

Privacidade de Dados - 2025-2

Trabalho 2 - Supressão

Javam Machado

1 Objetivo

O trabalho consiste em implementar um algoritmo que realiza a supressão de parte dos valores de municípios de um conjunto de dados e avalia o impacto em um estudo de frequência com base neste atributo.

2 Especificação

Considere o conjunto de dados “**dados_covid-ce_trab02.csv**”. Você deverá implementar um algoritmo que inicialmente suprime aleatoriamente os valores do atributo `MunicipioCaso` com percentual de supressão em 3 faixas diferentes, 25%, 50% e 75%. Para cada faixa você deverá gravar um dataset resultado com o nome “**dados_covid-ce_trab02_X.csv**” onde “**X**” é o valor do percentual de supressão no dataset. No passo seguinte, seu algoritmo vai calcular a frequência dos valores do atributo `MunicipioCaso` para cada um dos dataset, o original sem supressão e os outros três com a supressão realizada. Gere um arquivo com as frequências encontradas e grave este arquivo com o nome de “**frequencias_trab2.csv**”. Sua estrutura deve prever um registro para cada valor de município, relacionados em ordem alfabética, e o valor da frequência para cada dataset, ou seja o município e a frequência no dataset original, no dataset com supressão de 25% e assim vai. Para finalizar, seu algoritmo irá ordenar cada dataset “**dados_covid-ce_trab02_X.csv**” com base na frequência dos valores do atributo `MunicipioCaso` e plotar o histograma dos *TopK* municípios, com $k = 20$, gravando o histograma em disco para cada dataset.

3 Requisitos

- Linguagens: C++ ou Python;
- Trabalho em duplas;
- Comentar cada função implementada. Escreva um *Readme.txt* descrevendo o projeto;
- Comprimir em um único pacote o seu projeto (código-fonte e executável) e todos os arquivos gerados e submeter via **Classroom**;
- O trabalho deverá ser entregue até as 10h da quarta-feira, dia 15/10/2025, e apresentado pela dupla na sala de aula. A dupla deve providenciar o computador necessário para apresentar os resultados.

4 Avaliação

Na avaliação serão considerados os seguintes indicadores:

- **Corretude** do programa;
- **Pontualidade e documentação/qualidade** do código-fonte.