



Faculdade de Informática e Administração Paulista (FIAP)

Data Science

Gabriel Franco J Silva RM: 97574

Giuliano Sambianchi da Silva RM: 97676

Kelvin Mathias da Silva RM: 98137

Gabriel Barros Soares RM: 551232

Carlos Augusto Neto RM: 95856

ARCHITECTURE ANALYTICS & NOSQL

Checkpoint 5 - “Big Data Architecture & Data Integration” e “Data Governance & Data Security”

São Paulo – SP  
2023

Gabriel Franco J Silva RM: 97574

Giuliano Sambianchi da Silva RM: 97676

Kelvin Mathias da Silva RM: 98137

Gabriel Barros Soares RM: 551232

Carlos Augusto Neto RM: 95856

## ARCHITECTURE ANALYTICS & NOSQL

### Checkpoint 5 - “Big Data Architecture & Data Integration” e “Data Governance & Data Security”

Checkpoint 5 - “Big Data Architecture & Data Integration” e “Data Governance & Data Security”, apresentada a Faculdade de Informática e Administração Paulista (FIAP) como requisito para recebimento de nota das disciplinas de Big Data Architecture & Data Integration” e “Data Governance & Data Security”.

Orientador: Prof.º Renê de Ávila

São Paulo – SP  
2023

Bases de dados baixadas

## **Íris**

Número de Instâncias: 150

Quantidade m de atributos: 4

Características dos atributos (real, inteiro, nominal etc.):

Sepal lenght: real

Sepal width: real

Petal length: real

Petal width: real

Há atributos com valores ausentes? não há.

## **Adult**

Número de Instâncias: 48.842

Quantidade m de atributos: 14

Características dos atributos (real, inteiro, nominal etc.):

Age: inteiro

Workclass: nominal

Final weight: inteiro

Education: nominal

Education num: inteiro

Marital status: nominal

Occupation: nominal

Relationship: nominal

Race: nominal

Sex: nominal

Capital gain: inteiro

Capital loss: inteiro

Hours per week: inteiro

Native country: nominal

Há atributos com valores ausentes? Sim.

## **Breast Cancer Wisconsin (Original)**

Número de Instâncias: 699

Quantidade m de atributos: 11

Características dos atributos (real, inteiro, nominal etc.):

Clump Thickness: inteiro

Uniformity of Cell Size: inteiro

Uniformity of Cell Shape: inteiro

Marginal Adhesion: inteiro

Single Epithelial Cell Size: inteiro

Bare Nuclei: inteiro

Bland Chromatin: inteiro

Normal Nucleoli: inteiro

Mitoses: inteiro

Class: nominal

Há atributos com valores ausentes? sim

## **Wine**

Número de Instâncias: 178

Quantidade m de atributos: 13

Características dos atributos (real, inteiro, nominal etc.):

Alcohol: real

Malic acid: real

Ash: real

Alcalinity of ash: real

Magnesium: inteiro

Total phenols: real

Flavanoids: real

Nonflavanoid phenols: real

Proanthocyanins: real

Color intensity: real

Hue: real

OD280/OD315 of diluted wines: real

Proline: inteiro

Há atributos com valores ausentes? não há.

## **Poker Hand**

Número de Instâncias: 1.025.010

Quantidade m de atributos: 10

Características dos atributos (real, inteiro, nominal etc.):

S1: inteiro

C1: inteiro

S2: inteiro

C2: inteiro

S3: inteiro

C3: inteiro

S4: inteiro

C4: inteiro

S5: inteiro

C5: inteiro

Há atributos com valores ausentes? não há.

## **Abalone**

Número de Instâncias: 4.177

Quantidade m de atributos: 8

Características dos atributos (real, inteiro, nominal etc.):

Sex: nominal

Length: real

Diameter: real

Height: real

Whole weight: real

Shucked weight: real

Viscera weight: real

Shell weight: real

Rings: inteiro

Há atributos com valores ausentes? não há.

## **Internet Advertisements**

Número de Instâncias: 3.279

Quantidade m de atributos: 1.558

Características dos atributos (real, inteiro, nominal etc.):

Por serem 1558 atributos segue o resumo:

Os atributos são majoritariamente categóricos nominais. Por se tratar de anúncios de internet, URL's, palavras-chave e informações sobre páginas, as características variam também para valores contínuos ou categóricos codificados numericamente.

Há atributos com valores ausentes? Sim.

## **Yeast**

Número de Instâncias: 1.484

Quantidade m de atributos: 9

Características dos atributos (real, inteiro, nominal etc.):

Sequence Name: nominal

mcg: real

gvh: real

alm: real

mit: real

erl: real

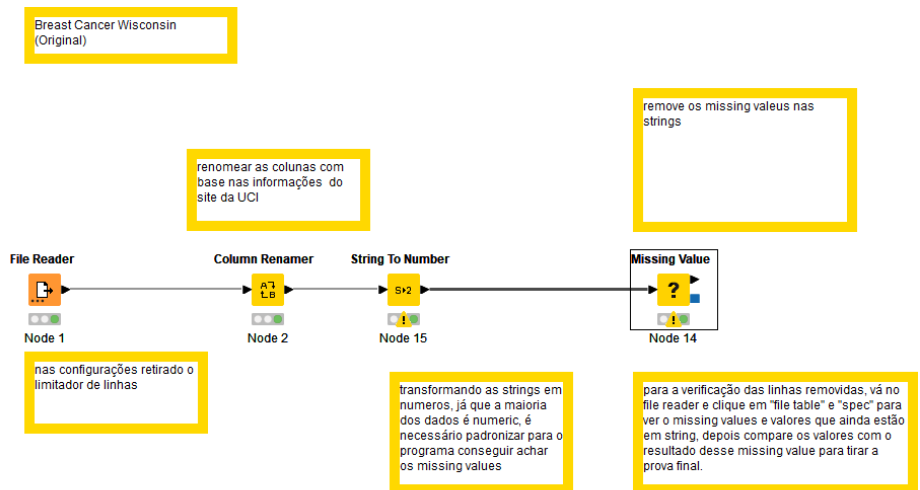
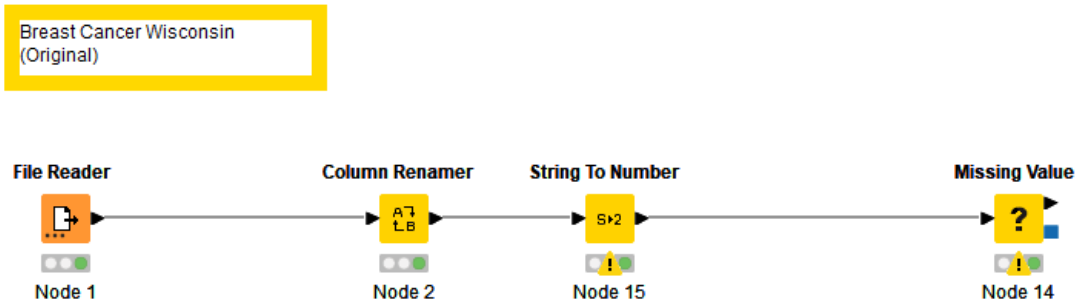
pox: real

vac: real

nuc: real

Há atributos com valores ausentes? não há.

Tratando a Base de dados Breast Cancer Wisconsin (Original), removendo os Missings Values.



Tratando a Base de dados Internet Advertisements, removendo os Missings Values.

