**AC2a**

Nome: Thiago Faro Ribeiro

RA: 235553

Nome: Giovanna Andreoli

RA: 235831

**Predição de Câncer**

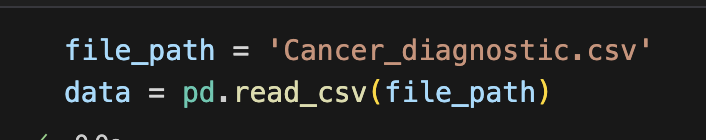
Informações do atributo:

a) ID Número de identificação

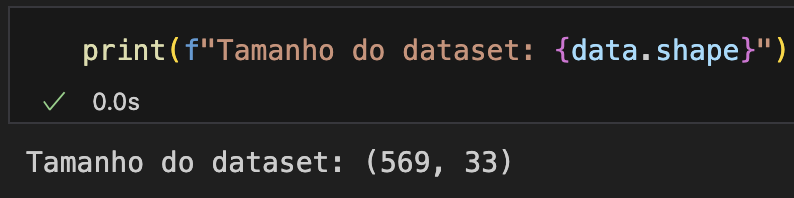
b) Diagnóstico (M = maligno, B = benigno) (a variável para predefinir)

Lista de exercícios:

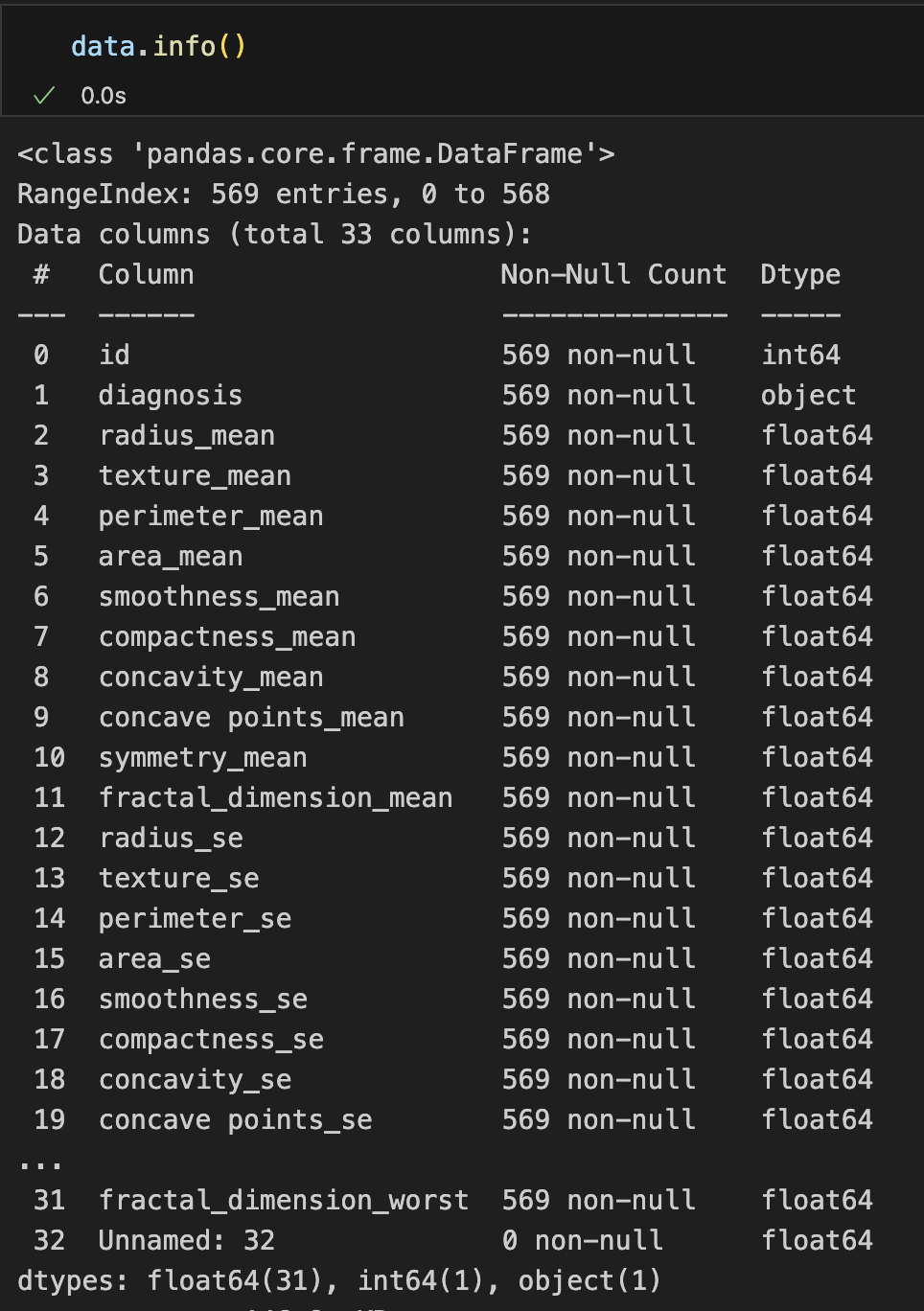
1. Ler o dataset “Cancer\_diagnostic.csv” e segui os seguintes passos.



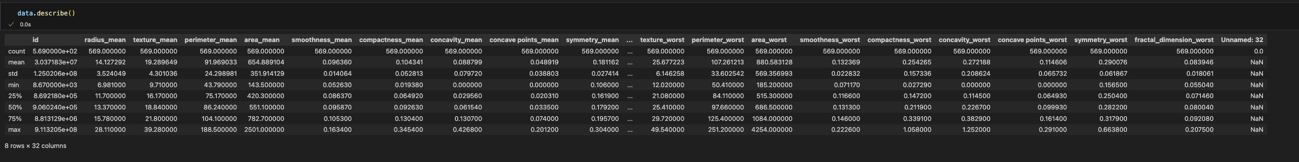
1. Qual é o tamanho do dataset?



1. Verifique colunas e tipos de dados



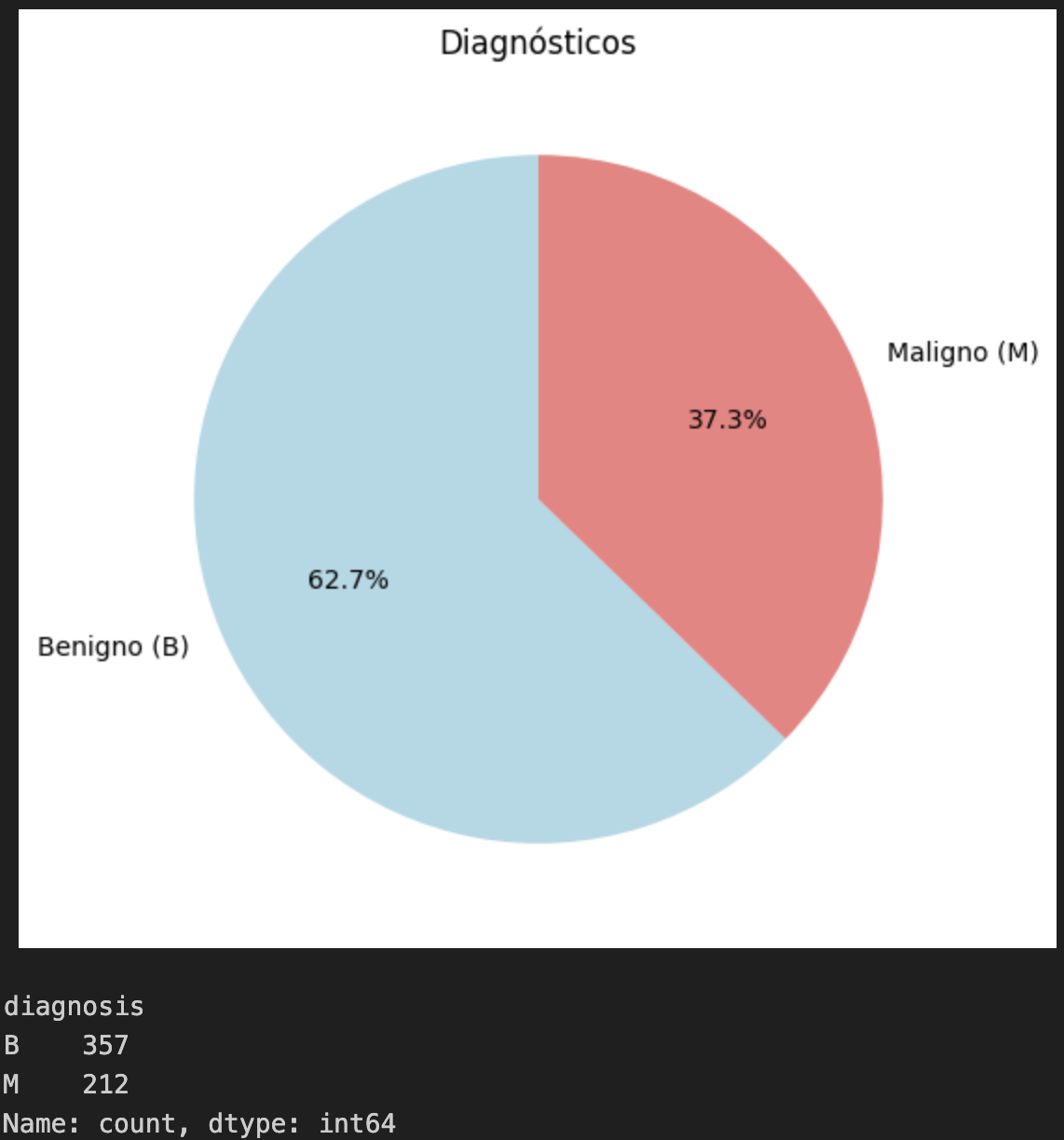
1. Mostra o resumo estatístico do dataset



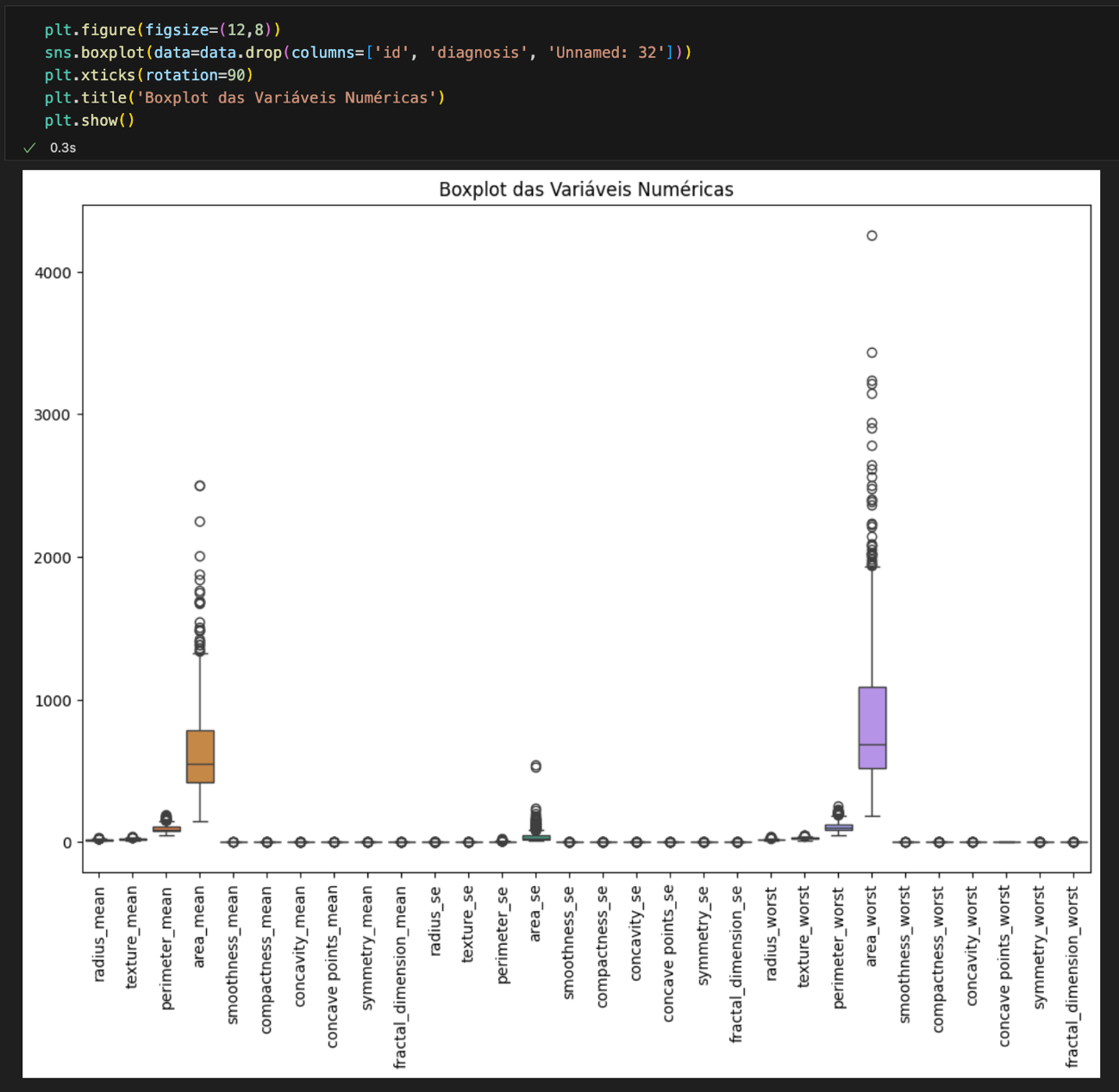
1. O dataset está balanceado?

O dataset não está balanceado.

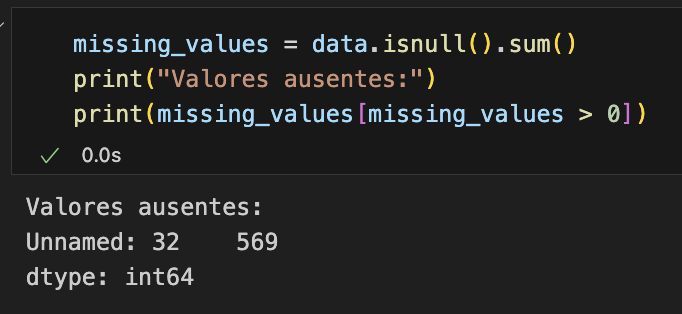
1. Mostrar um gráfico de pizza do dataset



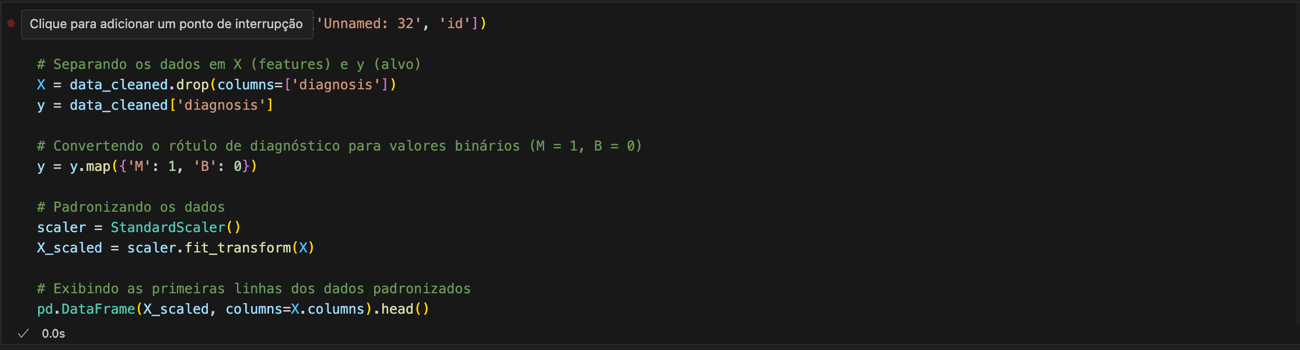
1. Mostra o gráfico de boxplot do dataset

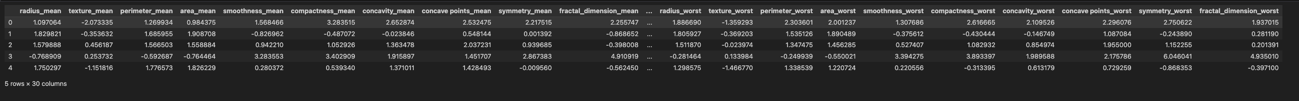


1. Tem valores ausentes? Se sim quais?

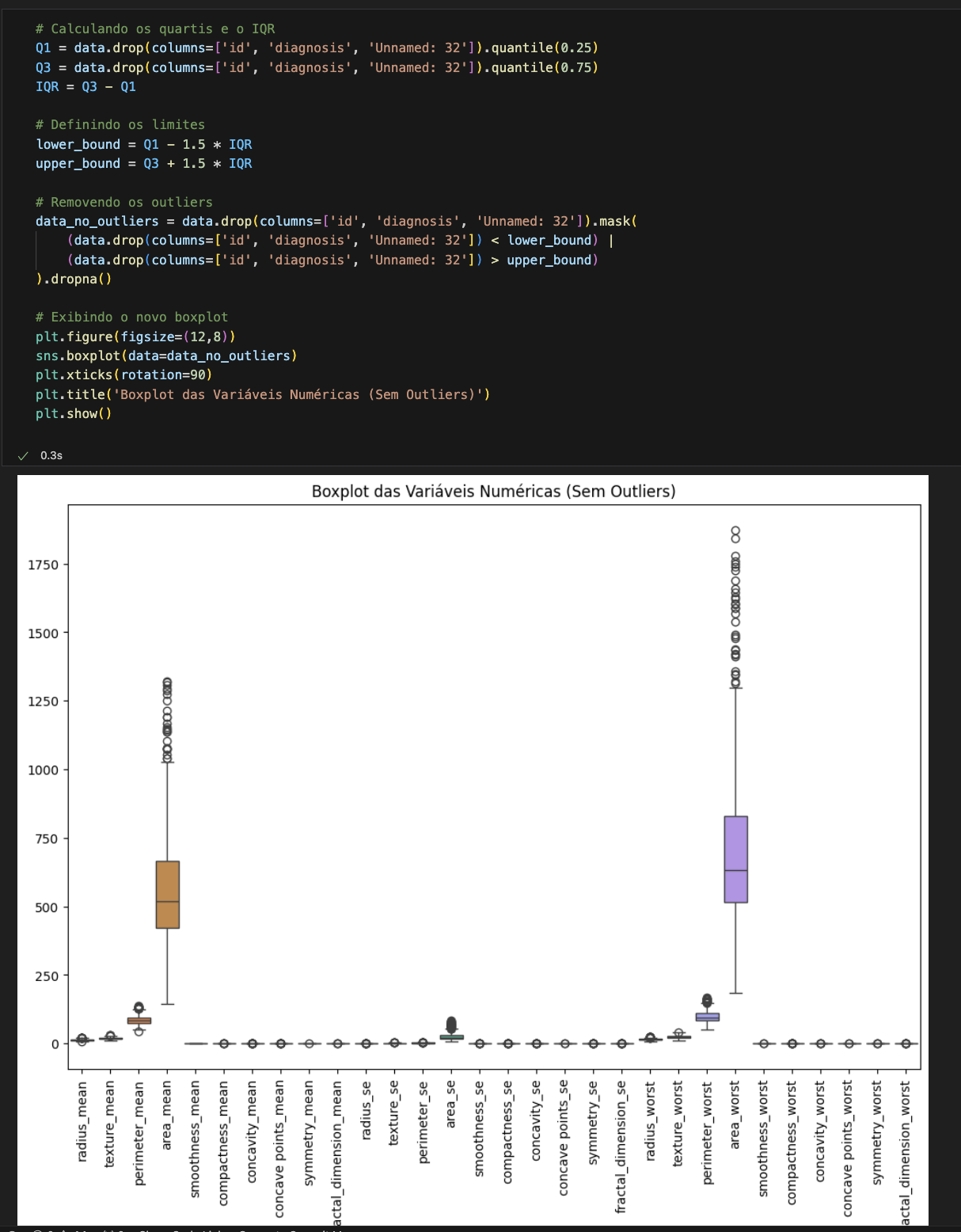


1. Realiza a padronização (standardization)

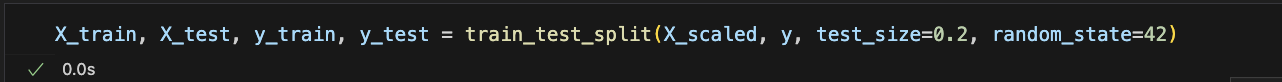




1. Se tiver outlier fazer a remoção deles e mostrar o boxplot do dataset.



1. Dividir o dataset em training (80%) e testing (20%)



1. Faz a classificação usando o algoritmo KNN com K = 3, K = 5, K = 7 e mostrar as métricas de accuracy, recall e precision para cada K.

