

7th KnowHow Meeting (ONLINE) ... 27 Aug 2020

(2) Data Science ... Data Mining ... Analysen ... Visualisierung ... KI

www.DataCampus.eu

Data requires the human factor

Martin Treder

(ReturnOnData)

Data Requires The Human Factor

What you need beyond hardware, software and algorithms

Martin Treder

Intro

Who am I?

- **Martin Treder, Mathematician, Data Consultant and author, with 30 years of data experience.**
- **Focus Areas: Data Governance, MDM, Data Quality, Analytics, Data Science, business cases**

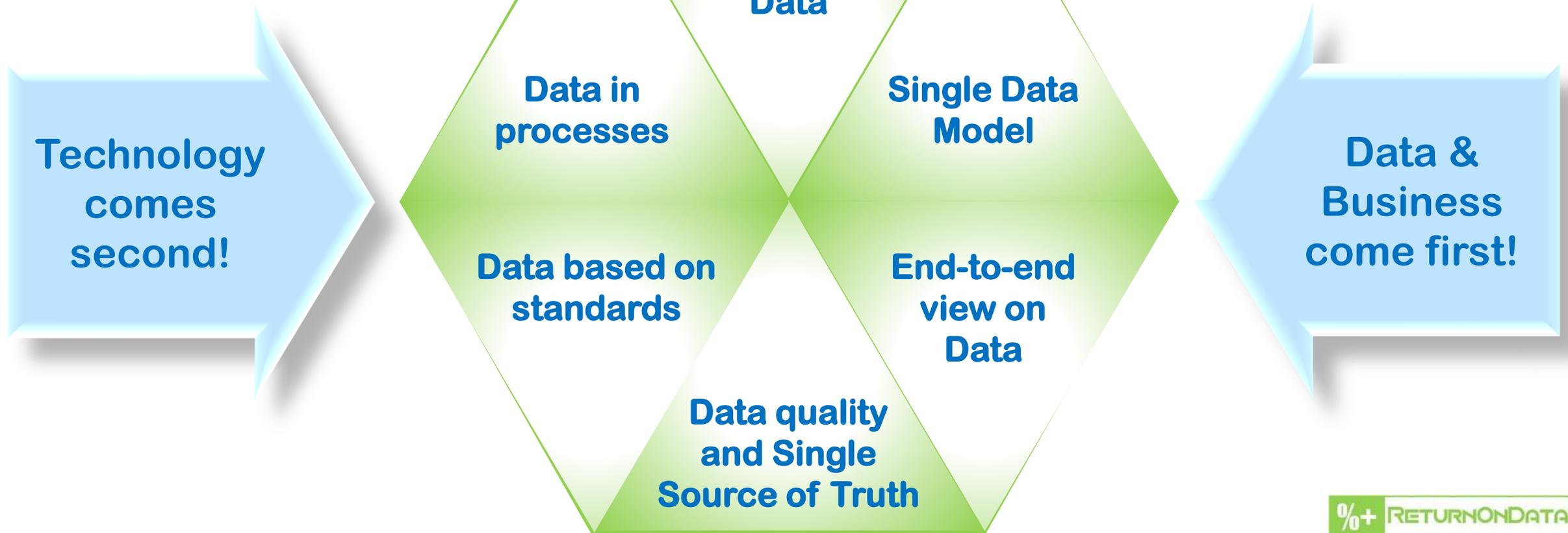
What is my message?

- **Effective management of data requires more than specialist knowledge and technology.**
- **Be a communicator, a salesperson, a diplomat and a psychologist!**



Data In Organisations – Where Are We Today?

What do we need?
The *Magic Hexagon*:



Data In Organisations – Where Are We Today?

But are we well-prepared?



Data In Organisations – Where Are We Today?

Don't forget the *Human factor!*

Science

*Maths, algorithms,
data models*

Hardware

*GPUs, in-memory,
Cloud, bandwidth*

Me...

Tools

*Databases. Languages
like C++, R or Python.
TensorFlow and PyTorch.*

- 1) **How do I personally deal with all the data?**
- 2) **How do I interact with the people around me?**

Data

*Masterdata, Reference Data,
Big Data, IoT Data, Web data,
Social Media data, ...*

The Human Factor in Data Management

1) How do I deal with the data?

2) How do I interact with others?

"It looks Scientific!"

Simple Example: ANOVA: The data is available, the algorithms are precisely defined and implemented, the result looks precise. Yet... may there be room for human bias...?

“This test can only [be] use[d] when a number of preconditions are in place. These are that all groups must either contain more than 30 observations or be normally distributed also that there must be comparable variances across groups.

Furthermore, the variable describing the groups should be nominal-scaled while the dependent variable (variable we compare with, or just 'the other! ") should be interval-scaled. In some textbooks, you will also be able to read that the groups should be similar size, do not have to worry about in SPSS, as it has a procedure to correct for unequal sample sizes.”

Business Analytics for Managers, <http://www.ba-support.com/doc/stat/Content/anova/anova.htm> (called 2019-07-18)

"It looks Scientific!"

This example indicates
our responsibilities as
data people!

Simple Example: ANOVA

And here it is, the human bias...

“This test can only [be] use[d] when a number of preconditions are in place. These are that all groups must either contain **more than 30 observations** or **be normally distributed** also that there must be **comparable** variances across groups.

Furthermore, the variable describing the groups **should** be nominal-scaled while the dependent variable (variable we compare with, or just 'the other! ") **should** be interval-scaled. In some textbooks, you will also be able to read that the groups **should** be similar size, do **not have to worry** about in SPSS, as it has a procedure to correct for unequal sample sizes.”

Business Analytics for Managers, <http://www.ba-support.com/doc/stat/Content/anova/anova.htm> (called 2019-07-18)

8 Data Handling Principles

(1) Consider Data Lineage

No data without provenance: This applies to all of its aspects:

Content

- Where does the data come from?
- Who guarantees its correctness?
- Is it *still* accurate?

Logic

- Have data sources been merged, joined, filtered?

Verbiage

- Did the creator use the same language?

Model

- Is the data model clear and known?

Resist the temptation to send data files around on request
Resist the temptation to use data files sent by someone

8 Data Handling Principles

(2) Understand your maths

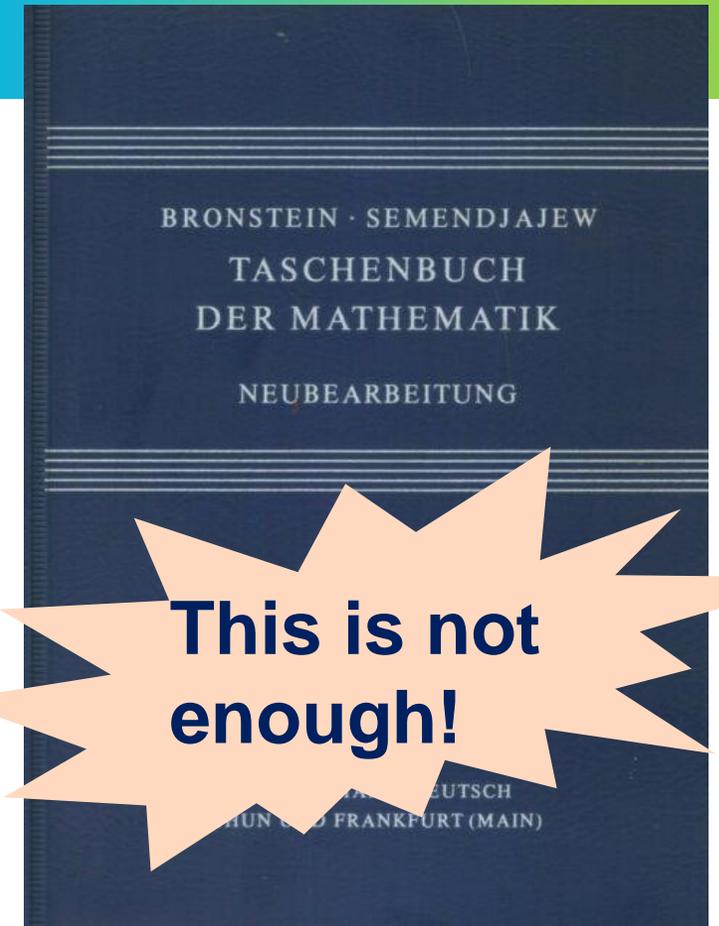
Are the right formulas sufficient?

What is the most important maths book for physicists?

Bronstein / Semendjajev!

What is their most important maths knowledge?

On which page do I find the formula?



If you don't understand the formulas you are using, how would you find out if you got it wrong???

8 Data Handling Principles

(3) Check the preconditions and assumptions

Most models are perfect – *as long as all preconditions and assumptions are given* (which is never the case, of course).

So, what is ruining the show? It's not the algorithms! It's reality!

*A function is **not** normal-distributed as soon as a plot of the data points looks like a Gaussian Bell.*

*A sample is **not** representative just because there was no conscious bias*

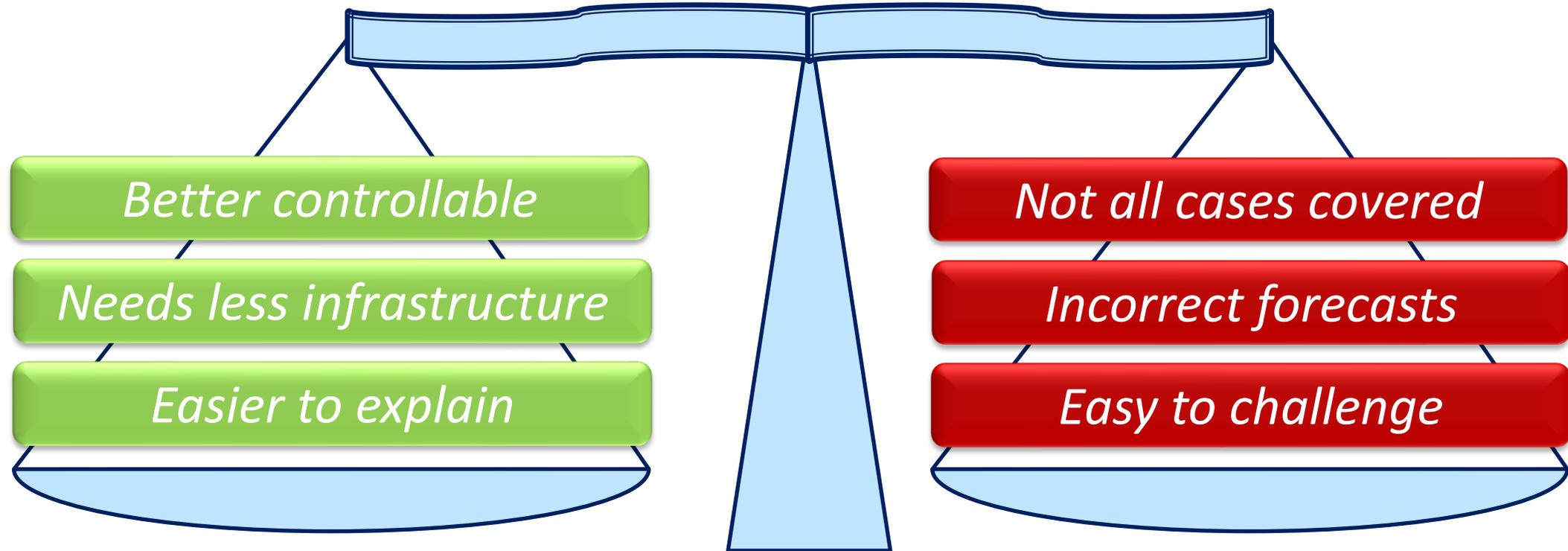
Data Scientists need to invest in this step. Nobody else can!

8 Data Handling Principles

(4) Simplicity is not a target in its own right

Let's always balance simplification with accuracy!

Reality is complex. Simplification moves models away from reality. Always!



As simple as possible – as complex as necessary!

8 Data Handling Principles

(5) Develop A Common Language

What is a
"customer"?!?

Babylon effect: The lack of a common languages kills projects

Data are more complex than code

- Most programmers speak the same language, and open-source collectives mainly solve technical problems
- People in charge of data often come from different industries without a common vocabulary and talk business

(The Economist, 23rd April 2020)

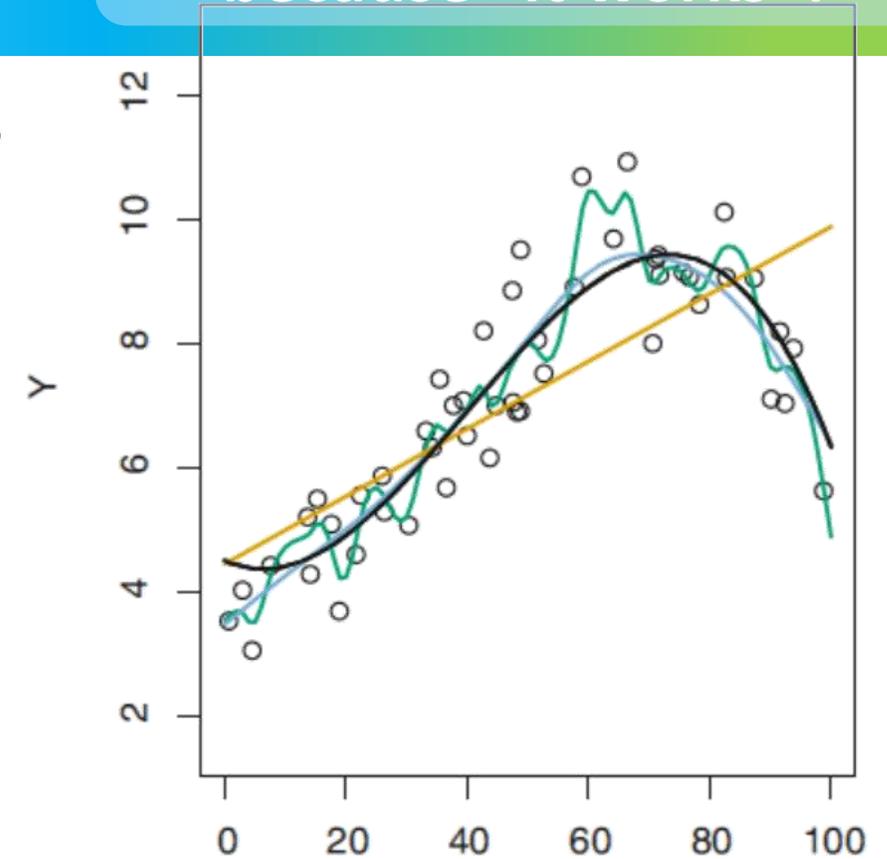
Work on unambiguous terminology, together with the business owners of the data domains!

8 Data Handling Principles

(6) Invest In The Model

- a) Simplistic exponential functions quickly go through the roof
- b) Polynomials are tremendously volatile outside the observed range
- c) Linearity is often limited to a certain range
- d) Overfitting often leads to plausible but wrong single results
- e) Underfitting is worst in combination with extrapolation

Don't use a formula just because "it works"!



From: Gareth James et.al. "An Introduction to Statistical Learning", Springer Nature

Clarify acceptance of simplifications and limitations!

8 Data Handling Principles

(7) Don't Convey A False Impression Of Preciseness

The scientifically calculated length of each side of this square is “1.41421356 inches”. How precise is this figure?



SQUARE

Let's look at the history behind this result:

- *Person A estimated the square's surface area to be 2 square inches*
- *Person B has applied the (mathematically correct) formula that "the length of a square's side is the square root of the surface area"*
- *The pocket calculator of Person B said: **1.41421356***

Always disclose the weakest link of the chain!

8 Data Handling Principles

(8) Minimise The "Bias Of The Investigator"

- You probably agree: *The result of an assessment should be independent of the person designing the model.*
- The important mathematical principle of "reproducibility" is useless if not applied to ALL steps of data exploitation!
- Do you always apply this principle consciously???

Become aware of your (subconscious) favourite approaches, and avoid them consciously!

The Human Factor in Data Management

1) How do I deal with the data?

2) How do I interact with others?

The Seven Interpersonal Secrets

(1) Collaborate

The Optimisation Principle:
The best data approach for the entire company
will always be better than
the sum of all functions'
best data approaches

- We should share our experience
- We should share best practices
 - We should leverage synergies
 - We need to get to consistent messages
 - We should not re-invent the wheel
- Different functions should work together
- Locations and entities should work together

The Seven Interpersonal Secrets

(2) Improve Your Storytelling

You always need a story.

Actually, you need multiple stories...

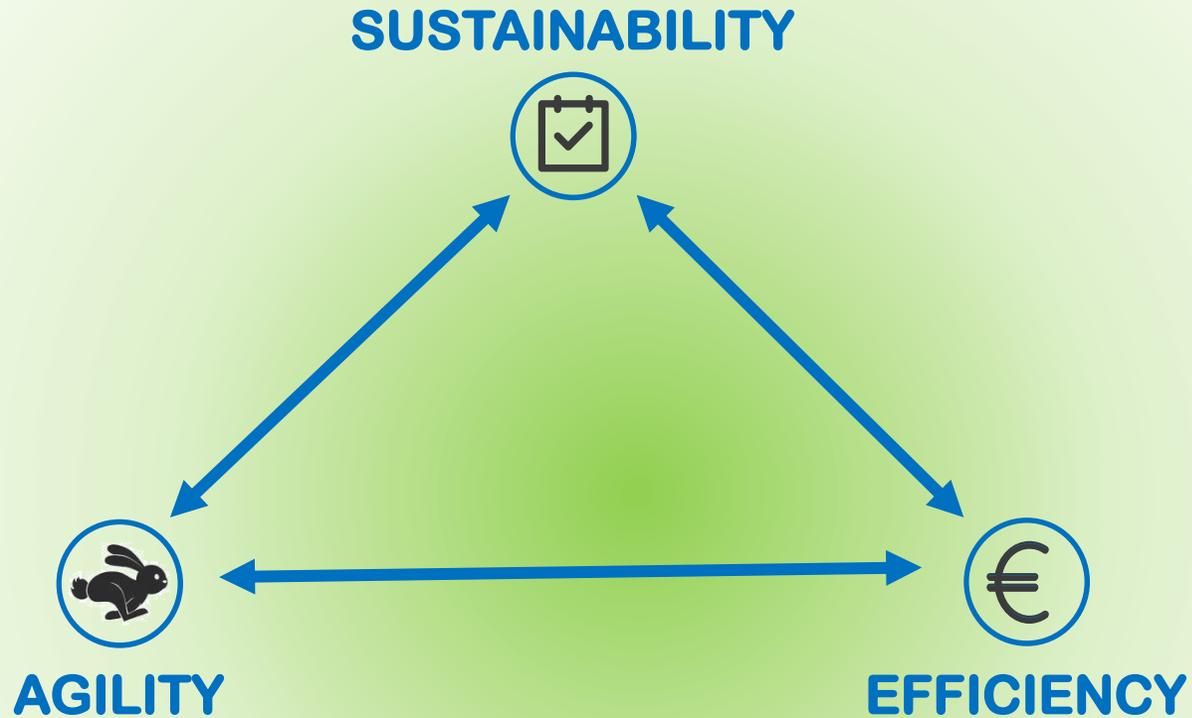
- Each stakeholder values different benefits
- Each group needs a different language:
Upper management, IT people, other data folks, staff, ...



**Your success depends on your
perceived power to solve people's problems**

The Seven Interpersonal Secrets

(3) Learn To Work With Conflicting Targets



The Seven Interpersonal Secrets

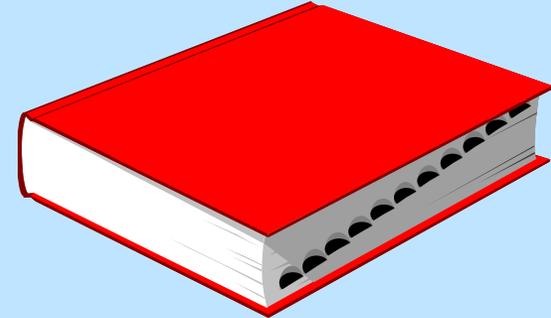
(4) Ask The Right Questions

~~Is...?
Has...?
Could...?
Should...?
Will...?~~

WHY...?

WHAT...?

HOW...?



The Seven Interpersonal Secrets

(5) Share Your Insights

Explain ...

... The idea of your model and algorithm

- Use business language
- Use drawings and two- or three-dimensional examples

... Your assumptions

- It is not a weakness to state *"My statement is only valid under the following assumptions..."*

... The quality of your data

- Including its origin, age and reliability

... The preciseness of your model

- It is not a weakness to state *"This is all we can say..."*

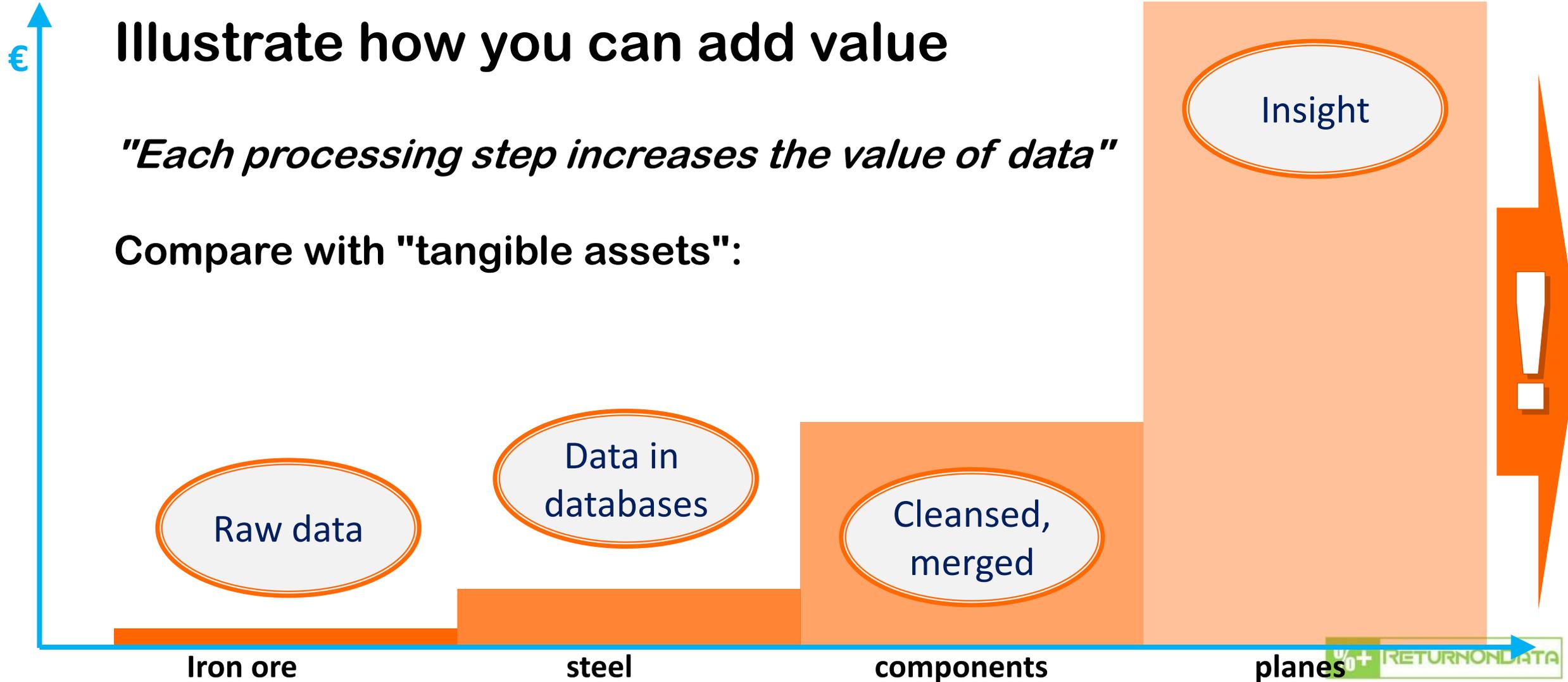
The Seven Interpersonal Secrets
(5) Share Your Insights

Explainable AI:

**Data Science should never appear as a
black box!**

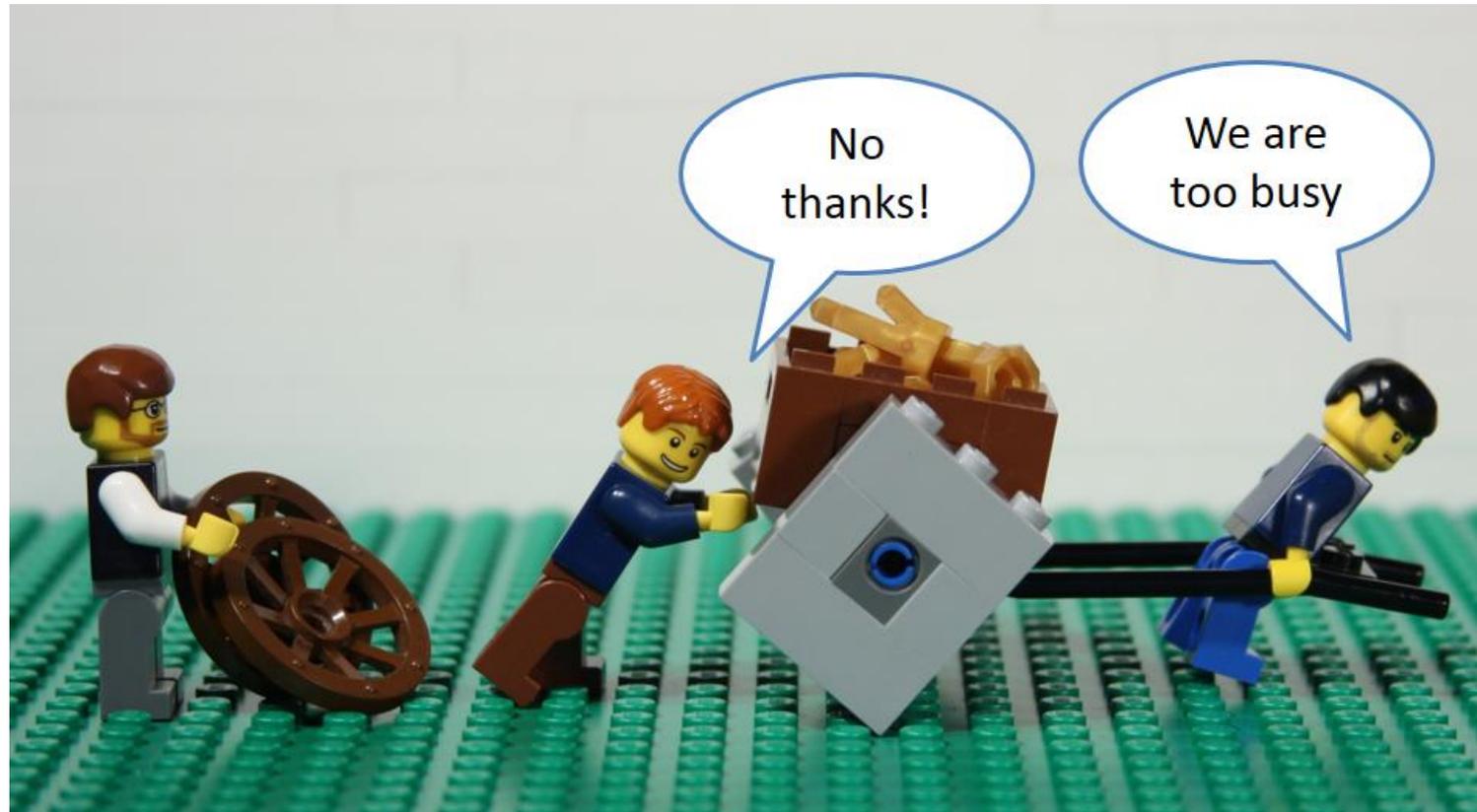
The Seven Interpersonal Secrets

(6) Explain The Commercial Opportunities



The Seven Interpersonal Secrets

(7) Offer Your Support – Be Modest And Patient...



Data Management Organisation

Thank you!

Advertorial 😊

You find my first book at Amazon, or directly via the publishing house

amazon

Books

Deliver to Treder, Königswinter 53639

Today's Deals Martin's Amazon.com Help Browsing History Registry Buy Again

Shop Happy H

Books Advanced Search New Releases Best Sellers & More Children's Books Textbooks Textbook Rentals Sell Us Your Books Best Books of the Month

Books > Computers & Technology > Networking & Cloud Computing

Look inside

Becoming a data-driven Organisation: Unlock the value of data Paperback – December 16, 2019

by Martin Treder (Author)

> See all 2 formats and editions

Kindle \$40.68	Paperback \$54.99
-------------------	------------------------------------

Read with Our **Free App** 1 New from \$54.99

Share

Pre-order
Qty: 1

This title has r
Ships from an

Pre

53639

Add to List

%+ RETURNONDATA

addressed the handling of data so far.

Yet, more and more business leaders have become aware of the topic. They recognize the increasing relevance of data, and the need to act now. Those leaders will welcome this book as it guides them through the first steps in their journey towards a data-driven organization.