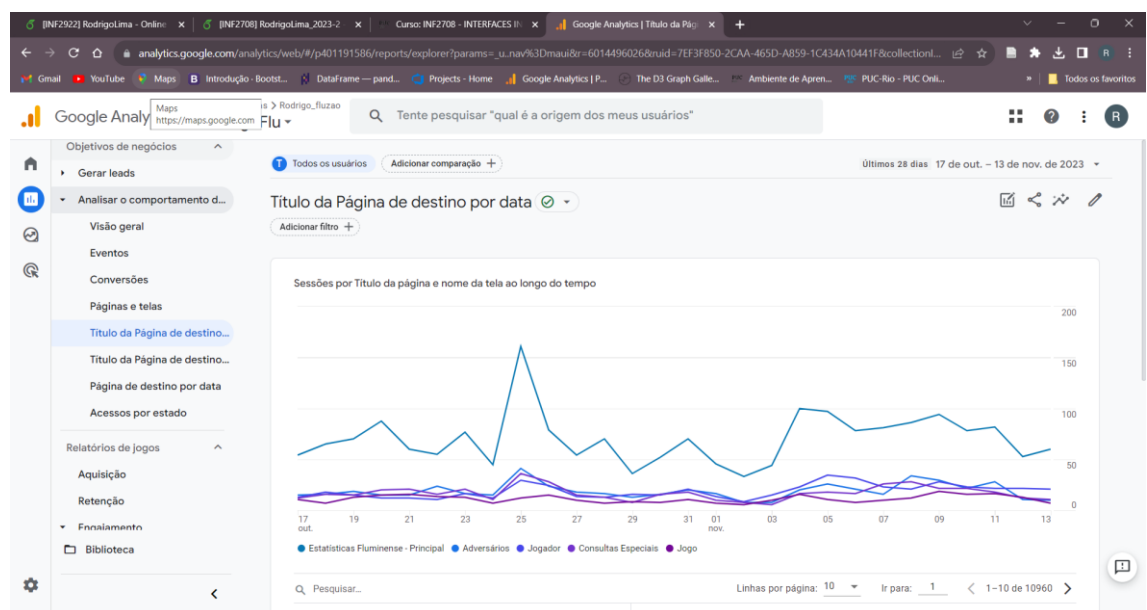


Figura 10 - Relatório de títulos acessados por data

8) Título das páginas por data e hora

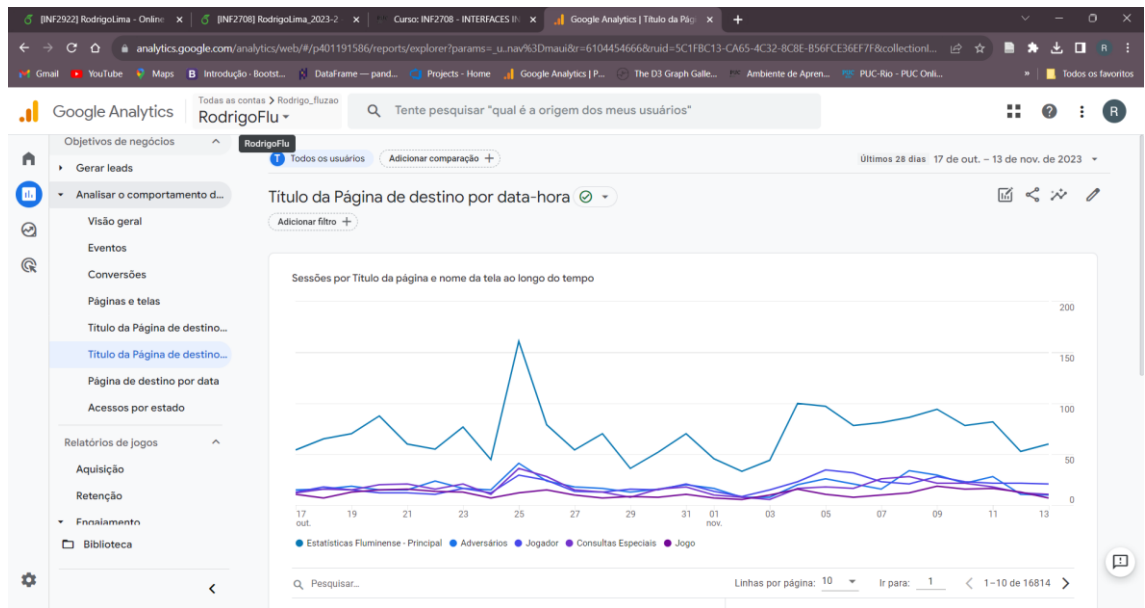


Figura 11 - Relatório de títulos acessados por data e hora

9) Acesso por estado

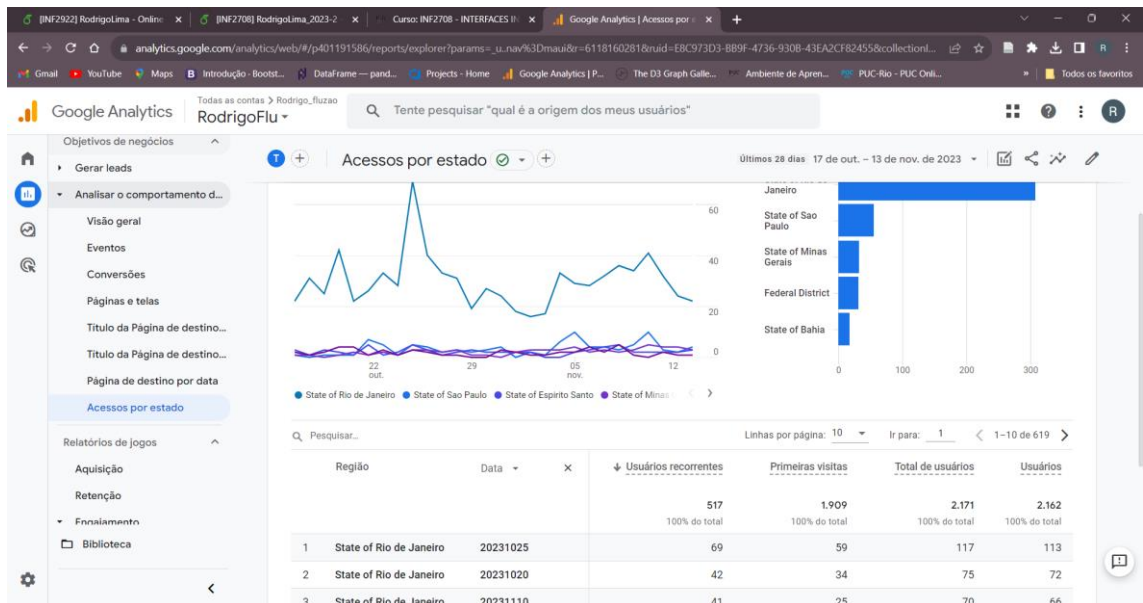


Figura 12- Relatório de acessos por estado



Os relatórios de Páginas e Títulos, itens 6, 7 e 8 da lista acima não são fornecidos inicialmente pelo Google Analytics e precisam ser criados, 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 sobre o que os dados se referem e as datas de corte, recomendo que sejam guardados dentro de uma pasta Logs do projeto principal para futura referência ou auditoria.

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

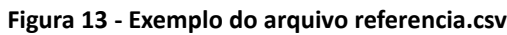
Após a formatação os arquivos devem ser abertos e concatenados com os seus respectivos arquivos completos que vão estar armazenados na pasta Arquivos, a formatação do nome deve ser igual a dos logs sem a data de corte, isso facilita a correlação para a concatenação.

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, ainda é realizado manualmente, já existe uma melhoria listada para fazer a transformação automaticamente na execução das análises.

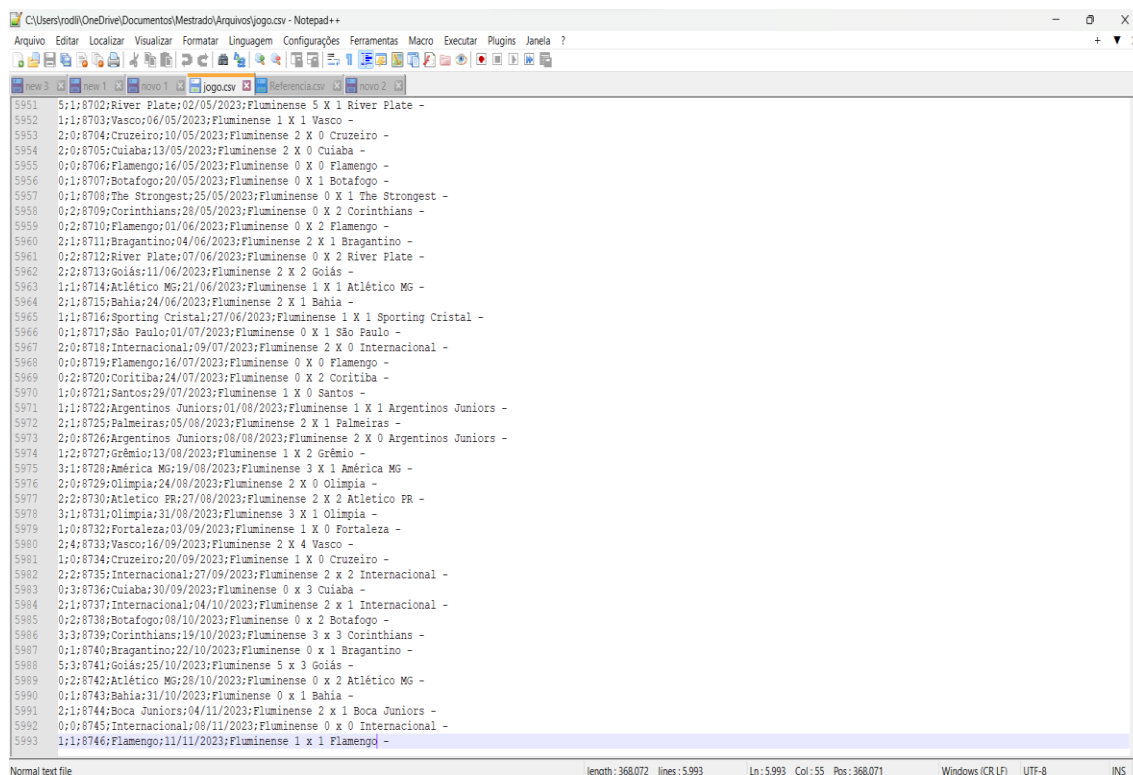
Ao final deste processo são gerados os arquivos de entrada para as análises, os arquivos CSV gerados podem ser salvos em uma cópia XLSX para melhor documentação, porém isso é opcional.

- 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

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 do que deve ser populado neste arquivo.



O arquivo abaixo jogo.csv também foi gerado para auxiliar a análise dos logs relacionados aos jogos de futebol mais acessados da história, segue abaixo um exemplo do arquivo, este arquivo é exclusivo para a análise do site Fluzao.xyz.



```
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 2:1:8723:Palmeiras;05/08/2023;Fluminense 2 X 1 Palmeiras -
5973 2:0:8724:Argentinos Juniors;08/08/2023;Fluminense 2 X 0 Argentinos Juniors -
5974 1:2:8725:Grêmio;13/08/2023;Fluminense 1 X 2 Grêmio -
5975 3:1:8726:América MG;19/08/2023;Fluminense 3 X 1 América MG -
5976 2:0:8727:Olimpia;24/08/2023;Fluminense 2 X 0 Olimpia -
5977 2:2:8728:Atletico PR;27/08/2023;Fluminense 2 X 2 Atletico PR -
5978 3:1:8729:Olimpia;31/08/2023;Fluminense 3 X 1 Olimpia -
5979 1:0:8730:Fortaleza;03/09/2023;Fluminense 1 X 0 Fortaleza -
5980 2:4:8731:Vasco;16/09/2023;Fluminense 2 X 4 Vasco -
5981 1:0:8732:Cruzeiro;20/09/2023;Fluminense 1 X 0 Cruzeiro -
5982 2:2:8733:Internacional;27/09/2023;Fluminense 2 X 2 Internacional -
5983 0:3:8734:Cuiaba;30/09/2023;Fluminense 0 X 3 Cuiaba -
5984 2:1:8735:Internacional;04/10/2023;Fluminense 2 X 1 Internacional -
5985 0:2:8736:Botafogo;08/10/2023;Fluminense 0 X 2 Botafogo -
5986 3:3:8737:Corinthians;19/10/2023;Fluminense 3 X 3 Corinthians -
5987 0:1:8738:Bragantino;22/10/2023;Fluminense 0 X 1 Bragantino -
5988 5:3:8739:Goias;25/10/2023;Fluminense 5 X 3 Goias -
5989 0:2:8740:Atlético MG;28/10/2023;Fluminense 0 X 2 Atlético MG -
5990 0:1:8741:Bahia;31/10/2023;Fluminense 0 X 1 Bahia -
5991 2:1:8742:Boca Juniors;04/11/2023;Fluminense 2 X 1 Boca Juniors -
5992 0:0:8743:Internacional;08/11/2023;Fluminense 0 X 0 Internacional -
5993 1:1:8744:Flamengo;11/11/2023;Fluminense 1 X 1 Flamengo -
```

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 cortar os datasets entre as datas escolhidas e fazer uma análise mais precisa de um período de tempo, essas análises podem evidenciar o comportamento de acesso ao sistema reagindo a eventos externos e trazendo valiosa contribuição para as sugestões de ajuste, criação de sistema de recomendação.

A única exceção a regra de ordem de execução se encontra nas pesquisas em que o arquivo de referência (referencia.csv) for criado, neste 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 somente uns ajustes de formato, outra opção é gerar uma primeira versão executando o programa Identificando\_links.ipnyb, este programa identifica todos os links chamados dentro dos HTMLs do sistema 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, fornecendo dados que serão explorados nos questionários, entrevistas e mais importante na formulação de sugestões de melhoria, ajudam 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 ou pode ser deixado em branco mesmo com 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.

Uma breve análise de cada programa criado, entradas, saídas e objetivo principal.

#### 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 com títulos únicos criados em todas as páginas e que tenham significado
- 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.