ASTROPHYSICS TO DATA SCIENCE

February 26th, 2016

Dr. Ali Vanderveld

Senior Data Scientist @ Groupon

ali.vanderveld@gmail.com

Background

- BS in physics from UIUC in 2001
- PhD in astrophysics from Cornell in 2007
- Two postdocs
  - Caltech / NASA JPL
  - KICP fellowship at the University of Chicago
- Research
  - Theoretical late-universe cosmology
  - “Dark sector”
  - Weak gravitational lensing
  - SNAP (now WFIRST), HALO, Euclid
- At Groupon since July 2013

“Data science”

- Answering often complex business / policy / medical / etc. questions with data
- Using machine learning and data mining techniques to detect trends and to optimize systems, e.g.
  - Relevance engines
  - Website widgets, like “Customers Also Viewed”
  - Optimization systems like Quantum Lead @ Groupon
- Large overlaps with computer science and econometrics
- A physics background is (mostly) excellent preparation

Why I changed fields

- Wanted to stay in Chicago, and generally to live where I want
- Did not want to be a professor
- Not willing to work more than 40 hours a week
- Was bored with astrophysics
- Wanted something more people-oriented
- Hoping to one day (soon) work in public policy

Data science @ Groupon

The Chicago wing of the Data Science team focuses on sales and the Local business. My projects have included:

- Predicting demand as a function of service and location using sales and search query data
- Determining user travel patterns from historical order and mobile GPS data (see Groupon Engineering blog post)
- Sales analytics, e.g. determining best balance between breadth and depth
- Predicting user value over the next 90, 180, and 365 days based on past behavior

Day-to-day work

- Much like research at a postdoc level
- Very autonomous
- Lots of coding, lots of SQL, lots of data
- Answering questions and building models
- Working with engineering to upgrade production systems
- Google Docs and visualizations in Tableau
- Legal team helps patent everything
- Annual conferences and papers, e.g. KDD
- Explaining results to executives :(

Benefits of data science in tech

- Extremely plentiful jobs
- Excellent compensation
- Autonomy and flexibility
- Informal atmosphere
- Good work-life balance
- Wide variety of problems to work on
- Fast pace
- Immediate, tangible results

Downsides

- Relative lack of job security
- Large employee churn
- Executives
- Selling a product vs studying nature
- Horrible business lingo, e.g. “algos,” “touch base,” “deep dive”

Getting a job

- Preparation
  - Coursera – data science, data mining, machine learning, R, etc.
  - Side projects with public data – census, Chicago open gov data
  - Personal website and LinkedIn
  - Knowing people helps – meetups and networking
- Need experience in
  - Programming – SQL, Python, R
  - Working with data
  - Statistics
  - Machine learning and data mining

Getting a job

- Finding jobs
  - Glassdoor, company websites
  - Network via Meetups, LinkedIn, friends of friends, anything
- Interviewing
  - Extremely different from academic interviews
  - On-the-spot problem solving and coding
  - Looking for a good colleague
  - Not expecting 100% correct answers