

## Tratando os dados: Consequências

Nesta aula, além de criarmos funções para remover TAG's HTML e trechos de códigos, aplicamos estas funções nos dados utilizados para criação de Modelos de linguagens, que diferencia diversos idiomas.

De acordo com o que foi estudado em aula, qual a importância de remover as TAG's e os códigos dos nossos dados?

*Selecione uma alternativa*

**A**

Como o objetivo é detectar idiomas através de modelos de linguagem, é importante diferenciar ao máximo os modelos de cada idioma. Manter TAG's e códigos, que são comuns independente do idioma, aumenta a similaridade entre os modelos, podendo prejudicar a diferenciação entre os idiomas.

**B**

Como o objetivo é detectar idiomas através de modelos de linguagem, é importante diferenciar o mínimo possível os modelos de cada idioma. Manter TAG's e códigos, que são comuns independente do idioma, aumenta a similaridade entre os modelos, podendo aumentar a capacidade de diferenciação entre os idiomas.

**C**

Como o objetivo é detectar idiomas através de modelos de linguagem, é importante diferenciar ao máximo os modelos de cada idioma. Manter TAG's, códigos e remover textos, aumenta a similaridade entre os modelos, potencializa a diferenciação entre os idiomas.