EMERITUS

Learn. From the world's best.

# Industry Guidance

www.emeritus.org

EMERITUS

Learn. From the world's best.

# Data Science and Industrial Interpretation

# How does a Data Scientist fit in?

Most of the corporate world doesn't really know what it wants. This is good and bad.

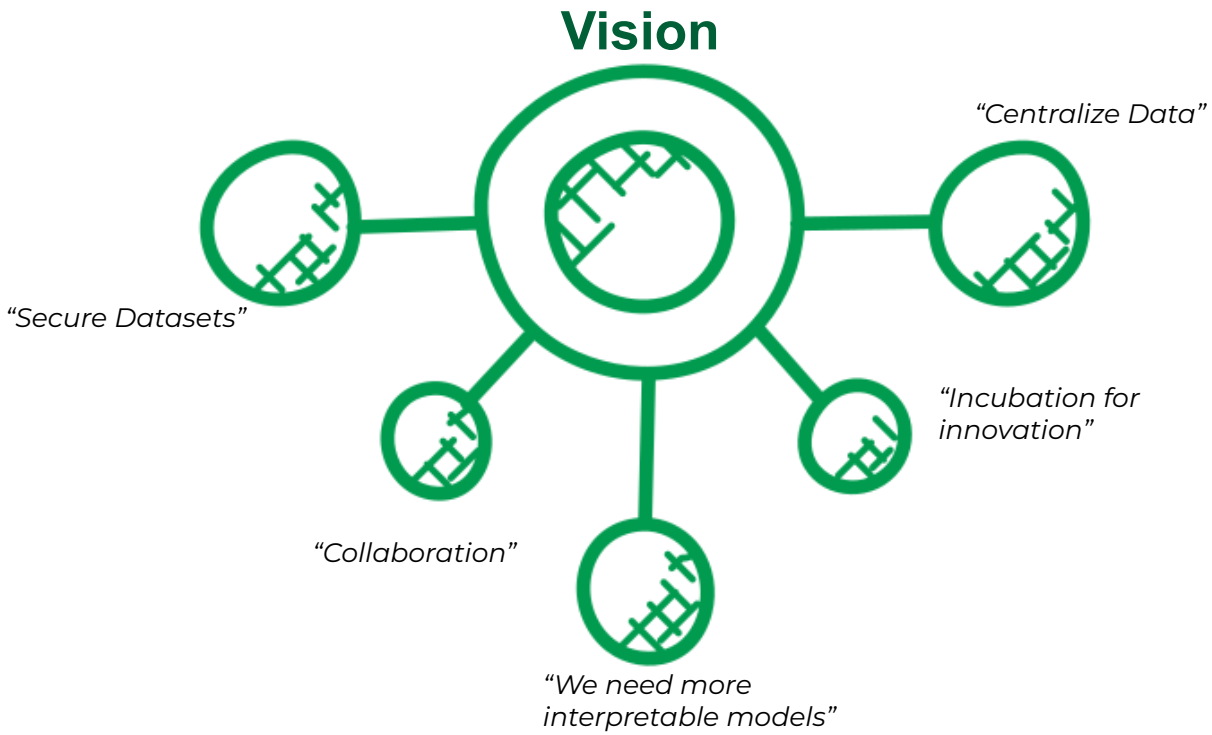
As a data scientist, it is your hefty responsibility to thoughtfully reinterpret the contradictions of the industrial AI landscape into valuable outcomes.

**The trick is to handle this discordance from the inside out, not just as some siloed observer**


"We are looking for a trailblazer who will help us start the transition to advanced analytics"
"If you are someone who enjoys getting their hands dirty with data, this is the right spot"
"10+ years of experience"
"Strong preference to work independently"
"You should thrive working with others and enjoy the collaborative approach"
"1 - 2 years of working with data"
"The ideal data scientist is ready to join an established advanced analytics enterprise"

Different postings from the **same** organization

# Industrial Data Science Volatility




Organizations do not yet have the over arching strategy well formulated

The larger the org, the more probable vision factions exist for data science.

Different goals  Different hiring  Different value

The data scientist is absorbed into an environment of conflicting visions

EMERITUS

Learn. From the world's best.

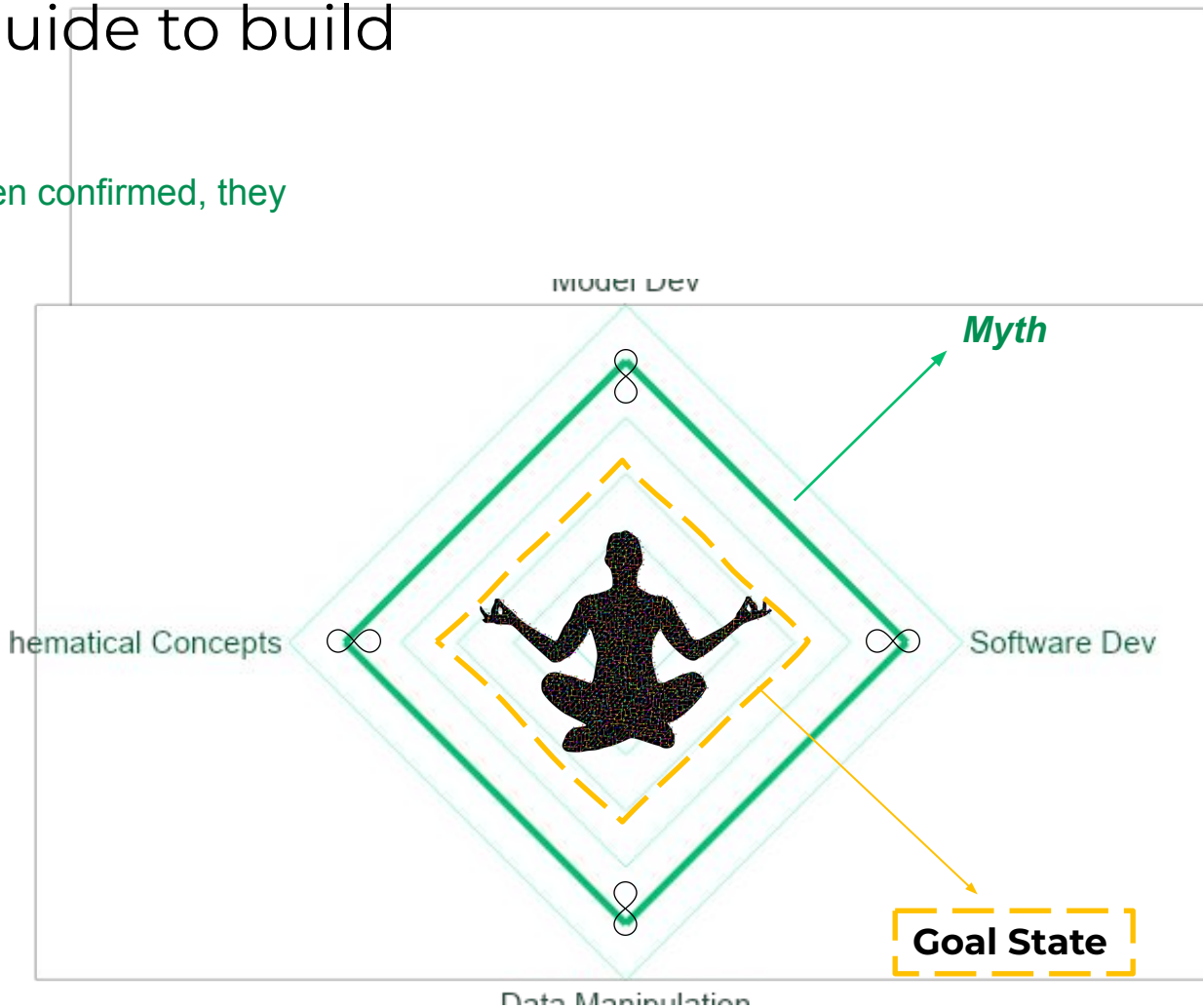
# Important aspects of getting a gig

# Use the Full Stack Data Scientist as a guide to build your skill identity

While the myth of the Full Stack Data Scientist of Silicon Valley has not been confirmed, they remain a useful idea

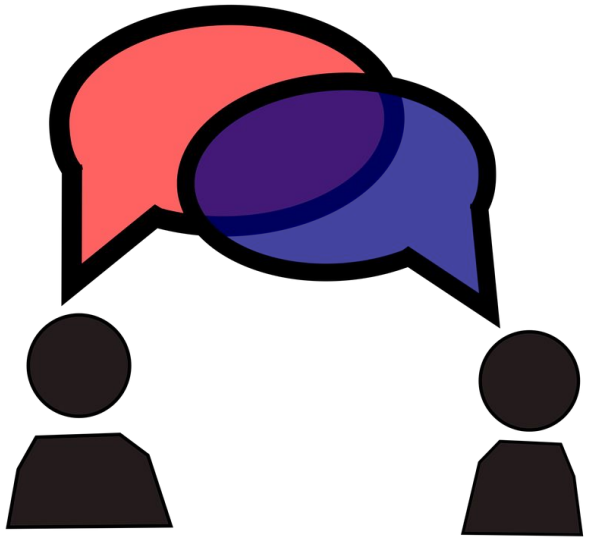
| Myth |
| --- |
| ∞ PhD in Math/Stats |
| ∞ Software dev professional |
| ∞ Trained the GPT-n series |
| ∞ Was a DB engineer |
| ∞ Has read all the relevant lit |
| ∞ Solved all past problems |

| Goal State |
| --- |
| ◇ Solid exposure to Stats |
| ◇ Familiar with dev env |
| ◇ Implemented ML/DL architectures |
| ◇ Know multiple DB langs |
| ◇ Keeps up with ML literature |
| ◇ Has learned from failed solutions |



# The Interview

## The 3 Main Categories



### Coding / Problem Solving

This will ultimately be the most straightforward, yet significant part of a DS interview. Topics range from ML specific to computer science. Questions may consider:

- Algorithms
- Data Manipulation
- ML models

### Business Case / Situational:

Where do your skills lead you in interacting with an organization? Can you formulate best ways to achieve the outcomes a team is aiming for? You need to demonstrate your ability to conceive of the data science problem statements within real organizations, beyond the toy problems of academic settings

### Personal Activity :

How does an experience you had inform you for new ones? What did you do to solve a relevant problem in your past experiences/workplace? What is a memorable time you would point to which

- Portfolio
- Past work/projects

# Coding / Problem Solving

There is no trick to technical interviews
You have it or you don't

**The Development Cycle**

Get Interview

Fail Interview

Identify Failure

Get Interview

**Algorithmic Practice**

- Project Euler
- Leet Code
- Code Forces

**Data Manipulation & Science Practice**

- Links attached in Appendix

# Business Case / Situational

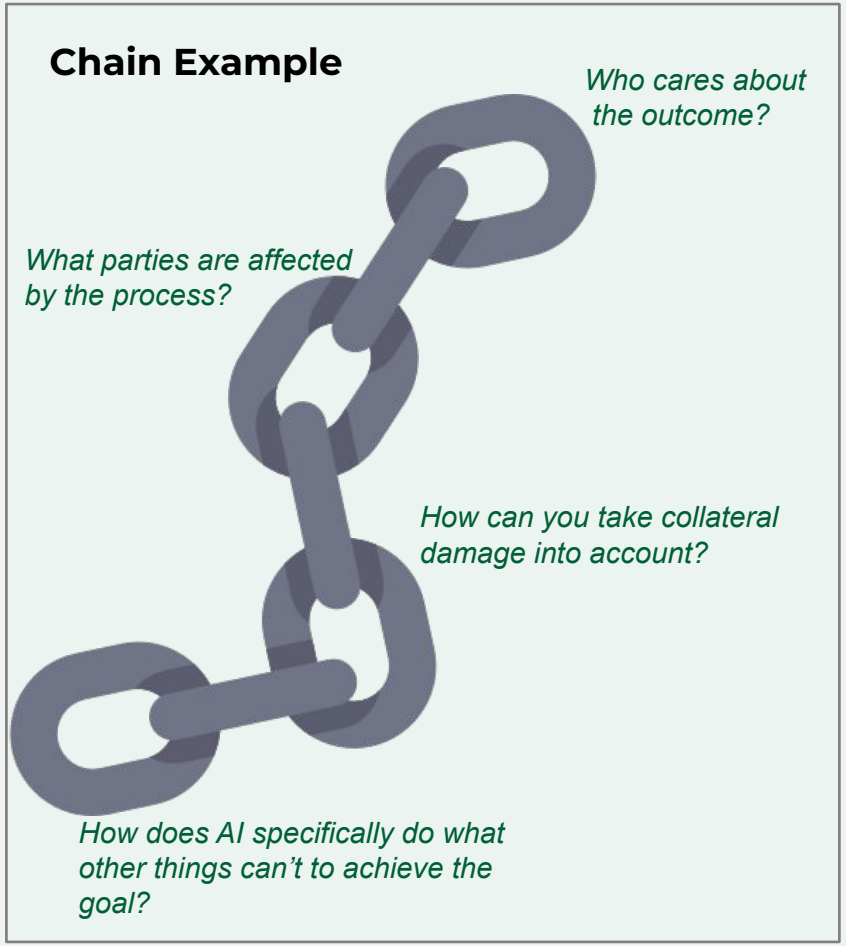
What is the simplest framework?
The Value Chain

## What does the interview care about?

- Did your research on the company make anything explicit
- Are they giving anything away in their language
- Is there an opportunity to inquire in the beginning?

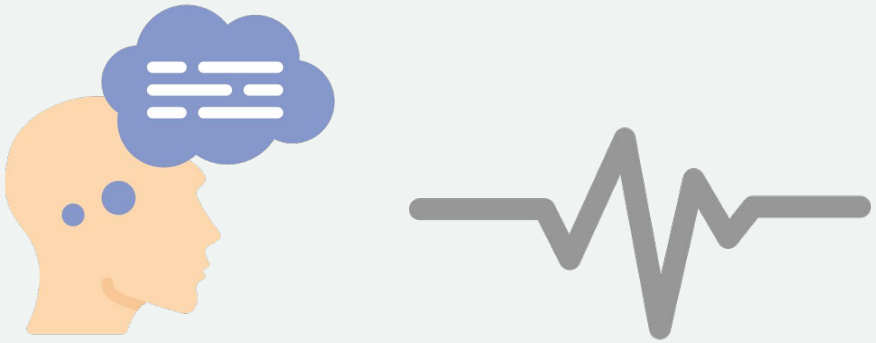
## Demonstrate AI as the Means to an Ends

- Ai is not the point; it is your best way to get to the point
- Show the ability to talk about what matters without AI as well
- Show situational awareness; talk about working with people



# Personal Activity

You will almost certainly be asked for retrospection on your experience throughout your interview



## Retro Guide

- What are my key stories?
- What are their value chains?
- How will I relay them?
- Make them more comprehensive

# Group Activity

Do a 2 minutes elevator pitch on your experience on:

- coding
- problem solving
- your portfolio

EMERITUS

Learn. From the world's best.

Soft skills for soft solutions

# Soft skills in Data Science

People talk about soft skills

Are they valuable?

Does anyone have examples?

What is the use-case?


Even more so than Tech, Data Science needs to **build robust bridges** to the user or client story

# What do soft skills really mean?


LinkedIn's list for 2021

*Creativity as a goal is tough without processes. Nonetheless a data scientist is just that – a scientist. This aspect, while often overshadowed, is an important feature of how and why we do our work.* Start with:

- Creating hypothesis statements
- Thinking of every problem as requiring a unique hypothesis
- Leveraging the insight of the non-data scientists involved
- Prototyping ideas in python notebooks with simple plug & play demos

# What do soft skills really mean?


*LinkedIn's list for 2021*

*As a data scientist it is always important to remember you are not the arbiter of truth, but more aptly, you can tell the best data driven stories. Utilize this through:*

- Showing the business outcomes as the focus
- Finding ways to put data behind existing stakeholder beliefs
- Making sure you can explain your results AND assumptions clearly
- Completely covering the value landscape for stakeholders in your assumptions
- Combination of a fundamental understanding of the business story with your data driven perspective

# What do soft skills really mean?


*LinkedIn's list for 2021*

*Without collaboration, data scientists easily get siloed. Assuming your hard skills, the most important work you have is to understand the business case and environment your role sits in. You can start this through:*

- Adopting language & concepts of various teams
- Slowly Introducing your own language & concepts for others to use when thinking through
- Sharing your sprints/project schedule with the right people for better alignment
- Not being afraid to ask for meetings which aim to achieve alignment at a high level

# What do soft skills really mean?


*LinkedIn's list for 2021*

*Believe it or not, most of the world is still not ready to completely incorporate AI and data driven workflows (including data scientists). This sets out a new challenge of adaption; from clean toy problems to messy "real' problems of a business. To increase your adaptability, try:*

- Expecting political obstacles; you are not IT, you need to have a nuanced view
- Preparing for delays by creating gap filling objectives
- Looking to have your beliefs(weights) continuously updated about the work.

# What do soft skills really mean?


LinkedIn's list for 2021

*We can not just be the "technical folk", no matter how hard people try to make us that, we have a responsibility to get the whole story. The use case also considers the user---their process and goals. EQ in data science is half of the battle. Otherwise, we are limited to mechanistic analogy.*

- Use technical jargon as needed by the conversant, not you.
- Do not underestimate the contorted AI-world narrative; help simplify it.
- Feel free to express your own weaknesses where relevant
- Seek HITL solutions where they are wanted

# Group Activity


Discuss with your peer which of the 5 soft skills you want to develop (select 1)
Identify 2 actions you are going to take to develop this skill

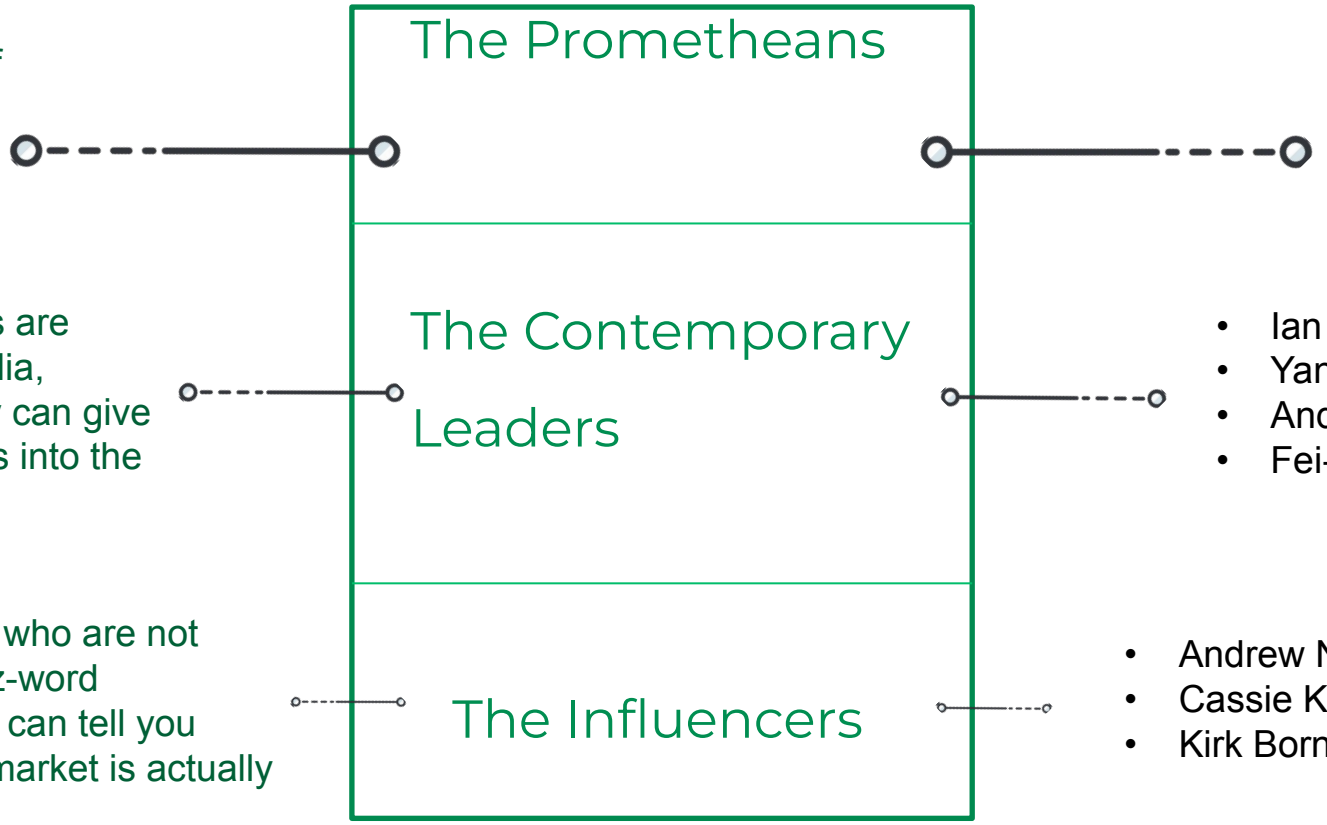
EMERITUS

Learn. From the world's best.

# People To Be Aware Of

# Thought Leaders

If you are taking the study of AI seriously, then it is very important to understand the foundation




- Alan Turing
- Claude Shannon
- Marvin Minsky
- Geoffrey Hinton

The current leaders are often on social media, YouTube, etc. They can give you the key insights into the field

- Ian Goodfellow
- Yann LeCun
- Andrej Karpathy
- Fei-Fei Li

Influencers who are not simply buzz-word marketeers can tell you where the market is actually going

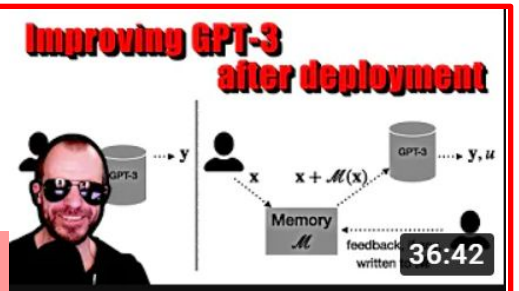
- Andrew Ng
- Cassie Kozyrkov
- Kirk Borne

Yannich kilcher(?) news, keeping up with lit

# Good Signals

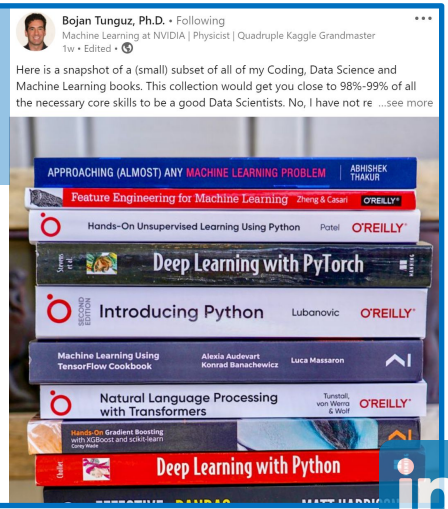
"One of the best sources for ML news and literature review"

Yannic Kilcher



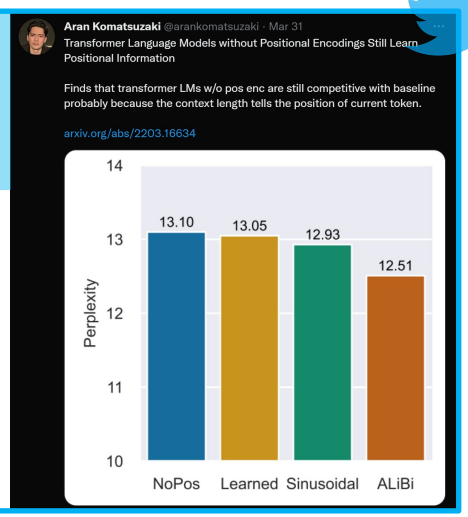

Bojan Tunguz

"A great filter for meaningful AI posts on LinkedIn (not easy to do)"




Aran Komatsuzaki

"Active ML Scientist with great takes on current state & usage"



EMERITUS

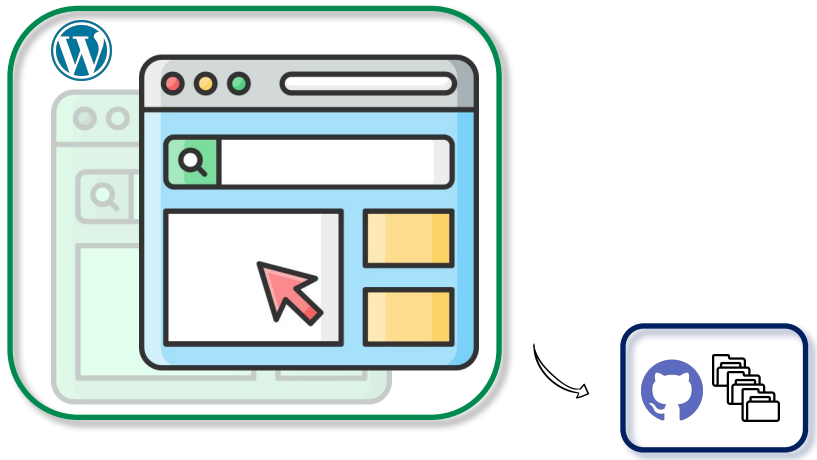
Learn. From the world's best.

# Industry Portfolio

# Where should I build my industry portfolio?

Keep it simple; it is the **content story** that counts

Website



Or


*If you have an idea of what areas you want to work in, one of the best things you can do is find correlated public problems/datasets to work on and add to your portfolio*

# Types of Content

## “Toy” Problems

*Any activity with **no** dependencies on the output*

- *Portfolio content which shows a personal imperative to perform data science activity suggests a candidate is capable of **self-motivated** growth*
- *Portfolio content demonstrating activity in an academic setting suggests a candidate’s **cooperative experience** with other data scientists*

## “Real” Problems

*Any activity **with** dependencies on the output*

- *Portfolio content displaying a role in a larger project with business outcomes communicates a familiarity with the **larger goals** of a business*
- *Portfolio content indicating a candidate’s experience working for a client suggests a **practical capacity** to provide value*

# “Toy” Problems

The goal of toy problems in the data science field is to establish the data scientist in question as having a wide range of theoretical prowess

## Class projects & Lab Assignments

For many who are breaking into the professional world, or pivoting from an adjacent field, these will be the bulk of material in their portfolio and personal story

### Tips

- Don’t over-sell practice problems from class as more than they are
- Try to find correlations of this work and work being described in industry

## Personal Work

Personal work is your opportunity to separate yourself from the pack with messier data and unique solutions. General skill building exercises give you the opportunity to get familiar with contemporary models/architecture

### Tips

- Find data which maps to tasks you want to do in your potential job, not just generic exercises.
- Don’t be afraid to work on things which do not have an existing solution.

## Classic Public Data Sources

Reddit Datasets
Google Dataset Search
Data.gov

# “Real” Problems

For those fortunate enough to have former professional roles, the task is to describe how those roles may compliment you to be a better data scientist

## Professional work from other fields

It is important to recognize the manner in which your previous work is related and informative to understanding how work gets done. Not just data science in a vacuum. This is an advantage

### Tips

- Don’t try too hard to make your previous field of focus the same as data science
- Try to provide why your previous role describes value-based aspects of data science workflows

## Cross-over work

If you have some experience working on, or, with data scientific outputs in your previous role(s), then make sure you accurately describe how you contributed and perhaps how it could be improved with your new knowledge

### Tips

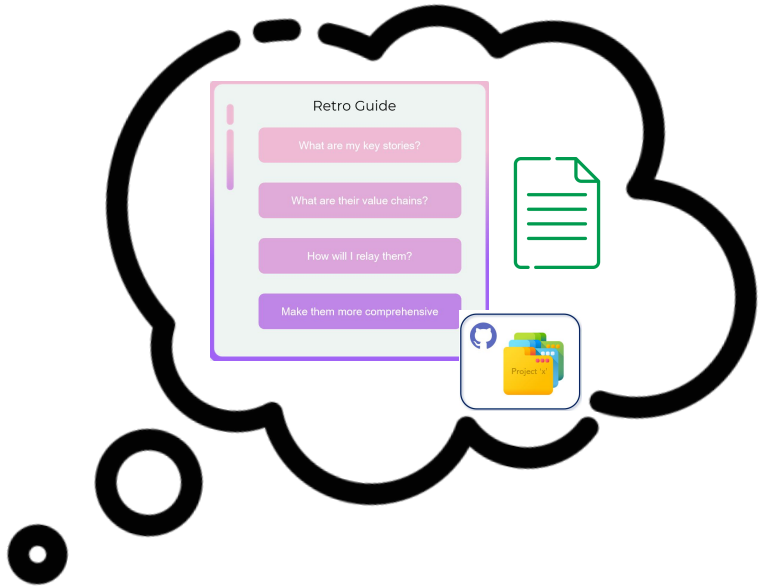
- Don’t overdue your contribution. If you owned part of the data engineering say that.
- Show how working with the modelling effort allows you to understand the larger process now as a data scientist

# Group Activity

Take a few minutes to think and share ideas about the kind of projects you would like to showcase in your portfolio.

# The Industry Portfolio is for you too

How did you set up the problem statement in this project?




## The generative approach

The experienced interviewee hasn't memorized their script, they have worked with it enough to be able to apply it to any situation.

EMERITUS

Learn. From the world's best.

# Index

www.emeritus.org

***Data Science-Specific Interview Questions***

https://www.simplilearn.com/tutorials/data-science-tutorial/data-science-interview-questions

https://hackr.io/blog/data-science-interview-questions

# Useful Links