

Disciplina: Tópicos Avançados em Bancos de Dados

Professora: Lizianne P. Marques Souto

Grupo: Ítalo Felipe dos Santos Galdino,

João Vítor Diniz Costa,

Renaly Ferreira Barbosa,

Guilherme Pires

**Trabalho 1**

1. (2,5 pontos) Com base nos repositórios de dados online apresentados, realize a Análise

Exploratória dos Dados. Siga os passos:

a. Escolha uma base de dados

# **Diabetes Health Indicators Dataset**

b. Classifique as variáveis como

**● Quantitativas: Discreta ou Contínua**

**BMI** Quantitativas Contínua

**MentHlth** Quantitativas Discreta

**PhysHlth** Quantitativas Discreta

**● Qualitativa: Nominal ou Ordinal**

**Diabetes\_binary** Qualitativa Nominal

**HighBP** Qualitativa Nominal

**HighChol** Qualitativa Nominal

**CholCheck** Qualitativa Nominal

**Smoker** Qualitativa Nominal

**Stroke** Qualitativa Nominal

**HeartDiseaseorAttack** Qualitativa Nominal

**PhysActivity** Qualitativa Nominal

**Fruits** Qualitativa Nominal

**Veggies** Qualitativa Nominal

**HvyAlcoholConsump** Qualitativa Nominal

**AnyHealthcare** Qualitativa Nominal

**NoDocbcCost** Qualitativa Nominal

**GenHlth** Qualitativa Ordinal

**DiffWalk** Qualitativa Nominal

**Sex** Qualitativa Nominal

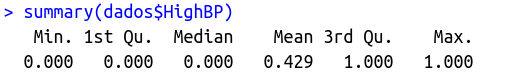
**Age** Qualitativa Ordinal

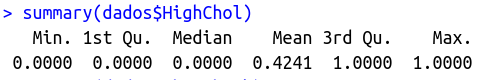
**Education** Qualitativa Ordinal

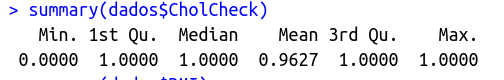
**Income** Qualitativa Ordinal

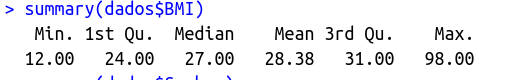
c. Usando o R realize a análise exploratória dos dados:

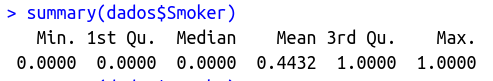
* Média, moda, mediana, máximo, mínimo, quartis.

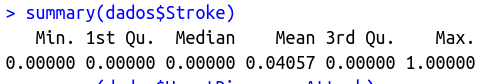


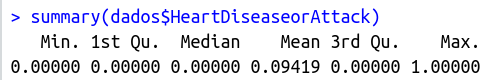
****

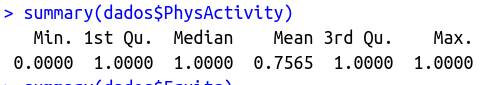


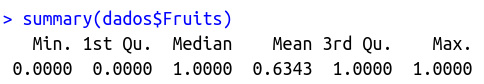


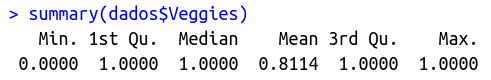


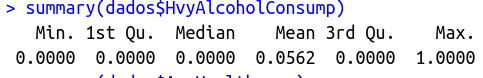


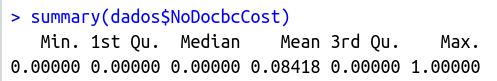


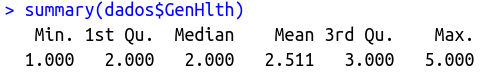


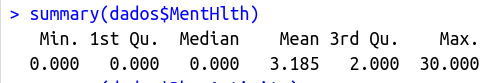


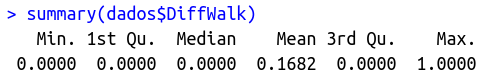
****

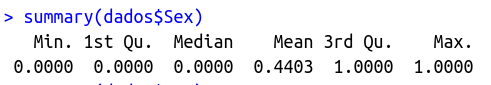


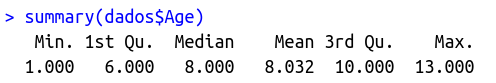


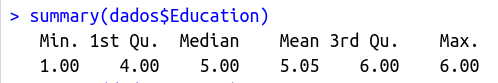


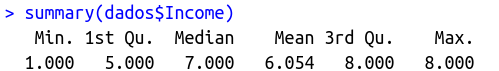












2. (2,5 pontos) Após a análise exploratória dos dados, realize os procedimentos necessários para o Pré-processamento:

**a. Limpeza dos dados: preenchimento de campos vazios e null.**

A base de dados não apresenta campos vazios ou null.

**b. Eliminação de ruídos (outliers).**

Eliminamos valores acima do BMI 65 devido a certeza que já teriam algum grau de diabetes.

**c. Transformação dos dados: eliminação/criação de variáveis, generalizar/agregar.**

Criamos a coluna class para ajudar o weka na classificação.