



CENTRO UNIVERSITÁRIO CARIOCA

Verusca Taiane Schimmels da Silva

João Pedro Pessanha Salomão Campello

MINERAÇÃO DE DADOS DO MINISTERIO DA SAÚDE UTILIZANDO  
ORANGE DATAMINING

RIO DE JANEIRO  
Verusca Taiane Schimmels da Silva

João Pedro Pessanha Salomão Campello

## MINERAÇÃO DE DADOS DO MINISTERIO DA SAÚDE UTILIZANDO ORANGE DATAMINING

Trabalho de Conclusão de Curso submetido ao  
Centro Universitário Carioca, como pré-  
requisito para a obtenção do grau de Bacharel  
em Ciência da Computação

Orientador: Prof. D. Sc Sérgio Assunção Monteiro

RIO DE JANEIRO  
Verusca Taiane Schimmels da Silva

João Pedro Pessanha Salomão Campello

## MINERAÇÃO DE DADOS DO MINISTERIO DA SAÚDE UTILIZANDO ORANGE DATAMINING

Trabalho de Conclusão de Curso submetido ao  
Centro Universitário Carioca, como pré-  
requisito para a obtenção do grau de Bacharel  
em Ciência da Computação

Aprovado em \_\_\_\_ de \_\_\_\_\_ de 2023

---

Orientador: Prof. D.Sc. Sérgio Assunção Monteiro  
Centro Universitário Carioca – UNICARIOCA

---

---

Orientador: Prof. D.Sc. Alberto Tavares  
Centro Universitário Carioca – UNICARIOCA

---

---

Prof. M.Sc. André Luiz Avelino Sobral  
Centro Universitário Carioca – UNICARIOCA

---

RIO DE JANEIRO

# Agradecimentos

Gostaria de expressar minha sincera gratidão aos os meus pais, Sergio Alves da silva e Valeria schimmels da silva e minha irmã Rachel schimmels da silva por me apoiarem incondicionalmente em todos os momentos, e por me darem oportunidade de estudar e me desenvolver academicamente. Agradeço aos professores, que gentilmente me forneceu materiais de pesquisa, conselhos e ideias que me ajudaram a aprimorar este trabalho. Gostaria de expressar minha sincera gratidão ao meu orientador, Sérgio Assunção Monteiro, por sua orientação valiosa e pelo suporte que ele me deu durante todo o processo de elaboração deste trabalho. Por fim, gostaria de agradecer a todos que contribuíram, em especial meus colegas de universidade Leonardo Ramos Alves e João Campello que foram importantes para realização deste trabalho, que representa um importante passo em minha jornada acadêmica e profissional.

Verusca Schimmels

# Agradecimentos

Gostaria de agradecer meu pai o Sr “Júnior Campello” por dar todo o suporte financeiro e sempre me incentivando a estudar e buscar meus objetivos e sonhos. Agradecer a minha mãe a Sr “Silvana Pessanha” pelos conselhos e suporte. Um agradecimento especial ao meu Ex padrasto o Sr “Antônio Nérís de Aguiar Filho” que me incentivou a escolher o curso Ciência da Computação e me mostrou toda a trajetória que eu iria encontrar pelo caminho, mostrando que com os estudos e trabalho árduo conquistamos tudo que sonhamos.

Agradecer minha colega de TCC a Verusca silva, pelo nosso trabalho em equipe, e por estar fazendo parte dessa conquista.

João Campello

# Resumo

Minerar os dados do Ministério da Saúde é muito importante para fazer análises e pesquisas em saúde pública. O Orange DataMining é um software livre de mineração de dados que permite a extração e análise de dados de diversas fontes. Neste contexto, o Orange pode ser utilizado para explorar e visualizar dados de saúde, incluindo dados epidemiológicos, dados de mortalidade, entre outros. Processo de extração de dados do Ministério da Saúde utilizando o Orange envolve a coleta dos dados em formato CSV ou XLS, a importação desses dados para o Orange, e a análise exploratória dos dados utilizando diversas técnicas de mineração de dados, como análise de componentes principais, clusterização, e associação de regras. Por meio da extração de dados do Ministério da Saúde com o Orange, é possível identificar padrões e tendências em relação a doenças e outras questões de saúde pública, bem como realizar previsões e análises de risco. Essas informações são de grande valor para os profissionais de saúde pública, governos e pesquisadores na tomada de decisões e desenvolvimento de políticas públicas. O número de casos de covid -19 no Brasil caiu. Porém tivemos, de acordo com dados do Ministério da Saúde, mais de 17 milhões de casos de COVID-19 foram confirmados no Brasil. O número de mortes causadas pelo vírus ultrapassou a marca de 400 mil, o que representa um dos maiores índices de mortalidade no mundo. As regiões mais afetadas pelo COVID-19 no Brasil são o Sudeste e o Nordeste. São Paulo é o estado com o maior número de casos confirmados, seguido pelo Rio de Janeiro, Minas Gerais e Bahia.

**Palavras-chaves:** extração de dados, Ministério da Saúde, Orange DataMining, análise de dados, saúde pública, epidemiologia, mortalidade, padrões, tendências, previsões, risco, COVID-19, casos confirmados, mortes, regiões afetadas, Sudeste, Nordeste, São Paulo, Rio de Janeiro, Minas Gerais, Bahia.

# Abstract

Mining data from the Ministry of Health is very important for conducting analyses and research in public health. Orange Data Mining is a free data mining software that allows for extraction and analysis of data from various sources. In this context, Orange can be used to explore and visualize health data, including epidemiological data, mortality data, among others. The process of extracting data from the Ministry of Health using Orange involves collecting the data in CSV or XLS format, importing it into Orange, and performing exploratory analysis of the data using various data mining techniques such as principal component analysis, clustering, and rule association. By extracting data from the Ministry of Health with Orange, it is possible to identify patterns and trends related to diseases and other public health issues, as well as make predictions and risk analyses. This information is of great value to public health professionals, governments, and researchers in decision-making and policy development. The number of COVID-19 cases in Brazil has decreased. However, according to data from the Ministry of Health, over 17 million cases of COVID-19 have been confirmed in Brazil. The number of deaths caused by the virus has surpassed the 400,000 mark, representing one of the highest mortality rates in the world. The regions most affected by COVID-19 in Brazil are the Southeast and the Northeast. São Paulo is the state with the highest number of confirmed cases, followed by Rio de Janeiro, Minas Gerais, and Bahia.



## SUMÁRIO

### 1.CAPITULO

1.....Erro! Indicador não definido.

1. INTRODUÇÃO.....Erro! Indicador não definido.

1.1 Objetivo do Projeto .....Erro! Indicador não definido.

1.2 Justificativa para escolha do tema .....Erro! Indicador não definido.

1.3 Metodologia .....Erro! Indicador não definido.

1.4 Organização do trabalho .....Erro! Indicador não definido.

**2.CAPITULO 2.....15**

2.Orange Datamining e os Conceitos De Data Science e Machine Learning ..... **Erro! Indicador não definido.**

2.1 Data mining ou Mineração de dados .....Erro! Indicador não definido.

2.2 Data Science .....Erro! Indicador não definido.6

2.3 Machine Learning ..... 17

2.4 Machine Learning Na Area Da Saúde ..... 17

**3.CAPITULO 3 .....15**

3.Importância Do Orange Datamining Na Mineração De Dados ..... 18

3.1 Usabilidade da Ferramenta Na Construção de Gráficos. .... 18

3.2 Aplicação Da Ferramenta Para Mineração De Dados. .... 21

3.3 Ferramenta Orange Datamining..... 24

**4.CAPITULO 4..... 27**

4. Estudo De Caso ..... 28

4.1 .Estudo De Caso Com Orange Datamining .....Erro! Indicador não definido.

4.1 .1 Estudo de casos dados do Brasil .....Erro! Indicador não definido.

4.2 Portal Coronavírus Covid-19..... 37

4.3 Tratamento dos dados ..... 40

<b>5.CAPITULO 5.....</b>	<b>41</b>
5.1 Conclusão .....	41
Bibliografia.....	42

## FIGURAS

Figura 1 Situação epidemiológica da covid 19 abril 2023 Fonte: Secretaria Estadual de saúde.	13
Figura 2 Processo de Mineração de dados Fonte: Marcia e Anny sliderplayer.com .....	16
Figura 3 Apresentação gráfica da ferramenta Orange data mining aonde estão os componentes da criação de base de dados, tabelas fonte :Autor.....	18
Figura 4 Apresentação gráfica dos gráficos gerados através dos dados carregados pela ferramenta Fonte: Autor .....	19
Figura 5 modelos De Data mining, e redes neurais. Fonte: Autor .....	19
Figura 6 Técnicas e modelos criados Fonte: Autor .....	20
Figura 7 modelos de aprendizado. Fonte: Autor .....	20
Figura 8 carregamento dos dados na tabela Fonte: ORANGE DATAMINING .....	21
Figura 9 interface da ferramenta Orange data mining Fonte: ORANGE DATAMINING .....	21
Figura 10 Aplicação para gerar o Mapa de Calor Covid-19 Fonte: ORANGE DATAMINING.....	22
Figura 11 Mapa de calor Fonte: ORANGE DATAMINING .....	23
Figura 12 aplicações para analisar Dados específicos Covid-19 Fonte: ORANGE DATAMINING	23
Figura 13 Dados minerados Covid-19 Brasil Fonte: ORANGE DATAMINING .....	24
Figura 14 Dados de novos casos e de mortes por covid 19 Fonte: O autor.....	25
Figura 15 Número de mortes e novos casos covid 19 Brasil Fonte: O autor .....	25
Figura 16 Gráfico de atividades do estudo de caso FONTE:AUTOR.....	26
Figura 17 Carregando os Dados com Orange Datamining FONTE: Autor .....	27
Figura 18 Número de casos nos municípios do Rio de Janeiro Fonte: Autor .....	28
Figura 19 previsões dos municípios Fonte: autor .....	28
Figura 20 Correlação de dados Fonte: Autor .....	29
Figura 21 gráfico de calor dos municípios do rio de janeiro Fonte: autor .....	30
Figura 22 Analise exploratória dados do município Fonte: Autor .....	30
Figura 23 Figura 23 Gráfico de dispersão Fonte: Autor .....	31
Figura 24 Gráfico de dispersão Fonte: Autor .....	32
Figura 25 Gráfico bar plot Fonte: Autor .....	32
Figura 26 Gráfico bar plot casos Fonte: Autor .....	33
Figura 27 t-sne método de incorporação Fonte: Autor .....	34
Figura 28 t-sne método de incorporação Fonte: Autor .....	34

Figura 29 PCA grau de variância Fonte: Autor .....	35
Figura 30 PCA grau de variância Fonte: Autor .....	35
Figura 31 Gráfico de dispersão dos estados Brasileiros Fonte: Autor .....	36
Figura 32 sites do governo do estado do Rio de janeiro Autor: Secretaria estadual de saúde. .	36
Figura 33 sites do governo do estado do Rio de janeiro Autor: Secretaria estadual de saúde. .	37
Figura 34 Painel coronavírus Brasil Fonte: Secretaria Estaduais de Saúde do Brasil ,2020.....	38
Figura 35 Gráfico de óbitos Fonte: Secretaria Estaduais de Saúde do Brasil ,2020.....	38
Figura 36 Dados retirados do site Fonte: Secretaria Estaduais de Saúde do Brasil ,2020 .....	39
Figura 37 Dados retirados do site Fonte: Secretaria Estaduais de Saúde do Brasil ,2020 .....	39

## **1. INTRODUÇÃO**

A saúde pública sempre foi um tema delicado no Brasil e Mundo. Enfrenta grandes desafios, como falta de investimento em infraestrutura, equipamentos, escassez de recursos humanos qualificados, falta de medicamentos e a crescente demanda por serviços de saúde. Além disso o país tem enfrentado epidemias de doenças como a pandemia do Covid-19. Em 31 dezembro de 2019 a Organização Mundial da pneumonia na cidade de Wuhan, China. Tratava-se de uma nova cepa (tipo) de coronavírus que não havia sido identificada antes em seres humanos. Uma semana depois, em 7 de janeiro de 2020, as autoridades chinesas confirmaram que haviam identificado um novo tipo de coronavírus. (Covid-19). O primeiro caso confirmado de pessoa com o novo coronavírus no Brasil ocorreu em 26 de fevereiro de 2020, de acordo com a (UNASUS, 2020).

A expressão Datamining surgiu no começo da década de 1990 em comunidades de bases de dados, mas desde os anos 60, já se falava desse conceito. A expressão se refere à etapa de análise de dados do processo conhecido como KDD (*Knowledge Discovery in Data-base*) “Descoberta de Conhecimento em Base de Dado”.

O uso dessa tecnologia Datamining é de extrema relevância, através dela conseguimos extrair grandes quantidades de dados, ajuda a controlar número de casos confirmados, número de mortes, pessoas vacinadas. Essa ferramenta ajudou o governo a controlar e tomar decisões sobre a Covid -19, diminuindo o número de casos e mortes.

### **1.1 Objetivo do Projeto**

Este trabalho tem como objetivo apresentar a ferramenta Orange Datamining na extração de dados do ministério da saúde SUS e secretarias estaduais de saúde para analisar as estimativas de novos casos no Brasil nos próximos anos. Orange Datamining é uma opção útil para auxiliar na extração de dados do Ministério da Saúde SUS e do governo do estado do Rio de Janeiro na criação de modelos preditivos de novos casos de COVID-19 no Brasil nos próximos anos. Porém, é importante ter em mente que a precisão das previsões dependerá da qualidade dos dados, como exemplo, nesse trabalho, iremos coletar dados do ministério da saúde de acordo com a figura 1 apresentada abaixo.

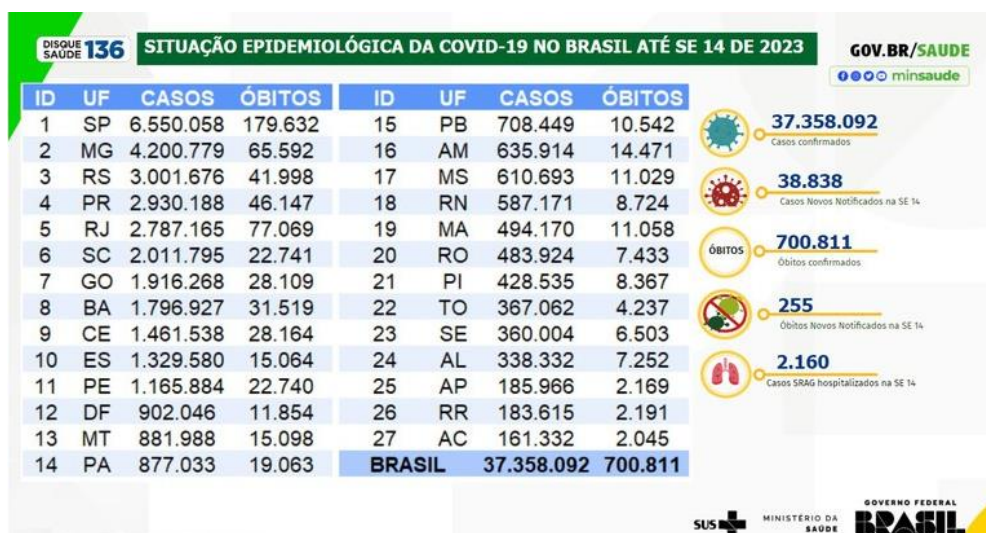


Figura 1 Situação epidemiológica da covid 19 abril 2023 Fonte: Secretaria Estadual de saúde.

## 1.2 justificativa para escolha do tema

Antes da pandemia de COVID-19, a saúde do Brasil já enfrentava desafios significativos. O país tinha um sistema de saúde pública insuficiente para atender às necessidades de uma população de mais de 200 milhões de pessoas. O sistema de saúde privado era caro e inacessível para muitos brasileiros, deixando milhões sem acesso a cuidados básicos de saúde. Além disso, o Brasil enfrentava uma crise financeira que afetou os investimentos em saúde. O país gastava menos de 10% do seu Produto Interno Bruto (PIB) em saúde, o que era considerado insuficiente para atender às necessidades da população.

Apesar desses desafios, houve alguns avanços na área de saúde no Brasil. O país implementou programas de prevenção e controle de doenças, como o Programa Nacional de Imunizações, que ajudou a reduzir a incidência de doenças infecciosas. Segundo o "Programa Nacional de Prevenção e Controle de Infecções Relacionadas à Assistência à Saúde (PNPCIRAS) 2021 a 2025" da Agência Nacional de Vigilância Sanitária (ANVISA), o governo busca promover ações efetivas para prevenir e controlar infecções relacionadas à assistência à saúde (ANVISA, 2021). Além disso, o Brasil conseguiu reduzir significativamente a mortalidade infantil nas últimas décadas.

No entanto a falta de informação sobre a COVID -19 Poderia ter reduzido muito a mortalidade e o tratamento, a mineração de dados pode ser utilizada para auxiliar no monitoramento e controle da pandemia de COVID-19 pela saúde pública verificação de caso, com o objetivo de identificar padrões e tendências epidemiológicas, bem como contribuir para a tomada de decisão em saúde.

## 1.3 Metodologia

Este trabalho realizara coleta de dados do ministério da saúde sobre os casos confirmados, recuperados e óbitos pós pandemia e as localidades com maior índice de infecção e mais atingidos pela pandemia de covid-19 utilizaremos a ferramenta Orange Datamining para ilustrar os dados e relacionar os índices de infecção e relacionar as informações, mostrar gráficos de dispersão.

## **1.4 Organização do trabalho**

Este trabalho está estruturado da seguinte maneira:

No Capítulo 2, será discutido o panorama geral da Covid-19, bem como os conceitos fundamentais de Datamining, Data Science e Machine Learning.

No Capítulo 3, apresentaremos a ferramenta Orange Datamining e suas funcionalidades para a aplicação de gráficos e gerenciamento de bases de dados.

No Capítulo 4, será descrito um estudo de caso que se concentra na utilização de dados do Ministério da Saúde e das Secretarias Estaduais de Saúde Pública, além da aplicação da ferramenta Orange Datamining na obtenção e análise desses dados.

No Capítulo 5, faremos uma conclusão geral do trabalho e apresentaremos algumas sugestões de temas para trabalhos futuros.

## **CAPITULO 2**

### **2. Orange Datamining e os Conceitos De Data Science e Machine Learning**

Coletar dados sobre a pandemia da Covid-19 é fundamental para disseminação da doença e avaliar o impacto das políticas públicas de saúde na prevenção da Covid-19. Mas a coleta manual de dados é demorada e sujeita a erros, então na utilização do Orange data mining ajuda a automatizar o processo.

Através do uso dessa ferramenta se pode coletar dados de diversas fontes, como relatórios oficiais de hospitais e clínicas, e até mesmo dados de mídias sociais. Usar os recursos de pré-processamento de dados do Orange para limpar e formatar os dados transformá-los em um formato padrão para análise. Também pode construir modelos de previsão para ajudar a prever disseminação da doença em diferentes regiões do país. Usar dados históricos para treinar modelos e, em seguida, aplicar dados em tempo real para fazer previsões precisas.

Graças ao uso dessa ferramenta que é capaz de coletar e analisar dados com muito mais rapidez e precisão do que manualmente, isso nos permite tomar melhores decisões e se prevenir da disseminação da COVID-19 e proteger a saúde pública.

#### **2.1 Datamining ou Mineração de Dados**

A mineração de dados tem suas raízes na década de 1960, quando empresas começaram a coletar e armazenar grandes quantidades de dados em computadores. No entanto, o termo "mineração de dados" só começou a ser usado na década de 1990, quando os pesquisadores começaram a desenvolver técnicas para extrair informações úteis desses grandes conjuntos de dados. Durante os anos 90, a mineração de dados era usada principalmente em áreas como finanças e marketing, para ajudar as empresas a identificar padrões em seus dados que pudessem ser usados para melhorar a tomada de decisões. Mas à medida que a tecnologia

evoluiu e a quantidade de dados disponíveis aumentou, a mineração de dados se tornou uma, ferramenta cada vez mais importante em áreas como ciência, medicina e governo.

Hoje, a mineração de dados é uma das principais áreas da inteligência artificial e está sendo usada para uma ampla variedade de aplicações, desde a previsão de eventos climáticos até a análise de padrões de consumo em tempo real.

De acordo com (Fayyad,1996) a mineração de dados é uma técnica que pode ser usada para extrair conhecimento útil a partir de grandes conjuntos de dados. Ele acredita que a mineração de dados pode ser usada para descobrir padrões e tendências que podem ser usados para tomar decisões mais informadas em várias áreas, como negócios, ciência e governo.

Fayyad também destaca a importância da qualidade dos dados para a mineração de dados; Ele enfatiza que a qualidade dos dados é essencial para garantir que as informações extraídas sejam precisas e úteis, a necessidade de uma abordagem sistemática para a mineração de dados que inclui a seleção cuidadosa de dados relevantes, pré-processamento e limpeza de dados e a aplicação de técnicas de mineração de dados apropriadas para obter resultados significativos. Além disso, Fayyad também discute a importância da privacidade dos dados e da ética na mineração de dados. Ele acreditava que a mineração de dados deve ser realizada de forma responsável e ética, respeitando a privacidade dos indivíduos e cumprindo as leis e regulamentações aplicáveis.

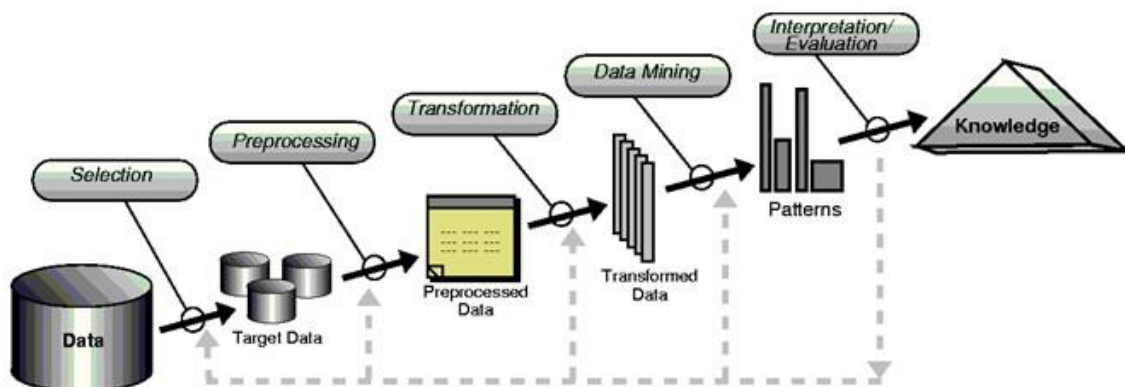


Figura 2 Processo de Mineração de dados Fonte: Marcia e Anny sliderplayer.com

O Datamining é uma ferramenta de gerenciamento de informações cada vez mais popular que tem como objetivo identificar estruturas de conhecimento para guiar decisões em condições de incerteza. Para atingir esse objetivo, técnicas analíticas têm sido desenvolvidas especificamente para lidar com questões relacionadas ao Datamining; No entanto, o Datamining ainda é baseado em princípios conceituais da Análise de Dados Exploratória e modelagem.

O processo de Datamining faz parte de um processo maior de conhecimento chamado Knowledge Discovery in Database (KDD). Esse processo envolve a estruturação do banco de dados, seleção, preparação e pré-processamento dos dados, transformação, adequação e redução da dimensionalidade dos dados, processos de Datamining e análise, assimilação, interpretação e uso do conhecimento extraído do banco de dados.



## **2.2 Data Science**

Com a evolução da internet no século XXI, permitiu que através de dispositivos eletrônicos pessoas, instituições e empresas possam se comunicar em todo mundo. Tendo o crescimento exponencial na quantidade de dados gerados pelo mundo.

O aumento da quantidade de dados disponíveis, surgiu a necessidade de encontrar formas eficientes de armazenar, processar e analisar esses dados para gerar insights úteis, assim, a ciência de dados se tornou uma disciplina essencial para empresas e organizações que buscam tomar decisões baseadas em dados, otimizar seus processos e melhorar seus resultados.

A ciência de dados surgiu como uma disciplina interdisciplinar que combina conhecimentos de áreas como estatísticas, ciência da computação e matemática aplicada para extrair insights e conhecimentos a partir de dados.

## **2.3 Machine Learning**

O Machine Learning é um ramo da IA que utiliza algoritmos e técnicas estatísticas para ensinar as máquinas a reconhecer padrões em grandes volumes de dados. O processo de Machine Learning envolve a coleta de dados, o treinamento de modelos e a geração de previsões ou classificações. O objetivo é permitir que as máquinas aprendam e melhorem sua performance ao longo do tempo, sem a necessidade de programação explícita. Uma das principais características do Machine Learning é a sua capacidade de lidar com dados complexos e identificar padrões que podem não ser perceptíveis aos humanos. Por exemplo, em um projeto de diagnóstico médico, um modelo de Machine Learning pode analisar milhares de registros de pacientes e identificar correlações sutis entre sintomas e doenças, o que pode ser difícil para um médico identificar apenas com base na experiência clínica.

Existem diversos tipos de algoritmos de Machine Learning, como os supervisionados, não supervisionados, de reforço e deep learning. Os algoritmos supervisionados são treinados com dados rotulados, ou seja, com exemplos de entrada e suas respectivas saídas esperadas. Já os algoritmos não supervisionados não têm informações prévias e são usados para agrupar ou identificar padrões nos dados. Os algoritmos de reforço são usados em situações onde um agente precisa aprender a tomar ações em um ambiente para maximizar uma recompensa. E o deep learning é uma subárea do Machine Learning que utiliza redes neurais artificiais com múltiplas camadas para aprendizado de representações complexas. Aplicações do Machine Learning são amplas e variadas, na medicina, por exemplo, pode ser usado para auxiliar em diagnósticos, identificar riscos em pacientes, planejar tratamentos personalizados e monitorar a eficácia de medicamentos. Na indústria, pode ser utilizado para otimizar a produção, prever falhas em equipamentos e melhorar a eficiência energética. No marketing, pode ser aplicado para análise de dados de consumidores, personalização de campanhas publicitárias e recomendação de produtos.

## **2.4 Machine Learning na Área da saúde**

Desde o início da pandemia de COVID-19, o aprendizado de máquina (ML) tem desempenhado um papel importante no combate à propagação do vírus e no tratamento dos pacientes. Com o volume de dados e informações disponíveis sobre o COVID-19, os algoritmos de ML podem

ser usados para analisar e identificar padrões nos dados, ajudando os profissionais de saúde a tomar decisões mais precisas e informadas. Uma das aplicações mais promissoras da IA e do ML na luta contra o COVID-19 é o diagnóstico.

Os algoritmos de ML podem ser treinados em grandes conjuntos de dados de pacientes infectados com o vírus, permitindo que os médicos identifiquem rapidamente os sintomas e tomem medidas de isolamento e tratamento. Isso é particularmente importante em casos de transmissão assintomática, onde os pacientes não apresentam sintomas óbvios, mas ainda podem transmitir o vírus. Além disso, os algoritmos de ML podem ser usados para analisar dados epidemiológicos e identificar padrões de propagação do vírus. Isso pode ajudar os governos e autoridades de saúde a tomar decisões informadas sobre políticas de distanciamento social, restrições de viagem e alocação de recursos médicos.

Outra aplicação importante de IA e ML na luta contra o COVID-19 é a pesquisa de medicamentos, com o uso de algoritmos de ML, os pesquisadores podem analisar grandes conjuntos de dados de moléculas e compostos para identificar candidatos a medicamentos promissores que possam tratar o vírus; Isso pode acelerar significativamente o processo de descoberta de medicamentos, ajudando a salvar vidas e a conter a pandemia.

## **CAPITULO 3**

### **3. Importância Do Orange Data Mining Na Mineração De Dados**

A mineração de dados com o Orange Data Mining tem se mostrado uma ferramenta valiosa na análise dos dados da COVID-19, oferecendo uma abordagem visual e interativa que permite explorar, modelar e compreender informações relevantes relacionadas a essa pandemia global. A quantidade de dados disponíveis sobre a COVID-19 é imensa, incluindo informações sobre casos confirmados, mortes, hospitalizações, testes, além de dados demográficos e epidemiológicos. O Orange Data Mining facilita a importação e integração desses dados de fontes, possibilitando a identificação de diferentes padrões e tendências importantes para o enfrentamento dessa crise. Uma das grandes vantagens do Orange é a sua capacidade de lidar com grandes volumes de dados de forma eficiente. Com uma pandemia, é fundamental analisar uma enorme quantidade de informações em tempo hábil. O Orange oferece diversas técnicas de análise de dados, como visualização, estatística, aprendizado de máquina e mineração de texto, que permitem a identificação de insights relevantes.

A proposta é mostrar a forma gráfica da ferramenta na extração de dados em estados e municípios do país, realizando simulações que ajudem na tomada de decisão priorizando as visualizações dos dados para melhor tomada de decisão.

#### **3.1 Usabilidade da ferramenta na construção de gráficos.**

O Orange Data Mining é uma ferramenta open source que permite criar projetos de Data Mining sem escrever código, sendo ideal para quem quer praticar machine learning sem programar ou para iniciantes que desejam aplicar conceitos. Com sua interface drag and drop, o usuário pode arrastar elementos até a tela e aplicar funções a eles, simplificando o processo de construção do projeto. A figura 3 apresenta um exemplo da interface e suas funcionalidades

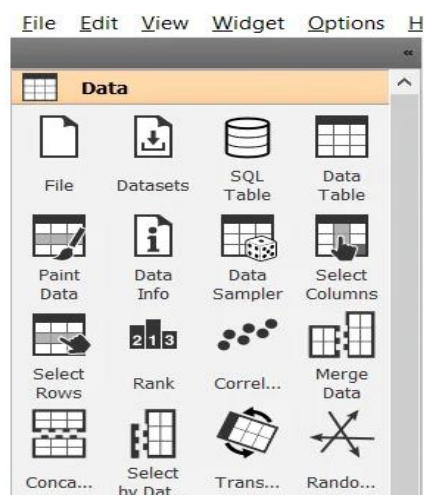


Figura 3 Apresentação gráfica da ferramenta Orange data mining aonde estão os componentes da criação de base de dados, tabelas fonte :Autor

Na figura 3 Cada ícone (ou widget) representa uma ferramenta para ser utilizado no processo de carga e preparo dos dados. É possível desde incluir dados ausentes, criar features, discretizar variáveis contínuas entre tantas outras técnicas comumente aplicadas;

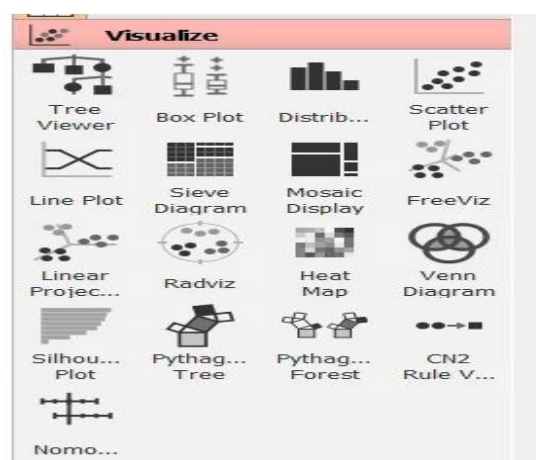


Figura 4 Apresentação gráfica dos gráficos gerados através dos dados carregados pela ferramenta Fonte: Autor

Na figura 4 Como o nome sugere, trata-se do grupo de ferramentas para visualização de dados. É possível fazer gráficos de dispersão, box-plots, gráficos de linha e etc.

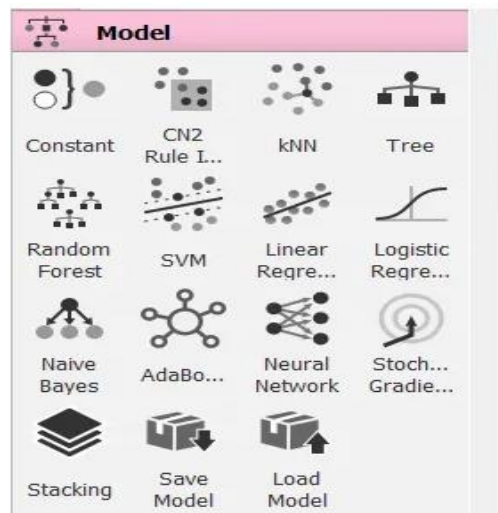


Figura 5 modelos De Data mining, e redes neurais. Fonte: Autor

Na figura 5 Essa aba contém os principais modelos que podemos utilizar em Datamining. Todos eles são modelos supervisionados. Temos desde os modelos mais simples, como a Regressão Linear, até modelos mais complexos, como Redes Neurais;

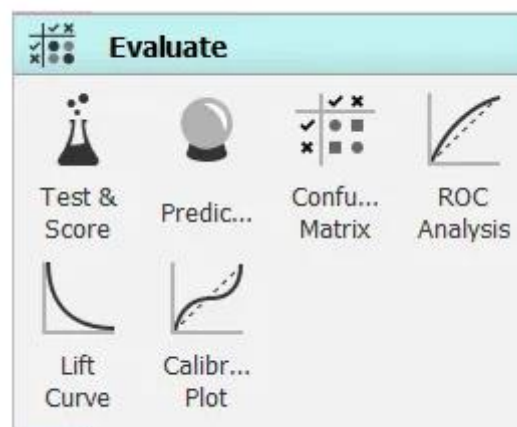


Figura 6 Técnicas e modelos criados Fonte: Autor

Na figura 6 Aqui podemos contar com as técnicas para avaliar o desempenho dos modelos criados. Matriz de confusão, análise AUC e validação cruzada estão disponíveis aqui;



Figura 7 modelos de aprendizado. Fonte: Autor

Na figura 7 Por fim, temos os modelos de aprendizado não supervisionado como K-means e alguns algoritmos comumente utilizados para transformação dos dados, como por exemplo a PCA. Para carregar um conjunto de dados na ferramenta e simples basta clicar a sua

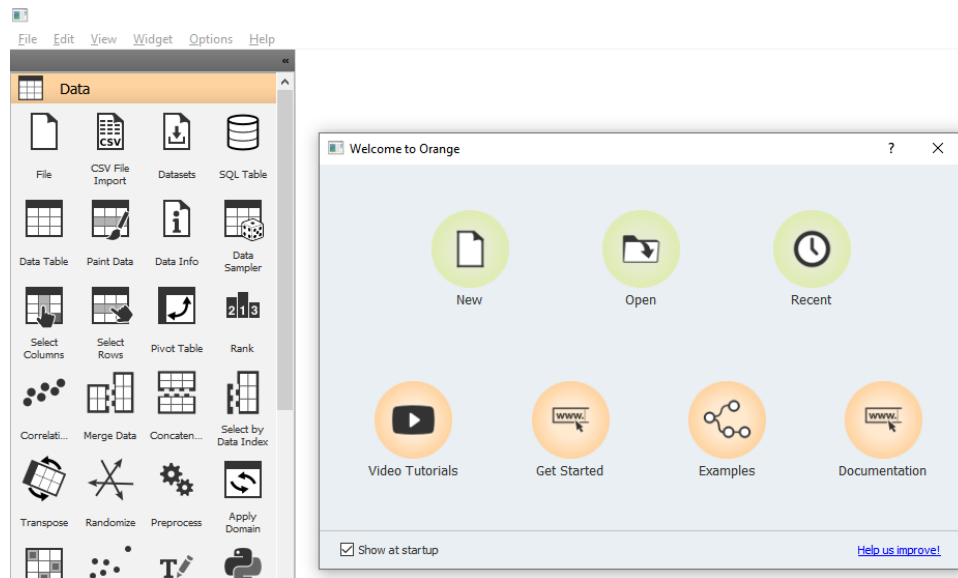
Para carregar um conjunto de dados na ferramenta e simples basta clicar a sua esquerda na aba no widget “File” e selecionar o conjunto de dados na pasta onde ele está salvo no caso os dados serão captados do site do ministério da saúde, após importar os dados, podemos visualizá-los de forma tabular. Isso também é muito simples. Basta clicar no ícone “Data Table”, e em seguida conectá-lo com “File”, então, clicamos duas vezes em “Data Table” e pronto e já podemos ter a visualização e as conexões dos dados.



Figura 8 carregamento dos dados na tabela Fonte: ORANGE DATAMINING

### 3.2 Aplicação da ferramenta para mineração de dados.

A seguir estão algumas aplicações da ferramenta Orange data mining minerando dados da Covid-19:



*Figura 9 interface da ferramenta Orange data mining Fonte: ORANGE DATAMINING*

A interface do Orange é composta por vários componentes:

**Tela de dados:** A área principal onde os usuários podem arrastar e soltar conjuntos de dados e executar várias operações de análise de dados.

**Widgets:** Ferramenta e funções pré-construídas que podem ser adicionadas à tela para executar várias tarefas, como visualização de dados, agrupamento, classificação, regressão e muito mais.

**Caixa de Ferramenta:** uma coleção de widgets categorizados por tipo de dados ou tipo de análise, o que facilita a localização de widget certo para uma tarefa específica.

**Janela de saída:** Exibe os resultados da análise, incluindo tabelas, gráficos e estatísticas.

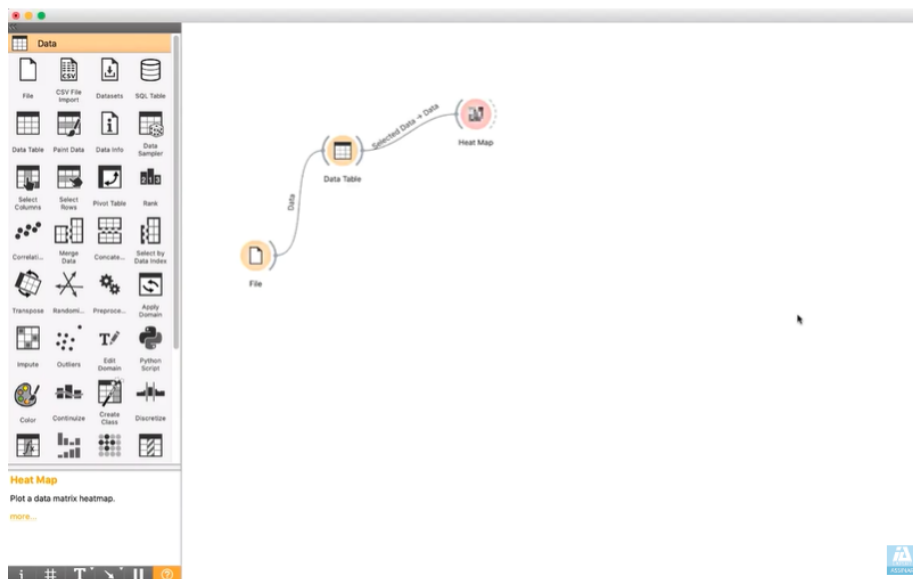


Figura 10 Aplicação para gerar o Mapa de Calor Covid-19 Fonte: ORANGE DATAMINING

Para gerar um mapa de Calor, crie o fluxo de trabalho, carregue os dados em um mapa de calor, adicione um widget de “Mapa de Calor” arrastando-o do painel de widgets à esquerda e soltando no espaço de trabalho, conecte o widget de “Mapa de Calor” aos dados do ministério da saúde soltando o nó de dados no widget de “Mapa de Calor”, configure as opções do widget, execute o fluxo de trabalho e visualize o mapa de calor gerado pelo widget de “Mapa de Calor”

Exemplo abaixo, indicando os maiores valores a direita na cor amarelo informando o último registro e na esquerda o primeiro registro na cor azul.

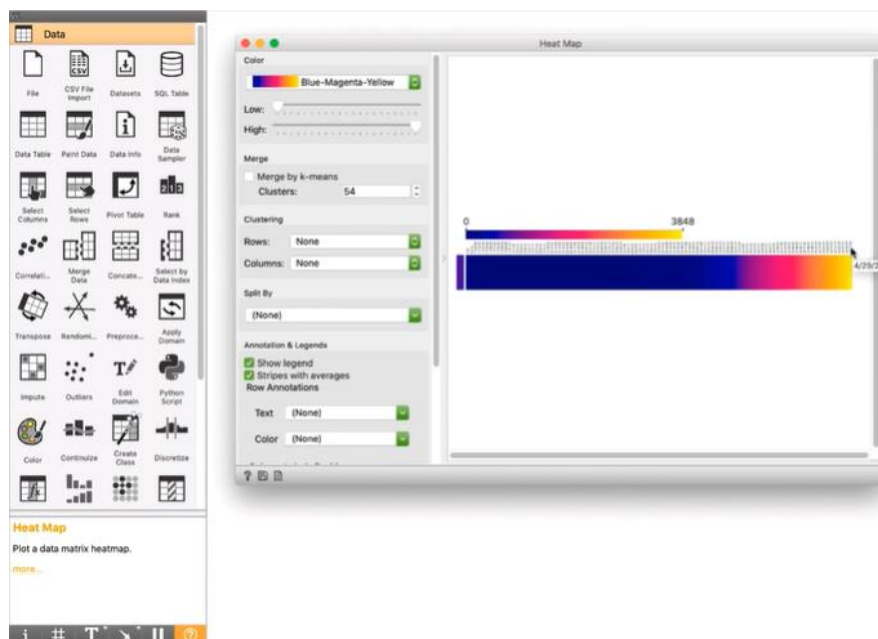


Figura 11 Mapa de calor Fonte: ORANGE DATAMINING

O mapa de Calor é uma ferramenta de visualização de dados que ajuda a identificar padrões e tendências em conjuntos de dados. Ele pode ser usado para exibir a distribuição dos valores de uma variável em relação a outras variáveis, ajudando a entender a correlação entre elas. O mapa de calor usa uma escala de cores para representar os valores de uma variável em diferentes partes do conjunto de dados. Os valores mais altos são geralmente representados por cores mais escuras e os valores mais baixos por cores mais claras. Isso permite visualizar facilmente onde ocorrem os valores mais altos e baixos e identificar as áreas do conjunto de dados que requerem mais atenção.

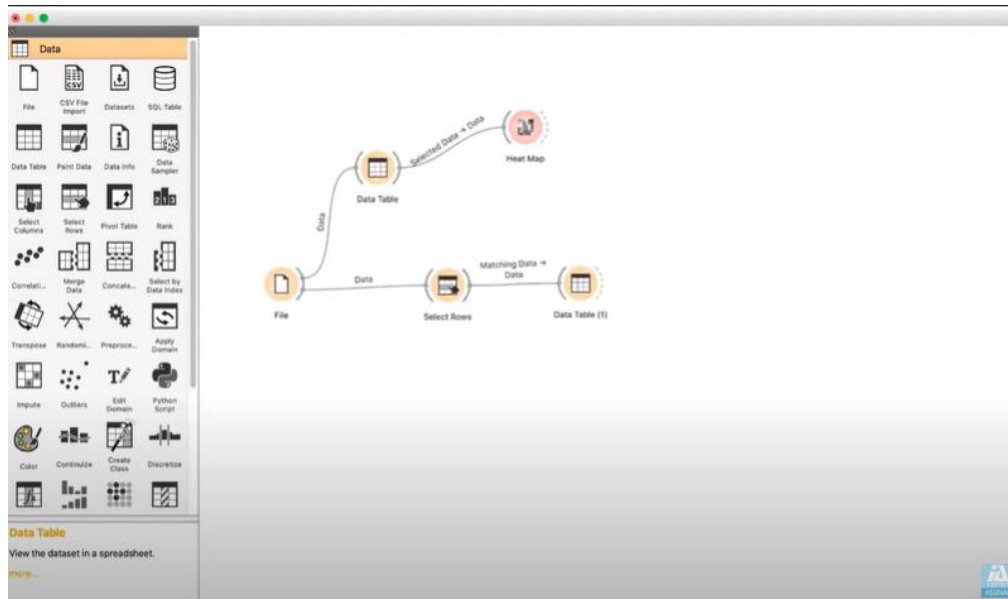


Figura 12 aplicações para analisar Dados específicos Covid-19 Fonte: ORANGE DATAMINING

Utilizando a função “select Rows” permite selecionar um subconjunto específico de linhas em um conjunto de dados maior, é possível criar condições personalizadas para selecionar apenas linhas que atendem critérios específicos, permitindo uma análise mais detalhada de subconjuntos específicos de dados.

Exemplo abaixo, análise específica do País Brasil sobre os dados covid-19



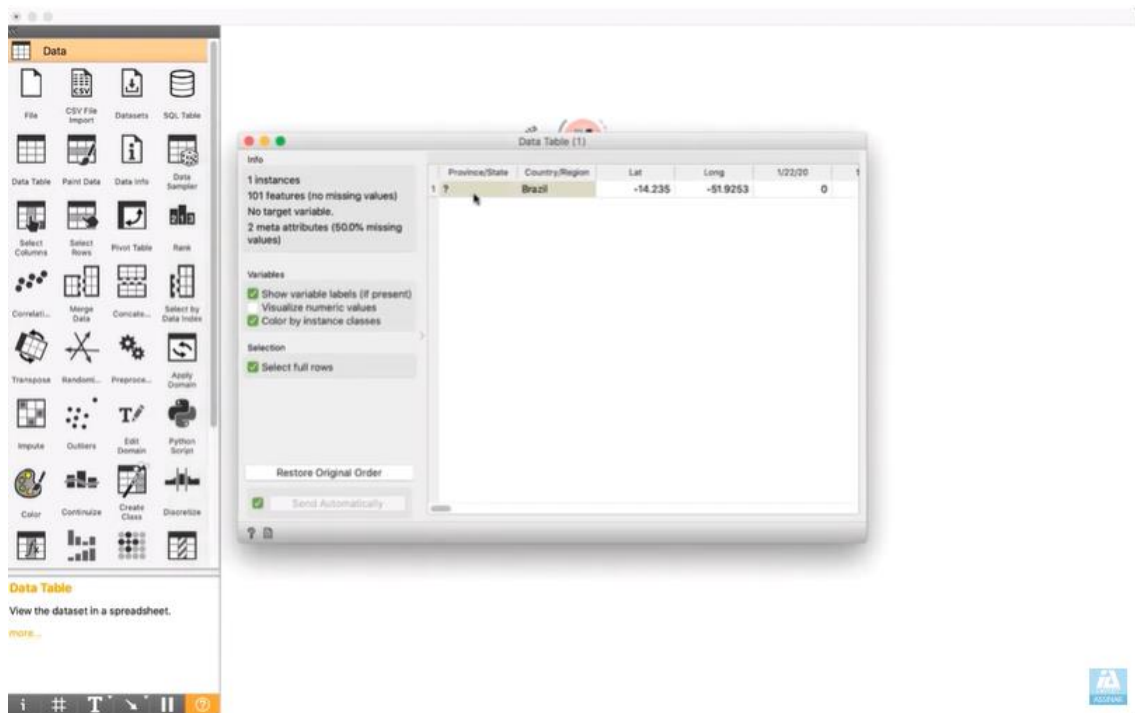


Figura 13 Dados minerados Covid-19 Brasil Fonte: ORANGE DATAMINING

### 3.3 Ferramenta Orange Datamining

A ferramenta Orange Datamining é uma plataforma de software de análise de dados que permite aos usuários realizar diversas atividades relacionadas à análise de dados, desde a preparação e visualização dos dados até a modelagem preditiva e a aprendizagem de máquina. Com o objetivo de auxiliar na exploração e interpretação de dados de diversas fontes, a ferramenta Orange possibilita a realização de tarefas como limpeza de dados, eliminação de valores ausentes, seleção de recursos relevantes, análise exploratória de dados, modelagem preditiva e aprendizagem de máquina.

A plataforma oferece muitas técnicas de modelagem preditiva, como regressão linear, árvores de decisão, redes neurais e análise de séries temporais, permitindo aos usuários criar modelos para prever resultados futuros com base em dados históricos. Além disso, com as diversas técnicas de aprendizado de máquina disponíveis, como clustering, classificação e associação, é possível automatizar tarefas e identificar padrões em dados complexos. Outra vantagem da ferramenta Orange é a integração com outras ferramentas de análise de dados, como Python e R, o que permite aos usuários utilizar as funcionalidades dessas ferramentas diretamente na interface do Orange. Portanto, o Orange Data Mining é uma ferramenta poderosa e flexível para análise de dados, que oferece diversas opções para explorar, interpretar e visualizar dados, possibilitando a criação de modelos preditivos e a automação de tarefas por meio de técnicas de aprendizado de máquina.

Tomada de decisão é o processo de escolha entre opções disponíveis, envolvendo a identificação do problema, coleta de informações, avaliação das alternativas e seleção da melhor opção com base em critérios específicos. A importância da tomada de decisão com o Orange Data Mining reside no fato de que ele permite que os usuários tomem decisões informadas com base em evidências, em vez de depender de suposições ou intuição, além

disso, o Orange Data Mining ajuda a identificar padrões e relações complexas nos dados que podem ser óbvios à primeira vista. Com sua interface gráfica fácil de usar e recursos avançados. O Orange Data Mining é uma ferramenta valiosa para qualquer pessoa que precise tomar decisões com base de dados. Desde cientista de dados e analistas de negócios até pesquisadores acadêmicos e estudantes, essa ferramenta é útil para a tomada de decisão informada e eficaz, Veja abaixo gráficos gerados pela ferramenta que mostram a redução de casos no mundo da covid 19.

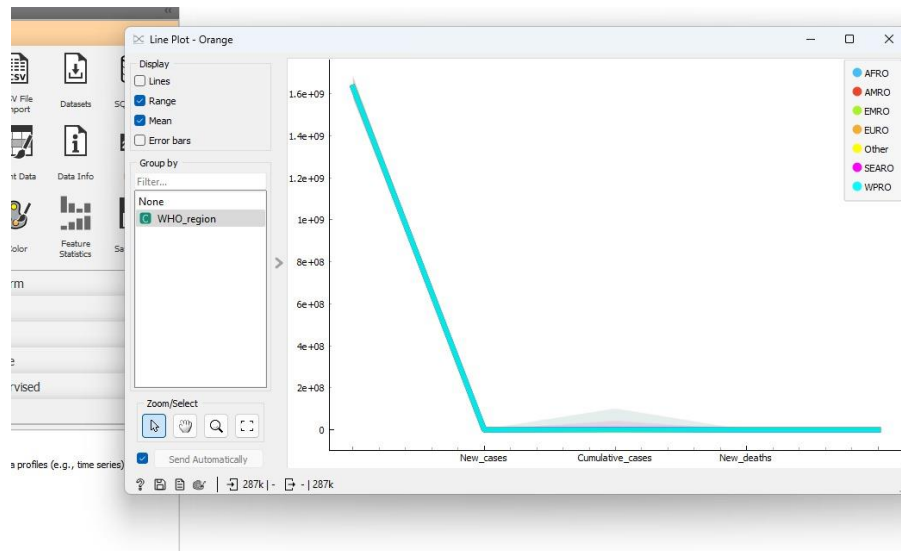


Figura 14 Dados de novos casos e de mortes por covid 19 Fonte: O autor

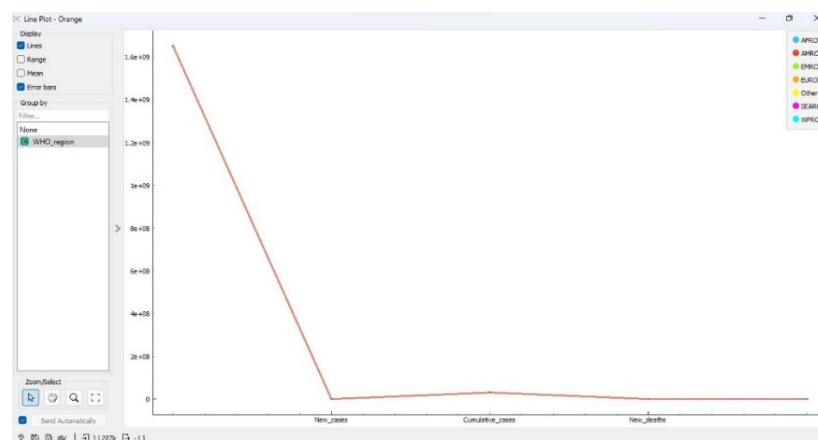


Figura 15 Número de mortes e novos casos covid 19 Brasil Fonte: O autor

## CAPITULO 4

### 4. ESTUDO DE CASOS

O objetivo deste estudo de caso é analisar dados relacionados à COVID-19 para entender melhor a disseminação da doença, as áreas mais afetadas e as tendências ao longo do tempo. Para realizar a pesquisa do covid 19 no Brasil após a retirada do status de pandemia pela organização mundial da saúde (OMS), foi necessário coletar os dados do Ministério da saúde e do governo do estado do Rio de janeiro e preparar os dados para que pudesse ser gerado na ferramenta Orange Datamining. Através da ferramenta iremos mostrar gráficos da pandemia dos estados brasileiros, números encontrados de casos, o número de recuperados e óbitos nas regiões do município carioca, abaixo iremos relacionar as atividades a serem apresentadas de acordo com diagrama de atividades.

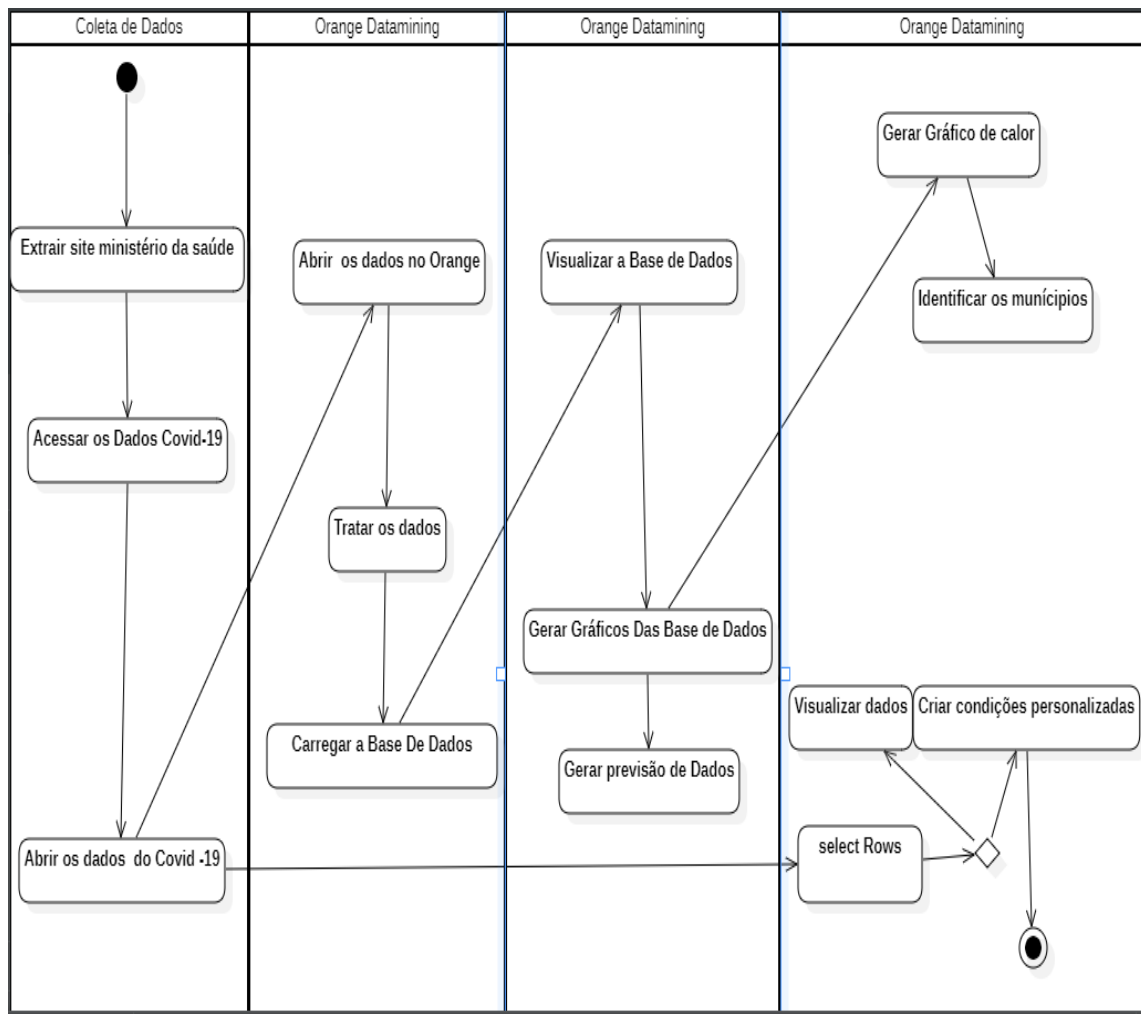


Figura 16 Gráfico de atividades do estudo de caso FONTE:AUTOR

#### 4.1. ESTUDO DE CASOS COM ORANGE DATAMINING

Na Figura 16, é apresentada a utilização da ferramenta Orange Datamining, que foi empregada nesse processo. Para carregar os dados, a opção "file" na aba "data" foi selecionada e os dados foram carregados a partir de um arquivo CSV. Em seguida, os dados foram carregados na aba "data table", onde foi possível estruturar os dados em uma base de dados de acordo com o tipo.

Todo conteúdo deste estudo de caso está disponível no GitHub, através do endereço <https://github.com/Vesilva33/Orange-Data-mining>

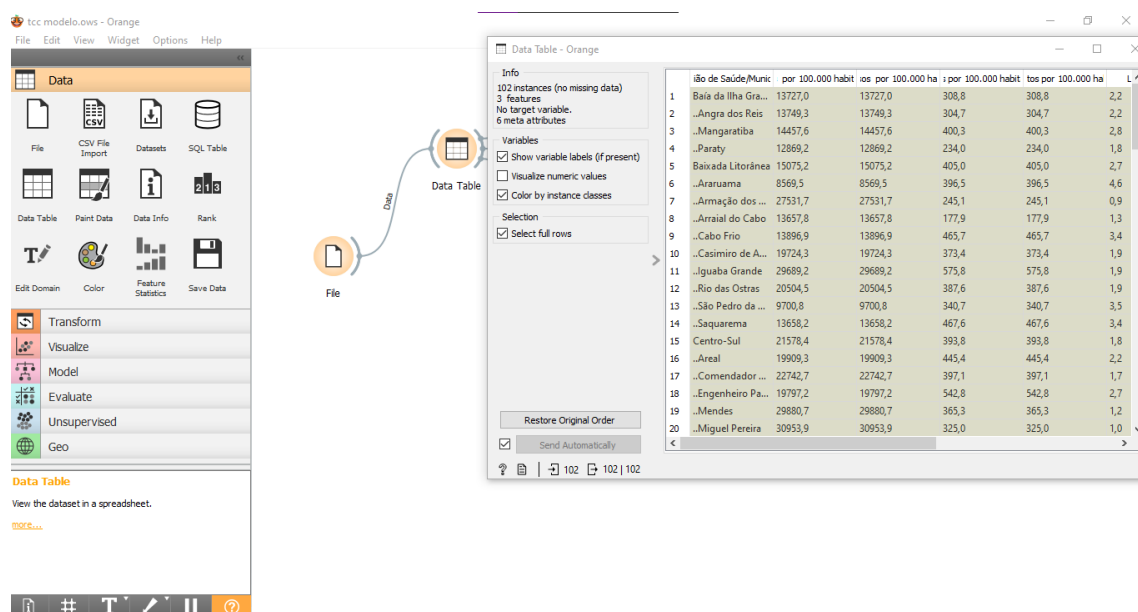


Figura 17 Carregando os Dados com Orange Datamining FONTE: Autor

Após o carregamento dos dados podemos abrir a aba data table e visualizar toda a base de dados carregada dentro do aplicativo, após o carregamento de dados podemos gerar a visualização dos municípios e gráficos com os dados da população no caso a base de dados foi gerada com a população de 2019 nos municípios do estado do rio de janeiro, ao clicar em data table podemos anexar a aba bar plot e gerar um gráfico que mostra nos municípios os casos e óbitos e infecção por 100.000 de habitantes conforme a figura 18 , Podendo alterar para verificação da letalidade do vírus nas regiões, óbitos nas regiões o maior índice.

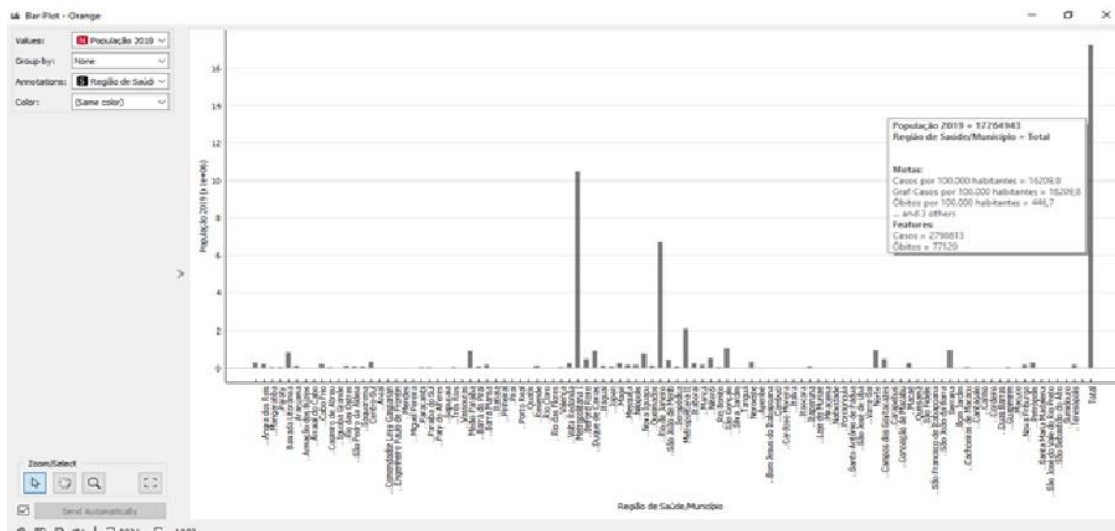


Figura 18 Número de casos nos municípios do Rio de Janeiro Fonte: Autor

Com a ferramenta Orange Datamining e possível gerar e mostra as previsões dos modelos sobre os dados, O widget recebe um conjunto de dados e um ou mais preditores (modelos preditivos, não algoritmos de aprendizado). Ele produz os dados e as previsões, adicionando a aba predictions na base de dados e carregando ela exibe os dados e ao clicar no município correspondente ela gera previsões de dados como esta exibido abaixo.

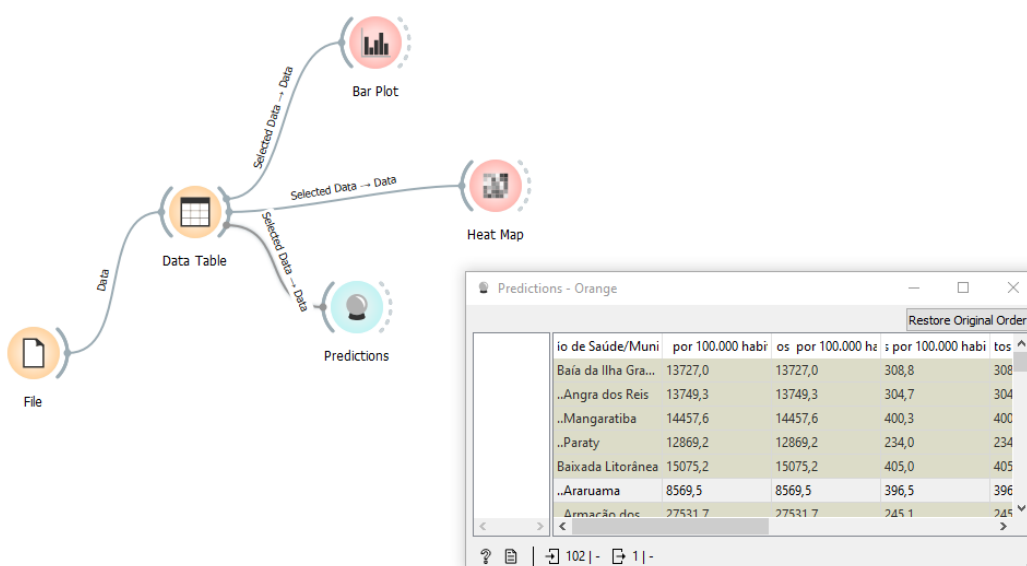


Figura 19 previsões dos municípios Fonte: autor

Além das previsões a ferramenta nos possibilita a realizar as correlações, determinar (mensurar) o grau de relacionamento entre duas variáveis; A análise de correlação só pode ser realizada em recursos numéricos contínuos. Para ilustrar esse conceito, utilizaremos o conjunto de dados de habitação como exemplo. Importe-o no widget Arquivo e conecte-o ao widget Correlações. Os pares de recursos com correlação positiva serão exibidos no topo da lista, enquanto os pares com correlação negativa serão exibidos na parte inferior.

Dirija-se ao par com a correlação mais negativa, DIS-NOX. Em seguida, conecte o widget de Gráfico de Dispersão ao widget de Correlações e defina as saídas para Data to Data e Features to Features. Observe como o par de recursos é instantaneamente plotado no gráfico de dispersão. Parece que os dois recursos realmente têm uma correlação negativa.

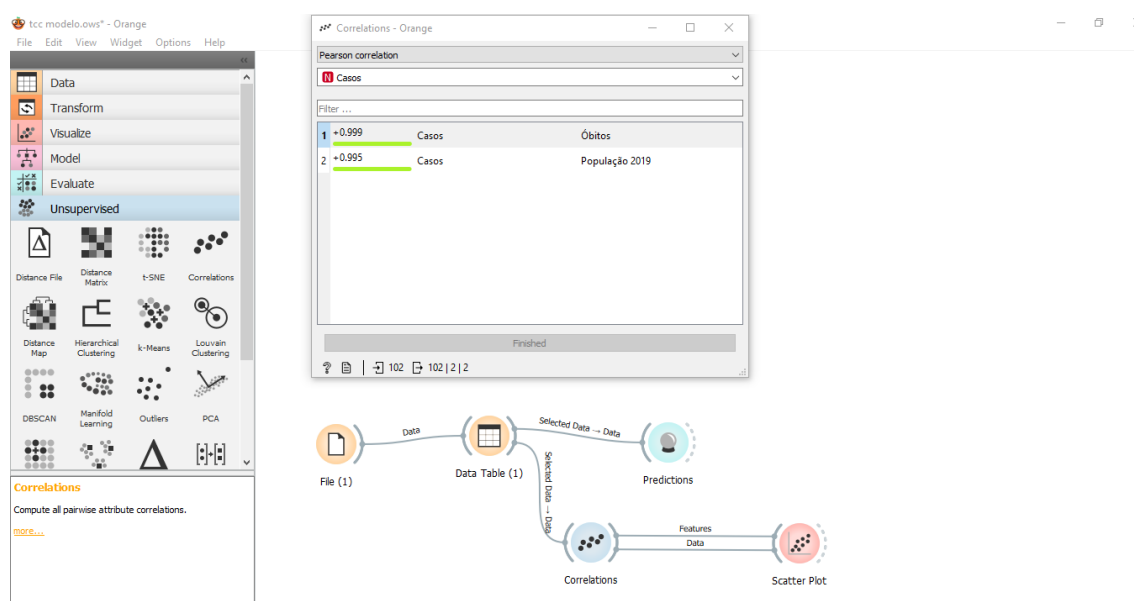


Figura 20 Correlação de dados Fonte: Autor

Na figura 19 podemos ver o gráfico de calor dos municípios do Rio de Janeiro, O heat map é uma ferramenta gráfica altamente eficaz para visualizar os valores dos atributos em uma matriz bidimensional. Esse método funciona exclusivamente em conjuntos de dados que contêm variáveis numéricas, e os valores são representados por cores, escolhidas de uma paleta de cores personalizada. Ao plotar as variáveis de classe e os atributos nos eixos x e y, é possível identificar com facilidade as áreas onde os valores dos atributos são mais fortes e mais fracos, permitindo a identificação de características típicas para cada classe. De maneira geral, o heat map é uma ferramenta valiosa para explorar padrões e tendências em grandes conjuntos de dados com ela podemos visualizar os dados do município do rio que apresentam casos mais intensos e, mas infecção.

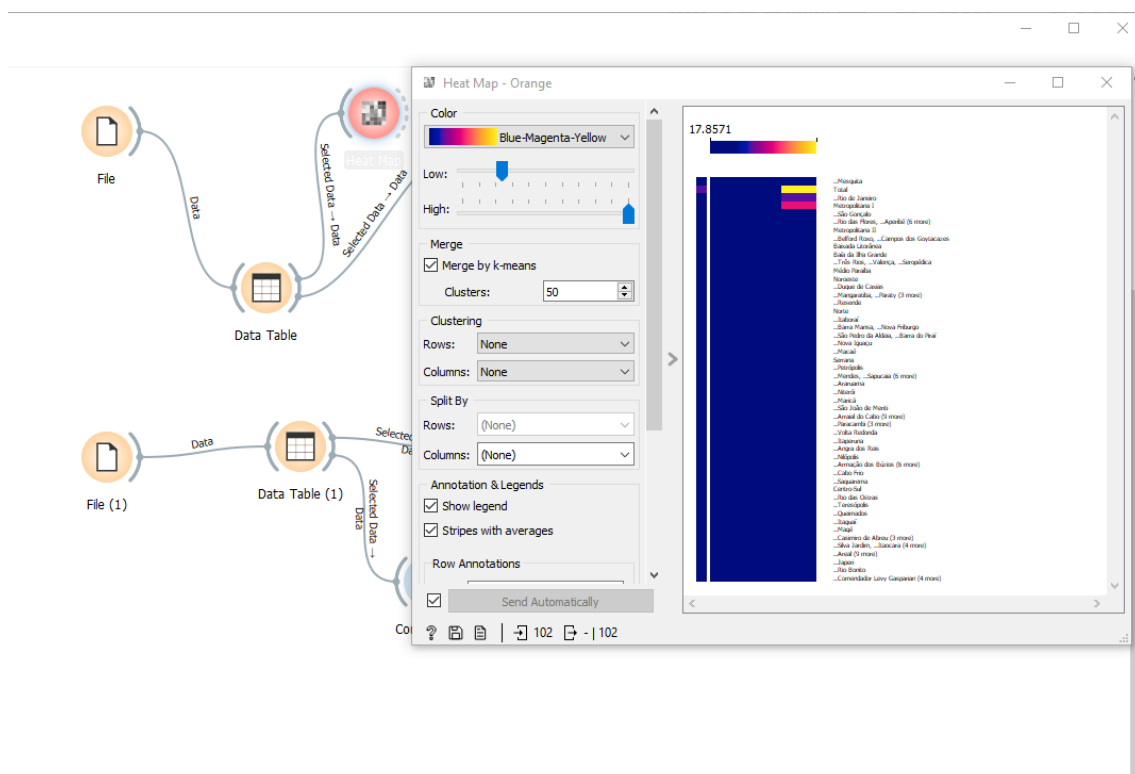


Figura 21 gráfico de calor dos municípios do rio de janeiro Fonte: autor

Na figura 22, podemos observar a função “select full Rows” permite selecionar todas as colunas de um conjunto, visualizar todas as instâncias, entender a distribuição dos valores, identificar outliers ou realizar análises exploratórias, analisando o exemplo acima a região de cabo frio, buscando uma análise mais detalhada da região selecionada.

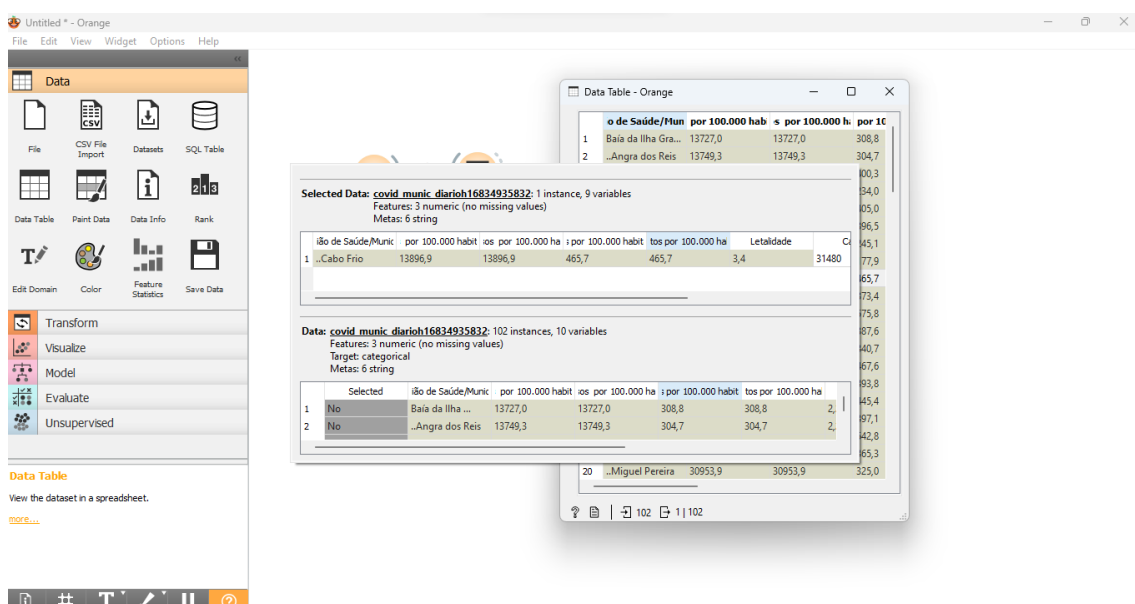


Figura 22 Analise exploratória dados do município Fonte: Autor

É uma ferramenta de visualização que permite criar gráficos de dispersão bidimensionais. Esses gráficos exibem dados como uma coleção de pontos, em que o valor de um atributo determina a posição no eixo horizontal (x) e o valor de outro atributo determina a posição no eixo vertical (y). No gráfico, é possível ajustar várias propriedades, como cor, tamanho e forma dos pontos, títulos dos eixos, tamanho máximo do ponto e "jittering", na figura 21 abaixo podemos visualizar os municípios afetados com os casos e o que se destacam, mas no índice de casos e mortes, clicando nas setas abaixo podemos ver melhor os números de óbitos como na figura 23 os dados do gráfico podem ser verificados na própria tela.





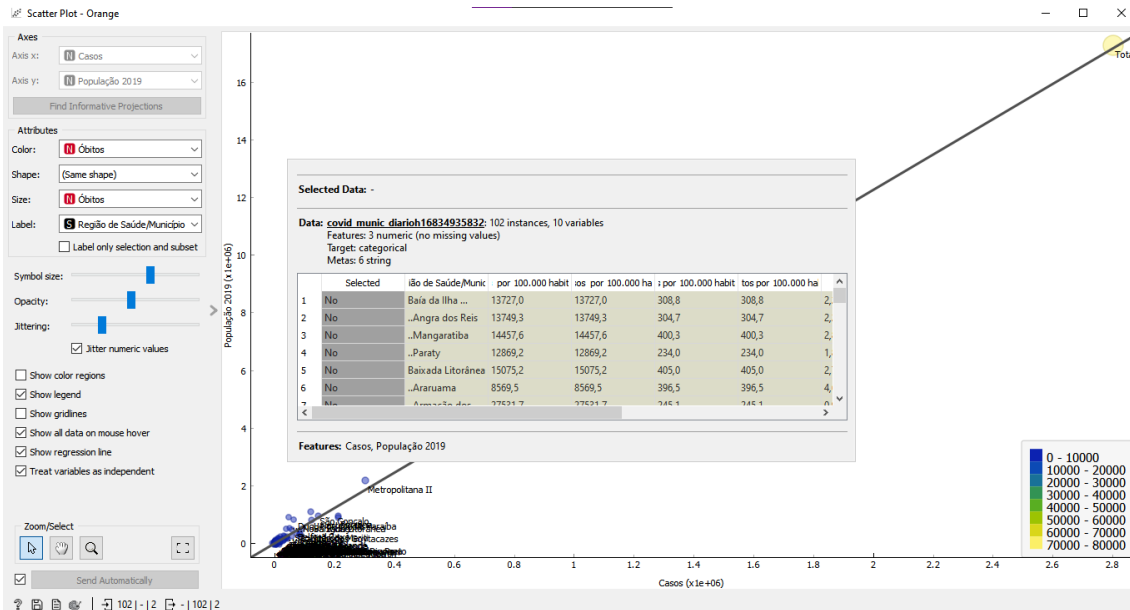


Figura 24 Gráfico de dispersão Fonte: Autor

#### 4.1.1 Estudo de casos dados do Brasil

A pandemia da Covid-19 continua a impactar o Brasil e o mundo, exigindo a coleta constante de dados para entender e monitorar. Foi essencial para pesquisa obter os números de casos confirmados e mortes relacionados à Covid-19. Esses dados foram fundamentais para avaliar a gravidade da pandemia. Obter informações atualizadas através de fontes confiáveis, como o Ministério da Saúde do Brasil, secretarias estaduais de saúde e órgãos internacionais de saúde, como a Organização Mundial da Saúde (OMS), nos ajuda a realizar uma pesquisa de qualidade. Através desses dados conseguimos acompanhar as taxas de transmissão para entender a velocidade de propagação do vírus. Podemos visualizar abaixo com os dados do Brasil os estados que ainda tem índice preocupantes de transmissão do covid-19 segundo a ferramenta Orange Datamining.

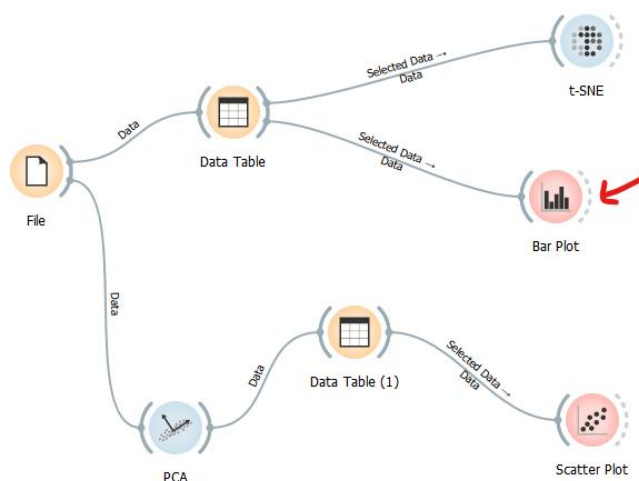


Figura 25 Gráfico bar plot Fonte: Autor

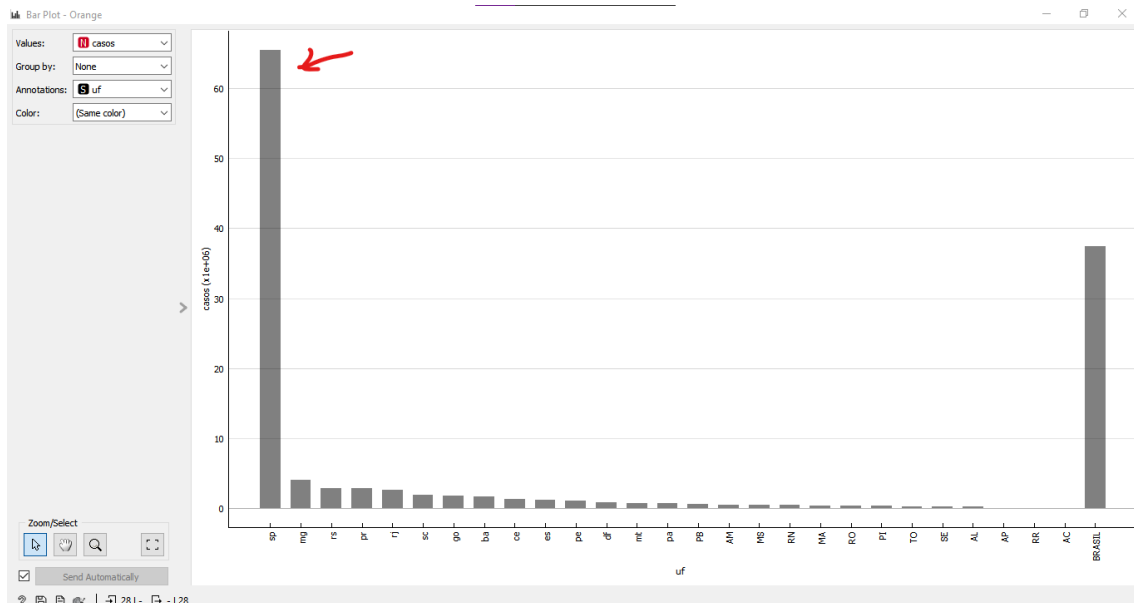


Figura 26 Gráfico bar plot casos Fonte: Autor

O widget t-SNE é usado para plotar os dados utilizando um método de incorporação de vizinho estocástico distribuído denominado t-SNE. Essa técnica de redução de dimensionalidade, semelhante ao MDS, tem como objetivo mapear os pontos para um espaço bidimensional (2D) com base na sua distribuição de probabilidade, utilizamos essa técnica para visualizar os estados e seus índices de casos e óbitos podendo também analisar esses dados em relação a outros estados do Brasil e seus índices conforme na figura 27 e figura 28 que mostra o estado de são Paulo com maior índice de casos.

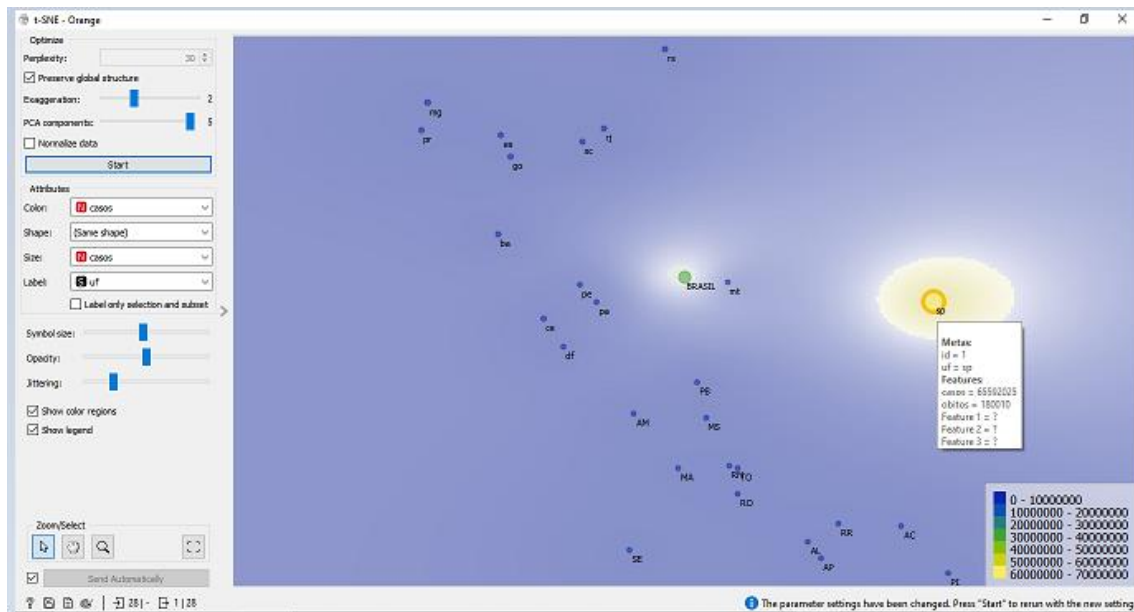


Figura 27 t-sne método de incorporação Fonte: Autor

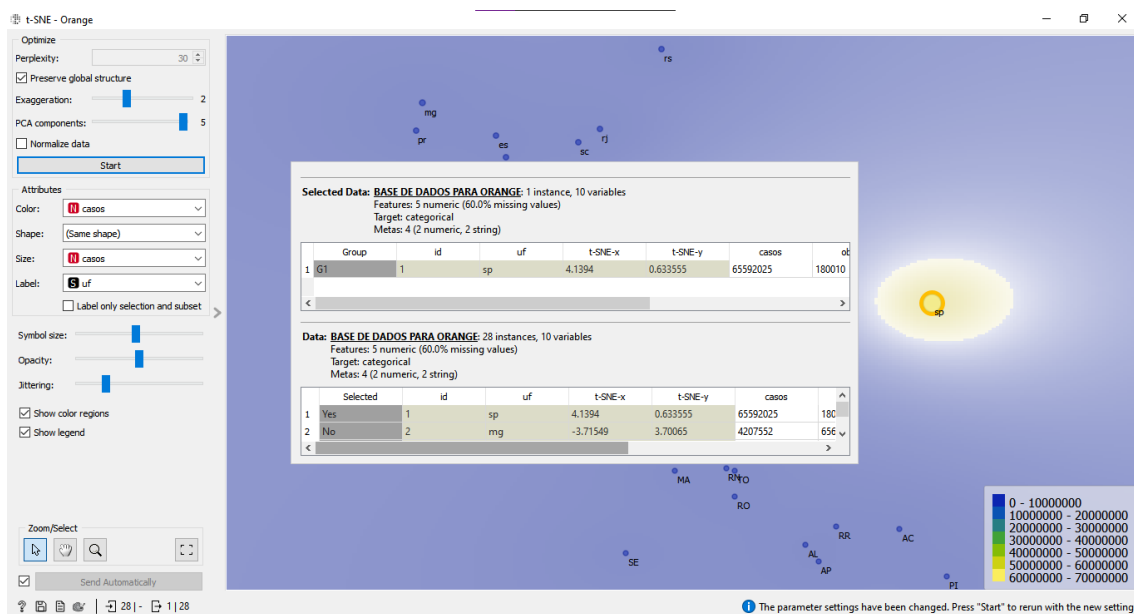


Figura 28 t-sne método de incorporação Fonte: Autor

O PCA (Análise de Componentes Principais) é uma técnica que transforma os dados em um novo conjunto de variáveis não correlacionadas, conhecidas como componentes principais. O widget PCA apresenta um gráfico que mostra o grau de variância explicada pelos melhores componentes principais e permite ao usuário interativamente definir o número de componentes a serem incluídos no conjunto de dados resultante. Neste fluxo de trabalho, é possível observar a transformação tanto na tabela de dados quanto no gráfico de dispersão, na figura 29 podemos visualizar esse exemplo do gráfico.



Figura 29 PCA grau de variância Fonte: Autor

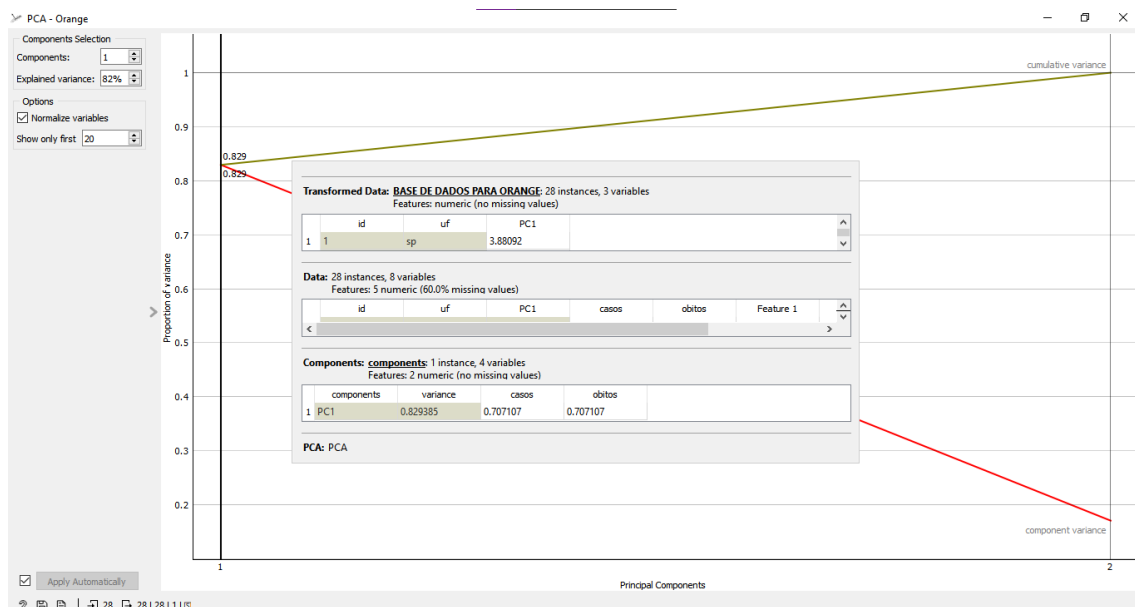


Figura 30 PCA grau de variância Fonte: Autor

E com o gráfico de dispersão podemos pôr fim identificar os estados do Brasil Que elevam o número de casos Acompanhar essas informações é crucial para entender a situação atual da pandemia e contribuir para a tomada de decisões informadas para proteger a saúde e o bem-estar da população, É importante acompanhar as taxas de transmissão para entender a velocidade de propagação do vírus. A taxa de reprodução ( $R_0$ ) é um indicador-chave, representando o número médio de pessoas infectadas por um indivíduo infectado, conforme a figura abaixo podemos visualizar no gráfico os casos em todos os estados do Brasil.

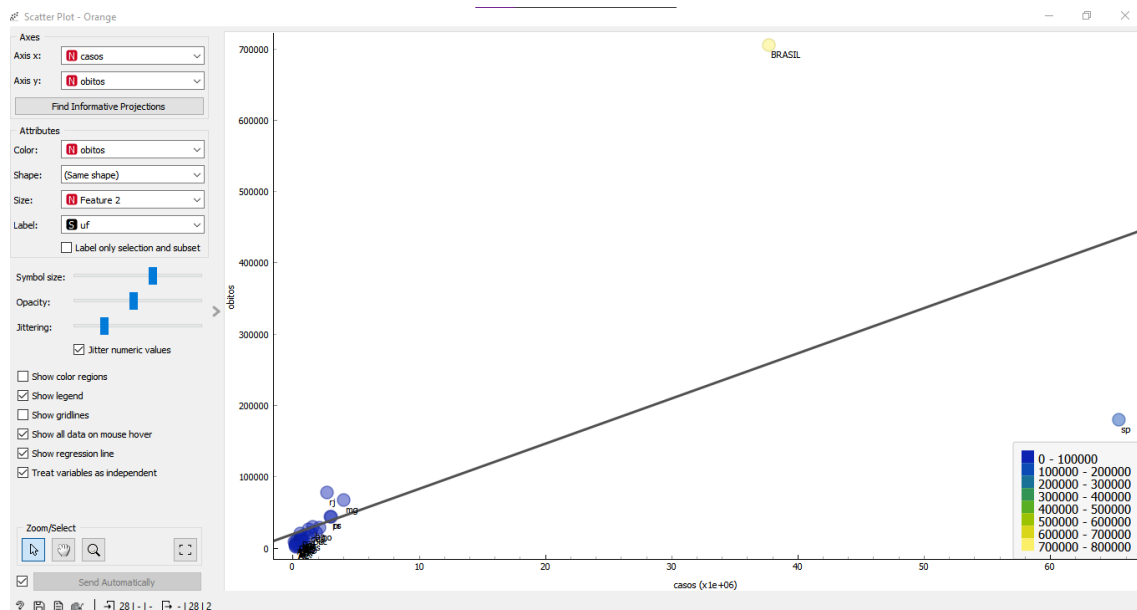


Figura 31 Gráfico de dispersão dos estados Brasileiros Fonte: Autor

## 4.2 Portal Coronavírus Covid-19

Os dados do covid 19 foram retirados do site do governo do estado do rio de janeiro painel coronavírus covid-19 da secretaria estadual de saúde do estado do rio de janeiro, mostrando os casos de confirmados, óbitos, incidência por 100.000 habitantes, letalidade da doença e as internações por covid nos municípios do estado do rio de janeiro, conforme ilustrado na figura 32 <https://painel.saude.rj.gov.br/monitoramento/covid19.html>

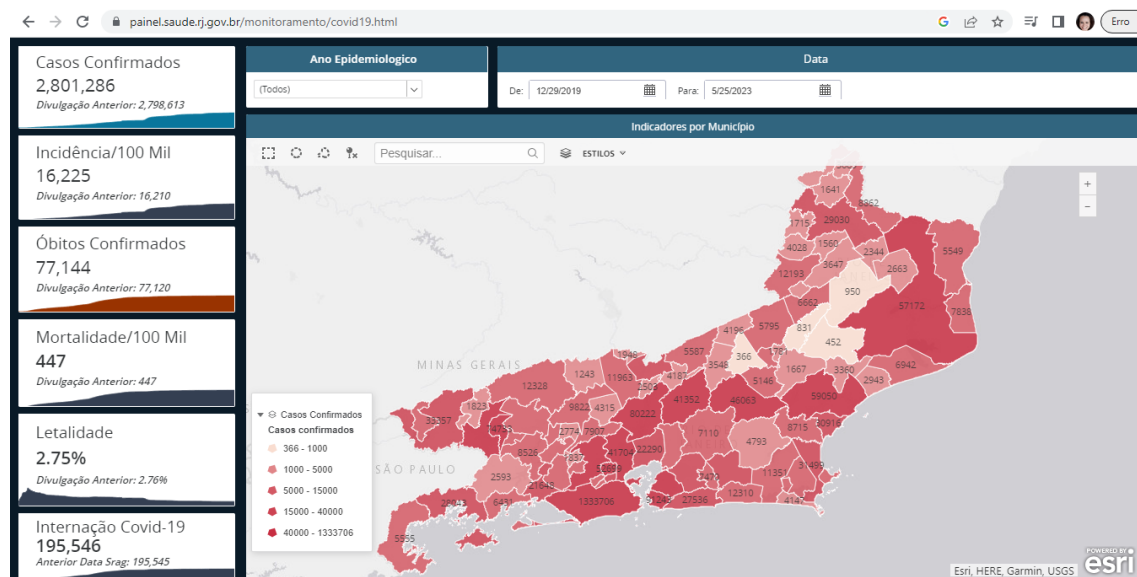


Figura 32 sites do governo do estado do Rio de janeiro Autor: Secretaria estadual de saúde.

Como citado acima os dados foram extraídos do portal coronavírus disponibilizados toda a segunda-feira de acordo com o ministério da saúde do estado do Rio de janeiro desde 2019 até 2023 mesmo após o estado de pandemia ter sido atualizado pela OMS ainda continuamos com dados atualizados pela secretaria estadual de saúde conforme mostra a figura 28 casos confirmados, óbitos confirmados e a letalidade da pandemia de covid 19 ainda são uma preocupação para o estados e municípios.

#### 4.2.1 Coleta dos Dados

A coleta de dados no portal do coronavírus do estado do Rio de Janeiro foi crucial para monitorar e compreender o impacto da pandemia na região. O portal, criado pelas autoridades de saúde, oferece uma fonte centralizada de informações atualizadas sobre casos confirmados, óbitos, taxas de casos e outras estatísticas relacionadas ao COVID-19. O sistema nos permitiu baixar as informações necessárias arquivo formato xlsx ou CVS (Valores separados por Vírgula), além de disponibilizar dados atualizados semanalmente.

Vale ressaltar que a coleta de dados no portal do coronavírus do estado do Rio de Janeiro segue protocolos e diretrizes rigorosas de privacidade e proteção de dados pessoais. Todas as informações são tratadas com o devido sigilo e anonimato, garantindo a segurança e a confidencialidade das pessoas envolvidas.

Ao acessar o portal, é possível acompanhar de forma transparente e confiável os dados relacionados ao COVID-19, contribuindo para a saúde e bem-estar da população fluminense, conforme na figura 33 abaixo podemos filtrar ano epidemiológico e verificar os dados de casos confirmados mortalidade e internação por covid 19 no estado do rio de janeiro

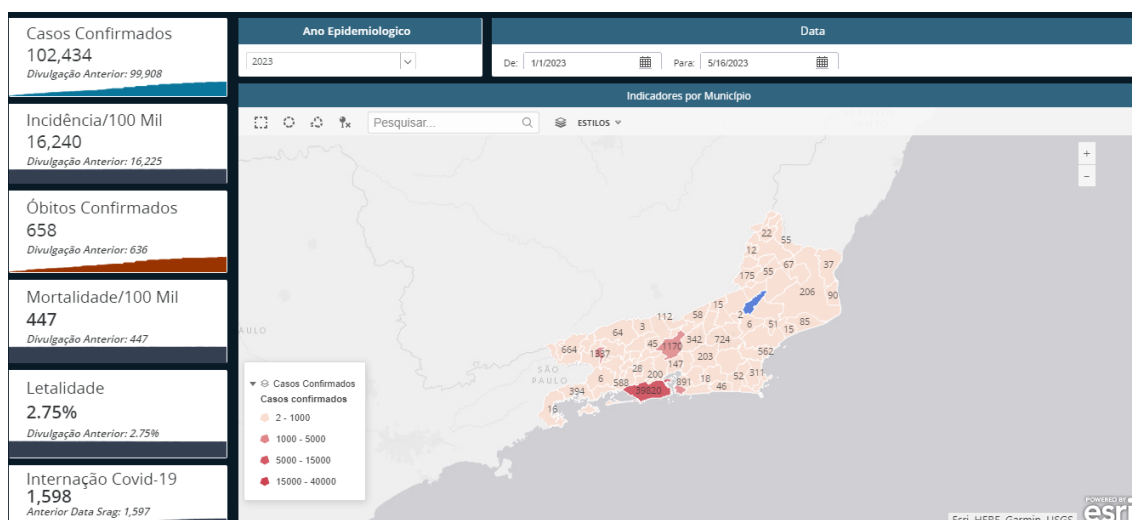


Figura 33 sites do governo do estado do Rio de Janeiro Autor: Secretaria estadual de saúde.

Os dados também foram retirados do ministério da saúde sus que disponibiliza dados do Brasil e seus estados o painel coronavírus foi criado para ser um veículo oficial de comunicação sobre a situação epidemiológica da Covid 19 no Brasil.

O processo de atualização das informações nos municípios, estados e esfera federal é um procedimento dinâmico e complexo. É importante ressaltar que os dados divulgados diariamente estão sujeitos a alterações. Isso se deve à diversidade existente entre os municípios brasileiros, que possuem diferenças em termos de tamanho populacional, infraestrutura e organização dos serviços de saúde. Além disso, é preciso considerar os desafios enfrentados pela pandemia de COVID-19. Nesse contexto, é possível ocorrer variações no número de casos confirmados e óbitos devido a eventuais erros ou atrasos no repasse das informações segundo o próprio **site do ministério da saúde**.

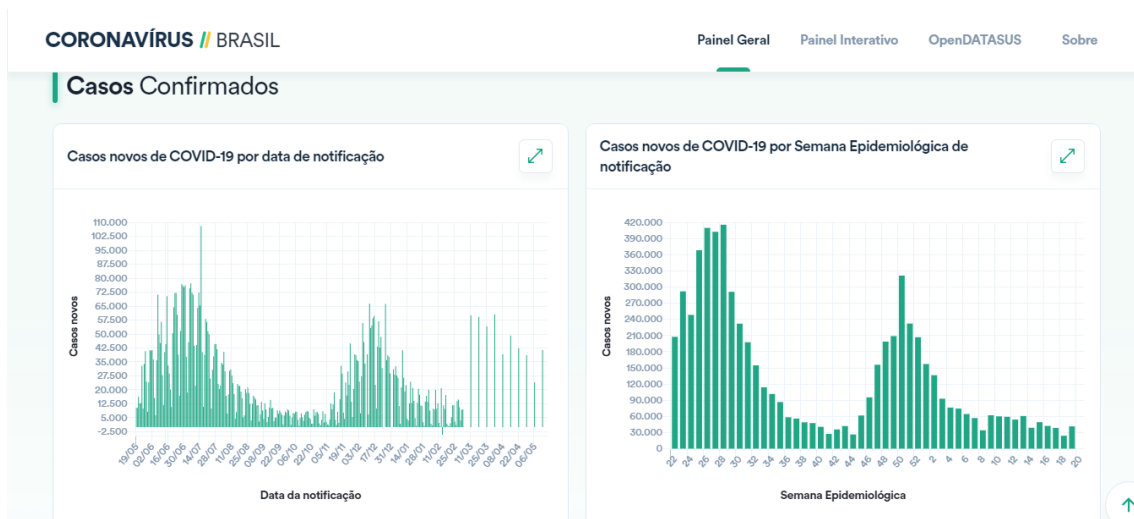


Figura 34 Painel coronavírus Brasil Fonte: Secretaria Estaduais de Saúde do Brasil ,2020

Em nível municipal, é importante destacar que o dado do dia atual pode ser menor do que o do dia anterior. Isso ocorre devido ao fato de que o município responsável pela notificação não necessariamente corresponde ao local de residência do caso ou óbito registrado. Portanto, ao realizar a investigação individual de cada ocorrência, as Secretarias Municipais e Estaduais de Saúde têm autonomia para corrigir as informações e repassá-las ao Ministério da Saúde.

É relevante ressaltar que os casos e óbitos são atualizados com base na data de notificação. O Ministério da Saúde está trabalhando em conjunto com as Secretarias Estaduais e Municipais de Saúde para divulgar esses indicadores de forma colaborativa, como podemos analisar nos gráficos disponibilizados **SITE PAINEL CORONAVIRUS** (Página que contém os dados do ministério da saúde).

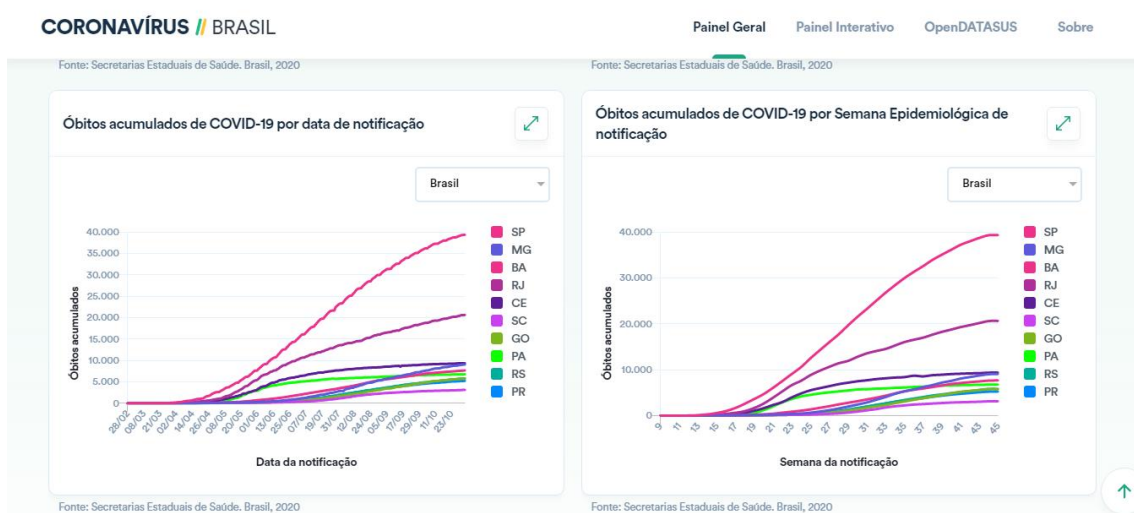


Figura 35 Gráfico de óbitos Fonte: Secretaria Estaduais de Saúde do Brasil ,2020

### 4.3 Tratamento dos dados.

A figura 36 mostra os dados brutos extraídos do site do ministério da saúde secretaria estadual foi realizado a preparação e a transformação dos dados brutos em um formato adequado para análise, na ferramenta Orange Datamining.

Baía da Ilha Grande	40003	900	13727	13727	308,8	308,8	291418	2,2
..Angra dos Reis	28019	621	13749,3	13749,3	304,7	304,7	203785	2,2
..Mangaratiba	6429	178	14457,6	14457,6	400,3	400,3	44468	2,8
..Paraty	5555	101	12869,2	12869,2	234	234	43165	1,8
Baixada Litorânea	126625	3402	15075,2	15075,2	405	405	839958	2,7
..Araruama	11346	525	8569,5	8569,5	396,5	396,5	132400	4,6
..Armação do Braz	9325	83	27531,7	27531,7	245,1	245,1	33870	0,9
..Arraial do Cabo	4145	54	13657,8	13657,8	177,9	177,9	30349	1,3
..Cabo Frio	31480	1055	13896,9	13896,9	465,7	465,7	226525	3,4
..Casimiro de Abreu	8715	165	19724,3	19724,3	373,4	373,4	44184	1,9
..Iguaba Grande	8405	163	29689,2	29689,2	575,8	575,8	28310	1,9
..Rio das Ostras	30895	584	20504,5	20504,5	387,6	387,6	150674	1,9
..São Pedro da Aldeia	10135	356	9700,8	9700,8	340,7	340,7	104476	3,5
..Saquarema	12179	417	13658,2	13658,2	467,6	467,6	89170	3,4

Figura 36 Dados retirados do site Fonte: Secretaria Estaduais de Saúde do Brasil ,2020

Para garantir a correta elaboração dos gráficos, foi necessário realizar ajustes nos títulos, destacando-os, além de corrigir as informações nos municípios que apresentavam pontos no início da planilha. Para facilitar a visualização dos dados, as colunas foram ampliadas, e os IDs foram removidos da planilha na figura 37 podemos verificar. Além disso, foram implementadas outras técnicas para aprimorar o processo de análise e apresentação dos dados, já com os dados do ministério da saúde em relação aos estados foi necessário a criação da base de dados extraindo manualmente as informações pois se tratava de dados muito misturados com datas que não estavam de acordo com o tempo do estudo apresentado.

Região de Saúde/Município	Casos	Óbitos	Casos por 100.000 habitantes	Gráf. Casos por 100.000 habitantes	Óbitos por 100.000 habitantes	Gráf. Óbitos por 100.000 habitantes	População 2019	Letalidade
Baía da Ilha Grande	40003	900	13727	13727	308,8	308,8	291418	2,2
Angra dos Reis	28019	621	13749,3	13749,3	304,7	304,7	203785	2,2
Mangaratiba	6429	178	14457,6	14457,6	400,3	400,3	44468	2,8
Paraty	5555	101	12869,2	12869,2	234	234	43165	1,8
Baixada Litorânea	126625	3402	15075,2	15075,2	405	405	839958	2,7
Araruama	11346	525	8569,5	8569,5	396,5	396,5	132400	4,6
Armação do Braz	9325	83	27531,7	27531,7	245,1	245,1	33870	0,9
Arraial do Cabo	4145	54	13657,8	13657,8	177,9	177,9	30349	1,3
Cabo Frio	31480	1055	13896,9	13896,9	465,7	465,7	226525	3,4
Casimiro de Abreu	8715	165	19724,3	19724,3	373,4	373,4	44184	1,9
Iguaba Grande	8405	163	29689,2	29689,2	575,8	575,8	28310	1,9
Rio das Ostras	30895	584	20504,5	20504,5	387,6	387,6	150674	1,9
São Pedro da Aldeia	10135	356	9700,8	9700,8	340,7	340,7	104476	3,5
Saquarema	12179	417	13658,2	13658,2	467,6	467,6	89170	3,4

Figura 37 Dados retirados do site Fonte: Secretaria Estaduais de Saúde do Brasil ,2020



## Capítulo 5

### 5. Conclusão do trabalho

Este trabalho demonstrou a aplicação eficaz do Orange Data Mining na análise e extração de informações valiosas a partir de conjuntos de dados relacionados ao COVID-19. Através do uso dessa poderosa ferramenta, foram exploradas diferentes técnicas de mineração de dados, como pré-processamento, seleção de características, modelagem e avaliação de desempenho, resultando em descobertas significativas.

Ao longo deste estudo, foi possível identificar padrões e tendências relevantes nos dados do COVID-19, permitindo a compreensão mais aprofundada da propagação da doença, fatores de risco, impacto das medidas de contenção e outras informações essenciais para a tomada de decisões e formulação de políticas de saúde pública.

O Orange Data Mining se mostrou uma ferramenta robusta, flexível e de fácil utilização, permitindo que usuários com diferentes níveis de habilidade em mineração de dados possam explorar e analisar os dados do COVID-19 de maneira eficiente. A sua interface intuitiva e recursos abrangentes facilitaram a manipulação dos dados, a criação de modelos e a visualização dos resultados, contribuindo para uma análise mais abrangente e compreensível.

Foi possível observar que políticas de saúde pública voltadas a tratamento precoce poderiam ter influenciado muito, mas positivamente em relação aos casos e sua prevenção, também ressaltou a importância de fontes de dados confiáveis e atualizadas, bem como a necessidade de um cuidadoso processo de pré-processamento para garantir a qualidade dos resultados obtidos.

Em suma, a utilização do Orange Data Mining neste estudo proporcionou uma abordagem eficiente para explorar conjuntos de dados do COVID-19, permitindo a descoberta de informações valiosas que podem auxiliar na compreensão e no combate a essa pandemia. A aplicação dessa ferramenta em outros contextos e conjuntos de dados pode ampliar ainda mais a sua utilidade e contribuir para avanços significativos na área da mineração de dados aplicada à saúde pública.

## REFERÊNCIAS BIBLIOGRÁFICAS

BRASIL. Agência Nacional de Vigilância Sanitária. Programa Nacional de Prevenção e Controle de Infecções Relacionadas à Assistência à Saúde (PNPCIRAS) 2013 – 2015. Brasília, 2013.

Disponível em:

<https://www.gov.br/anvisa/ptbr/centraisdeconteudo/publicacoes/servicosdesaude/publicacoes/pnpciras-2013-2015>. Acesso em: 07/05/2023

BRASIL. Agência Nacional de Vigilância Sanitária. Boletim Segurança do Paciente e Qualidade em Serviços de Saúde nº 14, 2016.

BRASIL. Agência Nacional de Vigilância Sanitária. Programa Nacional de Prevenção e Controle de Infecções Relacionadas à Assistência à Saúde (PNPCIRAS) 2016 – 2020. Brasília, 2016.

Disponível em:

<https://www.gov.br/anvisa/ptbr/centraisdeconteudo/publicacoes/servicosdesaude/publicacoes/publicacoes/pnpciras-2016-2020.pdf/view>. Acesso em: 07/05/2023

BRASIL. Agência Nacional de Vigilância Sanitária. Programa Nacional de Prevenção e Controle de Infecções Relacionadas à Assistência à Saúde (PNPCIRAS) 2021 a 2025

BRASIL. Agência Nacional de Vigilância Sanitária. Boletim Segurança do Paciente e Qualidade em Serviços de Saúde nº 22: Avaliação Nacional dos indicadores de IRAS e RM - 2019, 2020.

BRASIL. Agência Nacional de Vigilância Sanitária. Plano Nacional para a Prevenção e o Controle da Resistência Microbiana nos Serviços de Saúde, 2017. Disponível em:

<https://www.gov.br/anvisa/ptbr/centraisdeconteudo/publicacoes/servicosdesaude/publicacoes/publicacoes/planonacional-para-a-prevencao-e-o-controle-da-resistencia-microbiana-nos-servicos-desade.pdf/view>. Acesso em: 07/05/2023

BRASIL. Agência Nacional de Vigilância Sanitária. Diretriz Nacional para Elaboração de Programa de Gerenciamento do Uso de Antimicrobianos em Serviços de Saúde, 2017.

Disponível em:

<https://www.gov.br/anvisa/ptbr/centraisdeconteudo/publicacoes/servicosdesaude/publicacoes/publicacoes/diretriz-nacional-para-elaboracao-de-programa-de-gerenciamento-do-uso-deantimicrobianos-em-servicos-de-saude.pdf/view>. Acesso em: 07/05/2023

BRASIL. Agência Nacional de Vigilância Sanitária. Plano Integrado para a Gestão Sanitária da Segurança do Paciente em Serviços de Saúde 2021-2025, 2021. Disponível em:

<https://www.gov.br/anvisa/ptbr/centraisdeconteudo/publicacoes/servicosdesaude/publicacoes>. Acesso em: 07/05/2023

Centro de Informações Estratégicas e Resposta de Vigilância em Saúde (CIEVS-RJ). Dados organizados diariamente pelo CIEVS-RJ da Secretaria de Saúde do Estado do Rio de Janeiro, a partir do sistema esus-VE e SIVEP-Gripe, em articulação com as vigilâncias das secretarias municipais de saúde do estado. Atualizado em 16/05/2023 englobando todos os casos e óbitos confirmados até a semana 2023/19. Acesso em: 16/05/2023

ORANGE Data Mining. Disponível em: <https://orangedatamining.com>. Acesso em: 01/03/2023.

YOUTUBE. Data Mining: The Future of Work. Disponível em: [https://youtu.be/jJER5E6JA\\_Q](https://youtu.be/jJER5E6JA_Q). Acesso em: 10/04/2023.

Ministério da Saúde. Covid-19. Disponível em: <https://covid.saude.gov.br/>. Acesso em: 20/04/2023.

Secretaria de Estado de Saúde do Rio de Janeiro. Painel de Monitoramento da Covid-19. Disponível em: <https://painel.saude.rj.gov.br/monitoramento/covid19.html>. Acesso em: 21/04/2023.

SliderPlayer. Disponível em: <https://sliderplayer.com/>. Acesso em: 05/03/2023.

UNA-SUS. Disponível em: <https://www.unasus.gov.br/>. Acesso em: 05/03/2023.