An Ode To Data,

or, Careers in Data Science for Physicists

WHAT IS DATA SCIENCE?

Some takes:

- Data science is just statistics
- A data scientist is a business analyst who lives in San Francisco
- Better at statistics than most software engineers and better at software engineering than most statisticians
- A jobs program for STEM PhDs

…I'll come back to this

WHY NOW?

- "Big data": Increasingly huge amounts of data available.

- "Cloud computing": Cheap, scalable storage and processing power.

- Companies are more aware of data as an asset

"We've got our big data running on the cloud, now get a data scientist to give me some insights!" – CEO Business

WHY PHYSICISTS?

- Comfortable with math
- Comfortable with programming
- Collaboration & communication in a complex subject
- Mix of intuition and rigor, a sense of where your model/assumptions may be wrong but still "good enough"

MY EXPERIENCE

Silicon Valley Data Science

"It's what we do and where we do it"

- Data science consulting
- Software engineers, statisticians, physicists, …
- Strong collaboration of data science and data engineering
- "Agile" development methods

A (day, month, year) in the life

- Get asked a question by business
  - *"Who are our most loyal customers?"*
- Break it down into something you can model
  - *Which customers are most likely to make a purchase in the future?*
- Dig in to the data
  - *What data do we have and how do we get it?*
  - *How reliable is the data (missing/bad values)?*
  - *How is the data distributed?*

A (day, month, year) in the life

- Explore models
  - *Start simple, maybe logistic regression classifier*
- Communicate results and iterate on modelling
  - *Is the simple model good enough?*
- Scale and deploy the model
  - *Code clarity, efficiency, reliability, testing, error handling, edge cases*
- Lather/rinse/repeat
  - *How do we get more loyal customers?*

Is this for you?

- Analyzing data and writing code is fun!

- Much less time to get satisfying results

- Normal day job and all the benefits that come with that (income, people value your time, work/life balance)

- *But,* you're working for the man

  - Less freedom to choose problems

  - You're a cog in the global capital allocation machine

Sold! How do I get there?

- If you haven't done much programming, start learning now! (Python is a good place to start)
- Brush up on applied statistics and machine learning (books or courses)
- Pick a fun side project that forces you to use the above skills and get to work!
  - See it through to the end: put code on Github, make a blog post and/or demo website
- Do an internship (best), or Insight (good), or similar program (careful)

Kevin Zielnicki
kevin@svds.com

SILICON VALLEY
DATA SCIENCE

svds.com/hiring