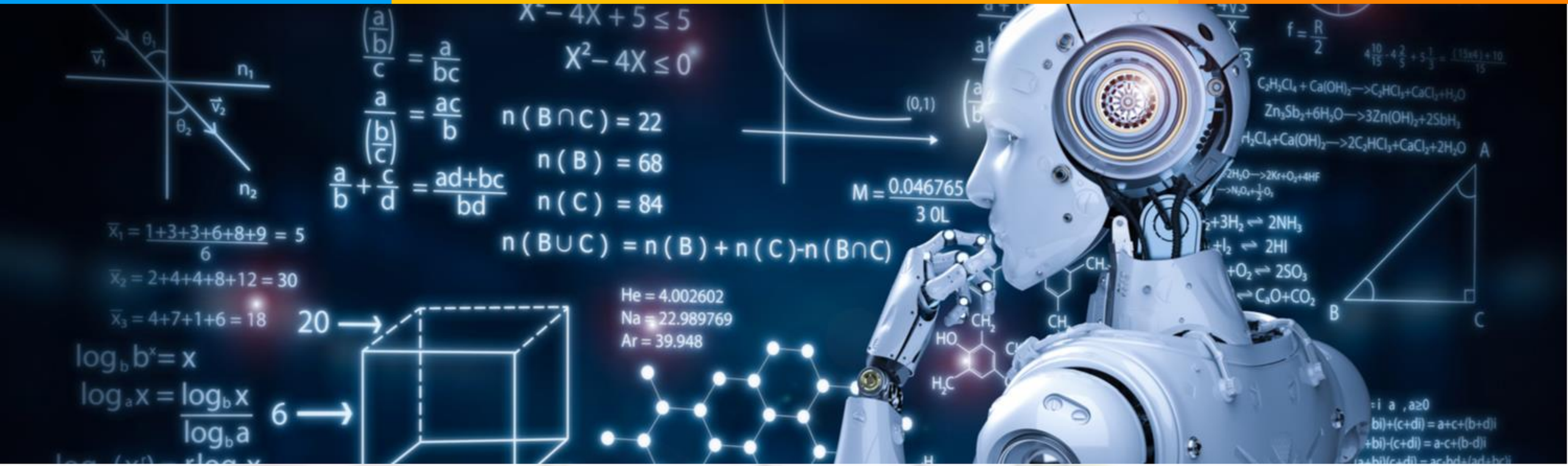
Duration: 90 min | Srikanth Shenoy | Orientation | AI & ML / BDA

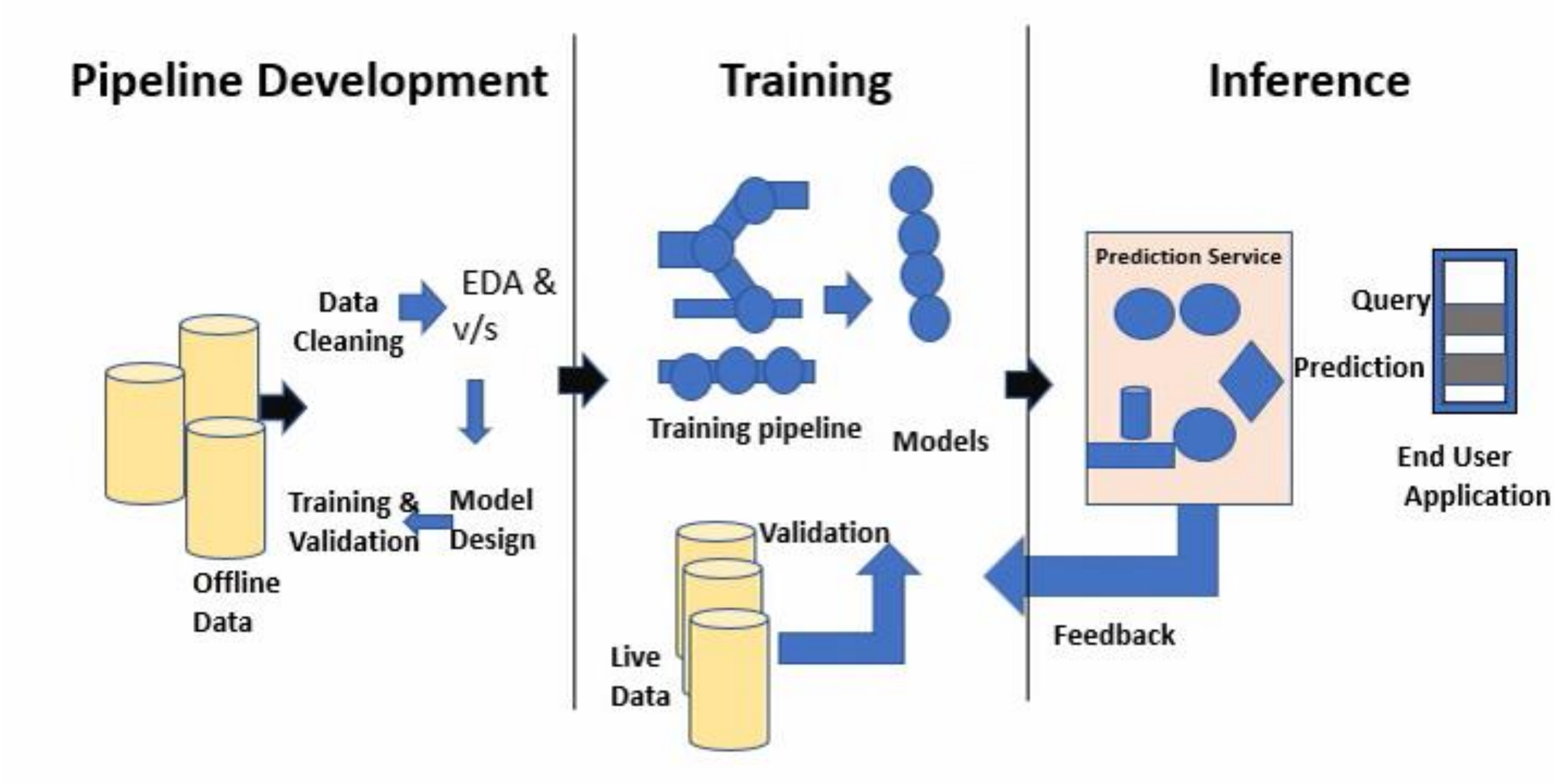



# Reality check & tips for thriving in Data Science & Machine Learning jobs landscape

# Realist or idealist?

- Are you okay with any internship/software job?
- Adamant on Data science/ ML job?

- What is your understanding of data science/ ML job?


Pipeline Development
Training
Inference
Data Cleaning
EDA & v/s
Training & Validation
Model Design
Offline Data
Training pipeline
Models
Validation
Live Data
Prediction Service
Query
Prediction
End User Application
Feedback

# Data Engineer


# Data Scientist

# ML Engineer


# Data Analyst

- Real jobs are  a mixture of all these plus
- Other must do things – can be interesting or just laborious

# Employability Statistics

- Sad state of affairs ☹
  - Only 7% of engineering graduates are employable
  - Only 3% qualify for startup tech role
- What is the root cause?

# How companies work?

- Hire based on THEIR requirements
- Hires should be able to produce results
- Results should benefit organization
- No guarantee of work in DS/ML. Could be any of:
  - Data fetching, cleaning, loading (ETL)
  - Data visualization
  - Development in some other language or UI
  - Full stack
  - Testing

# Realist or idealist?

- Are you okay with any internship?
- Or Adamant on Data science/ ML role?

- Placement facts
- VLSI first, SDE next, DS and ML may be at the end
- Some DS and ML companies may not show up at all

- Are you willing to give up on earlier SDE roles?
- What if “that” DS or ML company didn’t show up?
- Desperation when friends class/roommates placed

# Moral of the story

Nature of Internship/job is ALSO a matter of luck

# Question to ponder

- Suppose you got 2 jobs – SDE & Data science

- Will you pick Data Science role even if SDE role paid higher?

# Harsh truth

**Companies have freedom to renege on their good faith agreement of data science role.**
**Job offer is not a contract binding the company.**

**No budget for data science role. Take SDE role instead.**

Given same pay,
which company will you choose?

# Big company or a startup

**You'd think the answer is a no brainer**

**I've got news for you**

Future is gloomy

Why should I study?

ChatGPT is going to take a lot of jobs away

# Lessons from history

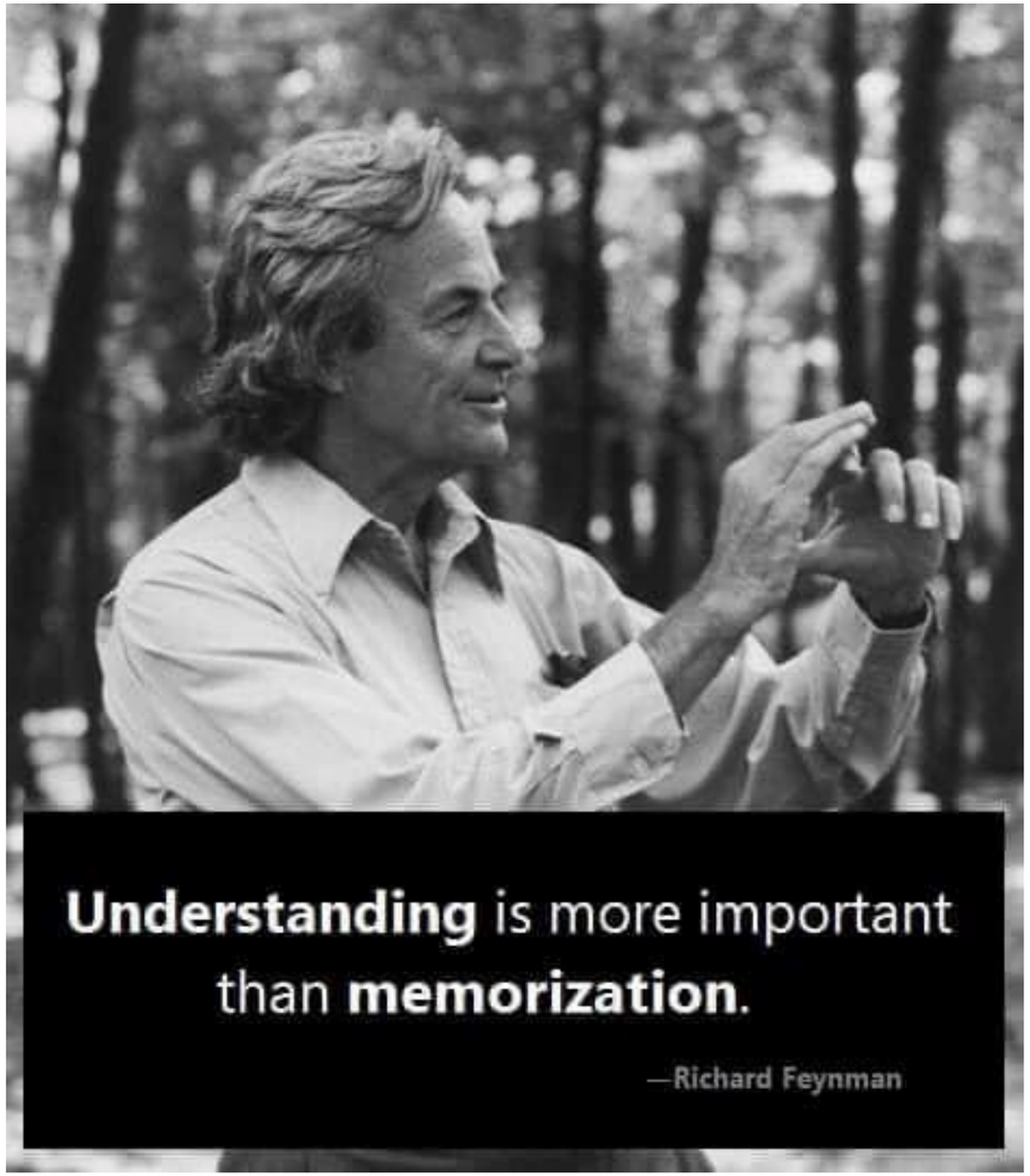
- How did people react when calculator first came?
- How did people react when computer first came?

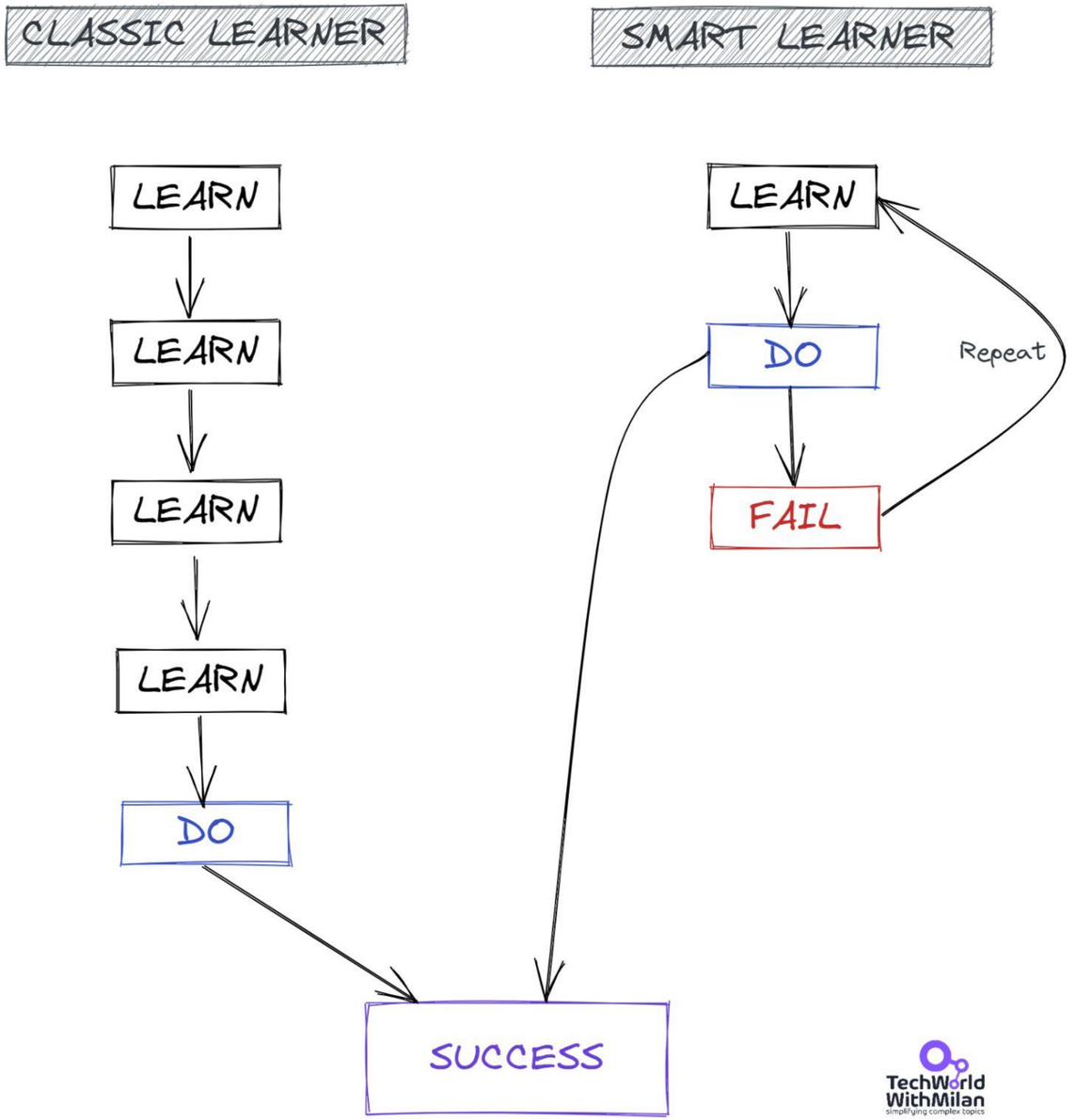
- Commodity jobs go away. Value adding jobs stay
  - E.g. those treating sklearn, keras as api go away
- Your way out -
  - Interdisciplinary - domain + technical expertise
  - Deep technical expertise alone
- You have an advantage

# Study smart & practice

- Get your basics and concepts straight
- Don’t pour all efforts into just “remembering”
- 80% of “remembered” stuff is forgotten in months
- Understanding & applying is important than memorization
- Key for career success – Intuition, problem solving
- Intuition -> Maths -> Programming

# Learning philosophy



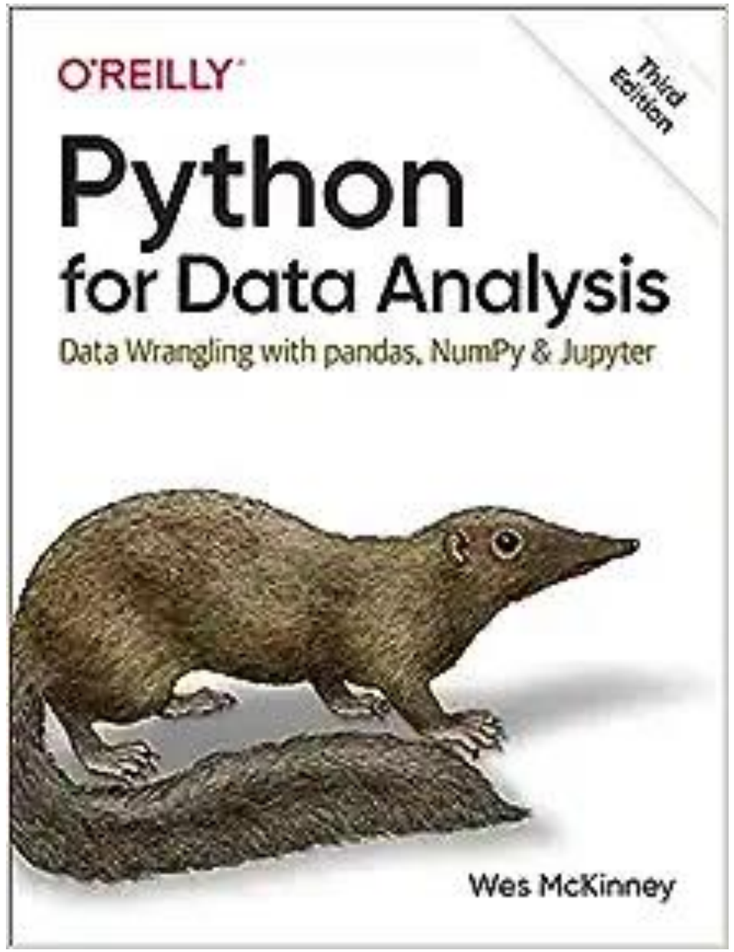



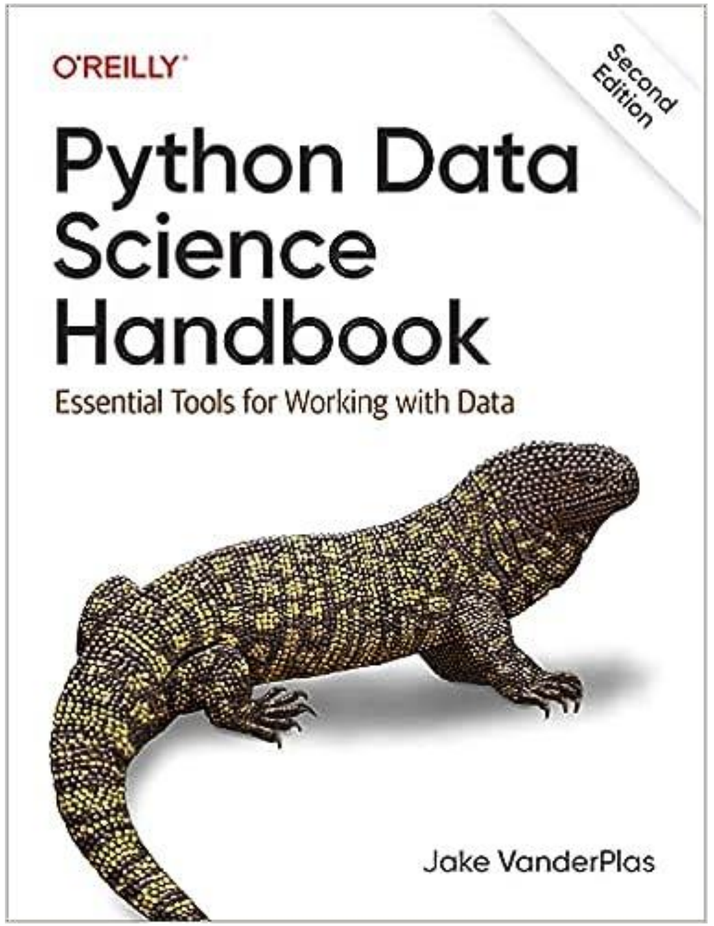

TechWorld WithMilan
simplifying complex topics

# Tips for placement prep

- Build github profile, keep updating throughout sem
- Put in your CV
- Keep coding - Practice Practice practice
- Set goals – Finish Numpy, Pandas in first 10-20 days
- Do a lot of hands-on EDA
- Don’t copy somebody else’s work or fork and claim as yours
- Get comfortable in using product documentation

# Pick any




- My preference – Second book
- Be systematic, annotate, revise

# Tips for placement prep

- Stress and showcase important skills
- For e.g. if you showcase a lot of visualization, you will get Data Analyst jobs
- Use Copilot, ChatGPT while coding
- Follow this:
  - Understand object oriented programming well
  - Learn to understand existing code from others
    - E.g. Look at scikit-learn repo on best practices
  - Rinse and repeat

# Tips for placement prep

- C, Python
- Linux - scripting
- SQL
- Nice to haves
  - Containerization (after everything else)
- Mental aptitude, reasoning

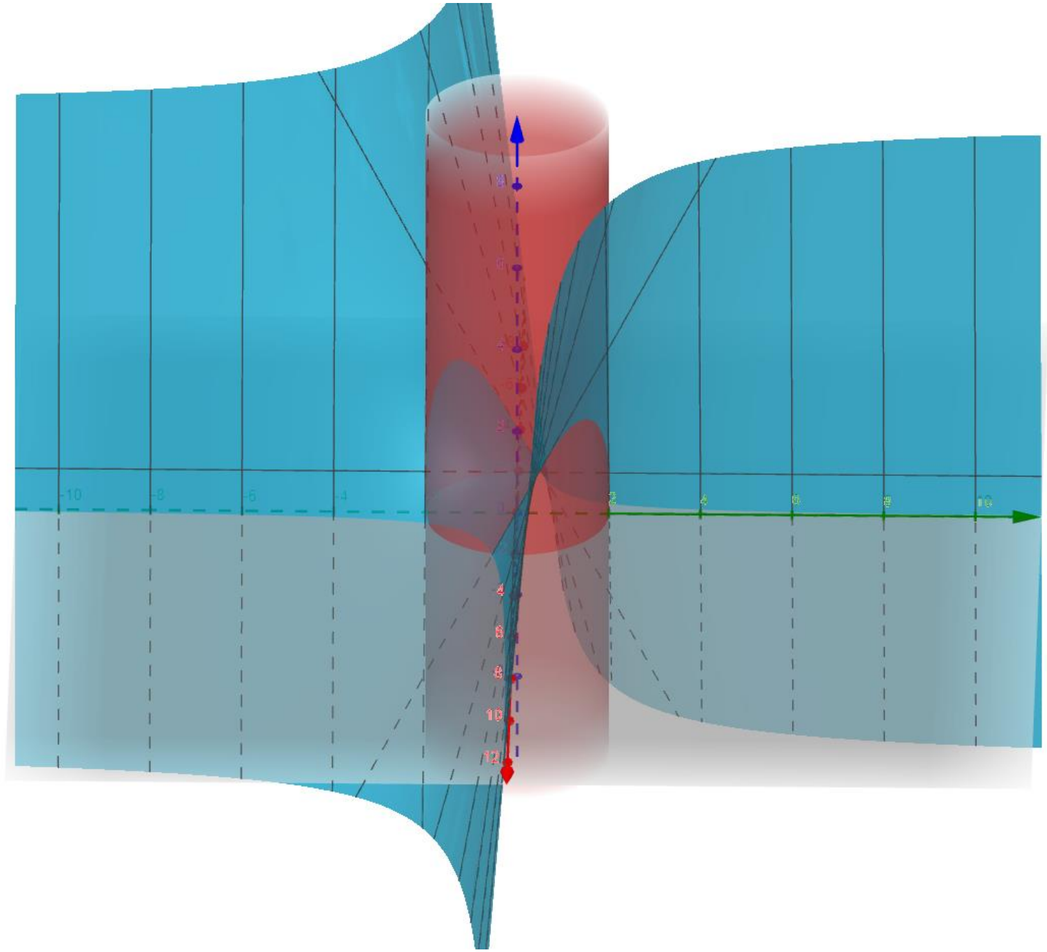
# Sample interview questions for EDA role

- Pandas join merge resampling grouping windowing
- Pandas 1.x versus 2.x
- Polars
- Have you encountered PyArrow?
- Have you tried to execute numpy ops using GPU?
- Sklearn is slow, How will you make it faster?
- Can sklearn run in GPU?

# Mantras for the semester

- Failing to plan = planning to fail
- Learn to learn – SQ3R, Feynman technique

- Having fun is an exercise in constrained optimization




QUESTIONS


ThankYou!