

# PYSPARK PROJECT: DISTRIBUIÇÃO DE CASOS DE VÍRUS DE MOSQUITO

# DATASET

- São 9 arquivos CSV, 3 arquivos para cada vírus, cada arquivo pode ser identificado pelo ano da coleta e o vírus. Os anos que trabalhei com os registros são 2019, 2020 e 2021.
- Disposição dos dados:

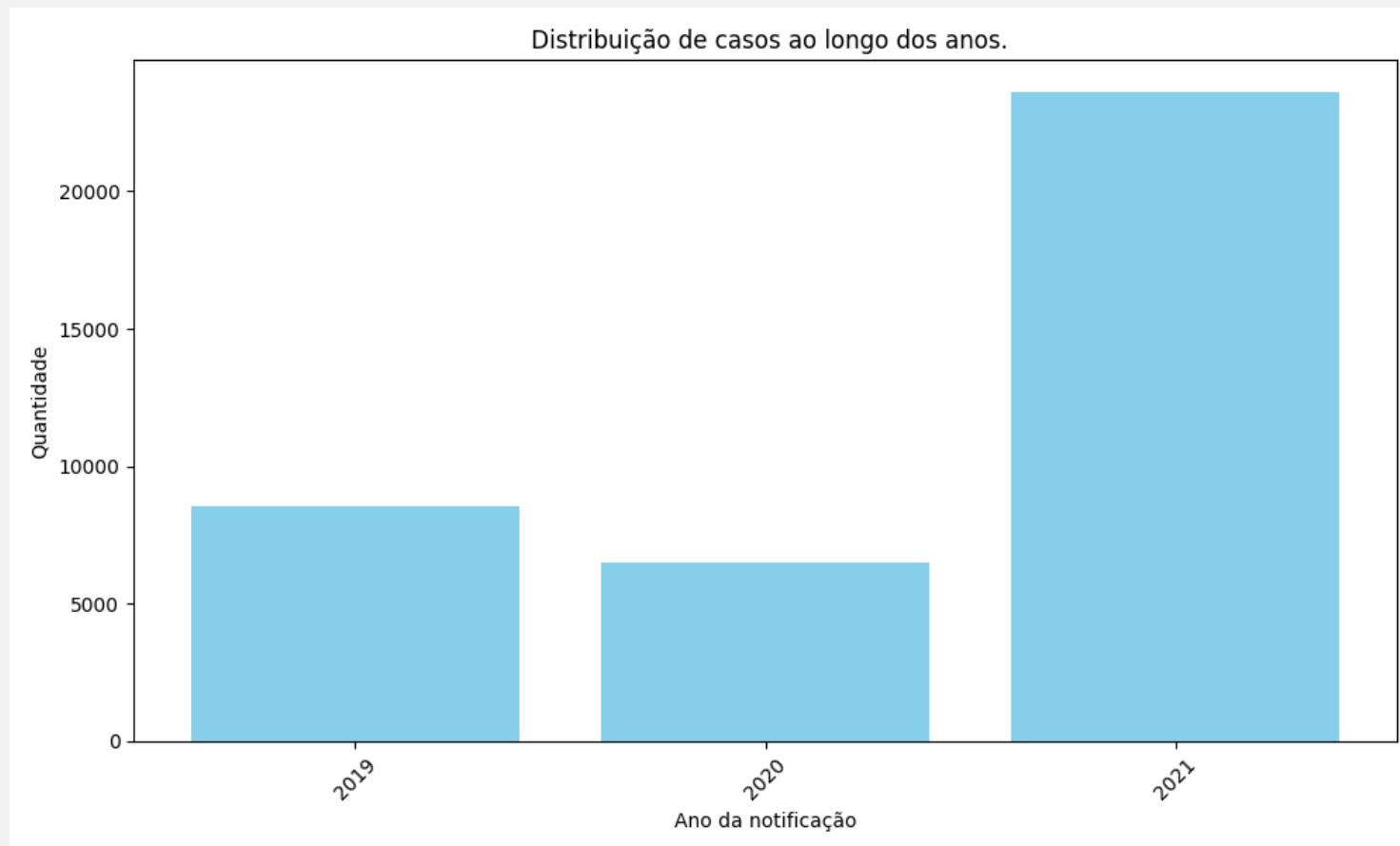
dt_notificacao	notificacao_ano	dt_nascimento	tp_sexo	no_bairro_residencia
2019-01-04	2019	2006-10-01	F	SANTO AMARO
2019-05-21	2019	2017-11-01	F	BOA VIAGEM
2019-01-09	2019	2012-03-01	M	IPSEP
2019-01-03	2019	2014-02-01	F	JORDAO
2019-01-02	2019	2018-05-01	M	IBURA

- Fonte: <http://dados.recife.pe.gov.br/it/dataset/casos-de-dengue-zika-e-chikungunya>

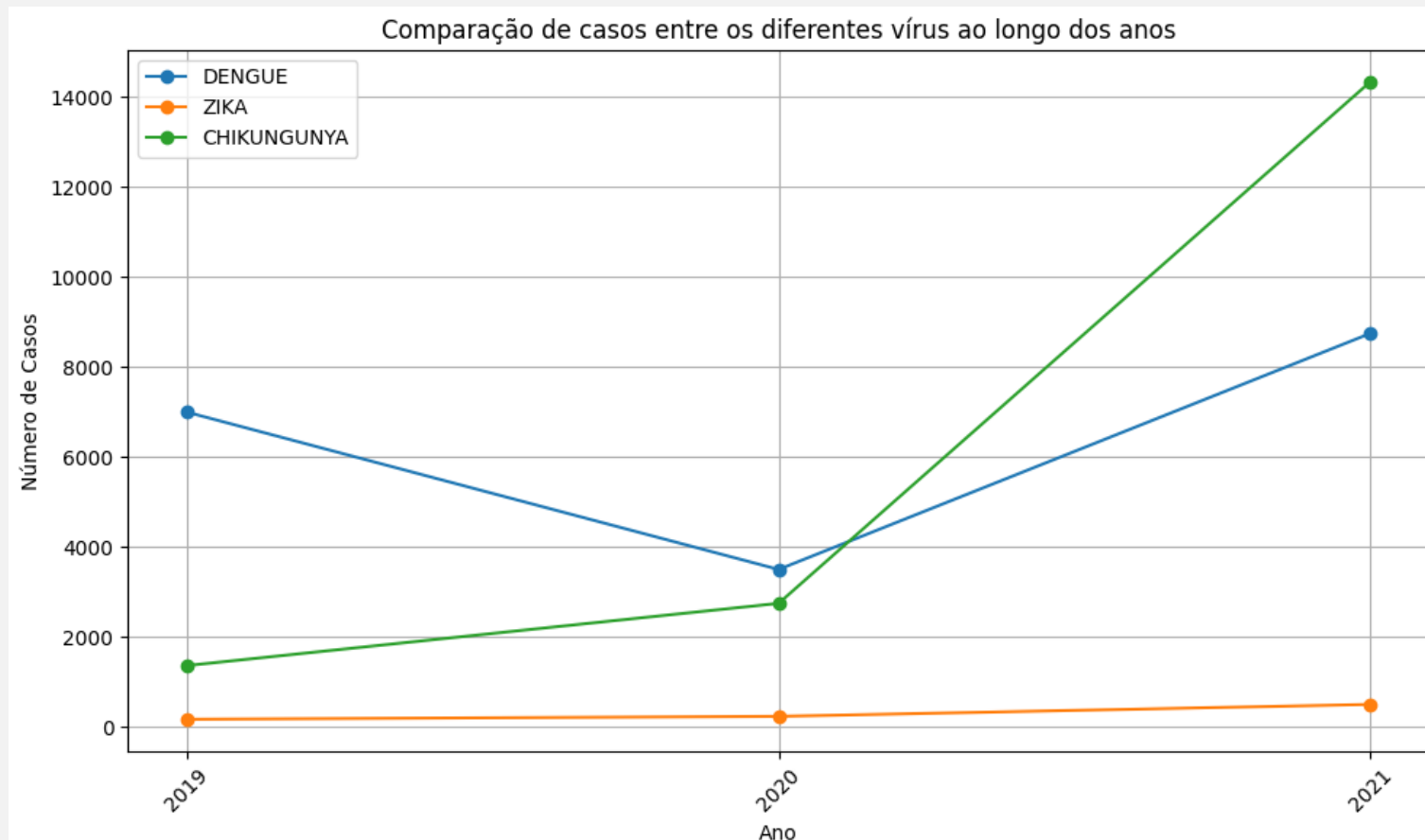
## DETALHES DOS DADOS

- 6 Arquivos CSV
- 126 Colunas
- 38626 Linhas (concatenando os 6 arquivos)
- 1726 Registros nulos
- 3 Vírus (Dengue, Chikungunya, Zika)

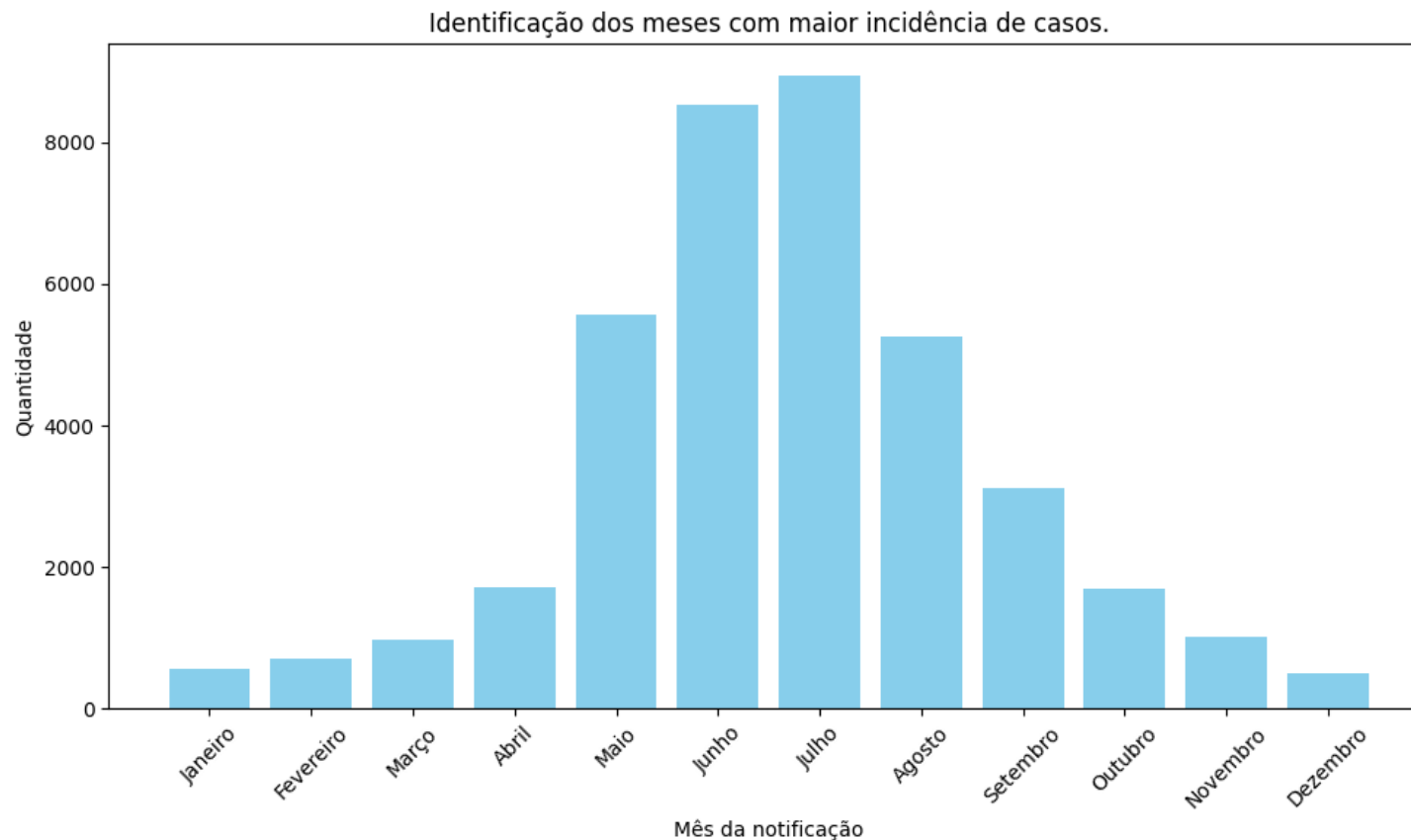
# DISTRIBUIÇÃO DE CASOS AO LONGO DOS ANOS



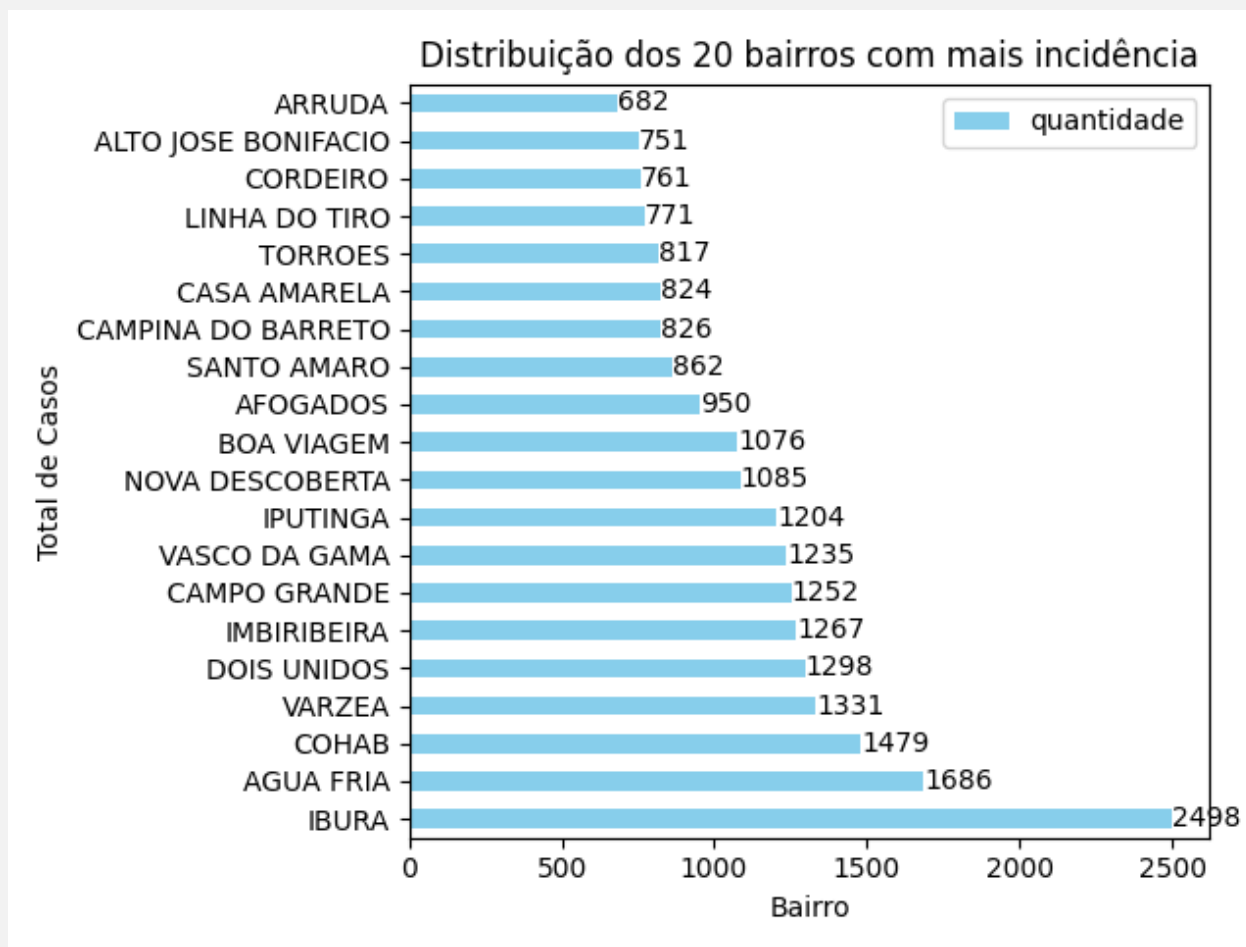
# COMPARAÇÃO DE CASOS ENTRE OS DIFERENTES VÍRUS AO LONGO DOS ANOS



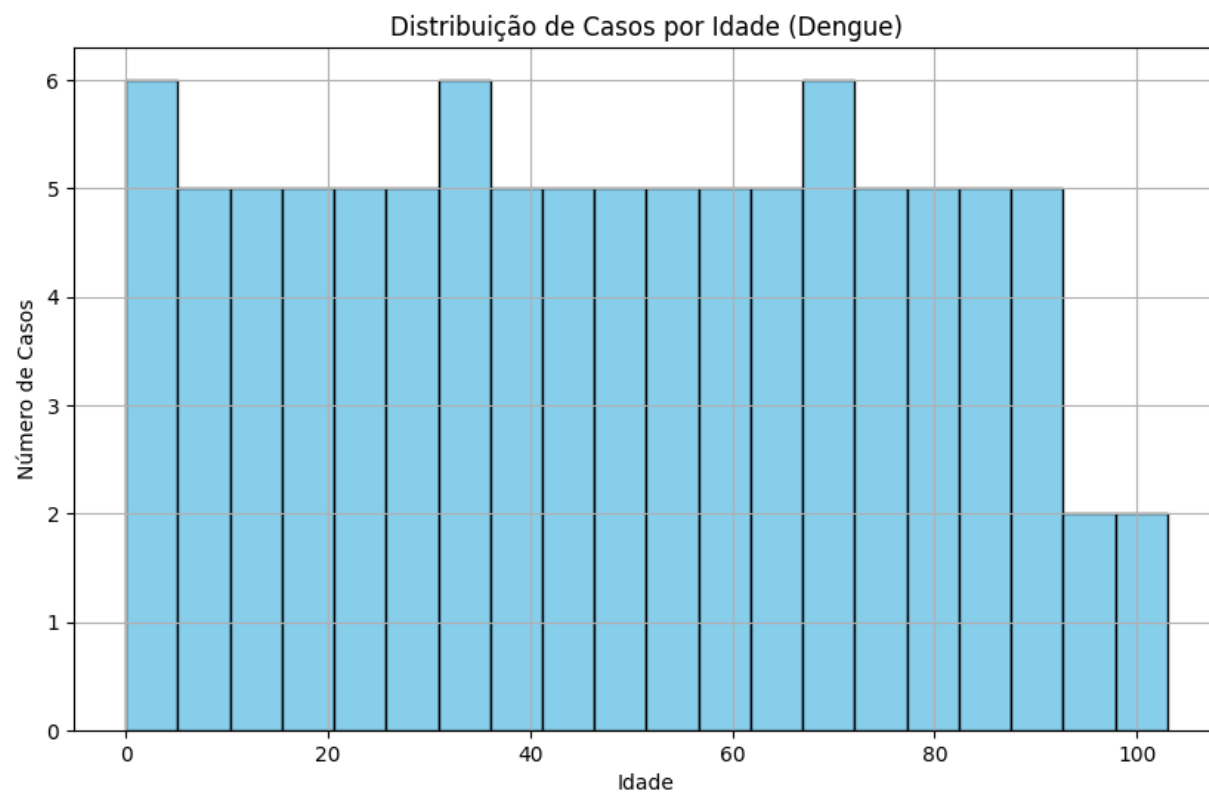
# IDENTIFICAÇÃO DOS MESES COM MAIOR INCIDÊNCIA DE CASOS



# DISTRIBUIÇÃO DE CASOS AGRUPADOS PELO BAIRRO

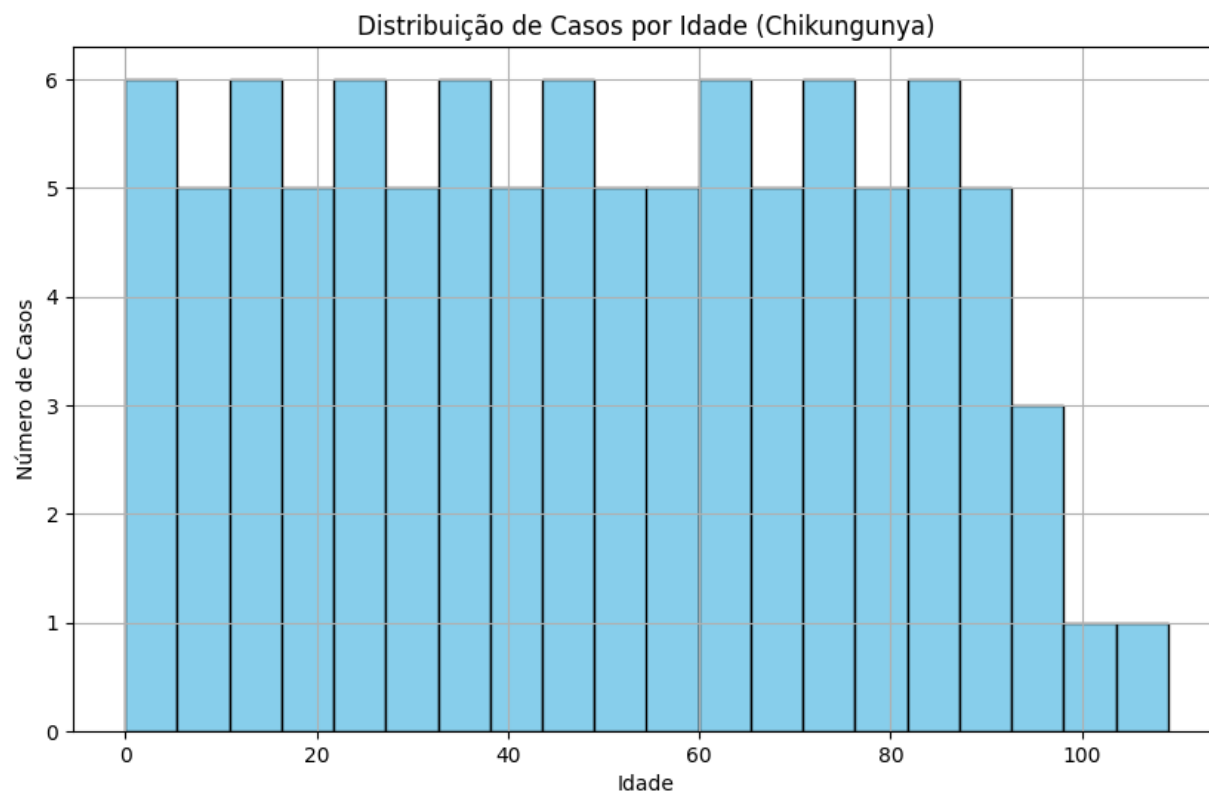


# DISTRIBUIÇÃO DE CASOS POR FAIXA ETÁRIA

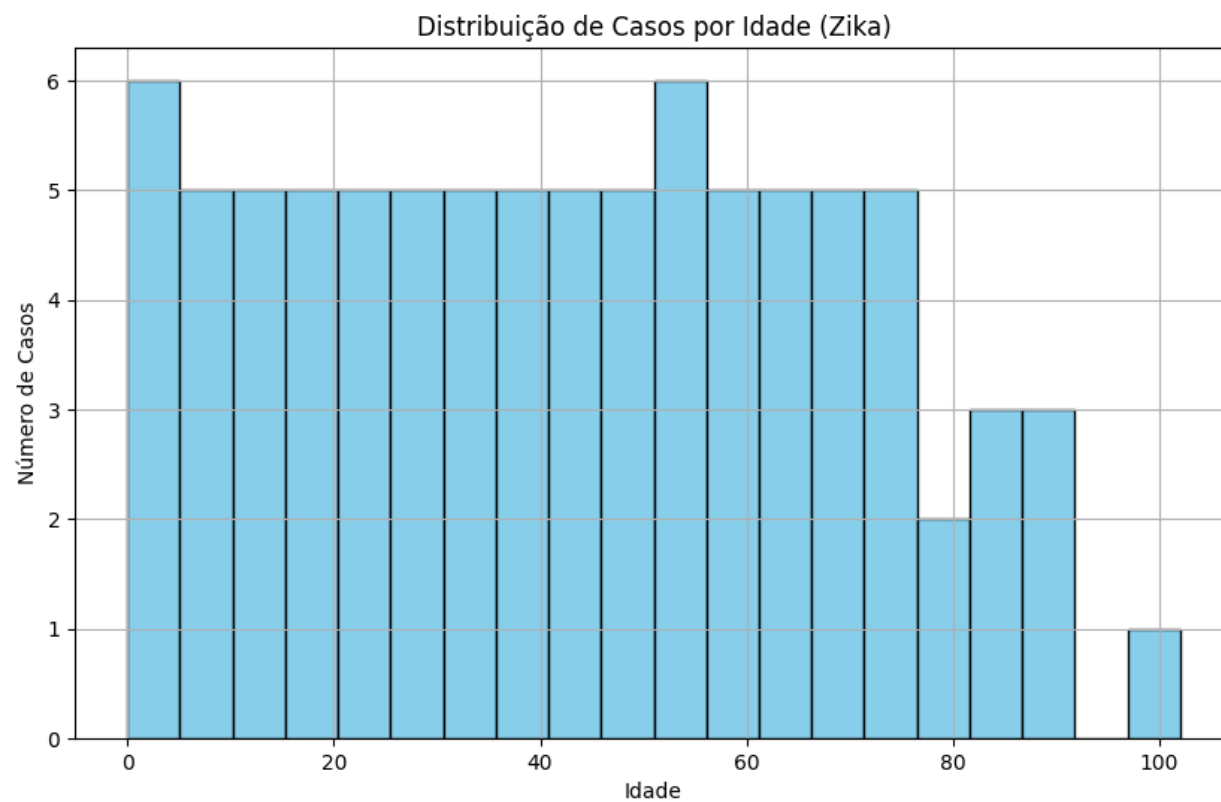




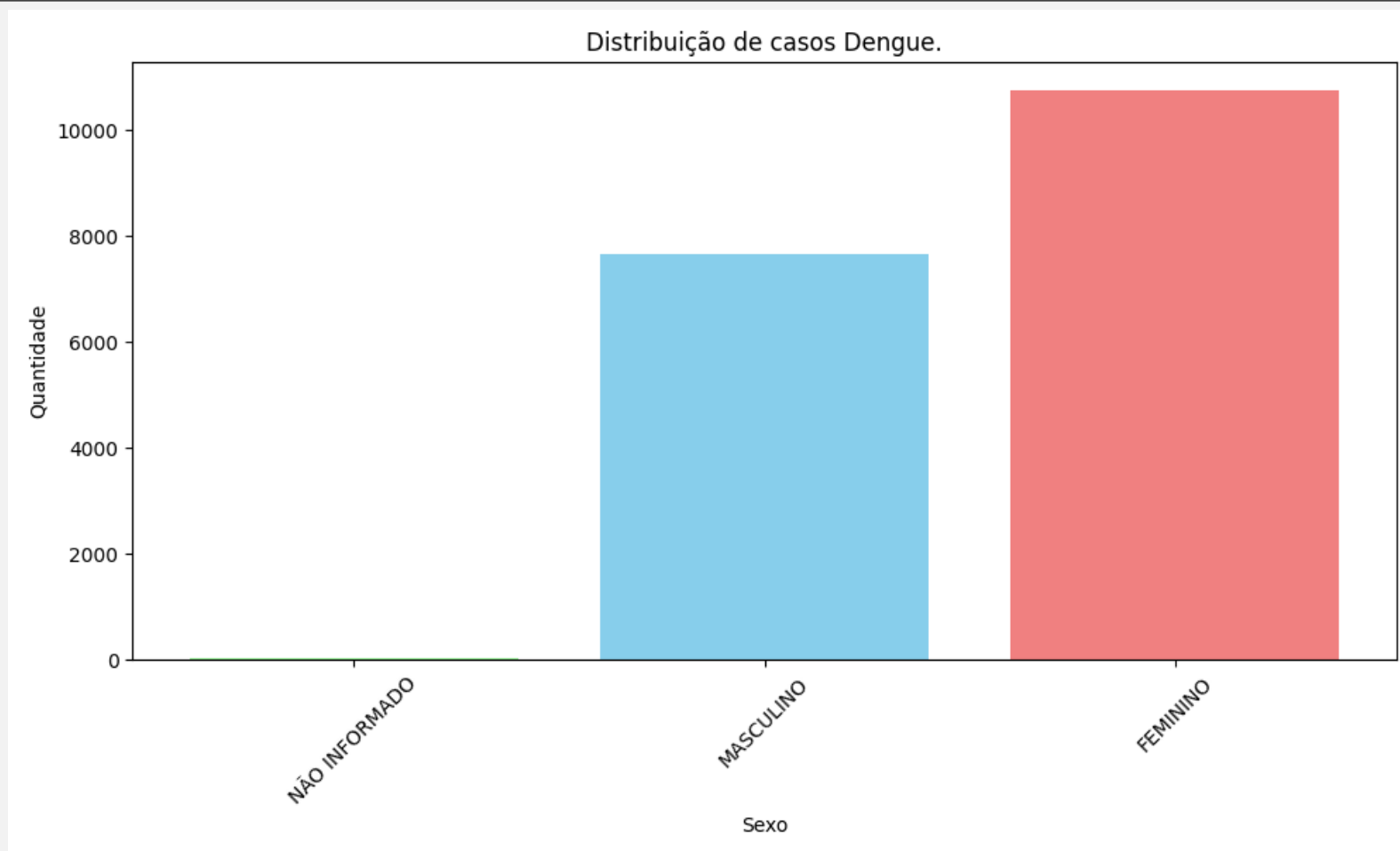
# DISTRIBUIÇÃO DE CASOS POR FAIXA ETÁRIA



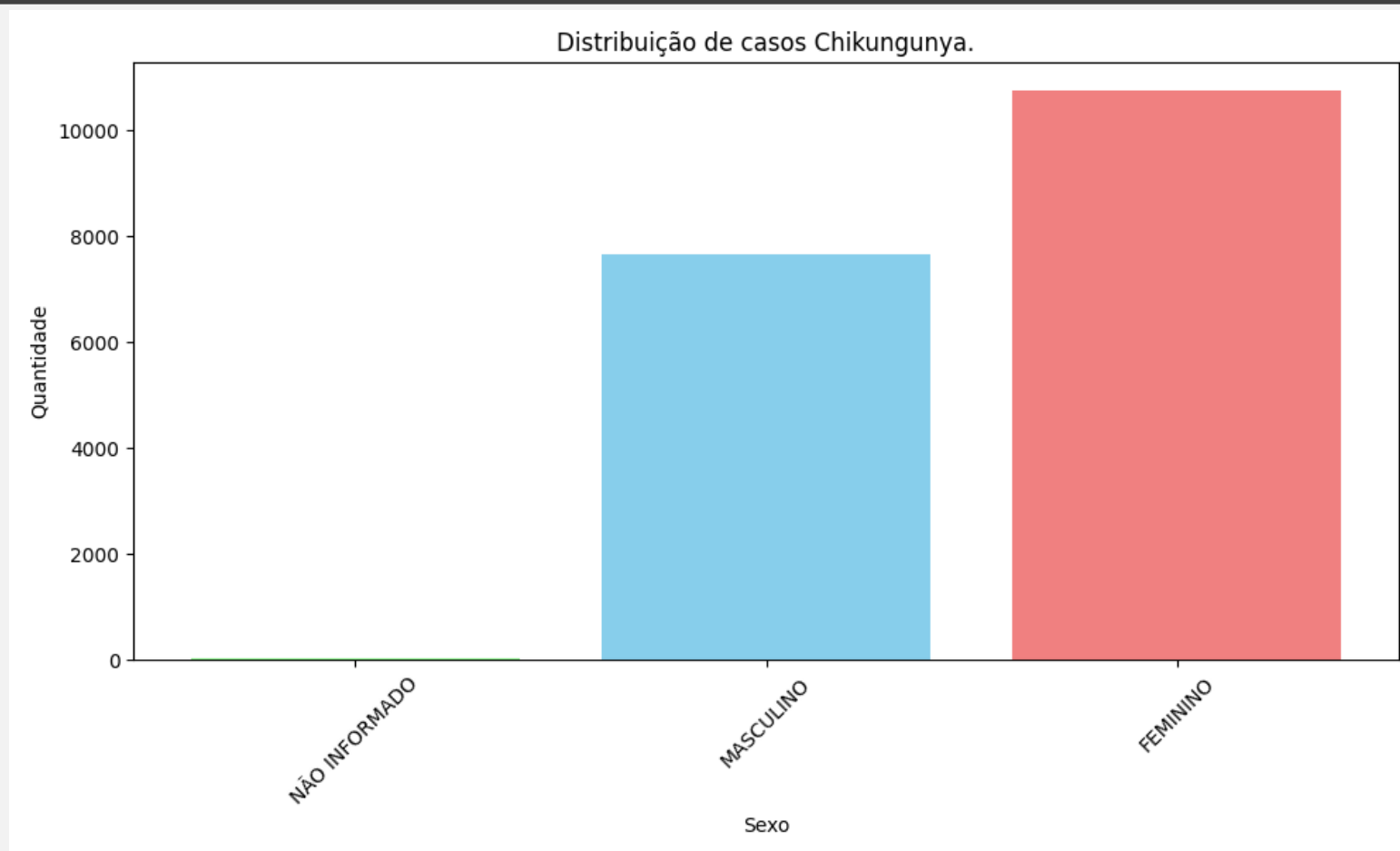
# DISTRIBUIÇÃO DE CASOS POR FAIXA ETÁRIA



# DISTRIBUIÇÃO DE CASOS POR SEXO E VÍRUS



# DISTRIBUIÇÃO DE CASOS POR SEXO E VÍRUS



# DISTRIBUIÇÃO DE CASOS POR SEXO E VÍRUS

