

# **ANÁLISE DE DADOS SOBRE INGRESSO DE MULHERES NO ENSINO SUPERIOR NO CURSO DE CIÊNCIAS DE DADOS**

Rebeca Caroline Souza da Silva

Matrícula: 01508561

Superior de Tecnologia em Data Science

A análise foi construída a partir do Microdados do Censo da Educação Superior do Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). São dados abertos disponíveis no Portal de Dados Abertos do Governo Federal. Foram utilizadas as bases do cadastro por cursos dos períodos de 2020 e 2021.

Apesar do crescente número de participação de mulheres na área de tecnologia, inclusive no nincho da ciência de dados, o mercado ainda é muito desigual tendo sua predominância masculina.

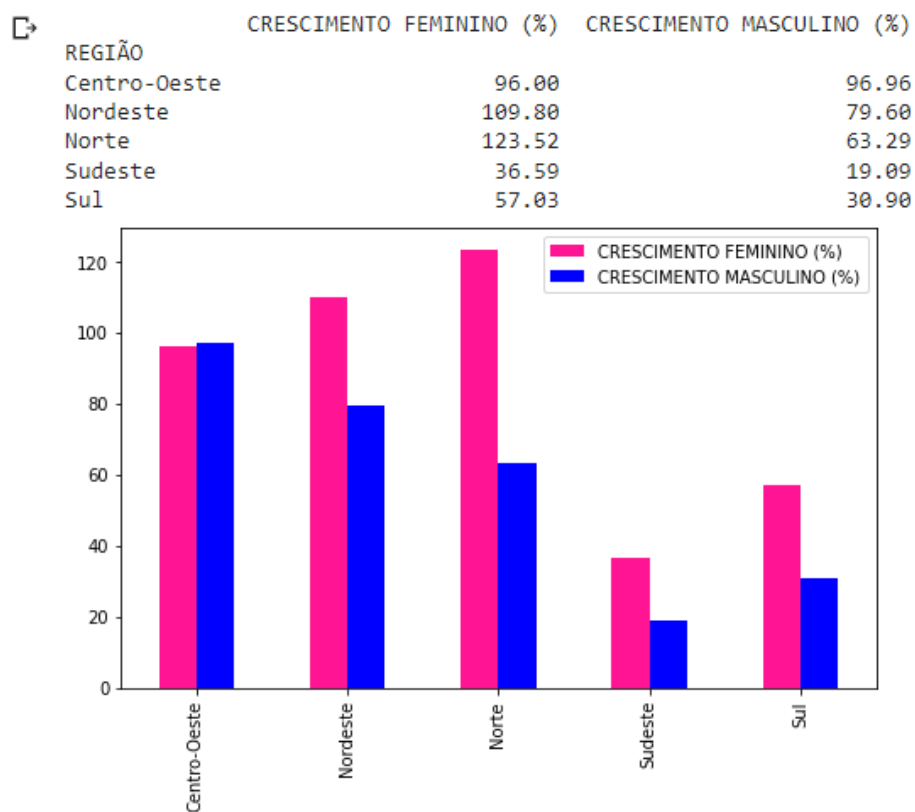
Baseada nessa tendência que a análise foi assentada. Tendo o intuito de observar dentro do período dos dois últimos anos o ingresso de mulheres na educação superior no cursos de ciências de dados por região do Brasil para que instituições educacionais e empresas sejam privadas ou públicas possam estimular a adesão do público feminino na área de dados, garantindo não somente uma formação superior de qualidade, como também, o ingresso dessas mulheres como profissionais no mercado de trabalho. E percebe-las por região pode impulsionar programas educacionais e capacitações profissionais em territórios com menores adesões de forma mais estratégica.

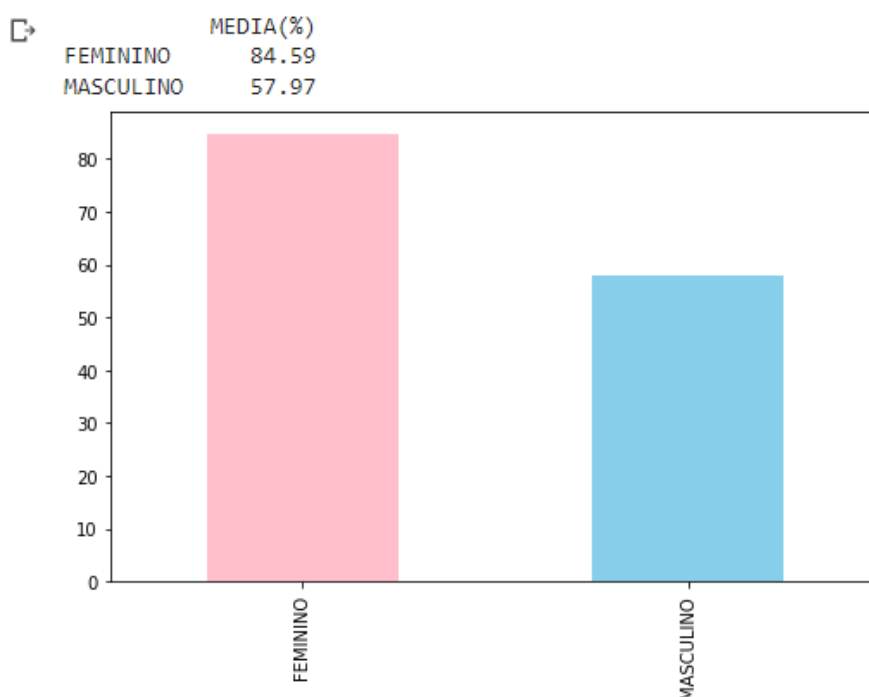
Foi feita uma análise descritiva dos dados com a ferramenta Python. Toda a análise foi comparar o ingresso na educação superior no curso de ciências de dados entre mulheres e homens.

Iniciou importando as bases, concatenando-as e trabalhando com filtros onde foi verificado que em quantidade a predominância masculina é sobressalente. Contudo, após o tratamento mais apurado e profundo dos dados aplicando algumas funções e métodos da biblioteca pandas, foi possível, compreender as informações que indicam que no ano de 2021 na área de dados houve maior adesão feminina em praticamente em todas as regiões do país.

	QT_ING_FEM		QT_ING_MASC	
	sum	mean	sum	mean
NO_REGIAO				
<b>Centro-Oeste</b>	148	0.99	490	3.29
<b>Nordeste</b>	158	0.49	699	2.17
<b>Norte</b>	55	0.45	208	1.69
<b>Sudeste</b>	2948	1.71	8639	5.03
<b>Sul</b>	329	0.77	919	2.14

Nesse dataframe vizualizamos a quantidade e a media de forma bem intuitiva os ingresso feminino (QT\_ING\_FEM) em todas as regiões do Brasil (NO\_REGIAO) que se comparados os ingresso masculino (QT\_ING\_MASC) são em números bem menores.





Em resumo, os gráficos em barras acima demonstram os comparativos feitos na análise contendo os resultados dos cálculos das médias em porcentagem (%), trazendo assim maior clareza e compreensão da informação extraída desses dados.

A média do crescimento do ingresso feminino em ensino superior no curso de ciências de dados no ano de 2021, atingiu 84.59 %. Ficando o crescimento do ingresso masculino em 57.97 %. Sendo assim, em comparação ao ano de 2020, houve 26.62 % mais adesão em 2021 por parte das mulheres na área de dados. Esse foi o resultado encontrado na análise de dados sobre ingresso de mulheres no ensino superior no curso de ciências de dados.

## REFERÊNCIA BIBLIOGRÁFICA

INEP. **Microdados do Censo da Educação Superior**, 04 de nov, de 2022. Disponível em: <https://www.gov.br/inep/pt-br/acesso-a-informacao/dados-abertos/microdados/censo-da-educacao-superior> . Acessado em: 26 Dez. de 2022

## ANEXO

### Código Desenvolvido:

```
import pandas as pd

base2020 = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/MICR
ODADOS CADASTRO CURSOS 2020.CSV', sep=';', encoding='latin1', useco
ls=['NU_ANO_CENSO', 'NO_REGIAO', 'NO_CURSO', 'QT_ING_FEM', 'QT_ING_M
ASC'])
base2021 = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/MICR
ODADOS CADASTRO CURSOS 2021.CSV', sep=';', encoding='latin1', useco
ls=['NU_ANO_CENSO', 'NO_REGIAO', 'NO_CURSO', 'QT_ING_FEM', 'QT_ING_M
ASC'])
analise = pd.concat([base2020, base2021])

display(analise)

analise_limpa = analise = analise.dropna()
analise_limpa = analise[analise['NO_CURSO'] == 'CIÊNCIA DE DADOS']
display(analise_limpa)

analise_qtda = analise_limpa.groupby('NO_REGIAO')[['QT_ING_FEM', 'QT
_ING_MASC']].agg(['sum', 'mean'])
analise_qtda.round(2)

analise_fem = analise_limpa.groupby(['NU_ANO_CENSO', 'NO_REGIAO'])['
QT_ING_FEM'].sum()
Fem = [analise_fem.filter(like='Centro-
Oeste').diff()/50*100], [analise_fem.filter(like='Nordeste').diff()
/51*100], [analise_fem.filter(like='Norte').diff()/17*100], [analise
_fem.filter(like='Sudeste').diff()/1246*100], [analise_fem.filter(l
ike='Sul').diff()/128*100]
display(analise_fem, Fem)

analise_masc = analise_limpa.groupby(['NU_ANO_CENSO', 'NO_REGIAO'])[
'QT_ING_MASC'].sum()
Masc = [analise_masc.filter(like='Centro-
Oeste').diff()/165*100], [analise_masc.filter(like='Nordeste').diff()
()/250*100], [analise_masc.filter(like='Norte').diff()/79*100], [ana
lise_masc.filter(like='Sudeste').diff()/3943*100], [analise_masc.fi
lter(like='Sul').diff()/398*100]
display(analise_masc, Masc)
```

```

import matplotlib.pyplot as plt

analise_conciliadafem = {

    'REGIÃO': ['Centro-
Oeste', 'Nordeste', 'Norte', 'Sudeste', 'Sul'],

    'CRESCIMENTO FEMININO (%)': [96.00, 109.80, 123.52, 36.59, 57.03]
,}

resultado_fem = pd.DataFrame(analise_conciliadafem)

graficofem = resultado_fem[['CRESCIMENTO FEMININO (%)']].plot(figsize=(8,5), color = 'deeppink', kind='bar')
graficofem.set_xticklabels(resultado_fem['REGIÃO'])

resultado_fem.set_index(['REGIÃO'], inplace=True)

print(resultado_fem)

analise_conciliadamasc = {

    'REGIÃO': ['Centro-Oeste', 'Nordeste', 'Norte', 'Sudeste', 'Sul'],

    'CRESCIMENTO MASCULINO (%)': [96.96, 79.60, 63.29, 19.09, 30.90],}

resultado_masc = pd.DataFrame(analise_conciliadamasc)

graficomasc = resultado_masc[['CRESCIMENTO MASCULINO (%)']].plot(figsize=(8,5), color = 'blue', kind='bar')
graficomasc.set_xticklabels(resultado_masc['REGIÃO'])

resultado_masc.set_index(['REGIÃO'], inplace=True)

print(resultado_masc)

conciliado = {

    'REGIÃO': ['Centro-
Oeste', 'Nordeste', 'Norte', 'Sudeste', 'Sul'],

    'CRESCIMENTO FEMININO (%)': [96.00, 109.80, 123.52, 36.59, 57.03]
,

    'CRESCIMENTO MASCULINO (%)': [96.96, 79.60, 63.29, 19.09, 30.90],
}

analise_final = pd.DataFrame(conciliado)

```

```

graficoconc = analise_final[['CRESCIMENTO FEMININO (%)', 'CRESCIMENTO MASCULINO (%)']].plot(figsize=(8,5), color = ['deeppink', 'blue'], kind='bar')
graficoconc.set_xticklabels(analise_final['REGIÃO'])

analise_final.set_index(['REGIÃO'], inplace=True)

print(analise_final)

media_fem = resultado_fem['CRESCIMENTO FEMININO (%)'].mean().round(2)
media_masc = resultado_masc['CRESCIMENTO MASCULINO (%)'].mean().round(2)

analise_mdfinal = pd.DataFrame([media_fem, media_masc])
analise_mdfinal.rename(columns={0: 'MEDIA(%)'}, inplace = True)
analise_mdfinal.rename(index={0:'FEMININO', 1:'MASCULINO'}, inplace = True)
print(analise_mdfinal)

graficomed = analise_mdfinal['MEDIA(%)'].plot(figsize=(8,5), color = ['pink', 'skyblue'], kind='bar')

print('A média do crescimento do ingresso FEMININO em ensino superior no curso de Ciências de Dados no ano de 2021 foi de:', media_fem, '%')
print('A média do crescimento do ingresso MASCULINO em ensino superior no curso de Ciências de Dados no ano de 2021 foi de:', media_masc, '%')
print('Em 2021 o ingresso foi de: {:.2f}'.format(media_fem - media_masc), '% maior para as mulheres na área de dad

```