How to successfully transition

From Physics to Data Science
Resources

• **Insight Data Science**
  • Past fellows
  • Blog: preparing for Insight (full of resources)

• **Portfolio:**
  • Github: personal projects, clean research code
  • LinkedIn: make a profile, add us to your network
  • Resume: ask us to share ours – Resume ≠ CV
  • Technical blog or other relevant online activity

• **Data Science News:**
  • Data Science Weekly, Data Science Central
  • Data Tau
  • Follow Data Science on LinkedIn, Medium, Quora
Resources

• **Python:**
  • If you’ve never done any python before: [Codecademy](https://www.codecademy.com)
  • Intermediate: [Google’s python class](https://www.google.com)
  • Book: [Data Science From Scratch](https://www.datafromscratch.org)
  • Pandas: [Python For Data Analysis](https://www.python-for-data-analysis.org)
  • Google, StackOverflow, StackOverflow, StackOverflow

• **R:**
  • [Data Science Specialization on Coursera](https://www.coursera.org)
  • Google, StackOverflow, StackOverflow, StackOverflow

• Forget about Matlab, Mathematica, ROOT, Excel, etc.
Resources

• SQL:
  • SQLZOO
  • Mode Analytics SQL tutorial
  • SAMS Teach Yourself: SQL in 10 Minutes

• Algorithms and Data Structures:
  • Python for Data Structures, Algorithms, and Interviews! On Udemy
  • HackerRank, Leetcode: coding challenges
  • Pramp: practice coding interviews with other candidates
  • Kaggle: practice ML on real-world data (data cleaning/analysis with R, pandas)
  • Practice whiteboarding with other people
Resources

• Machine Learning:
  • Andrew Ng’s machine learning class
  • Data Science Specialization on Coursera
  • Elements of Statistical Learning - textbook
• Topics to know:
  • Linear and logistic regression
  • Tree ensemble models: random forests, boosted decision trees
  • Bias-Variance trade-off
  • Common ML issues: overfitting, class imbalance, sparsity, collinearity
  • How to address them: regularization, dimensionality reduction, sampling, cross-validation
  • Performance metrics: precision/recall, ROC/AUC, accuracy
  • Purpose of algorithm: prediction or interpretation?
Resources

• Deep Learning / AI:
  • **Tensorflow Neural Networks Playground**
  • Michael Nielsen’s [deep learning online book](#)
  • Convolutional Neural Networks, Recurrent Neural Networks

• Statistics / Probability:
  • A/B Testing
  • P-values, confidence intervals
  • Statistical Tests (e.g. t-test, f-test, ANOVA, Chi² test, non-parametric tests)
  • Power Analysis
  • Dice and Coin probability questions (Glassdoor interview reviews)
  • Bayes’ Theorem
  • Multiple Comparisons (Bonferroni correction)
• **Business / Product Sense:**
  • What is the company’s data like? Where did it come from?
  • Understand who they serve (e.g. internal clients, B2B, B2C)
  • Familiarize yourself with the company’s values + how they make $
  • Translate business questions into data problems
  • Browse company’s technical blogs

• **Miscellaneous Tips:**
  • The recruiter is your friend – exhibit interest + ask all questions
  • Prepare questions to ask at the end of the interview:
    • What’s a problem that’s really interesting at the company you think hasn’t been looked into enough?