Tips and Perspective on Starting Your Data Science Career

JUSTIN CLIFFORD SMITH, PHD

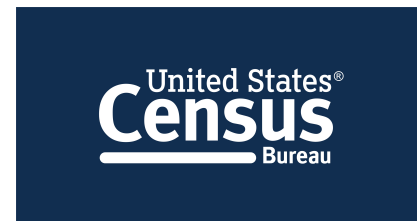SENIOR DATA SCIENTIST, OPTUM

Who am I?

PhD in theoretical condensed matter physics from UC Irvine.

One year as a postdoc at UCLA where I spent much of my free time learning data science skills.

Then, 15 months at the US Census Bureau as a data scientist in the Center of Optimization and Data Science working on NLP and ML for economics surveys.

One year at Optum as a Senior Data Scientist in Risk and Quality Decision Intelligence using NLP and ML to make various healthcare programs more efficient.

For fun, I'm a competitive épée fencer.

Optum is a leading healthcare company

UnitedHealth Group is a Fortune 7 company and the leading healthcare and health insurance provider.

Optum is a division of UnitedHealth Group founded in 2011 and is composed of many parts such as:

- OptumRx
- OptumCare
- OptumInsight (where I work!)

OptumInsight is the technology arm of Optum with a focus on using data science to improve healthcare services.

Data science is comprised of many facets

Data science roles can demand many skills:

- Data acquisition
- Data cleaning
- Data visualization
- Model building
- Software engineering
- Research and experimentation
- Subject matter expertise (SME)/business knowledge
- Verbal and written communication

Which facets you pursue indicates the job you should pursue

The variety of skills leads to a proliferation of job titles and roles:
- Data analyst
- Data scientist
- Data engineer
- Data visualization engineer
- Machine learning engineer

No job is better than another. They are all different and important.

Approaches to learning data science – take initiative

Take online courses.
- Fundamentals such as Andrew Ng's Machine Learning Course or the Deep Learning Specialization
  - Check for up to date reviews!
  - Take notes, don't just stare at the videos.

Participate in Kaggle/competitions.
- Great for developing model building abilities.
- Read through older competitions to learn foundation skills that are likely to appear on the job.

Volunteer with data driven organizations.

Work on personal projects/build a portfolio.

Take initiative! The beauty of DS is many skills can be acquired on the side.

What does data science look like in practice? Part 1

Many skills can be acquired on your own, but some don't fully develop until in a job.

Data acquisition can be laborious process. You may:
◦ Query APIs.
◦ Hunt down stakeholders across the organization.
◦ Even make it yourself.

Data cleaning and EDA is a crucial first step.
◦ Visualization and aggregate statistics are your friend.
◦ Make more plots than you feel are reasonable.
◦ Build intuition.
◦ Date cleaning can feel like grunt work, but its not.
  It builds biases and assumptions into your data.

Credit: H Alberto Gongora

What does data science look like in practice? Part 2

Compute environments come in all shapes and sizes.

- Workstations
- Cloud
- HPC (CPU, GPU, Spark)

Productionizing models depends on environment and business need.

- Containerization and reproducibility
- Batch inference, streaming inference, etc.

Software best practices will make your life easier.

- Version control
- Modularization
- Unit testing

Cultivate a growth mindset.

A snapshot of data science in practice – compute resources

Where you work can have a strong influence on what technologies are available to you.

In academia, you tend to have complete freedom over your personal workstation and a fair amount of autonomy on shared compute resources.

In government, the opposite is true in government where software provisioning can be strict.

In industry, this varies often by company size and tends to fall between the extremes.

💡 Be flexible and adapt to changing technologies.

💡 Master transferable tools and techniques such as software best practices, Unix, and Vim

A snapshot of data science in practice – data access and security

How data is accessed and used will often depend on your use case and sector.

Be mindful of data licenses! Some allow commercial use, others do not.

Data security is not often touched upon but is crucially important.

Security varies such as governmental data being protected by the US Code, or healthcare data being protected by HIPPA.

Remember that data relates back to people, and we are entrusted to both keep it safe and be ethical.

Credit: Margaret Hagan

A snapshot of data science in practice – teamwork and stakeholders

How you work changes by workplace, but also can follow broad trends:

- Academia tends to be independent.

- Government and industry tend to be more meeting driven.

- Both rely on collaboration to often produce final outputs.

The stakeholders/work output of your work changes drastically.

- Academia and R&D settings focus on papers and conferences.

- Industry and government focus on programs and products that make impact and/or drive revenue.

Find work environments that fit your working style and desired impact.

Take initiative.

Cultivate a growth mindset.

Be flexible and adapt to changing technology.

Master transferable tools and techniques.

Find work environments that fit your working style and desired impact.

Connect with me on LinkedIn!

https://www.linkedin.com/in/justincliffordsmith/

OPTUM

www.optum.com

https://careers.unitedhealthgroup.com