# Decision Science

## 1. Introduction and Course Overview


Mihai Bizovi

Head of Data Science @AdoreMe


2023-12-08

# A long and rewarding journey

**Why should you stick with the course?**

- Effective problem solving & **decision-making**
- Bridge the gap between **theory** and **practice**
  - between math world and real world
- **Reframe** what you already studied to make it useful – e.g. hypothesis testing, regression
- Lots of new methods for your **toolbox**
- Clear **understanding** and **precision** in:
  - data science terms and concepts
  - communication with clients, decision-makers, and stakeholders

Navigate the complexities of the field and choose your path. *(Source: Generated with DALL-E)*

# You meant “data science”?

- Confusing umbrella term, but for sure interdisciplinary
- Excitement about the potential of Analytics, Big Data, Data Mining, ML/DL/AI, Generative AI:
  - to improve the business outcomes 🚀
  - be at the core of business models in themselves
  - therefore, the hype of “sexiest job”
- These technologies are pervasive
  - ethical concerns about AI misuse in tech/gov
  - concerns about how we do Stats, ML, AI today 😈

# Decision Science – explicit goal

- Focused on **Decision-Making Under Uncertainty at Scale**
- For that, we need insights and tools from:
  - **Applied mathematics**: Linear Algebra, Prob/Stats, …
  - **Computer Science**: Numerical methods, Optimization, …
  - **Business Economics**: Marketing, Management, Finance
  - Cognitive Science, Cybernetics, Philosophy, …

> This reframing makes the learning roadmap much more clear and actionable – coherent set of tools, methods, ideas

# The gap between theory and practice

> Think of youself as a business person with superpowers

### Engineering in the trenches

- Develp skills – *a hard, but rewarding path from novice to expert*
- Pragmatic data-driven software
- Modeling workflow and pipelines
- Simulations as a safe playground
- This course is NOT a bootcamp

### Contemplating in the library

- Cultivating understanding and insight into *fundamental theoretical ideas*
- Understanding your domain & clients
- Apply the right model for the job
- Awareness of pitfalls and mistakes
- Rigorous, but NOT detail-oriented

This is the course I wish I had when getting started in the field.

# Principle nr. 1

Contemplating in the library, engineering in the trenches

*DALL-E helps with storytelling and visuals across the deck*

Dreyfus Model (Hunt 2008), Learning how to Learn (Oakley, Sejnowski, and McConville

# Principle nr. 2

We'll go a long, long way by mastering the fundamentals

True for music, painting, sports, math, business, even GenAI

A virtuoso violonist does the scales masterfully

## Principle nr. 3

# Do not confuse a tool with the end goal: Content ≠ Process

Knowing how to perform a t-test is different from designing an experiment

# Why did you learn all of that?

- **Linear Algebra** - language and geometry of data
- **Mathematical Analysis** - formalism of change
- **Probability** - logic of uncertainty
- **Statistics** - changing your mind and action under evidence
  - inference = data + assumptions
- **Econometrics** - does $\mathrm{X} \longrightarrow \mathrm{Y}$? Causes or associations?
- **Operations Research** - optimization with constraints

All of the above is useless without being used in software products or implemented in

# Principle nr. 4

We weave a golden thread connecting disciplines

Encouragement: what you studied is actually useful!

Choose the appropriate level of analysis and the practically useful tools and ideas

# But why? Business environment!

> VUCA: Volatile | Uncertain | Complex | Ambiguous

- A problem well formulated is halfway solved
- Limited resources, conflicting objectives
  - therefore, have to make tradeoffs
- Decisions informed by robust predictions and evidence
- Do you have what it takes to navigate this environment?
  - the tools, understanding and skills
  - attitude: learning and growth mindset

## Principle nr. 5

# Data Science is the art of solving the wrong problem

We often get an answer to a different question, not the one asked

# Principle nr. 6

## Calling Bullshit: reasoning and critical thinking


Advertisement is a prime example. It makes stuff supersalient

H. Frankfurt - Responsibility to call out BS and not generate more

# ~whoami: as engineer and leader

- Skin in the game: Head of data science @AdoreMe (VS)
- AI strategy, decision-making, ML systems design
- Statistics, ML/Data engineering, Full stack data apps
- Critical applications along the value chain:
  - demand forecasting & inventory optimization systems
  - recommender systems, NLP for try-at-home
  - marketing: acquisition, CRO experiment design, CRM

Why such systems are needed? Watch this video about Target's disaster launch in Canada,

# ~whoami: as teacher and researcher

- Maven *(yid: meyvn)* – experience to help you find your path
  - help you develop skills and understanding
- Graduate of Cybernetics and Quantitative Economics
  - Thesis & Dissertation on Bayesian Microeconometrics
  - Research in Probabilistic Methods for Time Series
  - Started in systems' dynamics and economic complexity
- Speaking at conferences
- Teaching since 2021 (ASE, Google Atelier) – 4th iteration

# ~whoami: as a person

- Cultivating wisdom – philosophy as a way of life
- Participating in meaningful friendships
- Painting, hiking, coffee, 14 years of pro-ish chess
- Art appreciation: blues, jazz, indie, opera, cinema
- Reading: Cognitive Science, Evolution, Economics
- How it fits together: synoptic integration [1]

1. synoptic is to take the best of multiple perspectives, reduce conflict and contradiction – then see where the evidence converges

# Conversation: fields & use-cases

- What are the **fields** in which data science methods are extensively used? e.g. finance, genomics, psychology, …
- What are some **products** that use AI, data science, data-driven systems? What are their **use-cases**? e.g. uber …

**Think in terms of reverse engineering**

When using those products, how do you think those systems were designed?

- What were the goals and user/client needs? What were the firm's objective?
- What constraints did they hit? Why is it a difficult problem?
- What are some potential approaches they settled on? What is a naive solution?

# The easiest and most difficult course

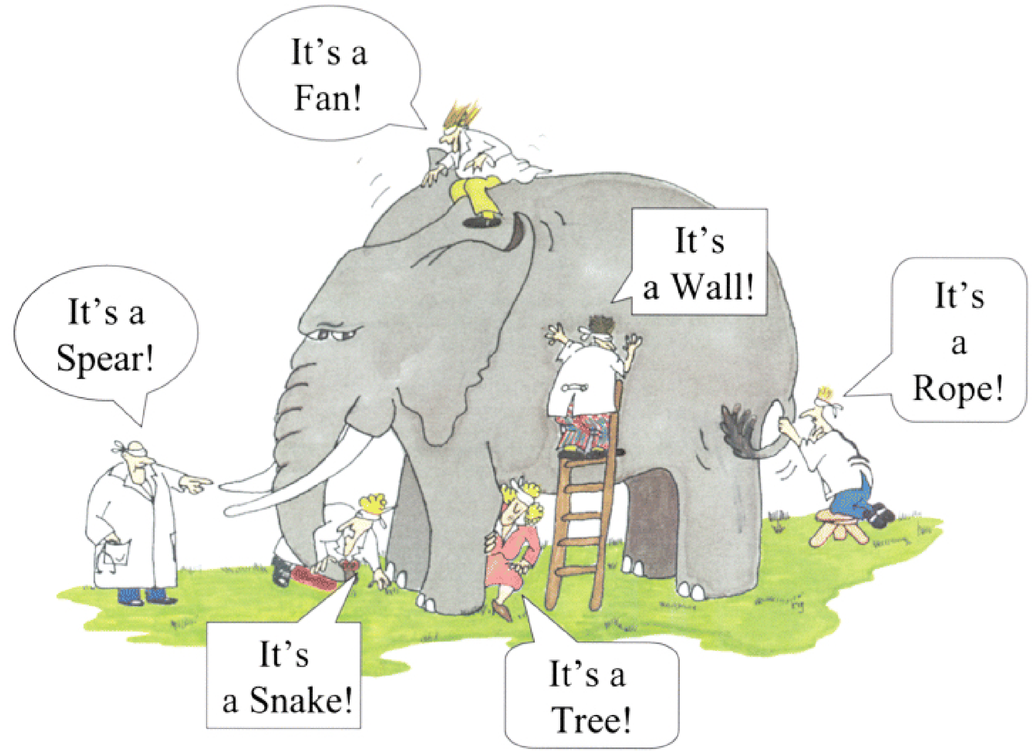
> “It was the best of times, it was the worst of times, it was the age of wisdom, it was the age of foolishness, it was the epoch of belief, it was the epoch of incredulity, it was the season of light, it was the season of darkness, it was the spring of hope, it was the winter of despair.” — A Tale of Two Cities, (Dickens 1859),

- No memorization, proofs or solving on paper
- You have to really understand, justify choices, write code
- No single, unique solution, but many tradeoffs

# Is this stuff for you?

- If yes – roadmap for a journey from novice to expert
- If no – you will still benefit, maybe I change your mind
    - touches every aspect of businesses, opportunities
    - know how to talk to analysts, engineers, clients, experts
- Everyone wins: engineers, PMs, analysts, data scientists, economists, managers and entepreneurs
    - in any domain $\oplus$ role
- Don’t go into it just because it pays very well

# Simplification and Relations

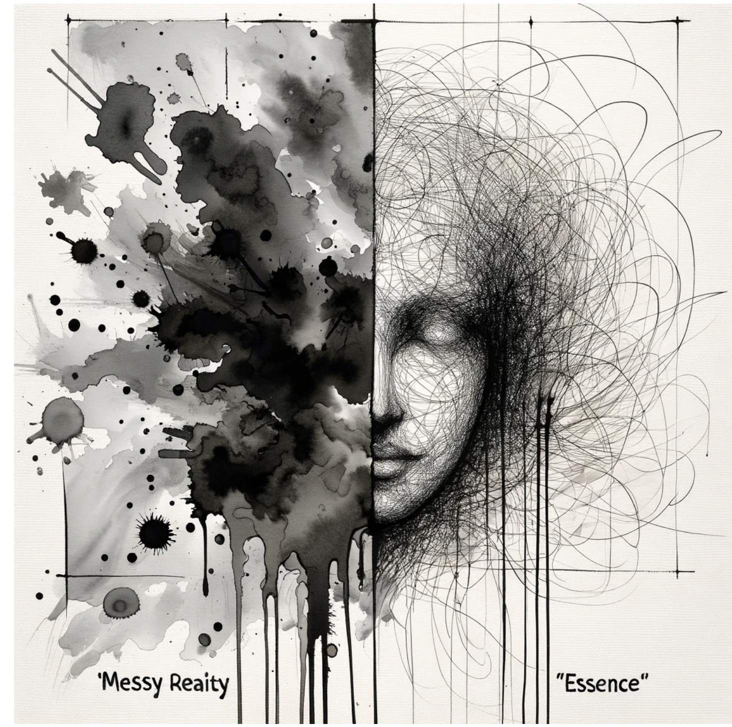



We see Pollock's **messy reality**, which is the data. We want to get to Picasso's **essence**

**Big picture** course, how it all fits together and how it is done in practice to be useful for firms

# Model = simplified representation

All models are wrong,
but some are useful




Never forget model limitations and assumptions!

Models are also golemns (of Prague)

# What we'll cover for sure

- **Business Economics**: the real firms, not theory
- Probability: filling in the gaps, lots of simulation
- Mathematical Statistics: the actually useful theory
- **Experiment Design** and A/B testing
- Bayesian statistics and **hierarchical models**
  - Think of it as advanced econometrics
- **R or Python** – your individual choice
  - Competence in programming is essential

# What we might cover

- Machine Learning and Tree-based Models
- Time Series and Demand Forecasting
- Unsupervised Learning
  - Dimensionality reduction
  - Clustering

# Project: problem you care about

- For engineers: if you like programming
  - Interactive, data-driven application to support decisions
- For analysts: if you like data analysis
  - Answer an interesting question, with some inference
- For data scientists and statisticians:
  - Solve a business problem with models

**Reuse from other classes NOT allowed**

Original and risky projects will be highly rewarded, even if not successful or flawed!

# Conversation

- What are you passionate about?
- What was the last thing you “read”?
  - paper, book, course, podcast
- What are your aspirations and goals for the carreer?

# Principle nr. 7

There is too much content about data science

Where do you even begin?

Lots of it is just bad, but there is too much good stuff 🤷🏻‍♀️

# Too much content

- Curated selection of excellent resources
  - Still, years’ worth of study
- Roadmap and conceptual frame to navigate by yourself
  - Choose what is relevant for your problem
  - Come back to a topic with greater understanding later
- No shortcuts that work, patience needed
- Master fundamentals (opinionated on how)

# References

Dickens, Charles. 1859. *A Tale of Two Cities*. New York: Chelsea House Publishers.

Hunt, A. 2008. *Pragmatic Thinking and Learning: Refactor Your "Wetware"*. Pragmatic Bookshelf Series. Pragmatic Bookshelf. https://books.google.ro/books?id=Pq3_PAAACAAJ.

Oakley, B., T. Sejnowski, and A. McConville. 2018. *Learning How to Learn: How to Succeed in School Without Spending All Your Time Studying; a Guide for Kids and Teens*. Penguin Publishing Group. https://books.google.ro/books?id=A3djDwAAQBAJ.