



Cannabis Data Science

Meetup

January 5<sup>th</sup>, 2022

# Data Science Pioneers



**Peter Naur** - Began using "*data science*" as an alternative name for computer science as early as 1974.

Author: Eriktj at English Wikipedia  
CC BY-SA 3.0  
<https://creativecommons.org/licenses/by-sa/3.0/>  
No changes were made to the image.



**C.F. Jeff Wu** - Used "*data science*" as an alternative name for statistics in 1985.

Commissioned by the owner, C. F. Jeff Wu  
CC BY-SA 3.0  
<https://creativecommons.org/licenses/by-sa/3.0/>  
No changes were made to the image.

Latanya Sweeney (Harvard Professor) has suggested 3 trends in data science:

- **Collect more** - Expansion of the number of fields being collected.
- **Collect specifically** - Replace an existing aggregate data collection with a person-specific one.
- **Collect it if you can** - Gather information by starting a new person-specific data collection.

## The Data Deluge - Collect it if you can.

- 1 Data design
- 2 Data collection
- 3 Data analysis

## What are your predictions for cannabis sales in 2022?

- Which states will permit medicinal or adult-use?
- How much in sales do you expect in New York in 2022?
- How much in sales do you expect in the U.S. in 2022?

# Possible Entrants in 2022

- **Delaware** - Will “revisit the bill during the 2022 legislative session” (2022-01-11).
- **Oklahoma** - Initiatives for adult-use cannabis for Oklahoma’s November 2022 election.
- **Mississippi** - A special session is expected to consider medicinal cannabis.
- **Maryland** - Adult-use legislation expected on the 2022 general election ballot.
- **Ohio** - Signatures being collected for adult-use cannabis legislation.
- **Wyoming** - Signatures being collected for medical cannabis legislation in 2022.
- **Pennsylvania** - Efforts to permit adult use cannabis.
- **Rhode Island** - Adult use waiting for special session.
- **Nebraska** - Possible medical use initiative for the 2022 ballot.
- **Arkansas** - Advocates working to get an adult-use initiative on the 2022 ballot.
- **Florida** - Movement on an adult-use initiative for the 2022 ballot.
- **Missouri** - 2 possible adult-use legalization measures in 2022.
- **North Carolina** - Medical cannabis legislation moving slowly.
- **South Dakota** - Trying to get adult-use on the ballot.
- **Idaho** - Working to get medical the state’s 2022 ballot.

Source: Cannabis Business Times *15 States That Could Legalize Cannabis in 2022*

<https://www.cannabisbusinesstimes.com/article/states-likely-legalize-cannabis-2022/>

## Cannabis Sales Predictions

- Legal U.S. sales forecasted to be \$31B (+28%) in 2022.
- New York City market estimated to be over \$4B (estimated worth).

CNBC *Potential new regulations could reshape the U.S. cannabis market in 2022*

<https://www.cnn.com/video/2021/12/30/potential-new-regulations-could-reshape-the-u-s-cannabis-market-in-2022.html>

# Forecasting the Long-Term with Structural Regression Models

- For long-term forecasting, a structural model may be superior to an atheoretical model because it can capture economic relationships.
- **Keep it sophisticatedly simple** - *“The approach, in our view, should not be too technical/heavily mathematical and ignores economic theory or practical realities, or too simple and ignores the econometric principles.”* (1)
- *“In the long run almost anything is possible.”*
  - John Maynard Keynes, April 1942 (2)

References: 1. Silvia, J, Iqbal, A, et. al (2014), 'Economic and Business Forecasting'.

2. Carter, Z.D. (2020) 'The Price of Peace'.



# Structural Regression Model

$$\begin{aligned} Sales_{i,t} = & \beta_0 + \beta_1 \text{Population}_{i,t} + \beta_2 \text{Adult Use}_{i,t} \\ & + \beta_3 \text{Months Medicinal}_{i,t} \\ & + \beta_4 \text{Months Adult Use}_{i,t} \times \text{Months Adult Use}_{i,t} \\ & + \beta_5 \text{Month}_{i,t} \end{aligned}$$

where

- **Population** is the total number of people in a state.
- **Adult Use** is a dummy variable, 1 if the state has adult use and 0 otherwise.
- **Months Medicinal** is the number of months since permitting medicinal cannabis.
- **Months Medicinal** is the number of months since permitting medicinal cannabis.
- **Month** is a dummy variable for the current month.



**Thank you for coming.**

This is the first, albeit crude, open source forecast of cannabis sales in 2022. You can create your own forecasts and see who's model predicts better (based on RMSE)!

$$RMSE = \sqrt{\frac{1}{T} \sum (Y_{t+1} - \hat{Y}_{t+1})^2}$$