Guide To Creating Kick-Ass Data Science & ML Portfolio Projects

Follow These Tips To Create Projects That Get Recruiters Hooked & Also Ace Your Take-Home Data-Science Interview Projects

Written By Nick Singh, Noah Yonack, Ryan Squire, & Anonymous

Chapter 1

Introduction

First off - this guide is NOT a Data Science 101 guide. This guide won't help you answer technical interview questions like "What is k-folds cross-validation?" or "When should you use linear regression over logistic regression?"

This guide aims to give high-level tips on how to craft a data-science or machine learning portfolio project that gets recruiters HOOKED. We're helping you position, market, & sell your data science project.

While this guide is primarily focused on portfolio projects intended to get you interviews, these same tips can also be applied to take-home data science projects that companies assign. You might be wondering - *Why are we even qualified to write this guide?*

We've been in the data & software engineering industry for a while now, having worked at a top Silicon Valley tech companies, VC-backed startups, and Wall Street firms in data-related roles. Three of us have also worked together at SafeGraph - a company 100% focused on data. Here, we've assigned and evaluated take-home data science projects.

Quick Background About The Authors

Nick Singh - Previously a Software Engineer at Facebook. Did a Data Engineering internship at Google's Nest Labs. Currently doing Growth & Marketing at SafeGraph.

Noah Yonack - Previously did Data Science & Software Engineering internships at Bridgewater Associates & LinkedIn. Currently doing Machine Learning & Data Science at SafeGraph.

Ryan Squire - Worked as a Data Scientist at Lumosity and a YC-backed startup. Stanford Neuroscience PhD dropout. Currently doing Product & Data Science at SafeGraph.

Anonymous Friend - Works as a quant at a top hedge fund in NYC. No other details to protect privacy :)

Chapter 1
The High-Level Philosophy

Our goal is to get you interviews for data jobs. So we need to impress the interview gatekeeper - the recruiter.

Truth is - a recruiter won't dive deep into your Jupyter Notebook, look at line #94, see the clever model you choose, and then offer you an interview. That's not how recruiting (or people!) work.

They'll just read the project description for 15 seconds in your cold email to them or when reviewing your resume. Maybe - if you're lucky - they click a link to look at a graphic or demo of the project. At this point, usually in under 30 seconds, they think to themselves - "hey this is neat and relevant!" and decide to give you an interview.

Thus, we're optimizing our data science projects to be sexy, exciting, and impressive to the decision maker in this process - the recruiter.

We're optimizing for projects that are easily explainable via email (because cold emails work really really well). We're optimizing for ideas that are 'Tweetable' - those that can be simply conveyed in 140 characters or less.

You can get more details about this concept by checking out the 36 Software Engineering Resume Tips Guide and reading all the tips related to "Principle #2: Build your Resume to Impress the Recruiter". Just replace the word "Resume" with "Data Science Project" - most of those tips directly carry over.

Overall, once you internalize this concept, that you're trying to impress a recruiter, you'll understand why the bulk of our advice is more high-level and top of the funnel. More marketing & sales oriented. We save some of the more specific technical tips towards the end of the guide.

Chapter 2
High-Level Tips

Make Your Project About You're Passion

Passion is contagious. When you work on something you're passionate about, it becomes much easier to talk and pitch the project on a phone call or in an interview. Your application as a whole becomes more memorable.

For example, Nick loves hip hop music & DJ'ing. That's why in college, Nick worked on RapStock.io - fantasy football but for hip hop artists. When talking about the project to recruiters, it was really easy for him to come across as passionate about data science and pricing algorithms because really the underlying passion for hip hop was shining through.

Another thing - if you're passionate about the work you are doing - it's less of a chore to get a project done. Work becomes play! And getting the project done really really matters... more on that later.

Work With Interesting Data

Working with stock ticker data or Twitter data? Boring!

Don't work on the same datasets everyone has worked with. Stand out. Be different. Don't work with datasets that people have worked with in their school work, such as the classic Iris Plant dataset or the MNIST Digit Recognition dataset.

Try scraping your own data, if you have the programming chops. Packages like BeautifulSoup and Scrapy in Python can help or Rvest for R users.

Quick Plug: You can also use SafeGraph data. Our foot-traffic dataset is derived from anonymized mobile location data. It covers foot-traffic to 5 million businesses in the U.S.

Work With Interesting Data (Continued)

Hedge funds use SafeGraph data to trade stocks. Retailers use it to determine the best location to open up a new chain store. Real estate companies use it to buy/sell commercial real estate.

People have used this type of data to understand the effect of E. Coli outbreaks on Chipotle restaurant foot-traffic, and how that correlates with quarterly earnings reports. Noah used the SafeGraph data to find food desserts in Brooklyn.

We can guarantee you that not many student projects have been done with the data since it's for enterprises and very expensive.

But you're in luck - download a free sample of foot-traffic insights for 5,000 businesses ($500 worth of free data) by using the coupon code **KickAssDataScienceProjects** on shop. safegraph.com before **9/30** (no credit card required).

Tell An Interesting Story

Your success as a data scientist is contingent on both your technical data science ability but also your ability to communicate. Try your best to tell an interesting story. This doesn't mean write a ton of text. It just means to have the elements and structure of a story. Make sure there is an introduction (exploratory data analysis), a hypothesis you pose (conflict), some build-up to an answer, and then a cool result (resolution).

One way of telling a compelling story is making sure your project has a quality visualizations. Bonus points if it's a cool GIF or chart that can be embedded in your cold-emails.

Done > Perfect. And Prove You Are Done.

As long as your work is reasonably correct, the actual technical details don't matter for getting an interview. Again, as mentioned, a recruiter is not going to dig into your project and notice that you didn't remove some outliers in the data. But a recruiter is able to determine how complete a project is! So make sure you go the extra mile in 'wrapping up a project'.

See if you can 'productionize' the project. Turn the data science analysis into a product. For example, if your project was training a classifier to predict age from a picture of a face, go the extra step and stand up a web-app that allows anyone to input a photo and predict their age using Computer Vision.

If you did an exploratory data analysis of neighborhood income vs. neighborhood homeless rates, try to make and host an interactive map visualization so that folks can explore and visualize the data for themselves. I like D3.js for visualizations. rCharts is also pretty cool for R users.

Make sure you put your code on Github, or use Google Collab to host & share the Jupyter notebook publicly. Even if no one sees the code (which is likely!), just having a link to it sends a signal that you are proud enough with your work to publish it openly. It also shows that you actually did what you said your resume did, and you just didn't make up something to pad the resume.

Demonstrate Business Value

Demonstrate business value with your data science project. Try to make it concrete, over theoretical. This is crucial for PhD folks who are breaking into industry, especially if the academic is trying to break into smaller companies or startups. Ultimately, you are being hired at a business, so talk in business terms. Show how your technical skills can drive business value.

Demonstrate Business Value (Continued)

It's okay to skip out on demonstrating business value IF you work with interesting enough data and can tell a good story. An example project we find interesting, creative, but is technically simple and not obviously a driver of business value:
A Highly Scientific Analysis of Chinese Restaurant Names.

And, in DJ Khaled voice, ANOTHA ONE
Analyzing Rap Lyrics To Find The Most Popular Fashion Brands.

Specific Technical Tips

An assortment of specific technical tips. These tips are also very relevant to take-home projects assigned by companies.

• Have a process. Approach the problem in a structured way. A useful framework for this is Joe Blitztsein's (Harvard Stat Professor) 7-step framework, which is described in detail by Ryan on Quora
• Make sure you explore the data, before jumping into it and using it in models. Look for outliers and bugs. Disclose these. It's okay to ignore the outliers/bugs, but good to call it out and say that it's not in scope and you will ignore it. So much of real data science is filtering out bad data.
• You don't need to use the most complicated models. Simple models are okay. Students, in particular, think using very complicated techniques will help them when in reality can make your analysis confusing if not warranted.
• Try multiple models. But state the trade-offs and hypothesis for why you are trying the models, based on the size, shape, and type of data.
• Have humility - okay to show limits of your approach.
• If you're solving a problem others have solved before, talk about why your approach is different or better. Cite other people's work.
• Its' good to talk about future work, and things you would have done with more time & resources.

Chapter 3

Conclusion

Go forth and craft that perfect data science project. Then cold email recruiters and decision makers at your dream companies and watch the interviews roll in! Don't forget to email me your success story if this strategy helped you land an interview or job!

Inspired And Want Free SafeGraph Data?

Use SafeGraph data in your data science projects! Just use coupon code **KickAssDataScienceProjects** on shop.safegraph.com before **9/30** to get $500 worth of free data (no credit-card required).

Hedge funds, retailers, real estate companies, & management consulting firms use SafeGraph's foot-traffic dataset (derived from anonymized mobile location data) to trade stocks, plan where to buy/sell commercial real estate, and determine the best location to open up a new chain store.

Join Nick Singh's Newsletter For More Guides

I send out a newsletter ~once a month to 16,200+ subscribers in 37 countries.
The newsletter covers topics such as Software Engineering Career Advice, Data Science, Growth Engineering, Tech Strategy, and Startups. Click here to join the newsletter!

Some of the most popular guides previously sent to the newsletter:
7 Cold-Emailing Tips To Land Your Next Job, 36 Tips on Software Engineering Resumes
Win Hackathons: A How-To Guide , 2019 Events For Summer Bay Area Interns Guide

Let's Connect!

Connect with me on LinkedIn and on Twitter and email me feedback/ praise / hate-mail!