**PARA PERGUNTAR:**

Dataset 7 está mesmo com problemas

A frase mais longa tem 667 characters

**PERGUNTADO:**

Remoção de uma story PrincessPeas do excel porque estava com colunas extra -> Olhar para as colunas e, se fizerem sentido, reter.

A cena das palavras repetidas com significados a frente -> Manter as palavras com as outras em parêntesis à frente

Interessa o encoding em que leio ou só dar match? -> Internamente o Python usa UTF-8. Gravar os dados depois em UTF-8