MAKING A CAREER IN DATA SCIENCE

Michael Becker
Senior Data Scientist @ Penn Medicine

#DataJawn

@beckerfuffle

So you want to be a Data Scientist?

- What's in a Job Title?
- Get Educated
- Data Science Skills
- The Portfolio
- The Job Search
- Finding a Good Employer
- Practice Makes Perfect
- Story Time!

bit.ly/SurveyJawn

TITLES

#DataJawn

@beckerfuffle

ALTERNATE ROLES

- Data Analyst
- Data Engineer
- Machine Learning Engineer

By Saad Andalib; CC BY 2.0, flickr.com/photos/saadandalib/8536426163

EDUCATION

#DataJawn

@beckerfuffle

#DataJawn

@beckerfuffle

#DataJawn

@beckerfuffle

EDU Wishes
Math
Machine Learning
Coding
Communication Skills
Visualization
Data Analysis
64%
38%
38%
32%
26%
15%
0.0
0.1
0.2
0.3
0.4
0.5
0.6

By I, Luc Viatour, CC BY-SA 3.0, commons.wikimedia.org/w/index.php?curid=4632541

SKILLS

#DataJawn

@beckerfuffle

Hours/Week Building Models
I don't create models
<1 hour per week
1-10 hours per week
10-20 hours per week
20-40 hours per week
4%
17%
64%
13%
2%
0
5
10
15
20
25
30
count

Models used
Linear/Logistic models
Tree-based models
Bayesian/Probabilistic models
Neural Nets/Deep learning
MCMC
89%
72%
35%
24%
11%
0.0
0.2
0.4
0.6
0.8

© Tom White, drib.net

PORTFOLIO

Set status

Michael Becker
mdbecker

PRO

Edit profile

Overview | Repositories 57 | Projects 0 | Stars 88 | Followers 96 | Following 56

Popular repositories

Customize your pins

word2vec_demo
word2vec demo for #hourofcode using gensim
JavaScript ★ 21 ⑂ 3

pydata_2013
PyData Boston 2013 talks: "Intro to scikit-learn" & "Realtime Predictive Analytics: Using scikit-learn and RabbitMQ"
★ 11 ⑂ 3

pycon_2014
Stuff for my Pycon talk "Realtime predictive analytics using scikit-learn & RabbitMQ"
https://us.pycon.org/2014/schedule/presentation/224/
★ 11 ⑂ 5

community_twitter_bot
Twitter bot that allows approved users to tweet from a shared account using direct messages
Python ★ 4

wikilanguages
Data set of popular wikipedia articles across 41 languages

pydatanyc2014
Materials for my pydata 2014 NYC talk

PUT EVERYTHING ON GITHUB

POST ON STACKOVERFLOW

PERSONAL PROJECTS

JOB SEARCH

Industry
Medical/Healthcare
Software
Consulting
Non-profit
Retail/E-commerce
Education
Telcom
Manufacturing/Industrial
Advertising/Marketing
32%
11%
11%
9%
9%
6%
6%
2%
2%
0
2
4
6
8
10
12
14
count

#DataJawn

@beckerfuffle

#DataJawn

@beckerfuffle

“if you apply for a technical job through Indeed, you’re very unlikely to get anywhere"

–Edouard Harris

"start with LinkedIn"

–Jeff Hale

"I look on LinkedIn and find if there is anyone in my network that can recommend me"

–Schaun Wheeler

"start embedding yourself in engineer-driven machine learning meetups"

–Edouard Harris

"If you meet 50% of the requirements, that might be enough. If you meet 70%, you're good to go. If you meet 100%, there's a good chance you're overqualified."

–Jeremie Harris

RED FLAGS

"oddly non-specific posting"

–Jeremie Harris

"a PhD is a decent proxy for skill, if you're not sure what actual skill looks like."

–Jeremie Harris

PRACTICE

kaggle
SOLD
House Prices: Advanced Regression Techniques
Predict sales prices and practice feature engineering, RFs, and gradient boosting

KAGGLE

PERSONAL PROJECTS

STORY TIME!

"**I got my PhD in Physics** in 2018, took ~9 months of networking and interviewing before getting two offers"

"I started out more or less as a **data analyst**, at a company that also had a large data science team. After about 8 months I was able to transition into an entry-level DS role on a team closely related to my analyst work"

"Started as a **Data Engineer**, did Science-y projects in my spare time, transitioned internally through an established program."

"I was an **epidemiologist** and got frustrated at how much data is collected by the government and how little bandwidth they give analysts to be creative with it so I moved into the **private industry** and figured out how to use my stats background to do that work."

“started as a **data analyst** right after undergrad. It wasn't fulfilling and my career prospects were limited so I **went back to school** to get my master's in information and data science”

"I always considered myself an **applied social scientist**, and I've been doing data science in some form for 25 years. I **call myself a data scientist now** because people seem to understand what that means a little better."

bit.ly/SurveyJawn
github.com/mdbecker/datajawn2019

WE'RE HIRING!

bit.ly/UPHSData

bit.ly/UPHSEngineer