

COLUNAS

DT_SIN_PRI
SG_UF_NOT
CS_SEXO
NU_IDADE_N
TP_IDADE
CS_GESTANT
CS_RACA
CS_ESCOL_N

FEBRE
TOSSE
GARGANTA
DISPNEIA
DESC_RESP
SATURACAO
DIARREIA
VOMITO
DOR_ABD
FADIGA
PERD_OLFT
PERD_PALA

FATOR_RISC
PUERPORA
CARDIOPATI
HEMATOLOGI
SIND_DOWN
HEPATICA
ASMA
DIABETES
NEUROLOGIC
PNEUMOPATI
IMUNODEPRE
RENAL
OBESIDADE

VACINA_COV
DOSE_1_COV
DOSE_2_COV
DOSE_REF
FAB_COV1
FAB_COV2
FAB_COVREF

HOSPITAL
DT_INTERNA
UTI
SUPORT_VEN

AN_SARS2
PCR_SARS2
CLASSI_FIN
EVOLUCAO
DT_EVOLUCA

Carregamento dos dados:

- 1) Baixar as bases SRAG 2021 e 2022 para diretório local
- 2) Carregar as bases com as colunas especificadas na página anterior
- 3) Unir as bases em um único objeto DataFrame

Mapeamento dos dados:

- 1) Sintomas e fatores de risco (olhar notebook)
1: 1 (positivo)
2: 0 (negativo)
9: np.nan (ausente)
- 2) Se FATOR_RISCO == 2, preencher todas os fatores de risco com 0 (olhar notebook)
- 3) Criar colunas de quantidade de sintomas (NUM_SYM) e fatores de risco (NUM_COM) (olhar notebook)
- 4) Criar coluna 'age' (olhar notebook)
- 5) Criar coluna de 'age group' (olhar notebook)
- 6) Criar coluna 'outcome' baseado na coluna EVOLUCAO
1: 'cure' (sobreviveu)
2: 'death' (óbito)
3: np.nan
9: np.nan
- 7) Criar coluna 'region' baseado na coluna SG_UF_NOT
- 8) Criar a coluna 'race' baseado em CS_RACA (olhar notebook)
- 9) Criar a coluna 'education' baseada em CS_ESCOL_N (olhar notebook)
- 10) Não fazer nenhuma alteração para consertar datas
- 11) Converter datas para datetime (olhar notebook)
São 6 datas
- 12) Cria a coluna 'icu' baseado na coluna UTI
1: 1
2: 0
9: np.nan
- 13) Criar a coluna 'imv' baseado na coluna SUPORT_VEN
1: 1
2: 0
3: 0
9: np.nan
- 14) Criar a coluna 'niv' baseado na coluna SUPORT_VEN
1: 0
2: 1
3: 0
9: np.nan
- 15) Criar a coluna 'disease severity' baseada nos sintomas (olhar notebook)

Filtragem dos dados:

1) Manter apenas pacientes com outcome sendo 'cure' ou 'death'
N = ?

2) Manter apenas pacientes ≥ 18 anos
N = ?

3) Manter apenas pacientes com sexo feminino ou masculino
N = ?

4) Manter apenas pacientes hospitalizados
N = ?

5) Manter apenas pacientes internadas (DT_INTERNA) entre 05/09/2021 e 02/04/2022
N = ?

6) Manter apenas pacientes com COVID-19 (CLASSI_FIN == 5):
N = ?

7) Manter apenas pacientes testados com antígeno ou PCR (PCR_SARS2 == 1 ou AN_SARS2 == 1)
N = ?

8) Manter apenas pacientes com status de vacinação (VACINA_COV sendo 1 ou 2)
N = ?

Mapeamentos dos fabricantes:

1) Os nomes dos fabricantes das vacinas devem ser padronizados em:
CoronaVac, Vexzevria, Comirnaty e Janssen

Mapear os campos FAB_COV1, FAB_COV2 e FAB_COVREF em um desses fabricantes, se a data da dose correspondente (DOSE_1_COV, DOSE_2_COV e DOSE_REF) estiver preenchida e entre os dias 18/01/2021 até o dia de hoje. Caso contrário, colocar np.nan.

2) Verificar se existem pacientes com status vacinado (VACINA_COV == 1) que não tem nenhuma data (DOSE_1_COV, DOSE_2_COV e DOSE_REF) preenchida.

3) Criar uma coluna nova que indique o número de doses de vacina que o paciente tomou, de acordo com as datas (DOSE_1_COV, DOSE_2_COV e DOSE_REF) estarem ou não preenchidas.

4) Criar uma coluna 'main vac' que indica qual a vacina tomada com a maior frequência pelo paciente (aquela que ele tomou mais doses).

Separação das bases:

1) Exportar as bases de dados nos dois períodos
P1: DT_INTERNA entre 05/09/2021 a 18/12/2021
P2: DT_INTERNA entre 19/12/2021 a 02/04/2022