

# Relatório de Análise de Salários

## 1. Introdução

Este relatório analisa um conjunto de dados sobre salários de profissionais de tecnologia, com foco em três principais dimensões: níveis de experiência, proporção de trabalho remoto e salários médios por ano e nível de experiência.

## 2. Importação de Bibliotecas e Conjunto de Dados

As bibliotecas necessárias foram importadas para realizar a análise, e o conjunto de dados foi carregado a partir de um arquivo CSV:

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import warnings
warnings.filterwarnings('ignore')

df = pd.read_csv('/content/ds_salaries.csv')
```

### Exibição de Amostra

Uma amostra aleatória de 10 registros do conjunto de dados foi exibida para visualização:

df.sample(10)

	work_year	experience_level	employment_type	job_title	salary	salary_currency	salary_in_usd	employee_residence	remote_ratio	company_location	company_size
2215	2022	MI	FT	Data Analyst	80000	USD	80000	US	100	US	L
3731	2020	MI	FT	Data Scientist	55000	EUR	62726	FR	50	LU	S
161	2023	SE	FT	Data Analyst	168400	USD	168400	US	0	US	M
8	2023	SE	FT	Data Scientist	141000	USD	141000	CA	0	CA	M
2129	2022	SE	FT	Data Scientist	185900	USD	185900	US	0	US	M
1949	2022	SE	FT	Data Engineer	95000	USD	95000	US	0	US	M
384	2023	SE	FT	Data Engineer	160000	USD	160000	US	0	US	M
1106	2023	SE	FT	Data Scientist	245000	USD	245000	US	0	US	M
3678	2021	MI	FT	Data Scientist	2500000	INR	33808	IN	0	IN	M
3482	2022	SE	FT	Research Scientist	144000	USD	144000	US	50	US	L

## 3. Distribuição dos Níveis de Experiência

A contagem dos níveis de experiência foi obtida e apresentada em um gráfico de barras:

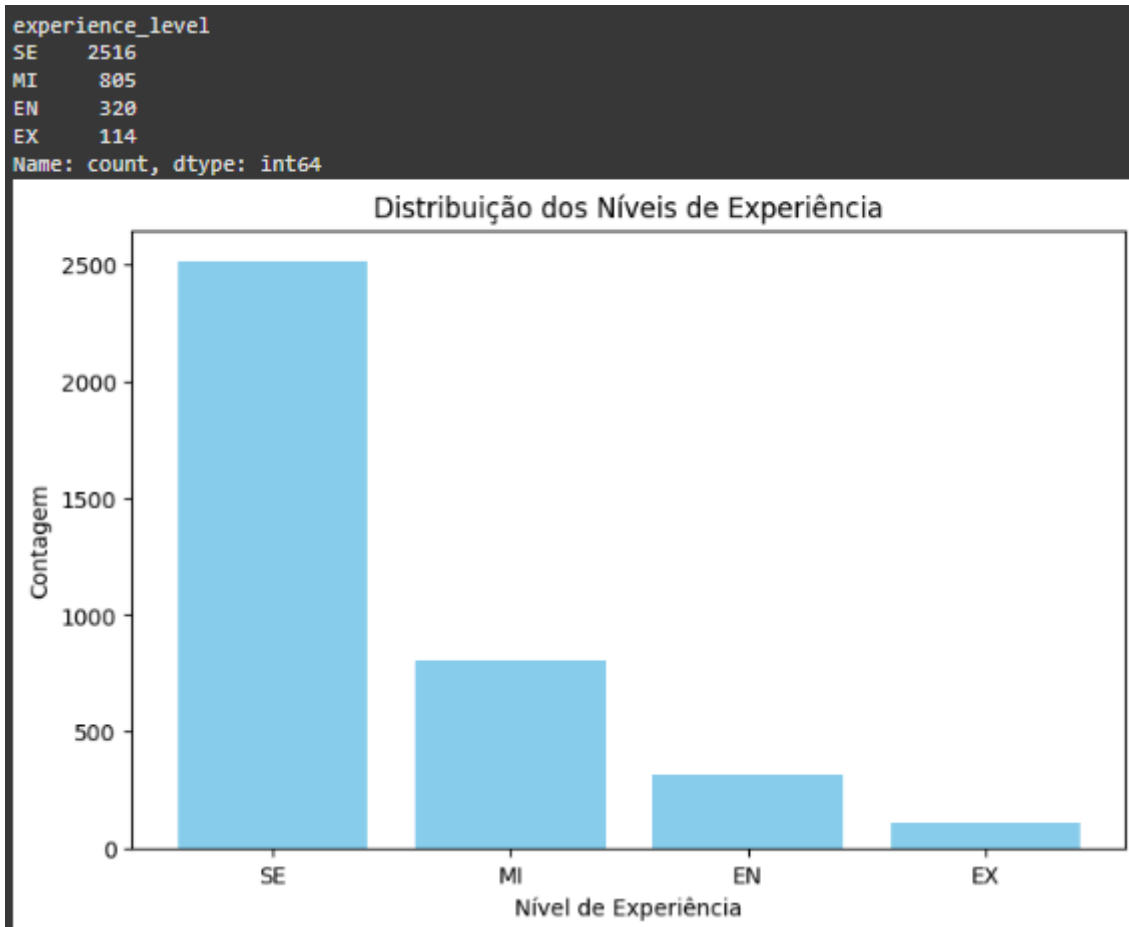
```
niveis_experiencia = df['experience_level'].value_counts()
print(niveis_experiencia)

categorias = niveis_experiencia.index
contagens = niveis_experiencia.values

plt.figure(figsize=(8, 5))
plt.bar(categorias, contagens, color='skyblue')
plt.xlabel('Nível de Experiência')
plt.ylabel('Contagem')
plt.title('Distribuição dos Níveis de Experiência')
plt.show()
```

## Resultados

A distribuição dos níveis de experiência mostra quantos profissionais estão em cada nível, permitindo identificar qual é o mais comum entre os entrevistados.



## 4. Proporção de Trabalho Remoto

A coluna `remote_ratio` foi convertida para o tipo `str` e a contagem da proporção de trabalho remoto foi realizada:

```
df['remote_ratio'] = df['remote_ratio'].astype(str)
proporcao_remos = df['remote_ratio'].value_counts()
print(proporcao_remos)

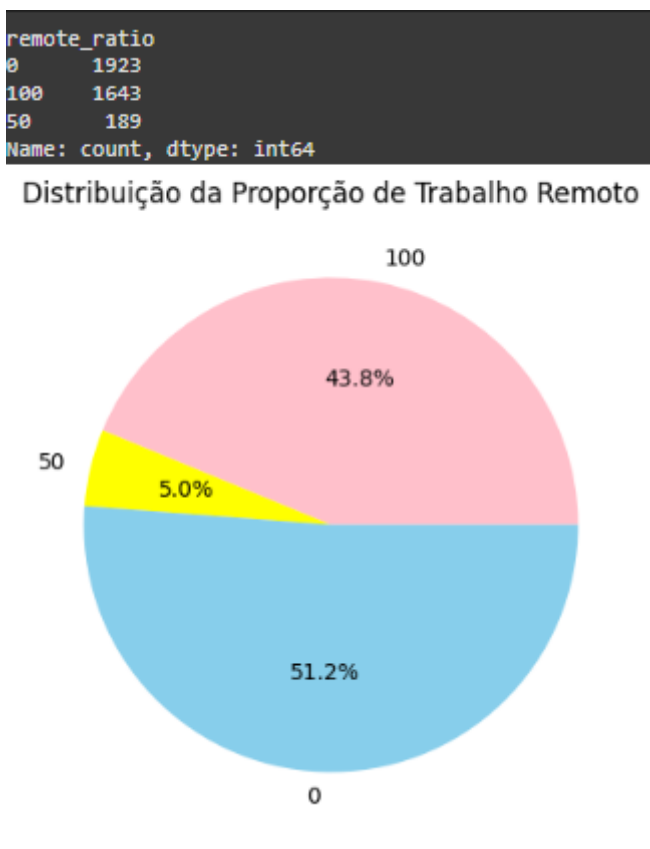
categorias = ['100', '50', '0']
contagens = [proporcao_remos.get(categoria, 0) for categoria in categorias]

# Criando o gráfico de pizza
plt.figure(figsize=(8, 5))
plt.pie(contagens, labels=categorias, autopct='%1.1f%%', colors=['pink', 'yellow', 'skyblue'])
plt.title('Distribuição da Proporção de Trabalho Remoto')

# Exibindo o gráfico
plt.show()
```

## Resultados

O gráfico de pizza apresenta a distribuição percentual dos diferentes níveis de trabalho remoto, permitindo entender a prevalência do trabalho remoto total (100%), parcial (50%) e nenhum (0%).



## 5. Distribuição dos Anos de Trabalho

A contagem dos anos de trabalho foi obtida e apresentada em um gráfico de barras:

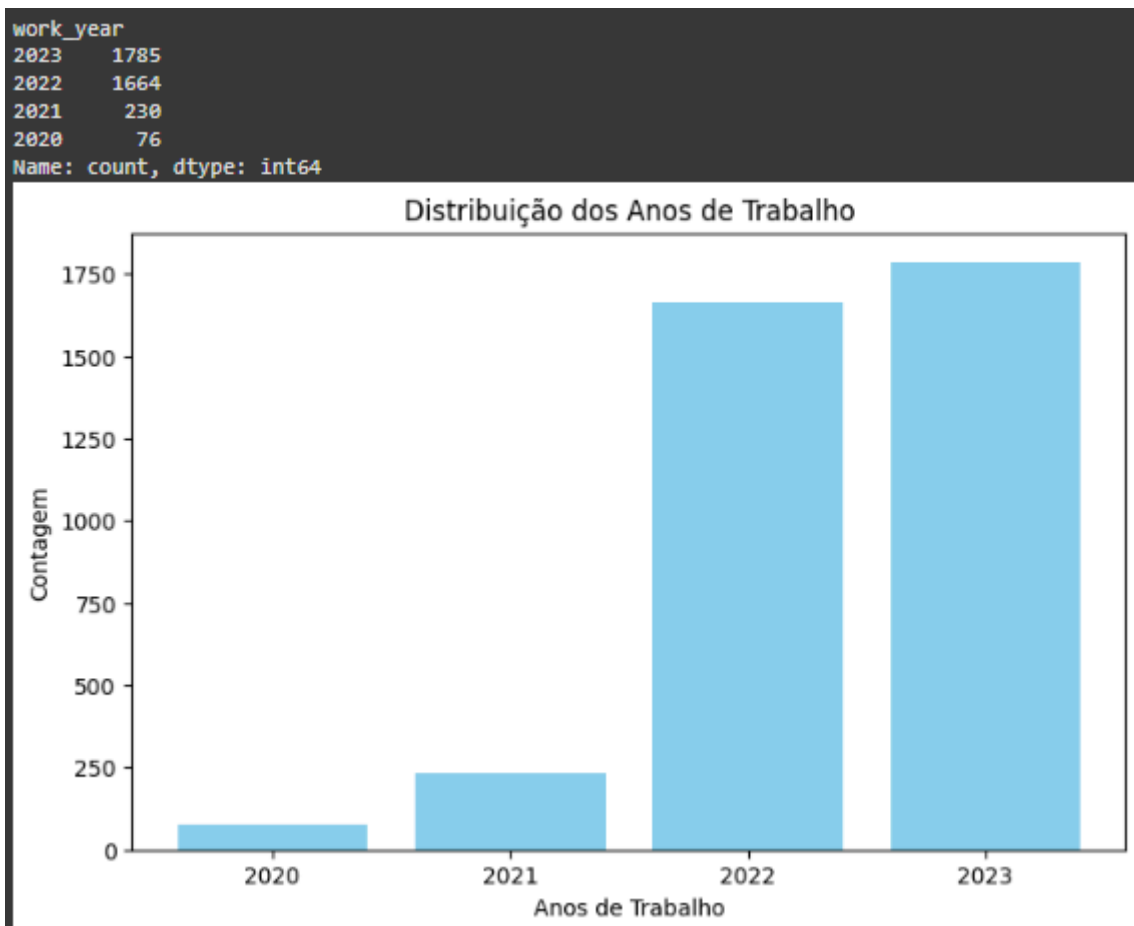
```
ano_count = df['work_year'].value_counts()
print(ano_count)

categorias = ano_count.index
contagens = ano_count.values

plt.figure(figsize=(8, 5))
plt.bar(categorias, contagens, color='skyblue')
plt.xlabel('Anos de Trabalho')
plt.ylabel('Contagem')
plt.title('Distribuição dos Anos de Trabalho')
plt.xticks(categorias)
plt.show()
```

## Resultados

O gráfico ilustra quantos profissionais estão em cada faixa de anos de trabalho, permitindo identificar a experiência geral no setor.



## 6. Salário Médio por Ano e Nível de Experiência

O salário médio em USD foi calculado por ano e nível de experiência e apresentado em um gráfico de linhas:

```
salario_ano_experiencia = df.groupby(['work_year', 'experience_level'])['salary_in_usd'].mean().reset_index()
print(salario_ano_experiencia)

plt.figure(figsize=(12, 6))

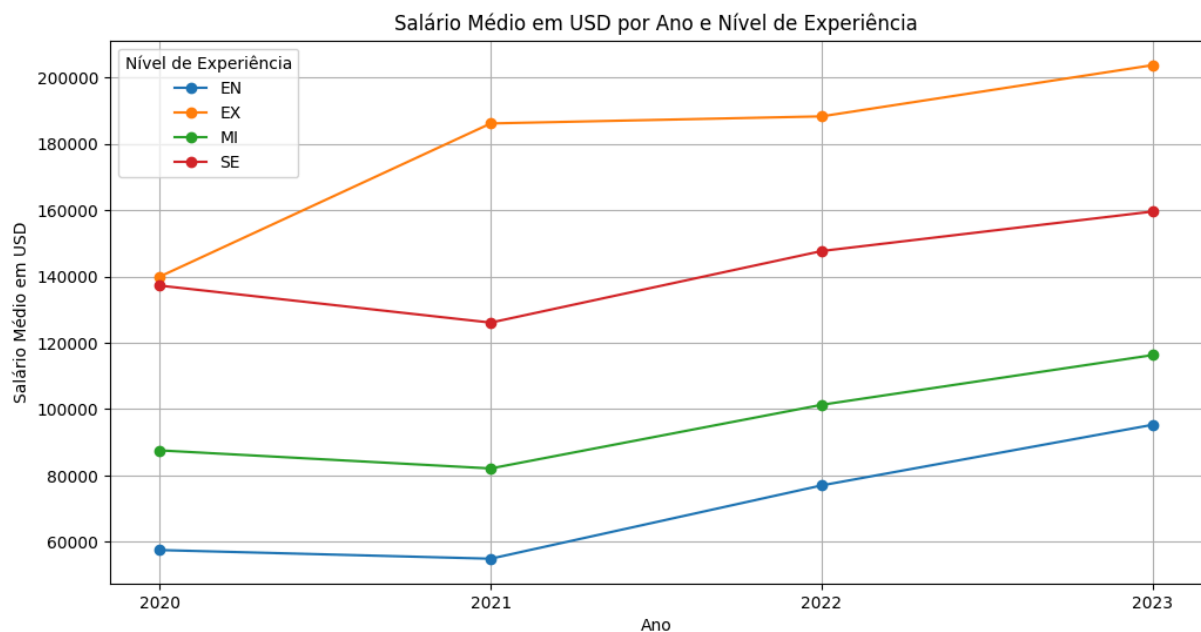
for level in salario_ano_experiencia['experience_level'].unique():
    subset = salario_ano_experiencia[salario_ano_experiencia['experience_level'] == level]
    plt.plot(subset['work_year'], subset['salary_in_usd'], marker='o', label=level)

plt.title('Salário Médio em USD por Ano e Nível de Experiência')
plt.xlabel('Ano')
plt.ylabel('Salário Médio em USD')
plt.xticks(subset['work_year'])
plt.grid(True)
plt.legend(title='Nível de Experiência')
plt.show()
```

### Resultados

O gráfico mostra a evolução do salário médio em USD ao longo dos anos, segmentado por níveis de experiência. Isso fornece insights sobre como os salários mudaram ao longo do tempo em diferentes níveis de experiência.

	work_year	experience_level	salary_in_usd
0	2020	EN	57511.608696
1	2020	EX	139944.333333
2	2020	MI	87564.718750
3	2020	SE	137240.500000
4	2021	EN	54905.254545
5	2021	EX	186128.000000
6	2021	MI	82116.934783
7	2021	SE	126085.356164
8	2022	EN	77006.024194
9	2022	EX	188260.292683
10	2022	MI	101305.598338
11	2022	SE	147659.688049
12	2023	EN	95283.966102
13	2023	EX	203705.683333
14	2023	MI	116297.596875
15	2023	SE	159568.928516



## 7. Distribuição do Porte da Empresa

A contagem do porte das empresas foi obtida e apresentada em um gráfico de barras:

```
porte_empresa = df['company_size'].value_counts()
print(porte_empresa)

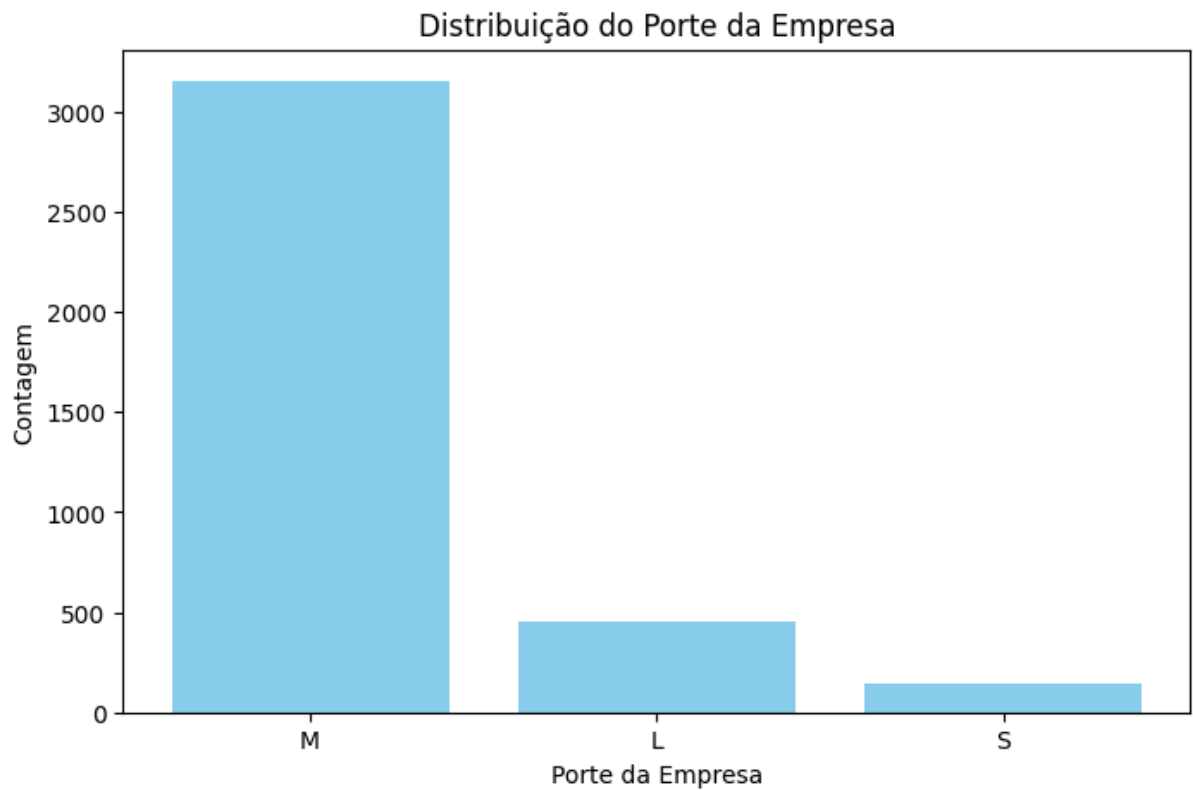
categorias = porte_empresa.index
contagens = porte_empresa.values

plt.figure(figsize=(8, 5))
plt.bar(categorias, contagens, color='skyblue')
plt.xlabel('Porte da Empresa')
plt.ylabel('Contagem')
plt.title('Distribuição do Porte da Empresa')
plt.show()
```

## Resultados

O gráfico de barras exibe a distribuição do porte das empresas, permitindo visualizar a quantidade de profissionais que trabalham em empresas de diferentes tamanhos (pequenas, médias e grandes).

```
company_size
M    3153
L     454
S     148
Name: count, dtype: int64
```



## 8. Conclusão

A análise realizada permitiu uma compreensão mais profunda da dinâmica salarial entre profissionais de tecnologia, destacando a importância do nível de experiência, da modalidade de trabalho (remoto) e do porte das empresas nas remunerações. As visualizações ajudam a identificar tendências e padrões, contribuindo para a tomada de decisões informadas.