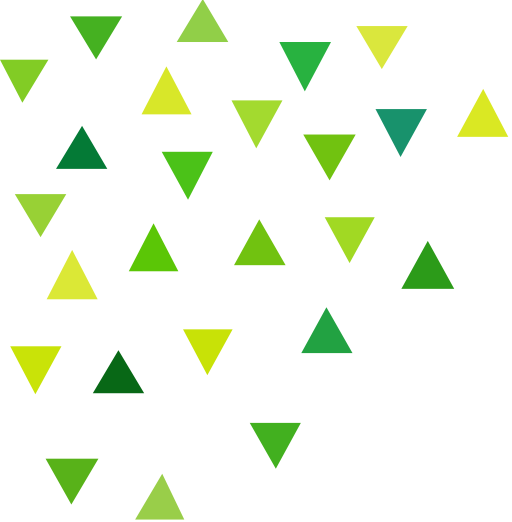

Cannabis Data Science #81

August 31st, 2022

Cannabis Data Science Application

Question of the Day

- What is your favorite strain?
  - Given consumer's history, $\overline{X}$, can we predict the product(s) with the most similar chemical profile(s) $\hat{Y}$?

Cannabis Data Science Tools

The data science tools we'll use include:

- **Web scraping**: Selenium and BeautifulSoup;
- **PDF scraping**: PDFPlumber and OCR (Tesseract);
- **Cryptography (SHA256)**: For creating unique IDs that encapsulate changes in results;
- **NLP** for standardization (coming soon);
- **Matching Models**: k-nearest neighbor model for product recommendations;
- **Optimization**: Gradient descent functions for finding optimal product mixes.

Recommendation Engines

Algorithm:

1. Calculate average product concentrations by consumer, $\bar{X}$;
2. Find k-nearest products, $\hat{Y}$, given the average concentrations;
3. Repeat when the consumer makes a purchase.

Possible extensions:

- Moving averages (i.e. only use m most recent products in $\bar{X}_m$)
- Creating blends, mixing. E.g. After finding 2-nearest products, y_1 and y_2, find α^* where

$$\min_{\alpha^*} = \bar{X} - (\alpha y_1 + (1-\alpha) y_2)\,.$$

 **Thank you for coming.**

Insights of the Day

- Start somewhere, then **iterate**.

What is on your mind for next week?