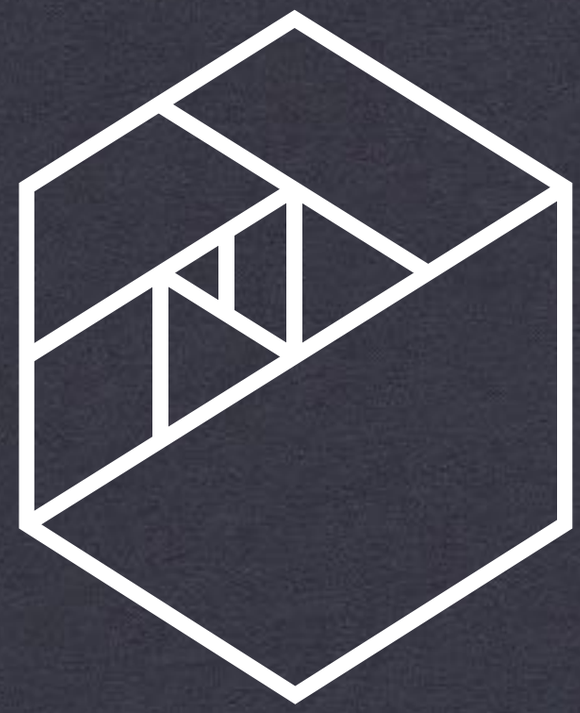
TRANSITIONING FROM MATH TO DATA SCIENCE

TAYLAN BILAL
SENIOR DATA SCIENTIST - METIS
TAYLAN@THISISMETIS.COM



METIS

Contents

1. From academia to industry
2. What is Data Science?
3. Skill Set of a Data Scientist
4. How to acquire these skills?
5. Data Science Bootcamps and Metis

About me

- I am a Senior Data Scientist @ Metis, a data science training company.
- I have been working as a data scientist in the industry for 4 years.
- I was a Ph.D. Student @ USC, 2008-13
- Dissertation in Pure Math, Algebraic Geometry.
- I decided to pursue a career in the industry, specifically Data Science, during my final year of Ph.D.

From academia to industry

- A few challenges awaiting you…
  - OPT / Visa if you are not authorized to work.
  - Putting together a good enough resume.
  - “Able and eager to learn” is not going to convince anyone.
  - Degrees are attractive, but perhaps not as much as experience / portfolio.

From academia to industry

- Lifestyle is different.
  - More structured; 9 to 5 every day, less flexibility, grind.
  - More structured; work ends… No guilt of “not working” on the weekends.
  - Coworkers with a wider spectrum of backgrounds.
  - Building something (practical) vs research (theoretical)
  - Fact of life: Industry pays better than academia.

From academia to industry

- Lifestyle also differs from company to company.
  - Startups: More demanding, instant impact, responsibility, ups and downs, no playbook.
  - Established companies: Mostly well-defined, professional, smaller role, less risk, less independence.

From **Math** to industry

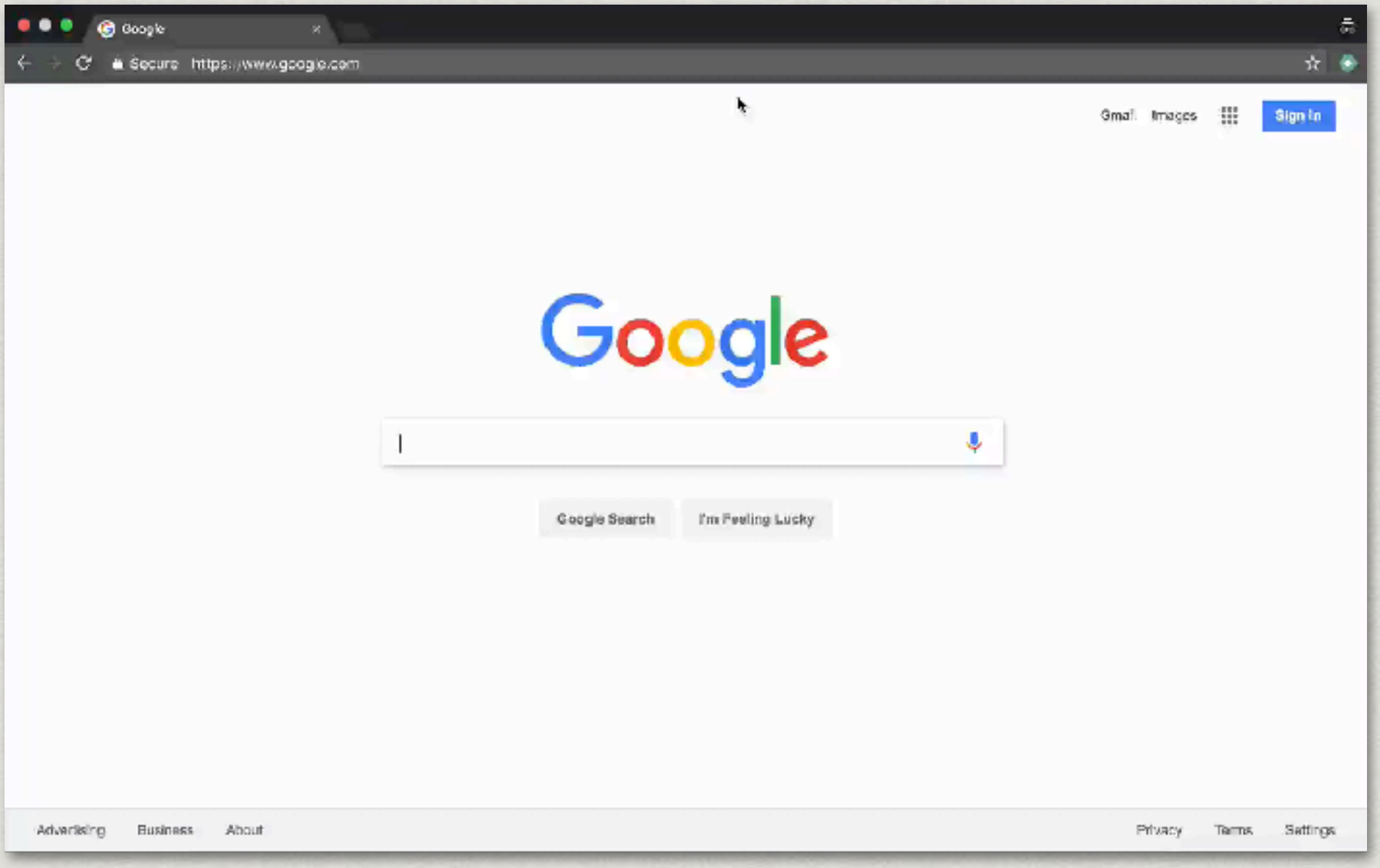
- Most likely destinations for math grads in the work force are;
    - Finance / Actuary
    - Software Engineering
    - **Data Science**

From **Math** to industry

- Most likely destinations for math grads in the work force are;
  - Finance / Actuary
  - Software Engineering
  - **Data Science** - Perhaps the easiest to transition to.

What is Data Science?

- **A LOT** of data is being created, constantly.



What is Data Science?

- **A LOT** of data is being created, constantly.
  - User activity on a webpage.
    - How long did they stay on a webpage?
    - Did they read the FAQs?
    - Did they leave a favorable review for their purchase?
    - …

What is Data Science?

- **A LOT** of data is being created, constantly.
  - User activity on a webpage.
  - Social media posts
  - Search strings
  - Location / movement
  - …

What is Data Science?

- **A LOT** of data is being created, constantly.
- Data is extremely valuable for companies.
  - Better recommendations: movies, articles, interesting ads
  - Better predictions: medical diagnoses, stopping malicious online behavior such as fraudulent transactions, less wasteful resource/ inventory management
  - Artificial Intelligence: chatbots (Siri, Alexa, etc), self-driving cars (less accidents, less traffic), speech recognition, computer vision
  - More…

What is Data Science?

- **A LOT** of data is being created, constantly.
- Data is extremely valuable for companies.
- Stories / visuals created from data are powerful.
  - Hans Rosling - Best Stats You've Ever Seen
- PROBLEM!

What is Data Science?

- **A LOT** of data is being created, constantly.
- Data is extremely valuable for companies.
- Stories / visuals created from data are powerful.
    - Hans Rosling - Best Stats You've Ever Seen
- PROBLEM! Abundance of data is new. Things are messy.

What is Data Science?

- Problem: Abundance of data is new. Things are messy.
    - Data is raw. It does not organize itself.
    - Not all generated data relates to a given problem. Problems do not define themselves in terms of data.
    - Meaningful information does not derive itself.

What is Data Science?

- Problem: Abundance of data is new. Things are messy.
- Dealing with this requires a broad set of skills.
  - Typical CS major does not have great understanding of mathematics / statistics.
  - Typical Math / Stats major does not have great computer skills to “do” things.

Data Science Skill Set

1. Mathematics, Probability, Statistics
    - Distributions, parameter estimation, derivatives, gradients, optimization…
2. SWE: Algorithms and Data Structures
    - Arrays, hash tables, complexity…
3. Programming: Unix and Python
    - pandas, scikit-learn, numpy, matplotlib, jupyter notebooks…
4. Version control:
    - git, maintaining code, writing clean and understandable code
5. Machine Learning
    - Classification, Regression, NLP, which also to use when…

Data Science Skill Set

6. Databases
    - SQL, indexes, inserting / retrieving data…
7. Data Manipulation
    - Cleaning, transforming, sampling data
8. Big Data tools
    - Spark, Hive, Hadoop…
9. Cloud computing tools
    - Amazon Web Services, Microsoft Azure…
10. Communication
    - Turning data into stories, visualization, explaining findings and challenges to the business side of the company.

That is overwhelming, but…

- This is why Data Scientists are in such high demand.
- Some of these skills are not necessary for some data related jobs.
- Some of these skills can be acquired / polished on the job.
- It all starts from **MATHEMATICS.**

Mathematics

- Math gives you the most important skill of all: How to learn.
- Combine that with grit, you can do it!
- Math - Prob - Stats is the foundation.
- It is okay if you don’t know. You can learn.

Learning the Skills

- There is a plethora of online resources (of varying quality)
    - r/datascience, r/machinelearning, r/learnpython
    - Youtube, google, stackoverflow.
    - MOOCs - Coursera, Udacity, Udemy
    - https://github.com/donnemartin/data-science-ipython-notebooks
    - codecademy.com
- https://datascopeanalytics.com/blog/how-do-i-become-a-data-scientist-an-evaluation-of-3-alternatives/

Learning the Skills - Bootcamps

- Bootcamps! **Metis,** General Assembly, Galvanize…
- Bootcamps are typically 12 week intense programs, covering everything a modern data scientist must know.
- Not just theoretical education.
- Design and build data science projects end to end.
- Build an online portfolio — very attractive to employers.
- Help with job hunting.

Learning the Skills - Bootcamps vs MOOCs

Bootcamps

- Costs ~15K
- Lasts ~3months
- Immersive, collaborative
- Project oriented
- Selective high quality content
- Help from experienced data scientists
- Job hunt help from the careers team and company's connections.

MOOCs

- Costs 0
- Lasts > 3 months
- Isolated
- Theory oriented
- Content of varying quality
- Help from static sources like google, stackoverflow…
- Self executed job search, preparing resume, interview preparation etc.

Some facts about Metis

- Only accredited Data Science bootcamp.
- Operates in NYC, San Francisco, Chicago, Seattle.
- Wide range of student backgrounds - 50% have advanced degrees; 75% come to Metis having previously worked in industry
- Graduates work at companies such as Facebook, Tesla, Apple, Spotify, IBM…
- Here are some real-world results:

Some facts about Metis SF Winter 2016 Cohort

GRADUATE 6
Data Scientist

GRADUATE 11
Data Scientist

GRADUATE 3
Data Scientist Researcher

GRADUATE 4
Senior Business Intelligence Analyst

GRADUATE 5
Contributing Data Science Writer

GRADUATE 1
CTO

GRADUATE 7
Core Faculty

GRADUATE 8
Data Scientist

GRADUATE 9
Data Scientist/Engineer

GRADUATE 10
Data Scientist

GRADUATE 2
Data Science Intern

GRADUATE 12
Data Scientist

GRADUATE 13
Lead Data Scientist

GRADUATE 14
Data Scientist

GRADUATE 15 *Waived career support*

NY Winter 2016 Cohort

GRADUATE 1 *Data Scientist*	GRADUATE 2 *Data journalist*	GRADUATE 3 *Data Scientist*	GRADUATE 4 *Data Scientist*	GRADUATE 5 *Associate Product Analyst*
GRADUATE 6 *Data Analyst*	GRADUATE 7 *Associate Data Scientist*	GRADUATE 8 *Senior Data Analyst*	GRADUATE 9 *Analytics & Insights Senior Analyst*	GRADUATE 10 *Senior Data Scientist*
GRADUATE 11 *Junior Data Scientist*	GRADUATE 12 *Junior Data Scientist*	GRADUATE 13 *Data Scientist*	GRADUATE 14 *Senior Manager of Retail Analytics*	GRADUATE 15 *Senior Statistical Analyst*
GRADUATE 16 *Physics Teacher*	GRADUATE 17 *Associate Director of Analytics*	GRADUATE 18 *Data Scientist*		

GRADUATE 19	*Working in non data science-related field*
GRADUATE 20	*job searching*
GRADUATE 21	*Waived career support*
GRADUATE 22	*job searching*

Get Started

- Get a Mac / Linux computer.
- Sign up for a GitHub account.
- Pick your favorite programming text editor and start coding. Learn how to execute your code.
- Install anaconda (it's a collection of cool packages).
- Check out the scikit-learn's documentation page. Try to run the examples on your machine in a Jupiter notebook.
- Metis Bootcamp **pre-work** repo:

https://github.com/thisismetis/dsp

Thank You!
Questions?

taylan@thisismetis.com

linkedin.com/in/taylanbil

METIS

www.thisismetis.com