

Documentação Técnica de Projeto de Pesquisa

Chamada	Projeto de Pesquisa - Edital nº 03/2021		
Programa	FAP - Fundação de Apoio a Pesquisa		
Termo de Outorga	Nº 510/2021	Número do Protocolo	00193-00001045/2021-88
Nome do Outorgado	Claynor Fernando Mazzarollo	Período	04/2022 - 04/2023
Título do Projeto	Implantação de Ambiente de BI como Serviço para Apoio às Atividades Gerenciais da Saúde Pública do GDF		
Instituição	IBTI - Instituto Brasília de Tecnologia e Informação		
Área de Conhecimento	Business Intelligence		

Descrição do projeto

O objetivo é promover o uso gerencial da informação em serviços de saúde para a melhoria da qualidade da atenção à saúde no Distrito Federal.

Envolve, fundamentalmente, a elaboração de serviços mediadores e de transformação para realizar uma integração dos conceitos de forma a atender não só a agregação / consolidação de dados em regime de inteligência de negócios (Business Intelligence - BI) - mas também atender a necessidades de saídas gerenciais - bem como suportar relações de transformação de dados visando a coleta e uso em sistemas locais de dados oriundos do Ministério da Saúde; permitindo assim que os elementos da arquitetura sejam fortemente desacoplados, visando o uso de estruturas díspares para atender à solução em ambientes de computação em nuvem.

Assim, como objetivo principal, o projeto proposto visa atender a disponibilização de informações gerenciais de forma consistente, prevendo suportar a discrepância entre os modelos canônicos de dados entre elementos partícipes do sistema de saúde e os sistemas existentes junto aos serviços que ora estão sendo ofertados pelo Ministério da Saúde.

Para entendimento, o modelo de dados canônico refere-se à prática de definir esquemas únicos para representar tipos globalmente, e reusar estes tipos referenciando-os nos esquemas projetados localmente para o serviço, isto é, integrando os modelos entre os diversos partícipes, podendo utilizar as nomenclaturas para operadoras governamentais e não-governamentais de saúde do DF.

Introdução: Uma breve descrição do projeto, incluindo seus objetivos e sua relevância para a empresa ou organização.

Documentação Técnica de Projeto de Pesquisa

Conjunto de Dados

A fonte primária de dados é o servidor FTP do DataSUS, que pode ser acessado utilizando através do link <ftp.datasus.gov.br> utilizando ferramentas de conexão de servidores.

O servidor é onde são publicados todos os dados da saúde pública do Brasil, incluindo os que foram utilizados neste projeto. Esses dados são disponibilizados em arquivos comprimidos no formato DBC devido a grande quantidade de informações.

A coleta, a limpeza e o pré-processamento dos dados são realizados automaticamente através do pipeline ETL incluído no pacote “IBTIsus”, criado durante o projeto utilizando linguagem de programação R com o intuito de facilitar o download e execução da aplicação.

Além do pipeline ETL, o pacote também pode organizar os dados em diferentes zonas lógicas, sendo elas Raw, Curated e Refined. Cada pasta é utilizada para uma etapa do processo.

A etapa de coleta consiste em acessar o servidor do DataSUS para fazer download das bases de dados selecionadas e convertê-las do formato DBC para o formato CSV, a fim de facilitar a limpeza e o tratamento das mesmas. Os dados nos formatos DBC e CSV são armazenados na zona lógica Raw para serem utilizados na próxima etapa.

Após a coleta das informações, é feita a limpeza dos dados. Colunas que contenham informações irrelevantes ou não preenchidas são descartadas. Posteriormente, os dados limpos são armazenados na zona lógica Curated para serem utilizados na próxima etapa.

A fase de pré-processamento é realizada após a limpeza e tem como objetivo aprimorar as bases de dados, fazendo o preenchimento das colunas com informações faltantes, convertendo variáveis, entre outros. Por fim armazenando os dados tratados na zona lógica Refined.

Após a conclusão do pipeline, os dados são enviados para o banco de dados PostgreSQL para serem acessados pelo Power BI, ambiente de Business Intelligence escolhido para apresentar os dados de maneira gráfica.

Simplificando, os dados passam pelo seguinte fluxograma:



Documentação Técnica de Projeto de Pesquisa

Guia do Usuário

Execução do Código

Primeiramente é necessário fazer a instalação do pacote IBTIsus e executar o código para realizar o processo de ETL e envio das bases para o banco de dados.

A instalação é feita a partir do seguinte código:

```
> install.packages("remotes")  
> remotes::install_github("IBTI-DF/SIGFAPDF/IBTIsus")
```

O uso do pacote depende de, basicamente, uma única função: o pipeline automático de coleta e tratamento dos dados.

Exemplo:

```
library(IBTI)  
pipeline(pasta = getwd(), periodo = 0)
```

Onde o argumento `pasta` é o diretório onde será criado o Data Lake (se não especificado, o Data Lake será criado no diretório vigente do R). E o argumento `periodo` é referente à data de atualização dos dados (nesse caso, ao se especificar uma data, a função só fará o download dos dados atualizados na data definida), ao não se definir uma data, será realizado o download de todos os dados do DataSUS.

Esta etapa, se for de preferência do usuário, pode ser automatizada para ser executada periodicamente. Assim, os dados se mantêm atualizados sem necessidade de execução manual.

Após o envio das tabelas para o banco de dados, o Power BI as identifica e inclui nos gráficos já criados para análises aprofundadas.

Conexão com o Banco de Dados

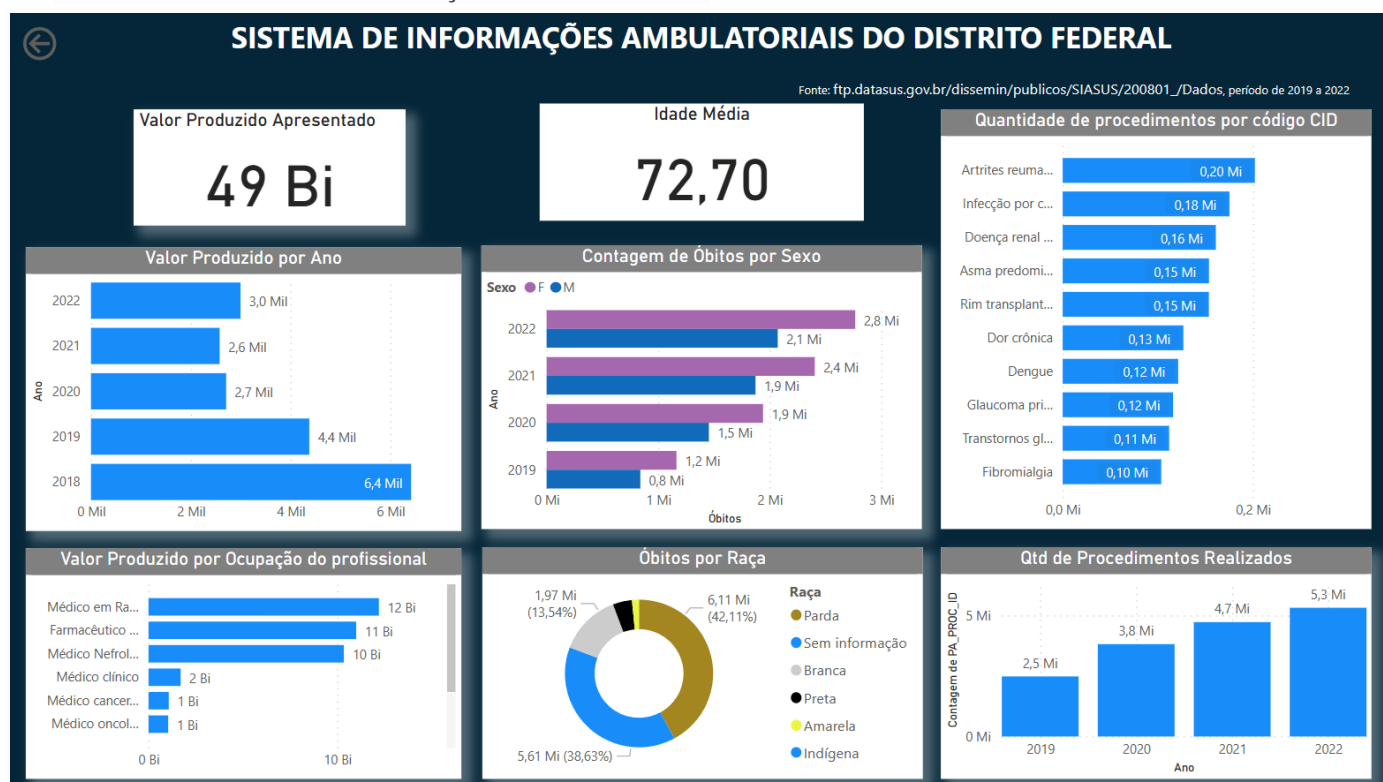
A conexão entre o banco de dados é feita através do Power BI. Após a criação do banco e upload dos dados, é necessário realizar a conexão com o host através da ferramenta Power BI. Dessa forma, os gráficos gerados serão abastecidos com os dados baixados.

Documentação Técnica de Projeto de Pesquisa

Dashboards

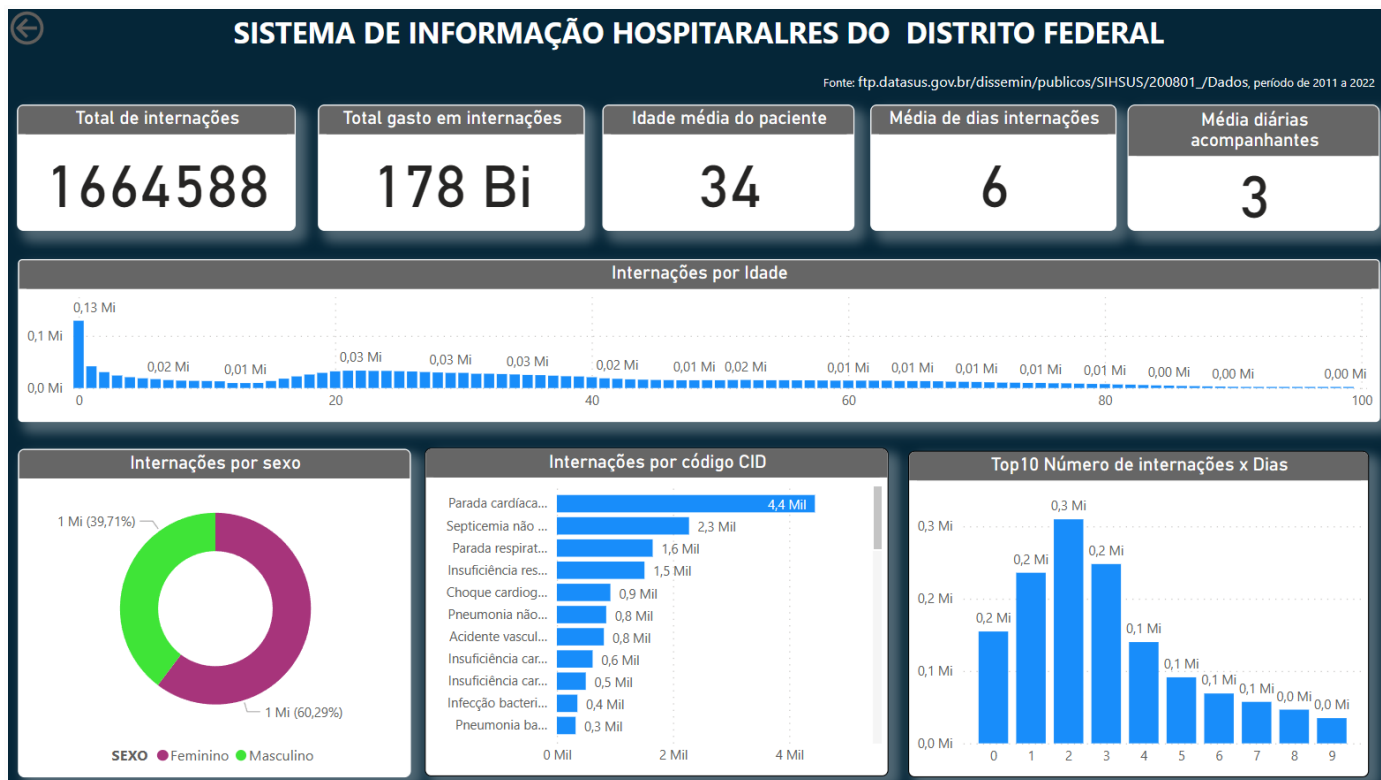
Para a visualização dos dados, foram criados cinco painéis gráficos no PowerBI. Após o processo de coleta e limpeza dos dados, foram selecionadas as métricas e filtros relevantes para cada área, são eles:

- SIA - Sistema de Informações Ambulatoriais

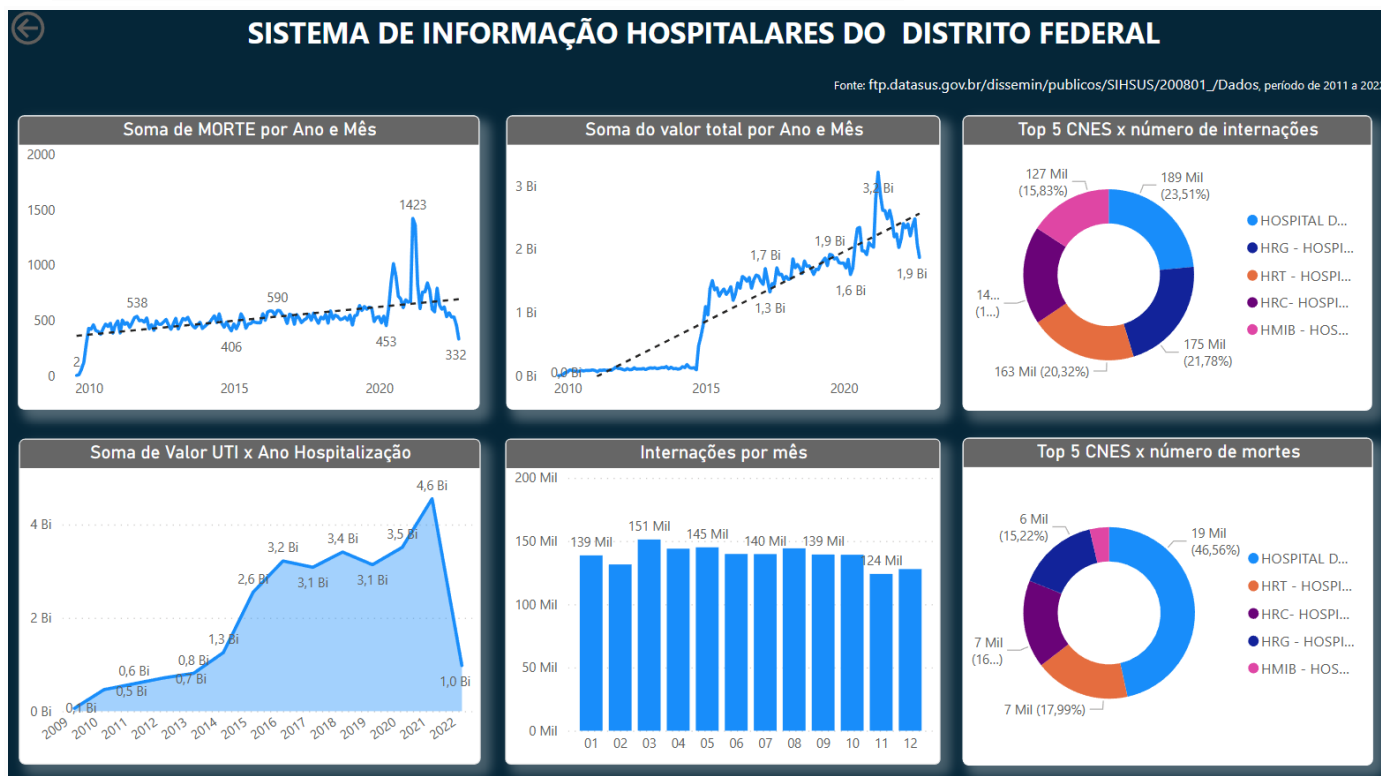


Documentação Técnica de Projeto de Pesquisa

- SIH - Sistema de Informações Hospitalares

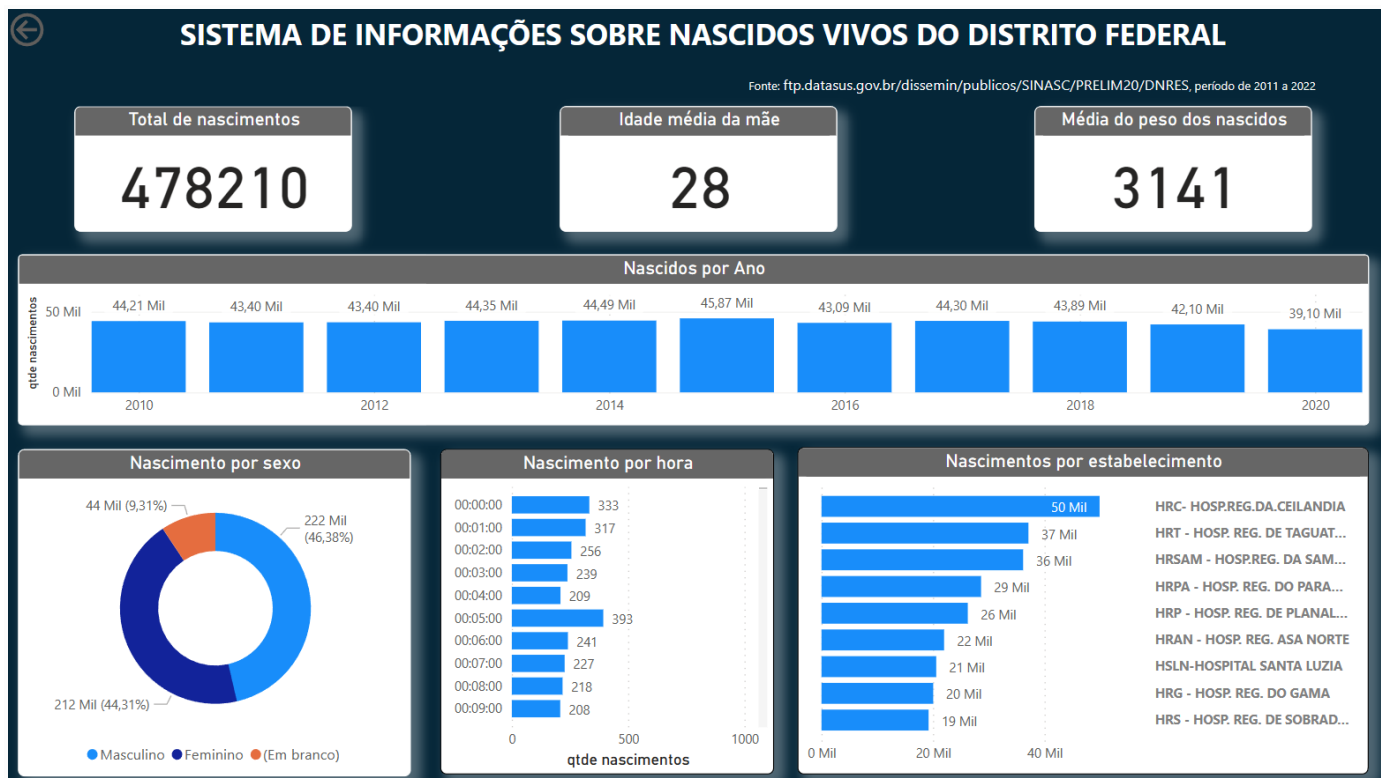


- SIH - Indicadores do Sistema de Informações Hospitalares

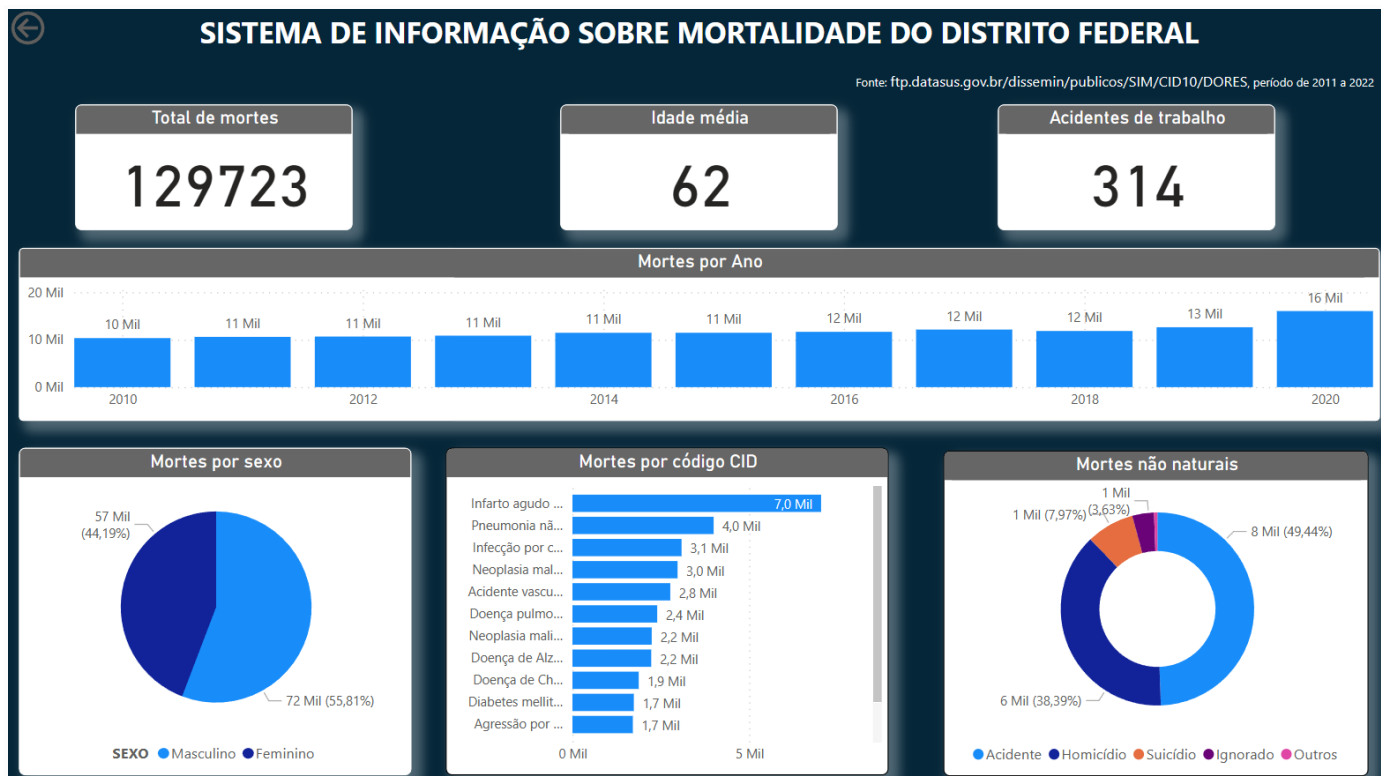


Documentação Técnica de Projeto de Pesquisa

- SINASC - Sistema de Informações sobre Nascidos Vivos



- SIM - Sistema de Informações sobre Mortalidade



Todos os gráficos são interativos e detalham informações importantes da saúde do Distrito Federal.

Documentação Técnica de Projeto de Pesquisa

Conclusão

Após a finalização de todos os passos citados anteriormente, será possível analisar os dados da saúde disponibilizados pelo DataSUS de maneira gráfica, com painéis personalizados e divididos de acordo com a área de especialidade.

As informações extraídas poderão gerar insights inovadores para a Gestão de Saúde do DF além de trazer informações atualizadas periodicamente.