

UNIVERSIDADE FEDERAL DE SANTA MARIA
COLÉGIO POLITÉCNICO
CURSO DE TECNOLOGIA EM SISTEMAS PARA INTERNET

Rodrigo Balest Pinto de Oliveira

**ESTUDO SOBRE O USO DE FERRAMENTAS DE INTELIGÊNCIA DE
NEGÓCIOS PARA ANÁLISE DE DADOS ABERTOS DA ÁREA DA
SAÚDE**

Santa Maria, RS
2022

Rodrigo Balest Pinto de Oliveira

**ESTUDO SOBRE O USO DE FERRAMENTAS DE INTELIGÊNCIA DE NEGÓCIOS PARA
ANÁLISE DE DADOS ABERTOS DA ÁREA DA SAÚDE**

Trabalho de Conclusão de Curso apresentado ao Curso Superior de Tecnologia em Sistemas para Internet da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do grau de **Tecnólogo em Sistemas para Internet**.

Orientador: Prof. Dr. (UFSM) Daniel Lichtnow

Ficha gerada com os dados fornecidos pelo autor

Rodrigo Balest Pinto de Oliveira

**ESTUDO SOBRE O USO DE FERRAMENTAS DE INTELIGÊNCIA DE NEGÓCIOS PARA
ANÁLISE DE DADOS ABERTOS DA ÁREA DA SAÚDE**

Trabalho de Conclusão de Curso apresentado ao Curso Superior de Tecnologia em Sistemas para Internet da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para a obtenção do grau de **Tecnólogo em Sistemas para Internet**.

Aprovado em 09 de fevereiro de 2022:

Daniel Lichtnow, Dr. (UFSM)
(Presidente/Orientador)

Marcos Alexandre Rose Silva, Dr. (UFSM)

Leandro Oliveira Freitas, Dr. (UFSM)

DEDICATÓRIA

À minha mãe Beni, pelo esforço que sempre nos dedicou.

À minha esposa Mara e meu filho Eduardo, que são a força que me impulsiona a sempre ir em frente.

AGRADECIMENTOS

Agradeço aos professores do curso de Tecnologia em Sistemas para Internet que, com paciência e dedicação, buscaram ensinar não somente a mim, mas todos os alunos do curso da melhor forma possível, mesmo com as dificuldades e restrições impostas pela pandemia.

Agradeço ao Professor Dr. Daniel Lichtnow pelo apoio e pelas orientações não somente durante o desenvolvimento deste trabalho, mas durante todo o curso.

Agradeço ao Emerson Alexandre Mortari, Chefe do Setor de Gestão de Processos e Tecnologia da Informação do Hospital Universitário de Santa Maria, pelo apoio e incentivo dados desde antes mesmo da minha inscrição para uma vaga neste curso.

Agradeço à minha família, que muitas vezes abriu mão de momentos de lazer, atenção e companhia para que o objetivo da conclusão deste curso fosse atingido.

Por fim, agradeço à Deus, aos Orixás e aos guias espirituais pelas forças que me deram, as quais me ajudaram a seguir adiante apesar de todas as dificuldades.

*Sabemos o que somos, mas não sabemos o que
poderemos ser.*

William Shakespeare

RESUMO

ESTUDO SOBRE O USO DE FERRAMENTAS DE INTELIGÊNCIA DE NEGÓCIOS PARA ANÁLISE DE DADOS ABERTOS DA ÁREA DA SAÚDE

AUTOR: Rodrigo Balest Pinto de Oliveira

ORIENTADOR: Daniel Lichtnow, Dr.

Com as atuais leis de acesso à informação, qualquer cidadão pode visualizar informações relacionadas à administração pública, trazendo a possibilidade de que qualquer pessoa interessada possa realizar análises sobre estes dados, fornecendo um panorama mais claro sobre as ações das diferentes esferas de governo. Já existem algumas iniciativas onde dados públicos são trabalhados de forma a transformá-los em informações úteis à sociedade e/ou dar publicidade a estes dados. Uma das áreas da administração pública onde isto é necessário é a área da saúde. A visualização destes dados está relacionada a uma área denominada Business Intelligence (BI, ou Inteligência de Negócios). Inteligência de negócios é um termo “guarda-chuva” que abrange arquiteturas, ferramentas, banco de dados, aplicações e metodologias. O objetivo da área de Inteligência de Negócios é prover informações que facilitem a análise e auxiliem na tomada de decisão. Existem várias ferramentas de inteligência de negócios, entre elas, o Power BI da Microsoft e o Data Studio da Google. Desta forma, este trabalho apresenta um estudo comparativo entre estas ferramentas, utilizando-se de dados abertos da área da saúde. Para o desenvolvimento deste estudo, primeiramente é realizada uma apresentação teórica de conceitos de Inteligência de Negócios relacionados a este trabalho. A seguir, é descrita a obtenção de dados e sua utilização no desenvolvimento de um *data warehouse*, relatando as transformações realizadas sobre os mesmos. Na etapa seguinte é documentado o uso do *data warehouse* no desenvolvimento de painéis de visualização de dados nas duas ferramentas estudadas, onde buscou-se manter o máximo de similaridade entre os painéis de cada ferramenta a fim de que fosse possível perceber as diferenças em cada ferramenta. Por fim, as diferenças encontradas são descritas e analisadas, indicando as características, pontos fortes e fracos de cada uma das ferramentas estudadas.

Palavras-chave: Inteligência de negócios. Dados abertos.

ABSTRACT

STUDY ON THE USE OF BUSINESS INTELLIGENCE TOOLS FOR OPEN DATA ANALYSIS IN THE HEALTH AREA

AUTHOR: Rodrigo Balest Pinto de Oliveira

ADVISOR: Daniel Lichtnow, Dr.

With the current access to information laws, any citizen can view information related to public administration, bringing the possibility that any interested person can carry out analyzes on this data, bringing a clearer picture of the actions of the different spheres of government. There are already some initiatives where public data are processed in order to transform them into useful information for society and/or publicize this data. One of the areas of public administration where this is necessary is the area of health. The visualization of this data is related to an area called Business Intelligence (BI, or Business Intelligence). Business Intelligence is an umbrella term that encompasses architectures, tools, databases, applications and methodologies. The objective of the Business Intelligence area is to provide information that facilitates analysis and helps in decision making. There are several Business Intelligence tools, including Microsoft's Power BI and Google's Data Studio. Thus, this work presents a comparative study between these tools, using open health data. For the development of this study, firstly, a theoretical presentation of Business Intelligence concepts related to this work is carried out. Next, obtaining data and its use in the development of a data warehouse is described, reporting the transformations performed on them. In the next step, the use of the data warehouse in the development of data visualization panels in the two studied tools is documented, where we sought to maintain the maximum similarity between the panels of each tool so that it was possible to perceive the differences in each tool. Finally, the differences found are described and analyzed, indicating the characteristics, strengths, and weaknesses of each of the tools studied.

Keywords: Business Intelligence. Open data.

LISTA DE FIGURAS

Figura 1 - Exemplo de um modelo estrela.....	21
Figura 2 - Exemplo de um esquema snowflake de um data warehouse.....	22
Figura 3 - Exemplo de tabelas com granularidade diferente.....	23
Figura 4 - A página inicial do CNES na internet.....	25
Figura 5 - Arquivos da base de dados oriunda do CNES.....	26
Figura 6 – Modelo lógico dos dados brutos oriundos do CNES.....	27
Figura 7 - Exemplo de um comando COPY utilizado.....	28
Figura 8 - Diagrama do modelo dimensional.....	28
Figura 9 - Scripts SQL de criação das tabelas do modelo dimensional.....	29
Figura 10 - Scripts SQL de migração de dados para o modelo dimensional.....	30
Figura 11 - Scripts SQL de criação e geração de dados da tabela de distâncias.....	32
Figura 12 - Os elementos que compoem o Power BI.....	33
Figura 13 - A interface do Power BI Desktop.....	34
Figura 14 - Seleção de tabelas no Power BI.....	35
Figura 15 - Painel de leitos por tipo sendo exibido no Power BI Desktop.....	36
Figura 16 - Painel de leitos por especialidade no Power BI Desktop.....	37
Figura 17 - Fórmula da medida adicionada à tabela bi.fato_leitos.....	37
Figura 18 - Exemplo de dica da ferramenta ao se manter o mouse sobre um dos pontos do mapa.....	38
Figura 19 - Icone que permite ativar a interação entre as visualizações.....	38
Figura 20 - Painel de leitos SUS e não SUS por UF no Power BI Desktop.....	39
Figura 21 - Painel de distâncias de leitos por especialidade no RS no Power BI Desktop.....	40
Figura 22 - Painel de leitos por estabelecimento no Power BI Desktop.....	41
Figura 23 - Tela inicial do Google Data Studio.....	42
Figura 24 - Painel do usuário da conta do Ngrok.....	43
Figura 25 - Ngrok em execução, exibindo as conexões ativas.....	43
Figura 26 - Interface web do Ngrok exibindo informações das conexões.....	44
Figura 27 - Tela de pesquisa de conectores do Data Studio.....	45
Figura 28 - Dados da conexão e listagem das tabelas do banco de dados na criação de uma fonte de dados no Data Studio.....	46
Figura 29 - Mensagem de erro exibida ao utilizar uma fonte de dados que usa uma tabela do PostgreSQL.....	47
Figura 30 - Tela de edição de campos da fonte de dados.....	47
Figura 31 - Combinação de dados para obtenção da quantidade de leitos por tipo.....	48
Figura 32 - Configuração do painel de leitos por tipo.....	49
Figura 33 - Combinação de dados para obtenção da quantidade de leitos por especialidade.....	50
Figura 34 - Configuração do painel de leitos por especialidade.....	50
Figura 35 - Consulta personalizada para a fonte de dados de leitos por UF.....	51
Figura 36 - Configuração do campo personalizado para quantidade de leitos não SUS.....	51
Figura 37 - Configuração do campo personalizado <i>localizacao</i>	52
Figura 38 – Dica sendo exibida ao passar o mouse sobre um dos pontos do mapa.....	52
Figura 39 – Painel de leitos SUS e não SUS por UF.....	53
Figura 40 – Consulta personalizada para a fonte de dados de distâncias de leitos por especialidade.....	54
Figura 41 – Painel de distâncias de leitos por especialidade.....	54
Figura 42 – Consulta personalizada para o painel de leitos por estabelecimento.....	55
Figura 43 – Painel de leitos por estabelecimento.....	55

Figura 44 – Resultado de busca por conectores do Data Studio com o termo “banco”.....	58
Figura 45 – Algumas das opções de conectores de dados do Microsoft Power BI.....	58
Figura 46 – Aba de dados do Microsoft Power BI.....	60
Figura 47 – Aba de modelo do Microsoft Power BI.....	61
Figura 48 – Exemplo de uma combinação de dados no Google Data Studio.....	61
Figura 49 – Exemplo de consulta personalizada para uma fonte de dados no Google Data Studio.....	62
Figura 50 – Painéis de seleção de visualizações no Data Studio (esquerda) e Power BI (direita).....	62

LISTA DE ABREVIATURAS E SIGLAS

UFSM	Universidade Federal de Santa Maria
3FN	Terceira Forma Normal
BD	Banco de Dados
BI	<i>Business Intelligence</i> , ou Inteligência de Negócios
CNES	Cadastro Nacional de Estabelecimentos de Saúde
EIS	<i>Executive Information System</i> , ou Sistema de Informações Executivas
ETL	<i>Extraction, Transformation and Loading</i> , ou Extração, Transformação e Carga
NAT	<i>Network Address Translation</i>
SEM	<i>Search Engine Marketing</i> , ou Marketing para Mecanismos de Busca
SEO	<i>Search Engine Optimization</i> , ou Otimização para Mecanismos de Busca
SGBD	Sistema Gerenciador de Banco de Dados
SIG	Sistemas de Informações Gerenciais
SQL	<i>Structured Query Language</i> , ou Linguagem de Consultas Estruturada
URL	<i>Uniform Resource Locator</i>

SUMÁRIO

1	INTRODUÇÃO	15
1.1	Objetivos	15
1.1.1	Objetivo geral	15
1.1.2	Objetivos específicos	15
1.2	Justificativa	16
1.3	Organização do trabalho	16
2	INTELIGÊNCIA DE NEGÓCIOS	18
2.1	Histórico	18
2.2	Processos de extração, transformação e carga de dados (ETL)	18
2.3	Data Warehouse	18
2.3.1	Modelagem dimensional	19
2.3.1.1	Normalização	19
2.3.1.2	Star Schema, ou modelo estrela	20
2.3.1.3	Snowflake schema, ou modelo floco de neve	21
2.3.1.4	Granularidade	22
2.3.2	Projeto de Data Warehouse	23
2.4	Ferramentas de visualização	24
3	ESTUDO DE CASO	25
3.1	Importação dos dados	27
3.2	Transformação dos dados	28
3.3	Power BI	32
3.3.1	Carga de dados	34
3.3.2	Leitos por tipo	35
3.3.3	Leitos por especialidade	36
3.3.4	Leitos SUS e não SUS por UF	37
3.3.5	Distâncias de leitos por especialidade	39
3.3.6	Leitos por estabelecimento	40
3.4	Data Studio	41
3.4.1	Acesso à base de dados local a partir da web	42
3.4.2	Carga de dados	45
3.4.3	Leitos por tipo	48
3.4.4	Leitos por especialidade	49
3.4.5	Leitos SUS e não SUS por UF	51
3.4.6	Distâncias de leitos por especialidade	53
3.4.7	Leitos por estabelecimento	55

4	COMPARATIVO ENTRE AS FERRAMENTAS	56
4.1	Licença de uso	56
4.2	Facilidade de uso	57
4.3	Fontes de dados	57
4.4	Transformação de dados	59
4.5	Visualizações	62
5	CONCLUSÃO	63

1 INTRODUÇÃO

Com a ubiquidade tecnológica dos dias atuais, muitos dados são produzidos diariamente nos mais diversos meios, incluindo-se a área da administração pública. Estes dados, sendo catalogados, tratados e apresentados corretamente, fornecem indicadores valiosos que podem ser usados pelos gestores no apoio a tomadas de decisão. É disso que trata a área de Inteligência de Negócios: a “transformação de dados em informações, depois em decisões e por fim em ações” (SHARDA; DELEN; TURBAN, 2019, p. 15).

Este processo de transformação passa por várias etapas, como a extração dos dados, sua limpeza e transformação e seu armazenamento em estruturas pré-planejadas conhecidas como armazéns de dados, ou *data warehouses*. Uma importante etapa do processo de análise de dados é a sua visualização. Fayad (2002 apud TURBAN, 2009, p. 124) esclarece que “visualização de dados refere-se às tecnologias que dão suporte à visualização e, algumas vezes, à interpretação de dados e informações em vários pontos ao longo da cadeia de processamento de dados”.

Neste trabalho são apresentados os conceitos sobre o que é Inteligência de Negócios, ou *Business Intelligence (BI)*, bem como um estudo comparativo sobre o uso das ferramentas Microsoft Power BI e Google Data Studio para visualização de dados abertos sobre a área da saúde em âmbito nacional. Também é desenvolvido um *data warehouse* com os dados disponíveis no sites do Ministério da Saúde, documentando-se os passos de seu desenvolvimento.

1.1 OBJETIVOS

1.1.1 OBJETIVO GERAL

O objetivo deste trabalho é estudar o uso das ferramentas Microsoft Power BI e Google Data Studio utilizando dados abertos da área de saúde, a fim de fornecer ao leitor informações que o auxiliem na escolha de uma ferramenta de Inteligência de Negócios.

1.1.2 OBJETIVOS ESPECÍFICOS

- Estudar os conceitos de Inteligência de Negócios;
- Desenvolver um *data warehouse*, documentando seu desenvolvimento;
- Estudar as ferramentas Microsoft Power BI e Google Data Studio;
- Desenvolver relatórios e painéis interativos sobre dados abertos utilizando Microsoft Power BI e Google Data Studio;
- Fazer um comparativo entre Microsoft Power BI e Google Data Studio;

1.2 JUSTIFICATIVA

A área de Inteligência de Negócios visa facilitar à tomada de decisões gerenciais sendo diversas ferramentas usadas no ambiente corporativo, mas que também podem ser úteis no campo da administração pública, como educação, cultura e saúde. Além disto, existe a necessidade de dar publicidade aos dados da administração pública.

Neste sentido, a Constituição Federal, no inciso XXXIII do art. 5^o, no inciso II do art. 37² e no § 2^o do art. 216³ garante e a Lei de Acesso à Informação (Lei nº 12.527 de 18 de novembro de 2011⁴) regulamenta o acesso de qualquer cidadão a informações relacionadas à administração pública, trazendo a possibilidade de que qualquer pessoa interessada possa coletar e realizar análises sobre estes dados, fornecendo um panorama mais claro sobre as ações das diferentes esferas de governo. Como exemplo deste tipo de iniciativa, temos o site Mobilidados⁵, que disponibiliza aos visitantes painéis de indicadores sobre mobilidade urbana. Diversas entidades da administração pública utilizam painéis de visualização de dados para fornecer informações aos cidadãos, por exemplo o Painel Lei de Acesso à Informação⁶ e os relatórios da Pró-Reitoria de Planejamento da UFSM⁷.

Este trabalho mostra um exemplo de tal ação, demonstrando a coleta de dados públicos do âmbito da saúde, sua transformação em informações e sua apresentação de uma forma mais amigável, ao mesmo tempo que faz um estudo comparativo entre as ferramentas Microsoft Power BI e Google Data Studio, para exibição das informações. Desta forma, espera-se que este trabalho auxilie na escolha de uma ferramenta de Inteligência de Negócios e também possa despertar o interesse pelo assunto, incentivando-o a contribuir com a divulgação das informações públicas de uma forma que seja útil tanto à administração pública como à população em geral.

1.3 ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado em cinco capítulos, conforme segue:

- O primeiro capítulo apresenta os objetivos e a justificativa do trabalho;
- O segundo capítulo expõe e fundamenta alguns conceitos sobre Inteligência de Negócios relevantes para o entendimento do estudo desenvolvido;
- O terceiro capítulo descreve o trabalho prático realizado, contendo a aquisição e transformação dos dados e o uso das ferramentas sobre estes dados;
- O quarto capítulo apresenta de forma comparativa as características das ferramentas Microsoft Power BI e Google Data Studio que de alguma forma influenciaram o desenvolvimento prático deste trabalho;

¹ http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm#art5xxxiii

² http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm#art37%C2%A73ii

³ http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm#art216%C2%A72

⁴ http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/112527.htm

⁵ <https://mobilidados.org.br/>

⁶ <http://paineis.cgu.gov.br/lai/index.htm>

⁷ <https://www.ufsm.br/pro-reitorias/proplan/relatorios-power-bi-ufsm/>

- O quinto capítulo expõe as considerações finais deste trabalho.

Já foi vista uma introdução ao trabalho, bem como seus objetivos e sua justificativa. A seguir são apresentados conceitos sobre Inteligência de Negócios importantes para o entendimento da atividade desenvolvida.

2 INTELIGÊNCIA DE NEGÓCIOS

Segundo Sharda (et al., 2019), *Business Intelligence* (BI, ou Inteligência de Negócios) “é um termo “guarda-chuva” que inclui arquiteturas, ferramentas, bancos de dados, aplicações e metodologias”. O objetivo do *Business Intelligence* é prover informações que facilitem a análise e auxiliem na tomada de decisão. Não se trata, pois, de apenas uma ferramenta, mas sim de vários componentes utilizados com um objetivo definido.

2.1 HISTÓRICO

Sharda (et al., 2019) descreve que apesar do Gartner Group ter começado a utilizar o termo BI na década de 1990, o seu conceito vem desde os Sistemas de Informações Gerenciais (SIG), utilizados nos anos 1970. Os relatórios produzidos por estes sistemas eram estáticos e bidimensionais, e ainda careciam de recursos de análise. Mais tarde, nos anos 1980, surgiram os Sistemas de Informações Executivas (EIS), expandindo o suporte computadorizado aos gerentes e executivos de nível superior. Estes sistemas adicionaram mais funcionalidades, como relatórios dinâmicos multidimensionais e recursos de análise.

Com o tempo, estes recursos foram apresentados em vários produtos comerciais até o meio da década de 1990. Mais tarde, estes recursos e outros novos surgiram no mercado com o nome de BI, dando o nome atual ao conceito originário dos Sistemas de Informações Executivas.

2.2 PROCESSOS DE EXTRAÇÃO, TRANSFORMAÇÃO E CARGA DE DADOS (ETL)

Com a necessidade de coletar informações de diversas fontes e armazená-los de forma consistente, surgiram os processos de ETL. Estes processos, conforme o próprio nome diz, possuem três etapas:

- Extração – Os dados são extraídos das diversas fontes onde se encontram (arquivos de texto, páginas web, bancos de dados transacionais, bancos de dados NoSQL) e são armazenados em um ambiente de transição ou *staging* (KIMBALL, 2002, p.8);
- Transformação – Os dados do ambiente de transição são transformados para o formato esperado pelo *data warehouse*, corrigindo inconsistências;
- Carga – A inclusão propriamente dita dos dados transformados no *data warehouse*.

2.3 DATA WAREHOUSE

Sharda (et al., 2019, p. 154) define um *data warehouse* (armazém de dados) como “uma coleção de dados produzidos para embasar a tomada de decisões; trata-se também de um repositório de dados históricos e correntes de potencial interesse para gestores de toda a organização”. Ele centraliza os dados de fontes diversas em um formato que dê suporte às atividades de BI.

Pode-se complementar com a definição de Inmon (1997), que diz que *data warehouse* “é um conjunto de dados baseado em assuntos, integrado, não volátil, e variável em relação ao tempo, de apoio às decisões gerenciais”. Esta definição indica quatro características importantes:

- São orientados para o assunto, ou seja, são organizados de forma a refletir um determinado assunto ou atividade da empresa. O exemplo mais comum seriam as vendas;
- São integrados, pois representam dados de diversas fontes de uma mesma maneira, criando consistência;
- Não volatilidade. Uma vez que os dados são inseridos em um *data warehouse*, eles não sofrem alterações;
- Variáveis de acordo com o tempo. Os dados informam sobre um determinado assunto em um momento específico, assim permitem a análise da variação destes dados de acordo com a variação do tempo.

2.3.1 MODELAGEM DIMENSIONAL

A modelagem dimensional é a abordagem preferencial para projetos de *data warehouse*, sendo largamente utilizada. Ela divide os dados entre tabelas fato e tabelas auxiliares chamadas de tabelas dimensões (Kimball, 2002, p.16). As tabelas dimensões armazenam as características de um evento (por exemplo, hora, dia, cidade, região, categoria), enquanto as tabelas fato armazenam o próprio fato ocorrido, relacionando-o com duas ou mais dimensões. Pode-se dizer que as tabelas fato guardam as medidas de um evento enquanto as dimensões guardam o contexto dos eventos.

Nas tabelas dimensão, cada linha possui uma chave substituta, ou *surrogate key*, e pode possuir uma chave natural, ou *natural key*. A chave substituta é uma chave numérica iniciada a partir do valor 1, a qual é referenciada pelas tabelas fato. Ela faz o papel de chave primária da dimensão dentro do *data warehouse*, não tendo significado fora dele. Já a chave natural tem a função de referenciar o dado na sua origem, fora do *data warehouse*. Assim, pode ter vários significados, como um CPF, um e-mail ou um código de produto.

Existem dois tipos de metodologias de modelagem de dados utilizadas na construção de um *data warehouse*: a Star Schema e a Snowflake. Para entender as diferenças entre ambas, é necessário primeiro compreender o conceito de normalização.

2.3.1.1 NORMALIZAÇÃO

Uma das características em que as metodologias Snowflake e Star Schema diferem é a sua normalização. A normalização é a aplicação de regras à modelagem de um banco de dados a fim de que se evite problemas com relação a desempenho, integridade e manutenção. Estas regras são chamadas de formas normais.

Para que se possa entender algumas das formas normais, primeiro é preciso entender os conceitos de dependência funcional, dependência parcial e dependência transitiva. Heuser (2009) diz que “em uma

tabela relacional, diz-se que uma coluna C_2 depende funcionalmente de uma coluna C_1 (ou que a coluna C_1 determina a coluna C_2) quando, em todas as linhas da tabela, para cada valor de C_1 que aparece na tabela, aparecer o mesmo valor de C_2 ”, e que “uma dependência (funcional) parcial ocorre quando uma coluna depende apenas de parte de uma chave primária composta”. Sobre dependência transitiva, diz que “uma dependência funcional transitiva ocorre quando, além de depender da chave primária da tabela, depende também de outra coluna ou conjunto de colunas da tabela.

A fim de melhor entendermos os modelos dimensionais vamos explicar apenas as três formas normais básicas:

- **Primeira Forma Normal (1FN).** Uma tabela está na primeira forma normal quando ela não contém tabelas aninhadas (HEUSER, 2009).
- **Segunda Forma Normal (2FN).** Uma tabela esta segunda forma normal quando, além de estar na primeira forma normal, não contém dependências parciais.
- **Terceira Forma Normal (3FN).** Uma tabela encontra-se na terceira forma normal quando, além de estar na segunda forma normal, não contém dependências transitivas.

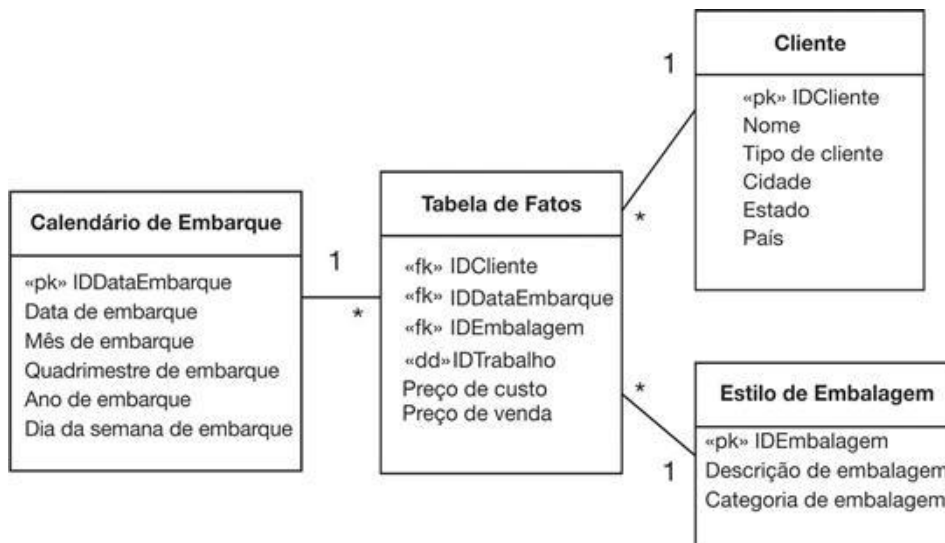
2.3.1.2 STAR SCHEMA, OU MODELO ESTRELA

No modelo estrela, a tabela fato está ligada a duas ou mais tabelas dimensão, ficando o diagrama de relacionamentos parecido com uma estrela, onde as tabelas dimensão são as pontas. A tabela fato é composta de atributos de dimensão e atributos de medidas. Os atributos de dimensão podem ser chaves estrangeiras para as tabelas dimensão, ou podem ser a própria dimensão em si. Este tipo de dimensão que não possui uma tabela própria é conhecido como dimensão degenerada.

Percebe-se que nesta organização, a normalização das tabelas (precisamente a 3FN) não é importante, pois muitos atributos das tabelas de dimensão se repetem. Isto é aceitável porque as consultas sobre tabelas desnormalizadas não precisam de tanto processamento para serem executadas, pois menos operações de junção precisam ser feitas. E, como em um *data warehouse* se trabalha com um grande volume de dados, o desempenho é mais desejável que a normalização.

Na Figura 1 temos um exemplo de modelo estrela, onde são armazenados dados sobre envio de produtos (embalagens) para clientes. Neste exemplo a tabela de fatos possui atributos de dimensão e atributos de medidas. As medidas são *Preço de custo* e *Preço de venda*, enquanto os atributos de dimensão são *IDCliente*, *IDDataEmbarque*, *IDEmbalagem* e *IDTrabalho*. Enquanto o atributo *IDTrabalho* é uma dimensão degenerada, os demais atributos de dimensão são referências a linhas nas tabelas de dimensão, que são *Cliente*, *Estilo de Embalagem* e *Calendário de Embarque*.

Figura 1 - Exemplo de um modelo estrela.



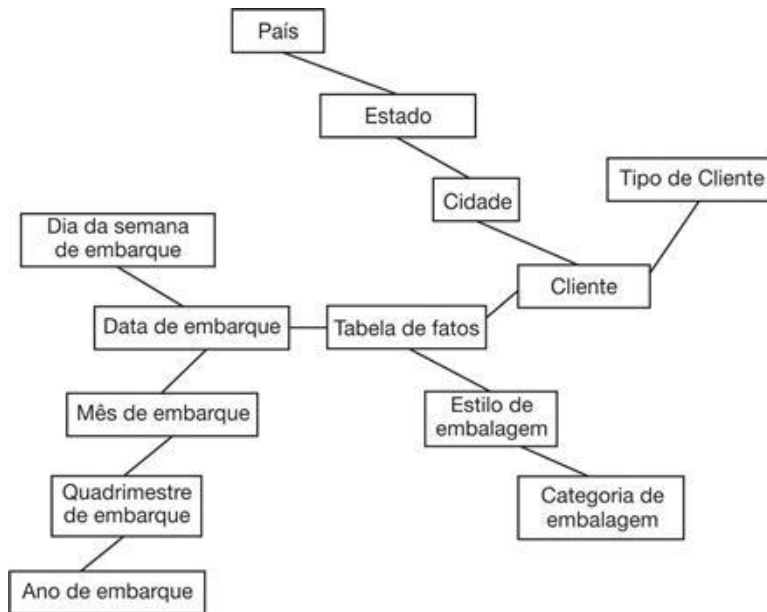
Fonte: (TEOREY, 2014).

2.3.1.3 SNOWFLAKE SCHEMA, OU MODELO FLOCO DE NEVE

No modelo *Snowflake*, busca-se aplicar a 3FN (terceira forma normal) às tabelas, deixando o diagrama de relacionamentos com a aparência de um floco de neve. Com isso, ganha-se mais espaço, mas perde-se em desempenho e o entendimento do usuário fica prejudicado. Por estas razões, o modelo *Snowflake* é muito menos usado do que o modelo estrela.

A Figura 2 mostra um exemplo de modelo *Snowflake*. Nele, o mesmo exemplo da Figura 1 é exibido, porém na 3FN. Pode-se perceber que informações que estariam repetidas no primeiro exemplo, como o tipo do cliente, cidade, estado e país do cliente, foram extraídas para tabelas próprias. O mesmo ocorre com as informações de embarque: os campos dia da semana, mês, quadrimestre e ano do embarque foram colocados em suas próprias tabelas, havendo relacionamentos entre elas de acordo com a 3FN.

Figura 2 - Exemplo de um esquema snowflake de um data warehouse.



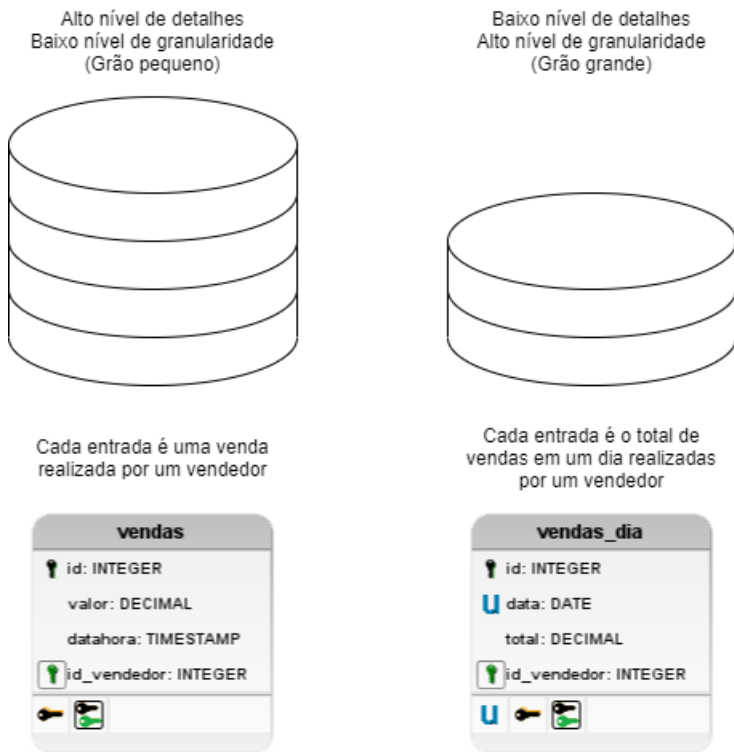
Fonte: (TEOREY, 2014)

2.3.1.4 GRANULARIDADE

A granularidade trata do nível de detalhamento da informação armazenada. Inmon (1997, p.43) diz que granularidade se refere ao nível de detalhe ou de sumarização das unidades de dados no *data warehouse*. O nível de granularidade tem relação inversa com o nível de detalhamento - quanto mais detalhes houver, menor será o nível de granularidade, ou seja, menor será o grão de dados. A Figura 3 - Exemplo de tabelas com granularidade diferente exhibe um exemplo, onde temos duas tabelas. A primeira armazena os dados sobre vendas realizadas por vendedor, onde cada linha representa uma venda, e na segunda tabela cada linha representa as vendas de um dia. Neste exemplo, a primeira tabela possui menor granularidade (grão menor) e maior nível de detalhamento, mas em contrapartida, também ocupa mais espaço de armazenamento.

Ao se definir a granularidade de um projeto, define-se o menor nível de detalhamento possível que as consultas àquele sistema podem ter. Por exemplo, na primeira tabela da Figura 3, pode-se saber o somatório vendido em uma semana, um dia ou uma hora. Caso a granularidade mínima tivesse sido definida para as vendas de um dia (como na segunda tabela da mesma figura), não seria possível obter os detalhes das vendas em determinada hora do dia, respondendo a perguntas como por exemplo, “quantas vendas são realizadas diariamente entre 8 e 10 horas da manhã?”.

Figura 3 - Exemplo de tabelas com granularidade diferente



Fonte: Elaborado pelo autor

2.3.2 PROJETO DE DATA WAREHOUSE

Kimball (2002, p.30) define quatro etapas para o processo de modelagem dimensional, as quais são brevemente revistas a seguir.

- 1. Escolher o processo de negócios.** O processo de negócios é a atividade realizada a qual se quer mensurar. Um exemplo clássico é a venda de um produto. Outros exemplos podem ser o transporte de mercadorias (para uma transportadora), atendimentos realizados (para um consultório), ou energia elétrica captada (para um sistema de painéis solares). No caso dos dados deste trabalho, será a quantidade de leitos hospitalares.
- 2. Declarar o grão do processo de negócios.** A definição da granularidade dirá qual será o nível de detalhamento dos dados. O nível de granularidade escolhido também afeta a forma como os processos de ETL deverão coletar e transformar os dados. Nesta etapa, decide-se, por exemplo, se será preciso armazenar dados sobre cada venda realizada, ou sobre as vendas de um dia. No exemplo do transporte de mercadorias citado acima, pode-se definir a localização de um transportador a nível de bairro, cidade ou estado. Para este trabalho, a granularidade definida foi o estabelecimento de saúde.

3. **Escolher as dimensões que se aplicam a cada linha da tabela fato.** As dimensões definem o contexto do evento. Assim, nesta etapa, decide-se quais informações sobre o contexto serão necessárias. Alguns exemplos são hora, data e localização. Neste caso, serão o estabelecimento de saúde, a especialidade do leito e o tipo de leito.
4. **Identificar os números que irão popular cada linha da tabela fato.** São os dados que se quer medir. Kimball (2002) diz que os fatos são determinados pela pergunta “O que estamos medindo?”. Fatos típicos são quantidades vendidas ou valor total de venda. Neste caso, é a quantidade de leitos hospitalares.

2.4 FERRAMENTAS DE VISUALIZAÇÃO

A visualização de dados é a apresentação de informações de forma imagética ou gráfica, e deve ser realizada de uma forma que torne eficaz o entendimento de seu conteúdo, facilitando o processo de tomada de decisão. Abaixo vemos alguns exemplos de ferramentas utilizadas para este fim.

- **Microsoft Excel**⁸ – Apesar de ser um editor de planilhas eletrônicas, o Microsoft Excel permite o acesso a diversas fontes de dados, bem como sua exibição não somente de forma tabular, mas também em formato de gráficos.
- **Microsoft Power BI** – Segundo a própria Microsoft⁹, O Power BI é uma coleção de serviços de software, aplicativos e conectores que trabalham juntos para transformar suas fontes de dados não relacionadas em informações coerentes, visualmente envolventes e interativas.
- **Google Data Studio**¹⁰ – O Data Studio é uma ferramenta online gratuita para exibição de dados em relatórios e painéis informativos personalizáveis.
- **MicroStrategy**¹¹ – A plataforma MicroStrategy, fornecida pela empresa de mesmo nome, combina a descoberta e análise visual de dados de autoatendimento com análises e relatórios corporativos adequados para sistemas de registro em larga escala.
- **SAS Visual Analytics**¹² – O SAS Visual Analytics é a ferramenta de visualização de dados da SAS, empresa conhecida no ramo de BI. Ela fornece relatórios interativos, descoberta visual, análise de autoatendimento, escalabilidade e governança, usando um ambiente de memória, bem como análise preditiva.
- **Tableau**¹³ - O Tableau é uma plataforma de análise como um serviço com forte descoberta de dados visuais.

⁸ <https://www.microsoft.com/pt-br/microsoft-365/excel>

⁹ <https://docs.microsoft.com/pt-br/power-bi/fundamentals/power-bi-overview>

¹⁰ <https://support.google.com/datastudio/answer/6283323?hl=pt-BR>

¹¹ <https://www.microstrategy.com/pt>

¹² https://www.sas.com/pt_br/software/visual-analytics.html

¹³ <https://www.tableau.com/pt-br/why-tableau/what-is-tableau>

3 ESTUDO DE CASO

Nesta seção é descrito o processo realizado para desenvolver os painéis de visualização com os dados abertos sobre saúde. Para construir as visualizações foram utilizados o Power BI, ferramenta da Microsoft e o Data Studio da Google. O objetivo é a partir deste uso poder realizar um comparativo entre as duas ferramentas de BI.

Os dados abertos sobre o sistema de saúde foram obtidos a partir do site do CNES – Cadastro Nacional de Estabelecimentos de Saúde. A página wiki do Ministério da Saúde o define da seguinte forma:

O Cadastro Nacional de Estabelecimentos de Saúde (CNES) é o sistema de informação oficial de cadastramento de informações de todos os estabelecimentos de saúde no país, independentemente de sua natureza jurídica ou de integrarem o Sistema Único de Saúde (SUS). Trata-se do cadastro oficial do Ministério da Saúde (MS) no tocante à realidade da capacidade instalada e mão-de-obra assistencial de saúde no Brasil em estabelecimentos de saúde públicos ou privados, com convênio SUS ou não.

Figura 4 - A página inicial do CNES na internet

The screenshot shows the homepage of the CNES website. At the top, there is a navigation bar with links for 'BRASIL', 'CORONAVÍRUS (COVID-19)', 'Simplifique!', 'Participe', 'Acesso à informação', 'Legislação', and 'Canais'. Below this is a search bar and social media icons. The main content area features a large banner for 'Wiki CNES' with the text 'Já acessou nossa wiki?' and the URL 'wiki.saude.gov.br/cnes'. To the left of the banner is a 'ACESSO RÁPIDO' menu with options like 'Consultas', 'Cronograma', and 'Obter CNES'. Below the banner are two search sections: 'Consulta Estabelecimento' and 'Consulta Profissional', each with a search input field and a 'Pesquisar' button. On the right side, there are sections for 'INFORMES' and 'LEGISLAÇÃO' listing recent updates and regulations.

Fonte: <http://cnes.datasus.gov.br/>

A Figura 4 exibe a página do CNES na internet. Na opção *Downloads* > *Base de Dados* do menu principal, foi baixado o arquivo *BASE_DE_DADOS_CNES_202105.ZIP* com os dados do banco para o mês de maio de 2021, o mais recente no momento. O arquivo contém vários outros arquivos no formato *.csv*, um para cada tabela do sistema. A Figura 5 exibe este arquivo aberto no programa WinRAR.

Também foi necessário baixar o arquivo *SCNES_DICIONARIO_DE_DADOS.ZIP* da página disponível em *Downloads* > *Documentação*. Este arquivo contém o dicionário de dados das tabelas baixadas anteriormente em formato *.docx*.

Figura 5 - Arquivos da base de dados oriunda do CNES

Name	Size	Packed	Type	Modified	CRC32
Pasta de arquivos					
..					
rlAdmGerenciaCnes202105.csv	199.118	40.353	Arquivo de Valores...	14/06/2021 09:08	02068004
rlAtividadeObrigatoria202105.csv	221	111	Arquivo de Valores...	14/06/2021 09:08	E1F655D5
rlCooperativa202105.csv	136.565	21.358	Arquivo de Valores...	14/06/2021 09:07	C176FB66
rlEquipeAldeia202105.csv	20.943	2.320	Arquivo de Valores...	14/06/2021 09:08	BA902065
rlEquipeNasfEsf202105.csv	11.036.770	2.664.087	Arquivo de Valores...	14/06/2021 09:08	B48BBA75
rlEstabAtendPrestConv202105.csv	40.329.184	9.043.949	Arquivo de Valores...	14/06/2021 09:05	4C5D670A
rlEstabAtenPsico202105.csv	137.970	42.543	Arquivo de Valores...	14/06/2021 09:08	C5B619F8
rlEstabAvaliacao202105.csv	86.101	20.925	Arquivo de Valores...	14/06/2021 09:08	FF90AE37
rlEstabCentralReg202105.csv	555.742	133.893	Arquivo de Valores...	14/06/2021 09:08	9B768AAF
rlEstabColetaSelRejeito202105.csv	30.446.301	7.368.202	Arquivo de Valores...	14/06/2021 09:05	1095B724
rlEstabComissaoOutro202105.csv	4.710.227	985.040	Arquivo de Valores...	14/06/2021 09:05	8E74915E
rlEstabComplementar202105.csv	4.451.645	831.395	Arquivo de Valores...	14/06/2021 09:05	E8AC848B
rlEstabEndCompl202105.csv	860.806	255.349	Arquivo de Valores...	14/06/2021 09:08	8C1535BD
rlEstabEqpEmbarcao202105.csv	18.853	4.421	Arquivo de Valores...	14/06/2021 09:08	6FBE4167
rlEstabEqpUnidApoio202105.csv	16.464	1.979	Arquivo de Valores...	14/06/2021 09:08	3051889E
rlEstabEquipamento202105.csv	62.546.277	14.464.241	Arquivo de Valores...	14/06/2021 09:07	2205DBA7
rlEstabEquipeMun202105.csv	212	137	Arquivo de Valores...	14/06/2021 09:08	ACC91622
rlEstabEquipeProf202105.csv	97.536.134	13.570.826	Arquivo de Valores...	14/06/2021 09:07	98110EA7
rlEstabInstFisiAssist202105.csv	45.639.520	11.288.629	Arquivo de Valores...	14/06/2021 09:05	705811C7
rlEstabOrgParc202105.csv	2.669	1.143	Arquivo de Valores...	14/06/2021 09:08	E26DAA77
rlEstabPoloAldeia202105.csv	67.614	17.989	Arquivo de Valores...	14/06/2021 09:08	C17E9E16
rlEstabProfComissao202105.csv	2.463.264	440.253	Arquivo de Valores...	14/06/2021 09:08	261D1538
rlEstabProgFundo202105.csv	26.639.230	6.380.405	Arquivo de Valores...	14/06/2021 09:05	500F3517
rlEstabRegimeRes202105.csv	19.364	5.906	Arquivo de Valores...	14/06/2021 09:08	DD6F4D6B
rlEstabRepresentante202105.csv	289.129	88.750	Arquivo de Valores...	14/06/2021 09:08	C6042F1C
rlEstabSamu202105.csv	980.501	236.967	Arquivo de Valores...	14/06/2021 09:08	B2975B2C
rlEstabServClass202105.csv	95.332.251	12.627.820	Arquivo de Valores...	14/06/2021 09:07	282A97F6
rlEstabServicoApoio202105.csv	23.563.999	5.642.501	Arquivo de Valores...	14/06/2021 09:05	A14E37AD
rlEstabSipac202105.csv	8.228.878	1.107.160	Arquivo de Valores...	14/06/2021 09:08	502EC2FF
rlEstabSubTipo202105.csv	5.078.342	1.203.604	Arquivo de Valores...	14/06/2021 09:08	F5F8CBFD
rlEstabTeleCnes202105.csv	1.285.360	226.500	Arquivo de Valores...	14/06/2021 09:08	6CF6D558
rlEstabUnidAcolhim202105.csv	51.979	17.249	Arquivo de Valores...	14/06/2021 09:08	7D816877
rlMunAtenPsico202105.csv	50.419	7.508	Arquivo de Valores...	14/06/2021 09:08	AB3A7F3F
rlMunRegimeRes202105.csv	2.555	640	Arquivo de Valores...	14/06/2021 09:08	CF067DE7
rlMunUnidAcolhim202105.csv	8.268	1.499	Arquivo de Valores...	14/06/2021 09:08	A2CE0E18
rlNasfEsf202105.csv	1.493	607	Arquivo de Valores...	14/06/2021 09:08	3BD67948
rlTipoEstabAtividade202105.csv	377	162	Arquivo de Valores...	14/06/2021 09:08	C8822326

Total 2.038.523.506 bytes in 107 files

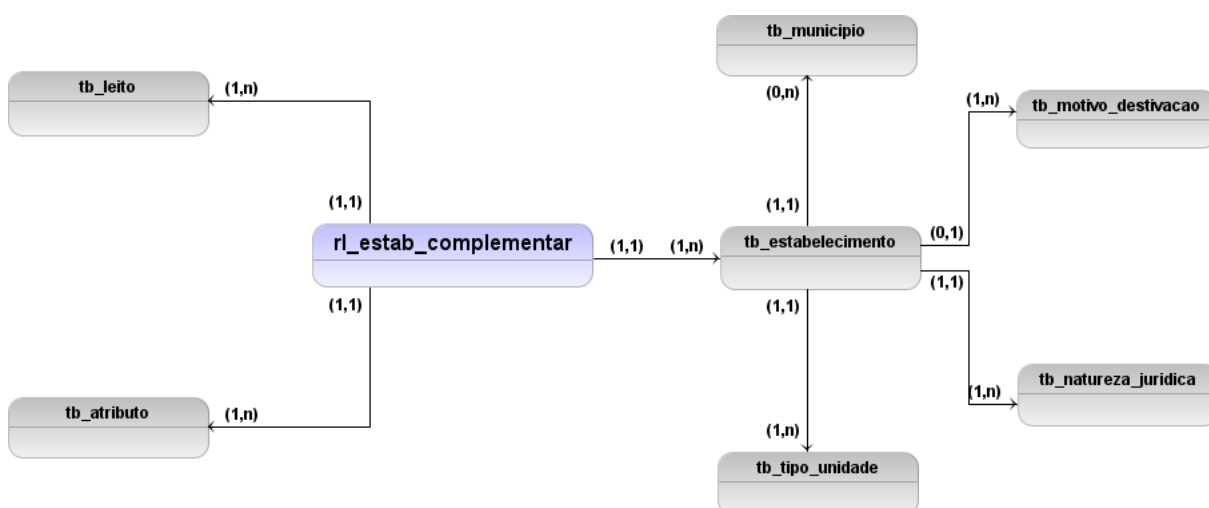
Fonte: Elaborado pelo autor

Após analisar o dicionário de dados, chegou-se às seguintes conclusões:

- A tabela *rl_estab_complementar* contém os dados sobre a quantidade de leitos no campo *qt_exist* e a quantidade de leitos SUS no campo *qt_sus*;

- Esta tabela também possui relacionamento para as tabelas *tb_leito* (que tem dados sobre a especialidade do leito), *tb_atributo* (com dados sobre a descrição do leito), e a tabela *tb_estabelecimento* (com dados sobre o estabelecimento);
- A tabela *tb_estabelecimento* possui relacionamentos com as tabelas *tb_municipio* (o município a que pertence o estabelecimento), *tb_tipo_unidade* (o tipo de estabelecimento de saúde), *tb_natureza_juridica* (a natureza jurídica do estabelecimento) e *tb_motivo_desativacao* (o motivo da desativação do estabelecimento);
- Pelas observações acima, concluiu-se que a tabela com as informações que irão alimentar a tabela fato do modelo dimensional seria a tabela *rl_estab_complementar*. Os dados das tabelas *tb_leito*, *tb_atributo* e *tb_estabelecimento* serão usados para alimentar as tabelas dimensão do modelo relacional. Ainda as tabelas *tb_municipio*, *tb_tipo_unidade*, *tb_natureza_juridica* e *tb_motivo_desativacao* contém dados para subdimensões da tabela *tb_estabelecimento*. O modelo lógico inferido a partir da análise dos dados pode ser visto na Figura 6.

Figura 6 – Modelo lógico dos dados brutos oriundos do CNES



Fonte: Elaborado pelo autor

3.1 IMPORTAÇÃO DOS DADOS

A seguir, foi necessário importar os dados dos arquivos *.csv* para um banco de dados de *staging* (uma área provisória – ver seção 2.2), a fim de realizar a extração e transformação dos dados. Para isso, utilizou-se o PostgreSQL como banco de dados. Ele conta com a função COPY, que permite realizar a cópia dos dados de um arquivo de texto para uma tabela do BD. A Figura 7 esta função sendo utilizada para importar os dados do arquivo *rl_estab_complementar* para uma tabela no banco de dados com o mesmo nome, onde os campos são separados pelo caractere “;” e a primeira linha do arquivo é uma linha de cabeçalho, contendo os nomes dos campos.

Figura 7 - Exemplo de um comando COPY utilizado.

```

COPY rl_estab_complementar
FROM 'C:\cnes\rl_estab_complementar.csv'
DELIMITER ';'
CSV HEADER;

```

Fonte: Elaborado pelo autor

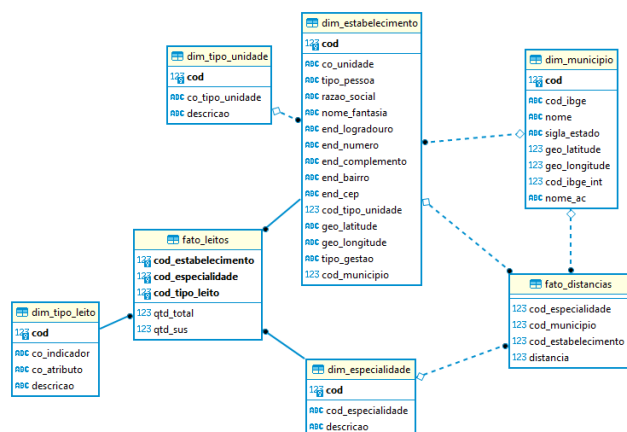
Antes de importar os dados, foi necessário criar as tabelas no banco de *staging*. Foi feita uma análise das colunas das tabelas nos arquivos *.csv* para definir os tipos de dados. As tabelas então foram criadas utilizando-se a ferramenta pgAdmin na versão 4. As tabelas *tb_tipo_unidade* e *tb_motivo_desativacao* não foram copiadas, pois viu-se que não seriam úteis para o objetivo proposto.

O comando COPY foi utilizado então para copiar os dados de cada uma das tabelas. Devido a má formatação de alguns arquivos *.csv* (como por exemplo, presença dos caracteres de separação de campos dentro do valor dos campos, ou arquivos com codificação diferente de UTF-8), alguns erros ocorreram durante a importação, sendo necessário realizar corrigir o arquivo manualmente e executar novamente o comando para realizar a importação.

3.2 TRANSFORMAÇÃO DOS DADOS

Com os dados importados, foi possível realizar a sua transformação e posterior inserção no modelo dimensional. Antes da transformação dos dados, as tabelas do modelo dimensional foram criadas. A Figura 8 mostra o diagrama do modelo dimensional, enquanto a Figura 9 mostra os scripts SQL utilizados para sua criação.

Figura 8 - Diagrama do modelo dimensional



Fonte: Elaborado pelo autor

Figura 9 - Scripts SQL de criação das tabelas do modelo dimensional

```

1 CREATE TABLE dim_tipo_unidade (
2     cod SERIAL,
3     co_tipo_unidade VARCHAR(2) NOT NULL,
4     descricao VARCHAR(60) NOT NULL,
5     PRIMARY KEY (cod)
6 );
7
8 CREATE TABLE dim_municipio (
9     cod SERIAL,
10    cod_ibge VARCHAR(7) NOT NULL,
11    nome VARCHAR(60) NOT NULL,
12    sigla_estado VARCHAR(2) NULL NULL,
13    PRIMARY KEY (cod)
14 );
15
16 CREATE TABLE dim_estabelecimento (
17    cod SERIAL,
18    co_unidade VARCHAR(31),
19    tipo_pessoa CHAR(2), --tp_pfpj(case 1 PF case 3 PJ)
20    razao_social VARCHAR(60), --no_razao_social
21    nome_fantasia VARCHAR(60), --no_fantasia
22    end_logradouro VARCHAR(60), --no_logradouro
23    end_numero VARCHAR(10), -- nu_endereco
24    end_complemento VARCHAR(20), --no_complemento
25    end_bairro VARCHAR(40), --no_bairro
26    end_cep VARCHAR(8), --co_cep
27    cod_tipo_unidade INTEGER, --tp_unidade
28    geo_latidade VARCHAR(30), --nu_latidade
29    geo_longitude VARCHAR(30), --nu_longitude
30    tipo_gestao CHAR(1), --tp_gestao
31    cod_municipio INTEGER, --co_municipio_gestor
32    PRIMARY KEY (cod),
33    FOREIGN KEY(cod_tipo_unidade) REFERENCES dim_tipo_unidade(cod),
34    FOREIGN KEY(cod_municipio) REFERENCES dim_municipio(cod)
35 );
36
37 CREATE TABLE dim_especialidade (
38    cod SERIAL,
39    cod_especialidade VARCHAR(2) NOT NULL, --tp_leito.co_leito
40    descricao VARCHAR(60) NOT NULL,
41    PRIMARY KEY (cod)
42 );
43
44 CREATE TABLE dim_tipo_leito (
45    cod SERIAL,
46    co_indicador VARCHAR(3) NOT NULL,
47    co_atributo CHAR(2) NOT NULL,
48    descricao VARCHAR(60) NOT NULL,
49    UNIQUE(co_indicador, co_atributo),
50    PRIMARY KEY(cod)
51 );
52
53 CREATE TABLE fato_leitos (
54    cod_estabelecimento INTEGER,
55    cod_especialidade INTEGER,
56    cod_tipo_leito INTEGER,
57    qtd_total INTEGER,
58    qtd_sus INTEGER,
59    PRIMARY KEY(cod_estabelecimento, cod_especialidade, cod_tipo_leito),
60    FOREIGN KEY(cod_estabelecimento) REFERENCES dim_estabelecimento(cod),
61    FOREIGN KEY(cod_especialidade) REFERENCES dim_especialidade(cod),
62    FOREIGN KEY(cod_tipo_leito) REFERENCES dim_tipo_leito(cod)
63 );
64

```

Fonte: Elaborado pelo autor

A seguir, foi feita a inserção dos dados no modelo dimensional a partir do banco de *staging*. Foi necessário realizar algumas modificações nos dados de origem para que fossem adequadamente adaptados às tabelas de destino. Estas modificações foram feitas nas próprias instruções SQL de importação de dados. Um exemplo é a instrução CASE, utilizada para transformar o código do tipo de pessoa para texto. As instruções de inserção no modelo dimensional são exibidas na Figura 10. A instrução CASE citada pode ser vista na linha 26.

Figura 10 - Scripts SQL de migração de dados para o modelo dimensional

```

1  INSERT INTO bi.dim_tipo_unidade (co_tipo_unidade, descricao)
2  SELECT co_tipo_unidade, ds_tipo_unidade FROM PUBLIC.tb_tipo_unidade;
3
4
5  INSERT INTO bi.dim_municipio (cod_ibge, nome, sigla_estado)
6  SELECT co_municipio, no_municipio, co_sigla_estado FROM PUBLIC.tb_municipio;
7
8
9  INSERT INTO bi.dim_estabelecimento (
10  co_unidade,
11  tipo_pessoa, --tp_pfpj(case 1 PF case 3 PJ)
12  razao_social, --no_razao_social
13  nome_fantasia, --no_fantasia
14  end_logradouro, --no_logradouro
15  end_numero, --nu_endereco
16  end_complemento, --no_complemento
17  end_bairro, --no_bairro
18  end_cep, --co_cep
19  cod_tipo_unidade, --tp_unidade
20  geo_latitude, --nu_latitude
21  geo_longitude, --nu_longitude
22  tipo_gestao, --tp_gestao
23  cod_municipio --co_municipio_gestor
24  ) SELECT
25  co_unidade,
26  CASE
27  WHEN tp_pfpj LIKE '1' THEN 'PF'
28  WHEN tp_pfpj LIKE '3' THEN 'PJ'
29  END AS tipo_pessoa,
30  no_razao_social,
31  no_fantasia,
32  no_logradouro,
33  nu_endereco,
34  no_complemento,
35  no_bairro,
36  co_cep,
37  dtu.cod AS cod_tipo_unidade,
38  nu_latitude,
39  nu_longitude,
40  tp_gestao,
41  dm.cod AS cod_municipio
42  FROM PUBLIC.tb_estabelecimento te
43  LEFT JOIN bi.dim_tipo_unidade dtu
44  ON dtu.co_tipo_unidade = LPAD(te.tp_unidade, 2, '0')
45  LEFT JOIN bi.dim_municipio dm
46  ON dm.cod_ibge = te.co_municipio_gestor;
47
48
49  INSERT INTO bi.dim_especialidade (cod_especialidade, descricao)
50  SELECT co_leito, ds_leito FROM PUBLIC.tb_leito;
51
52
53  INSERT INTO bi.dim_tipo_leito (co_indicador, co_atributo, descricao)
54  SELECT co_indicador, co_atributo, ds_atributo
55  FROM PUBLIC.tb_atributo
56  WHERE co_indicador LIKE '006';
57
58
59  INSERT INTO bi.fato_leitos (
60  cod_estabelecimento,
61  cod_especialidade,
62  cod_tipo_leito,
63  qtd_total,
64  qtd_sus
65  ) SELECT
66  destab.cod AS cod_estabelecimento,
67  desp.cod AS cod_especialidade,
68  dtleito.cod AS cod_tipo_leito,
69  rec.qt_exist::INTEGER,
70  rec.qt_sus::INTEGER
71  FROM PUBLIC.rl_estab_complementar rec
72  LEFT JOIN bi.dim_estabelecimento destab
73  ON destab.co_unidade = rec.co_unidade
74  LEFT JOIN bi.dim_especialidade desp
75  ON desp.cod_especialidade = LPAD(rec.co_leito, 2, '0')
76  LEFT JOIN bi.dim_tipo_leito dtleito
77  ON dtleito.co_atributo = rec.co_tipo_leito
78

```

Posteriormente foram adicionados os nomes das cidades com acentuação à coluna *nome_ac* da tabela *bi.dim_municipio* utilizando o código IBGE como referência. Esta medida foi tomada porque os nomes das cidades na fonte de dados do CNES não possuem acentos e estão em caixa alta. Desta forma seria possível uma melhor visualização dos nomes nos painéis.

As coordenadas geográficas das cidades foram obtidas a partir de um repositório no GitHub¹⁴ que compilou os dados oriundos do IBGE. Os dados do repositório foram então cruzadas com os dados do CNES através do código IBGE de cada cidade e adicionados à tabela *bi.dim_municipios*. Este repositório também foi utilizado para obter o nome das cidades com acentuação.

Também foi criada a tabela *bi.fato_distancias*, a fim de armazenar as distâncias entre cidades e leitos por especialidade. As cidades e estabelecimentos foram limitados aos existentes no estado do Rio Grande do Sul, a fim de se gerar uma amostra de dados suficiente sem necessitar de uma quantidade considerável de espaço em disco e tempo de processamento. Para calcular as distâncias entre as coordenadas geográficas, foram utilizadas as funções *ST_DistanceSphere* e *ST_MakePoint* do PostGIS, extensão geoespacial para o SGBD PostgreSQL. Os dados desta tabela foram gerados a partir de outras já existentes, conforme pode ser visto no script SQL da Figura 11.

¹⁴ <https://github.com/kelvins/Municipios-Brasileiros>

Figura 11 - Scripts SQL de criação e geração de dados da tabela de distâncias

```

1 CREATE TABLE bi.fato_distancias (
2     cod_especialidade int4 NOT NULL,
3     cod_municipio int4 NOT NULL,
4     cod_estabelecimento int4 NOT NULL,
5     distancia float4 NOT NULL
6 );
7
8 ALTER TABLE bi.fato_distancias
9 ADD CONSTRAINT fato_distancias_fk_especialidade
10 FOREIGN KEY (cod_especialidade)
11 REFERENCES bi.dim_especialidade(cod)
12 ON DELETE RESTRICT ON UPDATE RESTRICT;
13
14 ALTER TABLE bi.fato_distancias
15 ADD CONSTRAINT fato_distancias_fk_estabelecimento
16 FOREIGN KEY (cod_estabelecimento)
17 REFERENCES bi.dim_estabelecimento(cod)
18 ON DELETE RESTRICT ON UPDATE RESTRICT;
19
20 ALTER TABLE bi.fato_distancias
21 ADD CONSTRAINT fato_distancias_fk_municipio
22 FOREIGN KEY (cod_municipio)
23 REFERENCES bi.dim_municipio(cod)
24 ON DELETE RESTRICT ON UPDATE RESTRICT;
25
26 INSERT INTO bi.fato_distancias (cod_municipio, cod_estabelecimento, cod_especialidade, distancia)
27 SELECT
28     mun.cod_municipio,
29     leito.cod_estabelecimento,
30     leito.cod_especialidade,
31     ST_DistanceSphere(
32         ST_MakePoint(mun.geo_longitude, mun.geo_latitude),
33         ST_MakePoint(est_geo.lon, est_geo.lat)
34     ) / 1000 AS distancia
35 FROM
36     bi.dim_municipio mun,
37     bi.dim_estabelecimento est,
38     bi.fato_leitos leito,
39     (
40         SELECT
41             sub_mun.geo_longitude AS lon,
42             sub_mun.geo_latitude AS lat,
43             sub_est.cod,
44             sub_mun.sigla_estado AS uf,
45             sub_mun.cod AS cod_mun
46         FROM
47             bi.dim_estabelecimento sub_est
48         LEFT JOIN
49             bi.dim_municipio sub_mun
50             ON sub_est.cod_municipio = sub_mun.cod
51         WHERE
52             sigla_estado = 'RS'
53     ) est_geo
54 WHERE
55     est.cod = leito.cod_estabelecimento
56     AND est_geo.lat IS NOT NULL
57     AND est_geo.lon IS NOT NULL
58     AND mun.geo_latitude IS NOT NULL
59     AND mun.geo_longitude IS NOT NULL
60     AND est_geo.cod = est.cod
61     AND mun.sigla_estado = 'RS'
62     AND est_geo.uf = 'RS'

```

Fonte: Elaborado pelo autor

3.3 POWER BI

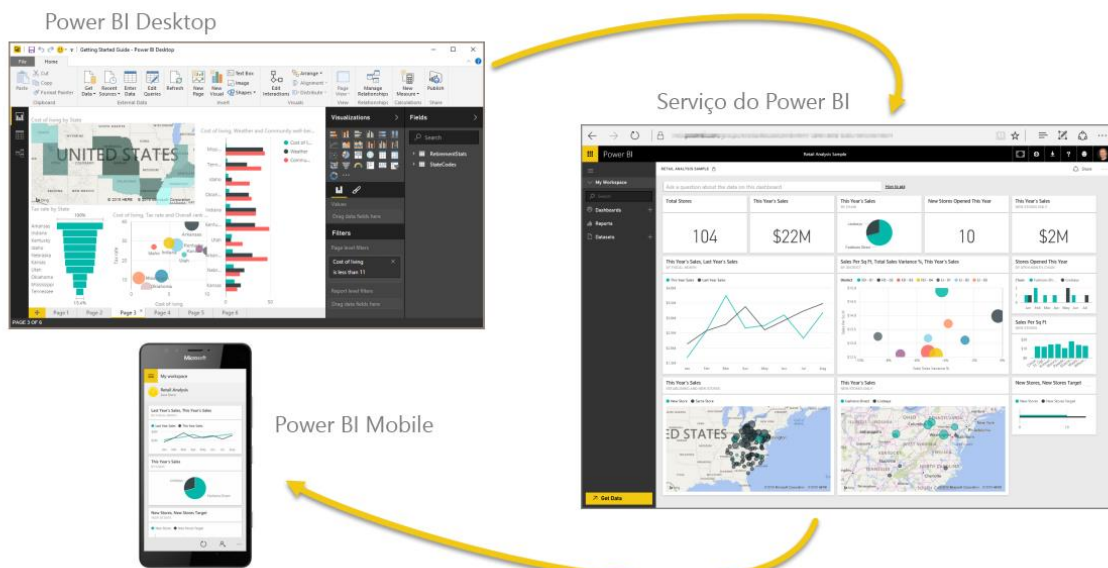
O Power BI é a ferramenta de Inteligência de Negócios da Microsoft para visualização de dados. O website da Microsoft descreve o Power BI como “uma coleção de serviços de software, aplicativos e conectores que funcionam juntos para tornar suas fontes de dados não relacionadas em informações coerentes, visualmente imersivas e interativas.” Ele permite a agregação de dados de diversas fontes, como

planilhas, arquivos do Excel, websites e bancos de dados, a transformação e limpeza destes dados, e sua apresentação em painéis que permitem interatividade. Todas estas etapas podem ser realizadas na mesma ferramenta. A Figura 13 exibe a interface do Power BI Desktop no desenvolvimento de um painel.

O Power BI é composto por três elementos distintos, conforme a Figura 12, cada um com um objetivo específico:

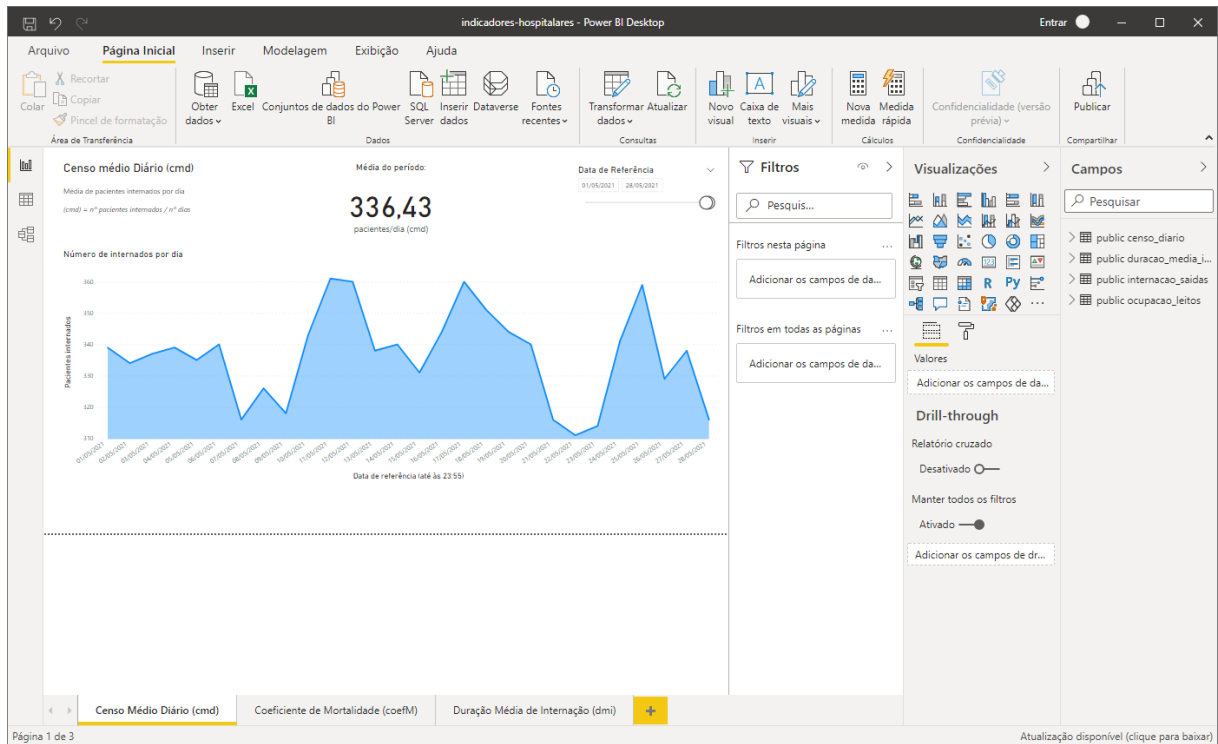
- **O Power BI Desktop** é o aplicativo da área de trabalho onde os relatórios e painéis são criados;
- **O serviço do Power BI** é um serviço SaaS (*software as a service*) online que proporciona a publicação e visualização dos painéis e relatórios com controle de acesso;
- **Os aplicativos móveis do Power BI** oferecem a apresentação dos painéis em dispositivos móveis com recursos próprios, como alertas e anotações.

Figura 12 - Os elementos que compõem o Power BI



Fonte: Microsoft (2021)

Figura 13 - A interface do Power BI Desktop



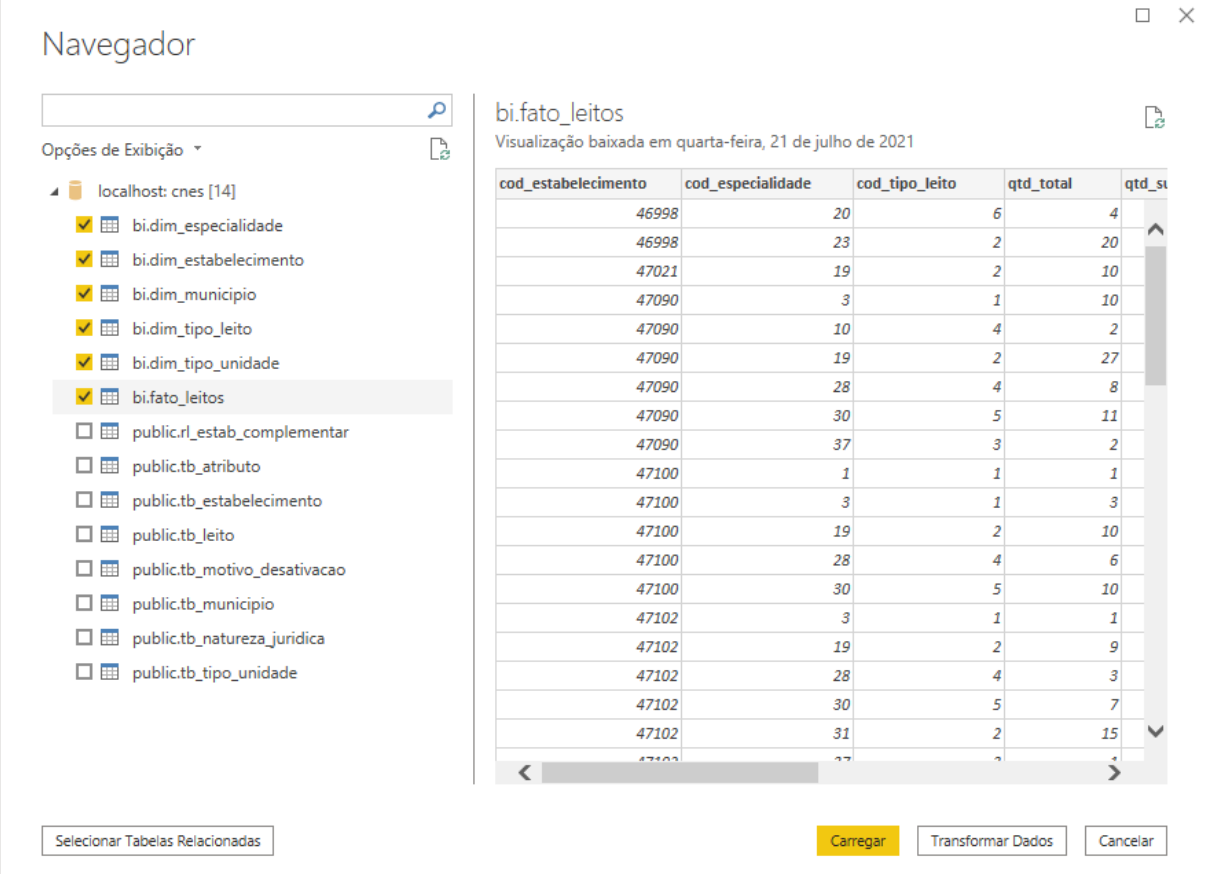
Fonte: Elaborado pelo autor

As seções seguintes descrevem como foi feita a carga de dados no Microsoft Power BI, bem como o desenvolvimento dos painéis. Os gráficos de rosca e de pizza foram escolhidos por serem representações simples e comumente utilizadas em representações visuais de dados, enquanto os demais gráficos foram escolhidos para demonstrar recursos das ferramentas, como interatividade entre visualizações, filtragem de dados e junção de dados de fontes diferentes.

3.3.1 CARGA DE DADOS

No Power BI, foi criado um novo relatório. A fonte de dados escolhida foi o banco PostgreSQL onde foi criado o modelo dimensional descrito na seção 3.1. Foi selecionada a opção *importar* como modo de conectividade de dados. Na seleção das tabelas, foram escolhidas apenas as do modelo dimensional, ficando as tabelas de *staging* (ver seção 2.2) fora da seleção. Como os dados já foram transformados anteriormente (ver seção 3.2), foi selecionada a opção *Carregar* ao invés de *Transformar Dados*. A Figura 14 exhibe as tabelas selecionadas na tela de importação de dados.

Figura 14 - Seleção de tabelas no Power BI



The screenshot shows the Power BI Navigator interface. On the left, a list of tables is displayed under the folder 'localhost: cnes [14]'. The table 'bi.fato_leitos' is selected, indicated by a checkmark and a highlighted row. Below the list is a button labeled 'Selecionar Tabelas Relacionadas'. On the right, a preview of the 'bi.fato_leitos' table is shown, with the title 'Visualização baixada em quarta-feira, 21 de julho de 2021'. The table has the following columns: 'cod_estabelecimento', 'cod_especialidade', 'cod_tipo_leito', 'qtd_total', and 'qtd_su'. The data rows are as follows:

cod_estabelecimento	cod_especialidade	cod_tipo_leito	qtd_total	qtd_su
46998	20	6	4	4
46998	23	2	20	20
47021	19	2	10	10
47090	3	1	10	10
47090	10	4	2	2
47090	19	2	27	27
47090	28	4	8	8
47090	30	5	11	11
47090	37	3	2	2
47100	1	1	1	1
47100	3	1	3	3
47100	19	2	10	10
47100	28	4	6	6
47100	30	5	10	10
47102	3	1	1	1
47102	19	2	9	9
47102	28	4	3	3
47102	30	5	7	7
47102	31	2	15	15
47102	37	3	2	2

At the bottom of the preview, there are three buttons: 'Carregar', 'Transformar Dados', and 'Cancelar'.

Fonte: Elaborado pelo autor

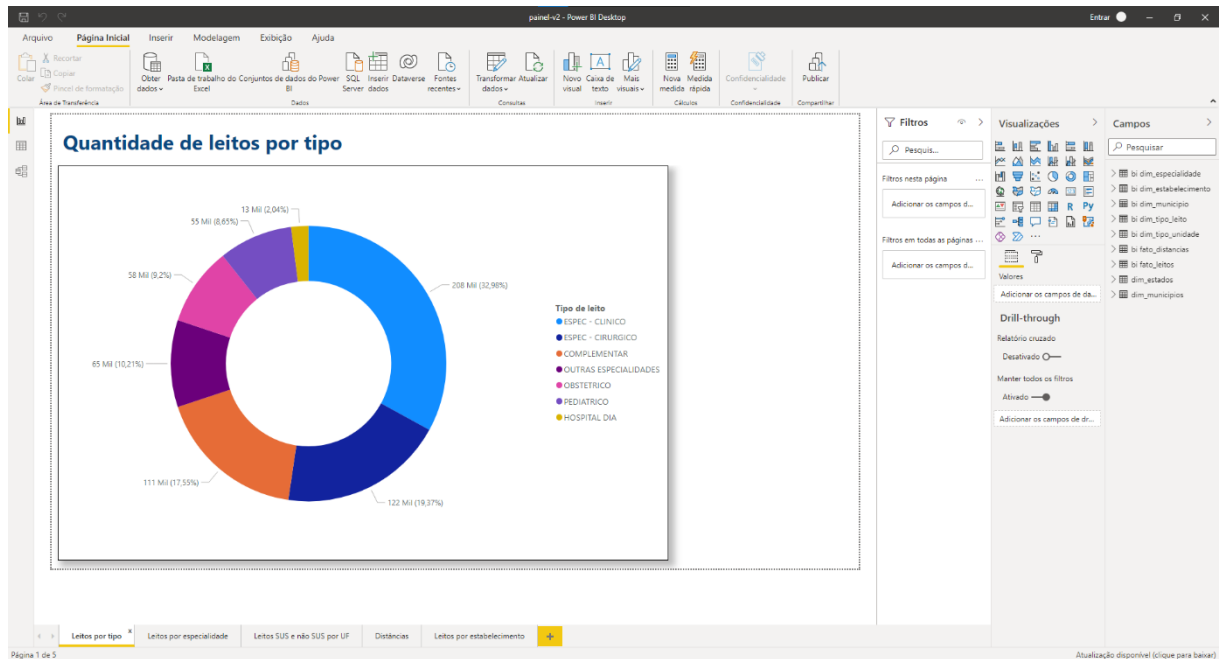
Cinco diferentes painéis foram criados, como descrito a seguir.

3.3.2 LEITOS POR TIPO

Este painel exibe a quantidade total de leitos (SUS e não SUS) em todo território nacional separados por tipo de leito. Para representar as informações, foi escolhido o gráfico de rosca. No campo legenda foi selecionada a coluna *descricao* da tabela *bi.dim_tipo_leito*, a qual foi renomeada para *Tipo de leito*, a fim de tornar sua apresentação mais clara. No campo valores, foi selecionada a coluna *qtd_total* da tabela *bi.fato_leitos*, com a opção *Soma* selecionada, a fim de garantir que os somatórios da quantidade de leitos por tipo de leito fossem exibidos no gráfico. O título do gráfico foi alterado para *Quantidades de leito por tipo*, ficando claro para o usuário sobre o que trata o gráfico.

Na Figura 15 temos a visualização do Power BI Desktop exibindo este painel.

Figura 15 - Painel de leitos por tipo sendo exibido no Power BI Desktop



Fonte: Elaborado pelo autor

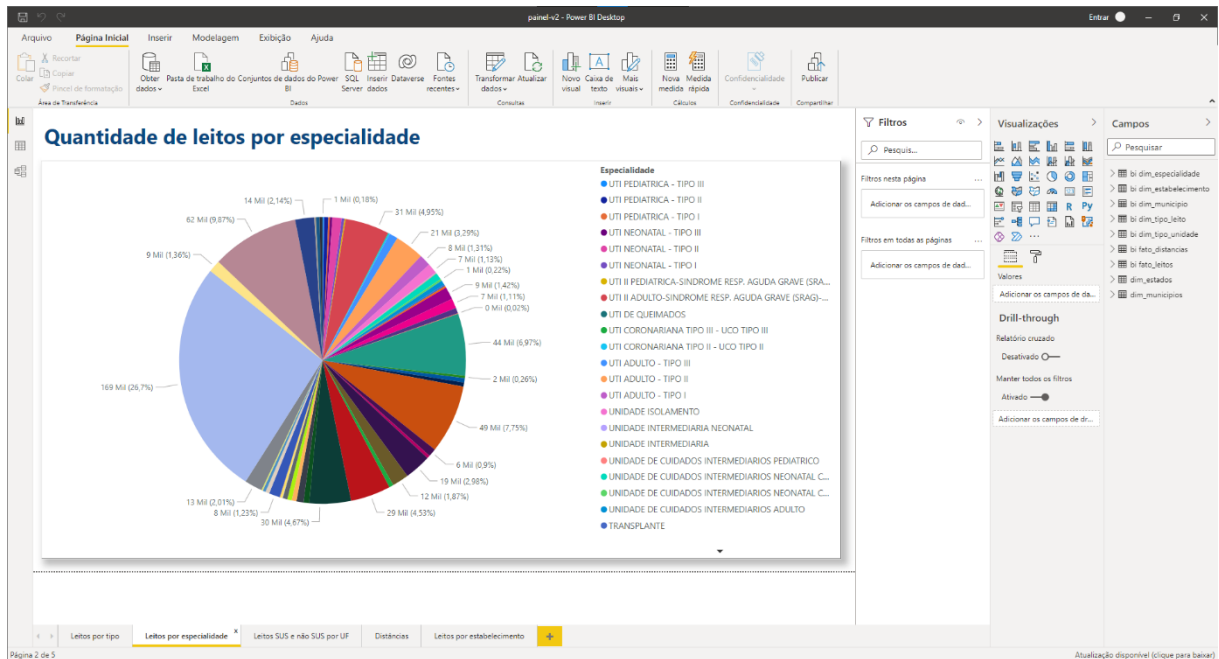
3.3.3 LEITOS POR ESPECIALIDADE

Este painel exibe a quantidade total de leitos (SUS e não SUS) em todo território nacional separados por tipo especialidade. Desta vez o gráfico selecionado foi o gráfico de pizza.

No campo legenda a coluna selecionada foi a *descricao* da tabela *bi.dim_especialidade*, sendo renomeada para *Especialidade*. Similarmente ao painel anterior, foi selecionada a coluna *descricao* da tabela *bi.dim_tipo_leito* no campo valores. Aqui também se verificou que sua opção *Soma* estava selecionada.

O título deste gráfico também foi alterado, ficando com o valor *Quantidade de leitos por especialidade*. O resultado pode ser visto na Figura 16.

Figura 16 - Painel de leitos por especialidade no Power BI Desktop



Fonte: Elaborado pelo autor

3.3.4 LEITOS SUS E NÃO SUS POR UF

Este painel exibe a quantidade de leitos SUS, não SUS, e total por UF. Foram utilizadas as visualizações de mapa e de barras empilhadas. Como os dados fornecem apenas a quantidade de leitos SUS e leitos totais, foi necessário criar uma nova medida na tabela *bi.fato_leitos* com o nome *qtd_total menos qtd_sus*. A fórmula usada pela medida é dada pela Figura 17.

Figura 17 - Fórmula da medida adicionada à tabela *bi.fato_leitos*

```
1 qtd_total menos qtd_sus =
2 SUM('bi fato_leitos'[qtd_total]) - SUM('bi fato_leitos'[qtd_sus])
```

Fonte: Elaborado pelo autor

A visualização do mapa foi configurada com o campo *Localização* utilizando a coluna *sigla_estado* da tabela *bi.dim_municipio* renomeada para *Estado*. No campo *Tamanho* foi utilizada a soma da coluna *qtd_total* da tabela *bi.fato_leitos*, renomeada para *Total*. No campo *Dicas de ferramentas* foram selecionados a soma da coluna *qtd_sus* da tabela *bi.fato_leitos* renomeada para *SUS*, e a medida *qtd_total menos qtd_sus* da mesma tabela. Esta última configuração permite que os valores dos campos selecionados possam ser visualizados ao se passar o mouse sobre um dos pontos do mapa (além do valor da coluna designada no campo *Tamanho*) conforme demonstra a Figura 18.

Figura 18 - Exemplo de dica da ferramenta ao se manter o mouse sobre um dos pontos do mapa

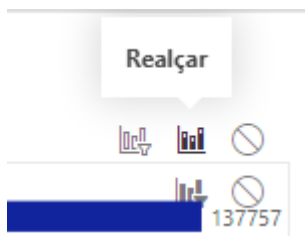


Fonte: Elaborado pelo autor

A visualização do gráfico de barras empilhadas foi configurada com o campo *Eixo* utilizando a coluna *sigla_estado* da tabela *bi.fato_leitos*, renomeada para UF. O campo *Valores* foi configurado com a soma da coluna *qtd_sus* da tabela *bi.fato_leitos* renomeada para *SUS*, e a medida *qtd_total menos qtd_sus* da mesma tabela. A visualização também foi configurada para exibir os valores totais em cada barra.

Ao clicar em um local no mapa ou em uma barra no gráfico, o estado selecionado é destacado na outra visualização. Isto foi possível utilizando o recurso *Editar interações* do Power BI. Para utilizá-lo, deve-se selecionar uma visualização, acessar a opção *Formato* no menu e ativar a opção *Editar interações*. Sobre as demais visualizações será exibido um ícone (ver Figura 19) que permite ativar a interação entre elas, realçando os dados relacionados.

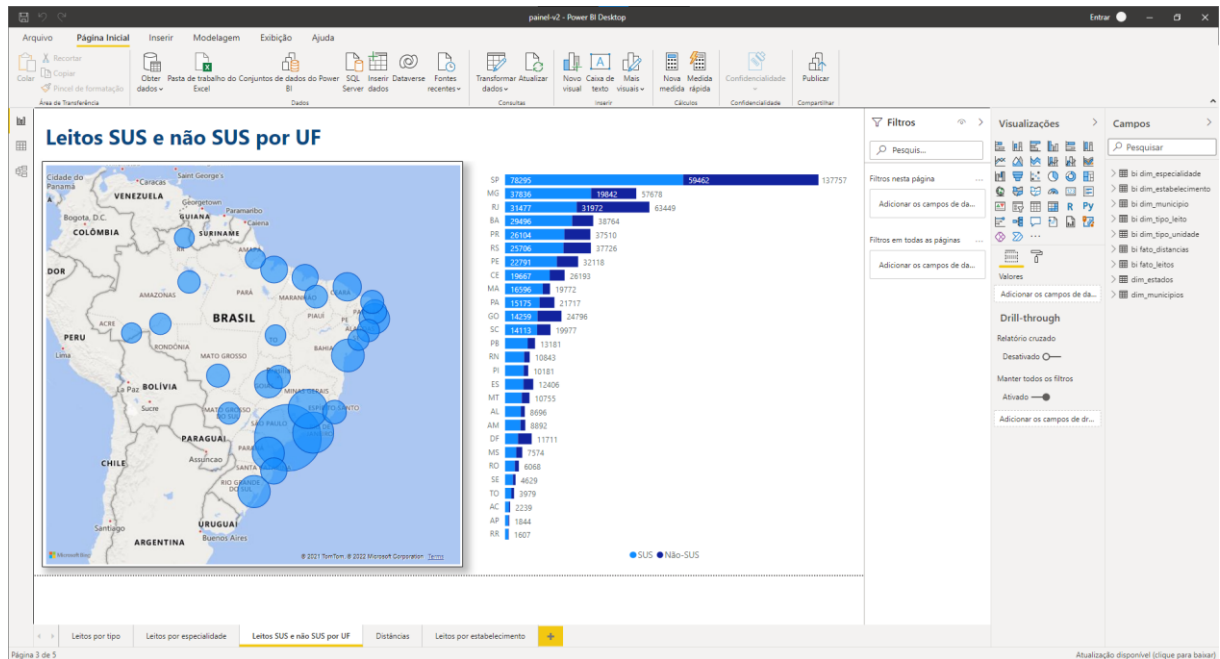
Figura 19 - Ícone que permite ativar a interação entre as visualizações



Fonte: Elaborado pelo autor

Nestas duas visualizações foi necessário filtrar os dados da coluna *sigla_estado* da tabela *bi.dim_municipio* para que os valores *AZ*, *TE* ou *UF* não fossem exibidos. Durante a construção das visualizações, percebeu-se que estes valores faziam parte do conjunto de dados utilizado, apesar de não se referirem a nenhum estado brasileiro. O painel criado pode ser visto na Figura 20.

Figura 20 - Painel de leitos SUS e não SUS por UF no Power BI Desktop



Fonte: Elaborado pelo autor

3.3.5 DISTÂNCIAS DE LEITOS POR ESPECIALIDADE

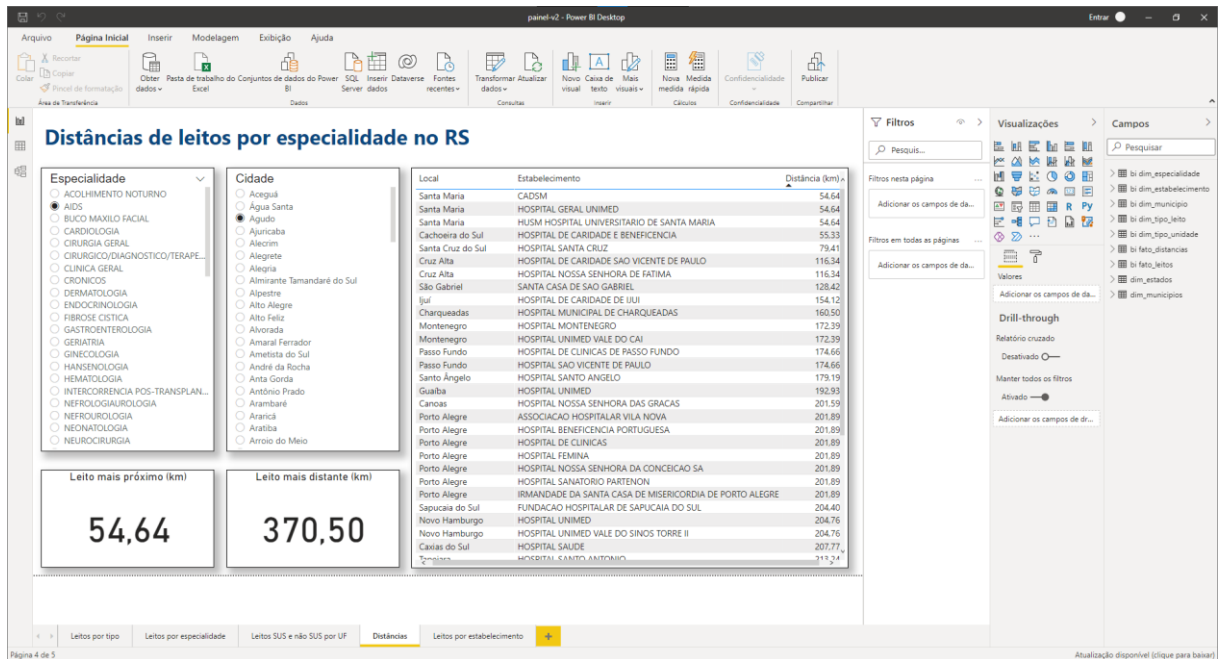
O painel de distâncias de leitos por especialidade possui duas segmentações de dados, a primeira sendo um seletor de especialidades e a segunda um seletor de cidade. Dada a especialidade e a cidade selecionada, uma visualização de tabela mostra todos os estabelecimentos de saúde que possuem um leito da especialidade escolhida, bem como o município onde o estabelecimento se localiza e a distância entre o município selecionado e o município do estabelecimento. Também são exibidos dois cartões, contendo os valores da menor e maior distância para a seleção atual.

No próprio Power BI foram criadas relações entre a tabela *bi.fato_distancias* e as tabelas referenciadas *bi.dim_municipio* e *bi.dim_especialidade*, criando duas novas colunas, *bi.dim_especialidade.descricao* e *bi.dim_municipio.nome_ac*, contendo respectivamente a descrição da especialidade do leito e o nome da cidade. As visualizações de segmentação de dados utilizam estes campos para exibir suas opções.

A filtragem dos dados entre as visualizações é feita utilizando-se a ferramenta *Editar interações* como no gráfico anterior. A diferença é que, ao invés de apenas realçar os dados, eles são filtrados com base no estado dos seletores de cidade e de especialidade.

As visualizações de cards foram configuradas para exibir o valor máximo e mínimo da coluna *distancia* da tabela *bi.fato_distancias*, também utilizando o filtro de interação. O resultado pode ser visto na Figura 21.

Figura 21 - Painel de distâncias de leitos por especialidade no RS no Power BI Desktop



Fonte: Elaborado pelo autor

3.3.6 LEITOS POR ESTABELECIMENTO

O painel de leitos por estabelecimento é uma visualização em forma de tabela, contendo o nome fantasia do estabelecimento, sua razão social, a cidade e estado onde se localiza, e a quantidade de leitos SUS e total. O propósito deste painel é listar os dados de diferentes tabelas em uma mesma visualização, a fim de que fosse possível estudar como cada ferramenta trabalha com dados de um relacionamento muitos para muitos.

O campo de valores da visualização deste painel foi configurado para exibir as colunas *nome_fantasia* e *razao_social* da tabela *bi.dim_estabelecimento* (renomeados para *Nome fantasia* e *Razão social*, respectivamente), as colunas *nome_ac* e *sigla_estado* da tabela *bi.dim_municipio* (renomeadas para *Cidade* e *UF*, respectivamente), e o somatório das colunas *qtd_total* e *qtd_sus* da tabela *bi.fato_leitos* (renomeadas para *Total* e *SUS*, respectivamente). O painel pode ser visto na Figura 22.

Figura 22 - Painel de leitos por estabelecimento no Power BI Desktop

Nome fantasia	Razão social	Cidade	UF	Total	SUS
HC DA FALUSP HOSPITAL DAS CLINICAS SAO PAULO	FUNDACAO FACULDADE DE MEDICINA MEC MPAS	São Paulo	SP	1698	1626
SANTA CASA DE BELO HORIZONTE	SANTA CASA DE MISERICORDIA DE BELO HORIZONTE	Belo Horizonte	MG	1287	1287
IMIP	INSTITUTO DE MEDICINA INTEGRAL PROF FERNANDO FIGUEIRA	Recife	PE	1218	1168
HOSPITAL DE BASE DE SAO JOSE DO RIO PRETO	FUNDACAO FACULDADE REGIONAL DE MEDICINA DE SAO JOSE DO RIO P	São José do Rio Preto	SP	1207	759
SANTA CASA DE SAO PAULO HOSPITAL CENTRAL SAO PAULO	IRMANDADE DA SANTA CASA DE MISERICORDIA DE SAO PAULO	São Paulo	SP	1168	1150
HOSPITAL NOSSA SENHORA DA CONCEICAO SA	HOSPITAL NOSSA SENHORA DA CONCEICAO SA	Porto Alegre	RS	1106	1032
IRMANDADE DA SANTA CASA DE MISERICORDIA DE PORTO ALEGRE	IRMANDADE DA SANTA CASA DE MISERICORDIA DE PORTO ALEGRE	Porto Alegre	RS	1052	591
HOSPITAL DO ROCIO	MATERNIDADE E CIRURGIA NOSSA SENHORA DO ROCIO S A	Campo Largo	PR	1062	997
HOSPITAL SANTO ANTONIO	ASSOCIACAO OBRAS SOCIAIS IRMA DULCE	Salvador	BA	954	954
HOSPITAL DAS CLINICAS FAIPA RIBEIRAO PRETO	FUNDACAO DE APOIO AO ENSINO PESQUISA E ASSISTENCIA HCFMRP	Ribeirão Preto	SP	942	895
HOSPITAL SAO PAULO HOSPITAL DE ENSINO DA UNIFESP SAO PAULO	S'PM ASSOCIACAO PAULISTA PARA O DESENVOLVIMENTO DA MEDICINA	São Paulo	SP	915	871
HOSPITAL BP	REAL E BENEFICENTE ASSOCIACAO PORTUGUESA DE BENEFERENCIA	São Paulo	SP	907	252
HOSPITAL DE CLINICAS	HOSPITAL DE CLINICAS DE PORTO ALEGRE	Porto Alegre	RS	904	828
HOSPITAL DO SERV PUB EST FCO MORATO DE OLIVEIRA SAO PAULO	INSTITUTO DE ASSIST MEDICA DO SERVIDOR PUBLICO ESTADUAL SP	São Paulo	SP	886	6
HOSPITAL DA RESTAURACAO	SECRETARIA DE SAUDE	Recife	PE	830	822
INSTITUTO AMERICO BAIRRAL DE PSIQUIATRIA	FUNDACAO ESPIRITA AMERICO BAIRRAL	Rapira	SP	808	491
REAL HOSPITAL PORTUGUES	REAL HOSPITAL PORTUGUES DE BENEFERENCIA EM PE	Recife	PE	804	240
HOSPITAL DE BASE DO DISTRITO FEDERAL	HOSPITAL DE BASE DO DISTRITO FEDERAL	Brasília	DF	791	763
HOSPITAL DE CAMPANHA COVID 19 HOSP CAMP MG UNID EXPOMINAS	POLICIA MILITAR DO ESTADO DE MINAS GERAIS	Belo Horizonte	MG	768	768
HGF HOSPITAL GERAL DE FORTALEZA	SECRETARIA ESTADUAL DE SAUDE DO CEARÁ	Fortaleza	CE	754	750
UF INSTITUTO DR JOSE FROTA CENTRAL	INSTITUTO DR JOSE FROTA CENTRAL	Fortaleza	CE	745	735
HOSPITAL GERAL ROBERTO SANTOS	BAHIA SECRETARIA DE SAUDE DO ESTADO	Salvador	BA	737	717
HOSPITAL SANTA MARCELINA SAO PAULO	CASA DE SAUDE SANTA MARCELINA	São Paulo	SP	730	600
HOSPITAL SAO VICENTE DE PAULO	ASSOCIACAO HOSPITALAR BENEFICENTE SAO VICENTE DE PAULO	Passo Fundo	RS	712	407
COMPLEXO HOSPITAL DE CLINICAS	UNIVERSIDADE FEDERAL DO PARANA	Curitiba	PR	711	711
HOSPITAL GOVERNADOR JOAO ALVES FILHO	FUNDACAO HOSPITALAR DE SAUDE	Araçáçú	SE	702	701
HOSPITAL UNIVERSITARIO EVANGELICO MCKENZIE	INSTITUTO PRESBITERIANO MCKENZIE	Curitiba	PR	696	569
SANTA CASA	ASSOCIACAO BENEFICENTE SANTA CASA DE CAMPO GRANDE	Campo Grande	MS	687	567
Total				632062	413173

Fonte: Elaborado pelo autor

3.4 DATA STUDIO

O Data Studio é a ferramenta do Google para visualização de dados, fazendo parte da *Google Marketing Platform*¹⁵. Ele é disponibilizado gratuitamente e, por ser uma plataforma *web*, pode ser acessado de qualquer computador conectado à internet, bastando que o usuário tenha uma conta do Google.

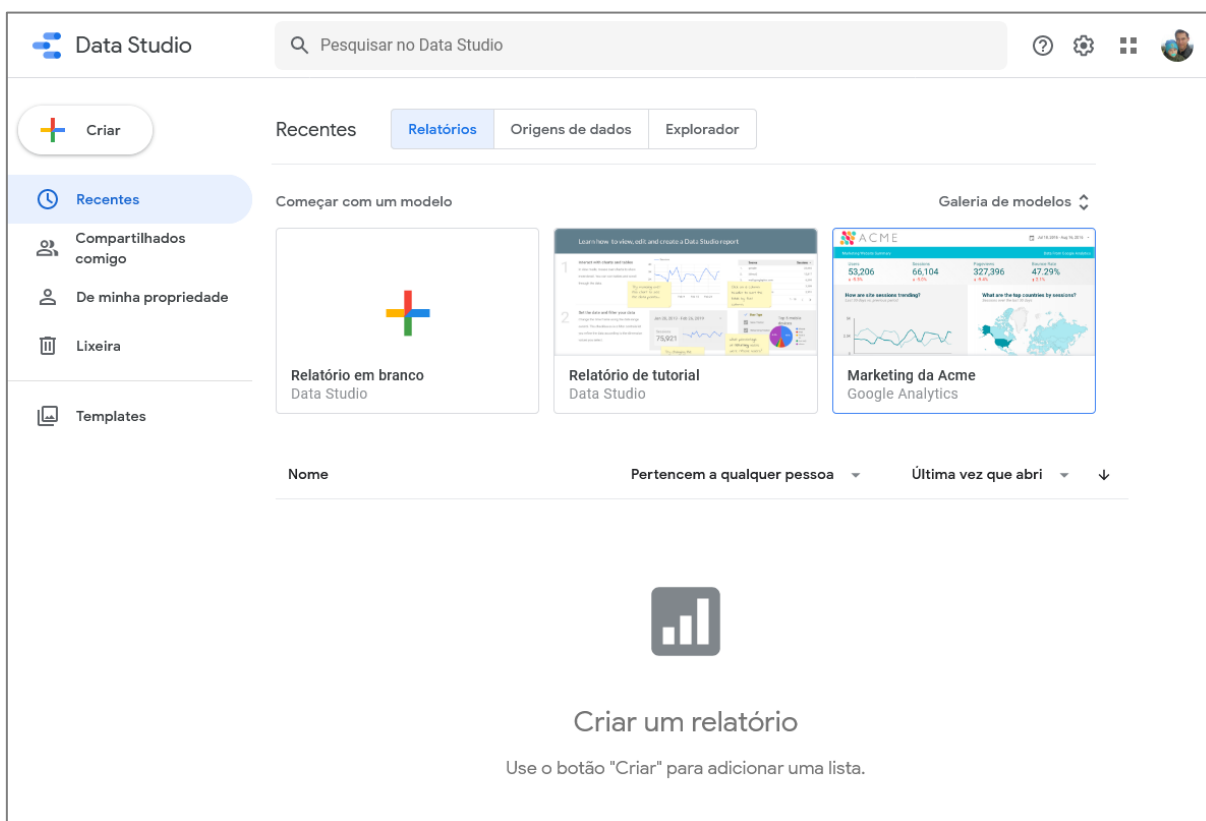
Segundo o website da *Google Marketing Platform*, “o Data Studio foi criado para usuários casuais, mas apresenta tantos recursos fáceis de usar que até as equipes de análise experientes gostarão de conhecer”. Desta forma, pode-se perceber o foco da ferramenta na facilidade de uso principalmente por usuários casuais.

Ele se conecta a diferentes fontes de dados através de conectores, que podem ser desenvolvidos pelos usuários da plataforma. O Google fornece diversos conectores nativos para fontes de dados mais comuns, como banco de dados MySQL e PostgreSQL, bem como para outras ferramentas da empresa, como YouTube Analytics, Search Console, Google Sheets e BigQuery, facilitando a integração entre elas.

Como se pode ver na Figura 23, a interface gráfica do Data Studio segue a mesma identidade visual das demais ferramentas da empresa, sendo familiar ao usuário de outras soluções de uso geral, como o Gmail.

¹⁵ <https://marketingplatform.google.com/about/>

Figura 23 - Tela inicial do Google Data Studio



Fonte: Elaborado pelo autor

3.4.1 ACESSO À BASE DE DADOS LOCAL A PARTIR DA WEB

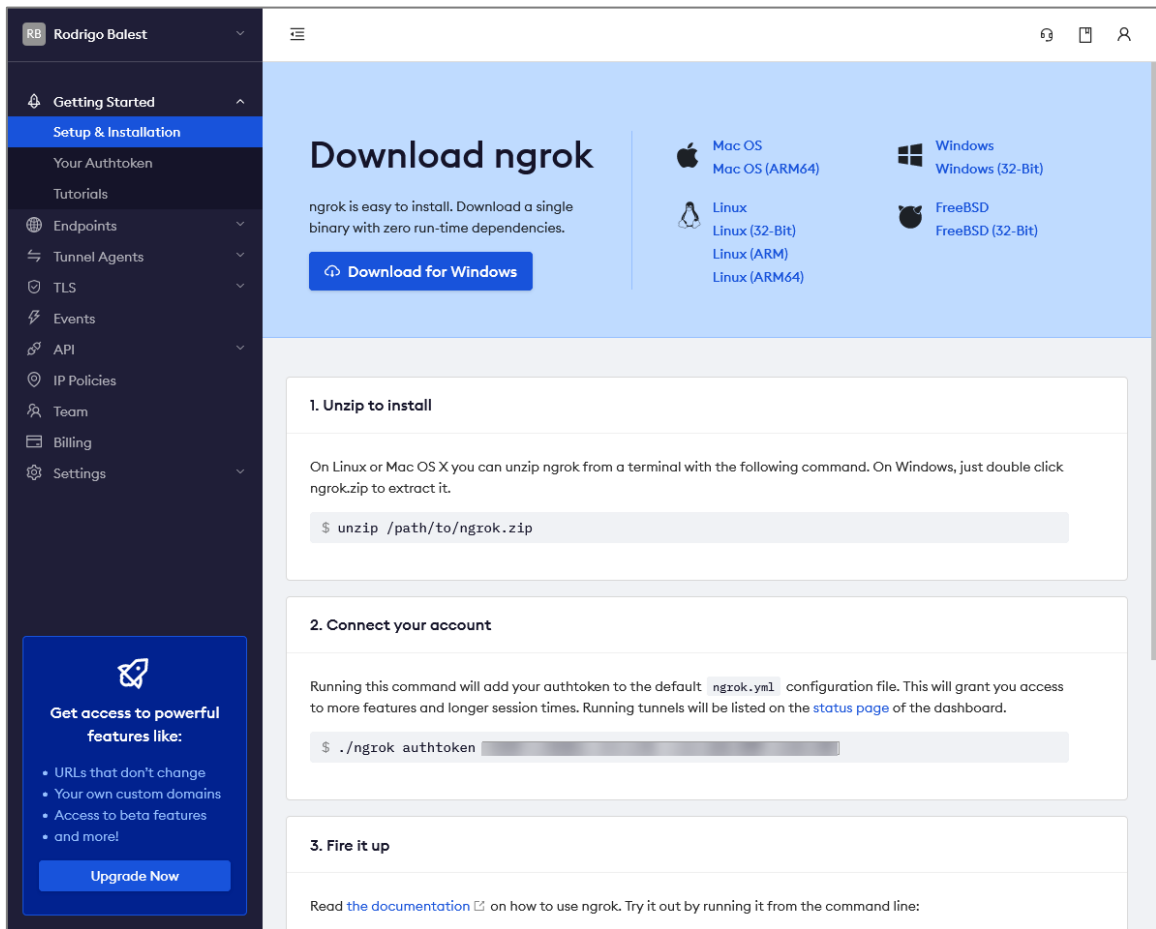
Como o Data Studio é uma ferramenta baseada na web e acessível através do navegador, as fontes de dados que serão acessadas devem estar disponíveis também na web. Neste caso, os dados estão em um banco de dados em uma máquina local. Por isso fez-se necessário encontrar uma forma de tornar o banco de dados acessível a partir da web. Para isso, utilizou-se o aplicativo Ngrok¹⁶.

O Ngrok é um aplicativo que expõe servidores rodando em uma máquina local atrás de NATs¹⁷ ou firewalls possam ser acessados na Internet através de túneis de rede seguros. Assim, foi criada uma conta no site do Ngrok e foi baixado o aplicativo para Windows, que é um executável de linha de comando. A seguir, conforme instruções no painel do usuário do Ngrok (vide Figura 24), foi executado o comando `ngrok authtoken <token de autenticação>` para associar a máquina local com a conta de usuário. O token de autenticação é fornecido pelo painel do usuário.

¹⁶ <https://ngrok.com/>

¹⁷ *Network Address Translation* é uma técnica que roteamento que traduz endereços IP entre uma rede pública e uma rede privada através de um roteador, permitindo a comunicação entre essas redes.

Figura 24 - Painel do usuário da conta do Ngrok



Fonte: Elaborado pelo autor

Com o servidor local do PostgreSQL em execução, foi executado o comando `ngrok tcp 5432`, o qual expõe a porta 5432 pelo protocolo tcp. A saída do terminal pode ser vista na Figura 25, enquanto na Figura 26 pode-se ver a interface web do aplicativo (acessível pelo endereço `http://127.0.0.1:4040`), exibindo algumas informações sobre as conexões ativas.

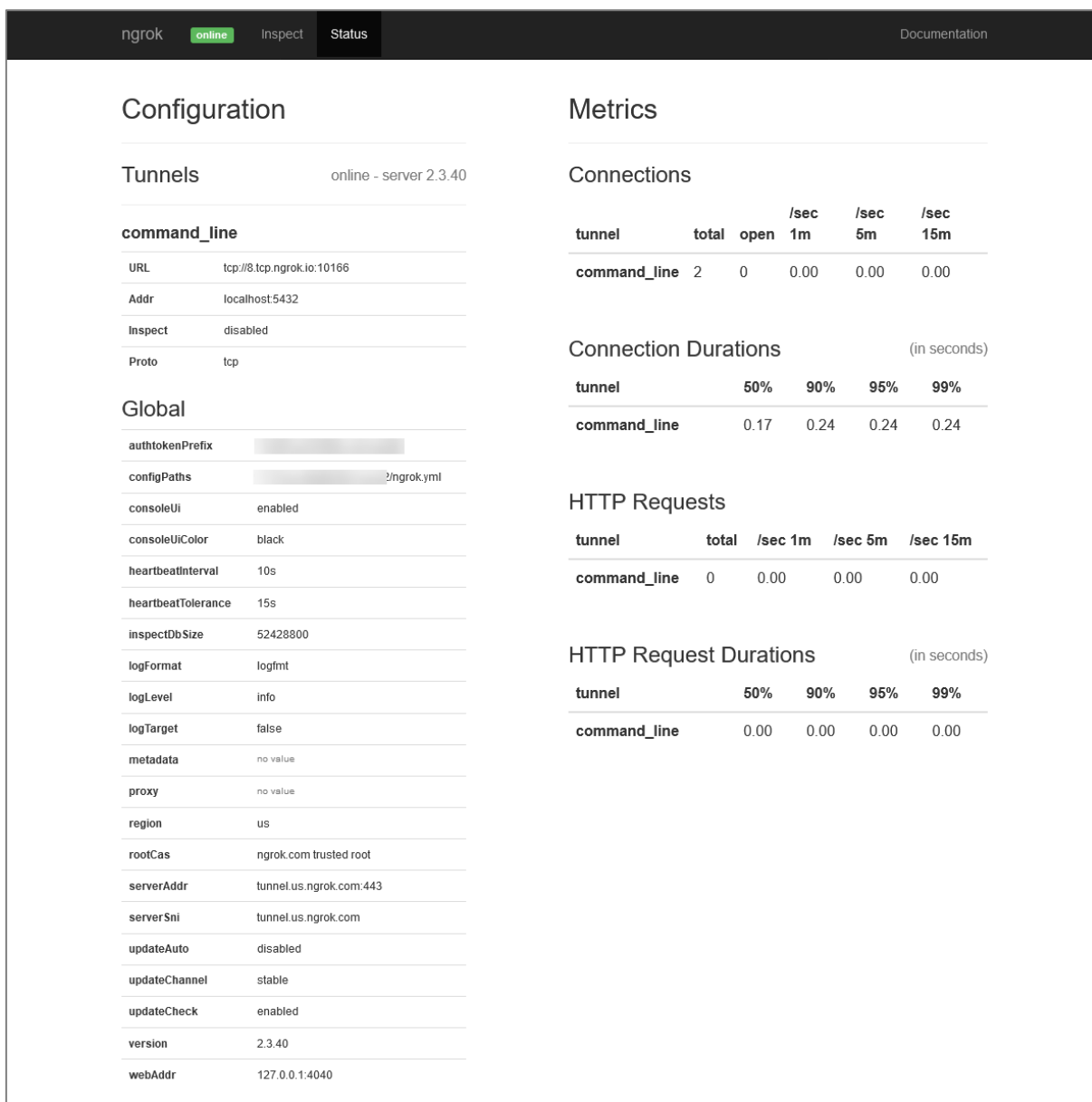
Figura 25 - Ngrok em execução, exibindo as conexões ativas.

```
ngrok by @inconshreveable (Ctrl+C to quit)
Session Status      online
Account             Rodrigo Balest (Plan: Free)
Version             2.3.40
Region              United States (us)
Web Interface        http://127.0.0.1:4040
Forwarding           tcp://8.tcp.ngrok.io:10166 -> localhost:5432

Connections         ttl    opn    rt1    rt5    p50    p90
0                   0      0.00  0.00  0.00  0.00
```

Fonte: Elaborado pelo autor

Figura 26 - Interface web do Ngrok exibindo informações das conexões



Fonte: Elaborado pelo autor

Importante ressaltar que nas contas gratuitas como a utilizada no desenvolvimento deste trabalho, o Ngrok gera uma URL pública diferente cada vez que o comando é executado. Assim, a URL exibida nas figuras deste trabalho pode diferir de uma figura para outra. Nos planos pagos é possível ter domínios reservados, bem como subdomínios na URL pública.

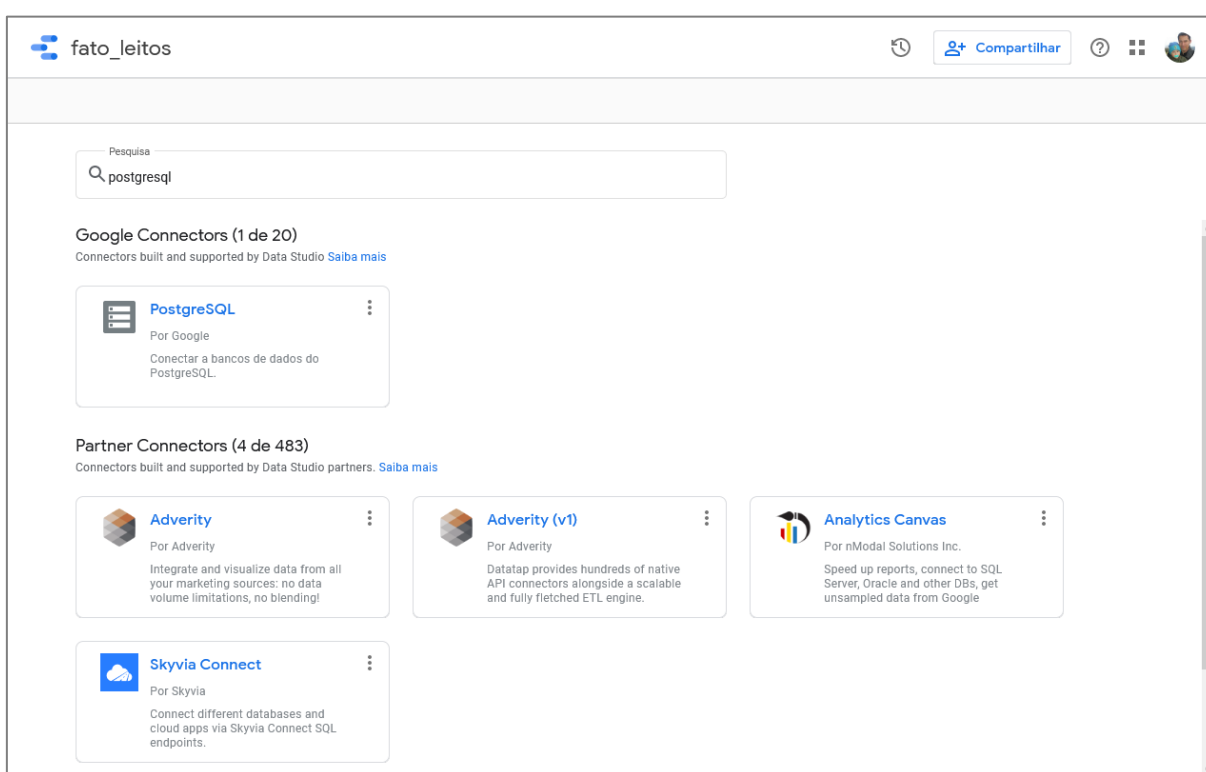
Como no plano gratuito não é possível utilizar uma URL estática, esta abordagem causou o inconveniente de precisar que as URLs das fontes de dados fossem alteradas frequentemente. Por isso, esta abordagem foi utilizada durante parte do desenvolvimento deste trabalho. Mais tarde, preferiu-se a utilização de um serviço de banco de dados Postgres online. Diferentemente do Power BI, o Data Studio precisa de conexão constante às fontes de dados durante o desenvolvimento dos painéis.

As seções seguintes descrevem como foi feita a carga de dados no Google Data Studio, bem como o desenvolvimento dos painéis.

3.4.2 CARGA DE DADOS

A primeira ação necessária foi a criação de uma origem de dados no Data Studio. Para isso, na interface inicial do Data Studio¹⁸ foi clicado o botão *Criar* e selecionada a opção *Origem de dados*. Na tela seguinte, exibida na Figura 27, foi pesquisado e selecionado o conector PostgreSQL. No topo da página o nome da fonte de dados foi alterado de *Fonte de dados sem título* para *fato_leitos*.

Figura 27 - Tela de pesquisa de conectores do Data Studio



Fonte: Elaborado pelo autor

Na próxima tela foram informados os parâmetros da conexão. Ao clicar no botão *Autenticar*, as tabelas do banco de dados são exibidas e podem ser selecionadas como fonte de dados. Uma diferença que se pode notar aqui é que, diferentemente do Power BI, o Data Studio permite que apenas uma tabela ou consulta SQL possa ser uma fonte de dados. No Power BI é possível selecionar várias tabelas e definir os relacionamentos entre elas. A Figura 28 mostra esta tela com as tabelas listadas.

¹⁸ <https://datastudio.google.com/u/0/navigation/reporting>

Figura 28 - Dados da conexão e listagem das tabelas do banco de dados na criação de uma fonte de dados no Data Studio.

The screenshot shows the Data Studio interface for creating a PostgreSQL data source. The top bar includes the 'fato_leitos' logo, a 'Compartilhar' button, and a 'CONECTAR' button. The main area is titled 'SELECIONAR CONECTOR' and features the PostgreSQL connector card. Below this, the configuration is split into two panels: 'BÁSICO' and 'TABELAS'.

BÁSICO

Autenticação do banco de dados

URL JDBC

IP ou nome do host: 0.tcp.ngrok.io

Porta (opcional): 17521

Banco de dados: cnes

Nome de usuário: postgres

Senha: ●●●

Ativar SSL

AUTENTICAR

TABELAS

CONSULTA PERSONALIZADA

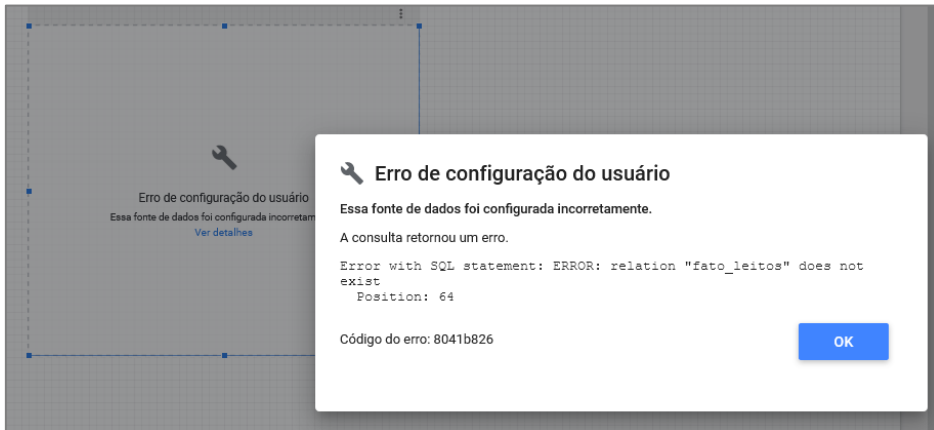
Tabela

- dim_especialidade
- dim_estabelecimento
- dim_municipio
- dim_tipo_leito
- dim_tipo_unidade
- fato_leitos
- rL_estab_complementar
- tb_tributo
- tb_estabelecimento
- tb_leito
- tb_motivo_desativacao
- tb_municipio
- tb_natureza_juridica
- tb_tipo_unidade

Fonte: Elaborado pelo autor

Um problema encontrado aqui é que, ao selecionar uma tabela, ocorreram erros ao criar um gráfico com a fonte de dados (Figura 29). Para que o erro não ocorresse, foi necessário informar uma consulta personalizada que seleciona todos os dados da tabela (por exemplo, *select * from bi.fato_leitos*). Não fica claro qual a razão do erro, mas a mensagem parece indicar que o Data Studio não seleciona o *schema* do banco de dados ao qual a tabela pertence.

Figura 29 - Mensagem de erro exibida ao utilizar uma fonte de dados que usa uma tabela do PostgreSQL.



Fonte: Elaborado pelo autor.

A seguir é exibida a tela de definição de campos e parâmetros. Nesta tela é possível criar novos campos na fonte de dados a partir dos campos já existentes. No caso deste trabalho, foram mantidos os campos originais e nenhum campo novo foi criado. Para encerrar a edição dos campos, clica-se no botão *Criar relatório*, que leva para uma página de criação de relatório. Esta tela pode ser vista na Figura 30.

Figura 30 - Tela de edição de campos da fonte de dados.

Campo	Tipo	Agregação padrão	Descrição
DIMENSÕES (5)			
cod_especialidade	123 Número	Soma	
cod_estabelecimento	123 Número	Soma	
cod_tipo_jeito	123 Número	Soma	
qtz_sus	123 Número	Soma	
qtz_total	123 Número	Soma	
MÉTRICAS (1)			
Record Count	123 Número	Automático	

Fonte: Elaborado pelo autor

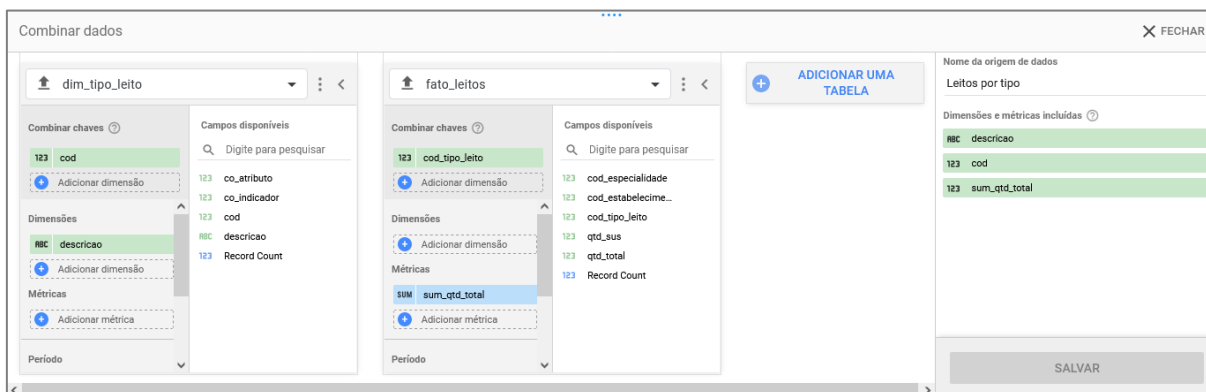
Estas etapas foram utilizadas para criar uma fonte de dados para cada uma das tabelas do modelo dimensional.

No Data Studio foram replicados os mesmos painéis criados no Power BI. A seguir é relatado como foi feito o desenvolvimento deles.

3.4.3 LEITOS POR TIPO

O primeiro painel é um gráfico em forma de rosca que exibe a quantidade de leitos por tipo. Para isso, foi selecionado o gráfico de rosca na barra de ferramentas, na opção *Adicionar gráfico*. No campo *Origem dos dados* foi selecionada a opção *Combinar dados*. Da fonte de dados *dim_tipo_leito* foram selecionadas as colunas *cod* como chave e *descricao* como dimensão, e da fonte de dados *fato_leitos* foram selecionadas as colunas *cod_tipo_leito* como chave e a coluna *qtd_total* como métrica, sendo renomeada para *sum_qtd_total*. Esta combinação de dados foi nomeada *Leitos por tipo*. A Figura 31 mostra a configuração desta combinação de dados.

Figura 31 - Combinação de dados para obtenção da quantidade de leitos por tipo.

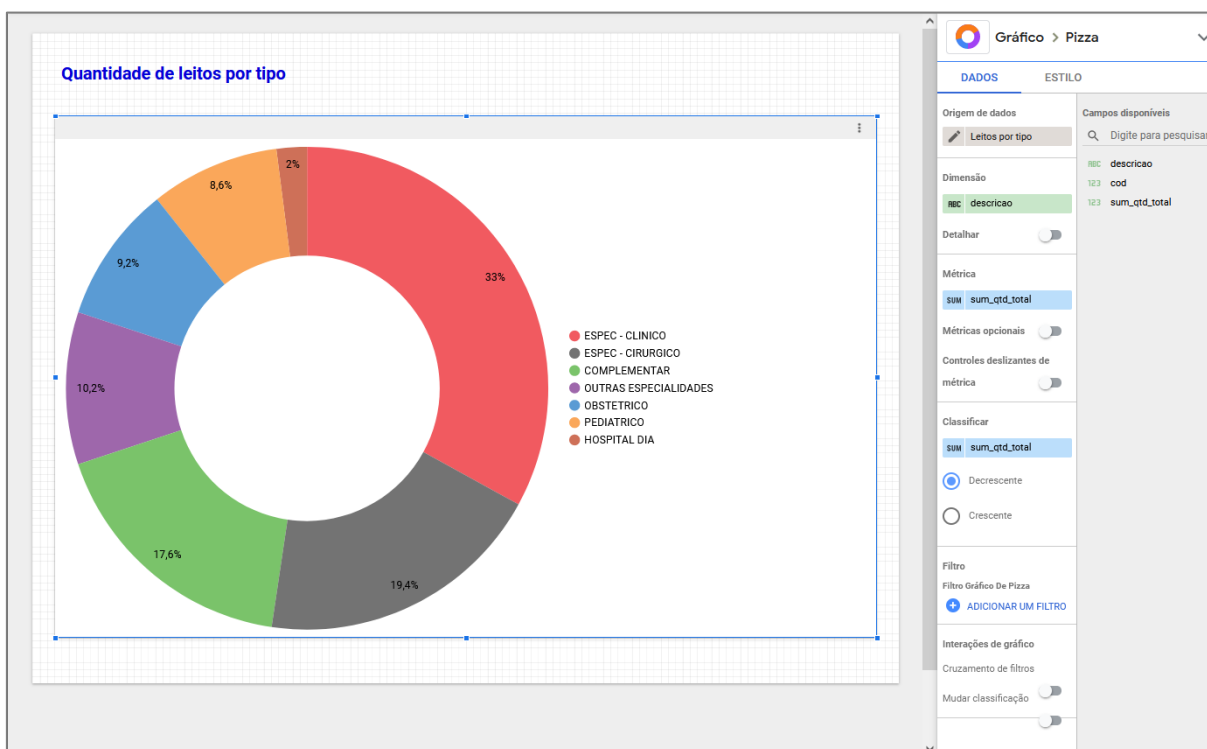


Fonte: Elaborado pelo autor.

Aqui é possível notar uma outra diferença do Google Data Studio em relação ao Power BI. Enquanto no Power BI as relações entre tabelas são inferidas a partir do modelo de dados pelos nomes dos campos (que pode ser configurado na aba *Modelo*), no Data Studio é necessário criar as combinações de dados manualmente, formando uma nova fonte de dados. No Power BI é possível utilizar várias tabelas na composição de um gráfico, já no Data Studio cada gráfico possui apenas uma fonte de dados.

A fonte de dados *Leitos por tipo* foi então utilizada para a criação do primeiro gráfico, o qual foi configurado para utilizar o campo *descricao* como dimensão enquanto o campo *sum_qtd_total* foi utilizado como métrica e como classificador dos dados. O painel de leitos por tipo pode ser visto em modo de edição na Figura 32, onde pode-se observar a configuração dos campos.

Figura 32 - Configuração do painel de leitos por tipo.

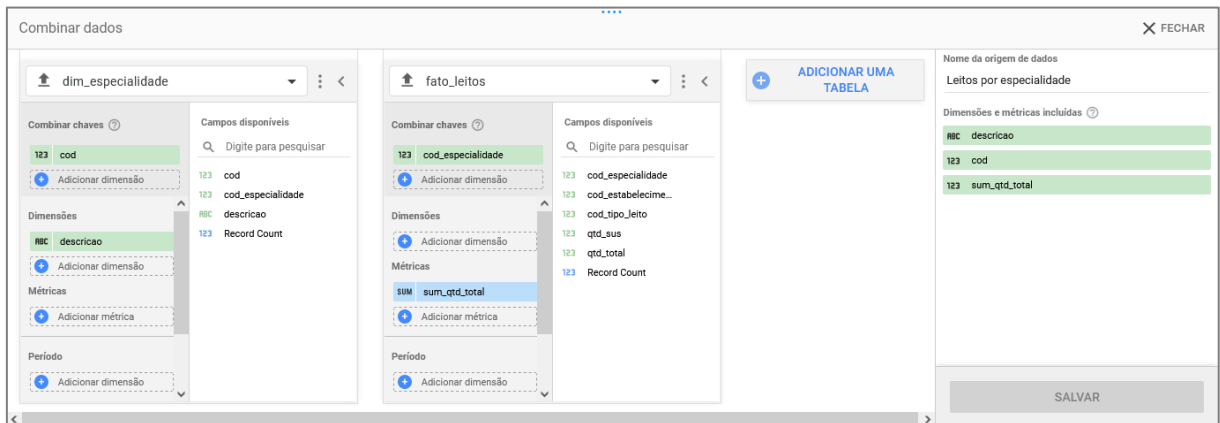


Fonte: Elaborado pelo autor.

3.4.4 LEITOS POR ESPECIALIDADE

O painel criado a seguir exibe a quantidade de leitos por especialidade. Da mesma forma que no painel anterior, foi criada uma combinação de dados. Nesta combinação, nomeada *Leitos por especialidade*, foram selecionadas as fontes de dados *dim_especialidade*, com o campo *cod* como chave e o campo *descricao* como dimensão, e a fonte de dados *fato_leitos* com o campo *cod_especialidade* como chave e o campo *qtd_total* como métrica, renomeado para *sum_qtd_total*. A Figura 33 exibe esta configuração.

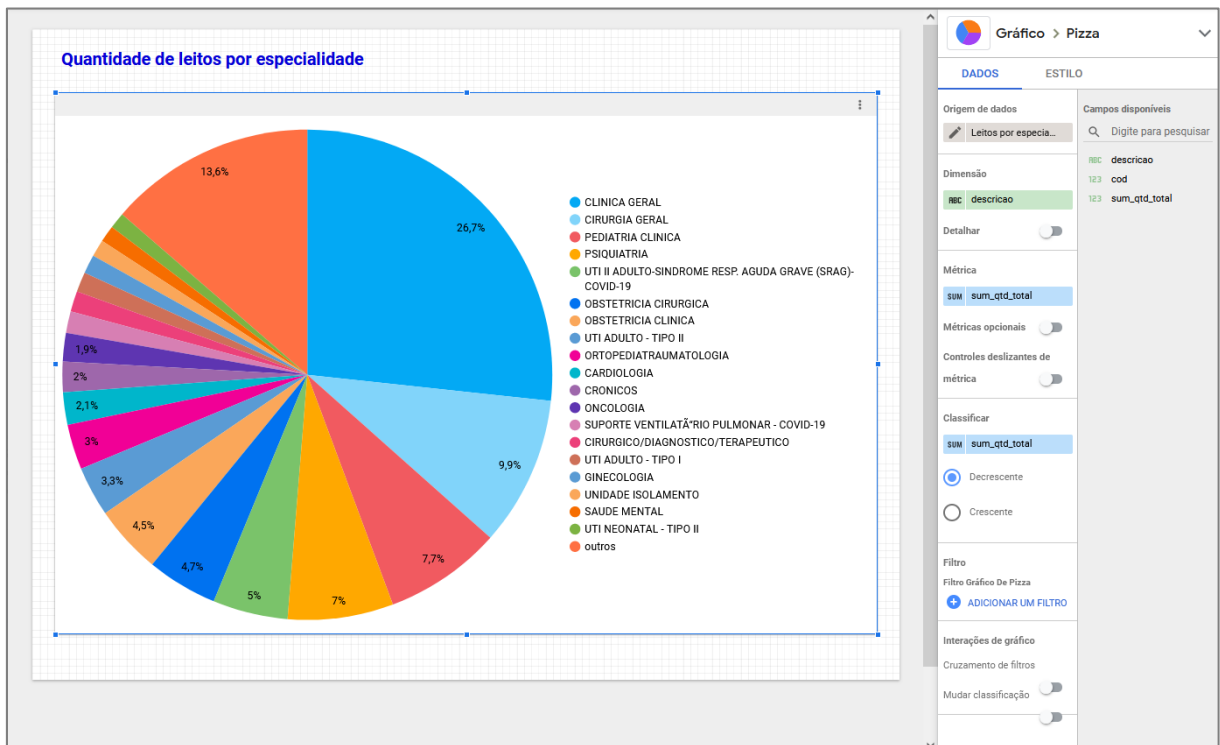
Figura 33 - Combinação de dados para obtenção da quantidade de leitos por especialidade.



Fonte: Elaborado pelo autor.

O gráfico criado é do tipo gráfico de pizza. Ele foi configurado para usar a combinação de dados *Leitos por especialidade* como fonte de dados, com o campo *descricao* como dimensão e o campo *sum_qtd_total* como métrica e como campo de classificação. A configuração deste painel pode ser vista abaixo, na Figura 34.

Figura 34 - Configuração do painel de leitos por especialidade.



Fonte: Elaborado pelo autor.

3.4.5 LEITOS SUS E NÃO SUS POR UF

Para exibir este painel, foi necessário relacionar os dados de três tabelas: *bi.fato_leitos*, *bi.dim_estabelecimento* e *bi.dim_municipio*. Como as tabelas não possuem uma chave em comum, foi necessário utilizar uma consulta personalizada como fonte de dados, a qual pode ser vista na Figura 35.

Figura 35 - Consulta personalizada para a fonte de dados de leitos por UF.

TABELAS	Enter Custom Query
CONSULTA PERSONALIZADA	<pre> 1 SELECT 2 sum(fl.qtd_sus) sum_qtd_sus, 3 sum(fl.qtd_total) sum_qtd_total, 4 dm.sigla_estado 5 FROM 6 bi.fato_leitos fl 7 LEFT JOIN 8 bi.dim_estabelecimento de 9 ON de.cod = fl.cod_estabelecimento 10 LEFT JOIN 11 bi.dim_municipio dm 12 ON de.Cod_municipio = dm.cod 13 GROUP BY dm.sigla_estado 14 </pre>

Fonte: Elaborado pelo autor.

Este painel exibe a quantidade de leitos não SUS, o que não faz parte da fonte de dados. Para isso foi criado um campo personalizado, que obtém o valor desejado subtraindo a quantidade de leitos SUS da quantidade total de leitos. Esta operação poderia ter sido realizada diretamente na consulta personalizada, mas foi feita através de um campo personalizado para demonstrar esta funcionalidade do Data Studio. A Figura 36 mostra a configuração deste campo personalizado.

Figura 36 - Configuração do campo personalizado para quantidade de leitos não SUS.

<p>Campos disponíveis</p> <ul style="list-style-type: none"> REC localizacao REC sigla_estado 123 sum_qtd_sus 123 sum_qtd_total 123 Record Count 	<p>Nome do campo</p> <p>sum_qtd_ao_sus</p> <p>Código do campo</p> <p>calc_p9opxdq6qc</p> <p>Fórmula ?</p> <p>FORMATAR FÓRMULA</p> <p>1 sum_qtd_total - sum_qtd_sus</p> <p><input checked="" type="checkbox"/></p> <p>FECHAR ATUALIZAR</p>
---	--

Fonte: Elaborado pelo autor.

Um dos gráficos deste painel exibe um mapa do Brasil com a quantidade de leitos em cada estado, que é uma visualização do Google Maps. Para a localização dos estados, foi utilizada a coluna da sigla do

estado, mas o Google Maps não conseguiu reconhecer algumas delas. A solução foi a criação de outro campo personalizado, cuja configuração pode ser vista na Figura 37. Com este campo, a sigla de alguns estados é substituída pelo nome do estado por extenso, e todos os valores são sufixados com o texto “, Brasil”.

Figura 37 - Configuração do campo personalizado *localizacao*.

The screenshot shows the configuration interface for a custom field in Power BI. On the left, a list of available fields includes 'sigla_estado', 'sum_qtd_nao_sus', 'sum_qtd_sus', 'sum_qtd_total', and 'Record Count'. The main configuration area shows the field name 'localizacao' and the code 'calc_4ikeo0q6qc'. The formula editor contains the following DAX formula:

```

1 CASE sigla_estado
2   WHEN 'AP' THEN 'Amapá, Brasil'
3   WHEN 'PE' THEN 'Pernambuco, Brasil'
4   WHEN 'AL' THEN 'Alagoas, Brasil'
5   WHEN 'PR' THEN 'Paraná, Brasil'
6   WHEN 'ES' THEN 'Espírito Santo, Brasil'
7   ELSE CONCAT( sigla_estado, ", Brasil")
8 END

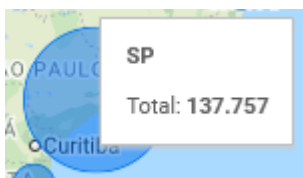
```

Buttons for 'FECHAR' and 'ATUALIZAR' are visible at the bottom right.

Fonte: Elaborado pelo autor.

A visualização do Google Maps foi configurada com a coluna *localizacao* como valor do campo *Local* e a coluna *sigla_estado* como valor do campo *Dica*. A coluna *sum_qtd_total* foi atribuída ao campo *Tamanho* e renomeada para *Total*. Diferentemente do Power BI, somente foi possível atribuir uma coluna ao campo *Dica*. Assim, ao passar o mouse sobre os pontos do mapa, apenas a sigla do estado e a quantidade total de leitos no estado é exibida, conforme visto na Figura 38.

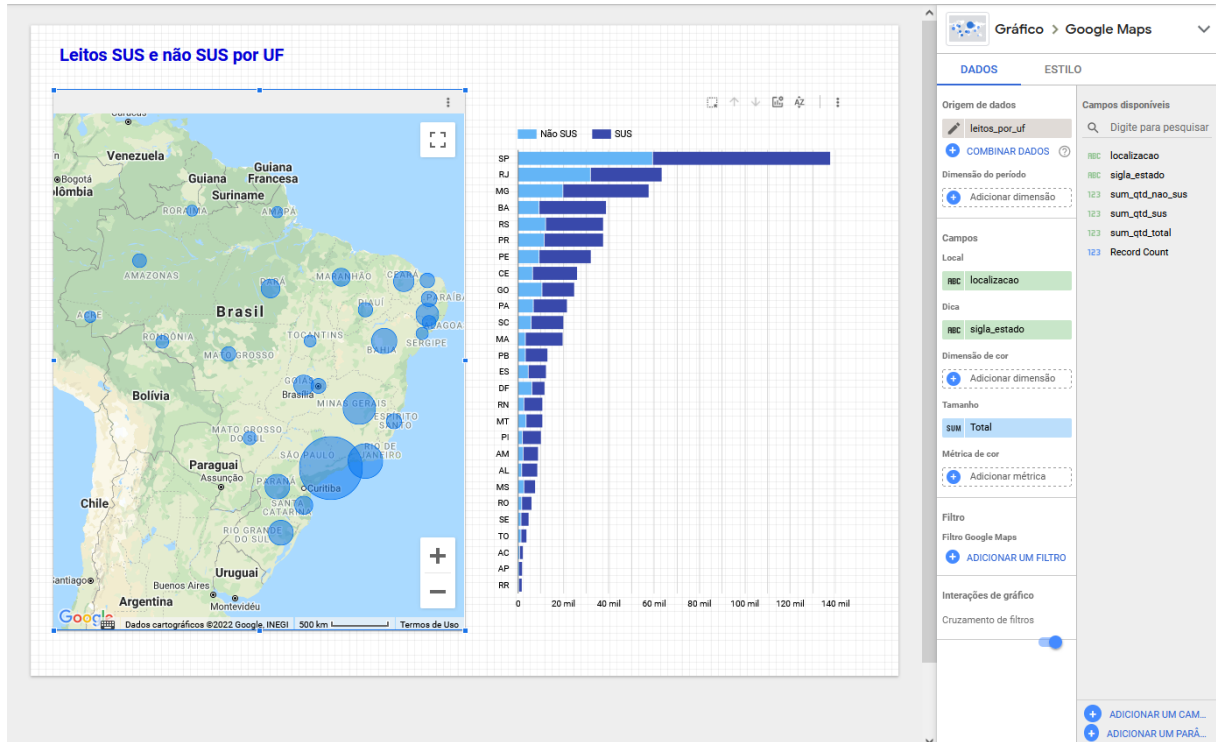
Figura 38 – Dica sendo exibida ao passar o mouse sobre um dos pontos do mapa.



Fonte: Elaborado pelo autor.

Já a configuração do gráfico de barras empilhadas foi feita utilizando a coluna *sigla_estado* como dimensão e as colunas *sum_qtd_nao_sus* e *sum_qtd_sus* como métricas, respectivamente renomeadas para *Não SUS* e *SUS*. A classificação foi feita pela coluna *sum_qtd_total*. O painel pode ser visto na Figura 39.

Figura 39 – Painel de leitos SUS e não SUS por UF.



Fonte: Elaborado pelo autor.

3.4.6 DISTÂNCIAS DE LEITOS POR ESPECIALIDADE

Este painel também possui dados originados em mais de uma tabela, e por este motivo a fonte de dados utiliza uma consulta personalizada, que pode ser vista na Figura 40.

Figura 40 – Consulta personalizada para a fonte de dados de distâncias de leitos por especialidade.

TABELAS

CONSULTA PERSONALIZADA

Enter Custom Query

```

1 SELECT
2   esp.descricao AS especialidade,
3   cid.nome_ac AS cidade,
4   estab.nome_fantasia AS estabelecimento,
5   local.nome_ac as local,
6   dist.distancia
7 FROM
8   bi.fato_distancias dist
9 LEFT JOIN
10  bi.dim_especialidade esp
11  ON dist.cod_especialidade = esp.cod
12 LEFT JOIN
13  bi.dim_municipio cid
14  ON dist.cod_municipio = cid.cod
15 LEFT JOIN
16  bi.dim_estabelecimento estab
17  ON dist.cod_estabelecimento = estab.cod
18 LEFT JOIN
19  bi.dim_municipio local
20  ON estab.cod_municipio = local.cod

```

Fonte: Elaborado pelo autor.

Foram utilizados dois controles do tipo *Lista de tamanho fixo*, um para seleção da cidade de origem e outro para seleção da especialidade desejada. Também foram utilizadas duas visualizações do tipo *Visão geral* para exibir o leito mais próximo e o mais distante, e uma visualização de tabela para exibir os dados filtrados de acordo com a opção selecionada em cada um dos controles. O painel resultante pode ser visto na Figura 41.

Figura 41 – Painel de distâncias de leitos por especialidade.

Distâncias

Especialidade (1)

- ACOLHIMENTO NOTURNO
- AIDS
- BUCO MAXILO FACIAL
- CARDIOLOGIA
- CIRURGIA GERAL
- CIRURGICO/DIAGNOSTICO
- CLINICA GERAL
- CRONICOS
- DERMATOLOGIA
- ENDOCRINOLOGIA
- GASTROENTEROLOGIA
- GERIATRIA
- GINECOLOGIA
- HANSENOLOGIA
- HEMATOLOGIA
- INTERCORRENCIA POS.

Cidade (1)

- Aceguá
- Agudo
- Ajuricaba
- Alcirim
- Alegrete
- Alegria
- Almirante Tamandaré do Sul
- Alpetre
- Alto Alegre
- Alto Feliz
- Alvorada
- Amaral Ferrador
- Ametista do Sul
- André da Rocha
- Anta Gorda
- Antônio Prado

Local	Estabelecimento	Distância
10. Charqueadas	HOSPITAL MUNICIPAL DE CHARQUEADAS	160.49803
11. Montenegro	HOSPITAL MONTENEGRO	172.38528
12. Montenegro	HOSPITAL UNIMED VALE DO CAI	172.38528
13. Passo Fundo	HOSPITAL SAO VICENTE DE PAULO	174.66284
14. Passo Fundo	HOSPITAL DE CLINICAS DE PASSO FUNDO	174.66284
15. Santo Ângelo	HOSPITAL SANTO ANGELO	179.1877
16. Guaíba	HOSPITAL UNIMED	192.9322
17. Canoas	HOSPITAL NOSSA SENHORA DAS GRACAS	201.5874
18. Porto Alegre	IRMANDADE DA SANTA CASA DE MISERICORDIA DE ...	201.88756
19. Porto Alegre	ASSOCIACAO HOSPITALAR VILA NOVA	201.88756
20. Porto Alegre	HOSPITAL SANATORIO PARTENON	201.88756
21. Porto Alegre	HOSPITAL NOSSA SENHORA DA CONCEICAO SA	201.88756
22. Porto Alegre	HOSPITAL FEMINA	201.88756
23. Porto Alegre	HOSPITAL DE CLINICAS	201.88756
24. Porto Alegre	HOSPITAL BENEFICENCIA PORTUQUESA	201.88756
25. Sapucaia do Sul	FUNDACAO HOSPITALAR DE SAPUCAIA DO SUL	204.39961
26. Novo Hamburgo	HOSPITAL UNIMED	204.76134
27. Novo Hamburgo	HOSPITAL UNIMED VALE DO SINOS TORRE II	204.76134
28. Caxias do Sul	HOSPITAL SAUDE	207.76796
29. Tapejara	HOSPITAL SANTO ANTONIO	213.23907
30. Gravataí	HOSPITAL DOM JOAO BECKER	221.00002
31. Erechim	FUNDACAO HOSPITALAR SANTA TEREZINHA DE ERE...	242.99605
32. Alegrete	SANTA CASA DE ALEGRETE	246.14528
33. Pelotas	HOSPITAL SANTA TEREZINHA DE PELotas	261.10000

Leito mais próximo (km): **54,64**

Leito mais distante (km): **370,50**

Data da última atualização: 20/02/2022 21:41:44

Fonte: Elaborado pelo autor.

3.4.7 LEITOS POR ESTABELECIMENTO

O painel de leitos por estabelecimento exibe uma listagem com dados de mais de uma tabela do banco de dados em uma visualização de tabela simples. Então foi utilizada a consulta personalizada exibida na Figura 42 para a fonte de dados utilizada por esta visualização.

Figura 42 – Consulta personalizada para o painel de leitos por estabelecimento.

TABELAS

CONSULTA PERSONALIZADA

Enter Custom Query

```

1 SELECT
2   estab.nome_fantasia,
3   estab.razao_social,
4   mun.nome_ac as nome,
5   mun.sigla_estado as uf,
6   lto.qtd_total,
7   lto.qtd_sus
8 FROM
9   bi.fato_leitos lto
10 LEFT JOIN
11   bi.dim_estabelecimento estab
12   ON lto.cod_estabelecimento = estab.cod
13 LEFT JOIN
14   bi.dim_municipio mun
15   ON estab.cod_municipio = mun.cod

```

Fonte: Elaborado pelo autor.

As colunas selecionadas para exibição na tabela foram *nome_fantasia*, *razao_social*, *nome*, *uf*, *qtd_total* e *qtd_sus*, renomeadas respectivamente para *Nome fantasia*, *Razão social*, *Cidade*, *UF*, *Total* e *SUS*. O resultado pode ser visto na Figura 43.

Figura 43 – Painel de leitos por estabelecimento.

Leitos por estabelecimento

	Nome fantasia	Razão social	Cidade	UF	Total	SUS
1.	IPSO	INSTITUTO PRONTO SOCORRO DE OTORRINOLARINGOLOGIA LTDA	null	DF	5	0
2.	SER CLINICA DE SAUDE MENTAL	CLINICA INTERDISCIPLINAR EM SAUDE MENTAL LTDA	null	DF	43	0
3.	SANTA CASA DE ABADIA DOS DOURADOS	SANTA CASA DE MISERICORDIA ABADIA DOS DOURADOS	Abadia dos Dourados	MG	3	2
4.	SANTA CASA DE ABADIA DOS DOURADOS	SANTA CASA DE MISERICORDIA ABADIA DOS DOURADOS	Abadia dos Dourados	MG	6	5
5.	HOSPITAL MUNICIPAL NOSSA SENHORA DA ABADIA	MUNICIPIO DE ABADIA DOS DOURADOS	Abadia dos Dourados	MG	9	9
6.	SANTA CASA DE ABADIA DOS DOURADOS	SANTA CASA DE MISERICORDIA ABADIA DOS DOURADOS	Abadia dos Dourados	MG	2	1
7.	HOSPITAL MUNICIPAL NOSSA SENHORA DA ABADIA	MUNICIPIO DE ABADIA DOS DOURADOS	Abadia dos Dourados	MG	3	3
8.	SANTA CASA DE ABADIA DOS DOURADOS	SANTA CASA DE MISERICORDIA ABADIA DOS DOURADOS	Abadia dos Dourados	MG	2	2
9.	UNIDADE MISTA	UNIDADE MISTA	Abadânia	GO	9	9
10.	CLINICA CIRURGICA E MATERNIDADE DR EDILSON SOUZA	EDILSON M DE SOUZA	Abaetetuba	PA	2	0
11.	HOSPITAL JULIA SEFER	HOSPITAL JULIA SEFER SA LTDA	Abaetetuba	PA	16	12
12.	HOSPITAL JULIA SEFER	HOSPITAL JULIA SEFER SA LTDA	Abaetetuba	PA	46	43
13.	HOSPITAL JULIA SEFER	HOSPITAL JULIA SEFER SA LTDA	Abaetetuba	PA	4	4
14.	HOSPITAL REGIONAL DO BAIXO TOCANTINS HOSPITAL SANT...	SECRETARIA EXECUTIVA DE SAUDE PUBLICA	Abaetetuba	PA	11	0
15.	HOSPITAL GERAL DE ABAETETUBA	HOSPITAL GERAL DE ABAETETUBA SC LTDA	Abaetetuba	PA	21	19
16.	CEMA	CENTRO MEDICO DE ABAETETUBA SC LTDA	Abaetetuba	PA	1	0
17.	SINTESE MEDICA SERVICOS	SINTESE MEDICA SERVICOS LTDA EPP	Abaetetuba	PA	2	0
18.	HOSPITAL REGIONAL DO BAIXO TOCANTINS HOSPITAL SANT...	SECRETARIA EXECUTIVA DE SAUDE PUBLICA	Abaetetuba	PA	16	16
19.	HOSPITAL REGIONAL DO BAIXO TOCANTINS HOSPITAL SANT...	SECRETARIA EXECUTIVA DE SAUDE PUBLICA	Abaetetuba	PA	4	4
20.	HOSPITAL REGIONAL DO BAIXO TOCANTINS HOSPITAL SANT...	SECRETARIA EXECUTIVA DE SAUDE PUBLICA	Abaetetuba	PA	8	8
21.	CLINICA CIRURGICA E MATERNIDADE DR EDILSON SOUZA	EDILSON M DE SOUZA	Abaetetuba	PA	5	0
22.	HOSPITAL JULIA SEFER	HOSPITAL JULIA SEFER SA LTDA	Abaetetuba	PA	5	5
23.	CLINICA CIRURGICA E MATERNIDADE DR EDILSON SOUZA	EDILSON M DE SOUZA	Abaetetuba	PA	6	0
24.	HOSPITAL REGIONAL DO BAIXO TOCANTINS HOSPITAL SANT...	SECRETARIA EXECUTIVA DE SAUDE PUBLICA	Abaetetuba	PA	5	5

Gráfico > Tabela

DADOS ESTILO

Origem de dados
leitos_por_estabel...

Campos disponíveis
Digite para pesquisar

COMBINAR DADOS

Dimensão de período
Adicionar dimensão

Dimensão

Nome fantasia

Razão social

Cidade

UF

Total

SUS

Adicionar dimensão

Record Count

Métrica

Adicionar métrica

Métricas opcionais

Controles deslizantes de métrica

Linhas por página

100

Linha de resumo

Exibir a linha de resumo

ADICIONAR UM CAM...

ADICIONAR UM PARÁ...

Fonte: Elaborado pelo autor.

4 COMPARATIVO ENTRE AS FERRAMENTAS

Após o uso prático, foi possível identificar as semelhanças e diferenças entre as ferramentas utilizadas, com seus pontos fortes e pontos fracos. O Microsoft Power BI demonstrou ser mais robusto, possuindo mais recursos, podendo então ser utilizado em uma maior variedade de possibilidades. Já o Google Data Studio apresentou uma simplicidade maior de uso, mas esta vantagem também traz algumas limitações, que são apresentadas no decorrer deste capítulo. O Quadro 1 apresenta os itens analisados e um resumo do resultado da avaliação.

Quadro 1 - Resumo dos resultados da avaliação

Item avaliado	Power BI	Data Studio
Licença de uso	Possui versão gratuita, mas com algumas restrições.	Possui apenas versão gratuita, com todos os recursos disponíveis.
Facilidade de uso	Possui mais recursos, porém, é mais intimidador para iniciantes.	Interface mais simples, familiar para usuários de outras ferramentas do Google
Fontes de dados	Maior diversidade de conectores de dados.	Possui muitos conectores de dados, mas com objetivos semelhantes. Foco em dados de ferramentas de marketing.
Transformações de dados	Possui mais possibilidades de realizar transformações de dados.	Menos possibilidades. Junções de dados limitadas a <i>left outer join</i> .
Visualizações	Visualizações comuns e algumas específicas. Possibilidade de desenvolver visualizações.	Visualizações comuns.

Fonte: Elaborado pelo autor

4.1 LICENÇA DE USO

Para poder utilizar o Google Data Studio, o usuário precisa apenas possuir uma conta do Google. Todos os seus recursos são oferecidos de forma gratuita. Existem apenas alguns conectores de terceiros que cobram pela sua utilização, como os conectores da Supermetrics¹⁹, mas o mesmo não ocorre com os conectores nativos desenvolvidos pelo Google.

Já o Microsoft Power BI oferece as licenças Pro, Premium e Free. A licença gratuita dá acesso às ferramentas de criação de painéis, mas oferece algumas restrições. Nela, não é possível compartilhar relatórios para edição com outros usuários, nem é permitido editar relatórios criados por uma conta Pro ou Premium. Os relatórios podem ser compartilhados somente para visualização na web, mas não há controle de acesso, o que significa que qualquer pessoa pode acessar os dados. Para a utilização com dados abertos,

¹⁹ <https://supermetrics.com/product/data-studio>

esta restrição não é um problema, mas pode ser um impeditivo caso o usuário queira utilizar a ferramenta com dados privados.

4.2 FACILIDADE DE USO

O Microsoft Power BI oferece mais recursos e opções de personalização do que o Google Data Studio, o que dá mais liberdade e possibilidades para o usuário, mas pode ser mais intimidador para usuários iniciantes, se tornando um pouco mais confuso. Já o Google Data Studio possui uma interface com menos controles e mais familiar para quem já é usuário de outras ferramentas do Google.

Por outro lado, algumas operações com dados de fontes diferentes são mais fáceis de serem feitas no Microsoft Power BI. As suas ferramentas de formatação e modelos de dados permitem que os gráficos consigam interpretar os relacionamentos de fontes diferentes automaticamente, o que facilita sua criação. Estas ferramentas são descritas de forma mais detalhada na seção 4.4 *Transformação de Dados*.

A forma de acesso às ferramentas também difere. O Power BI Desktop é um aplicativo desktop, ou seja, precisa ser baixado e instalado na máquina do usuário. A última versão utilizada (2.100.1401.0) necessita de 1,8 GB de espaço em disco segundo a página do produto na Microsoft Store²⁰. O Google Data Studio é um aplicativo baseado na web, logo pode ser acessado através de um navegador a partir de qualquer computador com acesso a internet.

4.3 FONTES DE DADOS

O Google Data Studio realiza acesso a dados através de conectores. Atualmente²¹ existem 21 conectores desenvolvidos pelo Google e aproximadamente 487 conectores desenvolvidos por parceiros.

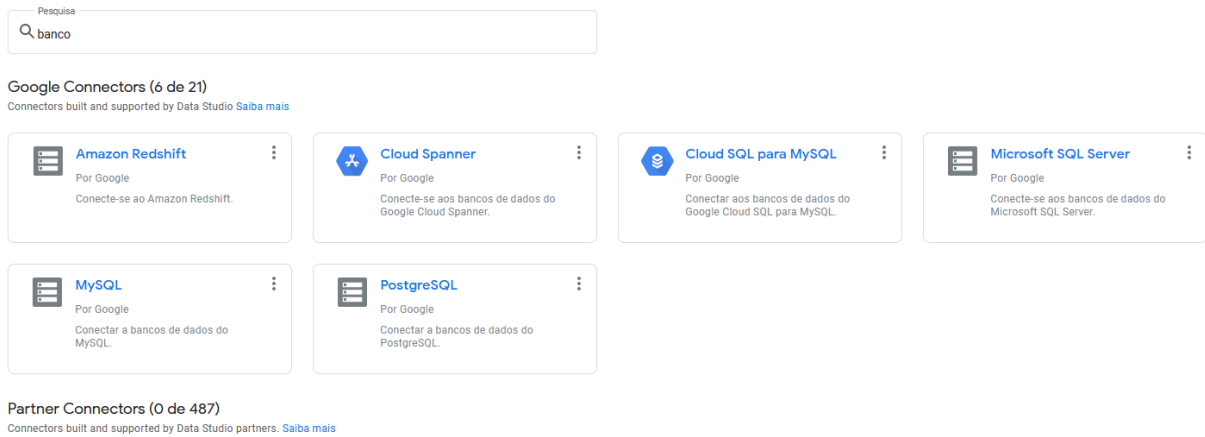
Pode-se perceber que a grande maioria dos conectores tem o propósito de conectar dados a ferramentas de marketing, ou para acesso a dados de mídias sociais. Como exemplo, uma busca pelo termo “banco” (a fim de encontrar conectores para bancos de dados) retornou apenas 6 resultados, sendo todos desenvolvidos pelo Google. Esta busca pode ser vista na Figura 44. Já uma busca pelo termo “facebook” retornou 96 resultados, todos desenvolvidos por parceiros. O termo “planilha” retorna apenas um resultado, que é um conector para as planilhas do Google. Não foi encontrado nenhum conector para planilhas do Excel. Caso a fonte de dados desejada seja uma planilha do Excel, é necessário primeiro exportar o arquivo para o formato CSV, e então utilizar o conector que permite upload de arquivos.

O Microsoft Power BI possui uma grande variedade de conectores, incluindo diversos bancos de dados que não são uma opção para o Google Data Studio, como Sybase e Oracle. Existem também opções para acessar dados em outros serviços da Microsoft, como pastas e listas do Sharepoint, Exchange Online, e serviços do Azure. Ainda é possível buscar dados em planilhas do Excel, arquivos PDF, scripts R e Python, e realizar *scraping* de páginas web. Algumas das opções de conectores para o Power BI podem ser vistas na Figura 45.

²⁰ <https://www.microsoft.com/store/productId/9NTXR16HNW1T>

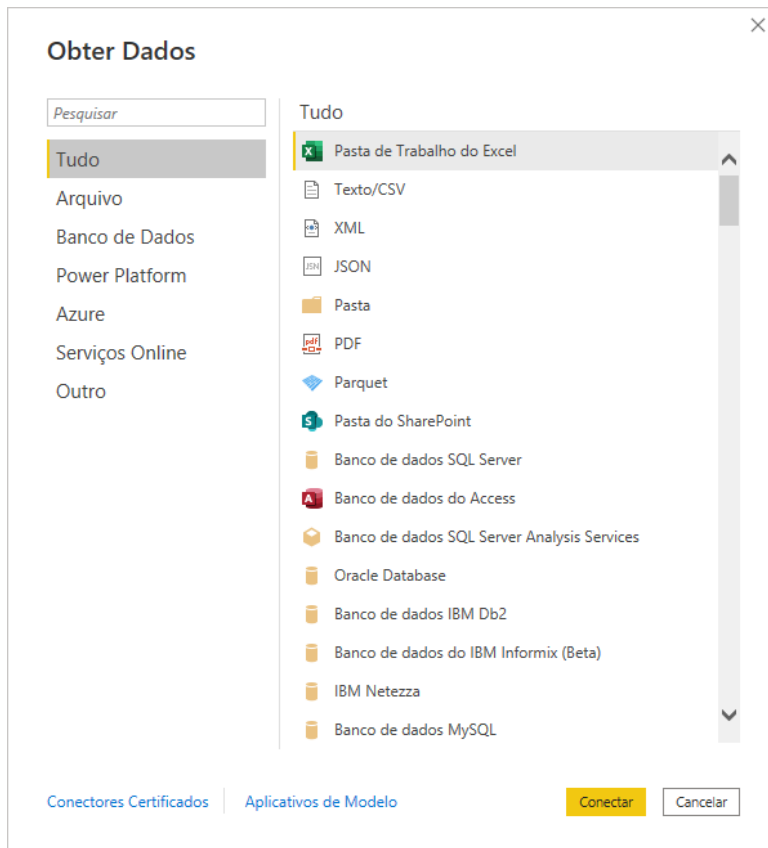
²¹ Informações verificadas em 24/01/2022 no endereço <https://datastudio.google.com/u/0/datasources/create/>.

Figura 44 – Resultado de busca por conectores do Data Studio com o termo “banco”.



Fonte: Elaborado pelo autor.

Figura 45 – Algumas das opções de conectores de dados do Microsoft Power BI.



Fonte: Elaborado pelo autor.

4.4 TRANSFORMAÇÃO DE DADOS

O Microsoft Power BI possui recursos que permitem a transformação de dados na própria ferramenta, como as abas de dados e de modelo. A aba de dados exibe as informações em forma de tabela, permitindo que dados de tabelas diferentes possam ser interpretados como uma única tabela, facilitando a organização dos campos no momento de criação dos gráficos. A Figura 46 mostra este recurso em uso, onde os campos *bi.dim_especialidade.descricao* e *bi.dim_municipio.nome_ac* foram agregados logicamente à tabela *bi.fato_distancias*.

Já a aba de modelo do Power BI pode ser vista na Figura 47. Nesta aba os relacionamentos entre as fontes de dados podem ser criados, não importando a origem destas fontes. Assim, os dados oriundos de um arquivo do Excel podem ser relacionados com os dados de um banco Postgres, por exemplo. Basta indicar as colunas correspondentes em cada fonte e a direção do relacionamento.

No Google Data Studio, o relacionamento entre os dados é mais limitado. Os relacionamentos são realizados através do recurso de combinação de dados, como no exemplo da Figura 48. Embora intuitivamente pareça ser possível criar diversos tipos de relacionamentos através deste recurso (inclusive relacionamentos muitos para muitos), só é possível realizar combinações equivalentes a *left outer join*, onde todas as fontes precisam ter um campo em comum. Inclusive, esta limitação é informada na própria documentação do Data Studio²². Também não é possível realizar a combinação de dados sobre dados já combinados, o que acaba impedindo que esta limitação possa ser contornada de outra forma.

Para fontes de dados que utilizem conexão com banco de dados, é possível utilizar uma consulta personalizada ao invés de selecionar uma tabela (ver Figura 49), sendo, portanto, possível realizar o relacionamento utilizando consultas com junções de dados. Mas no caso de outras fontes, como arquivos CSV, não há solução.

Esta é uma grande desvantagem do Google Data Studio, já que a modelagem dimensional utilizada nos projetos de *data warehouse* considera que exista uma tabela fato relacionada a várias tabelas dimensões, conforme apresentado nas seções 2.3.1.2 e 2.3.1.3, que tratam dos modelos estrela e floco de neve. Sem a possibilidade de realizar um relacionamento um para muitos, é impossível exibir os dados relacionados no Data Studio, a não ser que os dados passem por uma etapa de modificação antes de serem consumidos pela ferramenta.

Observação: Após o término deste trabalho, o Google publicou uma atualização do Data Studio com melhorias na combinação de dados. Esta atualização foi publicada no dia 17 de fevereiro de 2022²³.

²² <https://support.google.com/datastudio/answer/9061420?hl=pt-BR>

²³ Informações disponíveis no endereço <https://support.google.com/datastudio/answer/11542817>. Último acesso em 21/02/2022.

Figura 46 – Aba de dados do Microsoft Power BI.

Nome: bi fato_distancias

Arquivo | Página Inicial | Ajuda | Ferramentas da tabela

Gerenciar relações | Nova medida | Nova medida rápida | Nova coluna | Nova tabela

Calendários | Relações | Cálculos

cod_especialidade	cod_municipio	cod_estabelecimento	distancia	bi.dim_especialidade.descricao	bi.dim_municipio.nome_ac
19	3914	223784	0	CLINICA GERAL	Porto Xavier
19	2859	308288	0	CLINICA GERAL	Taquara
19	4843	80378	0	CLINICA GERAL	David Canabarro
19	4935	292268	0	CLINICA GERAL	Erechim
19	3867	223868	0	CLINICA GERAL	Piratini
19	4183	345203	0	CLINICA GERAL	São Pedro do Sul
19	3939	340972	0	CLINICA GERAL	Sarandi
19	4157	385614	0	CLINICA GERAL	Santa Maria
19	4176	291996	0	CLINICA GERAL	São Leopoldo
19	4184	292236	0	CLINICA GERAL	São Sebastião do Caí
19	2556	94706	0	CLINICA GERAL	Bom Jesus
19	1327	406804	0	CLINICA GERAL	Bagé
19	1327	334209	0	CLINICA GERAL	Bagé
19	3929	226657	0	CLINICA GERAL	São Francisco de Assis
19	3861	473909	0	CLINICA GERAL	Passo Fundo
19	4195	353159	0	CLINICA GERAL	Sobradinho
19	3916	445599	0	CLINICA GERAL	Quaraí
19	5241	151174	0	CLINICA GERAL	Guabiju
19	4858	151799	0	CLINICA GERAL	Estrela
19	5247	117616	0	CLINICA GERAL	Ivoti
19	1431	144772	0	CLINICA GERAL	Pinheiro Machado
19	5326	201334	0	CLINICA GERAL	Caxias do Sul
19	4155	92661	0	CLINICA GERAL	Santa Bárbara do Sul
19	4196	342375	0	CLINICA GERAL	Soledade
19	3852	155805	0	CLINICA GERAL	Novo Hamburgo
19	4174	342579	0	CLINICA GERAL	São José do Ouro
19	1652	224214	0	CLINICA GERAL	Canela
19	3601	226155	0	CLINICA GERAL	Aratiba
19	4929	201211	0	CLINICA GERAL	Cruz Alta
19	1364	140478	0	CLINICA GERAL	Guaíba
19	3897	221352	0	CLINICA GERAL	Não-Me-Toque
19	5240	79988	0	CLINICA GERAL	Frederico Westphalen
19	3604	452602	0	CLINICA GERAL	Arroio Grande

Campos

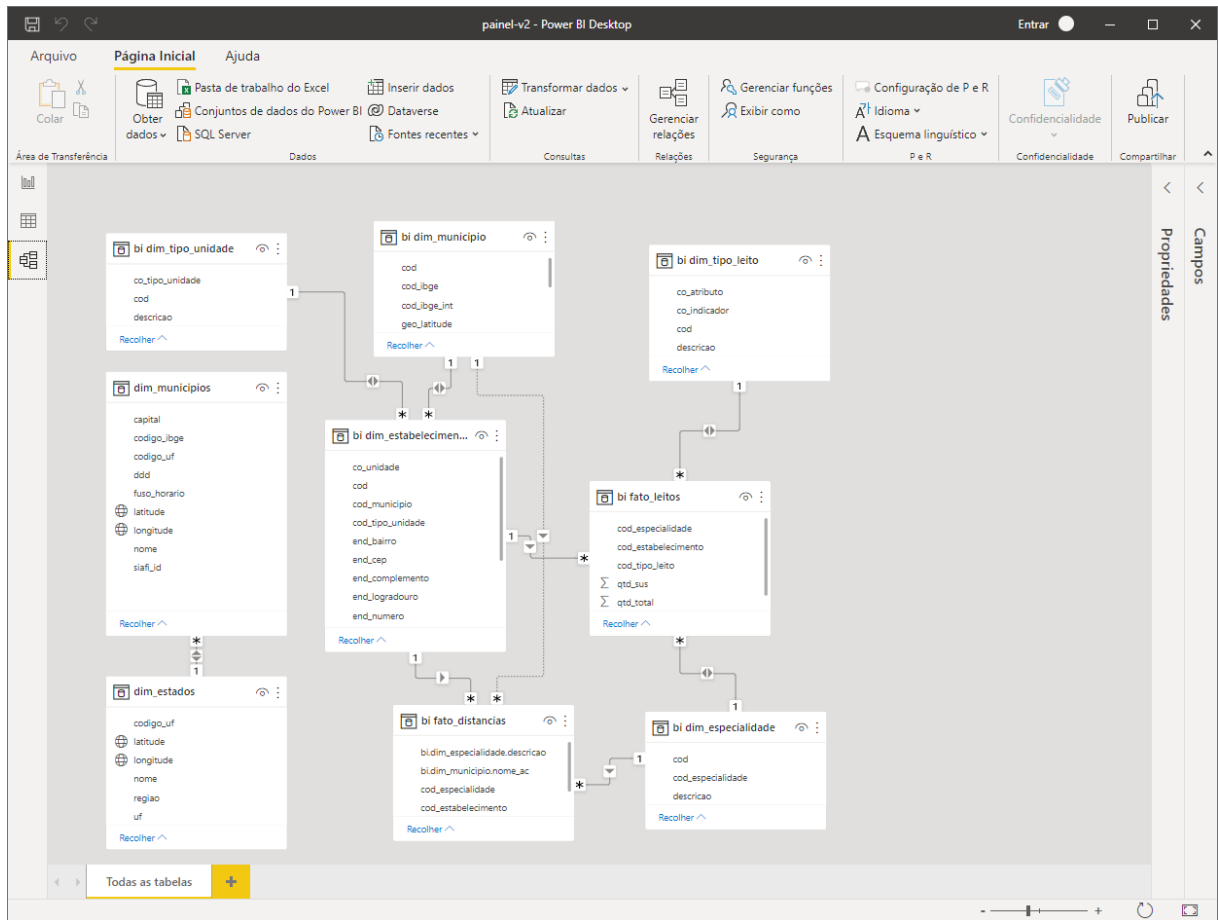
Pesquisar

- bi dim_especialidade
- bi dim_estabelecimento
- bi dim_municipio
- bi dim_tipo_leito
- bi dim_tipo_unidade
- bi fato_distancias
- bi fato_leitos
- dim_estados
- dim_municipios

Tabela: bi fato_distancias (1.780.751 linhas)

Fonte: Elaborado pelo autor.

Figura 47 – Aba de modelo do Microsoft Power BI.



Fonte: Elaborado pelo autor.

Figura 48 – Exemplo de uma combinação de dados no Google Data Studio.

Fonte: Elaborado pelo autor.

Figura 49 – Exemplo de consulta personalizada para uma fonte de dados no Google Data Studio.

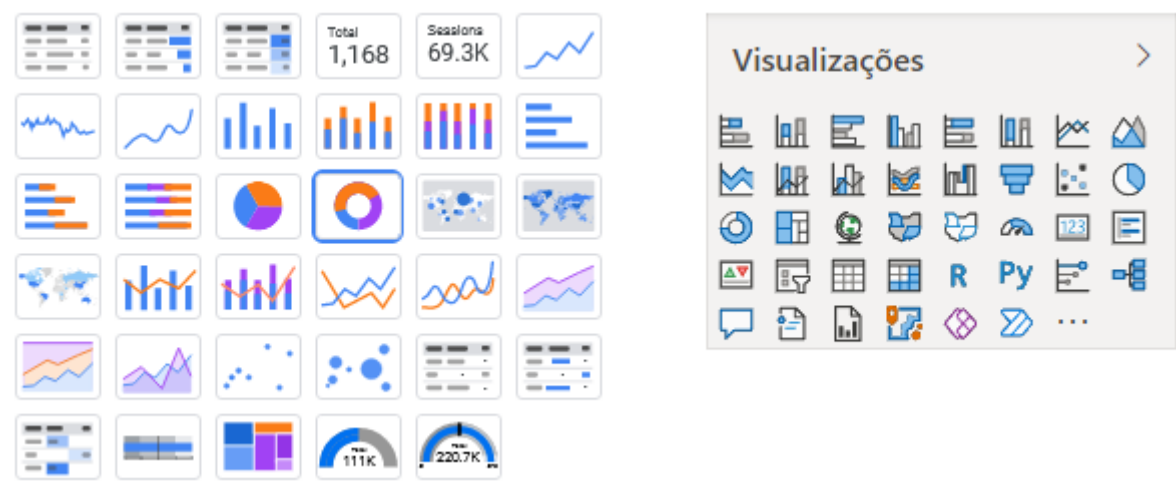
TABELAS	Enter Custom Query
CONSULTA PERSONALIZADA	<pre> 1 SELECT 2 estab.nome_fantasia, 3 estab.razao_social, 4 mun.nome_ac as nome, 5 mun.sigla_estado as uf, 6 lto.qtd_total, 7 lto.qtd_sus 8 FROM 9 bi.fato_leitos lto 10 LEFT JOIN 11 bi.dim_estabelecimento estab 12 ON lto.cod_estabelecimento = estab.cod 13 LEFT JOIN 14 bi.dim_municipio mun 15 ON estab.cod_municipio = mun.cod </pre>

Fonte: Elaborado pelo autor.

4.5 VISUALIZAÇÕES

Ambas as ferramentas possuem visualizações comumente encontradas em gráficos de exibição dados, como tabelas, barras verticais e horizontais, linhas, cartões e mapas. O Microsoft Power BI se destaca neste aspecto, apresentando, além das opções comuns, a possibilidade de incluir scripts R e Python, mapas ArcGIS e recursos do Power Apps e Power Automate. Ainda é possível utilizar visuais desenvolvidos por terceiros²⁴ e desenvolvidos pelo próprio usuário²⁵. A Figura 50 mostra as opções de visualizações do Google Data Studio e do Microsoft Power BI lado a lado.

Figura 50 – Painéis de seleção de visualizações no Data Studio (esquerda) e Power BI (direita).



Fonte: Elaborado pelo autor.

²⁴ <https://appsource.microsoft.com/pt-BR/marketplace/apps?product=power-bi-visuals>

²⁵ <https://docs.microsoft.com/pt-br/power-bi/developer/visuals/develop-power-bi-visuals>

5 CONCLUSÃO

O objetivo deste trabalho foi estudar o uso das ferramentas de Inteligência de Negócios Microsoft Power BI e Google Data Studio para visualização de dados abertos da área da saúde, buscando diferenças e semelhanças bem como pontos fortes e fracos de cada uma. Objetivo este que, como visto no decorrer do trabalho, foi atingido.

Para o seu desenvolvimento, foram apresentados conceitos importantes de Inteligência de Negócios. Também foi documentada a aquisição dos dados, sua transformação e seu uso na criação de um pequeno *data warehouse*, o qual foi utilizado para alimentar os painéis criados nas ferramentas estudadas.

Foram criados cinco painéis em cada uma das ferramentas, replicando os dados, a aparência e as funcionalidades de acordo com as possibilidades de cada uma. O objetivo da criação destes painéis é buscar diferenças e semelhanças entre as ferramentas, bem como as dificuldades encontradas no uso de cada uma. As análises foram realizadas sobre os aspectos de licença de uso, facilidade de uso, fontes de dados, transformação de dados e visualizações disponíveis.

No comparativo foi possível perceber que o Microsoft Power BI é a ferramenta mais completa entre as opções estudadas, mas por outro lado, é um pouco mais complexa para usuários iniciantes. O Power BI ainda possui algumas restrições no seu plano gratuito, conforme visto na seção 4.1. Já o Google Data Studio possui limitações com relação às suas fontes de dados, o que pode causar muitas dificuldades no uso dos modelos dimensionais floco de neve ou estrela, mas se mostra mais fácil de usar em outros casos, além de sua licença de uso ser gratuita sem limitações.

Devido a sua licença de uso e conectores de dados disponíveis, o Google Data Studio demonstra ser a melhor ferramenta entre as opções estudadas para análise e visualização de dados de marketing e de outras plataformas do Google, como o YouTube, Analytics e Search Console. Já para análises de conjuntos de dados maiores, informações empresariais privadas, e usuários que buscam mais recursos, o Microsoft Power BI se apresenta como a melhor alternativa. Importante ressaltar que, no caso de informações sigilosas, é importante a aquisição de uma licença Pró ou Premium, já que na licença gratuita os painéis são publicados na web sem restrições de acesso.

Como continuidade deste trabalho e sua ampliação, propõe-se:

- O uso do Microsoft Power BI e do Google Data Studio com outras fontes de dados abertos;
- A avaliação das ferramentas em suas versões futuras, onde os problemas encontrados atualmente possam ter sido resolvidos;
- Estudar e comparar o desenvolvimento de conectores de dados no Google Data Studio e Microsoft Power BI;
- A aplicação dos conhecimentos adquiridos durante o desenvolvimento deste trabalho nas atividades rotineiras do Hospital Universitário de Santa Maria.

REFERÊNCIAS

- BRASIL. Ministério da Saúde. **Wiki CNES**. Disponível em: <https://wiki.saude.gov.br/cnes/index.php/P%C3%A1gina_principal#Principais_T.C3.B3picos_do_CNES> Acesso em: 9 ago. 2021.
- FIVE ACTS. **Granularidade de dados: o que é, importância e erros para evitar**. Disponível em: <<https://www.fiveacts.com.br/granularidade-de-dados/>>. Acesso em: 6 ago. 2021.
- GOOGLE. **Principais recursos dos relatórios de dados e dos painéis: Google Data Studio**. Disponível em: <https://marketingplatform.google.com/intl/pt-BR_br/about/data-studio/features/>. Acesso em: 18 out. 2021.
- HELLER, M. **10 ferramentas de BI para visualização de dados**. Disponível em: <<https://cio.com.br/tendencias/10-ferramentas-de-bi-para-visualizacao-de-dados/>>. Acesso em: 9 ago. 2021.
- HEUSER, C. A. **Projeto de Banco de Dados**. 6. ed. Porto Alegre: Bookman, 2009. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9788577804528/>>. Acesso em: 02 out. 2021.
- INMON, W. H. **Como construir o data warehouse**. Tradução Ana Maria Netto Guz. Rio de Janeiro: Campus, 1997.
- KIMBALL, R. **The data warehouse toolkit: practical techniques for building dimensional data warehouses**. New York: John Wiley & Sons, 1996.
- KIMBALL, R.; ROSS, M. **The data warehouse toolkit: the complete guide to dimensional modeling**. 2. ed. New York: John Wiley & Sons, 2002.
- MARIANO, D. C. B. **Data mining**. Porto Alegre: Sagah, 2020. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9786556900292/>>. Acesso em: 6 ago. 2021.
- MICROSOFT. **Introdução ao Power BI**. Disponível em: <<https://docs.microsoft.com/pt-br/learn/modules/introduction-power-bi/2-what-power-bi>>. Acesso em: 9 ago. 2021.
- MICROSOFT. **O que é Power BI?** Disponível em: <<https://docs.microsoft.com/pt-br/power-bi/fundamentals/power-bi-overview>>. Acesso em: 9 ago. 2021.
- SHARDA, R.; DELEN, D.; TURBAN, E. **Business Intelligence e análise de dados para gestão do negócio**. Tradução Ronald Saraiva de Menezes. Porto Alegre: Bookman, 2019.
- TEOREY, T. J.; LIGHTSTONE, S.; NADEAU, T.; JAGADISH, H. D. **Projeto e modelagem de banco de dados**. Tradução de Daniel Vieira. Rio de Janeiro: Elsevier, 2014. Disponível em: <<https://integrada.minhabiblioteca.com.br/#/books/9788595156432/>>. Acesso em: 5 ago. 2021.