# *Data Science*
# *Project Portfolio Guide*

*By* ***Sebastian Gutierrez***

*www.DataScienceWeekly.org*

# Table of Contents

# 1

# Introduction

## 1.1 How To Use This Guide

**Read This Guide All The Way Through At Least Twice**

This guide will help you create the best possible data science project portfolio.

The first read through you do is to help you understand the reasons behind the decisions and suggestions in this guide.

You will learn what all the specific parts you should have in your portfolio, all the things you will do, and all the things you should expect to do.

The second pass is for when you start implementing and working on your data science project portfolio.

**Read This Guide After *The Data Science Getting Started Guide***

Though you can read this guide and implement you project portfolio without reading *The Data Science Getting Started Guide*, you'll miss out on a good deal of strategy that will drive superior results.

By reading the getting started guide first you'll have a solid understanding of what data science job opportunities appeal to you and which you want to pursue.

By knowing these things you'll be able to concentrate and better spend your time on things that matter while ignoring things that don't.

**Read This Guide Before *The Data Science Resume Guide***

Though you can craft your resume before putting together your portfolio, you'll be handicapping yourself because you won't have as many amazing targeted data science projects to include in your resume.

You cannot change the past, so that part of your resume will remain the same.

You can change the present and the future so you can make it apparent in your resume and cover letter what data science you can do through your work on your data science project portfolio work.

## 1.2 Form Follows Function

The idea of form follows function is that the form or shape of a thing should be based on its intended function.

## The Functions Of Your Project Portfolio

A Data Science Project Portfolio functions as a living representation of your experience with various data sets, technologies, tools, data science techniques, business problems, communication strategies, and general interest.

A project portfolio further functions as an expanded resume (i.e. based on the work you've done in the past here is work you can do now and in the future).

Lastly, a project portfolio functions as a complementary advertisement of you.

## Your Portfolio Is An Advertisement For You

While creating, building, and sharing your project portfolio will cause you to learn and develop many skills, you will treat your portfolio like an advertisement for you and your data science capabilities.

Everything in the portfolio should be on message and never stray from communicating that you are an awesome and very capable data scientist.

From the overall snapshot of the portfolio to the intricate detailed write-up of various data science projects, every single thing must be filtered thorough the idea that the portfolio is being used to sell your capabilities.

# 1.3 Showing Your Work

**Your Portfolio Consists Of Your Work**

Your project portfolio is an advertisement of what you are capable of doing and achieving.

Your portfolio shows what you are capable of doing and achieving.

This is where people fall down – they think that having links to projects or a GitHub account is all they need.

After all – isn't that your work?

Yes – it's your work.

No – it is not enough because 99.999% percent of the people who ever look at a GitHub account will not care enough to spend hours wading through code files let alone spending 15 minutes trying to decipher what you did or how you did it.

The 0.001% of people who do take the time to decipher your GitHub profile will be navigating blindly while looking at files randomly as they try to take it all in.

Which means they are not being guided through your work.

Which means they are not seeing the best side of you.

Which means they are not seeing what you think they should be seeing.

This means you need to show your work AND do it in a very deliberate manner.

## 1.4 Practice Makes Permanent

**Building A Data Science Project Portfolio Means Doing Data Science**

This shouldn't surprise you.

What it does mean is that you get to proverbially kill two birds with one stone – you learn data science and get a project portfolio from it.

And because you will want a well-rounded portfolio, it means you'll be doing lots of data science.

Which means that you'll be internalizing processes, techniques, tools, and know-how as you build your portfolio.

**Opinions Backed By Experience Are Very Valuable**

You'll naturally develop opinions and intuition into how to do many things as you internalize your data science work.

This is incredibly valuable because it means that when someone hires you, you'll be able to proverbially hit the ground running.

Which means they can pay you to solve problems rather than paying you to learn how to eventually solve problems.

The more you practice doing data science, the better data scientist you become.

The better data scientist you become, the easier time you will have getting interviews, doing very well at interviews, and getting hired.

## 1.5 Perfection Is The Enemy

**Quantity Begets Quality**

By focusing on quantity of small projects in your portfolio, carefully chosen of course, you'll do enough data science to develop quality skills.

In order to do a large amount of projects you will have to live with the fact that they will not be perfect.

There's a saying in software that the first 90% of a project takes 90% of the time and the last 10% takes another 90% of the time as well.

Knowing that you'll be doing lots of projects frees you from having to have one perfect project.

Taking care of quantity will make quality happen.

**Your Data Science Project Portfolio Is A Living Document**

This means that as you get better and better, you can revisit older projects to rebuild them to better showcase your new skills.

There are no rules against this, so you can always be improving the whole portfolio.

Rather than trying to make everything perfect the first time, you can focus on moving forward and learning to do great work.

By moving forward and doing data science you are also practicing for your interviews, test-projects, and doing the actual work once you get hired to do data science.

## 1.6 Have Fun

**Building A Portfolio Is Being A Data Scientist**

Have fun and enjoy the process.

If you learn to appreciate the process and enjoy what you are doing then you will have a very fulfilling career.

**Do What Is Fun For You**

Enthusiasm is contagious and will come across the projects you choose and how you communicate what you did.

It will make it easier to sit down to do the actual work and it will make it easier to continue to build your portfolio past the initial set of projects you choose to do.

Very few people ever have to force themselves to have dessert.

Figure out what data science work you like enough that it feels like it is eating dessert.

### What If I Can't Find The Fun

The goal is to find the intersection of what you find fun and what you can do in data science.

Building a Data Science Project Portfolio is a good way to explore and try to find this intersection.

However, if you can't find it there is no reason to despair.

It just means that you should find something else to do.

You show be thankful you discovered this before either struggling to find a job or getting a job and then struggling to do it.

Life is too short and precious to spend doing something that has zero fun for you.

# 2

# What Is A Data Science Project Portfolio And What Does It Do

## 2.1 What Is A Data Science Project Portfolio?

A portfolio is a living document showcasing your data science interests, abilities, and work.

It is living because it's something that grows and evolves as you do.

As you learn and encounter new things doing small examples projects for the portfolio will help cement your newfound knowledge.

"Showcasing" is important because the portfolio will be read, studied, and dissected by other people.

Not only will other people learn from you, they will also learn about you.

The primary reason most people put together a project portfolio is for use during their job search.

Whether it's an internship, a part-time job, or a full time job, the goal is reasonably clear.

And, in many ways, it's a multi-stage goal.

The portfolio gets you an interview.

The portfolio gets you prepared for an interview.

The portfolio gets you prepared for the small-project part of the interview.

The portfolio gets you prepared for the technical interview questions.

The portfolio gets you prepared for small talk before, during, and after the interview.

The portfolio gets you prepared for deciding whether the offer is right for you or not.

The portfolio gets you prepared to do the actual work.

Later the portfolio will prepare you to continue moving up in the same organization or switch to a new role at another organization.

Equally helpful is that as you are building the portfolio and going through these steps, it will help you determine your interest in the type of work, techniques, data, and environment.

Given all these areas where a portfolio helps, you can further break it down by who is being helped by your portfolio – employers, yourself, and others.

## 2.2 What Your Project Portfolio Does For Employers

**De-Risk, De-Risk, De-Risk**

There's a joke that the three most important words in real estate are: location, location, and location.

From an employer's point of view, the three most important words are de-risk, de-risk, and de-risk.

An employer wants to know that you know what the job entails.

An employer wants to know that you can actually do what the job entails.

An employer wants to know that you are actually interested in what the job entails.

If they can find someone applying to their data science job that checks of any, some, or ask if those things they'll be thrilled.

This is because of the word "de-risk".

Each of these statements takes part of the risk of hiring someone out of the equation.

Employers are afraid of hiring and then having to fire someone.

Most employers would rather wait to find the right candidate.

Your portfolio will be constructed with this in mind.

**Your Portfolio Shows You Know What The Job Entails**

By constructing your portfolio in such a way that you use and demonstrate knowledge of all the parts of the job you are applying to, you show that you know what the job entails.

This is why it's important to read the *Getting Started In Data Science Guide* first as it will walk you through really figuring out what the jobs you are really interested in entail.

Your portfolio will differentiate you from those who are applying to data science jobs indiscriminately.

**Your Portfolio Shows You Can Do What The Job Entails**

By constructing your portfolio in such a way that you use and demonstrate knowledge of all the parts of the job you are applying to, you show that you can do what the job entails.

Being knowledgeable about the techniques, software, and data used in the job are all very welcome and valuable to the employer.

What is equally valuable is your ability to communicate all of this knowledge through writing.

A surprisingly large part of a data science job entails communicating the work you will do, the work you are doing, and the work you have done.

Your portfolio demonstrates that you can communicate through written text.

Your portfolio shows you are able to create reports.

Your portfolio shows you are able to communicate technical information.

Your portfolio shows you are able to work through a whole problem and then describe what you accomplished.

Your portfolio shows you are able communicate what you learned from doing a project.

Your portfolio demonstrates that you have opinions about data science and all parts of the data science process.

**Your Portfolio Shows You Are Interested In What The Job Entail**

An employer has chosen to spend time and energy in a certain field in a certain type of job doing a particular kind of work.

They enjoy it enough to be actively employed in it.

They have been successful enough in it that they are a hiring manager.

Can you imagine how excited they are going to be when they find out you are interested in the same things they are?

That's what you're portfolio does – show them that you are willing and able to spend a significant amount of time in a certain topic area.

If this topic area weren't intriguing to you it would have been very hard to accomplish putting together your portfolio.

Thus your portfolio will differentiate you from other candidates who either have no indication that they are interested in the topic area or are asking to be taken at their word that they are interested.

Your extended work and interest in the area will win hands down every time.

**Your Portfolio Shows You Have Clarity Of Thought**

Clarity of thought is important to hiring managers.

Showing your clarity of thought through your project write-up de-risks the hiring process for you future employers.

Striving to communicate with clarity will help communicate clearly and with purpose.

This will help you prepare for interview settings.

Practicing clarity of thought will help you put your stories together for interviews since you'll have already thought about what you did, what you learned, what insights you achieve, and what opinions you formed.

## 2.3 What Your Project Portfolio Does For You

### Your Portfolio Helps You Focus

Having a portfolio with a goal helps you focus on your areas of interest.

Since you have chosen themes to work on, it keeps you from getting distracted.

Every time you read a new blog post, article, or tutorial you'll be able to quickly decide whether it is relevant to your career and job search.

This helps you spend more time on things that will help you rather than general knowledge building you may need in the future.

### Your Portfolio Helps You Build A Narrative

Having a focused portfolio allows you to build a narrative of where you have been and where you are going.

This in turn helps you during the job search process in three ways.

First, if things aren't going well you can feel very demoralized and lost.

By having a narrative you'll be able to see the bigger picture and let the small letdowns flow over you.

Second, when things do go well they will help to reinforce the narrative you have already built up therefore strengthening your resolve.

Everyone has a story in their head about what is happening and why.

As you see evidence of your story coming true your self-image and self-confidence will improve.

This self-confidence will help your interviewing.

Third, when you do interview, the narrative will be easy for the interviewers to understand and thus be able to make the case for you to others.

By having a legible narrative that makes sense and others understand, they will have less to fear about hiring you.

This in turn makes it easy to extrapolate what you will be like in the future.

**Your Portfolio Helps You Talk About Yourself**

In professional settings you will succeed based on your ability to do great work while simultaneously telling other people about that great work that you are doing and have done.

Job interview success is based on your ability to talk about yourself and what value you can provide to the employer.

In gatherings like conferences, meet-ups, lunch-and-learns, and similar settings, a foolproof way to connect to others is to find out how you can help them.

Having an ongoing portfolio gives something to talk about and to share.

It allows you to talk about yourself through your work without having to resort to telling people you are great without providing any evidence.

By working on your portfolio you are also developing thoughts and opinions on data, data science, and the subject areas you are working on.

It also gives you something to share on various social networks.

This further allows people to interact with you and get your opinions on work you have done.

**Your Portfolio Will Help You Build Your Web And Social Media Presence**

Since you know what key words, topics, techniques, languages, and libraries you are responsible for showing you know, you can use the portfolio to do a bit of self-promotion.

Always remember to understand the written and unwritten rules of a community before submitting self-promotional material.

Communities to look into (will differ based on key words and topics you are looking for):

- LinkedIn – various data / analytics / data science groups
- GitHub
- Gitter
- Stack Overflow – various topics
- Twitter and various hash tags
- Medium
- Hacker News
- DataTau
- Reddit and various sub-reddits

The communities you join need to be related to the information and types of jobs you are targeting.

Sharing your knowledge and answering questions about these topics will help you build your presence in these communities.

**Your Portfolio Helps You Build Your Network**

A big part of building a great network is finding the right people.

While being an acquaintance of a famous data scientist maybe be interesting, much more useful to you will be making contact with someone in the particular job you want with the particular skills you are developing.

This is because they will have a much better understanding of the hiring landscape and can potentially point you in the right direction.

Also, in case they are not famous, they will have less people trying to "network" with them and so it will be easier for you to connect with them (as long as you do so in a non-spammy way).

As you construct your portfolio, look for work to replicate, and see what others are doing you well naturally run into the types of people whom you should get in contact with.

**Your Portfolio Helps You Connect And Interact With Mentors**

As you encounter people's work in your area of interest send them a nice note telling them you enjoyed their work.

Every once in a while you can also send them your work with a note along the lines of "you did this great thing and I took it and tweaked and wanted to share with you in case you were interested".

A few people will engage (not everyone has the time or energy or scope).

As you build these relationships you may find that you click with a few people.

These people have seen that you are doing actual work and have something to show.

These people can become an unofficial mentor to you.

Show you are doing the work and people will take you seriously.

# 2.4 What Your Project Portfolio Does For Others

### Your Portfolio Helps Others Learn From You

You will learn things that were not obvious to you as you build your portfolio.

When you write these things down others will learn from you.

You will have made their lives better.

You will have helped them achieve a goal.

And because of this you will be seen in a positive light.

### Your Portfolio Helps Others Learn About You

In addition to setting you in a positive light, they will get curious about you.

They will want to learn more about you and from you.

They will be more inclined to recommend you because you have helped them and they have an easy way to point to how you helped them – your portfolio.

### Your Portfolio Helps Others Promote You

When others share your work or tell others about what they learned from you they will be providing an endorsement of your work and you.

Having a well-written portfolio that helps others learn serves as marketing material for you and your work.

By sharing your work, that is good and helps people learn, others will look good to their followers, colleagues, and friends.

In this way your portfolio work pays off twice – you help someone else learn while at the same time you help him or her look knowledgeable to his or her peers.

## 2.5 What Your Data Science Project Portfolio Should Not Do

**Your Portfolio Should Not Be Unprofessional**

Yes – is your website portfolio and you can do what you want to.

However, you need to remember that you are creating your portfolio to achieve getting a job, which means you should use safe for work language.

Do not use not-safe-for-work (NSFW) content unless specifically applying to an adult company that expects NFSW content.

Do not use humor unless you are 100% sure it will not offend.

It's not that you should be bland; it's more that you shouldn't disqualify yourself from a job before you even get considered.

Along the same lines you should refrain from being overly negative.

**Your Portfolio Should Not Misrepresent You**

Don't lie about what you have done.

Don't lie about your ability.

Don't lie about what you can do.

Don't misrepresent your history.

**You Portfolio Should Not Be Hard To Understand**

It should be spell-checked.

It should be grammar-checked.

It should be in the same language as the potential audience.

It should not talk about things outside of your data science projects and interests.

It should not be messy.

It should not be hidden in your website.

### Your Portfolio Should Not Make You Impossible To Contact

You are constructing the portfolio to show your work.

If someone wants to talk to you about your work make it easy for him or her to contact you.

Regardless of where someone is in your portfolio, it should be easy for him or her to find your contact information.

# 3

# How To Build Your Portfolio Quickly And Effectively With The Right Goals And Direction

## 3.1 The Main Goal

There are many goals that you could achieve with your data science portfolio.

You will only focus on the overarching goal of using your Data Science Project Portfolio to obtain a Data Science Job for the purposes of the rest of this guide.

What is now needed is the general direction in which to start off.

As they say, "the journey of ten thousand steps starts with one."

Which is all well and good except that if you are facing the wrong direction you'll just end up ten thousand steps in the wrong direction.

# 3.2 Choosing The Direction Of Your Portfolio

**General To Focused**

When choosing the direction of your portfolio work it's best to start out very general and build up to a very specific direction.

This allows you to start with a general theme and slowly discover what appeals to you.

To find the theme to start with, look at what jobs appeal to you.

Look at what industries appeal to you.

Look at what tools appeal to you.

Look at what techniques appeal to you.

Look at what end goals appeal to you.

As you then work through projects you'll develop a "taste" about what you like, don't like, and what areas you want to dig deeper into.

It is this "taste" that employers will be hiring for.

**Small To Big**

Since you are going to start generally and work towards highly specific projects it is best to start with very small projects.

This ensures you don't spend too much time and energy on general projects.

Doing small projects means you will cycle through them quickly and thus develop opinions about things faster than if you are doing only large projects.

As you get deeper into the theme you are exploring you can do bigger projects to further cement your knowledge and expertise of the particular area.

**Easy To Advanced**

It is important to start working on easy projects.

It means you'll be more likely to complete them.

It means you'll gain confidence by seeing a project to completion.

It means that you won't get discouraged from having to learn too many things at once.

Approaching each project as an opportunity to learn only one thing at a time will ensure that the project is small and easy to complete.

As you work through many projects you'll develop the skillset to work and learn more advanced topics.

**Taken All Together**

This means that you should start with very small, very general, very easy projects and work up to very specific advanced projects.

This ensures the direction your portfolio grows is organic and towards the type of work you would enjoy the most.

## 3.3 How To Start

Down the road your portfolio will be full of projects.

When you start however, you will only have one project so it will make you look like a beginner, which is fine since you are a beginner.

**Start By Setting An Achievable Schedule Of Wins**

The best way to make your data science portfolio useful from day one is to use it as a forcing mechanism to start doing data science projects.

Your goal is to learn and do data science.

So treat your portfolio as if it is already full.

This means that instead of trying to do huge projects that will make your portfolio look full, focus instead on very small projects that focus on one thing and can be done in one sitting.

You'll eventually be doing that anyway so better to start doing it right away.

Make your project easy to do and easy to finish.

Master the fundamentals first.

These things will make it easier for you to complete projects and thereby increase the numbers of wins you get.

## 3.4 Make It Really Easy To Win

You want to build your portfolio as quickly and effectively as possible.

**Decide On A Time Budget**

For each project set a time limit during which you must finish.

Denote and measure this time limit in minutes, not hours or days.

This will force you to not delude yourself about how long you are actually working on any one project.

This forces you to scale down the type of project you do as well as keeps you in check.

It also allows you to add small wins even if you work 5 minutes or there.

**Work With Purpose**

As you work within your time budget, it's important to only spend time on things that move your project forward.

This means that you should already know exactly what you are doing and to what end whenever you sit down to work.

Having an end goal in mind from the beginning allows you to make a better plan and course correct if you stray from the plan.

**Use Mind Maps**

Once you have an end goal in mind and are working backwards, the best way to organize the sequential steps is a one-directional mind map.

This type of mind map layout makes it clear what steps are next.

This layout also makes it clear what other steps you could take to keep moving forward if you are stuck on a particular step.

This layout also helps you establish a structure for working on future projects.

**Don't Waste Your Time On Things That Don't Matter**

Of course don't waste your time on things that don't matter.

This means that you have to have a very good idea of what things actually matter for achieving your goal of getting a data science job through the use of your data science project portfolio.

This is where the previous guide, *Data Science Getting Started Guid*e, can help.

## 3.5 Refer Back To Data Science Getting Started Guide

**Pairwise Comparison Winners**

In the *Data Science Getting Started Guide* you spent time shifting through and finding data science jobs that appeal to you.

Whether it was the tools, the data, the statistical techniques, the programming work, the math, or the business goals, you found a set of jobs you liked and a reason for liking those jobs.

In addition to the general direction advice above, it is highly recommended to keep the pairwise comparison winners in mind when building your portfolio.

You do this because you are building a portfolio to become the perfect candidate for the jobs you liked.

Having a very clear goal will help you achieve that goal.

Additionally, because most jobs list nice-to-haves and must-have, you can methodically work through them to make sure you are as competitive as possible for those types of jobs.

### Methodical Coverage Rather Than All Encompassing Projects

When you look at the list of nice-to-haves and must-haves across all the jobs you'll find lots of duplicate items.

You will also find lots of unique items.

Regardless of the balance between duplicate and unique things, you'll still end up with a long list of things you want to show mastery on.

Start with the duplicate items as they cover the most ground.

What is less obvious is that the best strategy is to do many very small projects covering individual skills rather than doing a few all-encompassing projects.

You want to do the small projects that focus on one skill because it is very clear to someone looking at that project what it was that you were trying to accomplish and whether you actually accomplished it or not.

# 4

# How To Structure Your Data Science Project Portfolio Landing Page

## 4.1 Why Structure Matters

**Easy To Get A Broad Overview**

You want to communicate as much information as you can as fast as possible.

You don't know how long someone is going to look at your portfolio page.

Whether they look at the landing page of your data science portfolio 5 or 60 seconds you want them to take away a positive impression of you and your work.

Someone who only looks for 5 seconds will see a well-structured page and a few key words.

The longer a person stays on your website, the clearer a picture they will have of what you know and what you don't know.

To make people stay longer you have to be memorable as well as entice people to look further into your work.

To that end, your data science project portfolio landing page needs to be 100% scan-able.

From a brief look a reader should be able to answer basic questions about the data, algorithms, techniques, tools, and projects you have used.

**Easy To Navigate**

If a potential employer comes to your website to see if you have a particular experience with a data set, or a tool, or an algorithm, or a technique they should be able to find the relevant information as soon as possible.

Most people will not have the time, energy, or inclination to dig through a poorly structured portfolio to find a specific piece of information.

If you make your portfolio easy to navigate you will have a big advantage of your competition.

Making it easy to navigate also means that all of your work will be visible.

This ensures that when someone looks at the portfolio they are not going to miss seeing something because it was hidden or not obvious to them for how to find it.

**Easy To Explore**

Along the same lines, you should make it as easy as possible for someone to explore your work once they have an interest to do so.

This is the way you turn a casual browser into a serious reader of your portfolio.

By making it easy to explore you raise the probability someone will look at more than one of your projects.

Creating a structure that makes it easy to explore makes it easy for people to be exposed to more of your work.

**Easy For You To Build Against**

Up to now the reasons for a great structure have been based on the needs of others.

An inwardly focused reason is that once you have a great structure, it is very easy for you to see what comes next and what gaps you have in your work.

As you workout what structure makes the most sense for you and your projects, you'll develop templates you can use in the future.

Which means the next time you start, you'll already have a structure in place you can just fill in.

Additionally it gives you a small nudge to continue working to fill out the portfolio, as your portfolio will look somewhat empty at the beginning so you'll be incentivized to continue working on it.

This structure will also help you keep your projects small as the smaller they are the quicker you can finish them thereby filling up and fleshing out your portfolio.

## 4.2 Overall Structure

**InfoVis Mantra Drives Overall Structure**

In 1996 an Information Visualization researcher, Ben Shneiderman, wrote a paper titled *The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations*.

From this academic paper comes Ben Shneiderman's InfoVis "mantra": "Overview first, zoom and filter, then details-on-demand".

It is this "mantra" that drives the structure of your portfolio structure.

You want anybody coming to your portfolio for the first time to get an overview first then be able to zoom and filter based on the details they want to review.

**List Of Specific Sections In Landing Page Structure**

This is the overall structure of the landing page:

- Portfolio Title Section
    - Your name

    - Data Science Portfolio
- Your Introduction Section
    - Who you are sentence
    - What you are interested in sentence
- Portfolio Introduction Section
    - Portfolio introduction sentence
    - Portfolio theme sentence
- Portfolio Project Section
    - Projects listed by Data Science Process
    - End-to-end large project(s)
- Portfolio Project Tag Section
    - Tools
    - Techniques
    - Data
- Your Contact Details Section
    - Full name
    - Best email to reach you on
    - Main personal social media profile
    - Social Media profiles related to data science
    - GitHub if available
    - Personal Blog if available

Let's take a closer look at each of these sections.

# 4.3 Portfolio Title Section

**Make It Clear Who Created This Portfolio**

The first line should be your full name.

This makes it obvious that it is your portfolio.

This helps in case your website URL does not make it clear who you are.

It is helpful to make it abundantly clear that this is in fact your page and your portfolio because some people will have their website on github.com pages, squarespace.com, WordPress, Wix, or another website hosting company.

The second line should say Data Science Project Portfolio.

This makes it clear to you and your audience what is on the page.

Remember, the less confused someone is when looking at the page the greater chance they will get value from what you wrote.

Then put an empty line before the next section.

# 4.4 Your Introduction Section

**Introduce Yourself**

In this section you want to introduce yourself to make out perfectly clear who you are and what you are interested about in the data science realm.

The first sentence will say who you are and contain basic and pertinent information.

You'll want to write something close to the following:

> [Your full name] is a data scientist with a background in [educational and/or work background] who focuses on solving problems in the [industry name] industry.

This sentence says right away that you are a data scientist.

You want to fake it until you make it.

Note, if think you are qualified to apply to "senior" data scientist positions then you would add the "senior" qualifier in front of data scientist.

Then it covers your educational background if it is relevant.

As well as covers your work background if it is relevant and available.

You then make it apparent that you understand that data scientists are hired to solve problems.

The last part of the sentence communicates that you have an idea of what industry appeals to you.

This helps to frame the full project portfolio.

The next sentence then establishes the major theme(s) of your project portfolio.

It covers tools, techniques, data, and specific things inside the industry you mentioned in the previous sentence.

Something close to the following:

> Specifically, [first name] uses [very broad programing tool or tools] to solve problems and generate insight in [line(s) of business] lines of business using [targeted tools, algorithms, and techniques].

Then it has your name to further establish that this is your portfolio.

Having your name here again also serves to establish a link between the specific things you are about to cover and your name.

You then specify a few broad technologies to setup the latter part of the sentence.

Here you could use examples like "Python", "R", "Scala", "Java", "Julia", "Clojure", etc., as it will help the reader place you within their universe of data science knowledge.

Note that there are some jobs where multiple languages will be expected.

If a tool is central to the job, like Tableau or Spark or AWS or a proprietary data store like Vertica then you'll want to mention it here as well.

Word of warning: this part of the sentence is meant to be specific though general so no need to stuff it with every possible key word you want related to you.

Another thing to pay attention to: make sure the words are related in a general manner.

If they are drastically different it will lead to confusion as to why and how you have chosen to concentrate on these tools.

Yes, for the right person esoteric tools will seem like a positive thing, for the rest of the population it will only seem strange and confusing.

The goal is to be narrow but not so narrow that only one group in the world can hire you.

The next part of the sentence reinforces that you are capable and understand that a data scientist solves problems and generates insight.

It is a good idea to keep hammering this point as a worry of employers hiring well educated people is that they'll have their heads in the cloud most of the time and won't actually solve the business problems they are getting paid to solve.

The next part dives deeper into the particular industry and line of business that you are interested in.

Perhaps you are interested in startups and growth hacking, or in human resources and employee happiness, or in sales, customer growth and retention, or in logistics and supply-chain, or in improving customer satisfaction, or something else.

By further clarifying what you are interested in, you are helping yourself and the reader to understand where you are coming from and what the portfolio will contain.

The next and last part of the sentence gets into the nitty-gritty of what you do: the specific algorithms, statistical techniques, technologies, and math that you know and demonstrate in your portfolio.

Maybe you like using natural language processing (NLP) to understand customer service requests so in this section you would talk about using Java and the Stanford Parser or Google's Parsey McParseFace with SyntaxNet or using Python and spaCy.

Maybe you like to model the impact of marketing on customer acquisition, retention, and churn or to predict disease risk and susceptibility in patients using Scikit-Learn and Random Forests.

This will be inline with the theme of your portfolio and the types of data science jobs you have identified as being the ones you want.

This sentence is highly personalized to you so there is no wrong answer.

**Am I Pigeonholing Myself Too Much And Hurting Myself?**

No.

A perfectly rational fear is that by defining and labeling yourself you are limiting your job possibilities.

This fear is reserved for when you are just starting out.

When you first come to data science it does make sense to be open to anything.

What you'll find as you work through projects is that you'll have a natural inclination to certain types of problems using certain types of tools to achieve certain types of goals.

You must embrace this.

Put yourself in a hiring manager's shoes and imagine they have two candidates in front of them: one who is a generalist and one whose portfolio and all supporting application material are line with the hiring manager's business needs.

For the generalist the best case is that they may have at some point touched upon the hiring manager's business.

The average and worst case is that they haven't and are therefore clueless about what the manager needs.

Which means lots of risk for the manager.

On the other hand you are a candidate whose portfolio and all supporting application materials have and show many examples of knowledge and know-how in the hiring managers business.

Your represent a much smaller risk because you're already up to speed on many things and have shown interest in the highly specific things the hiring manager cares about.

So no, pigeonholing yourself is not detrimental to your career prospects, if anything it is necessary.

## 4.5 Actual Portfolio Introduction

**Introduce What The Portfolio Is About**

At this point the reader knows your name, the industry vertical you are interested in, specific parts of the industry you do work in, and some detailed information about your technical data science skills.

Now it is time to introduce your portfolio.

Why introduce your portfolio?

You introduce your portfolio to help the reader build a narrative about what they are going to find in your data science projects.

The first sentence communicates that the projects cover the data science process and gives a count of how many projects are found in your portfolio.

You will write something like:

> This data science project portfolio is composed of [number of projects] projects covering the full spectrum of the data science process.

Though slightly generic it covers exactly what is found further down the page.

Additionally, because you are focusing on doing small projects the number of projects will look impressively large.

By mentioning that the portfolio encompasses the full spectrum of the data science process you are communicating with the reader that you are aware that data science is more than just "fancy math and algorithms".

This sentence also sets up the following sentence, which is where you talk about the theme(s) of your work and what it will cover.

The portfolio theme sentence reiterates the themes, techniques, and technologies you introduced in your personal introduction.

What you are doing here is teeing those things that described you to your portfolio.

You'll write:

> This portfolio's projects apply [very broad programing tool or tools] within a data science context to provide value to [line(s) of business] lines of business using [targeted tools, algorithms, and techniques].

This is very similar language to your introduction.

One difference here is that you talk about applying the tools to provide value.

You want to communicate at all times that you understand you are there to further the organization's goals.

## 4.6 Portfolio Project Section

**Showcase Your Data Science Projects**

This section is the core of your Data Science Project Portfolio.

Here you will list all of the data science projects you have done.

Section 4.9 will look into greater depth of how to structure each individual project entry.

For now you will ignore how each individual project is written and concentrate on how all of your various projects are listed.

The most obvious way to list the projects is by date of completion – newest entry at the top and older entries at the bottom of the page.

This ordering makes sense in that as you do new projects you should be getting better and more knowledgeable so your audience should only really see your latest work to get a feel of where you are and what skills you have.

The downside to this organization is twofold: a) unless your last project uses everything you know it will make it look like you have big gaps in your knowledge and b) unless the last project was a full on end-to-end data science project it won't be clear to the audience what parts of the data science process you are familiar with and what you have done.

Thus, the organization that makes the most sense is to break up the projects based on where they fall within the data science process.

This way it is very clear what parts you are familiar with as well as provides a very clear structure to your work.

Further, because it is not organized by date, you will not run into issues with having to cram everything you know into the latest entry.

As a reminder the data science process is:

1. Import – Get the data from an original source
2. Store – Store the data into a data store
3. Extract – Get the data from the data store
4. Organize – Organize the data into a usable subset
5. Tidy – Scrub / Clean the data
6. Transform – Change data into something your program will understand
7. Visualize – Descriptive statistics exploration
8. Model – Statistics + Machine Learning + Experiment Design
9. Coding – Machine Learning + Algorithms + Code
10. Understand – Understand insights achieved
11. Communicate – What was the approach, what worked, what didn't work, what assumptions were made, what would you do different, why do insights matter

12. Next Steps – Where you go from here
13. Document – Code + Thought process + Replication steps

Where steps 8 & 9 encompass:

A. Feature engineering – Variables to use
B. Metric selection – How is model evaluated
C. Algorithm selection – What algorithms to use
D. Parameter optimization – Tune model/algorithm
E. Deployment – Get it into production
F. Evaluation – Did it work? If so, how well?

Is this overkill?

No.

The goal of your portfolio is to showcase what you can do as convincingly as possible.

Showing that you have (small) projects in each of these areas unequivocally communicates that you have the full process down.

This organization also lets people explore the parts of your knowledge they are most interested in.

The last part of this section is where you highlight end-to-end projects.

Whether they are standalone projects or projects put together from the small projects listed above the in the data science process section, it is helpful to show that you can do a full project as well.

Our recommendation is to forget about stand-alone projects for now and build something together from all of the small projects you did above.

In this way each project serves two purposes – the individual project and the stand alone one.

One way to do this is to come up with a standalone project and then break it down into the small component projects.

This enables you to work quickly and efficiently towards the completion of the large project.

## 4.7 Portfolio Project Tag Section

**Categorize Your Projects Based On Key Terms**

Each of your projects will use some of the same types of tools, techniques, data, data sources, business goals, etc.

This section of the overview page gets you to list each of those things in one specific place so that it is easy to see.

Using the InfoVis mantra, someone could click one of these tag links and see all of the projects where you use that one specific thing.

For example lets say that you used *dplyr* in a few of your data science projects to do data manipulation.

If you didn't know, *dplyr* is a tool for the R statistical programming language that helps you work with R data frames.

What you would do is "tag" the projects where you use this and add this tool to the list of tools in this section.

Each tool would be hyperlinked to a list of projects where you used this tool.

In this way the more technically knowledgeable people looking at your data science project portfolio would be able to see for each specific thing they were interested in how you had actually used it.

The tag lists should include:

- Tools
- Techniques
- Data Sources and Types

## 4.8 Your Contact Details Section

**Be Easy To Reach**

This section gives people a way to know whose work this is and a way to contact you.

Isn't this repetitive?

Yes and very necessary.

People are generally busy.

You want to make their lives as easy as possible.

If they want to reach out to contact you make if easy for them to do so.

Don't make then find a contact page or about page.

The first thing should be your name.

Then you want to include your email.

It can either be specific to your portfolio or your personal one.

Do not hide it behind a clever scheme – just write out your email.

Yes you'll get some spam but you don't want to make harder for someone getting to reach out to you.

Next is the web address of your data science portfolio which links back to this page.

The reason you put a link to your portfolio first rather than a personal website or GitHub profile is that you want them to think your portfolio is more important than your GitHub profile.

Additionally, if they are interested in your technical work it is better for you that they look at the portfolio rather than whatever projects you have floating around in your GitHub.

Then include your Twitter account if you are active and Tweet about data science.

Otherwise it's not helpful to have it on your portfolio front page.

Then include your personal website if you are active and write / blog about data science.

Otherwise is not helpful to have it on your portfolio page.

Lastly, a GitHub profile if you have one and are active in it.

GitHub goes last because it has the least organization and as you learned earlier you always want to come across as super structured in your communication, writing, and thinking for your data science work.

## 4.9 Project Specific Structure For Your Data Science Project Portfolio Front Page

It's time to dig deeper into how each project should be written up.

The goal for each project small-write up is quickly describe what you did while not overwhelming the reader.

The secondary goal is to get them to click into the full project write up.

Each small write-up needs to serve as click bait.

Further, even from this small description it should show what you know and learned from the project.

Finally, it should also show that you know how to communicate clearly by virtue of having summarized a project into a short readable description.

**Project Title**

The project title should be as descriptive as possible.

While you want to interest the reader so that they click to see the full project write up, it's important accurately represent the project.

A good structure is:

> "[Action verb] [tool and/or technique] to [data science process step] so that [business related reason]"

The best way to keep your data science projects small is to write this title sentence first, as it will aid you in keeping the projects focused and well defined.

The title should be hyperlinked to the full project report for ease of navigation.

## Project Description

Here you want to emphasize results and what you learned.

The explanation is a continuation and expansion of the title.

A good structure is:

> "[Action verb] [business /data science goal] using [action verb] [tool and/or technique] on [type of data] [name of data] to [data science process step]"

The best way to approach your small data science projects is to write the title sentence and description first, as it will aid you in keeping the projects small and well defined.

First come up with the title.

Then come up with the description.

Then outline the project.

This helps you get started quickly and easily.

## Project Tags

Project tags help a reader understand the data, technologies, and techniques used in a specific project.

Though they are treated as separate things in the actual protect write-up, here you will list all of them in the same place.

No need to spend time organizing them, as you shouldn't have too many tags per each project because all the projects should be small, self-contained, and hyper-specific.

Having too many tags is evidence of the project being too large.

The tags here and in the actual full write up should cover a) data and data type, b) technologies, libraries, and tools, and c) techniques (statistical, mathematical, and computer science / programming).

**Project Tag: Data**

You have three main things to communicate for the data tag.

First whether the data is labeled or unlabeled as this lets the reader know whether you'll eventually use supervised or unsupervised machine-learning techniques.

The second thing you should communicate is where the data comes from.

The third thing is what the data actually is.

**Project Tag: Technologies**

For the technologies tag you want to communicate from most basic to most esoteric.

This allows you to cover everything you need to cover without having to specify over and over again certain parts of the technologies.

For non-specialists, they should recognize the first or second tags listed here.

For the specialists they should recognize the more specialized tools.

If you were to list them with specialized technologies first, non-specialists would blank out and skip things they might recognize.

No need to spend time organizing them, as you shouldn't have too many tags per each project.

Having too many technology tags is evidence of the project being too large.

**Project Tag: Techniques**

Similar to the technologies tag, you want to communicate from most basic to most esoteric.

This allows you to cover everything you need to cover without confusing non-specialist readers.

No need to spend time organizing them, as you shouldn't have too many tags per each project.

Having too many technique tags is evidence of projects being too large.

**Order Of Projects In Each Data Science Step Section**

The order matters for each data science step section.

You want to put your best work first.

The more projects listed per step, the probability they are looked at decreases as you get to the end of them.

So put your best work first for each step and don't be afraid to switch and or replace projects as you do better work.

## 4.10 Data Science Project Portfolio Website Hosting

All of what you have covered and will cover assumes that you have or will setup a website.

Can you become a data scientist without having a website?

Yes, of course – many people have and will continue to do so.

We recommend having one because we've found that it really helps.

Not only does it help you learn, it also helps others get to know you better and at what you are capable of.

So how should you build it and where should you host it?

Up to you and your knowledge.

Our suggestions are as follows:

- Basic HTML to get started
- Static website over dynamic website
- Website focused only on your data science work
- You don't need your own domain name to start, but it can be helpful when sharing your work with others
- GitHub pages are a good place to start as it has myriad tutorials
- Don't learn a ton of web programming just to build your portfolio
- Keep it simple.

Remember that you are going to become a data scientist, not a web programmer, so spend your time accordingly.

Get something up as quickly as possible and improve over time.

# 5

# How To Choose The Right Projects For Your Portfolio

## 5.1 Project Creation Versus Project Replication

**Replicate Projects Not Create New Projects At The Beginning**

You are starting out in data science and are learning new mental models, new methods, new technologies, new approaches to problem solving, and new terminology all at the same time.

When you sit down to create a portfolio project it can be overwhelming trying to figure out where to start as well as what steps to do to complete a project.

This is because in addition to learning all the things listed above, you have to create the next steps based on your current knowledge.

New knowledge creation is very difficult.

At this point in your burgeoning data science career you want to focus on learning not proverbially "inventing the wheel".

A great number of things you need to learn how to do for a data science job have already been created, discussed, and written about.

Rather than seeking to create new knowledge, you should be seeking to replicate other people's knowledge.

**Learn By Solving Solved Problems**

You want to learn as quickly and effectively as possible.

The best way to do this is to work through problems that have already been solved and have their solution written up.

This allows you to see what steps someone included and in what order.

This approach forces you to work through someone else's thought process to solve the problem.

**How To Properly Replicate Someone Else's Work While Learning At The Same Time**

First off – when you replicate someone's work as one of your projects for your portfolio, you want to make it very clear it's a replication and what the original source was.

This lets people know that you know how to share credit.

This lets people know you are aware of work other people are doing.

Second – do not copy and paste from the work you are replicating.

You want to do one part at a time by hand.

The process should be as follows:

1. Read the section you are going to do next
2. Close the reading
3. Think in your head what you just read and learned
4. Try to do it yourself in your computer
5. If it doesn't quite work spend a minute or two working on it before looking at the original source
6. If it works great – spend a minute comparing and contrasting what you ended up doing with what you originally read and sought to replicate.
7. Once it works spend a minute summarizing what you learned from this step in the project report
8. Move onto the next section of the project you are replicating

Because you are replicating something that's already been done and spelled out for you there should be very few things that leave you stuck.

By using the steps above, if you do get stuck you'll be stuck on a small part of the project that you can reach out to the original author with the hyper-specific question or use a forum to get the answer.

Communication wise, replicating someone else's work will cause you to have to explain things in a slightly different way because you aren't copy and pasting from someone else.

This act of figuring out how to say something in your own words will not only help you master the material better and more effectively, it will also improve your communication skills.

Lastly, this strategy will be fruitful for your entire career as it teaches you how to replicate techniques and tools such that when you are in a job and are tasked with doing a new project, you can take what the community already solved and make it relevant to your group's particular situation and goals.

### Replicating Can Be Useful To Original Authors As Well

The original authors will be delighted to hear from you given that you thought highly enough of their work to replicate it.

One way you can provide value to them is to extend their work in a direction of your choosing.

This will give them some food for thought and is always appreciated.

Another thing authors find helpful is if you find a bug in their code or some spelling mistakes.

As long as you communicate with kindness, your help will be greatly appreciated.

### When You Should Stop Replicating?

Once you've replicated a good amount of work for a particular step in the data science process, you my start wondering when to generate your own work.

This is highly specific to each individual and their situation.

For you, since you are working on becoming a data scientist as quickly, effectively, and efficiently as possible, is best to leave the original work until later when you have more of an idea of what you are doing, what you are trying to do, and what you are going to be shooting for with a specific type of project.

**Finding Work To Replicate**

Between now and when you start creating original work, one of the more important things for you to do is to find highly useful and valuable work to replicate.

The great thing about data science and machine learning is that it is a very open field where you can find myriad examples, blog posts, conference talks, books, and videos.

As you work through projects your taste will develop and you'll have a better idea of what projects are worth your time replicating and which you can skip.

Furthermore, as you work through different projects, you'll also have a better idea of the people on the community and what types of things they are working on.

This is helpful for your network, knowing where to get help and advice if need be, and of course job hunting.

In the next sections you'll learn specifically how to think about choosing data science projects for your portfolio that will help you work towards your goals.

# 5.2 Data Science Getting Started Guide

**Find Projects You Would Do In Your Future Job**

In the *Data Science Getting Started Guide* you found many jobs that you were interested in.

Then you took copious notes on the pairwise comparison winners.

You noted what tools they wanted.

You noted what data they wanted.

You noted what machine learning techniques they wanted.

You noted what programming languages and libraries they wanted.

You even noted what business problems and goals they wanted solved.

You want to do a small project for each thing mentioned above so that you can have familiarity with it, knowledge for how to do it, and why it is important to that particular type of data science job.

Once you have all of your notes detailing the different things hiring managers mentioned they wanted, you should combine it with the data science process to see where potential portfolio projects would fit.

# 5.3 Data Science Process

### Close Knowledge Gaps In The Data Science Process

Recall that the data science process is as follows:

1. Import – Get the data from an original source
2. Store – Store the data into a data store
3. Extract – Get the data from the data store
4. Organize – Organize the data into a usable subset
5. Tidy – Scrub / Clean the data
6. Transform – Change data into something your program will understand
7. Visualize – Descriptive statistics exploration
8. Model – Statistics + Machine Learning + Experiment Design
9. Coding – Machine Learning + Algorithms + Code
10. Understand – Understand insights achieved
11. Communicate – What was the approach, what worked, what didn't work, what assumptions were made, what would you do different, why do insights matter
12. Next Steps – Where you go from here
13. Document – Code + Thought process + Replication steps

Where steps 8 & 9 encompass:

A. Feature engineering – Variables to use
B. Metric selection – How is model evaluated

C. Algorithm selection – What algorithms to use
D. Parameter optimization – Tune model/algorithm
E. Deployment – Get it into production
F. Evaluation – Did it work? If so, how well?

What you'll want to do now is figure out what small projects you can do with the key words from the pairwise-comparison-winning jobs and the specific steps in the data science process.

Remember that you are setting up the structure for your data science portfolio home page based on the data science process so it should be evident to you what gaps in your list you have to fill.

You also want to choose projects that will strengthen your weak skills.

Later when you've filed out all the gaps, you'll want to start again by adding even stronger projects to specific sections where you feel there still exists some weakness.

As you develop your skills you'll find you have a tendency to do very well in some areas.

You should also do projects in these areas to make your strengths even stronger.

## 5.4 End-To-End Overall Project

**The Long Run**

Once you've filed out the various parts of your portfolio, it is time to think about doing an end-to-end project.

In the long run, you'll want to become a senior data scientist.

A big part of being a senior data scientist is the ability to structure, execute, and communicate an end-to-end data science project.

By starting to do and practice this skillset now, you will experience doing a full project.

You can do a full replication project or tie together many of the data science process steps into one larger project.

This type of project, as long as it's based on the pairwise comparison jobs and follows the data science process steps, will be very powerful for your job search and interviews.

This type of project will show you can do independent thinking.

You will build experience putting all the steps and pieces together.

If it works, terrific!

If it fails, you have achieved some insight, experience, and perspective into a failing data science project.

Which is valuable for a future employer because you will have matured and developed opinions and thoughts about the right ways to do things and how things can go wrong.

# 6

# Project Specific Write-up For Your Data Science Project Portfolio

## 6.1 Why You Should Use A Specific Consistent Structure For Your Data Science Project Write-ups

**Build Once, Use Multiple Times**

You want to be as clear as possible to readers of your portfolio.

To ensure clarity it is imperative that all of your work uses the same structure.

This way, after the reader reads one of your projects, they will fully understand how all of your other projects are written out.

**Save Time And Energy**

By having a consistent structure and template for your project write-ups it will save you time and energy when starting a new project.

Rather than starting with a blank page you'll find it easy to start each and every project because you already have a set of sections and paragraphs to fill out.

### Does The Same Structure Really Work For All The Different Projects?

Yes.

Because you are doing very focused small projects you will have to convey very similar information – scope, goals, results, references, etc.

### Templates Evolve As You Do

As you develop your knowledge, skills, opinions, and tastes in how to communicate the data science work you have done, your taste and structure will evolve.

This is great news – it means you have strong opinions about something with data to back it up.

What follows is a structure we recommend from having read hundreds of data science articles, posts, and project write-ups.

Start with this structure and then make it your own.

You will now look at how to individually structure and write up each data science project for your portfolio.

## 6.2 Title

**Use The Same Title As The Project Title On The Portfolio Front Page**

The first reason for the title resuse is that you already spent time and energy devising a great title so you should reuse it.

The second reason is that when a reader clicks from the front page and lands at this portfolio write up page, it will be extremely confusing if the page title and the link they clicked on are not the same.

The third reason comes from copywriter's advice – the job of the first sentence is to get you to the 2nd sentence.

If the title and description were good enough to cause someone to click into the project, it's in your best interest to keep the interest going.

## 6.3 TL;DR Paragraph

**Use The Same Description Paragraph As The Project Description Paragraph On The Portfolio Front Page**

Like the project title, you've already thought about it so you should use it again.

Though it may seem repetitive there will be many people who land in a particular project rather than the portfolio front page.

This brief intro-paragraph will help them figure out if the rest of your work is worth reading.

## 6.4 Long Description

**Two For One**

This long description is both for the reader and you.

For the reader, it goes into more depth of what you did.

For you, it gives you a chance to write out a longer version description of what your project entailed.

Think of the small description as a 10 second snippet you would share with someone else.

This longer description is your 1 to 2 minute version of what you did.

This is very helpful to think through for when people ask you what you have done in either social settings or interview settings.

This means you'll have to cover the following things:

- Data overview
- Technologies overview
- Techniques overview
- Software overview
- What you learned
- How it applies to business

This covers all the skills and thought process you learned through this particular project.

### Write Your Long Description After You Finish Your Project

Because of the amount of information this description covers it's important to leave it until the end of the project.

This allows you to truly reflect on what you did and clearly communicate what you learned.

Note that this is the opposite of the portfolio home page paragraph which you should write first as you want to limit the scope of the project by setting up the goals at the beginning before you dive into doing work on it.

## 6.5 Table Of Contents

**People Get Lost Easily**

If your project report is longer than 2 paragraphs people will get lost.

Giving them a very clear intra-page linked table of contents lets them quickly navigate.

Is it overkill?

No, because the Table of Contents is really a very high level outline of what you did.

Note that the Table of Contents also helps you when you start the project because it serves as an outline for what you are about to do.

**Readers Skims**

Most readers will skin to get the basic gist of what you did.

Readers do this by reading section headings as they scroll down the page.

You save the reader the pain and hassle of doing that process by aggregating all of the section headings into an easy to use table of contents.

**Outline Your Project Before And During**

By starting with the table of contents after writing the front-page title and the front-page small description paragraph you are setting yourself up for success.

First, you choose the project to replicate.

Then you make a brief outline of all of the steps.

And – Bingo!

You now have your table of contents.

Then, as you are doing the project, you continue to refer to and update the outline as need be so that it's clear what sections you have and what you are doing.

This keeps the table of contents up to date and helps keep you on track as you work through it.

**An Example From Data Scientist Sebastian Rashka**

A great example of this comes from an article on PCA (principle component analysis) by data scientist Sebastian Rashka.

The link to the permanent saved website (in case the original gets lost) is here:

http://web.archive.org/web/20160823061830/http://sebastianraschka.com/Articles/2015_pca_in_3_steps.html

The table of contents for this particular article is as follows:

- Sections
- Introduction
    - PCA Vs. LDA
    - PCA and Dimensionality Reduction
    - A Summary of the PCA Approach
- Preparing the Iris Dataset
    - About Iris
    - Loading the Dataset
    - Exploratory Visualization
    - Standardizing
- 1 - Eigendecomposition - Computing Eigenvectors and Eigenvalues
    - Covariance Matrix
    - Correlation Matrix
    - Singular Vector Decomposition
- 2 - Selecting Principal Components
    - Sorting Eigenpairs
    - Explained Variance
    - Projection Matrix
- 3 - Projection Onto the New Feature Space
- Shortcut - PCA in scikit-learn

This very clearly lays out what we'll be covered and how.

## 6.6 Project Background

**Where Are You And What Are You Doing**

As you start the project, paint a picture for the reader of where you are and what you are doing.

Start as close to the action as possible of where you will begin rather than doing a long history.

For instance, let's say you are doing a natural language processing project of removing "stop words" from tweets.

First – assume the reader had no idea about what any of the above mean.

Second – don't start with how NLP (natural language processing) came to be and early research results.

Instead start with a basic description like:

> Natural Language Processing (NLP) uses algorithms to process human language in order to understand things. For this particular project, [Title of your project], I will remove "Stop Words" from the tweets in the [dataset name/description]. "Stop Words" are very common words that appear in almost every tweet and thus provide little value. Words like "the", "a", "it", and others fall into this category.

Notice how this gives you enough background to have a sense of where this project is going.

This gives the reader some key terms that will come up in the project.

## 6.7 Step-by-step Explanation

**True Mastery Is The Mastery Of Fundamentals**

Your portfolio is a tool that sells your capabilities and your knowledge.

Don't mess it up by skipping steps or being sloppy with what you write.

Go step by step to show what you did.

Over communicate at all times.

Go through every single step even if you think it's obvious and not worth writing down.

This will show how much you know and are able to communicate.

This will also shows your thought process which readers love to get an insight into.

Remember that this is not an academic paper it is a portfolio.

So make it clear what you learned and what you would do differently if you were doing it again at every step.

This also gives the reader insight into your data science opinions.

Lastly, as you are going through and writing the entire process, remember to include section headings for easy skimming that match up with the "Table of Contents".

# 6.8 Conclusion

### A Note To Your Future Self

Conclusions are hard to write because they are written at the end of a long and laborious process.

What normally gets written in this section is nothing more than a regurgitation of the introduction.

While this has some value, it is important to go above and beyond, especially if it only takes an extra 10 to 15 minutes of work.

You need to write your conclusion as a note to your future self.

Start out with the introduction.

Then talk about any interesting references you found while doing the work.

Talk about any "today I learned" epiphanies you had while completing the project.

Then have a link to the GitHub or Jupyter code the reader can follow.

Then talk about mistakes you made.

This keeps you from making them in the future.

And lastly, talk about what you will do differently the next time you do this type of project.

This style of conclusion to a project write up is incredibly useful to your future self.

And, not surprisingly, it is incredibly useful to future employers and future colleagues because they have an open window to what you did, what you learned, what mistakes were made, and what you will do differently the next time you do a project of this type.

## 6.9 Related Projects

### Multiple Projects For Each Data Science Process Step

You will eventually do many small projects for each of the data science process steps.

### Help The Reader

If the reader is interested in a project you did on a particular step then they may be interested in other projects you have done in the same data science process step.

The way to help the reader is to help them find what they want easily.

Hence you should include in this section a list of titles and links to other projects from the same data science step.

## 6.10 Your Bio

**Sell Yourself**

You did great work.

Now that the reader has finished reading your project write up, give them a brief reminder of what you do and who you are.

**Reuse Front Page Bio**

You want to have an awesome bio that you use everywhere.

This makes it so that when you update it and improve it, you can paste it everywhere.

Since you already wrote a great bio for the front page of your Data Science Project Portfolio, you should use it here as well.

## 6.11 Contact Information

### Make It Easy To Contact You

Just like the information on the front page of your portfolio, you want to be as clear as possible for how to contact you.

Because you already thought about what to put on the front page, just use the same information and write up.

7

# Conclusion

## 7.1 Own It

**You Are A Data Scientist**

Yes, most people will only treat you as one when you have the actual job title.

However, being a data scientist is a multi-disciplinary approach you choose to use in your life.

So choose to use it and use it.

The closer your practice is to the real thing, the better results you will get.

Hence, if you decide that you are a data scientist and start doing data science projects, you will take it more seriously.

# 7.2 What To Do If You Get Stuck

### Ask For Help

As a first step break your learning steps down further and focus on doing even smaller projects.

If this doesn't work or you find that you aren't progressing because you are stuck, reach out to the person whose work you are replicating.

Tell them what you have done and where you are stuck.

Ask, very nicely, if you have missed a step.

Most people will reply and if they don't, by formulating a question you will have been forced to think through what your problem is and where you are stuck.

Which means you can then try search engines as well as asking on different forums for help.

Because you can show what work you've done, where you are stuck, and what the next step should be, you'll be able to find help easily and quickly.

Remember, as a beginner your focus should be on replicating, not creating, so you shouldn't run into too many projects where you fail to make any forward progress.

# 7.3 How To Know If You Are On The Right Track

### One's A Dot, Two's A Line, and Three's A Trend

The first project you do for any one of the data science steps, you'll be learning 100% of the time.

The second project you do for the same step, you'll start to notice similarities in approaches, steps, and what gets done.

By the time you do your third small project for the same step, you'll automatically start to guess what is going to come next.

This is where you'll know that you are on the right track – when you develop an intuition on what the next steps should be.

And because of the way you are doing the replicating of projects – read, close source, think about it, and try to do it without the original source – you'll already have been practicing predicting what is next.

This is very valuable because you are encoding the thought process into your mind so that when you replicate other projects for the same step or when you create your own projects from scratch for this step, you'll be able know what to do intuitively.

Hiring managers and future data science colleagues will be able to spot this from a mile away and will love you for it.

## 7.4 Have Fun And Good Luck

You are learning data science because you want to.

If you stop having fun or find that it's hard to get motivated to start, figure out where the fun starts and where it stops.

Then focus on more fun.

Yes, eventually most of your data science projects in the portfolio should match up as close as possible to your dream data science job, but in the mean time, make sure to take time out to have fun and do fun projects.

As they say, the only loosing choice is not to play.

So have fun and best of luck!