

# Time is Money

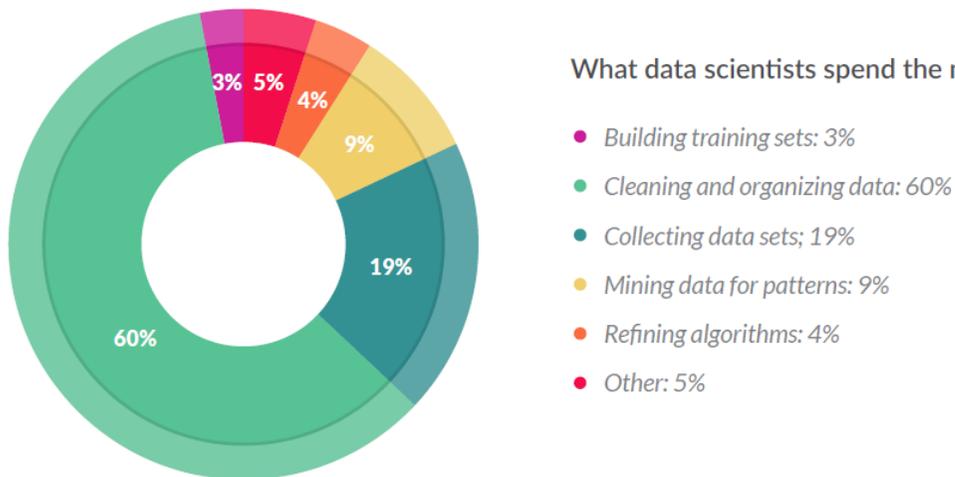
FAIR Time Effectiveness

**AXVECO**

*Sustainable Innovation*

# Reality versus Opinion

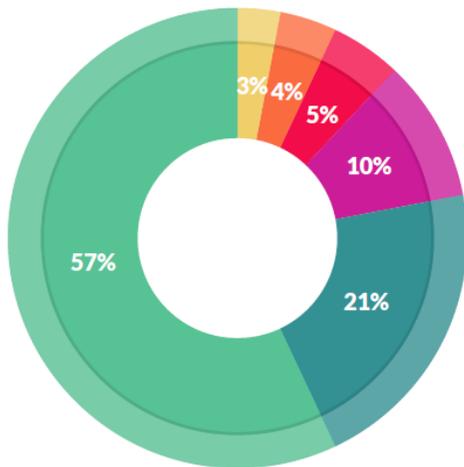
Here's where the popular view of data scientists diverges pretty significantly from reality. Generally, we think of data scientists building algorithms, exploring data, and doing predictive analysis. That's actually, not what they spend most of their time doing, however.



As you can see from the chart above, 3 out of every 5 data scientists we surveyed, spend most time cleaning and organizing data. You may have heard this referred to as “data wrangling” or compared to digital janitor work. Everything from list verification to removing commas to debugging databases—that time adds up and it adds up immensely. Messy data is by far the more time-consuming aspect of the typical data scientist’s workflow. And nearly 60% said they simply spent too much time doing it.

# Reality versus Preference

Simply put, data wrangling isn't fun. It takes forever. In fact, a few years back, the New York Times estimated that up to 80% of a data scientist's time is spent doing this sort of work. The problem here is twofold. Data scientists simply don't like doing this kind of work, and, this kind of work takes up most of their time. We asked our respondents what was the least enjoyable part of their job. They had this to say:



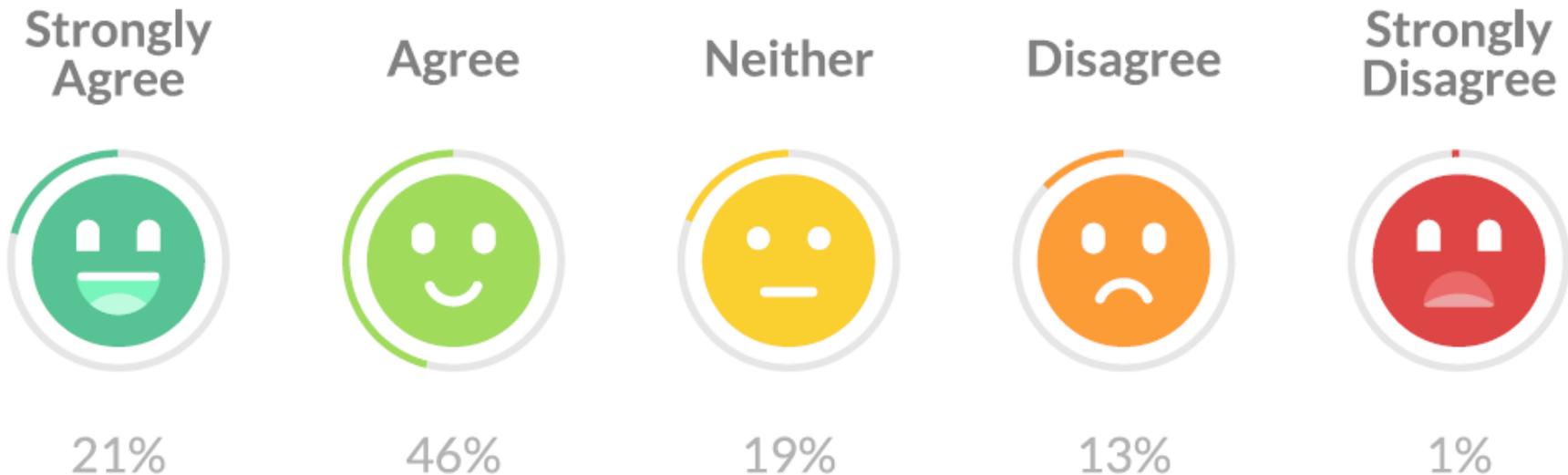
What's the least enjoyable part of data science?

- Building training sets: 10%
- Cleaning and organizing data: 57%
- Collecting data sets: 21%
- Mining data for patterns: 3%
- Refining algorithms: 4%
- Other: 5%

Note how those last two charts mirror each other. The things data scientists do most are the things they enjoy least. Last year, we found that respondents far prefer doing the more creative, interesting parts of their job, things like predictive analysis and mining data for patterns. That's where the real value comes. But again, you simply can't do that work unless the data is properly labeled. And nobody likes labeling data.

# Enabling existing staff

We asked our respondents to agree or disagree with the following statement: I have access to the tools I need to do my job effectively. Here's what they said:



It's notable that only 14% of respondents felt they were being held back by their tools. That evidences that, while there may not be enough data scientists, their organizations are committed to giving them the best possible chance at success. And that's never a bad thing.

# Recommendation

As more and more organizations adopt data as a key driver of decision making, the importance of streamlined, well-oiled data science teams is going to remain paramount. But the current status quo probably isn't sustainable.

On the one hand, we see a shortage of data scientists while on the other, they're spending too much time cleaning and munging data. This is time that could be much better served doing predictive analysis and building out machine learning practices. That's not to say that cleaning and labeling data isn't important, of course. Analysis on bad data is a garbage-in, garbage-out sort of scenario.

Rather, organizations that want to get the most of their data should aim to fix the problems their teams have now. They should talk to them and find out exactly what takes up their time. By mitigating the effort their teams spend doing janitorial data work, they'll be able to empower their teams to do the valuable tasks that data scientists actually enjoy doing.

# Other cost indicators

Apart from the availability of sufficient staff, their effective time spending and availability of their tools there are several other cost components that can easily be investigated and surely improved. This is a template from a PWC market study.

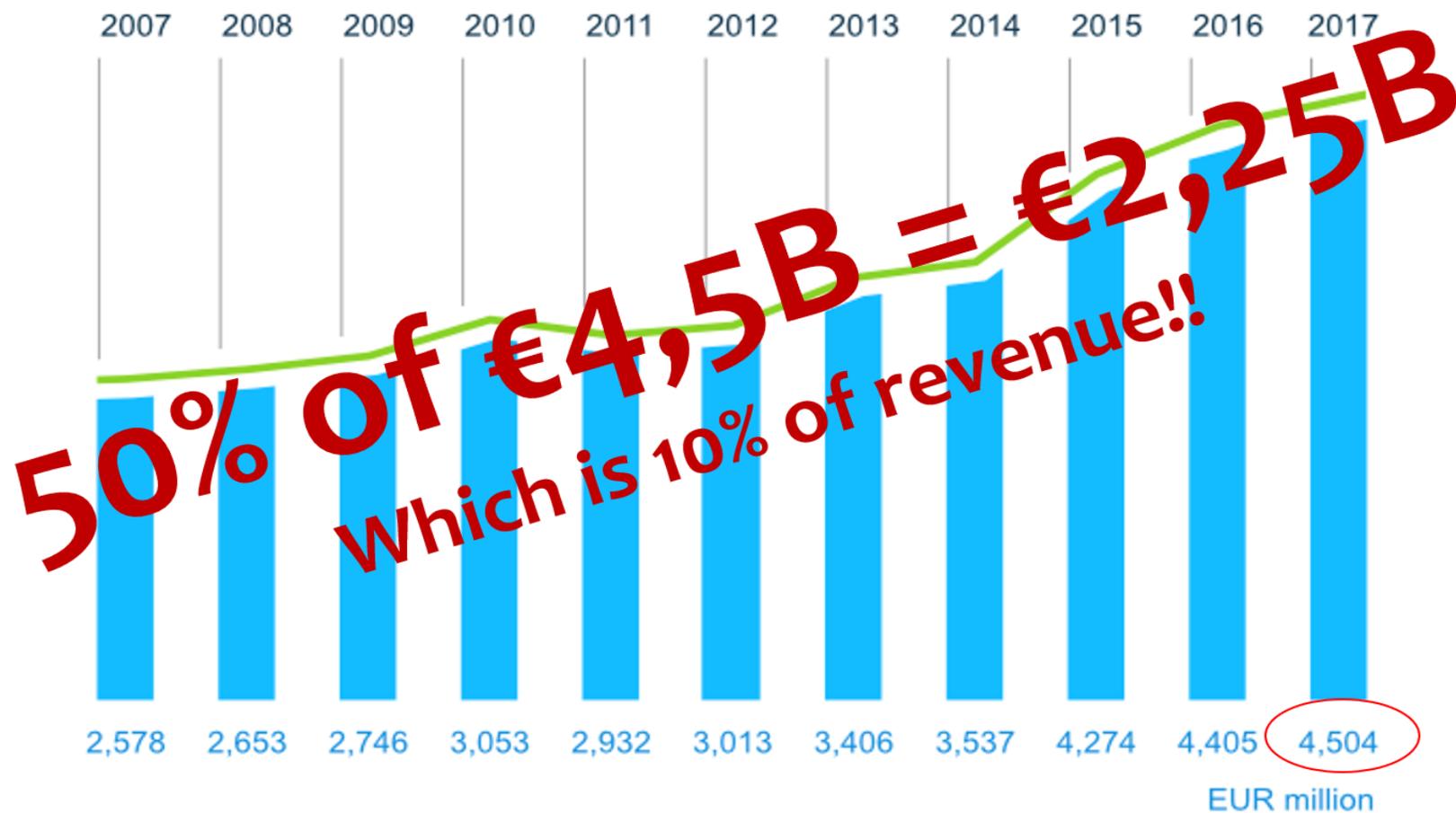
<b>Indicator #1 : Time spent</b>	<b>Indicator #2 : Cost of storage</b>	<b>Indicator #3 : Research retraction</b>	<b>Indicator #4 : Research duplication</b>	<b>Indicator #5 Potential economic growth</b>
<ul style="list-style-type: none"> <li>• Split academics &amp; non-academics</li> <li>• EU28 Average researcher salary</li> <li>• Number of researchers</li> <li>• Time <u>dedicated</u> to research</li> <li>• Time <u>wasted</u></li> </ul>	<ul style="list-style-type: none"> <li>• Data volume per researcher/year</li> <li>• Cost per TB</li> <li>• Number of repositories where data is stored</li> <li>• Data retention period</li> <li>• <u>Reduction</u> of the number of repositories where data is stored (FAIR)</li> </ul>	<ul style="list-style-type: none"> <li>• Split academics &amp; non-academics</li> <li>• EU28 Average researcher salary</li> <li>• Number of researchers</li> <li>• Time <u>dedicated</u> to research</li> <li>• Time <u>wasted</u> due to retraction</li> </ul>	<ul style="list-style-type: none"> <li>• Funds allocated to research grant in EU28</li> <li>• Number of research grant in EU28</li> <li>• Number of suspicious overlaps</li> </ul>	<ul style="list-style-type: none"> <li>• Positive impact on the number of citations</li> <li>• Societal value of research increased</li> <li>• Improve the availability of research data</li> <li>• Benefits for making research more accessible</li> </ul> <div data-bbox="1632 1039 1943 1115" style="border: 1px solid red; padding: 5px; text-align: center;"> <p><i>Insufficient quantitative data</i></p> </div>



$$\#1 \text{ in } \text{€} + \#2 \text{ in } \text{€} + \#3 \text{ in } \text{€} + \#4 \text{ in } \text{€} = \text{Total } \text{€}$$

# Cumulated cost and impact - example

## Research and Development Expenses 2017



# Recommendation – increase 10% in revenue

Make an inventory of your cost elements and pick the most relevant for your strategic goals. Define that area for further investigation and note what the budget cost are for research and development.

Set a goal for savings and discuss with your best capable staff that is responsible for the defined area.

Read the various subjects on this process and define the project on a small experiment inhouse.

Let us know if and how you have taken this challenge and what the findings were, we like to be surprised.

