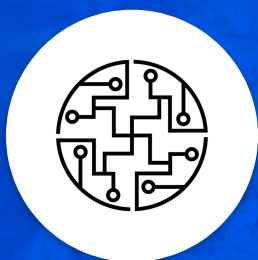
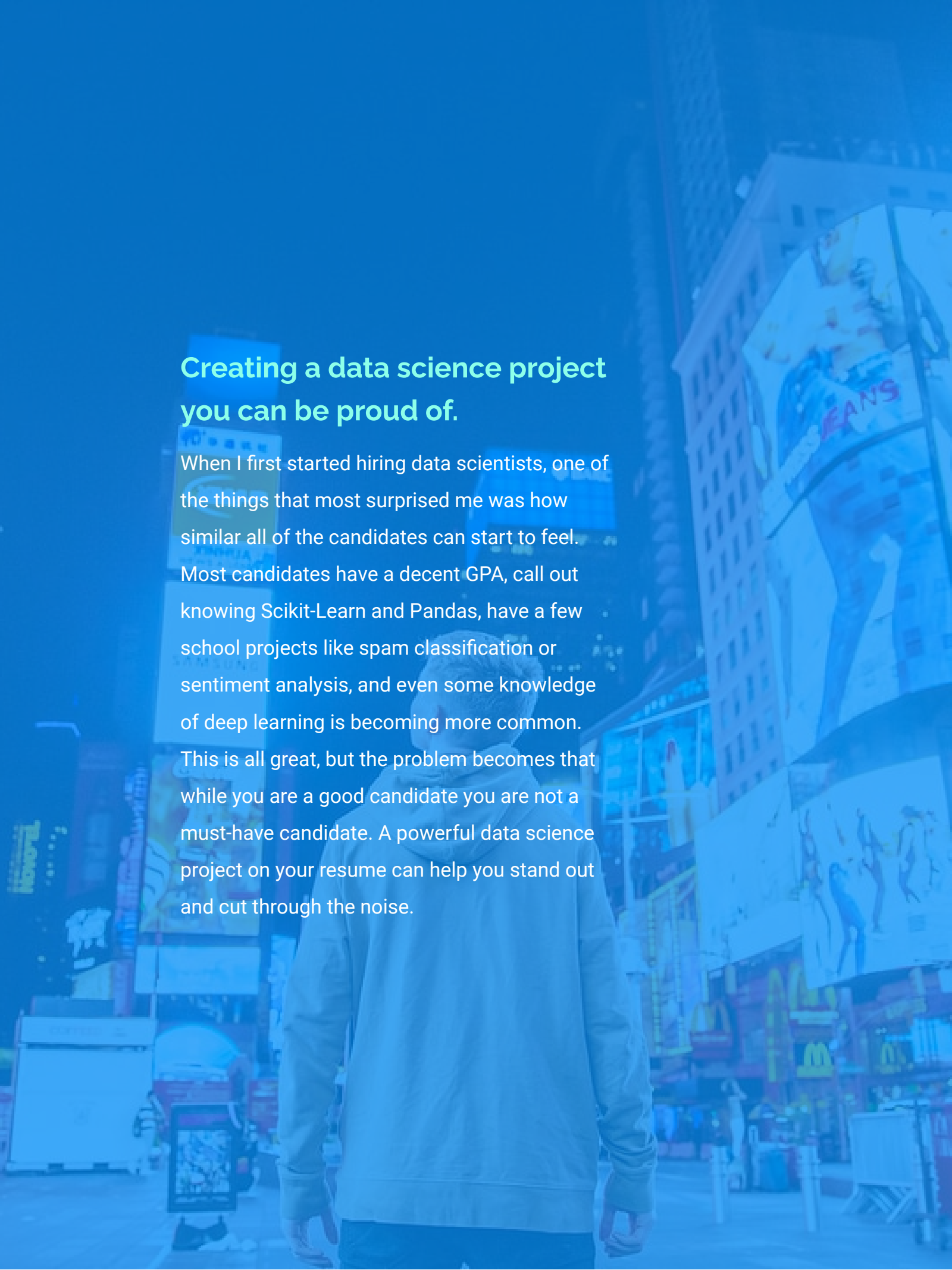


My 5-Step Process for Creating Amazing Data Science Projects



By Tyler Folkman
Learning With Data

A person wearing a light-colored hoodie is walking away from the camera down a busy city street. The street is lined with tall buildings and large billboards. One billboard on the right shows a person in a blue shirt with the word "JEANS" in red. Another billboard on the left shows a person in a yellow shirt. In the background, there are several McDonald's signs. The overall scene is a busy urban environment.

Creating a data science project you can be proud of.

When I first started hiring data scientists, one of the things that most surprised me was how similar all of the candidates can start to feel.

Most candidates have a decent GPA, call out knowing Scikit-Learn and Pandas, have a few school projects like spam classification or sentiment analysis, and even some knowledge of deep learning is becoming more common.

This is all great, but the problem becomes that while you are a good candidate you are not a must-have candidate. A powerful data science project on your resume can help you stand out and cut through the noise.



Step 1

Passion

It is unlikely you are sitting on ample time spare time, so committing to building your data science project will surely take commitment and sacrifice. In my experience, you will only be successful in doing so if you pursue a project for which you have a lot of passion. The passion doesn't necessarily have to be obvious either. Personally, I am not passionate about writing at all, but I am passionate about sharing ideas. Writing is a good medium by which I can pursue my passion for ideas. Maybe you would really like to do a project with deep learning, but are lacking the drive. But perhaps you are passionate about music. You might start your portfolio by using deep learning to [create music](#).

Focusing your effort towards your passions can greatly help you push through when it is tempting to give up.





Define Your Own Problem

It is incredibly tempting to build side projects around predefined problems on platforms such as Kaggle. While this certainly makes the process easier, it effectively removes one of the most important parts of the data science process: defining the problem. In industry, often one of the biggest challenges is converting a business problem to a data science problem. Before writing any code, think through the following:

- What problem do I want to solve?
- How do I think I could use data science to solve this problem?
- If I could solve this problem, what value would that create?

This step is essential because it sets the stage for the story of your project. **It will help you better explain to others why you chose this project and show your strategic thinking when tackling a problem.**





Step 3

Gather Your Own Data

If you defined your own problem, this step will almost certainly be mandatory. Your problem will probably be unique and thus you will need to spend some time gathering data. This is great! In my experience, I have found that I spend a significant amount of time thinking about the best way to gather data to help solve my problem. You can now showcase that skill in your project. For our music example, it might include examining the [Free Music Archive](#) which includes high-quality, legal audio downloads. I promise that by exploring how to acquire and gather your own data, you will learn a crucial step of the data science process and one that is not often taught in school.





Showcase Data Exploration

As Andrej Karpathy [said](#), "Become one with the data."

One of the first steps of any machine learning project is to spend time inspecting and analyzing your data. Don't skip this step. Not only is it important, but it also allows you to create some really great visualizations. Dig into your data and look at the following:

- Are there any outliers?
- What are the distributions of your features?
- Plot correlations between features and the target
- Look at actual examples of your data

There is a lot more you can do during this step, but those are a good place to start. Use [seaborn](#) to make your plots prettier or if you are feeling ambitious, try and make the visualizations interactive with something like [Plotly](#). **The goal here is to show others how you analyze data to uncover nuggets of wisdom others might have missed, which will make your models even better.**



Build Multiple Models

Too often, I see projects only show the best model. A great portfolio project allows people to understand your thought process, so please show us! To do this effectively, I would recommend the following process:

- First, create a non-machine learning baseline. This baseline should be something reasonable like a historical average. This is a crucial step for evaluating your first machine learning-based model.
- Second, create your first machine learning model. Describe why you choose to start with that model and compare it to your baseline.
- Third, build your second machine learning model. **The crucial part of this step is to clearly explain why this was the next best step to take.** Was your model overfitting and thus you needed to use a less complex model or add regularization? Maybe you used the same model, but developed additional features based on error analysis.
- Fourth, repeat step three until you feel comfortable with the results.



Bonus Step

Tell A Story

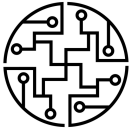
At this point, you have a lot of the key components in place and you might feel like you are almost done. Not so fast! You now need to go back and connect all your work and tell a great story.

Great data scientists are great storytellers

This is the most important step in creating an amazing project for your portfolio. **If you skip this step, you probably just have a bunch of code on GitHub. That is not a project.** Use a blogging platform such as Medium or you could even create your own blog, and explain your journey. Write about the goal of the project, highlight the key exploratory analyses, include your modeling results and thought process, and tell us how your project created value.

Think of this as how you would present your project to executives.





By Tyler Folkman
Learning With Data

Want more? I'll Show You How to Move Beyond Machine Learning Hype

According to Gartner, **over 85%** of data science projects **fail**.

The biggest problem I find people have when they land a data science job is having the skills to actually create business value with machine learning. In my course, you will learn everything you need to know to go from a business problem to business value using machine learning so your value doesn't stop at a Jupyter notebook. So - stop failing to create value and start becoming insanely valuable. And get **50% off** by just clicking the button below!

GET MY COURSE >>

Attract'

Created with Growth Tools in Partnership with Leadpages