



DATA VADER

STEP 4 OF 7 STEPS STARTER PACK

---

# BUILDING YOUR PORTFOLIO

BUILD A GROWING AND FULFILLING CAREER IN DATA SCIENCE





# Step 4: Building your portfolio

Hi there! Welcome to the fourth chapter of my eBook, 7 steps guide to build a growing and fulfilling career in Data Science.

I hope you have received and read Step 1, Step 2, and Step 3 of the guide.

Step 1 was about understanding Data Science, knowing why we see Data Science everywhere, and diving deeper into the components of Data Science: Mathematics/Statistics, Programming, and Business understanding

Step 2's content helped in knowing about various roles in Data Science and also help you in choosing the right role for you.

Step 3 was about getting to know how to acquire the skills required to land your dream career. I compartmentalized into three routes: doing a master's, doing a certification program, or taking online courses.

This chapter is your guiding light if you are looking to transition into Data Science or looking to grow your career in Data Science. This is a guide on building a portfolio.

When I was starting my career in Data Science, I used to hear all the time, build a portfolio, do projects but these words did not work. I had to explore. I am writing the things I have learned over the years.

So you are pursuing a master's or doing a certification program or you have recently finished some online courses. You are searching for jobs but you are seen as an inexperienced data professional.

This is like a paradox out here. *You need a job to get experience and you need the experience to get a job.*

How do you break the loop then?

By building a portfolio!

## **What is a portfolio?**

In most simple terms, a portfolio is a collection of your works that showcase your expertise in a particular domain.

It is your chance to impress your recruiter with the skills you possess and make them believe that you know what you are getting into.

It would be a one-stop destination where someone would get all information about you as a data professional. It is an extended online resume.

In certain professions like graphic designing, architecture, it is sort of mandatory to have a portfolio. Last month, when I was hiring a graphic designer, the first thing I looked for was their portfolio. I wanted to know what they are good at.



I had to hire two graphic designers, one for website UI and another for logo and graphics, depending on their expertise. I did this by looking at their portfolio alone!

This might happen to you as well. If you have a lot of projects on NLP in your portfolio, chances are that startups or companies requiring someone in NLP might contact you.

Your portfolio can be made out of various components:

- Github profile
- Kaggle competitions/notebooks
- Blogs on medium, substack or WordPress
- Personal website

The format in which you deliver your expertise can be different on different platforms.

Your Github profile will be full of code-heavy with a readme file and few dropdowns here and there.

Your blog will have an in-depth explanation of topics with codes. They have to be detailed with the technical stuff logically structured. I can promise you that you will learn the most when you write a blog. You will be giving it out to the world and you will be attentive to the details.

Kaggle is one of the best places on the internet to practice data science. Your consistency there will help grab the recruiter's attention.

Then comes your personal website which has everything in one place. All the projects that you have coded, the blogs that you have written, and your Kaggle titles.

You can actually create a personal website out of a Github profile, no need to spend a lot on building a website. If you are a web designing fan like me, please create one (and share the template with me :p).

The intention is the same: **To keep everything in one place.**

Now comes the disclaimer!

It is absolutely not mandatory that you will have to create 100+ projects on Github to get hired. Or that this will make you the best data scientists out there.

This is far from true!

A lot of awesome data scientists that I know, don't have repositories on Github. So, you know that it is not mandatory. But I will give you some reasons to work on your portfolio:

- It will help you break into data science as a fresher
- It will make you stand out as a data scientist
- It enhances your understanding of topics
- It motivates you to pursue some passion projects which could actually be lifesavers.

If you are in a job and don't want to continue doing data science in your free time, that's okay. That is your day job already :)



# **What are projects?**

A project is basically a sequence of tasks done to accomplish a certain outcome. Data Science projects are similar.

They use different tools and techniques to get results and solve a problem statement. With data science projects, you can demonstrate:

- Problem-solving skills
- Proficiency with certain data science tools
- Mathematical and statistical understanding
- Ability to communicate

This gives you even more reasons to do a project.

After you have done a project, you can essentially upload the code on Github with an elaborate readme file. Create a blog post around your solution, the challenges faced, and how you overcame them.

One step further, you can even create a YouTube video to explain it all. As we know, the best way to learn something is to teach one, so teach, reach out to people.

You can also do a fun project like a meme-generating Bot and share it with the world. Learning Data Science while having fun: double treat. Are you up for it?

# **How to proceed with a project?**

While doing a Data Science project, you will encounter these steps:

- **Data Collection**

You can collect data from various sources. You can run your own survey, download some publicly available datasets, or you can scrape data from the web. Also, there are APIs.

Scraping has some legal and ethical concerns so before you do that, be conscious and read all the details.

- **Data cleaning**

This is a very important step in each and every data science project, including the ones that you will do after coming into a job.

Sorting, segregating, making data consistent, imputing missing values are some of the things which you will do in this step.

- **Feature engineering**

This step is about finding feature importance, feature selection, and merging features. Trust me if you do this step correctly, half of the issues during modelling or analysis are resolved. Again, this step is an important yet ignored step, just like data cleaning.



- **Data Analysis**

This is the step where you generate insights from the data and use the data to derive actions that can be taken to improve metrics. This I have discussed already in earlier chapters.

- **Modeling**

You would potentially end your project by creating a dashboard if you want to stop at analysis.

However, if you are looking forward to performing Machine Learning or Deep Learning then this is the step for you.

- **Deployment**

It is very essential to serve your model. This you can do by creating web endpoints and then deploying over some cloud service.

That's it! You are done.

Now go and document your work as a blog and show the world!!

You can find a lot of tutorials and datasets on Kaggle, medium, and other blogs. YouTube is a great source. If you need personalized learning with mentorship, DataVader is here for you!



# **How to choose a project?**

It is up to you to choose a project. But, make sure that your projects are able to demonstrate these things:

- Your competency with mathematical and statistical concepts
- Your familiarity with tools required
- Your ability to solve complex problems. It could either be a novel problem that involves saving the world or some industry-grade problem.
- Your domain expertise. Let's say you are analyzing stocks then you should be able to decipher the P/E ratio, NAV, Dividend payout ratio, etc.
- Your ability to live with the project. Keep updating and integrating best practices, as it happens in companies.

Now, a couple of things more which I would like to discuss before we end this chapter.

Explore as much as you can. Find a domain that interests you and start doing projects. Don't only focus on state-of-the-art models and the highest accuracy, also focus on documenting your work and exploring the data properly.

Doing projects on the titanic dataset or the iris flower dataset will not add much value as it does not help you distinguish from others. Pick an interesting dataset.



# How DataVader can help you build portfolio

As we discussed, the portfolio is a very important aspect of being a data scientist.

At DataVader, I help you create real-world projects with real-world data. We skill you in all the end-to-end process, including:

- Data collection
- Data cleaning
- Data analysis
- Creating dashboards
- Modeling
- Deployment

These come with my 1:1 mentorship. I work closely with all the participants and have a small group.

I encourage you to solve business problems than following codes as usually done in online courses.

With DataVader, you get hands-on experience, a steep learning curve in Data Analytics and Machine Learning.

With that, I would like to conclude this chapter. I hope you have gotten an idea of how to build your portfolio and land your dream job.

If you need any help, I am here!