

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE COMPUTAÇÃO
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

RODRIGO DA COSTA PASSOS
MAURICIO BORGES PEREIRA JUNIOR

PAINEL ANALÍTICO SOBRE DADOS CLÍNICOS ABERTOS DE PACIENTES DA
COVID-19

RIO DE JANEIRO

2022

RODRIGO DA COSTA PASSOS
MAURICIO BORGES PEREIRA JUNIOR

PAINEL ANALÍTICO SOBRE DADOS CLÍNICOS ABERTOS DE PACIENTES DA
COVID-19

Trabalho de conclusão de curso de graduação apresentado ao Instituto de Computação da Universidade Federal do Rio de Janeiro como parte dos requisitos para obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Ph.D. Maria Luiza Machado Campos

Co-orientador: M.Sc. Vânia Jesus de Araújo Soares Borges

RIO DE JANEIRO

2022

P289p

Passos, Rodrigo da Costa

Painel analítico sobre dados clínicos abertos de pacientes da COVID-19 / Rodrigo da Costa Passos, Mauricio Borges Pereira Junior. – Rio de Janeiro, 2022.

106 f.

Orientadora: Maria Luiza Machado Campos.

Coorientadora: Vânia Jesus de Araújo Soares Borges.

Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação) - Universidade Federal do Rio de Janeiro, Instituto de Computação, Bacharel em Ciência da Computação, 2022.

1. COVID. 2. Inteligência de negócio. 3. Ambiente analítico. 4. Decisão médica. I. Pereira Junior, Mauricio Borges. II. Campos, Maria Luiza Machado (Orient.). III. Borges, Vânia Jesus de Araújo Soares (Coorient.). IV. Universidade Federal do Rio de Janeiro, Instituto de Computação. V. Título.

AGRADECIMENTOS

Gostaríamos de agradecer à Vânia Borges pelas revisões e recomendações durante a elaboração deste trabalho e, em especial, por sua disponibilidade. Também agradecemos pela participação de integrantes do projeto VODAN BR em nossas primeiras reuniões, nos ajudando a ter uma visão melhor do projeto como um todo.

Por último, e principalmente, agradecemos à professora Maria Luiza Campos por nos passar conhecimento e orientar na elaboração deste trabalho.

Mauricio Borges Pereira Junior

Gostaria de agradecer aos meus amigos Yuri Lopes de Abreu e Ygor da Hora que, por mais que não tenhamos tanto contato quanto antigamente, me auxiliaram durante o início e meados deste percurso universitário e, sem os quais, muito provavelmente não teria conseguido chegar aonde cheguei. Agradeço também a todos os professores que me saciaram daqueles conhecimentos com os quais pude me impulsionar em minha jornada profissional.

E, em especial, agradeço aos meus pais que me apoiaram em todos os momentos, mesmo os não tão bons, e me ajudaram a encontrar forças para continuar. Também agradeço à Julia Felix, minha namorada, pelo suporte e pela insistência para que eu terminasse este trabalho.

Espero um dia poder retribuir em dobro tudo o que vocês fizeram por mim!

Rodrigo da Costa Passos

Quero deixar meu agradecimento especial aos meus pais que sempre batalharam bastante para que eu pudesse estar realizando este sonho e por terem me dado toda força e suporte necessário para que eu pudesse sobrepujar os desafios dessa caminhada. Além disso, sou grato a Isabella Grossi, minha esposa, a qual se mostrou sempre pronta para me ajudar a ter forças, persistência e motivação para poder seguir nesse caminho tão árduo, mas tão recompensador.

‘Teremos todas as informações do mundo na ponta de nossos dedos não facilita nossa comunicação: dificulta. Quanto mais informações temos de lidar, mais difícil é filtrar o que é mais importante’

Lazlo Bock

RESUMO

Nesses dois últimos anos, vivemos algo novo na história recente da humanidade. As gerações atuais se encontravam confortáveis com os grandes avanços medicinais dos últimos tempos, onde pesquisas e estudos foram responsáveis pela melhoria da qualidade e da expectativa de vida no Brasil e no mundo em geral. No entanto, apesar desses avanços, percebemos que a sociedade não estava preparada para a pandemia do COVID-19. Esse novo surto tornou transparente a nossa fragilidade no que tange à resposta médica frente a novas patologias (em especial, as altamente contagiosas como a gripe). Durante esse período, a sociedade passou por diversas incertezas (e, por isso, conflitos) em seus âmbitos político, médico e social. Tratamentos e protocolos foram alterados com frequência, em tentativas muitas vezes infrutíferas de evitar o avanço da doença e a morte da população. E, dessa forma, são levantadas questões como: “há uma forma de estarmos mais bem preparados no futuro?”, ou “é possível tornar a pesquisa médica mais assertiva e eficiente?”. Este trabalho apresenta um exemplo de ambiente analítico, visando contribuir para a discussão do papel da inteligência de negócio (*Business Intelligence*) dentro das áreas de decisão médica. Seu objetivo é introduzir esse conceito e a sua utilização em contextos de tomada de decisão. Além disso, ele propõe uma organização dos dados médicos de forma mais propícia à consulta analítica, empregando um *Data Mart*, e expõe formas de análise e visualização dos dados médicos para uma visão compilada e objetiva acerca da eficácia de tratamentos e da evolução de quadros clínicos para a infecção por SARS-COV-2 por meio da construção de um *Dashboard*. De forma complementar, mas não menos importante, procura enfatizar a importância da disponibilidade de dados abertos para que, em um próximo evento ou pandemia, informações possam estar mais facilmente disponíveis e interligadas. Isso é feito em conjunto com a iniciativa VODAN, um projeto internacional com um piloto no Brasil, que visa coletar e disponibilizar um repositório de dados médicos sobre atendimentos referentes à COVID-19.

Palavras-chave: COVID; inteligência de negócio; ambiente analítico; decisão médica.

ABSTRACT

During the last two years, we have experienced something new in recent human history. Current generations were living comfortably with the recent times great medical advances, whereas researches and studies were responsible for the increase of life quality and expectations in Brazil and in the whole world. However, despite such advances, we noticed society was not prepared for COVID-19 pandemics. This new outbreak made transparent our debilities about the medical response before new pathologies (in particular, highly contagious diseases such as the flu). In this period, society went through several uncertainties (and conflicts, therefore) on political, medical, and social scopes. Treatments and protocols were changed frequently during often unsuccessful attempts to prevent disease's progress and the population's death. And all that raises some questions like: "is there a way to become more prepared for future diseases?", or "is it possible to make medical research more assertive and efficient?". This work presents an analytical environment example, aiming to contribute to the discussion about the role of business intelligence in the medical decision-making area. The goal is to introduce this concept and its use in decision-making contexts. Moreover, it proposes a medical data rearrangement to make it friendly to analytical querying by using a Data Mart. It also explores ways to analyze and visualize medical data aiming to provide a more compiled and objective view for treatments efficacy and for clinical condition evolution of SARS-COV-2 infection through building a dashboard. In addition, but not less important, this work aims to emphasize the importance of open data availability in order to get information more easily available and connected on a next event or pandemic situation. This is done in conjunction with the VODAN initiative, an international project with national work fronts that aims to collect and make available a massive repository of COVID-19 treatments related medical data.

Keywords: COVID; business intelligence; analytical environment; medical decision.

LISTA DE ILUSTRAÇÕES

Figura 1: Exemplo de modelagem estrela	27
Figura 2: Exemplo de gráfico de barra	29
Figura 3: Exemplo de gráfico de dispersão	30
Figura 4: Exemplo de gráfico de linhas	31
Figura 5: Exemplo de gráfico de área	32
Figura 6: Exemplos de gráficos de pirâmide etária	33
Figura 7: Exemplo de Cartão	33
Figura 8: Exemplo de Mapa de Árvore (TreeMap)	34
Figura 9: Exemplo de mapa de calor	35
Figura 10: Destaque do Ambiente Analítico na Arquitetura do Projeto VODAN BR	39
Figura 11: Extrato da Base de Dados do projeto VODAN BR	40
Figura 12: Extrato do Módulo de Acompanhamento destacando a questão sobre “Resultados Laboratoriais”	41
Figura 13: Extrato do Módulo de Acompanhamento destacando a questão sobre “Cuidados”	42
Figura 14: Tabelas e relacionamentos do modelo estrela	42
Figura 15: Campos da tabela fato "Resultado"	44
Figura 16: Campos da dimensão "Tempo"	45
Figura 17: Campos da dimensão "Paciente"	45
Figura 18: Campos da dimensão "Hospital"	45
Figura 19: Campos da dimensão "Questionário"	46
Figura 20: Campos da dimensão "Diagnóstico"	46
Figura 21: Campos da dimensão "Condição"	46
Figura 22: Campos da dimensão "Medicação"	47
Figura 23: Campos da dimensão "Cuidado"	47
Figura 24: Campos da dimensão "Sintoma"	47
Figura 25: Campos da dimensão "Exame"	47
Figura 26: Campos da dimensão "Avaliação"	48
Figura 27: Campos da dimensão "Atendimento"	48
Figura 28: Arquitetura ajustada aos dados do Repositório da FAPESP	49
Figura 29: Modelo lógico das Tabelas Empregadas no Ambiente Analítico	50

Figura 30: Fluxo de processamento ETL para os dados de pacientes	54
Figura 31: Configuração dos passos “Extrai Pacientes (...)”	54
Figura 32: Configuração do passo “Add constants 4 2”	55
Figura 33: Configuração do passo “Prep. Campos”	56
Figura 34: Configuração do passo “Null if”	56
Figura 35: Configuração do passo “Calcula Idade”	57
Figura 36: Configuração do passo “Mapeia Faixa Etária”	57
Figura 37: Configuração do passo “Mapeia Sexo”	57
Figura 38: Configuração do passo “Mapeia Cidade”	58
Figura 39: Configuração do passo “OUTPUT stg_002_fapesp_pacientes”	59
Figura 40: Fluxo de processamento ETL para os dados de exames	59
Figura 41: Configuração dos passos “Extrai Exames (...)”	60
Figura 42: Configuração do passo “Add constants 4”	61
Figura 43: Configuração do passo “Select values 2”	61
Figura 44: Configuração do passo “OUTPUT stg_002_fapesp_exames_total”	62
Figura 45: Configuração dos passos “Filtra COVID (...)”	63
Figura 46: Configuração dos passos “Add constants (...)” referentes aos filtros	63
Figura 47: Configuração do passo “OUTPUT stg_002_fapesp_exames_covid”	66
Figura 48: Fluxo de processamento ETL para os dados de Desfecho	67
Figura 49: Configuração dos passos “Extrai Desfechos (...)”	67
Figura 50: Configuração do passo “Add constants 4 2 4”	68
Figura 51: Configuração do passo “Prep. Campos 4”	68
Figura 52: Configuração do passo “OUTPUT stg_002_fapesp_desfechos”	69
Figura 53: Configuração do passo “Preenche EXAMES”	69
Figura 54: Destaque do SQL Executado	70
Figura 55: Colunas da Tabela de Desfechos	71
Figura 56: Colunas da Tabela de Exames	72
Figura 57: Colunas da Tabela de Pacientes	73
Figura 58: Colunas da Tabela de Exames COVID	73
Figura 59: Capa do dashboard	74
Figura 60: Esboço dos Painéis	75
Figura 61: Página de Dados Demográficos	76
Figura 62: Página de Internações Hospitalares	80
Figura 63: Página da Análise Comparativa entre Instituições	83

Figura 64: Página Exames COVID-19	85
Figura 65: Página Exames OMS	87
Figura 66: Exemplo de visualização do gráfico Resultados x Horas Corridas	88

LISTA DE TABELAS

Tabela 1: Relação das Tabelas e total de registros tratados por Instituição

70

LISTA DE QUADROS

Quadro 1: Gráfico de Barras	77
Quadro 2: Pirâmide Etária	78
Quadro 3: Gráficos de Colunas	79
Quadro 4: Mapa de Árvore	79
Quadro 5: Cartões	79
Quadro 6: Matriz	81
Quadro 7: Gráfico de Barras 100% Empilhadas	81
Quadro 8: Gráfico de Linhas	82
Quadro 9: Gráfico de Colunas Empilhadas	82
Quadro 10: Gráfico de Barras Empilhadas	83
Quadro 11: Gráfico de Colunas Empilhadas e Linha	84
Quadro 12: Gráfico de Barras Clusterizado	84
Quadro 13: Gráfico de Área Empilhado	86
Quadro 14: Gráfico de Barras Empilhado	86
Quadro 15: Gráfico de Dispersão	88

LISTA DE SIGLAS

SARS-CoV-2 - Severe Acute Respiratory Syndrome - Corona Virus - 2

COVID-19 - Corona Virus Disease - 2019

VODAN - Virus Outbreak Data Network

FAIR - Findable, Accessible, Interoperable and Reusable

CRF - Case Report Form

OMS - Organização Mundial da Saúde

HSL - Hospital Sírio-Libanês

FAPESP - Fundação de Amparo à Pesquisa do Estado de São Paulo

BI - Business Intelligence

SGIs - Sistemas de Gerenciamento de Informação

SGBDs - Sistemas de Gerenciamento de Banco de Dados

RPA - Robotic Process Automation

ETL - Extract, Transform and Load

ROI - Return On Investment

PDI - Pentaho Data Integration

SSIS - SQL Server Integration Services

SQL - Structured Query Language

DW - Data Warehouse

EUA - Estados Unidos da América

DDL - Data Definition Language

DAX - Data Analysis Expressions

HSL - Hospital Sírio-Libanês

USP - Universidade de São Paulo

UTI - Unidade de Terapia Intensiva

FRC-OMS - Fatores de Risco Comportamentais - Organização Mundial da Saúde

CSV - Comma-separated values

CEP - Código de Endereçamento Postal

SUMÁRIO

1 INTRODUÇÃO	15
1.1 MOTIVAÇÃO	16
1.2 OBJETIVO	16
1.3 CONTRIBUIÇÃO	17
1.4 ESTRUTURA	18
2 A INTELIGÊNCIA DE NEGÓCIO E SUA FUNDAMENTAÇÃO TEÓRICA	19
2.1 ASPECTOS HISTÓRICOS	19
2.2 CONCEITUAÇÃO E PROCESSOS	20
2.3 TRABALHOS RELACIONADOS	22
2.4 FASES DO DESENVOLVIMENTO	23
2.4.1 Planejamento	24
2.4.2 Execução	25
2.5 <i>DATA WAREHOUSES, DATA MARTS</i> E A MODELAGEM DIMENSIONAL	26
2.6 VISUALIZAÇÃO DA INFORMAÇÃO	28
2.6.1 Benefícios	28
2.6.2 Tipos de visualização	29
3 DESENVOLVIMENTO DA SOLUÇÃO ANALÍTICA	36
3.1 FERRAMENTAS	36
3.1.1 SQL Power Architec	36
3.1.2 Pentaho Data Integration (PDI)	37
3.1.3 MySQL	37
3.1.4 Power BI	38
3.2 DESENVOLVIMENTO DO <i>DATA MART</i> DO PROJETO VODAN BR	39
3.2.1. A tabela fato	44
3.2.2. As tabelas dimensão	44

3.3 DESENVOLVIMENTO DO AMBIENTE ANALÍTICO PARA VISUALIZAÇÃO DE DADOS SOBRE COVID-19	49
3.3.1 O Planejamento	49
3.3.2 A Fase de Execução	51
3.3.2.1 O processo de ETL	53
3.3.2.2 Desenvolvimento dos <i>dashboards</i>	70
4 CONCLUSÃO	89
4.1 DIFICULDADES ENCONTRADAS	90
4.1.1 Definição de escopo	90
4.1.2 Tecnologia	90
4.2 TRABALHOS FUTUROS	91
REFERÊNCIAS	92
APÊNDICES	95

1 INTRODUÇÃO

A situação pandêmica iniciada em 2020 ainda mobiliza pessoas e instituições em iniciativas relacionadas ao avanço e à melhoria de medidas de saúde para a sua prevenção e tratamento, buscando novas formas de combate ao SARS-CoV-2, o vírus causador da COVID-19. Essas iniciativas globais se tornam evidentes com a rápida resposta oferecida, seja na criação de vacinas (KESSELHEIM, DARROW, *et al.*, 2021) e medicamentos específicos para o combate à doença em tempo recorde, como também na realização de pesquisas (LI, SUN, *et al.*, 2021) e campanhas voltadas ao entendimento e à conscientização da sociedade.

Essa mobilização também se beneficia do fato de que vivemos em uma época na qual dados têm se tornado os principais ativos de empresas e instituições, tornando visíveis aspectos outrora desconhecidos ou ignorados por suas gerências. Toda informação obtida através dos dados é relevante e torna-se vantagem competitiva se corretamente utilizada. Além disso, permite melhorias em processos internos que eventualmente beneficiam todas as partes envolvidas. Os dados relacionados à saúde, nesse sentido, se tornaram ferramenta fundamental para o bom desempenho de instituições da área de saúde. Aplicações complexas como predição de doenças (CHEN, HAO, *et al.*, 2017) ou detecção de janelas em cirurgias críticas (BAUMGARTEN, HAEGELEN, *et al.*, 2018) começam a se tornar viáveis e passíveis de discussão.

É dentro dessas iniciativas que se situa a iniciativa *Virus Outbreak Data Network* (VODAN). Essa iniciativa foi criada com o intuito de gerar uma “comunidade de comunidades” para modelar e desenvolver rapidamente uma rede de dados capaz de suportar decisões baseadas em evidências no que tange ao combate contra o novo coronavírus e possíveis ameaças futuras (MONS, 2020). Essa rede de dados segue os princípios FAIR (WILKINSON, DUMONTIER, [...], MONS, 2016), um conjunto de critérios levantados pela comunidade GO FAIR (GO FAIR, [s.d.]) que orientam a modelagem e disponibilização de dados. Tais princípios visam a geração de dados que sejam de “fácil localização” (*findable*), acessíveis (*accessible*), interoperáveis (*interoperable*) e reutilizáveis (*reusable*).

No Brasil, o projeto VODAN BR ganhou notoriedade pelo potencial de integração entre os dados e metadados produzidos em hospitais e clínicas, e disponibilizados segundo o formulário de registro de caso (em inglês, *Case Report Form - CRF*) proposto pela Organização Mundial da Saúde (OMS) (WORLD HEALTH ORGANIZATION, 2021). Nesse contexto, o projeto VODAN BR (VEIGA *et al.*, 2021) visa fornecer uma resposta, ainda que em estágio

inicial, aos investimentos realizados pelas instituições, desenvolvendo uma visualização dos dados que permita aos diretores dessas instituições uma melhor visibilidade das análises possíveis de serem efetuadas a partir da massa de dados obtida.

Desta forma, é parte integrante deste trabalho um protótipo de tal visualização de dados, aplicando conceitos conhecidos da teoria de construção de ambientes analíticos para alcançar o resultado esperado. Para o desenvolvimento dessa visualização de dados, foram estudados e utilizados *softwares* e soluções técnicas para implementação de inteligência de negócio, destacando-se dentre elas técnicas arquiteturais como (i) a modelagem dimensional; (ii) a utilização da suíte de produtos Pentaho para a limpeza e o tratamento dos dados, e (iii) o emprego do Power BI como ferramenta de construção e automação de relatórios visuais.

1.1 MOTIVAÇÃO

A principal motivação deste trabalho reside na contribuição para com a área da saúde, buscando facilitar o trabalho realizado pelos médicos e apresentar meios que os auxiliem a salvar vidas. Some-se a isso a crise imposta pelo novo coronavírus que subtraiu a vida de centenas de milhares de pessoas (SANCHEZ, MOURA, *et al.*, 2021), modificando substancialmente a rotina de toda a sociedade. Essa crise passou a exigir uma análise mais efetiva das informações, visando prover insumos para a tomada de decisões mais assertivas pelas partes responsáveis. Esse é o cenário que temos atualmente, e são esses problemas que este trabalho visa auxiliar a mitigar.

1.2 OBJETIVO

Este trabalho, como primeiro passo, busca introduzir a teoria sobre a criação de ambientes analíticos com base na bibliografia estudada e tornar acessível a utilização da inteligência de negócio na área de decisões médicas. Uma vez explorada a teoria da construção de ambientes analíticos, o trabalho propõe, como primeiro entregável, uma modelagem em esquema estrela de um *Data Mart*, segundo a arquitetura proposta por Ralph Kimball. Tal modelagem visa permitir uma análise com alta granularidade no nível de analito¹ de exame

¹ Analito é todo componente discriminado em uma amostra obtida pela realização de um exame. Por exemplo, um exame de sangue possui como analitos a oxigenação, o nível de glicose, a concentração de hemácias, entre outros.

observado (possibilitando visualizar os resultados de cada exame a nível de analito) para os dados provenientes do preenchimento dos formulários elaborados pela OMS.

Como segundo entregável, o trabalho apresenta a construção de um relatório para visualização das informações médicas, empregando para isso os dados dos atendimentos realizados no Hospital Sirio Libânes (HSL) e dos exames do Grupo Fleury, ambos disponibilizados no repositório da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) COVID-19 Data Sharing/BR (FAPESP, 2020).

1.3 CONTRIBUIÇÃO

Este trabalho visa contribuir às áreas da computação e da medicina, às instituições parceiras do projeto VODAN BR, à Universidade Federal do Rio de Janeiro e, em última instância, à sociedade.

À área da computação fornece um estudo sobre o uso da inteligência de negócio, disciplina interna desta área, com o devido aprofundamento de suas fundamentações teóricas e a aplicação prática de seus conceitos. Aqui, busca-se também apresentar um processo de desenvolvimento de inteligência de negócio frequentemente utilizado em meios empresariais, fornecendo um estudo de caso relevante para aplicações futuras.

A área da medicina, por sua vez, entendida aqui como a área que estuda o funcionamento do organismo humano, suas patologias e o uso de substâncias e procedimentos para diagnosticá-las e corrigi-las, se beneficia há muito tempo do uso de tecnologia para cumprir com seus objetivos. A interseção entre este trabalho e a área da medicina torna-se ainda mais clara ao compreendermos que dados coletados são capazes de gerar informações estatísticas importantes para basear uma decisão médica. E esta, entre outras, é uma das contribuições que visamos com o desenvolvimento do estudo aqui produzido.

Com relação às instituições parceiras, este trabalho contribui com a produção de relatórios decorrentes da aplicação da inteligência de negócio no projeto VODAN BR. Deste modo, busca um retorno palpável dos investimentos de tempo e esforço em curto e médio prazo a tais instituições.

À Universidade Federal do Rio de Janeiro, este trabalho contribui fortalecendo o seu papel no desenvolvimento do projeto VODAN BR e somando mais um exemplar de estudo de caso às suas prateleiras.

E, finalmente, mais do que apenas a núcleos e instituições, este trabalho visa iniciar um projeto de inteligência de negócio capaz de fornecer à sociedade algumas respostas analíticas provenientes da massa de dados processada, tornando-se fonte confiável de informações sobre o novo coronavírus.

1.4 ESTRUTURA

Este primeiro capítulo contém uma breve introdução do trabalho, apresentando de modo geral os pormenores que levaram à sua idealização, assim como o seu objetivo, sua motivação e as contribuições a que se pretende.

No capítulo 2, é apresentada uma revisão da literatura e dos conceitos relacionados ao uso e implementação da inteligência de negócio em seus âmbitos histórico e atual. A partir disso, são abordados de forma mais técnica alguns tópicos de construção de *Data Warehouses* (e *Data Marts*, sendo estes os grãos que os compõem), as fases do processo de implementação dessa tecnologia e os formatos visuais para composição de painéis gerenciais para visualização da informação. Esse capítulo também contextualiza a importância deste trabalho entre outros que com ele se relacionam.

Com a base teórica estabelecida, o capítulo 3 se ocupa em discorrer sobre a implementação do projeto, apresentando o processo, soluções encontradas durante o desenvolvimento dos entregáveis, as ferramentas que foram utilizadas para tal fim e o resultado final obtido.

O capítulo 4 evidencia pontos positivos e negativos do desenvolvimento do projeto, contrastando o seu resultado final com o objetivo inicialmente proposto, e termina por apresentar propostas para contribuições futuras com base no projeto aqui desenvolvido.

Por último, nos apêndices, pode ser encontrado o formulário disponibilizado pela OMS a partir do qual foi desenvolvida a modelagem no esquema estrela.

2 A INTELIGÊNCIA DE NEGÓCIO E SUA FUNDAMENTAÇÃO TEÓRICA

A tecnologia, com sua capacidade de viabilizar o armazenamento massivo de dados e aumentar a praticidade e a rapidez de sua obtenção, de certa forma democratizou a possibilidade de análise e tratamento dos dados. Em 2010, por exemplo, o Google já beirava as 100 milhões de buscas por dia e o Twitter alcançou a marca de 15 bilhões de posts publicados em sua rede social (BARBIERI, 2011, p. 4). Isso é muito mais do que qualquer biblioteca do mundo seria capaz de armazenar e, ainda assim, os números vêm crescendo exponencialmente desde então. Onde outrora se utilizavam espiões para obter informações relevantes para decisões estratégicas militares, hoje, com a tecnologia, é possível extrair facilmente informações de redes sociais de concorrentes.

Mesmo antes de alcançar tal quantidade de dados, as empresas já perscrutavam formas de trabalhar as informações obtidas para aumentar sua vantagem competitiva, originando a expressão “Inteligência de Negócio”, em inglês *Business Intelligence* (BI). Cunhada pela primeira vez no contexto de sistematização do tratamento de dados em 1958 por Hans Peter Luhn (hoje tido como o pai da Inteligência de Negócio) (PAVKOV, POSCIC e JAKSIC, 2016), essa expressão busca proporcionar um acesso interativo aos dados, possibilitando a sua manipulação e fornecendo meios aos gerentes, diretores e analistas para conduzir análises apropriadas (TURBAN, SHARDA e DELEN, 2013). Tal mecânica possibilita que decisões mais assertivas para o crescimento do negócio possam ser efetivadas.

2.1 ASPECTOS HISTÓRICOS

O conceito de “inteligência” é notoriamente antigo e remonta desde os primeiros conflitos militares (PAVKOV, POSCIC e JAKSIC, 2016), onde a informação era extremamente relevante para a vitória por viabilizar decisões e estratégias que garantissem a superioridade sobre a nação rival. A título de exemplo de obtenção de dados para tomada de decisões, temos relatos famosos sobre censos realizados durante toda a história. É suficiente lembrar-se da famosa história de um menino que nasceu em Belém da Judéia, devido à viagem que seus pais realizaram para responder ao censo ordenado pelo imperador romano César Augusto (BIBLIA, 2018).

O surgimento da computação, por volta da década de 40 (CAMPBELL-KELLY, ASPRAY, *et al.*, 2014), tornou possível a automatização desta inteligência que outrora se dava a partir de métodos mecânicos e lentos. Em meados da década de 60, ocorreu o movimento para o desenvolvimento de sistemas eletrônicos que utilizassem dados para suporte à decisão. Nessa época vimos o lançamento de *mainframes* poderosos por um custo viável (POWER, 2003). Esses sistemas, conhecidos como “Sistemas de Gerenciamento de Informação” (SGIs), inauguraram a corrida que culmina hoje na chamada “era da informação”.

Em meados da década de 70, os SGIs evoluíram para os tradicionais Sistemas de Gerenciamento de Banco de Dados em arquitetura relacional (SGBDs). Esses sistemas se mostraram capazes de estruturar os dados de forma a facilitar seu processamento, consumo e organização, sendo utilizados até os dias de hoje.

Novas tecnologias de armazenamento surgiram após os SGBDs com o objetivo de implementar funcionalidades específicas. Com essas tecnologias, técnicas mais apuradas foram elaboradas para possibilitar a extração, o tratamento e a visualização de informações a partir dos dados. Os desafios, porém, permaneceram e se amplificaram com o crescimento da quantidade de dados: como processá-los de modo eficiente e agregar valor a partir da inteligência de negócio?

Durante um tempo as análises e a produção de relatórios foram realizadas nos próprios sistemas SGBDs. Essas análises eram comuns, considerando o uso dos bancos de dados relacionais para o armazenamento dos dados provenientes das aplicações. Porém, com o aumento da quantidade de dados e a possibilidade de sua obtenção através de fontes distintas, tais análises começaram a se tornar custosas e impactar o ambiente e a execução das aplicações.

Por esse motivo, por volta dos anos 90, viu-se a necessidade de separar os ambientes de armazenamento de dados das aplicações (ambiente transacional) e o ambiente de armazenamento de dados a serem consumidos pelos relatórios e análises (ambiente analítico), além da criação de uma metodologia que viabilizasse uma centralização dos dados e o seu processamento mais eficiente (PAVKOV, POSCIC e JAKSIC, 2016).

2.2 CONCEITUAÇÃO E PROCESSOS

De acordo também com Turban, Sharda e Delen (2013, p. 8): “*Business Intelligence*” (BI) é um termo ‘guarda-chuva’ que inclui arquiteturas, ferramentas e banco de dados,

aplicações e metodologias. É uma expressão livre e, portanto, significa coisas diferentes para pessoas diferentes”.

Dessa forma percebemos que a inteligência de negócio é entendida atualmente como um conjunto de técnicas de identificação, extração e análise de dados (ZENG, XU, *et al.*, 2006) baseadas em tecnologia da informação. Ela tem o objetivo explícito de gerar um método eficiente para acompanhamento das métricas e indicadores do negócio e, em última análise, resultados que tragam uma visão nova e criativa sobre um determinado contexto empresarial. Isso é possível pela automação do fluxo de extração, tratamento e carga dos dados, associados ao trabalho criativo de analistas responsáveis pela construção de novas análises e pelo desenvolvimento de formatos visuais práticos, intuitivos e completos para suporte à tomada de decisões corporativas.

Sendo assim, a inteligência de negócio tem possibilitado a descoberta em tempo hábil de problemas corporativos e/ou oportunidades de mercado a partir de um embasamento determinístico e transparente, viabilizando uma vantagem competitiva considerável para com os concorrentes que não gozam da mesma tecnologia. Algumas aplicações bastante interessantes da utilização da inteligência de negócio nesse contexto são, por exemplo, a identificação prematura de tendências de mercado, o acompanhamento de indicadores da empresa, a construção de benchmarks de desempenho e concorrência, a medição do contentamento de consumidores com a marca, entre outras (RANJAN, 2009).

Com a evolução desse conceito, passou-se a discutir e discriminar processos e atividades mais especializadas com a finalidade de alcançar uma melhoria do desempenho geral das técnicas utilizadas. A título de exemplo, consideramos aqui uma parcela de tais processos:

- Extração de dados: tem por finalidade obter os dados relevantes para decisões gerenciais e/ou corporativas;
- *Robotic Process Automation* (RPA): metodologia de obtenção de informações para sistemas cuja extração necessita da simulação de interações humanas;
- Tratamento de dados: responsável por limpar e organizar os dados de forma a prepará-los para outras fases do processo;
- Mineração de dados: possui o objetivo de explorar e correlacionar os dados armazenados em banco de dados a fim de verificar padrões ainda não conhecidos;
- *Benchmarking* e métricas de desempenho: permite a comparação entre dados atuais e históricos no intuito de verificar o desempenho do negócio em relação às metas anteriormente definidas;

- Análise descritiva: expõe, de forma sumarizada e completa, as informações referentes ao conjunto de dados;
- Análise estatística: constrói um conhecimento do dado dentro do todo, ou seja, viabiliza a identificação de padrões e facilita a visualização dos “pontos fora da curva”;
- Visualização de dados: transforma as informações em formatos visuais, como gráficos, sinalizações ou sumarizações, que facilitam o consumo e o entendimento dos dados coletados;
- Geração de relatórios: viabiliza o compartilhamento de análises de dados a partir de visualizações previamente concebidas com a finalidade de contribuir com uma base sólida e confiável para as conclusões e decisões das partes interessadas do negócio;
- Análise visual: refina as informações provenientes de gráficos e formatos visuais a fim de transformá-los em conhecimento e torná-los úteis à discussão e à tomada de decisão.

2.3 TRABALHOS RELACIONADOS

Este trabalho contextualiza-se dentro do conjunto de projetos e pesquisas que têm por objetivo contribuir para um melhor entendimento e resposta à pandemia do novo coronavírus. Mais especificamente, se relaciona àqueles que se utilizam de processo de extração, transformação e carga, em inglês *Extract, Transformation and Load* (ETL) e Visualização da Informação como ferramentas de análise para cumprir esse objetivo.

Como iniciativas, é importante destacar que a OMS disponibiliza, desde 2020, diversas visualizações elaboradas com o objetivo de dar transparência à sociedade com relação ao quadro da pandemia (WORLD HEALTH ORGANIZATION, 2020). Essas visualizações sumarizam informações básicas sobre a situação da COVID-19, utilizando formatos visuais diversificados como o mapa de calor para mostrar qual é o contexto de cada país diante das métricas de casos, mortes e vacinas. O trabalho aqui proposto se diferencia deste com relação à esfera abordada, pois trata dados provenientes dos próprios hospitais responsáveis pelos atendimentos dos pacientes com COVID-19, e à sua granularidade, por tratar dados clínicos que possibilitam não apenas compreender o volume para cada uma das métricas abordadas, mas também entender quais os perfis de pacientes que estão contidos em tais métricas. Este trabalho também busca detalhar a origem, o hospital e o método de detecção de COVID-19 utilizados, entre outras informações contidas nos dados de tratamento dos pacientes.

Dentre as iniciativas brasileiras, destaca-se a visualização desenvolvida pelo governo federal por meio do Ministério da Saúde (MINISTÉRIO DA SAÚDE, 2020). Essa visualização foi construída com um objetivo similar à da OMS, mas em uma esfera nacional, empregando dados disponibilizados pelo sistema de saúde do país e utilizando uma granularidade a nível de município.

Além das fontes de dados oficiais dos países, durante a pandemia, diversos conjuntos de dados foram divulgados em plataformas públicas de dados como o *Kaggle*. Diante disso, pesquisadores e instituições fizeram suas próprias análises, fornecendo uma compreensão valiosa sobre o comportamento do novo coronavírus. Nosso trabalho é um desses casos. Nele realizamos um tratamento sobre dados obtidos em um desses repositórios e disponibilizamos um conjunto de visualizações que possibilitem uma melhor compreensão deles.

Khanan, Nowrin e Mondal (2020) se aproximam bastante deste trabalho no que tange a nível de granularidade, porém empregam fontes diferentes. Além disso, seu trabalho não visa obter como resultado uma visualização dinâmica para ser utilizada futuramente por outras instituições.

Dessa forma, é possível constatar que os trabalhos relacionados buscam explorar um nível mais macro dos dados enquanto este trabalha os dados em um nível mais granular. Isso permite apresentar para todos os profissionais da área (administradores, enfermeiros, médicos, entre outros) um maior número de possibilidades de visualizações capazes de atender necessidades mais específicas através de análises detalhadas sobre a situação de um paciente, por exemplo.

2.4 FASES DO DESENVOLVIMENTO

O processo de construção de um fluxo de Inteligência de Negócio é organizado em quatro fases: descobrir, definir, desenvolver e validar (BONEL, 2021). Neste trabalho, essas fases são agrupadas em duas: planejamento e execução.

A fase de planejamento envolve: (i) compreender a ordem de grandeza da quantidade de dados que precisarão ser processados e a sua frequência de atualização; (ii) conhecer as métricas envolvidas na criação das análises e indicadores; (iii) modelar a arquitetura dos dados com base em tais métricas; e (iv) entender os desejos e necessidades das partes interessadas (*stakeholders*).

A fase de execução, por sua vez, visa: (i) implementar a infraestrutura necessária; (ii) tratar as informações; (iii) efetuar a carga no ambiente analítico; e (iv) construir relatórios para visualização das métricas construídas.

Cabe destacar que as fases não acontecem em sequência, mas compõem um ciclo na qual a execução exige, frequentemente, a revisão do planejamento. Essas revisões podem ocorrer por requisição dos *stakeholders* ou pelo surgimento, através de ciência de dados, de novas análises e métricas a serem integradas ao modelo.

2.4.1 Planejamento

Dentro do planejamento, a primeira etapa (compreender a ordem de grandeza da quantidade de dados) é relevante para a escolha da infraestrutura correta e para a criação de um plano de ação mais assertivo, dado que uma alta quantidade de dados exige técnicas mais apuradas e otimizadas de tratamento, além de um provisionamento escalável.

Para quantidades menores de dados, característica da maioria das empresas, é suficiente o provisionamento de um servidor ou mesmo realizar a execução dos fluxos de inteligência de negócio a partir de computadores pessoais. Para quantidades maiores, normalmente são utilizadas as técnicas de *Big Data*.

Hoje, *Big Data* é um termo muito conhecido na área de análise de dados, pois engloba a extração, tratamento, limpeza e carga de grandes quantidades de dados, geradas rapidamente, ou difíceis de tratar (MADDEN, 2012). Geralmente, esses dados não conseguem ser processados por uma máquina, ou mesmo um servidor, em tempo útil. Para o processamento desses dados, é necessário utilizar arquiteturas distribuídas e algoritmos de baixo custo de processamento e memória.

Esse tipo de arquitetura exige uma modelagem diferenciada, de forma a minimizar os custos de armazenamento e prover um alto desempenho de acesso aos dados. Atualmente, a modelagem mais utilizada e conhecida para ambientes analíticos é o esquema estrela (*Star Schema*) (CHAUDHURI e DAYAL, 1997), um modelo composto por tabelas dimensionais e uma tabela de fatos (explicado na seção 2.5).

As métricas, por sua vez, são definições essenciais do processo para construção das visões corporativas. Nelas, entendemos quais são os conteúdos úteis para as partes interessadas do projeto e quais são as regras de negócio da empresa. Elas são, normalmente, o “coração” do

processo como um todo. Mas, por meio da utilização das técnicas indicadas e das análises provenientes da mesma, é possível chegar a outras métricas úteis ainda não especificadas ou conhecidas.

Por último, e o mais importante, é necessário compreender as expectativas das partes interessadas do projeto, ou seja, o que esperam de retorno em questão de praticidade, eficiência, assertividade e ROI (retorno sobre investimentos), e de que forma pretendem visualizar os dados.

2.4.2 Execução

Para a execução do projeto, é necessário prover a infraestrutura necessária. Essa infraestrutura envolve a instalação local das ferramentas para construção do fluxo de dados (como Spark, PowerBI, Tableau, etc.) ou a utilização de arquiteturas provisionadas em nuvem. Esta última normalmente garante maior facilidade para o monitoramento e a escalonagem dos recursos. Além disso, todos os provedores mais famosos de infraestrutura em nuvem possuem recursos especializados em tratamento de dados e *Big Data*, o que costuma diminuir o investimento inicial necessário para as empresas que desejam se aventurar na área de *Business Intelligence*.

No tratamento de dados, as ferramentas de ETL costumam ser o principal método para implementar as regras de negócio. Elas permitem a limpeza e o tratamento dos dados, organizando-os de acordo com o modelo criado e as regras estabelecidas. Tecnologias como: *Pentaho Data Integration*, SSIS, Python ou mesmo SQL (para abordagens ETL), são comumente utilizadas como alternativas de baixo custo.

Com os dados tratados e armazenados de acordo com o modelo proposto em local adequado como, por exemplo, um *Data Warehouse*, passamos à etapa de exibir os dados e realizar consultas sobre eles. Com este intuito, as ferramentas de visualização de dados como PowerBI ou Tableau se encontram disponíveis no mercado, viabilizando a construção de relatórios e dashboards altamente personalizados.

2.5 DATA WAREHOUSES, DATA MARTS E A MODELAGEM DIMENSIONAL

Para operar dados em ambientes analíticos, Bill Inmon e, posteriormente, Ralph Kimball desenvolveram o conceito de *Data Warehousing*, cada um com técnicas específicas, visando contornar as dificuldades e desafios da manipulação de grandes volumes de dados (BARBIERI, 2011, p. 112). Ambas as abordagens utilizam a ideia de um repositório central com o objetivo de suportar o processamento massivo de consultas para formação de relatórios de negócios. Tais ideias contrastam, entretanto, pela sua metodologia de aplicação: enquanto Inmon se refere à construção desse repositório de forma monolítica para, após isso, serem gerados *Data Marts* (conjuntos de dados modelado segundo a perspectiva de áreas de negócio específicas), Kimball já propõe iniciar a construção do repositório pela implementação de cada *data mart* separadamente. O *Data Warehouse* seria então, segundo Kimball, Ross, *et al.* (2008), “nada mais do que a união de todos os *Data Marts*”.

O *Data Warehouse*, enquanto repositório de dados, é, segundo Turban, Sharda e Delen (2013, p. 51), “um sistema distinto que possui por finalidade prover armazenamento aos dados que serão utilizados para a execução de análises”. Devido à capacidade de sumarização e processamento desse grande volume de informação, ele se torna o núcleo dos sistemas de informações gerenciais e apoio à decisão das principais soluções de inteligência de negócio do mercado.

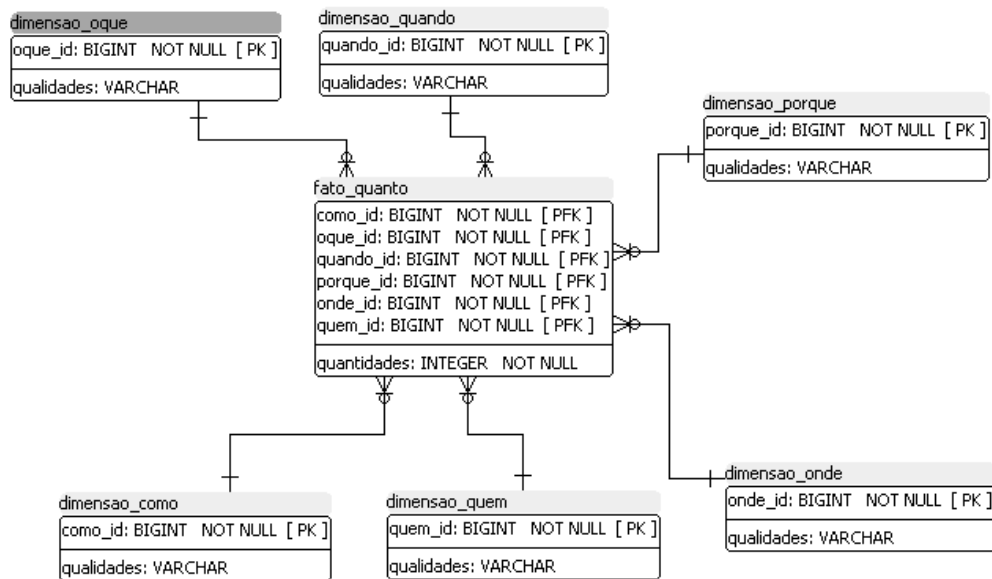
O DW, abreviação pela qual os *Data Warehouses* são mais comumente conhecidos, é o conceito-base para a montagem de um sistema de dados com grande propensão à sua utilização em contextos gerenciais e executivos. Sendo assim, é possível compreender de forma mais intuitiva o motivo pelo qual sua estrutura e implementação muitas vezes se confunde com o próprio conceito de “Inteligência de Negócio” e viabiliza uma introdução coerente para o melhor entendimento do termo.

A organização dos dados dentro do DW é primordial para agilizar o processamento de grandes volumes de dados. Para isso são propostos formatos de modelagem normalizados, sumarizados e discriminados por dimensão (contexto) na construção de cada *Data Mart* (BARBIERI, 2011, p. 96). Esses formatos, agrupados sob a alcunha de “modelagem dimensional”, permitem ao analista: (i) diminuir a necessidade de junções de tabelas, operações custosas na hora de construir análises; (ii) utilizar a mesma dimensão para diferentes *Data Marts*, construindo um relacionamento entre eles; e (iii) armazenar os resultados de operações

decorrentes da interseção entre as diferentes dimensões em diversos níveis de granularidade, agilizando cálculos que seriam realizados na geração de um relatório.

A modelagem estrela é o principal expoente da “modelagem dimensional”. Ela é um modelo de estruturação de ambientes analíticos que viabiliza a sua organização conceitual aliada à diminuição de custos de armazenamento e, ao mesmo tempo, à otimização da execução de consultas. Idealizada e exposta inicialmente por Ralph Kimball, hoje é aceita de forma quase unânime na área de inteligência de negócio como boa prática de modelagem. Para a sua construção, é necessário separarmos as informações em tabelas “fato” e “dimensão”, conforme exibido na Figura 1.

Figura 1: Exemplo de modelagem estrela



Fonte: Flávio Ceci, 2012

As tabelas do tipo “dimensão” contêm todas as informações periféricas ao dado, relacionando-o a um período, um local ou uma pessoa, por exemplo. Muitos utilizam um processo de gestão chamado 5W2H para modelar tais informações, ou seja, buscam responder para o dado as questões: “o quê” (*what*), “quando” (*when*), “por quê” (*why*), “onde” (*where*), “quem” (*who*) e “como” (*how*). Entretanto, o último H do processo, “quanto” (*how much*), é reservado para a tabela fato.

A tabela do tipo “fato”, por sua vez, é responsável por armazenar as métricas e os valores que deverão ser agrupados conforme utilizamos as informações das dimensões para filtrá-la. Ela se conecta unicamente com um registro de cada dimensão, devendo possuir, portanto, a informação em quantidade granular referente à interligação de registros específicos de cada

dimensão. Com a evolução desse modelo, pode-se relacionar as mesmas dimensões a outras tabelas “fato”, construindo assim uma espécie de “constelação” de dados.

2.6 VISUALIZAÇÃO DA INFORMAÇÃO

A visualização de dados refere-se às técnicas usadas para transmitir conceitos de modo universal por meio de gráficos e diagramas. O objetivo principal é auxiliar o ser humano a processar informações de forma eficaz, assegurando um correto entendimento para uma melhor tomada de decisão. Dessa forma, podemos dizer que, quando estamos construindo o processo de visualização de dados, “elaboramos uma narrativa convincente e contamos uma história.” (KNAFLIC, 2017).

2.6.1 Benefícios

Existem quatro motivos pelos quais a visualização de dados tem se tornado uma estratégia cada vez mais presente nos mais variados ambientes de trabalho pelo mundo. São eles: (i) permitir o controle de tendências ao longo do tempo; (ii) identificar a significância do dado para a empresa; (iii) viabilizar o embasamento de decisões; e (iv) ter visão do panorama geral (O que é visualização de dados? [s.d.]).

O primeiro ponto, “permitir o controle de tendências ao longo do tempo”, é essencial para fazer previsões assertivas, tendo por base o histórico do negócio e o seu comportamento ao longo do tempo. Dessa forma, o resultado esperado para o futuro do negócio é construído a partir do dado, ao invés de simples especulações.

Já o segundo ponto, “identificar a significância do dado para a empresa”, trata de encontrar as informações que caracterizam o seu negócio, entendendo como cada fator do seu projeto se relaciona entre si. E a partir dessas conexões, as partes interessadas ficam aptas para escolher quais os elementos mais prováveis de impactar no cumprimento dos objetivos propostos.

O próximo ponto trata de “viabilizar o embasamento de decisões”. A partir do fluxo de inteligência de negócio construído, especialmente se há uma governança de dados corretamente implementada, é possível mapear todo o caminho do dado, desde o momento em que ele é

gerado até chegar ao relatório gerencial. Com a validação desse fluxo, certifica-se a assertividade da informação que será utilizada para a tomada de decisões.

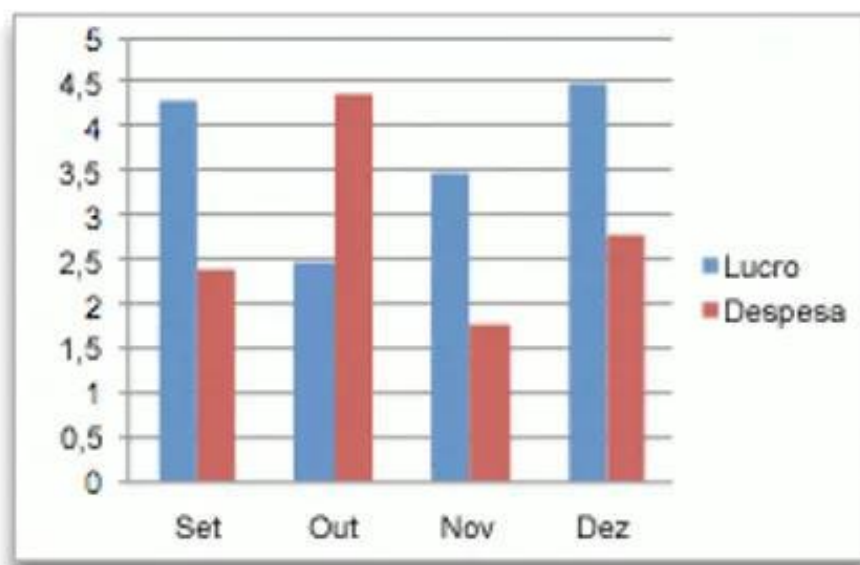
Por último, é necessário “ter visão do panorama geral”. O objetivo principal nesse ponto é contar a história e transmitir a mensagem da empresa e/ou instituição de forma tempestiva, clara e adequada por meio de diagramas, gráficos ou outras representações visuais. Assim garantimos que todas as partes interessadas, mesmo que minimamente, entendam o contexto geral do processo.

2.6.2 Tipos de visualização

O grande desafio da estratégia de visualização dos dados é validar quais dos mais diversos tipos de gráficos e diagramas são necessários para que a mensagem por trás da imagem seja passada de forma clara e objetiva. Pois, apesar de todos eles poderem apresentar o mesmo conjunto de dados, a quantidade de informações extraídas em cada um deles é diferente. Dentre esses gráficos e diagramas, os principais gráficos são os seguintes (KNAFLIC, 2017):

a) Gráfico de barra ou colunas

Figura 2: Exemplo de gráfico de barra



Fonte: Gabriela Xavier, 2019

O gráfico de barra ou coluna apresenta um conjunto de dados categóricos resumidamente. Nesse gráfico, cada um dos retângulos representa uma nova categoria e todos

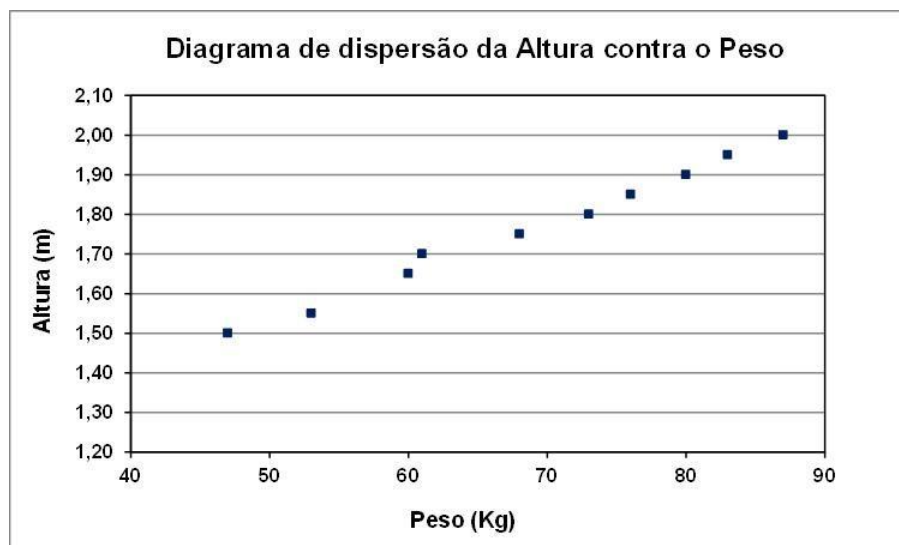
apresentam a mesma largura. Já a sua altura é proporcional a uma associação específica, como por exemplo a soma de valores de uma determinada categoria, conforme apresentado na Figura 2.

Esses gráficos são utilizados com grande frequência por revistas, jornais e outros meios de comunicação. São mais indicados para uso quando se necessita um gráfico de fácil análise e compreensão.

b) Gráfico de Dispersão

O gráfico de dispersão é um tipo de visualização circular que busca representar uma associação entre pares de dados. Geralmente, deseja-se determinar se existe uma relação entre duas variáveis.

Figura 3: Exemplo de gráfico de dispersão



Fonte: Marcelo Toledo, 2019

Eles geralmente buscam fornecer evidências visuais sobre as causas e efeitos de determinadas variáveis. Essas evidências podem ser positivas, como é o caso da Figura 3 que mostra que conforme o peso da pessoa aumenta sua altura também aumenta. Mas também, elas podem ser negativas (quando o valor de uma das dimensões diminui enquanto o outro aumenta) ou nula (quando elas não apresentam relações entre si).

c) Gráfico de linhas

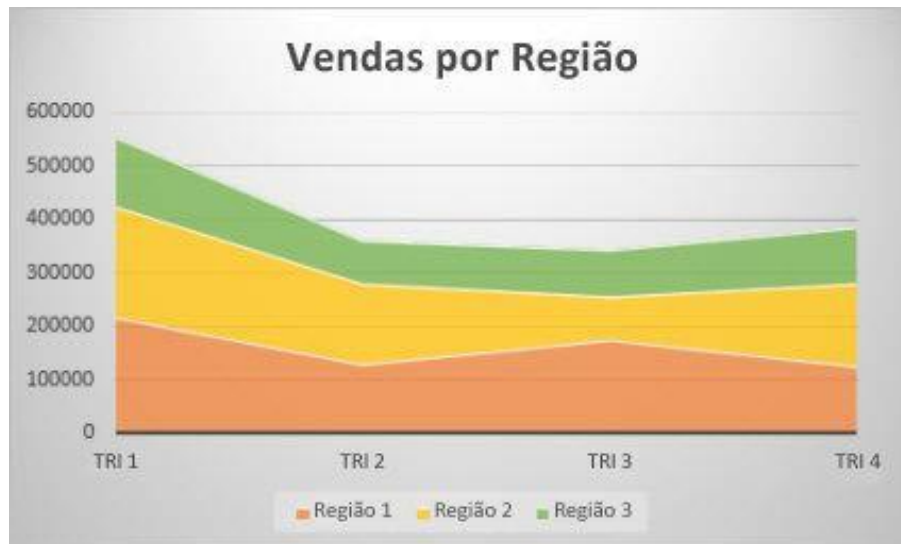
Figura 4: Exemplo de gráfico de linhas

Fonte: Virgílio F. M. dos Santos, 2019

O gráfico de linhas é um tipo de gráfico utilizado para demonstrar a evolução ou regressão de um certo dado ao longo do tempo. Eles facilitam muito o usuário a identificar tendências ou anomalias, como na Figura 4, onde é possível identificar que o ponto de máximo de vendas se encontra no mês de março e o ponto de mínimo no mês de julho.

São muito utilizados pelo setor financeiro, como no caso da bolsa de valores, identificando assim a variação de determinada ação. Além disso, são comumente exibidos por empresas e instituições com o fim de identificarem o atingimento de metas, pelo fato de serem ótimos para representar sequências de dados em uma escala de tempo dividida em períodos iguais.

d) Gráfico de área

Figura 5: Exemplo de gráfico de área

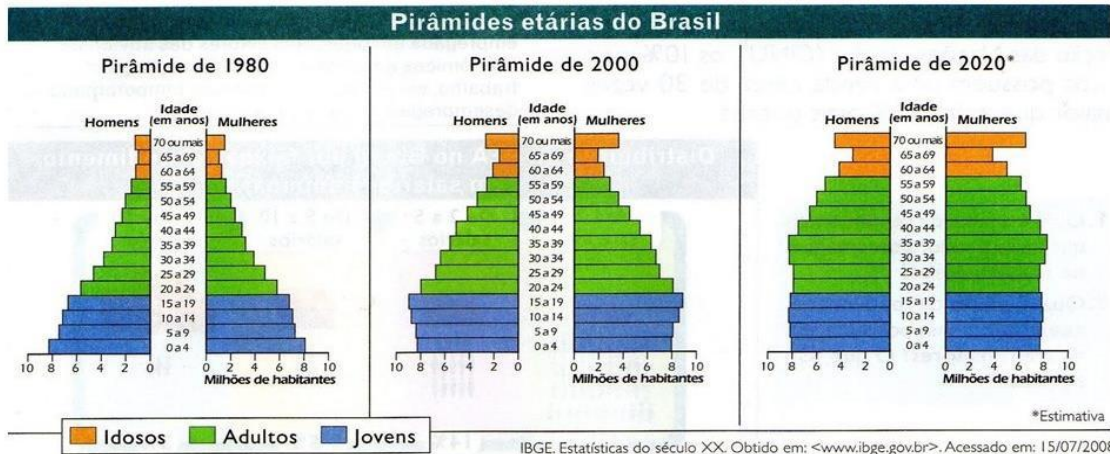
Fonte: Microsoft, 2020

O gráfico de área é representado por uma área preenchida cujo tamanho é determinado pelos valores dos dados. Suas duas principais funções são destacar as oscilações de uma categoria de acordo com variáveis como o tempo; e também nos ajudam a evidenciar uma noção de proporção sobre o todo.

As empresas que trabalham com a venda de determinados produtos costumam se beneficiar bastante com esse tipo de gráfico, pelo fato dele ser muito bom em mostrar a tendência de vendas durante um determinado período de tempo, conforme apresentado na Figura 5. Além disso, ele também tem muita utilidade quando se deseja medir o desempenho de diversos segmentos e a tendência de custos.

e) Pirâmide Etária

Figura 6: Exemplos de gráficos de pirâmide etária



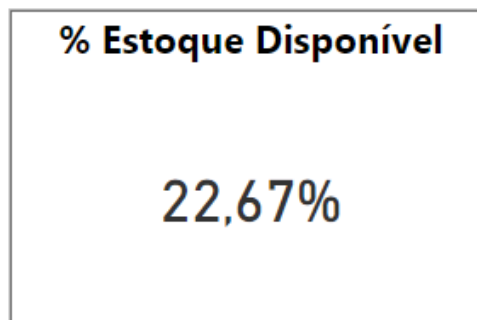
Fonte: IBGE, 2008

As pirâmides etárias são gráficos que apresentam a população de um determinado país ou região com relação às faixas etárias da mesma, separada por gêneros. Elas são organizadas de forma a facilitar a visualização da informação em formato de barras superpostas, mas com os eixos posicionados de forma simétrica. A Figura 6 destaca os dois gêneros em diferentes lados, mostrando a proporção de pessoas por cada faixa de idades, sendo fácil a identificação de qual a faixa etária contém uma quantidade maior de pessoas e se olhamos os gráficos dos diversos anos, conseguimos entender a evolução entre as faixas etárias da população.

Esse tipo de gráfico, quando utilizado em sequenciamento, como apresentado na Figura 6, auxilia a entender a evolução da população (avaliando as taxas de natalidade). Nesse contexto, podem, por exemplo, ajudar a elaborar um planejamento público a médio e longo prazo.

f) Cartões

Figura 7: Exemplo de Cartão



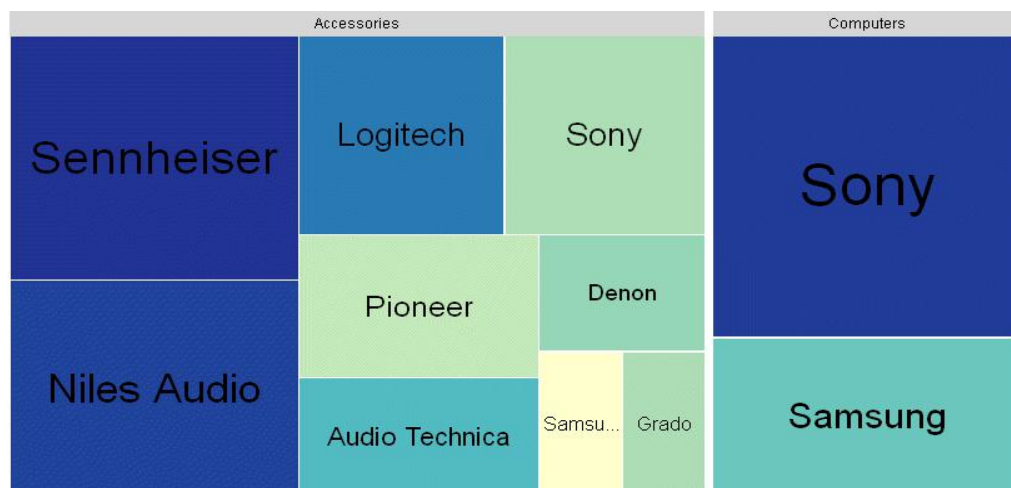
Fonte: Meirieli Ribeiro, 2019

Os cartões de informação são empregados quando se deseja destacar uma única informação, resumindo dados em uma saída que torne o processo de leitura “instantâneo”, ou porque aquela informação é a mais importante (em detrimento das outras). O formato de gráfico "Cartão" é um formato simples e eficiente. A Figura 7, por exemplo, destaca o percentual de estoque disponível naquele momento, exemplificando o caso de um painel que busca fazer a gestão de um estoque e sempre está informando a disponibilidade dos seus produtos como um todo.

g) Mapa de Árvore

O Mapa de Árvore (mais conhecido como *TreeMap*) é um formato de visualização para comparação quantitativa entre categorias. Nele, é possível fazer comparação tanto entre os tamanhos das áreas dos retângulos (maiores possuem valores mais altos que os menores) como entre as cores, onde os mais escuros possuem valores mais altos que os mais claros. Com ele, também é possível discriminar diferentes grupos de categorias através da utilização de cores contrastantes. No exemplo da Figura 8, vemos duas grandes categorias - acessórios e computadores - e as diversas proporções para cada marca dentro de cada categoria. O gráfico torna claras as marcas com maior representatividade em cada categoria.

Figura 8: Exemplo de Mapa de Árvore (TreeMap)



Fonte: *Information Builders*, 2017

Sua utilização é conveniente quando não é relevante saber a quantidade exata de uma categoria, mas sim o quanto ela se diferencia das categorias restantes. Também deve-se levar em consideração ao utilizar este formato que as informações de menor valor não estarão muito visíveis. É especialmente recomendado para comparação entre poucas categorias.

h) Mapa de Calor

Figura 9: Exemplo de mapa de calor

Fonte: Paulo Alves, 2020

Enfim, para apresentar uma visão geográfica dos dados podemos empregar os mapas de calor. Eles nos permitem contrastar informações entre as regiões, se utilizando de marcas com diferentes tamanhos, cores e intensidades (e, às vezes, simplesmente pinturas). A Figura 9 está presente em muitos jogos eletrônicos e filmes, destacando quais foram os lugares do mundo impactados por determinado fato, como por exemplo uma pandemia.

Podem ser utilizados em cima de quaisquer dados que possuam dimensão geográfica e costumam ser úteis para efeito de comparação entre países, estados ou cidades (entre outros).

3 DESENVOLVIMENTO DA SOLUÇÃO ANALÍTICA

Com base nas técnicas e conceitos discutidos no capítulo anterior, foram desenvolvidos os entregáveis deste trabalho. Este capítulo destina-se a apresentar o desenvolvimento dos entregáveis de acordo com as fases de planejamento e execução, destacando as ferramentas empregadas.

Para facilitar o entendimento, o desenvolvimento de cada entregável é apresentado em uma seção. Deste modo, após a apresentação das ferramentas adotadas no trabalho, será apresentado o desenvolvimento do nosso primeiro entregável, uma modelagem em esquema estrela de um *Data Mart*. Esse *Data Mart* tem por objetivo facilitar a análise dos dados obtidos a partir do formulário da OMS, adotado no Projeto VODAN BR.

O desenvolvimento do segundo entregável é apresentado em seguida. Esse entregável refere-se a um relatório de visualização de dados, desenvolvido a partir de dados clínicos de pacientes com COVID-19 disponibilizados no repositório da FAPESP.

3.1 FERRAMENTAS

Durante o desenvolvimento do projeto, nos deparamos com diversas possibilidades de ferramentas a serem utilizadas nas mais variadas fases da implementação. Essa diversidade tornou necessária a seleção de ferramentas para cada um dos momentos do processo. Os próximos tópicos abordam as principais ferramentas utilizadas, a perspectiva de suas escolhas e como elas atendem a arquitetura projetada .

3.1.1 SQL Power Architect

Para a construção de uma modelagem dimensional para o banco de dados, foi adotada a ferramenta *SQL Power Architect Community Edition*. A ferramenta foi desenvolvida por projetistas de DW e dispõe de recursos exclusivos para arquitetos de *Data Warehouse*.

Essa ferramenta foi selecionada por permitir a construção de diagramas de banco de dados de forma facilitada e já ser conhecida pelos autores deste trabalho. Além disso, provê integração com diversos sistemas gerenciadores de banco de dados, agilizando a geração dos *scripts DDL (Data Definition Language)* para criação das tabelas. Outro aspecto que

influenciou na escolha foi a interface visual simples do *software*. Cabe destacar que neste trabalho foi adotada a versão gratuita da ferramenta.

3.1.2 Pentaho Data Integration (PDI)

De acordo com Bouman e Dangen (2009), o Pentaho é uma suíte de inteligência de negócio, e não um único produto. Ele é formado por um conjunto de programas de computador que trabalham juntos para criar e oferecer soluções de inteligência de negócio.

O Pentaho Data Integration, também conhecido como PDI, kettle ou spoon, é uma ferramenta poderosa e flexível de ETL fornecida sob licença gratuita e código aberto. A simplicidade na construção de fluxos de inteligência de negócio, através do sequenciamento visual de passos de tratamento de dados, permite que mesmo um leigo realize o acompanhamento da execução de um processo. Além disso, é possível executá-lo tanto em uma máquina como em arquiteturas distribuídas.

Na construção do fluxo de tratamento de dados, é possível criar dois tipos de arquivo: *Job* (executor) e *Transformation* (transformação). O *Job* é utilizado para controle do fluxo a nível macro, executando sequencialmente as transformações - e verificando suas saídas - e as tomadas de decisão. Também possui passos específicos para controle do fluxo e transformações monolíticas. A *Transformation*, por sua vez, executa seu fluxo de forma paralela e possui todos os passos necessários para efetuar extração de dados de diversas fontes, junção de linhas, cargas em tabelas, agrupamento de informações e etc.

3.1.3 MySQL

O MySQL é um banco de dados relacional bem conceituado no mercado. Ele oferece uma solução robusta, sem custos, e com uma grande comunidade.

Apesar de ser mais indicada para arquiteturas transacionais, esse banco de dados foi selecionado por não apresentar custos e dispor de boa documentação na internet, tanto de seu fornecedor quanto da própria interação entre usuários nos fóruns. Outro fator relevante refere-se ao fato deste ser o banco de dados adotado pela instituição coordenadora do projeto VODAN BR como base de dados relacional.

Convém ressaltar que existem soluções mais adequadas ao viés analítico. Porém, essas soluções são pagas ou exigem um esforço relevante de aprendizado, sendo, por isso, consideradas fora do escopo deste trabalho.

3.1.4 Power BI

Para gerar a visualização, foram analisadas três ferramentas de visualização de dados conhecidas no mercado. Essas ferramentas são similares no sentido de proporcionarem uma série de facilidades se comparadas a outras. São elas: *Data Studio*², *Tableau*³ e *Power BI*⁴.

O *Data Studio* é uma ferramenta gratuita do Google que ajuda na exibição dos dados através de relatórios informativos, fáceis de ler, fáceis de compartilhar e totalmente personalizáveis. A grande desvantagem reside nas conexões com fontes de dados de terceiros, que às vezes podem se desconectar, gerando visualizações com dados tratados como “ausentes” no painel. Além disso, ele apresenta um alto nível de dificuldade no desenvolvimento de gráficos e análises mais complexas.

O Tableau, por sua vez, é fácil de ser conectado com as mais variadas fontes de dados e produz painéis extremamente personalizáveis. A desvantagem na ferramenta está em seu emprego com arquivos de dados maiores. Nesses casos, os arquivos levam um tempo considerável para serem carregados e os seus painéis, apesar de facilmente personalizáveis, são limitados com relação à quantidade e aos tipos de gráficos disponíveis.

O Power BI, ferramenta provida pela Microsoft sob uma licença paga para usos corporativos e comerciais, é uma das principais ferramentas do mercado para a construção de visualização de dados. Ela provê uma integração nativa com toda a suíte de aplicativos da Microsoft. Além disso, trabalha muito bem com a modelagem dimensional e possibilita, entre outras coisas, cálculos complexos a partir da utilização da linguagem DAX. A ferramenta dispõe de um número considerável de tipos de gráficos para a construção das visualizações, possibilitando também efetuar o *download* de novos formatos visuais, caso necessário. Com essas características, ela consegue tornar viável e prática a construção de relatórios altamente personalizados.

² <https://datastudio.google.com/>

³ <https://www.tableau.com/pt-br>

⁴ <https://powerbi.microsoft.com/pt-br/>

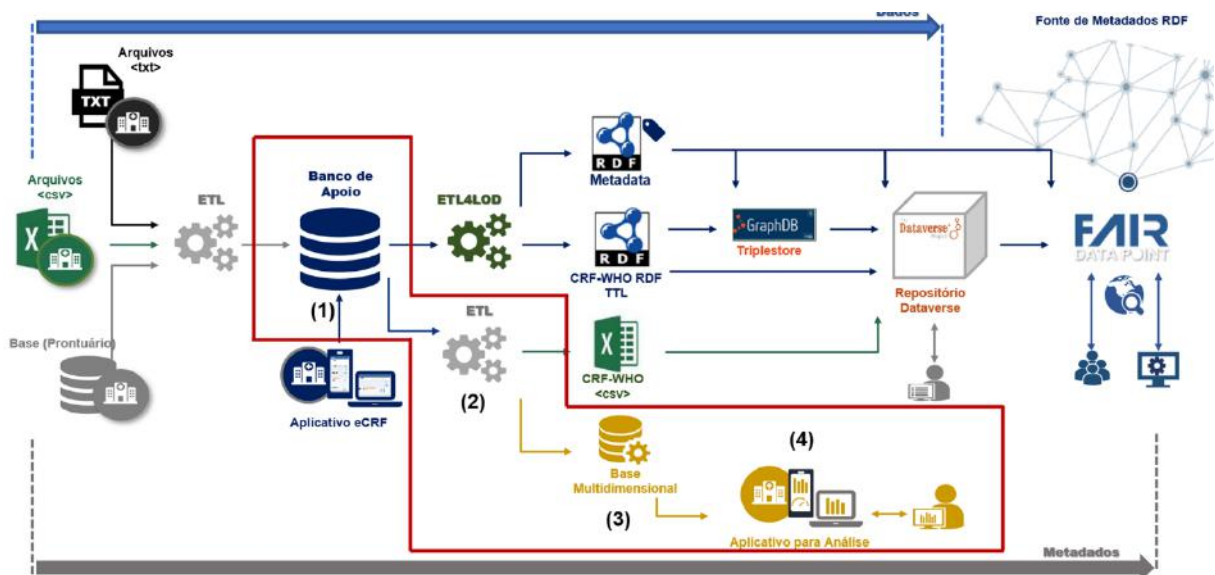
Outro ponto relevante sobre o Power BI é a facilidade em criar um único modelo com várias bases de dados diferentes, podendo elas serem extraídas das mais diversas fontes de dados. A diversidade de gráficos e de manipulações possíveis é grande. Esse mesmo fato se torna um dos poucos pontos negativos que se trata da necessidade de possuir conhecimento em Excel e em suas funções para que os dados assim possam ser manipulados de forma correta.

Dessa forma, o Power BI foi a ferramenta selecionada para a utilização no projeto VODAN BR devido a sua alta gama de funcionalidades e, principalmente, ao fato dos integrantes do projeto estarem mais familiarizados com essa ferramenta. A familiaridade com a ferramenta permitiu a sua utilização sem um tempo prévio para estudos e adaptação/entendimento.

3.2 DESENVOLVIMENTO DO DATA MART DO PROJETO VODAN BR

De acordo com a Arquitetura do Projeto VODAN BR, o ambiente analítico deve empregar o banco de apoio (*staging database*) para a extração dos dados de análise. Desse modo, a arquitetura do ambiente analítico, apresentada no destaque em vermelho da Figura 10, deve empregar um processo ETL (2) para extrair os dados do banco de apoio (1), transformando-os de acordo com o modelo multidimensional estabelecido e armazenando-os na base de dados multidimensional (3). A partir dessa base (1) são estabelecidos os painéis de visualização (4) com as análises produzidas sobre os dados.

Figura 10: Destaque do Ambiente Analítico na Arquitetura do Projeto VODAN BR



Fonte: Projeto VODAN BR, 2020

O esquema na Figura 11 representa a modelagem para atender a pesquisas clínicas de pacientes (anonimizados) de acordo com um formulário de pesquisa clínica previamente definido. Nesse esquema:

- Entidades na cor verde representam a modelagem de um questionário de pesquisa, contemplando os módulos que o compõem, as questões associadas e as respostas padronizadas, quando houver; e
- Entidades na cor laranja representam o registro das informações coletadas. Dentre essas entidades, a `tb_Participant` registra o prontuário médico designado a um paciente em um hospital; `tb_Hospital` representa o cadastro de hospitais participantes; `tb_assessmentQuestionnaire` identifica para qual questionário está sendo coletada as informações; `tb_FormRecord` registra os módulos do questionário selecionado que será respondido; e `tb_QuestionGroupFormRecord` armazena as informações para cada questão de um módulo de um questionário.

O Formulário da OMS é constituído por três módulos: (i) o módulo de admissão, destinado à coleta das informações no primeiro dia do paciente no hospital; (ii) o módulo de acompanhamento, destinado a registrar as informações dos pacientes durante o período de hospitalização; e (iii) o módulo de desfecho, destinado a registrar a conclusão do atendimento do paciente.

Cada módulo apresenta um conjunto de questões sobre o paciente (referentes aos atendimentos) organizadas em grupos. Alguns desses grupos são comuns a mais de um módulo como, por exemplo, os exames laboratoriais. Estes são colhidos na admissão e durante o período de internação. A Figura 12 apresenta o grupamento referente aos “Resultados Laboratoriais” enquanto a Figura 13 apresenta aquele referente aos “Cuidados”, ambos associados ao módulo de acompanhamento.

Figura 12: Extrato do Módulo de Acompanhamento destacando a questão sobre “Resultados Laboratoriais”

2c. LABORATORY RESULTS (*record units if different from those listed)					
Parameter	Value*	Not done	Parameter	Value*	Not done
Haemoglobin (g/L)		<input type="checkbox"/>	Creatinine (µmol/L)		<input type="checkbox"/>
WBC count (x10 ⁹ /L)		<input type="checkbox"/>	Sodium (mEq/L)		<input type="checkbox"/>
Haematocrit (%)		<input type="checkbox"/>	Potassium (mEq/L)		<input type="checkbox"/>
Platelets (x10 ⁹ /L)		<input type="checkbox"/>	Procalcitonin (ng/mL)		<input type="checkbox"/>
APTT/APTR		<input type="checkbox"/>	CRP (mg/L)		<input type="checkbox"/>
PT (seconds)		<input type="checkbox"/>	LDH (U/L)		<input type="checkbox"/>
INR		<input type="checkbox"/>	Creatine kinase (U/L)		<input type="checkbox"/>
ALT/SGPT (U/L)		<input type="checkbox"/>	Troponin (ng/mL)		<input type="checkbox"/>
Total bilirubin (µmol/L)		<input type="checkbox"/>	ESR (mm/hr)		<input type="checkbox"/>
AST/SGOT (U/L)		<input type="checkbox"/>	D-dimer (mg/L)		<input type="checkbox"/>
Urea (BUN) (mmol/L)		<input type="checkbox"/>	Ferritin (ng/mL)		<input type="checkbox"/>
Lactate (mmol/L)		<input type="checkbox"/>	IL-6 (pg/mL)		<input type="checkbox"/>

Figura 13: Extrato do Módulo de Acompanhamento destacando a questão sobre “Cuidados”

2e. SUPPORTIVE CARE *At any time during this 24-hour hospital day, did the patient receive:*

ICU or High Dependency Unit admission? Yes No Unknown
 Date of ICU/HDU admission [D][D]/[M][M]/[2][0][Y][Y] Unknown
 ICU/HDU discharge date [D][D]/[M][M]/[2][0][Y][Y] Not discharged yet Unknown

Oxygen therapy? Yes No Unknown **If yes, complete all below:**
O₂ flow: 1-5 L/min 6-10 L/min 11-15 L/min >15 L/min Unknown
Source of oxygen: Piped Cylinder Concentrator Unknown
Interface: Nasal prongs HF nasal cannula Mask Mask with reservoir CPAP/NIV mask Unknown

Non-invasive ventilation? (e.g. BIPAP, CPAP) Yes No Unknown **Prone position?** Yes No Unknown

Invasive ventilation (Any)? Yes No Unknown **If yes, what were the following values closest to 0800:**
 PEEP (cm H₂O) ____; F_iO₂ (%) ____; Plateau pressure (cm H₂O) ____; P_aCO₂ ____; P_aO₂ ____

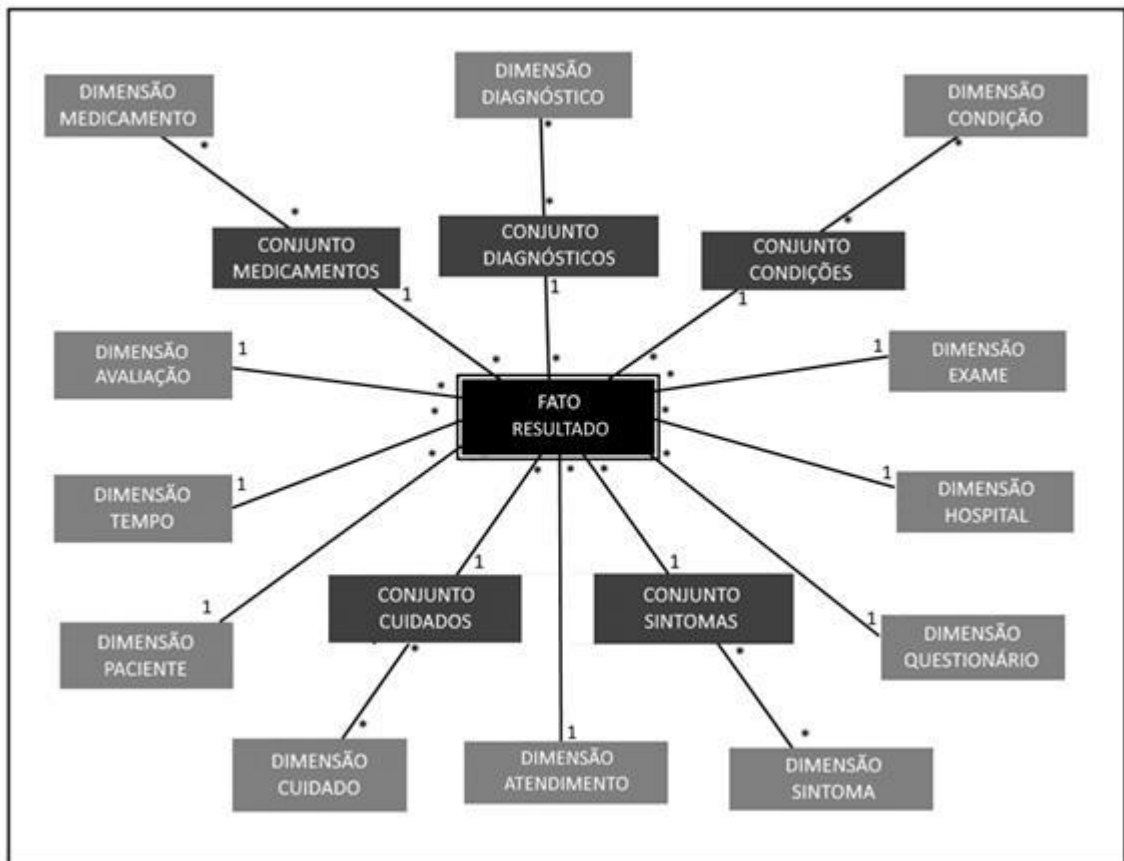
Extracorporeal (ECMO) support? Yes No Unknown **Inotropes/vasopressors?** Yes No Unknown

Renal replacement therapy (RRT) or dialysis? Yes No Unknown

A partir desse formulário, foi possível propor uma modelagem dimensional, segundo o modelo estrela, conforme ensinado por Barbieri (2011), com o objetivo de maximizar as análises relativas à linha temporal e a resposta dos pacientes a diversos estímulos (positivos e negativos) do tratamento.

Convém ressaltar que, dado o momento do projeto VODAN BR, não foi possível obter dados diretamente dos hospitais participantes ou utilizar máquinas clusterizadas para trabalhar grandes volumes de dados. Portanto, o modelo proposto neste trabalho deverá ser revisitado quando da conclusão do processo de carga dos hospitais.

Figura 14: Tabelas e relacionamentos do modelo estrela



Na modelagem-base, apresentada na Figura 14, trabalhamos com a granularidade de “resultado”. Entenda-se aqui o resultado como sendo o valor (quantitativo ou qualitativo) referente ao resultado de um exame (analito de exame, para ser mais assertivo) de uma “avaliação”. É interessante notar que, conforme esta modelagem, qualquer verificação acerca da situação do paciente ou do correto funcionamento do seu organismo é considerada um exame (ex.: exames laboratoriais, medição de temperatura, frequência cardíaca etc.). Entendemos que esse modelo não contempla todas as possibilidades de análise que podem ser extraídas a partir do formulário da OMS. Esperamos, no futuro, alguns desdobramentos com tabelas-fato que permitam análises de outras métricas do atendimento.

A captação de dados pelas diferentes dimensões propostas na modelagem pode ser compreendida de forma mais fácil ao entrar a fundo em uma das dimensões, como por exemplo a dimensão avaliação. A avaliação, é a instância de preenchimento (total ou parcial) do “questionário” realizado pelo funcionário de um “hospital”. A avaliação é feita diversas vezes durante o “atendimento” do “paciente”. Cada preenchimento acontece em uma instância de “tempo” específico e reflete a situação do paciente (quais “sintomas” está sentindo, quais as suas “condições” – entenda-se aqui como fatores relevantes que podem influenciar o quadro de saúde do paciente -, quais “cuidados” está recebendo, com quais “medicamentos” está sendo medicado e quais quadros de “diagnóstico” estão sendo tratados).

Como exemplo, digamos que uma mulher (“paciente”) grávida (“condição”) foi admitida (gera-se um “atendimento”) no HSL (“hospital”) às 13 horas de uma sexta-feira (“tempo”) se sentindo fraca (“sintoma”) e foi realizado um preenchimento (“avaliação”) do formulário (“questionário”) onde constou o exame laboratorial (“exame”) no qual o resultado da sua quantidade de hemoglobinas veio aquém do ideal (“resultado”). Em conjunto com outros exames, foi diagnosticado um quadro de anemia (“diagnóstico”) e administrado um remédio (“medicamento”). Além disso, a paciente precisou de transfusão de sangue (“cuidado”). Neste exemplo constam todas as tabelas apresentadas, ainda que provavelmente seus dados seriam preenchidos em uma sucessão de avaliações, e não em uma única. O medicamento, por exemplo, não estaria sendo administrado na primeira avaliação.

3.2.1. A tabela fato

A tabela fato “Resultado”, conforme exposto anteriormente, busca sumarizar o resultado de um exame. Para isso, julgamos importante incluir como métricas tanto o valor quantitativo quanto o valor qualitativo, sendo um ou outro preenchidos conforme o tipo de métrica do exame (apresentados ao final da Figura 15).

Figura 15: Campos da tabela fato "Resultado"

FATO RESULTADO	
tempo_id: INTEGER	CHAVES DE RELACIONAMENTO
paciente_id: INTEGER	
hospital_id: INTEGER	
questionario_id: INTEGER	
diagnosticos_id: INTEGER	
condicoes_id: INTEGER	
medicacoes_id: INTEGER	
cuidados_id: INTEGER	
sintomas_id: INTEGER	
avaliacao_id: INTEGER	
exame_id: INTEGER	
atendimento_id: INTEGER	
<hr/>	
valor_quant: NUMERIC(16, 9)	MÉTRICAS
valor_quant_normalizado: NUMERIC(16, 9)	
valor_qualit: VARCHAR	
obs: VARCHAR	

Essa tabela fato possui um nível tal de granularidade que é capaz de expor a situação de um paciente no tempo, tornando possível analisar a sucessão de acontecimentos para um dado grupo ou pessoa com relação a outros fatores como medicamentos utilizados, sintomas relatados e, até mesmo, comorbidades existentes. Desta forma, se tornaria trivial a construção de análises estatísticas visuais para certos assuntos que geraram até mesmo conflitos políticos na história recente do Brasil, como o uso de hidroxiquina para o tratamento de pessoas infectadas pelo novo coronavírus, entre outros.

3.2.2. As tabelas dimensão

Conforme relacionadas anteriormente, podemos enumerar 12 tabelas-dimensão: tempo, paciente, hospital, questionário, diagnóstico, condição, medicação, cuidado, sintoma, exame, avaliação e atendimento. Destas, cinco possuem relacionamento MxN entre a tabela fato e a dimensão, exigindo uma tabela ponte: diagnóstico, condição, medicação, cuidado e sintoma.

Figura 16: Campos da dimensão "Tempo"

DIMENSÃO TEMPO	
tempo_id:	INTEGER
<hr/>	
ano:	INTEGER
mes:	INTEGER
dia:	INTEGER
periodo:	INTEGER
hora:	INTEGER
seq:	INTEGER

A primeira dimensão a ser destacada é a dimensão tempo (vide Figura 16), normalmente utilizada para toda e qualquer modelagem analítica devido à sua alta incidência nos conjuntos de análise relacionados a tomadas de decisão. Aqui, preferimos trazer o nível de granularidade máximo possível, horário, permitindo dessa forma analisar a situação do(s) paciente(s) no decorrer do mesmo dia. Isso é essencial para se analisar variações de estado mais abruptas ou decorrentes de medidas emergenciais.

Figura 17: Campos da dimensão "Paciente"

DIMENSÃO PACIENTE	
paciente_id:	INTEGER
<hr/>	
identificador:	VARCHAR(50)
sexo:	VARCHAR(50)
faixa_idade:	INTEGER
idade:	INTEGER
altura_cm:	INTEGER
peso_kg:	INTEGER
circunferencia_braquial_mm:	INTEGER

A dimensão “Paciente”, por sua vez, busca trazer informações censitárias e classificatórias, mapeando os diferentes fatores que podem influenciar na resposta do paciente à doença e aos medicamentos. É de substancial interesse, por exemplo, compreender a influência da idade (apresentado entre as métricas da Figura 17) nos resultados analíticos

Figura 18: Campos da dimensão "Hospital"

DIMENSÃO HOSPITAL	
hospital_id:	INTEGER
<hr/>	
pais:	VARCHAR
estado:	VARCHAR
cidade:	VARCHAR
bairro:	VARCHAR
nome:	VARCHAR
latitude:	NUMERIC(9, 6)
longitude:	NUMERIC(9, 6)

A dimensão “Hospital”, apresentada na Figura 18, já busca nos trazer uma visão mais regional da informação. É interessante notar que ela nos permite definir os locais de maior busca de atendimento e, em nível macro, as regiões mais afetadas pelo coronavírus.

Figura 19: Campos da dimensão "Questionário"

DIMENSÃO QUESTIONÁRIO	
questionario_id:	INTEGER
versao:	INTEGER
descricao:	VARCHAR

A dimensão “Questionário” (vide Figura 19) visa estabelecer a origem dos dados. Esse tipo de informação é interessante para comparar os meios de obtenção e catalogá-los conforme necessidade. Dessa forma, é possível futuramente inserir informações advindas de outras fontes e, ao mesmo tempo, manter a base organizada. Além de entender se as versões do questionário ajudam a entender determinado comportamento.

Figura 20: Campos da dimensão "Diagnóstico"

DIMENSÃO DIAGNÓSTICO	
diagnostico_id:	INTEGER
descricao:	VARCHAR

A partir da dimensão “Diagnóstico”, podemos mapear e entender a correlação entre as alterações fisiológicas e sintomas junto às patologias conhecidas, de forma a facilitar uma correspondência estatística entre ambos e possibilitar diagnósticos mais assertivos. Além de futuramente ser possível realizar análises textuais no campo “descrição”, presente na Figura 20, para entender determinado padrão dado um diagnóstico.

Figura 21: Campos da dimensão "Condição"

DIMENSÃO CONDIÇÃO	
condicao_id:	INTEGER
cronica:	BOOLEAN
descricao:	VARCHAR

A dimensão “Condição”, embutida na Figura 21, tem importância vital no mapeamento da influência de comorbidades e outros fatores como gravidez, sedentarismo e quadros psicológicos (entre outros) no desenvolvimento da patologia.

Figura 22: Campos da dimensão "Medicação"

DIMENSÃO MEDICAÇÃO	
medicacao_id:	INTEGER
<hr/>	
uso_continuo:	BOOLEAN
descricao:	VARCHAR
dosagem_mg:	INTEGER

A dimensão “Medicação” (Figura 22) visa fornecer os meios de compreender em larga escala o efeito do uso de uma medicação para o quadro clínico do paciente, tanto no combate a sintomas quanto no combate à própria doença.

Figura 23: Campos da dimensão "Cuidado"

DIMENSÃO CUIDADO	
cuidado_id:	INTEGER
<hr/>	
descricao:	VARCHAR

A dimensão “Cuidado”, mostrada na Figura 23, busca pontuar a utilização de abordagens mais intensivas no tratamento de alguns pacientes, como a utilização de respiradores, leito em UTI, etc.

Figura 24: Campos da dimensão "Sintoma"

DIMENSÃO SINTOMA	
sintoma_id:	INTEGER
<hr/>	
descricao:	VARCHAR

A dimensão “Sintoma” (Figura 24) é responsável por mapear a ocorrência de sintomas e distúrbios fisiológicos apresentados pelo paciente para a patologia diagnosticada.

Figura 25: Campos da dimensão "Exame"

DIMENSÃO EXAME	
exame_id:	INTEGER
<hr/>	
tipo_origem:	VARCHAR
tipo_metrica:	VARCHAR
nome:	VARCHAR
analito:	VARCHAR
unidade_medida:	VARCHAR
ref_quant_inferior:	NUMERIC(16, 9)
ref_quant_superior:	NUMERIC(16, 9)
ref_qualit:	VARCHAR

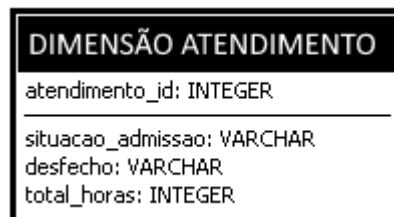
A dimensão “Exame” busca conciliar todo tipo de medição possível da situação do paciente em sua granularidade máxima. Desde exames laboratoriais (como um hemograma com resultados para cada analito) até exames de triagem (como medição de temperatura e pressão). Também categoriza por origem (laboratorial ou situacional, a princípio) e provê campos (presentes na Figura 25) para a inclusão de valores de referência.

Figura 26: Campos da dimensão "Avaliação"



A dimensão “Avaliação” (Figura 26) é a instância de preenchimento do questionário durante um atendimento. Dessa forma, ela é responsável por situar o resultado do exame dentro de um atendimento.

Figura 27: Campos da dimensão "Atendimento"



Por último, a dimensão “Atendimento”, na Figura 27, engloba toda a estadia do paciente no hospital e possui informações referentes ao desfecho do tratamento. É durante o atendimento que são feitas as avaliações responsáveis por gerar os dados para o fluxo de BI.

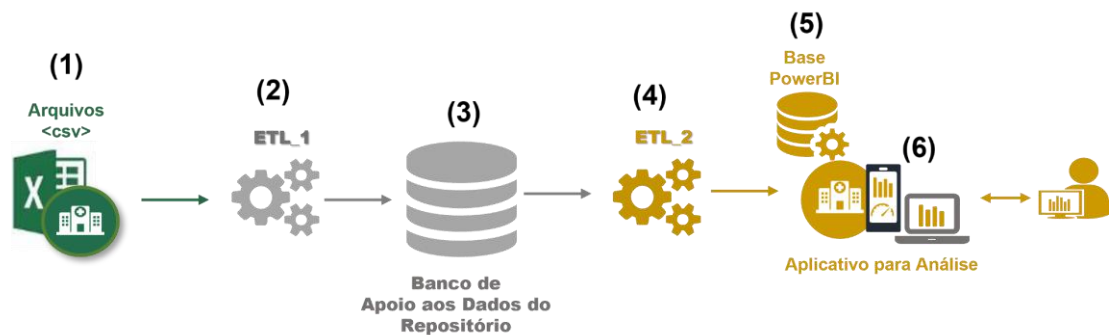
3.3 DESENVOLVIMENTO DO AMBIENTE ANALÍTICO PARA VISUALIZAÇÃO DE DADOS SOBRE COVID-19

Em virtude dos dados provenientes do formulário da OMS ainda não se encontrarem disponíveis no início deste projeto, este trabalho optou pelo emprego de dados clínicos disponibilizados em repositórios abertos. Deste modo, a arquitetura foi ajustada conforme apresentado na Figura 28. Nessa arquitetura, os dados (1) do HSL e do grupo Fleury são

coletados por meio de um processo ETL desenvolvido com a ferramenta PDI (2), responsável pela limpeza, tratamento e armazenamento em uma base de apoio (3). Os dados da base de apoio (3) são processados pelo ETL disponibilizado pela ferramenta Power BI (4). Esse processo (4) é responsável por extrair os dados da base de apoio, preparar métricas sobre eles e armazená-los na base do PowerBI (5). A partir dessa nova base (5) as visualizações são geradas (6).

Cabe ressaltar que não foi realizado um esforço com o objetivo de adequar os novos dados provenientes de repositórios à modelagem proposta anteriormente, baseada no formulário da OMS. Foram realizadas apenas alterações/ajustes pertinentes no sentido de permitir a construção das análises e visualizações pretendidas por este trabalho. Todas as análises e visualizações aqui tratadas, porém, podem ser construídas a partir da modelagem proposta, como veremos adiante.

Figura 28: Arquitetura ajustada aos dados do Repositório da FAPESP



3.3.1 O Planejamento

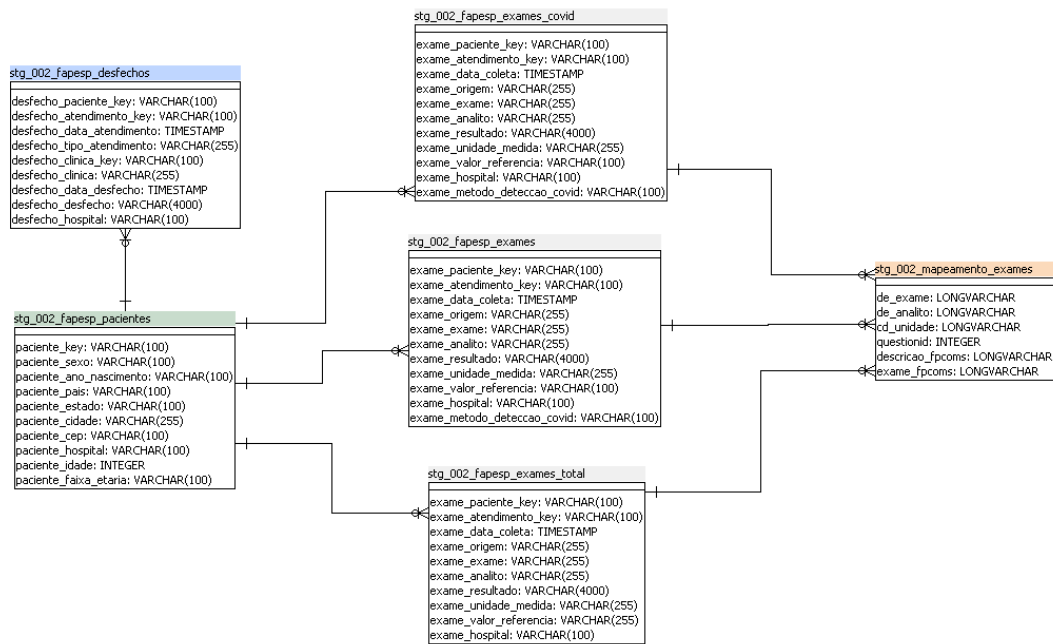
Para este entregável, a fase de planejamento exigiu a análise dos dados disponibilizados no Repositório da FAPESP e a avaliação quanto a melhor forma de empregá-los. Essa tarefa foi facilitada em virtude do dicionário de dados utilizado por todas as instituições que participam da iniciativa COVID-19 Data Sharing/BR.

Convém destacar que a FAPESP ficou conhecida como uma das principais agências de fomento à pesquisa científica e tecnológica do país a disponibilizar dados clínicos de pacientes com COVID-19. Em parceria com a Universidade de São Paulo (USP), ela implantou um repositório onde reuniu um conjunto de dados (*datasets*) de hospitais, inicialmente de São Paulo.

A análise do dicionário de dados permitiu identificar a existência de informações sobre: (i) pacientes como, por exemplo, sexo, estado de origem e idade; (ii) exames realizados durante os atendimentos desses pacientes; e (iii) desfechos dos atendimentos realizados. Este último apenas disponibilizado pelo HSL.

A partir desse dicionário, foi estabelecido o modelo apresentado na Figura 29. Esse modelo considera a estrutura dos dados dos hospitais envolvidos e adiciona outros atributos considerados importantes como, por exemplo, o nome do hospital e faixa etária para paciente. Com base no modelo, foram analisadas as métricas de interesse para a visualização de dados que permitiram identificar os conceitos relevantes a serem tratados.

Figura 29: Modelo lógico das Tabelas Empregadas no Ambiente Analítico



Com base no DM estabelecido na seção 3.2 e no modelo apresentado na Figura 29, foram estabelecidas as seguintes propostas para análise e visualização:

1 - Análise dos dados demográficos referentes aos pacientes atendidos nas instituições:

Esta análise considera o emprego das dimensões “Tempo”, “Paciente”, “Hospital” e “Exame”. Por meio dessas dimensões é possível estabelecer, para cada hospital, o total de pacientes que realizaram pelo menos um teste para COVID-19 que tenha sido positivado (reagente).

Para ampliar a análise, com o apoio da dimensão “Tempo”, realiza-se uma avaliação de pacientes atendidos mês a mês. Com o apoio da dimensão “Paciente”, essa análise pode abranger sexo, faixa etária e fases da vida.

Deste modo, as visualizações a serem geradas devem oferecer meios para identificar se, durante a evolução da doença, alguma faixa-etário ou sexo foi mais atingida do que outra.

2 - Análise das Internações Hospitalares

Esta análise tem por objetivo fornecer um instrumento analítico para os gestores de hospitais. Neste trabalho, essa análise se aplica apenas ao HSL, tendo em vista que o Grupo Fleury é uma instituição voltada a análises laboratoriais e de imagens, ou seja, exames.

Para essa análise, consideramos as dimensões “Tempo”, “Paciente”, “Hospital”, “Atendimento”, “Cuidado”, “Tipo de Atendimento”. Por meio dessas informações é possível analisar o fluxo de entrada e saída de pacientes para cada hospital, com base nas datas de admissão e de desfecho. Outra análise importante, refere-se à necessidade de Cuidados para paciente internados como, por exemplo, internação em UTI. Essa informação é registrada no FRC-OMS e foi disponibilizada como “Origem” de exames dos pacientes do HSL. Dessa forma é possível avaliar, mês a mês, do total de pacientes atendidos, quantos tiveram necessidade de UTI.

Ainda quanto aos atendimentos hospitalares, podemos considerar com base nas dimensões estabelecidas, análises sobre o tipo de atendimento procurado pelo paciente e os desfechos desses atendimentos. Este último, considerando como o paciente deixou o hospital, ou seja, se foi por Alta, Alta Administrativa, Óbito ou Transferência para outro hospital.

3 - Análise Comparativa entre Instituições:

Para essa análise, consideramos a possibilidade de trabalhar com as dimensões “Tempo”, “Hospital”, “Pacientes”, “Exames” de forma a gerar comparações entre duas instituições. Por exemplo, gerar uma comparação, mês a mês, do total de pacientes atendidos, permitindo filtros em ambos por sexo e fases da vida.

4 - Análise de Tipos de Exames COVID-19

Essa análise considera as mesmas dimensões da análise anterior, com ênfase para os testes de COVID-19, considerando os diferentes métodos de detecção empregados. Esse tipo de análise pode auxiliar os gestores quanto a aquisição de kits de exames, priorizando aqueles mais procurados ou mais aplicados pelos profissionais de saúde.

5 - Análise de Exames estabelecidos pela OMS

Para esta análise, será necessário utilizar as dimensões “Tempo”, “Pacientes”, “Exames” e a partir daí entender como foi a evolução da realização de cada um desses exames por pacientes distintos, dimensionando o seu tamanho pelo total de exames realizados.

A próxima parte do trabalho consistiu em reunir os dados necessários para serem utilizados nessa estrutura.

3.3.2 A Fase de Execução

Conforme abordado na seção 2.3.2, a fase de execução consiste na utilização dos dados disponibilizados publicamente no repositório da FAPESP para gerar visualizações que servirão como um primeiro modelo a ser apresentado aos hospitais parceiros do Projeto VODAN BR.

Ao mesmo tempo, para aumentar o universo de análises e entender a expectativa do usuário final, decidimos criar um formulário no Google Forms⁵ para ser preenchido pelos funcionários dos hospitais envolvidos e, assim, exibir para eles a riqueza de informações passíveis de serem extraídas dos dados e identificar de forma mais clara e objetiva quais eram suas principais necessidades.

Entretanto, devido ao momento do projeto e ao contato dificultado com os pontos focais dos hospitais, fizemos a opção de seguir em frente com o projeto utilizando as análises inicialmente pensadas. Foram, então, instalados e configurados em ambiente local (*on-premises*) os *softwares* necessários para prosseguimento do desenvolvimento.

Após serem extraídas as bases de dados disponibilizadas, foi escolhido um hospital que apresentasse de forma mais completa os dados necessários para a construção do painel proposto. O HSL foi a instituição que disponibilizou os dados de forma mais completa, ou seja, contemplando todos os itens do dicionário de dados. As informações sobre o desfecho dos pacientes ampliam a gama de análises a serem feitas a partir desses dados. Além do HSL, também foram tratados os dados do Grupo Fleury, proporcionando uma maior massa de dados de exames para análises.

De posse dos dados, foi necessário um tratamento, visando a padronização de informações como, por exemplo, nome de exames. Esse tratamento é essencial para facilitar as

⁵https://github.com/mborges-comp-ufrrj/TCC_Mauricio_Rodrigo/blob/main/Formul%C3%A1rio%20de%20Relev%C3%A2ncia%20de%20An%C3%A1lises%20-%20Formul%C3%A1rios%20Google.pdf

consultas e visualização dos dados. Para esta atividade empregamos o PDI. Por meio dessa ferramenta foi realizado o processo de ETL sobre os dados.

3.3.2.1 O processo de ETL

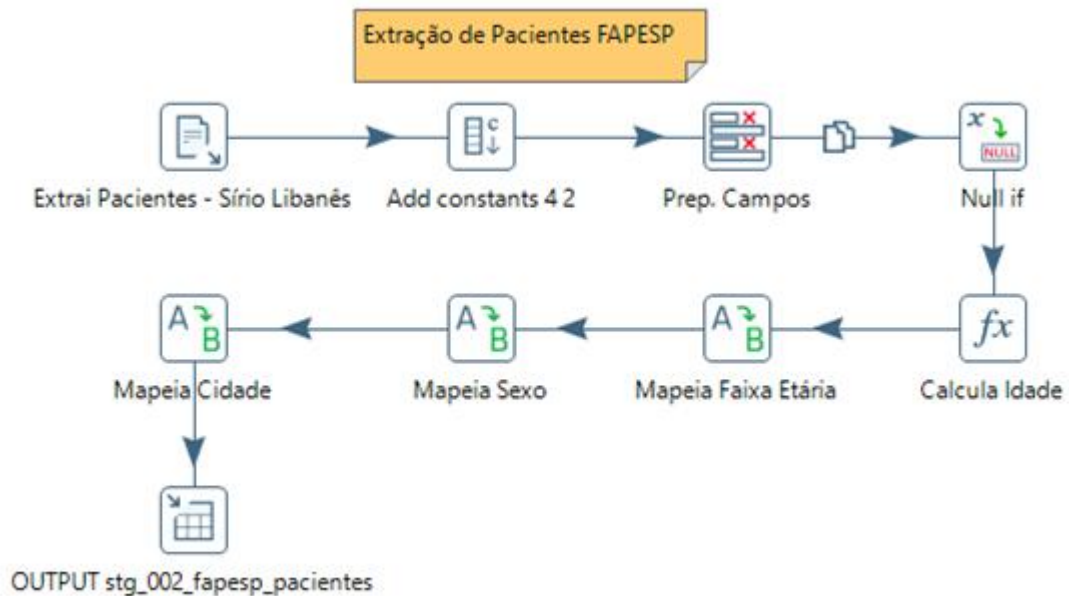
Desde o início do projeto, houveram diversas tentativas de implementação de fluxos de ETL, inclusive com análises razoavelmente complexas. Assim sendo, nossa opção foi por viabilizar a limpeza dos dados, filtrando-os conforme os resultados de exames demonstram a incidência do coronavírus naqueles pacientes, e mapear novos campos a partir das informações obtidas nos arquivos. Neste processo, foram tratados os dados do HSL e do Grupo Fleury. Para apresentar os passos (*steps*) executados na ferramenta PDI, selecionamos aqueles referentes ao tratamento dos dados do HSL. Esse tratamento foi o selecionado por apresentar um fluxo mais completo, compreendendo ações para dados demográficos, exames e desfechos dos atendimentos. Passos similares foram executados para o tratamento dos dados do Grupo Fleury. Entretanto, convém destacar o tratamento adicional para a chave de atendimento. Os arquivos do Fleury não apresentam a informação referente ao número do atendimento, sendo assim tornou-se necessária a criação de um identificador a partir da concatenação dos campos “paciente_key” e “exame_data_coleta”. Esse tratamento tornou possível a padronização das informações para os dois hospitais.

a) Obtenção das bases

As bases utilizadas para o trabalho de dados deste projeto foram obtidas a partir de repositório aberto mantido pela FAPESP, em parceria com diversas outras instituições, com dados clínicos e laboratoriais obtidos de atendimentos a pacientes reais (cuja confidencialidade foi mantida). Foram utilizados, mais especificamente, os dados publicados pelo HSL e pelo Grupo Fleury no repositório.

b) Extração dos dados de pacientes

Figura 30: Fluxo de processamento ETL para os dados de pacientes



Inicialmente foram tratados os dados de pacientes. O fluxo apresentado, vide na Figura 30, destaca os “steps” para a extração dos dados do arquivo CSV HSL_Pacientes_2.csv do HSL. Outro fluxo similar realiza o mesmo processo para os dados de paciente do grupo Fleury.

Figura 31: Configuração dos passos “Extrai Pacientes (...)”

A captura de tela mostra a configuração do passo "Extrai Pacientes - Sírio Libanês" em um software de ETL. O formulário contém os seguintes campos e opções:

- Step name:** Extrai Pacientes - Sírio Libanês
- Filename:** C:\Users\mauri\Google Drive\Faculdade\TCC\08.Bases\FAPESP_SirioLibar
- Delimiter:** |
- Enclosure:** -
- NIO buffer size:** 50000
- Lazy conversion?**
- Header row present?**
- Add filename to result:**
- The row number field name (optional):** (campo vazio)
- Running in parallel?**
- New line possible in fields?**
- Format:** mixed
- File encoding:** UTF-8

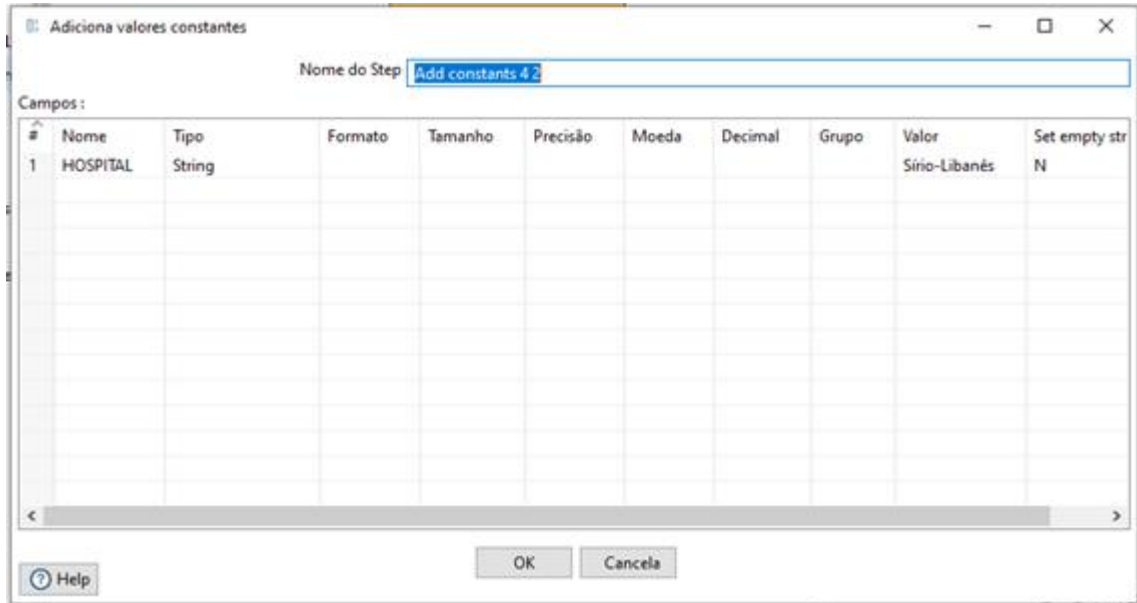
Abaixo do formulário, há uma tabela de mapeamento de campos:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim
1	ID_PACIENTE	String		100		RS	,	.	nenf
2	IC_SEXO	String		10		RS	,	.	nenf
3	AA_NASCIMENTO	String		10		RS	,	.	nenf
4	CD_PAIS	String		10		RS	,	.	nenf
5	CD_UF	String		10		RS	,	.	nenf
6	CD_MUNICIPIO	String		100		RS	,	.	nenf
7	CD_CEPREDUZIDO	String		10		RS	,	.	nenf

Na base da janela, há botões para: Help, OK, Obtem campos, Preview e Cancela.

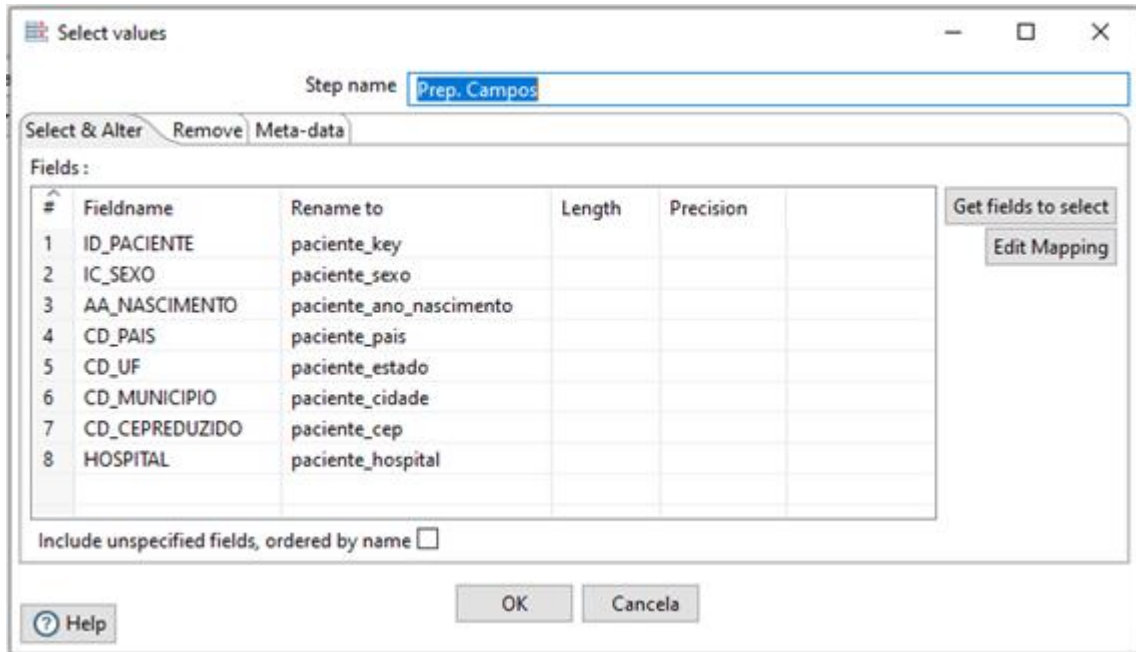
Nesta extração, conforme apresentado na Figura 31, foram obtidos dados de sexo, ano de nascimento, país, unidade federal e município dos pacientes do HSL. Apesar do processo coletar informações de CEP, estas foram irrelevantes, pois o campo “CD_CEPREDUZIDO” apresentava “null” para a maioria dos pacientes.

Figura 32: Configuração do passo “Add constants 4 2”



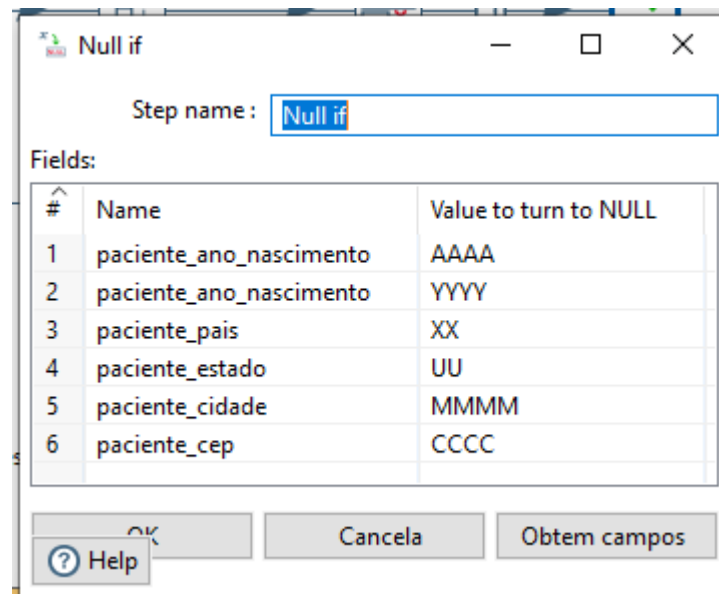
Ainda nesse processo, uma constante com o nome do hospital foi adicionada, conforme exposto na Figura 32. Essa informação permite separar os dados, possibilitando classificá-los, filtrá-los e até mesmo criar visualizações específicas para um dado hospital.

Figura 33: Configuração do passo “Prep. Campos”



Para melhor organização e entendimento, as colunas foram renomeadas destacando-se o nome da entidade/arquivo origem. Na Figura 33 são apresentadas as colunas com seus nomes originais e a forma como elas foram renomeadas, considerando que foram coletadas a partir do arquivo de pacientes.

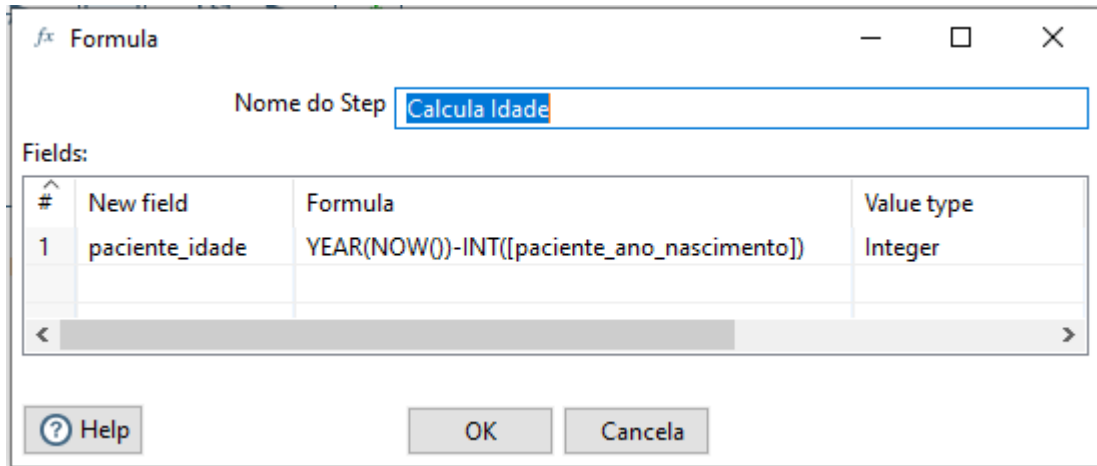
Figura 34: Configuração do passo “Null if”



Nessa etapa foi necessário o tratamento de valores de dados, conforme o próprio dicionário de dados do HSL já mencionava. Em especial, destacamos os pacientes cujo valor

de “paciente_ano_nascimento” é igual a “AAAA”, que deveriam ser associados à idade de 92 anos ou mais de acordo com as referências no dicionário de dados do HSL (foi feito dessa maneira para anonimizar os pacientes, dado que é uma faixa de idade a partir da qual o número de pacientes diminui drasticamente) e o mesmo foi feito dentro do ETL criado.

Figura 35: Configuração do passo “Calcula Idade”



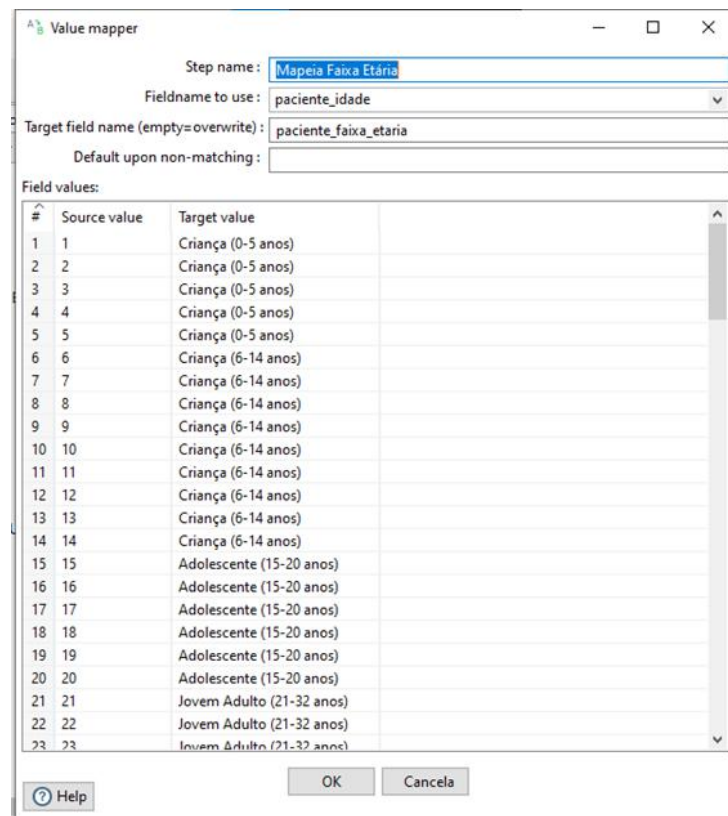
Nome do Step:

Fields:

#	New field	Formula	Value type
1	paciente_idade	YEAR(NOW())-INT([paciente_ano_nascimento])	Integer

Buttons:

Figura 36: Configuração do passo “Mapeia Faixa Etária”



Step name:

Fieldname to use:

Target field name (empty=overwrite):

Default upon non-matching:

Field values:

#	Source value	Target value
1	1	Criança (0-5 anos)
2	2	Criança (0-5 anos)
3	3	Criança (0-5 anos)
4	4	Criança (0-5 anos)
5	5	Criança (0-5 anos)
6	6	Criança (6-14 anos)
7	7	Criança (6-14 anos)
8	8	Criança (6-14 anos)
9	9	Criança (6-14 anos)
10	10	Criança (6-14 anos)
11	11	Criança (6-14 anos)
12	12	Criança (6-14 anos)
13	13	Criança (6-14 anos)
14	14	Criança (6-14 anos)
15	15	Adolescente (15-20 anos)
16	16	Adolescente (15-20 anos)
17	17	Adolescente (15-20 anos)
18	18	Adolescente (15-20 anos)
19	19	Adolescente (15-20 anos)
20	20	Adolescente (15-20 anos)
21	21	Jovem Adulto (21-32 anos)
22	22	Jovem Adulto (21-32 anos)
23	23	Jovem Adulto (21-32 anos)

Buttons:

A figura 35 apresenta o *step* para o cálculo da idade do paciente a partir do ano de nascimento. Como os dados foram obtidos em agosto/2020, o cálculo é realizado considerando

o ano de 2020. Essa informação é importante dado que o objetivo é gerar um retrato da situação no período específico da disponibilização dos dados.

Com base na idade foi realizado o mapeamento da faixa etária. As faixas foram estabelecidas em reunião de projeto e encontram-se listadas na Figura 36.

Figura 37: Configuração do passo “Mapeia Sexo”

Value mapper

Step name:

Fieldname to use:

Target field name (empty=overwrite):

Default upon non-matching:

Field values:

#	Source value	Target value
1	M	Masculino
2	F	Feminino

Figura 38: Configuração do passo “Mapeia Cidade”

Value mapper

Step name:

Fieldname to use:

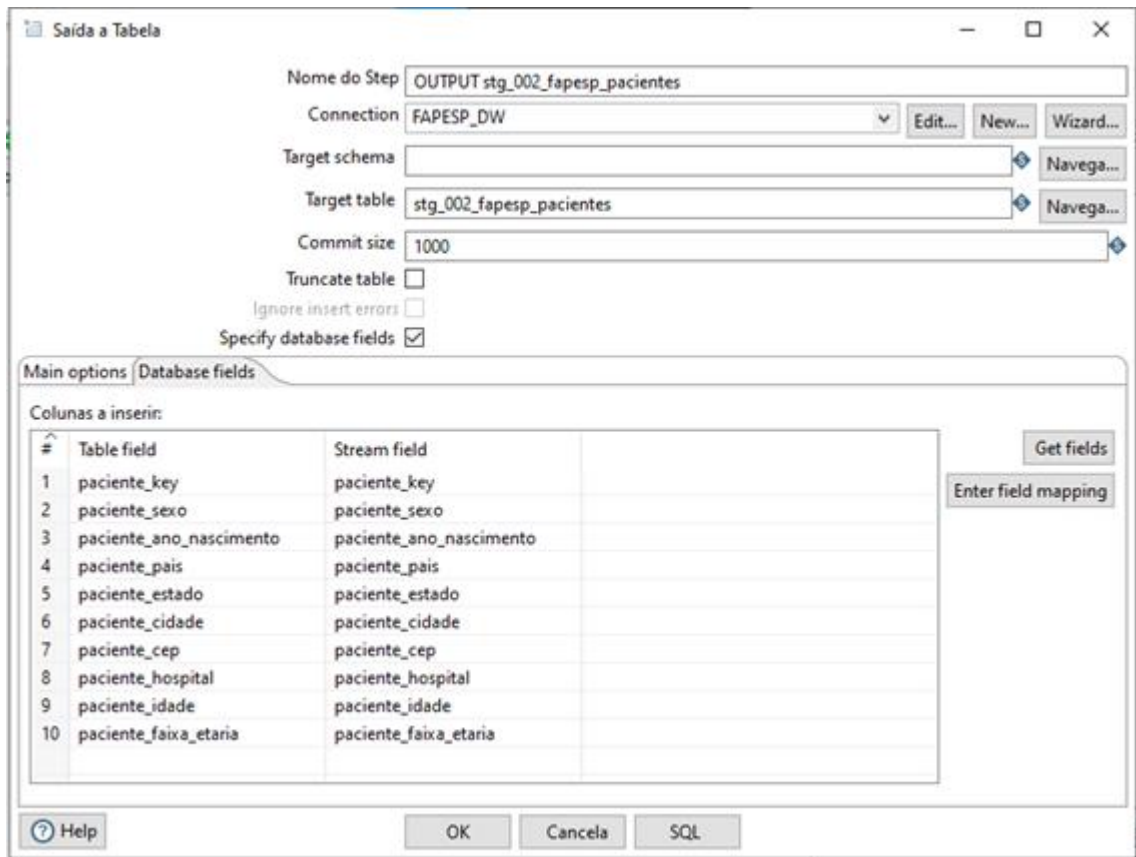
Target field name (empty=overwrite):

Default upon non-matching:

Field values:

#	Source value	Target value
1	BRASILIA	Brasília
2	CAMPOS DO JORDÃO	Campos do Jordão
3	GUARULHOS	Guarulhos
4	SAO PAULO	São Paulo

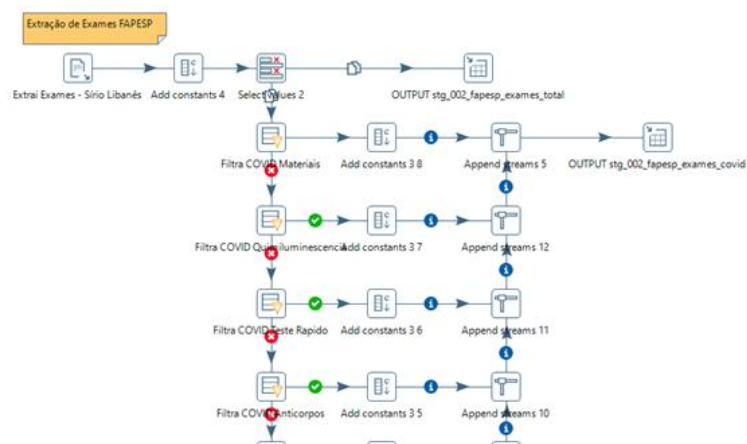
Figura 39: Configuração do passo “OUTPUT stg_002_fapesp_pacientes”



De forma similar, foi efetuado o mapeamento de sexo, apresentado na Figura 37, e o de cidade, representado na Figura 38. Realizadas as transformações necessárias, os dados tratados foram inseridos na tabela stg_002_fapesp_pacientes, utilizando o passo apresentado na Figura 40.

c) Extração dos dados de exames

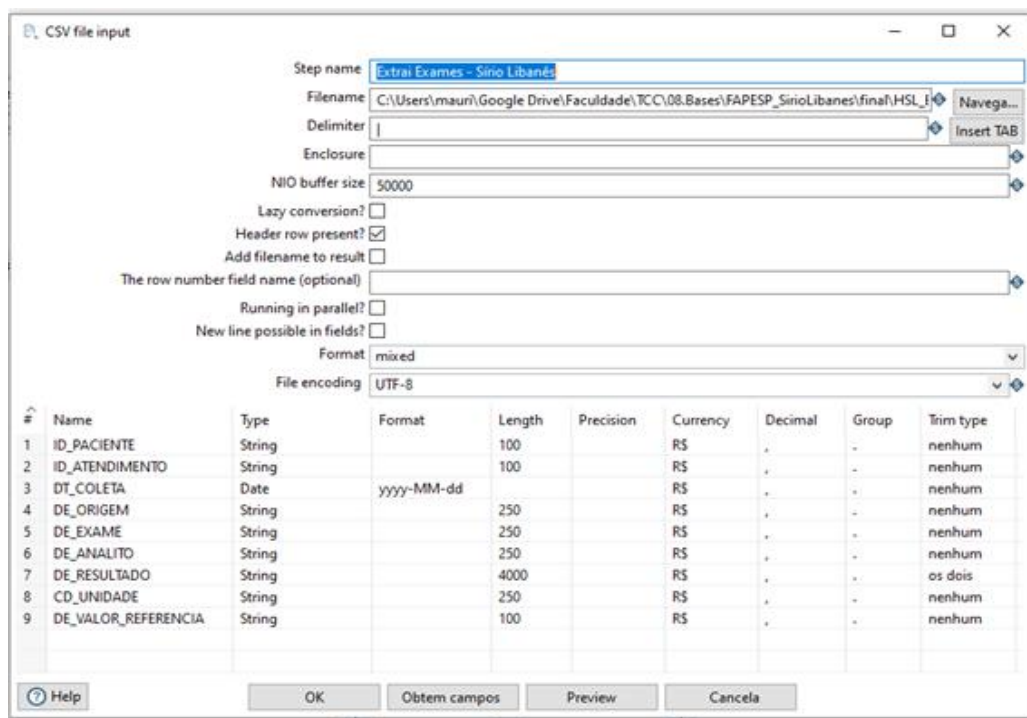
Figura 40: Fluxo de processamento ETL para os dados de exames



O processo de extração e tratamento dos dados de exames está representado na Figura 40. Em virtude do volume de dados, esse processo se apresentou como um dos maiores desafios deste trabalho.

O tratamento dos exames ocorre em dois momentos. Primeiramente é realizada a extração dos dados dos exames dos arquivos origem, com um tratamento de nomenclatura, e carga para a tabela “OUTPUT stg_002_fapesp_exames_total”. Em seguida são mapeados os exames de detecção de coronavírus. Esses exames, por sua vez, são rotulados conforme o método de detecção. Os *steps* para a realização desses procedimentos estão descritos a seguir.

Figura 41: Configuração dos passos “Extrai Exames (...)”



A Figura 41 apresenta o passo responsável pela extração do arquivo HSL_Exames_2.csv que contempla os exames do HSL. A partir desse arquivo se obtém informações sobre os exames realizados nos pacientes como data de coleta, proveniência do dado (laboratório, clínica, etc.), nome do exame, analito analisado, resultado medido, unidade de medida e valor de referência para o resultado.

Figura 42: Configuração do passo “Add constants 4”

Nome do Step:

Campos:

#	Nome	Tipo	Formato	Tamanho	Precisão	Moeda	Decimal	Grupo	Valor	Set empty str
1	HOSPITAL	String							Sírio-Libanês	<input checked="" type="checkbox"/>

Buttons:

Figura 43: Configuração do passo “Select values 2”

Step name:

Select & Alter Remove Meta-data

Fields:

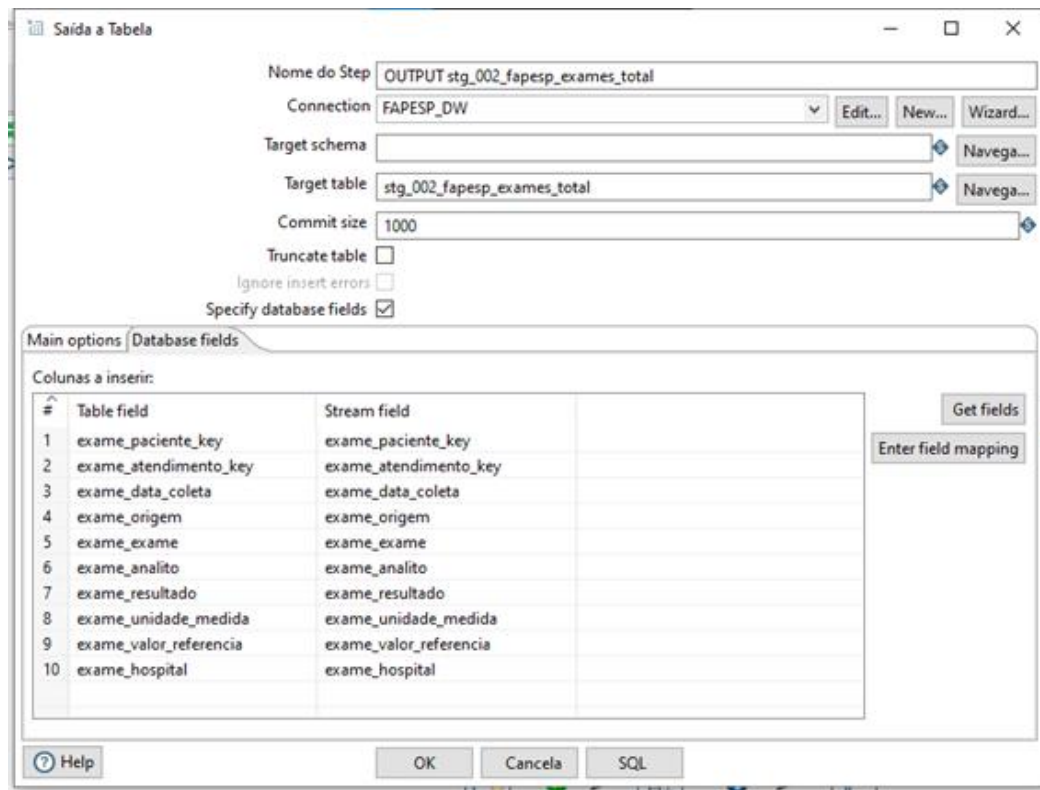
#	Fieldname	Rename to	Length	Precision
1	ID_PACIENTE	exame_paciente_key		
2	ID_ATENDIMENTO	exame_atendimento_key		
3	DT_COLETA	exame_data_coleta		
4	DE_ORIGEM	exame_origem		
5	DE_EXAME	exame_exame		
6	DE_ANALITO	exame_analito		
7	DE_RESULTADO	exame_resultado		
8	CD_UNIDADE	exame_unidade_medida		
9	DE_VALOR_REFERENCIA	exame_valor_referencia		
10	HOSPITAL	exame_hospital		

Buttons:

Include unspecified fields, ordered by name

Buttons:

Figura 44: Configuração do passo “OUTPUT stg_002_fapesp_exames_total”



A Figura 42 apresenta o passo responsável por adicionar o nome do hospital aos dados, como realizado anteriormente para os dados de paciente. Ainda de forma similar ao tratamento dos dados de paciente, após incluir a coluna com o nome do hospital, os campos são renomeados como descrito na Figura 43, facilitando a identificação da origem e o entendimento de seu conteúdo.

Neste ponto o fluxo do processamento é dividido. A Figura 44 apresenta o processo referente ao fluxo responsável por efetuar a inclusão de todos os exames na base de dados. Essa carga é realizada na tabela “OUTPUT stg_002_fapesp_exames_total”.

O segundo fluxo contempla os passos responsáveis por filtrar os exames conforme os resultados obtidos para o teste de detecção de coronavírus. De acordo com análise realizada previamente sobre os dados, vinte e um (21) testes diferentes para COVID-19 foram identificados. Para cada um foi especificado um filtro e a restrição imposta para a seleção de exames. A Figura 40 ilustra como um filtro foi estabelecido para cada um dos casos identificados.

Figura 45: Configuração dos passos “Filtra COVID-19 (...)”

Filter rows

Step name:

Send 'true' data to step:

Send 'false' data to step:

The condition:

```

exame_exame = [COVID-19-PCR para SARS-COV-2, Vários Materiais (Fleury)]
AND
exame_analito = [Coronavirus (2019-nCoV)]
AND
(
  OR
  exame_resultado = [DETECTADO (POSITIVO)]
  OR
  exame_resultado = [DETECTADO]
  OR
  exame_resultado = [DETECTÁVEL]
)

```

A Figura 46 ilustra a configuração do passo que associa a respectiva restrição de acordo com o filtro estabelecido, conforme a Figura 45.

Figura 46: Configuração dos passos “Add constants (...)” referentes aos filtros

Adiciona valores constantes

Nome do Step:

Campos:

#	Nome	Tipo	Formato	Tamanho	Precisão	Moeda	Decimal	Grupo	Valor	Set empty string?
1	exame_metodo_deteccao_covid	String							Material Coletado	N

Cabe ressaltar que a análise realizada considerou todas as possibilidades de resultado positivo para o coronavírus, capazes de serem catalogadas. Esses itens foram mapeados entre “Material Coletado”, “Anticorpos” ou “PCR”, de acordo com o método de detecção. A saber, os filtros executados foram:

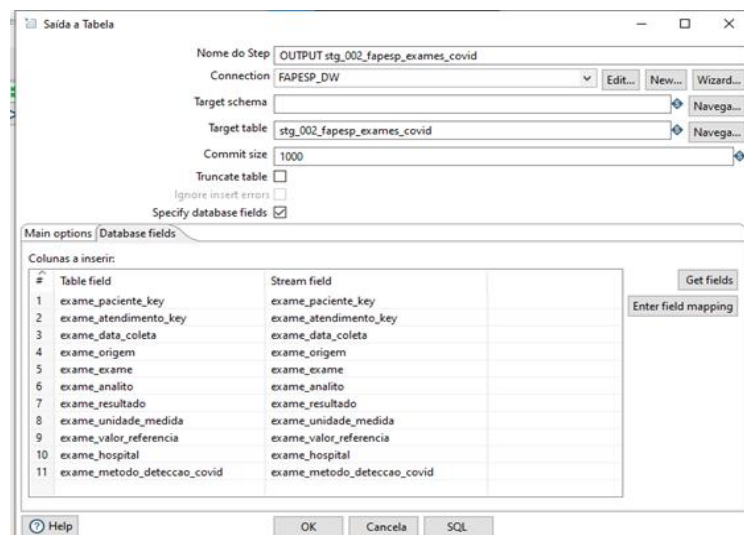
- Filtro 1 (PCR):
 - Exame: “COVID-19-PCR para SARS-COV-2, Vários Materiais (Fleury)”;
 - Analito: “Coronavírus (2019-nCoV)”;
 - Resultado: “DETECTADO (POSITIVO)”, “DETECTADO” ou “DETECTÁVEL”.
- Filtro 2 (Anticorpos):
 - Exame: “COVID-19-Sorologia IgM e IgG por quimiluminescência, soro”;

- Analito: “Covid 19, Anticorpos Totais, Eletroquimiolumi”, “Covid 19, Anticorpos IgM, Quimioluminescência” ou “Covid 19, Anticorpos IgG, Quimioluminescência”;
- Resultado: “REAGENTE”.
- Filtro 3 (Anticorpos):
 - Exame: “COVID-19-Teste Rápido (IgM e IgG), soro”;
 - Analito: “Covid 19, Anticorpos IgG, teste rápido” ou “Covid 19, Anticorpos IgM, teste rápido”;
 - Resultado: “REAGENTE”.
- Filtro 4 (Anticorpos):
 - Exame: “COVID-19, anticorpos IGA e IGG, soro”;
 - Analito: “Covid 19, Anticorpos IgA, Interpretação” ou “Covid 19, Anticorpos IgG, Interpretação”;
 - Resultado: “REAGENTE”.
- Filtro 5 (PCR):
 - Exame: “Detecção de Coronavírus (NCoV-2019) POR PCR (Anatomia Patológica)”;
 - Analito: “Detecção de Coronavírus (NCoV-2019) POR PCR (Anatomia Patológica)”;
 - Resultado: “DETECTADO”.
- Filtro 6 (PCR):
 - Exame: “Sars Cov-2, Teste Molecular Rápido Para Detecção, Vários Materiais”;
 - Analito: “Covid 19, Detecção por PCR”;
 - Resultado: “DETECTADO (POSITIVO)”.
- Filtro 7 (Anticorpos):
 - Exame: “Sorologia - Coronavírus, IgA”;
 - Analito: “Covid 19, Anticorpos IgA”;
 - Resultado: “REAGENTE”.
- Filtro 8 (Anticorpos):
 - Exame: “Sorologia - Coronavírus, IgG”;
 - Analito: “Covid 19, Anticorpos IgG”;
 - Resultado: “REAGENTE”.
- Filtro 9 (Anticorpos):

- Exame: “Anticorpos IgA e IgG SARS-CoV-2/COVID19”;
- Analito: “COVID IgA Intep” ou “COVID IgG Interp”;
- Resultado: “Reagente”.
- Filtro 10 (Anticorpos):
 - Exame: “COVID19, ANTICORPOS IgA, soro”;
 - Analito: “Covid 19, Anticorpos IgA, Elisa”;
 - Resultado: “REAGENTE”.
- Filtro 11 (Anticorpos):
 - Exame: “COVID19, ANTICORPOS IgG, soro”;
 - Analito: “Covid 19, Anticorpos IgG, Elisa”;
 - Resultado: “REAGENTE”.
- Filtro 12 (Anticorpos):
 - Exame: “COVID19, ANTICORPOS IgM, soro”;
 - Analito: “Covid 19, Anticorpos IgM, Quimioluminescência”;
 - Resultado: “REAGENTE”.
- Filtro 13 (Anticorpos):
 - Exame: “COVID19, ANTICORPOS TOTAIS, soro”;
 - Analito: “Covid 19, Anticorpos Totais, Eletroquimiolumi”;
 - Resultado: “REAGENTE”.
- Filtro 14 (Anticorpos):
 - Exame: “Empresarial: Anticorpos Totais IgG+IgM”;
 - Analito: “COVID19 IgG, teste rápido”, “COVID19 IgM, teste rápido” ou “COVID19, IgG + IgM”;
 - Resultado: “Reagente” ou “Reagente Fraco”.
- Filtro 15 (PCR):
 - Exame: “NOVO CORONAVÍRUS 2019 (SARS-CoV-2), DETECÇÃO POR PCR”;
 - Analito: “Covid 19, Detecção por PCR”;
 - Resultado: “DETECTADO” ou “DETECTADO (POSITIVO)”.
- Filtro 16 (PCR):
 - Exame: “PCR em tempo real para detecção de Coron”;
 - Analito: “Resultado COVID-19:”;
 - Resultado: “Detectado”.

- Filtro 17 (Anticorpos):
 - Exame: “SARS-COV-2, ANTICORPOS IgG, soro”;
 - Analito: “Covid 19, Anticorpos IgA, Elisa”;
 - Resultado: “REAGENTE”.
- Filtro 18 (Anticorpos):
 - Exame: “SARS-CoV-2, ANTICORPOS IgM E IgG, TESTE RÁPIDO”;
 - Analito: “Covid 19, Anticorpos IgG, teste rápido” ou “Covid 19, Anticorpos IgM, teste rápido”;
 - Resultado: “RESULTADO”.
- Filtro 19 (Anticorpos):
 - Exame: “Sorologia SARS-CoV-2/COVID19 IgG/IgM”;
 - Analito: “COVID IgG Interp” ou “COVID IgM Interp”;
 - Resultado: “Ragente”, “Reageente”, “Reagente” ou “Reagente Fraco”.
- Filtro 20 (PCR):
 - Exame: “TESTE MOLECULAR RÁPIDO PARA DETECÇÃO DE SARS-CoV-2 (COVID-19)”;
 - Analito: “Covid 19, Detecção por PCR”;
 - Resultado: “DETECTADO” ou “DETECTADO (POSITIVO)”.
- Filtro 21 (Anticorpos):
 - Exame: “Teste rápido Coronavirus COVID19 IgG/IgM”;
 - Analito: “COVID19 IgG, teste rápido” ou “COVID19 IgM, teste rápido”;
 - Resultado: “Reagente” ou “Reagente Fraco”.

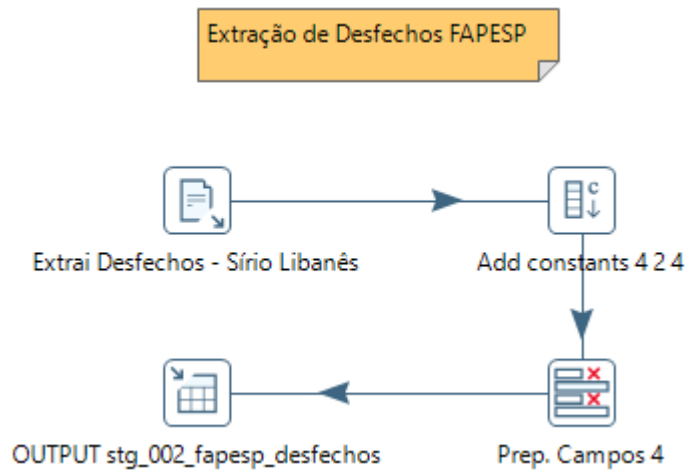
Figura 47: Configuração do passo “OUTPUT stg_002_fapesp_exames_covid”



Os dados obtidos a partir da aplicação dos filtros foram inseridos na tabela stg_002_fapesp_exames_covid, conforme apresentado na Figura 47.

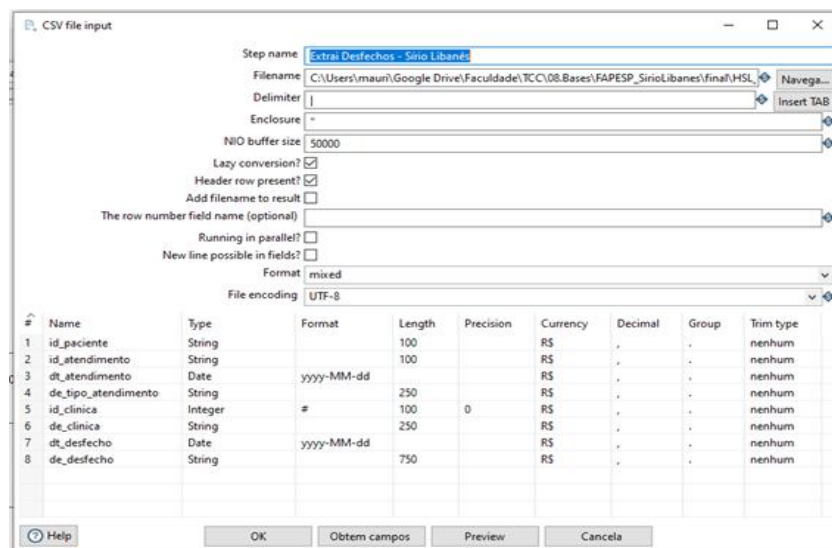
d) Extração dos dados de desfechos

Figura 48: Fluxo de processamento ETL para os dados de Desfecho



A extração e tratamento dos dados de desfechos acontecem exclusivamente para os arquivos gerados pelo HSL, única instituição que os disponibilizou no catálogo da FAPESP. O fluxo para obtenção destes dados encontra-se representado na Figura 48.

Figura 49: Configuração dos passos “Extrai Desfechos (...)”



Os dados são obtidos a partir do arquivo HSL_Desfechos_2.csv, conforme representado na Figura 49. Esses dados referem-se ao desfecho do atendimento realizado para o paciente e

contemplam data de atendimento, tipo de atendimento, ou seja, se foi um “Pronto Atendimento”, “Ambulatorial”, etc., clínica responsável pelo atendimento, data de desfecho e tipo de desfecho. Este último identificando se a conclusão do atendimento foi por óbito, alta ou outro caso.

Dessa forma similar aos processamentos de dados anteriores, a Figura 50 apresenta o passo responsável por inserir a coluna para distinguir o hospital.

Figura 50: Configuração do passo “Add constants 4 2 4”



Figura 51: Configuração do passo “Prep. Campos 4”

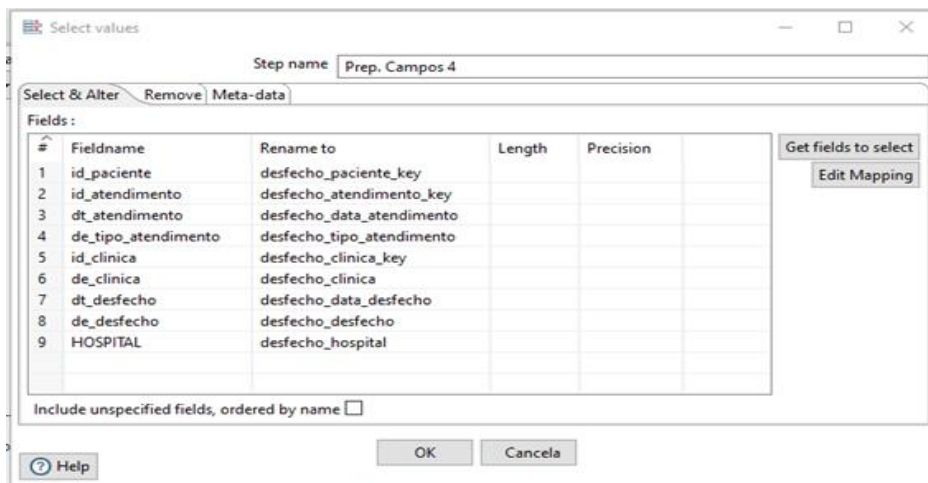
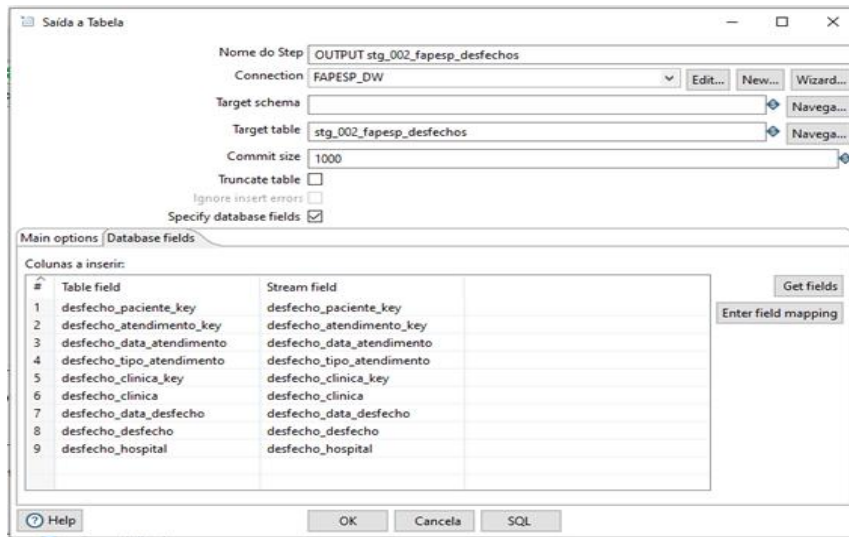
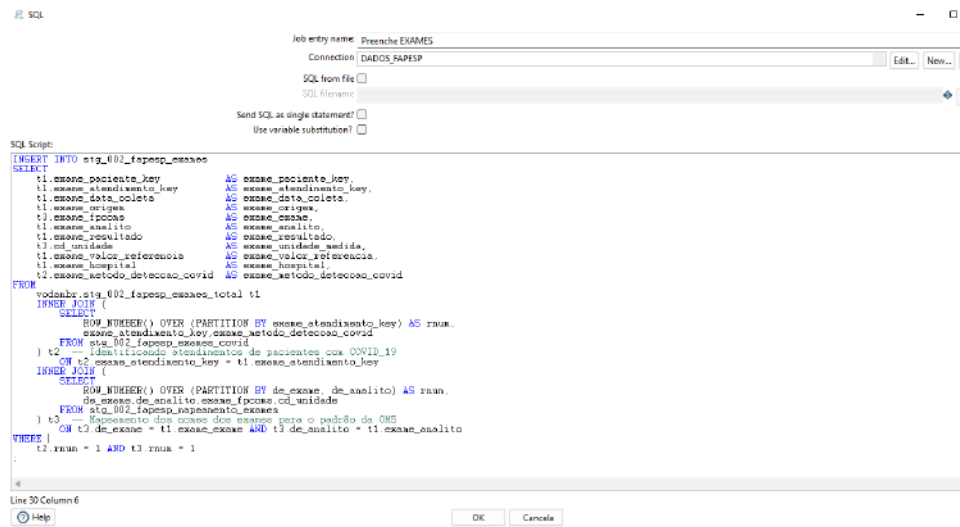


Figura 52: Configuração do passo “OUTPUT stg_002_fapesp_desfechos”



A Figura 51 apresenta a renomeação dos campos, associando-os ao "desfecho". Após esses tratamentos, os dados foram inseridos na tabela stg_002_fapesp_desfechos, como apresentado na Figura 52.

Figura 53: Configuração do passo “Preenche EXAMES”



Finalizada a carga dos dados do HSL e do Grupo Fleury, foi executada uma consulta com o SQL em destaque na Figura 53 e exibido no passo apresentado na Figura 54. O resultado dessa consulta é carregado para a tabela stg_002_fapesp_examess e contempla todos os exames realizados para os pacientes diagnosticados com COVID-19.

Figura 54: Destaque do SQL Executado

```

INSERT INTO stg_002_fapesp_examess
SELECT
t1.exame_paciente_key AS exame_paciente_key,
t1.exame_atendimento_key AS exame_atendimento_key,
t1.exame_data_coleta AS exame_data_coleta,
t1.exame_origem AS exame_origem,
t3.exame_fpcoms AS exame_exame,
t1.exame_analito AS exame_analito,
t1.exame_resultado AS exame_resultado,
t3.cd_unidade AS exame_unidade_medida,
t1.exame_valor_referencia AS exame_valor_referencia,
t1.exame_hospital AS exame_hospital,
t2.exame_metodo_deteccao_covid AS exame_metodo_deteccao_covid
FROM
vodaubr.stg_002_fapesp_examess_total t1
INNER JOIN (
SELECT
ROW_NUMBER() OVER (PARTITION BY exame_atendimento_key) AS rnum,
exame_atendimento_key, exame_metodo_deteccao_covid
FROM stg_002_fapesp_examess_covid
) t2 -- Identificando atendimentos de pacientes com COVID-19
ON t2.exame_atendimento_key = t1.exame_atendimento_key
INNER JOIN (
SELECT
ROW_NUMBER() OVER (PARTITION BY de_exame, de_analito) AS rnum,
de_exame, de_analito, exame_fpcoms, cd_unidade
FROM stg_002_fapesp_mapeamento_examess
) t3 -- Mapeamento dos nomes dos exames para o padrão da OMS
ON t3.de_exame = t1.exame_exame AND t3.de_analito = t1.exame_analito
WHERE |
t2.rnum = 1 AND t3.rnum = 1
;

```

3.3.2.2 Desenvolvimento dos dashboards

O tratamento dos dados e a carga foram responsáveis pela geração das tabelas apresentadas na Tabela 1. A partir dessas tabelas foram elaboradas as visualizações dos dados para as instituições parceiras, visando atender aos objetivos analíticos estabelecidos na fase de planejamento.

Tabela 1: Relação das Tabelas e total de registros tratados por Instituição

Tabelas	Sírio Libanês	Fleury
stg_002_fapesp_pacientes	4.273	247.981
stg_002_fapesp_desfechos	16.957	0
stg_002_fapesp_examess	87.116	42.392
stg_002_fapesp_examess_total	631.352	5.873.403
stg_002_fapesp_examess_covid	1.650	39.072

Como apresentado anteriormente, a ferramenta selecionada para o desenvolvimento dos dashboards foi o Power BI. Ele é conhecido por ser um conjunto de ferramentas analíticas de negócios e visualização de dados. O primeiro painel desenvolvido se baseou em algumas análises mais gerais focadas em determinar métricas mais qualitativas. Trabalhamos em duas análises relevantes no nosso primeiro painel: a primeira sobre os exames realizados por cada paciente e a verificação se ele estava acima ou abaixo do limite normal ou aceitável para aquele

determinado exame; já a segunda análise buscou identificar a quantidade de pessoas por faixa etária.

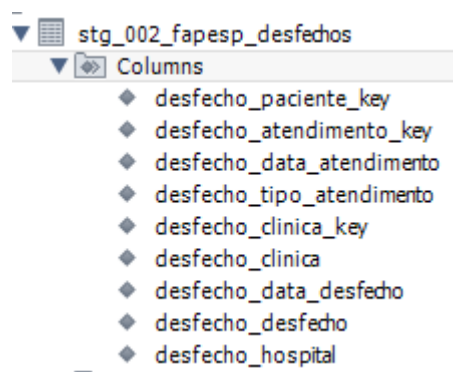
Em nossas reuniões, após alguns *feedbacks* sobre a apresentação de análises e visuais (que poderia ser melhorada), mudamos a estrutura do painel e filtramos os casos da base de dados de acordo com o resultado dos exames relativos à presença do novo coronavírus ou de seus anticorpos no paciente. Para realizar esse filtro, foi necessário olhar para a tabela de exames e identificar todos os pacientes que realizaram algum dos exames padrões de COVID-19 e todos os seus resultados possíveis. Esse passo do tratamento foi exposto na seção 3.3.2. Essa foi uma transformação custosa analiticamente e exigiu um grande esforço de mapeamento.

O grande desafio da construção da visualização de dados esteve em definir a melhor estratégia para “contar a história” dos dados para o público-alvo e, dessa forma, agregar valor ao projeto como um todo. Sendo assim, foram escolhidos três pontos-base para promover uma visualização de dados que apresentasse essa história de forma clara e objetiva:

1. Conhecer o perfil demográfico (sexo, fase da vida e faixa etária) dos pacientes ao longo do tempo proposto;
2. Descobrir os exames para detecção da COVID-19 solicitados/realizados de acordo com o setor de origem do atendimento do paciente;
3. Compreender a dinâmica de entrada e saída de pacientes e o volume mensal de exames realizados por hospital.

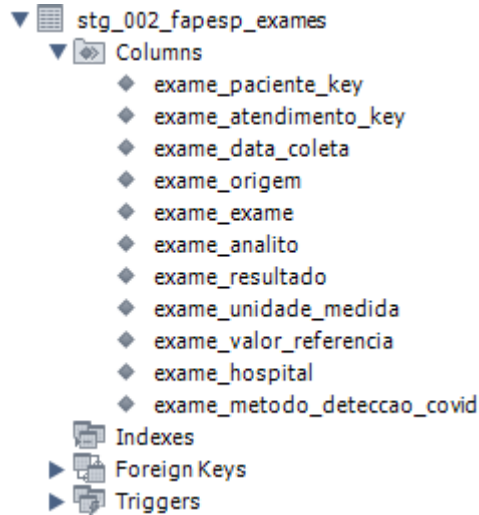
Após a construção de algumas versões preliminares, elaboramos um modelo final de painéis utilizando os dados tratados e filtrados para os pacientes diagnosticados com COVID-19. Para alcançar esse resultado, foram carregadas do MySQL para o Power BI quatro tabelas criadas nos passos de ETL: `stg_002_fapesp_desfechos`, `stg_002_fapesp_exames`, `stg_002_fapesp_pacientes` e `stg_002_fapesp_exames_covid`.

Figura 55: Colunas da Tabela de Desfechos

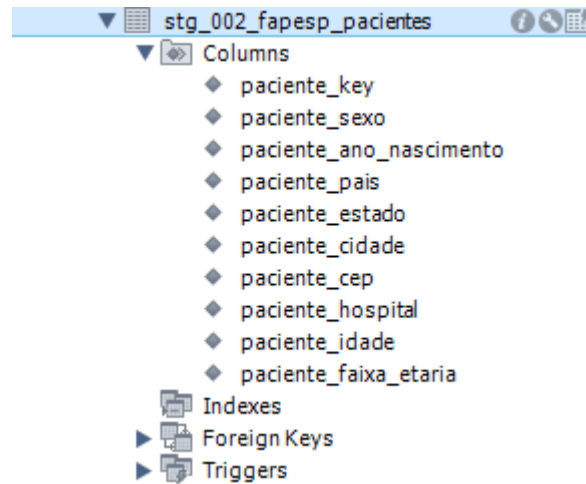


A Figura 55 apresenta as colunas da tabela stg_002_fapesp_desfechos. Seu modelo e estrutura se baseiam na presença de nove colunas. Elas buscam demonstrar as seguintes informações: identificador do paciente, identificador do desfecho do paciente, data do desfecho, tipo de atendimento, identificador da clínica, nome da clínica de desfecho, data do desfecho, nome do desfecho e hospital de desfecho.

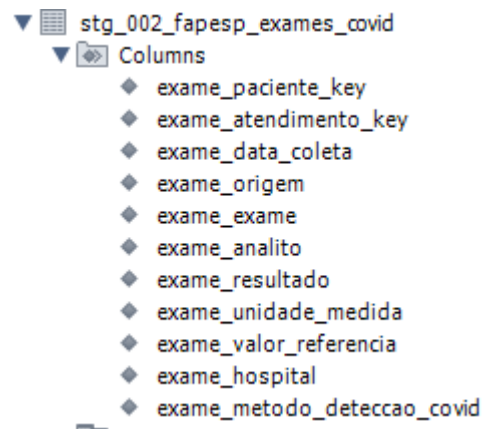
Figura 56: Colunas da Tabela de Exames



Já na tabela da Figura 56, a stg_002_fapesp_exames, o modelo e a estrutura estão baseados em onze colunas, sendo todas elas relacionadas às informações de exames que os pacientes realizaram, os quais foram mapeados dentre os exames do formulário da OMS. Ela possui as seguintes informações: identificador de pacientes, identificador de atendimento, data de coleta do exame, origem do exame, nome do exame, analito do exame, resultado do exame, unidade de medida utilizada, valor de referência do exame, hospital onde exame foi realizado e o método de detecção de COVID-19 .

Figura 57: Colunas da Tabela de Pacientes

O modelo da tabela stg_002_fapesp_paciente, conforme Figura 57, possui uma estrutura similar às demais. Ela é composta por dez colunas e apresentava as seguintes informações: identificador do paciente, sexo do paciente, ano de nascimento do paciente, pais do paciente, estado do paciente, cidade do paciente, cep do paciente, hospital do paciente, idade do paciente e faixa etária do paciente.

Figura 58: Colunas da Tabela de Exames COVID-19

O modelo da tabela stg_002_fapesp_exames_covid, conforme Figura 58, possui uma estrutura similar às demais, apresentando apenas os exames de COVID-19. Ela é composta por onze colunas e apresenta as seguintes informações: identificador de pacientes, identificador de atendimento, data de coleta do exame, origem do exame, nome do exame, analito do exame, resultado do exame, unidade de medida utilizada, valor de referência do exame, hospital onde exame foi realizado e o método de detecção de covid.

Nesse sentido, também foi criada uma identidade visual baseada nas cores do nosso Departamento e nomeamos o *Dashboard* com o título “ANALITO: Visualização de Análises Médicas Concebidas com base em Insumos Nacionais”. É interessante dar ciência de que o painel, apesar de ter sido desenvolvido a partir de uma modelagem de dados diferente da modelagem dimensional proposta, foi pensado e construído conforme as análises possíveis a partir desta. Dessa forma, o objetivo de se utilizar o painel para propor análises segundo o formulário proposto pela OMS se mantém.

Figura 59: Capa do dashboard



Assim, o *Dashboard* conta com a capa (Figura 59) e outras 5 páginas. A saber:

- Dados Demográficos: Apresenta o perfil dos pacientes atendidos.
- Internações Hospitalares: Destaca análises quantitativas e históricas para auxiliar o entendimento do fluxo de entrada e saída de pacientes no Hospital Sirio Libanês.
- Análise comparativa entre Instituições: Provê uma visão geral dos exames realizados pelos pacientes nas instituições envolvidas.
- Exames COVID- 19: Apresenta análises quantitativas com respeito aos diversos tipos de exames de COVID-19 realizados pelos pacientes.
- Exames OMS: Veicula análises quantitativas e históricas com base nos exames mapeados no formulário da OMS.

As páginas referentes às análises adotam como padrão:

Figura 60: Esboço dos Painéis



- Na parte superior, a Figura 60 apresenta: um link para retorno a página principal (Capa) (1), o título do painel (2), um filtro referente ao período dos dados apresentados (3) e o logo do instituto da computação (4). Para as páginas que apresentam dados do HSL e Fleury, nessa posição também é disponibilizado um filtro para as instituições (5). Na parte central, por sua vez: a parte superior à esquerda apresenta os filtros específicos (6) aplicáveis aos gráficos da página. Cada gráfico (7) na página está associado a um título (8) que auxilia seu entendimento.

a) Dados Demográficos

A página de dados demográficos, visualizada na Figura 61, conta com 6 visuais e 3 filtros. Como recurso do PowerBI é possível que os mesmos visuais se filtrem entre si.

Figura 61: Página de Dados Demográficos



Aspectos informativos do Painel de Dados Demográficos:

A Figura 61 apresenta a tela inicial desenvolvida para o dashboard proposto. Como o intuito de um dashboard analítico é permitir que o usuário realize operações que vão além dos dados apresentados na tela por meio de recursos de interatividade, foram criados três filtros para que o usuário possa navegar entre as visualizações presentes na tela. O modelo de filtro apresentado na Figura 61 será repetido nos próximos gráficos e todos os campos presentes naquele entorno servem para filtrar cada um dos gráficos apresentados nas telas.

Para a exibição dos dados apresentados nestes gráficos a métrica utilizada foi a quantidade de pacientes distintos. A partir dela, foi possível construir o cálculo do cartão com a quantidade de pacientes e estivemos trabalhando em realizar uma análise demográfica dos pacientes, identificando a região à qual cada um pertence. Essa análise em especial, porém, não gerou resultados tão significativos devido a todos os hospitais analisados se encontrarem na mesma região. Entretanto, é esperado que, com o crescimento futuro do projeto, possam ser incluídos diversos outros hospitais de outras regiões também.

Nesta mesma página, foi criada uma visualização em formato de pirâmide para mostrar por faixa etária e sexo o grau de incidência de infecções na população. Dessa forma, buscamos traçar o perfil de infecção do vírus segundo os dois principais vetores da sociedade. O gráfico em formato de pirâmide (seção 2.6.2.e) se encaixa nos mesmos conceitos de gráficos de barras da seção 2.6.2.a, mas dispostos de forma mais visual para essa informação específica. Foi escolhido devido à sua facilidade de análise para métricas numéricas agrupadas por um ou dois eixos.

Para tornar possível também a análise histórica, foi desenvolvido outro gráfico de barras (seção 2.6.2.a) evidenciando a quantidade total de pacientes atendidos por mês e por método de detecção de covid. A análise busca identificar o método de detecção comumente usada pelos pacientes e o volume de casos mês a mês.

O Mapa de Árvore (seção 2.6.2.g), situado ao fim do painel, busca evidenciar a proporção de pacientes presentes em cada fase da vida. A ideia é que sua inserção e posição, ajudem a entender mais detalhadamente o perfil do cliente introduzido nas visualizações anteriores.

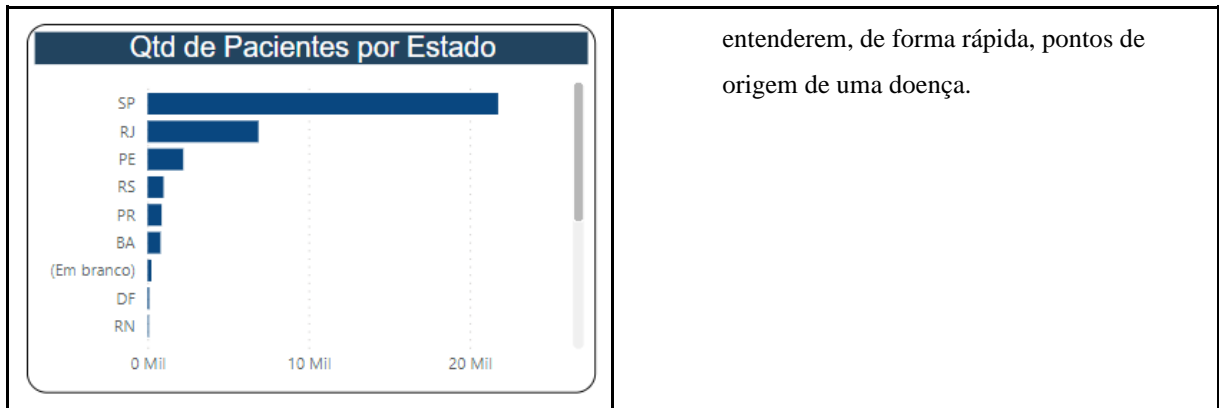
Para o desenvolvimento dessa página, foi utilizada principalmente a métrica de contagem distinta de pacientes, ou seja, a partir da modelagem dimensional proposta, seria possível conceber a visualização utilizando-se a contagem distinta do campo chave “paciente_id” na tabela fato Situação. Os dados utilizados para se compor os eixos e filtros seriam referentes aos campos “mês” e “ano” da dimensão Tempo, “sexo” e “faixa_idade” da dimensão Paciente, além do “nome” da dimensão Exame e “valor_qualit” da fato Resultado para compor o visual de método de detecção. Os valores de altas e óbitos são obtidos do campo “desfecho” na dimensão Atendimento.

Aspectos Técnicos do Painel de Dados Demográficos:

- Análise dos pacientes por Estado (Quadro 1):

Quadro 1: Gráfico de Barras

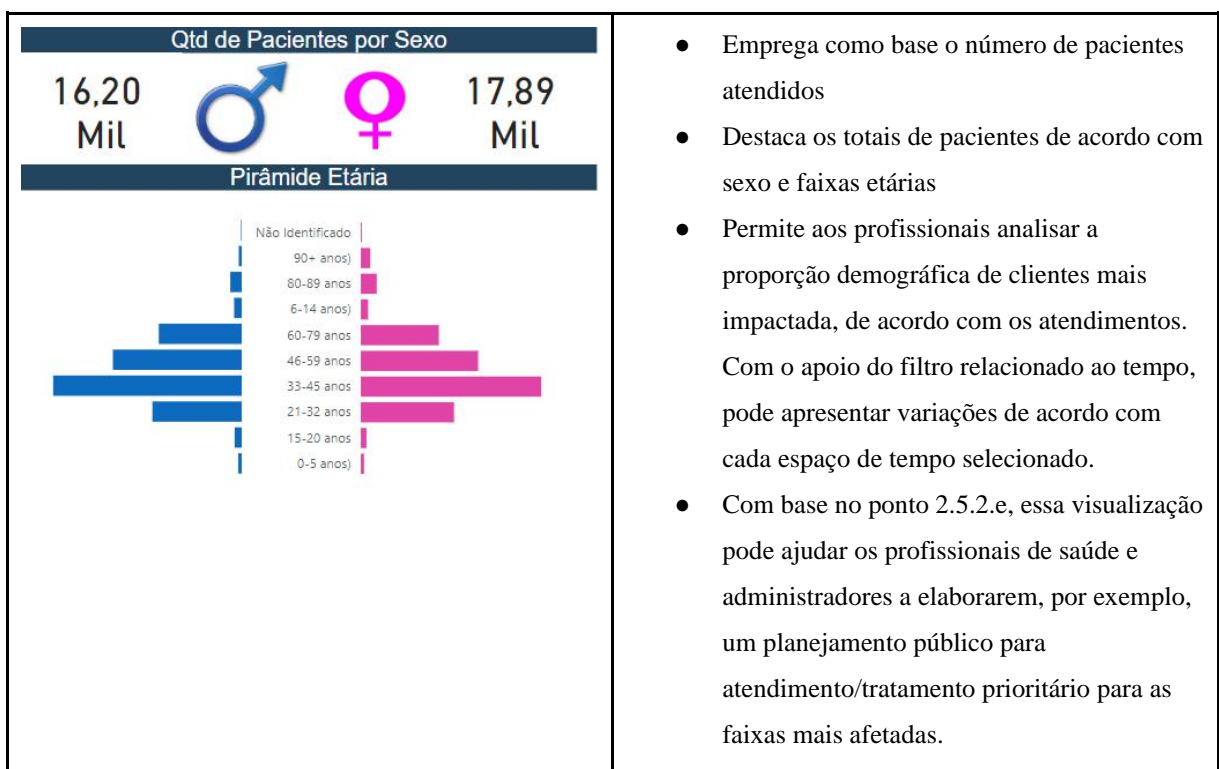
	<ul style="list-style-type: none"> ● Emprega como base o número de pacientes atendidos ● Apresenta o total de pacientes tratados nos hospitais, de acordo com o estado de origem ● Com base no ponto 2.6.1, essa visualização pode auxiliar os profissionais a
--	---



entenderem, de forma rápida, pontos de origem de uma doença.

- Análise Demográfica dos pacientes (Quadro 2):

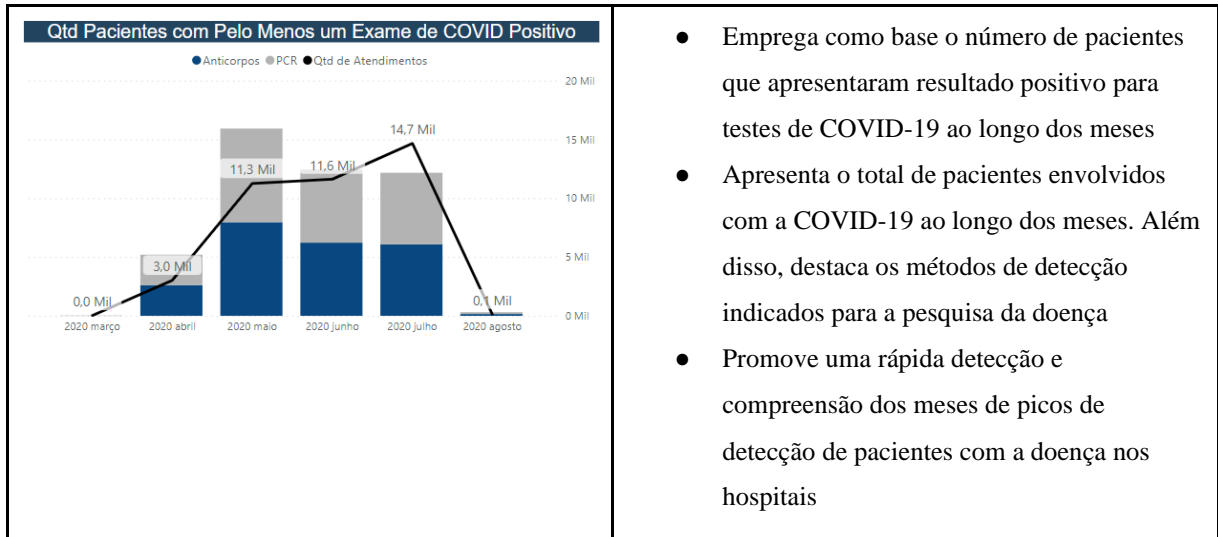
Quadro 2: Pirâmide Etária



- Emprega como base o número de pacientes atendidos
- Destaca os totais de pacientes de acordo com sexo e faixas etárias
- Permite aos profissionais analisar a proporção demográfica de clientes mais impactada, de acordo com os atendimentos. Com o apoio do filtro relacionado ao tempo, pode apresentar variações de acordo com cada espaço de tempo selecionado.
- Com base no ponto 2.5.2.e, essa visualização pode ajudar os profissionais de saúde e administradores a elaborarem, por exemplo, um planejamento público para atendimento/tratamento prioritário para as faixas mais afetadas.

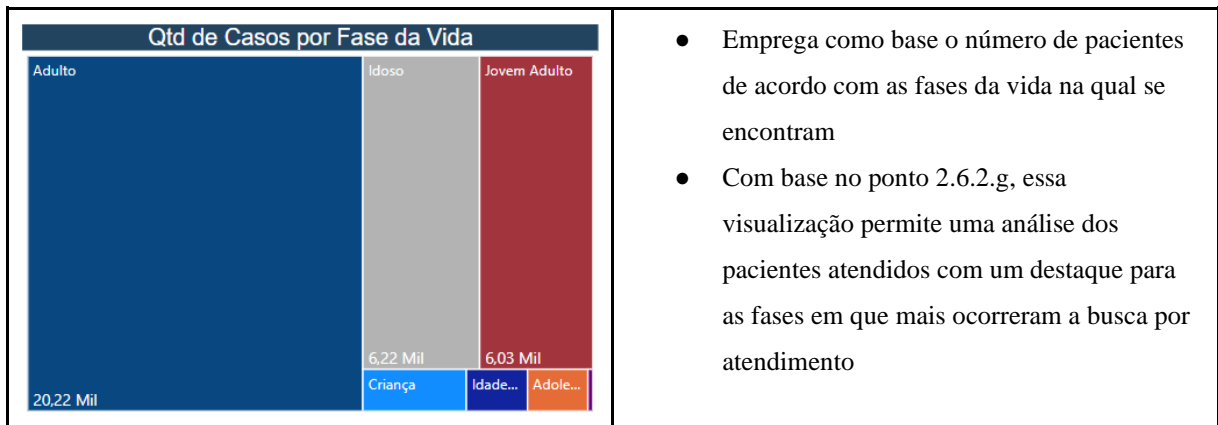
- Evolução dos pacientes ao longo do tempo por método de COVID-19 (Quadro 3):

Quadro 3: Gráficos de Colunas



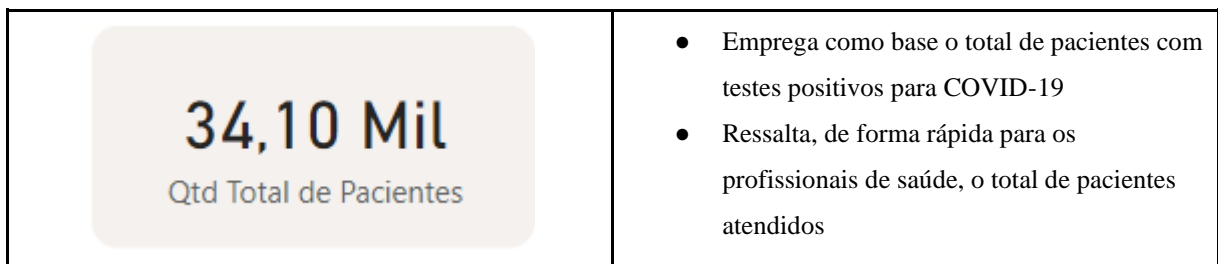
- Proporção dos Pacientes por Fase da Vida (Quadro 4):

Quadro 4: Mapa de Árvore



- Grandes Números dos Pacientes (Quadro 5):

Quadro 5: Cartões

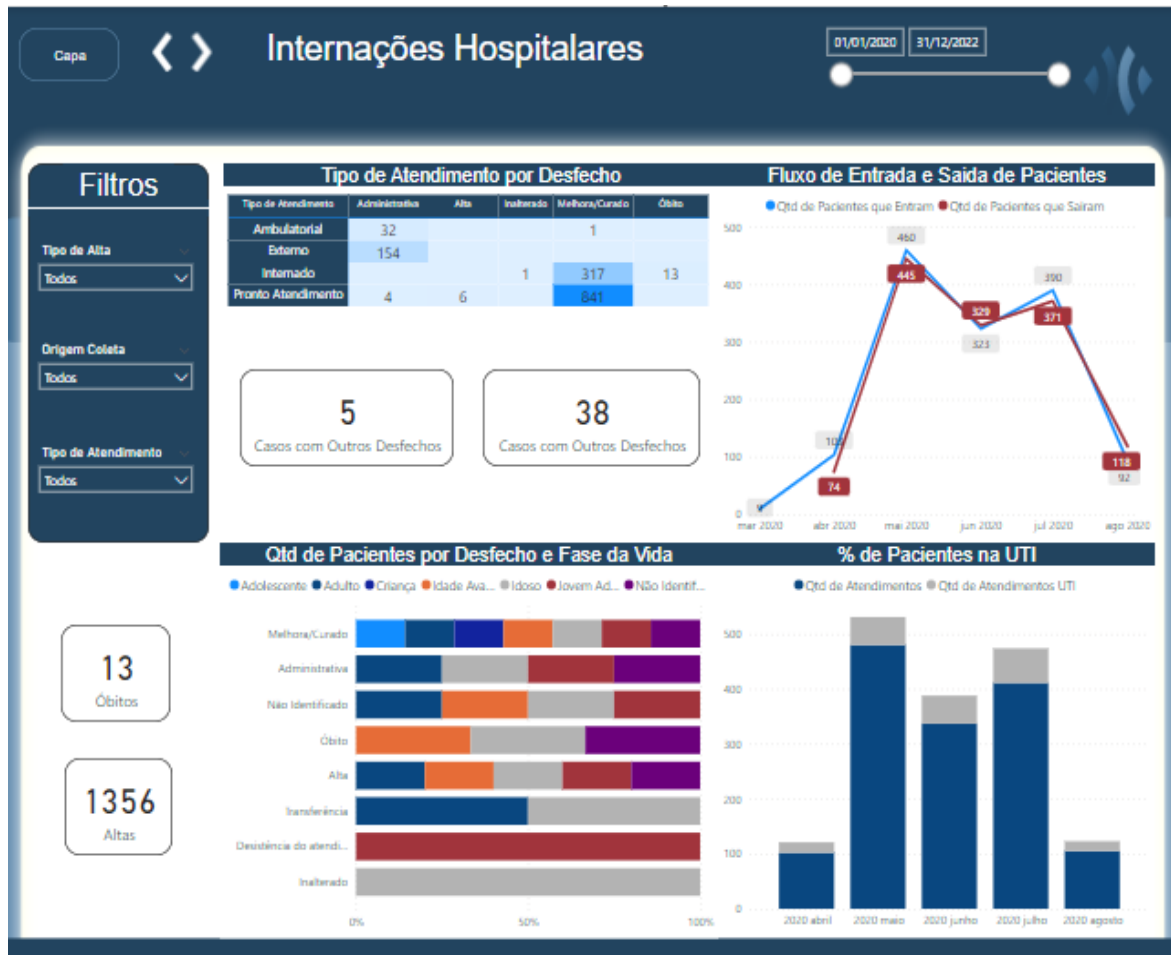


b) Internações Hospitalares

A página de Internações Hospitalares, apresentada na Figura 62, conta com cinco visuais e três filtros. Esta página destina-se a análise sobre pacientes internados, portanto não considera

os pacientes atendidos pelo Grupo Fleury. Os visuais elaborados se propõem a contar a história dos pacientes do HSL, identificando os casos mais graves que exigiram internações em UTI.

Figura 62: Página de Internações Hospitalares



Aspectos informativos do Painel de Informações Hospitalares:

Os visuais existentes nesta página proporcionam análises temporais dos casos do HSL, incluindo o final da história do paciente no hospital. Tais análises se tornam possíveis devido às informações de desfechos disponibilizadas pelo HSL.

Os totais de paciente e a dimensão Tempo são a base para os gráficos apresentados. Conforme destacado na seção 3.3.1, esta página do painel tem o intuito de ser o insumo para os gestores de hospitais. No caso do primeiro gráfico, por exemplo, ele busca mostrar por meio de um mapa de calor no formato de tabela qual é a correlação mais forte e mais fraca entre Tipo de Atendimento e Tipo de Desfecho.

Nesta mesma página, também utilizamos um gráfico de barras 100% empilhadas (apresentado na seção 2.6.2.a), para comparar o tipo de desfecho do paciente por fase da vida.

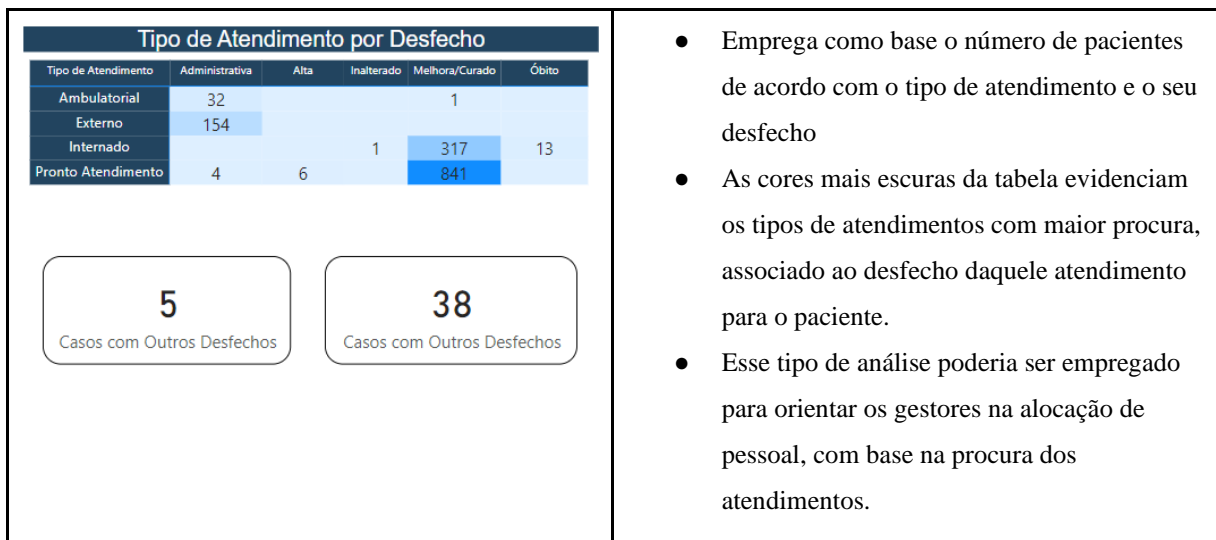
Dessa forma, fica claro o entendimento sobre qual é o momento de vida da maioria dos clientes que vem a óbito ou que tiveram seu atendimento finalizado com qualquer outro tipo de desfecho, além da possibilidade de encontrar padrões.

O segundo gráfico que utilizamos é o gráfico de linhas (apresentado na seção 2.6.2.c), responsável por entender qual é o fluxo de entrada e saída de pacientes durante o dia a dia desses casos presentes.

Aspectos Técnicos do Painel de Internações Hospitalares:

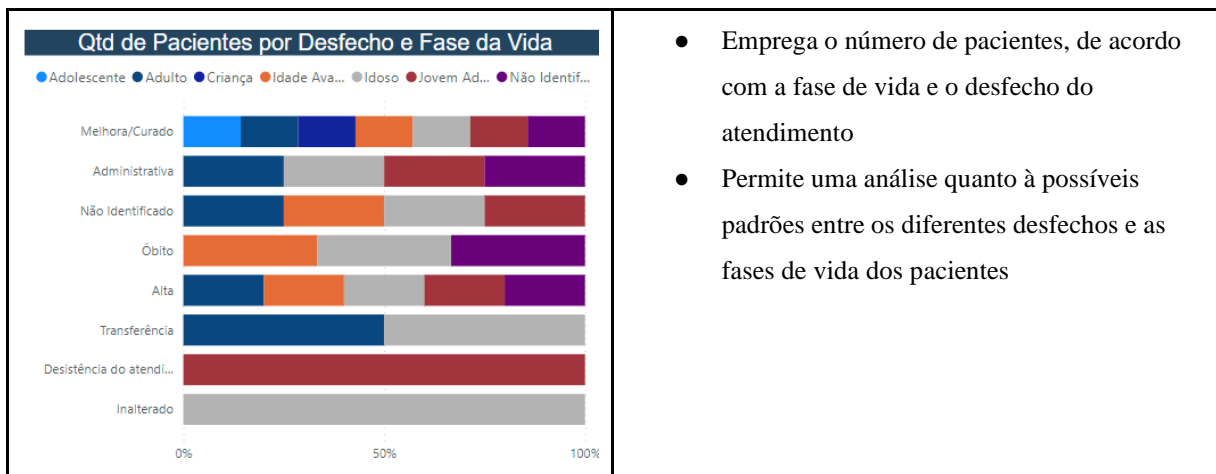
- Análise por tipo de atendimentos (Quadro 6):

Quadro 6: Matriz



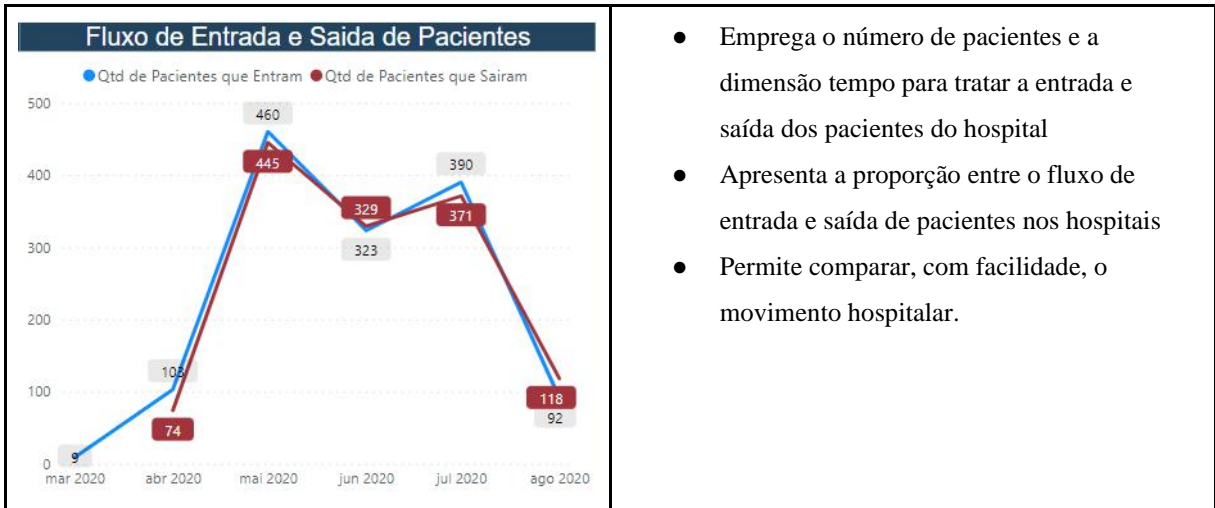
- Análise por Tipo de Desfecho (Quadro 7):

Quadro 7: Gráfico de Barras 100% Empilhadas



- Análise do Fluxo de Pacientes (Quadro 8):

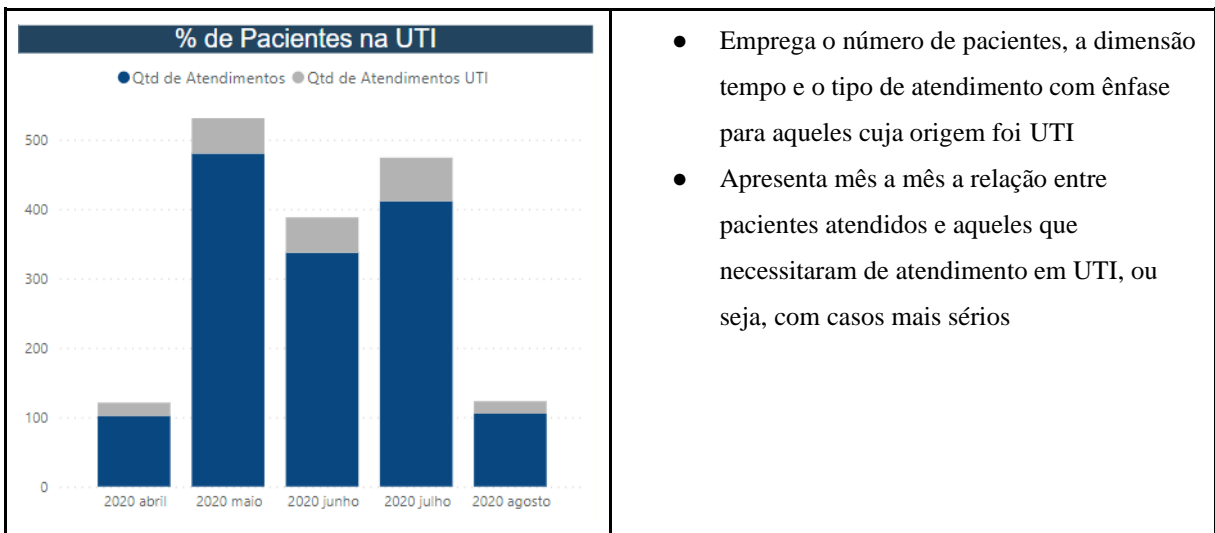
Quadro 8: Gráfico de Linhas



- Emprega o número de pacientes e a dimensão tempo para tratar a entrada e saída dos pacientes do hospital
- Apresenta a proporção entre o fluxo de entrada e saída de pacientes nos hospitais
- Permite comparar, com facilidade, o movimento hospitalar.

- Comparação entre os clientes que foram ou não para UTI (Quadro 9):

Quadro 9: Gráfico de Colunas Empilhadas

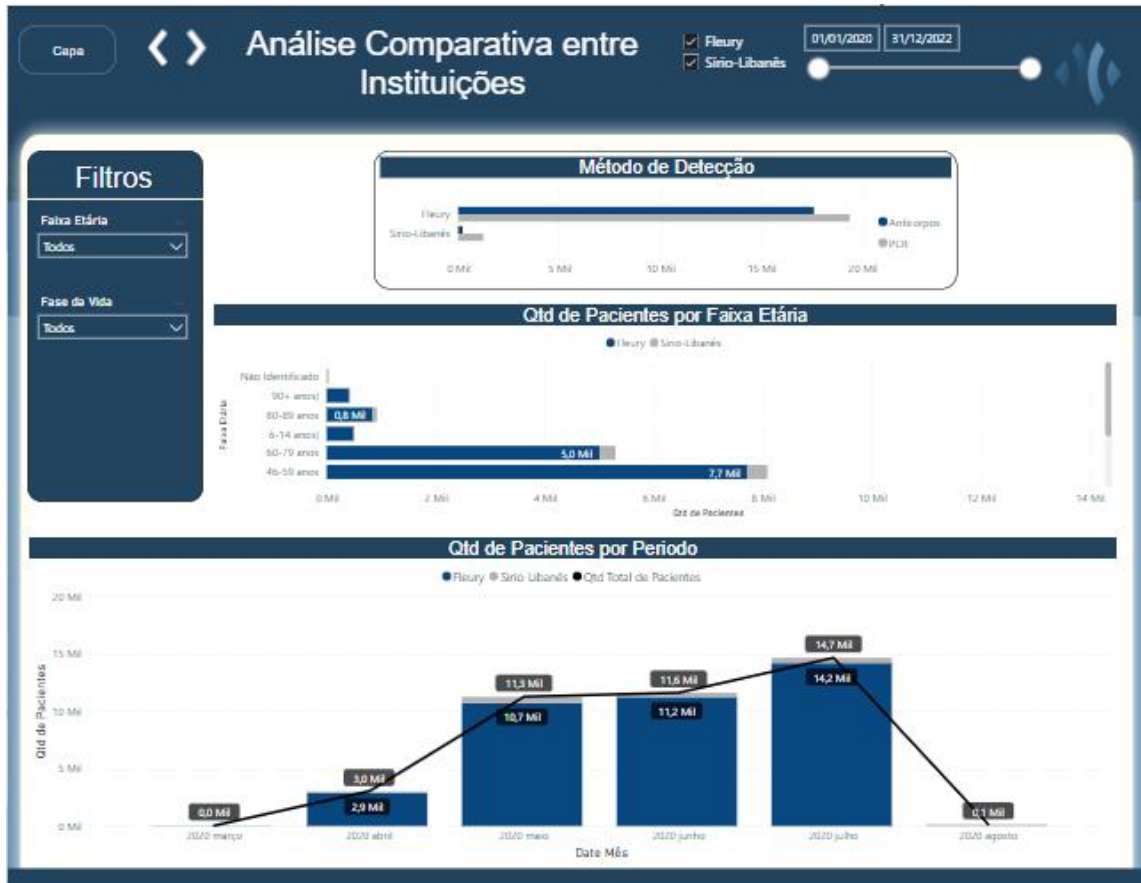


- Emprega o número de pacientes, a dimensão tempo e o tipo de atendimento com ênfase para aqueles cuja origem foi UTI
- Apresenta mês a mês a relação entre pacientes atendidos e aqueles que necessitaram de atendimento em UTI, ou seja, com casos mais sérios

c) Análise Comparativa entre Instituições

A página de Comparativa entre Instituições (Figura 63) conta com 3 visuais e 2 filtros, podendo os mesmos visuais filtrarem-se entre si.

Figura 63: Página da Análise Comparativa entre Instituições



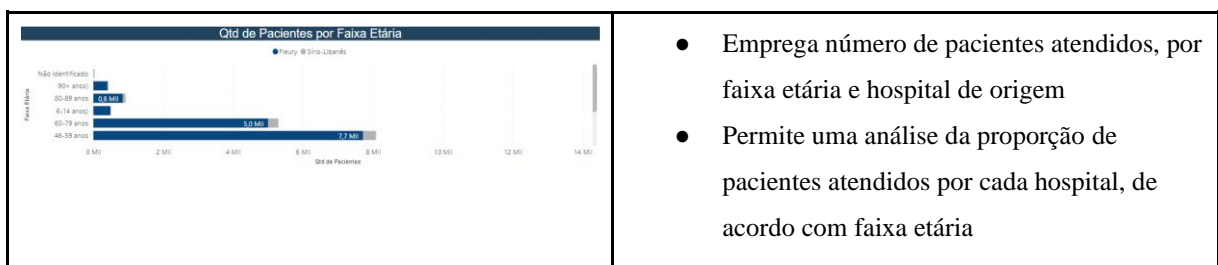
Aspectos informativos do Painel de Análise Comparativa:

Os três novos visuais aqui desenvolvidos visam dar uma visibilidade da quantidade de pacientes atendidos pelos dois tipos de hospitais e comparar dessa forma os perfis de clientes e os métodos de detecção utilizados. Dessa forma, é possível acompanhar com mais precisão o impacto que cada hospital faz na base total de pacientes. Por meio dos gráficos de barras e colunas apresentados no painel, a comparação entre as instituições se torna fácil de ser realizada.

Aspectos Técnicos do Painel de Exames:

- Análise da Quantidade de Pacientes por Faixa Etária e por Hospital (Quadro 10):

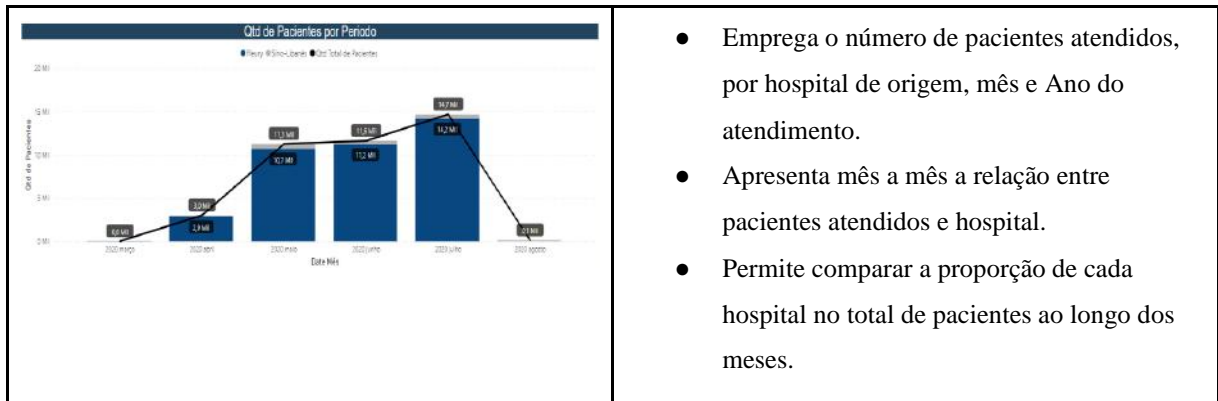
Quadro 10: Gráfico de Barras Empilhadas



- Emprega número de pacientes atendidos, por faixa etária e hospital de origem
- Permite uma análise da proporção de pacientes atendidos por cada hospital, de acordo com faixa etária

- Distribuição de pacientes por período (Quadro 11):

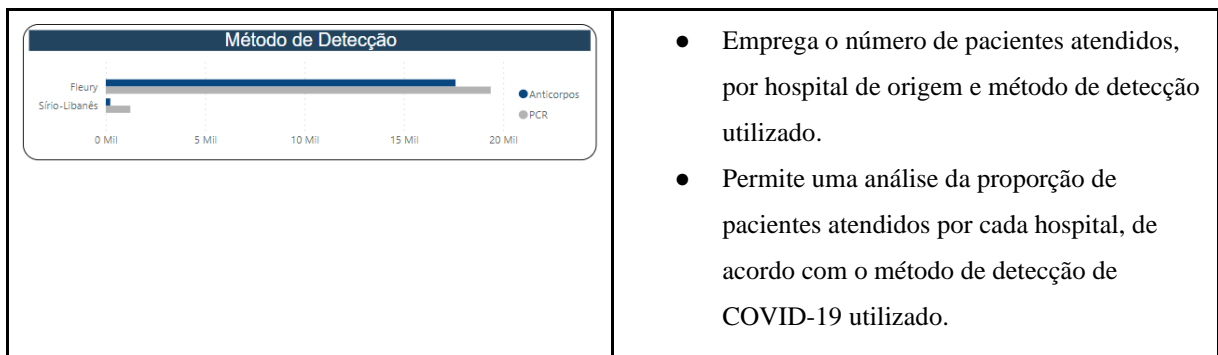
Quadro 11: Gráfico de Colunas Empilhadas e Linha



- Emprega o número de pacientes atendidos, por hospital de origem, mês e Ano do atendimento.
- Apresenta mês a mês a relação entre pacientes atendidos e hospital.
- Permite comparar a proporção de cada hospital no total de pacientes ao longo dos meses.

- Qtd de exames de COVID-19 realizados por Método e Hospital (Quadro 12):

Quadro 12: Gráfico de Barras Clusterizado



- Emprega o número de pacientes atendidos, por hospital de origem e método de detecção utilizado.
- Permite uma análise da proporção de pacientes atendidos por cada hospital, de acordo com o método de detecção de COVID-19 utilizado.

d) Exames COVID-19

A página Exames COVID-19 (Figura 64) conta com 3 visuais e 2 filtros, podendo os mesmos visuais filtrarem-se entre si.

Figura 64: Página Exames COVID-19



Aspectos informativos do Painel Exames COVID-19:

Os visuais construídos nesta página dão visibilidade aos exames de COVID-19 realizados. Os gráficos têm o intuito de correlacionar os métodos de detecção desses tipos de exames com a origem e com o período em que ocorreram.

O primeiro deles, o gráfico de barras (seção 2.6.2.a), utiliza os campos “exame_metodo_deteccao_covid” e “tipo_origem” da dimensão Exames COVID-19, além disso, realiza uma contagem distinta de “exame_atendimento_key” da mesma dimensão para contabilizar os atendimentos dentro das segmentações fornecidas pelos outros dois campos. Dessa forma, é possível identificar de onde provém os exames de COVID e qual desses exames é o mais utilizado por origem.

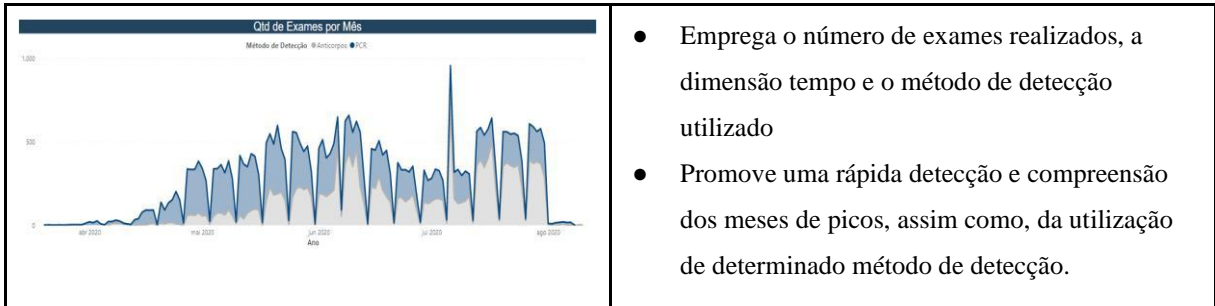
Já o segundo apresenta a volumetria de exames mês a mês a partir de um gráfico de área empilhada (seção 2.6.2.d), facilitando a análise de casos. Ele permite a visualização do total de exames de COVID-19 realizados a cada mês. Mas também, permite analisar de forma

segmentada por método de detecção, identificando quais são os picos de utilização de cada método.

Aspectos Técnicos do Painel de Exames COVID-19:

- Análise de Exames por Método de Detecção (Quadro 13):

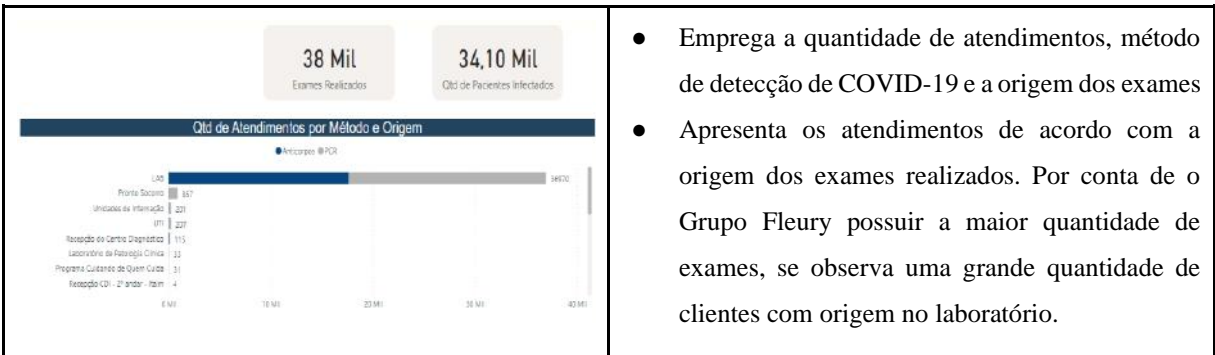
Quadro 13: Gráfico de Área Empilhado



- Emprega o número de exames realizados, a dimensão tempo e o método de detecção utilizado
- Promove uma rápida detecção e compreensão dos meses de picos, assim como, da utilização de determinado método de detecção.

- Análise do Número de Exames Realizados por Perfil Demográfico (Quadro 14):

Quadro 14: Gráfico de Barras Empilhado

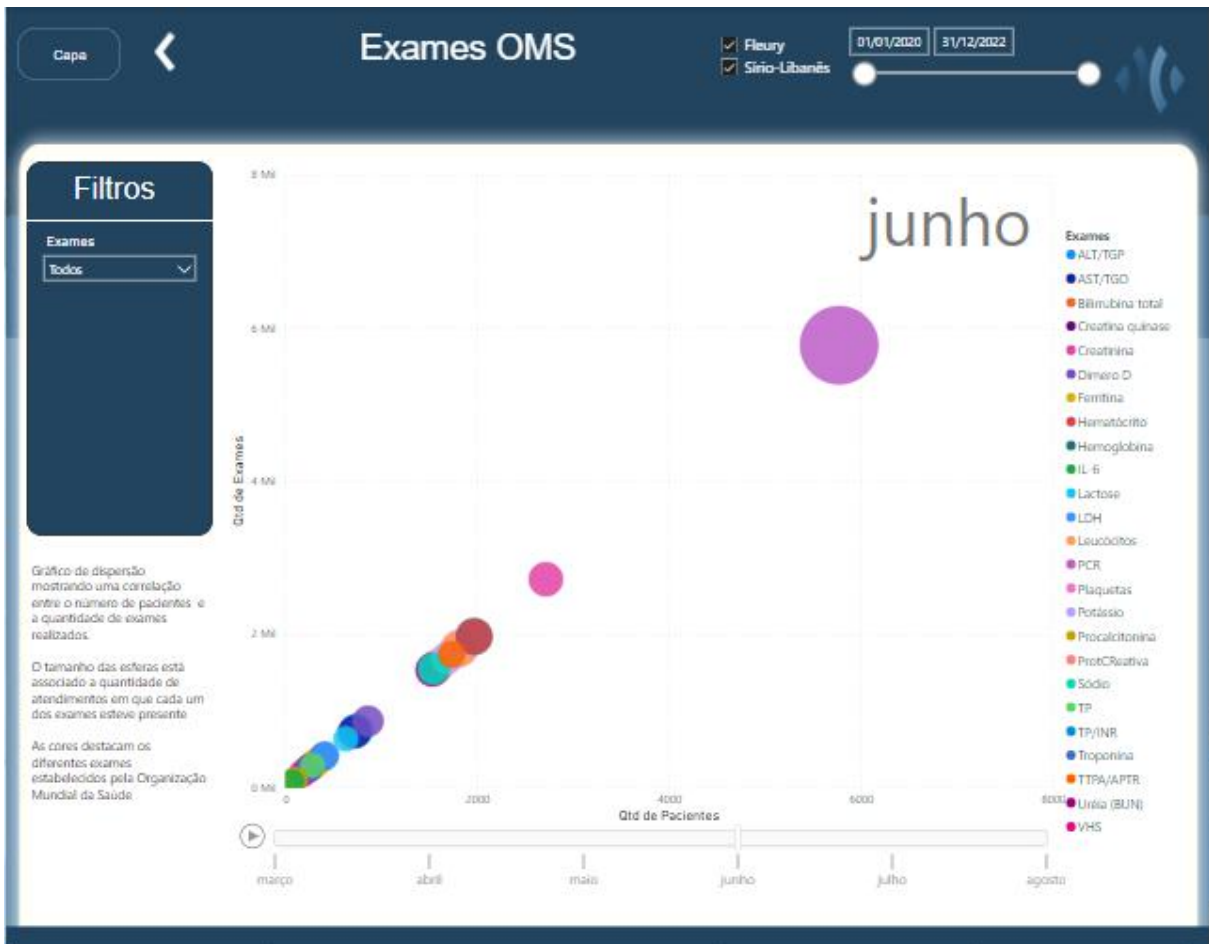


- Emprega a quantidade de atendimentos, método de detecção de COVID-19 e a origem dos exames
- Apresenta os atendimentos de acordo com a origem dos exames realizados. Por conta de o Grupo Fleury possuir a maior quantidade de exames, se observa uma grande quantidade de clientes com origem no laboratório.

e) Exames OMS

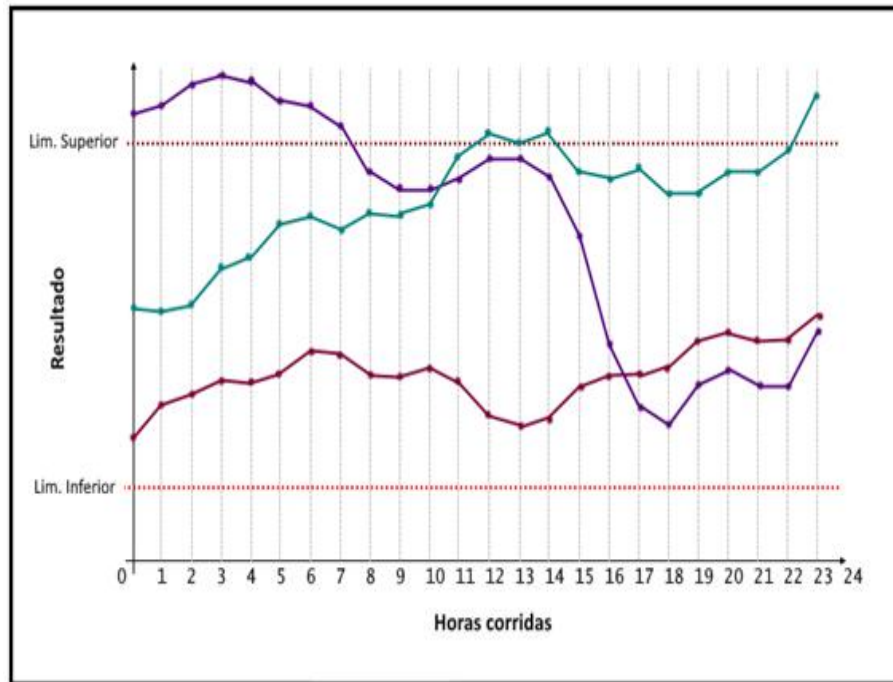
A página Exames OMS, representada na Figura 66, conta com 1 visual e 1 filtro. Ela se destina a apresentar a proporcionalidade mês a mês dos exames estabelecidos pela OMS para serem coletadas nos atendimentos de COVID-19, conforme destacado na Figura 11. Dessa forma, apresenta o total de exames realizados por mês e correlaciona-o com a quantidade de pacientes. Já o tamanho das bolhas é composto pela quantidade de atendimentos distintos, no qual cada um desses exames esteve presente.

Figura 65: Página Exames OMS



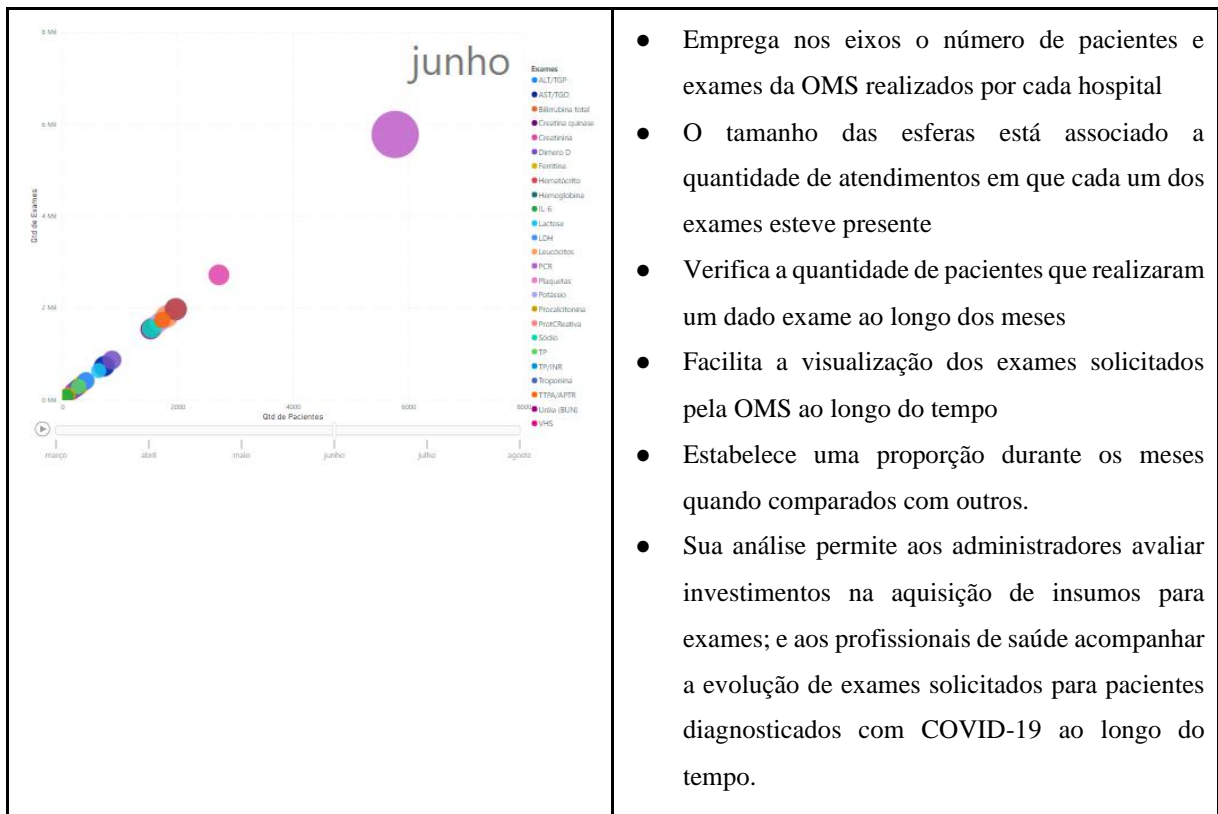
Reiteramos aqui que, devido à impossibilidade de realizar muitos tratamentos nos dados devido ao alto custo de processamento, não foi possível construir a série de análises que anteriormente haviam sido cogitadas. O universo de análises possíveis a partir dos dados de exames obtidos pelo projeto VODAN BR e a arquitetura dimensional construída é bem relevante e pode ser mais bem aproveitado em desenvolvimentos futuros. Dentre estas análises que não puderam ser implementadas, uma que nos chama particular atenção é o gráfico de linhas de Resultados x Horas Corridas. Neste gráfico (exposto na Figura 66) teríamos o valor normalizado “valor_quant_normalizado” (fato Resultado) de acordo com os valores de referência do analito em questão, permitindo assim acompanharmos a tendência de comportamento de quaisquer analitos no decorrer do tratamento de acordo com o seu valor de referência:

Figura 66: Exemplo de visualização do gráfico Resultados x Horas Corridas



- Distribuição de Exames Realizados ao longo do Tempo (Quadro 15):

Quadro 15: Gráfico de Dispersão



- Emprega nos eixos o número de pacientes e exames da OMS realizados por cada hospital
- O tamanho das esferas está associado a quantidade de atendimentos em que cada um dos exames esteve presente
- Verifica a quantidade de pacientes que realizaram um dado exame ao longo dos meses
- Facilita a visualização dos exames solicitados pela OMS ao longo do tempo
- Estabelece uma proporção durante os meses quando comparados com outros.
- Sua análise permite aos administradores avaliar investimentos na aquisição de insumos para exames; e aos profissionais de saúde acompanhar a evolução de exames solicitados para pacientes diagnosticados com COVID-19 ao longo do tempo.

4 CONCLUSÃO

O avanço da tecnologia tem determinado a direção para a qual cada aspecto mundial está se direcionando. E isso não é diferente na área da saúde: uma área que, a cada dia que passa, se torna mais dependente do consumo de dados estruturados para a criação de hipóteses e para o desenvolvimento de novos recursos (remédios, vacinas e etc) que auxiliem a alcançarmos maior longevidade da vida humana.

Os últimos anos de pandemia têm sido definitivos no que tange ao desenvolvimento de tecnologias de dados. Neste trabalho, assim como em outros, vemos que as decisões tomadas com base nos dados que foram disponibilizados por entidades e instituições ao redor do mundo forneceram os insumos necessários para análises estatísticas, monitoramento da doença, logística, entre outros. Mesmo a disponibilização de vacinas em um curto espaço de tempo só foi possível a partir de muitos testes e estudos comportamentais da doença no organismo humano, que gerou os dados necessários para que decisões assertivas pudessem ser tomadas.

Neste trabalho, especificamente, buscamos trazer de forma sucinta os passos que permitiram a criação da primeira versão do relatório para o projeto VODAN BR. É através da união emblemática de uma arquitetura de dados desenvolvida a partir das melhores práticas da área de inteligência de negócio, de uma transformação de dados realizada em aplicação conhecida no mercado e da visualização (relatório) trabalhada em cima de visuais-chave para o estudo proposto que alcançamos a finalidade proposta.

No tocante à arquitetura de dados modelada com a utilização da modelagem estrela, acreditamos ter possibilitado tanto para este como para futuros projetos uma gama de análises ainda não exploradas anteriormente para dados clínicos. Por ela permitir análises granulares como o comportamento de um paciente específico durante um atendimento, mas também permitir a obtenção de uma informação sumarizada através de quaisquer das dimensões do modelo, se torna uma modelagem extremamente versátil.

As visualizações produzidas com a ajuda do Power BI, de modo parecido, se tornam uma base comum para a produção de relatórios mais detalhados e análises mais complexas, além de conseguirem entregar informações úteis para a equipe médica com o que está pronto. Esta organização da informação para o controle efetivo de pesquisas médicas e logísticas administrativas na área da saúde permite compreender os padrões presentes no comportamento do novo coronavírus ao infectar os pacientes.

Infelizmente, entretanto, muitos hospitais ainda carecem de infraestrutura interna para tratamento e governança de dados, dificultando o avanço na implementação da automatização de relatórios e BI, impossibilitando assim que os médicos e pesquisadores tenham acesso a informações mais assertivas e rápidas.

Neste sentido, acreditamos que o propósito deste trabalho foi cumprido no que tange a gerar um relatório modelo a ser evoluído conforme o projeto VODAN BR cresce e toma forma perante as instituições de saúde brasileiras. Como próximos passos deste trabalho, propomos uma retomada do tratamento dos dados que se encontram em formato transacional (Figura 14) para adequação à modelagem estrela proposta, utilizando ferramentas escaláveis, que hoje já são disponibilizadas pelas instituições parceiras, para realizar o processamento massivo dos dados.

4.1 DIFICULDADES ENCONTRADAS

4.1.1 Definição de escopo

Por estarmos abordando um assunto muito novo no cenário mundial, as prioridades dentre as análises a serem feitas estavam em constante mudanças. Sendo assim, o projeto apresentou uma grande dificuldade a respeito de até onde poderíamos chegar com os nossos estudos.

Um agravante é que também precisávamos estar em constante contato com médicos e funcionários dos hospitais para poder entender sua real necessidade, mas durante a pandemia ficou muito difícil entrar em contato com esta classe. Dado que, a cada dia que passava, os números se tornavam cada vez maiores e mais variantes surgiam, então o contato, que era para ser constante, só foi possível de ser realizado em um único momento. Portanto, entendemos que a falta de *feedback* por parte dos pontos focais impossibilitou diversas melhorias que estavam previstas a serem realizadas durante este trabalho.

4.1.2 Tecnologia

Dentre as áreas de tecnologia, tivemos algumas dificuldades, como por exemplo a falta de uma máquina com um bom potencial de processamento, o que impossibilitou com que

podéssemos aplicar o modelo dimensional estruturado para os dados extraídos da FAPESP. Além disso, apesar do Power BI possuir uma grande variedade de gráficos disponíveis para uso, em alguns momentos é um pouco complexo encontrar de forma fácil para download a visualização desejada. Entretanto, após alguns estudos e pesquisas, torna-se possível sobrepujar esse desafio.

4.2 TRABALHOS FUTUROS

É possível notar que existem algumas análises relevantes para o negócio que podem ser exploradas à medida que o dado seja disponibilizado, como: o efeito da vacina no número de casos de COVID-19; as características encontradas nos pacientes nos diversos picos da pandemia à medida que novas variantes surgem; doenças que agravam as situações dos pacientes; além de outras que provavelmente vão surgir à medida que a doença sofre mutações e a medicina se torna mais especializada no assunto. Inclusive, com a possibilidade de utilizar computação distribuída para o processamento dos dados, é possível pensar e testar análises bem mais complexas.

A modelagem de dados aqui proposta pode também ser facilmente adaptada a outros diagnósticos diferentes do COVID-19, possibilitando assim a sua utilização para reagir a novos surtos epidêmicos e/ou diagnósticos usuais de rotina. Essa mesma modelagem pode ser utilizada para a construção de diversos outros relatórios e *dashboards* dada a sua versatilidade.

Enfim, também é importante uma reunião com pontos focais em instituições médicas para análise de eficácia e uso dos relatórios criados, e a geração de novas visualizações baseadas nos *feedbacks* adquiridos. Já têm sido empreendidos alguns esforços nesse sentido para com as equipes médicas dos hospitais parceiros a fim de identificar novas análises e possibilidades. Esperamos que, futuramente, muitas instituições de saúde venham se beneficiar do trabalho aqui produzido.

REFERÊNCIAS

- BARBIERI, C. **BI2 - Business intelligence: Modelagem & Qualidade**. [S.l.]: Elsevier Editora Ltda, 2011.
- BAUMGARTEN, C. et al. Data-driven prediction of the therapeutic window during subthalamic deep brain stimulation surgery. **Stereotactic and Functional Neurosurgery**, v. 96, n. 3, p. 142–150, 2018.
- BIBLIA. **Bíblia Sagrada: N.T. Lucas**. 9. ed. [S.l.]: Ave-Maria, 2018.
- BONEL, C. **Metodologia e Engenharia de Requisitos para projetos de Business Intelligence: Análise de dados levada a sério**. [S.l.]: Clube de Autores, 2021.
- BOUMAN, R.; DONGEN, J. V. D. **Pentaho Solutions: Business Intelligence and Data Warehousing with Pentaho and MySQL**. [S.l.]: Wiley, 2009
- BUILDERS, I. 2015. Disponível em: <https://infocenter.informationbuilders.com/wf80/index.jsp?topic=/pubdocs/charting/JSChart/source/topic106.htm>. Acesso em: 2 jun. 2022.
- CAMPBELL-KELLY, M. et al. **Computer: A history of the information machine**. [S.l.]: Routledge, 2018.
- CARRA, G. **Photo by Giacomo Carra on unsplash**. 2020. Disponível em: <https://unsplash.com/photos/gf6UDwpl0ac>. Acesso em: 2 jun. 2022.
- CECI, F. **Business Intelligence**. [S.l.: s.n.], 2012. ISBN 9788578174651.
- CHAUDHURI, S.; DAYAL, U. An overview of data warehousing and olap technology. **ACM Sigmod record**, ACM New York, NY, USA, v. 26, n. 1, p. 65–74, 1997.
- CHEN, M. et al. Disease prediction by machine learning over big data from healthcare communities. **IEEE Access**, v. 5, p. 8869–8879, Apr 2017.
- FAIR principles. 2022. Disponível em: <https://www.go-fair.org/fair-principles/>. Acesso em: 8 jun. 2021.
- FAPESP. 2016. Disponível em: <https://repositoriodatasharingfapesp.uspdigital.usp.br/>. Acesso em: 18 abr. 2022.
- GITLIN, J. **Quando e como usar os tipos de gráfico mais populares da SurveyMonkey**. 2022. Disponível em: <https://pt.surveymonkey.com/curiosity/when-how-to-use-surveymonkeys-most-popular-chart-types/>. Acesso em: 1 jun. 2022.
- GLOBAL covid-19 clinical platform: Rapid core case report form (CRF). World Health Organization, 2020. Disponível em: https://www.who.int/publications/i/item/WHO-2019-nCoV-Clinical_CRF-2020.4. Acesso em: 26 jan. 2022.
- IBGE. **Estatísticas do Século XX**. 2016. Disponível em: <https://ftp.ibge.gov.br/seculoxx/>. Acesso em: 1 jun. 2022.

- KESSELHEIM, A. S. et al. An overview of vaccine development, approval, and regulation, with implications for covid-19. **Health Affairs**, v. 40, n. 1, p. 25–32, 2021.
- KHANAM, F.; NOWRIN, I.; MONDAL, M. R. H. Data visualization and analyzation of covid-19. **Journal of Scientific Research and Reports**, v. 26, n. 3, p. 42–52, 2020
- KIMBALL, R.; ROSS, M. **The Data Warehouse Toolkit: The definitive guide to dimensional modeling**. [S.l.]: John Wiley & Sons, 2013.
- KNAFLIC, C. N. **Storytelling com dados: um guia sobre visualização de dados para profissionais de negócios**. [S.l.]: Alta Books, 2019
- LI, W. et al. Analyzing the research evolution in response to covid-19. **ISPRS International Journal of Geo-Information**, v. 10, n. 4, p. 237, 2021.
- MACKENZIE. **Redação, português, inglês, matemática, geografia, química, biologia, história e física: Grupos II - III, Tipo de Prova: A**. 2019.
- MADDEN, S. From databases to big data. **IEEE Internet Computing**, IEEE, v. 16, n. 3, p. 4–6, 2012.
- MICROSOFT. **O que é visualização de dados?** 2022. Disponível em: <https://powerbi.microsoft.com/pt-br/data-visualization/>. Acesso em: 22 jun. 2021.
- MONS, B. The vodan in: Support of a fair-based infrastructure for covid-19. **European Journal of Human Genetics**, v. 28, n. 6, p. 724–727, May 2020.
- PAINEL de casos de doença pelo coronavírus 2019 (COVID-19) no Brasil. 2020. Disponível em: <https://covid.saude.gov.br/>. Acesso em: 26 jan. 2022.
- PAVKOV, S.; POSCIC, P.; JAKSIC, D. Business intelligence systems yesterday, today and tomorrow - an overview. **Zbornik Veleucilista u Rijeci**, v. 4, n. 1, p. 97–108, 2016. Disponível em: <https://hrcak.srce.hr/160236>. Acesso em: 17 ago. 2022.
- POWER, D. J. A brief history of decision support systems. **DSSResources.com**, v. 3, 2007. Disponível em: <https://search.iczhiku.com/paper/CBucPxik4xSCfkbH.pdf>. Acesso em: 17 ago. 2022.
- RANJAN, J. Business intelligence: Concepts, components, techniques and benefits. **Journal of theoretical and applied information technology**, v. 9, n. 1, p. 60–70, 2009.
- SANCHEZ, M. et al. Mortalidade por covid-19 no brasil: Uma análise do registro civil de óbitos de janeiro de 2020 a fevereiro de 2021. **Scielo Preprints**, Mar 2021.
- SHARDA, R.; DELEN, D.; TURBAN, E. **Business intelligence: a managerial perspective on analytics**. [S.l.]: Prentice Hall Press, 2013.
- VEIGA, V. et al. Vodan br: A gestão de dados no enfrentamento da pandemia coronavírus. **Páginas a&b : Arquivos & Bibliotecas**, n. esp., p. 51–58, 2020.

WHO coronavirus (COVID-19) dashboard. World Health Organization, 2020. Disponível em: <https://covid19.who.int/>. Acesso em: 26 jan. 2022.

WILKINSON, M. D. et al. The fair guiding principles for scientific data management and stewardship. **Scientific Data**, v. 3, n. 1, Mar 2016.

XAVIER, G.; KATILO, L. **OS 13 Gráficos Mais utilizados por analistas e Como Interpretá-los**. 2022. Disponível em: <https://www.opuspesquisa.com/blog/tecnicas/tipos-de-graficos/>. Acesso em: 1 jun. 2022.

ZENG, L. et al. Techniques, process, and enterprise solutions of business intelligence. In: IEEE. 2006 IEEE International Conference on Systems, Man and Cybernetics. [S.l.], 2006. v. 6, p. 4722–4726.

APÊNDICES

FORMULÁRIO DA OMS



Global COVID-19 Plataforma Clínica

NOVO CORONAVÍRUS (COVID-19) - VERSÃO RÁPIDA

DESENHO DESTE FORMULÁRIO DE RELATO DE CASO (FRC)

Este FRC compreende 3 módulos:

Módulo 1 – preencher no primeiro dia de admissão à unidade de saúde.

Módulo 2 – preencher no primeiro dia de admissão à UTI ou Unidade de Cuidados Especiais (UCE). Este Módulo deve ser preenchido diariamente durante o maior número possível de dias, conforme a disponibilidade de mesmo. Continuar o acompanhamento dos pacientes transferidos para outras enfermarias.

Módulo 3 – preencher na alta ou em caso de óbito.

ORIENTAÇÕES GERAIS

- O FRC visa a coleta de dados de exames clínicos, entrevistas e prontuários médicos. Os dados podem ser coletados de forma retrospectiva se o paciente for cadastrado em data posterior a sua admissão.
- Número de Identificação do participante consiste de um código da localidade e um número do participante. Você pode obter o código do local e fazer o cadastramento no Sistema de Gerenciamento de Dados pelo email EDCARN@who.int. Os números dos participantes devem ser alocados sequencialmente para cada local, começando com 00001. Caso uma mesma localidade recrute participantes de enfermarias diferentes, ou caso seja difícil alocar números sequenciais, os números podem ser alocados em blocos ou incluir letras. Por exemplo, a Enfermaria X alocaria os números a partir de 00001 ou a partir de A0001 e assim por diante; a Enfermaria Y alocaria os números a partir de 50001 ou a partir de B0001 e adiante. Anote o número de identificação do participante no topo de cada página.
- Os dados são inseridos no banco de dados eletrônico central REDCap, disponível em <https://ncov.medsci.ox.ac.uk>, ou no banco de dados independente do seu local/sistema. Os FRCs impressos podem ser utilizados e os dados inseridos no banco de dados eletrônico posteriormente.
- Preencha todas as partes do formulário. As perguntas marcadas “Caso afirmativo...” devem ser deixadas em branco quando não forem relevantes (quando a resposta não for “Sim”).
- As opções marcadas com quadrados () são para respostas únicas (apenas uma resposta pode ser registrada).
- As opções marcadas com círculos () são para respostas múltiplas (todas as respostas aplicáveis devem ser marcadas).
- Marque “Desconhecido” para dados desconhecidos ou indisponíveis.
- Evite anotar dados fora das áreas designadas no formulário.
- Para o preenchimento dos FRCs impressos, recomenda-se escrever claramente com caneta, usando LETRAS DE FORMA.
- Marque as respostas com um (X). Para retificar as respostas, risque (-----) a resposta incorreta e anote os dados corretos acima. Rubrique e date toda rasura.
- Mantenha juntos todos os papéis de cada participante, utilizando um grampo ou uma pasta individualizada.
- Transfira todos os dados dos FRCs impressos para o banco de dados eletrônico. Todos os FRCs impressos podem ser arquivados pela instituição responsável. Todos os dados devem ser transferidos para o banco de dados eletrônico seguro.
- Insira os dados no sistema eletrônico disponível em: <https://ncov.medsci.ox.ac.uk>. Se sua localidade desejar coletar os dados de forma independente, podemos fornecer apoio para a criação de bancos de dados locais

COVID-19 CASE RECORD FORM RAPID version 23MAR2020

Adapted from SPRINT SARI CRF by ISARIC. Licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/) by ISARIC on behalf of Oxford University.

- Nosso contato para ajuda com os bancos de dados, para eventuais comentários e para nos informar sobre a utilização dos formulários é: EDCARN@who.int.



ID DO PARTICIPANTE | | | | | | | | | | | | | | | |

SINAIS E SINTOMAS NA HORA DA ADMISSÃO (Desc = Desconhecido)					
Histórico de febre	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc	Retração torácica	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc		
Tosse com expectoração	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc	Dor de cabeça	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc		
Tosse com hemoptise	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc	Consciência alterada / confusão	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc		
Dor de garganta	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc	Convulsões	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc		
Coriza (rinorreia)	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc	Dor abdominal	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc		
Respiração sibilante	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc	Vômitos / Náusea	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc		
Dor no peito	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc	Diarreia	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc		
Dor muscular (mialgia)	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc	Conjuntivite	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc		
Dor articular (artralgia)	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc	Erupções cutâneas	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc		
Fadiga / Mal estar	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc	Úlceras cutâneas	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc		
Falta de ar	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc	Linfadenopatia	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc		
Incapaz de andar	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc	Sangramento (hemorragia)	<input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc		
		Caso afirm: especifique o(s) local(is):			
Outros <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc Caso afirm, especifique: _____					
MEDICAÇÃO O paciente está recebendo algum dos seguintes medicamentos ATUALMENTE?					
Hidratação oral/orogástrica? <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc. Hidratação venosa? <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc					
Antiviral? <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc Caso afirm: <input type="radio"/> Ribavirina <input type="radio"/> Lopinavir/Ritonavir <input type="radio"/> Inibidor de neuraminidase					
<input type="radio"/> Interferon alfa <input type="radio"/> Interferon beta <input type="radio"/> Outro, especifique: _____					
Corticosteroide? <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc Caso afirm, via: <input type="radio"/> Oral <input type="radio"/> Intravenosa <input type="radio"/> Inalatória					
Caso afirm, informa agente e dose diário máximo: _____ Antibiótico? <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc					
Agente antifungal? <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc					
Agente antimalárico? <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc Caso afirm, especifique: _____					
Agente experimental? <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc Caso afirm, especifique: _____					
Anti-inflamatório não esteroide (AINE) <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc					
Inibidores da enzima de conversão da angiotensina (inibidores da ECA) <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc.					
Bloqueadores dos receptores de angiotensina II (BRAs) <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc					
CUIDADOS O paciente esta recebendo algum dos seguintes cuidados ATUALMENTE?					
UTI ou UCE? <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc					
Oxigenoterapia? <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc Caso afirm, complete todos os itens a seguir					
Vazão do O ₂ : <input type="radio"/> 1-5 L/min <input type="radio"/> 6-10 L/min <input type="radio"/> 11-15 L/min <input type="radio"/> >15 L/min <input type="radio"/> Desc					
Fonte do oxigênio: <input type="radio"/> canalizado <input type="radio"/> cilindro <input type="radio"/> concentrador <input type="radio"/> Desc					
Interface: <input type="radio"/> cânula nasal <input type="radio"/> cânula nasal de alto fluxo <input type="radio"/> máscara <input type="radio"/> máscara com reservatório <input type="radio"/> máscara CPAP/VNI					
Ventilação não-invasiva? (p.ex. BIPAP, CPAP) <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc.					
Ventilação invasiva (alguma)? <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc Inotrópicos/vasopressores? <input type="checkbox"/> Sim <input type="checkbox"/> Não <input type="checkbox"/> Desc					
RESULTADAS LABORATORIAIS NA ADMISSÃO (*anote as unidades quando diferente das especificadas)					
Parâmetro	Valor*	Não feito	Parâmetro	Valor*	Não feito
Hemoglobina (g/L)		<input type="checkbox"/>	Creatinina (µmol/L)		<input type="checkbox"/>
Leucócitos (x10 ⁹ /L)		<input type="checkbox"/>	Sódio (mEq/L)		<input type="checkbox"/>
Hematócrito (%)		<input type="checkbox"/>	Potássio (mEq/L)		<input type="checkbox"/>
Plaquetas (x10 ⁹ /L)		<input type="checkbox"/>	Procalcitonina (ng/mL)		<input type="checkbox"/>
TTPA/APTR		<input type="checkbox"/>	PCR (mg/L)		<input type="checkbox"/>
TP (segundos)		<input type="checkbox"/>	LDH (U/L)		<input type="checkbox"/>
INR		<input type="checkbox"/>	Creatina quinase (U/L)		<input type="checkbox"/>
ALT/TGP (U/L)		<input type="checkbox"/>	Troponina (ng/mL)		<input type="checkbox"/>
Bilirrubina total (µmol/L)		<input type="checkbox"/>	VHS (mm/hr)		<input type="checkbox"/>
AST/TGO (U/L)		<input type="checkbox"/>	Dimero D (mg/L)		<input type="checkbox"/>
Ureia (BUN) (mmol/L)		<input type="checkbox"/>	Ferritina (ng/mL)		<input type="checkbox"/>
Lactose (mmol/L)		<input type="checkbox"/>	IL-6 (pg/mL)		<input type="checkbox"/>



ID DO PARTICIPANTE | | | | | | | | | | | | | | | | | | | | | |

Admissão ao CTI ou UTI? Sim Não Desc **Caso afirm,** duração total: ____ dias
 Data de **admissão ao CTI:** [_D][_D][/][_M][_M][/][2][0][_Y][_Y] Desconhecido
 Data de alta da CTI/UTI [_D][_D][/][_M][_M][/][2][0][_Y][_Y] Ainda não teve alta Desc
Oxigenoterapia? Sim Não Desc **Caso afirm, complete todos os itens a seguir,**
 Vazão do O₂: 1-5 L/min 6-10 L/min 11-15 L/min >15 L/min
 Fonte do oxigênio: canalizado cilindro concentrador
Interface: cânula nasal cânula nasal de alto fluxo máscara máscara com reservatório máscara CPAP/VNI
Ventilação não-invasiva? (*p.ex. BIPAP, CPAP*) Sim Não Desc
Ventilação invasiva (alguma)? Sim Não Desc **Inotrópicos/vasopressores?** Sim Não Desc
Suporte extracorpóreo (ECMO)? Sim Não Desc **Posição prona?** Sim Não Desc

