



UNIVERSIDADE VEIGA DE ALMEIDA

CAMILE PIMENTA MONTEIRO - 1240201525

GABRIEL NATHAN DOS SANTOS PIRES - 1230118841

ITALLO VIDAL - 1230114462

LEONARDO PATROCINIO SILVA - S

**A1 – ANÁLISE DE NEGÓCIOS COM USO DE BIG DATA**

**RIO DE JANEIRO - RJ**

**2025**

CAMILE PIMENTA MONTEIRO - 1240201525

GABRIEL NATHAN DOS SANTOS PIRES - 1230118841

ITALLO VIDAL - 1230114462

LEONARDO PATROCINIO SILVA - 1230206925

### **A1 – Análise De Negócios Com Uso De Big Data**

Trabalho apresentado à Universidade Veiga de Almeida  
como requisito para a disciplina de Análise de Negócios  
com Uso de Big Data.

Orientador(a): Denis Gonçalves Cople

RIO DE JANEIRO - RJ

2025

## RESUMO

Este estudo aplicou técnicas de análise de big data em uma base com dados de idade, gênero, escolaridade, cargo, tempo de experiência e salário. Os resultados permitem observar tendências salariais de acordo com características individuais e profissionais, revelando correlações relevantes entre nível educacional, experiência e remuneração. A análise fornece evidências que podem apoiar tomadas de decisão em políticas de recursos humanos, gestão de carreiras e estudos sobre equidade salarial.

**Palavras-chave:** Big Data; Análise de dados; Salário; Escolaridade.

## **ABSTRACT**

This study applied big data analysis techniques to a database containing data on age, gender, education, position, experience, and salary. The results allow us to observe salary trends according to individual and professional characteristics, revealing relevant correlations between educational level, experience, and compensation. The analysis provides evidence that can support decision-making in human resources policies, career management, and studies on pay equity.

**Keywords:** Big Data; Data Analysis; Salary; Education.

## LISTA DE FIGURAS

Figura 1 - Trecho de código de extração dos dados.....	9
Figura 2 - Trecho de código para transformação dos dados.....	9
Figura 3 - Trecho de código para carregamento dos dados.....	10
Figura 4 - Salário médio por faixa etária.....	11
Figura 5 - Quantidade de Pessoas por Emprego com educação: Bachelor's Degree.....	11
Figura 6 - Quantidade de Pessoas por Emprego com educação: PhD.....	12
Figura 7 - Quantidade de Pessoas por Emprego com educação: High School.....	12
Figura 8 - Quantidade de Pessoas por Emprego com educação: Master's.....	12
Figura 9 - Distribuição por gênero com educação: Bachelor's Degree.....	13
Figura 10 - Distribuição por gênero com educação: High School.....	13
Figura 11 - Distribuição por gênero com educação: Master's Degree.....	13
Figura 12 - Distribuição por gênero com educação: PhD.....	14
Figura 13 - Salário Médio dos Empregos de Nível PHD.....	14
Figura 14 - Salário Médio dos Empregos de Nível Mestrado.....	15
Figura 15 - Salário Médio dos Empregos de Nível Bacharelado.....	15
Figura 16 - Salário Médio dos Empregos de Nível Bacharelado.....	16

## **LISTA DE ABREVIATURAS E SIGLAS**

ETL	Extract, Transform, Load (Extrair, Transformar, Carregar)
CSV	Comma-Separated Values (Valores Separados por Vírgulas)

## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>7</b>
<b>2. DESCRIÇÃO DA FONTE DE DADOS.....</b>	<b>8</b>
<b>3. AÇÕES DE ETL.....</b>	<b>9</b>
<b>4. GRÁFICOS E RESULTADOS ESTATÍSTICOS.....</b>	<b>11</b>
4.1. Salário médio por idade.....	11
4.2. Quantidade de pessoas por Emprego por nível de educação.....	11
4.3. Quantidade de pessoas por Gênero em nível de educação.....	13
4.4. Salário médio por cargo em nível de educação.....	14
<b>REFERÊNCIAS.....</b>	<b>15</b>

## **1. INTRODUÇÃO**

O presente trabalho realiza uma análise exploratória de dados sobre características profissionais e salariais de indivíduos. A base de conhecimento utilizada contém informações referentes à idade (Age), gênero (Gender), nível educacional (Education Level), cargo ocupado (Job Title), tempo de experiência (Years of Experience) e remuneração (Salary). O Objetivo é identificar padrões e relações entre variáveis demográficas, formação acadêmica, trajetória profissional e salários, fornecendo insumos para compreender como diferentes fatores influenciam a remuneração no mercado de trabalho.



## **2. DESCRIÇÃO DA FONTE DE DADOS**

### **2.1. Objetivo**

A base de dados tem como finalidade permitir a análise de fatores que influenciam a remuneração de profissionais, considerando aspectos demográficos e acadêmicos. O conjunto de informações possibilita explorar correlações entre idade, gênero, nível de escolaridade, experiência profissional e salários.

### **2.2. Domínio**

Os dados estão inseridos no domínio de recursos humanos e mercado de trabalho, sendo aplicáveis a estudos sobre desenvolvimento profissional, gestão de pessoas e políticas de equidade salarial.

### **2.3. Origem**

A base utilizada foi disponibilizada para fins acadêmicos, contendo dados estruturados com as seguintes variáveis:

- **Age:** idade do indivíduo
- **Gender:** gênero declarado
- **Education Level:** nível de escolaridade alcançado
- **Job Title:** cargo ocupado
- **Years of Experience:** anos de experiência profissional
- **Salary:** remuneração correspondente

### **2.4. Endereços**

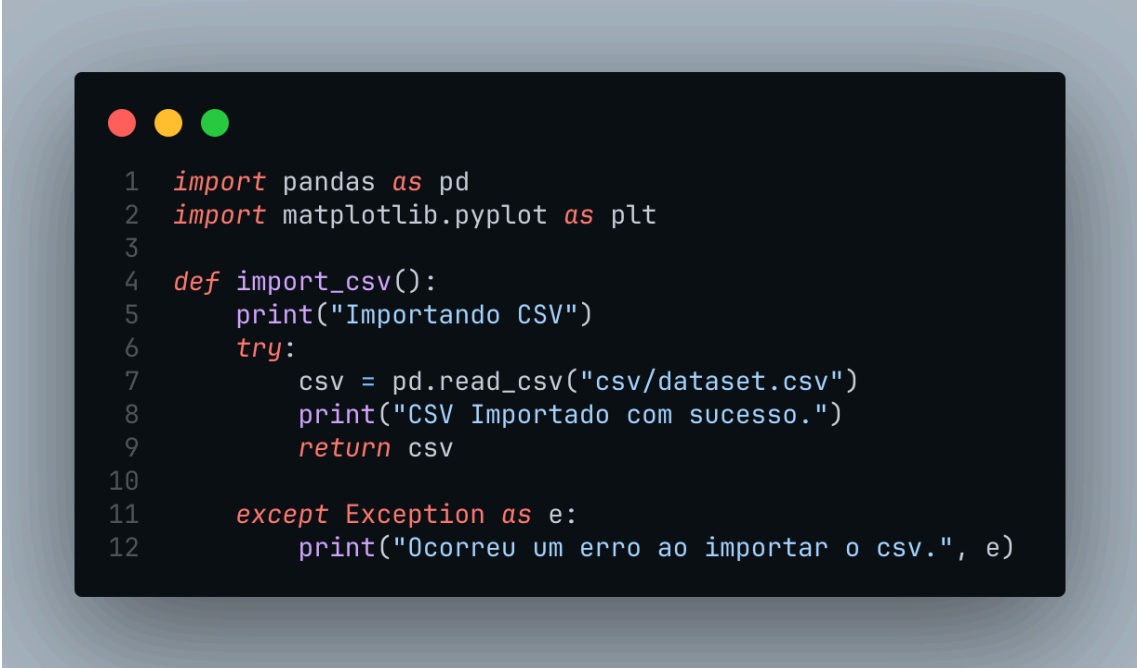
O conjunto de dados foi fornecido em formato digital (.csv) acessado localmente no ambiente de desenvolvimento para aplicação dos processos de ETL e posterior estatística.

### 3. AÇÕES DE ETL

#### 3.1. Extract

Neste trecho fazemos a coleta de dados.

Figura 1 - Trecho de código de extração dos dados



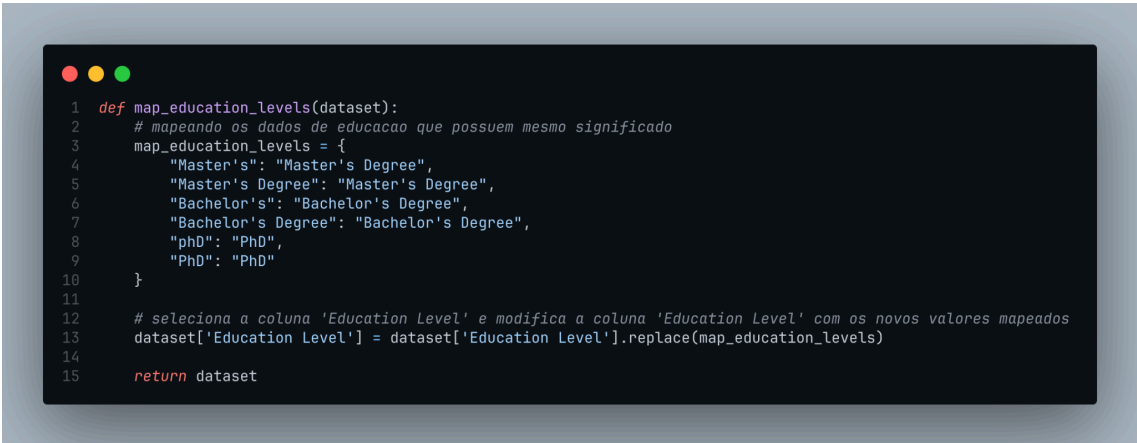
```
1 import pandas as pd
2 import matplotlib.pyplot as plt
3
4 def import_csv():
5     print("Importando CSV")
6     try:
7         csv = pd.read_csv("csv/dataset.csv")
8         print("CSV Importado com sucesso.")
9         return csv
10
11     except Exception as e:
12         print("Ocorreu um erro ao importar o csv.", e)
```

Fonte: Autor, 2025.

#### 3.2. Transform

Neste trecho fazemos o mapeamento das colunas do dataset para a análise de dados, devido a discrepâncias na forma como estão escritas ou significados ambíguos.

Figura 2 - Trecho de código para transformação dos dados



```
1 def map_education_levels(dataset):
2     # mapeando os dados de educacao que possuem mesmo significado
3     map_education_levels = {
4         "Master's": "Master's Degree",
5         "Master's Degree": "Master's Degree",
6         "Bachelor's": "Bachelor's Degree",
7         "Bachelor's Degree": "Bachelor's Degree",
8         "phD": "PhD",
9         "PhD": "PhD"
10    }
11
12    # seleciona a coluna 'Education Level' e modifica a coluna 'Education Level' com os novos valores mapeados
13    dataset['Education Level'] = dataset['Education Level'].replace(map_education_levels)
14
15    return dataset
```

Fonte: Autor, 2025.

### 3.3. Load

Este trecho apresenta as opções de carregamento de dados e, por fim, salva os dados no diretório “outputs/”.

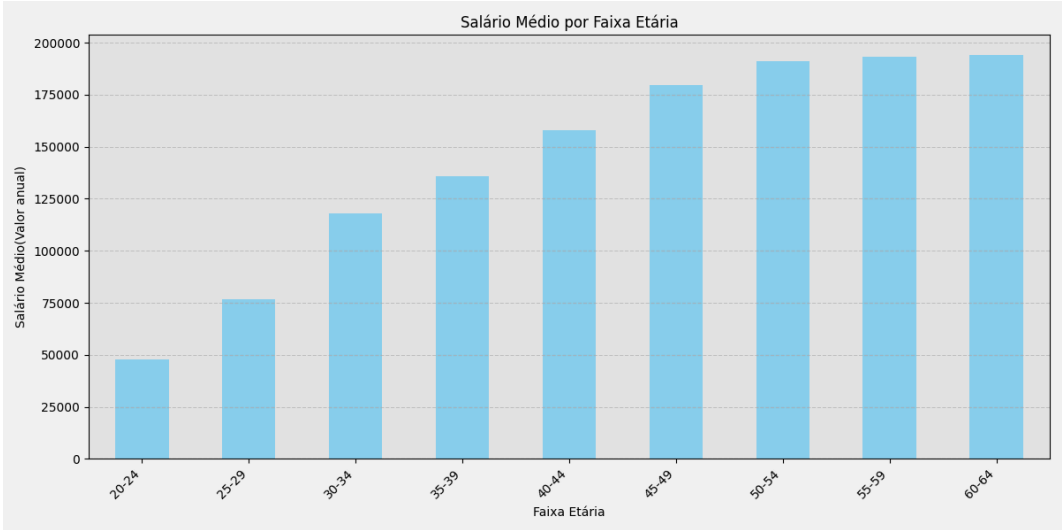
Figura 3 - Trecho de código para carregamento dos dados

```
1 import os
2 from time import sleep
3 import pandas as pd
4 from utils.map_education_levels import map_education_levels
5 from utils.clear_terminal import clear_terminal
6 from utils.menu import menu
7 from utils.education_and_job import analyze_education_vs_job
8 from utils.import_csv import import_csv
9 from utils.gender_and_education import analyze_gender_and_education
10 from utils.salary_by_age import salary_by_age
11
12 if __name__ == "__main__":
13
14     dataset: pd.DataFrame = map_education_levels(import_csv())
15
16     while True:
17
18         clear_terminal()
19
20         menu()
21
22         option = input("O que deseja fazer? ")
23
24         clear_terminal()
25
26         match option:
27             case "1":
28                 salary_by_age(dataset)
29             case "2":
30                 analyze_education_vs_job(dataset)
31             case "3":
32                 analyze_gender_and_education(dataset)
33             case "4":
34                 print("[4] Em desenvolvimento...")
35             case "0":
36                 print("Saindo...")
37                 break
38             case _:
39                 print("Opção inválida. Tente novamente.")
40                 continue
41
42     sleep(2)
```

4. GRÁFICOS E RESULTADOS ESTATÍSTICOS

4.1. Salário médio por idade

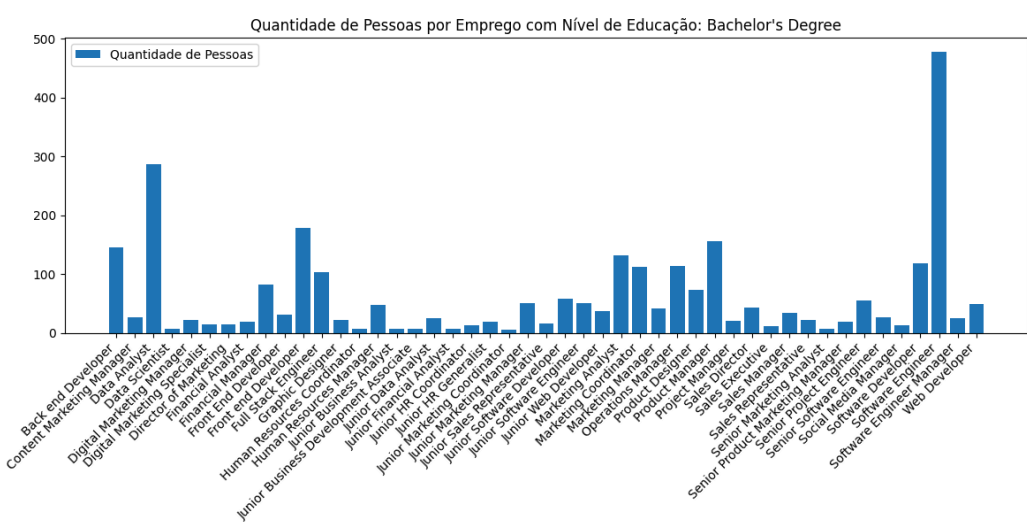
Figura 4 - Salário médio por faixa etária



Fonte: Autor, 2025.

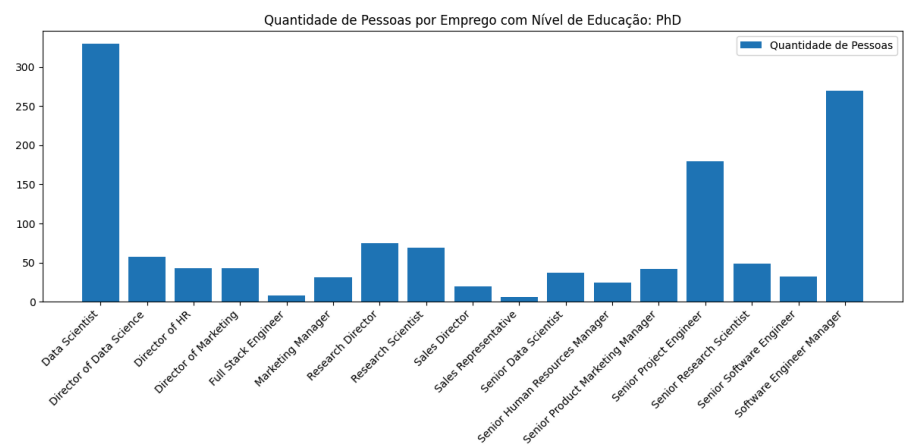
4.2. Quantidade de pessoas por Emprego por nível de educação

Figura 5 - Quantidade de Pessoas por Emprego com educação: Bachelor's Degree



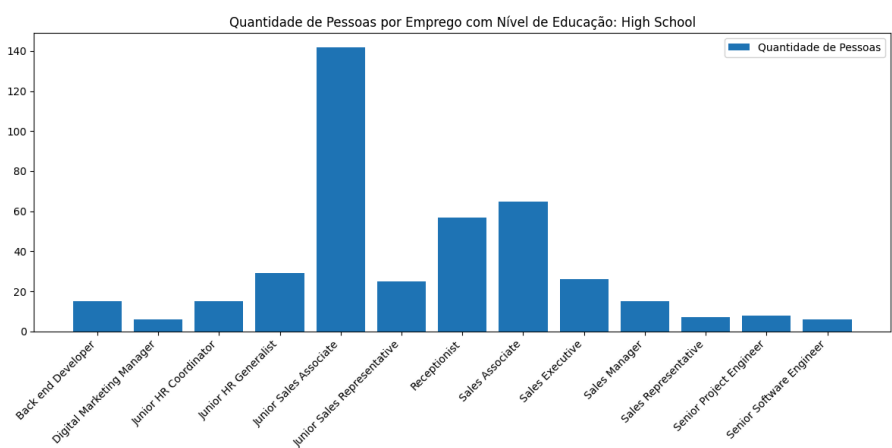
Fonte: Autor, 2025.

Figura 6 - Quantidade de Pessoas por Emprego com educação: PhD



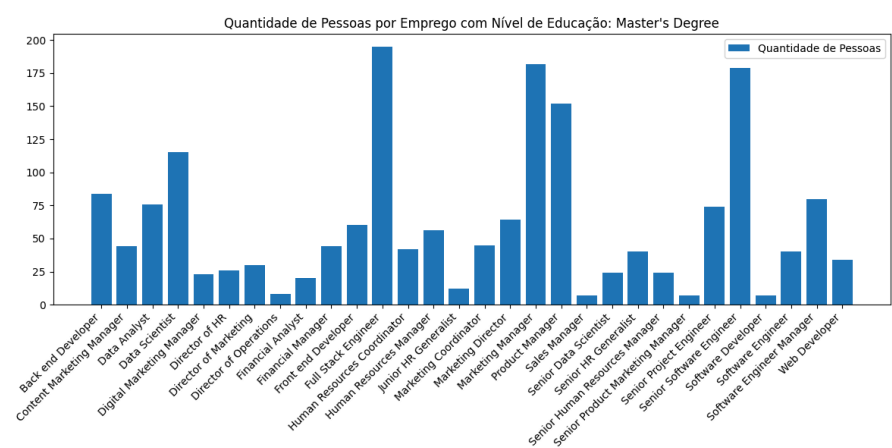
Fonte: Autor, 2025.

Figura 7 - Quantidade de Pessoas por Emprego com educação: High School



Fonte: Autor, 2025.

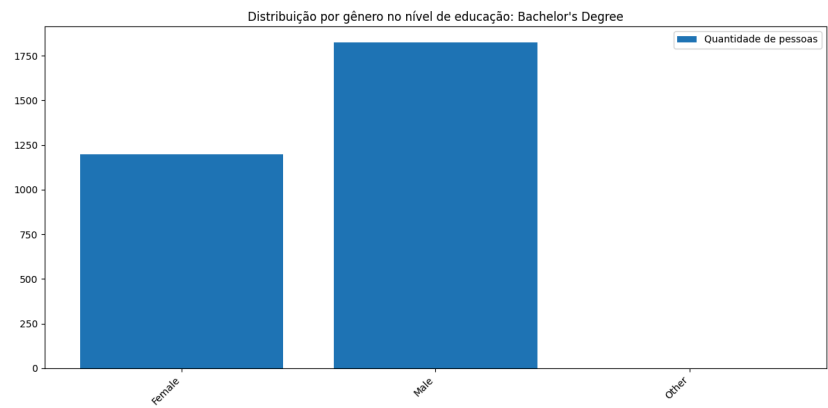
Figura 8 - Quantidade de Pessoas por Emprego com educação: Master's



Fonte: Autor, 2025..

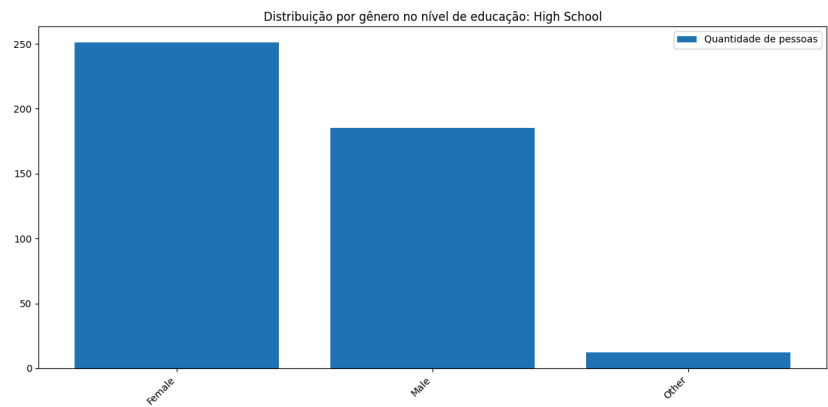
4.3. Quantidade de pessoas por Gênero em nível de educação

Figura 9 - Distribuição por gênero com educação: Bachelor's Degree



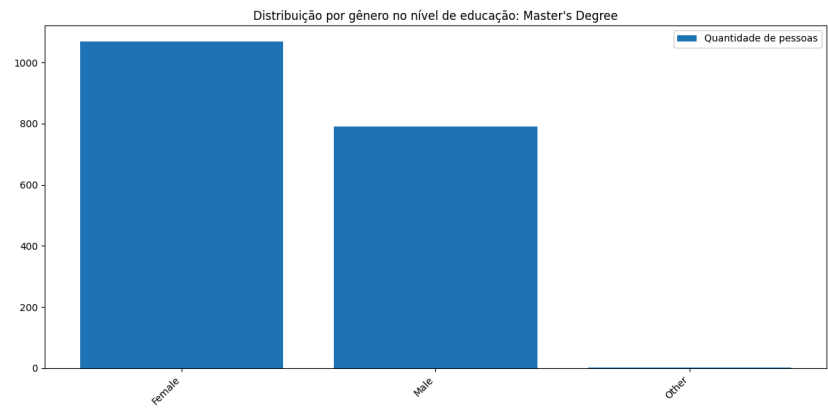
Fonte: Autor, 2025.

Figura 10 - Distribuição por gênero com educação: High School



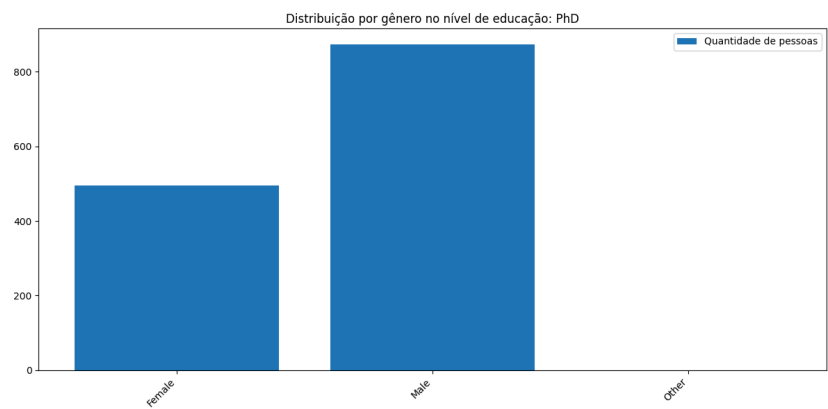
Fonte: Autor, 2025.

Figura 11 - Distribuição por gênero com educação: Master's Degree



Fonte: Autor, 2025.

Figura 12 - Distribuição por gênero com educação: PhD



Fonte: Autor, 2025.

4.4. Salário Médio por Cargo em cada Nível de Educação

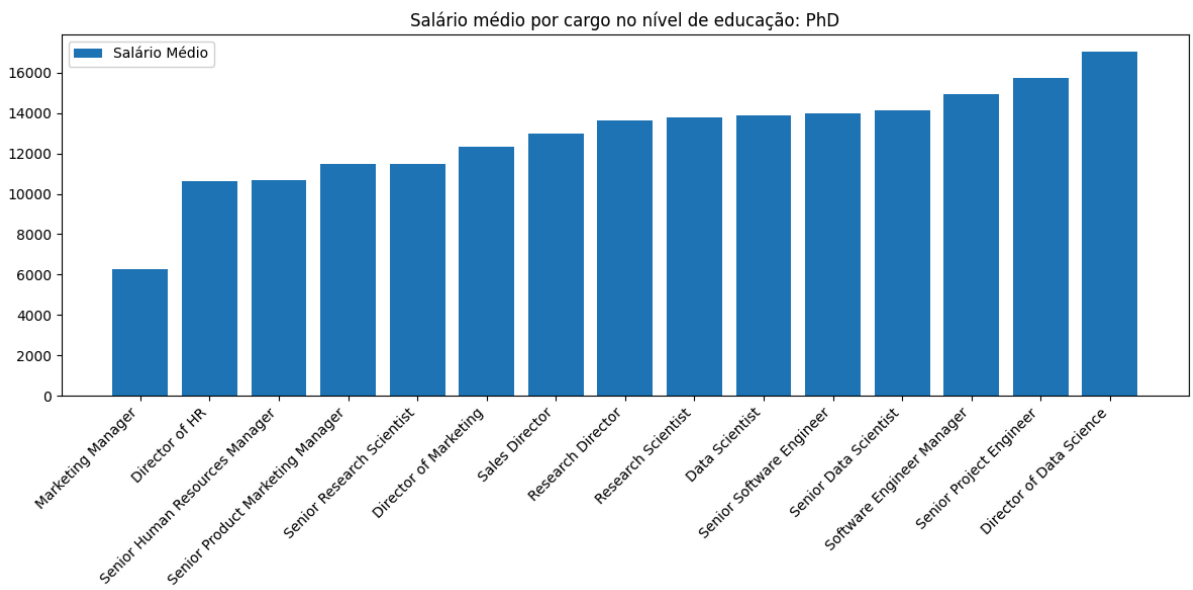


Figura 13 - Salário Médio dos Empregos de Nível PHD

Fonte: Autor, 2025.

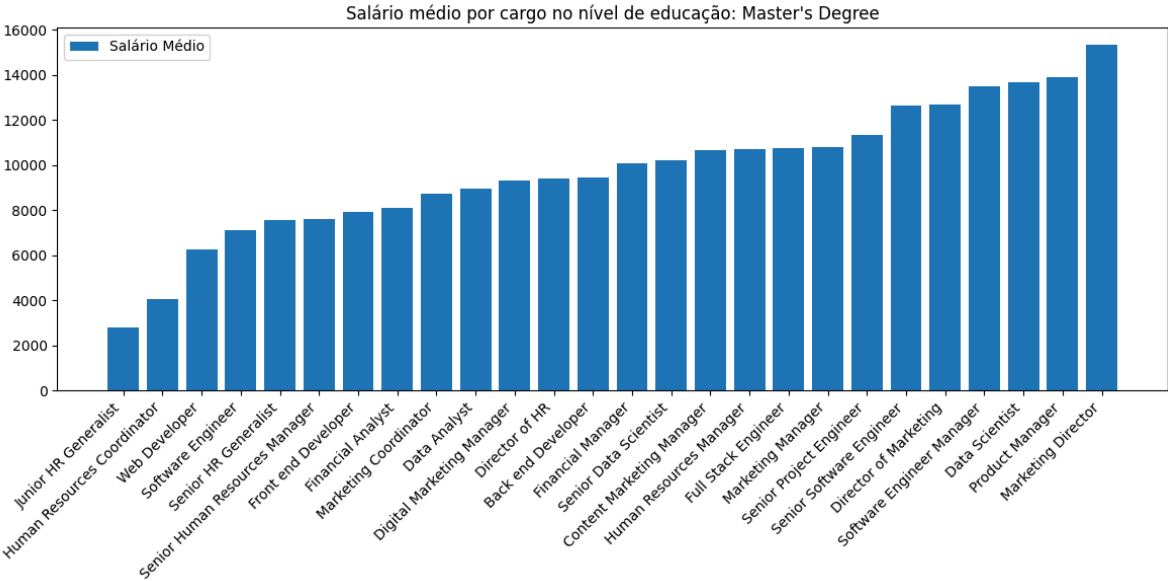


Figura 14 - Salário Médio dos Empregos de Nível Mestrado

Fonte: Autor, 2025.

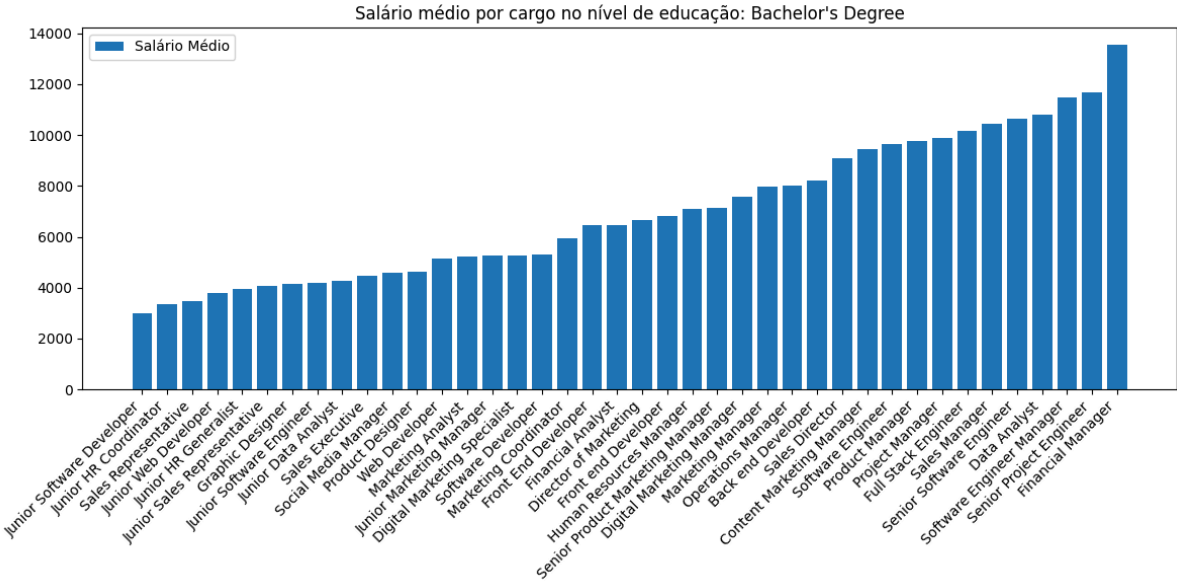


Figura 15 - Salário Médio dos Empregos de Nível Bacharelado

Fonte: Autor, 2025.



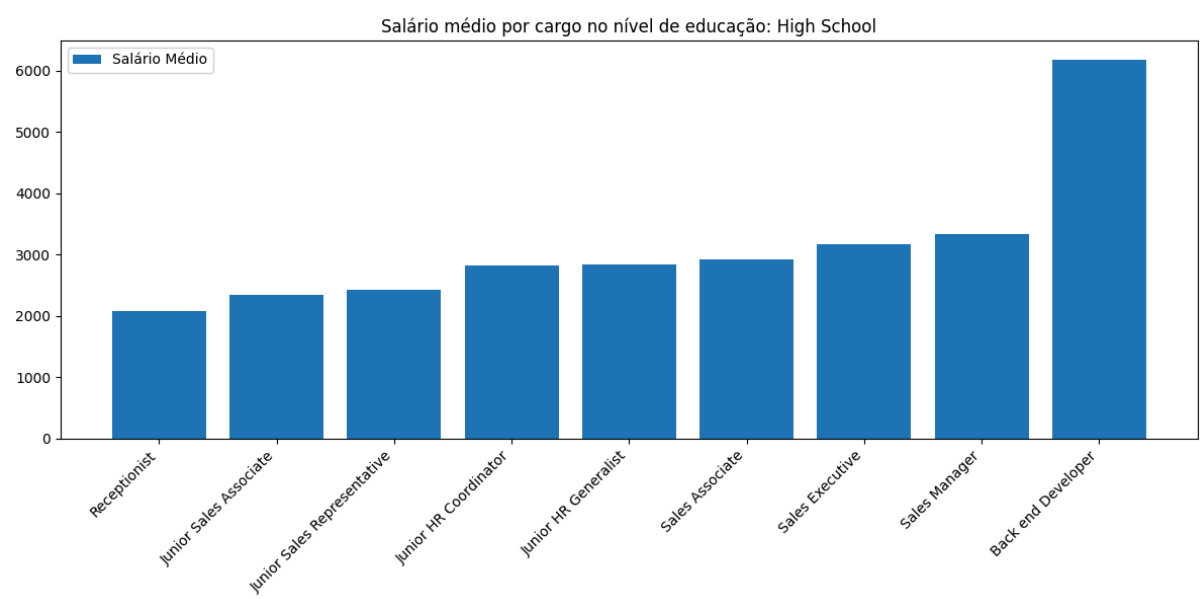


Figura 16 - Salário Médio dos Empregos de Nível Escolar

Fonte: Autor, 2025.

## **5. CONCLUSÃO**

A análise mostrou que idade, nível educacional, gênero e anos de experiência possuem impacto direto ou indireto na remuneração. Observou-se que maior escolaridade e maior tempo de experiência tendem a estar associados a salários mais altos, enquanto diferenças relacionadas a gênero e escolaridade evidenciam possíveis desigualdades no mercado de trabalho. Assim, os resultados reforçam a relevância de aplicar técnicas de big data para identificar padrões salariais e apoiar decisões em políticas de desenvolvimento profissional e equidade.

## REFERÊNCIAS

- W3SCHOOLS. **Pandas** **Tutorial**. Disponível em:  
<[www.w3schools.com/python/pandas/default.asp](http://www.w3schools.com/python/pandas/default.asp)>. Acesso em: 18 set. 2025.
- MATPLOTLIB. **Tutorials**. Disponível em:  
<<https://matplotlib.org/stable/tutorials/index.html>>. Acesso em: 18 set. 2025.
- W3SCHOOLS. **Matplotlib** **Tutorial**. Disponível em:  
<[https://www.w3schools.com/python/matplotlib\\_intro.asp](https://www.w3schools.com/python/matplotlib_intro.asp)>. Acesso em: 18 set. 2025.