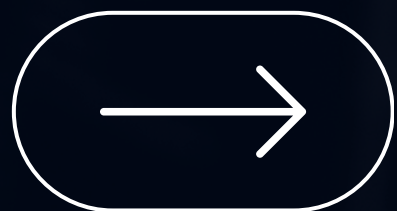


Mercado de Trabalho - Ciência de Dados



State of Data Brazil 2021

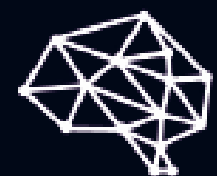
O dataset estudado fala sobre o **mercado de trabalho brasileiro** na área de dados, possuindo 2641 linhas e 356 colunas, divididas em 9 partes, desde a area demográfica, ate a parte técnica.

Previa do dataset:

id	Idade	Genero	Cargo Atual	Área de Formação
zyoz1s0ie1mva2zyoz1ybjvfvggdfz89i	32.0	Masculino	engenheiro de dados/data engineer	Computação / Engenharia de Software / Sistemas de Informação/ TI
zxwv1axhtjl3ccjczxwv0wb2q8r8r5	28.0	Feminino	cientista de dados/data scientist	Computação / Engenharia de Software / Sistemas de Informação/ TI
zuwpvmglkc46t88zuwplx92e9olkav38	46.0	Feminino	técnico	Computação / Engenharia de Software / Sistemas de Informação/ TI
zuvocreuesob1eg49awr9rxzuvocrq4z	35.0	Feminino	analista de bi/bi analyst/analytics engineer	Computação / Engenharia de Software / Sistemas de Informação/ TI
ztrtwtz05am4swq65l1wztrtwk6u4guz	27.0	Masculino	engenheiro de dados/data engineer	Outras Engenharias

Porque escolher este dataset?

- Relevância Atual: O dataset contém informações atualizadas sobre o mercado de trabalho na área de dados no Brasil em 2021, tornando-o relevante para pesquisas contemporâneas.
- Credibilidade da Fonte: A pesquisa foi realizada pela Data Hackers, uma das maiores comunidades de dados do Brasil, em colaboração com a Bain & Company, uma consultoria global respeitada.
- Potencial para Insights Originais: Dada a riqueza dos dados, existe a possibilidade de descobrir insights originais que podem contribuir para a compreensão do mercado de trabalho na área de dados.

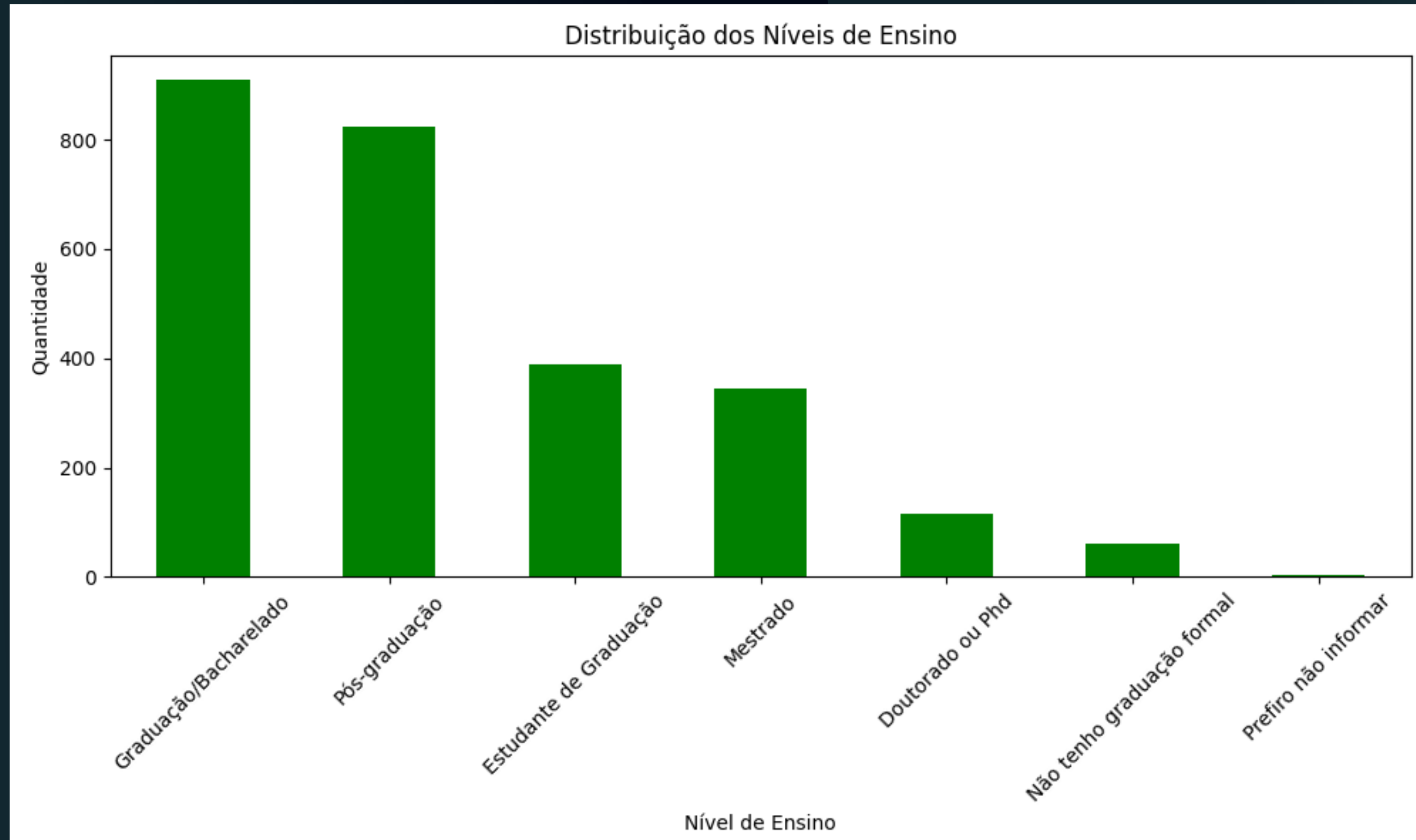


DATA HACKERS

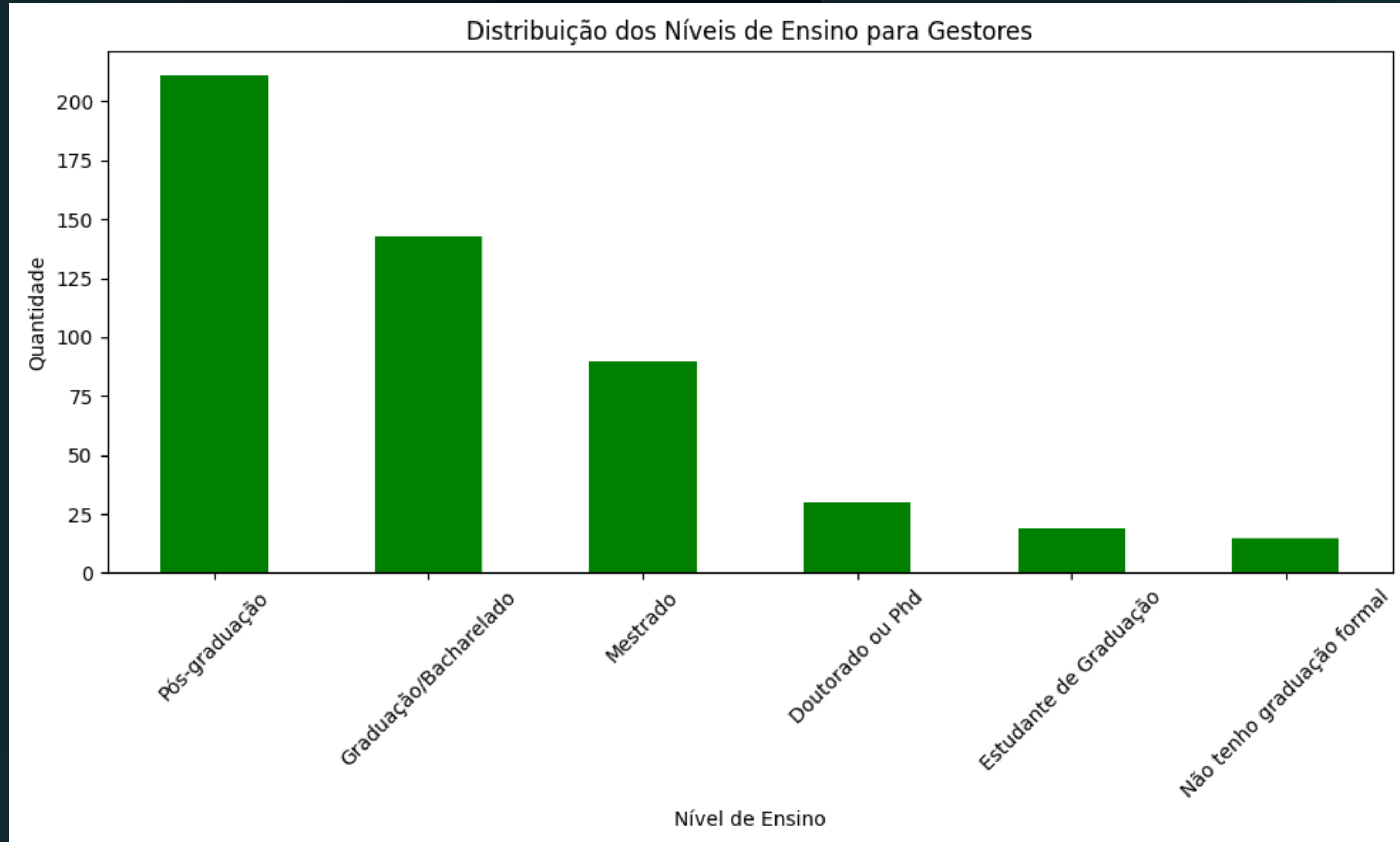
Pre-processamento – Etapas e Dificuldades

1. Renomeação das colunas
2. Substituição ou exclusão dos **NA** (Not a Number)
3. Transformação de valores que não se encaixavam como **Int** (números inteiros) ou **Float** (números quebrados)
4. **Padronização** de algumas colunas para letras minúsculas (para evitar problemas de case-sensitivity)

Escolaridade dos entrevistados



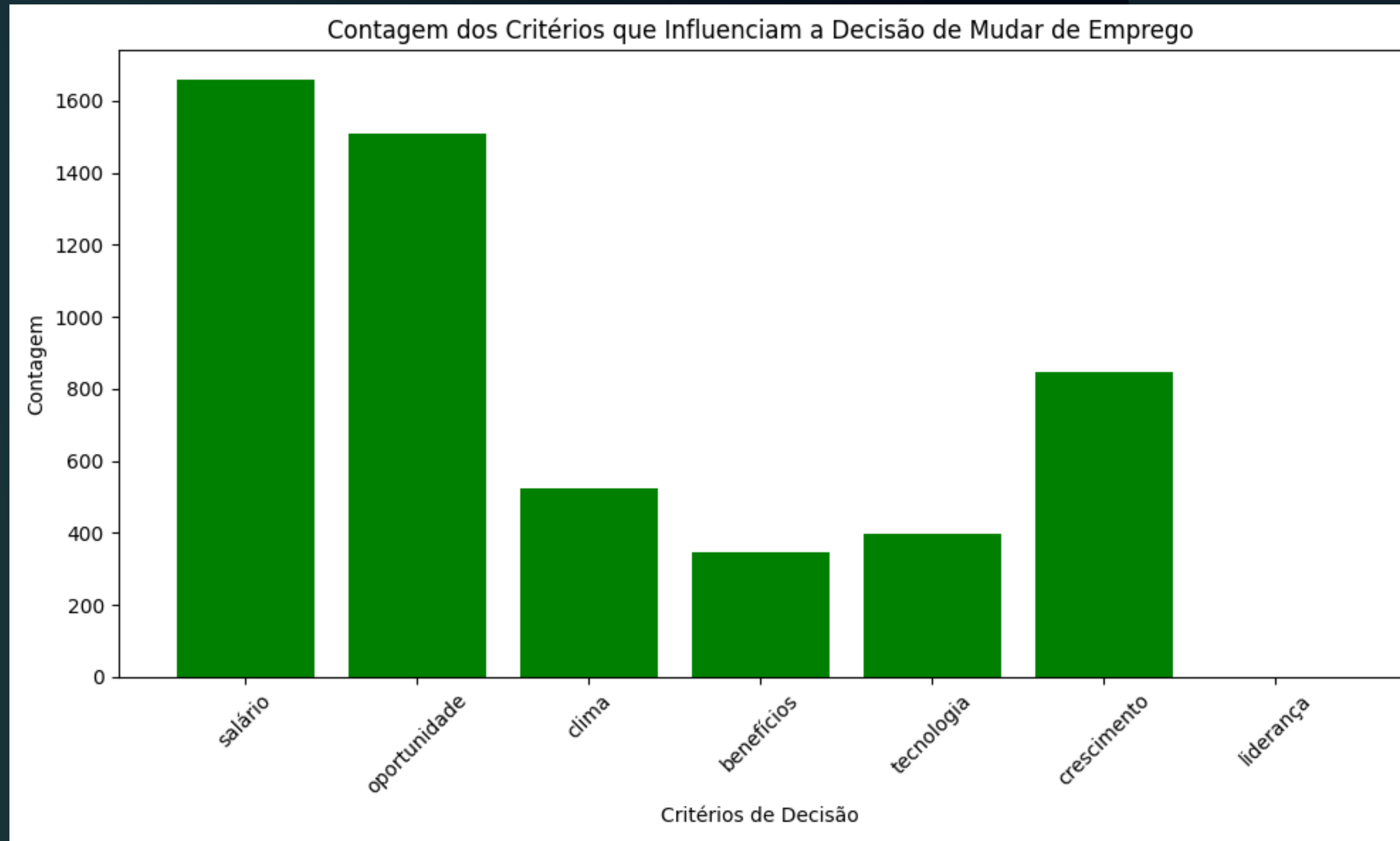
Escolaridade dos entrevistados que atuam como **gestores**



Quais cargos possuem mais
pessoas empregadas:

Quantidade de Pessoas por Cargo:	
Cientista de dados/data scientist	357
Analista de bi/bi analyst/analytics engineer	338
Analista de dados/data analyst	324
Engenheiro de dados/data engineer	300

Análises

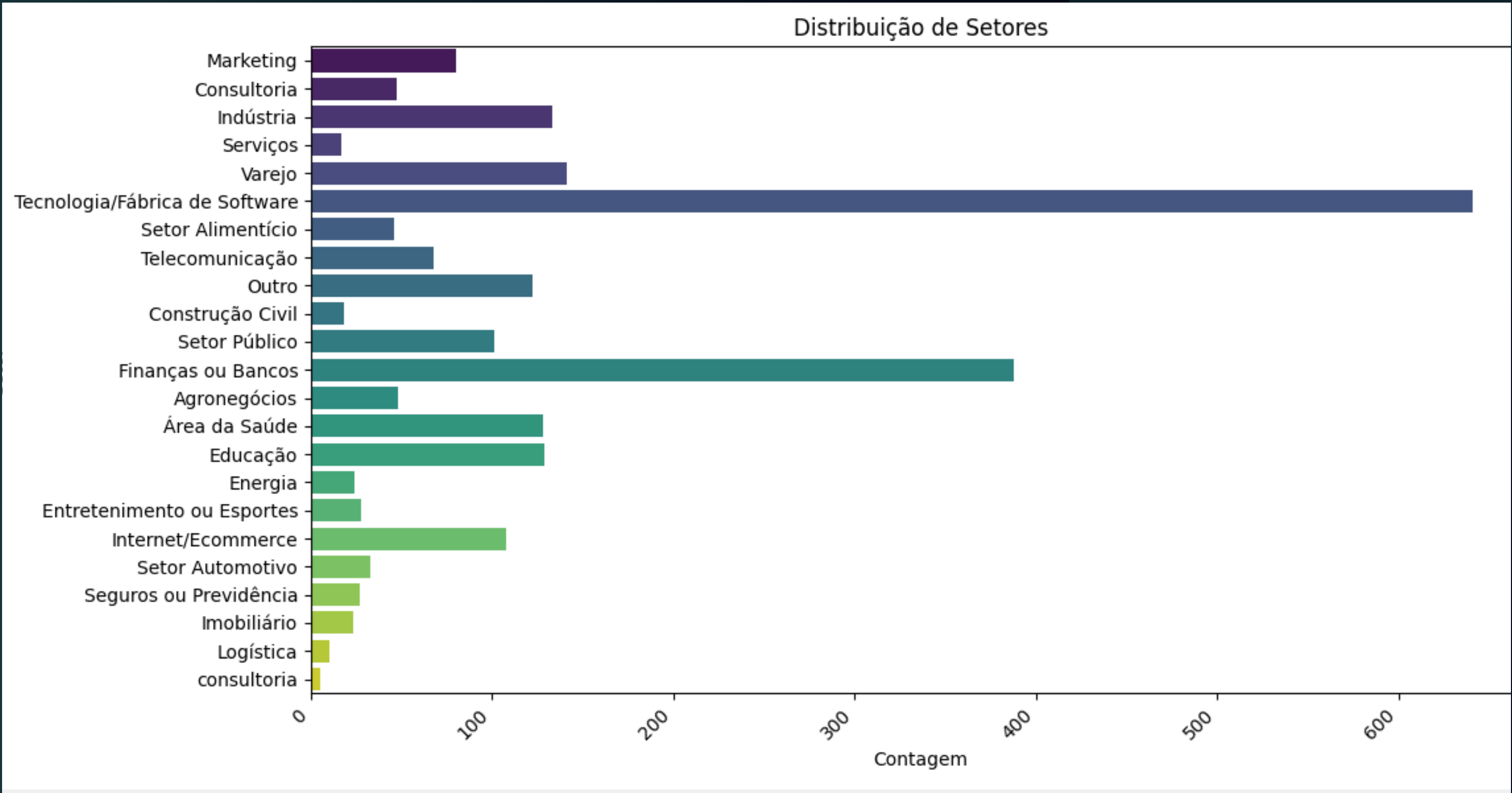


Salário: 1659 vezes
Oportunidade: 1509 vezes
Clima: 522 vezes
Benefícios: 345 vezes
Tecnologia: 399 vezes
Crescimento: 845 vezes
Liderança: 0 vezes

Analise sobre a idade dos entrevistados

Média de Idade:	31.16 anos
Mediana de Idade:	31.16 anos
Idade Mínima:	18.0 anos
Idade Máxima:	54.0 anos
Desvio Padrão de Idade	7.13 anos

Setores que mais empregam



Tecnologia/Fábrica de Software:

641

Finanças ou Bancos:

388

Varejo:

141

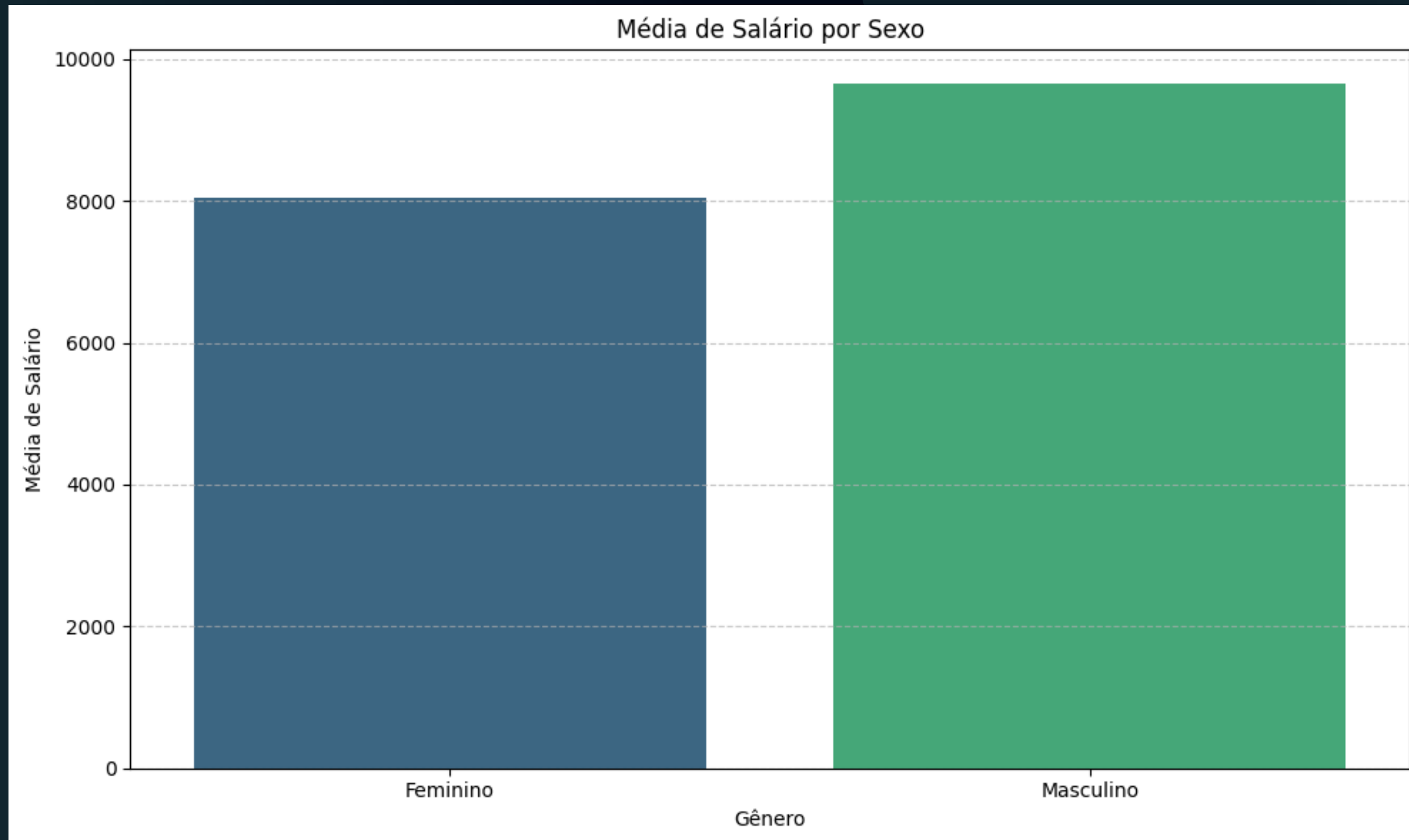
Indústria:

133

Área da Saúde:

128

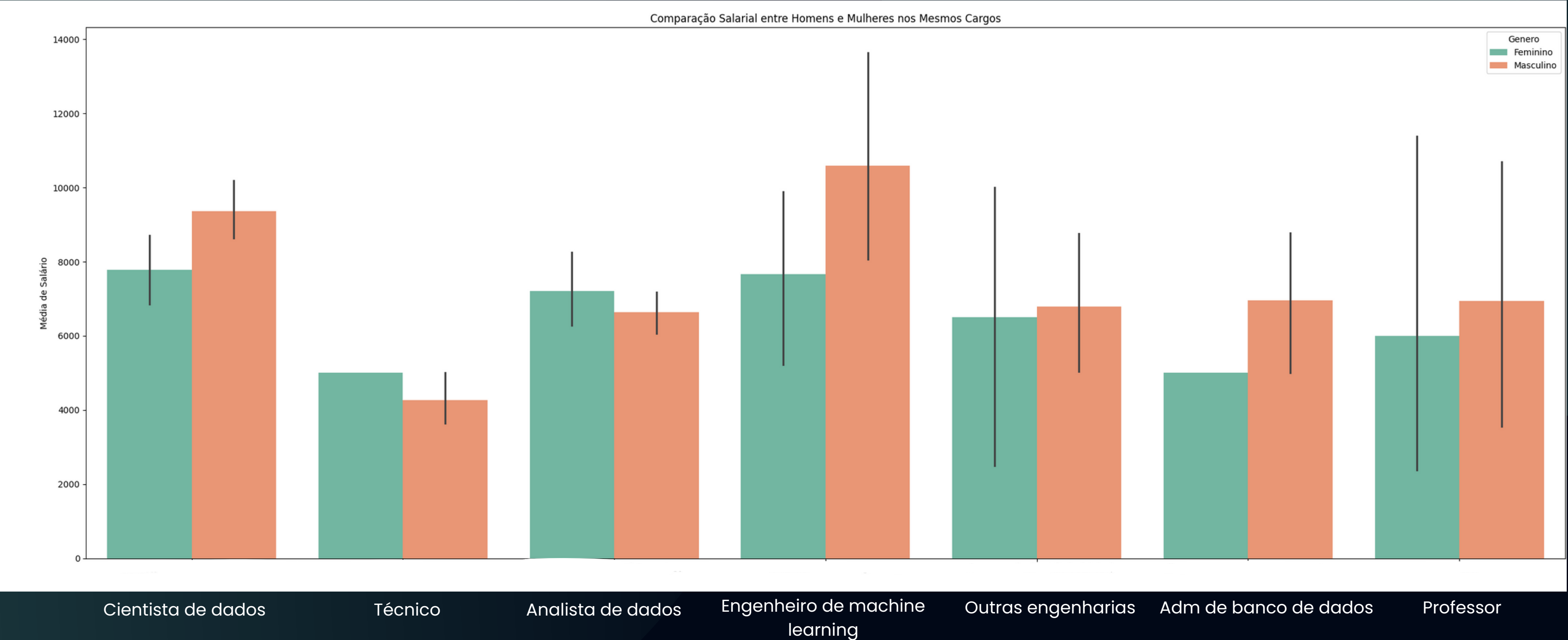
Média e **diferença** salarial entre homens e mulheres



Média e diferença salarial entre
homens e mulheres

Gênero	Quantidade	Média de Salário
Feminino	436	8036.048165
Masculino	1924	9659.544699
Diferença Percentual do salário		16.8071745055%
Diferença Percentual da quantidade		126.1017%

Média e diferença salarial entre homens e mulheres



REFERÊNCIAS BIBLIOGRÁFICAS

1. Data Hackers. Kaggle: Your Machine Learning and Data Science Community, 2021. Datasets. Disponível em: . Acesso em: 05 de out. de 2023
2. LVES-MAZZOTI, A. J.; GEWANDSZNAJER, F. O método nas ciências naturais e sociais: pesquisa quantitativa e qualitativa. 2.ed. São Paulo: Pioneira, 1998
3. BURREL, G.; MORGAN, G. Sociological paradigms and organizational analysis. Londres: Heinemann Books, 1979

Link de acesso aos códigos: https://github.com/ErickCoutinho/LAB_CD