

# Casamentos homoafetivos

Insights e previsões

# Introdução

Análise realizada por Gabriel Nunes M. Ghirardelli, como parte do processo seletivo para a vaga de Cientista de Dados na empresa Media.Monks.

O objetivo principal deste documento é simular uma entrega de relatório baseado em análise e predição de dados, realizada utilizando-se de habilidades na área de ciência de dados e conhecimento da linguagem Python.

O método utilizado para predição de dados foi a regressão linear, através da biblioteca sklearn. A regressão linear é uma técnica estatística utilizada para estabelecer a relação entre uma variável dependente em função de variáveis independentes. Esta técnica foi escolhida em função da existência de conhecimento prévio em relação à sua utilização, além do entendimento de que as variáveis disponíveis na problemática proposta permitiram a aplicação da técnica.

A predição dos dados e construção deste relatório se deu através das seguintes etapas:

- Consumo dos dados históricos fornecidos:  
Os dados brutos foram carregados em um repositório GitHub para melhor dinâmica de codificação e testagem.;
- Pré processamento dos dados:  
Preparação dos dados de forma a atingirem estrutura adequada para utilização com a técnica selecionado;
- Criação e treinamento dos modelos de regressão linear:  
Estruturação dos modelos de predição e inserção dos dados de treinamento;
- Criação de base de dados a ser predita:  
Processo de criação de base de dados com as combinações de atributos (ano, mês, estado e gênero) referentes aos valores a serem preditos;
- Predição e pós processamento de dados:  
Utilização dos modelos criados para predição dos dados desejado, além de tratamento para formato adequado para posterior análise;
- Criação de visualizações e relatório sobre os dados obtidos:  
Geração de tabelas e gráficos para representação visual dos dados preditos, e disponibilização através de relatório.

Dado a estrutura da técnica selecionada e a interpretação pessoal acerca da problemática, foi criado um modelo de predição para cada estado da base de dados.

Todo o código foi desenvolvido focando a modularidade e encapsulamento de funções e operações.

# Análise histórica

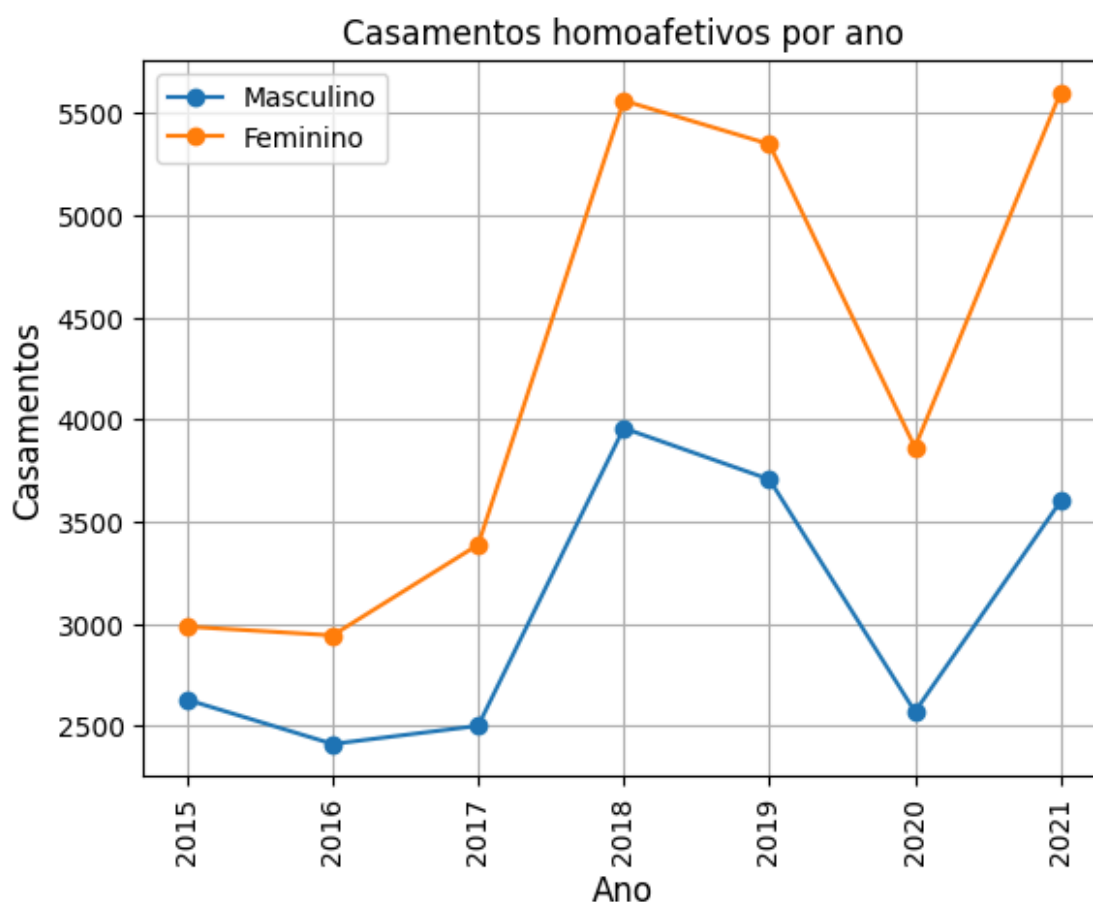
Dado a base de dados fornecida, a qual conta com dados dos anos 2015 a 2021, analisou-se o comportamento do número de casamentos homoafetivos, realizou-se as seguintes análises:

## Variação histórica ano a ano

A análise dos dados segmentados por ano mostram que de 2016 a 2018, o número de casamentos homoafetivos apresentou crescimento após um período de queda de 2015 a 2016.

Pode-se destacar o crescimento acentuado entre os anos 2017 a 2018, seguido por dois anos de decréscimo, o qual atinge seu vale em 2020. Este cenário pode estar relacionado ao cenário de pandemia global, iniciada em 2019 e acentuada em 2020.

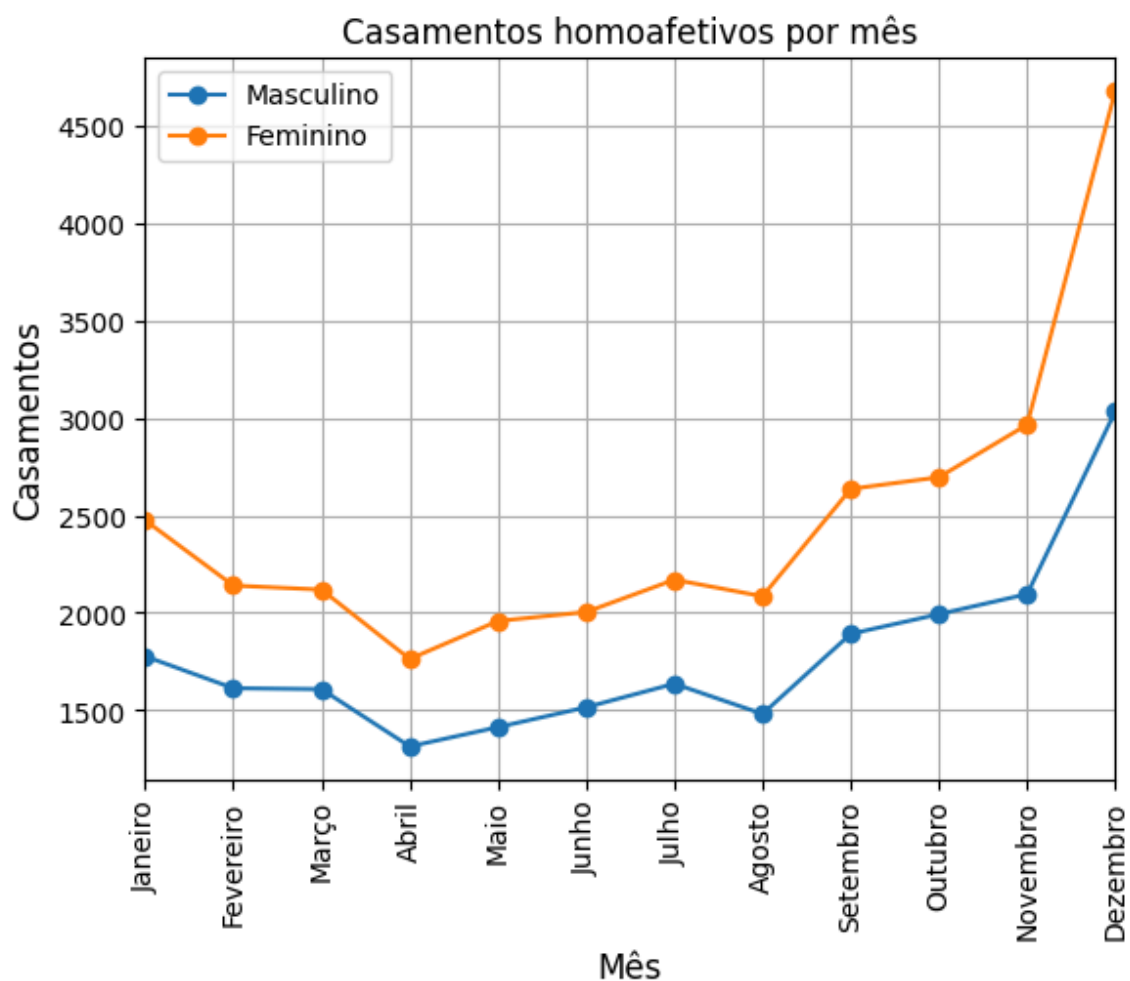
A partir de 2021 o número apresenta uma crescente de valores, indicando que os anos seguintes podem seguir esta tendência.



## Variação mensal

A análise dos dados segmentados por meses do ano pode indicar tendências em relação à escolha de datas. O gráfico abaixo ilustra a somatória dos casamentos para cada mês, deixando evidente uma baixa no número em abril, seguido de crescimento em quase todos os meses até dezembro, onde se observa o pico de casamentos.

Este acúmulo no número pode estar relacionado a fatores como épocas festivas, períodos de recesso, ou até mesmo escolhas relacionadas ao clima predominante de cada período do ano.



## Análise por gênero

Assim como em todas análises apresentadas anteriormente, o número de casamentos homoafetivos femininos se mostra historicamente quase 50% superior em relação aos masculinos. Apesar da população brasileira ser predominantemente do gênero feminino, esta predominância possui valores condizentes com os analisados, indicando que a proporção observada nos dados históricos pode possuir influência de demais fatores.

A análise por gênero possibilita ainda maior aprofundamento, de forma a identificar possíveis pontos de influência nos valores observados.

Gênero	Número
Feminino	29693
Masculino	21373

## Análise preditiva

Dado a base de dados fornecida, construiu-se um modelo de regressão para se realizar a predição do número de casamentos homoafetivos nos próximos 2 anos nos 3 maiores estados em número dos casamentos, gerando as seguintes análises:

### Visão Geral

Dado os 3 estados com maior número de casamentos homoafetivos no último ano (São Paulo, Rio de Janeiro e Minas Gerais), o número total predito para os próximos 2 anos é apresentado na imagem abaixo. Observa-se que os números apresentam ainda um comportamento de crescimento, estando de acordo com a tendência observada anteriormente.

Ano	Número
2022	5340
2023	5652

## Predição por estado

As imagens abaixo apresentam o total predito por estado, além do número por gênero de cada estado. Observa-se que o número de casamentos homoafetivos femininos apresentam comportamento semelhante ao histórico, sendo predominantes em todos os estados analisados.

UF	Gênero	Número
Minas Gerais	Feminino	958
Minas Gerais	Masculino	795
Rio de Janeiro	Feminino	843
Rio de Janeiro	Masculino	630
São Paulo	Feminino	4438
São Paulo	Masculino	3328

Pode-se pontuar também o grande volume de casamentos no estado de São Paulo, o qual representa mais de 70% do volume dos casamentos. Este grande volume pode estar relacionado a fatores como volume populacional e também a fatores culturais que possam influenciar o número dos casamentos.

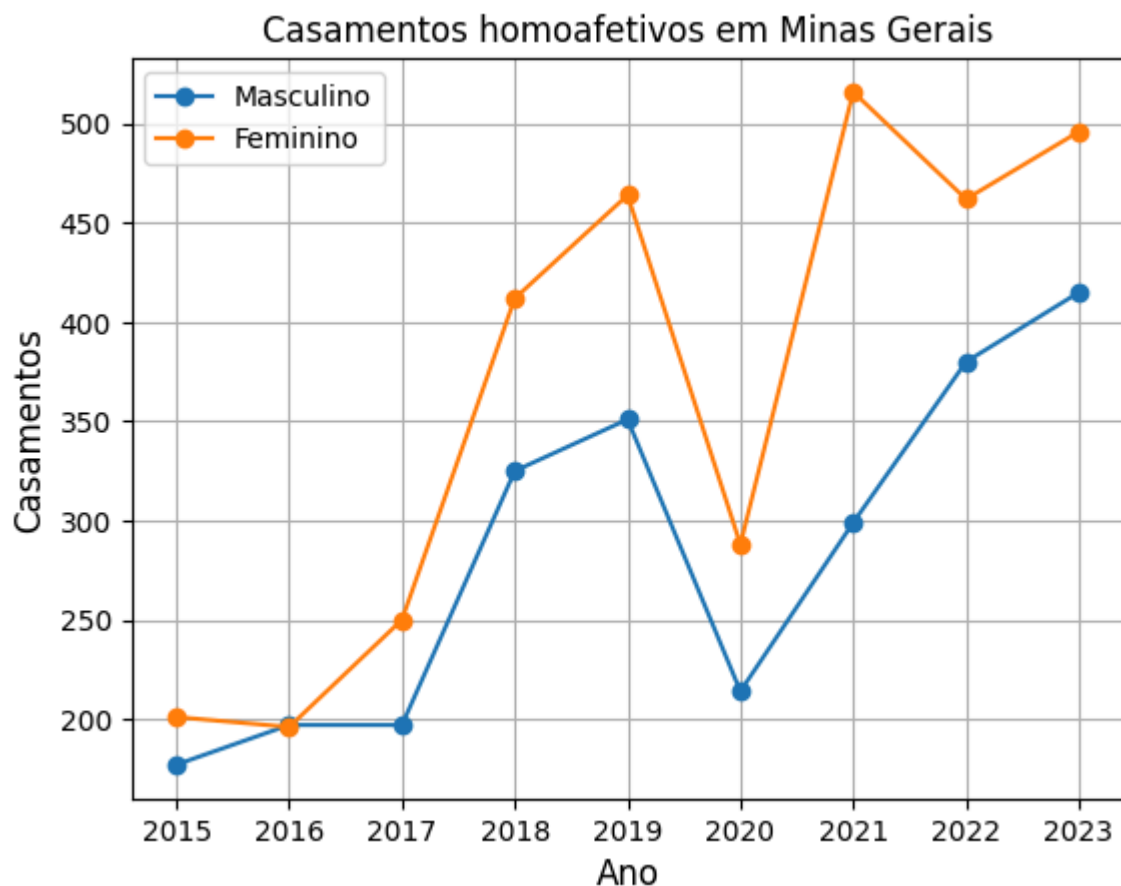
UF	Número
Minas Gerais	1753
Rio de Janeiro	1473
São Paulo	7766

## Comparativo histórico

Analisando-se os valores de cada estado individualmente ao longo dos anos, pode-se identificar tendências e comportamentos regionalizados.

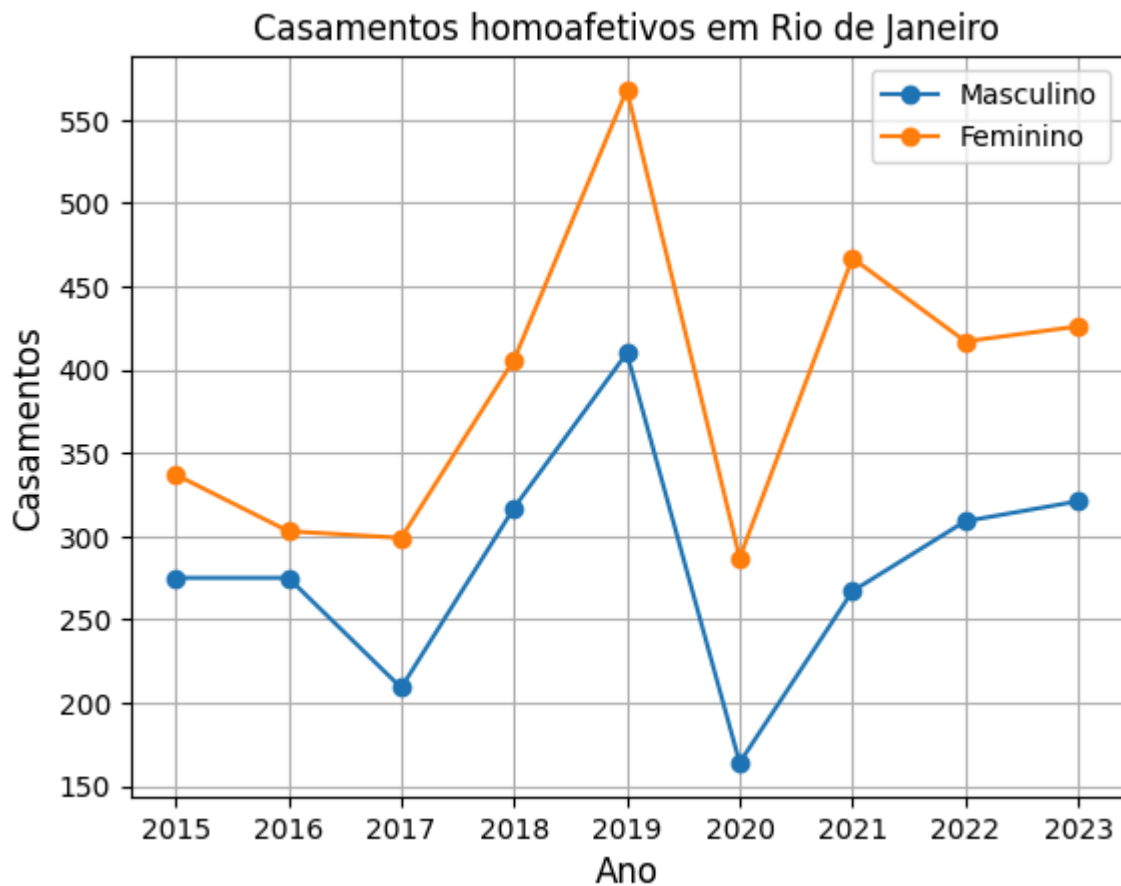
## Minas Gerais

Para o estado de Minas Gerais destaca-se o número de casamentos homoafetivos femininos que apresenta queda em 2022 em relação ao ano anterior. Este comportamento da predição pode estar relacionado ao fato da queda mais acentuada no número de casamentos femininos em 2020 ter causado um inflacionamento no ano seguinte (2021), desta forma atingindo uma certa estabilidade após o período. A curva de crescimento entre os anos 2022 e 2023 para ambos os gêneros se mostra muito similar, fortalecendo a hipótese anterior.



## Rio de Janeiro

De forma semelhante à predição de Minas Gerais, o número de casamentos homoafetivos apresentou queda acentuada no ano de 2020, atingindo menos de 50% do ano anterior, desta forma podendo também estar relacionado aos mesmos fatores apresentados anteriormente, causando o inflacionamento do número em 2021, para posterior estabilidade.

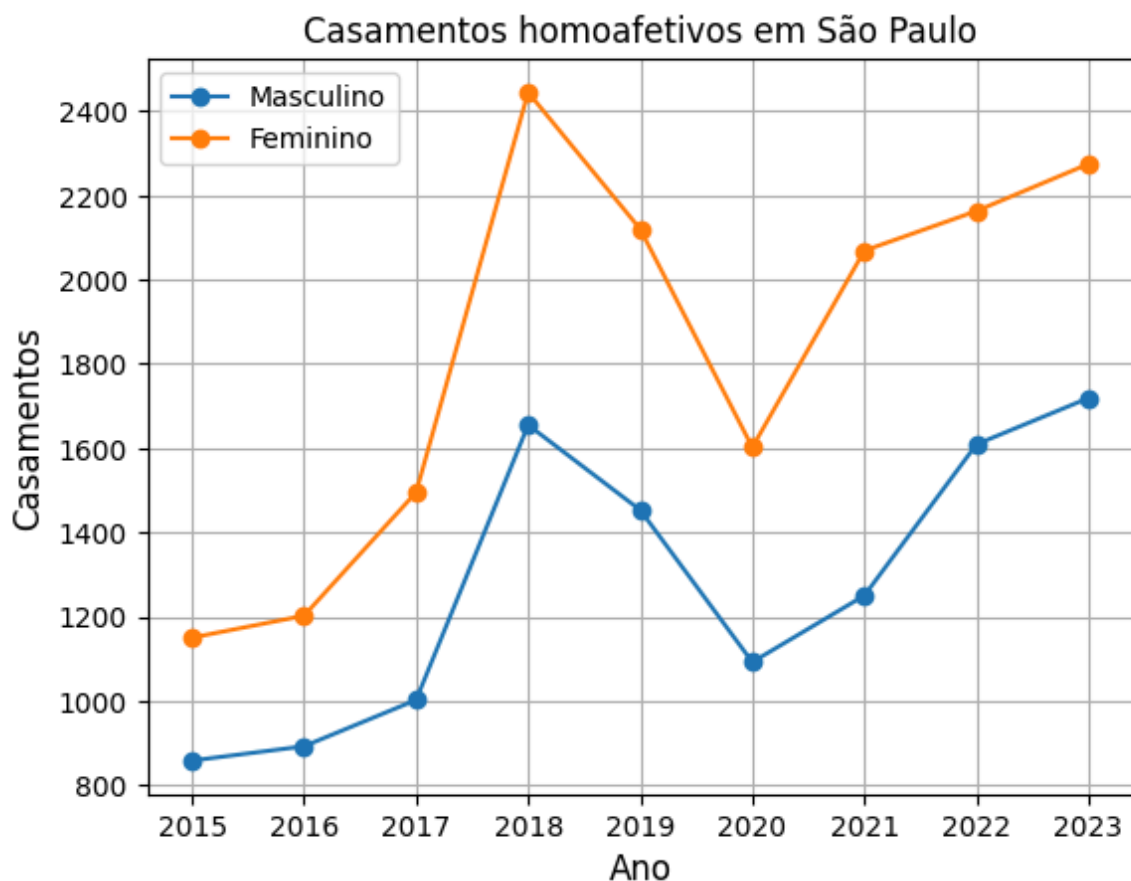




## São Paulo

O comparativo histórico x predito para o estado de São Paulo mostra que para ambos os gêneros, o número de casamentos se mantém crescente, porém em menor taxa quando comparados ao crescimento do ano anterior.

Ao se comparar a queda no número em 2020 com os demais estados analisados, nota-se que ambos os gêneros mantiveram números superiores a 50% do máximo histórico, ou seja, com menor taxa de decréscimo, da mesma forma que apresenta menor taxa de crescimento em 2021, seguido de estabilidade sem presença de queda.



# Conclusão

A análise histórica dos dados permitiu uma visão geral de tendência ao longo dos anos e dos meses, viabilizando a identificação de picos de casamentos principalmente no último trimestre do ano, além da identificação de correlação entre os valores observados e fatores externos.

Através dos dados preditos para os três maiores estados é possível identificar que a tendência segue sendo de crescimento no total de casamentos homoafetivos, porém este número estando sujeito a regionalização, dado que São Paulo é o local, com maior volume histórico e predito, e é o único estado que não apresentou queda em relação a 2022.

Os números preditos para Minas Gerais e Rio de Janeiro podem estar relacionados às maiores taxas de quedas históricas, mas também podem indicar regiões às quais ações sociais poderão causar maior efeito, dado o cenário de queda previsto inicialmente seguido de quase estabilidade dos valores.