What is really data about?

Luis Meazzini  Follow
May 15 · 6 min read ⭐️

Data science basic idea

Nowadays, most articles tend to show a lot of execution and a little exploration. Like Conor Deway wrote on his article:

Unlike engineers, designers, and project managers, data scientists are exploration-first, rather than execution-first.
Moreover, this is quite fair considering the history of Data Science.

**A little of data science history**

Data Science began with data mining on “From data mining to knowledge discovery in databases” [1] for discovering useful knowledge in data. Then in 2001, Willian S. Cleveland tried to combine the computation power available with data mining to improve statistical analysis, beginning the use of the “Data Science” term [2]. Today Data Science covers a lot, really many fields, there is Data Engineers, Data Analytics, Machine Learning Engineers, and so on. Also, recently Data Science is called “The Sexiest Job of the 21st Century”.

The primary tool of a Data Scientist is, of course, data. However, most everyone gets some data from Kaggle, runs a complex model such as a CNN and BOOM. We have a new data scientist. Using all these fancy tools without knowing the background and how to make sense out of data correctly and the model is a waste of time or just right for learning how to use those techniques.

**What is really data about?**

This opens a window to go a little further in the data question. What is

Data is not about Machine Learning, visualizations, statistics, and wrangling. Data is about understanding. Understanding the problem and how you can solve it using data with whatever tools or techniques you choose.

Understand the problem. Understand the data. And the rest will flow.

The Data Scientist goal’s is not making a good model

If there is a problem, a Data Scientist tries to solve it with Machine Learning. How do we know if he succeeded? The answer is probably using the right metrics, right? However, we must dig deeper. The real metric is not precision, recall, and all those statistical metrics.

The critical metric is impact.

As Data Scientists, we must use our data to make as much impact as possible to our company. The impact can be in the form of multiple things. These two are some of the essential forms of seeing the impact.

Insight: extract useful knowledge to guide your project development, providing direction. For example, with an exploratory analysis of real estate market data, you discovered that around 80% of the buyers are men. However, their wives are the ones that end choosing mostly everything in the house. That information is beneficial to direct advertisements to the right person.

Data products: developing products that use data to provide something in return (usually money). For example, a product that can predict the glucose levels of a diabetic person. With that, this person can use this knowledge to eat and use insulin at the right time.
What tool can I use to make an impact?

*Any, nobody cares about it.*

The most important thing to determine which technique to use is to understand the data. Then let the data talk about how it works. Instead of fitting a model to the data, a Data Scientist needs to know to fit the data into a model (the one that works better). Rachel Tatman has a quite good presentation on “So what method should you use?”, she ends her presentation with his phrase:

*If it is stupid but it works, it is not stupid.*

This means, if our model is straightforward (like hand-built rules from the 90s) and it still works great, then is not stupid at all.

Then, let’s imagine that I built an application that can make a huge impact. How to sell it? For example, you are working with a dataset from your company, and you discovered a pattern nobody saw. How to present it to your team?

**Data Storytelling**

As a combination of data, narrative, and visuals, Data Storytelling is one of the new, hot skill every Data Scientist will need in the near future.

Data Storytelling is a fundamental technique for a Data Scientist or Data Analyst who needs to explain his or her line of thinking to others, whether they have the technical knowledge or not. It is at that moment that your communication skills will be very important. After all, no one likes to see a lame presentation or read an annoying text, right?

Briefly, Data Storytelling is the act of explaining what you did, how you did it, and why you did it all in a way that keeps your reader or listener engaged.
People hear statistics, but they feel stories. (Forbes)

The goal of Data Storytelling is to tell the story of your data. It is making sense of the data humanly. Personifying the data as a living scene and let others understand the data with empathy, not only statistics. It may involve a combination of three key elements: data, visuals, and narrative.

When the narrative is coupled with data, it helps to explain to your audience what’s happening in the data and why a particular insight is important. Ample context and commentary are often needed to appreciate an insightfully. When visuals are applied to data, they can enlighten the audience to insights that they wouldn’t see without charts or graphs. Many interesting patterns and outliers in the data would remain hidden in the rows and columns of data tables without the help of data visualizations.

Finally, when narrative and visuals are merged, they can engage or even entertain an audience. It’s no surprise we collectively spend billions of dollars each year at the movies to immerse ourselves in different lives, worlds, and adventures. When you combine the right visuals and narrative with the right data, you have a data story that can
influence and drive **change**.

Mostly you must use the narrative, visuals, and data to make the complex simple. Using it to make everybody understand an obscure insight and think, “how I have not saw this before?”.

**Tips to be better at Data Storytelling**

There are already some articles and stories about Data storytelling tips. Some of those are below:

1. **10 Ways to Take Your Boss on a Journey Through the Data**
   
   “People hear statistics, but they feel stories.” Forbes, March 2016
   
   [towardsdatascience.com](https://towardsdatascience.com)

2. **Mastering Data Storytelling: 5 Steps to Creating Persuasive Charts and Graphs**
   
   Data storytelling is one of those "buzzwords" that in actuality is not really a buzzword-it's reflective of a necessary...
   
   [www.crazyegg.com](https://www.crazyegg.com)

3. **10 Tips for Better Data Storytelling**
   
   By Bill Shander.
I would add one last tip here. I believe it is the most important one.

Read, a lot.

Reading is essential to understand the function and history of our society. It is vital to develop our mind by discovering new things. Also, reading is a crucial aspect when we are talking about imagination and curiosity.

The more you read, the more you understand the subject you read. Now linking it back to Data Science, data is about everything. A reader can learn concepts faster about anything he reads. This can make a huge difference giving sense to data, giving the needed background to the numbers so that numbers can become ideas and insights.

The more that you read, the more things you will know. The more you learn the more places you will go. -Dr. Suess

References


---

**Data Scientists Are Thinkers**

Execution vs. exploration and what it means for you

towardsdatascience.com