

COVID-19: Integração de Dados Internacionais para Análise da Pandemia

MC356 - Projeto Final

Renan Hiroki Bastos - 176573

Vinicius Alves Mancine Dantas - 188092

Fernando dos Reis Santos Filho - 234471

1 - Introdução

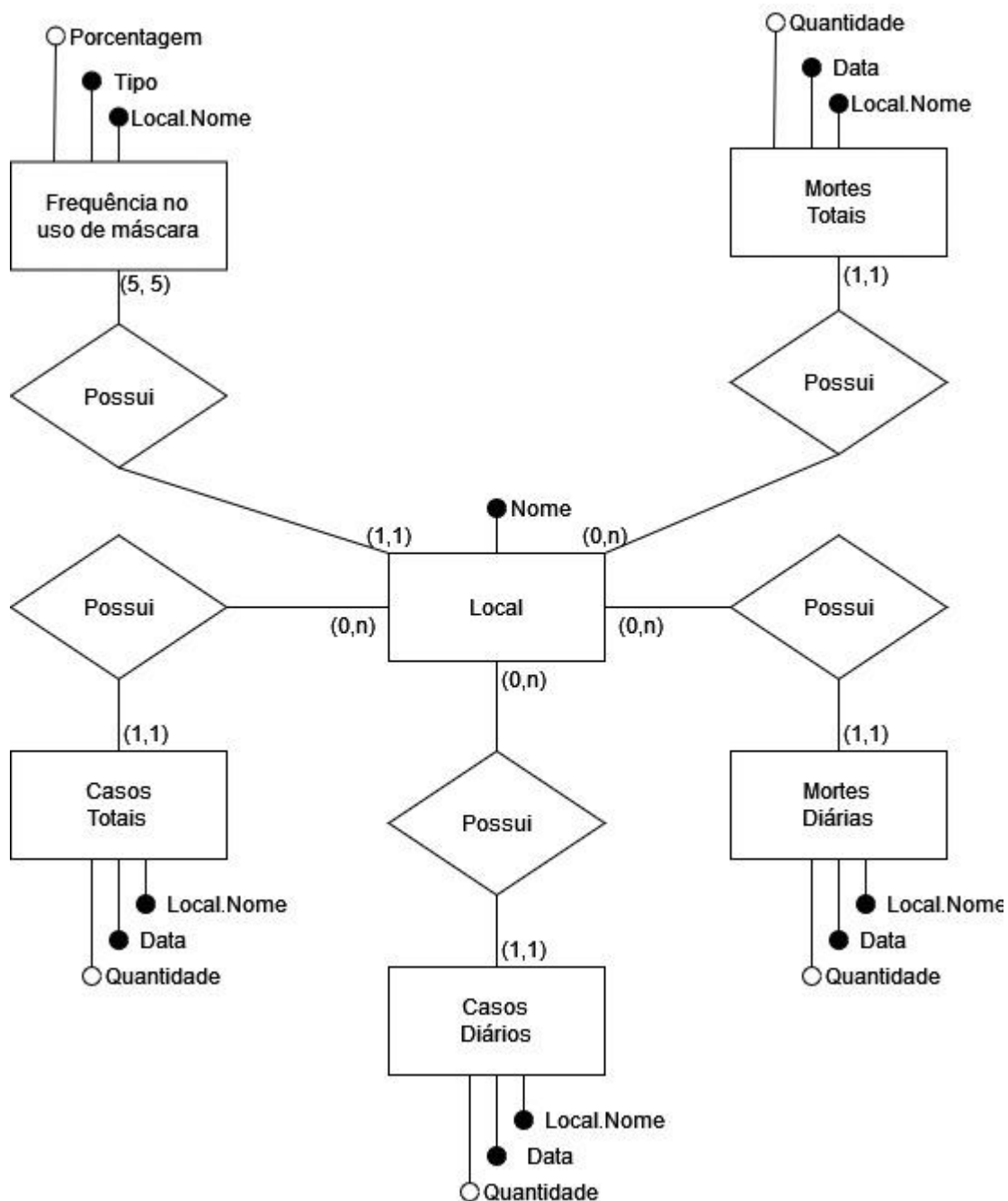
O final do ano de 2019 foi marcado pelo surgimento do vírus Sars-CoV-2, responsável pela doença Covid-19 e por uma intensa pandemia com consequências socioeconômicas e sanitárias gravíssimas, graças à sua alta taxa de transmissão. Hoje, quase dois anos depois, grandes bases de dados foram e ainda são construídas ao redor do mundo catalogando os milhões de casos e mortes confirmados por Covid-19.

Durante todo o período de pandemia, recomendações sobre lavar as mãos, evitar aglomeração e utilizar máscara foram fortemente disseminadas. Porém, ao redor do mundo tivemos inúmeros casos de resistência ao uso de máscara, atitude que contribuiu com a alta taxa de transmissão do vírus. Com isso, algumas dúvidas surgiram a respeito de quanto o uso da máscara influencia na disseminação do vírus e se lugares onde o uso de máscara não foi frequente apresentam maiores números de contaminados e mortos por Covid-19.

Logo, nosso objetivo com esse trabalho é extrair, tratar e compilar dados de fontes distintas relacionados à Covid-19 e a sua prevenção através do uso de máscaras em um único dataset público. Assim, será possível fazer análises sobre a pandemia, a adoção e a eficácia dos métodos de prevenção em um contexto global através de um único dataset.

2 - Data Set Design

Antes de qualquer integração de dados em um dataset, é necessário criar um modelo conceitual. Este modelo é responsável por apresentar ao usuário do banco como os dados se relacionam entre si, independentemente da sua implementação lógica. O modelo conceitual a seguir cumpre o papel de relacionar os casos de infecção e mortes por Covid-19 à frequência da adoção do uso das máscaras em seus respectivos países e/ou estados.



Após o modelo conceitual será feito um modelo lógico. Este tem como objetivo estruturar os dados de forma lógica, mantendo as relações estabelecidas no modelo conceitual, a fim de armazená-los em um banco de dados. O dataset explora as vantagens e desvantagens dos modelos relacionais e em grafos.

O modelo relacional possui a vantagem de ser o modelo mais comum e acessível, porém sua representação em tabelas é pouco intuitiva, dificultando, assim, a análise dos dados. Já o modelo em grafos possui uma representação mais visual e intuitiva para o usuário, facilitando a análise dos dados; por outro lado, trata-se de um modelo menos comum e acessível que o relacional.

A construção desses modelos e do dataset em si envolvem a integração dos dados em formato csv (Comma Separated Value) de fontes distintas, tratamento e transformação dos dados a fim de padronizar a representação dos dados nos modelos já apresentados.

3 - Análises

Concluindo o dataset será possível utilizar os dados disponíveis para realizar análises sobre a pandemia. Seguem perguntas a serem respondidas pelo dataset:

- Qual é a relação entre a frequência do uso de máscaras e a quantidade de casos por região?
- Há relação entre a frequência de uso de máscaras e a taxa de mortalidade?
- A taxa de infecção local influencia na quantidade de pessoas que usam máscaras?
- Onde houve o maior e o menor índice de uso de máscaras?
- Há diferenças na relação entre a frequência do uso de máscaras e a taxa de transmissão da Covid-19 entre EUA e Europa?

4 - Fontes de dados

Aqui listamos algumas possíveis fontes de dados que podemos utilizar:

- Coronavirus (Covid-19) Data in the United States
 - Link: <https://github.com/nytimes/covid-19-data>
 - Formato: CSV
- Data on COVID-19 (coronavirus) by Our World in Data
 - Link: <https://github.com/owid/covid-19-data/tree/master/public/data>
 - Formato: CSV, Json
- How often have you worn a face mask outside your home to protect yourself or others from coronavirus (COVID-19)?
 - Link: <https://www.statista.com/statistics/1114375/wearing-a-face-mask-outside-in-european-countries/>
 - Formato: XLS
- Personal measures taken to avoid COVID-19
 - Link: <https://today.yougov.com/topics/international/articles-reports/2020/03/17/personal-measures-taken-avoid-covid-19>
 - Formato: CSV, XLS
- Face Mask Usage by Demographic

- Link:
<https://www.statista.com/chart/22300/face-mask-usage-by-demographic/>
- Formato: Jpeg
- A Detailed Map of Who Is Wearing Masks in the U.S.
 - Link:
<https://www.nytimes.com/interactive/2020/07/17/upshot/coronavirus-face-mask-map.html>
 - Formato: CSV