

Casamentos Homoafetivos no Brasil

Uma análise feita por:

.monks

16/09/2024



SUMÁRIO

01. INTRODUÇÃO

02. METODOLOGIA

03. RESULTADOS

04. INSIGHTS

01.

INTRODUÇÃO

01. INTRODUÇÃO

Objetivo: Apresentar o processo de análise dos dados sobre a quantidade de casamentos homoafetivos no Brasil, detalhado as etapas realizadas, os resultados obtidos e principais insights.

Dados Utilizados:

- **Período:** 01/2015 a 12/2021;
- Estão divididos por **gênero**;
- Apresentam o **número de casamentos** de **cada estado** de acordo com o **período** e **gênero**.

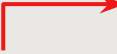


02.

METODOLOGIA

02. METODOLOGIA

Ferramentas Utilizadas:

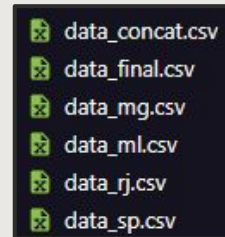
- Linguagem Python;
 - VSCode + Jupyter Notebook + Data Wrangler;
 - **Bibliotecas:**
 - Pandas;
 - Numpy;
 - Matplotlib;
 - Seaborn;
 - Scikit-Learn.
- Visualização das Tabelas
- 

02. METODOLOGIA

Preparação dos Dados:

1. Para Análise:

- Juntar as tabelas para formar um único dataframe ("*data_concat.csv*");
- Adição da coluna "Região" para análise dos dados ("*data_final.csv*");



2. Para Previsão (ML):

- Transformação dos dados categóricos para numéricos, adição da coluna "data" e "data_ordinal" e remoção da coluna "Região" ("*data_ml.csv*");
- Separação do Data Frame para os 3 Estados que foi solicitado a previsão;
- Remoção dos *Outliers*, através do método de Tukey ("*data_sp.csv*", "*data_mg*", "*data_rj*").

Não havia **nenhuma anormalidade** nos dados, então a única ação de limpeza de dados foi a **remoção** de **outliers** para treinamento do ML.

02. METODOLOGIA

Análise Descritiva:

- **Medidas Estatísticas:**

- Cálculo de **média, mediana, moda** e dispersão (**desvio-padrão, percentis**) para os dados de 2015 a 2021;

- **Proporção Total:** Por **região**, por **estado** e por **mês**.

- **Proporções por Gênero:**

- Análise das proporções de casamentos homoafetivos **masculinos e femininos** ao **longo dos anos**;
- Por **região** e por **estado**;

02. METODOLOGIA

Visualização dos Dados:

- **Gráfico de Linha:**
 - Casamentos Homoafetivos **Total, Feminino e Masculino** ao longo dos anos;
- **Gráfico de Barra:**
 - **Percentual** de Casamentos Homoafetivo **Feminino e Masculino**;
 - **Distribuição** de Casamentos Homoafetivo **Por Região**;
 - **Comparação** de Casamentos Homoafetivo **Por Gênero** em cada **Região**;
 - **Proporção Total e Anual** de Casamentos Homoafetivo **Por Estado**;
 - **Comparação** de Casamentos Homoafetivo **Por Gênero** em cada **Estado**;
 - **Distribuição mensal** de Casamentos Homoafetivo;
- **Gráficos Para identificação de Outliers:** *Box plot, Scatter* e Histograma

02. METODOLOGIA

Análise Preditiva:

- Para realizar a **previsão** dos dados para os **próximos 2 anos** nos **3 estados com maior número** de casamentos em **2021**. Foi utilizado o modelo ***Random Forest Regressor***, por conseguir **detectar não linearidades** e reduzir o **risco de overfitting**;
- Foi utilizado **20%** do dados para **Teste** e **80%** para **Treinamento**;
- Esses dados foram **transformados** para a **mesma escala** antes de entrar no modelo;
- Foi utilizado a métrica **Error Médio Absoluto (MAE)** para **avaliar** o modelo;
- Foram desenvolvidos **3 modelos, treinados e testados** para **cada estado**.

03.

RESULTADOS

03. RESULTADOS - ANÁLISE DOS DADOS

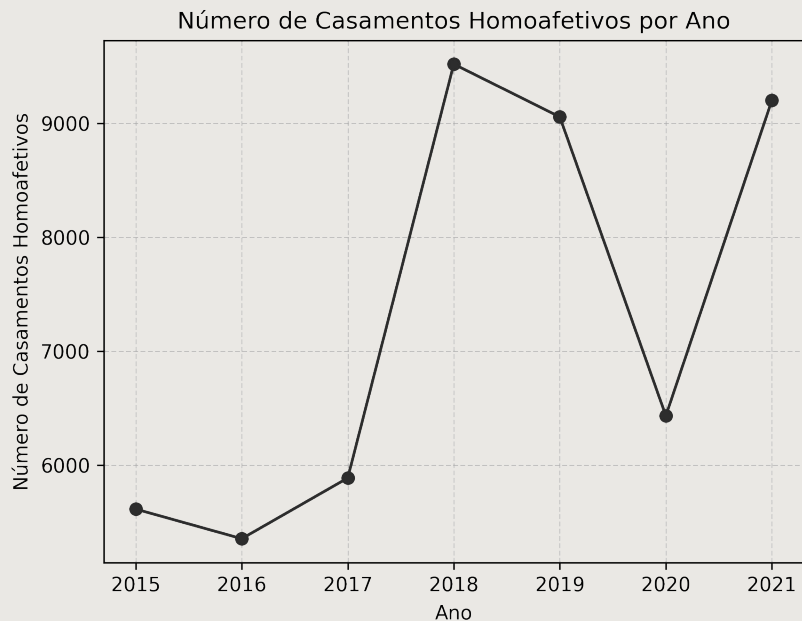
	ano	uf	genero	mes	numero	Região
0	2015	Acre	Masculino	Abril	0	Norte
1	2015	Acre	Masculino	Março	0	Norte
2	2015	Acre	Masculino	Fevereiro	0	Norte
3	2015	Acre	Masculino	Janeiro	0	Norte
4	2015	Rondônia	Masculino	Dezembro	1	Norte
...
4531	2021	Rondônia	Masculino	Maio	3	Norte
4532	2021	Rondônia	Masculino	Abril	1	Norte
4533	2021	Rondônia	Masculino	Março	1	Norte
4534	2021	Rondônia	Masculino	Fevereiro	0	Norte
4535	2021	Rondônia	Masculino	Janeiro	1	Norte

Data Frame utilizado para as análises
("data_final.csv")

Quantidade de Dados de "numero"	4536
Média	11,258
Desvio-Padrão	27,886
Menor Valor	0
25%	1
50%	4
75%	11
Maior Valor	827

03. RESULTADOS - ANÁLISE DOS DADOS

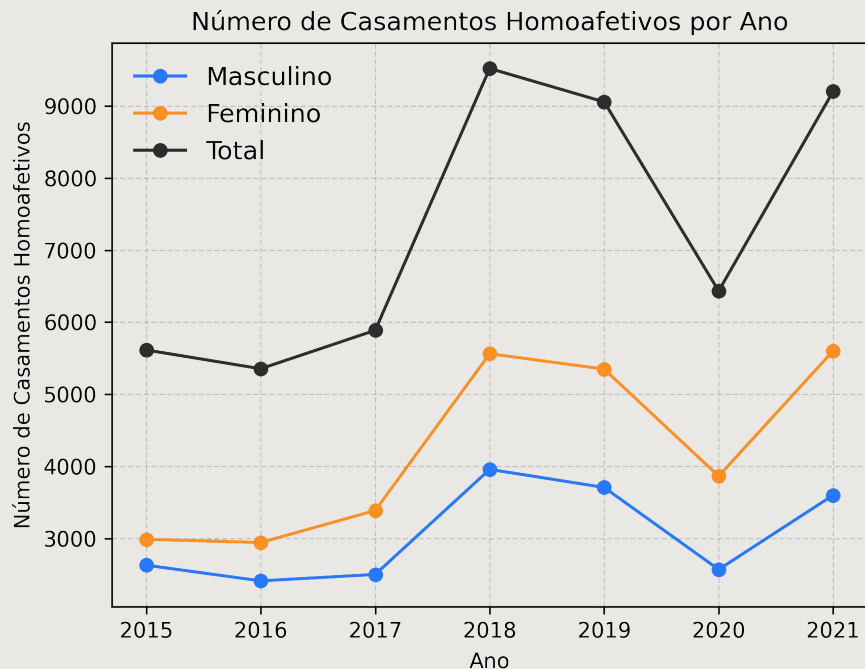
Quantidade de Casamentos Homoafetivos ao longo dos **anos** e a **variação** percentual anual



Ano	Total	Média	Variação (%)
2015	5614	8,66	-
2016	5354	8,26	-4,63
2017	5887	9,08	9,95
2018	9520	14,69	61,71
2019	9056	13,97	-4,87
2020	6433	9,93	-28,96
2021	9202	14,2	43,04

03. RESULTADOS - ANÁLISE DOS DADOS

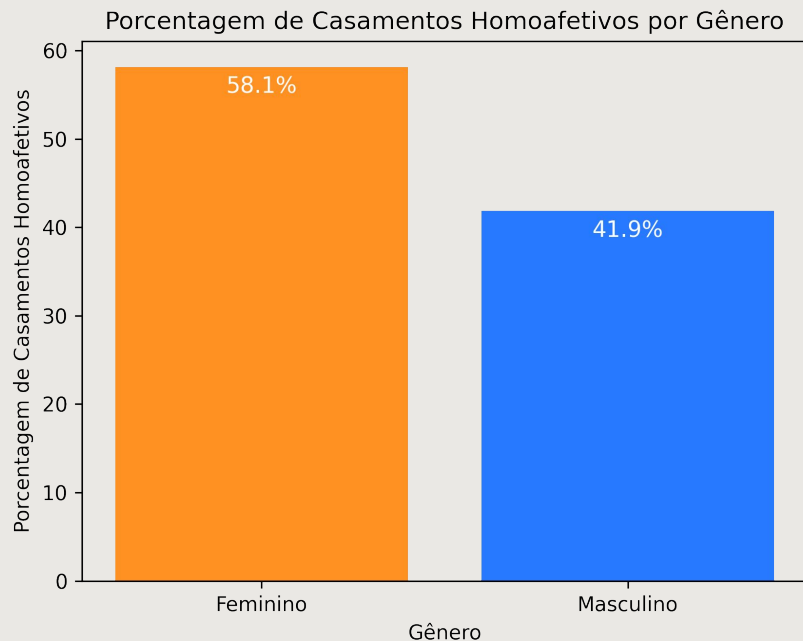
Quantidade de Casamentos Homoafetivos **Masculino** e **Feminino** ao longo dos **anos** e a **diferença** entre os gêneros



Ano	Fem.	Masc.	Diferença	Total
2015	2986	2628	358	5614
2016	2943	2411	532	5354
2017	3387	2500	887	5887
2018	5562	3958	1604	9520
2019	5349	3707	1642	9056
2020	3864	2569	1295	6433
2021	5602	3600	2002	9202

03. RESULTADOS - ANÁLISE DOS DADOS

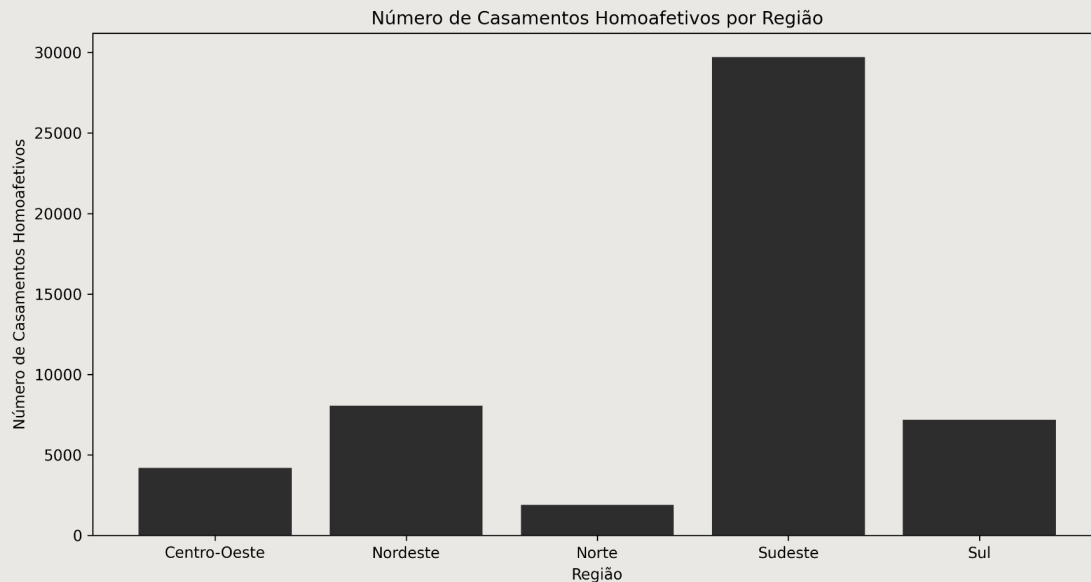
Porcentagem de Casamentos Homoafetivos Masculino e Feminino (2015-2021).



Gênero	Total	Média	%
Feminino	29693	13,09	58,14
Masculino	21373	9,42	41,85

03. RESULTADOS - ANÁLISE DOS DADOS

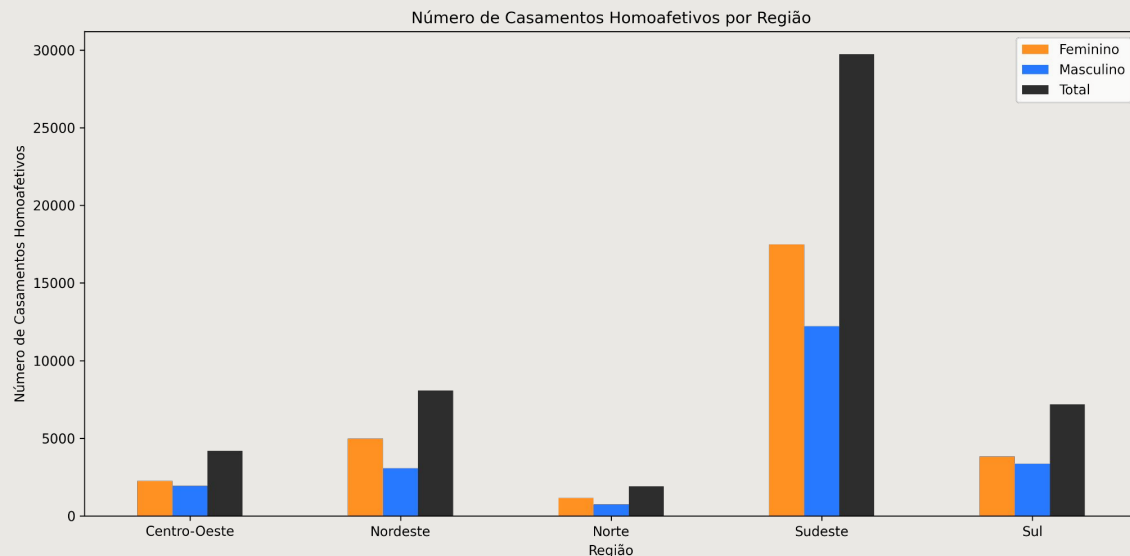
Quantidade de Casamentos Homoafetivos por Região (2015-2021).



Região	Total	Média
Sudeste	29714	44,22
Nordeste	8070	5,34
Sul	7192	14,27
Centro-Oeste	4192	6,24
Norte	1898	1,61

03. RESULTADOS - ANÁLISE DOS DADOS

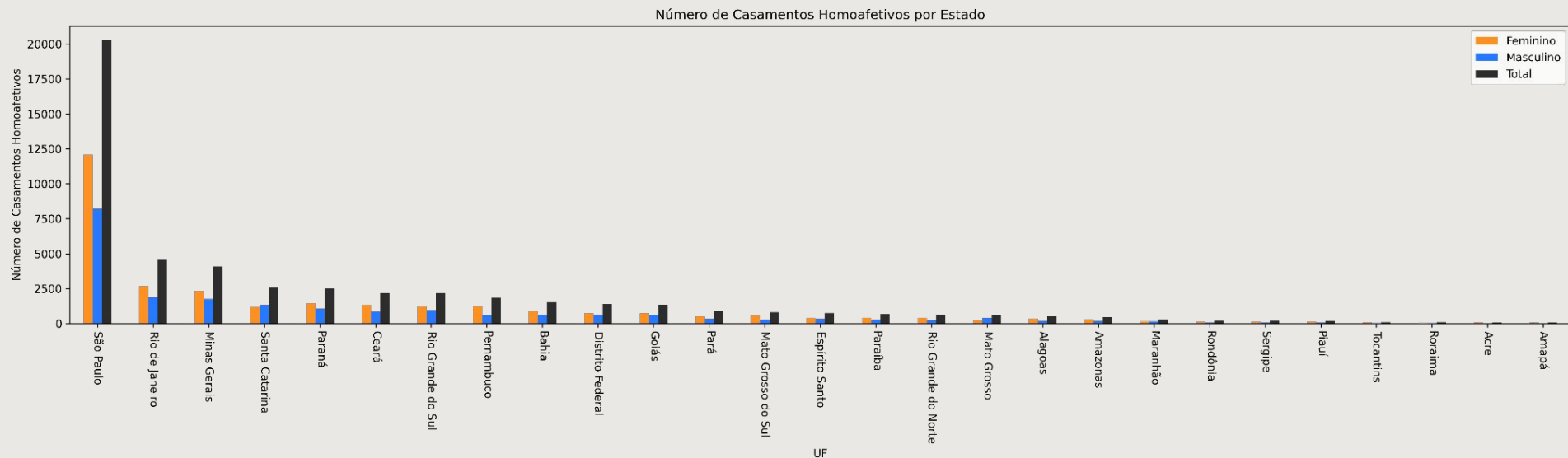
Quantidade de Casamentos Homoafetivos por **Região** e **Gênero** (2015-2021).



Região	Fem.	Masc.	Total
Sudeste	17482	12232	29714
Nordeste	4981	3089	8070
Sul	3826	3366	7192
Centro-Oeste	2250	1942	4192
Norte	1154	744	1898

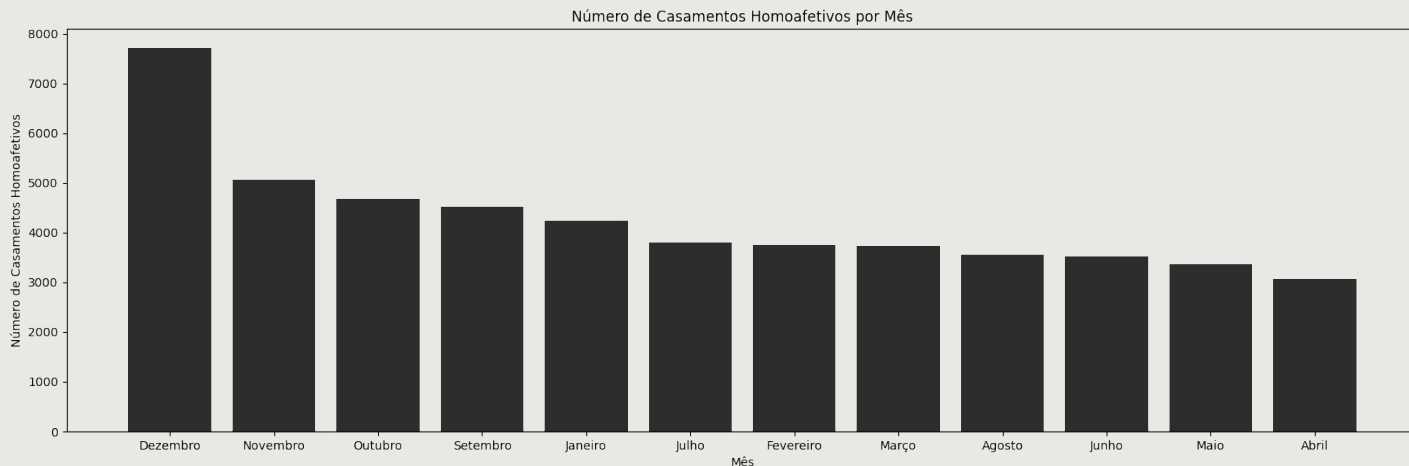
03. RESULTADOS - ANÁLISE DOS DADOS

Quantidade de Casamentos Homoafetivos por Estado e Gênero (2015-2021).



03. RESULTADOS - ANÁLISE DOS DADOS

Quantidade de Casamentos Homoafetivos por Mês (2015-2021).



Mês	Total
Dezembro	7716
Novembro	5062
Outubro	4689
Setembro	4528
Janeiro	4252

03. RESULTADOS - ANÁLISE DOS DADOS

Quantidade de Casamentos Homoafetivos por **Estado** e **Gênero**.

(2015-2021)

Estado	Feminino	Masculino	Total	Média
SP	12083	8202	20285	120,74
RJ	2667	1917	4584	27,29
MG	2327	1760	4087	24,33
SC	1175	1369	2544	15,14
PR	1446	1054	2500	14,88

(2021)

Estado	Total	Média
SP	3319	138,29
MG	815	33,95
RJ	734	30,58
PR	486	20,25
SC	404	16,83

Os estados que serão utilizados para a **previsão** dos dados, serão: **SP, MG e RJ**

03. RESULTADOS - PREVISÃO (2022 e 2023)

Data Frames utilizados:

```
df_sp = df[(df['uf_id'] == 35)]
df_sp
```

✓ 0.0s Open 'df_sp' in Data Wrangler

	ano	genero	mes	numero	uf_id	data	data_ordinal
176	2015	0	12	78	35	2015-12-01	735933
177	2015	0	11	86	35	2015-11-01	735903
178	2015	0	10	85	35	2015-10-01	735872
179	2015	0	9	76	35	2015-09-01	735842
180	2015	0	8	72	35	2015-08-01	735811
...
4451	2021	1	5	135	35	2021-05-01	737911
4452	2021	1	4	128	35	2021-04-01	737881
4453	2021	1	3	156	35	2021-03-01	737850
4454	2021	1	2	124	35	2021-02-01	737822
4455	2021	1	1	153	35	2021-01-01	737791

Dados de SP
("data_sp.csv")

```
df_mg = df[(df['uf_id'] == 31)]
df_mg
```

✓ 0.0s Open 'df_mg' in Data Wrangler

	ano	genero	mes	numero	uf_id	data	data_ordinal
148	2015	0	12	13	31	2015-12-01	735933
149	2015	0	11	25	31	2015-11-01	735903
150	2015	0	10	31	31	2015-10-01	735872
151	2015	0	9	14	31	2015-09-01	735842
152	2015	0	8	14	31	2015-08-01	735811
...
4419	2021	1	9	48	31	2021-09-01	738034
4420	2021	1	8	41	31	2021-08-01	738003
4421	2021	1	7	48	31	2021-07-01	737972
4422	2021	1	6	36	31	2021-06-01	737942
4423	2021	1	5	31	31	2021-05-01	737911

Dados do RJ
("data_rj.csv")

```
df_rj = df[(df['uf_id'] == 33)]
df_rj
```

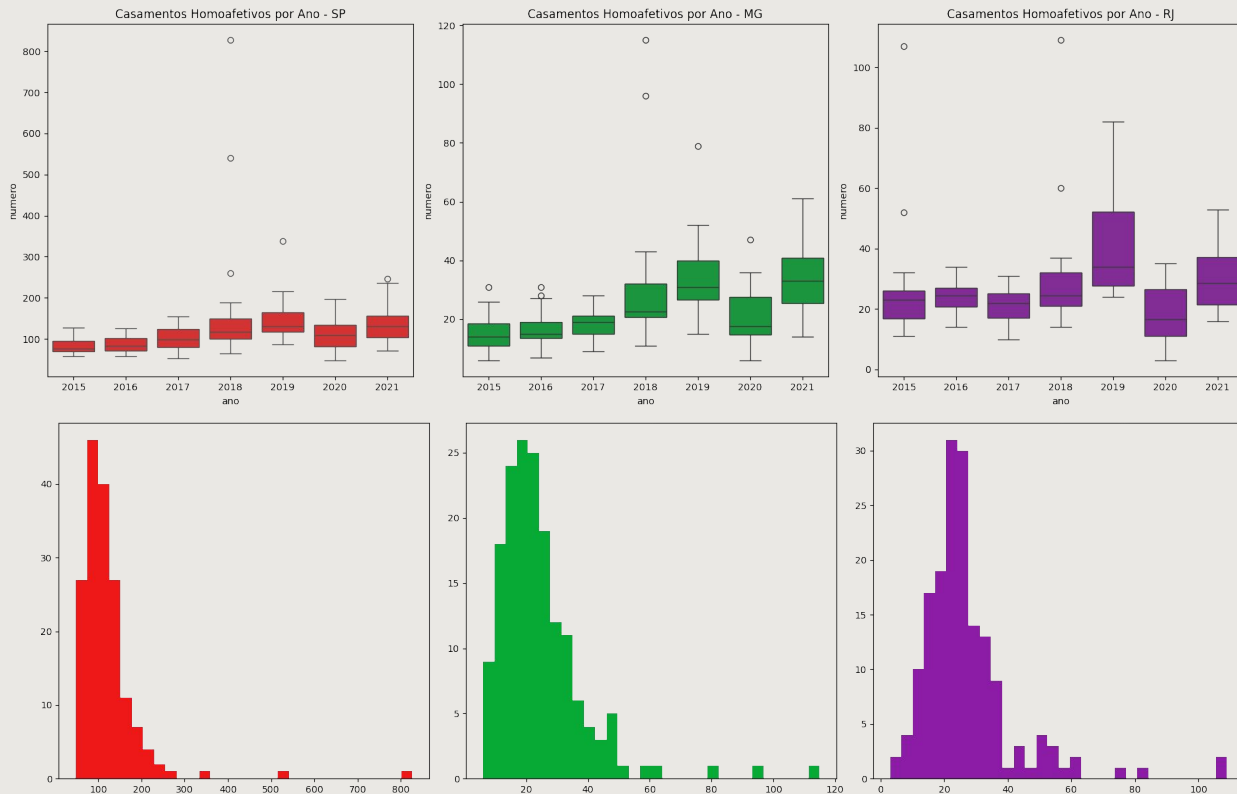
✓ 0.0s Open 'df_rj' in Data Wrangler

	ano	genero	mes	numero	uf_id	data	data_ordinal
160	2015	0	8	17	33	2015-08-01	735811
161	2015	0	7	20	33	2015-07-01	735780
162	2015	0	6	23	33	2015-06-01	735750
163	2015	0	5	16	33	2015-05-01	735719
164	2015	0	4	23	33	2015-04-01	735689
...
4431	2021	1	5	38	33	2021-05-01	737911
4432	2021	1	4	38	33	2021-04-01	737881
4433	2021	1	3	27	33	2021-03-01	737850
4434	2021	1	2	25	33	2021-02-01	737822
4435	2021	1	1	36	33	2021-01-01	737791

Dados de MG
("data_mg.csv")

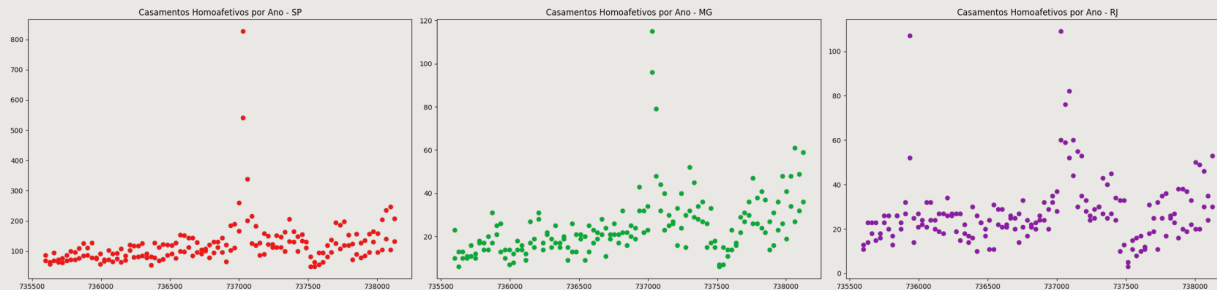
03. RESULTADOS - PREVISÃO (2022 e 2023)

Verificação dos *Outliers*

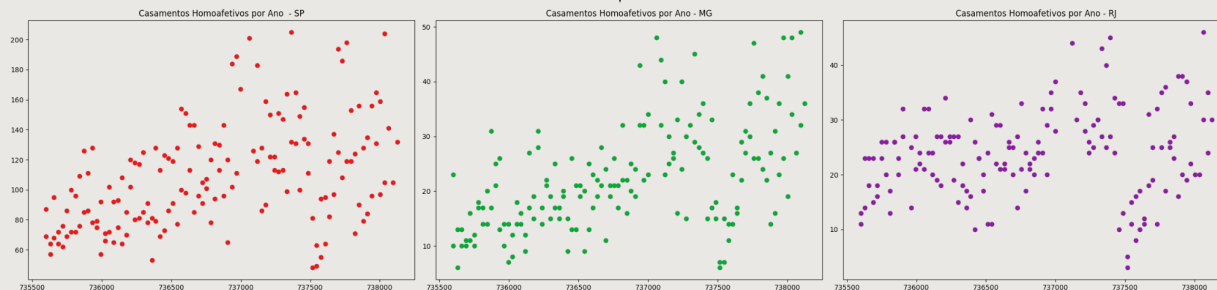


03. RESULTADOS - PREVISÃO (2022 e 2023)

Verificação dos *Outliers*



**Remoção dos Outliers
através do método de Tukey**



03. RESULTADOS - PREVISÃO (2022 e 2023)

Modelo RandomForestRegressor

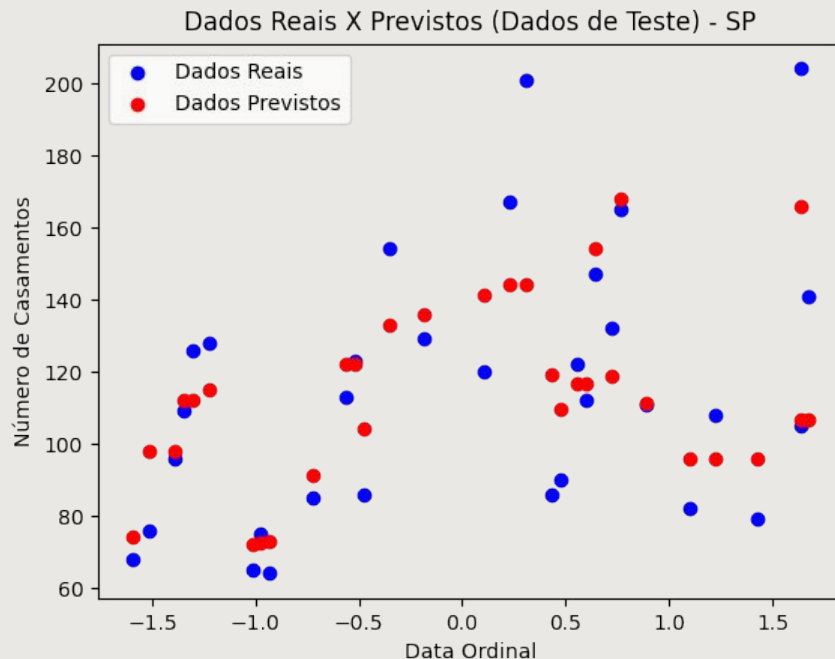
Melhores Hiperparâmetros - SP

```
{'bootstrap': True,  
'criterion': 'poisson', 'max_depth': 50,  
'max_features': 10, 'min_samples_leaf': 2,  
'min_samples_split': 4, 'n_estimators': 5}
```

**Erro Médio Absoluto
(MAE)**

13,92

Erro de aproximadamente
 ± 14 unidades.



03. RESULTADOS - PREVISÃO (2022 e 2023)

Modelo RandomForestRegressor

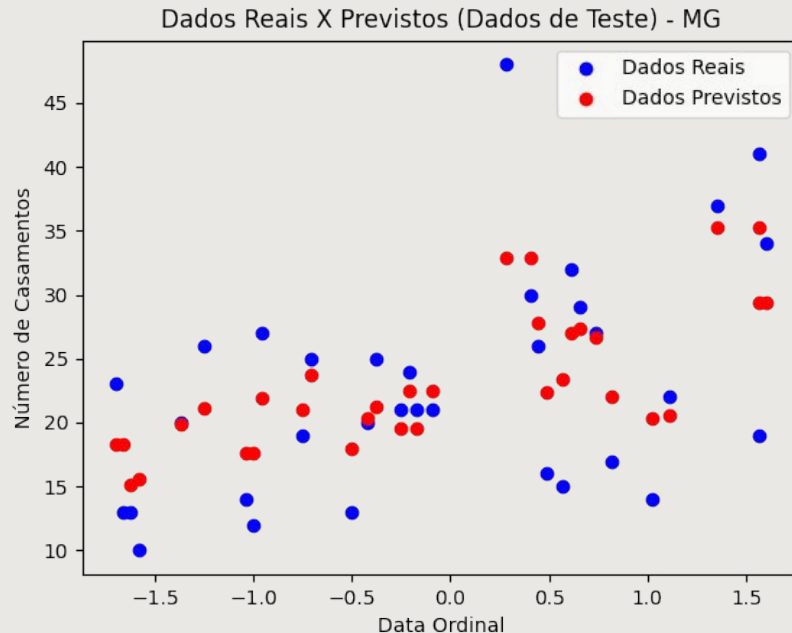
Melhores Hiperparâmetros - MG

```
{'bootstrap': True,  
'criterion': 'poisson', 'max_depth': 100,  
'max_features': 5, 'min_samples_leaf': 2,  
'min_samples_split': 8, 'n_estimators': 10}
```

**Erro Médio Absoluto
(MAE)**

4

Erro de aproximadamente 4 unidades para cima ou para baixo



03. RESULTADOS - PREVISÃO (2022 e 2023)

Modelo RandomForestRegressor

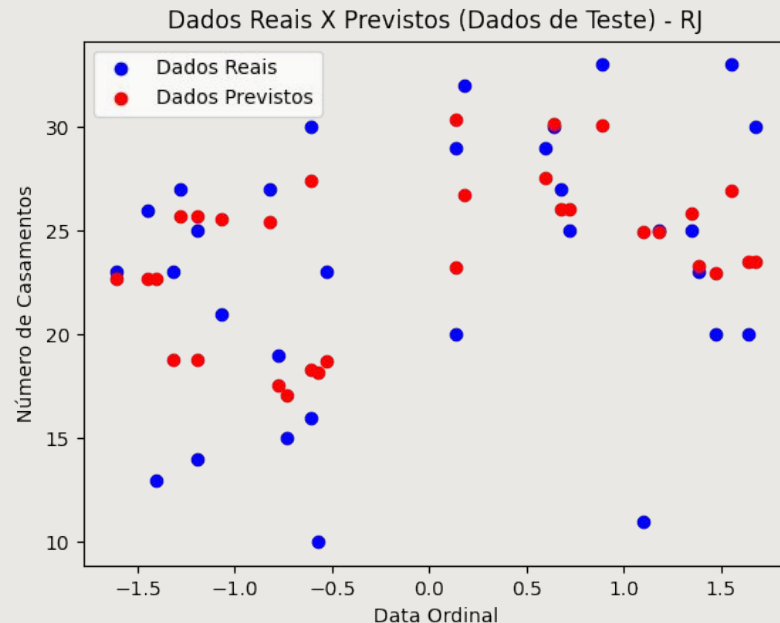
Melhores Hiperparâmetros - RJ

```
{'bootstrap': True,  
'criterion': 'poisson', 'max_depth': 100,  
'max_features': 5, 'min_samples_leaf': 2,  
'min_samples_split': 4, 'n_estimators': 10}
```

**Erro Médio Absoluto
(MAE)**

3,28

Erro de aproximadamente
3,28 unidades para cima ou
para baixo



03. RESULTADOS - PREVISÃO (2022 e 2023)

Valores Previstos no Modelo

ANO	São Paulo	Minas Gerais	Rio de Janeiro
2022	3273	781	618
2023	3310	781	622
Variação 2021-2022	-1,4%	-4,35%	-15,8%
Variação 2022-2023	2,96%	0%	0,64%



04. **INSIGHTS**

04. INSIGHTS

- A **média de casamentos** de **2015 a 2021**, considerando **todos os gêneros**, foi de **aproximadamente 11**. Apesar de ser um número pequeno, o **contexto social de aceitação** ao longo do tempo possui **grande influência** para essa média;
- **Em 2018**, houve um **aumento significativo** no número de casamentos homoafetivos. Esse fato **pode estar relacionado** com o **crescimento da aceitação pública, debates na mídia e eventos que trazem visibilidade para os direitos LGBTQIA+**.
- **No ano de 2020**, houve uma **queda acentuada** de aproximadamente **-29%** no número de casamentos, provavelmente causada pela **pandemia da COVID-19**. As restrições sanitárias, distanciamento social e *lockdown* impactaram diretamente a realização das cerimônias de casamento, fazendo com que muitos casais adiassem o evento, causando essa redução significativa.

04. INSIGHTS

- **Em 2021**, foi possível perceber uma **recuperação nos números de casamentos**, ultrapassando os valores de 2019. Uma possível causa para isso é a **volta das cerimônias** que haviam sido **adiadas** em 2020.
- Desde 2015, a **quantidade de casamentos homoafetivos femininos** se manteve **maior do que a de casamentos masculinos**, sendo **58,1% dos casamentos foram femininos**, enquanto **41,9% foram masculinos**, esse fato **se repete para uma análise regional**. Uma possível causa para isso é as mulheres estarem mais propensas a formalizar suas uniões. **Em 2021**, essa **diferença foi de 2002 casamentos** a mais entre **mulheres**, o maior valor entre 2015-2021.

04. INSIGHTS

- As regiões **Sudeste, Nordeste e Sul lideram** em **números absolutos** de **casamentos homoafetivos**. Porém, ao **observar as médias**, as regiões **Sudeste, Sul e Centro-Oeste** são as que **se destacam**. Embora o **Nordeste** tenha um **número total alto**, a **distribuição** de casamentos **por estados** no **Sul e Centro-Oeste** é **mais homogênea**, podendo indicar uma aceitação do contexto social ou política nesses estados.
- Há apenas **3 estados** onde os **casamentos homoafetivos masculinos** são **mais frequentes: Santa Catarina, Mato Grosso e Maranhão**. Esse dado pode ser reflexo de características culturais ou demográficas desses estados.
- Os estados com os maiores números casamentos homoafetivos são: **São Paulo, Rio de Janeiro e Minas Gerais**. Esses **estados são mais populosos, economicamente desenvolvidos** e possuem **maior aceitação social dos direitos LGBTQIA+**, influenciando **positivamente** os números.

04. INSIGHTS

- Já para o último ano, **2021**, Os **estados com os maiores números casamentos homoafetivos** são: **São Paulo, Minas Gerais e Rio de Janeiro**.
- A maior parte dos casamentos homoafetivos **ocorrem nos últimos 3 meses do ano**, sendo **Dezembro, Novembro e Outubro**. Uma possível causa para isso é o **clima mais favorável, recebimento de bônus ou 13º salário e férias** de fim de ano.
- O modelo de *Machine Learning* previu:
 - Queda de -1,4% para a quantidade de casamentos em São Paulo, entre 2021 e 2022. Ocorrendo uma recuperação em 2023;
 - Queda de -4,35% para os casamentos em Minas Gerais, entre 2021 e 2022;
 - Queda de -15,8% para os casamentos no Rio de Janeiro, entre 2021 e 2022. Ocorrendo uma pequena recuperação em 2023.

A vertical bar on the left side of the slide, composed of seven equal-width stripes of red, orange, yellow, green, blue, and purple, representing the rainbow flag.

Obrigada!

Uma análise feita por:

.monks