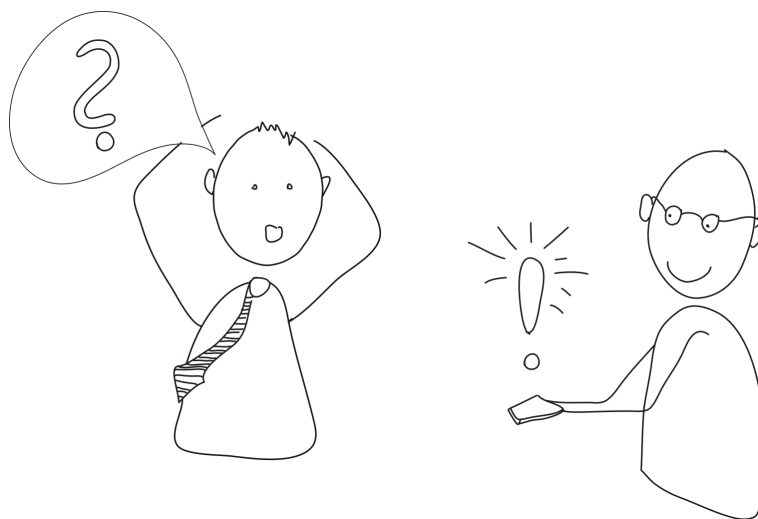


Hitchhiker's Guide to a Data Science Use Case

Introduction

The first step in a successful data science project is understanding the use case. This guide is based on over twelve years of experience in data science that showed us there is no simple one-fits-all solution! Therefore, this guide does not offer answers, but navigates you through your data science project, asking those questions that usually come up only after you have hit a dead end. Take some time to consider the following questions at or even before the start of your project – you will reap the rewards soon enough: You will not only be able to avoid costly misinvestments, but to gain a deeper understanding of your use case. This will enable you to see in which areas you might need further expertise. This guide will help to successfully carry out your project and establish data science in your business. Answering the questions should take about 15 minutes. You can either use the results yourself or contact us – based on your answers we can give individual advise and plan the next steps towards your data science project.



Use Case Definition

Type

- Research project
- Proof of concept
- Business scenario

Deployment

- In Scope
- Out of Scope
- What's this?

Organizational Prerequisites

The project is initiated by

- an individual person,
- a team,
- a department or
- management.

Initiator's experience

- First project
- First project in productive use
- One project among many

The company's data science team

- is involved,
- is responsible only for certain aspects,
- just stores the data,
- is not involved as of yet or
- does not yet exist.

The company's central IT dept.

- supports,
- is actively involved or
- does not know about the project.

Has the budget already been determined?

- Yes
- No

Has the technical infrastructure been determined?

- Fully
- Partly
- Not at all

Target Group

Which group will be using the project results?

Describe this group

- Professional experts
- Managers
- Workers
- Machines
- Other: _____

Significance of the results regarding this group's work

- Relevant for the business
- Decision support
- Nice to have
- Looks good
- Distractive



The results are to be presented and used in the following forms:

- Web services
- Interactive web applications
- Interactive reports
- Excel reports
- Powerpoints

Planned degree of automation

- Fully
- Decision support
- Feedback loop with humans
- Generating insight

Which problem the end user group has will be treated with this project?

Does the project need the end user group's expert knowledge in order to succeed?

- Yes
- No

Effects

How will the project's results support the end users' in their daily tasks?

Effects of the results

- Job loss due to automation
- Causes additional work
- Causes disruptive factors
- Simplifies daily tasks

How will the results be integrated into the end users work routine?



Value creation – savings per year [€]

- Inestimable
- > 10.000
- > 100.000
- > 1.000.000
- Business critical

Effects on the company

Effects on society and humankind

Data

Data sources

- I/someone in my team owns the data
- Public data
- Third-party data
- Data of a different department
- Data from the company's data lake

The data is/was generated by

- Sensors
- Machines
- Human interactions recorded by machines
- Manually created

The data is available as

- Raw data
- Aggregated data
- Reports

Ground Truth

- Available, as future events are to be predicted
- Available because manually created
- Does still have manually created
- What's this?

Historic data is available for

- > 5 Jahre
- > 1 year
- > 3 months
- Only just started gathering data
- No historic data available yet

Size of relevant data in total

- > 1 TB
- > 100 GB
- > 1 GB
- Less



Biggest data set for one aspects

For example, if you have a data set with informatio about many machines, how big is the part of data regarding just one machine as being a unit to be analysed?

- > 1 TB
- > 100 GB
- > 1 GB
- Less

Infrastructure

Available storage format for data:

- Hadoop
- NoSQL Databases
- SQL Databases
- Excel
- Other: _____

Experience with Hadoop:

- Good
- Beginner
- None

Experience with NoSQL:

- Good
- Beginner
- None

Experience with SQL:

- Good
- Beginner
- None

Preferred platform for analysis:

- Graphic
- Code-based

The data is comprised of

- Tables
- Texts
- Pictures
- Sound files
- Other: _____

Is the platform for analysis already decided upon?

- Yes
- No

Experience with existing platform:

- Good
- Beginner
- None

Available Hardware:

- Everything available
- Cloud-based
- Virtual Standard Server
- Office PCs
- No specific hardware

Requirements regarding deployment (applying the results)

- Central
- Spread out worldwide
- Virtually real-time deployment (Milliseconds to answer)
- Online deployment (Seconds to answer)
- Offline deployment

