**INTRODUÇÃO**

O projeto abaixo consiste na análise do dataset “Iris.csv”. É uma base de dados que contém 150 observações com as dimensões das pétalas e sépalas de 3 espécies diferentes de flores.

O intuito desse projeto foi de desenvolver habilidades e treinar as técnicas mostradas no decorrer do curso de Data Science. Serão utilizadas técnicas como analise exploratória, Data Wrangling, Análise fatorial e Predição a partir de uma Logística Multinominal.

**DESENVOLVIMENTO**

Foi realizada a análise exploratória inicialmente para visualizarmos os dados para podermos iniciar a estruturação e manipulação do dataset.



*Dataset “Iris.csv”. Inicialmente estruturado.*

*Descrição das Variáveis*

**Id** = Identificação da Observação

**SepalLenghtcm** = Comprimento da sépala em cm

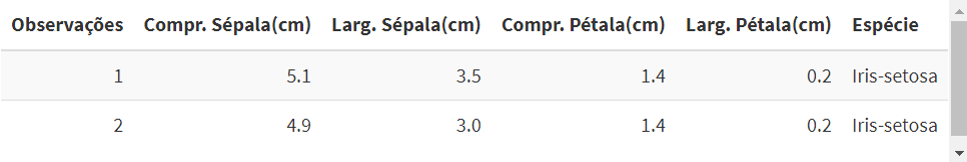
**SepalWidthcm** = Largura da sépala em cm

**PetalLenghtcm** = Comprimento da pétala em cm

**PetalWidthcm** = Largura da pétala em cm

**Species** = Espécie

Vemos um dataset com estrutura padrão, variáveis em colunas e observações em linhas. Não é uma base de dados que necessita passar por diversas modificações em sua estrutura para que possa ser utilizado. Apenas mudar os nomes das variáveis já é interessante.



*Dataset “Iris.csv”. Após alteração dos nomes das variáveis.*

O dataset contém 150 observações de 3 espécies diferentes. O gráfico abaixo ilustra o histograma de cada espécie.



*Histograma Quantidade de Observações x Espécies*