**How To Be a Good Data Analyst Without Good Data**

Musings on data quality

*“This is wrong.”* — key stakeholder

I dreaded hearing that phrase (and variations of it) when showing dashboards to project stakeholders. I would have a lot of anxiety around data analytics projects in the beginning because I didn’t want to fail. (I’m not great at handling failure— still working on that.)

But once I had more experience as an analyst, my perspective started to change. I wasn’t a bad analyst; rather, I was given bad data.

What can you do?

In this article, I will walk through some of the strategies I employed when the data quality for a project was… less than ideal.

A quick note before I start: a lot of the time, the data is out of your control! There could be business decisions that were made years ago that have negatively impacted the data quality at your organization. IT reductions, acquisitions, failed ERP transitions, etc. Long-term, sustainable change in data quality needs to be an initiative from leaders higher up than analysts.

But, at the end of the day, you often have to work toward a deliverable anyway. So, how do you make progress?

**Automate data cleaning tasks**

The term “bad data” can encompass a variety of issues. Let’s talk first about “messy data.”

You might get data that is disorganized, has spelling issues, has the wrong data types… the list goes on. The data isn’t “good,” but you know how to fix it.

After you’ve done some [exploratory data analysis](https://medium.com/gitconnected/cozy-up-with-your-data-6aedfb651172), you should be familiar with any issues in your dataset that you need to correct. Once you are ready to clean your data, here are some tips:

* If there is business logic that goes into the cleaning, don’t forget to document it (e.g., you have to change some department labels because you know they are wrong in the system).
* Make sure that your process is sustainable — if it involves creating a mapping file that has to be actively maintained, then it won’t be a good long-term fix.
* Spending more time on cleaning data at the beginning of the project will pay off in the end.
* Automate where possible.

If you have to manually clean the same dataset for different projects (let’s say, for reporting in Excel), chances are you are going to make a mistake. So whether it's building logic into PowerQuery to feed a PowerBI dashboard, using an Alteryx workflow to clean your data, or wrangling the data in Python, you can leverage automation to help error-proof the data cleaning process.

**Lean on SMEs for data validation**

Now let’s talk about when the data isn’t “good,” but you’re not sure how to fix it.

As a data analyst, you can’t know the ins and outs of every area of the business. The folks who work every day in pricing, supply chain, customer support — they are the ones who know every edge case and scenario that can affect their data. And they are the people who you need in your corner when it comes time to validate your analysis.

Here are a few takeaways I had from some tougher projects:

* Find the SMEs (Subject Matter Experts) early on in the project. The sooner you get their input and can walk through the data with them, the better.
* When the time comes to validate the results of your analysis, have them test out your dashboard. They will know all the edge cases in the data and how the business logic should work.
* Ask questions! You can use this project to deepen your knowledge of an area of the business. (This is what I did when I worked on supply chain data projects — soon, other analysts would come to me for help in understanding the database relationships and such.)

The subject matter experts can also help you push for better data quality if you hit roadblocks in your analysis. They might understand better how the data is collected or the history of certain datasets at the company.

**Repeat after me: No data is better than bad data**

Sometimes being a good data analyst means delaying a deliverable. If you are unsure about the data quality or are aware of any issues with the data, then it’s best not to publish the results of your analysis.

It’s not helpful to have a dashboard to make “data-driven decisions” if the data is driving people in the wrong direction.

This is really common sense, but when I was a new analyst, I sometimes felt pressure from stakeholders to deliver results when I wasn’t confident in the quality. I never liked pushing back deadlines and having to deliver bad news, but it was ultimately the right thing to do.

**In conclusion**

Being a good data analyst with bad data is hard! But you’re not alone in the struggle. Automating the tedious data cleaning tasks and then leaning on business experts for help in validating and fixing the data are some great ways to make progress on your projects. And even when the data doesn’t have much integrity, you can maintain yours by being transparent about the data quality and realistic project timelines.

If you have tips for dealing with bad data, please share them in the comments!