

Fundação Getúlio Vargas
EMAp - Escola de Matemática Aplicada
Ciência de Dados e Inteligência Artificial
Introdução à Computação

ANALISANDO DADOS DA SAÚDE COM PYSUS

Aluno(a): Vinícius Nascimento

Professor: Flávio Coelho

Rio de Janeiro
2022

1 Introdução

1.1 SINAN

A biblioteca PySUS permite extrair dados do DataSUS, o departamento de informática do SUS. O DataSUS armazena e disponibiliza dados sobre a saúde pública no Brasil. Diversos bancos de dados do DataSUS são possíveis de se acessar com o PySUS. Um destes bancos é o SINAN (Sistema de Informação de Agravos de Notificação), o qual é responsável por armazenar dados sobre doenças de notificação compulsória, ou seja, doenças que devem ser notificadas ao Ministério da Saúde. Ao notificar compulsoriamente, os profissionais de saúde contribuem para a vigilância epidemiológica, que é fundamental para monitorar a ocorrência de doenças e identificar padrões e tendências.

1.2 Agravado escolhido

Optei por trabalhar acerca da **hanseníase** (também conhecida como lepra). Ela é uma doença infecciosa crônica causada pela bactéria *Mycobacterium leprae*. Essa doença afeta a pele, os nervos periféricos, as mucosas do trato respiratório superior e, em alguns casos, outros órgãos do corpo. Os sintomas incluem manchas claras ou vermelhas na pele com diminuição da sensibilidade, dormência e fraqueza nas mãos e nos pés.

A hanseníase é uma das doenças mais antigas da humanidade e tem sido estigmatizada ao longo da história. No Brasil, é considerada uma doença muito rara, com menos de 15 mil ocorrências por ano.

O código CID10 (classificação internacional de doenças) para a hanseníase é A30.

2 Metologia

Primeiramente, foi necessário instalar o PySUS: `!pip install pysus`, depois, importei o SINAN da biblioteca PySUS: `from pysus.online_data import SINAN` e para criar um dataset em que eu possa usar quando for analisar meus dados:

```
✓ [107] df = SINAN.download('Hansenias', [2019, 2020])
5s

✓ [108] df19 = pd.read_parquet(df[0])
2s
      df19.to_csv('Hansenias.csv', index=False)
      df19.info()
      print(df19.columns)

      df19['DT_NOTIFIC'] = pd.to_datetime(df19['DT_NOTIFIC'])

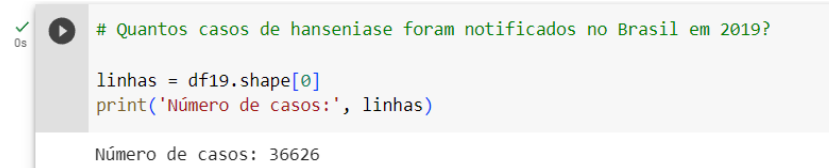
      df19.set_index('DT_NOTIFIC', inplace=True)
      df19
```

A biblioteca fornece funções e métodos convenientes para essa finalidade, facilitando a extração dos dados de forma eficiente. Os dados foram analisados com base em gráficos e tabelas, geradas por meio de comandos que estão disponibilizados no link do meu Google Colab (caso não esteja funcionando o hiperlink, disponibilizei o link no bloco de notas que veio junto com a pasta desse trabalho)

Após carregar os dados extraídos em um dataframe do Pandas, é possível realizar diversas operações de análise e manipulação, como filtragem, agrupamento, criação de gráficos e muito mais. O Pandas permite explorar e extrair insights dos dados de forma eficiente, facilitando a compreensão dos padrões e tendências presentes nos dados do SINAN.

3 Resultados

3.1 Quantos casos de hanseníase foram notificados no Brasil em 2019?




A screenshot of a Jupyter Notebook cell. On the left, there is a green checkmark and the text '0s'. To the right of the checkmark is a play button icon. The main area of the cell contains the following Python code:

```
# Quantos casos de hanseníase foram notificados no Brasil em 2019?  
  
linhas = df19.shape[0]  
print('Número de casos:', linhas)
```

Below the code, the output of the cell is displayed: 'Número de casos: 36626'.

Foram notificados 36.626 casos de hanseníase no Brasil, em 2019. Isso se deve a vários fatores, como fatores socioeconômicos, dificuldades no diagnóstico precoce e no acesso ao tratamento.

3.2 Quantos casos de hanseníase foram notificados no Brasil em 2019, por município?

✓  # Quantos casos de hanseníase foram notificados no Brasil em 2019, por município?

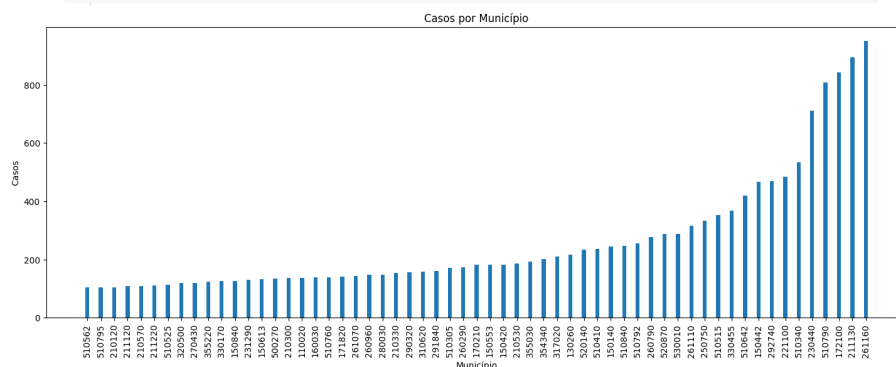
```
df_municipio = df19.groupby('ID_MUNICIP').size()
df_municipio = df_municipio[df_municipio > 100].sort_values()
df_municipio = pd.DataFrame(df_municipio)
df_municipio = df_municipio.rename_axis('Município')
df_municipio = df_municipio.rename(columns={0: 'Casos'})
df_municipio
```

✓ [111] # Estética

```
1s largura_barra = 0.3
n_barras = len(df_municipio.index)
fig, ax = plt.subplots()
ax.bar(df_municipio.index, df_municipio['Casos'], width=largura_barra)
fig.set_size_inches(n_barras * largura_barra, 6)

# Dados
plt.xlabel('Município')
plt.ylabel('Casos')
plt.title('Casos por Município')
plt.xticks(rotation=90)

plt.show()
```



São muitas cidades, o que impossibilitou a criação de um gráfico com todas. portanto, separei as com mais de 100 casos para a criação deste. "Traduzindo" o código dos municípios, o top 5 de cidades com mais casos de hanseníase em 2019 (da com mais caso para a com menos), é: Recife(PE), São Luís(MA), Palmas(TO), Sinop (MT) e Fortaleza(CE).

A hanseníase está associada a condições socioeconômicas desfavoráveis, como pobreza, falta de saneamento básico, moradia precária e acesso limitado a serviços de saúde. Nessas cidades, desigualdades sociais e econômicas são persistentes e podem contribuir para a incidência da doença, especialmente em áreas mais pobres e menos desenvolvidas.

3.3 Quantos casos de hanseníase foram notificados no Brasil em 2019, por estado?

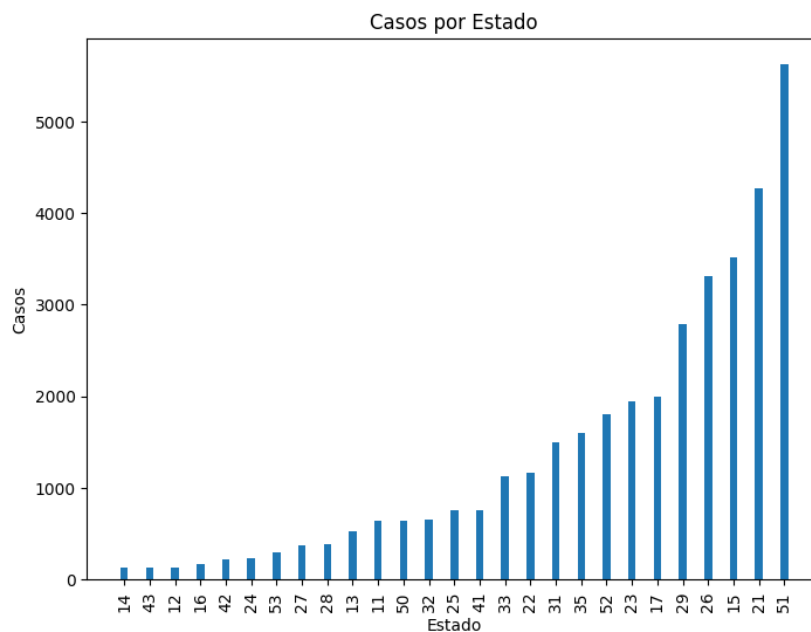
```
# Quantos casos de hanseníase foram notificados no Brasil em 2019, por estado?

df_estado = df19.groupby('SG_UF_NOT').size().sort_values()
df_estado = pd.DataFrame(df_estado)
df_estado = df_estado.rename_axis('Estado')
df_estado = df_estado.rename(columns={0: 'Casos'})
df_estado

[113] # Estética
largura_barra = 0.3
n_barras = len(df_estado.index)
fig, ax = plt.subplots()
ax.bar(df_estado.index, df_estado['Casos'], width=largura_barra)
fig.set_size_inches(n_barras * largura_barra, 6)

# Dados
plt.xlabel('Estado')
plt.ylabel('Casos')
plt.title('Casos por Estado')
plt.xticks(rotation=90)

plt.show()
```



”Traduzindo” o código dos estados, o top 5 de estados com mais casos de hanseníase em 2019 (da com mais caso para a com menos), é: Mato Grosso, Maranhão, Pará, Pernambuco e Bahia.

Provavelmente, pelo mesmo motivo explicitado no item 3.2.

3.4 Quantos casos de hanseníase foram notificados no Brasil em 2019, por mês?

```
[114] # Quantos casos de hanseníase foram notificados no Brasil em 2019, por mês?
```

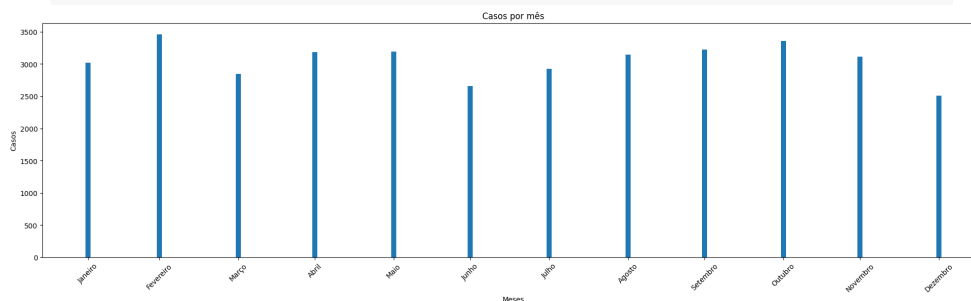
```
df19meses = df19.resample('M').count()
df19meses = pd.DataFrame(df19meses)

# Estética
largura_barra = 2
n_barras = len(df19meses.index)
fig, ax = plt.subplots()
ax.bar(df19meses.index, df19meses['TP_NOT'], width=largura_barra)
fig.set_size_inches(n_barras * largura_barra, 6)

# Dados
plt.xlabel('Meses')
plt.ylabel('Casos')
plt.title('Casos por mês')
plt.xticks(rotation=90)

meses = ['Janeiro', 'Fevereiro', 'Março', 'Abril', 'Maio', 'Junho',
         'Julho', 'Agosto', 'Setembro', 'Outubro', 'Novembro', 'Dezembro']
plt.xticks(df19meses.index, meses, rotation=45)

plt.show()
```

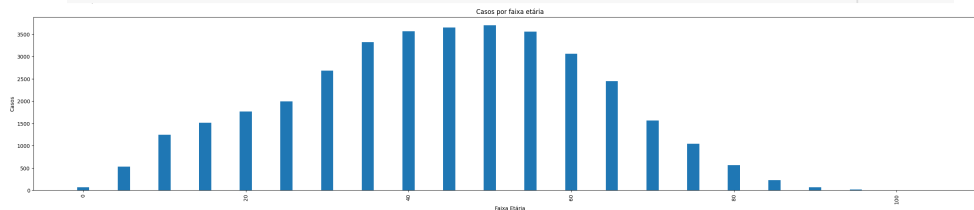


A doença se manteve constante no Brasil durante todo 2019. com um pico em fevereiro, tendo 3.459 casos nesse mês.

A hanseníase não depende do clima para se espalhar e sua transmissão é silenciosa, isso faz com que ela seja constante o ano todo.

3.5 Quantos casos de hanseníase foram notificados no Brasil em 2019, por faixa etária?

```
✓ [115] # Quantos casos de hanseníase foram notificados no Brasil em 2019, por faixa etária?  
0s  
  
df19['Idade'] = decodifica_idade_SINAN(df19.NU_IDADE_N)  
  
def agrupamento(idade):  
    return (idade // 5) * 5  
  
df19['Faixa Etária'] = agrupamento(df19['Idade'])  
  
df_faixa_etaria = df19.groupby('Faixa Etária').size()  
  
# Estética  
largura_barra = 1.5  
n_barras = len(df_faixa_etaria.index)  
fig, ax = plt.subplots()  
ax.bar(df_faixa_etaria.index, df_faixa_etaria.values, width=largura_barra)  
fig.set_size_inches(n_barras * largura_barra, 6)  
  
# Dados  
plt.xlabel('Faixa Etária')  
plt.ylabel('Casos')  
plt.title('Casos por faixa etária')  
plt.xticks(rotation=90)  
  
plt.show()
```



A doença prevalece entre os adultos, com ênfase entre 35 e 60 anos.

Isso se deve devido a vários motivos: período de incubação mais longo (o período de incubação da hanseníase pode ser bastante longo. Assim, os sintomas da hanseníase geralmente começam a se manifestar na idade adulta, uma vez que a infecção teve tempo suficiente para se desenvolver), maior exposição à bactéria e a dificuldade no diagnóstico em crianças.

3.6 Quantos casos de hanseníase foram notificados no Brasil em 2019, por sexo?

```
✓ 0s # Quantos casos de hanseníase foram notificados no Brasil em 2019, por sexo?
```

```
df_sexo = df19.groupby('CS_SEXO').size().sort_values()
df_sexo = pd.DataFrame(df_sexo)
df_sexo = df_sexo.rename_axis('Sexo')
df_sexo = df_sexo.rename(columns={0: 'Casos'})
df_sexo
```

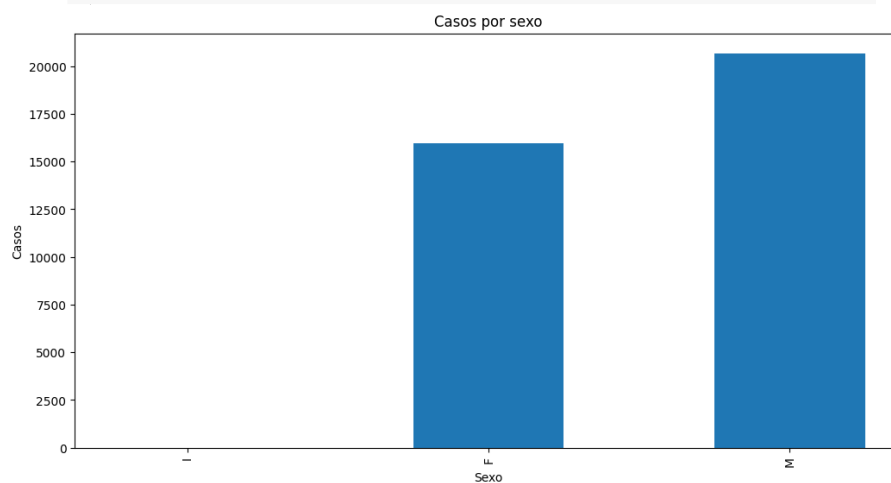
```
✓ [117] 0s # Estética
largura_barra = 0.8 # Reduzindo a largura das barras
n_barras = len(df_sexo.index)
fig, ax = plt.subplots()
ax.bar(df_sexo.index, df_sexo['Casos'], width=largura_barra)

# Espaçamento entre as barras
ax.set_xticks(range(n_barras))
ax.set_xticklabels(df_sexo.index, rotation=90)
ax.margins(x=0.05) # Espaço adicional entre as barras

fig.set_size_inches(n_barras * largura_barra, 6)

# Dados
plt.xlabel('Sexo')
plt.ylabel('Casos')
plt.title('Casos por sexo')

plt.show()
```



O portador da hanseníase é mais propenso a ser do sexo masculino.

De acordo com estudos, isso se deve aos homens serem expostos a mais atividade ocupacionais que envolvam o ambiente externo e diferenças hormonais.

3.7 Quantos casos de hanseníase foram notificados no Brasil em 2019, por raça/cor?

```
✓ [118] # Quantos casos de dengue foram notificados no Brasil em 2019, por raça/cor?

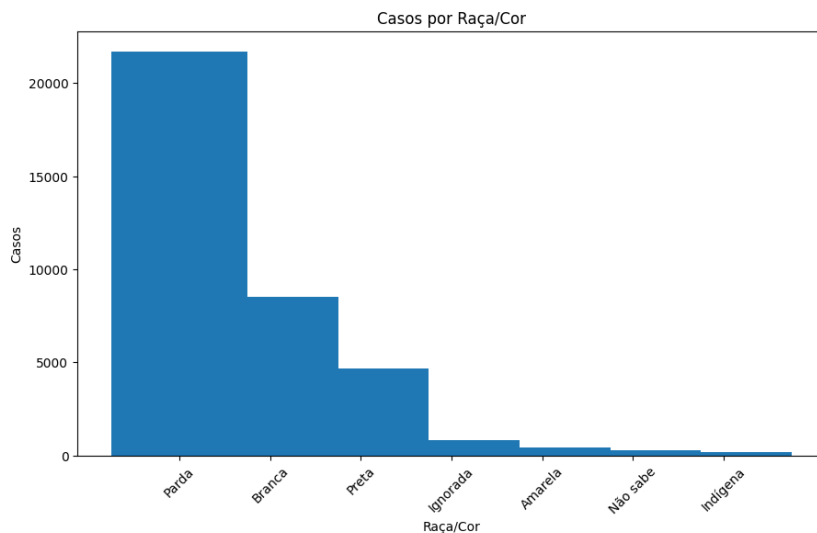
df_cor = df19.groupby('CS_RACA').size().sort_values()
df_cor = pd.DataFrame(df_cor)
df_cor = df_cor.rename_axis('Raça/Cor')
df_cor = df_cor.rename(columns={0: 'Casos'})
df_cor

✓ [126] # Estética
0s
largura_barra = 1.5
n_barras = len(df_cor.index)
fig, ax = plt.subplots()
ax.bar(df_cor.index, df_cor.values.flatten(), width=largura_barra)
fig.set_size_inches(n_barras * largura_barra, 6)

# Dados
plt.xlabel('Raça/Cor')
plt.ylabel('Casos')
plt.title('Casos por Raça/Cor')
plt.xticks(rotation=90)

cores = ['Parda', 'Branca', 'Preta', 'Ignorada', 'Amarela', 'Não sabe', 'Indígena']
plt.xticks(df_cor.index, cores, rotation=45)

plt.show()
```



Pessoas pardas foram mais expostas a doença. Brancos e pretos vêm logo depois, fechando o top 3 do ranking.

Pardos são mais de duas vezes mais atingidos pela hanseníase que os brancos e isso se deve ao fator socioeconômico, uma vez que a desigualdade social atinge, principalmente, pessoas não-brancas, aumentando a probabilidade dessas pessoas serem expostas a vulnerabilidade social.