

Getting into Data Science

Ivan Teo
Data Scientist, Stitch Fix
25 April 2019

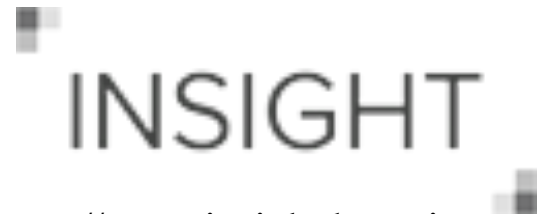


My Journey



Theoretical and
Computational
Biophysics Group
(Klaus Schulten Lab)

2009 - 2016



<https://www.insightdatascience.com/>

2016



2017 - 2018



STITCH FIX

2019

Life in data science

Key areas	Machine learning	Experimentation	Analytics
		AB testing Multi-armed bandits	
Business applications	Forecasting Predictive servicing Risk management Ops optimization	UI/UX Model effectiveness	ROI User engagement Operational efficiency

Life in data science

Key areas

Machine learning

Experimentation

Analytics

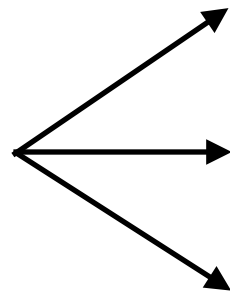
AB testing
Multi-armed bandits

Business applications

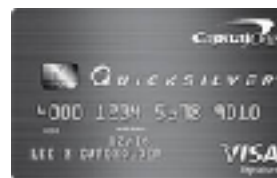
Forecasting
Predictive servicing
Risk management
Ops optimization

UI/UX
Model effectiveness

ROI
User engagement
Operational efficiency



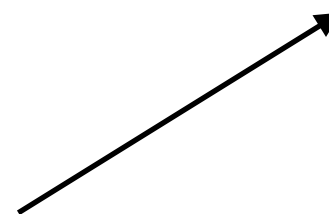
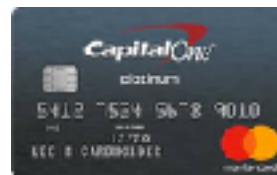
.00342



.00012



.00549



Life in data science

Key areas

Machine learning

Experimentation

Analytics

AB testing
Multi-armed bandits

Business applications

Forecasting
Predictive servicing
Risk management
Ops optimization

UI/UX
Model effectiveness

ROI
User engagement
Operational efficiency

Customer ID	Credit score	Delinquencies	...	Credit limit	APR	Default in 2 mths
1	675	0		5000	22.5	1
2	570	1		3000	26	0
3	720	0		4000	22	0
4	512	1		3000	27	0
...



“Customer A has 5% chance of defaulting within next 2 months”

Life in data science

Key areas

Machine learning

Experimentation

Analytics

AB testing
Multi-armed bandits

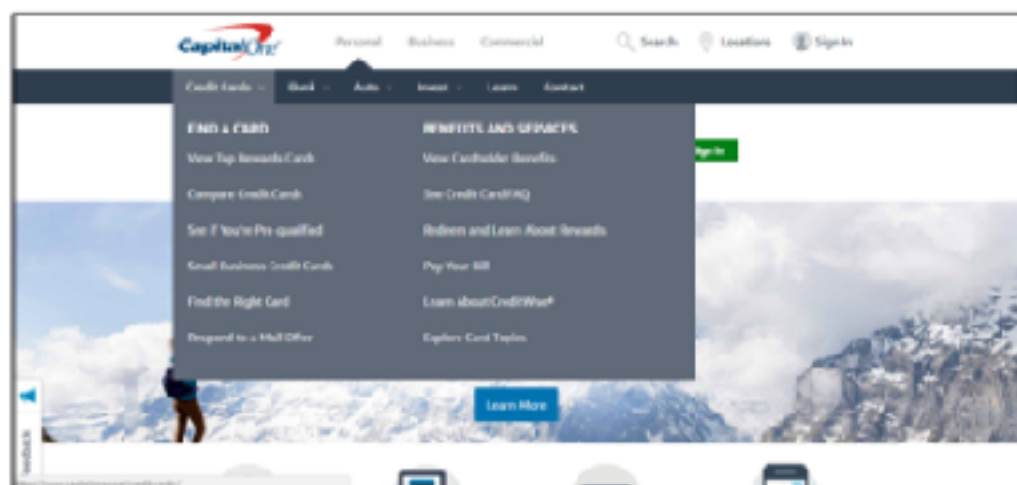
Business applications

Forecasting
Predictive servicing
Risk management
Ops optimization

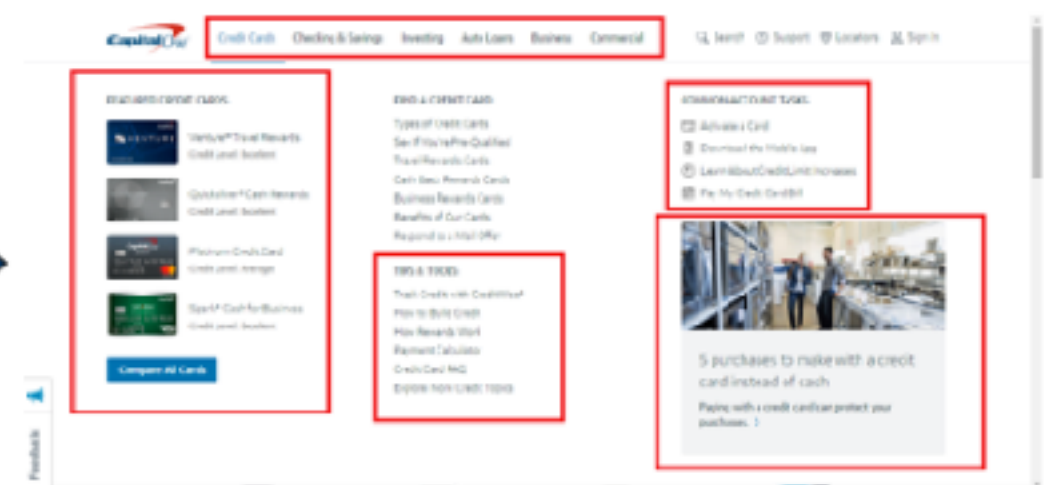
UI/UX
Model effectiveness

ROI
User engagement
Operational efficiency

Old Site Navigation, In use since 2015 Site Redesign



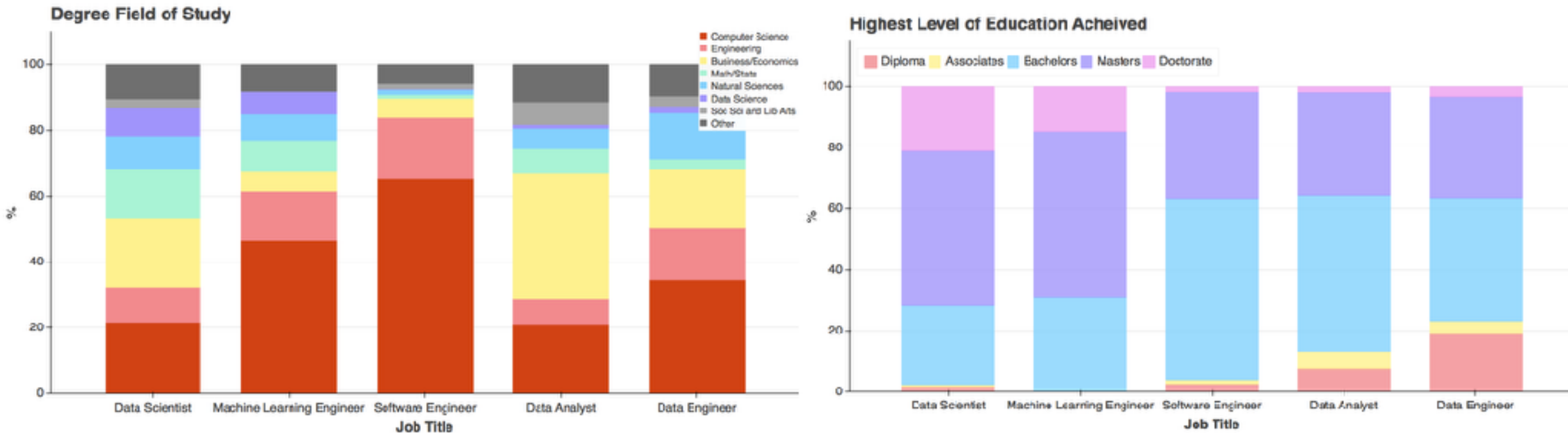
New Site Navigation Concept; A/B test launch 5/30



Physicists in data science

Data scientists come from diverse educational backgrounds, but are disproportionately represented by graduate degrees in math/natural sciences.

$$P(\text{math/science grad} \mid \text{DS}) \gg P(\text{math/science grad})$$



Your background will serve you well:

$$P(\text{DS} \mid \text{math/science grad}) \gg P(\text{DS})$$

Physicists in data science

Hard skills:

- Python
- SQL
- Stat 101
- Machine learning basics

- R
- Hadoop
- Spark
- Cloud computing

Soft skills:

- Communication
- Schmoozing
- Teamwork
- Problem-solving

Physicists in data science

Hard skills:	Soft skills:
<ul style="list-style-type: none">- Python- SQL- Stat 101- Machine learning basics	<ul style="list-style-type: none">- Communication- Schmoozing- Teamwork- Problem-solving
<ul style="list-style-type: none">- R- Hadoop- Spark- Cloud computing	

Python & SQL:

www.codecademy.com

www.hackerrank.com

leetcode.com

Machine learning:

www.coursera.org/learn/machine-learning

www.analyticsvidhya.com

towardsdatascience.com

Physicists in data science

Hard skills:	Soft skills:
<ul style="list-style-type: none">- Python- SQL- Stat 101- Machine learning basics	<ul style="list-style-type: none">- Communication- Schmoozing- Teamwork- Problem-solving
<ul style="list-style-type: none">- R- Hadoop- Spark- Cloud computing	

Ever done any:
Hypothesis testing?
Dimensionality reduction, e.g. PCA?
Clustering?
Signal processing?
Regression?
etc ...

Have you:
Worked in collaboration with others?
Given talks?
Independently defined and solved problems?
etc ...

How to get in the game

Resume:

- Short and sweet
- Translate your experience (previous slide) into DS terms
- Quantify the impact of your work
- Talk to Career Office and recruiters about resume tips

Talking to people:

- For career advice, quality > quantity, ask people with multiple years of experience in leading companies
- For opportunities, anything goes
- Do not ask people for a job, unless they are actively recruiting
- If seeking advice, explicitly ask for it, e.g. don't expect people to reach out just because you connected with them
- Expect to be ignored many times
- Maintain good relationships with recruiters
- Request mock interviews (important!!)

Signaling:

- Bootcamps
- Side projects
- Blogs
- Stat/ML courses

Applying to a role:

- Personalize your resume
- Cover letters are usually unnecessary
- Learn about the business
- Learn about how current employees feel (Glassdoor)
- Research typical compensation for the role and location (Comparably+Glassdoor+Google)