

Salário do desenvolvedor de dados em 2024 💰

Analisando salários de desenvolvedores de dados em 2024

Valor: 25 pontos

Data de entrega: 04/10/2024, via Canvas

Somente um aluno deve postar o trabalho no Canvas em arquivo formato .pdf

Nomes dos alunos do grupo:

- 1.
- 2.
- 3.
- 4.
- 5.

Qual(is) software(s) foi(ram) utilizado(s) para a realização do trabalho:

Sobre o conjunto de dados

Conteúdo:

O conjunto de dados de 2024 sobre salários e atributos de emprego dos desenvolvedores de dados oferece informações valiosas sobre o cenário em evolução dos desenvolvedores de dados. Inclui variáveis-chave como salário, cargo, nível de experiência, tipo de emprego, residência do funcionário, proporção de trabalho remoto, localização da empresa e tamanho da empresa. Esses dados permitem análises detalhadas de tendências salariais, padrões de emprego e variações geográficas nas funções de desenvolvedor de dados. Pesquisadores, analistas e organizações podem aproveitar esse conjunto de dados para compreender melhor as tendências de remuneração, a distribuição das funções de desenvolvedor de dados em diferentes regiões e o impacto do trabalho remoto e do tamanho da empresa no emprego nesta área.

Descrição das variáveis do dataset:

experience_level: Nível de experiência profissional (por exemplo, júnior, médio, sênior).

Employment_type: Tipo de contrato de trabalho (por exemplo, período integral, meio período, contrato).

job_title: A função ou título específico do funcionário (por exemplo, Engenheiro de Dados).

salário: A remuneração recebida, na moeda original.

salário_currency: A moeda em que o salário é pago.

salário_in_usd: O salário convertido em dólares americanos para comparação.

Employee_residence: O local onde o funcionário reside.

remote_ratio: Porcentagem de trabalho realizado remotamente.

company_location: A localização geográfica da empresa.

company_size: A escala da empresa, geralmente baseada na contagem de funcionários.

Tarefas:

- 1) (1pt) Construa uma tabela de frequências, de forma adequada, para representar a variável `salário_in_usd`.
- 2) (1pt) Construa uma tabela de frequências, de forma adequada, para representar a variável `job_title`.
- 3) (6pts) Construir um gráfico adequado para cada uma das seguintes variáveis:
 - `experience_level`
 - `Employment_type`
 - `salário_in_usd`
 - `remote_ratio`
 - `company_size`

Escreva um pequeno parágrafo citando os principais achados observados nos gráficos construídos.

- 4) (3pts) Calcule as medidas descritivas (média, mediana, mínimo, máximo, desvio padrão, coeficiente de variação, 1º quartil, 3º quartil) para a variável `salário_in_usd`. Construa um boxplot para a variável `salário_in_usd`. Escreva um pequeno parágrafo citando as principais informações obtidas por meio da análise das medidas descritivas calculadas e do gráfico construído.
- 5) (3pts) Construa uma tabela de contingência mostrando a frequência e um dos percentuais (da linha, da coluna OU do total geral) entre as variáveis `experience_level` e `remote_ratio`. Construa um gráfico adequado para representar os dados dessa tabela. Escreva um pequeno parágrafo citando os principais achados.
- 6) (3pts) Construa uma tabela de contingência mostrando a frequência e um dos percentuais (da linha, da coluna OU do total geral) entre as variáveis `Employment_type` e `company_size`. Construa um gráfico adequado para representar os dados dessa tabela. Escreva um pequeno parágrafo citando os principais achados.
- 7) (3pts) Calcule as medidas descritivas (média, mediana, mínimo, máximo, desvio padrão, coeficiente de variação, 1º quartil, 3º quartil) para a variável `salário_in_usd` estratificando pela variável `experience_level`. Construa um boxplot estratificado entre essas duas variáveis. Escreva um pequeno parágrafo citando as principais informações obtidas por meio da análise das medidas descritivas calculadas e do gráfico construído (compare as medianas, a variabilidade, a homogeneidade, a presença de valores discrepantes e as médias).
- 8) (3pts) Calcule as medidas descritivas (média, mediana, mínimo, máximo, desvio padrão, coeficiente de variação, 1º quartil, 3º quartil) para a variável `salário_in_usd` estratificando pela variável `Employment_type`. Construa um boxplot estratificado entre essas duas variáveis. Escreva um pequeno parágrafo citando as principais informações obtidas por meio da análise das medidas descritivas calculadas e do gráfico construído (compare as medianas, a variabilidade, a homogeneidade, a presença de valores discrepantes e as médias).
- 9) (1pt) Cite uma situação hipotética em que o gráfico de linhas seria adequado para representar tais dados. Quais informações podemos obter ao utilizar um gráfico de linhas?
- 10) (1pts) Pesquise a respeito do uso de mapa de calor em tabelas. Quando ele deve ser usado? Qual sua utilidade/benefícios? Mostre um exemplo hipotético.