



VINÍCIUS DOS SANTOS ARAÚJO

**Utilização de Ferramenta de *Business Intelligence* para Análise dos Registros de
Cânceres nos Hospitais do Sistema Único de Saúde (SUS)**

Sorocaba

2022

VINÍCIUS DOS SANTOS ARAÚJO

**UTILIZAÇÃO DE FERRAMENTA DE BUSINESS INTELLIGENCE PARA
ANÁLISE DOS REGISTROS DE CÂNCERES NOS HOSPITAIS DO SISTEMA
ÚNICO DE SAÚDE (SUS)**

Trabalho de Conclusão de
Curso apresentado ao Instituto de
Ciência e Tecnologia de
Sorocaba, Universidade Estadual
Paulista (UNESP), como parte
dos requisitos para obtenção do
grau de Bacharel em Engenharia
de Controle e Automação.

Orientador(es): Prof. Dr. Márcio
Alexandre Marques.

Sorocaba

2022

A663a

Araújo, Vinícius dos Santos

Utilização de ferramenta de Business Intelligence para análise dos registros de cânceres nos hospitais do sistema único de saúde (SUS) / Vinícius dos Santos Araújo. -- Sorocaba, 2022 70 p.

Trabalho de conclusão de curso (Bacharelado - Engenharia de Controle e Automação) - Universidade Estadual Paulista (Unesp), Instituto de Ciência e Tecnologia, Sorocaba
Orientador: Márcio Alexandre Marques

1. Business Intelligence. 2. Dashboard. 3. Tableau. 4. Neoplasias. 5. INCA. I. Título.



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Instituto de Ciência e Tecnologia
Câmpus de Sorocaba

UTILIZAÇÃO DE FERRAMENTA DE BUSINESS INTELLIGENCE PARA
ANÁLISE DOS REGISTROS DE CÂNCERES NOS HOSPITAIS DO SISTEMA ÚNICO
DE SAÚDE (SUS)

VINÍCIUS DOS SANTOS ARAÚJO

ESTE TRABALHO DE GRADUAÇÃO FOI JULGADO ADEQUADO
COMO PARTE DO REQUISITO PARA OBTENÇÃO DO GRAU DE
BACHAREL EM ENGENHARIA DE CONTROLE E AUTOMAÇÃO

Prof. Dr. Maurício Becerra Vargas
Coordenador

BANCA EXAMINADORA:

Prof. Dr. Márcio Alexandre Marques

Prof. Dr. Galdenoro Botura Júnior

Prof. Dr. Everson Martins

Junho de 2022

DEDICATÓRIA

de modo especial, à minha família que foram e são
minha base na busca por trilhar uma trajetória de
sucesso.

AGRADECIMENTOS

Ao Prof. Dr. Márcio Alexandre Marques que foi de grande suporte e responsável por grandes ensinamentos e dedicação desde o início da minha trajetória junto à Unesp Sorocaba até a realização deste trabalho de conclusão.

Ao Prof. Steven Frederick Durrant, pelo suporte em minha iniciação científica que me permitiu expandir o conhecimento aplicado à Engenharia.

Ao Instituto Nacional do Câncer pela disponibilização de forma pública e anonimizada dos dados utilizados como objeto de estudo do trabalho.

ARAÚJO, V. S. **Utilização de Ferramenta de *Business Intelligence* para Análise dos Registros de Cânceres nos Hospitais do Sistema Único de Saúde (SUS)**. 2022 (Bacharel em Engenharia de Controle e Automação) – Instituto de Ciência e Tecnologia - Câmpus de Sorocaba, Universidade Estadual Paulista, 2022.

RESUMO

O presente trabalho mostra as principais características e as tendências das ferramentas de *business intelligence* e uma aplicação, usando Tableau, a partir da base de dados do RHC (Registro Hospitalar do Câncer, gestão INCA). Além de estudar as diferentes ferramentas e suas tendências com base na visão da Gartner, foi feita uma avaliação dos benefícios de uma arquitetura robusta em nuvem de forma aplicada, identificando a qualidade dos dados da base e os benefícios da digitalização dos dados de gestão pública, principalmente focada em saúde. Para o desenvolvimento do caso de estudo, foi construída um *dashboard* por meio do Tableau Software integrado ao Microsoft MySQL. A partir do *dashboard* implementado, foram realizadas análises de características de ocorrências de algumas das neoplasias mais comuns em homens e mulheres no Brasil: próstata, mama, pulmões no Brasil. Os resultados mostraram a predominância de utilização de tabaco e álcool por homens, bem como as tendências de relação entre idade e a aparição das neoplasias. Conclui-se, que a implementação da ferramenta permite uma visualização e análise desses dados de forma rápida e flexível, a partir de filtros selecionados pelo usuário, analisando e visualizando os dados sob diferentes ópticas. Do ponto de vista da base de dados, a ferramenta permite ainda o acompanhamento da integridade da base de maneira contínua. Por fim, foram apresentados ainda os benefícios da digitalização dos bancos de dados em especial com uma visão integrada ao estado atual das bases de saúde no Brasil.

Palavras-chave: *Business Intelligence; Dashboard; Tableau; Neoplasias; INCA;*

ARAÚJO, V. S. Using Business Intelligence Tool to Analyze Cancer Registries in Hospitals of the Brazilian National Health System (SUS). 2022 (Bachelor's degree Automation and Control Engineering) – Science and Technology Institute – Sorocaba, UNESP 2022.

ABSTRACT

The paper presents main characteristics and trends of business intelligence tools and an application, using Tableau, from the RHC (Hospital Cancer Registry, INCA management) Brazilian National Cancer registers. Besides studying the different tools and their trends based on Gartner's vision, the benefits of a robust cloud architecture in an applied way were identified, by reviewing data's integrity and the benefits from the digitalization of public management data, mainly focused on health sector. For the development of the case study, a dashboard was built using Tableau Software integrated with Microsoft MySQL. From the implemented dashboard, analyses were performed on the occurrences' characteristics of the most common neoplasms in men and women in Brazil: prostate, breast, and lungs. The results showed the predominance of tobacco and alcohol use by men, as well as the tendencies of relationship between age and the appearance of neoplasms. We conclude, that the implementation of the tool allows a fast and flexible visualization and analysis, from filters selected by the user, analyzing and visualizing the data from different perspectives. From the database point of view, the tool also allows the monitoring of the database integrity in a continuous manner. Finally, the benefits of database digitalization were also presented, especially with an integrated view of the current state of health databases in Brazil.

Keywords: Business Intelligence; Dashboard; Tableau; Neoplasms;

LISTA DE FIGURAS

Figura 1: Arquitetura de BI na visão de Sherman	14
Figura 2: Quadrante mágico das plataformas <i>Business Intelligence</i> 2020	16
Figura 3: Visão parcial do <i>Dashboard</i> COVID-19 utilizando Tableau	22
Figura 4: Exemplo de utilização de Power BI Desktop.	30
Figura 5: Exemplo de <i>Explain Data</i> Tableau.....	31
Figura 6: Exemplo do mecanismo de busca do ThoughtSpot.....	33
Figura 7: <i>Analytics</i> e integração de dados de ponta a ponta.	34
Figura 8: Visão geral Integrador RHC	37
Figura 9: Visão geral TabWin (DataSUS)	38
Figura 10: Etapas de configuração SQL Server 2019.....	39
Figura 11: Propriedades de conexão do banco de dados local DESKTOP-4PLLNR 40	
Figura 12: Propriedades de conexão do banco de dados local DESKTOP-4PLLNR 41	
Figura 13: Seleção de conexão via Tableau	45
Figura 14: Informações de conexão ao servidor.....	46
Figura 15: Visão parcial dos dados importados ao Tableau	46
Figura 16: Edição de <i>aliases</i>	47
Figura 17: Visão parcial dos dados traduzidos	48
Figura 18: Esquemático dos objetivo do <i>dashboard</i>	49
Figura 19: Visualização da criação da planilha “Idade”	50
Figura 20: Visualização da criação da planilha “Idade”	51
Figura 21: Fluxograma de criação do <i>dashboard</i> no Tableau.....	53
Figura 22: Apresentação geral do <i>Dashboard</i> com informações do RHC	56
Figura 23: Distribuição proporcional dos casos analíticos.....	57
Figura 24: Tipos de câncer mais frequentes em homens e mulheres	57
Figura 25: Histórico de consumo alcóolico e tabaco	58

Figura 26: <i>Dashboard</i> – Neoplasia de Brônquios e Pulmões – Estado de SP	59
Figura 27: <i>Dashboard</i> – Neoplasia de Brônquios e Pulmões – Cidade de SP	60
Figura 28: <i>Dashboard</i> – Neoplasia de Brônquios e Pulmões – Estado de MG	61
Figura 29: Idade e etnia – Neoplasia de Brônquios e Pulmões – Estado de MG	62
Figura 30: Alcoolismo e tabagismo – Estado de MG	62
Figura 31: <i>Dashboard</i> – Neoplasia de Mama – Visão geral.....	64
Figura 32: <i>Dashboard</i> – Neoplasia de próstata – Visão geral	66

LISTA DE TABELAS

Tabela 1: Dicionário das variáveis dados do RHC	26
Tabela 2: Apresentação das planilhas criadas durante o desenvolvimento.....	52
Tabela 3: Número de casos analíticos registrados no RHC	54

SUMÁRIO

1. INTRODUÇÃO	9
2. REVISÃO BIBLIOGRÁFICA.....	11
2.1 <i>Big data</i>	11
2.2 Gestão de dados	12
2.3 <i>Business Intelligence</i> (BI)	13
2.4 Soluções de <i>BI</i> e o quadrante mágico Gartner	15
2.5 Tableau.....	22
2.6 Banco de dados	23
2.7 Bases de dados da saúde	24
2.8 Registros Hospitalares do Câncer (RHC)	25
3. DESENVOLVIMENTO	29
3.1 Metodologia	29
3.2 Comparação das ferramentas de <i>Business Intelligence</i>	29
a. Microsoft	29
b. Tableau	30
c. ThoughtSpot	32
d. Qlik	34
3.3 Análise de dados	35
3.4 Análise visual	36
3.5 Seleção e preparo do banco de dados	37
3.6 Integração do banco de dados	45
3.7 Estruturação do <i>Dashboard</i>	48
4. RESULTADOS.....	54
4.1 Visão geral	54
4.2 Câncer de brônquios e pulmões – Estado de São Paulo.....	58
4.3 Câncer de brônquios e pulmões – Estado de Minas Gerais	62

4.4 Câncer de mama – Brasil	63
4.5 Câncer de próstata – Brasil	65
4. CONCLUSÃO	67
REFERÊNCIAS	68

1. INTRODUÇÃO

Os diversos termos que agora ganham notoriedade relacionado à análise de dados são impulsionados pelo avanço e investimento em tecnologia de modo a suportar o crescimento dos negócios de forma competitiva. A taxa de crescimento na quantidade de informação disponível tornou a tomada de decisão cada vez mais complexa no que tange direcionamento de negócios. Os setores não estão mais em seus silos, mas sim totalmente integrados. É cada vez mais difícil identificar competidores, fornecedores, potenciais clientes e tendências. Dessa forma, as ferramentas de inteligência de negócios ganharam espaço, trazendo as análises de dados e o *Big Data* para o cotidiano dos gestores a partir da facilitação das análises e sua visualização. A combinação destes dois fatores torna o entendimento claro, e as tomadas de decisão mais assertivas e adotadas.

Segundo a Gartner [1], tanto *startups* como grandes empresas começaram a oferecer ao seu pessoal a possibilidade de usar os dados para tomar decisões embasadas, o que explica a adoção cada vez maior de soluções de *business intelligence*. *Business Intelligence* (BI) nada mais é do que a utilização da coleta, organização, análise e monitoramento de dados na tomada de decisão, trazendo visibilidade sobre variáveis acerca das decisões, permitindo assim que, independente do setor que esteja sendo utilizada o resultado esperado seja atingido.

Os resultados da análise visual profunda e eficiente são factíveis. Não faltam exemplos de empresas que foram impulsionadas pela adoção dessas ferramentas. Segundo com Michelle Wallace, o LinkedIn ao implementar uma estratégia de dados aumentou sua receita de serviços para empresas em 85%, ano após ano. Um componente importante dessa estratégia foi uma plataforma de análise de vendas nova e bastante simples que o LinkedIn desenvolveu com a Tableau [2].

Com o crescimento da procura e valorização das ferramentas de BI, grandes empresas mobilizaram-se em torno da criação e evolução de suas ferramentas. Segundo o quadrante mágico publicado em 2020 pela Gartner [1], divide-se o posto principal como líderes de mercado, as empresas Microsoft e a Tableau seguidas por Qlik e ThoughtSpot.

Apesar do termo *Business Intelligence* remeter ao ambiente corporativo, as ferramentas aplicam-se a diferentes contextos, sendo muitas vezes utilizadas pelo setor público. Um exemplo foi a utilização na identificação de alunos sob riscos de evasão escolar nas escolas públicas de “Des Moines” nos Estados Unidos [3]. A equipe de pesquisa e gerenciamento de dados da DMPS (*Des Moines Public School*) usou um modelo de regressão linear múltipla, que chamaram de “coeficiente de evasão”, para medir indicadores a fim de prever quais alunos poderiam estar sob o risco de evasão. Para potencializar o modelo, foi usada uma plataforma

de *business intelligence*. Com a visualização de dados, ficou mais fácil para a equipe identificar todos os alunos em risco e dedicar a atenção necessária a cada um deles [3].

Dr. Pat Hanrahan, cientista líder e cofundador da Tableau Software traz a seguinte definição para análise visual: “Começamos por aquilo que a análise visual não é: uma representação gráfica dos dados. Praticamente qualquer software pode criar um gráfico, um diagrama ou um painel. A análise visual oferece algo muito mais profundo. Análise visual é o processo de raciocínio analítico facilitado por interfaces visuais interativas.” [4]. Nos campos da Engenharia e, em específico em Controle e Automação, abordam-se inúmeras técnicas matemáticas para a resolução de problemas elétricos, mecânicos, da computação e de controle. De forma geral, tais técnicas trazem consigo o conceito de representações visuais por meio de gráficos que buscam facilitar a análise do problema ou apresentar o resultado final de seus cálculos. A análise visual é o processo que permite aproximar o público-alvo – o qual tende a ser pessoas de negócios com pouco conhecimento técnico e alto poder de decisão – da teoria que embasam as análises. A motivação deste trabalho tem, portanto, a conversão deste conteúdo muitas vezes teóricos à prática visual por meio da utilização de ferramentas de *Business Intelligence*. Assim, os objetivos desse trabalho são:

- i. Analisar as ferramentas de BI comerciais disponíveis no mercado;
- ii. Estudar as melhores práticas no que tange análises visuais;
- iii. Pesquisar e selecionar banco de dados brasileiro na área da saúde;
- iv. Analisar o banco de dados selecionado utilizando Tableau Prep;
- v. Desenvolver *dashboard* integrado em tempo real utilizando Tableau.

2. REVISÃO BIBLIOGRÁFICA

2.1 *Big data*

Vivemos na era dos dados. Se o grande advento da internet nos meados dos anos 90 foi conectar diferentes lugares do mundo, hoje temos a consolidação do avanço da conectividade e do poder de processamento culminando em diferentes tecnologias emergentes. Cada pequeno objeto conectado à internet é uma fonte de dados em potencial e a lista é extensa: celulares, *tablets*, computadores, *notebooks*, geladeiras, televisões, dispositivos inteligentes, lâmpadas [...] todos os objetos “smart” podem ser utilizados para compor uma grande massa de dados os quais tornam-se cada vez mais valiosos. É necessário portanto não só tecnologia, mas técnica para armazenar e processar todo esse volume. Não somente pela natureza do negócio, mas pela valorização e vantagem competitiva criada pela utilização dessas massas, as grandes companhias têm investido amplamente neste mercado.

‘Big Data’ era somente um grande volume de dados de forma dispersa que necessitava de processamento rápido para apropriada interpretação. Mais tarde, em 2012, incluiu uma nova perspectiva onde não somente a interpretação destes dados era importante, mas também a veracidade, variabilidade, visualização e valor [5]. Dessa forma, o termo deixa de ser somente definido por seu volume, mas por sua usabilidade, criando oportunidades para diferentes setores, tornando-se cada vez mais inerente a operação de grandes companhias.

Neste contexto, entram as grandes arquiteturas e técnicas de gerenciamento. Ora, se por um lado a regulamentação sobre esse mercado tem amadurecido, por outro a acirrada disputa comercial por experiências de usuário cada vez mais personalizadas faz grande valia deste setor. Não à toa grandes ferramentas têm se tornado cada vez mais relevantes no mercado como: Hadoop, Hive, Apache Mahout entre outros.

O grande desafio do *Big Data* – o qual é tipicamente composto por peta ou terabytes – é ser manuseado de forma teórica, técnica e tecnológica de forma eficiente de forma a gerar conclusões valiosas. Portanto, dada a complexidade deste contexto, torna-se ainda mais importante as ferramentas de *Business Intelligence*, gerando integração desde a origem dos dados até a sua visualização para tomadas de decisão, foco deste trabalho.

2.2 Gestão de dados

Dada sua importância e volume, a gestão de dados tem aumentado cada vez mais sua relevância. Tanto em termos de infraestrutura quanto de gestão, as equipes de tecnologia têm recebido cada vez mais aportes para criar ambientes robustos, íntegros, rápidos e seguros.

Em termos de estrutura, investimentos nas mais novas tecnologias de ponta não bastam para gerar vantagem competitiva, uma vez que estas podem ser rapidamente copiadas por competidores. No entanto, pesquisas recentes demonstram que aliar tecnologia de ponta com produtividade organizacional tornam as áreas de TI grande *driver* de ganho de *performance* [6], criando capacidades únicas. A utilização de tecnologia ultrapassada leva a três grandes problemas:

- a. **Segurança de dados:** como dito, os dados têm se tornado cada vez mais valiosos no mundo atual. Não à toa os governos têm criado legislações acerca do tema. No entanto, também tem crescido o número de ataques cibernéticos, e estes podem ser altamente custosos.
- b. **Perda de dados:** estruturas ultrapassadas tendem a dificultar a integração. A falta de integração leva a dificuldades de processo nativos pela infraestrutura das empresas, gerando assim setores que operam em silos. Além disso, o não investimento em integração de dados dificulta a aplicação de qualquer técnica de *Business Intelligence*, uma vez que a descentralização dos dados afeta a capacidade de gerar processamento de dados rápidos e precisos.
- c. **Produtividade:** considerando as tecnologias obsoletas e a dificuldade descrita no item anterior, há ainda a desmotivação causada nos colaboradores por conta da ineficiência de sistemas antigos. Assim, o contraste das novas gerações utilizando tecnologias de décadas passadas gera ruído na confecção de *insights* de ponta.

Dessa forma, observam-se não só os pontos positivos os quais são claros quando se discute a utilização das novas tecnologias de dados nos negócios, mas também é possível clarificar o impacto da não modernização.

2.3 Business Intelligence (BI)

Conforme visto anteriormente, o BI é um conjunto de conceitos e metodologias que suporta as tomadas de decisão de empresas e governantes a partir da conversão de grandes bases de dados em visualizações amigáveis e intuitivas. Segundo Angeloni e Reis [7], o conceito de *Business Intelligence* com o entendimento de que é a Inteligência de Negócios ou a Inteligência Empresarial, compõe-se de um conjunto de metodologias de gestão implementadas através de ferramentas de software, cuja função é proporcionar ganhos nos processos decisórios gerenciais e da alta administração nas organizações, baseada na capacidade analítica das ferramentas que integram em um só lugar todas as informações necessárias ao processo decisório. Reforça-se que o objetivo do *Business Intelligence* é transformar dados em conhecimento, que suporta o processo decisório com o objetivo de gerar vantagens competitivas [7].

Na visão de Batista, (2004, apud ANGELONI e REIS, 2006, p. 5) [7] as ferramentas de BI podem fornecer uma visão sistêmica do negócio e ajudar na distribuição uniforme dos dados entre os usuários, sendo seu objetivo principal transformar grandes quantidades de dados em informações de qualidade para a tomada de decisões. Através delas, é possível cruzar dados, visualizar informações em várias dimensões e analisar os principais indicadores de desempenho empresarial. Outro conceito muito utilizado para o BI é o de Barbieri [8] que define BI como: um guarda-chuva conceitual, visto que se dedica à captura de dados, informações e conhecimentos que permitam às empresas competirem com maior eficiência em uma abordagem evolutiva de modelagem de dados, capazes de promover a estruturação de informações em depósitos retrospectivos e históricos, permitindo sua modelagem por ferramentas analíticas. Seu conceito é abrangente e envolve todos os recursos necessários para o processamento e disponibilização da informação ao usuário.

Bezerra et al. [9] define os critérios básicos para seleção e apresentação das ferramentas, com base nos estabelecidos por Sá et al. (2012) [10], referem-se à *performance*, arquitetura, consultas e relatórios (agendamento, *dashboards*, exportação dos dados) para a organização. Alguns exemplos importantes de ferramentas de BI são apresentadas em conjunto ao seu desenvolvedor: Qlikview (QLIK, c1993-2020), que pertence à marca Qlik, Pentaho (PENTAHO..., c2005-2020) que é o mais popular e poderoso sistema de BI livre a nível mundial (BEZERRA et al., 2014) [9], BI Oracle (PÚBLIO, 2012); RM Bis.Net (TOTVS, 2012); BI Vision (PROCENGE, c2010-2015b); Power BI (MICROSOFT) e Tableau (TABLEAU) [10].

Sendo assim, diferentes organizações têm reconhecido a tecnologia e a informação como potenciador da tomada de decisão, incluindo aceleração e melhoria na resposta ao

mercado, aumento de eficiência, otimização de custos, centralização de dados e melhoria de processos.

Na literatura de Chou [11], as principais características das ferramentas de BI envolvem gestão, segurança, arquitetura, capacidade analítica, exploração visual de dados de forma interativa, compatibilidade com dispositivos móveis, capacidades analíticas avançadas, preparação de dados, publicação e compartilhamento. Todas essas características tendo como ponto comum a facilidade de utilização e a visualização da informação.

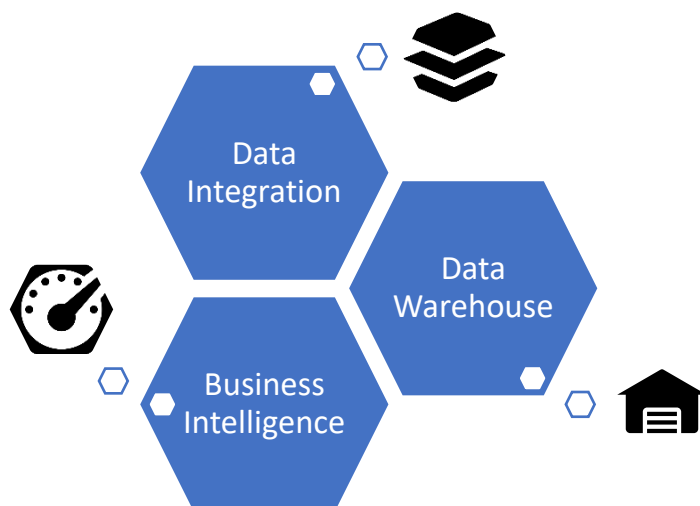
Por outro lado, com base no SAS Institute, o BI deve ser entendido como a capacidade de entregar a informação certa para a pessoa certa no tempo certo, para suportar melhores tomadas de decisão e ganhar vantagem competitiva. Esse conceito é fundamental no contexto onde as empresas visam cada vez mais atender às novas necessidades dos clientes aplicando e investindo em tecnologia de dados. No entanto, há necessidade de foco em resultados de forma pragmática tendo, portanto, o BI como ferramenta para atingir um (ou mais) objetivos e não como objetivo-fim. [12]

Em Sherman (2015) [12], a arquitetura de BI é composta por três grandes pilares:

- (i) *Data Integration (DI)*: aquisição e integração de dados.
- (ii) *Data Warehouse (DW)*: gestão e armazenamento de dados.
- (iii) *Business Intelligence (BI)*: processamento e informação dos dados.

A arquitetura e seu papel são apresentados pela figura 1.

Figura 1: Arquitetura de BI na visão de Sherman



Fonte: Autoria própria

Compreendendo o *Business Intelligence* como um sistema composto pelos três pilares definidos por Sherman, é possível entender que o *Data Integration* é o componente responsável

por adquirir os dados de entrada deste sistema. Quando fala-se de integração de dados, é importante imaginar diferentes sistemas utilizados nas empresas, por diferentes áreas, e que têm potencial ou compõe uma área estratégica dessa empresa. Dessa forma, para a criação de uma solução de *Business Intelligence* robusta, é necessário integrar dados de diferentes formatos, com diferentes fontes e periodicidade. Nesse contexto que entram as soluções de *ETL (Extract and Load)*. Tais soluções são utilizadas para manter a homogeneidade dos dados, extraindo informação de diferentes fontes, corrigindo-as para um mesmo formato e enviando para um centralizador e que podem ou não estar inseridas na solução de BI.

Entra então o segundo pilar, que é a gestão dos dados. O *Data Warehouse* diz respeito aos sistemas que centralizam as informações, ou seja, aos bancos de dados. Tais soluções são responsáveis por organizar e disponibilizar fácil acesso às informações, permitindo sua recuperação para utilização no futuro.

Por fim, as soluções de *Business Intelligence* se utilizam dos dois primeiros pilares e de seus recursos gráficos e de integração para permitir a criação de *Dashboards* e *Storylines* que enriquecerão a tomada de decisão com uma visão analítica robusta e de fácil entendimento. Além disso, permitem o compartilhamento dos *insights* gerados.

De uma maneira prática, as fontes de dados poderiam ser exemplificadas por sistemas CRM (*Customer Relationship Management*) – utilizados na gestão dos processos de vendas, ERP (*Enterprise Resource Planning*) – utilizados na gestão da cadeia produtiva e arquivos internos das áreas. Assim, as três fontes deveriam ser combinadas de modo a compor uma base robusta e homogênea. Seguindo esse exemplo, uma solução de *Business Intelligence* seria relacionar a demanda (vendas) com a oferta (produção). Sistemas que integram as diferentes variáveis e relacionam ambos processos são claros exemplos de como as soluções visuais de BI permitem uma rica visão sobre os possíveis caminhos.

2.4 Soluções de BI e o quadrante mágico Gartner

O desenvolvimento estratégico de empresas foi diretamente afetado pela ampliação da transformação digital. A implementação de soluções de BI foi peça chave para aumentar a eficiência e agilidade nas tomadas de decisão como resposta a variação constante dos diversos mercados. Nesse cenário, destaca-se o Gartner Group o qual é uma organização de consultoria com mais de 15 mil clientes no mundo e oferece serviços de tecnologia da informação ao redor do mundo. A Gartner não é somente um grupo de implementação de tecnologia, mas de análise e relatórios das principais tecnologias disponíveis em seu setor, sendo o seu principal o quadrante mágico das plataformas de *Analytics* e *Business Intelligence* apresentado na figura 2,

relatório no qual apresenta anualmente as principais tendências das ferramentas de BI de maneira gráfica considerando os pilares: compreensão de mercado, *marketing*, vendas, abordagem, modelos de negócios, estratégia e inovação.

Figura 2: Quadrante mágico das plataformas *Business Intelligence* 2020



Fonte: Gartner Group [13]

O quadrante mágico proporciona uma comparação gráfica dos provedores de tecnologia e serviços onde o mercado está crescendo e os diferentes concorrentes possuem diferenciais relevantes. A partir dos pilares de análise, as empresas de BI e seus respectivos produtos são classificados em Líderes, Visionários, Desafiante e Empresas de Nicho, sendo divididas em quatro quadrantes. Para estabelecer o posicionamento no gráfico, são utilizados quinze critérios relacionados às soluções que podem receber diferentes pesos de acordo com o mercado

avaliado. Por exemplo, características de alta relevância para empresas Desafiantes (como por exemplo a compreensão de mercado ou o modelo de negócios) recebe um peso maior do que a sua abordagem de vendas, uma vez que para essas empresas é mais relevante possuir um modelo de negócios e uma compreensão das tendências de mercado mais assertiva do que um modelo de vendas já estabelecido [14].

Dado os critérios de avaliação definidos, dentre as quatro classificações principais, os provedores ainda são posicionados considerando a capacidade de desenvolvimento de análises – que envolve velocidade, facilidade, recursos analíticos avançados, exploração visual, infraestrutura *cloud* e conectividade, segurança, manipulação dos dados e armazenamento. Neste quesito, quanto melhor o *player*, mais alto é posicionado considerando o eixo Y. O segundo macro critério de posicionamento é a visão de completude. Dado todo o poderio de análise, neste segundo quesito os provedores são posicionados em termos de integração dos conteúdos, publicação, compartilhamento e colaboração, facilidade de utilização, integração das ferramentas e *design* claro e atrativo [14].

O relatório “*How markets and providers are evaluated in Gartner Magic Quadrants*” [14] apresenta as definições das categorias utilizadas:

I. Completude da visão

- a. Compreensão do mercado: a habilidade do prestador de serviços em entender a necessidade dos clientes e traduzir essa necessidade em produtos e serviços. O vendedor deve apresentar grande nível de compreensão e entendimento do que o cliente quer e precisa, e usar essa informação para definir ou melhorar o relacionamento com o cliente.
- b. Estratégia de *marketing*: uma comunicação clara, diferenciada e consistente que passa por toda a forma com que a organização se posiciona de forma *online*, em anúncios, eventos com clientes e mídia.
- c. Estratégia de vendas: a estratégia para vender produtos ou serviços que utilizam de redes de vendas diretas ou indiretas, *marketing*, serviços e comunicações com filiados para estender o tamanho e a profundidade do funil de vendas, *skills* e *expertise*, tecnologia, serviços e base de clientes.
- d. Estratégia de produtos: a abordagem no desenvolvimento de produtos e serviços que enfatizam diferenciais, funções, metodologias e ferramentas em relação ao presente e aos requisitos futuros.

- e. Modelo de negócios: a validade e lógica da proposta de negócios do fornecedor frente ao mercado.
- f. Estratégia de indústria/vertical: estratégia para direcionar recursos, *skills* e produtos para atender as demandas de segmentos de mercados individuais, incluindo verticais da indústria.
- g. Inovação: concentração de recursos, *expertise* e capital para obter vantagem competitiva, investimento, consolidação ou se defender de aquisições.
- h. Geografia: estratégia para direcionar recursos, *skills* e produtos para atender necessidades fora da região origem do vendedor – de forma direta ou indireta a partir de parceiros e canais – de acordo com o apropriado para as diferentes regiões e mercados.

II. Habilidade para executar

- a. Produtos/serviços: principais produtos e serviços oferecidos pelo fornecedor que competem no mercado. Essa categoria inclui as capacidades dos produtos e serviços, qualidade, conjunto de características e habilidades, conforme definido no mercado.
- b. Viabilidade: avaliação da saúde financeira do fornecedor, do sucesso prático e financeiro dos negócios, e da probabilidade de a unidade seguir investindo e ofertando seu portfólio de produtos.
- c. Execução de vendas/preço: a capacidade do fornecedor nas atividades pré e de vendas e na sua estrutura que o suporta em seu mercado. Esse critério inclui gestão de negócios, precificação e negociação, suporte na venda e a efetividade do canal de vendas como um todo.
- d. Resposta ao mercado e histórico de vendas: capacidade do fornecedor de responder e mudar de direção, ser flexível e alcançar sucesso competitivo a medida que novas oportunidades se desenvolvem, as necessidades do cliente evoluem e a dinâmica do mercado muda. Também considera o quão responsivo o fornecedor tem sido ao longo do tempo.
- e. Execução de *marketing*: clareza, qualidade, criatividade e eficácia da execução de programas de *marketing* direcionados a entregar a mensagem do fornecedor ao mercado, promover a marca e os negócios, aumentar o conhecimento sobre produtos

e serviços e estabelecer uma identificação positiva do cliente com o produto, marca ou fornecedor. Podem incluir, entre outros elementos, a combinação de publicidade, promoções, influenciadores, suporte ao cliente de forma ativa e atividade de vendas.

- f. Experiência do cliente: relacionamentos, produtos, serviços e programas que permitem aos clientes terem sucessos com os produtos que estão utilizando. Esse critério inclui as formas pelas quais os clientes recebem suporte técnico ou administrativo. Também inclui as ferramentas adicionais, programas de suporte ao cliente e sua qualidade, disponibilidade de grupos de usuários e nível de serviço.
- g. Operações: a capacidade do fornecedor de cumprir seus objetivos e compromissos. Inclui a qualidade da estrutura organizacional, habilidades, experiências, programas e sistemas que permitem ao fornecedor operar de forma eficaz e eficiente.

Além dos macros critérios que visam compreender as capacidades do vendedor, também são analisados 15 critérios em relação a ferramenta de *Business Intelligence*.

III. Critérios específicos das ferramentas de BI

- a. Segurança: Capacidades que permitem a segurança da plataforma, administração dos usuários, auditoria de acesso e autenticação da plataforma.
- b. Capacidade de gerenciamento: Capacidades para rastrear o uso, gerenciar como a informação é compartilhada e por quem, realizar análises de impacto e trabalhar com aplicações de terceiros.
- c. Nuvem: A capacidade de apoiar a construção, implantação e gerenciamento de aplicações analíticas e analíticas na nuvem, com base em dados tanto na nuvem como no local, e através de implantações em várias nuvens.
- d. Conectividade com as fontes de dados: Capacidades que permitem aos usuários conectar-se e ingerir dados estruturados e não estruturados contidos em vários tipos de plataformas de armazenamento, tanto no local como na nuvem.
- e. Preparação de dados: Suporte para arrastar e soltar, combinação de dados de diferentes fontes orientada pelo usuário, e criação de modelos analíticos (como medidas, conjuntos, grupos e hierarquias definidos pelo usuário).
- f. Complexidade do modelo: Suporte a modelos de dados complexos, incluindo a capacidade de lidar com tabelas de fatos múltiplos, interoperar com outras plataformas analíticas e suportar implementações de gráficos de conhecimento.

- g. Catálogo: A capacidade de gerar e “curar” automaticamente um catálogo pesquisável dos artefatos criados e utilizados pela plataforma e suas dependências.
- h. Percepções automatizadas: Um atributo central da análise aumentada é a capacidade de aplicar técnicas *machine learning* para gerar automaticamente *insights* para os usuários finais (por exemplo, identificando os atributos mais importantes em um conjunto de dados).
- i. Análises avançadas: Capacidades analíticas avançadas que são facilmente acessadas pelos usuários, podendo ser contidas dentro da própria plataforma ABI (*analytics and business intelligence* – solução que combina modelos analíticos a plataforma de construção de *dashboards*) ou utilizáveis através da importação e integração de modelos desenvolvidos externamente.
- j. Visualização de dados: Suporte para painéis de controle altamente interativos e a exploração de dados através da manipulação de imagens de gráficos. Inclui uma variedade de opções de visualização que vão além das de gráficos de torta, barras e linhas, tais como mapas de calor e árvores, mapas geográficos, gráficos de dispersão e outros visuais de propósito especial.
- k. Consulta em linguagem natural: Isto permite que os usuários consultem dados usando termos comerciais que são digitados em uma caixa de busca ou falados.
- l. Contagem de histórias de dados: A capacidade de combinar a visualização interativa de dados com técnicas narrativas, a fim de empacotar e fornecer insights de forma atraente e de fácil compreensão para apresentação aos tomadores de decisão.
- m. Análises incorporadas: As capacidades incluem um SDK (*software development kit* – kit de desenvolvimento que permite a programação de soluções incorporadas) com APIs (*Application Programming Interface* – formato que simplifica integração de diferentes sistemas) e suporte para padrões abertos a fim de incorporar conteúdo analítico em um processo comercial, um aplicativo ou um portal.
- n. Geração de linguagem natural: A criação automática de descrições ricas de *insights* encontrados em dados. Dentro do contexto analítico, à medida que o usuário interage com os dados, a narrativa muda dinamicamente para explicar as principais descobertas ou o significado dos gráficos ou painéis de controle.
- o. Relatórios: A capacidade de criar e distribuir aos consumidores relatórios de *grid-layout*, com multi-páginas, e relatórios *pixel-perfect*.

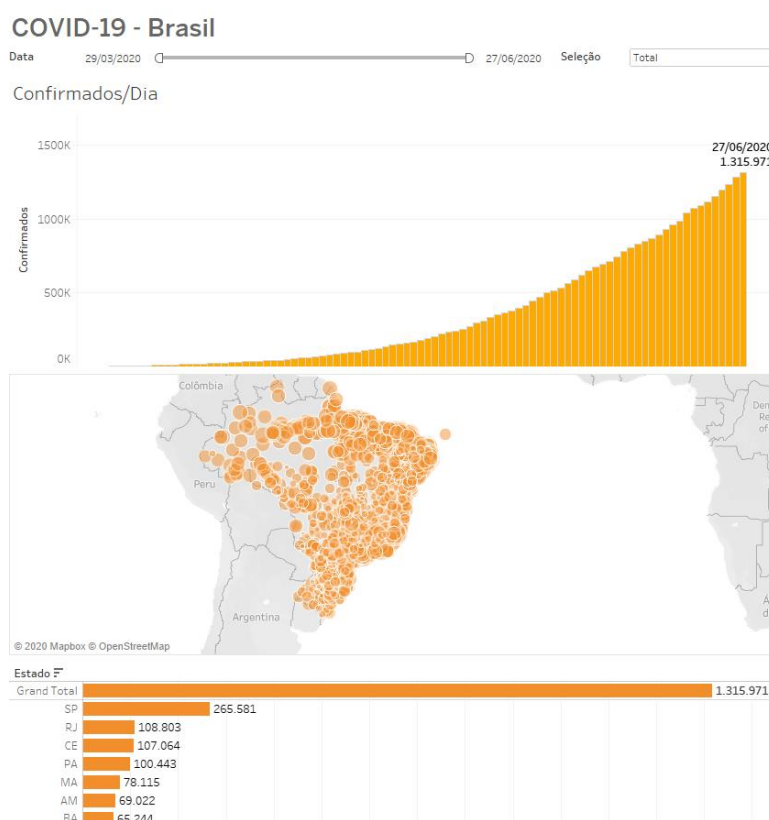
Todos os conceitos são analisados considerando cenários e exigências reais do mercado e as empresas analisadas considerando forças e fraquezas. Dessa forma, é possível compreender não só o atual posicionamento das soluções, como também tendências do setor.

2.5 Tableau

A consultoria de prestígio Gartner define o Tableau como líder no setor há mais de oito anos. Cada vez mais empresas comprovam e melhoram sua análise do fluxo de dados diariamente, gerando decisões e resultados mais assertivos a partir da utilização da ferramenta a qual ajuda a promover uma análise de dados confiável, rápida, segura e fácil. O grande diferencial das ferramentas líderes é o seu desacoplamento da área de TI para o gerenciamento de relatórios, o que é promovido dada a sua interface clara e intuitiva. Dessa forma, a criação dos relatórios fica a cargo da área de negócios, utilizando a área de TI como suporte de segurança e controle, evitando o acúmulo de pedidos para análise de dados, e gerando maior rapidez na identificação de tendências [14].

As soluções da Tableau permitem a criação de *dashboards* para análise profunda dos dados, integrando-os, manipulando e combinando-os de modo a criar análises visuais. Na figura 3, é possível observar um exemplo da combinação de diferentes bases de dados em relação à evolução da COVID-19 no Brasil em 2020. De forma interativa, o usuário pode selecionar o período e os estados em que deseja avaliar a evolução da doença.

Figura 3: Visão parcial do *Dashboard* COVID-19 utilizando Tableau



Fonte: Autoria própria

A solução Tableau é composta por três softwares: Tableau Desktop, Tableau Server e Tableau Prep. A versão *desktop* converte em segundos as consultas a qualquer banco de dados, em tempo real, possibilitando conexão direta com a fonte de informação ou a utilização de bancos de dados salvos em memória, desenvolvendo gráficos completamente visuais e intuitivos. Por outro lado, o Tableau Prep é a ferramenta de preparação das bases de dados relacionais. Já o Tableau Server permite compartilhar os *Dashboards* com as análises criadas. A ferramenta oferece ainda conexão com diferentes plataformas, podendo ser integrada às ferramentas Matlab, Azure, R, AWS. Também oferece a utilização de APIs [14].

De acordo com Elissa Fink, da Gartner, essa solução é intuitiva, orientada ao visual, com exploração dos dados interativa, enriquecendo a experiência do usuário. David Stone, gerente sênior da plataforma de *Analytics* do eBay, reforça o ponto de Fink. A companhia atualmente armazena 5 *Terabytes* de dados em seu *Data Warehouse (DW)* diariamente. Quando utilizava somente Excel, havia limitação de processamento e de dados, enquanto o Tableau permite a análise dessa grande base de forma intuitiva no cotidiano, o que enriqueceu fortemente as tomadas de decisão da empresa [15]. Quando se discute a análise de dados de forma massiva, é comum a confusão entre o poder de processamento das ferramentas. Tendo em vista a crescente demanda pela utilização de dados diários dos clientes nas tomadas de decisão, é fundamental soluções que manejem um grande volume de dados de forma rápida e com o mínimo esforço de processamento.

2.6 Banco de dados

Abraham [16] define um banco de dados “é uma coleção de dados inter-relacionados, representando informações sobre um domínio específico”, ou seja, quando é possível agrupar um conjunto de informações inter-relacionadas sobre um mesmo assunto, pode-se dizer que há um banco de dados. Sendo assim, um bando de dados nada mais é do que uma base que poderá ser consultada no futuro.

Os sistemas de gerenciamento de banco de dados (SGBD) são softwares que permitem a manipulação e armazenamento dos bancos de dados, além de interação com o usuário direta ou indiretamente. O principal objetivo de um SGDB é proporcionar um ambiente tanto conveniente quanto eficiente para a recuperação e armazenamento das informações do banco de dados. [16]

O Sistema Gerenciador de Banco de Dados disponibiliza recursos para definir, construir, manipular, compartilhar, proteger e manter bancos de dados. [17] Neste sentido, existem diferentes soluções que permitem o gerenciamento de banco de dados, sendo as principais

Oracle, SQL Server, PostgreSQL, MySQL, AWS e Access. Os sistemas de gerenciamento são, portanto, programados de modo a permitir o acesso, extração e manipulação por diferentes softwares terceiros. Cria-se assim uma interface entre os usuários e a informação.

2.7 Bases de dados da saúde

Conforme cita Luiz Felipe [18], de acordo com a lei 8.080 de 1990 que instituiu o Sistema Único de Saúde (SUS) no Brasil, é do Ministério da Saúde a atribuição legal de organização o Sistema Nacional de Informação em Saúde (SNIS). Entretanto, na década de 2000, a concepção de uma Política Nacional de Informação e Informática em Saúde era de que não se obteve sucesso em sua implementação [18]. Por outro lado, em ações mais recentes, em específico na Política Nacional de Informação e Informática (PNII) publicada em 2016 [19], instituiu-se o desenvolvimento da digitalização da saúde integrada às ferramentas disponibilizadas no e-Gov (programa de digitalização de soluções políticas responsável, por exemplo, pela criação de CNH/RG digital etc.). Apesar da dificuldade brasileira na digitalização da saúde, é tendência recente com exemplos internacionais e também nacionais a adoção de tecnologias de dados que integrem diferentes áreas da saúde, principalmente no setor privado. Muito além das grandes empresas do setor da saúde, surge o termo *Healthtech* para categorizar as *startups* que investem em tecnologia e saúde como seu modelo de negócio [20].

Considerando a dificuldade de integração e centralização de todos os dados de saúde de um país, surgem diferentes iniciativas que integram não só conceitos acadêmicos, mas também práticos. As bases de dados em saúde são ferramentas de extrema importância para uma contínua formação dos profissionais dessa área. Citam-se as mais relevantes:

- a. Portal de periódicos da Capes;
- b. *Medical Literature Analysis and Retrieval System Online (MEDLINE)*;
- c. Literatura Latino-Americana e do Caribe em Ciências da Saúde (LILACS);
- d. *Scientific Electronic Library Online (SciELO)*.

Os quatro portais são referências na integração de dados e artigos por conta da elevada exigência para publicação. Além de oferecerem uma base de pesquisa robusta, facilitando os estudos nessa área.

Com o objetivo de encontrar uma base prática para a implementação das análises propostas nesse trabalho, foram realizadas pesquisas tanto no portal de periódicos da Capes quanto na SciELO. Em ambas plataformas, encontramos um número significativo de trabalhos que faziam referência ao Registro Hospitalar de Câncer (RHC). Dessa forma, optou-se por esse banco de dados para realizar o estudo prático deste trabalho.

2.8 Registros Hospitalares do Câncer (RHC)

Os Registros Hospitalares do Câncer (RHC) são centros de coleta, armazenamento, processamento e análise - de forma sistemática e contínua - de informações de pacientes atendidos em uma unidade hospitalar, com diagnóstico confirmado de câncer. A informação produzida em um RHC permite o monitoramento da assistência prestada ao paciente. Sua principal função é clínica, sendo um recurso para acompanhar e avaliar a qualidade do trabalho realizado nos hospitais, incluindo os resultados no tratamento do câncer. Para consolidação das informações, a maioria dos RHC utilizam o SisRHC, sistema para informatização dos dados, desenvolvido e disponibilizado pelo INCA [21]. Os dados enviados ao SisRHC passam por um processo de validação – manual por um responsável além de uma crítica automatizada pelo sistema para garantir que as informações estão respeitando a codificação da base de dados. O SisRHC trabalha com um banco de dados traduzido, ou seja, para cada campo – ou coluna – a base considera uma informação numérica de forma a resumir a informação. Por exemplo, para o campo gênero, 1 representa Masculino, 2, representa Feminino. A prática é adotada de forma a comprimir as informações e reduzir o tamanho do banco de dados. A crítica realizada pelo SisRHC é realizada em cima dos campos chamados obrigatórios, de forma a garantir que uma opção válida foi inserida. A sistemática implementada prevê o envio trimestral das informações.

As bases de dados, consolidadas segundo o ano da primeira consulta realizada no hospital informante, são enviadas para compor a base nacional dos registros hospitalares de câncer, sob a guarda do INCA, por meio do IntegradorRHC [21].

O funcionamento de um RHC e o envio regular dos dados para o IntegradorRHC é obrigatório para os hospitais habilitados na Atenção Especializada em Oncologia do SUS e, facultativo, para os hospitais não habilitados [21].

A base de dados apresentada na tabela 1, tem o objetivo de cadastrar casos analíticos e não analíticos atendidos em diferentes hospitais localizados no Brasil. É fundamental ressaltar que o número de casos presentes na base do RHC não refletem a totalidade dos novos casos de câncer diagnosticados entre os residentes de cada estado, não devendo, portanto, ser usados para o cálculo de taxas de incidência de câncer. Os casos definidos como analíticos, objeto principal do RHC, referem-se aos pacientes que chegam aos hospitais sem tratamento oncológico prévio, com diagnóstico de câncer ou não [21]. Os dados são armazenados por *aliases*, ou seja, em formato de números de forma a economizar espaço de armazenamento e devem ser traduzidos a partir do domínio de cada campo.

Tabela 1: Dicionário das variáveis dados do RHC

Dicionário das variáveis da base de dados do SisRHC disponível para download no IRHC (atualizado 30/03/2020)			
Nº do campo na ficha de cadastro	Variável	Descrição	Domínio
44	ALCOOLIS	Histórico de consumo de bebida alcoólica	1.Nunca; 2.Ex-consumidor; 3.Sim; 4.Não avaliado; 8.Não se aplica; 9.Sem informação
22	ANOPRIDI	Ano do diagnóstico	aaaa
42	ANTRI	Ano da triagem	dd/mm/aaaa
base de dados SP	BASDIAGSP	Base mais importante para o diagnóstico do tumor	1.Exame clínico 2.Recursos auxiliares não microscópicos 3.Confirmação microscópica 4.Sem informação
	BASMAIMP	Base mais importante para o diagnóstico do tumor	1.Clínica; 2.Pesquisa clínica; 3.Exame por imagem; 4.Marcadores tumorais; 5.Citologia; 6.Histologia da metástase; 7.Histologia do tumor primário; 9. Sem informação
47	CLIATEN	Clínicas do primeiro atendimento - entrada do paciente	Codificação segundo Tabela de Clínicas do SisRHC
31	CLITRAT	Clínica de início do tratamento	Codificação segundo Tabela de Clínicas do SisRHC
SIS	CNES	Número do CNES do Hospital	Codificação segundo tabela do Cadastro Nacional de Estab. de Saúde
32	DATAINITRT	Data do início do primeiro tratamento específico para o tumor, no hospital	dd/mm/aaaa
36	DATAOBITO	Data do óbito	dd/mm/aaaa
21	DATAPRICON	Data da 1ª consulta	dd/mm/aaaa
23	DIAGANT	Diagnóstico e tratamento anteriores	1.Sem diag./Sem trat.; 2.Com diag./Sem trat.; 3.Com diag./Com trat.; 4.Outros; 9. Sem informação
22	DTDIAGNO	Data do primeiro diagnóstico	dd/mm/aaaa
32	DTINITRT	Ano do início do primeiro tratamento específico para o tumor, no hospital	aaaa
21	DTPRICON	Ano da 1ª consulta	aaaa
42	DTTRIAGE	Data da triagem	dd/mm/aaaa
28a	ESTADIAG	Estadiamento clínico do tumor (TNM) - Grupo	Codificação do grupamento do estágio clínico segundo classificação TNM
28a	ESTADIAM	Estadiamento clínico do tumor (TNM)	Codificação do grupamento do estágio clínico segundo classificação TNM
17	ESTADRES	UF de procedência (residência)	Sigla da UF de procedência
41	ESTCONJ	Estado conjugal atual	1.Solteiro; 2.Casado; 3.Viúvo; 4.Separado judicialmente; 5.União consensual; 9.Sem informação
35	ESTDFIMT	Estado da doença ao final do primeiro tratamento no hospital	1.Sem evidência da doença (remissão completa); 2.Remissão parcial; 3.Doença estável; 4.Doença em progressão; 5.Suporte terapêutico oncológico; 6. Óbito; 8. Não se aplica; 9. Sem informação
48	EXDIAG	Exames relevantes para o diagnóstico e planejamento da terapêutica do tumor	1.Exame clínico e patologia clínica; 2.Exames por imagem; 3.Endoscopia e cirurgia exploradora; 4.Anatomia patológica; 5.Marcadores tumorais; 8.Não se aplica; 9. Sem informação
43	HISTFAMC	Histórico familiar de câncer	1.Sim; 2.Não; 9.Sem informação
SIS	IDADE	Idade na 1ª consulta (calculada pela diferença entre a data da 1ª consulta e a data do nascimento)	Idade, em anos; valor igual a zero para crianças menores de 1 ano
11	INSTRUC	Escolaridade	1.Nenhuma; 2.Fundamental incompleto; 3.Fundamental completo; 4.Nível médio; 5.Nível superior incompleto; 6.Nível superior completo; 9.Sem informação
50	LATERALI	Lateralidade do tumor	1.Direita; 2. Esquerda; 3.Bilateral; 8.Não se aplica; 9.Sem informação
9	LOCALNAS	Local de nascimento	Sigla da UF de nascimento
25	LOCTUDET	Localização primária (Categoria 3d)	Código da CID-O, 3 dígitos
25	LOCTUPRI	Localização primária detalhada (Subcategoria 4d)	Código da CID-O, 4 dígitos
49	LOCTUPRO	Localização provável do tumor primário (somente para os casos em que a localização primária do tumor é desconhecida)	CID-O, 4 dígitos
51	MAISUMTU	Ocorrência de mais um tumor primário	1.Não; 2.Sim; 3.Duvidoso
MUUH	MUUH	Município da unidade hospitalar	Tabela de municípios do IBGE
12	OCUPACAO	Ocupação principal	Codificação pela Tabela da Código Brasileiro de Ocupações; mais de três 9 representa Ocupação ignorada
46	ORIENC	Origem do encaminhamento	1.SUS; 2.Não SUS; 3.Veio por conta própria; 8.Não se aplica; 9. Sem informação
28b	OUTROESTA	Outros estadiamentos clínicos do tumor	Codificação do grupamento do estágio clínico segundo outras classificações que não a TNM
34	PRITRATH	Primeiro tratamento recebido no hospital	1.Nenhum; 2. Cirurgia; 3.Radioterapia; 4.Quimioterapia; 5.Hormonioterapia; 6.Transplante de medula óssea; 7.Imunoterapia; 8.Outras; 9.Sem informação
13	PROCEDEN	Código do Município de procedência (residência)	Tabela de municípios do IBGE
10	RACACOR	Raça/cor	1.Branca; 2.Preta; 3.Amarela; 4.Parda; 5.Indígena; 9.Sem informação
33	RZNTR	Principal razão para a não realização do tratamento antineoplásico no hospital	1.Recusa do tratamento; 2.Tratamento realizado fora; 3.Doença avançada, falta de condições clínicas ou outras doenças associadas; 4.Abandono do tratamento; 5.Complicações de tratamento; 6.Óbito; 7.Outras razões; 8.Não se aplica; 9. Sem informação
6	SEXO	Sexo	1. Masculino; 2. Feminino
45	TABAGISM	Histórico de consumo de tabaco	1.Nunca; 2.Ex-consumidor; 3.Sim; 4.Não avaliado; 8.Não se aplica; 9.Sem informação
26	TIPOHIST	Tipo histológico do tumor primário	Codificação da morfologia do tumor pela CID-O
27	TNM	TNM	Codificação do estágio clínico segundo classificação TNM
38	TPCASO	Tipo de caso	1. Sim (Analítico); 2. Não (Não analítico)
SIS	UFUH	UF da unidade hospitalar	Sigla da Unidade da Federação da unidade hospitalar (IBGE)
SIS	VALOR_TOT	Text	-

Fonte: [21]

O SisRHC é um sistema gerido pelo DataSUS – órgão de tecnologia associado ao Ministério da Saúde. Esse núcleo é responsável pelo desenvolvimento de uma série de ferramentas associada ao acesso à informação da população. Uma das ferramentas utilizadas é o TabWin – ferramenta que permite a análise de diferentes banco de dados da saúde de gestão do Ministério. A ferramenta foi desenvolvida em 2010, conecta à diferentes FTPs e gera arquivos no formato TabWin – espécie de arquivo semelhante aos arquivos .CSV (*comma separated values file* – dados separados por vírgula) formato comumente associado ao Microsoft Excel [21]. Apesar de um movimento recente do DataSUS na disponibilização de dados em tempo real *online* para diversos banco de dados da saúde (como por exemplo a carteira de vacinação ou o acompanhamento de câncer de mama), atualmente a única forma de acesso ao banco de dados completo do RHC (incluindo diferentes patologias) é por meio do TabWin.

Como apresentado pela tabela 1, o banco de dados coleta uma série de informações associada aos pacientes de câncer de forma anonimizada e compreendendo os diferentes estados. A base contém desde informações sobre o diagnóstico – como tipo de câncer, método utilizado para o diagnóstico, clínica de tratamento, data de diagnóstico e etc. – à informações relacionadas ao paciente como gênero, idade, histórico familiar de câncer, tabagismo, alcoolismo.

A partir do manual disponibilizado pelo Ministério da Saúde acerca do RHC, o coordenador do RHC é responsável por avaliar a estrutura das informações recebidas dos diferentes hospitais de forma a garantir que o banco de dados esteja “limpo”, ou seja, com informações que respeitam o formato e domínio adequado.

Foram incluídas críticas no próprio SisRHC, para facilitar o processo de validação, que identificam, durante o processo de digitação de cada caso e no momento da solicitação de gravação do caso no banco de dados, as inconsistências e os casos que não se enquadram nas regras gerais de coerência de casos de câncer pré-estabelecidas. Algumas das críticas são referentes apenas às opções que constam em cada item das fichas, com o objetivo de evitar a inclusão de dados inexistentes para aquele item específico [22].

Como exemplo, em relação ao item sexo, só são aceitas as opções 1 – Masculino, 2 – Feminino. Para outros itens, no entanto, levam-se em consideração o cruzamento de determinados itens com outros da ficha. Isso ocorre na crítica que correlaciona sexo com topografia (não aceita tumor de testículo em mulheres ou tumor de colo uterino em homens) ou, ainda, quando relaciona topografia e histologia, entre outras críticas [23].

Ainda segundo o manual, apesar do sistema contar com críticas às informações inseridas, é de responsabilidade do coordenador do RHC que aplique o entendimento das regras de cadastro

das informações no RHC de forma a realizar uma análise crítica, uma vez que situações não previstas pela crítica do sistema pode acontecer durante o preenchimento das informações.

3. DESENVOLVIMENTO

3.1 Metodologia

A conceituação teórica, terá início no entendimento do mercado e das soluções disponíveis, estudando o último relatório do quadrante mágico da Gartner, analisando essencialmente como essas soluções são avaliadas, compreendendo os pontos mais relevantes com foco nas quatro soluções consideradas líderes pelo relatório. Os líderes serão estudados pois são as empresas que se destacam com as soluções do presente e que mais se adequam com as tendências de mercado para o futuro. Consequente, os conceitos de análise visual profunda serão abordados utilizando o próprio software Tableau como exemplo, descrevendo os métodos e comparando-os com contraexemplos.

Em termos práticos, a solução de *dashboard* será também desenvolvida utilizando Tableau. Neste contexto, é necessário selecionar e preparar o banco de dados para possibilitar a sua criação. Como foco deste estudo, será utilizado um banco de dados na área da saúde que permita explorar os principais recursos de um *dashboard*, afim de aplicar os conceitos aprendidos até então. Por fim, será estruturada o *dashboard* que permitirá uma análise visual e interativa do banco de dados selecionado.

3.2 Comparação das ferramentas de *Business Intelligence*

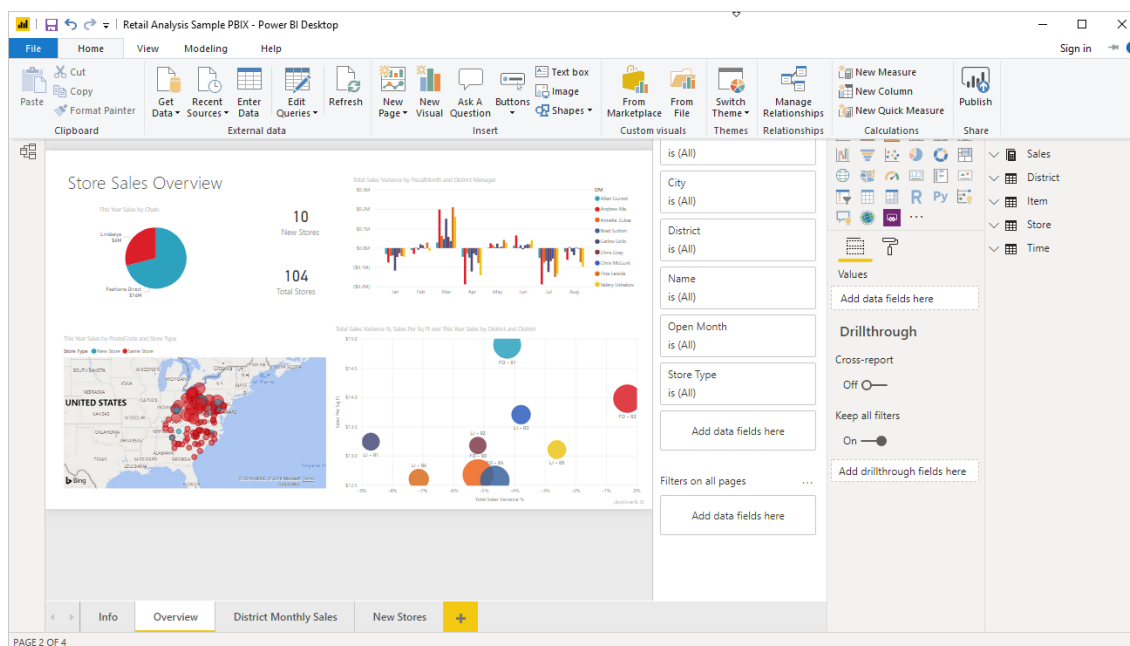
O princípio de seleção das ferramentas de BI é variado e muitas vezes está relacionado não somente às características dos softwares como também do posicionamento estratégico e parcerias das empresas. O quadrante mágico da Gartner busca proporcionar uma visão externa às estratégias, permitindo com que as ferramentas sejam avaliadas por seus pontos fortes, buscando proporcionar uma visão não enviesada. Com base nesse quadrante, quatro fornecedores se destacam como líderes: Microsoft, Tableau, Qlik e ThroughSpot.

a. Microsoft

De acordo com o quadrante de 2020 [3], a Microsoft possui uma linha de produtos compreensivo e visionário, além de uma capilaridade de mercado a partir do canal do Microsoft Office. São ofertadas ao cliente preparação de dados, descoberta de dados visuais, *dashboards* interativos e análises avançadas utilizando sua ferramenta chamada PowerBI. Um ponto importante na Microsoft é sua capacidade SaaS (*Software as a Service*) onde seus produtos podem ser utilizados a partir de um navegador e contratados como se fossem um serviço por assinatura. A instalação do software PowerBI Desktop – apresentado pela figura 4 - é necessária nos casos em que dados mais complexos estão sendo tratados. Ainda segundo o relatório, a

Microsoft possui como forças seu estímulo viral, a partir da integração com os produtos Office que são amplamente utilizados no mercado. A capacidade de utilização na nuvem, onde quase todas as análises podem ser realizadas sem a necessidade da utilização de um software instalado. Além do desenvolvimento de tecnologias emergentes, como a utilização de serviços com inteligência artificial para leitura de textos, sentimentos e análise de imagens.

Figura 4: Exemplo de utilização de Power BI Desktop.



Fonte: [24]

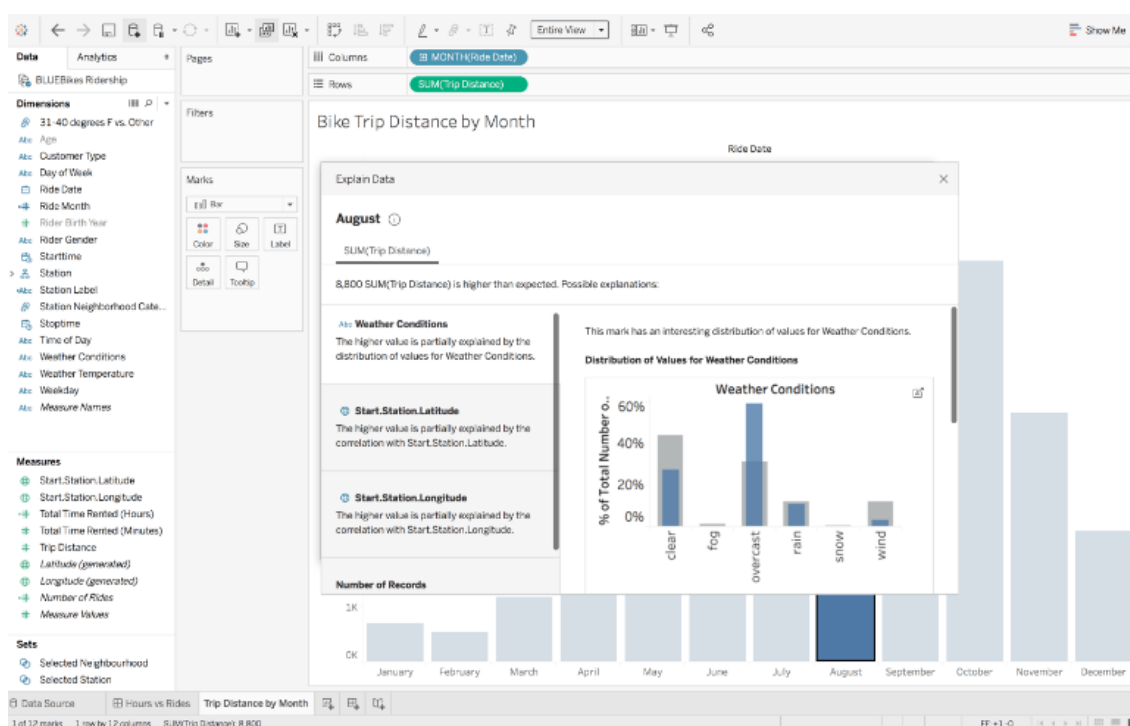
Como pontos de atenção se destacam a ausência de recursos importantes e da fluidez na utilização da ferramenta SaaS, incluindo *déficits* funcionais de *dashboard* e de análise, além das ferramentas de linguagem natural e alertas. Outro ponto de atenção importante é que a Microsoft não permite ao cliente utilizar o sistema *cloud*, adotando a utilização do Azure (também da Microsoft) o que restringe o cliente às soluções da empresa. Além disso, apesar da ferramenta oferecer uma ampla escala de fontes de dados, a ferramenta oferece problemas na conexão com SAP e HANA [14].

b. Tableau

O Tableau oferece uma experiência de exploração visual que permite os negócios acessar, preparar, analisar e apresentar conclusões de grandes massas de dados. A empresa possui uma estratégia de *marketing* poderosa e em expansão, mas há um certo receio em relação a recente aquisição pela empresa Salesforce.

Quanto aos softwares, a empresa tem investido fortemente no aumento do escopo de atuação, particularmente na utilização de análise analíticas avançadas e capacidade de governança de dados. Em relação às análises, o software possui as ferramentas “Ask data” e “Explain Data” que provém ao usuário em linguagem natural *insights* identificados pelo software automaticamente utilizando inteligência artificial para analisar as tendências. Um exemplo é apresentado na figura 5. Já em relação à governança de dados, o Tableau melhorou o Tableau Prep Builder para incorporar funções de agendamento e monitoramento de gestão de dados [24].

Figura 5: Exemplo de *Explain Data* Tableau.



Fonte: [25]

Como forças, o relatório da Gartner [14] destaca o entusiasmo da base de clientes, que conta com um gerenciamento de comunidade relevante no mercado contando com usuários que são chamados “champions” e estão espalhados pelo mundo, ofertando conteúdo da Tableau em diferentes canais. Além disso, o Tableau oferece uma fácil e rápida exploração e manipulação de dados de forma muito intuitiva, o que enriquece muito a experiência em relação à competidores.

Existem riscos em relação às mudanças do mercado, uma vez que é observada uma tendência de mercado para a utilização de ferramentas com inteligência artificial mais do que

foco em utilização visual, no entanto a Tableau já demonstrou estar equilibrando as prioridades do presente com as do futuro.

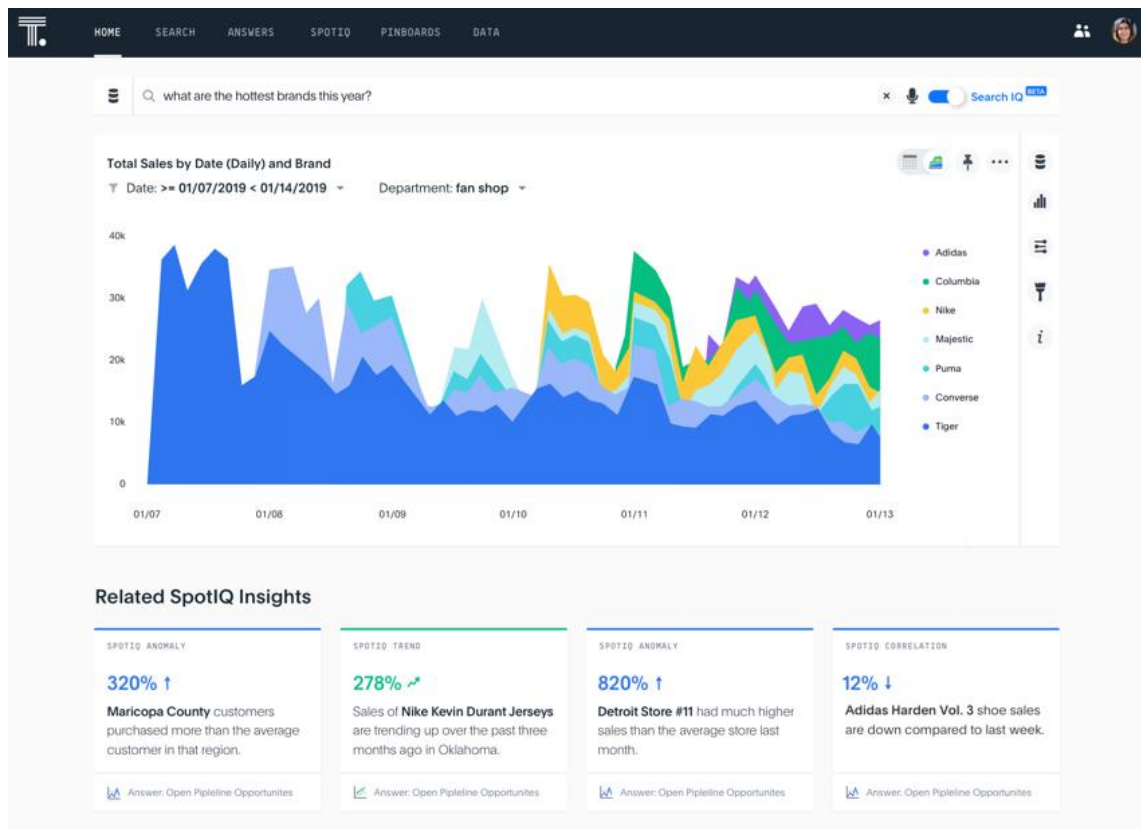
c. ThoughtSpot

O grande diferencial da ferramenta é baseado na mais recente tendência das ferramentas de BI: a utilização de inteligência artificial para a análise de dados. Através da ferramenta, é possível realizar perguntas ao software a respeito da base de dados que está sendo analisada com suporte a perguntas analíticas complexas [14]. O ThoughtSpot também utiliza da mesma ferramenta para fazer sugestões de análises ao usuário, o que enriquece os *insights*. Um outro ponto importante é sua constante busca por compatibilidade com diferentes *players* de base de dados, tendo recentemente incorporado as grandes nuvens como opção para importar dados: Amazon Redshift, Google BigQuery e Microsoft Azure.

Como destaques, o software apresenta a sua ferramenta de pesquisa utilizando linguagem natural como interface primária para consultar dados. Ou seja, o software transforma perguntas digitadas ou faladas em consultas ao banco de dados. Essa ferramenta suporta análises complexas e com uma grande base (maior de 1TB). Um exemplo é apresentado na figura 6, onde o cliente digita quais são as marcas mais “quentes” do ano e a ferramenta interpreta que o cliente gostaria de entender quais são as marcas que mais venderam no ano.

A ferramenta de análises complexas SpotIQ descobre anomalias, correlações e realiza análises comparativas sem a necessidade de programação a partir da inteligência artificial que constantemente analisa os dados inseridos na ferramenta. Um segundo ponto importante na companhia é a orientação ao consumidor e uma visão de produto direcionada a incorporação de inteligência artificial. De acordo com o relatório da Gartner [14], a ferramenta está buscando criar uma relação similar ao Facebook dentro de sua plataforma, incluindo *feeds* de conteúdos relacionados, automação de identificação de anomalias e alertas proativos. Como terceiro ponto forte, se destaca seu posicionamento de mercado, uma vez que apesar de uma companhia relativamente pequena, seus produtos tem seguido as tendências mais solicitadas pelos usuários do mercado.

Figura 6: Exemplo do mecanismo de busca do ThoughtSpot



Fonte: [26]

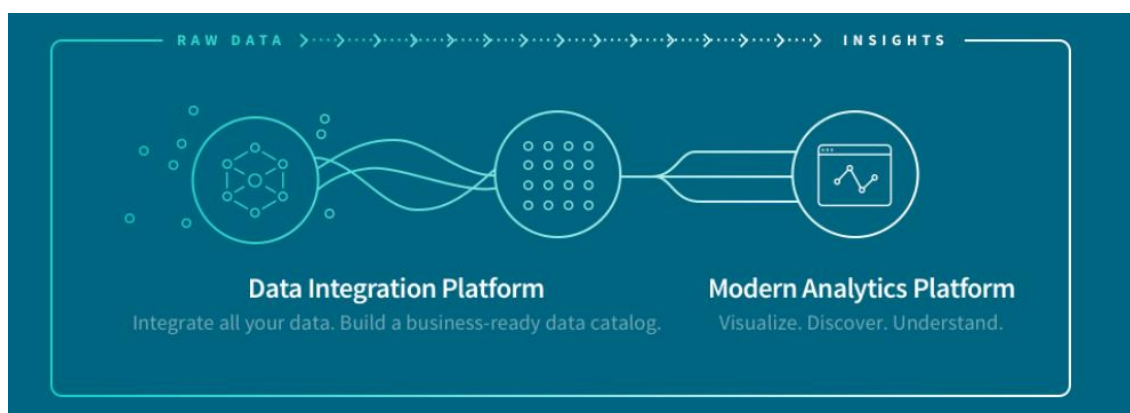
Como pontos fracos, se destacam os *gaps* na preparação dos dados, o alcance global limitado e a necessidade de uma estrutura de TI revisada antes da implementação da ferramenta já que essa é incompleta e não contempla o circuito todo de Data Integration e Data Warehouse [14]. Os pontos fracos da ferramenta a colocam em risco em relação os seus competidores, uma vez que a tendência é que as companhias busquem por parceiros que contemplem todo este circuito e é justamente o caminho em que o Tableau e a Microsoft vem trilhando nos últimos anos.

d. Qlik

De acordo com a Gartner [3], Qlik é o quarto líder em *Business Intelligence*. Sua ferramenta conta com foco na estrutura de microserviços e capacidade multinuvem, tendo a melhor experiência SaaS do mercado. Além de sua estrutura focada em utilização na nuvem, as soluções da Qlik também contam com um motor que utiliza de inteligência artificial e aprendizado de máquina para oferecer *insights* contextualizados e uma análise profunda dos dados. Essa ferramenta permite que usuários de todos os níveis combinem informações sem as limitações conhecidas das ferramentas baseadas em consultas (*queries*) que exigem conhecimento de sintaxe relacional.

Como pontos fortes se destacam a flexibilidade de desenvolvimento, sendo a Qlik um dos primeiros fornecedores a oferecer uma experiência fluida aos clientes e gestão a partir de diferentes nuvens. A flexibilidade de desenvolvimento e de arquitetura é um diferencial competitivo importante, e a Qlik tem focado fortemente em sua oferta de SaaS para facilitar as implementações. Nesse sentido, seu portfólio conta com um número vasto de soluções que abrangem todo o ciclo de vida de análise de dados. O Qlik Sense entrega ao cliente análises visuais profundas tanto para analistas de dados quanto para usuários de negócio e também comportam capacidade de desenvolvimento numa mesma plataforma. Já o Qlik Data Catalyst é utilizado para catalogar e desenvolver uma governança de dados dentro da organização. Além disso, o Qlik Data Integration Platform é uma oferta que adiciona um poderio de integração e movimentação de dados nesse ambiente. Assim, as ferramentas da Qlik cobrem desde o *Data Warehouse* e a integração dos dados até o usuário final como sugere a figura 7 [27].

Figura 7: *Analytics* e integração de dados de ponta a ponta.



Fonte: [27]

Como pontos fracos, se destacam um período de realinhamento de modelo de negócios que a empresa sofreu, passando por cortes de pessoas em 2019. Além disso, recentemente a

empresa tem focado em mover seus clientes da ferramenta QlikView (legado) para o Qlik Sense, no entanto tem sido observada uma resistência nessa transição. Por fim, embora a ferramenta seja desenvolvida para suportar uma análise visual, os clientes tem relatado que a ferramenta tem sido utilizada para realização de *dashboards* parametrizados, ou seja, que seguem sempre um mesmo modelo, limitando a capacidade de compreensão de novos *insights*.

3.3 Análise de dados

Existem diferentes termos que se referem às ferramentas de análise de dados. Um deles e mais comum é o *advanced analytics* o qual se refere às técnicas e ferramentas que vão além das ferramentas tradicionais utilizadas por negócios. Uma definição mais prática seria a compreensão de técnicas matemáticas – como séries temporais, previsão, clusterização, regressão linear – incumbidas em softwares para utilização no dia-a-dia, de forma prática.

As ferramentas de *Business Intelligence* focam em trazer soluções intuitivas para a implementação aplicação das ferramentas de análise de dados clássicas. Dessa forma, torna-se possível realizar análises rápidas e flexíveis – com a utilização de diferentes variáveis, filtros dinâmicos, diferentes cenários e dimensões. Assim, é possível visualizar os dados sob diferentes perspectivas.

O grande diferencial da análise de dados nas ferramentas de BI não está na aplicação do conceito, e sim na facilidade de criar os “por quês”, permitindo até com que não especialistas testem os cenários. A exemplo do Tableau, existem soluções incorporadas que combinam agregações básicas de dados às análises estatísticas como como desvio padrão, moda, assimetria (análise de frequência), enviesamento, correlação e covariância. Uma das ferramentas mais tradicionais da estatística são as ferramentas de análise de tendências. Como exemplo, podemos citar aquelas que, a partir de dados de temperatura e precipitação, visam estimar as condições climáticas.

Dentre as ferramentas de análise de tendência, há um destaque para diferentes formatos (linear, logarítmico, exponencial, potência e polinomial). Neste trabalho, para análise da tendência de casos, os dados foram ajustados ao **modelo linear**:

$$y = a(x) + b$$

Onde:

x – data de diagnóstico;

y – número de casos;

b – é o valor da reta ajustada para o tempo $t = 0$;

a – é o coeficiente angular

A tendência da série analisada foi obtida pelo comportamento da reta, onde:

$a > 0$ – tendência positiva

$a < 0$ – indica tendência negativa

3.4 Análise visual

Um dos grandes benefícios das ferramentas de *Business Intelligence* é o foco em garantir uma análise visual dos dados. Essa análise é dada a partir de painéis ou relatórios que permitem identificar perguntas “como”, por exemplo, “Quais foram os estados com maior incidência de casos de COVID-19?”. Essas perguntas ajudam a orientar as discussões, e são apresentadas de forma visual dentro dos *dashboards*. Dessa forma, a história a ser contada com os dados é facilitada, permitindo com que a audiência acompanhe o raciocínio e levante suas questões. A análise visual torna-se um processo dinâmico, onde as diferentes exibições levam a diferentes caminhos e questões em cima dos dados.

A análise visual é um método de explorar dados visualmente, em tempo real. Uma experiência produtiva de análise visual apresenta certas características. A qualquer momento, você pode:

- Alterar os dados que você está vendo imediatamente (com um clique). Isso é importante porque diferentes perguntas pedem diferentes dados.
- Alterar imediatamente a maneira como você está vendo os dados.

Isso é importante porque diferentes exibições de dados respondem a diferentes perguntas. Com cada pequena alteração, a exibição dos dados é atualizada imediatamente, facilitando a exploração de dados no formato correto.

3.5 Seleção e preparo do banco do dados

O desafio de trabalhar com o banco de dados dos Registros Hospitalares do Câncer (RHC) baseou-se em dois pilares:

- Obtenção dos dados:** Captação das informações via Integrador RHC;
- Banco de dados:** Preparação de um banco de dados *online* em tempo real utilizando MySQL para consumo via Tableau.

Para a obtenção dos dados, foi utilizado o sistema disponibilizado pelo INCA chamado Integrador RHC (figura 8): interface responsável por disponibilizar o *download* dos registros hospitalares de câncer dos hospitais associados (SUS e particulares).

Figura 8: Visão geral Integrador RHC

Integrador RHC
Registro Hospitalar de Câncer

Login: Senha: Esqueci minha senha. Clique [aqui](#).

Tabular Dados
Cadastre-se
Fale Conosco
Unidade Hospitalar

Tabular Dados
[Manual de uso do TabNet](#) [Notas Técnicas](#)

Tabulador Hospitalar: Informações originais, das bases dos RHC (casos analíticos e não analíticos). Acesso a todos os usuários, inclusive ao público em geral.
[Consultas](#)
[Download](#)

Relatório de Tempo: Aqui é possível emitir Relatório de Intervalo de Tempos entre Primeira Consulta -> Diagnóstico, Diagnóstico -> Tratamento e Primeira Consulta -> Tratamento.
[Consultar](#)

Últimas Atualizações

Data	Anos	Estado	Descrição
16/09/2021	2000 - 2019	Todos	Bases novas atualizadas no IntegradorRHC, recebidas do dia 27/01/2021 ao dia 15/09/2021.
26/01/2021	2000 - 2019	Todos	Bases novas atualizadas no IntegradorRHC, recebidas do dia 18/09/2020 ao dia 26/01/2021.
17/09/2020	2000 - 2018	Todos	Bases novas atualizadas no IntegradorRHC, recebidas do dia 27/03/2020 ao dia 17/09/2020.
26/03/2020	2000 - 2018	Todos	Bases novas atualizadas no IntegradorRHC, recebidas do dia 18/09/2019 ao dia 26/03/2020.
17/09/2019	2000 - 2017	Todos	Bases novas atualizadas no IntegradorRHC, recebidas do dia 24/01/2019 ao dia 17/09/2019.
23/01/2019	2000 - 2017	Todos	Bases novas atualizadas no IntegradorRHC, recebidas do dia 06/07/2018 ao dia 23/01/2019.
06/07/2018	2000 - 2016	Todos	Bases novas atualizadas no IntegradorRHC, recebidas do dia 20/01/2018 ao dia 05/07/2018.
14/03/2018	2000 - 2016	SP	Bases dos RHC de SP, geradas pela FOSP em dezembro de 2017.
23/01/2018	2005 - 2014	BA	Atualização do banco de dados de 2005 até 2014, Hospital Edgard Santos, CNES: 0003816
19/01/2018	2000 - 2016	Todos	Bases novas atualizadas no IntegradorRHC, recebidas do dia 29/06/2017 ao dia 19/01/2018.
28/06/2017	2000 - 2015	Todos	Bases novas atualizadas no IntegradorRHC, recebidas do dia 25/01/2017 ao dia 28/06/2017.
25/01/2017	2000 - 2015	Todos	Bases novas atualizadas no IntegradorRHC, recebidas do dia 30/05/2016 ao dia 24/01/2017.
30/05/2016	2000 - 2015	Todos	Bases novas atualizadas no IntegradorRHC, recebidas do dia 01/02/2016 ao dia 30/05/2016.
01/02/2016	2000 - 2015	Todos	Bases Novas Atualizadas no IntegradorRHC, recebidas do dia 06/07/2015 ao dia 01/02/2016.
11/08/2015	2011 - 2011	RJ	Atualização do banco de dados de 2011, Hospital do Câncer - HCI, CNES: 2273454
06/07/2015	1985 - 2014	Todos	Incluídos os Hospitais do estado de São Paulo: Banco de dados de 2000 até 2011
28/05/2015	2010 - 2010	SP	Atualização da base de 2010
08/12/2014	1985 - 2014	Todos	Atualização Inicial

Fonte: [28]

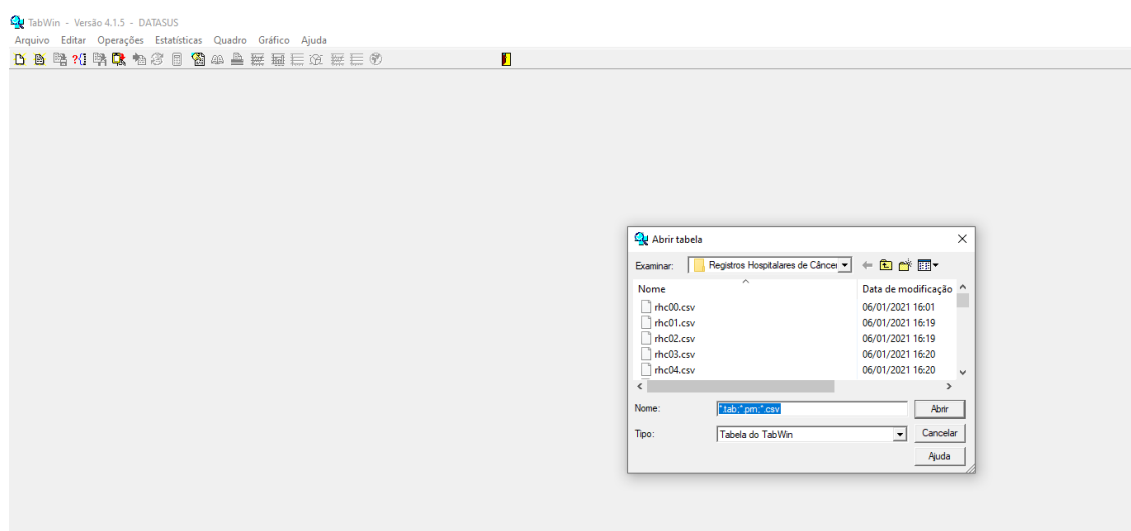
Nas notas técnicas do INCA [21], os dados disponíveis são oriundos dos Registros Hospitalares de Câncer, geridos pelas unidades hospitalares que prestam assistência de Alta Complexidade em Oncologia no Sistema Único de Saúde (SUS), que tem como principal fonte de informações os prontuários médicos de pacientes com câncer atendidos nas unidades hospitalares. As Secretarias Estaduais de Saúde (SES) são responsáveis pela construção de uma base estadual consolidada de casos analíticos, a partir da qual é construída a base nacional consolidada utilizada

no tabulador público. O tabulador hospitalar abrange todas as bases de dados originais encaminhadas pelos RHC, contemplando casos analíticos e não analíticos.

Conforme a figura 8, o sistema permite o *download* em formato tabwin* os arquivos que contém informações seguindo a estrutura da tabela 1. O formato é de conhecimento dos profissionais do meio da saúde e de autoria própria dos sistemas dataSUS e foi criado para padronizar a documentação de registros hospitalares. As ferramentas TABWIN e TABNET permitem sua conversão em formatos comerciais e possibilitam a integração de diversas bases de dados existentes no DATASUS e consequentemente facilitam a extração de informações relevantes para disseminação de informações em saúde, como por exemplo; calcular indicadores, taxas, dados estatísticos, visualização dos dados através de gráficos, mapas etc.

O TABWIN, apresentado na figura 9, é um programa desenvolvido para rodar na plataforma Windows, criado pelo DATASUS para facilitar o trabalho de tabulação e o tratamento de dados. Suas funcionalidades facilitam a construção e aplicação de índices e indicadores de produção de serviços, de características epidemiológicas (incidência de doenças, agravos e mortalidade) e de aspectos demográficos de interesse (educação, saneamento, renda etc.) - por estado e por município; a programação e o planejamento de serviços; a avaliação e tomada de decisões relativas à alocação e distribuição de recursos; a avaliação do impacto de intervenções realizadas nas condições de saúde [22].

Figura 9: Visão geral TabWin (DataSUS)



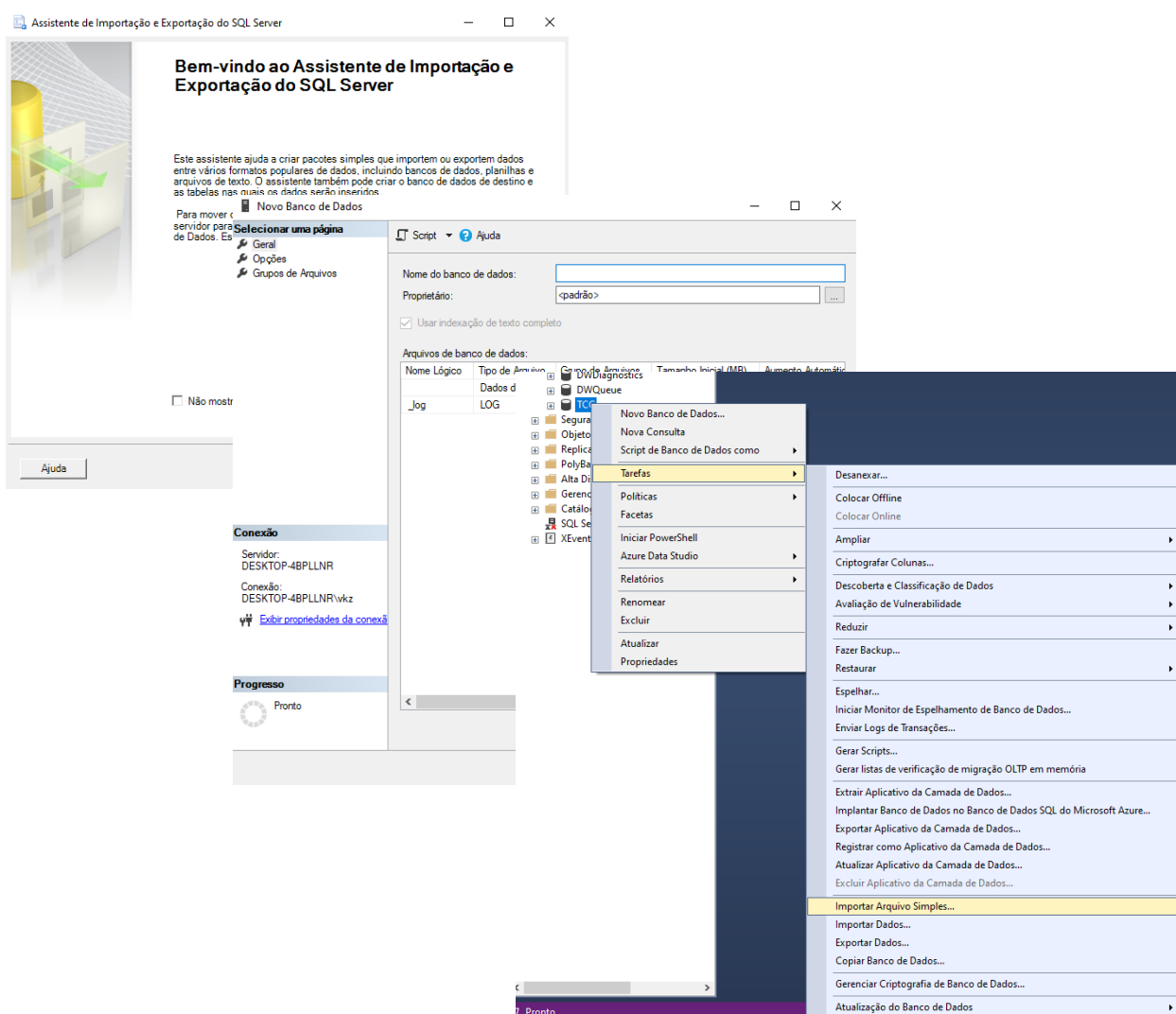
Fonte: Autoria própria

Com foco na criação do banco de dados, as informações foram extraídas do SisRHC em formato TABWIN e a plataforma foi utilizada para a conversão dos arquivos em formato comercial

.csv (*comma separated values* – dados separados por vírgulas). No total, para o período de 2000 à 2021, foram extraídos 33 arquivos.

Posterior a extração e conversão dos arquivos, considerando a busca por uma arquitetura que disponibilizasse as informações em tempo real, utilizou-se o Microsoft SQL Server para a criação de um servidor local conforme destacado pela figura 10. Nesse, criou-se uma instância de acesso local e posteriormente foi criada uma tabela com estrutura compatível à estrutura de informações disponibilizadas pelo DataSUS.

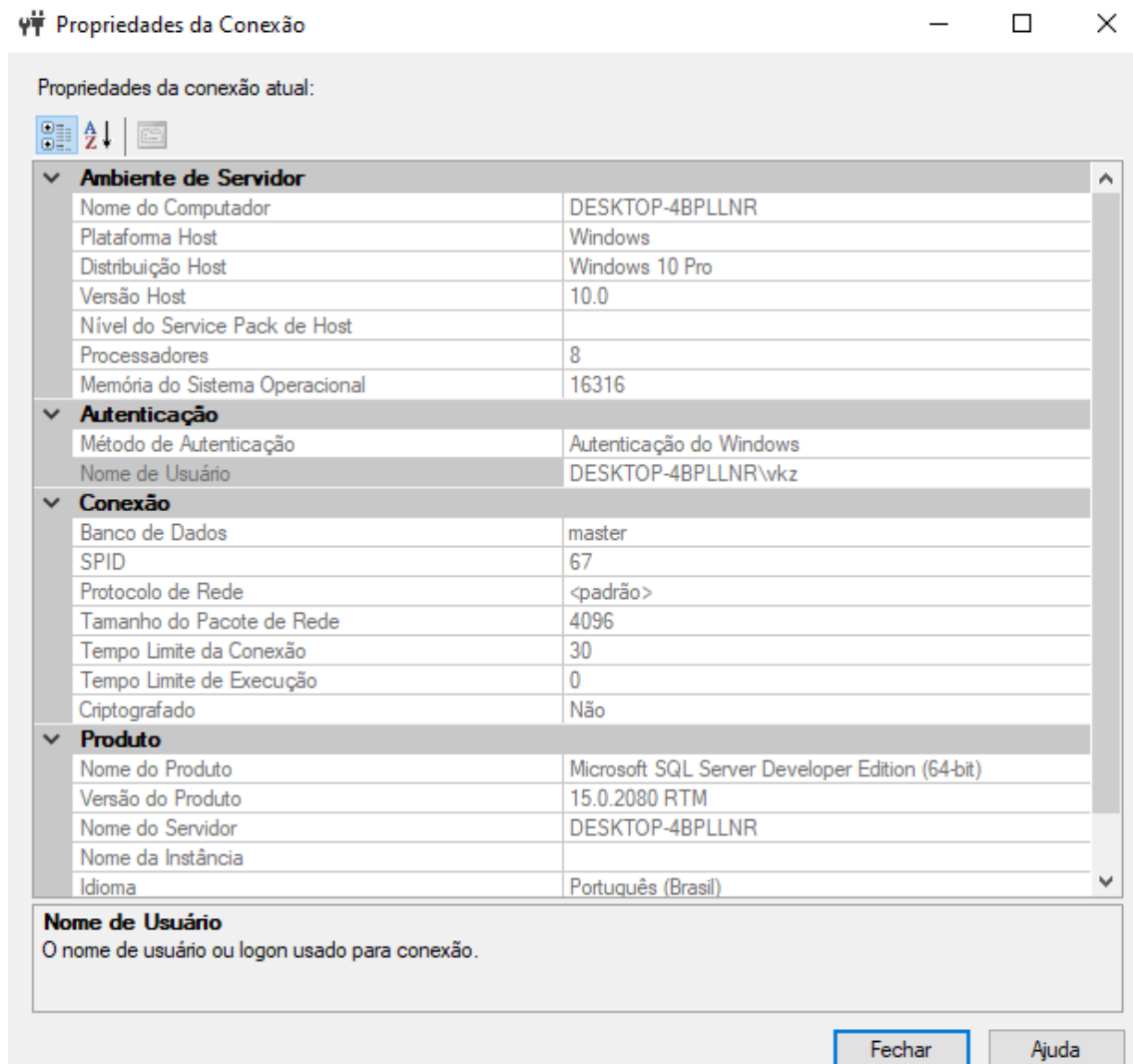
Figura 10: Etapas de configuração SQL Server 2019



Fonte: Autoria própria

O servidor local foi criado com as configurações de conexão apresentadas pela figura 11.

Figura 11: Propriedades de conexão do banco de dados local DESKTOP-4PLLNR

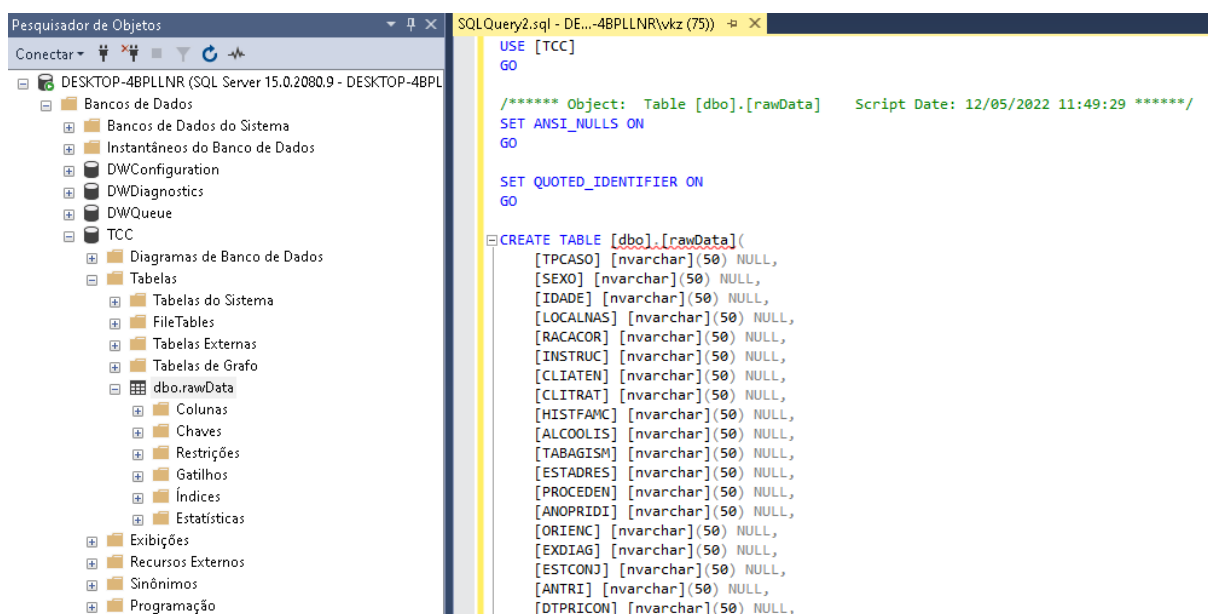


Fonte: Autoria própria

O modo utilizado para a criação do banco de dados foi o método de autenticação do Windows. Assim, serviços que busquem se conectar ao servidor, devem fornecer as informações de login do próprio Windows em protocolo SSL – *Secure Sockets Layer*, camada de soquete seguro, um protocolo de segurança que cria um link criptografado entre um servidor Web e um navegador Web.

Para a criação da tabela compatível ao formato dos dados, foram criadas colunas que seguem o formato indicado no dicionário de variáveis. Para tal, foi criada uma consulta em linguagem SQL (Structure Query Language) a partir do editor de consultas apresentado pela figura 12.

Figura 12: Propriedades de conexão do banco de dados local DESKTOP-4PLLNR



Fonte: Autoria própria

O primeiro passo foi a criação da tabela respeitando o formato das variáveis presentes no banco de dados do RHC, definindo o nome das colunas, seu tipo e restrição de preenchimento:

```

CREATE TABLE TCC.dbo.rawData (
    [TPCASO] nvarchar(50) NULL,
    [SEXO] int NULL,
    [IDADE] int NULL,
    [LOCALNAS] nvarchar(50) NULL,
    [RACACOR] int NULL,
    [INSTRUC] int NULL,
    [CLIATEN] int NULL,
    [CLITRAT] int NULL,
    [HISTFAMC] int NULL,
    [ALCOOLIS] int NULL,
    [TABAGISM] int NULL,
    [ESTADRES] nvarchar(50) NULL,
    [PROCEDEN] int NULL,

```

```

[ANOPRIDI] int NULL,
[ORIENC] int NULL,
[EXDIAG] int NULL,
[ESTCONJ] int NULL,
[ANTRI] int NULL,
[DTPRICON] int NULL,
[DIAGANT] int NULL,
[BASMAIMP] int NULL,
[LOCTUDET] nvarchar(50) NULL,
[LOCTUPRI] nvarchar(50) NULL,
[TIPOHIST] nvarchar(50) NULL,
[LATERALI] int NULL,
[LOCTUPRO] int NULL,
[MAISUMTU] int NULL,
[TNM] int NULL,
[ESTADIAM] int NULL,
[ESTADIAG] int NULL,
[RZNTR] int NULL,
[DTINITRT] int NULL,
[PRITRATH] int NULL,
[ESTDFIMT] int NULL,
[CNES] int NULL,
[UFUH] nvarchar(50) NULL,
[MUUH] int NULL,
[OCUPACAO] int NULL,
[DTDIAGNO] nvarchar(50) NULL,
[DTTRIAGE] nvarchar(50) NULL,
[DATAPRICON] nvarchar(50) NULL,
[DATAINITRT] nvarchar(50) NULL,
[DATAOBITO] nvarchar(50) NULL,
[OUTROESTA] int NULL,
[VALOR_TOT] int NULL,
[BASDIAGSP] nvarchar(50) NULL,
[FileName] nvarchar(200) NULL,
[UploadDatetime] datetime NULL
)
GO

```

Com a tabela criada no MySQL, o próximo passo foi a importação dos dados dos arquivos .CSV. Para tal, criou-se um procedimento – sub-rotina lógica para importação de dados para bancos relacionais.

--Criação do procedimento:

```
CREATE PROCEDURE [dbo].[usp_ImportDB]
```

```
    @Filepath varchar(500)
```

```
    ,@Pattern varchar(100)
```

```
    ,@TableName varchar(128)
```

```
    ,@ViewName varchar(128)
```

```
    ,@ResetTable bit = 0
```

```
AS
```

```
SET QUOTED_IDENTIFIER OFF
```

```
DECLARE @query varchar(1000)
```

```
DECLARE @numfiles int
```

```
DECLARE @filename varchar(100)
```

```
DECLARE @files TABLE (Name varchar(200) NULL)
```

-- Deleta o conteúdo da tabela criada de forma a garantir integridade da tabela

```
IF @ResetTable = 1
```

```
BEGIN
```

```
    PRINT 'Emptying table [' + @TableName + ']...'

```

```
    EXEC ('DELETE ' + @TableName)
```

```
END
```

--Sobe uma lista de arquivos CSV com o nome da pasta em que estão inseridos

```
SET @query = 'master.dbo.xp_cmdshell "dir ' + @filepath+@pattern + ' /b"'
```

```
INSERT @files(Name)
```

```
EXEC (@query)
```

```
DECLARE curs_files CURSOR FOR
```

```
SELECT Name FROM @files WHERE Name IS NOT NULL
```

--Para cada arquivo CSV, executa a consulta

```

SET @numfiles =0
OPEN curs_files
FETCH NEXT FROM curs_files INTO @filename
WHILE (@@FETCH_STATUS = 0)
BEGIN
    SET @numfiles+=1

```

--Inserção massiva de cada CSV na tabela 'rawData' com coluna contendo nome do arquivo origem e horário de upload

```

SET @query = ('BULK INSERT ' + @ViewName
+ ' FROM "' + @Filepath+@filename + '" WITH(
CODEPAGE = "65001",
DATAFILETYPE = "char",
FIRSTROW = 2,
FIELDTERMINATOR="",""',
ROWTERMINATOR="0x0a");'

+ ' UPDATE ' + @TableName
+ ' SET [FileName] = ' + "'" + @filename + "'"
+ ' WHERE [FileName] Is Null;'

+ ' UPDATE ' + @TableName
+ ' SET [UploadDatetime] = ' + "'" + CAST(GETDATE() as nvarchar(1000)) + "'"
+ ' WHERE [UploadDatetime] Is Null;'

)

PRINT 'Importing [' + @filename + '] into [' + @TableName + ']....'
EXEC (@query)

FETCH NEXT FROM curs_files INTO @filename

END

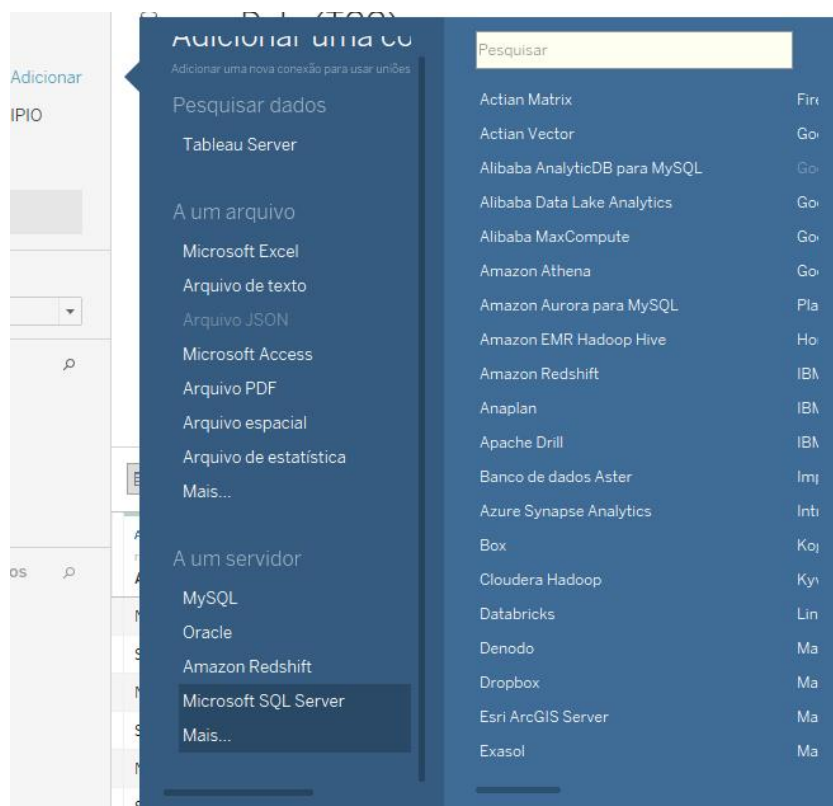
```

Dessa forma, os dados foram importados do CSV e disponibilizados na tabela online disponibilizada no servidor criado.

3.6 Integração do banco de dados

Tendo o banco de dados estruturado no Microsoft SQL Server, o primeiro passo foi a integração ao Tableau. Nesse sentido, a plataforma permite integração nativa para a integração conforme apresentado pelas figuras 13 e 14.

Figura 13: Seleção de conexão via Tableau



Fonte: Autoria própria

Figura 14: Informações de conexão ao servidor

Microsoft SQL Server

Geral SQL inicial

Servidor
DESKTOP-4BPLLNR

Banco de dados
TCC

Autenticação
Usar autenticação do Windows (preferencial)

☐ Exigir SSL

☐ Ler dados não confirmados

Fazer login

Fonte: Autoria própria

De maneira automatizada, o Tableau identifica o Microsoft SQL Server em andamento na máquina. Assim, basta realizar a *login* utilizando a autenticação do Windows para concluir a importação em tempo real da base para o Tableau. A figura 15 apresenta o banco de dados importado ao Tableau.

Figura 15: Visão parcial dos dados importados ao Tableau

Classificar campos: Ordem das fontes de dados

rawData Tp caso	rawData Sexo	rawData Idade	rawData Local nas	rawData Racacor	rawData Instruc	rawData Clísten	rawData Clítrat
"2"	2	071	RJ	1	9	10	88
"2"	1	000	99	9	9	27	88
"1"	2	049	RJ	1	9	20	20
"1"	1	066	ES	1	9	32	32
"1"	2	060	99	1	9	11	4
"2"	1	046	RJ	9	9	9	88
"2"	1	080	MG	1	9	29	88
"1"	2	047	RJ	2	9	4	4
"1"	2	058	99	4	9	15	15
"2"	1	055	RJ	1	9	10	88
"1"	1	075	RJ	9	9	32	32
"1"	1	076	EX	1	2	17	17
"1"	2	065	RJ	2	1	17	17
"1"	2	068	RJ	1	2	17	17
"1"	2	067	RJ	1	2	17	17

Fonte: Autoria própria

Com base no dicionário de variáveis apresentado pela Tabela 1, foi feita a tradução dos dados (numéricos) em informações qualitativas a partir da ferramenta de criação de *alias*, considerando-se a tradução do código numérico em um código qualitativo. Por exemplo, para o campo “Alcoolis”, a figura 16 apresenta a tradução do número presente no banco de dados (membro) na correspondência qualitativa (Valor (Alias)) de histórico de alcoolismo dos pacientes.

Figura 16: Edição de *alias*

Membro	Com alias	Valor (alias)
0		0
1	*	Nunca
2	*	Ex-consumidor
3	*	Sim
4	*	Não avaliado
8	*	Não se aplica
9	*	Sem informação

Fonte: Autoria própria

Para cada campo do banco de dados, foi criada a tradução de forma a permitir a criação do *dashboard* utilizando a representação das informações. Após a implementação de todas as traduções, o banco de dados tornou-se apto a utilização, conforme apresentado pela figura 17.

Figura 17: Visão parcial dos dados traduzidos

Abc rawData Alcoolis	Grupo Alcoolis (grupo)	Dataobito	Dtdiagno	# rawData Idade	Grupo Idade (grupo)	Abc rawData Loctudet	Abc rawData Muuh	Abc rawData Racacor
Nunca	Nunca	nulo	10/03/2002	71	70+	Outras Neoplasias M...	3304557	Branca
Sem informação	Sem informação	nulo	02/01/2002	0	0 - 9	Hipofaringe	3304557	Sem informação
Nunca	Nunca	nulo	21/07/2002	49	40 a 49	Tecido Conjuntivo e d...	3304557	Branca
Sem informação	Sem informação	nulo	27/05/2002	66	60 a 69	Próstata	3304557	Branca
Nunca	Nunca	nulo	21/08/2002	60	60 a 69	Glândula Tireóide	3304557	Branca
Sim	Sim	07/05/2002	08/05/2002	46	40 a 49	Fígado e Das Vias Bili...	3304557	Sem informação
Nunca	Nunca	31/05/2002	11/03/2002	80	70+	Brônquios e Dos Pulm...	3304557	Branca
Nunca	Nunca	nulo	16/10/2002	47	40 a 49	Glândula Tireóide	3304557	Preta
Sem informação	Sem informação	nulo	14/10/2003	58	50 a 59	Colo do Útero	3304557	Parda
Sem informação	Sem informação	nulo	20/05/2003	55	50 a 59	Outras Neoplasias M...	3304557	Branca
Sem informação	Sem informação	nulo	17/11/2003	75	70+	Próstata	3304557	Sem informação
Nunca	Nunca	04/01/2004	25/08/2003	48	40 a 49	Leucemias e morfolog...	3304557	Preta
Não se aplica	Não se aplica	nulo	09/10/2000	8	0 - 9	Leucemias e morfolog...	3304557	Parda
Nunca	Nunca	nulo	31/08/2000	60	60 a 69	Leucemias e morfolog...	3304557	Branca
Sim	Sim	11/07/2002	30/08/2000	46	40 a 49	Leucemias e morfolog...	3304557	Parda
Nunca	Nunca	nulo	23/08/2000	64	60 a 69	Leucemias e morfolog...	3304557	Preta
Nunca	Nunca	24/12/2001	24/08/2000	33	30 a 39	Leucemias e morfolog...	3304557	Branca
Sem informação	Sem informação	06/09/2000	23/08/2000	31	30 a 39	Leucemias e morfolog...	3304557	Branca
Nunca	Nunca	17/11/2000	22/08/2000	21	20 a 29	Leucemias e morfolog...	3304557	Branca
Nunca	Nunca	11/02/2003	24/08/2000	75	70+	Leucemias e morfolog...	3304557	Branca
Sem informação	Sem informação	16/10/2002	06/10/2000	80	70+	Leucemias e morfolog...	3304557	Branca
Nunca	Nunca	nulo	25/08/2000	64	60 a 69	Leucemias e morfolog...	3304557	Preta

Fonte: Autoria própria

3.7 Estruturação do Dashboard

O primeiro passo para o desenvolvimento do *dashboard* foi definir seus objetivos:

1. Permitir rápida navegação dentre as diferentes patologias;
2. Avaliar as características dos pacientes (sexo, idade, etnia, alcoolismo, tabagismo, ocupação);
3. Visualização da tendência de casos e óbitos.

Dessa forma, foi montado um esquemático para orientar o desenvolvimento do trabalho. Do ponto de vista de análise visual, dois pontos são fundamentais – a visualização rápida dos dados (utilizando orientação visual e cores que combinem conceitos – como apresentar dados masculino e feminino sempre com um padrão de cores específicos, assim como orientar o *dashboar* por regiões, como apresentado pela figura 18. Outro ponto fundamental é a determinação dos filtros – são estes que são utilizados para dar dinamismo as análises, permitindo a navegabilidade sob diferentes óticas.

Figura 18: Esquemático dos objetivo do *dashboard*

Fonte: Autoria própria

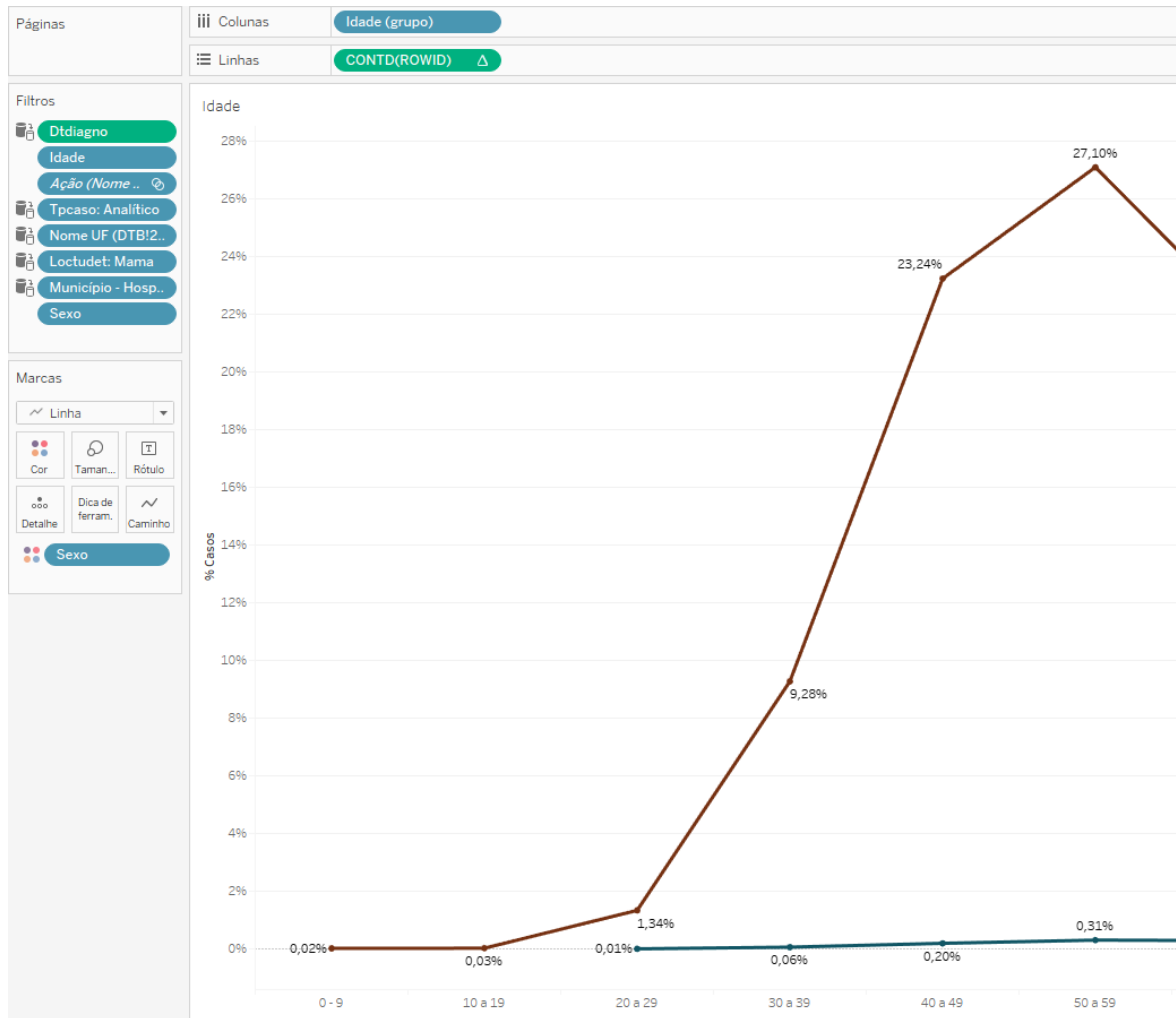
O desenvolvimento de *dashboards* no Tableau é dividido em três grandes elementos:

1. **Definição das planilhas:** criação das visualizações – gráficos de barras, linhas, gantt, mapa, dispersão etc. – a partir das variáveis presentes no banco de dados;
2. **Criação de painéis:** reunião dos diferentes gráficos com a configuração dos filtros a serem aplicados nas diferentes planilhas
3. **Histórias:** combinação de diferentes painéis, permitindo navegar por diferentes filtros pré-definidos – exemplo: visão Brasil, Estado de SP, cidade de São Paulo.

De forma a ilustrar o desenvolvimento do trabalho, a figura 19 apresenta o desenvolvimento da visualização da evolução da idade na incidência das neoplasias. Para a criação de uma visualização, é adicionado o dado a ser estudado como coluna. Nesse caso, a ideia era visualizar em faixas etárias a evolução da doença. Foram criados portanto grupos que avaliam o campo

“Idade” presente no registro do RHC, de associam a linha registrada a um dos grupos de faixa etária (0-9, 10 a 19, 20 a 29, 30 a 39, 40 a 49, 50 a 59 ou 60+).

Figura 19: Visualização da criação da planilha “Idade”



Fonte: Autoria própria

Além da definição das informações que compõe o gráfico, outro ponto fundamental é adicionar quais são as informações a serem filtradas pelo usuário. Nesse caso, foram adicionadas as informações apresentadas pela figura 20.

Figura 20: Visualização da criação da planilha “Idade”



Fonte: Autoria própria

Nesse caso, os filtros aplicados à informação foram:

1. Dtdiagno – Data de diagnóstico
2. Idade – Idade do paciente
3. Tpcaso – Visualização dos casos analíticos
4. Nome UF – Estado do paciente
5. Loctudet – Localização primária do tumor
6. Município – Hospitalar – Município em que foi registrado o caso.
7. Sexo – Sexo do paciente

Os filtros de idade e sexo foram implementados de forma a excluir da base dados não condizentes. Nessa mesma configuração, foi definido que os filtros se aplicariam à todas as planilhas criadas, de forma a manter o desenvolvimento integrado nas diferentes visualizações a serem criadas. A mesma lógica foi aplicada ao desenvolvimento das diferentes visualizações que compõe o *dashboard*, apresentadas pela tabela 2.

Tabela 2: Apresentação das planilhas criadas durante o desenvolvimento

Planilha	Dado	Visualização
Sexo	Sexo do paciente	Gráfico de pizza
Geografia	Estado de registro	Mapa geográfico
Idade	Sexo do paciente	Gráfico de linhas
Etnia	Etnia do paciente	Gráfico de pizza
Ocupacional – Geral	Ocupação principal	Gráfico de barras horizontais
Ocupacional – Masculino		Gráfico de barras horizontais
Ocupacional – Feminino		Gráfico de barras horizontais
Mais frequentes – Masculino	Localização primária do tumor	Gráfico de barras horizontais
Mais frequentes – Feminino		Gráfico de barras horizontais
Alcoolismo – Masculino	Histórico de consumo de bebida alcóolica	Gráfico de barras verticais
Alcoolismo – Feminino		Gráfico de barras verticais
Tabaco – Masculino	Histórico de consumo de tabaco	Gráfico de barras verticais
Tabaco – Feminino		Gráfico de barras verticais
Tendência – Geral	Nº de casos / ano	Gráfico de linha e tendência
Óbito	Nº de mortes / ano	Gráfico de linha e tendência

Fonte: Autoria própria

Do ponto de vista de criação de planilhas, o processo é similar para todos os elementos. Um diferencial está na análise de tendência de número de casos por ano. Ao criar o gráfico de linha, o Tableau permite a adição linhas de tendência.

Para a análise dos dados, optou-se pelo modelo linear uma vez que o coeficiente de determinação nesse modelo é próximo a 0,95 para o período analisado nas diferentes neoplasias. Vale ressaltar que os dados da base do RHC não devem ser utilizados para determinar a tendência do câncer como um todo, uma vez que representa apenas os registros dos hospitais associados.

A figura 21 apresenta um passo a passo iterativo para a construção do *dashboard* conforme as informações definidas e contidas na tabela 2. A partir desse modelo, é construído o *dashboard* final.

Figura 21: Fluxograma de criação do *dashboard* no Tableau

Fonte: Autoria própria

4. RESULTADOS

4.1 Visão geral

O primeiro passo analisado foi o “recorte” dos dados. Uma vez que o banco de dados do RHC é atualizado de forma anual com atualizações assíncronas – ou seja, no ano de 2022 aparecem novos registros de casos de câncer de anos passados, seja pela transferência de casos para hospitais do SUS, seja pela não atualização do caso na data de início de tratamento. A tabela 3 apresenta uma visão completa com o número de casos presentes de 2000 à 2021.

Tabela 3: Número de casos analíticos registrados no RHC

Ano de diagnóstico	Casos analíticos
2000	20.737
2001	69.925
2002	81.985
2003	84.926
2004	97.268
2005	116.358
2006	123.782
2007	139.107
2008	152.959
2009	162.933
2010	174.336
2011	187.081
2012	193.384
2013	206.192
2015	216.074
2016	214.475
2017*	183.865
2018*	106.505
2019*	4.934
2020*	26
2021*	0
Total	2.243.469

Fonte: Autoria própria

Observa-se que os números de casos registrados a partir de 2017 são consideravelmente menores. No entanto, essa redução não representa que o número de casos de câncer esteja caindo e sim ao fato da recomendação de pelo menos um ano para inclusão no sistema, sendo esse período destinado a obtenção de maiores informações sobre o tumor, bem como a forma de tratamento adotada, garantindo assim maior amplitude e qualidade dos dados. Dessa forma, portanto, o *dashboard* tratará dos dados de 2000 a 2015.

De maneira geral, as estatísticas sobre câncer são divulgadas com defasagem aproximada de três anos em relação ao ano vigente. No entanto, esse fato não é impeditivo para análises no campo de planejamento e controle de doença, uma vez que se trata de uma doença crônica que não apresenta flutuações abruptas em curtos períodos de tempo.

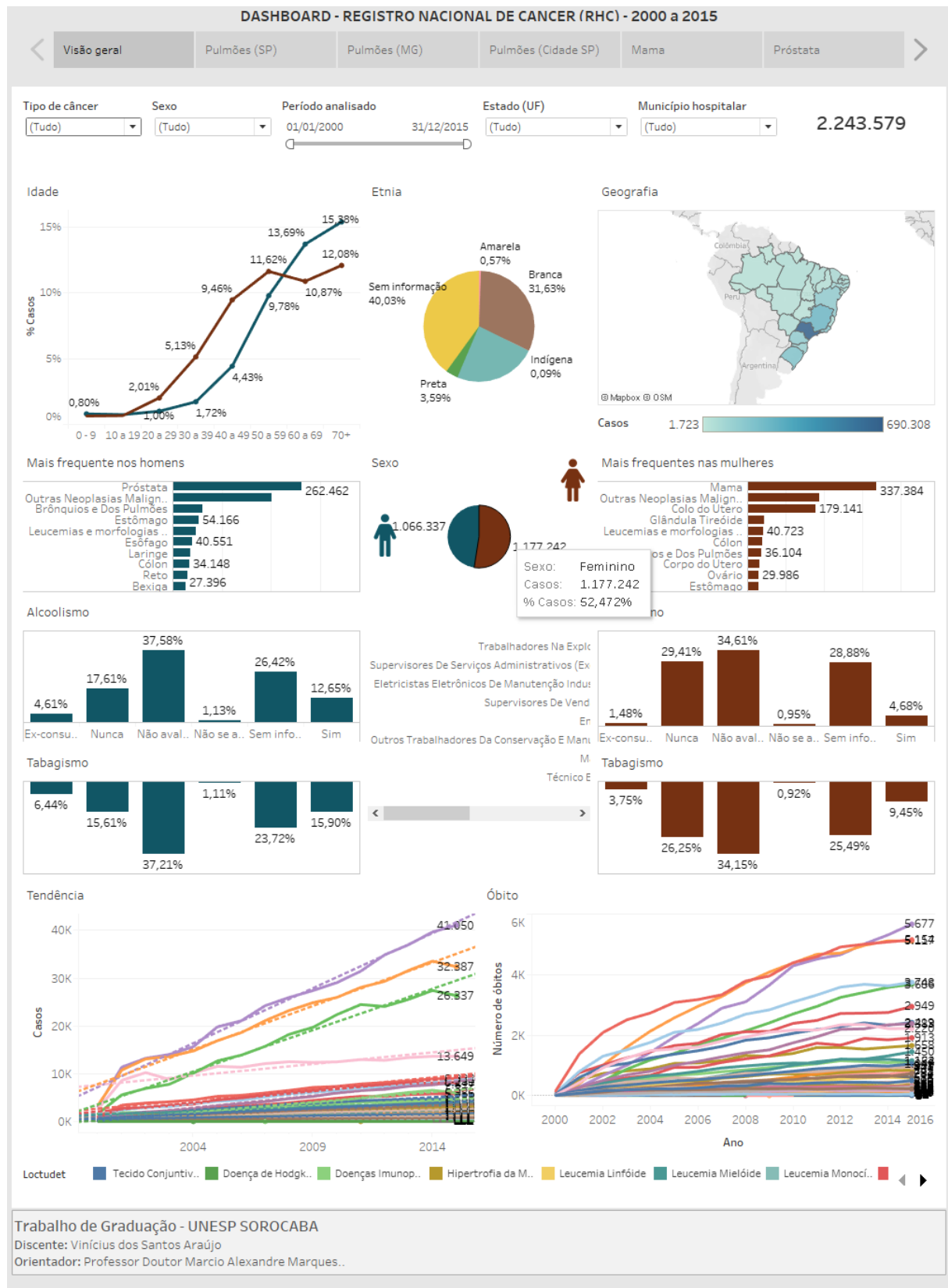
Uma visão geral do *dashboard* construído é apresentada na figura 22. O modelo de apresentação das informações foi baseado em seis visualizações:

- a. Visão geral: comportamento da base como um todo, com destaque para as neoplasias de maior incidência em homens e mulheres e suas características
- b. Pulmões (SP): recorte do câncer de pulmão em casos registrados no estado de São Paulo
- c. Pulmões (MG): recorte do câncer de pulmão em casos registrados no estado de Minas Gerais
- d. Pulmões (Cidade SP): recorte do câncer de pulmão em casos registrados na cidade de São Paulo
- e. Mama: recorte do câncer de mama registrado no Brasil
- f. Próstata: recorte do câncer de próstata registrado no Brasil

Além das diferentes visualizações criadas, também foram definidos os seguintes filtros que permitem uma análise personalizada dos dados:

- a. Tipo de câncer: seleção da(s) neoplasia(s)
- b. Sexo: seleção do sexo
- c. Período analisado: seleção do período de registro
- d. Estado: seleção do estado de registro
- e. Município hospitalar: seleção do estado município hospitalar de registro

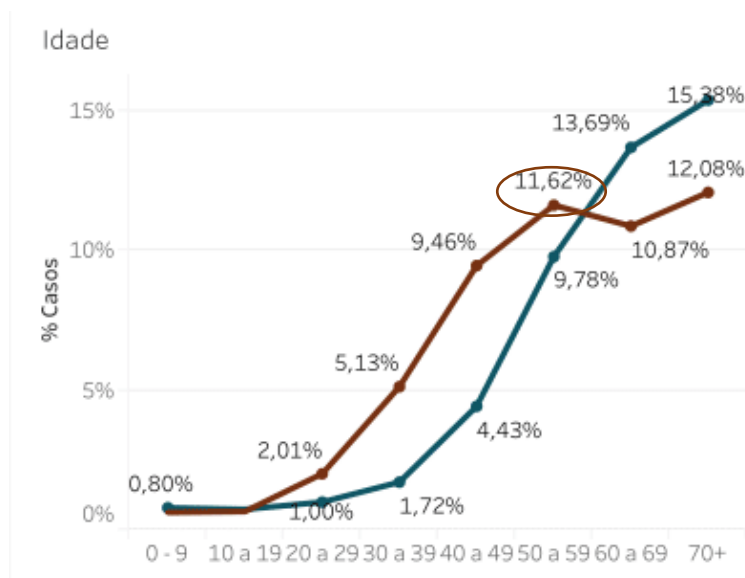
Dessa forma, além das visualizações determinadas nesse trabalho, a construção permite a navegação ‘sob demanda’ a partir dos filtros construídos.

Figura 22: Apresentação geral do *Dashboard* com informações do RHC

Fonte: Autoria própria

A distribuição dos tumores entre homens e mulheres ocorreu de forma proporcional, 47% em homens e 53% nas mulheres. Conforme visto na figura 23, a frequência da doença aumenta conforme a idade, aparecendo de forma mais marcante a partir da faixa de 50 a 59 em mulheres (vermelho), enquanto em homens a faixa com maior característica é a partir de 60 anos.

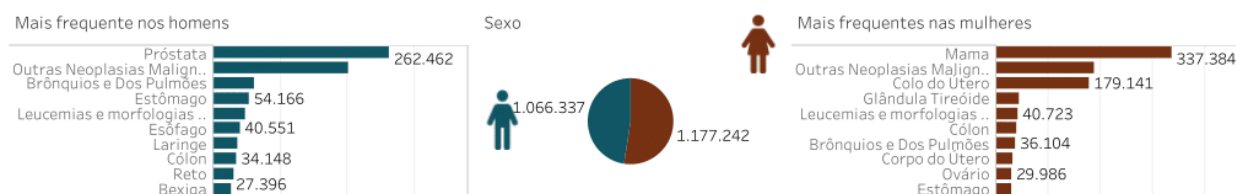
Figura 23: Distribuição proporcional dos casos analíticos



Fonte: Autoria própria

Conforme apresentado na figura 24, no sexo feminino as neoplasias de mama e colo do útero foram as mais frequentes. Já para o sexo masculino, as mais frequentes são próstata e brônquios e pulmões. É possível observar ainda um problema de registro, já que a segunda “neoplasia” nos dois sexos é caracterizada como “outras neoplasias malignas”.

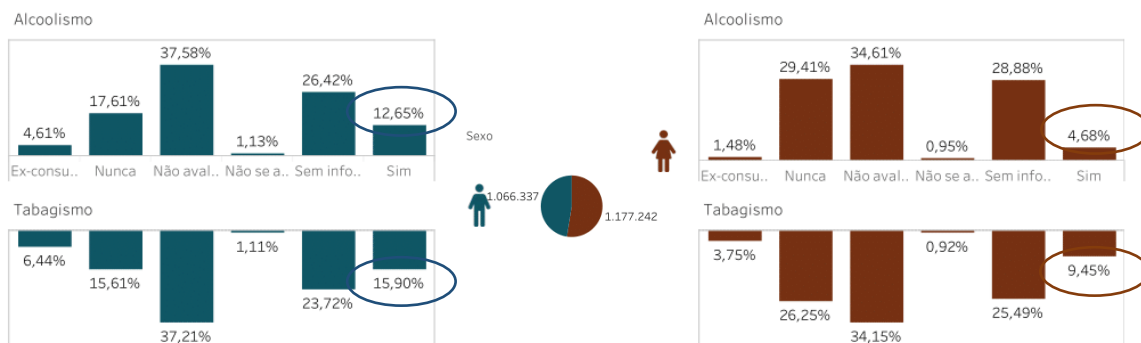
Figura 24: Tipos de câncer mais frequentes em homens e mulheres



Fonte: Autoria própria

Do ponto de vista de histórico de consumo alcoólico e de tabaco apresentada pela figura 25, é possível observar que o consumo de ambos é maior no sexo masculino, principalmente no consumo alcoólico.

Figura 25: Histórico de consumo alcóolico e tabaco



Fonte: Autoria própria

4.2 Câncer de brônquios e pulmões – Estado de São Paulo

Um segundo objetivo do trabalho foi realizar a análise específica de uma neoplasia em uma geografia definida. Dessa forma, inicialmente escolheu-se o câncer de brônquios e pulmões no estado e cidade de São Paulo. Ao construir o *dashboard*, apresentado nas figuras 25 e 26, é possível observar que não há completude dos dados de etnia, consumo alcoólico, ocupação. Dessa forma, somente é possível contabilizar o número de casos registrados, sua distribuição etária e a tendência.

Dentro do Estado de São Paulo, foram registrados 32.529 casos de neoplasia de brônquios nos hospitais associados ao RHC. Destes, 63,04% foram em homens e 36,96% em mulheres. Do ponto de registros de casos, é possível observar uma tendência linear essencialmente a partir do ano de 2009, tendo sido registrados 2680 casos em 2015.

No recorte da cidade de São Paulo, foram registrados 15.330 casos de neoplasia de brônquios nos hospitais associados ao RHC. Destes, 60,20% foram em homens e 39,80% em mulheres. Do ponto de registros de casos, é possível observar uma tendência linear essencialmente a partir do ano de 2009, tendo sido registrados 1.317 casos em 2015 e pico em 2014. É importante reforçar que a tendência de crescimento também está associada a ampliação da rede de hospitais cadastrados no RHC, não podendo ser utilizada para a determinação se o câncer está se tornando mais frequente ou não na população como um todo.

Dada a ausência de dados no estado de São Paulo, para a análise das características de uma neoplasia específica com recorte local, foi criada uma visão da neoplasia de brônquios e pulmões no estado de MG, conforme apresentado pela figura 27.

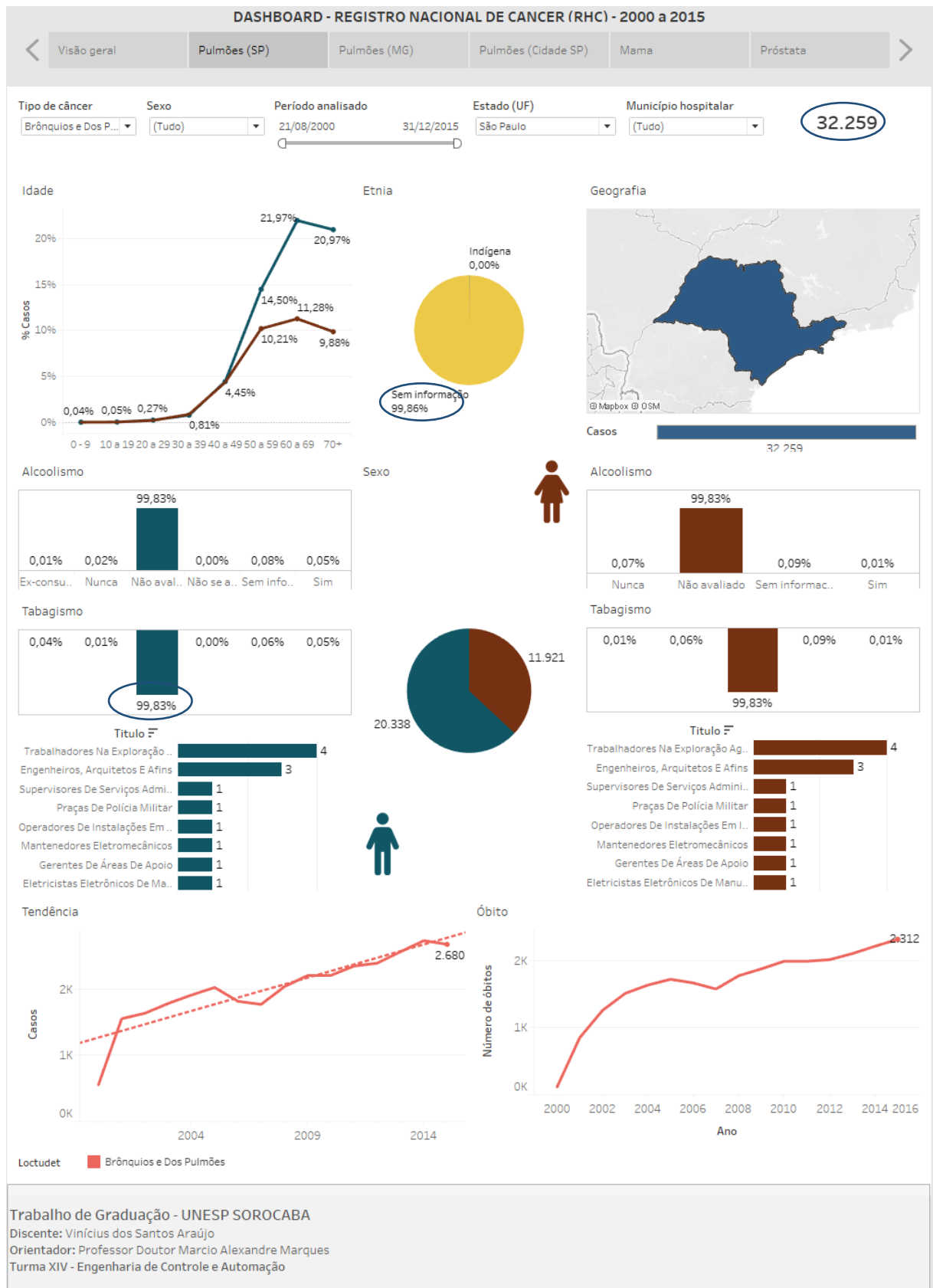
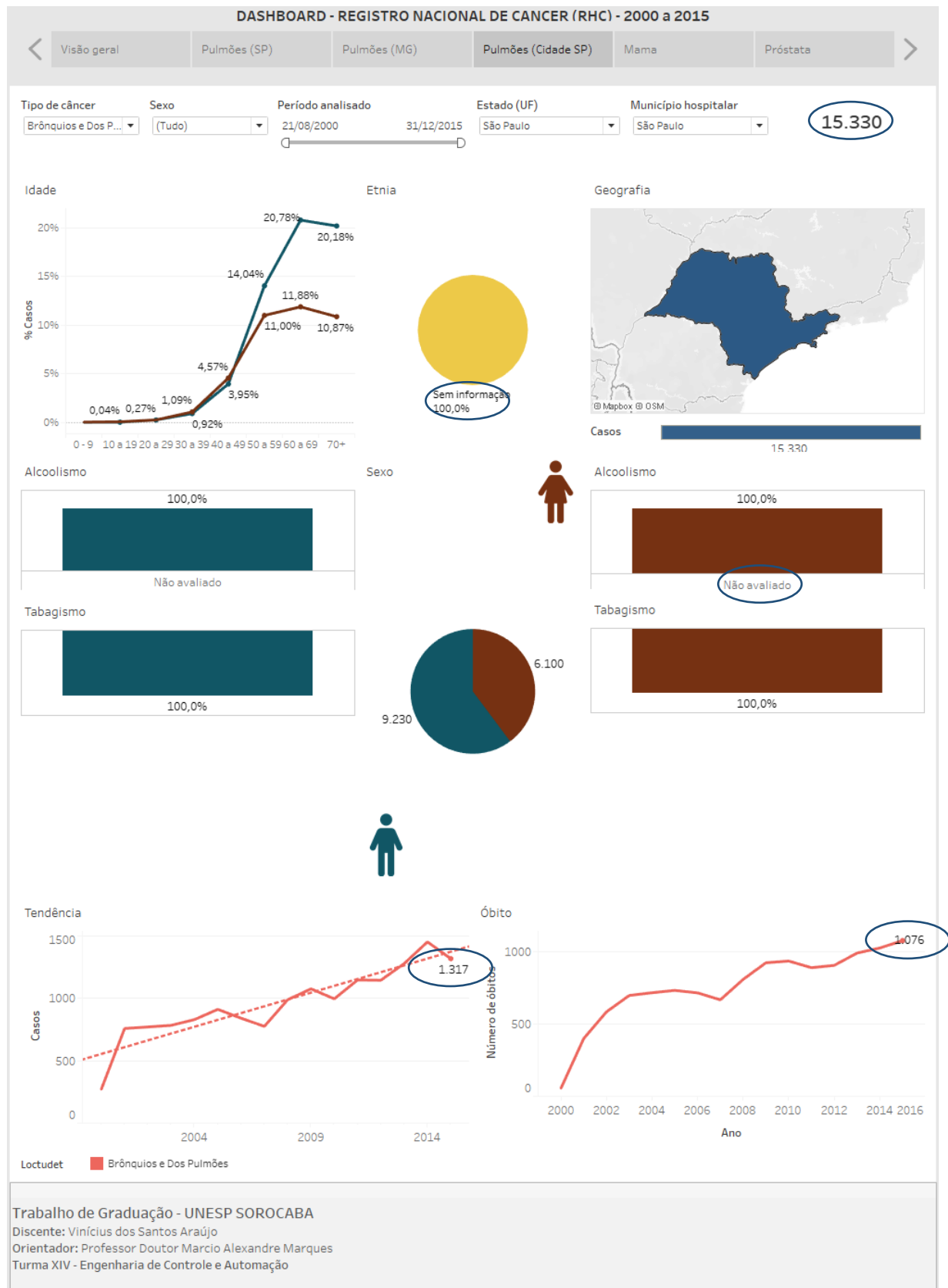
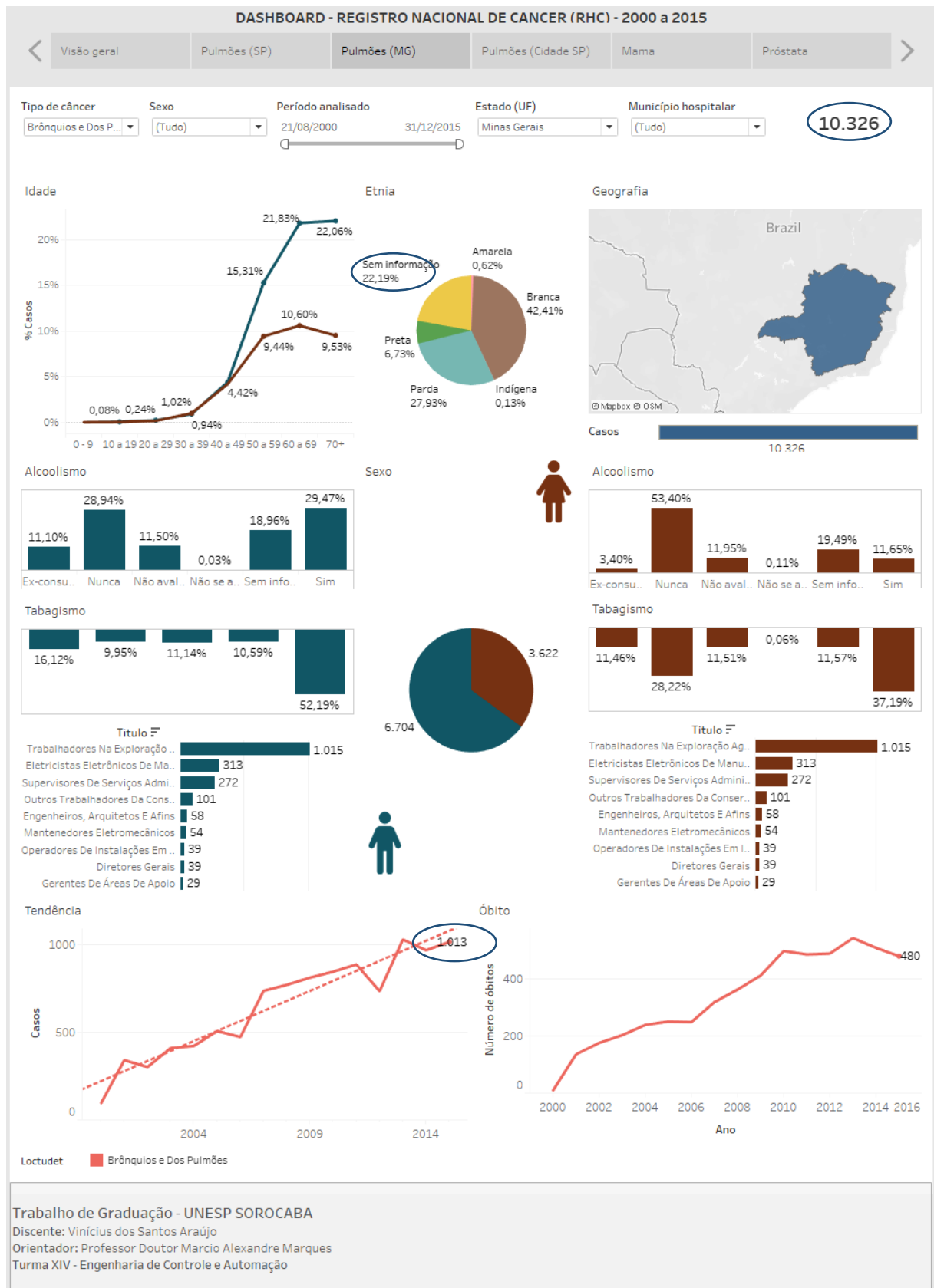
Figura 26: *Dashboard* – Neoplasia de Brônquios e Pulmões – Estado de SP

Figura 27: *Dashboard* – Neoplasia de Brônquios e Pulmões – Cidade de SP

Fonte: Autoria própria

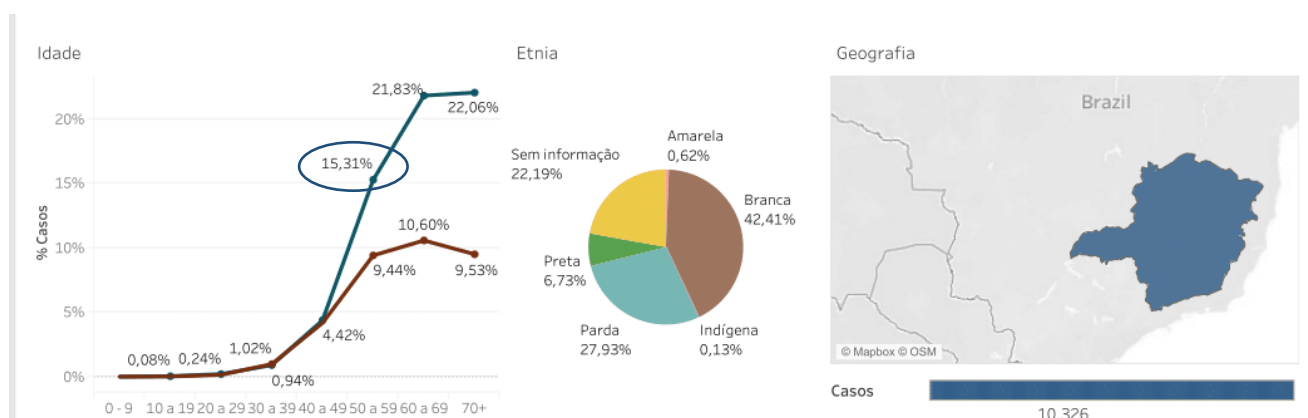
Figura 28: *Dashboard* – Neoplasia de Brônquios e Pulmões – Estado de MG

Fonte: Autoria própria

4.3 Câncer de brônquios e pulmões – Estado de Minas Gerais

Foram registrados 10.326 casos no período de 2000 a 2015. Do ponto de vista de características, conforme apresentado pela figura 29, o número de casos cresce fortemente a partir da faixa de 40 a 49 anos essencialmente nos casos do sexo masculino. É possível observar ainda que 22% da base não possui informação sobre etnia.

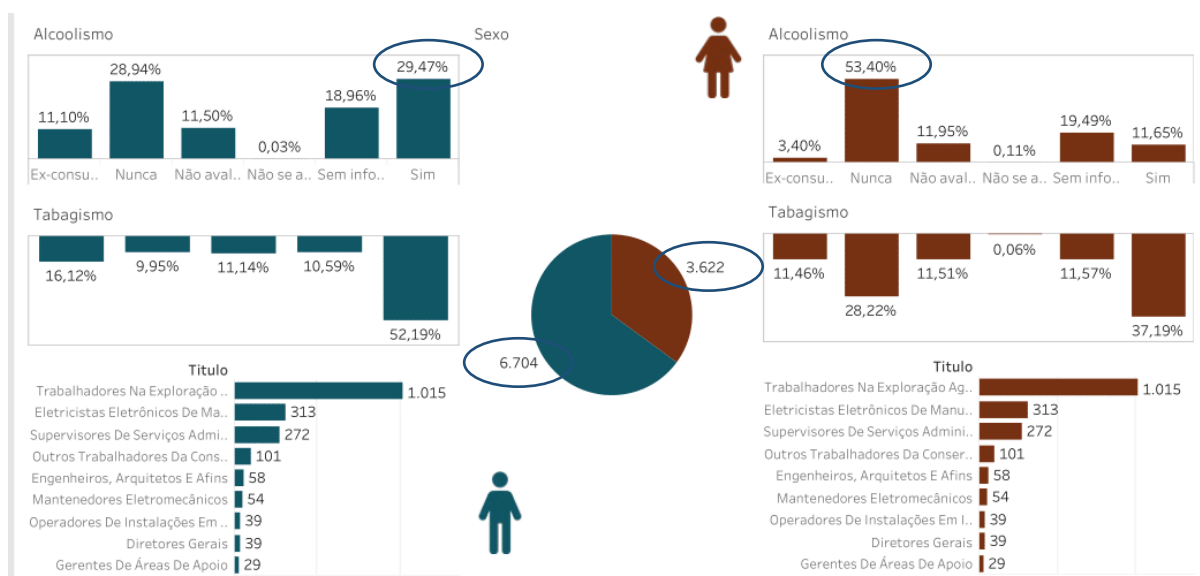
Figura 29: Idade e etnia – Neoplasia de Brônquios e Pulmões – Estado de MG



Fonte: Autoria própria

Como apresentado na figura 30, 29,47% dos casos reportados em homens possuem histórico de consumo alcoólico. Da mesma forma, como observado na amostra geral, as mulheres reportam um consumo alcoólico bem reduzido, observando 53,40%.

Figura 30: Alcoolismo e tabagismo – Estado de MG



Fonte: Autoria própria

Quanto ao tabagismo, 52,19% dos homens fazem uso, enquanto 16,12% são ex-consumidores. Do lado das mulheres, 37,19% fazem uso, enquanto 11,46% são ex-consumidores. Vale ressaltar o grande percentual de mulheres que relatam nunca terem feito uso, 28,22%.

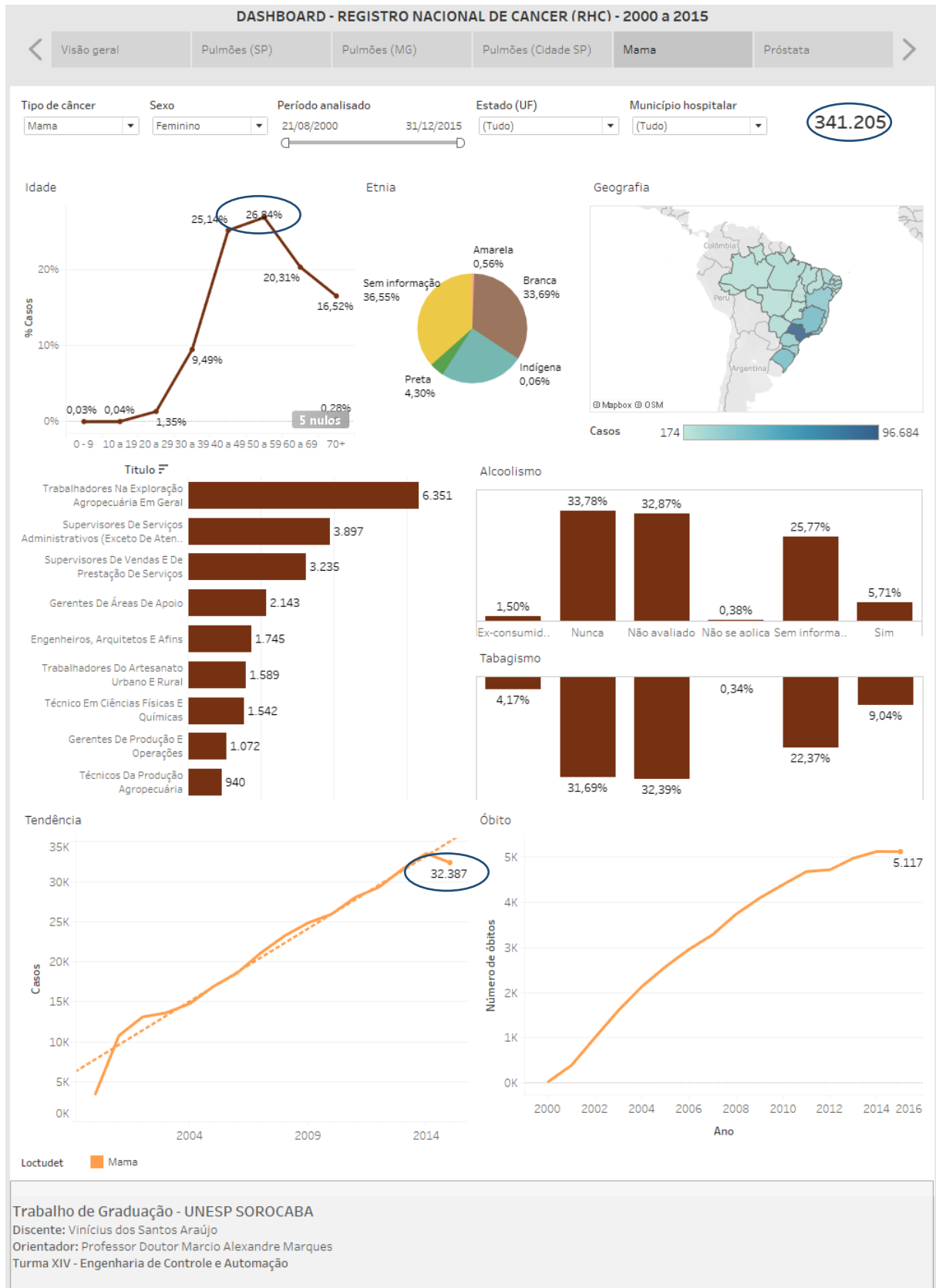
4.4 Câncer de mama – Brasil

A figura 31 apresenta o *dashboard* construído para analisar os registros de neoplasia de mama no período de 2000 a 2015. Foram registrados 341.205 casos distribuídos nos diferentes estados, sendo São Paulo o de maior registro. Esse comportamento está relacionado ao número de hospitais com foco no tratamento de câncer presentes no estado, com grandes referências brasileiras.

Ao analisar a distribuição na idade, 88,81% dos casos ocorrem na faixa de 40 a 70 anos, com destaque para 50 a 59 anos, com 26,84% dos casos. Na distribuição, 33% das mulheres se identificam brancas. Na informação de etnia, são observados 36,55% de registros sem informação. Em 2014, ocorreram 5100 óbitos associados a doença e 32 mil casos foram registrados no mesmo ano.

Ao analisar a base, observa-se novamente o baixo relato de consumo de álcool/tabaco, 5,71% e 9,04% respectivamente. Em média, 33% das mulheres reportam o não consumo de ambas substâncias, enquanto 24% não avaliadas sobre o tópico.

Por fim, do ponto de vista de ocupação profissional, as características que prevalecem são atividades agropecuárias e administrativas.

Figura 31: *Dashboard* – Neoplasia de Mama – Visão geral

Fonte: Autoria própria

4.5 Câncer de próstata – Brasil

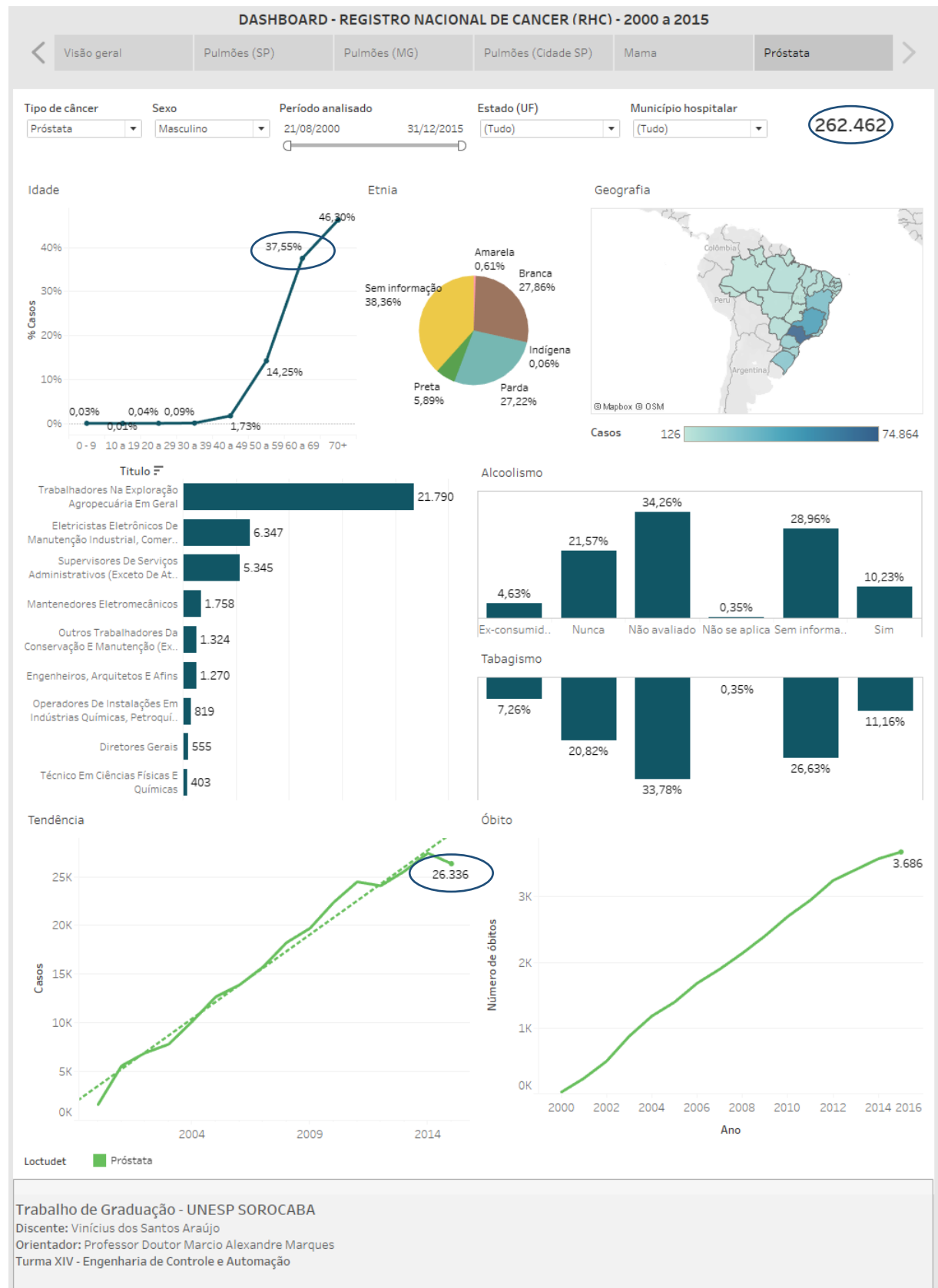
A figura 32 apresenta o dashboard construído para analisar os registros de neoplasia de próstata no período de 2000 a 2015. Foram registrados 262.462 casos distribuídos nos diferentes estados, sendo São Paulo o de maior incidência, seguido por Minas Gerais.

Ao analisar a distribuição na idade, 98,1% dos casos ocorrem na faixa de 50 a 70 anos, com destaque para 70+ anos, com 46,30% dos casos. Na distribuição, 27% dos homens se identificaram brancos. Na informação de etnia, são observados 38,36% de registros sem informação. Em 2015, ocorreram 3686 óbitos associados a doença e 26 mil casos foram registrados no mesmo ano.

Ao analisar a base, mais de 50% da base não informou ou foi avaliado sobre o uso de álcool e tabaco, enquanto cerca de 20% reportam não fazer consumo.

Por fim, do ponto de vista de ocupação profissional, as características que prevalecem são atividades agropecuárias seguidas por atividades industriais (eletricistas, manutenção etc.).

Figura 32: Dashboard – Neoplasia de próstata – Visão geral



Fonte: Autoria própria

4. CONCLUSÃO

Com o desenvolvimento desse trabalho, foi possível observar a flexibilidade proporcionada por uma plataforma de *business intelligence* para auxiliar a análise de dados. A criação do banco de dados e a facilidade na integração junto ao Tableau é um dos destaques da ferramenta. Essa integração *online* permite não só a extração e acompanhamento dos dados ao longo dos meses, como também uma rotina que permita um maior controle sobre a qualidade das informações inseridas.

Do ponto de vista do Tableau, a criação dos filtros e das histórias permitem um dinamismo interessante para criação do *storytelling* a partir dos dados. A combinação de diferentes fontes de informação também é um fator relevante, o que permite o desenvolvimento e a implementação de uma arquitetura robusta em termos de informações.

Com relação aos diferentes *dashboards*, sobretudo como exemplo a de câncer de pulmão, vale ressaltar a importância de garantir uma completude mais ampla das informações. Seja por motivos de controle ou de estudo, as características dos pacientes é fundamental para identificação de correlações que vão além da idade. Hoje, o banco de dados do RHC é falho em alguns níveis de informação, especialmente no estado de São Paulo, onde grande parte das informações adicionais não é inserida no sistema.

Como trabalhos futuros, a exemplo das ferramentas criadas em parceria ao DataSUS e os diferentes aprendizados – inclusive de arquitetura – decorrentes da COVID-19, cabe a avaliação da estruturação de uma ferramenta de *business intelligence* associada ao registro nacional de câncer, o que traz potencial não somente para tratar e ter dados de maior qualidade, como também expandir a um número mais amplo de hospitais do Brasil.

REFERÊNCIAS

ANGELONI, Maria T.; REIS, Eduardo S. *Business Intelligence* como Tecnologia de Suporte a Definição de estratégias para melhoria da qualidade do ensino. In: Encontro da ANPAD, 2006, Salvador. XXX Encontro Nacional de Pós-Graduação em Administração, 2006, 2006. v. 1. p. 16 páginas.

BARBIERI, C. *BI-Business Intelligence: modelagem e tecnologia*. Rio de Janeiro: Axcel Books, 2001

BEZERRA, A. A. *et al.* *Business Intelligence: uma perspectiva de soluções aplicadas no contexto da Gestão da Informação*. In: ENCONTRO DE ESTUDOS SOBRE TECNOLOGIA, CIÊNCIA E GESTÃO DA INFORMAÇÃO, 5., 2014, Recife. Anais... Recife: InFoco Consultoria Júnior/UFPE, 2014.

CHOU, L. Top 10 Key Features of BI Tools in 2020. Acessado em 18 de Outubro de 2020. Disponível em <<https://towardsdatascience.com/top-10-keyfeatures-of-bi-tools-in-2020-fd02c1dd4c71>>.

DISTRITO. *Healthtechs*: como elas ajudam na transformação do mercado. Acesso em 22 de outubro de 2020. Disponível em <distrito.me/healthtechs-o-que-sao-e-como-ajudam-na-transformacao-do-mercado/>.

LANEY, D. "The Importance of 'Big Data': A Definition". Gartner. Acessado 29 de Setembro de 2020.

ELMASRI, Ramez; NAVATHE, Shamkant B. *SISTEMAS DE BANCO DE DADOS*. 6. ed. São Paulo: Addison Wesley, 2011

GARTNER GROUP. Disponível em <<https://www.gartner.com/en>>. Acessado em 20 de Junho de 2020.

GARTNER GROUP. *Magical Quadrant for Analytics and Business Intelligence Platforms*. Disponível sob demanda.

GROENFELDT, T. 2013. A Tableau solution to those excel blues. FORBES. Acesso em 22 de outubro de 2020. Disponível em <<https://www.forbes.com/sites/tomgroenfeldt/2013/02/22/a-tableau-solution-to-those-excel-blues/#75cddea675c8>>

HANRAHAN, P.; STOLTE, C; MACKINLAY, J; *Selecting a Visual Analytics Application*. Disponível em <<https://www.tableau.com/pt-br/whitepapers/selecting-visual-analytics-application>>. Acessado em 30 de Junho de 2020.

INCA. IntegradorRHC (Registro Hospitalar do Cancer). Acesso em 12 de maio de 2022. Disponível em <<https://irhc.inca.gov.br/RHCNet/>>

INCA. Registros Hospitalares de Câncer. Disponível em <<https://www.inca.gov.br/numeros-de-cancer/registros-hospitalares-de-cancer-rhc>>. Acesso em 16 de novembro de 2021.

INCA. Notas técnicas sobre RHC. Disponível em <https://irhc.inca.gov.br/files/Notas_tecnicas_final.pdf >. Acesso em 11 de abril de 2022.

JACKSON, T. 2015. 18 key performance indicator examples defined. Clear-Point Strategy Published 5.3.2015. Acessado 29 de Setembro de 2020. Disponível em <https://www.clearpointstrategy.com/18-keyperformance-indicators/>.

MICROSOFT Power BI – Fundamentos. Acesso em 12 de maio de 2022. Disponível em <<https://docs.microsoft.com/pt-br/power-bi/fundamentals/desktop-what-is-desktop>>

MINISTÉRIO DA SAÚDE. Política nacional de informação e informática em Saúde. 2016. Acesso em 12 de maio de 2022.

MORAES I.H.S. Informação em Saúde para o fortalecimento do controle social do SUS. In: Seminário de Comunicação, Informação e Informática em saúde. Brasília: Editora do Ministério da Saúde; 2005.

PINTO, L. F. 2006. Estratégias de integração e utilização de bancos de dados nacionais para avaliação de políticas de saúde no Brasil.

QLIK. Acesso em 12 de maio de 2022. Disponível em < <https://www.qlik.com/pt-br/products> - IDEM>.

SÁ, T. *et al.* Uma Análise Comparativa entre as Ferramentas OLAP como Apoio a Soluções de BI nas Empresas. Em: CONGRESSO TECNOLÓGICO TI E TELECOM – INFOBRASIL, 5., 2012, Fortaleza. Fortaleza: INFOBRASIL, 2012.

SHERMAN, R. (2015). *Business Intelligence Guidebook: From Data Integration to Analytics*. Elsevier and Morgan Kaufmann.

SILBERSCHATZ, Abraham; KORTH, Henry F.; SUDARSHAN, S. SISTEMA DE BANCO DE DADOS. 3. ed. São Paulo: Makron Books, 1999

TABLEAU. Autor desconhecido. *Des Moines Public School District improves intervention programs with predictive analytics*. Disponível em <<https://www.tableau.com/solutions/customer/des-moines-public-school-district-improves-intervention-programs-predictive>>. Acessado em 21 de Junho de 2020.

TABLEAU – Fundamentos. Acesso em 12 de maio de 2022. Disponível em <[tableau.com/pt-br/products/our-platform](https://www.tableau.com/pt-br/products/our-platform)>.

THOUGHTSPOT – “Get answers to questions you didn’t think to ask.” Acesso em 12 de maio de 2022. Disponível em <<https://www.thoughtspot.com/product>>

WALLACE M. A estratégia de dados simples que ajudou o LinkedIn a aumentar sua receita de serviços para empresas em 85%. Disponível em <https://www.tableau.com/sites/default/files/media/Whitepapers/wp_linkedin_pt-br.pdf>. Acessado em 20 de Junho de 2020.

WEAVER, T. et al. 2016. How markets and providers are evaluated in Gartner Magic Quadrant. Acesso em 28 de dezembro de 2020. Disponível em <<https://www.gartner.com/en/documents/3188318/how-markets-and-fornecedores-are-evaluated-in-gartner-magic-q>>