Lizzie Walsh
Practical Python for News Investigations
November 8, 2025
Assignment 2: Text analysis

Pitch: Analyzing .txt files of FDA Adverse Event Reports

Assignment: Extract text or tables from multiple PDFs or other documents to quantify occurrences, search for names of people, or for entities like companies, etc. You may need to scrape a site to first acquire these documents. You may do that step in Assignment 1.

I'd like to analyze the total case listings for kratom reported through the FDA's FAERS Dashboard. Though this was exported in excel workbook form, a scrape of the .txt or .pdf version is more appropriate and useful here because column information varies widely, and I'd like to pull out specific words and phrases rather than analyze numbers. The .pdf version of these case reports are 240 pages.

Within the pdf, I'd like to analyze:
● The number of serious vs non-serious cases
● The counts for each year (beginning in 2015)
● The most common product names
● The most common reason for use
● The most common reactions
● The average patient age
● What Reference Literature is most commonly used for each case (where applicable)

Looking at these specific details of each case will help me understand how kratom cases are reported to the FDA and what the most common outcomes are. These cases are only related to kratom, not its isolate products 7-oh or mitragynine pseudoindoxyl.