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. Gabriel Fernandes de Oliveira
2. Italo Henrique Soares dos Santos
3. Pedro Reis de Souza
4. Rodrigo Reis do Valle
5.

Qual(is) software(s) foi(ram) utilizado(s) para a realização do trabalho:
Os softwares utilizados foram Jupyter Notebooks e Python.

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. Construir 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) Construir 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) Construir 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. Construir 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. Construir 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.

Relatório Exploratório - Dados Gerais - Processamento Excel

Diagnóstico

```
Linhas: 1.000.000
Colunas: 10
Memória: 1.000.000 linhas x 10 colunas
```

Processamento Excel

Tarefa 1

```
Tarefa 1: Contagem de Ocorrências
Coluna: Cidade
Total de categorias: 10
Total de ocorrências: 1.000.000

Top 10 categorias:
Cidade
Curitiba      100.000
Fortaleza     100.000
Manaus        100.000
Porto Alegre  100.000
Recife        100.000
Salvador      100.000
São Paulo     100.000
Belo Horizonte 99.999
Brasília       99.999
Rio de Janeiro 99.999
Name: count, dtype: int64

Salvando resultado em: /home/claude/tarefa_1_contagem.csv
Salvando estatísticas em: /home/claude/tarefa_1_estatisticas.csv
Concluído!
```

Tarefa 2

```
Tarefa 2: Tabela Cruzada
Linhas: Cidade
Colunas: Categoria
Total de combinações: 100

Tabela cruzada (primeiras 5 linhas x 5 colunas):
Categoria  Alimentação  Eletrônicos  Esporte  Livros  Roupas
Cidade
Belo Horizonte        9            9        9       9       9
Brasília              9            9        9       9       9
Curitiba              9            9        9       9       9
Fortaleza             9            9        9       9       9
Manaus                9            9        9       9       9
Porto Alegre          9            9        9       9       9
Recife                9            9        9       9       9
Rio de Janeiro        9            9        9       9       9
Salvador              9            9        9       9       9
São Paulo             9            9        9       9       9

Salvando tabela completa em: /home/claude/tarefa_2_tabela_cruzada.csv
Concluído!
```

Tarefa 3

```
Tarefa 3: Tabela Cruzada com Agregação
Linhas: Cidade
Colunas: Categoria
Valores: Valor_Total
Função: sum
Total de combinações: 100

Tabela cruzada (primeiras 5 linhas x 5 colunas):
Categoria  Alimentação  Eletrônicos  Esporte  Livros  Roupas
Cidade
Belo Horizonte        9            9        9       9       9
Brasília              9            9        9       9       9
Curitiba              9            9        9       9       9
Fortaleza             9            9        9       9       9
Manaus                9            9        9       9       9
Porto Alegre          9            9        9       9       9
Recife                9            9        9       9       9
Rio de Janeiro        9            9        9       9       9
Salvador              9            9        9       9       9
São Paulo             9            9        9       9       9

Salvando tabela completa em: /home/claude/tarefa_3_tabela_cruzada_agregada.csv
Concluído!
```

Tarefa 4

```
Tarefa 4: Gráfico de Barras
Coluna: Cidade
Tipo: horizontal
Top 15 categorias
Salvando gráfico em: /home/claude/tarefa_4_grafico_barras.png
Concluído!
```

Tarefa 5

```
Tarefa 5: Gráfico de Pizza
Coluna: Categoria
Top 5 categorias
Salvando gráfico em: /home/claude/tarefa_5_grafico_pizza.png
Concluído!
```

Tarefa 6

```
Tarefa 6: Gráfico de Linha
Coluna: Data_Venda
Tipo: temporal
Agrupamento: mês
Salvando gráfico em: /home/claude/tarefa_6_grafico_linha.png
Concluído!
```

Tarefa 7

```
Tarefa 7: Gráfico de Histograma
Coluna: Valor_Total
Bins: 30
Salvando gráfico em: /home/claude/tarefa_7_grafico_histograma.png
Concluído!
```

Tarefa 8

```
Tarefa 8: Gráfico de Barras Empilhadas
Linhas: Cidade
Colunas: Categoria
Tipo: empilhado
Salvando gráfico em: /home/claude/tarefa_8_grafico_barras_empilhadas.png
Concluído!
```

Tarefa 9

```
Tarefa 9: Gráfico de Dispersão
Eixo X: Valor_Unitário
Eixo Y: Valor_Total
Amostragem: 10.000 pontos
Salvando gráfico em: /home/claude/tarefa_9_grafico_dispersao.png
Concluído!
```

Tarefa 10

```
Tarefa 10: Gráfico de Boxplot
Coluna: Valor_Total
Salvando gráfico em: /home/claude/tarefa_10_grafico_boxplot.png
Concluído!
```

Tarefa 11

```
Tarefa 11: Gráfico de Área
Coluna: Data_Venda
Tipo: temporal
Agrupamento: mês
Salvando gráfico em: /home/claude/tarefa_11_grafico_area.png
Concluído!
```

Tarefa 12

```
Tarefa 12: Gráfico de Pizza
Coluna: Cidade
Top 5 categorias
Salvando gráfico em: /home/claude/tarefa_12_grafico_pizza.png
Concluído!
```

Análise Exploratória

Tarefa 13

```
Tarefa 13: Estatísticas Descritivas
Colunas numéricas: 4

Estatísticas:
       Valor_Unitário  Quantidade  Valor_Total  Desconto
count        1.000000    1.000000     1.000000  1.000000
mean         2.500000    3.000000     4.500000  0.050000
std          1.118034    1.414214     2.121320  0.031623
min          1.000000    1.000000     1.000000  0.000000
25%          1.750000    2.000000     3.000000  0.025000
50%          2.500000    3.000000     4.500000  0.050000
75%          3.250000    4.000000     6.000000  0.075000
max          4.000000    5.000000     8.000000  0.100000

Salvando em: /home/claude/tarefa_13_estatisticas_descritivas.csv
Concluído!
```

Tarefa 14

```
Tarefa 14: Matriz de Correlação
Colunas numéricas: 4

Correlações:
                Valor_Unitário  Quantidade  Valor_Total  Desconto
Valor_Unitário        1.000000    0.000000     0.000000  0.000000
Quantidade            0.000000    1.000000     0.000000  0.000000
Valor_Total           0.000000    0.000000     1.000000  0.000000
Desconto              0.000000    0.000000     0.000000  1.000000

Salvando em: /home/claude/tarefa_14_matriz_correlacao.csv
Concluído!
```

Tarefa 15

```
Tarefa 15: Heatmap de Correlação
Colunas numéricas: 4
Salvando gráfico em: /home/claude/tarefa_15_heatmap_correlacao.png
Concluído!
```

Tarefa 16

```
Tarefa 16: Análise de Outliers
Coluna: Valor_Total
Método: IQR

Resultados:
Total de valores: 1.000.000
Outliers: 0
Percentual: 0.00%
Limite inferior: -1.500,00
Limite superior: 7.500,00

Salvando em: /home/claude/tarefa_16_outliers.csv
Concluído!
```

Tarefa 17

```
Tarefa 17: Gráfico de Boxplot Múltiplo
Colunas numéricas: 4
Salvando gráfico em: /home/claude/tarefa_17_boxplot_multiplo.png
Concluído!
```

Tarefa 18

```
Tarefa 18: Gráfico de Pareto
Coluna: Cidade
Top 15 categorias
Salvando gráfico em: /home/claude/tarefa_18_grafico_pareto.png
Concluído!
```

Tarefa 19

```
Tarefa 19: Gráfico de Linha Múltiplas
Colunas numéricas: 4
Amostragem: 1.000 pontos
Salvando gráfico em: /home/claude/tarefa_19_grafico_linha_multiplas.png
Concluído!
```

Tarefa 20

```
Tarefa 20: Gráfico de Dispersão Múltiplo
Colunas numéricas: 4
Amostragem: 1.000 pontos
Salvando gráfico em: /home/claude/tarefa_20_grafico_dispersao_multiplo.png
Concluído!
```

Tarefa 21

```
Tarefa 21: Gráfico de Barras Agrupadas
Linhas: Cidade
Colunas: Categoria
Tipo: agrupado
Salvando gráfico em: /home/claude/tarefa_21_grafico_barras_agrupadas.png
Concluído!
```

Tarefa 22

```
Tarefa 22: Gráfico de Área Empilhada
Colunas: Cidade, Categoria, Data_Venda
Tipo: empilhado
Salvando gráfico em: /home/claude/tarefa_22_grafico_area_empilhada.png
Concluído!
```

Tarefa 23

```
Tarefa 23: Gráfico de Violino
Colunas numéricas: 4
Amostragem: 1.000 pontos
Salvando gráfico em: /home/claude/tarefa_23_grafico_violino.png
Concluído!
```

Tarefa 24

```
Tarefa 24: Gráfico de Pizza Múltiplo
Colunas: Cidade, Categoria, Data_Venda
Top 5 categorias
Salvando gráfico em: /home/claude/tarefa_24_grafico_pizza_multiplo.png
Concluído!
```

Tarefa 25

```
Tarefa 25: Gráfico de Linha Acumulada
Colunas numéricas: 4
Amostragem: 1.000 pontos
Salvando gráfico em: /home/claude/tarefa_25_grafico_linha_acumulada.png
Concluído!
```

Análise de Dados Ausentes

Tarefa 26

```
Tarefa 26: Análise de Dados Ausentes
Colunas analisadas: 10

Resumo:
Total de colunas: 10
Colunas com ausentes: 0
Total de ausentes: 0
Percentual geral: 0.00%

Detalhamento:
Coluna  Ausentes  Percentual
Salvando em: /home/claude/tarefa_26_analise_ausentes.csv
Concluído!
```

Tarefa 27

```
Tarefa 27: Gráfico de Ausentes
Colunas analisadas: 10
Salvando gráfico em: /home/claude/tarefa_27_grafico_ausentes.png
Concluído!
```

Tarefa 28

```
Tarefa 28: Gráfico de Heatmap de Ausentes
Colunas analisadas: 10
Amostragem: 1.000 linhas
Salvando gráfico em: /home/claude/tarefa_28_heatmap_ausentes.png
Concluído!
```