



Universidade Federal do Ceará
Centro de Ciências
Departamento de Computação

CKP9011 e CK0268 – Introdução à Ciência de Dados 2024.1

Lista 1

Exercício: Tratamento de Dados

Objetivos: Exercitar os conceitos referente à manipulação, tratamento e limpeza de dados.

Data da Entrega: 18/03/2024

OBS 1: Exercício Individual.

OBS 2: A entrega da lista deverá ser executada utilizando-se o SIGAA.

Questão 1

Crie um arquivo Jupyter Notebook e realize as seguintes operações:

- a) Ler o dataset covid19br_vt1.csv, o qual está disponível no link a seguir:
<https://drive.google.com/drive/folders/1x5LzOa357zzJ3Q0MTbZBUxjkUoHnzm-H?usp=sharing>
- b) Identificar e listar as posições (células) contendo valores faltantes.
- c) Contar quantas linhas possuem valores faltantes.
- d) Para cada coluna (feature), contar quantas linhas possuem valores faltantes.
- e) Identificar e listar as linhas repetidas (duplicadas).
- f) Identificar e listar as posições (células) contendo valores que não pertencem ao domínio (tipo de dados) esperado.
- g) Crie uma coluna chamada “caracteres” contendo a quantidade de caracteres da coluna “text”.
- h) Crie uma coluna chamada “words” contendo a quantidade de palavras da coluna “text”.
- i) Crie uma coluna chamada “viral” contendo o valor 0 se o texto da mensagem (valor do campo “text”) não for encontrado em outras linhas e 1, caso contrário.
- j) Crie uma coluna chamada “sharings” contendo a quantidade de vezes que o texto armazenado no atributo “text” aparece no dataset.
- k) Crie uma coluna chamada “sentiment” contendo os valores: -1 para textos negativos, 0 para textos neutros e 1 para textos positivos.
- l) Eliminar as linhas cujo valor da coluna “text” contenham “trava-zaps”.
- m) Identificar inconsistências entre os atributos (features).

“A Educação, qualquer que seja ela, é sempre
uma teoria do conhecimento posta em prática”.

Paulo Freire