

Especialização Gestão e Ciência de Dados
Dados, Conhecimento e Inteligência Artificial

Equipe:

Fernando Aranha

Eder Ferreira

Pedro Henrique

Análise do Perfil Eleitoral Brasileiro de 2020: Tendências, Padrões e Implicações

1. Introdução

Analisar o perfil do eleitorado é essencial para entender a composição demográfica dos eleitores, identificar padrões de comportamento eleitoral e planejar campanhas políticas eficazes. Este projeto visa analisar a base de dados fornecida pelo Tribunal Superior Eleitoral (TSE) referente ao perfil do eleitorado em 2020, buscando insights que possam orientar estratégias políticas e sociais.

1.1. Descrição do problema.

Como podemos prever a participação eleitoral nas próximas eleições presidenciais no Brasil com base em dados demográficos, como idade, gênero e escolaridade, utilizando o conjunto de dados do Tribunal Superior Eleitoral (TSE) sobre a população votante atualizado em setembro de 2020, visando entender padrões e tendências?

1.2. Perguntas de pesquisa.

1. Qual é a distribuição demográfica dos eleitores (idade, gênero, estado civil, escolaridade) no Brasil em 2020?
2. Quais grupos demográficos representam a maior proporção do eleitorado?
3. Como a escolaridade dos eleitores se distribui geograficamente?

1.3 Identificar as partes interessadas e suas necessidades, se houver.

Possíveis partes interessadas podem incluir políticos, cientistas políticos, agências de pesquisa, entre outros, com necessidades variadas de compreensão do eleitorado. Incluindo o próprio cidadão como forma de aumentar a transparência do processo.

1.4 Objetivos

1. Analisar e descrever a distribuição demográfica do eleitorado brasileiro em 2020.
2. Identificar e visualizar padrões regionais significativos.
3. Prover insights baseados nos dados que possam informar estratégias políticas e sociais.

2. Coleta de dados:

- Identificar as fontes de dados relevantes: O conjunto de dados do TSE sobre o perfil eleitoral de 2020.
- Coletar os dados necessários: Baixar e importar o conjunto de dados para análise do sitio: <https://www.kaggle.com/datasets/aryvini/tse-perfil-eleitorado-2020>
- Descrever e avaliar a qualidade e a quantidade dos dados: Avaliar a integridade dos dados, verificar se há dados faltantes ou inconsistentes, e avaliar a quantidade de registros disponíveis.

3. Pré-processamento dos dados:

No campo em constante evolução da Ciência de Dados e Inteligência Artificial, a modelagem de dados é um componente fundamental para transformar informações não estruturadas em insights significativos (Provost & Fawcett, 2013).

Este processo meticuloso permite que organizações e pesquisadores façam previsões acuradas, identifiquem padrões ocultos e tomem decisões informadas com base em grandes volumes de dados (Hastie, Tibshirani & Friedman, 2009).

A efetividade da modelagem de dados está fortemente atrelada a uma etapa inicial vital, que é o pré-processamento de dados (Kotsiantis, Kanellopoulos & Pintelas, 2006).

Esta fase prepara o cenário, assegurando a integridade, consistência e formatação adequada dos dados para serem processados por algoritmos de aprendizado de máquina (Brownlee, 2019).

- Combinar diferentes fontes de dados quando necessário: Não é necessário neste caso, já que estamos lidando com um único conjunto de dados.
- Limpar, filtrar e preparar os dados para a análise: Remover dados duplicados, tratar valores ausentes ou inconsistentes, se houver, e preparar os dados para análise.
- Tratar dados ausentes e inconsistentes: Aplicar técnicas como preenchimento de dados ausentes ou remoção de registros com inconsistências.
- Realizar a transformação de dados quando necessário: Converter tipos de dados, codificar variáveis categóricas, normalizar dados, etc.
- Criar o conjunto de dados final para análise: Preparar o conjunto de dados limpo e estruturado para análise subsequente.

4. Conclusão

Os dados mostram que a maior parte do eleitorado brasileiro em 2020 é composta por eleitores femininos, solteiros, na faixa etária de 35 a 39 anos, e com ensino médio completo. A escolaridade dos eleitores varia significativamente entre os estados, com algumas regiões apresentando maiores proporções de eleitores com ensino superior completo e outras com maiores proporções de eleitores analfabetos ou com ensino fundamental incompleto.

5.Referências

- Brownlee, J. (2019). Data Preparation for Machine Learning. Machine Learning Mastery.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
- Kotsiantis, S., Kanellopoulos, D., & Pintelas, P. (2006). Data preprocessing for supervised learning. International Journal of Computer Science, 1(2), 111-117.
- Provost, F., & Fawcett, T. (2013). Data Science for Business. O'Reilly Media.
- <https://www.kaggle.com/datasets/aryvini/tse-perfil-eleitorado-2020/code>

Resultados da Análise

1. Distribuição demográfica dos eleitores no Brasil em 2020

- **Distribuição por Faixa Etária:**

- A faixa etária com o maior número de eleitores é de 35 a 39 anos, com 4.090.971 eleitores.

DS_FAIXA_ETARIA	QT_ELEITORES_PERFIL
100 anos ou mais	15679
16 anos	137288
17 anos	309039
18 anos	512815
19 anos	659576
20 anos	787017
21 a 24 anos	3198957
25 a 29 anos	3956175
30 a 34 anos	4008753
35 a 39 anos	4090971
40 a 44 anos	3732945
45 a 49 anos	3167047
50 a 54 anos	2936341
55 a 59 anos	2598952
60 a 64 anos	2136032
65 a 69 anos	1638983
70 a 74 anos	1212601
75 a 79 anos	781292
80 a 84 anos	520602
85 a 89 anos	266748
90 a 94 anos	152486
95 a 99 anos	76678
Inválido	10927

- **Distribuição por Gênero:**

- Feminino: 19.428.653 eleitores
- Masculino: 17.474.054 eleitores
- Não Informado: 5.197 eleitores

DS_GENERO	QT_ELEITORES_PERFIL
FEMININO	19428653
MASCULINO	17474054
NÃO INFORMADO	5197

- **Distribuição por Estado Civil:**

- Solteiro: 22.534.103 eleitores
- Casado: 11.915.361 eleitores
- Divorciado: 1.187.549 eleitores
- Separado Judicialmente: 288.837 eleitores
- Viúvo: 972.885 eleitores
- Não Informado: 9.169 eleitores

DS_ESTADO_CIVIL	QT_ELEITORES_PERFIL
CASADO	11915361
DIVORCIADO	1187549
NÃO INFORMADO	9169
SEPARADO JUDICIALMENTE	288837
SOLTEIRO	22534103
VIÚVO	972885

- **Distribuição por Escolaridade:**

- Ensino Médio Completo: 9.403.400 eleitores
- Ensino Fundamental Incompleto: 8.512.914 eleitores
- Ensino Médio Incompleto: 5.483.920 eleitores
- Lê e Escreve: 3.969.041 eleitores
- Ensino Superior Completo: 3.497.232 eleitores
- Ensino Fundamental Completo: 1.989.519 eleitores
- Superior Incompleto: 1.748.360 eleitores
- Analfabeto: 2.295.947 eleitores
- Não Informado: 7.571 eleitores

DS_GRAU_ESCOLARIDADE	QT_ELEITORES_PERFIL
ANALFABETO	2295947
ENSINO FUNDAMENTAL COMPLETO	1989519
ENSINO FUNDAMENTAL INCOMPLETO	8512914
ENSINO MÉDIO COMPLETO	9403400
ENSINO MÉDIO INCOMPLETO	5483920
LÊ E ESCRIVE	3969041
NÃO INFORMADO	7571
SUPERIOR COMPLETO	3497232
SUPERIOR INCOMPLETO	1748360

2. Grupos demográficos que representam a maior proporção do eleitorado

- **Faixa Etária mais prevalente:**

- 35 a 39 anos

- **Gênero mais prevalente:**

- Feminino
- **Estado Civil mais prevalente:**
 - Solteiro
- **Escolaridade mais prevalente:**
 - Ensino Médio Completo

3. Distribuição da escolaridade dos eleitores geograficamente

A tabela a seguir mostra a distribuição da escolaridade por estado (SG_UF):

SG_UF	ANALFABETO	ENSINO FUNDAMENTAL COMPLETO	ENSINO FUNDAMENTAL INCOMPLETO	ENSINO MÉDIO COMPLETO	ENSINO MÉDIO INCOMPLETO	LÊ E ESCRIVE	NÃO INFORMADO	SUPERIOR COMPLETO	SUPERIOR INCOMPLETO
AC	40033	24776	98117	116368	75800	46289		49530	32409
AL	71201	42783	228960	193320	138009	90134		78618	47091
AM	51015	52932	223324	333985	173857	69910	17	86641	61792
AP	14955	26050	91301	127978	87217	23789		45099	34165
BA	793981	452289	2520825	2848466	1451835	1507239	1427	857016	400945
CE	520909	367429	1414623	1640377	963506	781806	1005	565325	312780
DF	16428	94053	254674	552088	232435	66705	1	437432	186985
ES	46883	79693	310629	269781	191132	76209	491	109565	48030
GO	39811	78597	301443	351861	192114	94374		163503	88546
MA	220083	143673	546160	627940	382155	337746	144	151397	83075
MG	215690	321912	1309152	1090731	861855	409996	3502	479144	196835
PA	5184	3203	13823	11070	12082	7704	18	5367	1572
PB	32239	21820	107055	106183	51978	61378		39828	33929
PE	108250	78472	403222	377866	228492	172784	265	137751	68204
PI	47008	29016	132090	116435	76242	82543	1	51236	22443
PR	5848	14904	48211	45865	20949	8247	4	15800	7842
RJ	8486	38094	110169	98088	89681	29228	694	27029	16355
RN	10248	6128	41712	37502	20718	17111		14548	8357
RO	2376	2689	12670	10613	7003	4234		3567	1663
RR	1176	2126	8949	13581	7402	2562		8162	3469
RS	80	213	675	276	180	108		222	67
SC	5058	36051	83456	93274	47099	15061	2	38418	25055
SE	5539	3570	26026	12446	13246	10157		3088	2392
SP	23441	59465	187057	265808	129181	36327		98256	44191
TO	10003	8934	37980	57238	28669	17178		23964	17426