



On Transitioning from Academia to Data Science

Noa Tamir, April 18th 2021, APS April Meeting,   @noatamir

About me

- B.Sc. Physics, Tel Aviv university
- M.Sc. Business & Management Science, Humboldt University Berlin
- M.Res. Economics, European University Institute
- Industries: Space, Design, Gaming, e-Learning, Automotive, e-Agriculture
- Independent Data Consultant and Educator
- External Data Science Lecturer at HTW Berlin, University of Applied Sciences
- Co-founded the Women in Machine Learning and Data Science Lab

On Transitioning from Academia to Data Science

The 5 Core Data Science Skills

The Gap in Skills, Processes, and Other Things

Teaching a Data Science Lab to Support Transitioning to Industry

The Skills

The 5 Core Data Science Skills

Science Toolkit

Science Toolkit:

Mathematics, Algorithms

Statistics, Experimental Design, Hypothesis Testing, Modelling

Data Platforms

Data Platform:

SQL

Wrangling, i.e. using SQL, or Python/R packages effectively

Using large, cloud based, data lakes and warehouses

Programming

Programming:

Good programming skills in Python / R:

documentation, testing, structure, IDEs, collaborative version control

Creating reproducible code

Understanding application/package deployment/release

Communication

Communication:

Listening skills

Data Visualisation

Technical writing and presentation to other domain experts

Receiving and giving feedback

Business and Domain Expertise

Business and Domain Expertise:

Understanding business goals, and pragmatic thinking

Timely and effective collaboration with other domain experts

Translating business problems into data/analytical problems

Mining domain specific insights / engineering domain specific features

The Gap

The Gap in Skills, Processes, and Other Things

Science Toolkit

Science Toolkit:

Exploratory Data Analysis (EDA), or the importance of a thorough investigation of a real world messy data set

Evaluation: business impact > effect size or model accuracy

Understanding research purpose and method selection

Data Platforms

Data Platform:

SQL

Using large, cloud based, data lakes and warehouses

Programming

Programming:

Writing and sharing reproducible and structured code

Code style and documentation

Working on an open / collaborative code base

Using version control

Communication

Communication:

Data Visualisation

Technical writing and presentation to other domain experts

Code reviews

Team/stakeholder communication, building effective work relationships

Business and Domain Expertise

Business and Domain Expertise:

Measuring and communicating business impact

Collaborating with domain experts

Processes

Processes:

Project management

prioritisation, good enough, agile, estimation

Research & *Development*, aka delivery

Familiarity with work culture, and operational departments

Other things

Other things...

Being perceived as overly precision oriented or not business oriented

Job Market variety and clarity

Lack of industry standards for roles and terminology

A professional network

Communicating your experience effectively across the gap

Teaching

Teaching a Data Science Lab to Support Transitioning to Industry

The lab is a program to apply existing skills, and hone them from on-the-job learning

The Goal of a Data Science Lab

Students gain familiarity with work processes, and practice the required communication skills at the core of a data scientist's job

A better understanding of what's expected from you at work

Real-world data
Weekly planning
Receiving and following guidance from leads
Sharing your ongoing work and working together on not-perfect code/ideas
Team collaboration
Peer 2 peer code review
Pair Programming
Communication to stakeholders
Documentation for project handover

Intro to Project
Client Pitch
Weekly mentored sprints:
EDA, Cleaning, Modelling, Evaluation
Client Presentation
Technical Handover/Presentation

Experience in industry processes and communication

A portfolio project for one's
résumé and presentation
interviews

also..

Practical Data Visualisation
Business Thinking
Data Tracking, Quality, and Modelling
Data Project Management
Engineering and Data Engineering
APIs and Application Deployment
Version Control and IDE workflows

MOOCs
FOSS Communities
Bootcamps
Competitions and Hackathons
Internships
Books, Videos, and Podcasts



Thank you.

Noa Tamir, April 18th 2021, APS April Meeting,   @noatamir



WiMLDS Github: “Berlin DS Lab”

Techpoint Charlie Podcast: “Let’s Make Sense of Data Science”, “What is the Role of a Data Scientist?”

YouTube: PyData Talk “Professional Development and Career Progression for Data Scientists”, eRUM Talk & AI in Practice Meetup “Data Culture in Practice”

Noa Tamir, April 18th 2021, APS April Meeting,   @noatamir



Links:

<https://www.youtube.com/watch?v=G3IW5RNSGcE>

<https://www.youtube.com/watch?v=Bt7-7yXG8A4>

<https://youtu.be/UrO8EN-BZmE?t=1529>

<https://github.com/wimlds/berlin-ds-lab>

<https://github.com/noatamir/Talks>

<https://techpointcharlie.blog/>

Noa Tamir, April 18th 2021, APS April Meeting,   @noatamir

Bonus topic, which I did not
present

The Keepers

What do people keep and cherish
about from their academic
experience

Independent work and self motivation

Continuous learning of new techniques

Analytical thinking and Pragmatism

(if one's supervisor was publication/funding oriented)

Statistics and Experimental Design:

Interpreting regression coefficients and being able to explain their meaning to others

Knowing several regression types

Knowing how to phrase a testable hypothesis and validate it

Understanding how biases evolve and stick

Advanced Mathematics and Algorithms

Communication and Collaboration, especially reviews

How to present data and results accurately and clearly (although perhaps not enough to a non-science audience).

Being able to write and communicate difficult concepts clearly - this is really important for team work and for collaborating with people who have varying degrees of data and statistical knowledge.

writing articles for multi-disciplinary academic journals, where you have to very clearly and very briefly justify to non-specialists why your research and finding are important.