

Análise do Mercado de Dados no Brasil: Relatório de Ciência de Dados

J. P. Baganha, Pedro Henrique Carneiro Silva, Luís Cláudio Sarmento, Paulo Ricardo Silva Fonseca, Lívio Marcel Cambraia

5º Período de Ciência da Computação

Prof. Dr. Isaac Elgrably

Disciplina Ciência de Dados

1 Escola de Negócios, Tecnologia e Inovação - ARGO, Centro Universitário do Estado do Pará, Brasil.

Belém, Pará. 2025.

1 INTRODUÇÃO

A atividade desenvolvida tem como base os dados da pesquisa *State of Data Brasil 2023*, realizada pela comunidade Data Hackers em parceria com a Bain & Company. Esse levantamento reúne mais de 5.200 respostas e fornece informações sobre perfil demográfico, situação profissional, escolaridade, experiência, distribuição regional, remuneração e diversidade de profissionais da área de dados no Brasil.

O objetivo do trabalho foi aplicar técnicas de ciência de dados, com foco em análise exploratória e visualização, para examinar os seguintes aspectos:

- Situação de trabalho por gênero e por etnia;
- Níveis hierárquicos por etnia, gênero e região;

- Distribuição de cargos por gênero;
- Faixa salarial por gênero nos principais cargos;
- Proporção de pessoas com deficiência por região;
- Nível de ensino por cargo e por região.

As análises foram feitas com base no processamento e limpeza do conjunto de dados original, seguido da criação de variáveis agregadas, tabelas cruzadas e representações visuais, a partir das quais foi possível extrair informações comparativas. A seguir, são apresentados os métodos utilizados, os resultados obtidos e sugestões para aprofundamento dos estudos.

2 METODOLOGIA

O trabalho foi desenvolvido em ambiente Colab, utilizando a linguagem Python com as bibliotecas Pandas, Matplotlib e Seaborn. As etapas metodológicas seguiram a seguinte ordem:

A. Leitura e carregamento dos dados

O conjunto de dados original foi lido a partir de um arquivo CSV fornecido na atividade, que foi convertido em um arquivo de planilha Excel. A leitura foi realizada com **pandas.read_excel()**.

B. Seleção e renomeação de colunas

Foram selecionadas colunas relevantes para as análises exigidas, com renomeação dos nomes originais para facilitar a manipulação.

C. Limpeza de dados

As seguintes ações foram aplicadas:

- a. Remoção de valores ausentes (null).
- b. Exclusão de categorias genéricas como “outros”.
- c. Filtro para manter apenas os gêneros “Masculino” e “Feminino”.

D. Criação de categorias agregadas

Cargos foram agrupados em categorias como “Cientista de Dados”, “Engenheiro de Dados”, “Analista de Dados”, “Analista de BI” e “Outros”.

E. Análises realizadas

- a. Distribuição geral da situação de trabalho.
- b. Situação de trabalho por gênero.
- c. Situação de trabalho por etnia.
- d. Top 10 cargos por gênero.
- e. Nível hierárquico por etnia.
- f. Tabela cruzada e heatmap da situação de trabalho por gênero e etnia.
- g. Níveis hierárquicos por região, etnia e gênero.
- h. Faixa salarial por cargo e gênero, com conversão para valor médio numérico.
- i. Proporção de profissionais com deficiência (PcD) por região.
- j. Nível de ensino por cargo e por região.

F. Criação de visualizações

Para cada análise, foram utilizadas técnicas de visualização como:

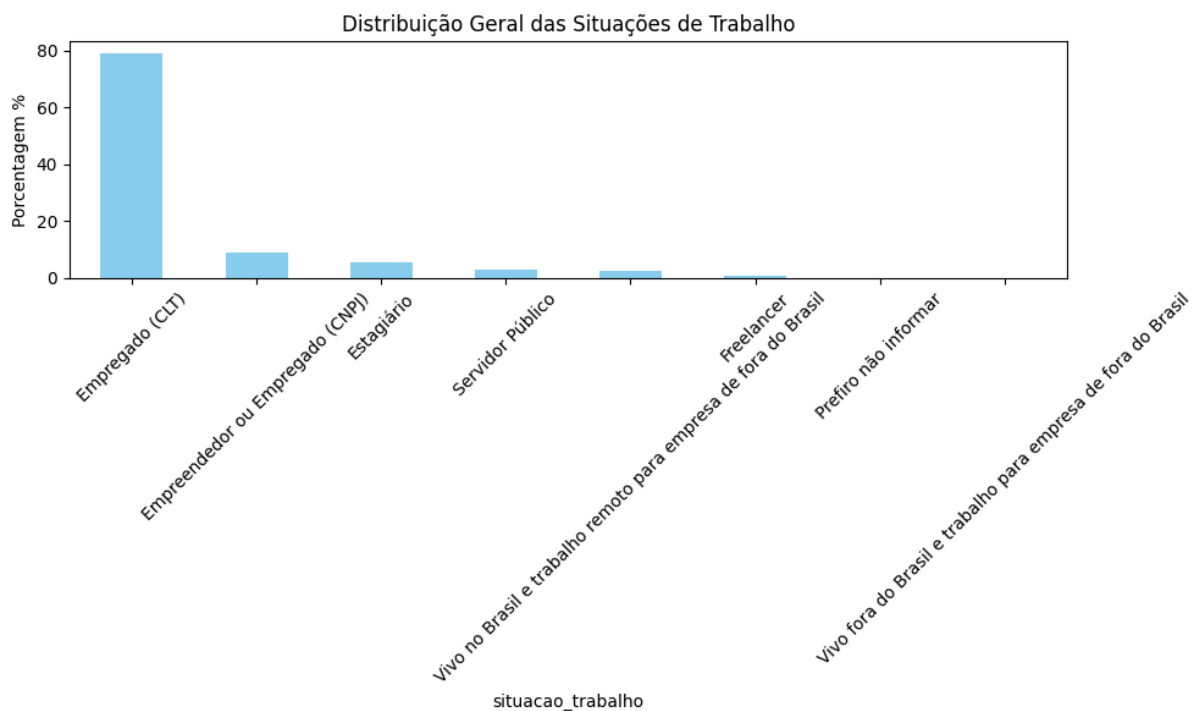
- a. Gráficos de barra (simples e empilhados).
- b. Countplots segmentados por categoria.
- c. Boxplots com faixa salarial.
- d. Heatmap com porcentagem por grupo.

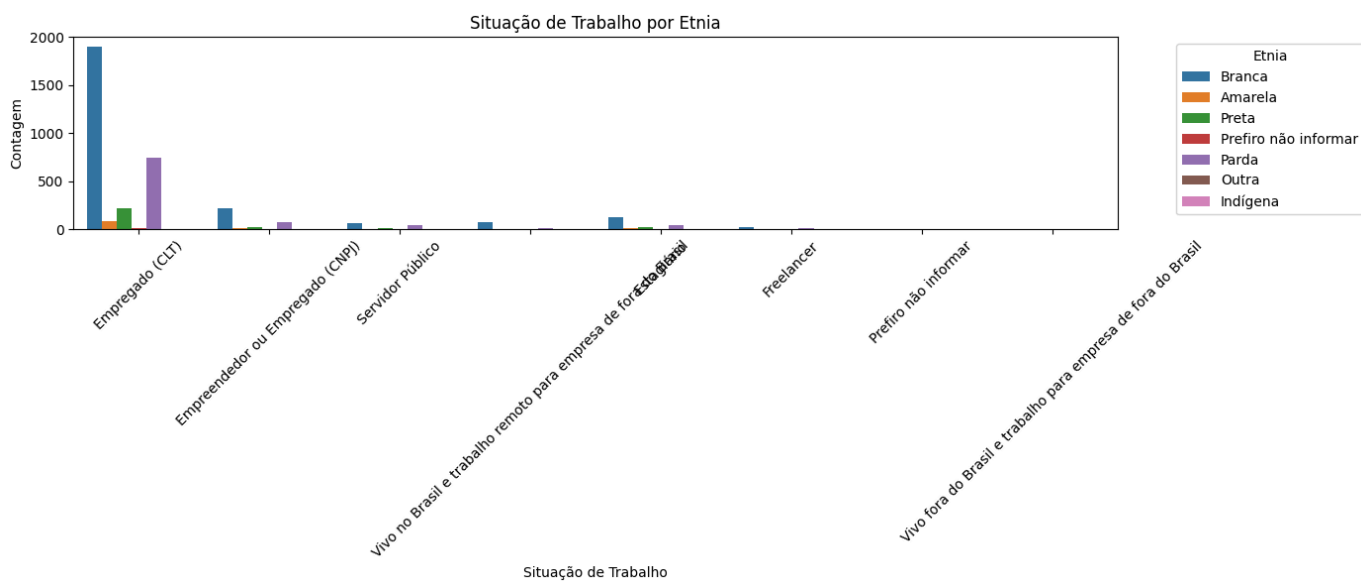
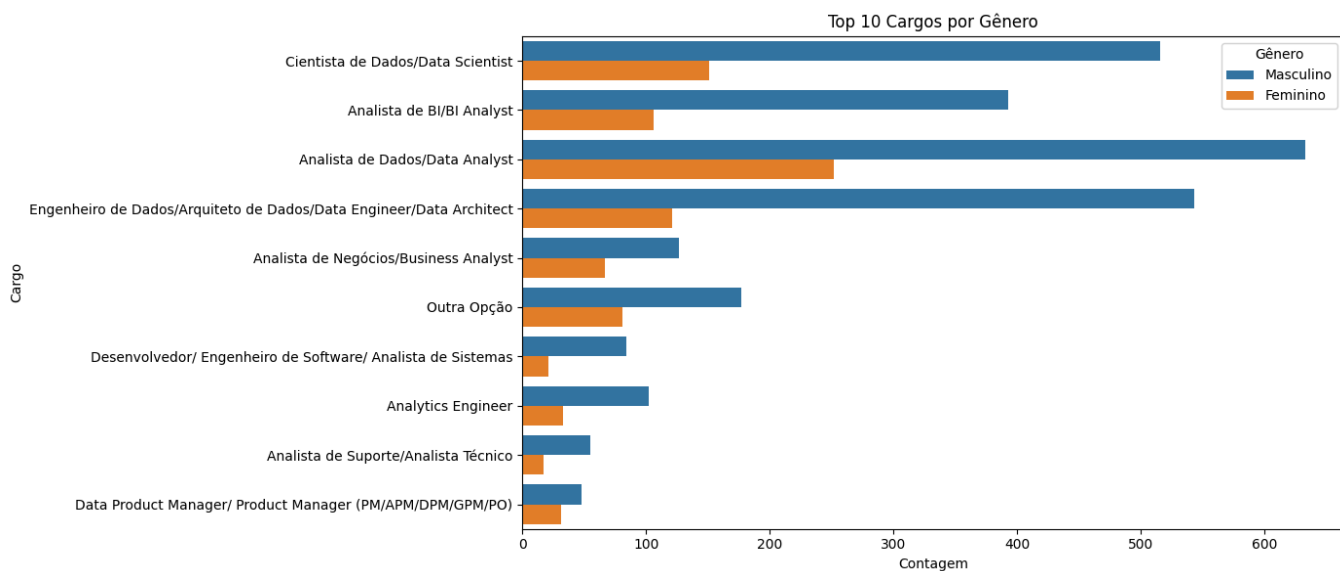
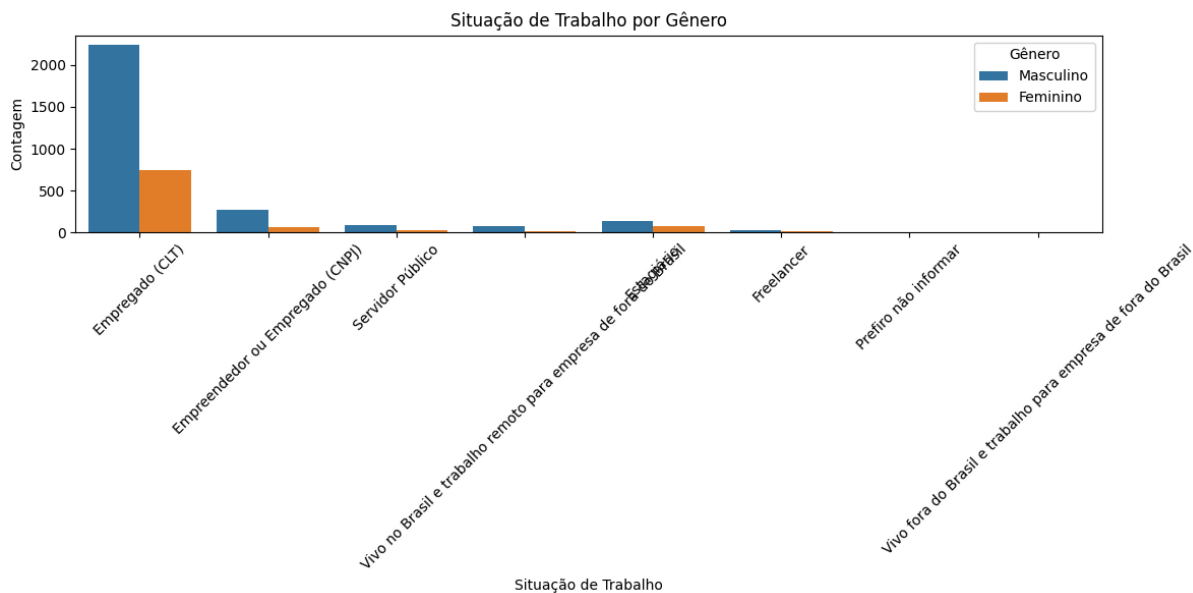
G. Padronização e ordenação

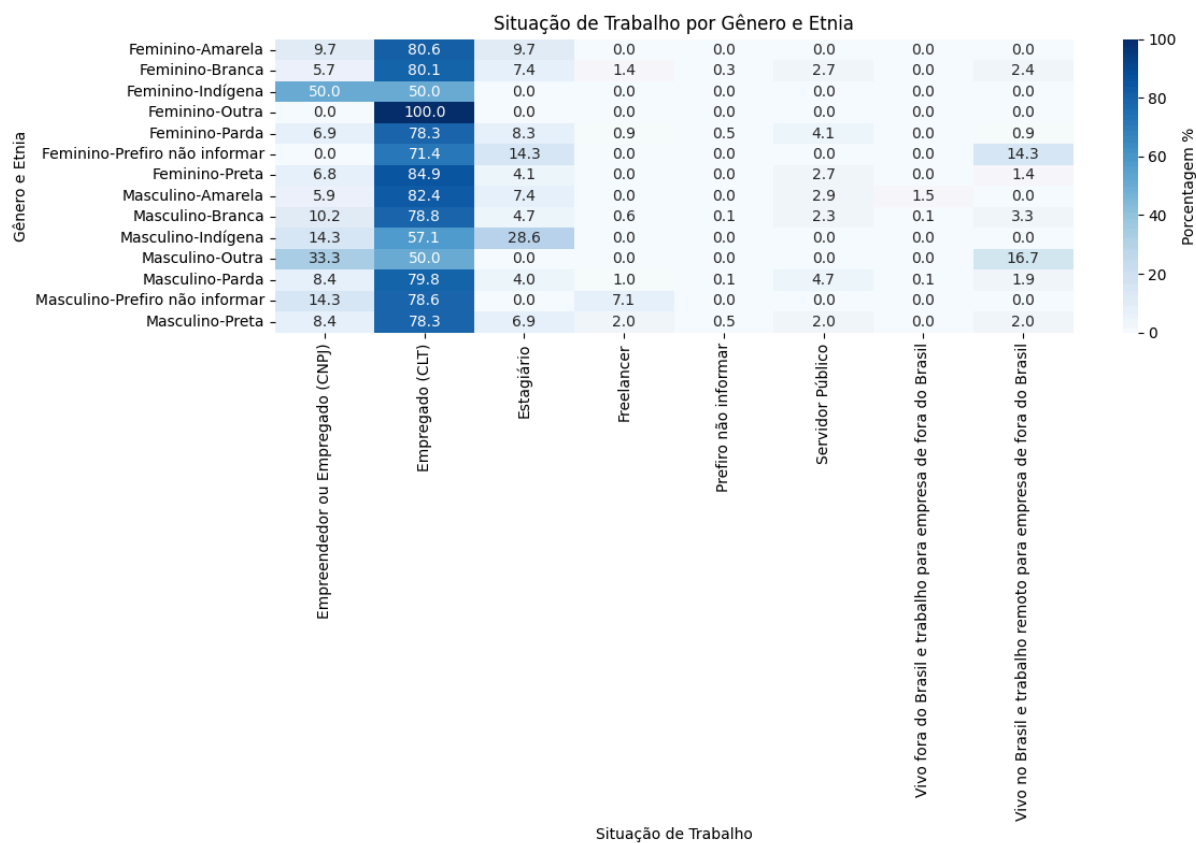
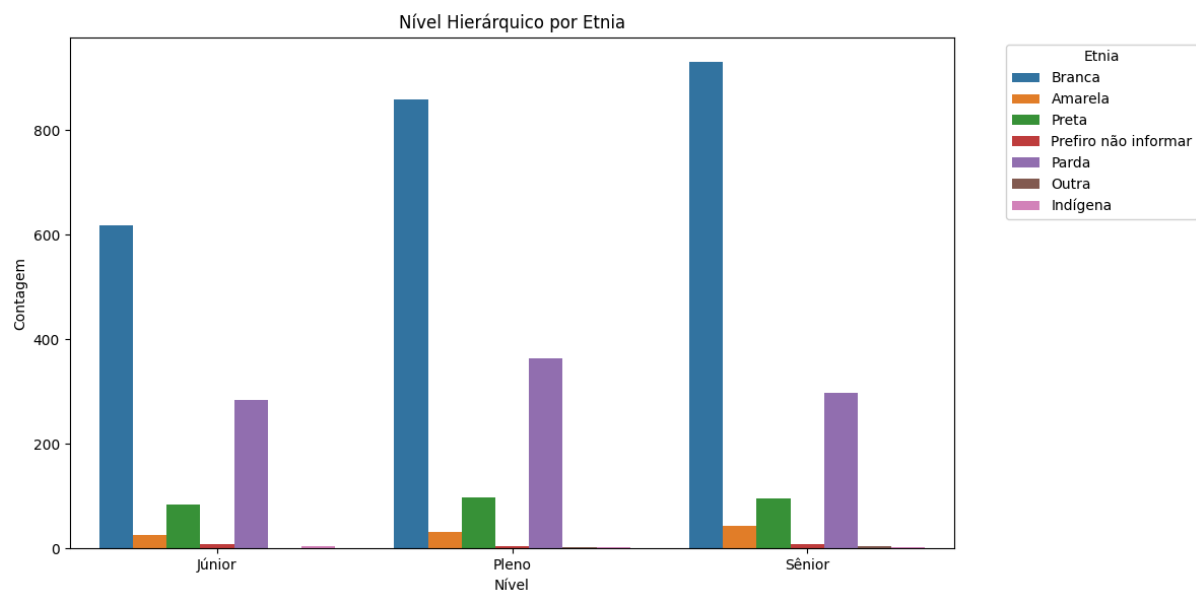
Algumas variáveis categóricas foram ordenadas de forma manual para garantir coerência visual e facilitar comparações (como faixa salarial e nível hierárquico).

Essa metodologia permitiu realizar cruzamentos entre múltiplas variáveis, respeitando os recortes exigidos e garantindo consistência no tratamento dos dados.

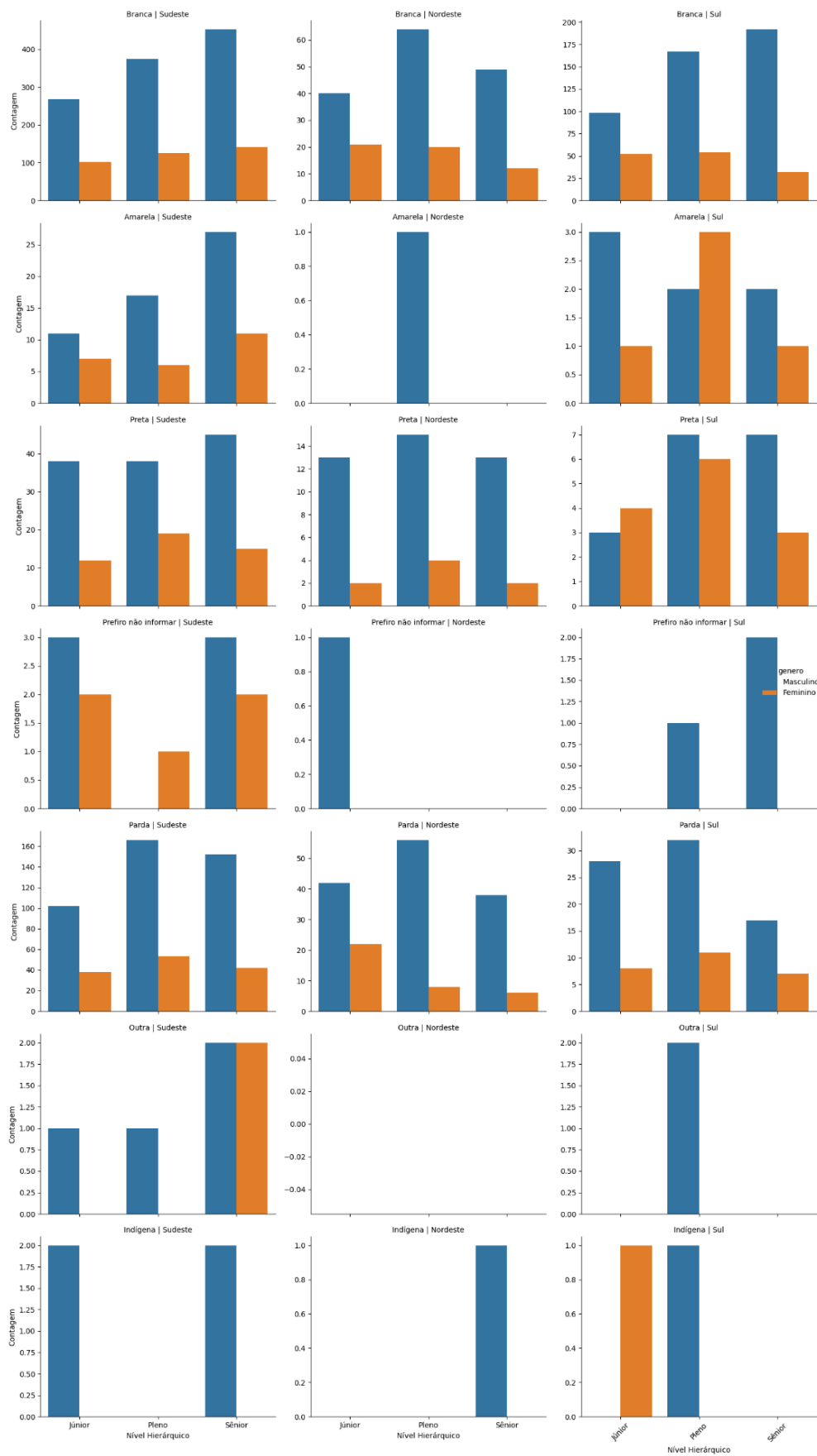
3 VISUALIZAÇÕES

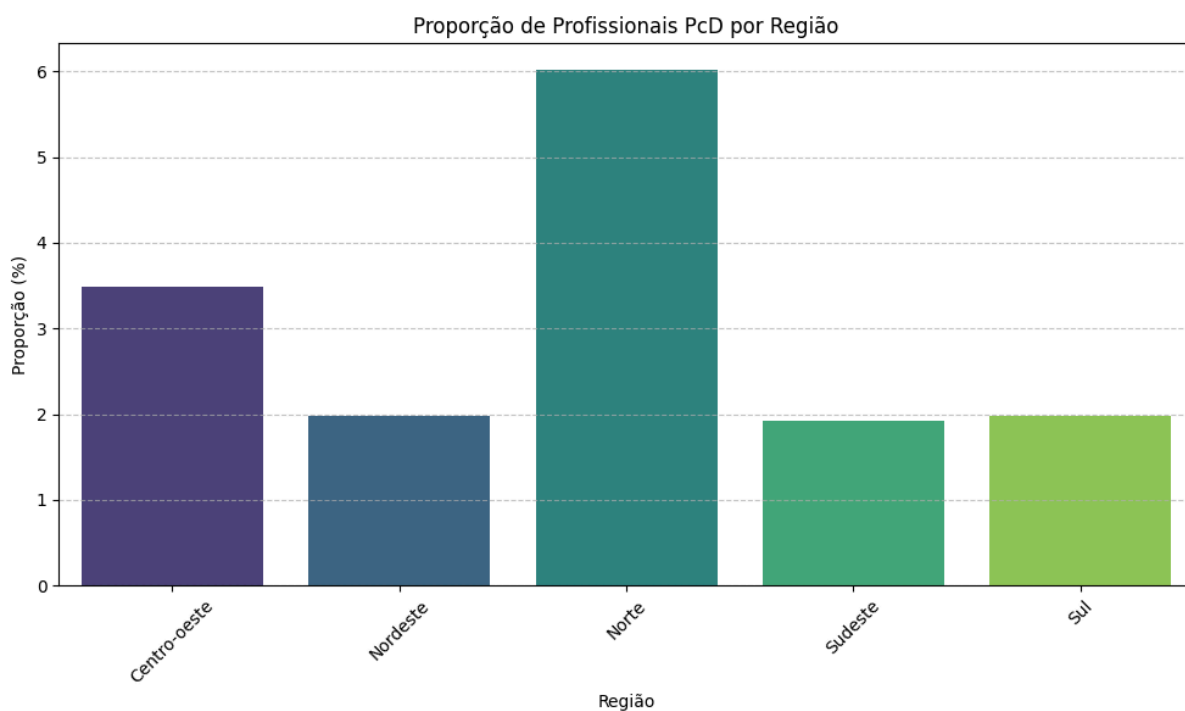
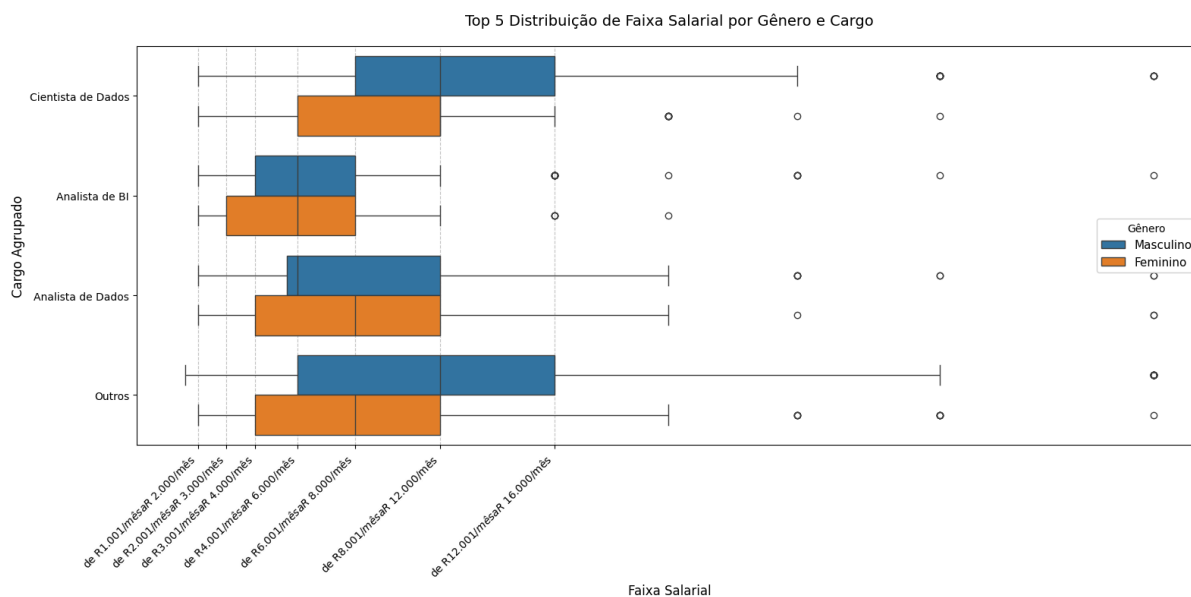






Distribuição de Níveis Hierárquicos por Gênero, Etnia e Região





4 INSIGHTS E TOMADAS DE DECISÃO

A partir das análises realizadas sobre os dados da pesquisa *State of Data Brasil 2023*, foram identificados os seguintes pontos de atenção, organizados por temática:

4.1 Situação de Trabalho por Gênero e Etnia

- Homens e mulheres apresentam padrões semelhantes em termos de alocação por categoria de trabalho, mas há maior concentração masculina nas categorias com vínculo formal de emprego.

- Pessoas negras e pardas têm proporção mais elevada de atuação como autônomos ou em modalidades menos formais, em comparação com pessoas brancas.
- A combinação de gênero e etnia mostra que mulheres negras possuem menor proporção de trabalho formal, indicando um padrão cruzado de desigualdade.

Tomada de decisão: Adotar políticas específicas de inclusão voltadas para mulheres negras em programas de contratação e retenção pode reduzir a disparidade identificada.

4.2 Níveis Hierárquicos por Região, Gênero e Etnia

- Pessoas brancas ocupam, proporcionalmente, mais posições de nível sênior.
- No Norte e Nordeste, a presença de profissionais em níveis mais altos é menor em todas as etnias.
- Mulheres, independentemente da região, têm menor representação em cargos seniores.

Tomada de decisão: Incentivar programas de aceleração de carreira em regiões periféricas e com foco em diversidade pode ampliar o acesso a cargos de liderança técnica.

4.3 Distribuição de Cargos por Gênero

- A maioria dos cargos mais comuns da área possui predominância masculina, com exceção do cargo de Analista de BI, que apresenta maior equilíbrio de gênero.

Tomada de decisão: Mapear os gargalos de entrada de mulheres em áreas técnicas e promover capacitação voltada para cargos com menor representatividade feminina.

4.4 Faixa Salarial por Gênero nos Principais Cargos

- Mulheres possuem menor mediana salarial em todos os cargos analisados, mesmo ocupando posições semelhantes às dos homens.

Tomada de decisão: Estabelecer revisões periódicas de faixas salariais com recorte de gênero pode apoiar a construção de planos de remuneração mais equilibrados.

4.5 Proporção de Profissionais com Deficiência por Região

- As regiões Norte e Nordeste apresentam maior percentual relativo de profissionais PcD em comparação com as demais.

Tomada de decisão: Investigar as condições de contratação e permanência dos profissionais PcD nessas regiões pode direcionar ações de fortalecimento de inclusão estruturada.

4.6 Nível de Ensino por Cargo e Região

- Regiões Sul e Sudeste concentram a maior proporção de profissionais com pós-graduação.
- Nas demais regiões, predomina o ensino superior como nível máximo de formação.

Tomada de decisão: Apoiar financeiramente o acesso à pós-graduação em regiões com menor proporção pode reduzir desigualdades na formação acadêmica do setor.