# Decision Science: Translating Data Insights to Corporate Decision Making

**Farshad L. Miraftab**
**November 2020**

# Data vs. Decision Scientist

Leveraging ML algorithms and data science principals to make:

### Machines / Systems Smarter

» Typically product/feature focused

» Too much data can be the enemy

» Performance / accuracy focused

### Humans / Teams Smarter

» Decision / human focused

» Not enough data can be the enemy

» Interperetability focused

# Lessons I've Learned

# #1

# Ignoring Uncertainty

## The Most Dangerous Equation

*Ignorance of how sample size affects statistical variation has created havoc for nearly a millennium*

Howard Wainer

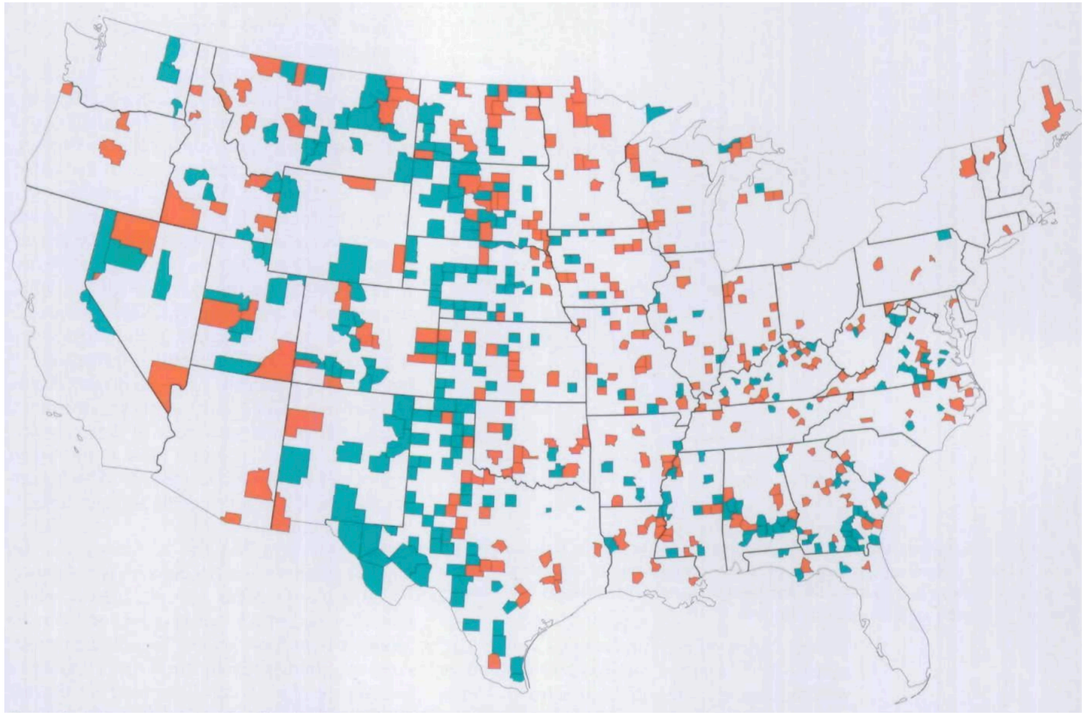
What constitutes a dangerous equation? There are two obvious interpretations: Some equations are dangerous if you know them, and others are dangerous if you do not. The first category may pose danger because the secrets within its bounds open doors behind which lies terrible peril. The obvious winner in this is Einstein's iconic equation $e = mc^2$, for it provides a measure of the enormous energy hidden within ordinary matter. Its destructive capability was recognized by Leo Szilard, who then instigated the sequence of

© The Goldsmiths' Company. Photograph: ImageWise

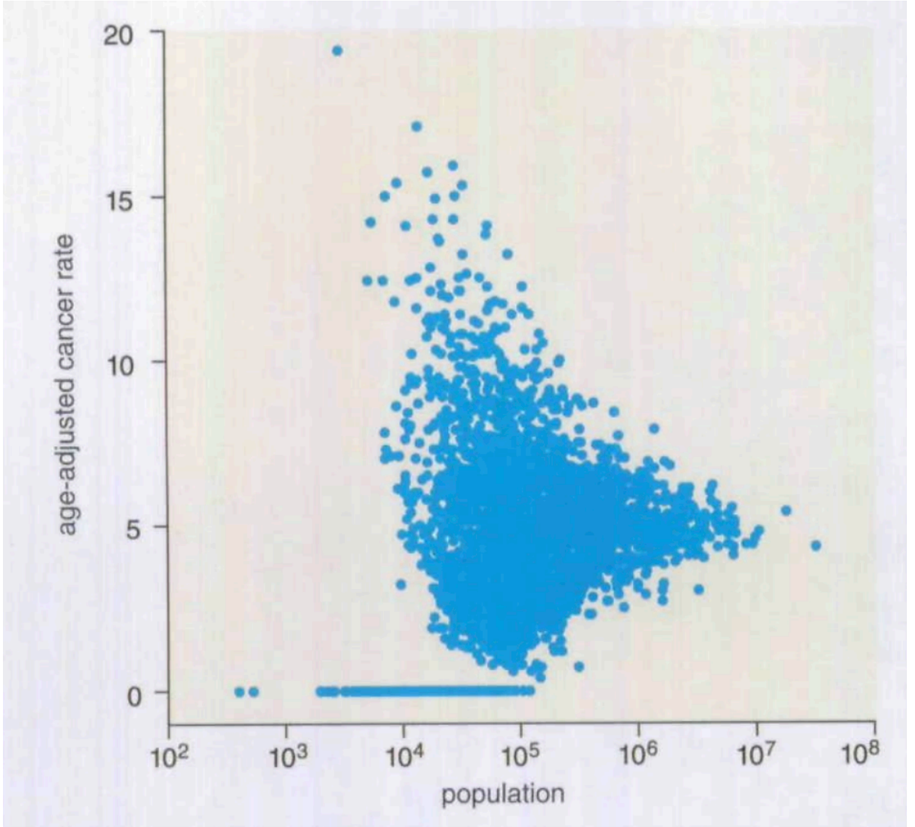
**Figure 1. Trial of the pyx has been performed since 1150 A.D. In the trial, a sample of minted coins, say 100 at a time, is compared to a standard. Limits are set on the amount that the sample can be over- or underweight. In 1150, that amount was set at 1/400. Nearly 600 years later, in 1730, a French mathematician, Abraham de Moivre, showed that the standard deviation does not increase in proportion to the sample. Instead, it is proportional to the square root of the sample size. Ignorance of de Moivre's equation has persisted to the present, as the author relates in five examples. This ignorance has proved costly enough that the author nominates de Moivre's formula as the most dangerous equation.**

$$\sigma_{\overline{x}} = \frac{\sigma}{n}$$

# Counties with highes/lowest cancer rates



# Cancer Rates vs County Population



# Where can sample sizes skew data?

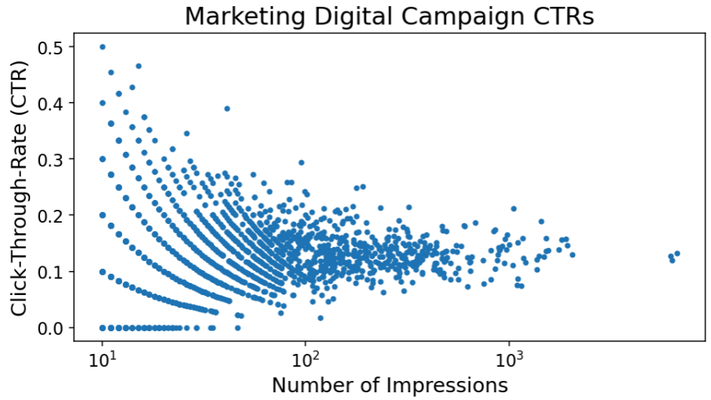
» Marketing
  - Funnel metrics

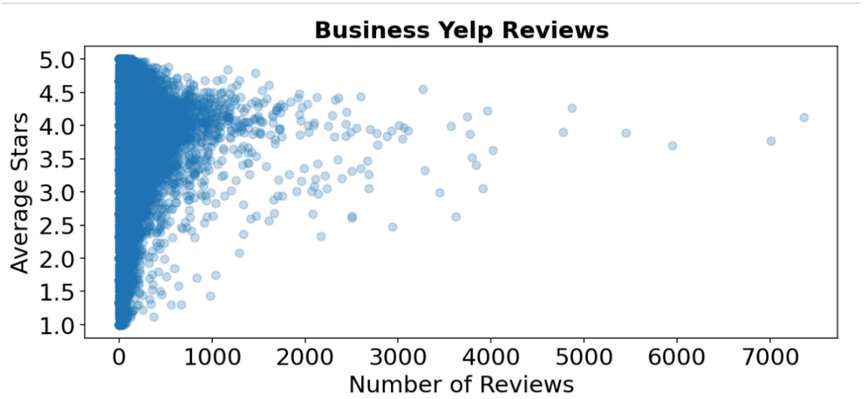
» Ratings
  - Yelp Reviews; Reddit upvotes

» Growth Product
  - Conversion Rates

» Sports
  - BA's, FT-%s, 3P%s

» Epidemiology
  - Effective Reproduction Rate ($R_t$)



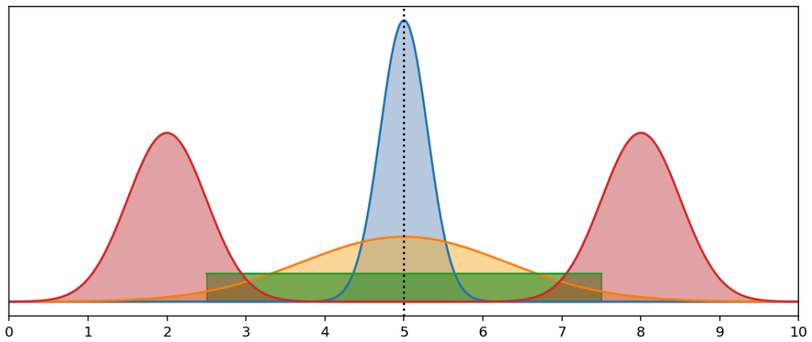



# Summarizing uncertainty to a single value

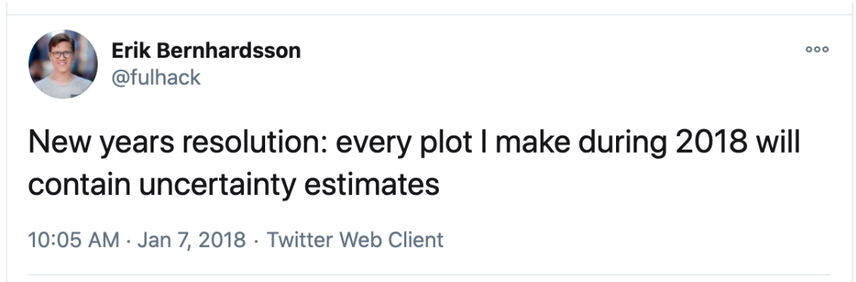
Avg age of diaper wearers in U.S.?

Average home price? household income?

What's in common with these distributions?



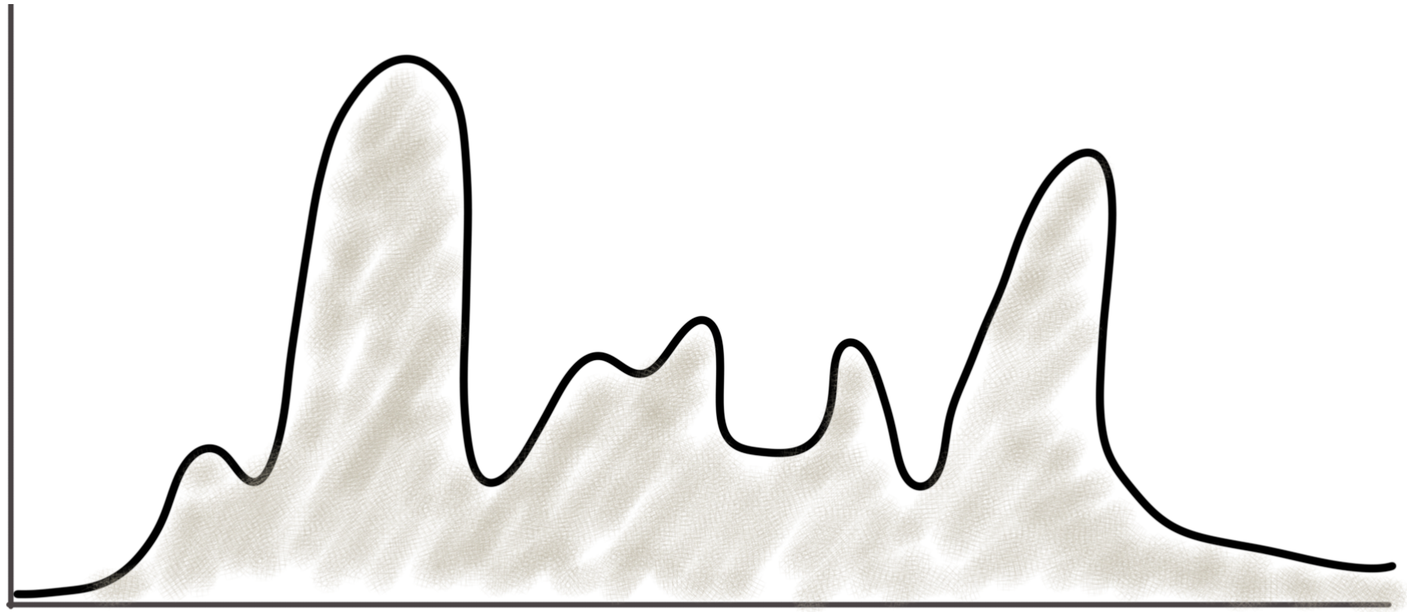

Can uncertainty be visualized?



"...*we don't have enough data*"

*The less you know, the more you gain*

# A random non-parametric distribution

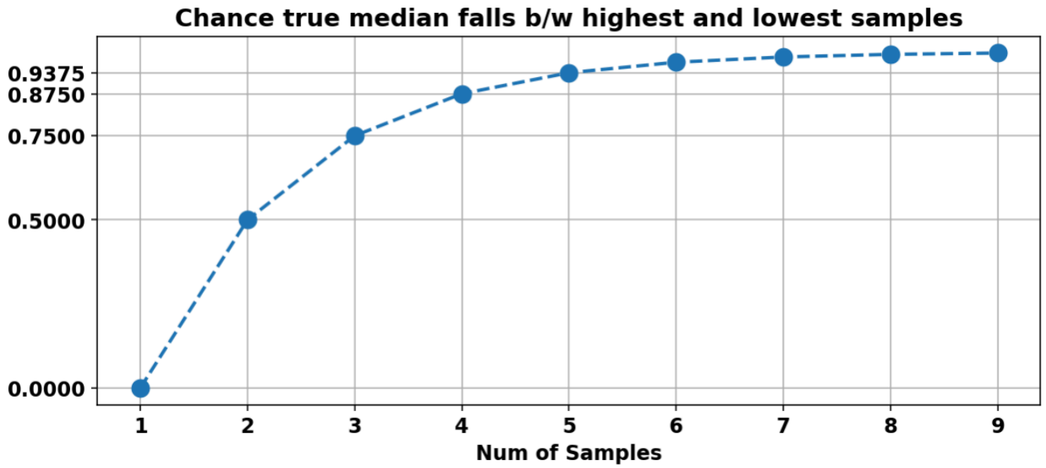


If you randomly sample 5 values from this distribution, what is the probability that the **true median** of this distribution falls between the largest and smallest values you sampled?

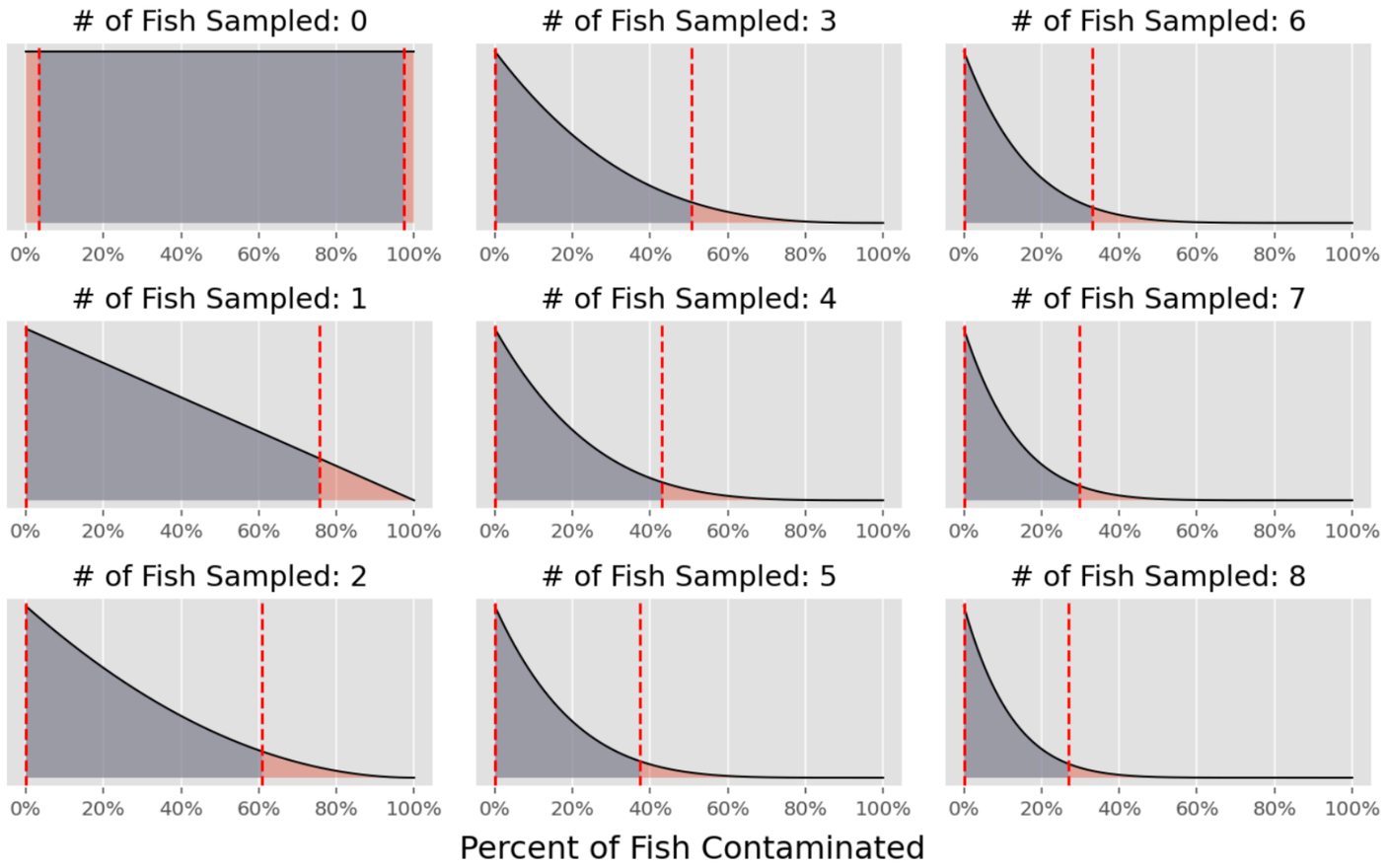
# Reducing Uncertainty

**Science is more about *gathering* data than about *having* data.**

$$= 1 - (2 \cdot (\frac{1}{2})^{samples})$$
$$= 1 - (2 \cdot (\frac{1}{2})^{5})$$
$$= 93.75\%$$

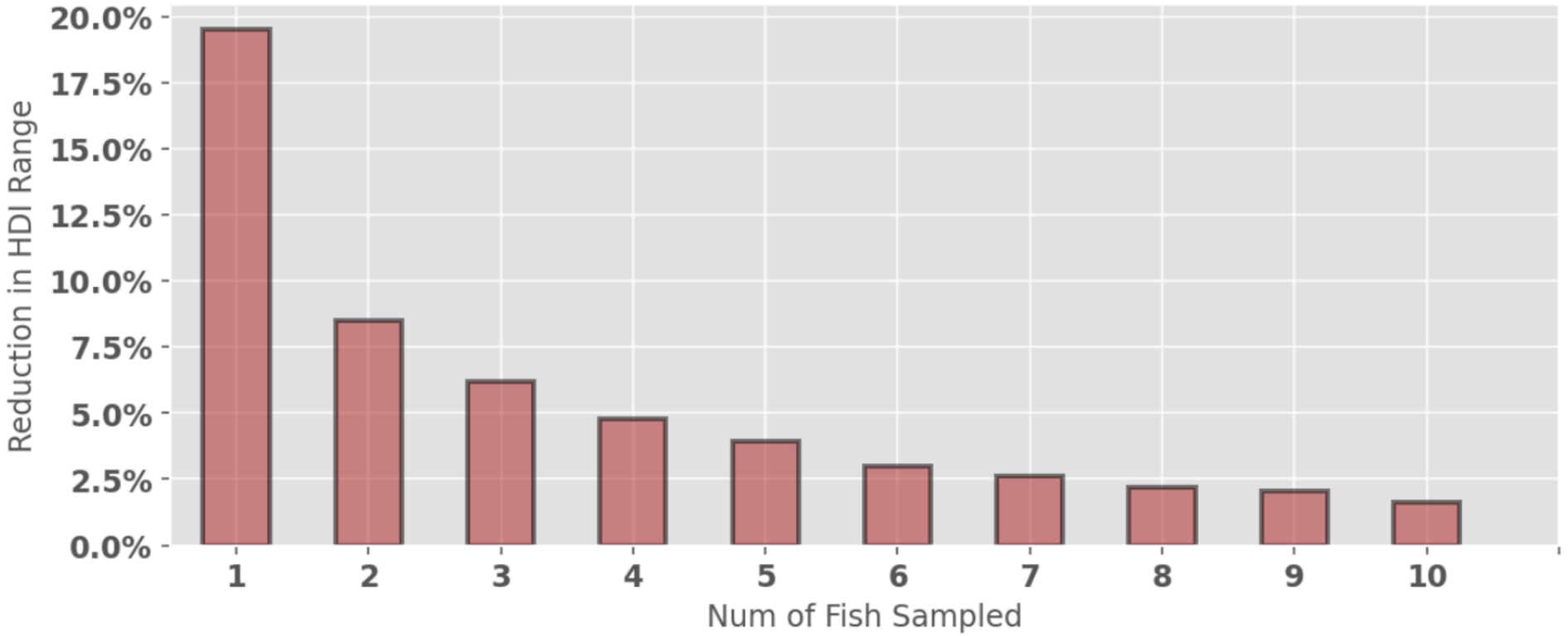


# Fish Contamination



# The value of information diminishes

## Reduction in uncertainty after each sampled fish



# #2
# Reports-Driven ≠ Data-Driven

# Dashboards are poorly executed

## Developers perspective

» Time-consuming to develop and maintain

» Ambiguous objective/goals

» In a constant demand

## Consumers perspective

» Information overload (no portion control)

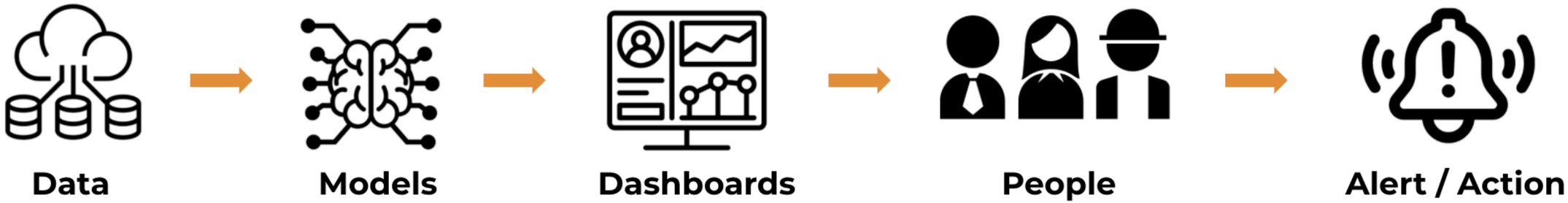
» Rarely used to drive decisions

» An at-a-glance summary of performance

» Inactionable

# Don't measure with a micometer and cut with an ax

In a typical analytics workflow, where lies the weakest link?



# What makes an effective dashboard?

## Know the Purpose

What question is aiming to be quantified?
Why is it important?

## Quantify your Trigger conditions

What are the critical thresholds that if violated, trigger an alert?

## Know your Audience

Who will be the primary consumers?
What is their general intuition for data?

## Define your Playbook

What course of action(s) need to be taken when an alert is fired?

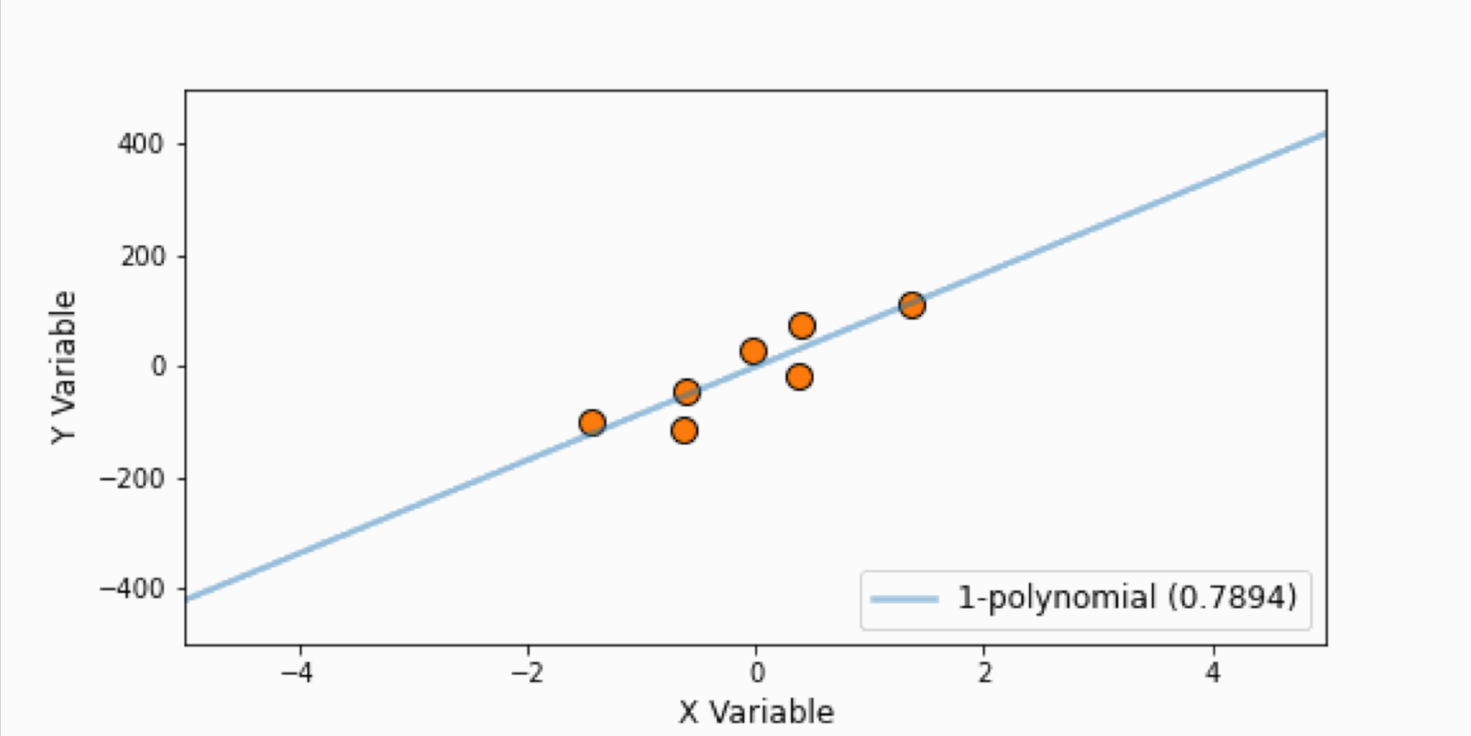
# #3

# "When you hear hoof beats, think horses, not zebras"

Farshad L. Miraftab ~ Director of Analytics and Data Science

# Simple > Complex

The fear or overfitting!

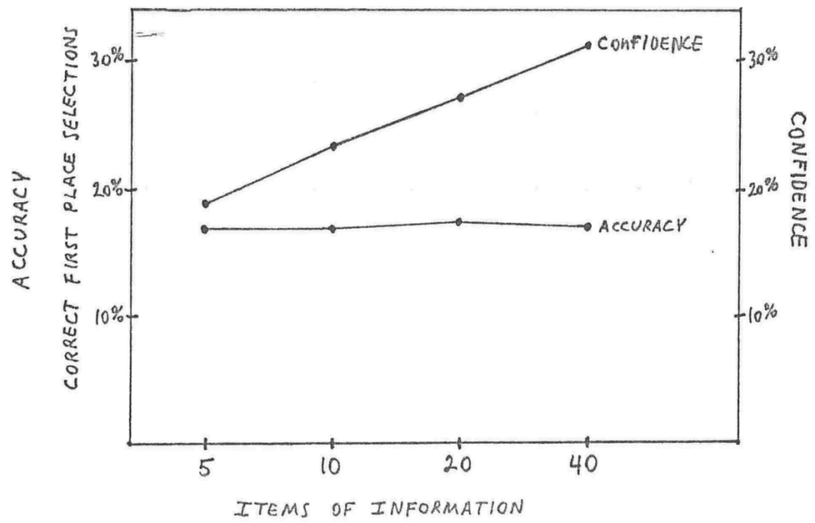


*"The more frequently you look at data, the more noise you are disproportionally likely to get"*

**— Nasim Taleb**

# There can be a *placebo effect* with analysis

More predictors we try to process can give us a false sense of confidence in our predictions without any improvement to the accuracy of the predictions[1]




[1] https://scholarsbank.uoregon.edu/xmlui/bitstream/handle/1794/23607/928.pdf?sequence=3&isAllowed=y

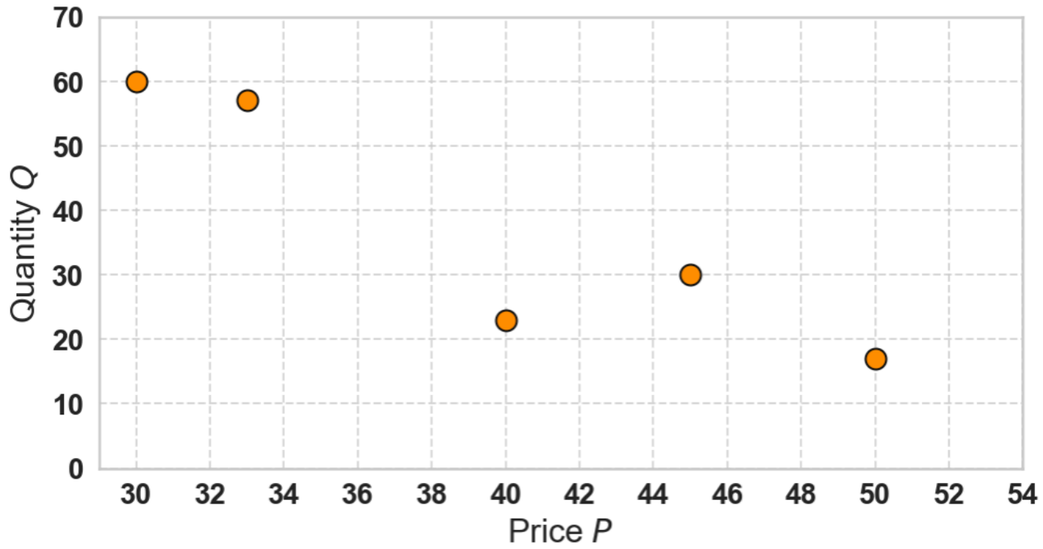
# Deep Learning is nice, not having to use one is *nicer*

## Pricing Optimization[2]

Tested a product at 5 different price points every week and collected the `Quantity Q`

Competiting Objectives?

1. **Maximize** Expected Profit?
2. **Minimize** Risk?
3. **Maximize** chance of achieving $625+ in Profit?



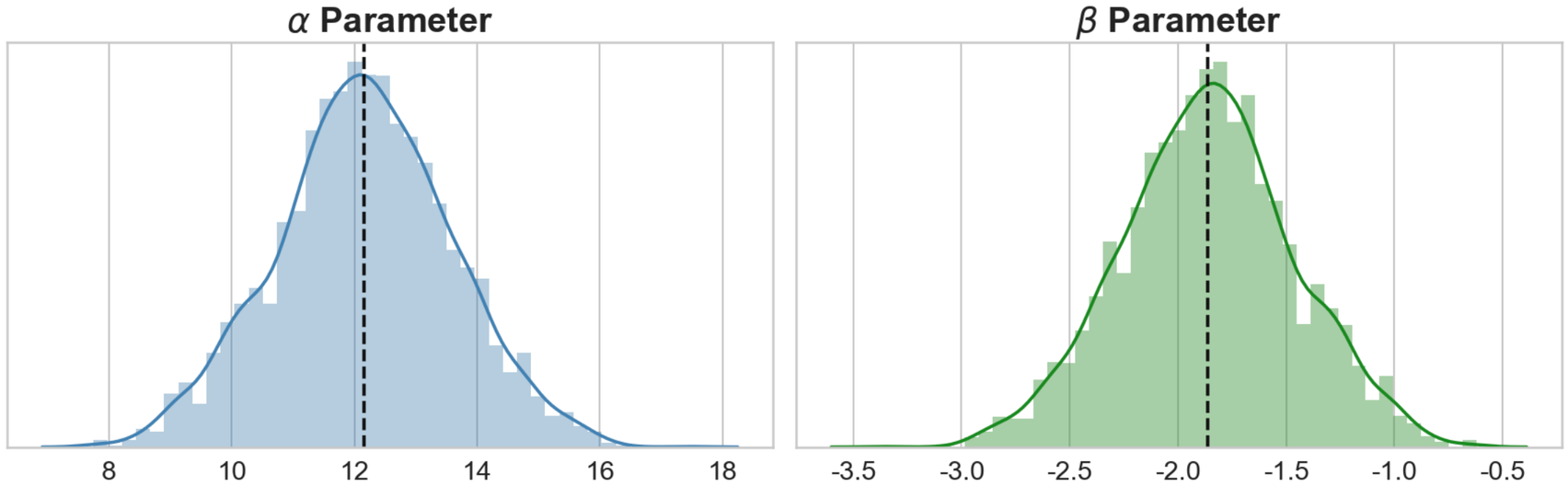

[2] https://cscherrer.github.io/post/max-profit/

# Price Elasticity of Demand

$$Q = c \cdot P^{\beta}$$
$$\log Q = \log c + \beta \log P$$
$$\log Q = \alpha + \beta \log P$$

**Goal: Estimate the values for $\alpha$ and $\beta$ that best fit our data**

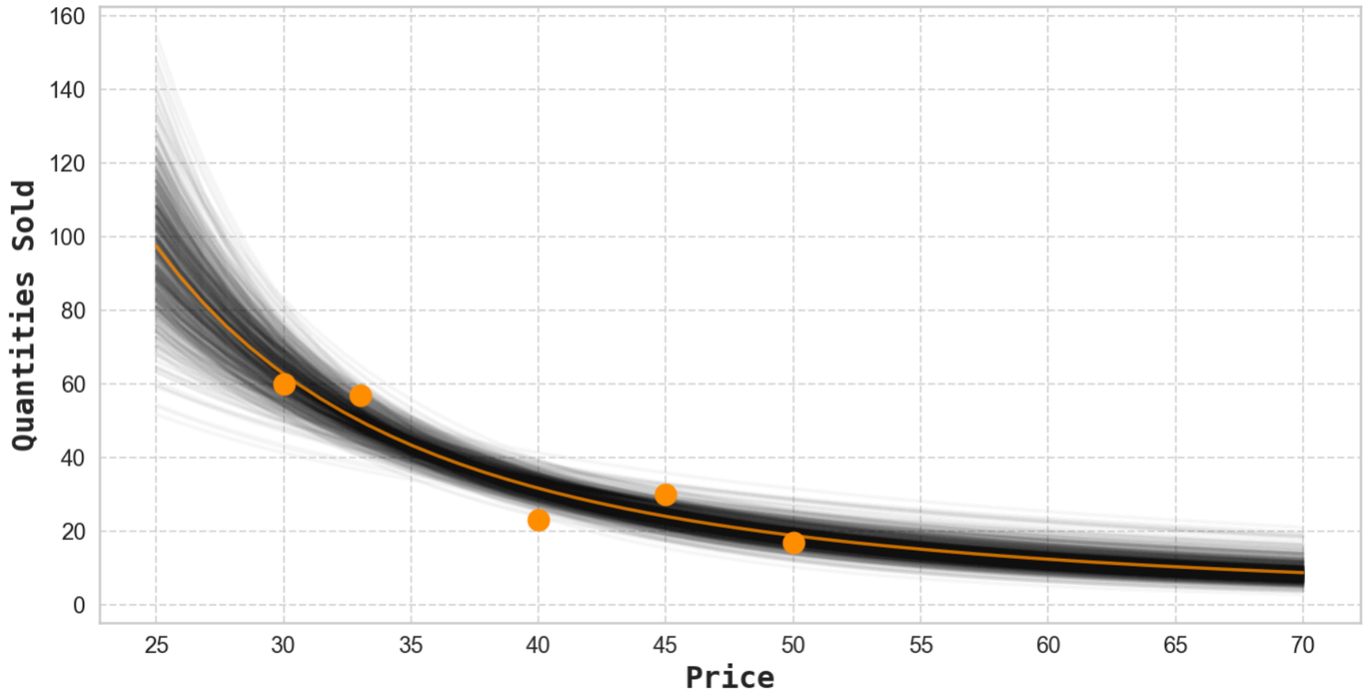
# Estimated Parameters




$$e^{\alpha+\beta \log P} \longrightarrow \mu$$
$$Poisson(\lambda = \mu) \longrightarrow Quantity$$

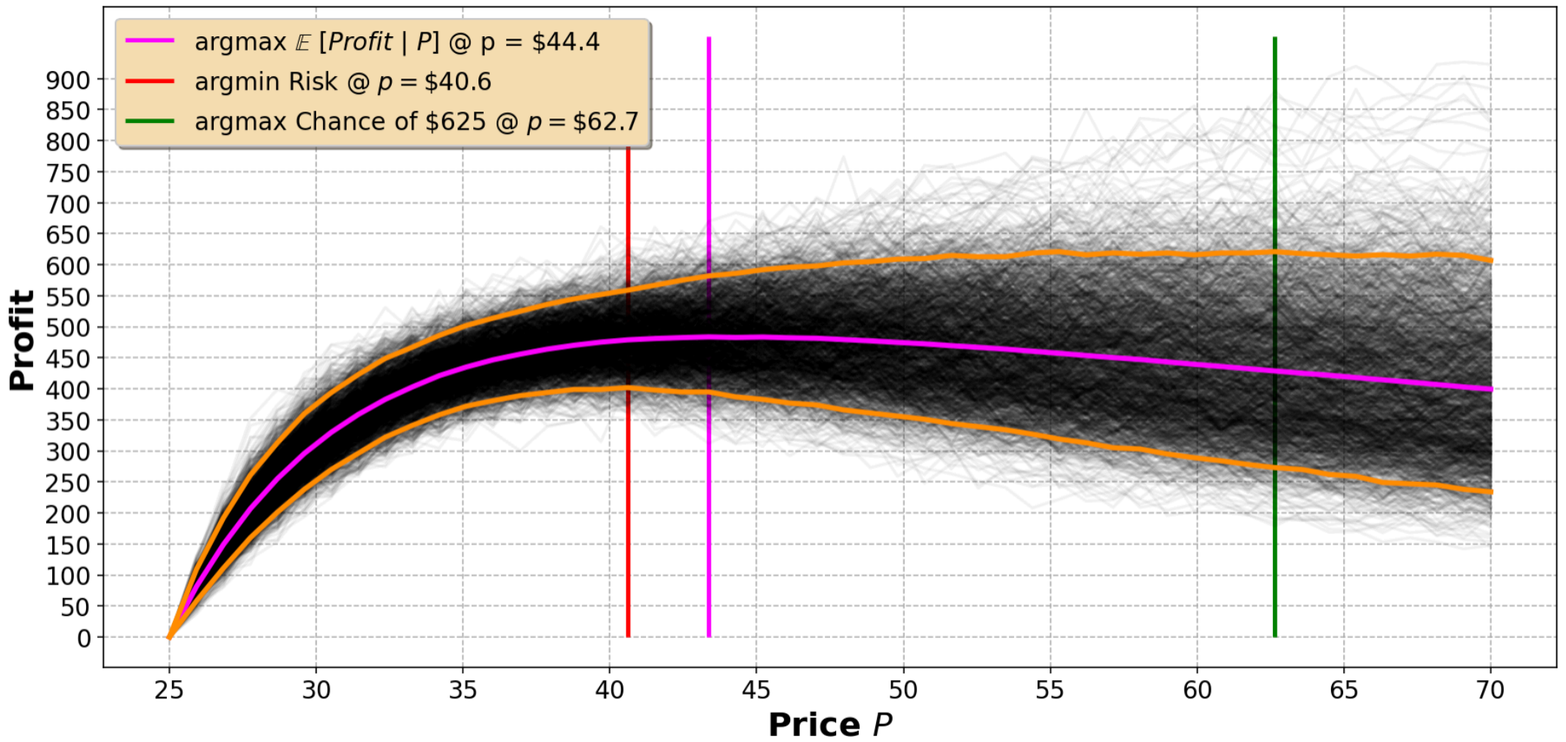
# Simulating Model Output

Each line represents $\mathbb{E}[Q|P]$ based on one posterior sample



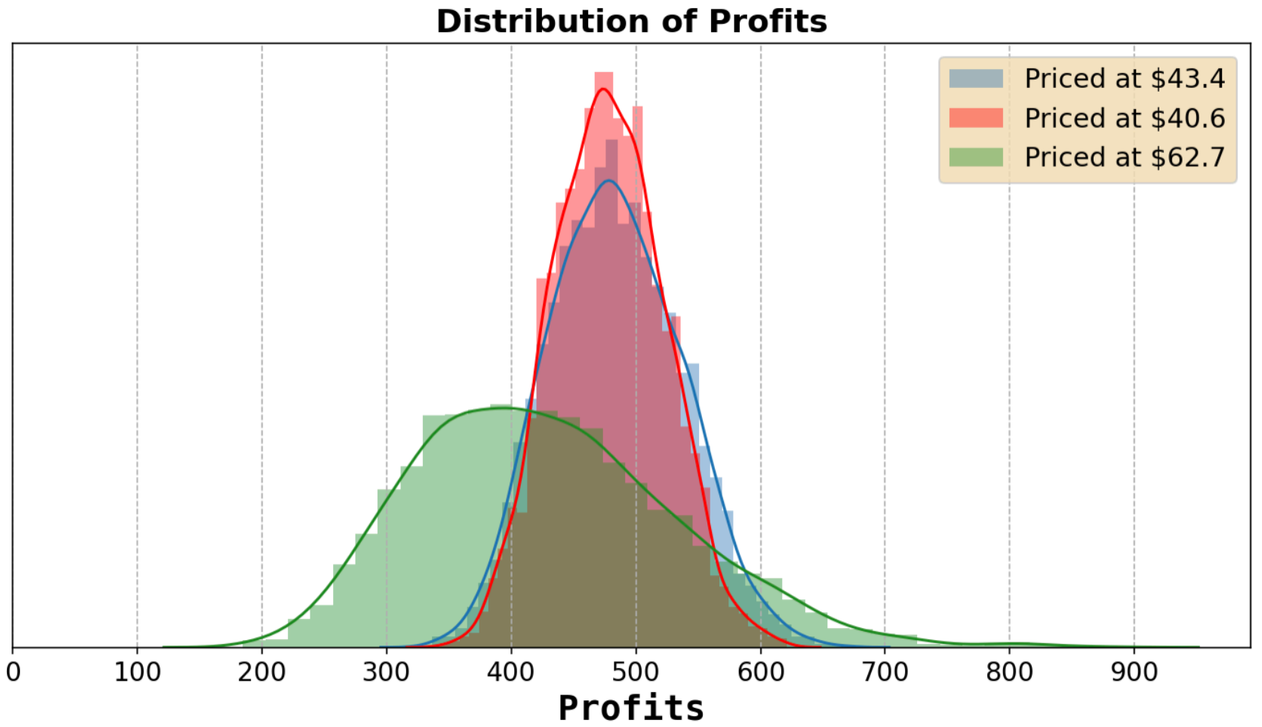
# Optimal Pricing

$$Profits = (Price - \$25) \cdot Quantity$$



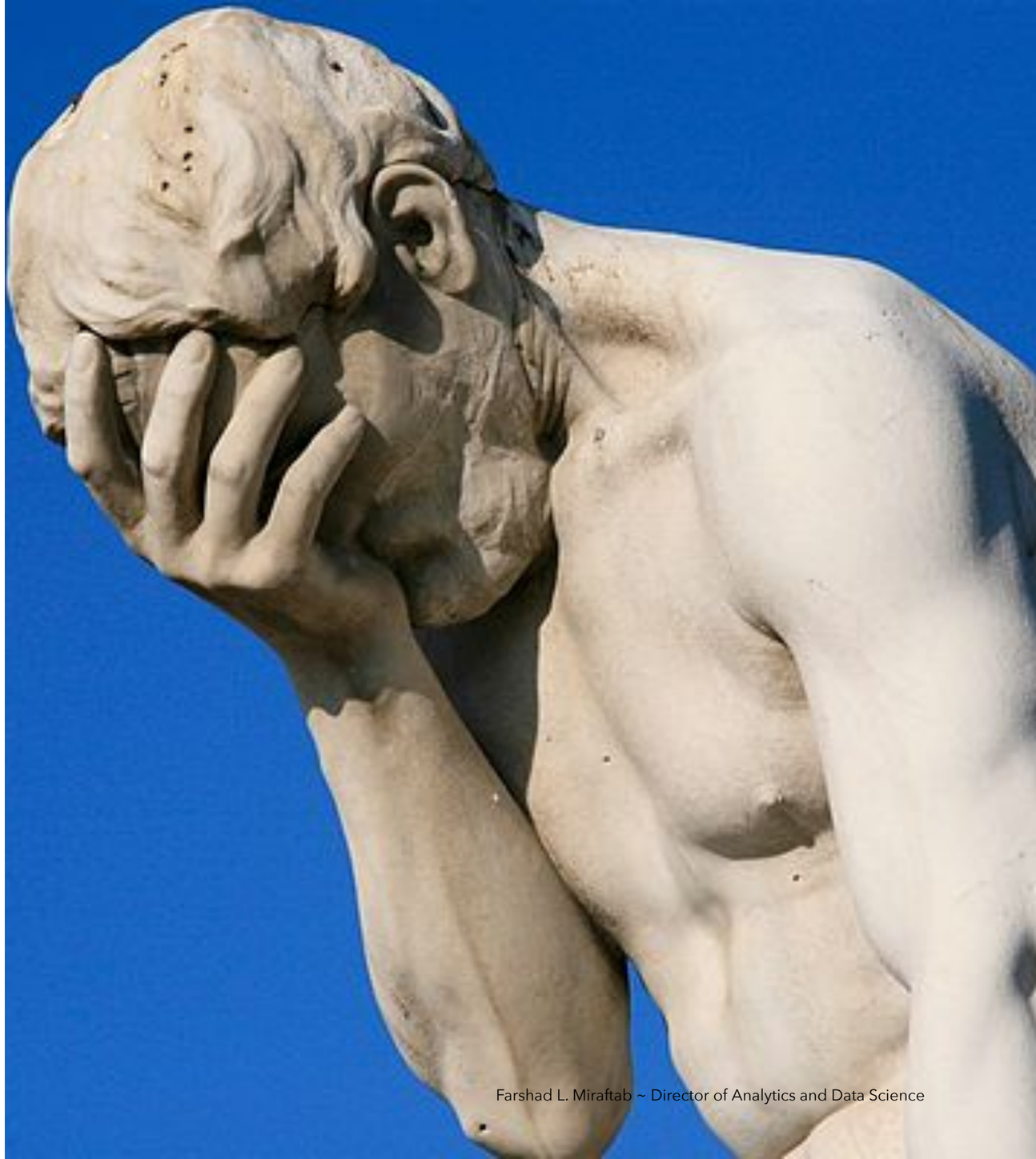
# Know your Tradeoffs

| Price | Avg Profit | Worse-Case | prob(>$625) |
|---|---|---|---|
| @$43.4 | **$483** | $399 | .4% |
| @$40.6 | $478 | **$408** | ~0% |
| @$62.7 | $429 | $277 | **4.1%** |



# #4

# Solving the *wrong* problem



Farshad L. Miraftab ~ Director of Analytics and Data Science

# The Law of Instrument

**"To a hammer, every problem looks like a nail"**

» We are often solution-backwards when instead, we should strive to be **problem-forward**

» Algorithms can actually be our **enemies**

» **Understand the fundamental problem** that and why it's important first

» Guide the discussions to **surface the right questions** and clearly define the problem

# Churn Modeling

## The Churn Problem in SaaS

Often, customer success and sales teams what to identify which of the customers are most likely to churn. Typically, a data scientist defines the problem as **can I build a classification algorithm to predict which customers are more or less likely to churn**

## How the Problem should be defined:

The problem isn't the inability trying to predict churn per se, but rather help CS and sales individuals prioritize which customers in their portfolios they should proactively outreach to mitigate the possibility of churn. A likelihood a churn may not be the only dimension used to prioritize but perhaps the size of the customer as well.

# THE COBRA EFFECT

A WELL-INTENTIONED MEASURE CAN OFTEN BACKFIRE AND HAVE THE OPPOSITE EFFECT TO INTENDED


**INTENTION**
REDUCE COBRA POPULATION

**ACTION**
A BOUNTY FOR DEAD COBRAS!

**EFFECT**
PEOPLE START COBRA FARMING

sketchplanations

# #5

# Explore your Curiosities

*In science if you know what you are doing you should not be doing it... In engineering if you do not know what you are doing you should not be doing it*

**— Richard Hamming**

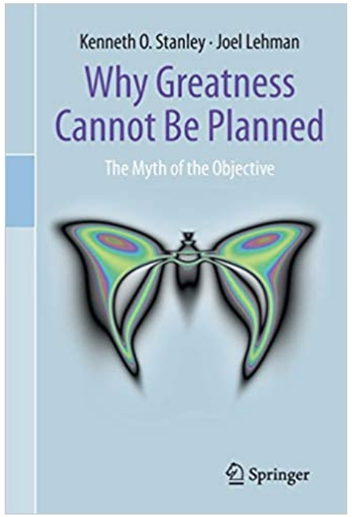
# Isn't every scientist a *data* scientist?

Science has always been about **data**, **observation**, and **exploration**. We should explore our curiosities, ask ourselves interesting questions about the domains we work in and follow our instincts with the data.

Objectives can stress, distract, and even regress us in our journey in achieving success

# The Myth of the Objective

"...if you're wondering how to escape the myth of the objective, just do things because they are interesting. Not everything needs to be guided by rigid objectives...follow the scent of interestingness, of novelty, or of whatever present clues we feel may propel the great treasure hunt of innovation"



# Thank You!