Road to a Data Science Career

@_jameslamb

/jameslamb

Part 1: Choosing a Job

- The title "data scientist" is not that important
- The job market is much friendlier now than it was in March / April
- Consider remote-only jobs
- Apply if you're 50% qualified

Part 2: Getting to an interview

- Put a personal statement on your resume
- Connect your experience to business value
- You have to prove that you can code
- You'll have to do a take-home case study

Part 3: Getting the job

- Get comfortable saying "I don't know"
- Be prepared for situational questions
- You probably won't need to do leetcode stuff
- Say "I", not "we"

Part 1

Choosing a Job

The title "data scientist" is not that important

data
scientist

data
analyst

research
scientist

machine learning
engineer

economist

statistician

data
journalist

business
intelligence
specialist

researcher

sales
engineer

analytics
consultant

actuary

data
engineer

quant

business
analyst

biostatistican

solutions
architect

The job market is much friendlier now than it was in March / April

Consider remote-only jobs

Apply if you're at least 50% qualified:
data science job descriptions are terrible

Position Qualifications

- Experience producing high quality, well-regarded insights & analytics work
- A higher degree in statistics, econometrics, machine-learning, physics, chemistry, applied mathematics or other related quantitative field is preferred
- Masters with 1+ years or Bachlors with 4+ years experience
- Strong coding skills, preferable python or other similar language
- Experience with relational databases / SQL
- Understanding AWS S3, HDFS, Hbase, or other non-relational DB
- Experience in working with un-cleaned data
- Modeling skills required, from machine learning to more ad-hoc methods
- Experience in Time Series and Forecasting

this is from a posting for "Data Scientist II", from ConAgra Foods

Part 2

Getting to An Interview

Put a personal statement on your resume

OVERVIEW

Data science professional with 5+ years of experience

Strong experience and interest in timeseries, productionizing machine learning models, and using software engineering principles to improve data science productivity

PROFESSIONAL EXPERIENCE

Amazon Web Services
IoT Data Scientist, August 2019-Present
- Delivered consulting projects involving moving industrial IoT data from edge to cloud and building applications and machine learning models on that data

Uptake

TECHNOLOGIES

R

- package creation, distribution
- rmarkdown, knitr
- data.table, httr, jsonlite, DBI
- caret, LightGBM, XGBoost
- rgeos, sp, sf, rgdal

Python

- package creation, distribution
- json, requests, pandas, psycopg2
- numpy, scikit-learn
- click, Flask, jinja2, dill

Connect your experience to business value

BAD

"Trained a CatBoost binary classification model with 98% recall"

"Wrote and managed a data pipeline using Python and Airflow, which ran every hour and processed 10 GB of data"

GOOD

"Created a model to detect pricing reports that needed to be reviewed by commodity analysts. Saved analysts 10,000 unnecessary reviews per quarter. This binary classification model was created in Python, using CatBoost."

"Wrote and managed a data pipeline that picked up new reviews posted on a travel site and filtered out profanity. Used Python and Airflow to allow this pipeline to run every hour and process 10 GB of data, reducing the average time to publish reviews by 5 hours"

You have to prove that you can code

How To Create A GitHub Profile README

Monica Powell

https://www.aboutmonica.com/blog/how-to-create-a-github-profile-readme

You'll have to do a take-home case study

Part 3

Getting the Job

Get comfortable saying "I don't know"

Problem 5 [*N* = 77]: Which of the following options do you prefer?

A. a sure win of $30 [78 percent]
B. 80% chance to win $45 [22 percent]

THINKING, FAST AND SLOW

DANIEL KAHNEMAN

WINNER OF THE NOBEL PRIZE IN ECONOMICS
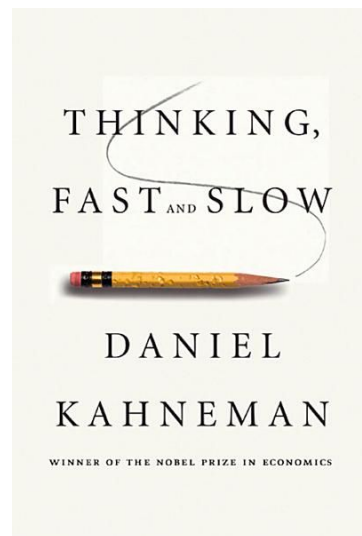
Kahneman & Taversky (1981), "The Framing of Decisions and the Psychology of Choice"

Be prepared for situational questions

Tell me about a time you….

...had to persuade someone else to change their minds

...failed to meet a deadline

...delivered an important project

You probably will not need to know stuff like algorithms, algorithmic complexity

Say "I", not "we"
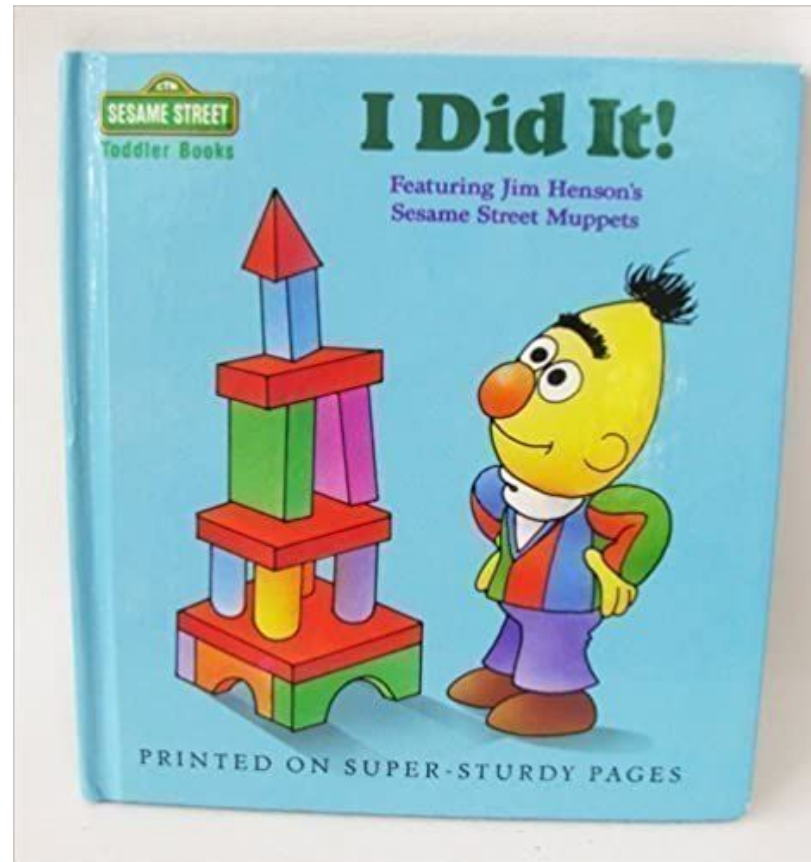
Thanks for your time!

Questions?

As a data scientist, there are many new skills to learn everyday (different software, coding languages and new data analytical methods, etc).

What's the key to learning new skills quickly and how to keep oneself in the right pace of learning?

For resumes and interviews, is it more important to make myself 'professional' (eg. know a lot of theories and have a lot of skills) or 'different' (eg. participate in unusual projects or have interdisciplinary knowledge), or these two are virtually the same?

How will the pandemic impact job search for graduates, especially international students?

And how should we adjust ourselves to it?

What type of data science personal projects have stuck out to you on resumes?

How can I land in a data scientist job If I don't have data scientist intern experience

It seems hard to get a data scientist interview when I only have course project experience.

I think I have enough skills such as programming, probability, etc. But I do not have any experience to do some jobs related data science.

How can I start my first step towards data science?

As students seeking for some new grad data science positions, do we need to master knowledge in computer science algorithms, such as dynamic programming, searching and sorting, etc.?

And need we be familiar with some data structures such as stack, queue?

I've participated in data science competitions where I didn't need to know these things, but I have been asked these things in some interviews.

I'm looking for a career in theoretical math but might fall back to working as a data scientist in industry.

What activities I need to do/courses I need to take/things I need to learn/capacity I need to develop for a job in the industry?