



**Data Science
Academy**

www.datascienceacademy.com.br

**Big Data Real-Time Analytics com
Python e Spark**

**NLTK - Processamento de Linguagem
Natural em Python**



Ao final do capítulo você encontra um Jupyter Notebook e um arquivo csv com uma introdução prática sobre como processar texto (limpar, transformar e manipular). Essa introdução é importante para que você compreenda o que faremos na atividade prática do capítulo seguinte, quando vamos analisar dados do Twitter em tempo real.

Execute o Jupyter Notebook (não é necessário usar o PySpark) e estude cada célula e cada parte do código de processamento de texto. Será usado o NLTK, um dos principais frameworks para Processamento de Linguagem Natural.

Bons estudos!

Equipe DSA