Data Analysis with AI

Ing. David Chudán, Ph.D.

Agenda

- Introduction, Context, and Safety
- Working with Sample Data – Exploration and Preparation for Further Analysis
- Data Visualization
- Building a Simple Machine Learning Model (Linear Regression)
- „The AI Data Analyst"– Julius
- Questions and Discussion

# Generative AI as a data analysis assistant

- Based on the ability of large language models to generate the programming code.

- The more well-known the programming language is (it was more represented in the training data), the better the outputs are.

- Python is the leader in data analytics / machine learning and models use it within the code interpreter, models are good at this functionality.

Risks & Best Practices in AI Data Analysis

- **Do not upload any sensitive data** to public models.
- If you work with data from real people, **always anonymize** it before uploading (remove names, IDs, addresses).
- **Valid code ≠ Correct logic.** AI can solve the wrong problem. Always review the code/formula, not just the summary.
- **Large datasets may be truncated**; models might "forget" earlier instructions in long conversations.
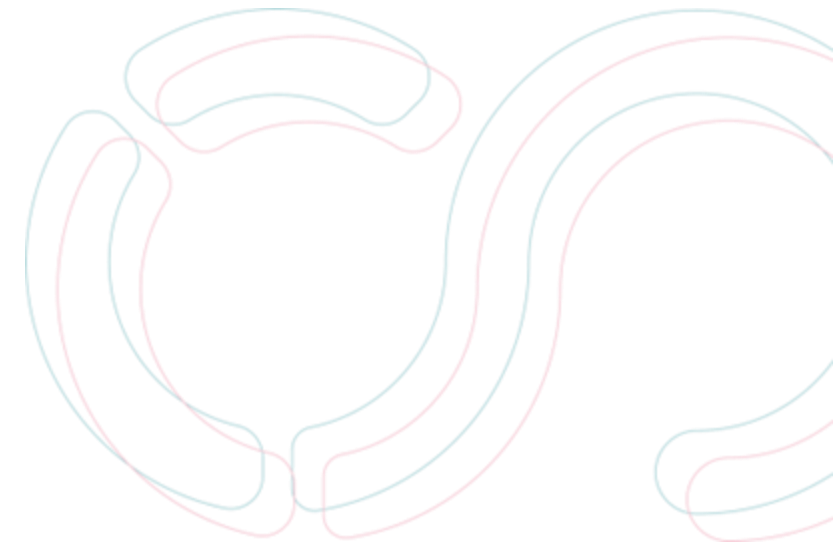- It is an assistant; **human oversight is mandatory**.

Tools – What we are going to use

- For working with sample data, you can use any (your favourite) chatbot – ChatGPT, Gemini, Claude…

- As a specialized tool, we will use the service julius.ai. For the purposes of this workshop, the free version is sufficient; the limit is 15 data-related conversations per month.
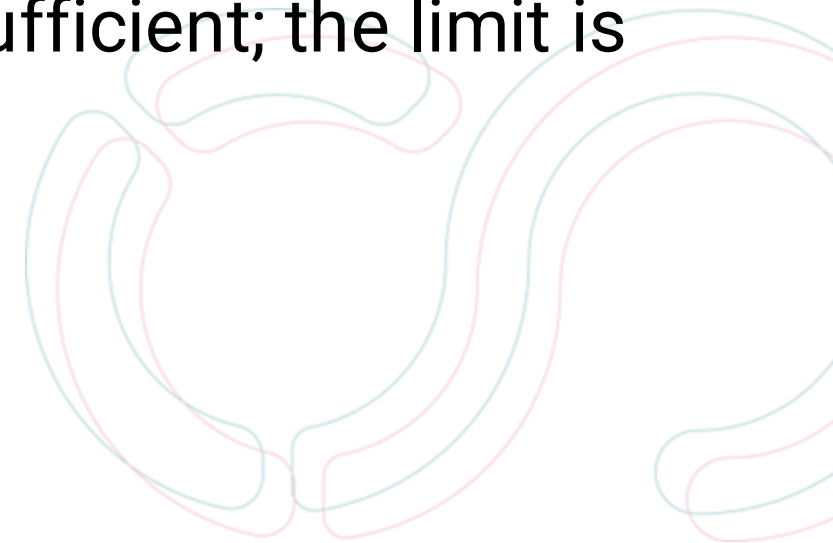
Tools – Excel support

- Copilot as an Office 365 add-in
  - Assistance with data import
  - Highlighting, sorting, and filtering
  - Generating and understanding (explaining) formulas
  - Visualization (summaries, charts, etc.)
- Emerging capabilities – Claude for Excel

Thank you