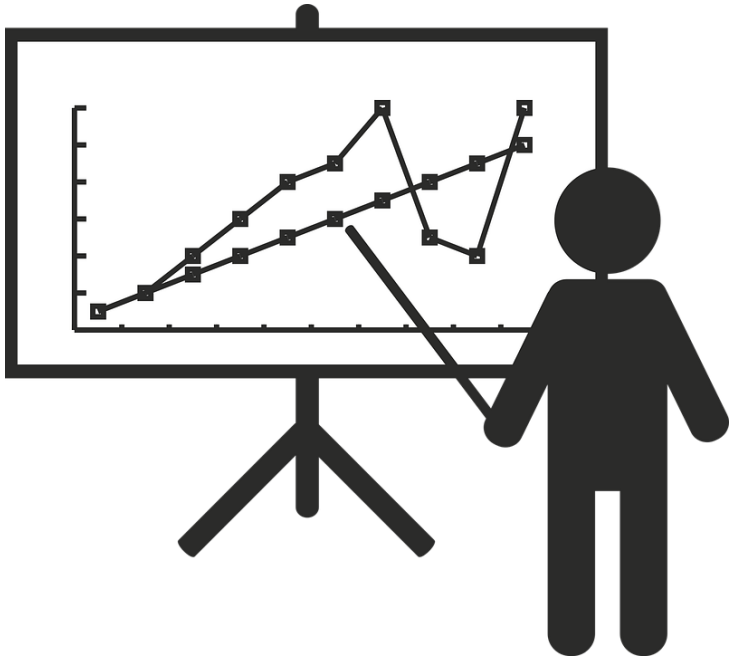
Task 02

Deadline 29.06.
Slides must be uploaded to Felix before the lecture.
Use email if there are technical problems with Felix.

Data sets for this exercise were generated with the help of chatGPT.

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 5 (→ 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
  - 5 persons: 28min

Background/Motivation for tasks

- You are a consulting team with expertise in data science
- You have been hired by EU-Park to explore the potential of data science technologies for the theme park's operations

More specific Background/Motivation for tasks

- Made-up theme park EU-Park
  - The theme park “EU-Park” operates several rides.
  - You are tasked with analyzing the wait times.
  - EU-Park aims zu understand
    - What impacts the wait times?
    - How and how well can wait times be predicted?
    - …
  - EU-Park aims to
    - Determine busing times to plan head with the needed amount of staff.
    - Optimize the days available for “limited” season tickets, where owners have access to the park 250 days a year.
    - Offer visitors a prime service with short term wait time predictions (so that they can optimize their visit).
    - …

Data for the Task

- EU-Park provides you with data for wait times in recent years. It was measured on 15min level and aggregated to hours, days, and weeks.
- The data contains information about weather and holidays. You can assume that a reliable weather forecast is available for 2 weeks in the future.
- Note that the data is synthetic. Hence the holidays may not match any real holidays. However, you can assume that holidays are known.
- The data contains information about the rides in three categories:
  - Rollercoasters: "Wadon", "Green Fire", "Gold Star", "ISS"
  - Water rides: "Fjord Fun", "Austrian white water", "Netpun"
  - Theme rides: "Pirates of Bavaria", "Dino Garden", "Ferry Boat"
- The data is included in the folder Task_02

Assessment of the Technical Solution

- Evaluation criteria
  - Overall quality 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 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 further forecasting models, considering different aggregation levels (e.g. hours or days) and different forecasting horizons (e.g. 1 hour ahead or 6 hours ahead[1]). You should have at least 2 analyses per person (different models, including evaluation with XAI)
- Explain the concepts you used for data analysis (for the target group)
- Evaluate and review your results (including the use of XAI techniques)
- **Note: If applicable, you can also show that a good prediction (using the models we have covered) is not possible.**

[1]For simplicity, you may forecast e.g. the 6th hour ahead, without forecasting 1-5 hours ahead.

Q1

- Conduct an exploratory data analysis for a chosen ride and temporal resolution. Conduct this analysis in the context of the forecasting task.

Q2

- Build a forecasting model for a chosen ride, forecasting horizon and temporal resolution. Evaluate the model and analyze it using XAI methods.

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***