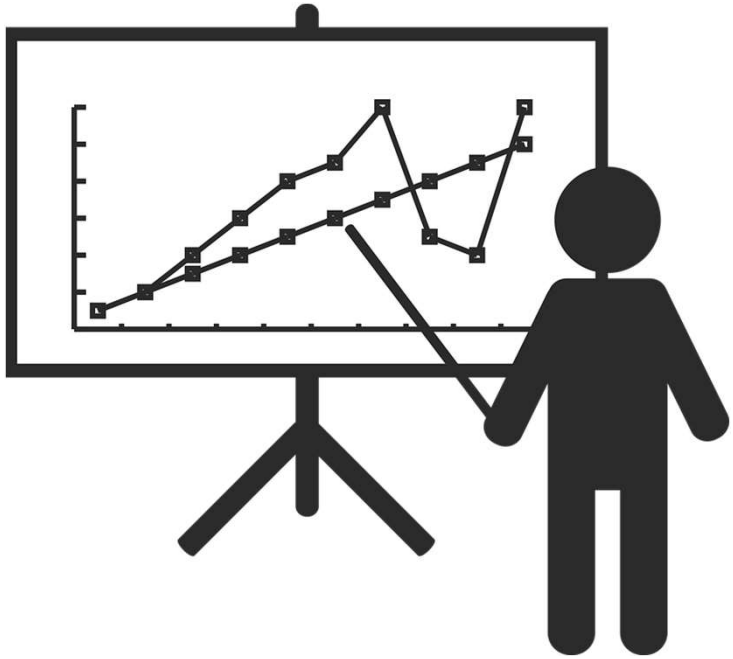
Task 1

Deadline/Presentation 11.05
(Upload at Felix before the lecture,
use email if this does not work)

Skills of a Data Scientist

Generating
Insights from Data



Communicating
Insights

Technology to use

- The task is to be implemented with Python and optionally other tools (e.g. for additional visualization).
- The majority of the work is to be implemented with Python.

Presentation requirements

- **Group size max 4 (→ register with Felix)**
  - Individual evaluation possible: → register in advance and clearly mark contribution of each group member
- **Insights gained through the analyses will be presented and interpreted**
  - Made-up target group: Board of Directors (CEO/CTO) of SmartBuild
  - Consider implication on SmartBuild's business.
  - Implemented solution is explained and demonstrated (if applicable)
  - Made-up target audience CTO/BI department of SmartBuild's
- **Technical concepts of the solution will be explained**
  - E.g. architecture diagrams, algorithms
  - Made-up target audience BI department SmartBuild
  - Hand in code+slides

Grading of Presentation

- **Timing!**
- Completnes of the results
- Reflection on results
  - What value can we derive from your insights?
  - Why was the analysis done in this particular way?
  - What could be improved (if you had more time and resources)?
- Is the presentation well tailored to the target group (easy to comprened)
- How you present
- Discussion

Duration of Presentations

- Duration of Presentations
  - 1 person: 15min
  - 2 persons: 20min
  - 3 persons : 22min
  - 4 persons: 25min

Background/Motivation for tasks

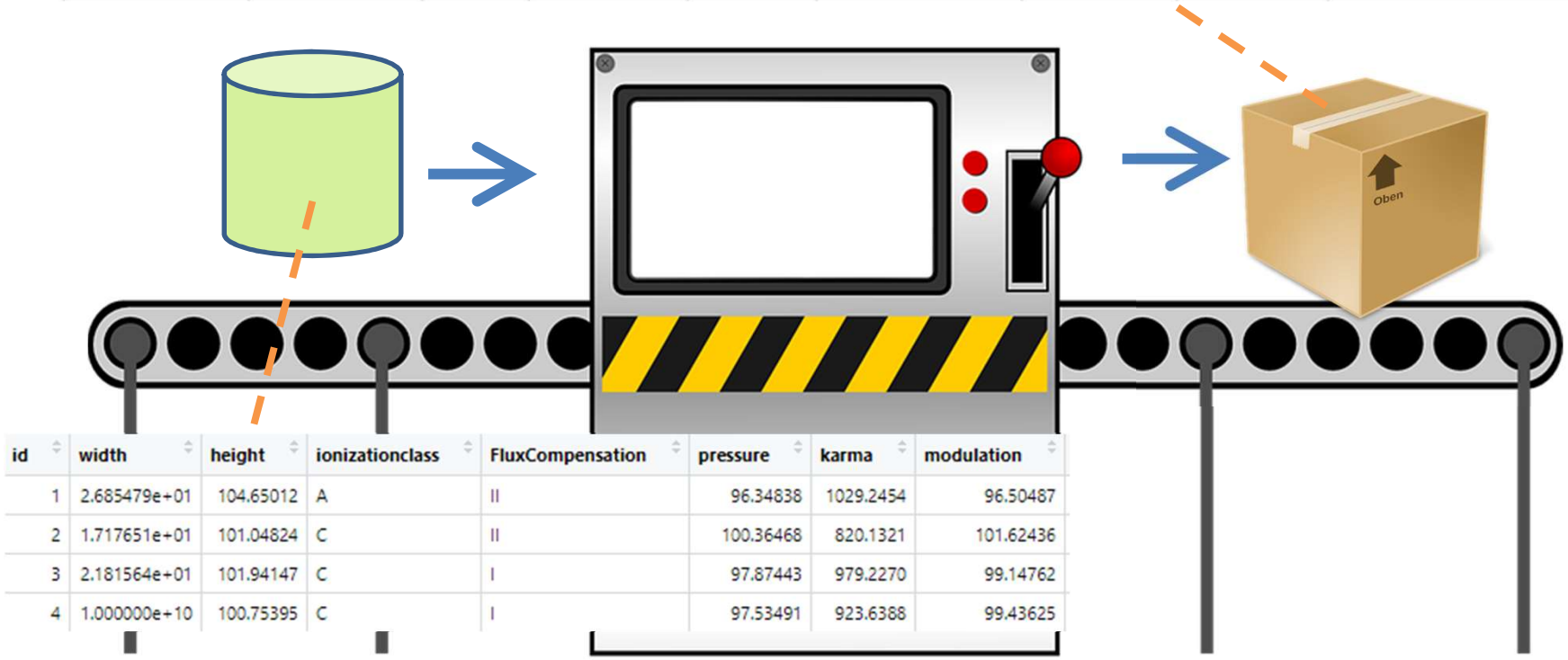
- Made-up manufacturer SmartBuild
  - SmartBuild builds a final product from raw material in production lines
  - The properties width, height, ionizationclass, FluxCompensation, pressure, karma, modulation of the raw material are measured on input
  - Properties of the resulting end product are also measured (other attributes in the data)
  - SmartBuild wants to check if the properties can be predicted at the input to optimize the production (e.g. avoid rejects/defects, increase quality).

Background/Motivation for tasks

- You are a consulting team with expertise in Data Science
- You have been hired by SmartBuild to explore the potential of Data Science technologies for the company's manufacturing operations

SmartBuild Data Records

weight_in_kg	weight_in_g	error	error_type	Quality	reflectionScore	distortion	nicesness	multideminsionality
7.548250e+04	7.548250e+07	yes	severe	95.54875	718.5242	51.73237	2.566950e+01	yes
2.982253e+04	2.982253e+07	no	None	99.48593	740.9946	51.34866	1.876470e+01	no
4.852630e+04	4.852630e+07	no	None	97.87388	684.9130	49.83995	2.030074e+01	yes



id	width	height	ionizationclass	FluxCompensation	pressure	karma	modulation
1	2.685479e+01	104.65012	A	II	96.34838	1029.2454	96.50487
2	1.717651e+01	101.04824	C	II	100.36468	820.1321	101.62436
3	2.181564e+01	101.94147	C	I	97.87443	979.2270	99.14762
4	1.000000e+10	100.75395	C	I	97.53491	923.6388	99.43625

Data for the Task

- The data is included in the file manufacturing.csv at Felix (folder Task_01)

Assessment of the Technical Solution

- Evaluation criteria
  - Overall qaulity of the solution
  - Depth of experiments/tests
  - Sound evaluation/interpretation of the results
  - Reasonable design of the models (features, parameters)
  - Solution for pre-processing of data (if available)

Assessment of the Presentation

- Evaluation criteria
  - Timing
  - Targeted to the Made-up target group CEO+CTO of the Made-up manufacturing company to be consulted.
  - Note: The company wants to understand your results and the technical foundation of your solution.
  - Potentially they would like to integrate your solution into their corporate IT

Tasks

- Address Q1,Q2
- Create predictive models for further product characteristics so that they come to at least 2 analyses per person (different models)
- Explain the concepts you used for data analysis (for the target group)
- Evaluate and review your results
- **Note: If applicable, you can also show that a good prediction (using the models we have covered) is not possible.**

Q1

- Create a predictive model for the attribute "weight_in_kg".

Q2

- Create a predictive model for the "error_type" attribute

Q2 Alternative

- Create a prediction model for the "error" attribute

Use of LLMs (e.g. ChatGPT)

- Currently there are no uniform guidelines for the handling of LLMs in teaching.
- For this course the following applies until further notice:
  - LLMs are treated like existing programming tools or help pages.
  - ***You must be able to explain your solution in detail (i.e. you must understand what you are doing). → Presentation discussion***