



Desafio: Análise Estatística de Salários em Dados

≡ Escola

Dados

:≡ Habilidades

Estatística descritiva

✓ Mais 2 propriedades



Desafio



Arquivos do Desafio:



salario_profissionais_dados.csv 269.7KB

Análise Estatística: Salário de Profissionais de Dados

Este projeto tem como objetivo aplicar técnicas de estatística descritiva em um conjunto de dados contendo informações sobre profissionais da área de dados ao redor do mundo. A ideia é utilizar ferramentas de análise exploratória e visualização para entender como diferentes fatores impactam o salário desses profissionais.

Contexto

Você foi contratado para apoiar uma equipe de RH especializada em **tecnologia e ciência de dados**. Essa equipe recebeu uma base de dados com informações salariais de profissionais ao redor do mundo e quer entender **como variáveis como cargo, país, nível de experiência e ano influenciam os salários**.

Sua missão é organizar e explorar os dados, aplicando técnicas de estatística descritiva para extrair **insights relevantes sobre a remuneração na área de dados**.

Sobre os dados

O dataset contém dados de salários de profissionais da área de dados de diversas partes do mundo.

Entre as colunas disponíveis, temos:

Coluna	Descrição
<code>work_year</code>	Ano de referência do salário
<code>experience_level</code>	Nível de experiência (ex: EN = iniciante, MI = intermediário)
<code>employment_type</code>	Tipo de contrato (CLT, freelancer, etc.)
<code>job_title</code>	Cargo ocupado
<code>salary_in_usd</code>	Salário anual em dólares
<code>employee_residence</code>	País de residência do profissional
<code>company_location</code>	Localização da empresa
<code>company_size</code>	Tamanho da empresa (P = pequena, M = média, L = grande)

Legenda das Categorias

Experience_level

Código	Descrição
EN	Entry-level / Iniciante
MI	Mid-level / Intermediário
SE	Senior-level / Sênior
EX	Executive-level / Executivo

Company_size

Código	Descrição
S	Pequena empresa
M	Média empresa
L	Grande empresa

Como começar?

- Importe os dados usando a biblioteca `pandas` e explore as primeiras linhas com `.head()`.
- Verifique os tipos de dados, identifique colunas importantes e confira se há valores ausentes.
- Comece com uma análise exploratória simples: frequências, distribuições e estatísticas básicas.
- Use gráficos para visualizar padrões e tendências.

Etapas de Desenvolvimento

Etapa 01) Importação e Exploração Inicial da Base

Antes de qualquer análise, é fundamental entender a estrutura dos dados com os quais você está lidando. Nesta etapa, seu foco será **importar o dataset** e realizar uma primeira exploração para verificar:

- Quais colunas estão disponíveis?
- Quantas linhas existem?
- Existem dados ausentes?
- Quais são os tipos de variáveis?



Dicas: Antes de começar a análise, entenda bem o que cada variável representa. Isso ajuda a guiar a leitura dos dados. Você pode usar métodos como `df.head()`, `df.shape`, `df.info()` e `df.isnull().sum()`

Etapa 02) Frequência e Distribuição das Categorias

Agora que você já conhece a base, vamos explorar as variáveis **categóricas** — aquelas que representam grupos ou classificações, como os **cargos dos profissionais**.

Seu objetivo aqui é responder perguntas como:

- Cargos mais comuns
- Nível de experiência

- Tamanho da empresa

Essas informações ajudam a entender o perfil predominante da base e fornecem contexto para análises salariais futuras.

💡 Dica: Use `.value_counts().head(10)` para ver os mais frequentes e `normalize=True` para proporções relativas (porcentagem)

Etapa 03) Estatísticas Descritivas

A principal variável numérica da base é `salary_in_usd`. Nesta etapa, você vai se aprofundar na análise estatística dessa variável para entender como os salários estão distribuídos.

- **Medidas de tendência central:** média, mediana
- **Medidas de dispersão:** desvio padrão, mínimo e máximo
- **Distribuição geral:** através de histogramas
- **Comparações por grupos:** especialmente por nível de experiência (usando boxplots)

💡 Dica: Use `groupby()` com `mean()` para comparar médias entre categorias. Ex:
`df.groupby('company_size')['salary_in_usd'].mean()`

Etapa 04) Comparações por País

Agora que você já entendeu os salários em geral, é hora de investigar **como eles variam entre diferentes grupos**.

Nesta etapa, seu objetivo será:

- Verificar os **10 países** com as maiores médias salariais

Essas comparações são fundamentais para entender onde estão os maiores salários — seja em termos de função ou localização geográfica.

💡 Dica: Use `groupby()` com `.mean()` para calcular médias por categoria, e depois ordene com `.sort_values()` para ver os maiores valores primeiro.

Etapa 05) Correlações e Tendências

Por fim, vamos analisar se existe alguma relação entre salário e outras variáveis **quantitativas** da base.

As perguntas principais aqui são:

- Existe alguma **tendência de aumento salarial com os anos?**
- Quanto maior o tempo de experiência, maior o salário?

Você irá utilizar uma **matriz de correlação** para investigar essas relações e interpretar os coeficientes gerados.

💡 Dica: Use `corr()` para criar uma matriz de correlação entre `salary_in_usd`, `work_year` e `years_of_experience`.

Critérios de Avaliação

Os critérios de avaliação mostram como você será avaliado em relação ao seu desafio.

Critério	Descrição do que será avaliado	Pontos
Importação e Exploração Inicial dos Dados	Leitura correta do arquivo com <code>pandas</code> , análise inicial com <code>.info()</code> , <code>.shape()</code> , <code>.head()</code> e verificação de valores nulos.	20
Análise de Frequências e Visualizações	Análise das variáveis categóricas com <code>value_counts()</code> e interpretação correta dos cargos mais frequentes. Uso adequado de filtros como <code>head(10)</code> .	20
Estatísticas Descritivas de Salários	Aplicação das principais medidas descritivas (média, mediana, desvio padrão, etc.), além de visualização com histograma e boxplot para comparação entre grupos (como nível de experiência).	20
Análises Comparativas e Correlações	Cálculo da média salarial por grupo usando <code>groupby()</code> e interpretação de quais cargos e países apresentam os maiores valores médios.	20
Correlação com Tempo e Experiência	Aplicação da matriz de correlação entre variáveis numéricas com <code>.corr()</code> e interpretação dos coeficientes, especialmente entre <code>salary_in_usd</code> , <code>work_year</code> e <code>years_of_experience</code> .	20

Entrega