Quando trabalhamos com dados em uma análise ou em um modelo de machine learning, é importante garantir que esses dados sejam precisos e completos. **Dados nulos**, também conhecidos como dados faltantes ou ausentes, são um problema comum que pode prejudicar a precisão de qualquer análise ou modelo de machine learning.

Os dados nulos ocorrem quando não há informações disponíveis para determinada observação em um conjunto de dados. Isso pode acontecer por várias razões, como falhas no registro dos dados, perda de informações ou até mesmo erros humanos. Quando esses dados nulos são inseridos em um modelo de machine learning, o modelo pode interpretar erroneamente a falta de informação como algo significativo, resultando em um modelo enviesado ou impreciso, ou muitas vezes o algoritmo nem ao menos roda.

Por isso, é importante tratar esses dados antes de inserir em um modelo de machine learning. Existem várias técnicas que podem ser utilizadas para tratar dados nulos, como: a exclusão das observações que possuem dados faltantes e a imputação de valores faltantes com base em outros dados do próprio conjunto de informações.

A exclusão de observações com dados faltantes pode ser uma opção viável quando sua quantidade é pequena em comparação ao tamanho do conjunto de dados, ou quando a exclusão não prejudica a representatividade dos dados. No entanto, essa abordagem pode levar a perda de informações importantes e diminuir a capacidade do modelo de aprender.

A imputação de valores faltantes é uma técnica comum para tratar dados nulos. Ela consiste em substituir os dados faltantes por um valor estimado com base em outros dados disponíveis. Existem vários métodos de imputação, como: média, mediana, moda ou até mesmo modelos de machine learning específicos para imputação.

Existem alguns métodos da biblioteca Pandas que nos auxiliam a identificar quais são dados nulos em nosso DataFrame, como:

1. isnull(): retorna um DataFrame de valores booleanos, no qual True indica o valor nulo e False indica o valor válido.
2. notnull(): retorna um DataFrame de valores booleanos, no qual True indica o valor válido e False indica o valor nulo.
3. isna(): retorna um DataFrame de valores booleanos, com True indicando o valor nulo e False o valor válido (a mesma funcionalidade do método isnull()).
4. notna(): retorna um DataFrame de valores booleanos, em que True indica o valor válido e False indica o valor nulo (a mesma funcionalidade do método notnull()).

Em resumo, o tratamento de dados nulos é fundamental para garantir que um modelo de machine learning tenha uma precisão elevada e uma capacidade de generalização adequada. O uso de técnicas adequadas de tratamento de dados nulos é um passo importante no processo de preparação de dados para qualquer análise ou modelo de machine learning.