

Integrantes:

André Caiado Ramos - RA: 822154512 - Leopoldina
Antônio Carlos Ruy Neto - RA: 822164957 - Leopoldina
Camila Teixeira Donola - RA:819224067 - Mooca
Gustavo Fernandes Pacheco - RA: 822144529 - Leopoldina
Lucas Basaglia Camanducci - RA: 822141662 - Leopoldina
Lucas Henrique de Oliveira - RA 821230388 - Leopoldina

Entrega 2: Documentação com a Explicação da Base de Dados: Os grupos devem apresentar a base de dados escolhida, descrevendo todas variáveis/colunas existentes, o que cada uma representa e quais os tipos de dado guardados em cada variável. Devem também explicar qual foi a variável target escolhida e justificar essa escolha.

A base de dados fornecida contém informações sobre salários em vários países e territórios ao redor do mundo. Vamos descrever as variáveis ou colunas existentes na base de dados:

- `country_name`: Esta coluna contém o nome do país ou território.
- `continent_name`: Aqui, é registrado o nome do continente ao qual o país ou território pertence.
- `wage_span`: Indica a periodicidade dos salários (por exemplo, mensal, anual, semanal, etc.), e, no caso desta base de dados, todos os valores estão em uma periodicidade mensal.
- `median_salary`: Esta é a variável target escolhida para nossa análise. Ela representa a mediana dos salários no respectivo país ou território.
- `average_salary`: Representa a média dos salários no país ou território.
- `lowest_salary`: Mostra o valor mínimo do salário nesse local.
- `highest_salary`: Apresenta o valor máximo do salário nesse local.

Nossa variável target escolhida para a análise é a `median_salary`, que representa a mediana dos salários. A escolha dessa variável se justifica pelo fato de a mediana ser uma medida de tendência central robusta e que não é sensível a valores extremos. Além disso, a mediana é útil para entender a distribuição dos salários em cada local, o que pode fornecer informações valiosas sobre a equidade salarial e as condições econômicas em diferentes países e continentes.

Agora, com a variável target definida, podemos realizar análises estatísticas e exploratórias da base de dados, como calcular estatísticas resumidas, identificar tendências salariais em diferentes regiões e criar visualizações para comunicar os resultados de forma clara.