



EDA Comparação Pré e Pós Pandemia

Link para o notebook:

https://colab.research.google.com/drive/1kFa_gWzuCydSGYw78DyYItIJ52Iopu7a?usp=sharing

1. Visão Geral

1.1. Quantidade de Linhas e Colunas

Após a limpeza e o tratamento dos dados brutos do DATASUS, o *dataframe* final resultou em **1.135.769 entradas**, cada uma composta por **23 variáveis** distintas para análise.

1.2. Colunas do Dataset

O conjunto de dados consolidado para o período pré-pandemia contém **1.135.769 linhas e 23 colunas**, divididas nas seguintes categorias de informações:

- **Geografia e Localização:** UF_ZI (Unidade Federativa), MUNIC_RES (Município de Residência) e MUNIC_MOV (Município de Movimentação/Hospital).
- **Demografia do Paciente:** NASC (Data de Nascimento), IDADE, SEXO e RACA_COR.
- **Temporalidade:** ANO_CMPT (Ano de competência), MES_CMPT (Mês de competência), DT_INTER (Data de Internação) e DT_SAIDA (Data de Saída).
- **Informações Clínicas e Diagnósticos:** DIAG_PRINC (Causa principal), DIAG_SECUN, CID_ASSO (CID Associado), CID_MORTE e ESPEC (Especialidade).
- **Gestão e Internação:** CAR_INT (Caráter da internação), IDENT, COBRANCA, COMPLEX (Complexidade), QT_DIARIAS (Quantidade de diárias) e DIAS_PERM (Dias de permanência).
- **Desfecho:** MORTE (Indicador binário de óbito).

1.3. Tipo dos dados

1.4. Valores nulos



Documentação Análise Exploratória de Dados

Consistência da Base: O dataset final não apresenta valores nulos (missing values) em nenhuma das 23 colunas analisadas.

1.5. Valores Duplicados

Para garantir a fidedignidade dos indicadores, a base de dados passou por um processo de limpeza e desduplicação antes das análises:

Volume Bruto: o conjunto de dados inicial continha **1.135.769 registros**.

Identificação de duplicatas: foram identificados **22.836** registros duplicados (aproximadamente **2,05%** da base total).

Justificativa: essas duplicatas decorrem de processamentos administrativos e correções de faturamento no sistema do SIH-SUS.

Tratamento: foi utilizado uma chave de internação para classificar o evento de uma internação, mostrado a seguir:

```
chave_internacao = [ 'UF_ZI', 'ANO_CMPT', 'MUNIC_RES', 'DT_INTER',
'DT_SAIDA', 'DIAG_PRINC', 'SEXO', 'IDADE', 'MORTE' ]
```

Universo de Estudo Final: após a limpeza, a análise foi conduzida sobre 1.112.933 internações únicas, assegurando que cada evento assistencial fosse contabilizado apenas uma vez.

1.6. Recorte Temporal

O período analisado foi de janeiro de 2014 a dezembro de 2024, composto por todos os períodos da pandemia do COVID-19.

2. Panorama Geral

2.1. Número total de internações

Após o processo de limpeza e desduplicação, o estudo consolidou um universo de **1.112.933 internações únicas** no estado de São Paulo entre 2014 e 2024.

2.2. Média Mensal de Internações

O número médio de internações por mês de cada período analisado é apresentado na tabela a seguir, calculado a partir da contagem mensal de internações únicas.

Período	Média
Pré-Pandemia	9.612.625,00



	0
Pandemia	6.628.666.66 7
Pós-Pandemia	7.591.333.33 3

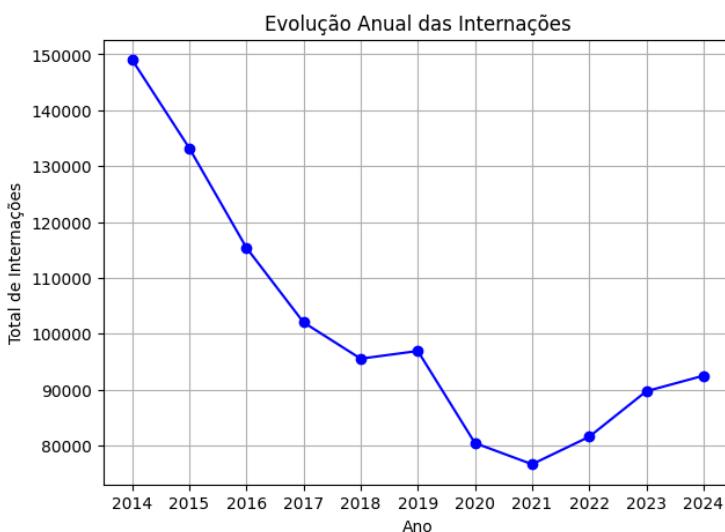
2.3. Quantidade de municípios residente e de internação

Enquanto os pacientes são provenientes de **1.247 municípios** diferentes, a rede hospitalar está concentrada em apenas **333 cidades**. Essa disparidade revela uma centralização dos serviços de saúde mental, obrigando a maioria dos pacientes a se deslocarem de suas cidades de origem para obter assistência hospitalar, evidenciando uma desigualdade no acesso à saúde. Vale analisar fluxo de evasão para identificar polos sobrecarregados.

3. Análise Temporal

3.1. Evolução Anual das Internações

O gráfico mostra queda nas internações por saúde mental de 2014 até 2018, com pequena oscilação em 2019. Em 2020 e 2021, período da pandemia, há uma redução mais acentuada. Isso pode estar ligado às restrições, medo de procurar hospitais e reorganização dos serviços de saúde. A partir de 2022, observa-se retomada gradual das internações. Mesmo assim, os números ainda permanecem abaixo do início da série histórica.



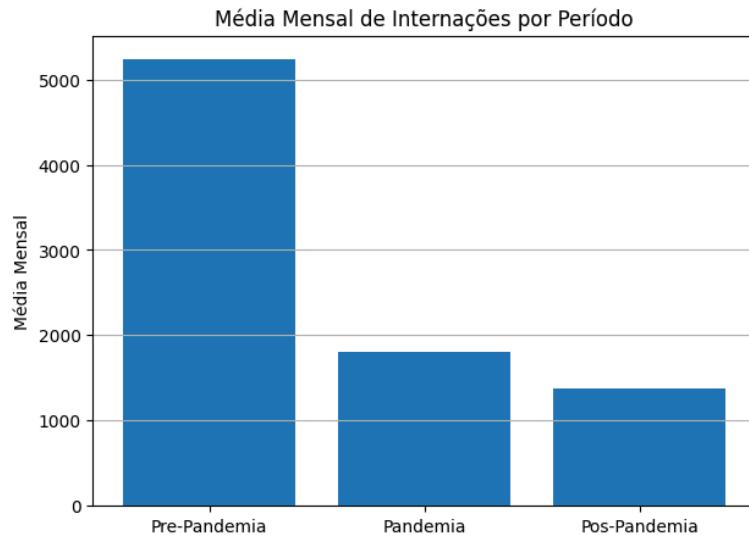
3.2. Média Mensal das Internações Por Período

O gráfico mostra que a média mensal de internações era bem mais alta no período pré-pandemia. Durante a pandemia, há uma queda expressiva na média. No



Documentação Análise Exploratória de Dados

pós-pandemia, observa-se uma leve recuperação em relação ao período pandêmico. Mesmo assim, os valores continuam bem abaixo do cenário anterior à pandemia. Isso reforça o impacto significativo da pandemia na dinâmica das internações em saúde mental.



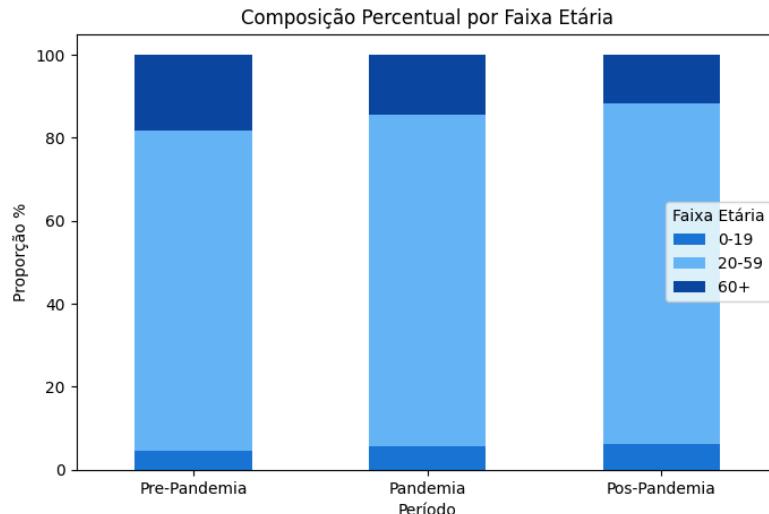
4. Perfil Sociodemográfico

4.1. Distribuição de Idade

Esse gráfico mostra que a faixa de 20–59 anos domina claramente as internações em todos os períodos, representando a maior parte dos casos. Durante a pandemia, há um leve aumento proporcional dos adultos, enquanto a participação de 0–19 e 60+ diminui um pouco. No pós-pandemia, percebe-se um pequeno crescimento na proporção de jovens (0–19) e idosos (60+), mas os adultos continuam concentrando a maioria absoluta. Ou seja, o que muda não é o perfil etário, mas sim o volume total de internações.

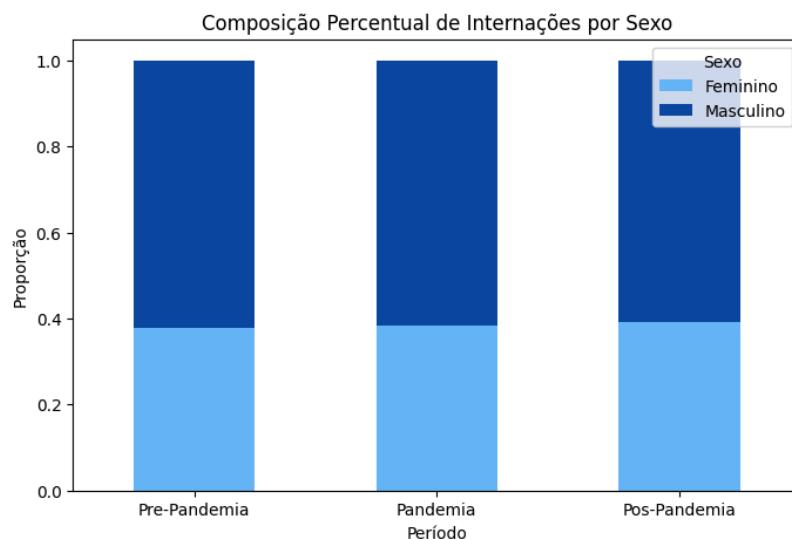


Documentação Análise Exploratória de Dados



4.2. Distribuição de Sexo

Os dados mostram que os homens representam a maioria das internações em todos os períodos, sempre em torno de 60% ou mais. Durante a pandemia, praticamente não há mudança relevante na distribuição por sexo, a diferença se mantém estável. No pós-pandemia, há um leve aumento proporcional das mulheres, mas ainda distante de um equilíbrio. Ou seja, assim como na faixa etária, a pandemia afetou mais o volume total do que o perfil por sexo, o padrão estrutural permaneceu muito parecido.

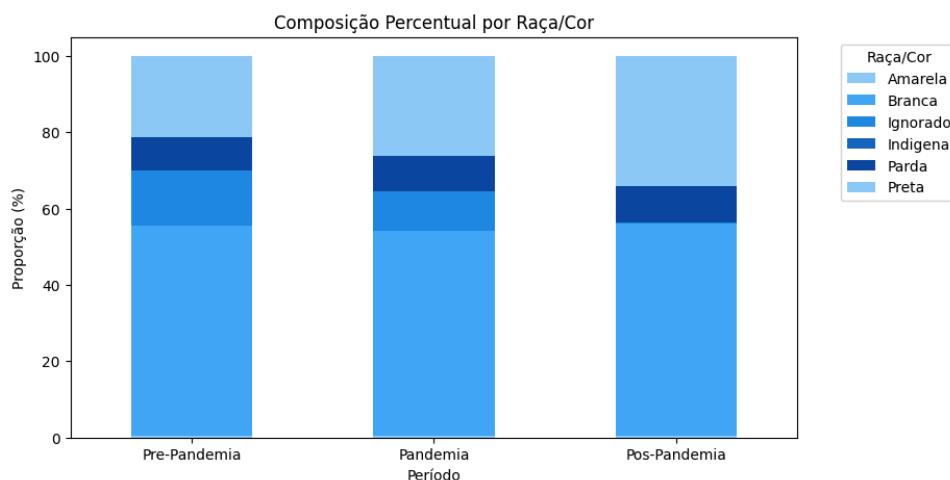


4.3. Distribuição de Raça/Cor



Documentação Análise Exploratória de Dados

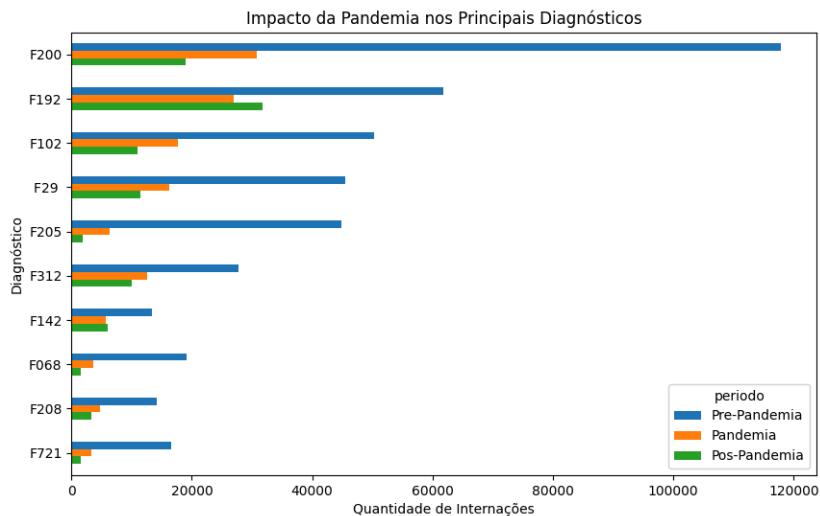
A população **branca** continua sendo a maioria em todos os períodos (cerca de 55%), com pouca variação. Durante a pandemia e principalmente no pós-pandemia, há um **aumento relevante na proporção de pessoas pretas**, passando de ~21% para mais de 34%, isso é uma mudança expressiva. A categoria “Ignorado” cai drasticamente no pós-pandemia, o que pode indicar melhoria no preenchimento dos dados (e isso pode ter influenciado o aumento proporcional em outras categorias). Ou seja, diferente de sexo e idade, aqui há uma alteração mais visível na composição, que merece análise mais crítica antes de tirar conclusões definitivas.



5. Diagnóstico

5.1. Top 10 Diagnóstico

Houve uma queda perceptível no número de internações, o que reforça a ideia de que o volume despencou na pandemia, não foi um CID específico. No pré-pandemia, o F200 (esquizofrenia) e o F192 (uso de múltiplas drogas) lideram com folga. Na pandemia, todos caem muito, principalmente F068, F205 e F721, sugerindo redução de internações eletivas ou maior barreira de acesso. No pós-pandemia, alguns diagnósticos ligados a uso de substâncias (F192, F102, F142) se mantêm relativamente altos comparados a outros, enquanto transtornos como F200 não retomam o patamar anterior. Ou seja, além da queda geral, pode estar havendo uma mudança no perfil clínico das internações, com maior peso relativo dos transtornos relacionados a substâncias no período mais recente.



6. Análise Geográfica

6.1. Fluxo de Origem-Destino

A taxa de evasão permanece **muito estável** entre pré-pandemia e pandemia (cerca de 39%). Isso sugere que, apesar da queda no volume de internações, o padrão de deslocamento entre municípios praticamente não mudou no auge da pandemia. No pós-pandemia, há um leve aumento para ~40,8%, indicando pequena intensificação da dependência externa. Ou seja, a pandemia impactou o volume total, mas não alterou de forma significativa a lógica estrutural da evasão: o sistema já era dependente antes e continuou sendo depois.

6.2. Municípios de Residência vs Internação

Os dados mostram que, mesmo antes da pandemia, cidades como São Paulo, Mogi Guaçu e Piracicaba apresentavam altos fluxos de evasão para municípios menores como Itapira e Jaú. Durante a pandemia, apesar da redução geral de deslocamentos, os mesmos trajetos continuam entre os mais frequentes, indicando uma dependência estrutural de determinados serviços de saúde fora do município de residência. No período pós-pandemia, observa-se uma leve diminuição na quantidade de evasões, mas os padrões se mantêm, sugerindo que o acesso local ainda é insuficiente. Itapira e Espírito Santo do Pinhal aparecem repetidamente como destinos, reforçando seu papel como polos de atendimento. Esses fluxos destacam desigualdades no acesso aos serviços de saúde mental, evidenciando a necessidade de ampliar a cobertura municipal.



6.3. Cidades que mais recebem pacientes de outros municípios

Os dados mostram que Itapira se manteve como o hub mais importante em todos os períodos, recebendo muito mais pacientes que os demais municípios, o que evidencia sua centralidade no atendimento. Outros polos como Presidente Prudente, Espírito Santo do Pinhal e Amparo também se destacam, mas com números significativamente menores. Durante a pandemia, houve uma redução geral no número de pacientes recebidos, refletindo tanto a menor mobilidade quanto possíveis restrições de atendimento. No período pós-pandemia, os fluxos começaram a se recuperar, mas ainda não atingem os níveis pré-pandemia, indicando uma retomada gradual. Esses hubs confirmam a concentração de serviços em alguns municípios, reforçando a necessidade de ampliar a cobertura local para reduzir deslocamentos.

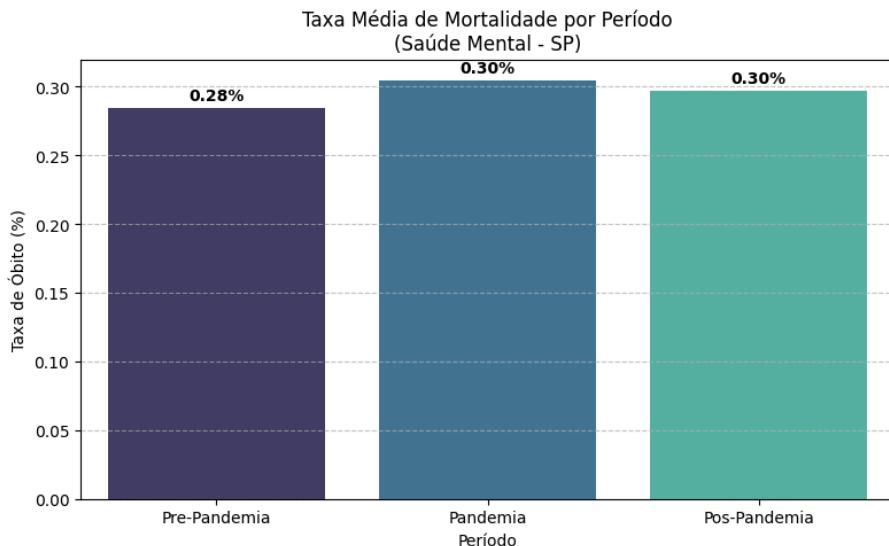
7. Óbito

7.1. Taxa de óbito por ano

O gráfico mostra que a taxa média de mortalidade em saúde mental em São Paulo se manteve relativamente estável ao longo dos períodos analisados. Houve um leve aumento durante a pandemia, passando de 0,28% no pré-pandemia para 0,30%, e essa taxa se manteve praticamente igual no pós-pandemia. Isso indica que, apesar das mudanças no acesso e no fluxo de pacientes durante a pandemia, a mortalidade não sofreu variações significativas. A estabilidade sugere que os serviços de saúde mental conseguiram manter um padrão de atendimento que evitou aumento expressivo no risco de óbito. Contudo, mesmo taxas aparentemente baixas merecem atenção, dada a vulnerabilidade do público atendido.

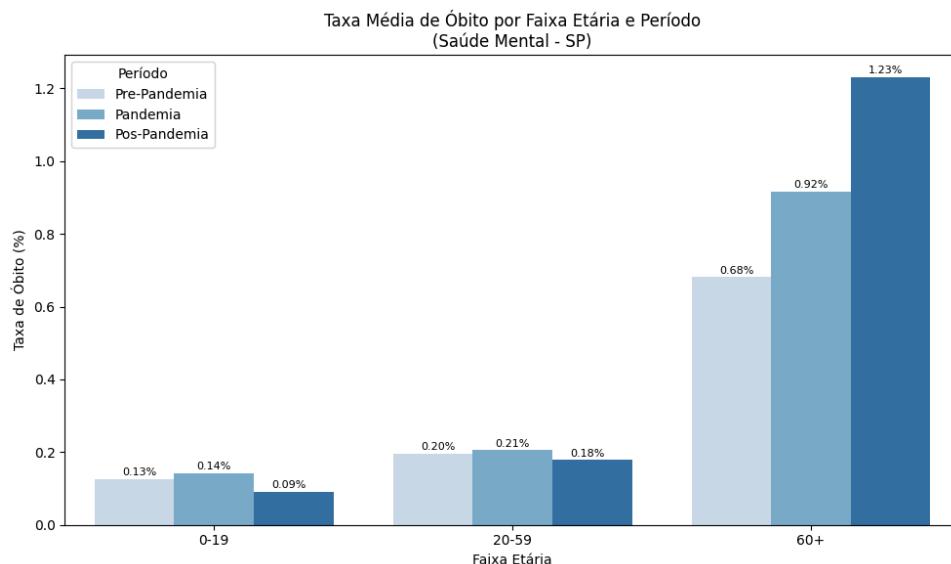


Documentação Análise Exploratória de Dados



7.2. Óbito por idade

O gráfico mostra que a taxa média de óbito em saúde mental aumentou significativamente na faixa etária 60+ durante a pandemia e pós-pandemia, chegando a 1,23%. Nas faixas mais jovens, 0-19 e 20-59, as variações foram pequenas, com ligeira queda ou estabilidade no pós-pandemia. Isso indica que o impacto da pandemia foi mais severo entre os idosos, possivelmente devido à maior vulnerabilidade e menor acesso a cuidados adequados de saúde mental nesse período.



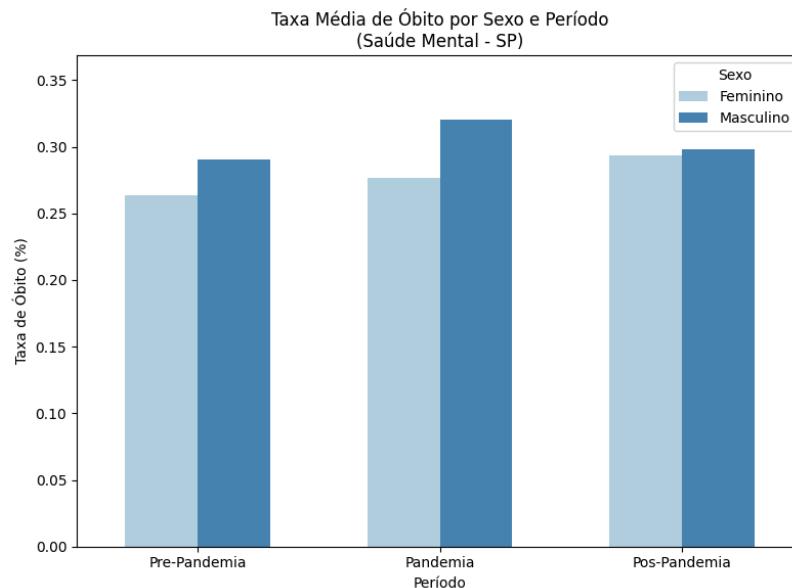
7.3. Óbito por sexo

O gráfico mostra que a taxa média de óbito por sexo aumentou durante a pandemia, com os homens apresentando consistentemente taxas mais altas que as mulheres. No período pré-pandemia, a diferença entre sexos era menor, mas



Documentação Análise Exploratória de Dados

durante a pandemia ela se acentuou. No pós-pandemia, observa-se uma leve redução para os homens, enquanto para as mulheres há um aumento discreto. Isso sugere que a pandemia teve impacto mais significativo nos óbitos masculinos em saúde mental. Em geral, os dados indicam a importância de políticas direcionadas de prevenção, considerando diferenças de sexo.



7.4. Óbito por raça/cor

Observando o gráfico, nota-se que a taxa de óbito variou bastante entre os períodos e raças/cores. A população amarela apresentou aumento expressivo durante a pandemia, mas reduziu no pós-pandemia. Pessoas de raça ignorada tiveram pico no pós-pandemia, atingindo a maior taxa do período analisado. As raças preta e parda mantiveram taxas relativamente estáveis, com leve aumento na pandemia. Esses padrões sugerem desigualdades no impacto da pandemia sobre diferentes grupos raciais.



Documentação Análise Exploratória de Dados

