**Three Ways to Gain Data Science Experience in a Non-Data Science Job**

**Develop your data science skills regardless of your field**

Now you’re everywhere, just like the skies. *Brendan’s Death Song — Red Hot Chilli Peppers*

Two years ago I was feeling restless about my professional future. I had just finished my Master’s, where I had had my first contact with data science and causal inference. The “problem”? I loved it!

How can loving data science be a problem, one could ask. Well, actually I am a civil servant and my work does not involve data science. This meant I had a challenging tradeoff ahead.

On the one hand, I had a safe job, but unfortunately not one that would allow me to develop as a data scientist. On the other hand, I could just quit my job and look for a data science position without having any previous professional experience in the field..

Both options offered risks: risk of unemployment or risk of unhappiness for not following my dreams.

Fortunately (or not) I am a big fan of self-reflection and decided not to make my choice before careful thought. One day I was reading Yuval Harari’s *Sapiens*¹and came across this unsettling idea that, from now on, we will have to get used to learning and developing constantly. The future welcomes professional instability and impermanence. In this new age, we have to be ready to learn new skills, forever. Here are Harari’s words:

We’ll have some big changes by 2025, even bigger changes by 2035 and an even bigger revolution in 2045. Old jobs will disappear, new jobs will emerge but the new jobs too will quickly change and vanish. People will have to retrain and reinvent themselves not just once but again and again throughout their lives.

If the time of completely safe jobs is over and if I would have to develop and learn constantly anyway, why couldn’t I find data science in my current job?

As in the Red Hot Chilli Peppers song I cited above, data science is now everywhere, just like the skies. Certainly, I would find gaps in my field that could be filled with my new analytical skills.

In the beginning, I felt pessimistic about the real data science opportunities I would find in my current work. It required some persistence and determination but eventually, I made it.

Below you will learn about the 3 initiatives I worked on to develop my data skills as a legislative advisor at the Brazilian Congress. For each experience, I prepared some recommendations based on my experience and the challenges I faced.

**1. Automation of reports with Python Pandas and Matplotlib**

Almost all jobs involve some kind of Excel report! Even if your tasks are bureaucratic, the automation of reports is an excellent opportunity to apply data science in your daily routine.

**Break through resistance to change in your team.** When I first suggested the automation of our reports to my manager, he was reluctant to allow me to use working hours to code the app. In this situation, the best approach is to show clearly how the automation of reports will add value to the team. To achieve that, you can show them examples of reports and dashboards similar to the one you want to develop, highlighting how the team will gain time and insights compared to the current reports. In my specific situation, one team member had to visit several pages on a website to get information about bills going through Congress. Most of this work could be done in milliseconds by a Python script!

**Search for existing solutions to make your task easier.** Perhaps you are lucky like I was and your company or institution has an API providing the data you need! In this case, all you need to do is code your API calls and get the data. But be careful! There are tons of data out there, your duty as a data scientist is to select, process, and communicate them effectively to your team. This can be done with visualizations and storytelling. Moreover, consider the actionability of the indicator you are providing, that is, what action your team is supposed to take based on the values reported in that indicator.

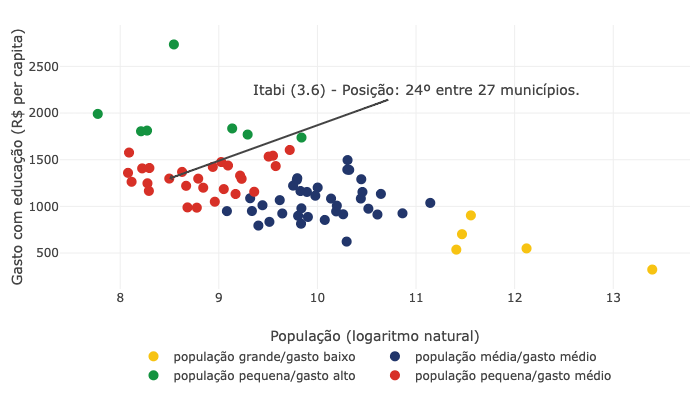
**Use established frameworks to develop your app**. [Streamlit](https://streamlit.io/), [Flask](https://flask.palletsprojects.com/en/2.2.x/), and [Shiny](https://shiny.rstudio.com/) are effective tools to build your reports and present your results. Especially if you do not have your data provided by an API and depend on your team feeding the app, it is essential to create a clear and intuitive interface for your user. You could learn more about CSS and HTML, which will allow you to customize and make your report more user-friendly.

**2. Publication of a Shiny Dashboard for external clients**

My second initiative was creating a [Shiny](https://shiny.rstudio.com/) app to process and display insights on educational data. The objective was to provide actionable information for mayors to improve their local schools in the Brazilian state of Sergipe. The app gathered data about student achievement, the infrastructure, and the management of municipal schools. It also compared and ranked municipalities to offer the low-performing municipalities examples of successful cases. The results can be seen [here](http://www.educa-sergipe.com.br/) (unfortunately only in Portuguese).

**Listen to your client's needs.** When you develop a dashboard, keep your user in mind. Are you adding a visualization because it adds value to your user or because you think it is beautiful? You have to talk to your target group to find out what will really help them and listen to their worries and wishes. When I developed this app, some mayors were worried about being ranked, because they feared being compared to larger and richer municipalities, which would not be fair. Based on that input, I developed a clustered ranking in which municipalities were divided into groups that shared similar characteristics regarding investment in education and population.

The image below shows a [plotly](https://plotly.com/) graph with the clusters computed by the k-nearest neighbor algorithm.



Clustered Ranking with KNN. Image created by the Author.

**Show your work.** No matter how effective and attractive your Shiny app is, it will not be useful if you do not share it with people who might benefit from it. Share it on your social media and check how your company or institution can help to promote your dashboard. Finally, consider making an online or onsite event to present your tool to your main target group.

**3. Workshop about your favorite data science topic**

This was by far the most challenging yet rewarding initiative for me! Why? Simply because I had a serious problem speaking in public. Sweating and extreme nervousness were some of the symptoms. Sometimes I could not even remember what I had presented and if my reasoning had made sense.

Despite that, I organized a workshop to present the synthetic control method to colleagues at Congress. Basically, it is a method for causal inference used for policy evaluation as well as by companies wishing to assess the impact of their actions. Uber, for instance, uses synthetic control to inform their decisions. Watch [this video](https://youtu.be/j5DoJV5S2Ao) if you want to know more about their case. Besides that, check out [this post published with TDS](https://towardsdatascience.com/understanding-synthetic-control-methods-dd9a291885a1), by

[Matteo Courthoud](https://medium.com/u/666130fb420f?source=post_page-----d1fa01496721--------------------------------)

, explaining the synthetic control method.

But what about my public speaking anxiety? Well, that's my first piece of advice.

**Get out of your comfort zone and look for help.** One month before the workshop I decided it was enough of suffering because of public speaking anxiety and decided to enroll in a public speaking course. I have to admit that the first meetings were really difficult for me. However, it got easier as I presented more and more in front of colleagues facing the same problem. I learned how to manage my emotions, eliminate noises in my communication and be objective and engaging. Unfortunately, by the time I had to present my synthetic control workshop, the public speaking course was still at the beginning, but still it helped me a lot!

**Engage your audience.** Synthetic control, as well as many data science topics, can be hard to explain and complex to understand. Do not lose your audience with excessive technical details and focus on the intuition behind complex concepts. Also, make quizzes and questions to engage them. I started my talk about causal inference with a quiz where participants answered if they had had coffee on that day and what was their level of concentration at that moment. With this simple activity, I could exemplify several challenges in trying to estimate the effect of coffee on concentration levels.

This workshop was very rewarding because I could feel I added value to my colleagues and incentivized them to learn more about data science. Many of them congratulated me on the initiative.