

# **Data Science**

# **Getting Started Guide**

*By Sebastian Gutierrez*

[www.DataScienceWeekly.org](http://www.DataScienceWeekly.org)

# Table of Contents

---

## **Chapter 1: Introduction ..... 4**

|  |          |
|--|----------|
| <b>1.1: How To Use This Guide.....</b>   | <b>4</b> |
| <b>1.2: How To Decide Where You Are Going .....</b>  | <b>6</b> |
| <b>1.3: The Process To Figure Out Your Destination Process And Answer All Of Your Questions.....</b> | <b>8</b> |

## **Chapter 2: What are Data Scientists and What Does A Data Science Job Actually Entail ..... 14**

|  |           |
|--|-----------|
| <b>2.1: What Is A Data Scientist?.....</b>   | <b>14</b> |
| <b>2.2: What Do Data Scientists Do?.....</b> | <b>19</b> |

## **Chapter 3: The Pairwise Comparison Process ..... 24**

|  |           |
|--|-----------|
| <b>3.1: The Pairwise Comparison Procedure Description .....</b>  | <b>24</b> |
| <b>3.2: Check Out Eight Different Job Posting Websites.....</b>  | <b>25</b> |
| <b>3.3: Look At Each Website's Offerings For Data Science Jobs.....</b>  | <b>26</b> |
| <b>3.4: Do Pairwise Comparison Between The Jobs On The First Results Page All The Way Through To The Bottom Of Each Page For Each Website.....</b> | <b>27</b> |
| <b>3.5: Do Pairwise Comparison Between The Jobs That Won Round 2 .....</b>   | <b>31</b> |
| <b>3.6: Do Pairwise Comparison Between The Jobs That Won Round 3 .....</b>   | <b>32</b> |
| <b>3.7: Extract Information From The Winners Of The Third Round .....</b>  | <b>34</b> |
| <b>3.8: Analyze Information From The Winners.....</b>  | <b>36</b> |
| <b>3.9: Create A Plan For What You Need To Learn And Do.....</b>   | <b>37</b> |

**Chapter 4: What To Look For During The In-Depth Analysis ..... 40**

|   |           |
|---|-----------|
| <b>4.1: Your Wants And Their Wants .....</b>  | <b>40</b> |
| <b>4.2: Analysis To Do In Each Step .....</b> | <b>43</b> |
| <b>4.3: Step Specific Analysis .....</b>      | <b>44</b> |
| <b>4.4: Intangibles And You .....</b>         | <b>51</b> |

**Chapter 5: Knowledge Gaps That Matter And How To Develop An Action Plan To Close Them ..... 56**

|   |           |
|---|-----------|
| <b>5.1: Knowledge Gap Evaluation .....</b>    | <b>56</b> |
| <b>5.2: Knowledge Gap Closing .....</b>       | <b>58</b> |
| <b>5.3: Planning Action Plan .....</b>        | <b>59</b> |
| <b>5.4: Constructing Action Plan .....</b>    | <b>61</b> |
| <b>5.5: Doing Action Plan In Public .....</b> | <b>65</b> |

**Chapter 6: Conclusion ..... 67**

|  |           |
|--|-----------|
| <b>6.1: Next Steps .....</b>                               | <b>67</b> |
| <b>6.2: What To Do If You Get Stuck .....</b>              | <b>67</b> |
| <b>6.3: How To Know If You Are On The Right Track.....</b> | <b>69</b> |
| <b>6.4: Have Fun.....</b>                                  | <b>70</b> |
| <b>6.5: Discipline .....</b>                               | <b>70</b> |
| <b>6.6: Good Luck .....</b>                                | <b>71</b> |

# Introduction

## I.I How To Use This Guide

---

The purpose of this guide is to help you become a Data Scientist faster than you think possible.

It will give you the building blocks necessary to figure out what you need to learn as well as what you don't need to.

In addition to using this guide right away, it's written and composed in a way that will be useful for the rest of your life.

The information you learn within will continue to be applicable as your career progresses.

Read all the way once through to understand process and thought process behind it

While you can skip doing the work, it won't help you as much as if you take the time and effort to read all the way through and internalize not only the how but the why.

**As you progress in your data science career, keep coming back to sections that will help you understand what you want, what employers want, and how to build a plan that takes you from A to B**

As you grow up, mature, and learn new things your viewpoints, wants, and needs will change – sometimes dramatically and sometimes slowly.

Change is normal and welcome.

What that does mean however is that you will outgrow your role and/or your company.

When that happens, as you start to feel the need for a change in your role, it is important to take a step back and understand what it is that you want, what it is that employers and job you would want to have want, and build a plan to get you from where you are to where you want to be.

**The payoff is greater than the amount of work you put into the process**

The payoff for going through this guide and doing the work will be greater than the amount of work you put into it.

In the absolute worst case you realize you don't want to be a data scientist and while sad because you bought this guide to help you become one, at least you won't have to go through all the work and pain to become one only to realize that it's not what you wanted.

In the best case, not only do you find an awesome role with a terrific company in an exciting field, it'll set you up nicely for a great career

progression that will keep paying dividends (monetary and psychological as well).

Before getting into the particulars the various processes described in this book it's important to point out while work is rewarding and you may find yourself spending 8 or many more hours a day at the office, the time away from the office during weekdays and weekends is equally if not more important.

If you're not happy at home, then you'll have a hard time staying happy at work.

So when you do this work, remember that you are looking for a job and career that matches your outside-of-work life as well.

=)

Remember – work works, get your pen and paper out, and have a blast!

## I.2 How To Decide Where You Are Going

---

Before you begin this awesome journey of learning data science to become a data scientist, it's worth covering a few things that will help you get to where you want to go faster.

First of, let's talk about knowing where you are going:

## Your destination

*“Would you tell me, please, which way I ought to go from here?”*

*“That depends a good deal on where you want to get to,” said the Cat.*

*“I don’t much care where—” said Alice.*

*Then it doesn’t matter which way you go,” said the Cat.*

- Lewis Carroll’s Alice’s Adventures in Wonderland, Chapter 6

The purpose of using this guide is to establish a destination as well as a path from where you are now to where you want to go.

With a destination, it will be much easier for you to decide what to learn, how to learn it, when to learn it, and why to learn it.

Through this process you’ll figure out what exactly you need to learn as well as what you can ignore.

Rather than being dismayed by the gargantuan number of options of things to learn to become a “Real Data Scientist”, you can focus exclusively on the only things you actually need to know.

Because it is a focused list of things you need to learn, you will be setting yourself up for success in becoming a data scientist, being prepared to apply for the specific types of data science jobs that you would want, and be able to follow the path to where you need to go.

By doing this research ahead of time, you’ll minimize the amount of time you spend on false paths.

Further, you can set yourself up for success with small achievable wins that help you understand when you are and aren’t on the right track.

From here you'll now know your goal of what to learn, so the next step is to think about how to achieve that goal.

### **How to start the journey to your destination**

In front loading the research and work to better understand what you want out of a data science job and what they will want from you, you can then work backwards.

That is, since you know exactly what you need to be able to show to a potential employer, you can look for educational courses/materials that provide an outline for mastery of those specific topics.

Since educational courses/materials are put together with the goal of the student achieve the education outcome, it means that you'll have a plan for what to learn and when.

## **I.3 The Process To Figure Out Your Destination Process And Answer All Of Your Questions**

---

In order to develop a coherent learning path based on your highly specific knowledge gaps, it is necessary to understand what you need to learn.

However, because data science is currently so such an amorphous term and field, depending on what industries you work in, what companies you work for, or even what departments you work in, the data, tools, technologies, and platforms you work on could vary drastically.

As an example, Data Science London (<http://datasciencecelondon.org/>), a non-profit organization dedicated to the free, open, dissemination of data science (and the largest data science community in Europe with more than 5,770 data scientists and data professionals), suggests this is the data scientist's toolkit:

## [Sort of a] Data Scientist Toolkit

- Java, R, Python... (bonus: Clojure, Haskell, Scala)
- Hadoop, HDFS & MapReduce... (bonus: Spark, Storm)
- HBase, Pig & Hive... (bonus: Shark, Impala, Cascalog)
- ETL, Webscrapers, Flume, Sqoop... (bonus: Hume)
- SQL, RDBMS, DW, OLAP...
- Knime, Weka, RapidMiner... (bonus: SciPy, NumPy, scikit-learn, pandas)
- D3.js, Gephi, ggplot2, Tableau, Flare, Shiny...
- SPSS, Matlab, SAS... (the enterprise man)
- NoSQL, Mongo DB, Couchbase, Cassandra...
- And Yes! ... MS-Excel: *the most used, most underrated DS tool*



You can see that the learning path is gargantuan at best and confusing at worst.

To solve that issue, the key to this guide is to do pairwise comparison of hundreds of jobs to figure out what types of jobs appeal to you and what those jobs entail.

### The pairwise comparison procedure overview

Before we talk about the why, how, who, when, where of the process and why it'll help you in your data science journey, let's nail down what the "pairwise comparison procedure" does.

From Wikipedia:

*"Pairwise comparison generally is any process of comparing entities in pairs to judge which of each entity is preferred"*

Pretty simple – you compare two things and decide which you prefer.

What you are going to do is do a pairwise comparison between two different data science job postings to see which you prefer.

Not what specific things you prefer from one or the other, which job in total you prefer between the two.

And to make sure you don't second guess yourself, you'll be doing these types of comparisons across a few hundred jobs.

This ensures that rather than hemming and hawing over which one you prefer, you can quickly figure out which you prefer.

### The goal of doing the pairwise comparison procedure

The reason you do the comparison between two jobs is to keep the mental load light – rather than thinking through hundreds of possibilities, you only have to take into account option A and option B.

Then after you've done several hundreds of these comparisons, you can dig deeper into all of the choices that won out.

By analyzing these choices, you'll be able to notice patterns that you may not have been aware of or were not comfortable being truthful to yourself about.

### **The reason why doing the pairwise comparison procedure works**

What this process is trying to illicit from you are your revealed preferences for what excites you about potential data science jobs.

Instead of asking you what you want, this process draws out of you what it is that you find interesting.

And because of the large number of pairwise comparisons you do, you won't get sidetracked or enchanted by any single data science job advert.

### **Why it works specifically for you**

Because you are you and unique in your own special way and because you'll be the one making the pairwise comparison decisions over and over again, it'll be 100% reflective of your individual choices.

As you grow and mature in your career, you'll be able to do this process over and over again as your tastes and wants change.

When you do this again in the future, it'll be useful for where you currently are in the scope of knowledge, skills, and goals.

## Choices and why you need to work backwards

Working backwards is highly suggested.

In thinking about what you want and how you are going to achieve it, you have two choices.

A – you can come up with the dream job in your head and then try to find it in the real world.

B – you can find the best possible job for you in the real world and then work to get it.

B is the more pragmatic because this job already exists and someone out there is trying to fill it.

Working backwards allows you to do option B.

## See what data scientists do and choose what things about those jobs are interesting to you

This is why rather than starting at the more common place people suggest you start – learning math, programming, statistics, machine learning, etc., this guide starts with jobs adverts.

You want to “figure out” what you want from the choices that already exist.

Once you have the jobs in mind that you like, you can develop your own list of things that excite you as well as don’t excite you.

To fully appreciate why this is the most pragmatic way to approach this, let's discuss what a data scientist is and what they do.

# 2

# What are Data Scientists and What Does A Data Science Job Actually Entail?

## 2.1 What Is A Data Scientist?

---

You start by asking, “what is a data scientist” because if you are going to learn how to become one, it’s important to understand what one is.

First, popular culture definitions.

### **Popular culture definitions**

"A data scientist is someone who is better at statistics than any software engineer and better at software engineering than any statistician."  
– Josh Wills

“A Data Scientist is statistician who lives in Silicon Valley / San Francisco”

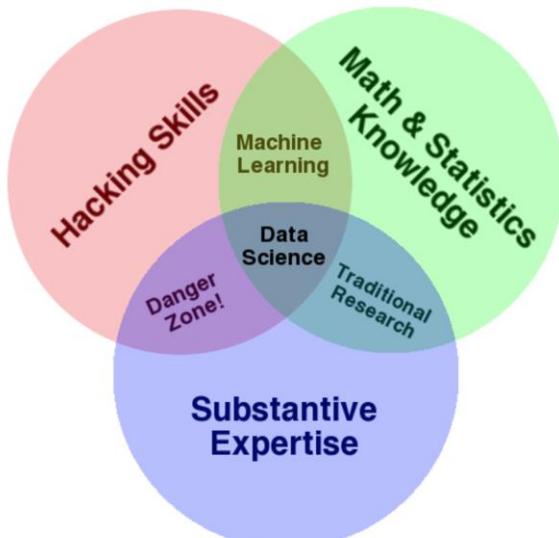
“A Data Scientist is a statisticians who calls themselves a Data Scientist to get a job”

“A Data Scientist is someone who does statistics on a Mac.”

“A Data Scientist is someone derives inference from data too big to fit on a single computer”

These definitions don't tell you what a data scientist is.

The standard fall back answer to “What is a Data Scientist” or “What is Data Science” comes from Drew Conway's 2013 article titled “The Data Science Venn Diagram” which shows the following picture:



<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>

Using this picture, Drew Conway says, “I hope this brief illustration has provided some clarity into what data science is and what it takes to get there. By considering these questions at a high level it prevents the

discussion from degrading into minutia, such as specific tools or platforms, which I think hurts the conversation.”

Which is terrific in that it’s pretty clear that while yes, statistics is involved, one doesn’t need to live in San Francisco or have a Mac.

Data Science is the intersection of “Substantive Expertise”, “Hacking Skills”, and “Math and Statistics Knowledge”.

This is helpful, but it doesn’t really tell you where to begin your journey or what MOOCs / Schools / Bootcamps / Books / etc. to start reading and studying.

To simplify things even further and really nail the definition once and for all, let’s look at the most pragmatic definition there is.

### **Most pragmatic Data Science definition**

“A Data Scientist is a person who holds the job title of Data Scientist.”

Is that really the best definition?

Yes.

But it’s tautological.

For sure!

But that’s the way it is.

Yes a person may have hacking skills, expertise, as well as math and statistics knowledge, but until that person is hired and given a title of a Data Scientist, they are not a Data Scientist.

## What about the technology?

You may be thinking to yourself – well, what if a Data Scientist:

- Uses Excel to analyze datasets that are less than 5 MB in size?
- Doesn't use Hadoop?
- Uses SQL instead of NOSQL?
- Doesn't use a Data Lake or even know what one is?
- Etc....
- Etc....
- Etc....

None of these questions matter to how you define a “data scientist”, because they all fall under the “No True Scotsman” information fallacy.

## The “No True Scotsman” fallacy

From Wikipedia:

*Philosophy professor Bradley Dowden explains the fallacy as an "ad hoc rescue" of a refuted generalization attempt. The following is a simplified rendition of the fallacy:*

*Person A: "No Scotsman puts sugar on his porridge."*

*Person B: "But my uncle Angus likes sugar with his porridge."*

Person A: "Ah yes, but no true Scotsman puts sugar on his porridge."

You can see how this applies to defining what data scientists do and what tools and technologies they use.

If you run into a person who says that real Data Scientists don't use Excel, please ignore them.

## **Your true north**

Which brings us back to learning Data Science and becoming a Data Scientist.

Because the best definition of a Data Scientist is someone who holds the job title of "Data Scientist", it means that to learn and become one means getting a job as one.

Which really means that everything you must learn and do will be centered on becoming hirable as a Data Scientist.

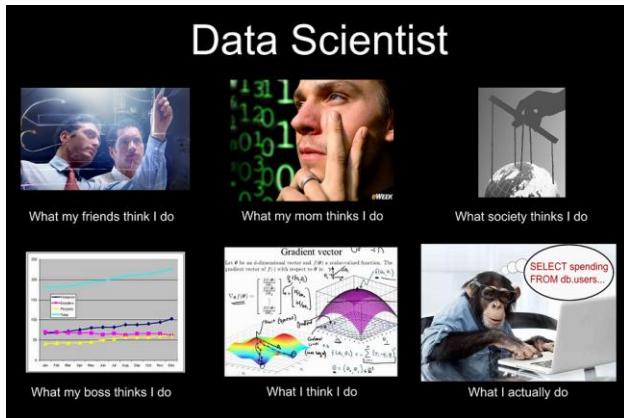
Becoming hirable as a Data Scientist will serve as your "True North", that is, everything you learn and do will be based on whether or not it will help you land a job as a Data Scientist.

Now that we've figured out the definition of a Data Scientist and why it's important to you in your quest to begin learning data science, let's look at what data scientists actually do.

## 2.2 What Do Data Scientists Do?

---

You may have seen this graphic before:



Source: <http://www.sintetia.com/wp-content/uploads/2014/05/Data-Scientist-What-I-really-do.png>

Depending on whom you ask and how you ask it, what a data science job entails will vary drastically.

Per the picture, even someone on the job, who is already a Data Scientist, will have many different ways to explain what the job entails.

When you are looking to get started in data science, it can be very confusing what you need to learn and how in-depth you need to learn the various subjects.

You may think that you can look at LinkedIn for guidance, as that is where you can see what real people have written about their data science jobs.

### **LinkedIn profiles can mislead**

On LinkedIn, you will find very flowery and jargon-y language about what different data science roles entail and what people do.

This is because LinkedIn is a public resume so people have to showcase the best parts of their job, as they never know who is going to be looking at their profile.

Further, because it is self-written, they will be projecting what they want others to think they are doing rather than what they are actually doing.

Which means that parts of what they actually do and how they do it are omitted from the description.

So when you look at data science roles and how the people who hold those roles describe themselves on LinkedIn, you will only be seeing only a small part of what the role entails.

Which means it is not helpful to figure out what you need to learn to become a data scientist.

### Recruiters can mislead

When communicating with recruiters, you'll find they use very flowery and sometimes even dishonest language about what specific roles entail.

This is because while they do get paid when some gets a job, they understand that it's partial luck at the end of the day, so they need to take a more shotgun-like approach to the situation.

Their job entails getting as many candidates as possible to sit down with hiring managers in order to be able close the deal.

To do this, they need to make the job sound as appealing as possible so that they can send as many people as they can to the top of the job interview funnel.

Which means that it's not helpful to speak with them about what you need to learn in order to become a data scientist.

### **People at meetups, conferences, and events can mislead**

Similar to LinkedIn and recruiters, most people at live events are going to be predisposed to presenting their best / most exciting selves to others.

Which means that when you meet people in events to try to figure out what they are doing, what you should be doing, and how you can get started in data science, you'll have to take everything they say with a grain of salt.

The information you learn will be filtered through people's egos and personalities.

Given all of this, how do you actually figure out what a data science job actually entails and what you need to learn?

### **Job advertisements are your best bet**

The one place where people cannot be hand-wavy are job advertisements.

A hiring manager runs into a problem they need helping solved so they decide to hire someone to help them solve the problem.

From a hiring manager: “I have a problem I’m trying to solve, and I’m looking for someone to help me solve it.”

They write up a job advert that describes who they need and want and what the person will be doing.

Barring some run-in with Human Resources about how to properly word it, the hiring manager needs to be as direct and concrete as possible because if they are not then they will have trouble finding someone to help them solve the problem they need help solving.

The hiring manager will be very specific on what educational background is a hard requirement and a soft requirement.

The hiring manager will spell out what technologies need to be mastered and for what purposes.

The hiring manager will spell out the various statistical, mathematical, machine learning, and data science techniques that need to be part of a candidate’s experience.

The hiring manager will spell out what the job entails and what the candidate will be doing for the manager, group, and company.

Which means that data science job adverts will, in aggregate, give you the clearest picture of what data scientists need to know and learn to be able to do the job.

And, since a data scientist is someone who has the job title of data scientist, this is the clearest way to figure out what you will need to learn.

Which means you can figure out the plan for how to get started learning it.

The next chapter will take you through how you can use job advertisements to focus on what you like and what those jobs entail.

## 3

# The Pairwise Comparison Process

## 3.1 The Pairwise Comparison Procedure Description

---

**A brief overview of how to do the pairwise comparison procedure**

By now you should have a hand-wavy understanding of what the process will entail – you look at a few hundred data science job postings and decide between a bunch of them on which you like better.

The steps will be the following:

1. Check out eight different job posting websites
2. Look at their offerings for data science jobs
3. Do pairwise comparison between fifty jobs returned in the search results
4. Do pairwise comparison between the jobs that won round 2
5. Do pairwise comparison between the jobs that won round 3
6. Extract information from the winners of the third round
7. Do an in-depth analysis of the information from the winners
8. Create a plan for what you need to learn and do based on what you learned in step 7

Let's now go through the steps one by one to flesh out the idea behind it, what you'll be doing, and what you will achieve at each step.

## 3.2 Check Out Eight Different Job Posting Websites

---

You will use the following US-based job websites for this process:

- <http://www.careerbuilder.com>
- <http://www.indeed.com>
- <https://www.linkedin.com>
- <http://careers.stackoverflow.com>
- <http://www.glassdoor.com>
- <https://angel.co>
- <http://www.startuphire.com/>
- <http://www.monster.com/>

Note that if you are not based in the US and/or are interested in non-US data scientist roles, these websites will still work for you.

This is because you are not looking for specific jobs to apply to; you are looking to learn about what you want and what they want.

Later you can take those learnings and find jobs specific to the locations you are actually interested in.

These eight websites were chosen to make sure there is a large selection of jobs across industries, company sizes, company maturities, and postings.

### 3.3 Look At Each Website's Offerings For Data Science Jobs

---

Though it will vary across websites, you'll want to search for jobs that have the title of "data scientist".

A way to do this is to put quotes around the two words so that the search engine of each website knows that you are looking for that specific phrase in the job posting.

#### **Modifiers for your specific situation**

Some of the websites allow you to put in a postal code (zip code) or location to narrow down your search. This can be an interesting way to drill down into the types of data science jobs you think you may be interested in.

The only word of warning here is that because some industries (at least in the US) are very heavily skewed location wise, if you do decide to narrow down your search by location, you will find some industries more prevalent than if you had done a site wide (country wide) search.

For instance, if you look at New York City, you will probably encounter a great deal more "data science in finance jobs" than if you had looked at Seattle as your location.

#### **Job search results**

Once you have searched for “data scientist” in each website, the website will return to you a paginated search results list.

Sometimes these are ordered by date or by perceived relevance of the specific phrase you were looking for and how closely it matched the job posting.

For the purposes of this guide, you don’t care how it’s ordered; all you care about is that you now have a list of jobs that you can explore.

For this process you are looking for at least fifty jobs in the results so that you have a large amount of data to work through.

### 3.4 Do Pairwise Comparison Between The Jobs On The First Results Page All The Way Through To The Bottom Of Each Page For Each Website

---

Now that you have a results page for each of the websites, you have a list of data science jobs to compare.

While you could automate this process, it’s recommended you do it by hand to ensure you are 100% present and focused on this very important part of the procedure.

#### Where you will save your work

Create a folder (where it's backed-up regularly) to save your work.

As you go through this process, you'll be working with a few different text files.

Within this folder, create a text file where you will save your results of this first step of your work.

Name this file, *pairwise-comparison-round-1.txt*

Open the text file and go back to your browser.

### **First website, first results page**

First chose one website you will focus on and leave the rest for later.

Now, look at the first page of data science job results.

Let's assume that the website you are on has ten results per page.

Start at the top and work your way down to the bottom of the web page by pairs of twos.

Open the first link in a new tab and then open the second link in a new tab.

Now you have two tabs open that you will compare.

Remember, you are looking, thinking about, and deciding only between these two jobs (tabs) and which one you would prefer.

Remember, read the whole description for both and chose one.

Don't spend too long on it.

The goal is to look at both jobs holistically and decide which one you would rather be doing today.

Since you'll be doing hundreds of other comparisons, diving into the nitty-gritty is not important.

What is important is to decide, holistically, which of the two you prefer.

Once you've picked it, copy the URL of the winner into the text file named *pairwise-comparison-round-1.txt*.

Close the two tabs and go back to search result listing page.

Open the next two candidates and do the same process again.

In this way, if there are ten results per page, you'll do five comparisons: 1 vs. 2, 3 vs. 4, 5 vs. 6, 7 vs. 8, and 9 vs. 10.

Of course, if there are more or less results, then you'll do a different number.

Once you've done the first page, then you'll go to the next page.

The goal here is to go through fifty comparisons on this specific website so the number of pages will depend on how many results are returned per page.

As we talked about in the beginning, the more you invest in this process, the better your eventual outcome, so if you want to compare more than fifty per website, please do so – you'll thank yourself later.

Since you are looking at eight different job websites, when you do this for each website you'll end up doing four hundred comparisons (fifty comparisons per website times eight websites).

As you go through this process, you'll be noting down the URLs of the jobs that won in the text file *pairwise-comparison-round-1.txt*.

This is important because it means that you will have all of the information from all of the websites in one place for the first round of comparisons.

Doing four hundred comparisons will give you two hundred URLs of jobs that won.

Now that you have two hundred URLs of jobs that won, it's time to start the second round of pairwise comparisons.

## 3.5 Do Pairwise Comparison Between The Jobs That Won Round 2

---

Create a second text file named *pairwise-comparison-round-2.txt*.

This will hold the results of the second round of pairwise comparisons.

This also allows you to keep track of what you have done and gives you a historical record you can later look back on to see if what you've found out would still hold true.

Open both files: *pairwise-comparison-round-1.txt* and *pairwise-comparison-round-2.txt* in preparation for round two.

What you'll now do is take the first URL from *pairwise-comparison-round-1.txt* and open a new tab in your web browser.

Then you'll take the second URL from *pairwise-comparison-round-1.txt* and open a new tab in your web browser.

You'll then do the pairwise comparison between those two (and only those two) jobs/tabs to see which one you prefer.

Remember, you'll be doing lots of these, so think holistically which role appeals to you more and chose that one.

Copy the URL of the comparison winner and paste it into text file *pairwise-comparison-round-2.txt*.

Open the next two candidates and do the same process again.

If you obtained two hundred URLs of jobs that won from the first round, this means that you'll be doing one hundred comparisons in this second round.

This means that you'll end up with one hundred URLs of jobs that won the second round in the text file *pairwise-comparison-round-2.txt*.

The more you invest in this process, the better your eventual outcome, so if you compared more than fifty per website for each of the websites, you'll have more comparisons to do and more URLs that will be present at the end of the second round.

Now that you have one hundred URLs of jobs that have won twice, it's time to start the third (and last) round of pairwise comparisons.

## 3.6 Do Pairwise Comparison Between The Jobs That Won Round 3

---

Create a third text file named *pairwise-comparison-round-3.txt*.

This will hold the results of the third (and final) round.

Open the following files: *pairwise-comparison-round-2.txt* and *pairwise-comparison-round-3.txt* in preparation for round three.

What you'll now do is take the first URL from *pairwise-comparison-round-2.txt* and open a new tab in your web browser.

Then you'll take the second URL from *pairwise-comparison-round-2.txt* and open a new tab in your web browser.

You'll then do the pairwise comparison between those two (and only those two) jobs/tabs to see which one you prefer.

Remember, you'll be doing lots of these, so think holistically which role appeals to you more and chose that one.

At this point, you should recognize the jobs, as you will have looked at them several times up to this point.

The reason for this third round is to compare winners versus winners – that is, in the first round the choices may have been exceedingly obvious, whereas in the second round, you had already chosen things that you preferred, so the third round is testing preferred options versus preferred options.

What you are doing is working to reveal your preferences rather than staring at a blank paper and writing out your preferences.

Which is also the reason why there are so many comparisons being done; because you want to give yourself the best chance of seeing patterns develop from the data.

If you obtained one hundred URLs from the second round, that means you'll do fifty comparisons to give you fifty URLs of third round job winners.

If you did more comparisons in round one and round two, then you'll have more URLs in the third round.

Once you have these fifty URLs saved in the *pairwise-comparison-round-3.txt* file, you are ready to dig deep into the job posting text to analyze what is similar between them and what preferences were revealed.

From here on in, you'll look to see commonalities across postings in the realms of tools, technologies, industries, programming languages, roles, etc.

## 3.7 Extract Information From The Winners Of The Third Round

---

Create a text file named *round-3-job-text.txt*.

This file will hold all of the actual text from each of the third round pairwise comparison winners.

In this section, you will open a tab for each of the URLs in the *pairwise-comparison-round-3.txt* file and copy / paste the text into the *round-3-job-text.txt* file.

Copy the first URL from the *pairwise-comparison-round-3.txt* file and open it in a tab.

Copy the text from this job advert and paste it into the

The structure of the *round-3-job-text.txt* file should be the following:

~~~~~

~~~~~

Job URL:

[ URL ]

Job Title:

[ Title ]

Job Text:

[ Text Goes Here ]

Note, square brackets are there to signify what goes there while the  
“~~~~~”

“~~~~~”

serves as a separator between each job posting.

Once you've done this for the first URL, go to the second one.

Continue until you run out of URLs in the *pairwise-comparison-round-3.txt* file.

While somewhat tedious, you now have searchable text all in one file.

Additionally, if you want to read / view what's there you can just scroll down rather than having to switch tabs over and over again.

Save the file and get ready to start doing the analysis.

## 3.8 Analyze Information From The Winners

---

You will now become very familiar with the text of these fifty or more job ads.

This will be a multi-step process in which you'll do a read through all of the ads and take notes on various things related to the goal of the particular read through.

The multi-step analysis process steps are as follows:

- 1.** Basic read through and take notes on what stood
- 2.** Read through looking at education information
- 3.** Read through looking at work experience information
- 4.** Read through looking at relevant experience information
- 5.** Read through looking at statistical knowledge information
- 6.** Read through looking at mathematical knowledge information
- 7.** Read through looking at programming knowledge information
- 8.** Read through looking at machine learning knowledge information
- 9.** Read through looking at data knowledge information
- 10.** Read through looking at technical information
- 11.** Read through looking at “data science tools” information
- 12.** Read through looking at industry information
- 13.** Read through looking at business information
- 14.** Read through looking at data science group information
- 15.** Read through looking at role responsibilities information
- 16.** Read through looking at communications information
- 17.** Read through looking at intangibles information

The next chapter, *Chapter 4 - What To Look For During The In-Depth Analysis*, will help you figure out what to look for when doing all of these read throughs.

Based on the information and data you collect from these steps, you will then create a plan for closing your knowledge gaps so that you can become a data scientist faster than you think possible.

## 3.9 Create A Plan For What You Need To Learn And Do

---

After having done multiple read-throughs of all of the jobs, you will have a very concrete understanding of what tools, techniques, math, statistics, machine learning, algorithms, etc. you need to learn to be able to apply to jobs that appeal to you.

The next thing you need to do is understand where your knowledge lies in respect to what the jobs you like expect of you.

Using these knowledge gaps you need to create a plan to close them in a way that creates material you can show to others that proves you know the material.

What you are going to do is to match your knowledge gaps with steps in the data process.

Then you will do small tutorials you find on the web to learn and show you can understand the missing bits of knowledge.

The data process and what parts you need to do projects on

The data process:

1. Import – Get the data from an original source
2. Store – Store the data into a data store (database is most common)
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 / algorithm / statistical package will understand
7. Visualize – Descriptive statistics exploration
8. Model – Statistics + Machine Learning + Experiment Design
9. Coding – Machine Learning + Algorithms + Code
10. Understand – Explain 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 + where everything came from (makes it so that you can replicate in the future with ease)

Given these steps, you can now plug in small projects into parts of the data process to work on / show you can do certain skills within certain parts of the data process.

Doing many small projects allows you to do many more examples than you would have done if you were doing a few large projects.

*Chapter 5 – Knowledge Gaps That Matter And How To Develop An Action Plan To Close Them* will help you figure out the best way for you to put a plan together to close the knowledge gaps that you have, show that you know what you are talking about, and start putting together a data science project portfolio.

Now that you've read through the whole process it's time to dig deeper into the research...

## 4

# What To Look For During The In-Depth Analysis

## 4.1 Your Wants And Their Wants

---

### What do you want?

At the most basic, what you want is what excites you.

Basic needs and wants that can be met by most jobs will not suffice to keep you on track to reaching your goal.

Through the pairwise comparison process you should have seen enough jobs that certain phrases, industries, or tools that piqued your imagination.

These are the things you want to hone in on and remember as you are going through the process of becoming a data scientist.

When you are successful and have become a data scientist, you want to make sure you are doing what you want which is doing something that excites you.

At the same time, doing what you want means not doing things you don't want to do.

### **What do you not want?**

It is equally important to think about and realize what doesn't excite you.

Through the pairwise comparison process you will have seen enough jobs that certain phrases, industries, or tools made you react negatively to the job posting.

These are things you want to be aware of so that you don't end up with a "dream job" that while stellar on paper, is actually a nightmare for you because it entails many things that you don't want.

It is important to keep both what excites you and what doesn't excite you in mind.

### **Given what you want, what do they want?**

Because you now have a group of data science jobs you would consider taking (given the starting set), you can now figure out given what you want, what is it that employers want from you.

On the surface, this is a relatively simple answer – they want what they have spelled out as wanting.

However, as you'll see below, there are many specific areas to consider when looking at what the job adverts are communicating.

## Performing the actual read throughs

For each read through, what you are looking for are the things that keep popping up again and again.

You read over four hundred job postings to get to this list of fifty, so it's useful to do a read through all of them to refresh your memory of what actually won.

Some things will stand out that you remember and you may find yourself surprised to see other things come up again and again that you didn't notice during previous readings.

As you go through the read-throughs, start a new section at the bottom of the *round-3-job-text.txt* text file to fill out the seventeen multi-step analysis process steps as well as the 6 basic questions (explained below) that you should ask in each step.

Finally, you'll also want to write down the other "step specific" things you find through the read throughs.

It is very important that you copy and paste the exact words, phrases, and/or sentence that contain the wording that kept coming up as the descriptive words are ones that you want to capture as well as the general context.

You will later use these to create your highly specific plan of action as well as much later, construct your data science resume.

## 4.2 Analysis To Do In Each Step

---

In this section, you will cover the analysis you will do for every single step.

You have the file *round-3-job-text.txt* that includes all of the text from the fifty or more job postings that won the three rounds.

You will go through the following seventeen steps reading through all of the job text to dig out information relevant to each step.

The multi-step analysis process steps are as follows:

1. Basic read through looking at anything that stands
2. Read through looking at education information
3. Read through looking at work experience information
4. Read through looking at relevant experience information
5. Read through looking at statistical knowledge information
6. Read through looking at mathematical knowledge information
7. Read through looking at programming knowledge information
8. Read through looking at machine learning knowledge information
9. Read through looking at data knowledge information
10. Read through looking at technical information
11. Read through looking at “data science tools” information
12. Read through looking at industry information
13. Read through looking at business information
14. Read through looking at data science group information
15. Read through looking at role responsibilities information
16. Read through looking at communications information
17. Read through looking at intangibles information

## Analysis to apply to each read through

For each of the seventeen steps above, you'll want to go through the following analysis:

- A. Note what is required
- B. Note what is recommended but not required
- C. Note what is mentioned but is not a hard or soft requirements
- D. Note specific words / phrases you are vaguely familiar with
- E. Note specific words / phrases you do not know
- F. Note action words related to specific step of the seventeen steps above
- G. Summarize your findings in your own words

In addition to this, you'll want to do some step specific analysis.

## 4.3 Step Specific Analysis

---

In addition to the analysis (A through G) above, you'll also want to do some specific analysis for some of the seventeen steps.

Note that not all of the seventeen steps will require extra analysis as for some it's pretty straight forward.

Let's go through the specific steps that have extra information you want to dig out of the job postings.

## #2 - Read through looking at education information

You are looking to see if you need a bachelor, masters, or PhD degree.

You are looking to see if you need to be a Post Doc.

You are looking to see if any special certifications are required.

You are looking to see if any special government clearances are required.

You are looking to see if they mention any non-traditional educational routes like MOOCs (massive open online course), Nano-degrees, data science bootcamps, and/or data science fellowships.

You are looking to see if directly related experience can or may offset degree requirements.

Note that you can find data science jobs that will ignore their requirements if you are good enough, but they are hard to find and it will be an uphill battle.

## #3 - Read through looking at work experience information

You are looking to see if educational projects can count as work experience.

For PhDs and Post Docs, your thesis projects are very close to work experience, whereas an undergrad project will be less likely to be considered work experience.

You are looking for action words here in addition to actual content.

Things like “assists”, “builds”, “utilizes”, “works with”, “models”, “identifies”, “communicates”, “educates”, “generates”, “maintains”, “updates”, “upgrades”, “manages”, etc.

These types of action words will give you a sense of what the job responsibilities will entail as well as what types of things could be relevant from your previous work experience.

#### **#4 - Read through looking at relevant experience information**

You are looking to see if there is a separation between relevant experience and work experience.

If there is, then there is a high likelihood that your out-of-work / out-of-school personal data science projects will be looked upon very positively.

It also means that you can close knowledge gaps in these areas with projects without having to go and get a job that gives you that experience.

#### **#7 - Read through looking at programming knowledge information**

You are looking for specific programming languages.

You are looking for specific programming libraries that sit on top of the languages.

You are looking for specific libraries that sit on top of other libraries that help abstract out common tasks.

For example: Python + NumPy + SciPy + Scikit-Learn + Matplotlib ...

You are looking for platforms that production code is hosted on and the languages (if mentioned) that are used in production, development, testing, and data science environments.

### **#8 - Read through looking at machine learning knowledge information**

You are looking for specific depth of techniques – general versus highly technical modeling.

You are looking for how the knowledge is being applied.

### **#9 - Read through looking at data knowledge information**

You are looking for specific vendor or open source projects.

You are looking for specific types of data that are used (web metrics, ecommerce stats, sensor data, etc.).

### **#10 - Read through looking at technical information**

You are looking for general technical information that appears in many of the jobs.

### **#11 - Read through looking at “data science tools” information**

You are looking for vendor specific tools that exist to help data scientists do data science.

### **#12 - Read through looking at industry information**

You are looking for general industry terms like “entertainment industry”, “finance”, “marketing”, “ad-tech”, “non-profit”, etc.

On a second reading, you are looking for hyper-specific industry terms.

### **#13 - Read through looking at business information**

You are looking for company size.

You are looking for company organizational structure.

You are looking for company age.

You are looking for company technical level.

You are looking to see if data science is the business the business is in or whether data science is more of a support function.

### **#14 - Read through looking at data science group information**

You are looking for group structure – embedded within organization teams or its own entity.

You are looking for whom the group interacts with – corporate functions (Sales, Marketing, Human Resources, etc.).

You are looking to see if group produces only external products, internal products, or a mixture.

You are looking for the group size.

You are looking for the age of the group.

You are looking for management structure.

You are looking for mention of other team members and their backgrounds.

### #15 - Read through looking at role responsibilities information

You are looking for how wide your knowledge needs to be.

You are looking for how deep your knowledge needs to be.

You are looking for the combination of how wide and deep your knowledge needs to be.

You are looking for how research-based the role will be.

You are looking for how much application the role will entail.

You are looking for how much autonomy the role provides.

You are looking for any job-related perks the role provides.

You are looking for evidence you will have to create your own tools.

You are looking for what already existing tools you will use.

You are looking for whom you will interact with.

You are looking for reporting structure.

You are looking for complexity of work.

You are looking for depth of work.

You are looking to see if you are supporting already existing data science work, starting new fresh work, or a mixture.

You are looking for any physical and/or mental requirements.

You are looking to see if this is a research type role.

You are looking to see if this is an application implementer type role.

You are looking to see if this is a hybrid research and researcher role.

You are looking to see if this role will require you to do data engineering work in addition to data science work (connecting and productizing data flows and data engineering systems)

## **#16 - Read through looking at communications information**

You are looking to see if presenting data science work is a small, large, or non-existent part of the role.

You are looking to see if you will do education for the rest of the organization.

You are looking to see if it is suggested/required that you present papers / give talks outside of the organization.

### **#17 - Read through looking at intangibles information**

You have done the research and made it all the way down to read through #17 – look for *intangible information*.

Here the word intangible means part of the job description that will affect whether the job is interesting to you, but it's not necessarily a hard or soft skill.

Because these things can be very central to the role, the next section is devoted to looking at them.

## **4.4 Intangibles And You**

---

### **#17 - Read through looking at intangibles information**

Let's break down this step even more to explore what kinds of things you should be looking for and what can affect what types of data science skills you learn and what types of data science jobs you decide to go after.

### **#17A – *Intangibles and you – PEOPLE***

You are looking at what types / roles of people are mentioned.

You are looking to see if the people mentioned in the role would excite you to be work-friends and/or associated with them.

You are looking at what types of diversity-related people descriptors are used.

You are looking at what genders of people are mentioned.

You are looking at what educational background people who are mentioned have.

You are looking at what race people who are mentioned have.

You are looking at what creed people who are mentioned have.

### **#17B – *Intangibles and you – IDEAS***

You are looking to see if the role / company / industry meets your ethical and moral considerations.

You are looking for what kinds of ideas the people who work there have.

You are looking for words / phrases related to possible ethics and morals of people.

You are looking at ways they view the world.

You are looking at societal value judgments.

You are looking at phrases that give you insight into the company culture.

### **#I7C – Intangibles and you – THINGS**

You are looking at things related to the role like meals at work.

You are looking to see if they have a book budget.

You are looking to see if they have a conference budget.

You are looking to see if they have “20%” time.

You are looking to see what other work-related perks they have (game rooms, massages, lunches provided, company offsite meetings, etc.)

### **#I7D – Intangibles and you – LOCATION**

You are looking to see where the company is located.

You are looking to see if the company has a work-from-home program.

You are looking to see if the company allows for remote working.

You are looking to see if the role will include travel.

You are looking to see if the locations the role(s) are located in have the amenities you want.

You are looking to see if the location the role is located in means you'll be doing commuting and if so, will it be public transportation or private cars.

### **#17E – Intangibles and you – MONEY & COMPENSATION**

You are looking to see if the role / company / industry would meet your compensation goals.

You are looking to see what compensation level is mentioned.

You are looking to see if there are signing bonuses.

You are looking to see if there is equity attached to the position.

You are looking to see if there are revenue sharing schemes.

You are looking to see if there are retirement plan options mentioned.

You are looking to see if they mentioned salary / equity / compensation progression.

You are looking to see if the compensation they are given will allow you to work in the location they are located in.

## #17F – *Intangibles and you – VACATION*

You are looking to see if a vacation policy is mentioned.

You are looking to see if a vacation is earned or automatically given.

You are looking to see if there is “unlimited” vacation offered.

**Whew**

You made it!

Now that you’ve done all the research and have figured out all of the possible things the job roles show, the next chapter will focus on taking all of this newfound knowledge and turning it into your own personal highly specific plan for becoming a data scientist.

## 5

# Knowledge Gaps That Matter And How To Develop An Action Plan To Close Them

## 5.1 Knowledge Gap Evaluation

---

***“The first principle is that you must not fool yourself – and you are the easiest person to fool.”***

***- Richard Feynman, Nobel Prize Winning Physicist***

The saying “knowing is half the battle” is directly applicable to you at this point.

You now know what you need to learn as well as what you do not need to learn.

You will now take your list of technical terms, mathematical terms, computer science terms, data science terms, etc. and do the following:

Using your favorite search engine, search for the following:

- <Key term> + beginners guide
- <Key term> + interview questions
- <Key term> + meetup
- <Key term> + python
- <Key term> + r
- <Key term> + Scala

Then you look through the top 5 to 10 search engine results for each search.

For each result:

- Read through it
- Close your eyes and imagine replicating it
- Note down words you do not understand
- Note down any steps you didn't recall or remember
- Note down anything that was surprising to you
- Note down anything that you missed
- Ask yourself if, given the data and goal, how much of this tutorial you could replicate without looking at it?

This list of notes is now a clear representation of knowledge gaps you need to fill.

Ask yourself for the tutorials, given data set and goal how much of this could I replicate without looking?

You now have a secondary source of things to learn and master.

## 5.2 Knowledge Gap Closing

---

### Working Backwards

In order to hit all the major points and minor points of what you found in your research of data science jobs and the knowledge gap evaluation it's important to write out exactly what it is that you found.

So you write out a few long sentences of what you need to be able to show / prove that you can do.

From here, you break down the process of what you need to be able to do into smaller pieces.

For instance, one your long sentences could be broken down into the type of problem you are trying to solve, the data you are using, the specific technique(s) you are going to use, the specific software you are going to use, as well as the specific tools you are going to use.

### Building the plan

Now that you have an idea of what you need to be able to show and prove, you have to create a plan for showing that.

Your first instinct may be to do large projects that tie all of the different knowledge gaps together.

If this is so, please ignore it.

The key to building the plan is to do a large number of small projects not a small number of large projects.

A large number of small projects taken together as a body of work can and will show that you can do all the things you need to do.

By having a large number of small projects it also gets you used completing small “data science” type projects.

To figure out where exactly these small projects fall and where to fit them in, take a look at how to plan the action plan.

## 5.3 Planning Action Plan

---

Your goal is to learn the bare minimum required to get a data science job.

You are not in it for glory; you are in it for getting a job.

Once you get the job, you can continue learning and doing and really excel at the job.

Until then, the goal is to learn the bare minimum.

This sounds like a bad thing, but it's not.

Your goal is to learn, as quickly and effectively as possible, what is necessary to be able to get a data science job, which means that you first

have to focus on the must haves and then can later focus on the nice to haves.

So you start out focusing on learning the hard requirements and then add in as interests arise the other things that they'd love for you to know, but don't mandate that you know it.

### **Requirement ladder gives you a sense of growth**

You can view this as a requirement ladder – the more rungs you go up, the better chance you have of getting a job.

It also lays out a path that you can focus on and points the way for you to grow your knowledge and experience.

### **Establishing a learning path**

Because hard requirements are that, hard and non-negotiable, it allows you focus your studies and your plan on attaining those skill sets and knowledge first.

Which is great because it makes it clear what you need to learn first – so rather than playing a guessing game and focusing on things that come easy to you or things that sound good (but you have no data telling you that you are right), you can focus on things that you have hard evidence for.

### **Time related questions to ask before starting to plan the Action Plan**

When do you need to have a job by?

How much time do you have to study?

How much time can you set aside each day to maintain momentum?

How much time can you realistically spend versus how much you wish you could spend?

Better to be realistic and do it than to be wishful and not do it.

### **Budget related questions to ask before starting to plan the Action Plan**

How much money do you have?

How much money can you spend on books?

How much money can you spend on training courses and materials?

How much money can you spend on educational opportunities?

## **5.4 Constructing Action Plan**

---

### **Mind map**

Build a mind map to help you visualize where you are going and if you are missing any steps.

Because you are working backwards, it is easier to build the mind map.

The layout makes it clear what steps need to be taken so it is easy to see.

Make each step super simple to understand

Make each step super simple to complete in one day

If steps are not getting done, then break them down further.

This will be helpful to see where you are getting stuck or where you are not making progress due to technical or motivation difficulties.

### **Inventing versus replicating**

For a vanishingly small amount of jobs, you'll be inventing algorithms and conducting original research.

For the rest, you'll be replicating already existing research and applying algorithms to particular data sets and domains for your specific company.

Because of this, you will want to focus on replicating tutorials (as you can check your work against them), rather than working on inventing the whole process by yourself from scratch.

Remember that the goal is to show you can do something and don't have knowledge gaps for a particular topic.

Replicating someone else's example lets you borrow their brain to learn and show how to something.

How then will you go from small projects to having work you can point to that shows what you know and what your capable of doing?

The data process and what parts you need to do projects on

You can think of the full data process as getting data from somewhere and eventually communicating an insight from this data.

To that end, in order to showcase you can do what is needed for the job, you'll do small projects that not only prove you can do the necessary skills, it will also help you learn and close the knowledge gaps that you have.

The data process:

1. Import – Get the data from 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 / algorithm / statistical package will understand
7. Visualize – Descriptive Statistics Exploration
8. Model – Statistics + Machine Learning + Experiment Design
9. Coding – Machine Learning + Algorithms + Code
10. Understand – Explain insights
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 do you go from here
13. Document – Code + Thought process + where everything came from (makes it so that you can replicate in the future with ease)

The highly personal plan then is a plan of how to show you know how to do the full data process by replicating different tutorials that fill your knowledge gaps.

### The highly personal plan

Given what you learned from your research, what gaps you have in your knowledge, and what gaps you need to close, you can now marry that with the data process to generate a series of small projects that will get you from where you are to becoming a data scientist.

The idea will be to have three **SMALL** projects per skill set you want to show case with the specifics you found from the pairwise comparison process research.

For each small project, you want to find three different tutorials that you can replicate.

By having three different tutorials, you'll see slightly different approaches to the same goal.

## 5.5 Doing Action Plan In Public

---

Since you know key words / topics / techniques / languages that will be in your data science knowledge, you can build your web presence from ground up

You want to show you can do a data science project from start to finish.

You want to show you can do each section of the data process:

1. Import – Get the data from 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 / algorithm / statistical package will understand
7. Visualize – Descriptive Statistics Exploration
8. Model – Statistics + Machine Learning + Experiment Design
9. Coding – Machine Learning + Algorithms + Code
10. Understand – Explain insights
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 do you go from here
13. Document – Code + Thought process + where everything came from (makes it so that you can replicate in the future with ease)

You also want to show that you have no knowledge gaps in relation to the research data you found in the pairwise comparison process.

To that end, you should do three tutorial data science project replications of each of the thirteen data process steps making sure to hit all of the key terms that you found in the pairwise comparison process.

Since you'll be doing lots of very small projects, you'll be able to post frequently on your website and have a web presence that shows you doing consistent work.

This shows future / potential employers that you value data science, that you are consistent, that you work through things, and that you have a grasp of all aspects of data science related to their field, data, role, and industry.

In doing this, you can also build your social media presence by sharing your work.

You also have something public to chat with data people about when you meet them in the real world.

Additionally when you finally get to put together your resume, cover letter, and emails to employers, you'll have a handy reference that will speak volumes to your work ethic and your data science work.

## 6

# Conclusion

## 6.1 Next Steps

---

Get to work 😊

Your goal should be to start and do a little bit of work everyday.

As was covered in the introduction, the more research and thought you put into this process, the easier and faster you'll be able to get a data science job because you'll know exactly what knowledge gaps you have and be able to close them effectively and efficiently while showcasing your work publically.

## 6.2 What To Do If You Get Stuck

---

**Stuck how?**

You need to ask yourself:

- Did I lose motivation for research?
- Did I lose motivation for planning the Action Plan?
- Did I lose motivation for doing the Action Plan?
- Did I lose motivation for doing small projects?

What you need to do will depend on where you are getting stuck.

If it's losing motivation for becoming a data scientist you have to remember and think about the original reasons you decided to start / switch to the field.

If it's one of the other reasons then the following two things will help:

**Break learning steps down further**

If you are having trouble getting started or continuing, it may be that you are taking on too much and need to break what you are doing into smaller steps.

Each small step should be able to be completed in one sitting.

This helps you with starting, practicing, doing, and finishing data science work.

This will also help you keep the scope small enough that you never feel overwhelmed.

**Switch to another small step on the mind map**

If at some point you really just can't get started or move forward it's very helpful to be able to easily see what other steps you can take to keep moving forward.

This is why you really need to invest in creating a mind map of your Action Plan.

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

---

### **Revisit *pairwise-comparison-round-3.txt***

By going through your Action Plan you are closing as many knowledge gaps as you can that were found in the jobs that won the pairwise comparison process.

As you do small projects it will be helpful to re-read the job posting text.

You'll understand more of what all of job entail.

### **Redo pairwise comparison process**

If you do it again, you'll notice that you have a much clearer understanding of how each potential job compares to jobs that you really like.

## 6.4 Have Fun

---

Have you ever struggled to eat your favorite dish?

Have you ever struggled to read your favorite book?

Have you ever struggled to finish watching your favorite movie?

The answer to all of these questions is “no” because all of these things are fun.

To achieve your goal of learning data science and becoming a data scientist you need to make the process fun and enjoyable.

Is all very well and good if you are doing it for the money or for the high profile and prestige of the position.

However, what will get you through every single step is having something to do that is enjoyable for you.

## 6.5 Discipline

---

*"Discipline is remembering what you want"*

- David Campbell, Founder Saks Fifth Avenue

Through the pairwise comparison research you figured out what it is that you wanted.

Through the analysis and thinking of Action Plan construction you thought about how you were going to fill your knowledge gaps in a way that you could demonstrate publically through project work.

Now you just have to remember the conclusions of your work and start having fun!

## 6.6 Good Luck

---

**You are awesome and you will do great!**