



FACULDADE DE TECNOLOGIA E INOVAÇÃO SENAC DF

DÉBORA FERREIRA
LORENA FREITAS
LUDMILLA CARVALHO

DATASET: SALÁRIO CIENTISTA DE DADOS 2021 A 2023

Deep Learning

Brasília, DF

2023

INDÍCE

1. Introdução
2. Objetivo
3. Rótulos
4. Tipos de Dados
5. Quantitativos
6. Número de Datasets
7. Relacionamentos
8. Formato dos Dados
9. Análise Crítica

1. INTRODUÇÃO

Nesta documentação, exploraremos um dataset obtido do site Kaggle, que se concentra nos salários de cientistas de dados no período de 2021 a 2023, comparando esses valores com outras profissões que lidam com análise de dados para tomada de decisões. O objetivo principal é analisar, compreender e preparar os dados para futuras análises. Para isso, empregaremos seis métodos essenciais:

1. Rótulos: Identificaremos e descreveremos os rótulos, que representam as variáveis no dataset.
2. Tipos de Dados: Examinaremos os tipos de dados, diferenciando entre valores numéricos, textuais e datas, para entender como as informações estão formatadas.
3. Qualitativos: Analisaremos atributos qualitativos, se referem a dados que são expressos em números e representam quantidades mensuráveis.
4. Número de Datasets: Investigaremos quantos conjuntos de dados estão incluídos na documentação.
5. Relacionamento: Vamos descrever como os dados em diferentes datasets estão conectados.
6. Formatos dos Dados: Descreveremos os formatos em que os dados estão apresentados, como CSV ou Excel.

Além disso, realizamos uma etapa crucial de limpeza de dados no ambiente Jupyter para eliminar entradas duplicadas, reduzindo a redundância e garantindo que as análises subsequentes sejam baseadas em dados confiáveis. Ao detalhar cada variável e conceito neste dataset, pretendemos facilitar a interpretação e a análise, possibilitando insights valiosos para quem utilizar esses dados.

2. OBJETIVO

Como de estudantes de Ciência de Dados, estamos empenhados em realizar uma análise abrangente que permita obter insights valiosos sobre o futuro de nossa profissão e seu contínuo crescimento no mercado de trabalho. Compreender como a Ciência de Dados evoluirá é essencial para tomar decisões bem fundamentadas sobre nossas carreiras.

Além disso, essa análise visa ir além da simples observação de tendências. Planejamos empregar técnicas avançadas de análise de dados para realizar uma análise preditiva, fornecendo uma base sólida para antecipar como a Ciência de Dados se posicionará no cenário profissional em comparação com outras profissões semelhantes. Este aspecto é especialmente relevante para os estudantes que desejam tomar decisões de carreira informadas.

Ao longo do processo, iremos coletar e explorar informações detalhadas sobre salários, comparando os ganhos dos profissionais de Ciência de Dados com aqueles em ocupações relacionadas. Além disso, investigaremos como a geolocalização dos profissionais afeta a distribuição salarial, oferecendo uma visão abrangente das oportunidades regionais em nosso campo.

Nosso objetivo é fornecer uma análise completa, informada e prospectiva que auxilie os estudantes de Ciência de Dados a navegar no cenário profissional em constante evolução, permitindo que tomem decisões de carreira sólidas e com conhecimento de causa. Acreditamos que esses insights serão um recurso valioso para nossas escolhas de carreira e um guia confiável para o futuro de nossa profissão.

3. RÓTULOS

WORK YEAR	ANO DE TRABALHO
EXPERIENCE LEVEL	NÍVEL DE EXPERIÊNCIA
EMPLOYMENT TYPE	TIPO DE EMPREGO
JOB TITLE	CARGO
SALARY	SALÁRIO
SALARY CURRENCY	MOEDA SALARIAL
SALARY IN USD	SALÁRIO EM DÓLARES
COMPANY LOCATION	LOCALIZAÇÃO DA EMPRESA
COMPANY SIZE	TAMANHO DA EMPRESA

Neste contexto, temos variáveis de diversos tipos, incluindo textuais, numéricas e categóricas

4. TIPOS DE DADOS

Nome Coluna	Tipo de Dado	Tamanho	Restrições	Descrição
Ano	Integer	4 bytes	Not Null	Representa o ano específico de coleta de dados salariais.
Nível Experiência	Varchar	2 bytes	Not Null	Representa o nível de experiência profissional dos funcionários, categorizado como EN (Nível Básico), EX (Experiente), MI (Nível Médio), SE (Sênior).
Tipo Emprego	Varchar	2 bytes	Not Null	Representa o tipo de emprego, rotulado como FT (tempo integral), CT (contratado), FL (freelancer), PT (tempo parcial).
Cargo	Varchar	50 bytes	Not Null	Representa os cargos dos funcionários, como "Cientista Aplicado", "Analista de Qualidade de Dados" etc.
Salário	Decimal	10 bytes	Not Null	Representa os valores salariais em seus respectivos formatos de moeda.
Salário Moeda	Decimal	10 bytes	Not Null	Representa o código da moeda que representa o salário.
Salário em USD	Decimal	10 bytes	Not Null	Representa os valores salariais convertidos em dólares americanos para comparação

				uniforme.
Local da Empresa	Varchar	5 bytes	Not Null	Representa a localização das empresas, especificada como códigos de país (por exemplo, "US" para os Estados Unidos e "NG" para a Nigéria).
Tamanho da Empresa	char	1 byte	Not Null	Representa o tamanho das empresas, classificado como "L" (Grande), "M" (Médio) e "S" (Pequeno).

5. QUANTITATIVOS

3761 linhas e 9 colunas.

6. NÚMERO DE DATASETS

Na presente pesquisa e análise, empregamos apenas um conjunto de dados.

7. RELACIONAMENTOS

Uma vez que estamos utilizando apenas uma tabela, não existem outras tabelas para estabelecer relacionamentos.

8. FORMATO DOS DADOS

Através do site Kaggle, encontramos este dataset que estava no formato CSV.

9. ANÁLISE CRÍTICA

Salários por Profissão:

- **Análise:** Comparando os salários entre cientistas de dados, engenheiros de dados e analistas, observamos que os cientistas de dados têm uma média salarial 88.57% maior do que os engenheiros de dados. Além disso, a média salarial para Cientista de Dados é 151.53% maior que a média para Analista de Dados.
- **Interpretação:** Isso sugere que, com base nos dados analisados, a profissão de Cientista de Dados é a mais bem remunerada em comparação com Engenheiro de Dados e Analista de Dados.

Distribuição Geográfica:

- **Análise:** Ao examinar a distribuição geográfica de cientistas de dados em 31 países, observamos que os quatro países com mais profissionais nesta área são, em ordem decrescente, Estados Unidos, Reino Unido, Canadá e Índia.
- **Interpretação:** Isso indica que esses países têm uma presença significativa de cientistas de dados e podem representar locais-chave para oportunidades nesta profissão.

Experiência Profissional:

- **Análise:** Ao avaliar as experiências dos cientistas de dados, notamos uma predominância de profissionais com níveis sênior, seguidos por níveis médio e básico, enquanto os profissionais experientes são menos representados.
- **Interpretação:** Isso sugere que, embora haja uma presença significativa de profissionais experientes e seniores, existe uma demanda crescente por cientistas de dados com diferentes níveis de experiência, com foco particular em profissionais seniores.