EMERITUS

Learn. From the world's best.

Industry Overview

www.emeritus.org

- [ ] What does the job market say the hard skills are?
- [ ] Where do AI trends come from?
- [ ] How should I prepare for this changing landscape?
- [ ] What is on the horizon for professional Data Science

Questions you should always be asking

Skills

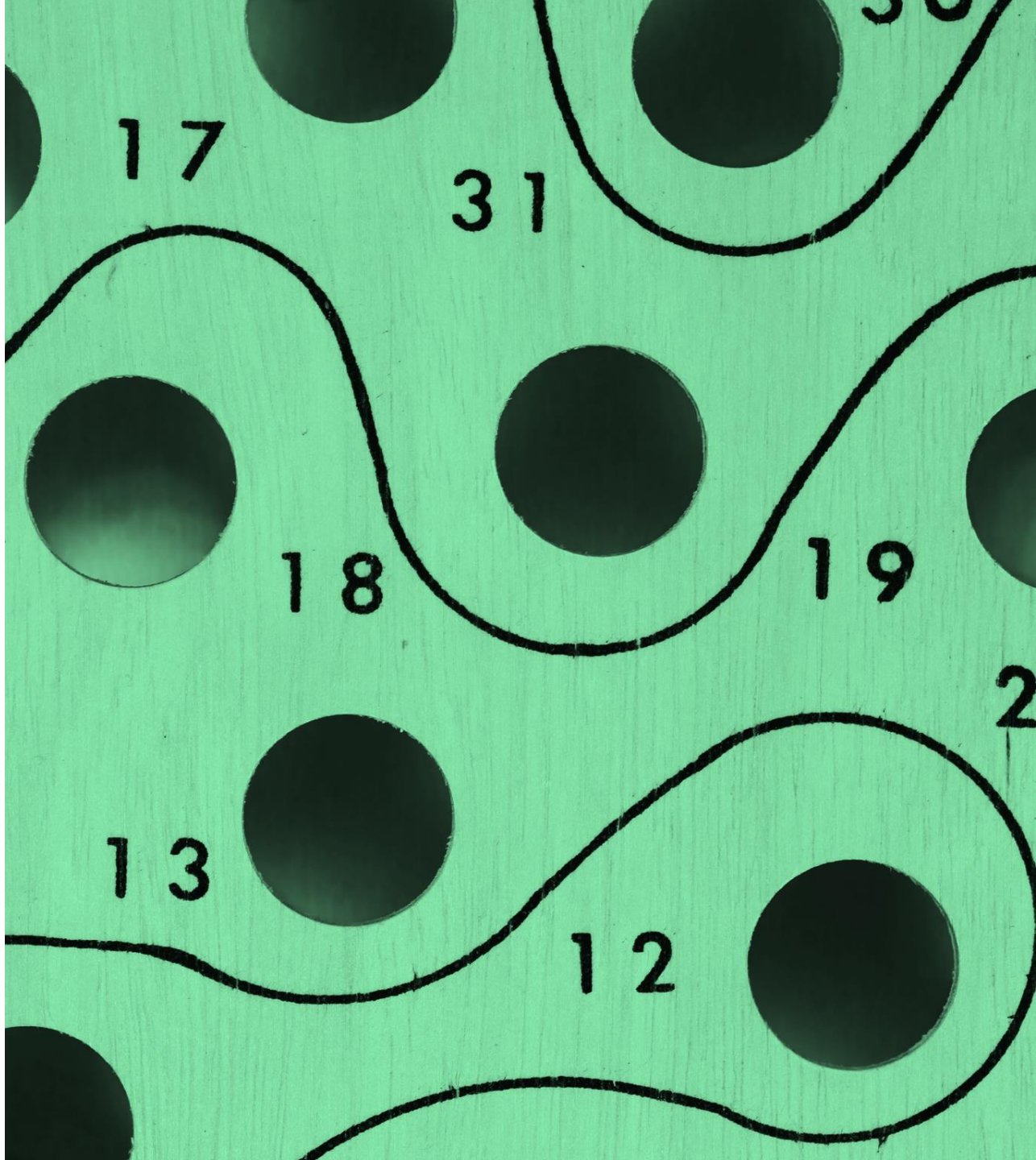
How can I understand which skills map to specific roles I see in the market?

Data Science still contains **more task diversity** than most other fields of work

The relevant skills alter highly from **domain-to-domain**

There exists **certain tools which cover more** territory of use

Trends in AI are hard to follow due to the wide range of performed functions. The **general trends move with Big Tech**



A Sample of the Tools Landscape

Data Manipulation	Software Development	Model Development	Mathematical Concepts
MySQL	Spark	Scikit Learn	Matrix Algebra
mongoDB	Kafka	Tensorflow	Information Theory
Pandas	AWS	Pytorch	Objective Functions
Python	Java	Python	Multivariate Differentiation
Tableau/Qlik	Docker	Spark	Probability Distributions
Ggplot2	Git	AWS	
Kafka	Kubernetes		
AWS			
D3.js			

not scraped

Observed Meta

This combo was the most frequent pair across posts

- One language
- Many common Sub-tools

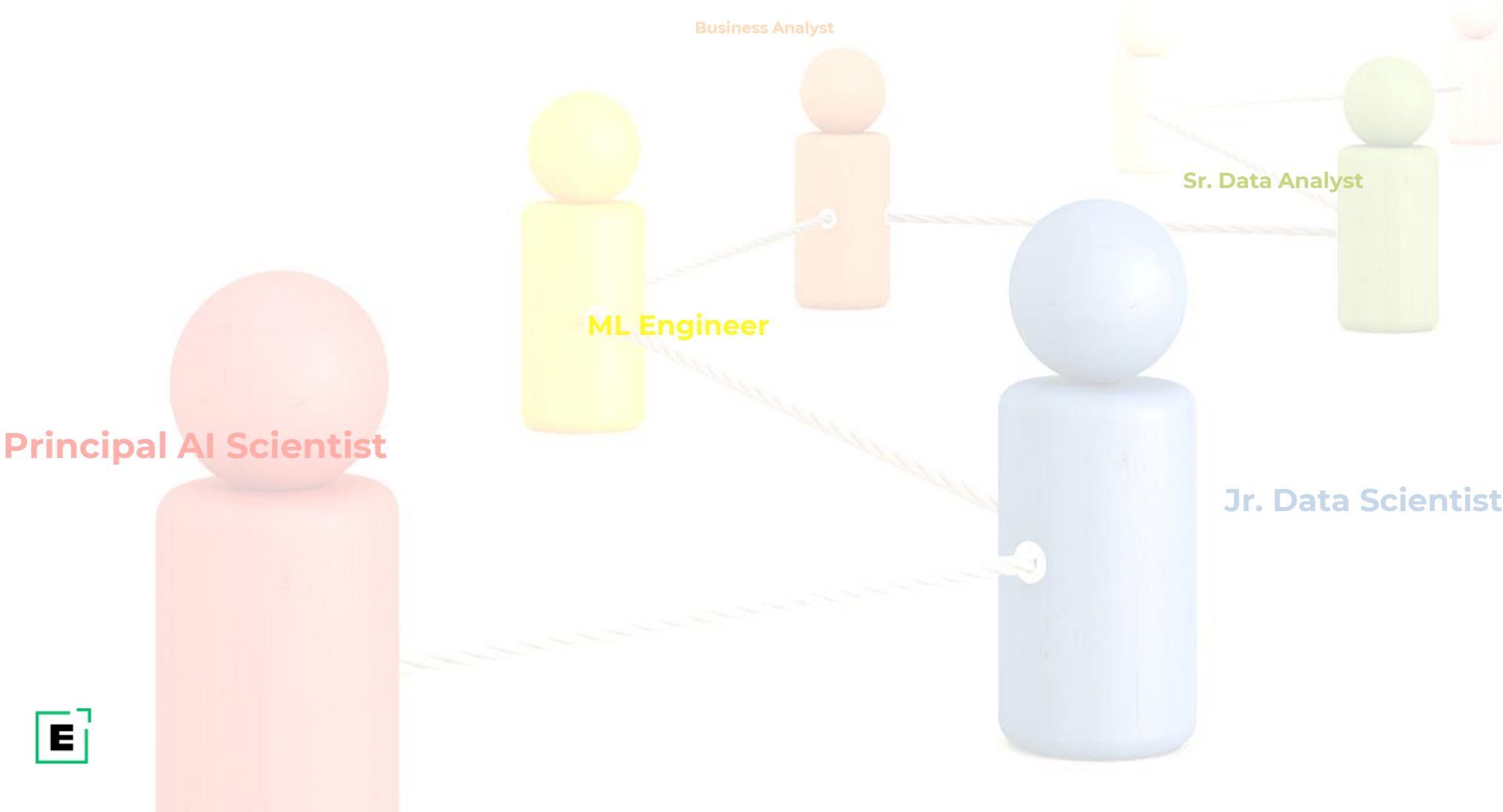
*Data collected from LinkedIn (January 2022)
*Sorting Algorithm produces a frequency weighted sample

EMERITUS

Learn. From the world's best.

From skills to roles

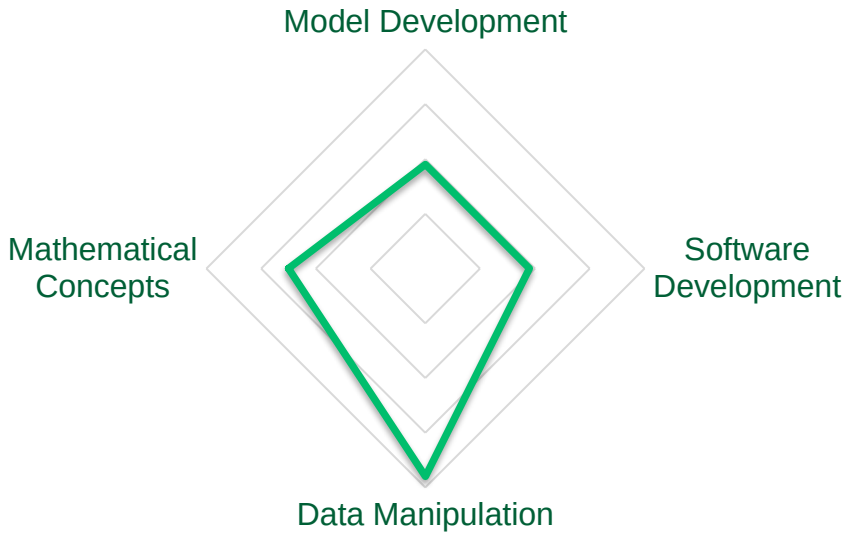
Navigating The Big Roles in Data Science



Data Analyst

The Data Analyst is arguably the most transparent role in the ambiguous world of data science. They provide value by analyzing data. This goes from tableau reports to EDA in python using models and statistical techniques

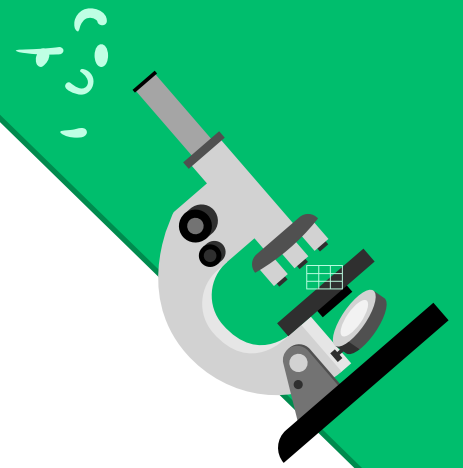
Skill Focus Map



If you are:

- Confident with various databasing languages
- In the process of learning some Data Visualization tools
- Have a feel for using stronger statistics, i.e. PCA/t-SNE, ANOVA, etc.

. . .Then there are roles waiting for you



*Data collected from LinkedIn (January 2022)

AI/ML Engineer

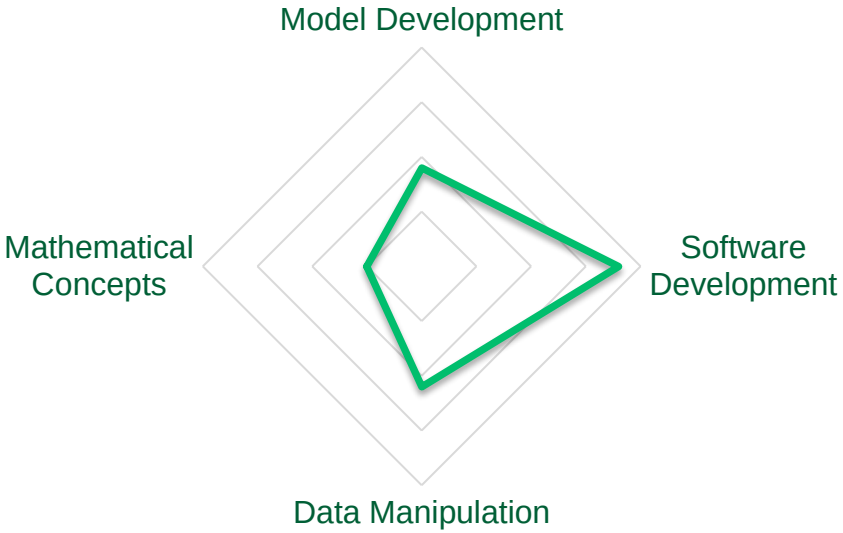
The roles designated by ML or DL engineer will be more solely focused on putting models into production. Depending on where you are, this can be entirely done in managed service or ground-up software development

If you are:

- Already engaged in the Dev/ML ops space community
- From a software engineering background
- Have any experience building pipelines with services such as SageMaker, GCP AI, H20, etc.

. . .then there are roles waiting for you

Skill Focus Map

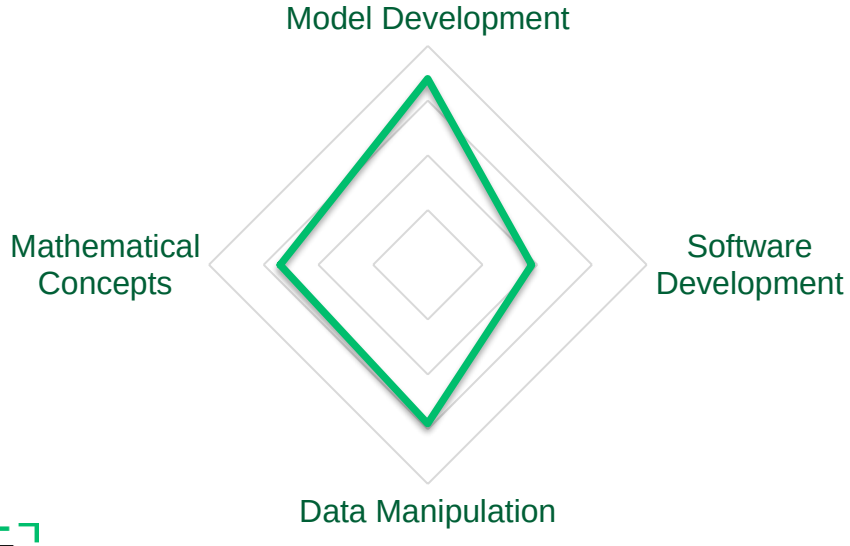


*Data collected from LinkedIn (January 2022)

Data Scientist

The range for roles designated as ‘Data Scientist’ is the broadest. You could work in a small-midsized company where you do everything, or you could be at a large company (or tech related org), where you are solely building Deep Learning models.

Skill Focus Map



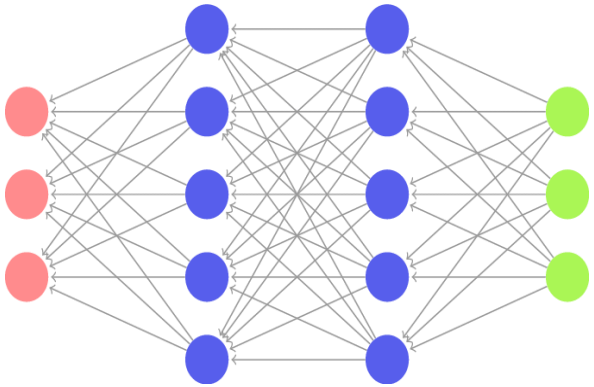
If you are:

- Having successfully trained ML/DL models on many types of problems
- Exhibit a track record of diverse problem solving
- Are familiar with reproducing various topics in the current research literature.

. . .then there are roles waiting for you

*Data collected from LinkedIn (January 2022)

The Data Scientist Role is the hardest to guarantee value



In the data science stack, the Data Scientist's role is most likely to be completing tasks related to AI. That is, they 'own' the model conception, building, and training phase of the pipeline.

The Bad News:

Higher Uncertainty

The Good News

AI is not the biggest obstacle

EMERITUS

Learn. From the world's best.

Organizations and AI Trends

Understanding the flow of tech

Big Tech

Big technology organizations, like those in FAANG(now MAANG), have a certain monopoly on AI. They specifically own and distribute the computational function of modelling. Most computational workloads are now being performed in the cloud, as opposed to on-prem servers. AI modelling is no different. Data Science roles within big tech are also **the most precisely outlined and coordinated for production**, since they remain software companies first

Corporate Roles

Corporate roles make up a large bulk of the consistently available jobs on the market. These roles are with large organizations which are not primarily producing broad use software for customers. Therefore, most active data scientists are familiar with applying their work to a specific domain. Corporate DS roles tend to **rely more heavily on communication** compared to others

Smaller Organizations

Roles at smaller companies make up the other large portion of available jobs. Smaller companies have the same large diversity of domains for which data scientist must apply their craft as corporate entities, but often even less technical structure. It is very common for a data scientist at a small company to be **wearing many hats** in the role of production

Influence

Big Tech Roles

Best suited for those with significant interest in software development. These roles are highly managed through agile frameworks and code management protocol

- **Technical architecture**
- **Multiple programming languages**
- **Large Dataset Experience**
- **Working on engineering/software teams**

Big Tech often sets the trends on workstyles and tools for the rest of the industry to adopt and integrate

Data Science Goals

Large scale B2B & B2C products

Areas to Focus

End-to-end product development

- Dev/ML ops
- Batch & Stream Processing

Current SoTA Research

- Deep Learning Architectures
- Transformer based NLP

Corporate Roles

Roles in larger corporations tend to suit those looking for consistent schedules and high interaction. A data scientist in these roles is often engaging internal teams & clients for their scope of work.

- **Lots of database querying**
- **Time management**
- **Understanding the client domain**

Corporate data science roles tend to apply trends from big tech in order to create solutions for a particular demographic or clientele

Data Science Goals

Client Services/Outcomes

Areas to Focus

Cloud ML Platforms

- AutoML Frameworks
- SageMaker, GAIP, etc

Big Data Mmgt (PaaS)

- SQL, PostGreSQL, MongoDB
- Hadoop/Spark,
- Visualization

Small Organization Roles

Roles in smaller companies can be very ideal for those looking for more independence and decision making in their workflow. It is not unusual to be the only data scientist, or among few.

- **Data engineering and analysis work**
- **Stakeholder management**
- **Focusing on internal objectives**
- **Environment setup and ownership**

Roles within smaller companies commonly require more vision establishing and comprehensive work

Data Science Goals

Internal Projects & Asset Development

Areas to Focus

Value Proposition

- Problem Statements
- Reporting Schema Creation
- Dataset Creation

Architectures and frameworks

- Matching cloud-to-AI services
- Lightweight python notebooks

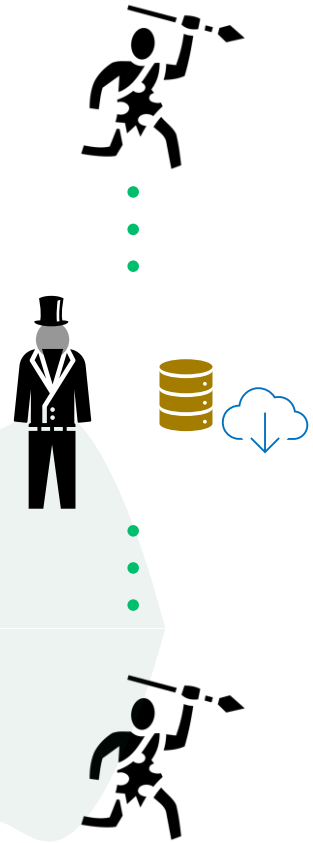
EMERITUS

Learn. From the world's best.

The Short History and Market Dynamics

Why the fuss?

Data Science is in such high demand because it is the most tangible way to extract value from an organization's data



- In the **1980s**, Relational Databases became ubiquitous along with the wider commercial use of computers
- IBM, Microsoft, Oracle, and others led the way into corporate use of DBs through the creation of **broad-use applications** with the general UI/UX we still have now
- After the data honeymoon of the early 2000s, organizations wanted more than base utility. They wanted to see their data as a renewable resource for intelligence
- In the **2010s**, companies across many industries started investing heavily in **Data Lakes**. Spurred on by current academic success in Machine Learning, there was now a significant promise of use for their heaps of data
- A little over 10 years later; Data Science and its models are everywhere, but the return on investment for data lakes is not clear for everyone yet

The Demand

564,978 Job postings in US asking for ML or AI skills from July 2020 to June 2021 (IDC)

112,198 Job postings in the UK asking for ML or AI skills from July 2020 to June 2021 (Emsi)

$145,000 Average US salary for an entry-level ML engineer with 0-2 years of experience (Emsi)

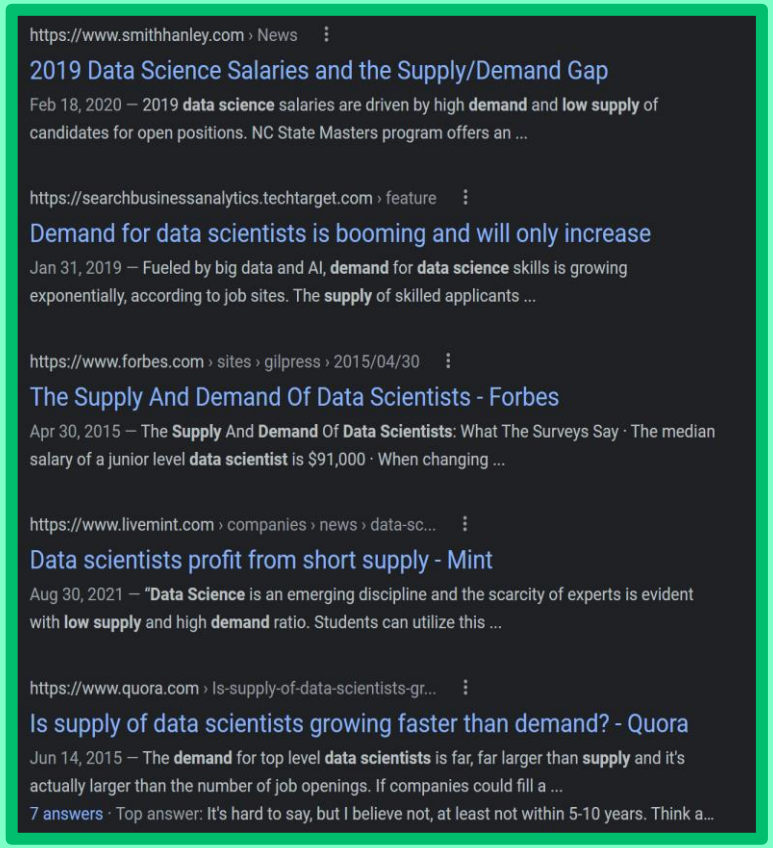
£60,000 Average UK salary for an entry-level ML engineer with 0-2 years of experience (Emsi)

A note on wage

It is extremely important to recognize the invalidity of collective statistics like mean and variance of salaries as reported. One company could pay twice as much, or little, for the same role for no tangible reason

Learn the submarkets of interest!

How's the Supply goin?



Still low...

Job postings 10 years ago

Hiring Manager reqs:
- PhD in Math or related field
- 10 years of experience
- R, Python, C, Java, etc
- Many business cases

While the job posts of the past may have been more authentic, they assumed a new industrial field somehow had years of experience, resulting in only PhDs being possible to hire

Applicants 10 years ago

" My undergrad was in computational biology; I recently had some success using decision trees to predict ingredient activity from raman-spectroscopic signal data"

Job postings now

Recruiter:
- Bachelors' degree or n years of experience
- Googles a handful of topical skills and hopes the hiring manager has input on specific info
- Copies half of info from another posting

The high AI demand and marketing may have made talent acquisition more generic, but it also has opened the field to **anyone** with the right skills

Applicants now

" I was able to get 99.1% accuracy on MNIST using VGG19 with dropout in 5 hours of training"

Learn the domain you are applying to!

Some still want PhD-SMEs
Many Do Not

Existing SME vs **Learning Subject Matter**

The biggest thing most burgeoning data scientists need to work on in this AI crazed world, is learning the client/user subject matter. Most AI projects are failing now because of a subject gap between the users and project scoping

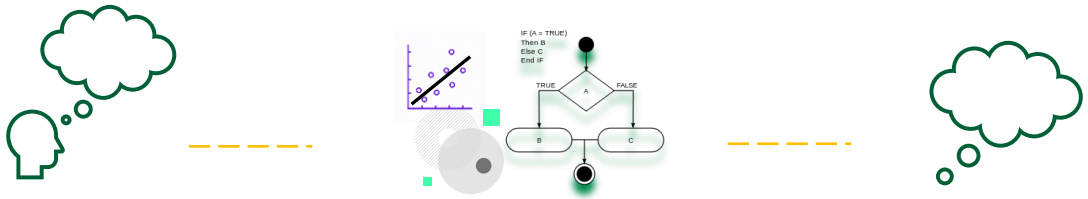
As you continue to learn new algorithms, frameworks, and skills, keep laser focused on developing meaningful understanding of subject matter. Focus on:

1. Use value **2. The problem statement** **3. User Workflow** **4. Storytelling**

The Contemporary Crisis for Data Science

2010s — *2020s*

Deployment Scarcity



Investment – Reward Ratio

Poor subject matter understanding
Poor project planning

No maintenance contracts
No deployment
No value

The Professional Problem Statement

We must increase the probability of success for AI initiatives

The Twilight of AI Value

How are we solving this current crisis?

- The academic field is continuing to accelerate
- We have established the business precedent or Data Science
- We are in the middle of generalizing workflows
- We have just started to identify client/user/subject engagement methods
- MLops is starting to mature
- Data Lakes are beginning to contain data for modelling

There are no **scientists** who begin investigating a subject by not attempting to understand it. Evidently, Data Science can be **no different**

Data Science is **not** about using data to have no bias. Data Science allows the exploration of **many biases**

Now is the time for those who don't have the most skills

→ Strong subject understanding beats all the skills in the world

EMERITUS

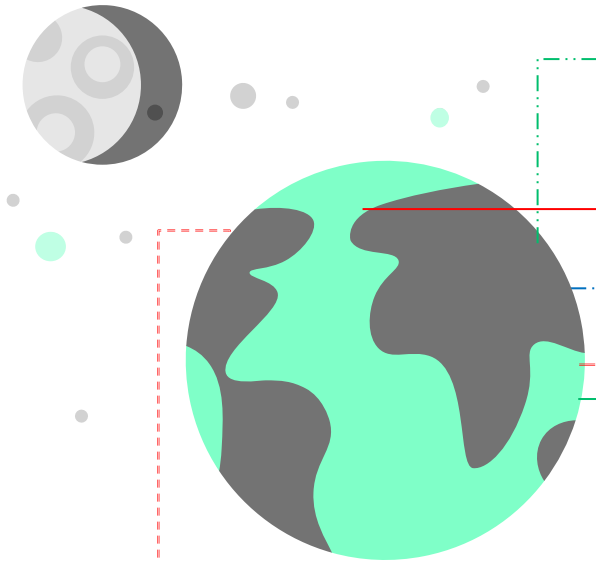
Learn. From the world's best.

Lifestyle Considerations

The Global Data Scientist

The EU are playing the long game and leading the regulatory framing. From east to west, the EU is also the highest % remote AI workforce

The US is known for having the highest compensation and trend setting due to Big Tech influence. Every state has increased hiring rates for data science roles



The UK, had a slower start but is commonly cited as now being one of the fastest growing DS hotspots

India's corporate tech boom has neatly lined up with recent data science growth, making them an essential international player

China has the largest organized investment of any country in the AI space. Leading in areas such as img-rec, China is paving a way for societal interaction

Canada has the most structural plans, behind China, to take their ML origins to scale

A Remote Possibility

In the post-pandemic world, data science is even more likely to continue its growth of remote roles across many industries

This sample workday is now very easily accomplished outside of an office. For those going into the future data science workforce, it is important to determine what environmental factors allow you to feel most effective in the On-Prem / Hybrid / Remote split

Something to consider before you go

On-Prem

Highest face-to-face communication

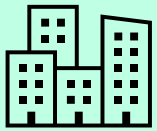
More reported creative meetings

Employees in the UK/US reported a better sense of community when on prem

Higher avg auditory stimulus

Daily transportation

A common complaint from DS employees on Glassdoor in 2019 was not having enough time for developing in the office



Hybrid

Mix of office vs home biases

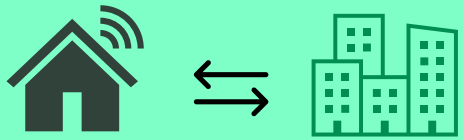
Work-life balance polls favor hybrid split

The tech-sector reported the hybrid model as preferred for team building

Less consistency for habit building

Individual's routines tend to diverge

While the hybrid model has done well within teams, there is evidence as to weakened interaction between teams



Remote

Increased time allocation

Transportation is unnecessary

Workers across technical disciplines cite remote work as accelerating their output

Less reinforcement signals for support

Increased emotional alienation of labor

There is a growing split of workers who find at-home environments