How to Build an Effective Data Science Portfolio in 2021

A comprehensive guide to creating a stellar Data Science portfolio

Harshit Tyagi [Follow]
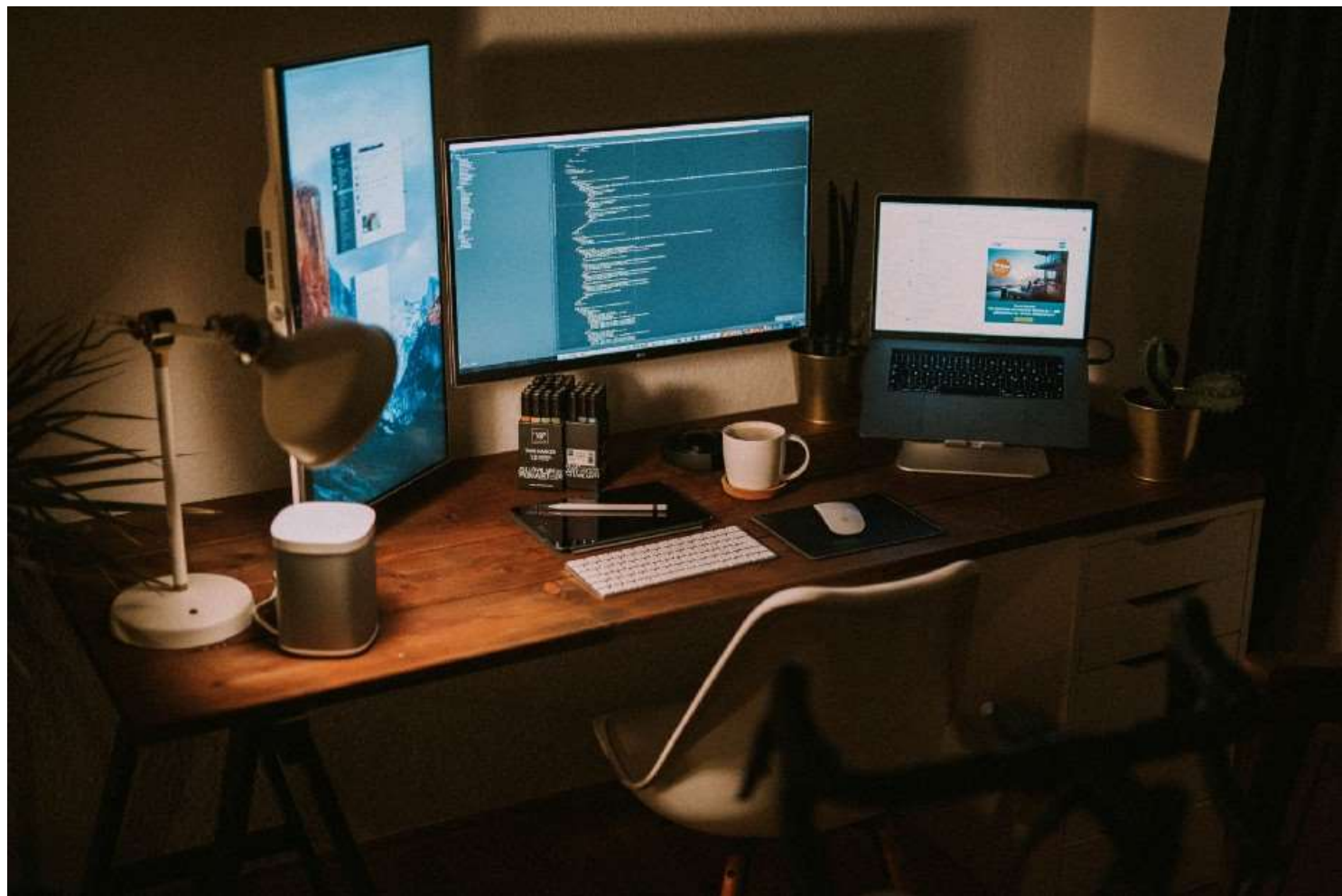
Jan 9 · 11 min read ★

Photo by Jannis Brandt on Unsplash

If the **Skills** section on your Resume states **Python**, **R**, **SQL**, **Machine Learning**, **Deep Learning** and you're wondering why you get **REJECTED** every time, you should keep

reading.

There are millions seeking a job in Data Science and the opportunities are limited. So, the important question is how would you stand apart from the pack?

The blog tries to capture everything you need to build a kickass portfolio — so good that they can't ignore you!

> *Wait, but why do I need a portfolio in the first place?*

Why should I build a Portfolio?

For someone who has received a Master's degree or a Ph.D. from a top tier college, getting a job might not be that difficult. The institute adds credibility to your profile which employers look for.

For someone who doesn't have a relevant degree or enough experience, that credibility needs to be established via a stellar portfolio showcasing your potential. The portfolio then works as evidence of your competencies.

There are numerous factors that can enhance your chances of getting noticed by the employer. With a smart strategy and consistent efforts, you'll be able to crack it.

Let's build a fool-proof plan right here to work towards landing a job!
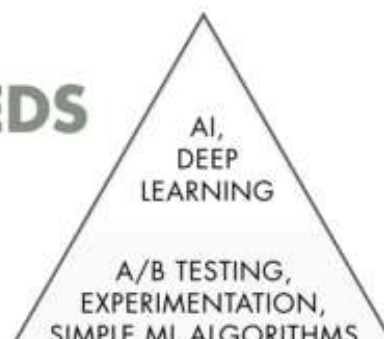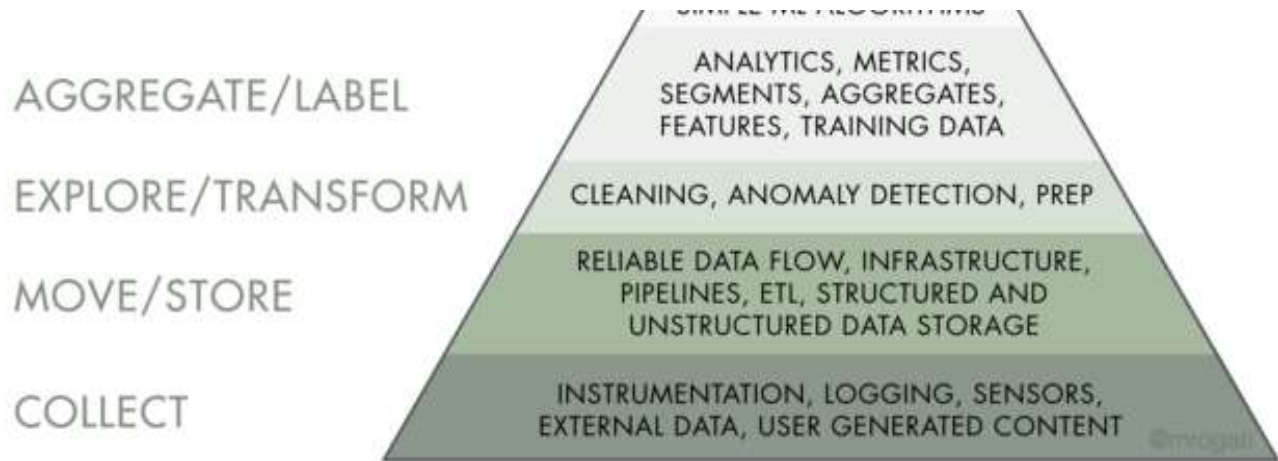
Step 1 — Identify YourSelf

Hopping from one career portal to another and applying for any job that mentions "Data" isn't a smart move. It would add to your stress and workload only to learn that they have rejected you.

Narrow Down Your Search

THE DATA SCIENCE
HIERARCHY OF NEEDS

AI,
DEEP
LEARNING

A/B TESTING,
EXPERIMENTATION,
SIMPLE ML ALGORITHMS

LEARN/OPTIMIZE

AGGREGATE/LABEL	ANALYTICS, METRICS, SEGMENTS, AGGREGATES, FEATURES, TRAINING DATA
EXPLORE/TRANSFORM	CLEANING, ANOMALY DETECTION, PREP
MOVE/STORE	RELIABLE DATA FLOW, INFRASTRUCTURE, PIPELINES, ETL, STRUCTURED AND UNSTRUCTURED DATA STORAGE
COLLECT	INSTRUMENTATION, LOGGING, SENSORS, EXTERNAL DATA, USER GENERATED CONTENT

By Monica Rogati — https://medium.com/hackernoon/the-ai-hierarchy-of-needs-18f111fcc007

The Data Science spectrum in itself is huge. Most people lie in one of the strata of the pyramid shown in the diagram. Only a few can master 2 or 3 of the layers.

A data-driven organisation today employs for various positions and here is a list with the difficulty level of the problems that these professionals solve:

- **Data Analysts** — Easy to Medium

- **Data Engineers** — Medium to Hard

- **ML Engineers** — Medium

- **Research /Data Scientists** — Hard

- **AI Engineers /Deep Learning Practitioners** — Very Hard

Obviously, no one individual can pull off all the tasks. The first thing that you got to do is identify the skillsets that you have(or want to) mastered. Based on that skillset you should shortlist the job description that you aim for.
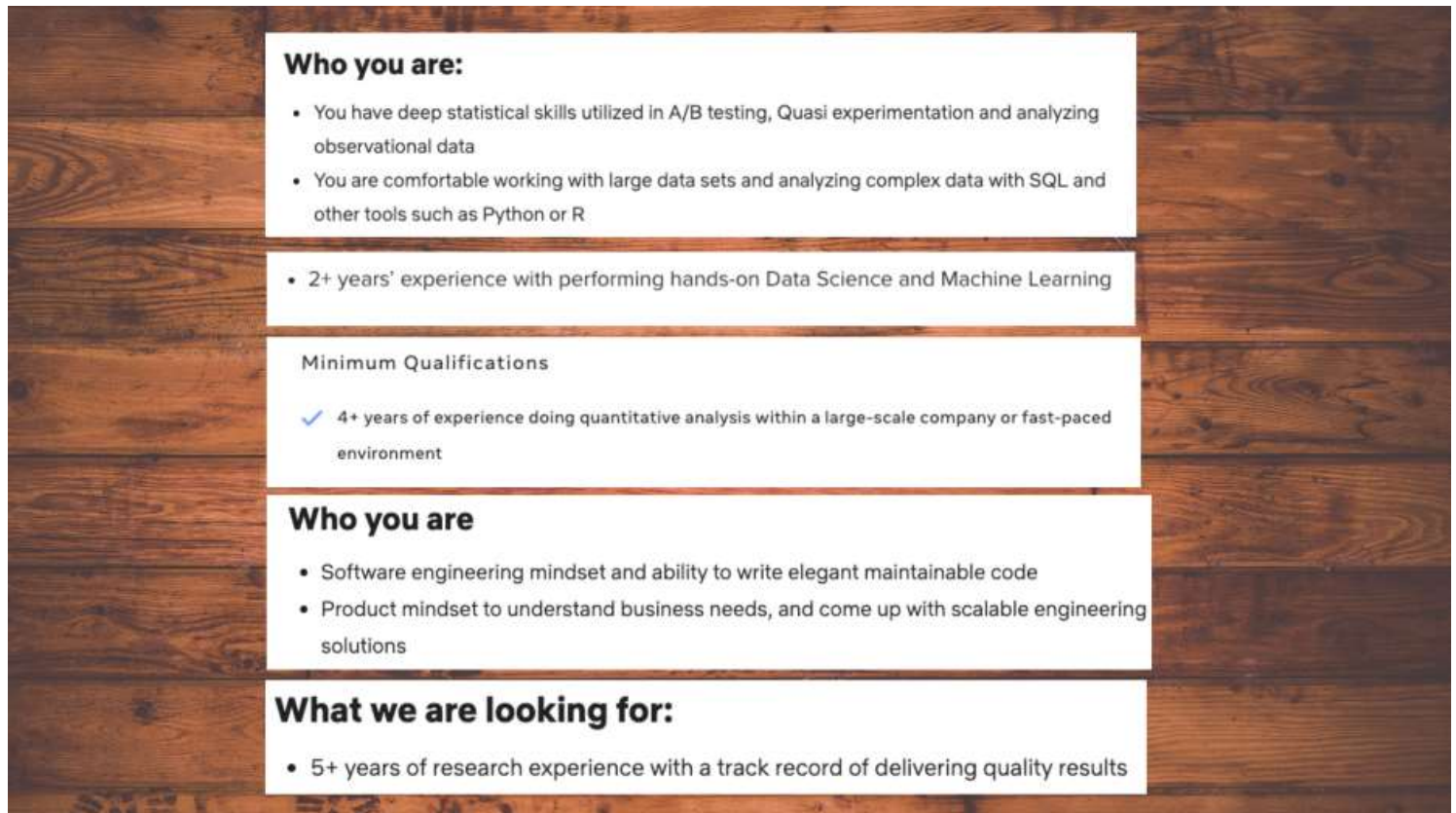
Step 2 — Studying the Job Description

If you spend enough time going through a bunch of job descriptions of various data profiles, you'll notice that they ask for the experience even if it's a fresher out of a college.

The second thing that you should understand is that there are jobs that have more generalist requirements like data analysis and then there are more focused and

dedicated areas of research like a research scientist at a hedge fund which is very math-heavy.

Here are a few screenshots that I've captured from a few big(Facebook, NetFlix) and mid-size(h20.ai) organizations look for in a candidate:

Screenshots from multiple job descriptions of data science roles.

Studying them takes us back to the very important and commonly asked question:

> *How do I compensate for the experience factor if I am a fresher?*

The answer is **Projects!**

> *Wait! I already knew that...*

Here is what you probably didn't know, these projects can't be your analysis of MNIST dataset or solving the titanic dataset classification problem.

> *So, what kind of projects? Where do I get these projects? What am I required to do?*

Let's deep dive into building the portfolio:

Step 3 — Showing Expertise via Projects

Projects are your only substitute for experience.

Chris Albon, when asked what freshers should have in their portfolio when they are seeking their first job in an interview with Datacamp, he said:

> *when someone applies, some of the best things that they can apply with are projects that they've done or something like, say, a boot camp or maybe their dissertation research or something like that, where we can take a look and say, oh, cool, like you've done some interesting stuff, you've worked with some data, some interesting ways.*

What should these projects reflect:

There are **4 major factors** that your project(s) should validate, no matter which profile you apply for:

1. Your firm grip over required **competencies**

2. The **complexity** of the problem you have solved or studied — it can either be a novel problem or a commonly asked enterprise-grade problem.

3. **Domain** expertise — the amount of research you did in order to find the answers to the questions or building data infrastructure.

4. **Your will** to go that extra mile and make the project stand out — Deploying your project for public use or writing a blog or publishing a video to explain your findings.

Types of Projects to Add to your Portfolio

Keeping in mind the above-mentioned factors, here's a list of project ideas that will require sincere efforts but will add weight to your portfolio.

- **Working with real data:** If you can show someone that you can work with raw data coming from different sources and answer interesting questions about social laws, finance, healthcare, or any scientific experiment, that would be highly regarded.

- **Exploring Publicly available datasets:**

Google Cloud Public Datasets

Uncover new insights with high-demand public datasets.

[View one-pager]   [Explore public datasets]

Explore public datasets in Google Cloud

Google Cloud Public Datasets facilitate access to high-demand public datasets, making it easy for you to access and uncover new insights in the cloud. By analyzing these datasets hosted in BigQuery and Cloud Storage, you can seamlessly experience the full value of Google Cloud with ease.

Making use of publicly available datasets, explore the data for several insights, define questions that have never been asked before, dig into journals and research papers to look for related material, and then uncover hidden patterns using statistical models. An in-depth analysis of a publicly available dataset is again a good place to start off.

- **Exploit your curiosity:** As a curious data professional, there must be products/services/questions that you find intriguing. Use this curiosity to dig into new problems. For example, a sports fanatic can go about building a dashboard or a data infrastructure that manages the statistics and performance patterns of all the players.

- **Contributing to Open Source Packages:** Every organisation highly regard open-source contributions to machine learning or scientific computing packages. Developing Free and Open-Source Software greatly enhances your chances to be recruited. You can try to contribute to packages like sklearn, numpy, pandas. It shows that you can work with huge and complex codebases and that you know your stuff well.

- **Building End-to-End projects:** A great way of proving that you are truly a generalist is to build end-to-end projects(more like products). Don't stop at finding the solution or creating a prototype for a recommendations system or a fintech chatbot, go the extra mile, deploy it, share it with your peers to use it, collect some

analytics. This shows how passionate you are about what you do and to what extent you can go to learn new technologies and methods.

- **Skill-specific projects:** There are people who are really good at cleaning data or creating insightful plots or automating data pipelines. You should consider developing your own Python packages that could automate those cleaning tasks or given a dataframe the package should create pair plots and all the other possibilities to expedite the EDA process.

List of some really cool portfolios for inspiration:

- https://nycdatascience.com/blog/student-works/improving-a-music-websites-user-experience/

- http://varianceexplained.org/r/trump-tweets/

- https://nycdatascience.com/blog/student-works/forecasting-cryptocurrencies-price-trends/

- https://nycdatascience.com/blog/student-works/web-scraping/covid19s-impact-on-preventable-cancer-risk-in-women-a-call-for-action/

Timeline for the project

The amount of time you spend on a project tells about the complexity, niche, and volume of work it requires. It should help you justify if the project is portfolio-worthy or not.

It would depend individually and how much effort you put into your project to take it to the next level.

Just to give you something to quantify, if you have picked up a nascent technology to work with, you should spend at least a month building something concrete.

How to add these projects to your portfolio

Once you have a few good projects that you can include in your portfolio, the next step is to package your work in the best possible manner.

Apple is known for its packaging and design. Be sincere about how you package your work before you display it. Here is how you can add more weight to your projects:

- **GitHub URL:** If you decide to add a link to your repo, make sure that the repo just doesn't contain a Jupyter notebook, it should have all the other files like requirements.txt .gitignore, a license if required, etc. You will be hired as a complete package and not just as a jupyter notebook expert.

- **Blogs:** Writing about what you've achieved is always a good practice and it builds the trust of the employer in your work and your ability to effectively communicate what you've done.

- **Deployed applications:** If you've deployed your ML-powered application, provide the link for the employer to play around with it.

- **Dashboards:** If you are proud of your analysis, you can go about creating a dashboard out of it. You may use Voila or Dash if you're working in Python. If you're a business analytics expert, you can add your Power BI, or **Tableau** dashboard to showcase your analytics skills.

Step 4 — Social Media Profiles

A good social media profile can help you land your next dream job. GitHub, LinkedIn, Twitter Kaggle, StackOverflow, Medium are the major platforms that people use to share their work/sentiments, network, consume information, and advertise.

Organizations and recruiters use these platforms to reach out to their next potential hire.

- **GitHub:** Having a good GitHub profile with a lot of contributions or stars on their repository makes you a competitive programmer.

- **Kaggle:** Participating in Kaggle competitions, creating useful notebooks and datasets can also help you build a good data analyst profile.

An excerpt from Reshama Shaikh's post To Kaggle or Not says:

> *It is true, doing one Kaggle competition does not qualify someone to be a data scientist. Neither does taking one class or attending one conference tutorial or analyzing one dataset or reading one book in data science. Working on competition(s) adds to your experience and augments your portfolio. It is a complement to your other projects, not the sole litmus test of one's data science skillset.*

- **LinkedIn:** I have personally used LinkedIn to land my first job, my first client, and many collaborators. A one-stop platform to connect with people who work at your dream companies, interact with them, find jobs, and follow interesting advancements. Do read <u>this complete data science LinkedIn Profile guide</u> to optimize your profile.
  Tip: You should be ready to offer something first before you ask for a favor.

- **Twitter:** All the big names in the data science space use Twitter quite frequently and you get to interact with people in your field. You learn about what these people are working on and their sentiments on social issues.
  You can promote your blogs, videos, and other findings with your Twitter. People have got job offers, invitations to conferences, freelancing work, and influencer marketing contracts for their work and good followership on Twitter.

Top Data Scientists to follow on Twitter:

- <u>Andreas Mueller</u> — Sci-kit Learn Developer

- <u>Yann LeCunn</u> — Chief AI Scientist at Facebook

- <u>Dean Abbott</u> — Chief Data Scientist SmarterHQ

- <u>Andrew Ng</u> — Co-Founder of Coursera

There are many others, you can look at my profile and the people I follow <u>on</u> my Twitter profile.

Step 5 — Condensing a Portfolio into a single page Resume

The most important element of your job application is your resume as it decides whether you're going to be shortlisted for the job or not.

Considering you have every other element in good shape, it's time to condense that information in an elegant and concise resume.

As you must know, the recruiters don't spend more than a couple of minutes to skim through your resume, so you need to convey everything you've done within a single page.

The most important sections after your name and contact info:

1. **Summary**: In 1–2 sentences, convey what you have been doing and what you intend to do.

2. **Skills**: Don't fill these up with all the random skills that come to mind. Don't mark yourself on a scale. A single line with all the major competencies should be enough.

3. **Projects**: This should be the major section for a fresher as you don't have much in your experience section. Be concise about what you've achieved, add hyperlinks to your work. Enlist capstone projects, Kaggle competitions, independent research, and projects. This section will be called your portfolio.

4. **Coursework**: Add relevant coursework only. You can mention your GPA if applicable.

5. **Experience**(if you have any): Add relevant job history along with the bullet points that speak of the major tasks you accomplished at the organisation.

6. **Social Media Links:** Don't forget to add links to your active social media profiles.

Here's an example of a good resume that was reviewed during Kaggle CareerCon2018.

An example of a good resume presented at Kaggle CareerCon

Call to Action

There would be a lot of questions regarding projects for each profile. Where to look for them? How to get started? How to prepare for interviews and many others?

I have been working on creating projects for each profile based on my experience working as an Instructional Designer for Web and Data Science tracks.

Based on your response to this post, I will create a Discord channel for each profile where I'll be sharing the projects and the instructions to complete them with the

timeline associated with each.

I strongly believe in project-based pedagogy and thus I would be creating a lot of content where project development would be covered. I'd be sharing the resources you can use to learn(some of which I'll create myself) and complete the projects successfully.

You can look at one of my examples here: COVID-19 Interactive Analysis Dashboard from Jupyter Notebooks.

The next part of this blog would cover **Interview prep** which I'll be publishing in a few weeks as I'd be sharing a few project-tutorials first.

Here's the video version of this blog post on my channel Data Science with Harshit:

Data Science with Harshit

With this channel, I am planning to roll out a couple of series covering the entire data science space. Here is why you should be subscribing to the channel:

- This series would cover all the required/demanded quality tutorials on each of the topics and subtopics like Python fundamentals for Data Science.

- Explained <u>Mathematics and derivations</u> of why we do what we do in ML and Deep Learning.

- <u>Podcasts with Data Scientists and Engineers</u> at Google, Microsoft, Amazon, etc, and CEOs of big data-driven companies.

- <u>Projects and instructions</u> to implement the topics learned so far. Learn about new certifications, Bootcamp, and resources to crack those certifications like this **TensorFlow Developer Certificate Exam by Google.**

If you liked my content and would love to see more of it, connect with me on <u>Twitter</u> or <u>LinkedIn</u>.

Data Science   Programming   Artificial Intelligence   Machine   Careers

About   Help   Legal

Get the Medium app