

## Disciplina de Visualização de Dados para a Área de Saúde

Projeto - Análise de condições clínicas respiratórias com base em anotações médicas e áudios de tosse

Brunna Raphaelly Amaral da Silva - 144566

Carolina Vieira Campos - 263081

Gabriel Bianchin de Oliveira - 230217

Taciana Alessandra Gomes Cruz - 107132

## INTRODUÇÃO

Na primeira etapa do trabalho foi explorado o entendimento das features do dataset Virufy COVID-19.

## DATASET

### INSPEÇÃO DOS DADOS

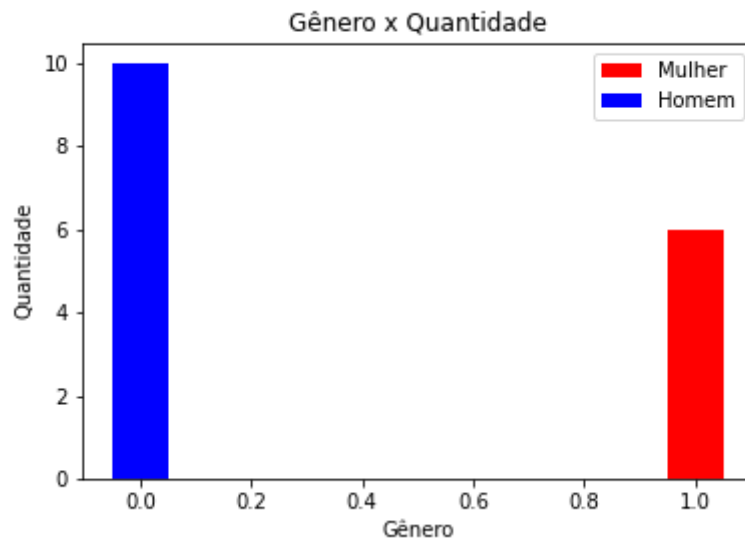
A base de dados contém ao todo 16 amostras. Para cada amostra, existem diversas informações: data da gravação (variável *date*), resultado do exame PCR de COVID-19 (variável *corona\_test*), idade do paciente (variável *age*), gênero (variável *gender*), o histórico médico (variável *medical\_history*, que pode conter diversas condições), se é um paciente fumante (variável *smoker*), os sintomas do paciente (variável *patient\_reported\_symptoms*, que pode conter diversos sintomas) e o arquivo de tosse (variável *cough\_filename*).

	date	corona_test	age	gender	medical_history	smoker	patient_reported_symptoms	cough_filename
0	4/21	0	53	0	none,	1	none,	neg-0421-083-cough-m-53.mp3
1	4/21	1	50	0	Congestive heart failure,	0	Shortness of breath,	pos-0421-084-cough-m-50.mp3
2	4/21	0	43	0	none,	0	Sore throat,	neg-0421-085-cough-m-43.mp3
3	4/21	1	65	0	Asthma or chronic lung disease,	0	Shortness of breath,New or worsening cough,	pos-0421-086-cough-m-65.mp3
4	4/21	1	40	1	none,	0	Sore throat,Loss of taste,Loss of smell,	pos-0421-087-cough-f-40.mp3
5	4/21	0	66	1	Diabetes with complications,	0	none,	neg-0421-088-cough-f-66.mp3
6	4/21	0	20	1	none,	0	none,	neg-0421-089-cough-f-20.mp3
7	4/21	0	17	1	none,	0	Shortness of breath,Sore throat,Body aches,	neg-0421-090-cough-f-17.mp3
8	4/21	0	47	0	none,	0	New or worsening cough,	neg-0421-091-cough-m-47.mp3
9	4/21	1	53	0	none,	0	Fever, chills, or sweating,Shortness of breath...	pos-0421-092-cough-m-53.mp3
10	4/21	1	24	1	none,	0	none,	pos-0421-093-cough-f-24.mp3
11	4/21	1	51	0	Diabetes with complications,	0	Fever, chills, or sweating,New or worsening co...	pos-0421-094-cough-m-51.mp3
12	4/22	0	53	0	none,	1	none,	neg-0422-095-cough-m-53.mp3
13	4/22	1	31	0	none,	0	Shortness of breath,New or worsening cough,	pos-0422-096-cough-m-31.mp3
14	4/22	0	37	0	none,	0	none,	neg-0422-097-cough-m-37.mp3
15	4/22	0	24	1	none,	0	New or worsening cough,	neg-0422-098-cough-f-24.mp3

Iniciamos a análise da base de dados verificando a quantidade de testes positivos e testes negativos para COVID-19. O gráfico a seguir detalha visualmente esta análise. Ao todo, a base possui 9 dados com a classificação negativa e 7 dados com a classificação positiva para COVID-19.



Em seguida, analisamos a quantidade de dados por gênero. Ao todo, 10 dos 16 dados são de homens e o restante são de mulheres. O gráfico abaixo apresenta visualmente o resultado.



Na sequência, verificamos a quantidade de amostras de pacientes do gênero masculino em relação ao teste de COVID-19. Existem 5 amostras com resultado positivo e 5 com resultado negativo na base de dados.



Assim como os pacientes do gênero masculino, analisamos os dados do gênero feminino. Constatamos 4 pacientes do gênero feminino com teste negativo e 2 com teste positivo.



Ao final, transformamos as variáveis *medical\_history* e *patient\_reported\_symptoms* no formato *one-hot encoding* e calculamos a correlação entre as variáveis. Em relação a variável *corona\_test*, o maior valor obtido foi igual a 0,49 com a variável *Shortness\_of\_breath* (variável originalmente transformada de *patient\_reported\_symptoms*).

	corona_test	age	gender	smoker	Congestive heart failure	Asthma or chronic lung disease	Diabetes with complications	Shortness of breath	Sore throat	New or worsening cough	Loss of taste	Loss of smell	Body aches	Fever, chills, or sweating
corona_test	1.00	0.16	-0.16	-0.33	0.29	0.29	0.05	0.49	0.22	0.36	0.43	0.43	-0.23	0.43
age	0.16	1.00	-0.54	0.28	0.14	0.40	0.42	0.05	-0.06	0.16	0.11	0.11	-0.44	0.25
gender	-0.16	-0.54	1.00	-0.29	-0.20	-0.20	0.10	-0.24	0.03	-0.33	0.10	0.10	0.33	-0.29
smoker	-0.33	0.28	-0.29	1.00	-0.10	-0.10	-0.14	-0.25	-0.25	-0.29	-0.14	-0.14	-0.10	-0.14
Congestive heart failure	0.29	0.14	-0.20	-0.10	1.00	-0.07	-0.10	0.38	-0.17	-0.20	-0.10	-0.10	-0.07	-0.10
Asthma or chronic lung disease	0.29	0.40	-0.20	-0.10	-0.07	1.00	-0.10	0.38	-0.17	0.33	-0.10	-0.10	-0.07	-0.10
Diabetes with complications	0.05	0.42	0.10	-0.14	-0.10	-0.10	1.00	-0.25	0.15	0.10	-0.14	-0.14	-0.10	0.43
Shortness of breath	0.49	0.05	-0.24	-0.25	0.38	0.38	-0.25	1.00	0.13	0.31	0.15	0.15	0.38	0.15
Sore throat	0.22	-0.06	0.03	-0.25	-0.17	-0.17	0.15	0.13	1.00	0.03	0.56	0.56	0.38	0.56
New or worsening cough	0.36	0.16	-0.33	-0.29	-0.20	0.33	0.10	0.31	0.03	1.00	0.10	0.10	-0.20	0.49
Loss of taste	0.43	0.11	0.10	-0.14	-0.10	-0.10	-0.14	0.15	0.56	0.10	1.00	1.00	-0.10	0.43
Loss of smell	0.43	0.11	0.10	-0.14	-0.10	-0.10	-0.14	0.15	0.56	0.10	1.00	1.00	-0.10	0.43
Body aches	-0.23	-0.44	0.33	-0.10	-0.07	-0.07	-0.10	0.38	0.38	-0.20	-0.10	-0.10	1.00	-0.10
Fever, chills, or sweating	0.43	0.25	-0.29	-0.14	-0.10	-0.10	0.43	0.15	0.56	0.49	0.43	0.43	-0.10	1.00