Landing your First ML Role

QUEENSLAND AI HUB

QUEENSLAND AI

Types of Data/ML Roles

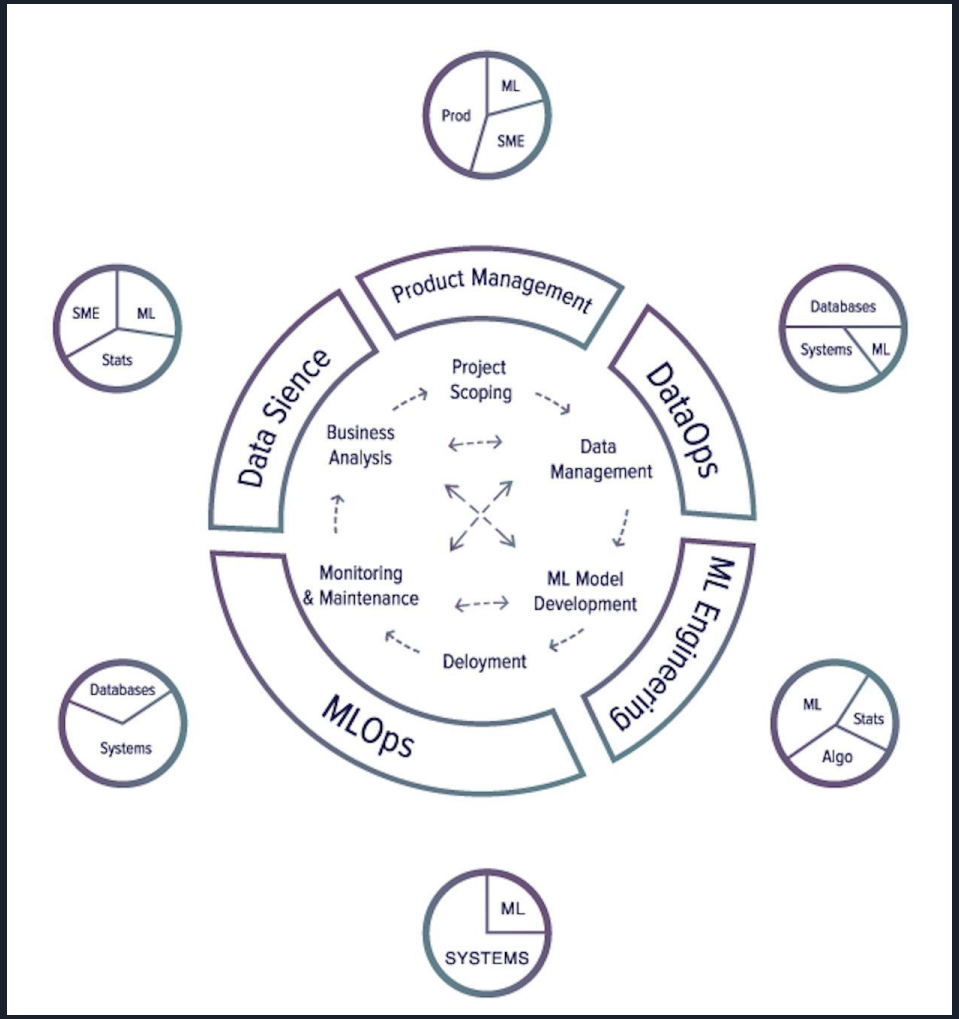
Data Analyst - Focussed on reporting & analysis of data. Often involves lots of SQL & BI tools (Tableau, PowerBI etc) and stats.

Data Engineer - Focussed on data/ETL pipelines, data platforms/warehouses/lakes.

ML Engineer - Focussed on model building & deployment, MLOps becoming an important skill

Data Scientist - Usually more about data analysis and model building, less on deployment/engineering. DS Title can mean anything.

Takeaway: Data analyst & data science roles can be good path to full-time ML engineering.



Source: [Chip Huyen]

Market Trends

- **Data Science roles are more common over past 10 years**
  - But title can mean anything
  - Many older companies have rebranded their analytics teams to data science teams, but doing the same old analytics, BI reporting, and basic modelling work.
- **Dedicated ML Engineer roles surging in popularity in past 2 years**
  - More companies ready to invest in building & deploying modern models
- **There are around 2x as many Data Engineering roles as DS + ML Engineering**
  - Comfort working with data pipelines.
  - Software engineering background a strong foundation.
  - Good path to jump to ML engineer - transition towards MLOps.

Types of Companies/Work

Consulting (Big 4, Boutique) - Wider variety, fast-paced. Great for early career, background in business & project work is highly valuable.

Research (Institutes, Research Orgs) - Narrow focus, heavy R&D, need to build strong skills/knowledge in your chosen field. Unis will require qualifications.

'Older' Businesses (e.g. Banks, Large Retail) - Majority of companies will be mostly analytics/analysis work, some will be ML engineering. Less likely to be cutting edge.

Internet/Data Businesses (Tech companies, SaaS) - Modern tech stacks, more agile/less organised, more freedom. Suits self-learners & motivated individuals.

Takeaway: Understand the exact projects you will be working on, focus on what you will learn in your first role.

ML Engineer Skills

Skill	Time Spent as Junior	Time Spent as Senior
Data Science	High	High
Model Building	High	Medium
Engineering (Python)	Medium	High
Deployment & Production	Low	High
Cloud	Low	Medium
Professional/Communication	High	High

Data Science Skills

Learn:

- Working with Data (object storage, databases), some SQL, Pandas
- Visualisation of data (Matplotlib, Plotly, BI tools)
- Statistical analysis (correlations, trends)
- Analysis of a problem (research, understanding a problem, communicating findings)
- Classical model building (simple models, PCA, linear regressions)

Data Science skills form the basis of everything else you will do. Learn enough of these points to know how they work, and continually sharpen your skills as required.

Some courses: General Data Science + ML, PostgreSQL, Pandas User Guide

Model Building Skills

Learn:

- Clustering/Unsupervised (K-means, PCA, tSNE, UMAP)
- Classical ML (Regression, Trees, SVMs)
- Deep Learning (Pick one framework, ideally PyTorch)
- All Core Domains (Tabular, Time-series, NLP, Computer Vision)
- Very important: Train/test set splitting, Testing & Validation, Metrics

Most deployed real-world models are classical models. Spend just as much time on these as you do deep learning.

Some courses: Hands-On ML (classical and DL), Fast.ai, DeepLearning.ai, ITSL, Scikit-learn User Guide

Software Engineering Skills

Learn:

- Git, collaboration, how CI/CD works
- Unix shell, command line, editors
- Defensive programming/validation, design patterns
- Debugging, logging & monitoring
- Testing, code quality

SE skills are the hardest to learn online and take the longest to master. Best way to learn is to join a high quality team with skilled engineers.

Some courses: Missing CS Semester, Pragmatic series, Fluent Python, Effective Python, PEP8, Google Python Style Guide, Python Design Patterns, Python Architecture Patterns

Deployment/Production Skills

Learn:

- Containerising Applications/Models
- Building APIs & batch jobs
- Model monitoring & retraining pipelines
- Reproducibility, Versioning, Artifact storage
- MLOps & managing workloads

ML Engineers tend to be focussed on deploying and managing models. Focus time on cloud & engineering if this is the direction you want to go.

Some courses: Deeplearning.ai MLOps course, MLOps guide

Cloud Skills

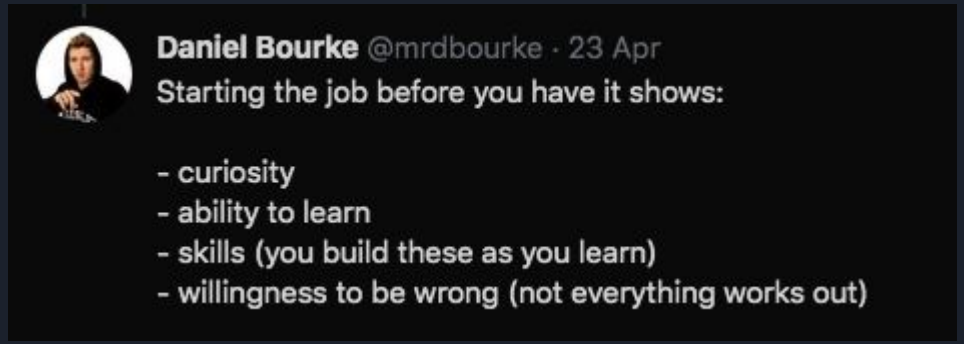
Learn:

- Cloud object storage, VMs, & basic security & IAM at a minimum
- General cloud benefits (scalability, redundancy, managed services)
- Cloud Machine Learning tools (e.g. Sagemaker platform, GCP Vertex ecosystem)
- A cloud certification is a great introduction/box-ticker.

Learn about the cloud and how to deploy things into it, especially if you are more interested in ML Engineers/MLOps.

Some certifications: AWS Machine Learning, AWS Developer Associate, AWS Solution Architect Associate, A Cloud Guru Courses

Start the job before you have it



Source: https://www.mrdbourke.com/how-can-a-beginner-data-scientist-like-me-gain-experience/

What employers look for in juniors

Companies expect juniors to cost more in resources than they produce for 3-6 months.

- De-risk yourself
  - Minimise the number of skills they will have to teach you
- Show that you can solve problems on your own, and will not require hand-holding. Show that you're hungry to learn.
  - Personal projects are great for this
- Show that you will be productive in weeks or days after joining the company
  - The best way is to show that you've done it before.

Strategy of landing your first role

1. Decide on what you want to work on
2. Research the current market
3. Create a list of skills to work on and develop
4. Build a portfolio to show off your capabilities
5. Share your work and network with others
6. Focus on learning, not salary
7. Always be learning & advocate for yourself
8. Give back to others

Strategy of landing your first role

1. Decide on the type of work and role that interests you most
    a. Would you prefer consulting, product-focussed, or research?
    b. Are you more interested in data science-y roles or engineering roles?
    c. What compliments your existing strengths?
    d. Where will you learn the most?
    e. What role right now will get you to where you want to go in the long-term?

Strategy of landing your first role

2. Research the current market
    a. What roles are available for what you want to do?
    b. What types of companies and problems are they typically looking for?
    c. Where are the companies located (are there more in Sydney/Melbourne?)

Strategy of landing your first role

3. Build a list of skills to work on and develop
    a. What skills do the roles you want typically ask for?
    b. What existing skills do you have that most competitors will not have? How can you leverage these for your benefit?
    c. What types of problems do these companies typically work on?
    d. Is there any other knowledge or skills you can develop to compliment your ML skills?
    e. How can you work these skills into your projects in order to develop them?

ML Interviews book

Strategy of landing your first role

4. Build a portfolio to show off your capabilities
    a. Start working on projects to hone your craft and showcase your talent
    b. Focus on projects that are interesting to you - you do your best work when you're most engaged
    c. Use a mix of project types to build and showcase different skills
        i. An analysis notebook/blog post on an interesting dataset
        ii. Model building on Kaggle competitions
        iii. A small web service with a front-end hosted in the cloud

Strategy of landing your first role

5. Share your work and network with others
    a. Blog about your work, sharing it with others. This will develop your communication and story-telling skills. Share your code publicly on github and your notebooks on Kaggle
    b. Share your achievements and the content you create in your networks (LinkedIn, reddit, TikTok, Twitter)
    c. Recruiters/HR may not understand the self-taught route. Reach out directly to the head of ML/data science and you'll have a much better chance.
    d. Consider applying for internships to build your skills and get to know an employer
    e. Join communities where you will meet other ML engineers or prospective employers. E.g. Queensland AI Meetup, Brisbane Data Science Meetup, QUT BANDS

Strategy of landing your first role

6. Focus on learning, not salary
    a. Your first role should be focussed on learning as much as you can
    b. Don’t worry about salary for the next 6 months, think about maximising it in 2-3 years time.
    c. Focus on roles with
        i. Existing teams and mentors (not where you’ll be operating alone)
        ii. Strong engineering culture
        iii. Time for training and development
        iv. Growth opportunities and challenging problems

Strategy of landing your first role

7. Always be learning & advocate for yourself
    a. This field requires constant learning to stay current. This is why it pays so well.
    b. Bring learning into your day-to-day work. Always be improving your work, don't cut corners.
    c. Communication is more and more important as you grow. Learn to communicate your work well, focus on why it is valuable and important to the relevant stakeholders.
    d. Ask for training, find a mentor, and be willing to present why you deserve your career progression.

Strategy of landing your first role

8. Give back to others
    a. The field is growing, many people need help
    b. Teaching and mentoring is
        i. A great way to solidify your knowledge
        ii. A key skill as you become more senior
        iii. Fun. You'll meet other great and brilliant people who will teach you new things
    c. Contributing to the community will help build your profile