

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
NÚCLEO DE EDUCAÇÃO À DISTÂNCIA
Pós-graduação Lato Sensu em Ciência de Dados e Big Data

Fábio Roberto Barros Padilha

**ANÁLISE DE DADOS NACIONAIS SOBRE A EVOLUÇÃO
DAS SÍNDROMES RESPIRATÓRIAS AGUDAS GRAVES
POR SARS-CoV-2, NO ANO DE 2020**

Belo Horizonte
2021

Fábio Roberto Barros Padilha

ANÁLISE DE DADOS NACIONAIS SOBRE A EVOLUÇÃO
DAS SÍNDROMES RESPIRATÓRIAS AGUDAS GRAVES
POR SARS-CoV-2, NO ANO DE 2020

Trabalho de Conclusão de Curso apresentado ao Curso de Especialização em Ciência de Dados e Big Data como requisito parcial à obtenção do título de especialista.

Belo Horizonte
2021

Sumário

1. Introdução	3
1.1 Contextualização	3
1.2 Objetivos	5
2. Coleta dos dados	6
2.1 Dados OpenDataSUS	6
2.2 Dados do Instituto Brasileiro de Geografia e Estatística	8
3. Processamento	9
3.1 Dados do Instituto Brasileiro de Geografia e Estatística	9
3.2 OpenDataSUS	10
3.3 Transformação	10
3.3.1 Notificações	11
3.3.2 Idade	12
3.3.3 Raça	13
3.3.4 Sexo	14
3.3.5 Comorbidades	15
3.3.6 Recursos do tratamento	16
3.3.7 Evolução	17
4. Análise Exploratória	19
4.1 Número absoluto de mortes de covid-19 no Brasil	19
4.2 Taxa de mortalidade da covid-19 no Brasil	21
4.3 Taxa de letalidade da covid-19 no Brasil	23
5. Inferência Estatística	26
5.1 Perfil do brasileiro com maior e menor risco de morte	26
6. Machine learning	29
6.1 Árvore de Decisão de Classificação	29
6.2 Árvore de Decisão e Seleção de Atributos	29
6.3 Árvore de Decisão de Classificação otimizado	30
6.4 Naive Bayse	31
7. Resultado	32
7.1 Dashboard-Covid19	32
8. Links	34
Referências Bibliográficas	35

Apêndice A - Taxa de Mortalidade Estados e Capitais	36
Anexo A - Ficha de Notificação	37
Anexo B - Dicionário de dados	39

1. Introdução

Nos últimos anos, o aumento de dados disponíveis digitalmente e o aparecimento de novas tecnologias têm favorecido a produção de conhecimento útil para sociedade.

De acordo com Gilson Medeiros de Oliveira Junior (2010):

O uso de técnicas de inteligência artificial, mais especificamente algoritmos de aprendizagem para classificação de padrões, tem crescido bastante nos últimos anos pelo fato de auxiliarem na resolução de problemas complexos do mundo real que podem envolver um elevado número de variáveis [...].
(JUNIOR, 2010, p.1)[4]

Esses algoritmos de aprendizagem estão relacionados ao aprendizado de máquina (machine learning), que é uma parte de Inteligência Artificial (IA) onde um algoritmo computacional cria, a partir dos dados disponíveis, modelos de aprendizado para a resolução de problemas (MEHTA, 2017).[5]

Dentre as técnicas de aprendizagem de máquina, pode-se citar o algoritmo de Naive Bayes, importante classificador probabilístico, que é, frequentemente, utilizado em processamento de diagnósticos médicos (cálculo de probabilidades em cima de diagnóstico de doenças), segundo Pedro César Tebaldi Gomes (2019) [2].

Outros importantes algoritmos de aprendizagem são as máquinas de vetores suporte - SVMs (Support Vector Machines), que “[...] têm a capacidade de resolver problemas de classificação e regressão, adquirindo com o aprendizado na etapa de treinamento a capacidade de generalização” (JUNIOR, 2010, p.5). [4]

Por fim, pode-se destacar a técnica de mineração de dados de árvore de decisão na aprendizagem de máquina. De acordo com Hosokawa (2011) [3], a árvore de decisão é um modelo preditivo, representado por um gráfico em forma de árvore, contendo as decisões a serem tomadas e seus possíveis riscos, custo e prejuízos.

Neste sentido, este trabalho utilizou as técnicas citadas acima, dentre outras, para análise de dados do Sistema Único de Saúde (SUS) relacionados às Síndromes Respiratórias Agudas Graves (SARS) relacionadas à covid-19, em 2020. De forma a identificar as regiões do Brasil mais impactadas pela covid-19, bem como os perfis com maior risco de morte, contribuindo com informações importantes para implementação de políticas de saúde pública, a exemplo da elaboração de um plano nacional de vacinação.

1.1 Contextualização

O ano de 2020 foi marcado por isolamento, medo e insegurança, causados pela pandemia do novo coronavírus SARS-CoV-2 e, com ela, pelo aumento considerável dos casos

de óbitos por covid-19.

Durante todo o curso da pandemia, um dos principais problemas que as autoridades e os profissionais de saúde enfrentaram foi a falta de informação: de recursos médicos, como respiradores, e de um plano adequado para sua distribuição.

Como o novo coronavírus é a principal causa para o aumento dos casos de mortes de Síndrome Respiratória Aguda Grave (SARS), identificar as regiões do Brasil mais afetadas, bem como ser capaz de prever os perfis com maior risco de morte, será de grande ajuda para as autoridades na elaboração de políticas públicas na área da saúde.

Dado o atual volume de informações relacionados ao SARS-CoV-2 (Covid-19) e às Síndromes Respiratórias Agudas Graves disponíveis, é possível se utilizar da tecnologia, como a inteligência artificial, para realizar previsões que auxiliem as autoridades da área de saúde e os médicos na elaboração de planos no combate ao novo coronavírus.

Sobre o uso das informações geradas com base nestas análises preditivas, Roberto Moreira Santos (2017) afirma que:

A aplicação desses resultados provenientes das análises é de suma importância para vários setores: como o setor da saúde, que consegue mapear os riscos de certas doenças que ocorrem em determinados grupos de pacientes, ou a resposta aos melhores medicamentos [...] (SANTOS, 2017, p.10)[7]

Neste contexto, “A analítica preditiva pode ser apoiada por técnicas da computação como aprendizagem de máquina (machine learning) e a mineração de dados para aumentar seu potencial de precisão” (WAYNE, 2007 apud SANTOS, 2017, p.11).

Dada a situação de calamidade pública provocada pelo novo Coronavírus SARS-CoV-2, que levou o Brasil a ultrapassar a marca de 180 mil mortes por covid-19, em 2020, e a chegada da vacina, buscou-se utilizar, neste trabalho, técnicas de mineração de dados e machine learning, com apoio do RStudio para o desenvolvimento de um projeto de ciência de dados com propósito de contribuir com a contenção da doença no Brasil. As bibliotecas utilizadas na elaboração do projeto estão descritas na Tabela 1 apresentada abaixo:

ggplot2	gridExtra	geobr	geobr	dplyr	latexpdf
stringr	openxlsx	rpart	e1071	randomForest	scales

Tabela 1: Pacotes R

Assim, espera-se que as previsões apresentadas com base no uso dessas tecnologias gerem conhecimentos úteis que auxiliem as autoridades públicas na implementação de políticas públicas na área da saúde e capacitem melhor o país na contenção da doença.

1.2 Objetivos

A pandemia do SARS-CoV-2 está impondo ao Brasil e ao mundo o maior desafio sanitário dos últimos 100 anos, tendo sido a Gripe Espanhola, ocorrida no ano de 1918, a última pandemia registrada de grande impacto mundial. Neste contexto, há uma atuação ativa por parte das autoridades de saúde na implementação de políticas públicas para conter o avanço da doença e o crescente número de óbitos.

Embora o Brasil, tenha um dos melhores sistemas de saúde da América Latina, a capacidade de atendimento em todo país é bastante desigual. Nas regiões mais pobres, onde a capacidade de atendimento de saúde é menor, têm-se observado a existência de situações críticas, como a falta de médicos, de recursos e de leitos hospitalares.

Neste contexto, este trabalho tem como objetivo geral entender, de forma regionalizada, a evolução dos casos graves da doença, a fim de contribuir com as autoridades da área de saúde na construção de políticas públicas voltadas para redução do número de óbitos e para contenção do novo coronavírus.

É importante destacar que, para este trabalho, são considerados casos graves de covid-19 aqueles existentes no arquivo CSV do OpenDataSUS.

Por fim, têm-se os seguintes objetivos específicos que irão ajudar na consecução do objetivo geral deste trabalho de conclusão de curso:

- identificar as regiões brasileiras com maior número de óbitos por covid-19;
- identificar a taxa de mortalidade por covid-19 dos estados e capitais;
- realizar teste de hipótese da taxa de letalidade do Brasil;
- apresentar a média móvel da taxa de letalidade de covid-19 do Brasil e regiões, ao longo do ano de 2020;
- encontrar os perfis dos pacientes com maiores e menores riscos de morte por covid-19;
- utilizar técnicas de machine learning para identificar o percentual de acerto do risco de morte por covid-19, a partir dos atributos selecionados na pesquisa.

2. Coleta dos dados

Nesse trabalho foram utilizada 2 (duas) fontes de dados, uma disponibilizada pela Secretaria de Vigilância em Saúde no portal OpenDataSUS e a outra disponibilizada no portal do Instituto Brasileiro de Geografia e Estatística (IBGE), cujo detalhamento será apresentado nas seções seguintes.

2.1 Dados OpenDataSUS

Dentre os dados utilizados neste trabalho, destacam-se os dados fornecidos pela Secretaria de Vigilância em Saúde, que está vinculada ao Ministério da Saúde, no portal OpenDataSUS. Sobre os dados coletados, o portal OpenDataSUS informa que:[6]

Esta página tem como finalidade disponibilizar o legado dos bancos de dados (BD) epidemiológicos de SRAG, da rede de vigilância da Influenza e outros vírus respiratórios, desde o início da sua implantação (2009) até os dias atuais (2021), com a incorporação da vigilância da COVID-19. Atualmente, o sistema oficial para o registro dos casos e óbitos de SRAG é o Sistema de Informação da Vigilância Epidemiológica da Gripe (SIVEP-Gripe).

O arquivo utilizado neste trabalho foi disponibilizado para [download em formato CSV](#). É importante observar que este arquivo é atualizado semanalmente e que, no momento do download, continha dados até 04/01/2021.

Como a Secretaria de Vigilância em Saúde atualiza os dados semanalmente, para garantir o acesso ao arquivo analisado neste trabalho, fez-se necessário manter uma cópia no [Google drive](#).

O arquivo contém 1.103.183 (um milhão cento e três mil e cento e oitenta e três) linhas e 154 colunas. Cada linha do arquivo corresponde ao registro de um paciente com a síndrome respiratória aguda grave. As 154 colunas são campos coletados a partir da [ficha de notificação do paciente](#), além de informações geradas pelo sistema. A ficha de notificação e o dicionário de dados estão disponíveis, respectivamente, nos anexos A e B. O trecho de código para importação do arquivo segue descrito abaixo:

```
arquivo = read.csv("dados/INFLUD-04-01-2021.csv", header = T, sep = ";",
na.strings = "", stringsAsFactors = T)
```

Após a importação dos dados, as colunas objeto da análise foram selecionadas e renomeadas, a fim de facilitar a identificação nas seções seguintes. Os trechos de código, para selecionar e renomear as colunas seguem descritos abaixo:

A) Selecionar:

```
pacientes = arquivo[ , c("DT_NOTIFIC", "SEM_NOT", "DT_SIN_PRI", "SEM_PRI", "SG_UF_NOT",
"ID_MUNICIP", "CO_MUN_NOT", "ID_UNIDADE", "CO_UNI_NOT",
"CS_SEXO", "NU_IDADE_N", "CS_GESTANT", "CS_RACA", "SURTO_SG", "NOSOCOMIAL", "FEBRE",
"TOSSE", "GARGANTA", "DISPNEIA", "DESC_RESP", "SATURACAO", "DIARREIA", "VOMITO",
"GORICA", "DOR_ABD", "FADIGA", "PERD_OLFT", "PERD_PALA", "FATOR_RISC", "PUERPERA",
"CARDIOPATI", "HEMATOLOGI", "SIND_DOWN", "HEPATICA", "ASMA", "DIABETES",
"NEUROLOGIC", "PNEUMOPATI", "IMUNODEPRE", "RENAL", "OBESIDADE", "OBES_IMC", "VACINA",
"ANTIVIRAL", "HOSPITAL", "DT_INTERNA", "UTI", "DT_ENTUTI", "DT_SAIDUTI",
"SUPPORT_VEN", "RAIOX_RES", "TOMO_RES", "AMOSTRA", "TP_TES_AN", "RES_AN",
"POS_AN_FLU", "TP_FLU_AN", "POS_AN_OUT", "AN_SARS2", "AN_VSR",
"AN_PARA1", "AN_PARA2", "AN_PARA3", "AN_ADENO", "PCR_RESUL", "POS_PCRFLU",
"TP_FLU_PCR", "PCR_FLUASU", "PCR_FLUBLI", "POS_PCRROUT", "PCR_SARS2", "PCR_VSR",
"PCR_PARA1", "PCR_PARA2", "PCR_PARA3", "PCR_PARA4", "PCR_ADENO", "PCR_METAP",
"PCR_BOCA", "PCR_RINO", "TP_AM_SOR", "TP_SOR", "RES_IGG", "RES_IGM", "RES_IGA",
"CLASSI_FIN", "CRITERIO", "EVOLUCAO", "DT_EVOLUCAO")]
```

B) Renomear:

```
colnames(pacientes) = c("DT_NOTIFICACAO", "SEMANA_NOTIFICACAO", "DT_PRIMEIROS_SINTOMAS",
"SEMANA_PRIMEIROS_SINTOMAS", "SG_UF", "MUNICIPIO",
"COD_MUNICIPIO", "HOSPITAL", "COD_HOSPITAL", "CS_SEXO", "NU_IDADE", "CS_GESTANTE",
"CS_RACA", "IN_SURTO_SINDROME_GRIPAL", "IN_NOSOCOMIAL", "IN_FEBRE", "IN_TOSSE",
"IN_GARGANTA", "IN_DISPNEIA", "IN_DESCONFORTO_RESP", "IN_SATURACAO", "IN_DIARREIA",
"IN_VOMITO", "IN_DOR_ABDOMINAL", "IN_FADIGA", "IN_PERDA_OLFTATO", "IN_PERDA_PALADAR",
"IN_FATOR_RISC", "IN_PUERPERA", "IN_CARDIOPATI", "IN_HEMATOLOGI", "IN_SINDROME_DOWN",
"IN_DOENCA_HEPATICA", "IN_ASMA", "IN_DIABETES", "IN_DOENCA_NEUROLOGICA",
"IN_PNEUMOPATIA_CRONICA", "IN_IMUNODEPRESSAO", "IN_DOENCA_RENAL", "IN_OBESIDADE",
"NU_OBESIDADE_IMC", "IN RECEBEU_VACINA_GRIP", "IN_USOU_ANTIVIRAL", "IN_HOSPITALIZADO",
"DT_INTERNACAO", "IN_UTI", "DT_ENTRADA_UTI", "DT_SAIDA_UTI", "CS_SUPORTE_VENTILATORIO",
"CS_RAIOX_RESULTADO", "CS_TOMOGRAFIA_RESULTADO", "IN_COLETOU_AMOSTRA",
"CS_TESTE_ANTIGENO", "CS_RESULTADO_ANTIGENO", "IN_POSITIVO_INFLUENZA",
"CS_TIPO_INFLUENZA", "IN_POSITIVO_OUT_VIRUS", "IN_POSITIVO_SARS2",
"IN_POSITIVO_VSR", "IN_POSITIVO_PARAINFLUENZA1", "IN_POSITIVO_PARAINFLUENZA2",
"IN_POSITIVO_PARAINFLUENZA3", "IN_POSITIVO_ADENOVIRUS", "CS_PCR_RESULTADO",
"IN_POSITIVO_PCR_INFLUENZA", "CS_TIPO_INFLUENZA_PCR", "CS_PCR_INFLUENZA_A_SUBTIPO",
"CS_PCR_FLU_B_LINHAGEM", "IN_PCR_POSITIVO_OUT_VIRUS", "IN_PCR_POSITIVO_SARS2",
"IN_PCR_POSITIVO_VSR", "IN_PCR_POSITIVO_PARAINFLUENZA1",
"IN_PCR_POSITIVO_PARAINFLUENZA2", "IN_PCR_POSITIVO_PARAINFLUENZA3",
"IN_PCR_POSITIVO_PARAINFLUENZA4", "IN_PCR_POSITIVO_ADENOVIRUS",
"IN_PCR_POSITIVO_METAPNEUMOVIRUS", "IN_PCR_POSITIVO_BOCAVIRUS",
"IN_PCR_POSITIVO_RINOVIRUS", "CS_TIPO_AMOSTRA_SOROLOGICA_SARS_COV2",
"CS_TESTE_SOROLOGICO", "IN_SOROLOGICO_POSITIVO_IGG", "IN_SOROLOGICO_POSITIVO_IGM",
"IN_RESULTADO_POSITIVO_IGA", "CS_CLASSICACAO_FINAL", "CS_CRITERIO_CONFIRMACAO",
"CS_EVOLUCAO", "DT_EVOLUCAO")
```

2.2 Dados do Instituto Brasileiro de Geografia e Estatística

Fez-se necessário enriquecer este trabalho com dados do Instituto Brasileiro de Geografia e Estatística - IBGE, tanto para o cálculo das taxas de mortalidades, bem como para elaboração dos mapas.

O IBGE disponibiliza o arquivo neste [link](#), no diretório: \textbf{Estimativa_de_Populacao/Estimativas_2020}, em formato xls, conforme Figura 1 abaixo:

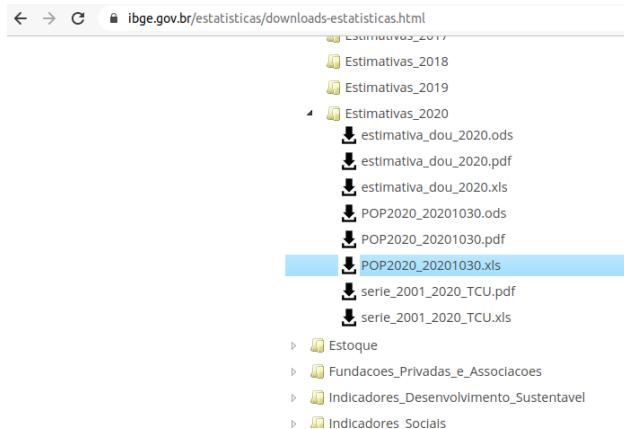


Figura 1: Arquivo IBGE

Após o download, o arquivo foi convertido para o formato xlsx, no LibreOffice, para facilitar a importação para RStudio. O arquivo contém duas abas, uma com a estimativa populacional por unidade federativa e outra aba com a estimativa populacional por municípios.

Para este trabalho, importamos apenas os dados da aba municípios, uma vez que a população por estado ou por região pode ser facilmente obtida a partir do somatório populacional dos municípios.

A importação dos dados foi iniciada a partir da terceira linha do arquivo, a fim de desconsiderar as linhas referentes ao cabeçalho, conforme trecho de código a seguir:

```
arquivoIBGE=read.xlsx("dados/POP2020_20201030.xlsx", sheet = "Municípios",
startRow = 3, colNames =FALSE, rowNames = FALSE)
```

3. Processamento

Este capítulo contempla as transformações realizadas nos data frames resultantes das importações dos arquivos CSV (disponibilizado no portal OpenDataSUS) e XLS (disponibilizado no portal do IBGE). Estes campos tratados são os principais atributos utilizados nos algoritmos e gráficos constantes nos capítulos seguintes.

3.1 Dados do Instituto Brasileiro de Geografia e Estatística

Os dados populacionais do IBGE, necessários para realização deste trabalho, estão presentes da linha 3 até a linha 5572, de acordo com as Figuras 2 e 3:

The screenshot shows the LibreOffice Calc interface with the title bar "POP2020_20201030.xlsx - LibreOffice Calc". The main area displays a table titled "ESTIMATIVAS DA POPULAÇÃO RESIDENTE NOS MUNICÍPIOS BRASILEIROS COM DATA DE REFERÊNCIA EM 1º DE JULHO DE 2020". The table has columns: UF, COD. UF, COD. MUNIC, NOME DO MUNICÍPIO, and POPULAÇÃO ESTIMADA. The data starts from row 3 and ends at row 351. The first few rows show data for RO (Rondonia), such as Alta Floresta D'Oeste (22.728), Ariquemes (109.523), and Cacoal (5.188).

ESTIMATIVAS DA POPULAÇÃO RESIDENTE NOS MUNICÍPIOS BRASILEIROS COM DATA DE REFERÊNCIA EM 1º DE JULHO DE 2020					
2	UF	COD. UF	COD. MUNIC	NOME DO MUNICÍPIO	POPULAÇÃO ESTIMADA
3	RO	11	00015	Alta Floresta D'Oeste	22.728
4	RO	11	00023	Ariquemes	109.523
5	RO	11	00031	Cacoal	5.188
6	RO	11	00049	Cahyba	85.938
7	RO	11	00056	Cerejeiras	16.204
8	RO	11	00064	Colorado do Oeste	15.544
9	RO	11	00072	Colniza	7.220
10	RO	11	00080	Crotas Macapéus	53.737
11	RO	11	00089	Espirito D'este	32.695
12	RO	11	00106	Guajará-Mirim	46.556
13	RO	11	00114	Itapuã	18.146
14	RO	11	00122	Jiparana	130.099
15	RO	11	00130	Machadinho D'este	40.867
16	RO	11	00148	Nova Bréscia D'este	20.489
17	RO	11	00156	Novo Progresso D'este	23.771
18	RO	11	00189	Pimenta Bueno	36.881
19	RO	11	00205	Porto Velho	539354 ⁽¹⁾
20	RO	11	00254	Presidente Médici	18.571
21	RO	11	00263	Rondônia	3.800
22	RO	11	00288	Roxol de Moura	55.407
23	RO	11	00296	Santa Luzia D'este	6.216
24	RO	11	00314	Santarem	162.222
25	RO	11	00329	São Miguel do Guaporé	23.077
26	RO	11	00338	Nova Mamoré	31.392
27	RO	11	00346	Alvorada D'este	14.106
28	RO	11	00371	Manoel Urbano dos Paredes	23.226
29	RO	11	00403	Alto Pará	21.847
30	RO	11	00452	Buritis	40.356
31	RO	11	00505	Porto Horizonte do Oeste	8.829
32	RO	11	00515	Cacaulândia	5.936
33	RO	11	00700	Campo Novo de Rondônia	14.266
34	RO	11	00809	Candeias do Jamari	27.388
35	RO	11	00908	Castanheira	2.987

Figura 2: Cabeçalho arquivo IBGE

The screenshot shows the LibreOffice Calc interface with the title bar "POP2020_20201030.xlsx - LibreOffice Calc". The main area displays a table with a single row of data. The table has columns A, B, C, D, E, and F. The data consists of six rows of information, starting from row 5569 and ending at row 5573. The data includes codes for GO (Goiás), DF (Distrito Federal), and other municipalities like Vila Boa, Vila Propício, and Brasília, along with their respective populations.

A	B	C	D	E	F
5569	GO	52	22054	Vicentípolis	8.873
5570	GO	52	22203	Vila Boa	6.312
5571	GO	52	22302	Vila Propício	5.882
5572	DF	53	00108	Brasília	3.055.149
5573					

Fonte: IBGE. Diretoria de Pesquisas - DPE - Coordenação de População e Indicadores Sociais - COPIS.

Notas:

(1) População judicial do município de Porto Velho-RO: 494.013 habitantes. Processo Judicial nº 12318-40.2018.4.01.4100 - Seção Judiciária de Rondônia.

(2) População judicial do município de Manaus-AM: entre 30.505 e 37.356 habitantes. Parecer de Força Executória nº 00010/2017/NUCOB-GE/ACIP/PA/MPF/AGU, em trâmite na 3ª VEM/AM.

(3) População judicial do município de Santa Isabel do Rio Negro-AM: entre 23.773 e 30.564 habitantes. Parecer de Força Executória nº 00007/2017/NUCOB-GE/ACIP/PA/MPF/AGU, em trâmite na 3ª VEM/AM.

Figura 3: Rodapé arquivo IBGE

A importação dos dados foi iniciada a partir da linha três do arquivo, descartando o cabeçalho. Para descartar o rodapé do data frame resultante da importação, foi selecionado somente até a linha 5.570 (cinco mil quinhentos e setenta). Por fim, foi criado um campo extra no data frame para armazenar o código do município sem o último dígito.

A necessidade de criar uma coluna extra para o município sem o dígito verificador ocorreu devido à ausência do dígito verificar no arquivo CSV disponibilizado no portal OpenDataSUS. O campo criado com o código do município sem o dígito verificador foi utilizado em JOINs nas seções seguintes, onde serão criados gráficos pelas diversas localidades do país.

É importante destacar que a informação das regiões do país (norte, nordeste, centro-oeste, sudeste e sul) não estavam presentes nos arquivos apresentados. Tornando-se necessária a criação de uma matriz com os estados e suas respectivas regiões. Essa matriz permitiu o enriquecimento dos dados e a produção dos gráficos summarizados, conforme trecho de código abaixo:

```
regiao = matrix(data = NA, nrow = 27, ncol = 2, byrow = TRUE)
regiao[1,] = c("AC", "Norte")
regiao[2,] = c("AP", "Norte")
regiao[3,] = c("AM", "Norte")
regiao[4,] = c("PA", "Norte")
regiao[5,] = c("RO", "Norte")
regiao[6,] = c("RR", "Norte")
regiao[7,] = c("TO", "Norte")
regiao[8,] = c("AL", "Nordeste")
regiao[9,] = c("BA", "Nordeste")
#Continua para todos os 27 estados...
```

```
pacientes = pacientes %>% left_join(regiao, c("SG_UF" = "SG_UF"))
```

3.2 OpenDataSUS

A transformação dos dados do SUS, disponibilizados no portal OpenDataSUS, foi realizada em duas etapas. Na primeira etapa, o processamento teve como finalidade favorecer a elaboração de gráficos no ggplot2 a fim de apresentar a distribuição dos dados. Já na segunda etapa, outras transformações aconteceram com o propósito de ajustar os dados para as execuções dos algoritmos de machine learning.

3.3 Transformação

A partir dessa seção será feita a limpeza e transformação dos dados, de maneira mais significativa. Nas seções anteriores, o processamento dos dados se restringiu a importar os arquivos objeto da análise e enriquecer um dos data frame resultante da importação com as regiões do país. Para melhor visualizar o que já foi realizado até o momento, segue Figura 4 com fluxo do pré-processamento dos dados:

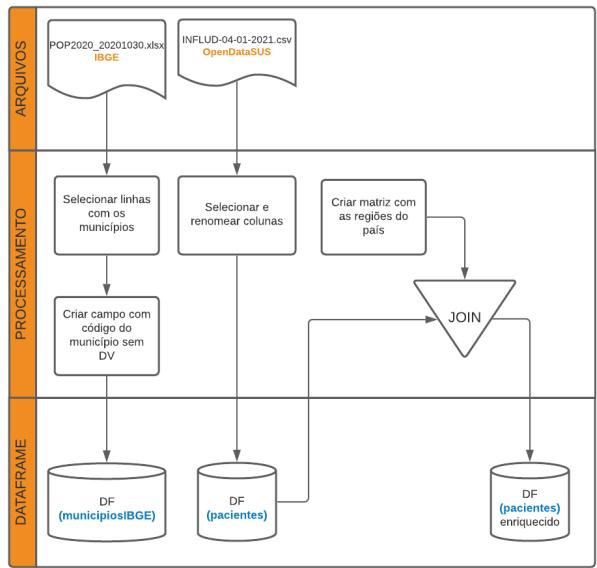


Figura 4: Pré-processamento dos dados

Nas subseções abaixo, será realizado o detalhamento das principais colunas do arquivo CSV disponibilizado no portal OpenDataSUS. Ressalta-se que não foi necessário detalhar a transformação do arquivo disponibilizado pelo IBGE, uma vez que o mesmo não precisou de transformações significativas, além das já mencionadas no item 3.1.

É importante enfatizar, ainda, que foi necessário elaborar gráficos preliminares nessas subseções, a fim de validar o processamento. No entanto, os gráficos mais elaborados e que trazem informações mais importantes para este trabalho estão dispostos no capítulo 4.

3.3.1 Notificações

O campo DT_NOTIFICACAO, que contém a data do preenchimento da ficha de notificação, foi ajustado para o formato dd/mm/aaaa. Após a conversão do campo, foi necessária a realização de algumas limpezas nos dados, como a remoção dos registros com data de notificação superior a 31/12/2020 do data frame, visto que o escopo desse trabalho é o ano de 2020, bem como a retirada das notificações que não foram classificadas como casos de covid-19. Conforme trecho de código abaixo:

```

#Redefinir tipo de dados da coluna
pacientes$DT_NOTIFICACAO = as.Date(pacientes$DT_NOTIFICACAO, format="%d/%m/%Y")
pacientes["MES_NOTIFICACAO"] = format(pacientes$DT_NOTIFICACAO, "%Y/%m")
#Removendo ocorrencia com datas de notificação superior a 31/12/2020
pacientes = pacientes[pacientes$DT_NOTIFICACAO < as.Date("01/01/2021", format="%d/%m/%Y"), ]

```

É importante destacar que não existem valores omissos na coluna DT_NOTIFICACAO. A partir deste processamento foi possível a criação do Gráfico 1 (Notificações de

Covid-19 por Mês) e verificar que no Brasil, em 2020, os primeiros casos de covid-19 foram registrados em fevereiro e que, no mês de julho, ocorreu o maior número de casos.

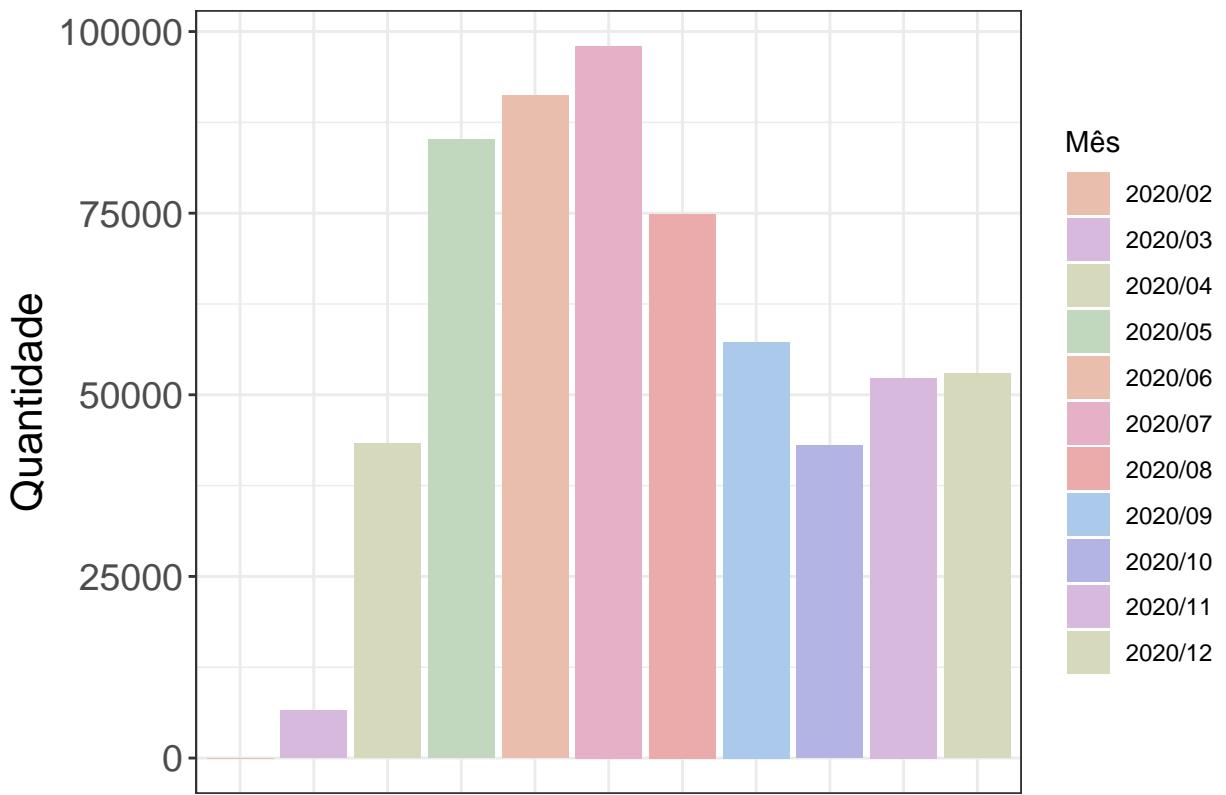


Gráfico 1: Notificações de Covid-19 por Mês

3.3.2 Idade

No processamento da coluna NU_IDADE, atribuiu-se a média da idade nacional dos casos de covid-19, que é de 59 (cinquenta e nove) anos, nas ocorrências com valores superiores a 110 (cento e dez) anos. Essa transformação teve como objetivo minimizar os valores outliers e os prováveis erros de digitação. Conforme trecho de código abaixo:

```
#Idades acima de 110 anos atribui a idade média.
pacientes$NU_IDADE=as(pacientes$NU_IDADE, "integer");
pacientes[pacientes$NU_IDADE>110,]$NU_IDADE=as.integer(mean(pacientes$NU_IDADE, na.rm=T))
```

Igualmente ao ocorrido no item anterior indentificou-se que não há valores omissos na coluna NU_IDADE.

Com base nesse processamento foi elaborado o Gráfico 2 (Boxplot de Idades dos casos graves de covid-19 por região), no qual fica evidenciado que:

- a região centro-oeste e a região nordeste apresentam respectivamente a menor e a maior média nacional de idade dos casos graves de covid-19;
- existem outlier abaixo de 15 anos nas regiões nordeste, suldeste e sul;

- aproximadamente 75% dos casos graves de covid-19 possuem idade acima dos 40 anos.

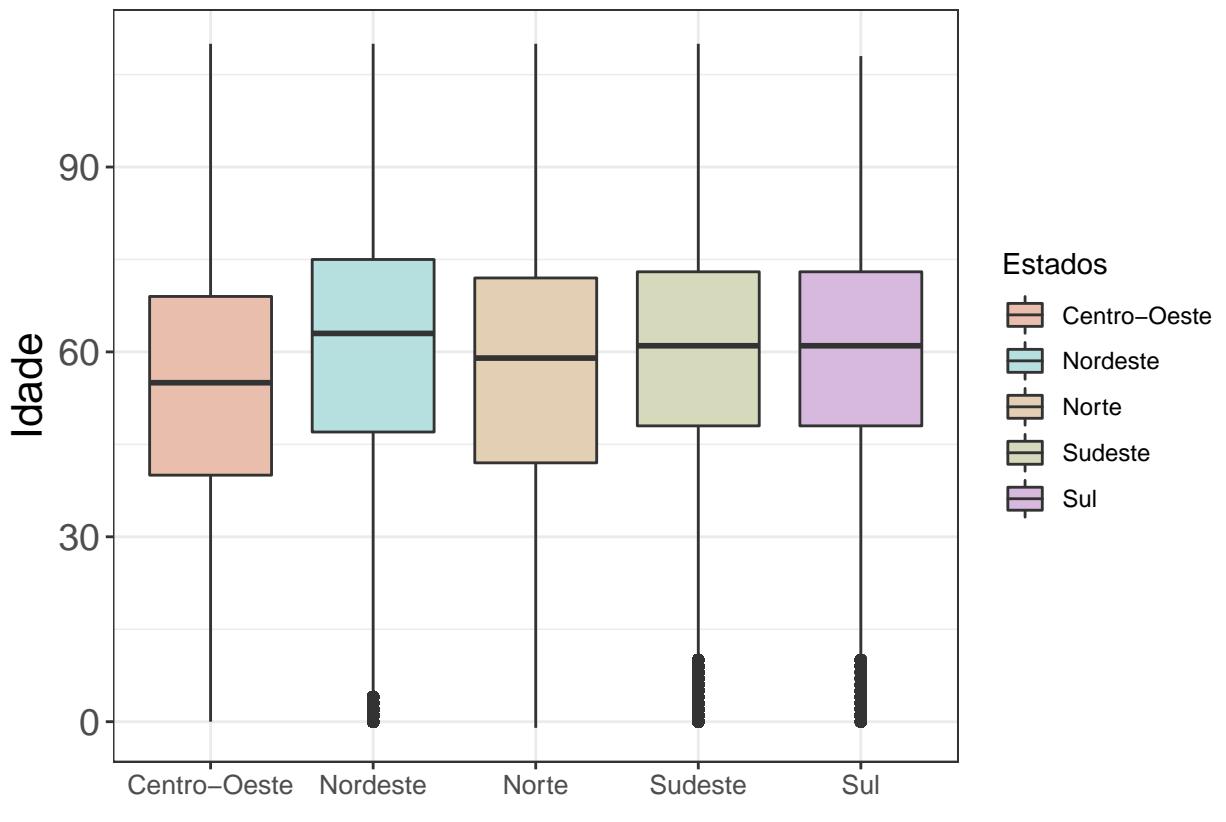


Gráfico 2: Boxplot de Idades dos casos graves de covid-19 por região

3.3.3 Raça

Na coluna CS_RACA, os valores numéricos foram substituídos pelas strings correspondentes, conforme descrito no dicionário de dados disponível na seção anexo. Essas alterações são verificáveis no trecho de código a seguir:

```
#Transformando domínios da coluna raça em valores mais significativos para o gráfico
pacientes[ complete.cases(pacientes$CS_RACA) & pacientes$CS_RACA==9,]$CS_RACA = NA
pacientes[ complete.cases(pacientes$CS_RACA) & pacientes$CS_RACA==1,]$CS_RACA = "Branca"
pacientes[ complete.cases(pacientes$CS_RACA) & pacientes$CS_RACA==2,]$CS_RACA = "Preta"
pacientes[ complete.cases(pacientes$CS_RACA) & pacientes$CS_RACA==3,]$CS_RACA = "Amarela"
pacientes[ complete.cases(pacientes$CS_RACA) & pacientes$CS_RACA==4,]$CS_RACA = "Parda"
pacientes[ complete.cases(pacientes$CS_RACA) & pacientes$CS_RACA==5,]$CS_RACA = "Indígena"
pacientes$CS_RACA = factor(pacientes$CS_RACA)
```

É importante destacar que, na coluna CS_RACA, o percentual de valores é de 22% (vinte e dois por cento). Esses valores foram removidos nas elaborações de gráficos e modelos de machine learning.

Este processamento possibilitou a construção do Gráfico 3 (Quantitativo dos casos graves de covid-19 por raça), revelando que:

- a raça com maior representatividade de casos graves de covid-19 é a branca seguida

- pela parda;
- a raça indígena foi a que registrou menos casos.

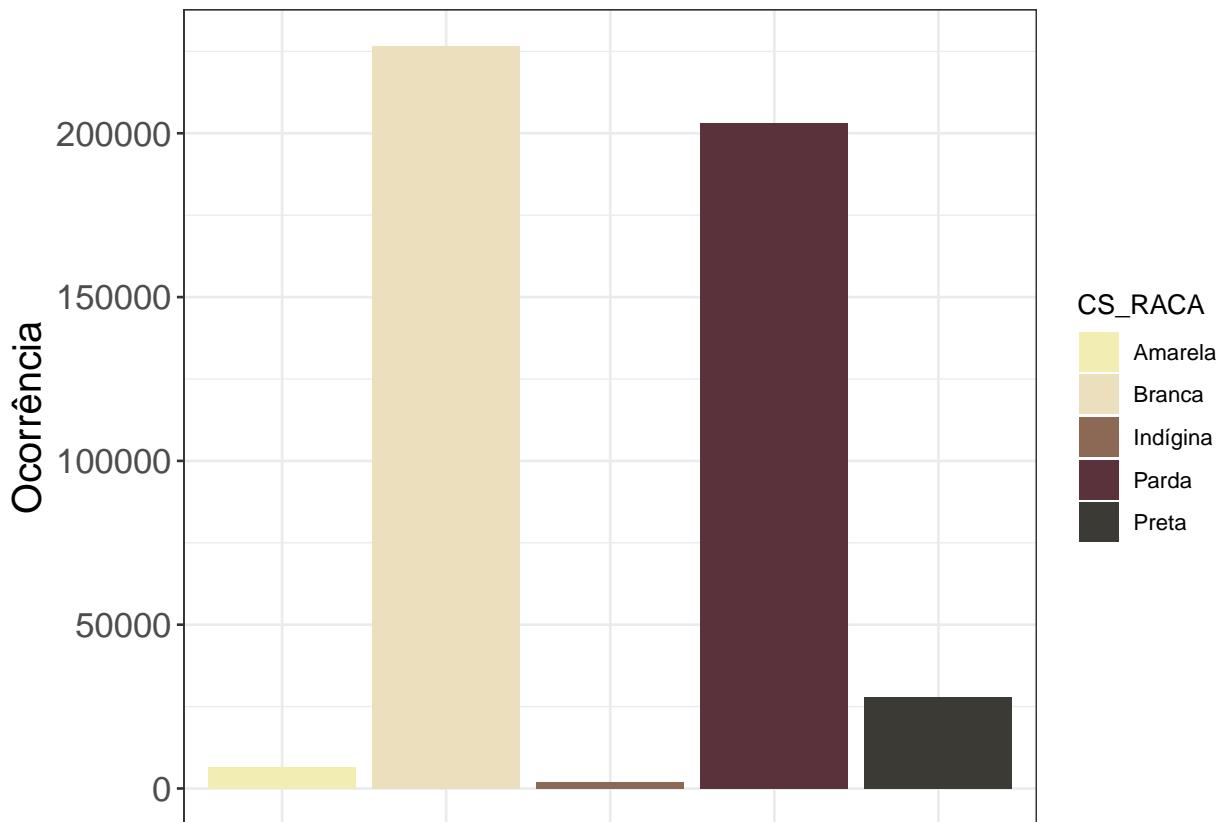


Gráfico 3: Quantitativo dos casos graves de covid-19 por raça

3.3.4 Sexo

O percentual de valores omissões na coluna CS_SEXO é de 0.01% (zero vírgula zero um por cento). Esses valores foram tratados a partir da coluna CS_GESTANTE, atribuiu-se sexo feminino, quando esta coluna informa existência de gestação. Em contrapartida, pra os valores omissões de sexo que não apresetam indicação de gestação, atribuiu-se sexo masculino, dado que a presença do sexo masculino é maior (vide Gráfico 4 abaixo) e a incidência de valores omissos é muito pequena.

Abaixo segue o trecho de código responsável por este processamento:

```
pacientes[is.na(pacientes$CS_SEXO) & pacientes$CS_GESTANTE>4, ]$CS_SEXO="M"
pacientes[is.na(pacientes$CS_SEXO) & pacientes$CS_GESTANTE<=4, ]$CS_SEXO="F"
pacientes$CS_SEXO = factor(pacientes$CS_SEXO)
```

Com base neste processamento foi gerado o Gráfico 4 (Percentual dos casos graves de covid-19 por sexo), que destaca a maior incidência de casos graves de covid-19 no sexo masculino.

Sexo

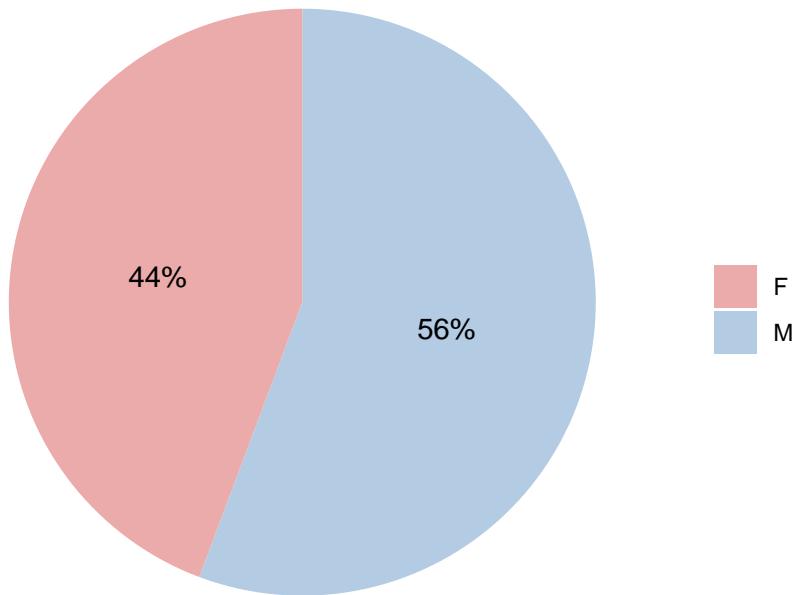


Gráfico 4: Percentual dos casos graves de covid-19 por sexo

3.3.5 Comorbidades

Nas colunas referente às comorbidades do paciente, foram convertidos os valores numéricos para as strings correspondentes, conforme descrito no dicionário de dados, disponível no Anexo B, deste trabalho.

A coluna IN_FATOR_RISC, informa a existência de comorbidade no paciente. Nela não foram identificados valores omissos.

No trecho de código abaixo verifica-se parte do processamento realizado nas colunas referente a comorbidades.

```
#Puerpera?  
pacientes[complete.cases(pacientes$IN_PUERPERA) &  
         pacientes$IN_PUERPERA==9,]$IN_PUERPERA=NA  
pacientes[complete.cases(pacientes$IN_PUERPERA) &  
         pacientes$IN_PUERPERA==1,]$IN_PUERPERA="Sim"  
pacientes[complete.cases(pacientes$IN_PUERPERA) &  
         pacientes$IN_PUERPERA==2,]$IN_PUERPERA="Não"  
pacientes$IN_PUERPERA=factor(pacientes$IN_PUERPERA)  
#Cardiopatia?  
pacientes[complete.cases(pacientes$IN_CARDIOPATI) &  
         pacientes$IN_CARDIOPATI==9, ]$IN_CARDIOPATI=NA
```

```

pacientes[complete.cases(pacientes$IN_CARDIOPATI) &
           pacientes$IN_CARDIOPATI==1, ]$IN_CARDIOPATI="Sim"
pacientes[complete.cases(pacientes$IN_CARDIOPATI) &
           pacientes$IN_CARDIOPATI==2, ]$IN_CARDIOPATI="Não"
pacientes$IN_CARDIOPATI=factor(pacientes$IN_CARDIOPATI)
#Continua para todos os campos do grupo comorbidades...

```

Com base no processamento acima foi possível gerar o Gráfico 5 (Comorbidades preeexistentes nos casos graves de covid-19) no qual verifica-se a maior incidência de comorbidades referente a cardiopatias e diabetes.

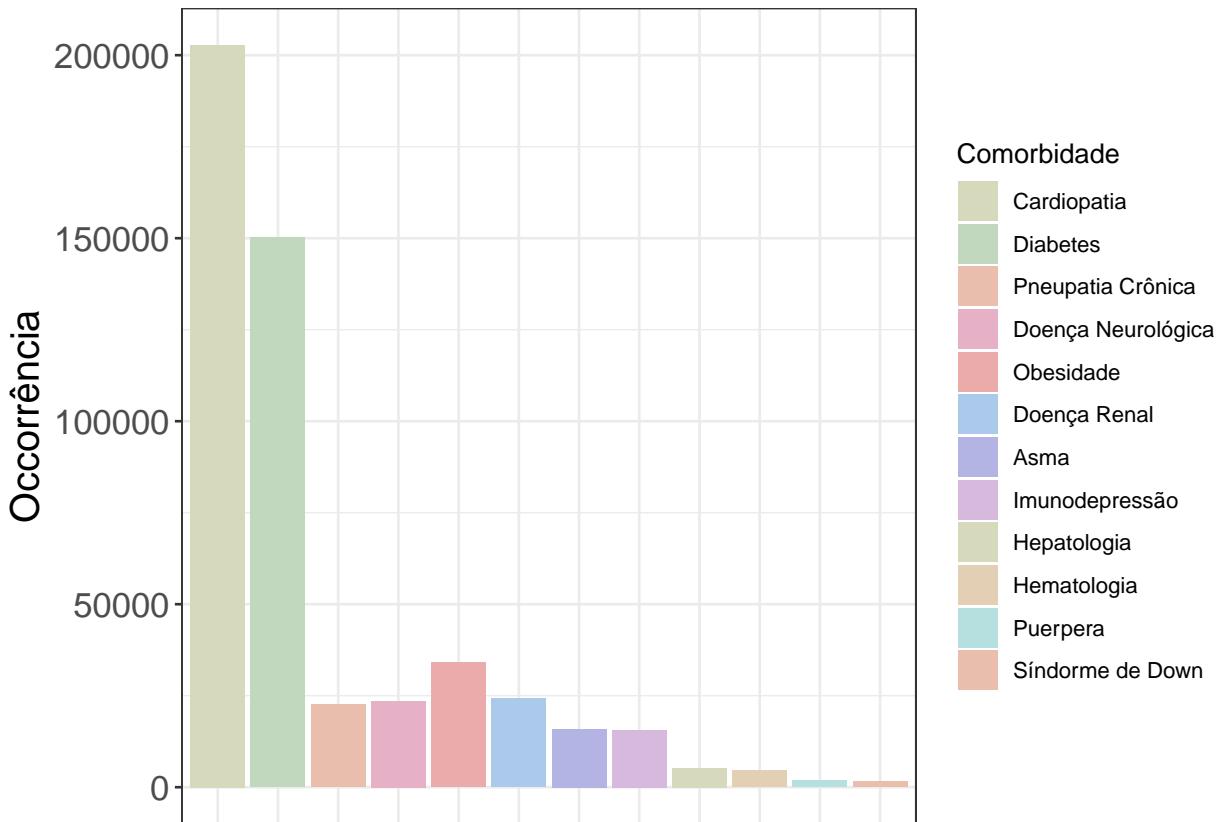


Gráfico 5: Comorbidades preeistentes nos casos graves de covid-19

3.3.6 Recursos do tratamento

Nesta seção constam os recursos que foram utilizados pelo paciente durante o tratamento. Os recursos são: se o paciente precisou de oxigênio, se precisou de suporte de ventilação pulmonar, se foi pra UTI, se tomou algum antiviral ou se foi hospitalizado.

Ressalta-se que a coluna mais utilizada deste grupo foi a IN_HOSPITALIZADO, cujo percentual de valores omitidos é de 2.61% (dois vírgula sessenta e um por cento), estes valores foram removidos na geração dos gráficos e modelos de machine learning.

Nas colunas referentes a recursos do tratamento do paciente, foram convertidos os valores numéricos para strings correspondentes, conforme descrito no dicionário de dados,

disponível no Anexo B. O trecho de código abaixo reflete este processamento:

```
#Foi Hospitalizado?  
pacientes[complete.cases(pacientes$IN_HOSPITALIZADO) &  
  pacientes$IN_HOSPITALIZADO==1, ]$IN_HOSPITALIZADO="Sim"  
pacientes[complete.cases(pacientes$IN_HOSPITALIZADO) &  
  pacientes$IN_HOSPITALIZADO==2, ]$IN_HOSPITALIZADO="Não"  
pacientes$IN_HOSPITALIZADO=factor(pacientes$IN_HOSPITALIZADO)  
#Continua para os demais campos...
```

Esse processamento permitiu a elaboração do Gráfico 6 (Recursos utilizados nos casos graves de covid-19), do qual conclui-se que aproximadamente metade dos pacientes hospitalizados precisaram de oxigênio não invasivo.

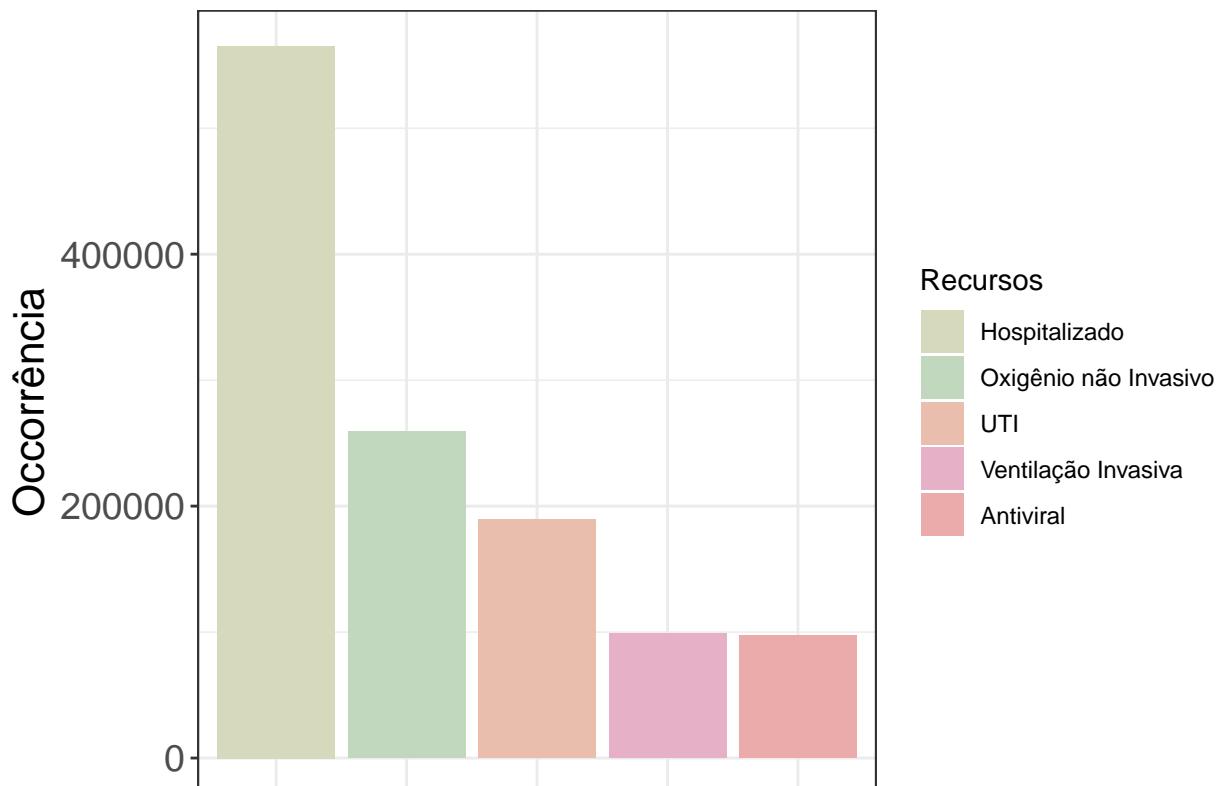


Gráfico 6: Recursos utilizados nos casos graves de covid-19

3.3.7 Evolução

Na coluna CS_EVOLUCAO, os valores numéricos foram convertidos para as strings correspondentes, conforme descrito no dicionário de dados, disponível no Anexo B, deste trabalho.

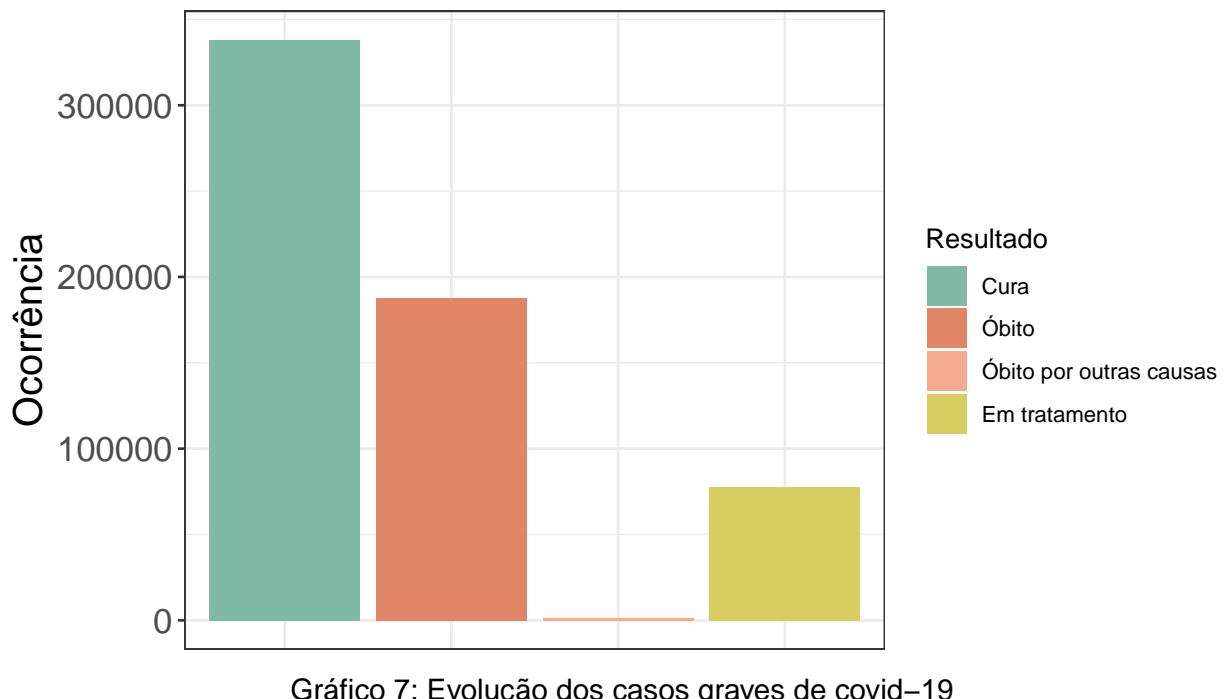
O percentual de valores omissos nesta coluna é de 12,31% (doze vírgula trinta e um por cento), para esses valores foram atribuído a valor “Em tratamento”.

O trecho de código abaixo reflete o processamento desta coluna:

```
#Data Evolução / Desfecho
pacientes$DT_EVOLUCAO = as.Date(pacientes$DT_EVOLUCAO, format="%d/%m/%Y")
pacientes["SEMANA_EVOLUCAO"] = format(pacientes$DT_EVOLUCAO, "%U");
pacientes$SEMANA_EVOLUCAO=as.integer(pacientes$SEMANA_EVOLUCAO)

#Resultado Evolução / Desfecho
pacientes[complete.cases(pacientes$CS_EVOLUCAO) &
  pacientes$CS_EVOLUCAO==1, ]$CS_EVOLUCAO="Cura"
pacientes[complete.cases(pacientes$CS_EVOLUCAO) &
  pacientes$CS_EVOLUCAO==2, ]$CS_EVOLUCAO="Óbito"
pacientes[complete.cases(pacientes$CS_EVOLUCAO) &
  pacientes$CS_EVOLUCAO==3, ]$CS_EVOLUCAO="Óbito por outras causas"
pacientes[is.na(pacientes$CS_EVOLUCAO), ]$CS_EVOLUCAO="Em tratamento"
pacientes[complete.cases(pacientes$CS_EVOLUCAO) & pacientes$CS_EVOLUCAO==9,]$CS_EVOLUCAO="Em tratamento"
pacientes$CS_EVOLUCAO=factor(pacientes$CS_EVOLUCAO)
```

A partir desse processamento foi elaborado o Gráfico 7 (Evolução dos casos graves de covid-19), do qual conclui-se que, aproximadamente, 1/3 (um terço) dos pacientes graves de covid-19 evoluem para óbito.



4. Análise Exploratória

Neste capítulo, serão detalhadas as taxas de mortalidade e letalidade, bem como o número de mortes absolutos decorrentes de covid-19 no Brasil, em 2020. É importante observar que essa análise utilizou a técnica de drill down, detalhando as informações até o nível de capitais.

Com intuito de facilitar as análises presentes neste capítulo, torna-se necessário nivelar alguns conceitos, tais como: taxas de mortalidade e letalidade.

A taxa de mortalidade é a razão entre o número total de óbitos e o total da população vezes uma constante utilizada como unidade de medida, neste trabalho será adotado 1.000.000 (um milhão) de habitantes.

Já a taxa de letalidade é a razão entre o número total de óbitos e o número total de casos graves.

Essas taxas foram calculadas exclusivamente a partir dos casos graves de covid-19. Deve-se relembrar que, para este trabalho, considera-se casos graves de covid-19 aqueles presentes no arquivo CSV do OpenDataSUS.

4.1 Número absoluto de mortes de covid-19 no Brasil

No ano de 2020, o Brasil teve 187.727 (cento e oitenta e sete mil e setecentos e vinte e sete) mortes por covid-19. Neste sentido, o Sudeste apresentou quase a metade dos casos de mortes por covid-19 do país, com 88.362 (oitenta e oito mil e trezentos e sessenta e duas) mortes. As demais regiões registraram juntas um número de mortes um pouco superior, totalizando 99.365 (noventa e nove mil e trezentos e sessenta e cinco) mortes, o que representa 52,93% (cinquenta e dois vírgula noventa e três por cento) do total de óbitos registrados no país por covid-19.

A este respeito, o Gráfico 8 (Mortes de covid-19 por região) apresenta a distribuição das mortes decorrente de covid-19, de forma regionalizada, conforme pode ser visto abaixo:

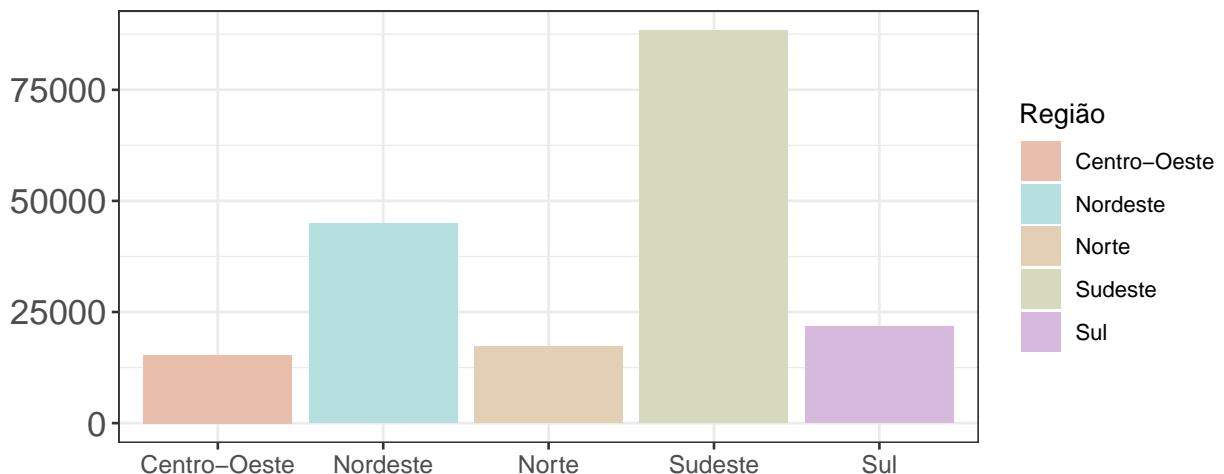


Gráfico 8: Mortes de covid-19 por região

Após a região sudeste, o Nordeste aparece como a segunda região com maior número de mortes do Brasil, totalizando 44.965 (quarenta e quatro mil e novecentos e sessenta e cinco) mortes, ou seja, 23,95% (vinte e três vírgula noventa e cinco por cento) do total de mortes por covid-19 do país.

No Sudeste, o Estado de São Paulo foi responsável por 53,45% (cinquenta e três vírgula quarenta e cinco por cento) das mortes, registrando 47.226 (quarenta e sete mil duzentos e vinte e seis) óbitos, ou seja, mais da metade das mortes na região, como pode ser visto no gráfico 9 (Mortes por covid-19 na região Sudeste) apresentado abaixo:

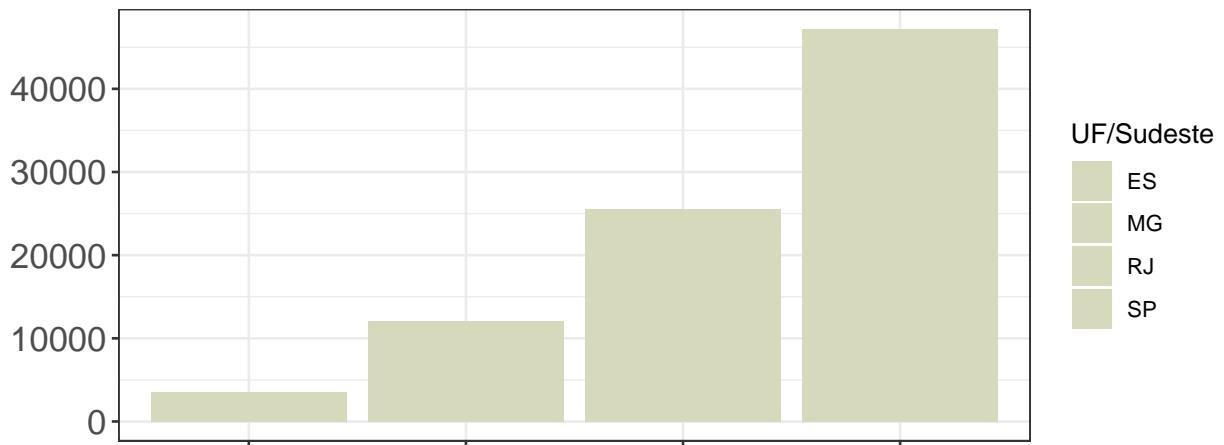


Gráfico 9: Mortes por covid-19 na região Sudeste

É importante observarmos que o Estado de São Paulo apresenta mais mortes do que a região nordeste inteira, que é a segunda maior região com casos de morte por covid-19 no país.

No que se refere ao Nordeste, o Estado do Ceará apresenta, aproximadamente, 1/4 (um quarto) do número de óbitos, no total de 10.281 (dez mil duzentos e oitenta e um) óbitos, seguido por Pernambuco e Bahia, com 9.597 (nove mil quinhentos e noventa e sete) e 8.321

(oito mil trezentos e vinte e um) óbitos, respectivamente. Esses três estados, juntos, têm quase 60% (sessenta por cento) das mortes por covid-19 no Nordeste. Sobre este assunto, o Gráfico 10 (Mortes por covid-19 na região Nordeste), exibido abaixo, apresenta a distribuição de mortes por covid-19 na região.

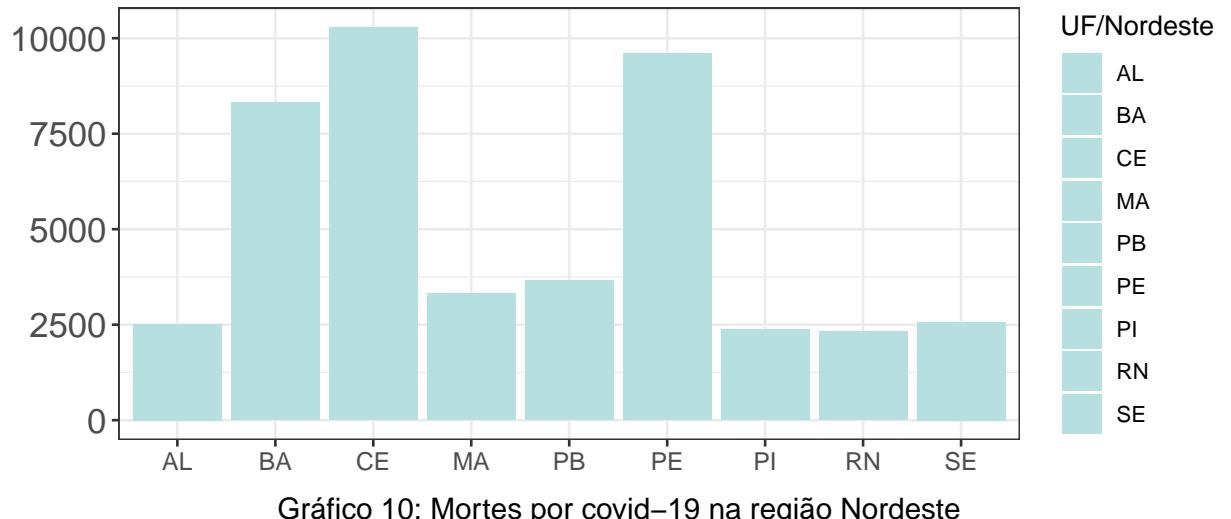


Gráfico 10: Mortes por covid-19 na região Nordeste

4.2 Taxa de mortalidade da covid-19 no Brasil

Apesar do estado de São Paulo liderar o números de mortes absoluto por covid-19 no Brasil, em relação ao tamanho da população, a situação é diferente. Verifica-se que 5 (cinco) estados do país ultrapassam São Paulo em mortes por milhão de habitantes. O Rio de Janeiro é o estado com a mais alta taxa de mortalidade do país, seguido pelo Distrito Federal. Para desmostrar segue a Tabela 2}.

RJ	DF	AM	CE	SE	SP
1.470	1.439	1.254	1.119	1.110	1.020

Tabela 2: Estados com maiores taxas de mortalidade por milhão de habitantes

Apesar da Tabela 2} apresentar apenas os 6 (seis) estados com as maiores taxas de mortalidade, pode-se visualizar no Gráfico 11 o mapa de calor das taxas de mortalidade por estado. De forma complementar, tem-se, ainda, a tabela completa com as taxas de mortalidade por covid-19 de todos os estados disponível no Apêndice A.

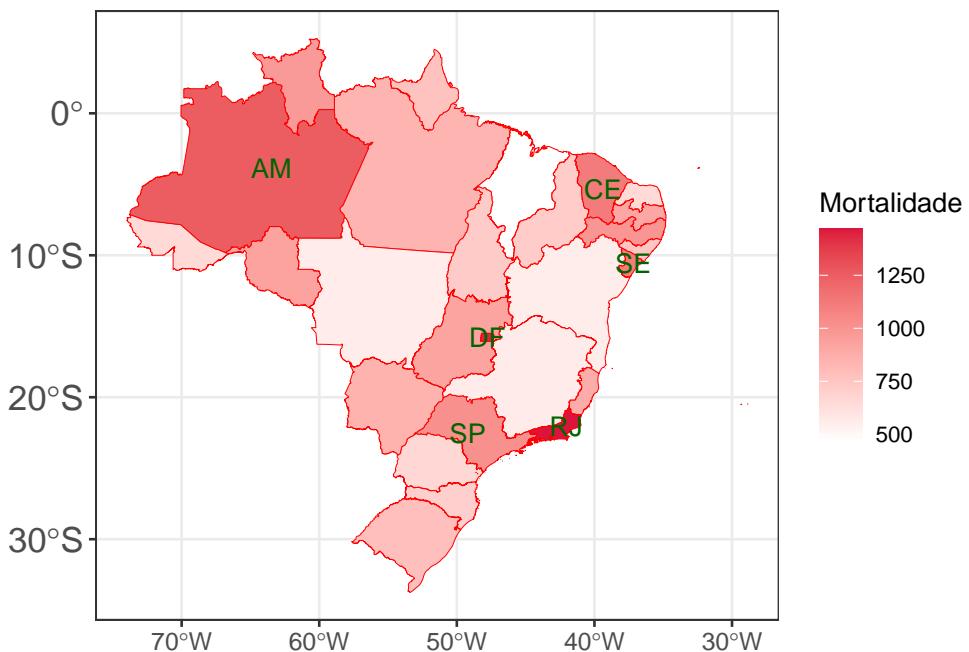


Gráfico 11: Mapa de calor das taxas de mortalidade por milhão de habitantes

No que se refere as capitais brasileiras, as taxas de mortalidade são ainda maiores que as estaduais. A média da taxa de mortalidade nas capitais é de 1.689 (um mil seiscentos e oitenta e nove) óbitos por milhão de habitantes, superior ao Estado do Rio de Janeiro, Estado com maior taxa de mortalidade do país (veja Tabela 2).

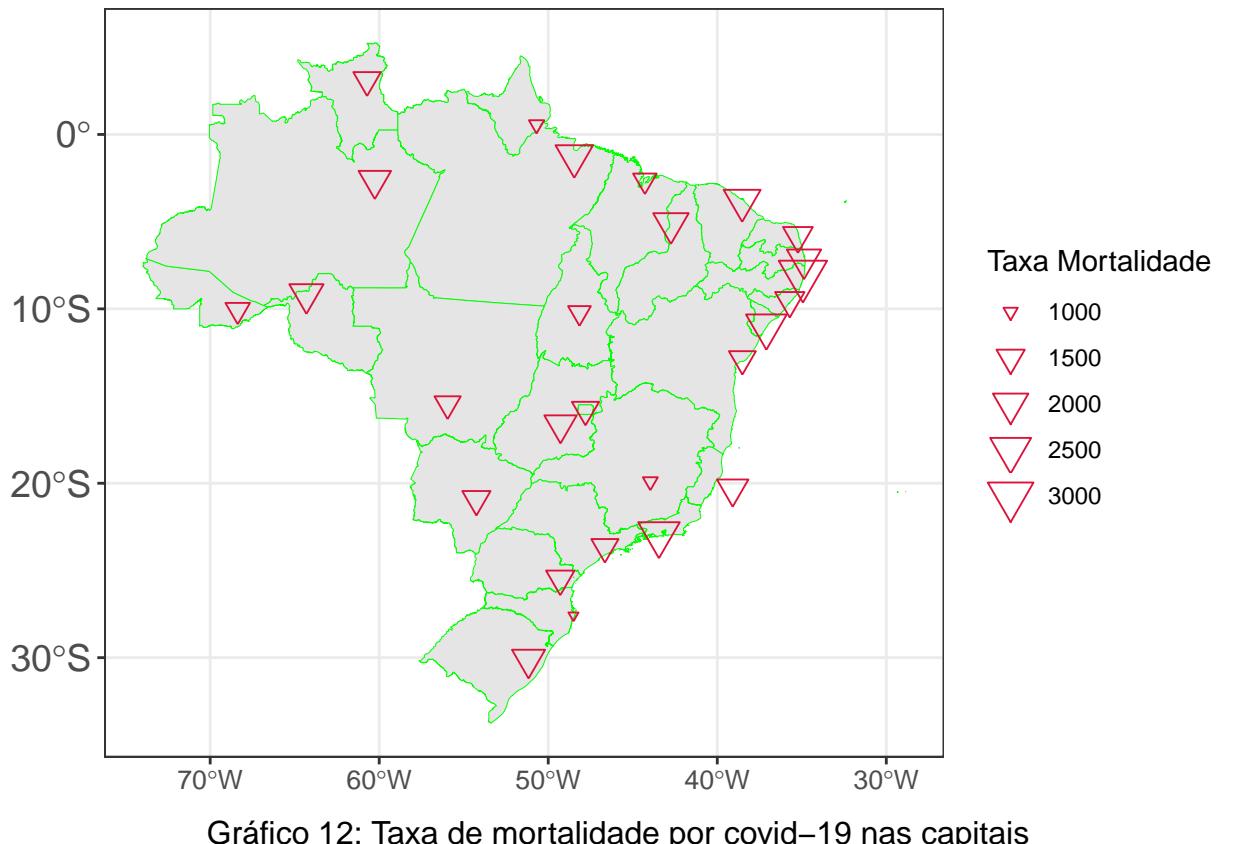
Recife é a capital com maior número de mortes por milhão de habitantes, com 3.313 (três mil trezentos de treze) óbitos por covid-19, o que representa uma taxa 96% (noventa e seis por cento) superior à média nacional das capitais.

Das 7 (sete) maiores taxas de mortalidade por covid-19 nas capitais as regiões norte e nordeste ocupam 6 posições no ranking, com exceção da cidade do Rio de Janeiro. Já o município de São Paulo tem taxa de mortalidade inferior a média nacional e ocupa a 19^a (décima nona) posição do ranking, com 1.436 mortes por milhão de habitantes, esse ranking consta na Tabela 3} abaixo.

Recife	Rio de Janeiro	Aracaju	Belém	Fortaleza	Teresina	João Pessoa
3313	2556	2472	2175	2126	1999	1913

Tabela 3: Capitais com maiores taxas de mortalidade

As taxas de mortlidade das capitais por covid-19 estão representadas em forma geométrica de triangulo invertido no Grafico 12 (Taxa de mortalidade por covid-19 nas capitais), sendo o maior trinangulo, a maior taxa de mortalidade, pondendo-se, ainda, visualizar mais detalhes das taxas de mortalidade por covid-19 nas capitais no Apêndice A, deste trabalho.



4.3 Taxa de letalidade da covid-19 no Brasil

A taxa de letalidade da covid-19 no Brasil, com verificação a partir de teste de hipótese, é de 30% (trinta por cento), com nível de significância de 95% (noventa e cinco por cento). Para este teste, utilizou-se uma amostra com 564,093 (quinhentos e sessenta e quatro mil e noventa e três) pacientes hospitalizados. Esta taxa é menor do que a taxa identificada no estudo realizado no Reino Unido, no qual comparou-se a taxa de letalidade da covid-19 com a do vírus ebola. A seguir, verifica-se o trecho de código para realização do teste de hipótese:

```

qtdObitos=dim(pacientesCovidTxLetalidade[pacientesCovidTxLetalidade$CS_EVOLUCAO=="Óbito",])[1]
totalPacientes=dim(pacientesCovidTxLetalidade[,])[1]
prop.test(qtdObitos, totalPacientes, 0.35, conf.level = 0.95)

##
## 1-sample proportions test with continuity correction
##
## data: qtdObitos out of totalPacientes, null probability 0.35
## X-squared = 4698.4, df = 1, p-value < 0.000000000000022
## alternative hypothesis: true p is not equal to 0.35
## 95 percent confidence interval:
## 0.3052663 0.3076743
## sample estimates:
## p
## 0.306469

```

A Figura 5 mostra a manchete da matéria da BBC NEWS, que faz referência ao estudo Britânico citado acima.



Figura 5: Matéria BBC NEWS

De acordo com James Gallagher (2020), repórter de ciéncia e saúde[1]:

Pessoas internadas em hospitais por conta do novo coronavírus têm a mesma probabilidade de morrer do que aquelas hospitalizadas pelo vírus do ebola, segundo nova pesquisa conduzida no Reino Unido.

No que se refere à média móvel das taxas de letalidade por covid-19 do Brasil e regiões, Norte e Nordeste apresentaram, no início da pandemia, as maiores médias do país, ambas acima de 50% (cinquenta por cento), atingindo pico na 20^a (vigésima) semana.

Nas semanas seguintes, a região Norte apresentou queda acentuada de sua taxa, chegando ao seu ponto mais baixo na 43^a (quadragésima terceira) semana. Nesta ocasião, chegou a apresentar a menor média do país. Contudo, à partir da semana 44^a (quadragésima quarta) semana voltou a subir, mantendo-se em ascensão até o término do ano de 2020, ocasião na qual ultrapassou a média das demais regiões.

Quanto ao Nordeste, este se manteve com média acima das apresentadas no Brasil e demais regiões, praticamente, durante todo ano de 2020, com exceção do mês de dezembro quando foi ultrapassada pela Região Norte.

É importante destacar que a Região Sul apresentou, durante quase todo ano de 2020, as menores médias do país. No entanto, no fim do ano, obteve média móvel superior a média nacional.

Estes resultado podem ser observados no Gráfico 13 (Média móvel semanal da taxa de letalidade), que está apresentado abaixo:

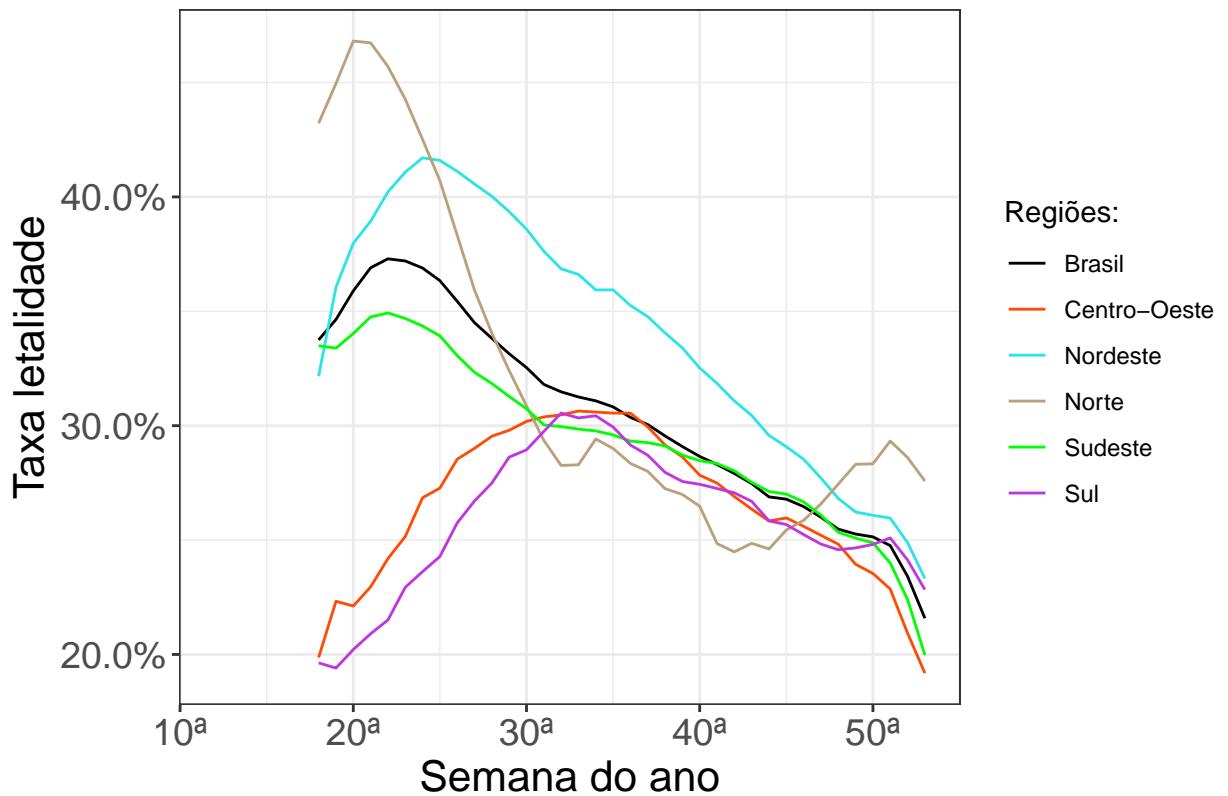


Gráfico 13: Média móvel semanal da taxa de letalidade

5. Inferência Estatística

No capítulo anterior, foram analisados os número absolutos de óbitos por covid-19, a média móvel da taxa de letalida nacional por região, bem como as taxas de mortalidade de covid-19 nas regiões, estados e capitais brasileiras.

Neste capítulo, o foco será identificar o perfil do paciente hospitalizado com maior e menor risco de morte por covid-19, no Brasil.

5.1 Perfil do brasileiro com maior e menor risco de morte

Dos atributos disponíveis no data frame, foram selecionadas as seguintes colunas para construção dos perfis de pacientes hospitalizados com maior risco de morte: sexo, idade, existência de comorbidades, raça e região.

Na tentativa de reduzir os número de perfis e torná-los mais representativos para este modelo, foram adotados alguns tratamentos que serão detalhados nos próximos parágrafos.

No que se refere à idade, os valores foram agrupados em intervalos, adotando-se os mesmos intervalos utilizados na piramide etária adotada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), conforme Tabela 4} abaixo:

0 a 4 anos	20 a 24 anos	40 a 44 anos	60 a 64 anos	80 anos ou mais
5 a 9 anos	25 a 29 anos	45 a 49 anos	65 a 69 anos	
10 a 14 anos	30 a 34 anos	50 a 54 anos	70 a 74 anos	
15 a 19 anos	35 a 39 anos	55 a 59 anos	75 a 79 anos	

Tabela 4: Faixa Etária

No que se refere à comorbidade, as colunas que detalham os tipos de comorbidades foram ignoradas para construção deste modelo, sendo adotada apenas a coluna fator de risco, que indica a existência de doença preexistente ou não.

As execuções necessárias para a otimização deste modelo foram realizadas em duas etapas:

A) Primeira execução:

Na primeira execução o modelo de regressão logística apresentou valores de baixa significância para:

- faixa etária: nos intervalos de 5 a 9 anos, 10 a 14 anos, bem como não exibiu a faixa de 0 a 4 anos;
- raça do paciente: na cor branca, bem como não exibiu a cor amarela.

O código utilizado na execução do modelo de regressão logística pode ser verificado no trecho de código abaixo, bem como o resultado da execução na Tabela 5.

```
modeloRegressaoLogistica = glm(formula = IN_OBITO_COVID ~ CS_SEXO +
FAIXA_ETARIA + CS_RACA + IN_FATOR_RISC + REGIAO,
data = dadosRegressaoLogistica, family = binomial(link = "logit"))
modeloRegressaoLogistica.table <- xtable(modeloRegressaoLogistica)
print(modeloRegressaoLogistica.table, type = "LaTeX" )
```

	β^1	Std. Error	z value	Pr(> z)
β^0	-3.3356	0.0731	-45.61	0.0000
Homem	0.2391	0.0071	33.49	0.0000
05 a 09 anos	-0.1248	0.1091	-1.14	0.2526
10 a 14 anos	0.1606	0.1144	1.40	0.1603
15 a 19 anos	0.4831	0.0934	5.17	0.0000
20 a 24 anos	0.4469	0.0802	5.57	0.0000
25 a 29 anos	0.4227	0.0745	5.68	0.0000
30 a 34 anos	0.5345	0.0706	7.57	0.0000
35 a 39 anos	0.6648	0.0686	9.69	0.0000
40 a 44 anos	0.8634	0.0677	12.75	0.0000
45 a 49 anos	1.0554	0.0672	15.70	0.0000
50 a 54 anos	1.2525	0.0668	18.76	0.0000
55 a 59 anos	1.5067	0.0665	22.67	0.0000
60 a 64 anos	1.8241	0.0663	27.52	0.0000
65 a 69 anos	2.0901	0.0662	31.57	0.0000
70 a 74 anos	2.3196	0.0662	35.03	0.0000
75 a 79 anos	2.5236	0.0663	38.04	0.0000
80 anos ou mais	2.9123	0.0660	44.13	0.0000
Branca	0.0187	0.0302	0.62	0.5363
Indígena	0.3564	0.0641	5.56	0.0000
Parda	0.1592	0.0302	5.28	0.0000
Preta	0.3490	0.0328	10.65	0.0000
Fator de Risco	0.5169	0.0082	62.95	0.0000
Nordeste	0.3166	0.0150	21.12	0.0000
Norte	0.3953	0.0173	22.87	0.0000
Sudeste	0.0907	0.0138	6.60	0.0000
Sul	-0.1248	0.0162	-7.69	0.0000

Tabela 5: Modelo com valores de baixa significância presentes

B) Segunda execução:

Após o ajuste do modelo e a reexecução em uma amostra com 436.361 (quatrocentos e trinta e seis mil e trezentos e sessenta e um) pacientes. Conclui-se que o perfil com maior risco de morte tem: idade de 80 (oitenta) anos ou mais (β^1 : 2.9126), já apresenta doenças

preexistentes (β^1 : 0.5170), reside na região norte (β^1 : 0.3953), é de cor preta (β^1 : 0.3310) e homem (β^1 : 0.2390), com probabilidade de óbito de 75% (setenta e cinco por cento).

Já o perfil com menor risco de morte no Brasil: reside na região sul (β^1 : -0.1241), tem idade de 0 (zero) a 14 (quatorze) anos, não tem doenças preexistentes, é de cor branca ou amarela e do sexo feminino, com probabilidade de óbito por covid-19 de 3% (três por cento). Abaixo segue o trecho de código responsável pela segunda execução do modelo.

```
modeloRegressaoLogistica = glm(formula = IN_OBITO_COVID ~ CS_SEXO +
FAIXA_ETARIA + CS_RACA + IN_FATOR_RISC + REGIAO,
data = dadosRegressaoLogistica, family = binomial(link = "logit"))
modeloRegressaoLogistica.table <- xtable(modeloRegressaoLogistica)
print(modeloRegressaoLogistica.table, type = "LaTeX" )
```

Os valores de β^0 e β^1 resultado da reexecução do modelo ajustado podem ser verificados na Tabela: 6}.

	β^1	Std. Error	z value	Pr(> z)
β^0	-3.3180	0.0482	-68.85	0.0000
Homem	0.2390	0.0071	33.48	0.0000
15 a 19 anos	0.4834	0.0809	5.98	0.0000
20 a 24 anos	0.4472	0.0653	6.85	0.0000
25 a 29 anos	0.4230	0.0580	7.30	0.0000
30 a 34 anos	0.5348	0.0530	10.10	0.0000
35 a 39 anos	0.6651	0.0502	13.24	0.0000
40 a 44 anos	0.8637	0.0490	17.62	0.0000
45 a 49 anos	1.0557	0.0483	21.85	0.0000
50 a 54 anos	1.2528	0.0477	26.26	0.0000
55 a 59 anos	1.5069	0.0473	31.89	0.0000
60 a 64 anos	1.8243	0.0470	38.81	0.0000
65 a 69 anos	2.0904	0.0469	44.58	0.0000
70 a 74 anos	2.3199	0.0469	49.45	0.0000
75 a 79 anos	2.5238	0.0471	53.59	0.0000
80 anos ou mais	2.9126	0.0466	62.50	0.0000
Indígena	0.3382	0.0572	5.91	0.0000
Parda	0.1413	0.0086	16.48	0.0000
Preta	0.3310	0.0149	22.21	0.0000
Fator de Risco	0.5170	0.0082	62.96	0.0000
Nordeste	0.3165	0.0150	21.11	0.0000
Norte	0.3953	0.0173	22.86	0.0000
Sudeste	0.0910	0.0137	6.62	0.0000
Sul	-0.1241	0.0162	-7.67	0.0000

Tabela 6: Modelo de regressão logística otimizado

6. Machine learning

Neste capítulo, utiliza-se técnicas de machine learning para identificar outros atributos com impacto no risco de óbito dos pacientes, além dos atributos já descritos nas seções anteriores novos atributos serão incorporados aos modelos. Os algoritmos, também serão diversificados.

A eficiência dos atributos selecionados, bem como o algoritmo utilizado, será mensurada a partir da taxa de acerto do algoritmo.

Os dados utilizados nos treinamentos dos modelos, bem como os dados utilizados nos testes, serão extraídos a partir de uma amostra aleatória com 1000 (um mil) registros. Sendo 70% (setenta por cento) da amostra destinada para o treinamento e 30% (trinta por cento) destinada para os testes.

6.1 Árvore de Decisão de Classificação

Na primeira execução, o algoritmo utilizado será o rpart e os atributos serão os mesmos utilizados nas seção de inferência estatística. Com essa configuração a taxa de acerto do modelo foi 64% (sessenta e quatro por cento).

```
arvoreClassificacao = rpart(formula = CS_EVOLUCAO ~ CS_SEXO +
FAIXA_ETARIA + CS_RACA + IN_FATOR_RISC + REGIAO,
data = dadosMachineLearningObito, method = "class")

teste = predict(arvoreClassificacao, newdata = dadosTeste)
previsao=cbind(teste, dadosTeste)
previsao["PREVISAO"] = ifelse(previsao$CURA > .5, "CURA", "OBITO")

confusao=table(previsao$CS_EVOLUCAO, previsao$PREVISAO)
taxaAcerto = (confusao[1] + confusao[4]) / sum(confusao)
taxaAcerto
## [1] 0.6417066
```

6.2 Árvore de Decisão e Seleção de Atributos

Com método randomForest e mantendo praticamente os mesmo atributos da execução anterior, com exceção do atributo REGIAO, substituído pelo atributo SG_UF a taxa de acerto referente ao risco de morte do paciente, apresenta uma pequena melhora. Depois da mudança, a taxa de acertou passou para 66% (sessenta e seis por cento).

```
selecaoAtributos = randomForest(CS_EVOLUCAO ~ CS_SEXO +
FAIXA_ETARIA + CS_RACA + IN_FATOR_RISC + SG_UF, data = dadosTreino)

teste = predict(selecaoAtributos, dadosTeste)
previsao=cbind(teste, dadosTeste)
previsao["PREVISAO"] = previsao$teste
```

```

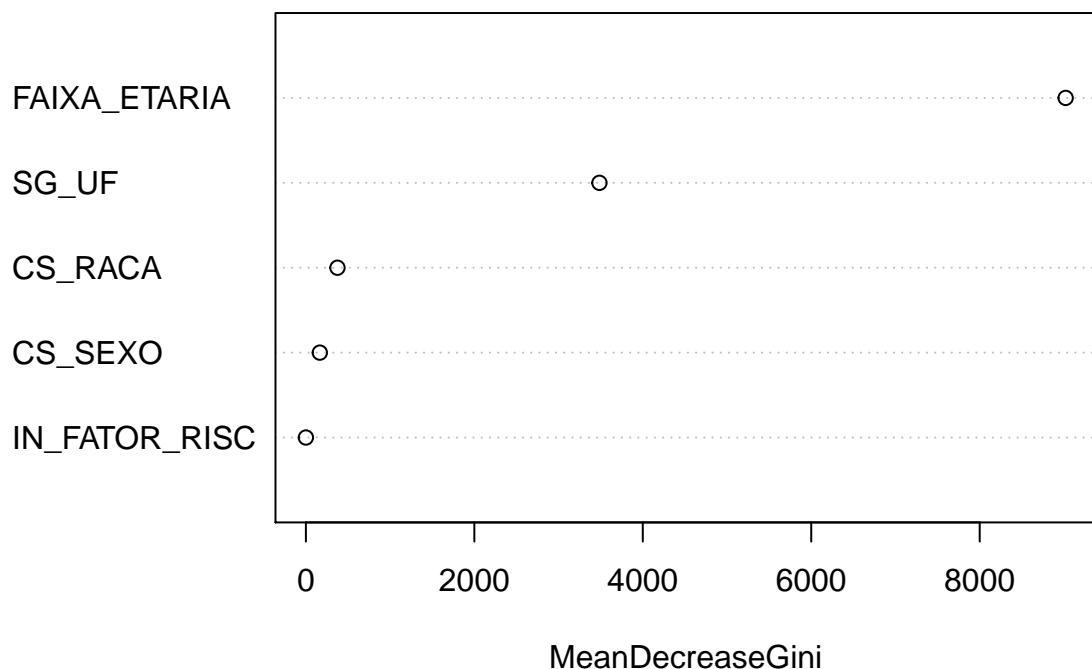
confusao=table(previsao$CS_EVOLUCAO, previsao$PREVISAO)
taxaAcerto = (confusao[1]+confusao[4])/sum(confusao)
taxaAcerto
## [1] 0.6613501

```

O gráfico abaixo apresenta os atributos mais significativos na classificação de risco de morte do paciente. A FAIXA_ETARIA do paciente é o atributo com maior peso na classificação do registro em relação ao risco de morte. Nas próximas execuções detalharemos mais os atributos FAIXA_ETARIA e SG_UF.

```
varImpPlot(selecaoAtributos, main="Atributos mais significativos")
```

Atributos mais significativos



6.3 Árvore de Decisão de Classificação otimizado

Utilizando novamente a função rpart agora com atributo NU_IDADE e MUNICIPIO, a taxa de acerto passou a ser de 67% (sessenta e sete por cento)

```
arvoreClassificacao = rpart(formula = CS_EVOLUCAO ~ NU_IDADE + MUNICIPIO,
                             data = dadosMachineLearningObito, method = "class")
```

```

teste = predict(arvoreClassificacao, newdata = dadosTeste)
previsao=cbind(teste, dadosTeste)
previsao["PREVISAO"] = ifelse(previsao$CURA > .5, "CURA", "OBITO")

```

```

confusao=table(previsao$CS_EVOLUCAO, previsao$PREVISAO)
taxaAcerto = (confusao[1]+confusao[4])/sum(confusao)
taxaAcerto

```

```
## [1] 0.678406
```

Quando adicionado o hospital de atendimento do paciente como atributo do modelo, a taxa de acerto passa a ser de 70% (setenta por cento).

```
arvoreClassificacao = rpart(formula = CS_EVOLUCAO ~ NU_IDADE + MUNICIPIO + HOSPITAL,
                             data = dadosMachineLearningObito, method = "class")

teste = predict(arvoreClassificacao, newdata = dadosTeste)
previsao=cbind(teste, dadosTeste)
previsao["PREVISAO"] = ifelse(previsao$CURA > .5, "CURA", "OBITO")

confusao=table(previsao$CS_EVOLUCAO, previsao$PREVISAO)
taxaAcerto = (confusao[1]+confusao[4])/sum(confusao)
taxaAcerto
## [1] 0.7021875
```

6.4 Naive Bayse

Com os mesmos atributos da seção anterior e o método naiveBayes do pacote e1071 a taxa de acerto foi menor, 68% (sessenta e oito por cento)

```
modeloNaiveBayes = naiveBayes(formula = CS_EVOLUCAO ~ NU_IDADE + MUNICIPIO + HOSPITAL,
                                data = dadosTreino)

teste = predict(modeloNaiveBayes, dadosTeste)
previsao=cbind(teste, dadosTeste)
previsao["PREVISAO"] = previsao$teste

confusao=table(previsao$CS_EVOLUCAO, previsao$PREVISAO)
taxaAcerto = (confusao[1]+confusao[4])/sum(confusao)
taxaAcerto
## [1] 0.682228
```

7. Resultado

Nas seções acima fica evidenciado que no ano de 2020 as regiões mais pobres do país, Norte e Nordeste, foram as regiões que mais sofreram com a pandemia, com as maiores taxas de mortalidade e letalidade. A exceção do estado do Rio de Janeiro que apresentou a maior taxa de letalidade do país.

No que se refere ao perfil do paciente foi verificado que a idade é o atributo que mais contribui para o aumento do risco de morte, seguido pela existência de comorbidades e sofrendo influência até do sexo. Os dados analisados demonstram que homens apresentam mais risco de óbito.

A partir dos atributos selecionados e dos algoritmo de machine learning utilizados neste trabalho foi verificado uma taxa de acerto de até 70% (setenta por cento) dos casos graves de covid-19 que evoluíram no ano de 2020 para óbito.

7.1 Dashboard-Covid19

De maneira complementar, a este trabalho, foi elaborado no RStudio um dashboard na linguagem R, com auxílio das libs: flexdashboard, shiny, shinyWidgets e também das libs já descritas no primeiro capítulo. O deploy do dashboard foi realizado no shinyapps.io podendo ser acessado no link <https://fabiopadilhatccpucminas.shinyapps.io/Dashboard-Covid19/>.

O dashboard amplia a possibilidade de consulta, indo além das informações apresentadas neste relatório. Ele permite também uma abordagem em drilldown sendo possível uma análise nacional, passando pelas regiões, estados chegando até o nível de municípios. A Figura 6 mostra a página principal do dashboard, a figura 7 exibe a página de sobre com os créditos.

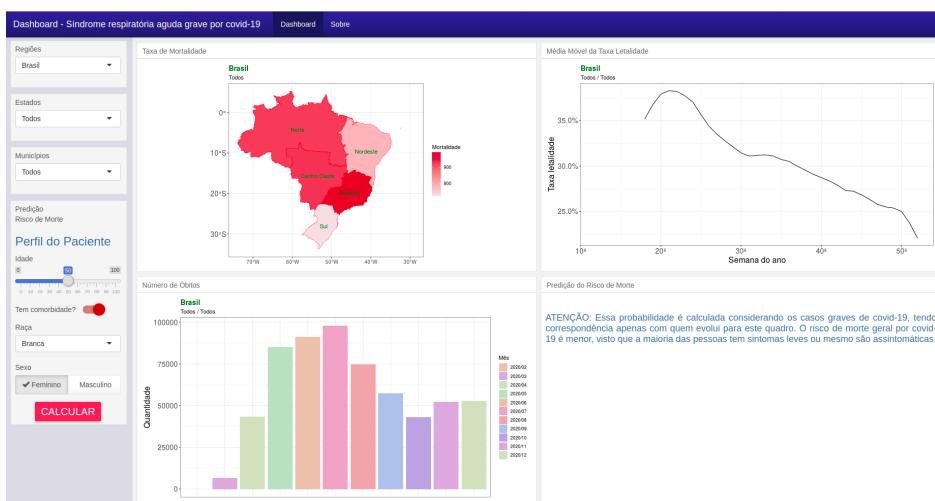


Figura 6: Dashboard

O shinyapps.io é uma plataforma de serviço na nuvem para o desenvolvimento de aplicações web usando R. A Figura 8 exibe o deploy do dashboard no shinyapps.io.

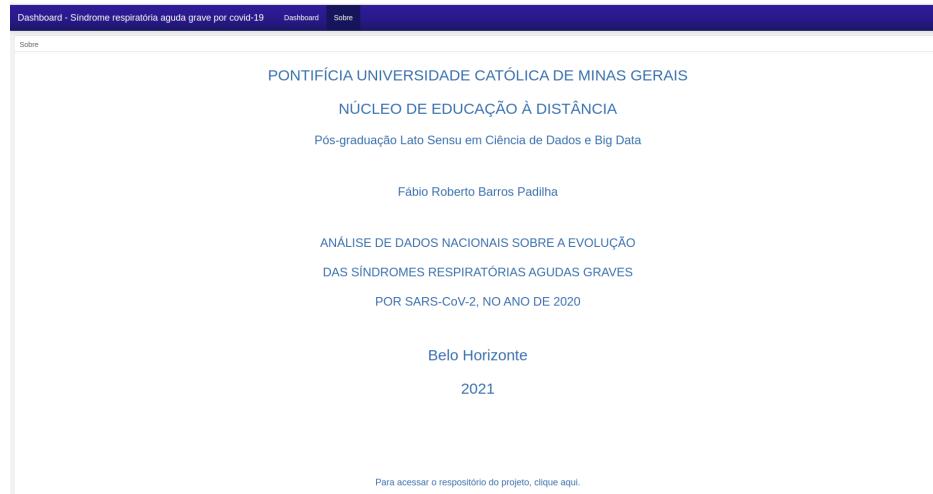


Figura 7: Sobre

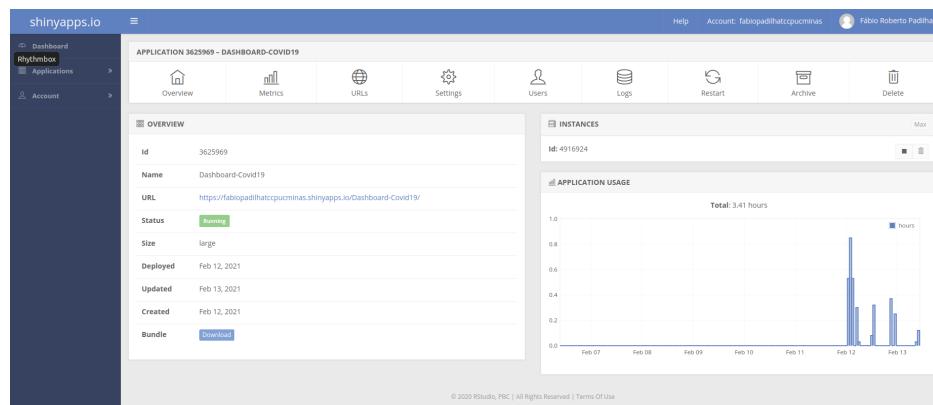


Figura 8: Shinyapps.io

8. Links

A Tabela 7 dispõe dos links com os código fontes implementados neste trabalho, do link do deploy da aplicação (dashboard implementado), bem como do link com o backup dos dados utilizados.

Git Hub	Repositório com o código fonte	https://github.com/fabio...
Google Driver	Backup dos Dados e Relatório Final	https://drive.google.com/drive...
Shinyapps.io	Dashboard Covid-19	https://fabiopadilhatccpucminas...

Tabela 7: Links TCC

Referências Bibliográficas

Referências

- [1] Gallagher, James. *Coronavírus é 'tão mortal quanto ebola' em casos de internação, diz estudo britânico.* DATA GEEKS, 29 abr. 2020. Disponível em: <https://www.bbc.com/portuguese/geral-52481818> - (Acessado em: 14 jan. 2021)
- [2] GOMES, Pedro César Tebaldi. *Classificação com Naive Bayes.* DATA GEEKS, 22 fev. 2019. Disponível em: <https://www.datageeks.com.br/naive-bayes/> - (Acessado em: 14 jan. 2021)
- [3] Hosokawa, Eric Ossamu. *Técnica de Árvore de Decisão em Mineração de Dados.* 2011. *Trabalho de Conclusão de Curso (Tecnólogo em Processamento de Dados)* - Faculdade de Tecnologia de São Paulo, São Paulo, 2011.
- [4] JUNIOR, Gilson Medeiros de Oliveira. *Máquina de Vetores Suporte: estudo e análise de parâmetros para otimização de resultado.* 2010. *Trabalho de Conclusão de Curso (Bacharelado em Ciência da Computação)* - Universidade Federal de Pernambuco, Recife, 2010.
- [5] MEHTA, R. *Big Data Analytics with Java.* Birmingham: Packt Publishing Ltd, 2017.
- [6] Portal OpenDataSUS, *SRAG 2021 - Banco de Dados de Síndrome Respiratória.* OpenDataSUS, 10 jan. 2021. Disponível em: <https://opendatasus.saude.gov.br/dataset/bd-srag-2021>
- [7] SANTOS, Roberto Moreira. *A importância do uso de ferramentas de analítica preditiva gratuitas para as empresas.* 2017. *Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação)* - Universidade Federal de Pernambuco, Recife, 2017.

Apêndice A - Taxa de Mortalidade Estados e Capitais

	UF	População	Mortalidade	Capital	População	Mortalidade
1	RJ	17.366.189	1.470	Rio de Janeiro	6.747.815	2.556
2	DF	3.055.149	1.439	Brasília	3.055.149	1.439
3	AM	4.2077.14	1.254	Manaus	2.219.580	1.774
4	CE	9.187.103	1.119	Fortaleza	2.686.612	2.126
5	SE	2.318.822	1.110	Aracaju	664.908	2.472
6	SP	46.289.333	1.020	São Paulo	12.325.232	1.436
7	PE	9.616.621	997	Recife	1.653.461	3.313
8	RR	631.181	969	Boa Vista	419.652	1.458
9	RO	1.796.460	927	Porto Velho	539.354	1.902
10	GO	7.113.540	924	Goiânia	1.536.097	1.764
11	PB	4.039.277	906	João Pessoa	817.511	1.913
12	ES	4.064.052	879	Vitória	365.855	1.697
13	MS	2.809.394	849	Campo Grande	906.092	1.498
14	PA	8.690.745	839	Belém	1.499.641	2175
15	RS	11.422.973	788	Porto Alegre	1.488.252	1.820
16	AP	861.773	770	Macapá	512.902	1.015
17	AL	3.351.543	751	Maceió	1.025.360	1.566
18	TO	1.590.248	728	Palmas	306.296	1.227
19	PI	3.281.480	723	Teresina	868.075	1.999
20	SC	7.252.502	706	Florianópolis	508.826	972
21	PR	11.516.840	666	Curitiba	1.948.626	1.508
22	RN	3.534.165	655	Natal	890.480	1.593
23	AC	894.470	655	Rio Branco	413.418	1.306
24	MG	21.292.666	565	Belo Horizonte	2.521.564	1.009
25	MT	3.526.220	560	Cuiabá	618.124	1.389
26	BA	14.930.634	557	Salvador	2.886.698	1.426
27	MA	7.8114.598	467	São Luís	1.108.975	1.264

Tabela 8: Taxa Mortalidade Estados e Capitais

Anexo A - Ficha de Notificação

Nº

MINISTÉRIO DA SAÚDE
SECRETARIA DE VIGILÂNCIA EM SAÚDE

SIVEP Gripe - SISTEMA DE INFORMAÇÃO DE VIGILÂNCIA EPIDEMIOLÓGICA DA GRIPE
FICHA DE REGISTRO INDIVIDUAL - CASOS DE SÍNDROME RESPIRATÓRIA AGUDA GRAVE HOSPITALIZADO - 27/07/2020

CASO DE SÍNDROME RESPIRATÓRIA AGUDA GRAVE (SRAG-HOSPITALIZADO): Indivíduo com *SG que apresente: dispneia/desconforto respiratório OU pressão persistente no tórax OU saturação de O₂ menor que 95% em ar ambiente OU coloração azulada dos lábios ou rosto. (*SG: Indivíduo com quadro respiratório agudo, caracterizado por pelo menos dois (2) dos seguintes sinais e sintomas: febre (mesmo que referida), calafrios, dor de garganta, dor de cabeça, tosse, coriza, distúrbios olfativos ou gustativos).

Para efeito de notificação no Sivep-Gripe, devem ser considerados os casos de SRAG hospitalizados ou os óbitos por SRAG independente de hospitalização.

1	Data do preenchimento da ficha de notificação:		2	Data de 1 ^{as} sintomas	
3	UF:	4	Município: Código (IBGE):		
5	Unidade de Saúde: Código (CNES):				
Dados do Paciente					
6	CPF do cidadão: _____				
7	Nome: _____				
9	Data de nascimento: _____	10	(Ou) Idade: _____ 1-Dia 2-Mês 3-Ano _____	8	Sexo: _____ 1-Masc. 2-Fem. 9-Ign
11	Gestante: _____ 1-1º Trimestre 2-2º Trimestre 3-3º Trimestre 4-Idade Gestacional Ignorada 5-Não 6-Não se aplica 9-Ignorado				
12	Raça/Cor: _____ 1-Branca 2-Preta 3-Amarela 4-Parda 5-Indígena 9-Ignorado				
13	Se indígena, qual etnia?				
14	Escolaridade: _____ 0-Sem escolaridade/Analfabeto 1-Fundamental 1º ciclo (1 ^a a 5 ^a série) 2-Fundamental 2º ciclo (6 ^a a 9 ^a série) 3-Médio (1 ^º ao 3 ^º ano) 4-Superior 5-Não se aplica 9-Ignorado				
15	Ocupação:		16	Nome da mãe:	
17	CEP: _____ - _____				
18	UF: _____	19	Município: _____ Código (IBGE): _____	22	Nº:
20	Bairro: _____		21	Logradouro (Rua, Avenida, etc.): _____	
23	Complemento (apto, casa, etc...): _____			24	(DDD) Telefone: _____ - _____
25	Zona: _____ 1-Urbana 2-Rural 3-Periurbana 9-Ignorado	26	País: (se residente fora do Brasil)		
Dados de Residência					
27	Paciente tem histórico de viagem internacional até 14 dias antes do início dos sintomas? _____ 1-Sim 2-Não 9-Ign				
28	Se sim: Qual país? _____ Em qual local? _____				
30	Data da viagem: _____	31	Data do retorno: _____		
32	É caso proveniente de surto de SG que evoluiu para SRAG? _____ 1-Sim 2-Não 9-Ignorado				
33	Trata-se de caso nosocomial (infecção adquirida no hospital)? _____ 1-Sim 2-Não 9-Ignorado				
34	Paciente trabalha ou tem contato direto com aves, suínos, ou outro animal? _____ 1-Sim 2-Não 3- Outro, qual _____ 9-Ignorado				
35	Sinais e Sintomas: 1-Sim 2-Não 9-Ignorado 1- Febre 2- Tosse 3- Dor de Garganta 4- Dispneia 5- Desconforto Respiratório 6- Saturação O ₂ <95% 7- Diarreia 8- Vômito 9- Dor abdominal 10- Fadiga 11- Perda do olfato 12- Perda do paladar 13- Outros _____				
36	Possui fatores de risco/comorbidades? _____ 1-Sim 2-Não 9-Ignorado Se sim, qual(is)? (Marcar X) 1- Puérpera (até 45 dias do parto) 2- Doença Cardiovascular Crônica 3- Doença Hematológica Crônica 4- Síndrome de Down 5- Doença Hepática Crônica 6- Asma 7- Diabetes mellitus 8- Doença Neurológica Crônica 9- Outra Pneumopatia Crônica 10- Imunodeficiência/Imunodepressão 11- Doença Renal Crônica 12- Obesidade, IMC _____ 13- Outros _____				
37	Recebeu vacina contra Gripe na última campanha? _____ 1-Sim 2-Não 9-Ignorado			38	Data da vacinação: _____
Se < 6 meses: a mãe recebeu a vacina? _____ 1-Sim 2-Não 9-Ignorado			Se sim, data: _____		
a mãe amamenta a criança? _____ 1-Sim 2-Não 9-Ignorado					
Se >= 6 meses e <= 8 anos: Data da dose única 1/1: _____ (dose única para crianças vacinadas em campanhas de anos anteriores) Data da 1 ^ª dose: _____ (1 ^ª dose para crianças vacinadas pela primeira vez) Data da 2 ^ª dose: _____ (2 ^ª dose para crianças vacinadas pela primeira vez)					

Dados de Atendimento	39	Usou antiviral para gripe? __ 1-Sim 2-Não 9-Ignorado	40	Qual antiviral? __ 1-Oseltamivir 2-Zanamivir 3-Outro, especifique: _____	41	Data início do tratamento _____ _____ _____ _____ _____ _____
	42	Houve internação? __ 1-Sim 2-Não 9-Ignorado	43	Data da internação por SRAG: _____ _____ _____ _____ _____ _____	44	UF de internação: _____ _____ _____ _____ _____ _____
	45	Município de internação: _____ _____ _____ _____ _____ _____		Código (IBGE): _____ _____ _____ _____ _____ _____		
	46	Unidade de Saúde de internação: _____ _____ _____ _____ _____ _____		Código (CNES): _____ _____ _____ _____ _____ _____		
	47	Internado em UTI? __ 1-Sim 2-Não 9-Ignorado	48	Data da entrada na UTI: _____ _____ _____ _____ _____ _____	49	Data da saída da UTI: _____ _____ _____ _____ _____ _____
	50	Uso de suporte ventilatório: __ 1-Sim, invasivo 2-Sim, não invasivo 3-Não 9-Ignorado	51	Raio X de Tórax: __ 1-Normal 2-Infiltrado intersticial 3-Consolidação 4-Misto 5-Outro: _____ 6-Não realizado 9-Ignorado	52	Data do Raio X: _____ _____ _____ _____ _____ _____
	53	Aspecto Tomografia __ 1-Típico COVID-19 2-Indeterminado COVID-19 3-Atípico COVID-19 4-Negativo para Pneumonia 5-Outro 6-Não realizado 9-Ignorado	54	Data da tomografia: _____ _____ _____ _____ _____ _____		
	55	Coletou amostra __ 1-Sim 2-Não 9-Ignorado	56	Data da coleta: _____ _____ _____ _____ _____ _____	57	Tipo de amostra: __ 1-Secréção de Naso-orofaringe 2-Lavado Bronco-alveolar 3-Tecido post-mortem 4-Outra, qual? _____ 5-LCR 9-Ignorado
	Dados Laboratoriais	58	Nº Requisição do GAL: _____	59	Tipo do teste para pesquisa de抗igenos virais: __ 1-Imunofluorescência (IF) 2-Teste rápido antigênico	
60		Data do resultado da pesquisa de抗igenos: _____ _____ _____ _____ _____ _____	61	Resultado da Teste antigênico: __ 1-positivo 2-Negativo 3-Inconclusivo 4-Não realizado 5-Aguardando resultado 9-Ignorado		
62		Laboratório que realizou o Teste antigênico: _____ _____ _____ _____ _____ _____		Código (CNES): _____ _____ _____ _____ _____ _____		
63		Agente Etiológico - Teste antigênico: Positivo para Influenza? __ Se sim, qual influenza? __ 1-Influenza A 2-Influenza B Se outros vírus respiratórios qual(is)? (marcar X) __ SARS-CoV-2 __ Virus Sincicial Respiratório __ Parainfluenza 1 __ Parainfluenza 2 __ Parainfluenza 3 __ Adenovírus __ Outro vírus respiratório, especifique: _____				
64		Resultado da RT-PCR/outro método por Biologia Molecular: __ 1-Detectável 2-Não Detectável 3-Inconclusivo 4-Não realizado 5-Aguardando resultado 9-Ignorado	65	Data do resultado RT-PCR/outro método por Biologia Molecular: _____		
66		Agente Etiológico - RT-PCR/outro método por Biologia Molecular: Positivo para Influenza? __ 1-Sim 2-Não 9-Ignorado Se sim, qual influenza? __ 1-Influenza A 2-Influenza B Influenza A, qual subtipo? __ 1-Influenza A(H1N1)pdm09 2-Influenza A/H3N2 3-Influenza A não subtipado 4-Influenza A não subtipável 5-Inconclusivo 6-Outro, especifique: _____				
		Influenza B, qual linhagem? __ 1-Victoria 2-Yamagatha 3-Não realizado 4-Inconclusivo 5-Outro, especifique: _____				
		Positivo para outros vírus? __ 1-Sim 2-Não 9-Ignorado Se outros vírus respiratórios, qual(is)? (marcar X) __ SARS-CoV-2 __ Virus Sincicial Respiratório __ Parainfluenza 1 __ Parainfluenza 2 __ Parainfluenza 3 __ Parainfluenza 4 __ Adenovírus __ Metapneumovírus __ Bocavírus __ Rinovírus __ Outro vírus respiratório, especifique: _____				
67		Laboratório que realizou RT-PCR/outro método por Biologia Molecular: _____ _____ _____ _____ _____ _____		Código (CNES): _____ _____ _____ _____ _____ _____		
68		Tipo de amostra sorológica para SARS-CoV-2: __ 1-Sangue/plasma/soro 2-Outra, qual? _____ 9-Ignorado	69	Data da coleta: _____ _____ _____ _____ _____ _____		
70	Tipo de Sorologia para SARS-CoV-2: __ 1-Teste rápido 2-Elisa 3-Quimioluminescência 4-Outro, qual? Resultado do Teste Sorológico para SARS-CoV-2: __ IgG __ IgM __ IgA 1-Positivo 2-Negativo 3-Inconclusivo 4-Não realizado 5-Aguarda resultado 9-Ignorado	71	Data do resultado: _____ _____ _____ _____ _____ _____			
Conclusão	72	Classificação final do caso: __ 1-SRAG por influenza 2-SRAG por outro vírus respiratório 3-SRAG por outro agente etiológico, qual _____ 4-SRAG não especificado 5-SRAG por COVID-19	73	Critério de Encerramento: __ 1-Laboratorial 2-Clínico Epidemiológico 3-Clínico 4-Clínico-Imagen		
	74	Evolução do Caso: __ 1-Cura 2-Óbito 3-Óbito por outras Causas 9-Ignorado	75	Data da alta ou óbito: __ __ __	76	Data do Encerramento: _____ _____ _____ _____ _____ _____
	77	Número D.O: __ __ __ __ __ __ - __				
78	OBSERVAÇÕES:					
79	Profissional de Saúde Responsável:			80	Registro Conselho/Matrícula: _____ _____ _____ _____ _____ _____	

Anexo B - Dicionário de dados



MINISTÉRIO DA SAÚDE
SECRETARIA DE VIGILÂNCIA EM SAÚDE

SIVEP Gripe
SISTEMA DE INFORMAÇÃO DE VIGILÂNCIA EPIDEMIOLÓGICA DA GRIPE

Dicionário de Dados

FICHA DE REGISTRO INDIVIDUAL – CASOS DE SÍNDROME RESPIRATÓRIA AGUDA GRAVE HOSPITALIZADOS

Este documento tem como finalidade descrever as variáveis exportadas para o banco de dados em DBF.

CAMPO OBRIGATÓRIO é aquele cuja ausência de dado impossibilita a inclusão do registro no sistema.
CAMPO ESSENCIAL é aquele que, apesar de não ser obrigatório, registra dado necessário à investigação do caso ou ao cálculo de indicador epidemiológico ou operacional.

CAMPO INTERNO é aquele que apesar de não constar na ficha e não aparecer no display da tela, é preenchido automaticamente pelo sistema.

CAMPO OPCIONAL é aquele que só deve ser preenchido caso seja necessário, aparece no display da tela e consta no banco de dados.

Nome do campo	Tipo	Categoria	Descrição	Características	DBF
Nº	Varchar2(12)		Número do registro	Campo Interno Número sequencial gerado automaticamente pelo sistema. Utilizar o padrão: 320120000123	NU_NOTIFIC
1-Data do preenchimento da ficha de notificação	Date DD/MM/AAAA		Data de preenchimento da ficha de notificação.	Campo Obrigatório Data deve ser <= a data da digitação.	DT_NOTIFIC
Semana Epidemiológica do preenchimento da ficha de notificação	Varchar2(6)		Semana Epidemiológica do preenchimento da ficha de	Campo Interno Calculado a partir da data dos Primeiros Sintomas. (SS)	SEM_NOT

			notificação.	
2-Data de 1ºs sintomas	Date DD/MM/AAAA	Data da 1º sintomas do caso.	Campo Obrigatório Da ta deve ser <= a data da digitação e data do preenchimento da ficha de notificação	DT_SIN_PRI
Semana Epidemiológica dos Primeiros Sintomas	Varchar(2(6)	Semana Epidemiológica do inicio dos sintomas.	Campo Interno Calculado a partir da data dos Primeiros Sintomas. (SS)	SEM_PRI
3-UF	Varchar(2(2)	Tabela com código e siglas das UF padronizados pelo IBGE.	Campo Obrigatório Unidade Federativa onde está localizada a Unidade Sentinelas que realizou a notificação.	SG_UF_NOT
			<p>Se usuário que está digitando a ficha for de nível:</p> <ul style="list-style-type: none"> ▪ <u>Unidade Sentinelas</u> - o campo é preenchido automaticamente pelo sistema com a UF, município e unidade onde está cadastrado o usuário. ▪ <u>Municipal</u> – o campo é preenchido automaticamente pelo sistema com a UF e município onde está cadastrado o usuário. ▪ <u>Estadual</u> – o campo é preenchido automaticamente pelo sistema com a UF do usuário. ▪ <u>Federal</u> - abre tabela com todas as UF que possuem unidades sentinelas cadastradas no sistema. 	
4-Município Código (IBGE)	Varchar(2 (6)	Tabela com código e nomes dos Municípios padronizados pelo IBGE.	Campo Obrigatório Município onde está localizada a Unidade Sentinelas que realizou a notificação.	ID_MUNICIP OU CO_MUN_NOT
			<p>Se usuário que está digitando a ficha for de nível:</p> <ul style="list-style-type: none"> ▪ <u>Unidade Sentinelas</u> – o campo é preenchido automaticamente pelo sistema com o Município onde está localizada a unidade de notificação. ▪ <u>Municipal</u> – o campo é preenchido automaticamente pelo sistema com o município do usuário. ▪ <u>Estadual ou Federal</u> – abre tabela com todos os municípios da UF selecionada no campo 3 que possuem unidades sentinelas cadastradas no sistema. 	
Regional de Saúde de Notificação Código (IBGE)	Varchar(2 (6)	Tabela com código e nomes das Regionais de Saúde dos municípios de notificação padronizados pelo IBGE.	Campo Interno Regional de Saúde onde está localizado o Município realizou a notificação.	ID_REGIONA OU CO_REGIONA

			<ul style="list-style-type: none"> ▪ <u>Unidade Sentinelas</u> – o campo é preenchido automaticamente pelo sistema com a Regional do Município onde está localizada a unidade de notificação. ▪ <u>Municipal</u> – o campo é preenchido automaticamente pelo sistema com a regional do município do usuário. 	
5-Unidade de Saúde Código (CNES)	Varchar(7)	Tabela com códigos CNES e nomes das Unidades Sentinelas cadastradas no sistema.	<p>Campo Obrigatório</p> <p><u>ID_UNIDADE_OU_CO_UNI_NOT</u></p> <p>Unidade Sentinelas que realizou coleta de amostra e registro docaso.</p> <p>Preenchendo o nome da unidade, o código é preenchido automaticamente, e vice-versa;</p> <p>Se usuário que está digitando a ficha for de nível:</p> <ul style="list-style-type: none"> ▪ <u>Unidade Sentinelas</u> - o campo é preenchido automaticamente pelo sistema. ▪ <u>Municipal</u> – abre tabela com as unidades sentinelas do município. ▪ <u>Estadual ou Federal</u> – abre tabela com as unidades sentinelas do município selecionado o campo 4. 	
6-CPF do paciente	Varchar(15)		<p>Campo Essencial</p> <p><u>NU_CPF</u></p> <p>Número do Cadastro de Pessoa Física (CPF) do paciente notificado</p> <p>Quando preenchido o número do CPF o sistema deve preencher o Nome, Sexo, Data de Nascimento, Idade, Raça/Cor e o Nome da mãe do paciente.</p>	
7-Nome	Varchar(70)		<p>Campo Obrigatório</p> <p>Nome completo do paciente (sem abreviações)</p>	<u>NM_PACIENT</u>
8-Sexo	Varchar(1)	1-Masculino 2-Feminino 9-Ignorado	<p>Campo Obrigatório</p> <p>Sexo do paciente.</p>	<u>CS_SEXO</u>
9-Data de nascimento	Date DD/MM/AAAA		<p>Campo Essencial</p> <p><u>DT_NASC</u></p> <p>Data de nascimento do paciente.</p>	
10-(ou) Idade	Varchar(3)		<p>Campo Obrigatório</p> <p><u>NU_IDADE_N</u></p> <p>Idade informada pelo paciente quando não se sabe a data de nascimento.</p> <p>Se digitado a data de nascimento, a idade é calculada e preenchida automaticamente pelo sistema; considerando o intervalo entre a data de nascimento e a data dos primeiros sintomas.</p> <p>Na falta desse dado é registrada a idade</p>	

(ou) Tipo/Idade	Varcha r2(1) 1-Dia 2-Mês 3-Ano	aparente.	Campo Obrigatório Se digitado a data de nascimento, o campo Idade/Tipo é calculado e preenchido automaticamente pelo sistema, considerando o intervalo entre a data de nascimento e a data dos primeiros sintomas. Se a diferença for de 0 a 30 dias, o sistema grava em Idade = (nº dias) e em Tipo = 1-Dia. Por exemplo: se Data de nascimento = 05/12/2012 e Data dos 1ºs sintomas = 11/12/2012, então Idade = 6 e Tipo = 1-Dia. Se a diferença for de 1 a 11 meses, o sistema grava em Idade = (nº meses) e em Tipo = 2-Mês. Por exemplo: se Data de nascimento = 05/10/2012 e Data dos 1ºs sintomas = 11/12/2012, então Idade = 2 e Tipo = 2-Mês. Se a diferença for maior ou igual a 12 meses, o sistema grava em Idade = (nº anos) e em Tipo = 3-Ano. Por exemplo: se Data de nascimento = 05/10/2011 e Data dos 1ºs sintomas = 11/12/2012, então Idade = 1 e Tipo = 3-Ano.	TP_IDADE
11-Gestante	Varcha r2(1) 1-1º Trimestre 2-2º Trimestre 3-3º Trimestre 4-Idade Gestacional Ignorada 5-Não 6-Não se aplica 9-Ignorado	Idade gestacional da paciente.	Campo Obrigatório Se selecionado categoria 2-Feminino no campo Sexo. Se selecionado sexo igual a Masculino ou a idade for menor ou igual a 9 anos, o campo é preenchido automaticamente com 6-Não se aplica. Se selecionado sexo igual a Feminino e idade for maior que 9 anos, o campo não pode ser preenchido com 6-Não se aplica.	CS_GESTANT
12-Raça/Cor	Varcha r2(2) 1-Branca 2-Preta 3-Amarela 4-Parda 5-Indígena 9-Ignorado	Cor ou raça declarada pelo paciente: Branca; Preta; Amarela; Parda (pessoa que se declarou mulata, cabocla, caufuzá, ma meluca ou mestiza de preto com pessoa de outra cor ou raça); e, Indígena.	Campo Obrigatório	CS_RACA
13-Se indígena, qual etnia?	Varcha r2(4)	Tabelado SIASI com código e nomes das etnias indígenas.	Nome e código da etnia do paciente,	Campo Essencial
				CS_ETINIA

			quando indígena.	Habilitado se campo 11-Raça/Cor for iguala 5-Indígena.	
14-Escolaridade	Varchar(2(1)	0-Sem escolaridade/ Ainda lêabeto 1-Fundamental 1º ciclo (1ª a 5ª série) 2-Fundamental 2º ciclo (6ª a 9ª série) 3-Médio (1º a o 3º ano) 4-Superior 5-Não se aplica 9-Ignorado	Nível de escolaridade do paciente. Para os níveis fundamental e médio deve ser considerada a última série ou ano concluído.	Campo Essencial Preenchido automaticamente com a categoria “não se aplica” quando idade for menor que 7 anos Quando idade for maior que 7 anos, o campo não pode ser preenchido com “não se aplica”.	CS_ESCOL_N
15- Ocupação	Varchar(2(6)	Tabela com código da Ocupação da Classificação Brasileira de Ocupações (CBO).	Ocupação profissional do paciente	Campo Essencial	PAC_COCBO ou PAC_DSCBO
16-Nome da mãe	Varchar(2(70)		Nome completo da mãe do paciente (se em abreviações).	Campo Essencial	NM_MAE_PAC
17-CEP	Varchar(2(8)		CEP de residência do paciente.	Campo Essencial	NU_CEP
18-UF	Varchar(2(2)	Tabela com código e siglas das UF padronizados pelo IBGE.	Unidade Federativa de residência do paciente.	Campo Obrigatório Validado a partir da tabela de CEP dos Correios. Se campo 25-País for Brasil.	SG_UF
Regional de Saúde de Residência Código (IBGE)	Varchar(2(6)	Tabela com código e nomes das Regionais de Saúde dos municípios de residência padronizados pelo IBGE.	Regional de Saúde onde está localizado o Município de residência do paciente.	Campo Interno Preenchendo o nome da regional de saúde de residência, o código é preenchido automaticamente, e vice-versa;	ID_RG_RESI CO_RG_RESI
19-Município Código (IBGE)	Varchar(2(6)	Tabela com código e nome dos Municípios padronizados pelo IBGE.	Município de residência do paciente.	Campo Obrigatório Se campo 25-País for Brasil.	ID_MIN_RESI CO_MUN_RES

			Se o CEP não for preenchido, o campo é habilitado depois de selecionada uma UF no campo 17. Nesse caso, o sistema abre ta belacom os municípios da UF.
			Preenchendo o nome do município, o código é preenchido automaticamente, ou vice-versa.
20-Bairro	Varchar(2)	Bairro de residência do paciente.	Campo Essencial Se preenchido o campo CEP, o Bairro é preenchido automaticamente pelo sistema.
21-Logradouro (Rua, Avenida, etc.)	Varchar(250)	Logradouro (rua, avenida, quadra, travessa, etc.) do endereço de residência do paciente.	Campo Essencial Se preenchido o campo CEP, o logradouro é preenchido automaticamente pelo sistema.
22-Nº	Varchar(2)	Nº do logradouro (nº da casa ou do edifício).	Campo Essencial NU_NUMERO
23-Complemento (apto, casa, etc.)	Varchar(15)	Complemento do logradouro (bloco, apto, casa, etc.).	Campo Essencial NU_COMPLEM
24-(DDD) Telefone	Varchar(2)	Código DDD e número de telefone para contato do paciente.	Código DDD e número de telefone para contato do paciente. NU_DDD_TEL OU NU_TELEFON
25-Zona	Varchar(2)	Zona geográfica do endereço de residência do paciente.	Campo Essencial CS_ZONA
26-País (se residente fora do Brasil)	Varchar(3)	País de residência do paciente.	Campo Obrigatório ID_PAIS OU CO_PAIS
27- Paciente tem histórico de viagem internacional até 14 dias antes do	Varchar(2)	Se preenchido CEP, ou se selecionada uma UF, o campo País é preenchido automaticamente pelo sistema e desabilitado para edição. Se selecionado País diferente de Brasil, os campos 17 a 23 são desabilitados.	Campo Obrigatório HISTO_VGM

início dos sintomas?	9-Ignorado			
28- Se sim: Qual país?	Varcha2(3)	Tabela com código e nome dos Países.	Campo Obrigatório Habilitado se campo 27- Paciente tem histórico de viagem, selecionado categoria 1-Sim em Paciente tem histórico de viagem internacional a té 14 dias antes do inicio dos sintomas?	PAIS_VGM OU CO_PS_VGM
29- Em qual local?	Varchar2(30)	Local (cidade, estado, província e outros) onde foi realizada a viagem	Campo Essencial Habilitado e selecionado categoria 1-Sim em campo 27- Paciente tem histórico de viagem internacional a té 14 dias antes do inicio dos sintomas?	LO_PS_VGM
30- Data da viagem	Date DD/MM/AAAA	Data e m que foi realizada a viagem	Campo Obrigatório Se preenchido campo 27- Paciente tem histórico de viagem internacional a té 14 dias antes do inicio dos sintomas? (campo 26), deve ser informada a data de viagem	DT_VGM
31- Data de retorno	Date DD/MM/AAAA	Data e m que retornou de viagem	Campo Obrigatório Se preenchido campo 27- Paciente tem histórico de viagem internacional a té 14 dias antes do inicio dos sintomas? (campo 26), deve ser informada a data de retorno	DT_RT_VGM
32- É caso proveniente de surto de SG que evoluiu para SRAG?	Varcha2 (1) 1-Sim 2-Não 9-Ignorado	Caso é proveniente de surto de SG?	Campo Essencial	SURTO_SG
33-Trata-se de caso nosocomial (infecção adquirida no hospital)?	Varcha2 (1) 1-Sim 2-Não 9-Ignorado	Caso de SRAG com infecção adquirida após internação.	Campo Essencial Quando o campo 33 for igual a 1, é permitido digitar data de inicio dos sintomas posterior a data de internação.	NOSOCOMIAL
34- Paciente trabalha ou tem contato direto com aves, suínos, ou outro animal?	Varcha2(1) 1-Sim 2-Não 9-Ignorado	Caso com contato direto com aves ou suínos.	Campo Essencial	AVE_SUINO
Paciente trabalha ou tem contato direto com aves, suínos/Outro animal (especificar)	Varchar2(60)	Informar o animal que o paciente teve contato se selecionado a opção 3.	Campo Essencial Habilitado de campo 34- Contato com outro animal = 3 (Outro).	OUT_ANIM
35-Sinais e Sintomas/Febre	Varcha2 (1) 1-Sim 2-Não 9-Ignorado	Paciente apresentou febre?	Campo Essencial	FEBRE
35-Sinais e Sintomas/Tosse	Varcha2 (1) 1-Sim 2-Não	Paciente apresentou tosse?	Campo Essencial	TOSSF

		9-Ignorado			
35-Sinais e Sintomas/Dor de Garganta	Varcha r2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou dor de garganta?	Campo Essencial	GARGANTA
35-Sinais e Sintomas/Dispneia	Varcha r2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou dispneia?	Campo Essencial	DISPNEIA
35-Sinais e Sintomas/Desconforto Respiratório	Varcha r2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou desconforto respiratório?	Campo Essencial	DESC_RESP
35-Sinais e Sintomas/Saturação O₂< 95%	Varcha r2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou saturação O ₂ < 95%?	Campo Essencial	SATURACAO
35-Sinais e Sintomas/Diarréia	Varcha r2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou diarreia?	Campo Essencial	DIARREIA
35-Sinais e Sintomas/Vômito	Varcha r2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou vômito?	Campo Essencial	VOMITO
35-Sinais e Sintomas/Dor abdominal	Varcha r2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou dor abdominal?	Campo Essencial	DOR_ABD
35-Sinais e Sintomas/Fadiga	Varcha r2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou fadiga?	Campo Essencial	FADIGA
35-Sinais e Sintomas/Perda do Olfato	Varcha r2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou perda do olfato?	Campo Essencial	PERD_OLF
35-Sinais e Sintomas/Perda do Paladar	Varcha r2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou perda do paladar?	Campo Essencial	PERD_PALA
35-Sinais e Sintomas/Outros	Varcha r2(1)	1-Sim 2-Não 9-Ignorado	Paciente apresentou outro(s) sintoma(s)?	Campo Essencial	OUTRO_SIN
35-Sinais e Sintomas/Outros (Descrição)	Varcha r2(30)		Listar outros sintomas e sintomas	Campo Essencial	OUTRO_DES

			apresentados pelo paciente.	Habilitado se selecionado categoria 1-Sim em Sinais e Sintomas/Outros.
36-Fatores de risco	Várcha/2 (1)	1-Sim 2-Não 9-Ignorado	Paciente apresenta algum fator de risco	Campo Essencial
36-Fatores de risco/ Puérpera	Várcha/2 (1)	1-Sim 2-Não 9-Ignorado	Paciente é puérpera ou parturiente (mulher que pariu recentemente – até 45 dias do parto)?	Campo Essencial Habilitado se selecionado no campo 8- Sexo Feminino.
36-Fatores de risco/ Doença Cardiovascular Crônica	Várcha/2 (1)	1-Sim 2-Não 9-Ignorado	Paciente possui Doença Cardiovacular Crônica ?	CARDIOPATI
36-Fatores de risco/ Doença Hematológica Crônica	Várcha/2 (1)	1-Sim 2-Não 9-Ignorado	Paciente possui Doença Hematológica Crônica ?	HEMATOLOGI
36-Fatores de risco/ Síndrome de Down	Várcha/2 (1)	1-Sim 2-Não 9-Ignorado	Paciente possui Síndrome de Down?	SIND_DOWN
36-Fatores de risco/ Doença Hepática Crônica	Várcha/2 (1)	1-Sim 2-Não 9-Ignorado	Paciente possui Doença Hepática Crônica ?	HEPATICA
36-Fatores de risco/ Asma	Várcha/2 (1)	1-Sim 2-Não 9-Ignorado	Paciente possui Asma ?	ASMA
36-Fatores de risco/ Diabetes mellitus	Várcha/2 (1)	1-Sim 2-Não 9-Ignorado	Paciente possui Diabetes mellitus?	DIABETES
36-Fatores de risco/ Neurológica Crônica	Várcha/2 (1)	1-Sim 2-Não 9-Ignorado	Paciente possui Doença Neurológica ?	NEUROLOGIC
36-Fatores de risco/ Pneumopatia Crônica	Outra	Várcha/2 (1)	Paciente possui outra pneumopatia crônica?	PNEUMOPATI
36-Fatores de risco/ Imunodeficiência	Várcha/2 (1)	1-Sim	Paciente possui	IMUNODEPRE

ou Imunodepressão		2-Não 9-Ignorado	Imunodeficiência ou Imunodepressão (diminuição da função do sistema imunológico)?		
36-Fatores de risco/ Doença Renal Crônica	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Paciente possui Doença Renal Crônica ?	Campo Essencial	RENAL
36-Fatores de risco/ Obesidade	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Paciente possui obesidade?	Campo Essencial	OBESIDADE
36-Fatores de risco/ Obesidade (Descrição IMC)	Varchar2(3)		Valor do IMC (Índice de Massa Corporal) do paciente calculado pelo profissional de saúde.	Campo Essencial	OBES_IMC
36-Fatores de risco/ Outros	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Paciente possui outro(s) fator(es) de risco?	Campo Essencial	OUT_MORBI
36-Fatores de risco/ Outros (Descrição)	Outros	Varchar2(30)	Listar outro(s) fator(es) de risco do paciente.	Campo Essencial	MORB_DESC
37-Recebeu vacina contra Gripe na última campanha?		Varchar2(1)	1-Sim 2-Não 9-Ignorado	Informar se o paciente foi vacinado contra gripe na última campanha, após verificar a documentação / caderneta.	VACINA

			resposta.		
38-Data da vacinação	Date DD/MM/AAAA		Da ta da última dose de vacina contra gripe que o paciente tomou.	Campo Essencial Habilitar campo 31-Recebeu vacina contra Gripe for igual a 1.	DT_UT_DOSE
Se < 6 meses: a mãe recebeu a vacina?	Varchar(2 (1) 1-Sim 2-Não 9-Ignorado	Se paciente < 6 meses, a mãe recebeu vacina?	Da ta deve ser <= a data da digitação (data atual).	Mae_VAC	
Se sim, data	Date DD/MM/AAAA	Se a mãe recebeu vacina, qual a data?	Da ta deve ser <= a data da digitação (data atual).	DT_VAC_MAE	
Se < 6 meses: a mãe amamenta a criança?	Varchar(2 (1) 1-Sim 2-Não 9-Ignorado	Se paciente < 6 meses, a mãe amamenta a criança?	Da ta deve ser <= a data da digitação (data atual).	M_AMAMENTA	
Se >= 6 meses e <= 8 anos: Data da dose única /1	Date DD/MM/AAAA	Se >= 6 meses e <= 8 anos, data da dose única para crianças vacinadas em campanhas de anos anteriores	Da ta deve ser >= 6 meses e <= 8 anos	DT_DOSEUNI	
Se >= 6 meses e <= 8 anos: Data da 1ª dose	Date DD/MM/AAAA	Se >= 6 meses e <= 8 anos, data da 1ª dose para crianças vacinadas pela primeira vez	Da ta deve ser >= 6 meses e <= 8 anos	DT_1_DOSE	
Se >= 6 meses e <= 8 anos: Data da 2ª dose	Date DD/MM/AAAA	Se >= 6 meses e <= 8 anos data da 2ª dose para crianças vacinadas pela primeira vez	Da ta deve ser >= 6 meses e <= 8 anos	DT_2_DOSE	
39-Usou antiviral para gripe?	Varchar(2 (1) 1-Sim	Fez uso de antiviral		ANTIVIRAL	

		2-Não 9-Ignorado	para tratamento da doença?		
40-Qual antiviral?	Varchar(2 (1)	1- Oseletamivir 2- Zanamivir 3- Outro	Qual antiviral utilizado?	Campo Essencial Habilitado se campo 39- Qual antiviral? for igual a 1.	TP_ANTIVIR
Qual antiviral /Outro, especifique	Varchar(2(30)		Se o antiviral utilizado não foi Oseletamivir ou Zanamivir, informar qual antiviral foi utilizado.	Campo Essencial Habilitado se campo 40- Qual antiviral? for igual a 3.	OUT_ANTIV
41-Data do início do tratamento	Date DD/MM/AAAA		Data em que foi iniciado o tratamento com o antiviral.	Campo Essencial Habilitado se campo 39-Usou antiviral para gripe? for igual a 1. Data deve ser <= a data da digitação (data atual).	DT_ANTIVIR
42-Houve internação?	Varchar(2(1)	1-Sim 2-Não 9-Ignorado	O paciente foi internado?	Campo Essencial Caso o campo não seja igual a 1 – Sim o sistema emitirá um aviso indicando que não atende a definição de caso.	HOSPITAL
43-Data da internação por SRAg	Date DD/MM/AAAA		Data em que o paciente foi hospitalizado.	Campo Obrigatório Data deve ser maior ou igual a 2- Data de 19s sintomas e menor ou igual a data da digitação (atual).	DT_INTERNIA
44-UF de internação	Varchar(2(2)	Tabela com código e siglas das UF padronizados pelo IBGE.	Unidade Federativa de internação do paciente.	Campo Essencial Habilitado se campo 42-Houve internação? for igual a 1	SG_UF_INTE
Regional de Saúde de Internação Código (IBGE)	Varchar(2 (6)	Tabela com código e nomes das Regionais de Saúde dos municípios de internação padronizados pelo IBGE.	Regional de Saúde onde está localizado o município de internação do paciente.	Campo Interno Preenchendo o nome da regional de saúde de internação, o código é preenchido automaticamente, e vice-versa.	ID_RG_INTE OU CO_RG_INTE
45-Município de internação/ Código(IBGE)	Varchar(2(20)	Tabela com código e nomes dos Municípios padronizados pelo IBGE.	Município onde está localizada a Unidade de Saúde onde o paciente internou.	Campo Essencial Habilitado se campo 42-Houve internação? for igual a 1	ID_MN_INTE OU CO_MU_INTE

46-Unidade de Saúde de internação/ Código CNES	Varchar(20)	Tabela com códigos CNES e nomes das Unidades de Saúde.	Campo Essencial Habilitado e campo 42-Houve internação? for igual a 1	ID_UN_INTE OU CO_UN_INTE
47-Internado em UTI?	Varchar(1)	1-Sim 2-Não 9-Ignorado	Campo Essencial	UTI
48-Data da entrada na UTI	Data DD/MM/AAAA	Data de entrada do paciente na Unidade de Terapia intensiva (UTI).	Campo Essencial Habilitado e campo 47-Internado em UTI? for igual a 1. Data deve ser maior ou igual a 2-Data de 1ºs sintomas da SRAG e menor ou igual a data da digitação (atual).	DT_ENTUTI
49-Data da saída da UTI	Data DD/MM/AAAA	Data em que o paciente saiu da Unidade de Terapia intensiva (UTI).	Campo Essencial Habilitado e campo 47-Internado em UTI? for igual a 1. Data deve ser maior ou igual a 48-Data da entrada na UTI e menor ou igual a data da digitação (atual).	DT_SAIDUTI
50-Uso de suporte ventilatório?	Varchar(1)	1-Sim, invasivo 2-Sim, não invasivo 3-Não 9-Ignorado	Campo Essencial O paciente fez uso de suporte ventilatório?	SUPPORT_VEN
51- Raio X de Tórax	Varchar(1)	1-Normal 2-Infiltrado intersticial 3-Consolidação 4-Misto 5-Outro 6-Não realizado 9-Ignorado	Campo Essencial Informar resultado de Raio X de Tórax.	RAIOX_RES
Raio X de Tórax/ Outro (especificar)	Varchar(2)(30)		Campo Essencial Informar o resultado do RX de tórax se selecionado opção 5-Outro.	RAIOX_OUT
52-Data do Raio X	Data DD/MM/AAAA		Campo Essencial Se realizou RX de Tórax, especificar a data do exame.	DT_RAIOX
53- Aspecto Tomografia	Number(3)	1-Típico COVID-19 2-Indeterminado COVID-19	Campo Essencial Informar o resultado da tomografia.	TOMO_RES

		3-Atípico COVID-19 4-Negativo para Pneumonia 5-Outro 6-Não realizado 9-Ignorado	tomografia.		
Aspecto Tomografia/Outro (especificar)	Varchar2(100)		Informar o resultado da tomografia se selecionado a opção 5-Outro	Campo Essencial Habilitado de campo 53- Aspecto Tomografia = 5 (Outro)	TOMO_OUT
54- Data da Tomografia	Date DD/MM/AAAA		Se realizou tomografia, especificar a data do exame.	Campo Essencial Habilitado se campo 53- Aspecto Tomografia for igual a 1, 2, 3, 4 ou 5.	DT_TOMO
55-Coletou amostra?	Varchar2(1) 1-Sim 2-Não 9-Ignorado		Foi realizado coleta de amostra para realização de teste diagnóstico?	Campo Essencial	AMOSTRA
56-Data da Coleta	Date DD/MM/AAAA		Data da coleta da amostra para realização do teste diagnóstico.	Campo Essencial Habilitado de campo 55-Coletou amostra? = 1. Data deve ser maior ou igual a 2-Data de 1's sintomas e menor ou igual a data da digitação (atual).	DT_COLETA
57-Tipo de amostra	Varchar2(30) 1-Secreção de Naso-orofaringe 2-Lavado Broco-alveolar 3-Tecido post-mortem 4-Outra, qual? 5-LLR 9-Ignorado		Tipo da amostra clínica coletada para o teste diagnóstico.	Campo Essencial Habilitado de campo 55-Coletou amostra? = 1.	TP_AMOSTRA
Tipos de amostra/Outra	Varchar2(30)		Descrição do tipo da amostra clínica, caso diferente das listadas nas categorias do campo.	Campo Essencial Campo habilitado se selecionado categoria 4-Outra, qual em Tipo de amostra.	OUT_AMOST
58-Nº da Requisição do GAL			Número da requisição de exames gerado	Campo Essencial	REQUI_GAL

			pelo sistema GAL.		
59- Tipo do Teste antígenico	Number(3)	1-Imunofluorescência (IF) 2- Teste rápido antígenico	Tipo do teste antígenico que foi realizado.	Campo Essencial	TP_TES_AN
60- Data do resultado teste Antígenico	Data DD/MM/AAAA		Data do resultado do teste antígenico.	Campo Essencial	DT_RES_AN
61- Resultado do Teste Antígenico	Varchar2(1)	1-positivo 2-Negativo 3- Inconclusivo 4-Não realizado 5-Aguardando resultado 9-Ignorado	Resultado do Teste Antígenico Esse campo só será marcado com 5-Aguardando Resultado e estará habilitado se o campo 55-Coletou amostra? = 1	Campo Essencial	RES_AN
62-Laboratório que realizou o Teste antígenico	Varchar2(70)	Nomes dos Laboratórios cadastrados no sistema	Laboratório responsável pela liberação do resultado do teste antígenico.	Campo Essencial	LAB_AN
62-Laboratório que realizou o Teste antígenico	Varchar2(7)	Table com códigos CNES		Campo Essencial	CO_LAB_AN
63-Agente etiológico – Teste Antígenico. Positivo para Influenza?	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Resultado do Teste Antígenico que foi positivo para Influenza	Campo Essencial	POS_AN_FLU
63-Agente etiológico – Teste Antígenico. Se sim, qual Influenza?	Varchar2(1)	1-Influenza A 2-Influenza B	Resultado do Teste Antígenico, para o tipo de Influenza.	Campo Essencial	TP_FLU_AN
63-Agente etiológico – Teste Antígenico. Positivo para outros vírus?	Varchar2(1)	1-Sim 2-Não 9-Ignorado	Resultado do Teste Antígenico, que foi positivo para outro vírus respiratório.	Campo Essencial	POS_AN_OUT
63-Agente etiológico – Teste Antígenico. SARS-CoV-2	Varchar2(1)	1-marca do pelo usuário Vazio - não o marcado	Resultado do Teste Antígenico, para SARS-CoV-2.	Campo Essencial	AN_SARS2
63-Agente etiológico – Teste Antígenico.	Varchar2(1)	1-marca do pelo usuário Vazio - não o marcado	Resultado do Teste Antígenico, para outros vírus? = 1.	Campo Essencial	AN_VSR

VSR			VSR.	Habilitado se campo 63-Agente etiológico, Teste Antigênico. Positivo para outros vírus? = 1.
63-Agente etiológico – Antigênico. Parainfluenza 1	Teste	Varcha1/2 (1)	1-marcado pelo usuário Vazio - não o marcado	Resultado do Teste Antigênico, para Parainfluenza 1.
63-Agente etiológico – Teste Antigênico. Parainfluenza 2	Teste	Varcha1/2 (1)	1-marcado pelo usuário Vazio - não o marcado	Resultado do Teste Antigênico. Para influenza 2.
63-Agente etiológico – Antigênico. Parainfluenza 3	Teste	Varcha1/2(1)	1-marcado pelo usuário Vazio - não o marcado	Resultado do Teste Antigênico. Para influenza 3.
63-Agente etiológico – Antigênico. Adenovírus	Teste	Varcha1/2(1)	1-marcado pelo usuário Vazio - não o marcado	Resultado do Teste Antigênico. Adenovírus.
63-Agente etiológico – Antigênico. Outro vírus respiratório	Teste	Varcha1/2 (1)	1-marcado pelo usuário Vazio - não o marcado	Resultado do Teste Antigênico. Outro vírus respiratório.
63-Agente etiológico – Outro vírus respiratório (Descrição)	Teste	Varcha1/2(30)		Nome do outro vírus identificado pelo Teste Antigênico.
64-Resultado da RT-PCR/outro método por Biologia Molecular		Varcha1/2 (1)	1-Detectável 2-Não Detectável 3-Inconclusivo 4-Não realizado 5-Aguardando Resultado 9-Ignorado	Resultado do teste de RT-PCR/outro método por Biologia Molecular.
65-Data do Resultado RT-PCR/outro método por Biologia Molecular		Da te DD/MM/AAAA		Data da Resultado RT-PCR/outro método por Biologia Molecular
				Campo Essencial Esse campo virá marcado com 5-Aguardando Resultado e estará habilitado se o campo 55-Coletou amostra? = 1.
				PCR_RESULT
				DT_PCR
				Data deve ser >= a data da coleta - campo 56.

66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Positivo para Influenza?	Várcha r2(1) 1-Sim 2-Não 9-Ignorado	Resultado da RT-PCR foi positivo para Influenza	Campo Essencial	POS_PCRFLU
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Se sim, qual Influenza?	Várcha r2(1) 1-Influenza A 2-Influenza B	Resultado diagnóstico do RT-PCR para o tipo de Influenza.	Campo Essencial	TP_FLU_PCR
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Se Influenza A, qual subtipo?	Várcha r2(1) 1-Influenza A(H1N1)pdm09 2-Influenza A(H3N2) 3-Influenza A não subtipado 4-Influenza A não subtipável 5-Inconclusivo 6-Outro, especifique:	Subtipo para Influenza A. Habilitado se campo 66-Agente etiológico – RT-PCR/outro método por Biologia Molecular: Positivo para Influenza? = 1.	Campo Essencial	PCR_FLUASU
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Se Influenza A, qual subtipo? Outro, especifique:	Várcha r2(30)	Outro subtipo para Influenza A. Habilitado se Agente etiológico – RT-PCR/outro método por Biologia Molecular: Se Influenza A, qual subtipo? = 6.	Campo Essencial	FLUASU_OUT
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Se Influenza B, qual linhagem?	Várcha r2(1) 1-Victoria 2-Yamagata 3-Não realizado 4-Inconclusivo 5-Outro, especifique:	Linhagem para Influenza B. Habilitado se campo 66-Agente etiológico – RT-PCR/outro método por Biologia Molecular: Se sim, qual Influenza? = 2.	Campo Essencial	PCR_FLUBLI
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Se Influenza B, qual linhagem? Outro, especifique:	Várcha r2(30)	Outra linhagem para Influenza B. Habilitado se 66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Se Influenza B, qual linhagem? = 5.	Campo Essencial	FLUBLI_OUT
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Positivo para outros vírus?	Várcha r2(1) 1-Sim 2-Não 9-Ignorado	Resultado da RT-PCR foi positivo para outro vírus respiratório	Campo Essencial	POS_PCRMOUT
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: SARS-CoV-2	Várcha r2(1) 1-imarcado pelo usuário Vazio - não o marcado	Resultado diagnóstico do RT-PCR para (SARS-CoV-2).	Campo Essencial	PCR_SARS2
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: VSR	Várcha r2(1) 1-imarcado pelo usuário Vazio - não o marcado	Resultado diagnóstico do RT-PCR para (VSR).	Campo Essencial	PCR_VSR

				Biologia Molecular: Positivo para outros vírus? = 1	
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Parainfluenza 1	Várcha/2 (1)	1-marcado pelo usuário Vazio -não marcado	Resultado diagnóstico do RT-PCR para Parainfluenza 1.	Campo Essencial Habilitado se campo 66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Positivo para outros vírus? = 1	PCR_PARA1
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Parainfluenza 2	Várcha/2 (1)	1-marcado pelo usuário Vazio -não marcado	Resultado diagnóstico do RT-PCR para Parainfluenza 2.	Campo Essencial Habilitado se campo 66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Positivo para outros vírus? = 1	PCR_PARA2
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Parainfluenza 3	Várcha/2 (1)	1-marcado pelo usuário Vazio -não marcado	Resultado diagnóstico do RT-PCR para Parainfluenza 3.	Campo Essencial Habilitado se campo 66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Positivo para outros vírus? = 1	PCR_PARA3
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Parainfluenza 4	Várcha/2 (1)	1-marcado pelo usuário Vazio -não marcado	Resultado diagnóstico do RT-PCR para Parainfluenza 4.	Campo Essencial Habilitado se campo 66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Positivo para outros vírus? = 1	PCR_PARA4
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Adenovírus	Várcha/2 (1)	1-marcado pelo usuário Vazio -não marcado	Resultado diagnóstico do RT-PCR para Adenovírus.	Campo Essencial Habilitado se campo 66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Positivo para outros vírus? = 1	PCR_ADENO
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Metapneumovírus	Várcha/2 (1)	1-marcado pelo usuário Vazio -não marcado	Resultado diagnóstico do RT-PCR para Metapneumovírus.	Campo Essencial Habilitado se campo 66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Positivo para outros vírus? = 1	PCR_METAP
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Bocavírus	Várcha/2 (1)	1-marcado pelo usuário Vazio -não marcado	Resultado diagnóstico do RT-PCR para Bocavírus.	Campo Essencial Habilitado se campo 66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Positivo para outros vírus? = 1	PCR_BOCA
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Rinovírus	Várcha/2 (1)	1-marcado pelo usuário Vazio -não marcado	Resultado diagnóstico do RT-PCR para Rinovírus.	Campo Essencial Habilitado se campo 66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Positivo para outros vírus? = 1	PCR_RINO

			Biologia Molecular: Positivo para outros vírus? = 1	
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Outro vírus respiratório, especifique:	Varchar2 (1)	1-marcado pelo usuário Vazio -não o marcado	Campo Essencial Resultado diagnóstico do RT-PCR para Outro vírus respiratório.	PCR_OUTRO
66- Agente etiológico – RT-PCR/outro método por Biologia Molecular: Outro vírus respiratório (Descrição)	Varchar2 (30)		Campo Essencial Nome do outro vírus respiratório identificado pelo RT-PCR.	DS_PCR_OUT
67-Laboratório que realizou RT-PCR/outro método por Biologia Molecular Código (CNES)	Varchar2 (7)	Tabela com códigos CNES e nomes dos laboratórios cadastrados no sistema.	Campo Essencial Laboratório responsável pela liberação do teste resultante do teste diagnóstico (RT-PCR) da amostra do paciente.	LAB_PCR OU CO_LAB_PCR
68- Tipo de Amostra Sorológica para SARS-CoV-2	Number(3)	1-Sangue/plasma/soro 2-Outra, qual? 9-Ignorado	Campo Essencial Tipo de amostra sorológica que foi coletada.	TP_AM_SOR
Tipo de Amostra Sorológica para SARS-CoV-2/Outra, qual?			Campo Essencial Descrição do tipo da amostra clínica, caso o diferente das listadas na categoriaum (1) do campo.	SOR_OUT
69- Data da coleta		Data DD/MM/AAAAA	Campo Essencial Data da coleta do material para diagnóstico por Sorologia.	DT_CO_SOR
70- Tipo de Sorologia para SARS-CoV-2	Number(3)	1-Teste rápido 2-Elisa 3-Quimiluminiscência 4-Outro, qual	Campo Essencial Tipo do Teste Sorológico que foi realizado	TP_SOR

70- Tipo de Sorologia para SARS-Cov-2	Varchar(100)		Descrição do tipo de Teste Sorológico Campo habilitado se selecionado categoria 4- Outro, qual?	OUT_SOR
70- Tipo de Sorologia para SARS-Cov-2/Outro, qual?	Varchar(100)	Outro tipo de amostra Sorológica		SOR_OUT
70- Resultado do Teste Sorológico para SARS-CoV-2:	Varchar(2(1))	1-Positivo 2-Negativo 3-Inconclusivo 4-Não realizado 5-Aguarda resultado 9-ignorado	Resultado da Sorologia para SARS-CoV-2	RES_SOR
70- Resultado do Teste Sorológico para SARS-CoV-2:	Varchar(2(1))	IgG	Resultado da Sorologia para SARS-CoV-2	RES_IgG
70- Resultado do Teste Sorológico para SARS-CoV-2:	Varchar(2(1))	IgM	Resultado da Sorologia para SARS-CoV-2	RES_IM
70- Resultado do Teste Sorológico para SARS-CoV-2:	Varchar(2(1))	IgA	Resultado da Sorologia para SARS-CoV-2	RES_IgA
71- Data do Resultado	Date DD/MM/AAAA		Data do Resultado do Teste Sorológico Data deve ser maior ou igual a 69- Data da Coleta	DT_RES
72-Classificação final do caso	Varchar(2(1))	1-SRAG por influenza 2-SRAG por outro vírus respiratório 3-SRAG por outro agente etiológico, qual: 4-SRAG não especificado 5-SRAG por COVID-19	Diagnóstico final do caso.	CLASSI_FIN
72-Classificação final do caso	Varchar(2(30))		Campo Obrigatório Se tiver resultados divergentes entre as metodologias laboratoriais, priorizar o resultado do RT-PCR.	
3-SRAG por outra causa , qual:		Descrição de qual outro agente etiológico foi identificado	Campo Obrigatório Se campo 72-Classificação final do caso = 3.	CLASSI_OUT

			Habilitado se campo 72-Classificação final do caso = 3.	
73-Critério de Encerramento	Varchar(2)	1. Laboratorial 2. Clínico Epidemiológico 3. Clínico 4. Clínico Imagem	Indicar qual o critério de confirmação.	CRTERIO
74-Evolução do caso	Varchar(2)	1-Cura 2-Óbito 3- Óbito por outras causas 9-Ignorado	Evolução do caso	EVOLUCAO
75-Data da alta ou óbito	Date DD/MM/AAAA	Data da alta ou óbito	Campo Essencial	DT_EVOLUCA
76-Data do Encerramento	Date DD/MM/AAAA	Data do encerramento do caso.	Campo Obrigatório Se o campo 72-Classificação final do caso estiver preenchido. Data do encerramento deve ser > ou = a data da digitação (atual).	DT_ENCERRA
77- Número D.O		Número da Declaração de Óbito	Campo Essencial Habilitado se o campo 74- Evolução do caso = 2	NU_DO
78-Observações	Varchar(999)	Outras observações sobre o paciente consideradas pertinentes.	Campo Opcional	OBSERVA
79-Profissional de Saúde Responsável	Varchar(60)	Nome completo do profissional de saúde (sem abreviações) responsável pela notificação.	Campo Essencial	NOME_PROF
80-Registro Conselho/Matrícula	Varchar(15)	Número do conselho ou matrícula do	Campo Essencial	REG_PROF

		profissional de saúde responsável pela notificação (Ex: CRM/RJ 1234)	
Data da digitação	Date DD/MM/AAAA	<p>Campo Interno</p> <p>Preenchido automaticamente pelo sistema com a data da digitação da ficha.</p> <p>Não é a data de preenchimento da ficha manualmente e sim a data em que é digitado no sistema.</p> <p>Não é atualizada se houver alterações posteriores de dados.</p>	DT_DIGITA