Analysis of Real-World Data from a Solar-Equipped Smart Home

A data-driven exploration of influence factors on solar energy generation and consumption patterns

Stephan Herbert

4.11.2025

GREENBOOTCAMPS
Your Career Booster in Digital Competence

Agenda

- About me
- Project context
- Planning & Technology Stack
- Data Sources & Energy Flow
- Research Hypotheses
- Hypotheses check
- Summary and recommendations

About Me

Link to LinkedIn

Stephan Herbert

- 37 years old
- Frankfurt am Main

Professional Background

- Master degree in Mechanical engineering (TU Darmstadt)
- 7+ years of experience in the automotive industry (Continental / FRA)
- Specialized in NVH / Acoustics
- Analysis of production line data, using KNIME and Power BI

Career Goal

- Transitioning into Data Analyst / BI Analyst roles to leverage analytical skills in data-driven decision making.
- Looking for jobs in Frankfurt / Rhein-Main area

Project Context

Topic

- Private smart home data provided by a friend → Real-world data
- Data set: PV / energy / heating + weather data (03/24 – 09/25)

PV = Photovoltaic

Stakeholder - Jürgen

- Retired engineer and tech enthusiast
- Continuously optimizing his smart home system

Motivation

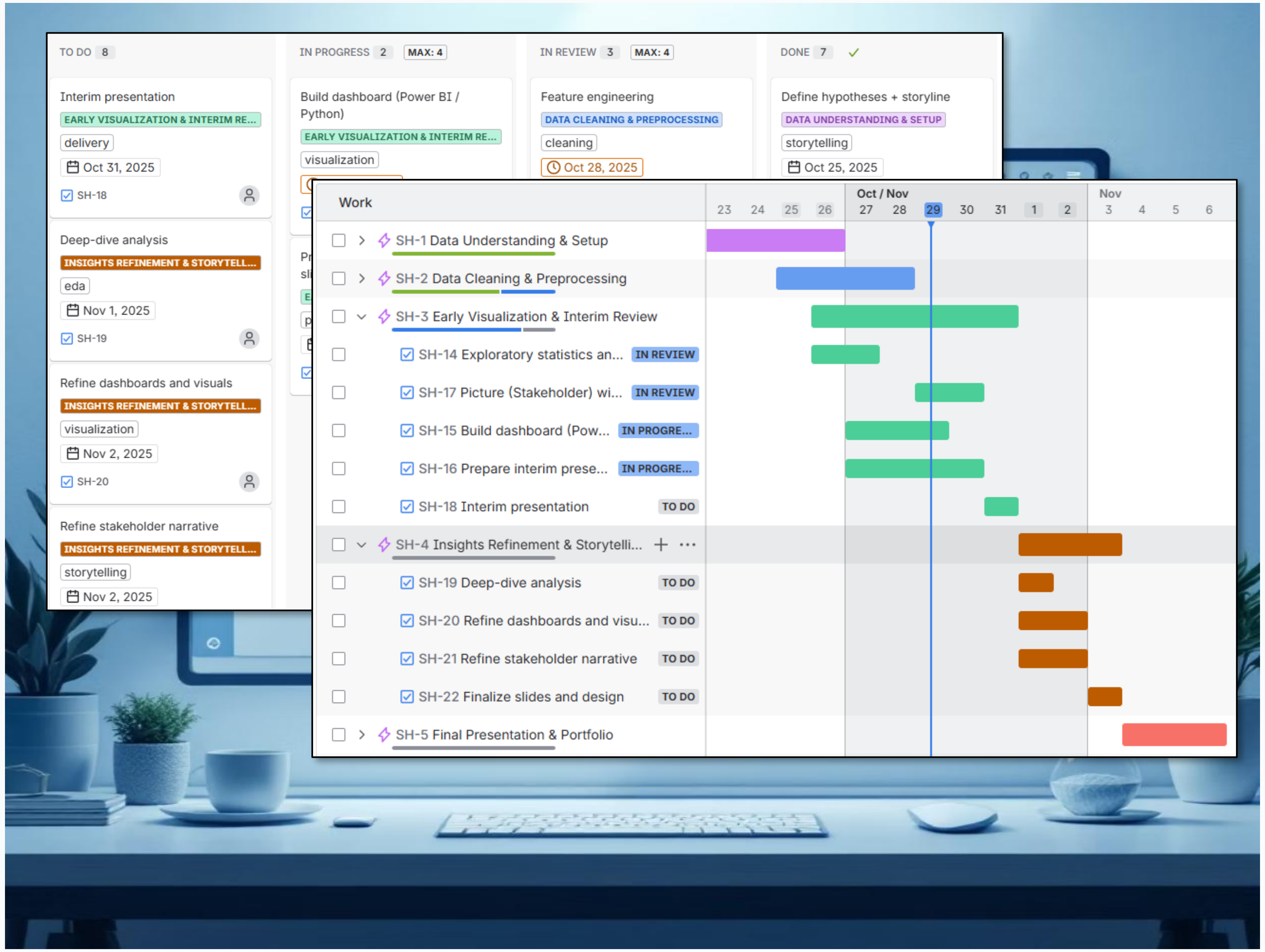
Help Jürgen understand:

- Key metrics and influence factors on his (smart) home
- Optimize his system for better financial returns

Planning & Technology Stack

Project Planning

JIRA Kanban board for task tracking and workflow organization.



Technology Stack

Python

Jira

Power BI

Streamlit

Gamma AI

PowerPoint

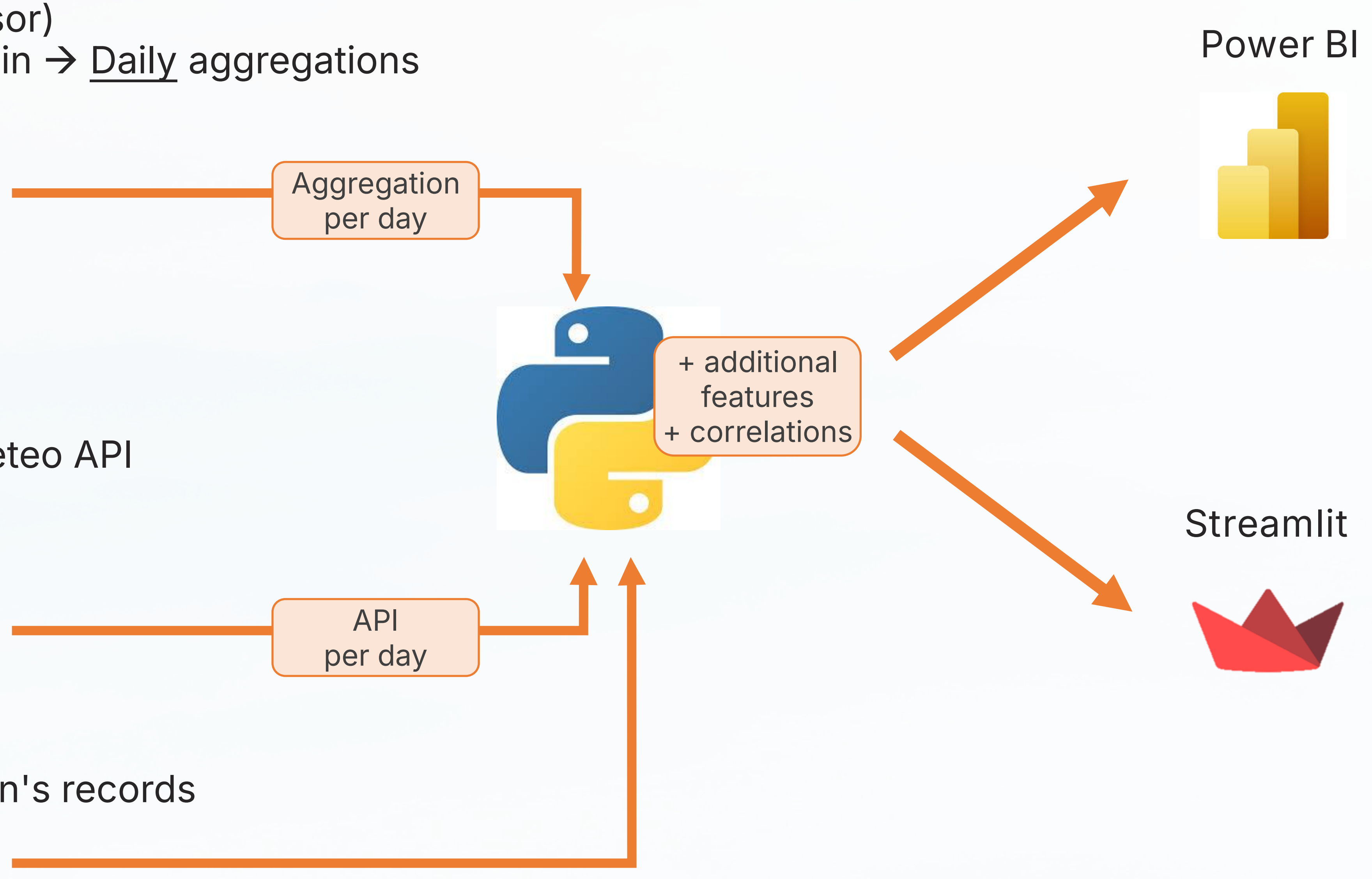
Data Sources & workflow

Data Sources

- Energy data from PV system and consumption
  - 9 CSV's (one per sensor)
  - 1.5 years – every 10 min → Daily aggregations

- Weather data via OpenMeteo API

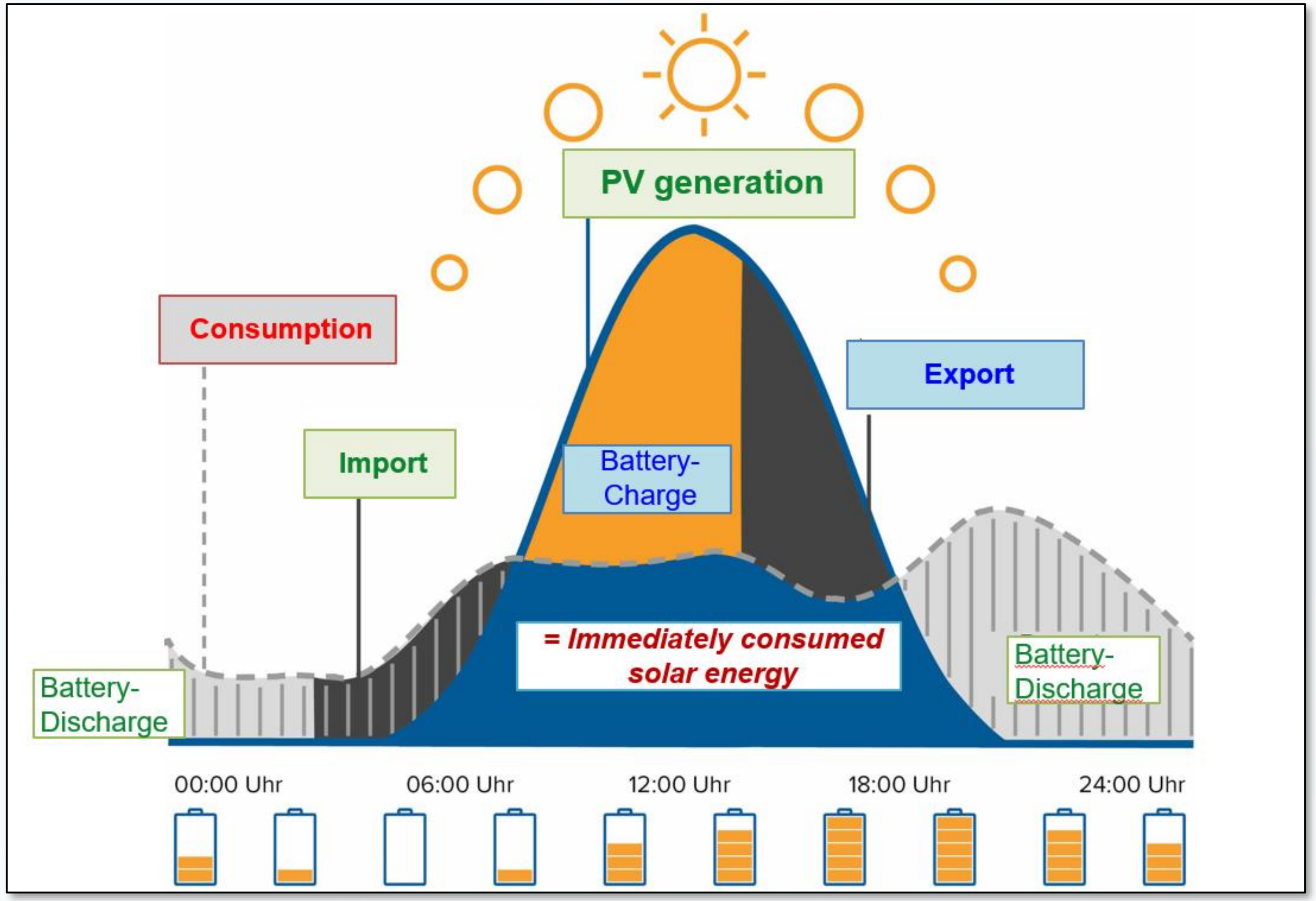
- Financial data from Jürgen's records



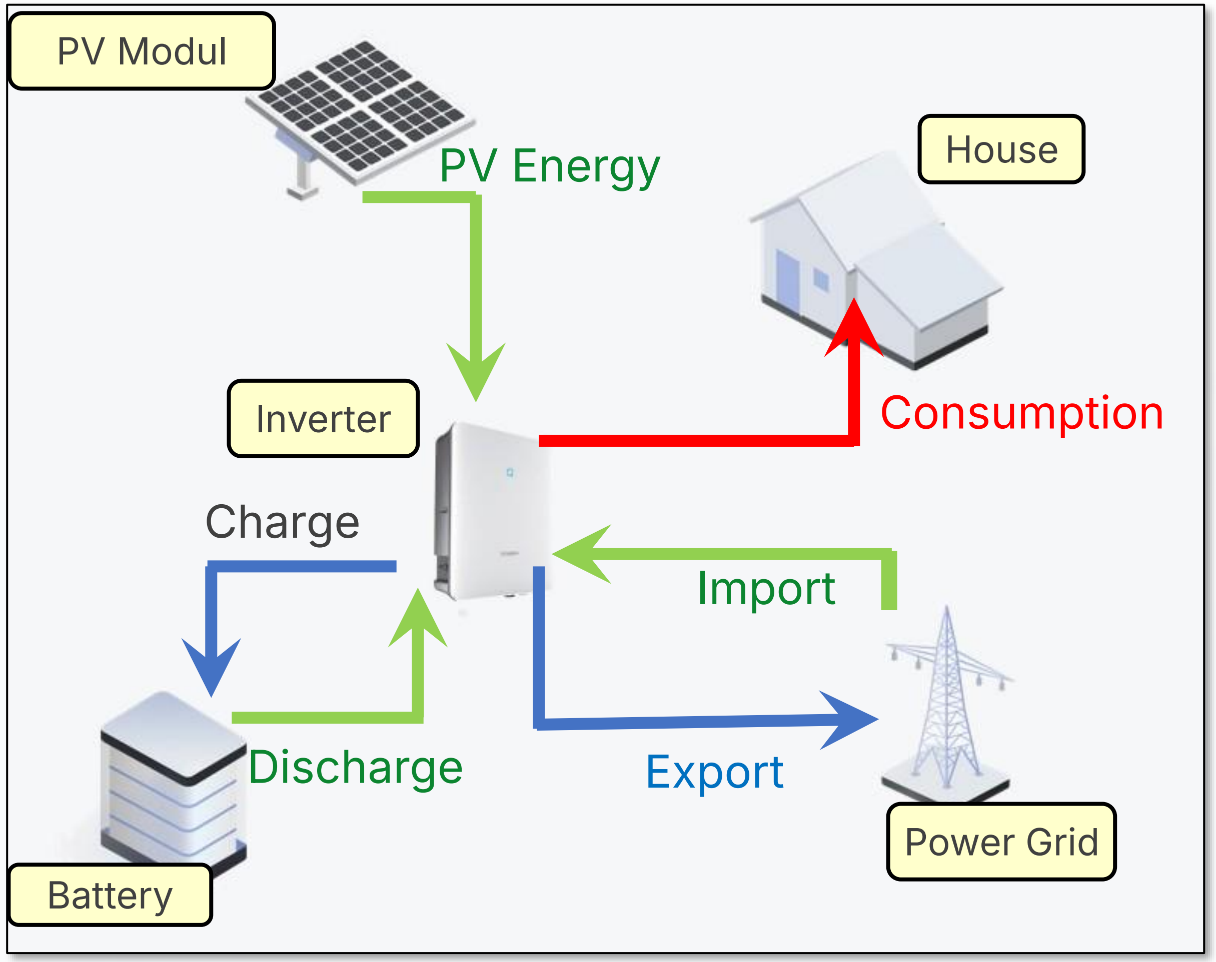
Energy Flow

PV-generation over the day

PV = Photovoltaic



Energy Flow System



Consumption = PV + Import – Export – (Charge – Discharge)

Research Hypotheses

1. Heating Usage

Heating is used even when outdoor temperature exceeds 15°C

2. Solar energy – Influence factors on PV-generation

Season/month has greater impact on PV generation than weather conditions

3. Energy Independence

Self-sufficiency is only achieved in Q2 and Q3

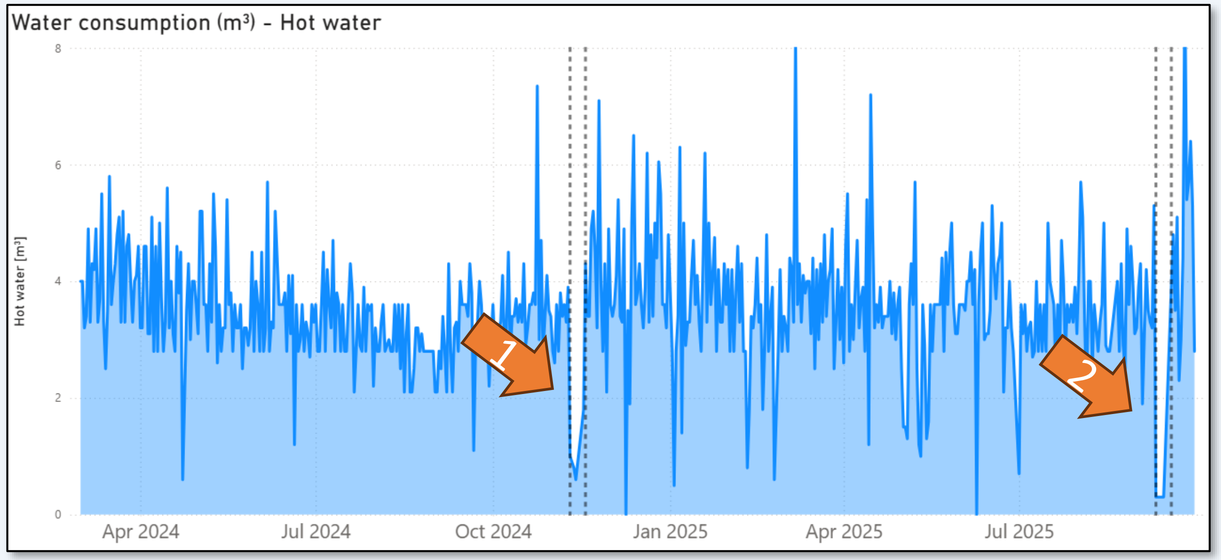
4. Optimization Potential ⇔ Payback Period

Even under best conditions, the system amortization time can only be reduced by 1 year compared to current conditions

Hypothesis 1: Hot water & Heating usage

1. Hot water

- Outliers of hot water usage:
  1. Vacation time?
  2. Sensor issue?
  3. “Only cold showers”?



1. 

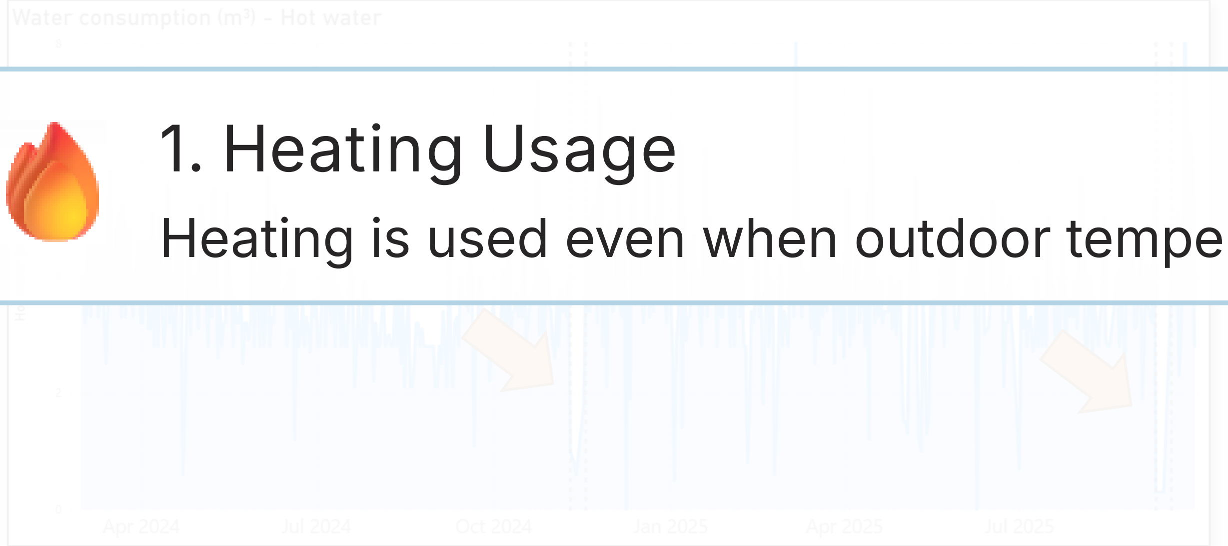
2. 

Data quality matters – "NaN“-values sometimes <u>zero consumption</u> ⇔ <u>missing data</u>

Hypothesis 1: Hot water & Heating usage

1. Hot water

- Outliers of hot water usage:
  1. Vacation time?
  2. Sensor issue?
  3. "Only cold showers"?



2. Heating

- Heating used between October – June
- Min-°C apparently more important than Mean-°C
- Small saving potential for time frames with > 10°C_min

1. Heating Usage

Heating is used even when outdoor temperature exceeds 15°C

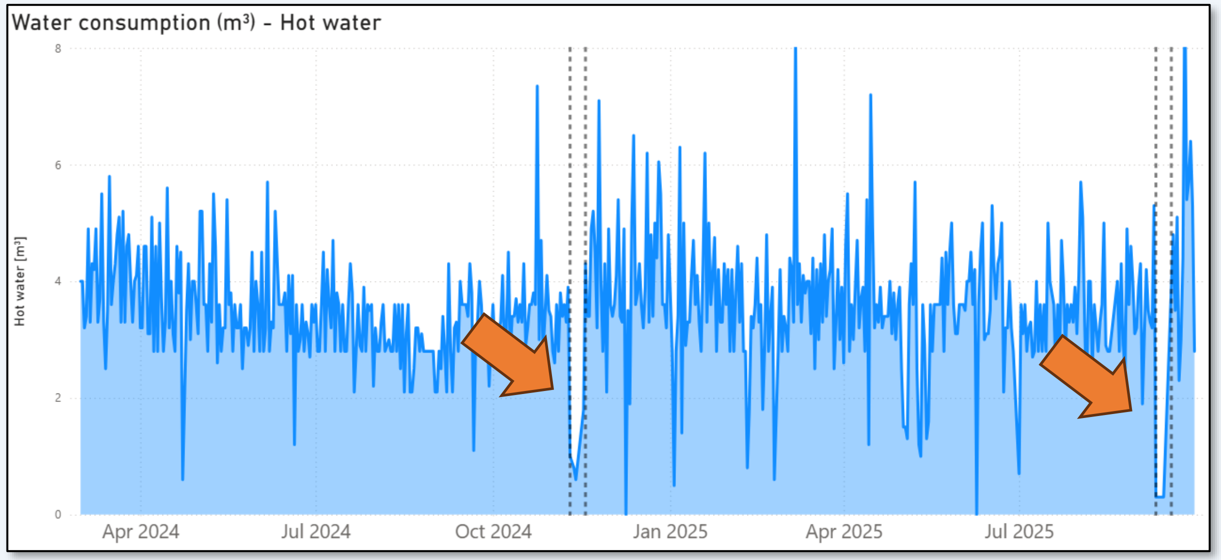
CONFIRMED

1. Data quality matters – "NaN"-values sometimes zero consumption ⇔ missing data
2. Heating occasionally used too long (when temperature > X °C), but only ~X % savings potential

Hypothesis 1: Hot water & Heating usage

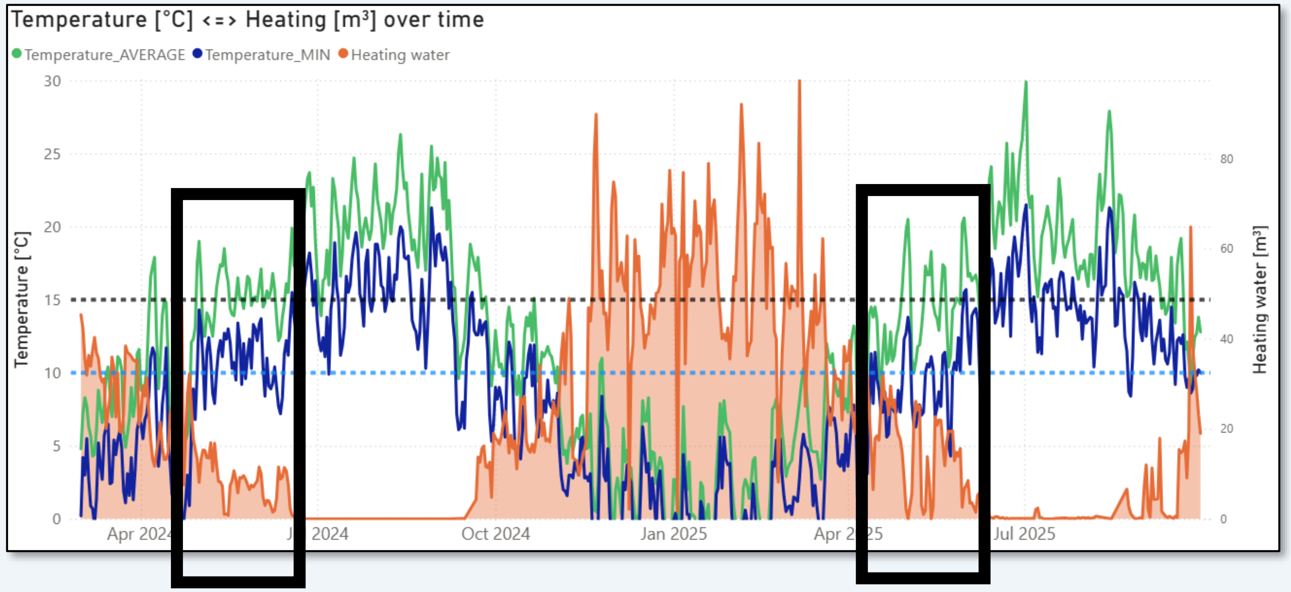
1. Hot water

- Outliers of hot water usage:
  1. Vacation time?
  2. Sensor issue?
  3. “Only cold showers”?



2. Heating

- Heating used between October – June
- Min-°C apparently more important than Mean-°C:
  - Heating & Min-°C > 10°C: 111 days
  - Heating & Mean-°C > 15 °C: 91 days



1. Data quality matters – "NaN“-values sometimes zero consumption ⇔ missing data
2. Heating used too long (on 78 “warmer” days), but only ~5 % savings potential (107 m³ out of 2181 m³)

Hypothesis 2: Solar energy – Influence factors on PV-generation

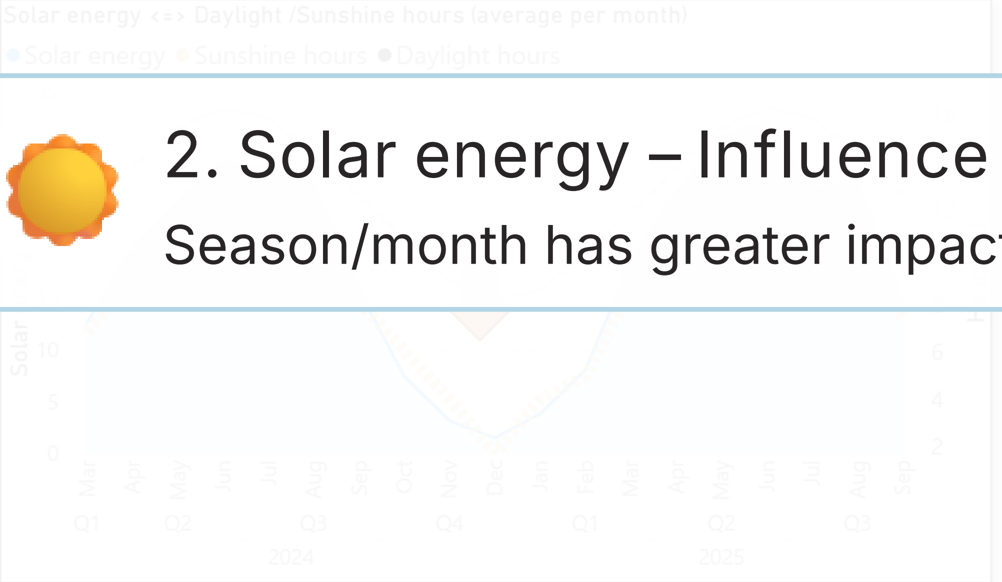
PV = Photovoltaic

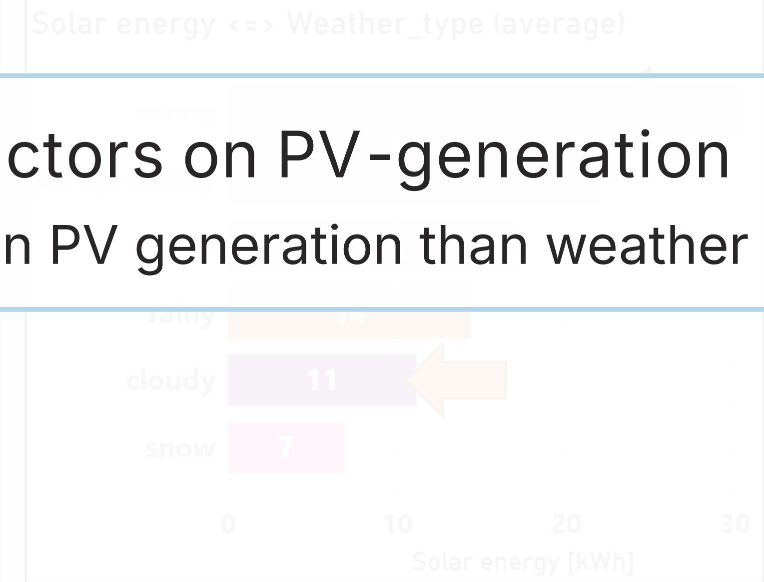
1. Season / sunshine hours

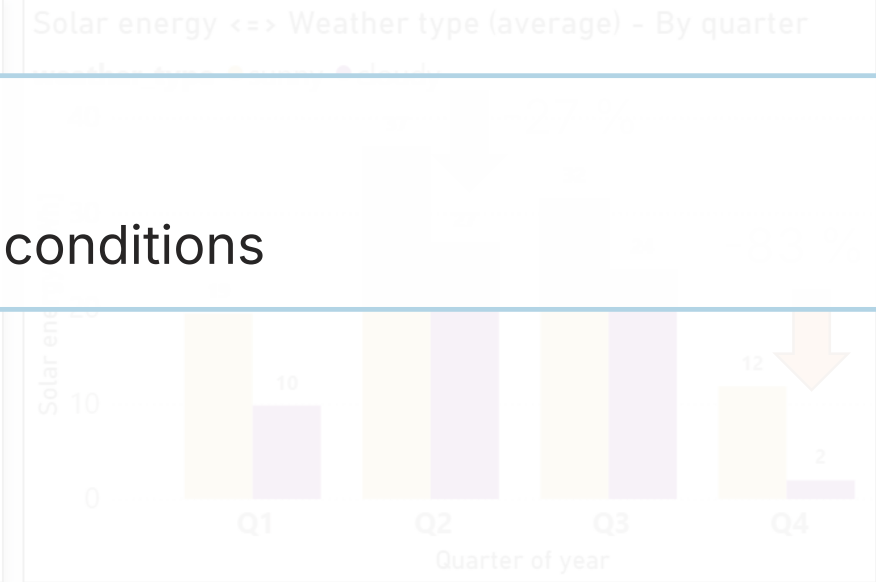
Solar energy highly correlates:
- to sunshine hours (0.87)
- to daylight hours (0.77)

2. Weather conditions

- Huge correlation to weather type (-0.47)
  cloud coverage per day (-0.62)
- Cloudy ⇔ sunny weather:
  - Q2 / Q3: up to -27 %
  - Q1 / Q4: up to -83 %







2. Solar energy – Influence factors on PV-generation

Season/month has greater impact on PV generation than weather conditions

CONFIRMED

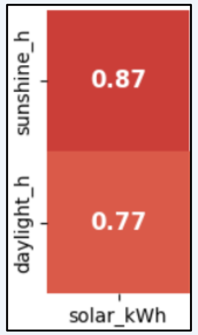
1. Higher correlation of solar energy to seasonal effects (daylight...) than to weather type / cloud coverage
2. But: Weather conditions have severe influence too!

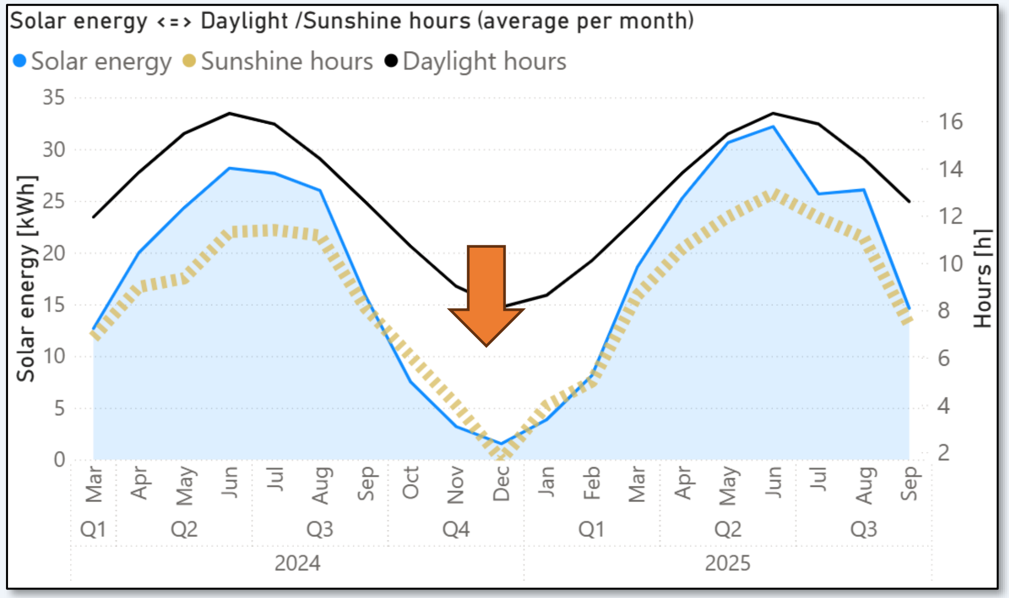
Hypothesis 2: Solar energy – Influence factors on PV-generation

PV = Photovoltaic

1. Season / sunshine hours

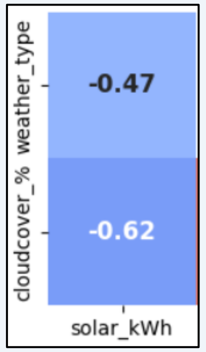
Solar energy highly correlates:
- to sunshine hours (0.87)
- to daylight hours (0.77)

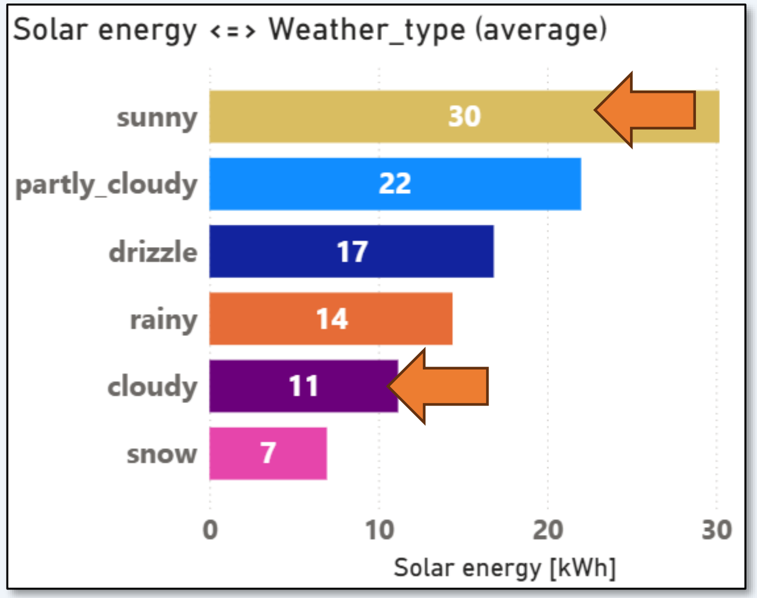


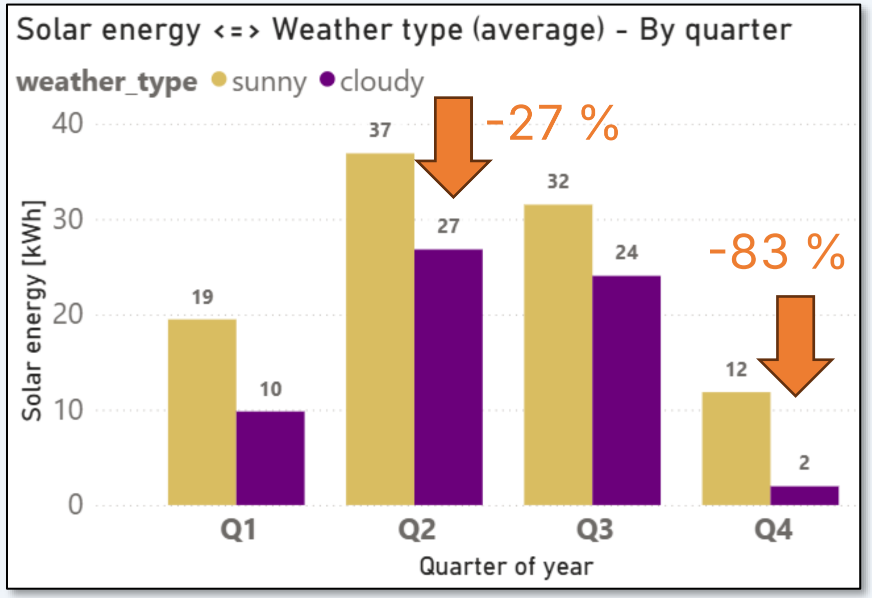


2. Weather conditions

- Huge correlation to weather type (-0.47)
  cloud coverage per day (-0.62)
- Cloudy ⇔ sunny weather:
  - Q2 / Q3: up to -27 %
  - Q1 / Q4: up to -83 %







CONFIRMED

1. Higher correlation of solar energy to seasonal effects (daylight...) than to weather type / cloud coverage
2. But: Weather conditions have severe influence too!

Hypothesis 3: Energy Independence

1. Energy flow + Self-sufficiency %

- Consumed energy is quite stable (~ 9.4 kWh/day)
- Import > 50% from October – February
- Self-sufficiency >90% from May – August

2. Energy import

- Days without imports even in November + February
- 63 % of days with < 1 kWh import

3. Energy Independence

Self-sufficiency is only achieved in Q2 and Q3

FALSIFIED

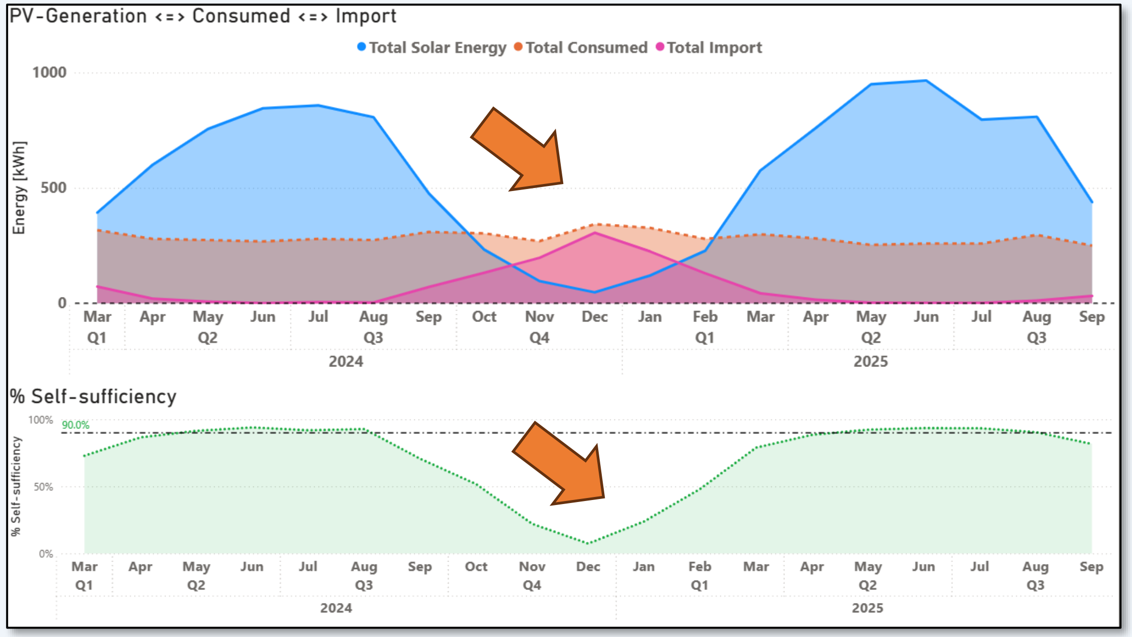
1. April-August: >80% of days have 0-1 kWh grid import *(note: "NaN" could be "no import" or sensor issue)*
2. Except Dec & Jan: Every month shows at least 4 days with 0-1 kWh import, even in Q1 and Q4

PV = Photovoltaic

Hypothesis 3: Energy Independence

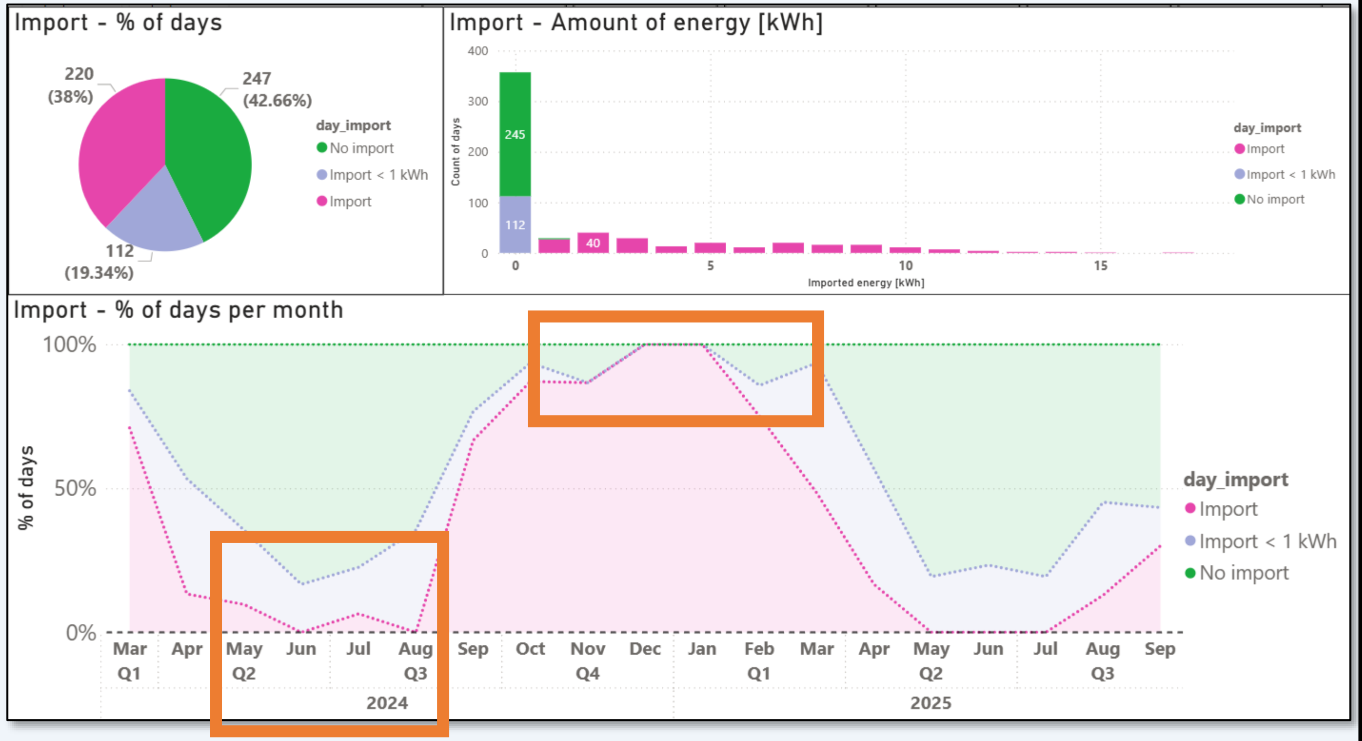
1. Energy flow + Self-sufficiency %

- Consumed energy is quite stable (~ 9.4 kWh/day)
- Import > 50% from October – February
- Self-sufficiency >90% from May – August



2. Energy import

- Days without imports even in November + February
- 63 % of days with < 1 kWh import



- Except Dec & Jan: Every month shows at least 4 days with 0-1 kWh import, even in Q1 and Q4

(note: "NaN" could be "no import" or sensor issue)

Hypothesis 4: Optimization Potential ⇔ Payback period

1. Hardware

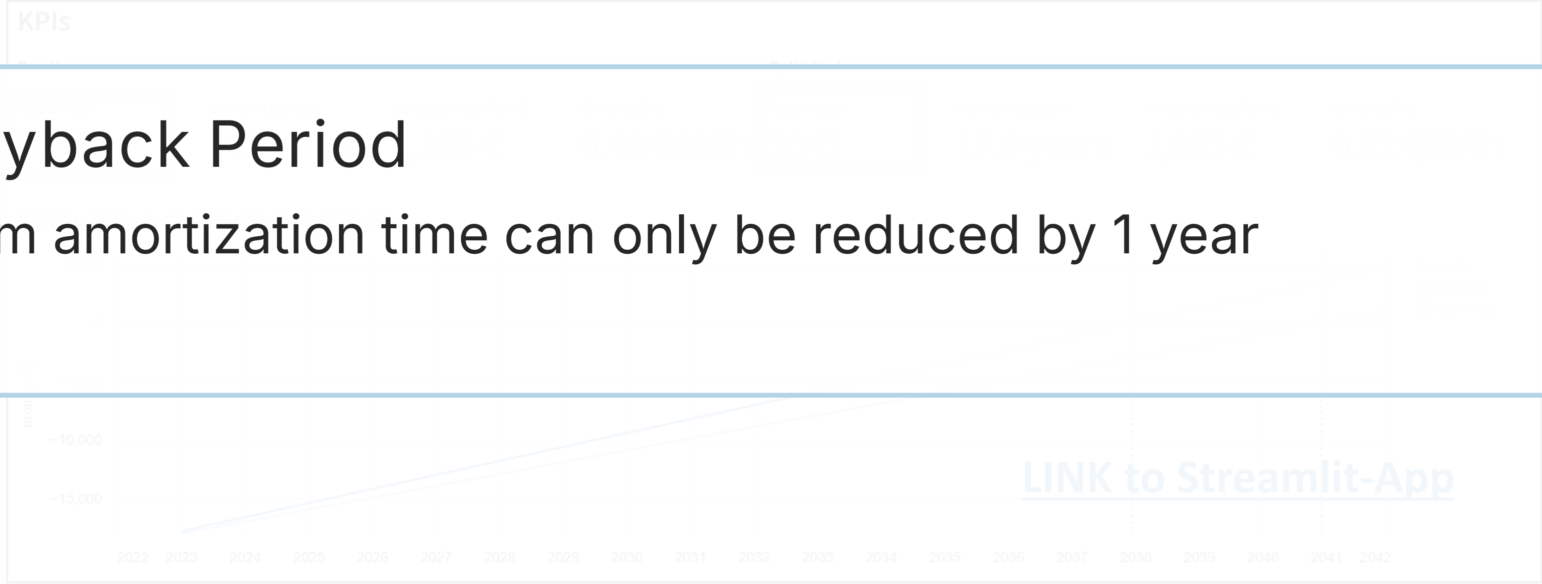
a.) More PV-generation:
- Solar panels ✗
- Shadow from tree ✗

c.) Electric Vehicle :
- Potential "big battery" in the future

2. Yearly energy costs (Import) ⇔ Feed-in tariff (Export)

a.) Feed-in tariff : Fixed for 20 years (0.082 € per kWh) ✗

b.) Import costs:
- Optimization potential: ~0.32 €/ kWh instead of 0.40 €/ kWh
- BUT: fewer yearly costs → longer amortization time



LINK to Streamlit-App

4. Optimization Potential ⇔ Payback Period

Even under best conditions, the system amortization time can only be reduced by 1 year compared to current conditions

FALSIFIED

1. No hardware optimizations (except electric vehicle)
2. Cheaper contract results in fewer yearly costs (- 90 €), but higher amortization time (> 3 years)

Hypothesis 4: Optimization Potential ⇔ Payback period

1. Hardware

a.) More PV-generation:
- Solar panels ✗
- Shadow from tree ✗

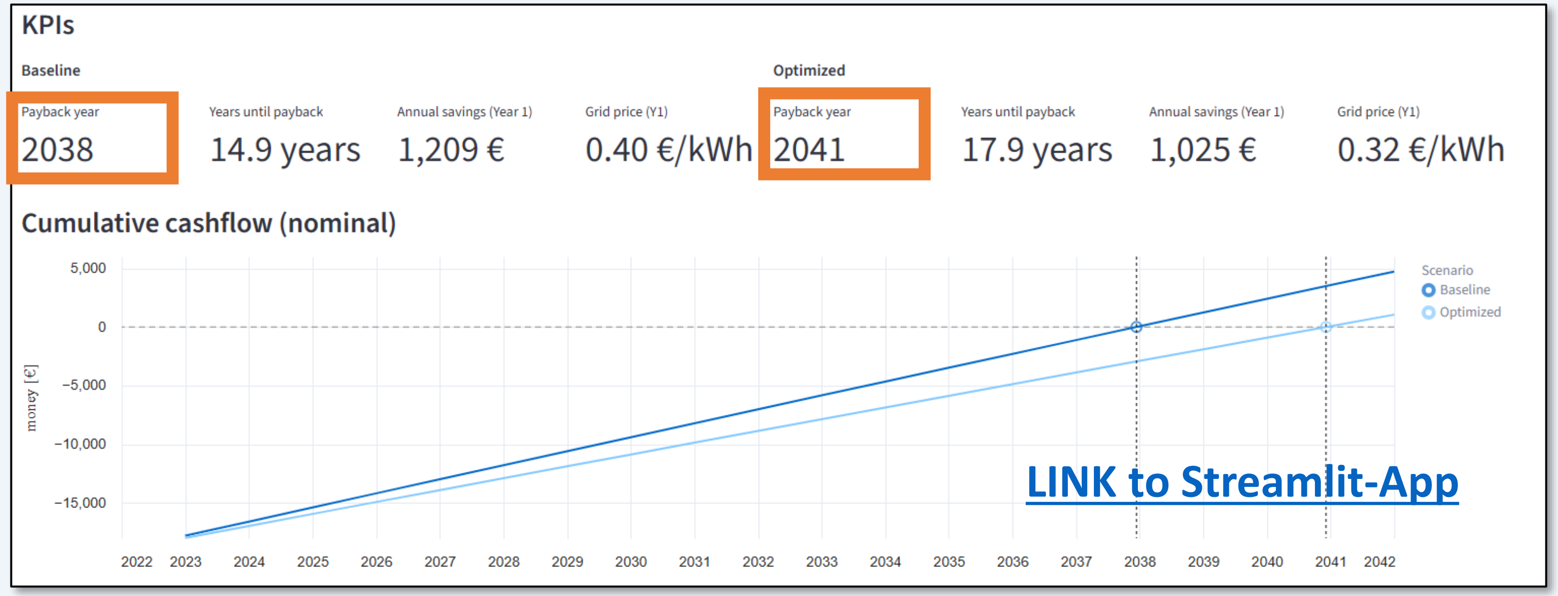
b.) Battery: ✗
- Current battery capacity = 9.6 kWh
- Consumption per day (avg) = 9.4 kWh
  ➔ Charged battery lasts for ~ 1 day
- ~ 5000 – 6000 € invest!

c.) Electric Vehicle :
- Potential "big battery" in the future

2. Yearly energy costs (Import) ⇔ Feed-in tariff (Export)

a.) Feed-in tariff : Fixed for 20 years (0.082 € per kWh) ✗

b.) Import costs:
- Optimization potential: ~0.32 €/ kWh instead of 0.40 €/ kWh
- BUT: fewer yearly costs → longer amortization time



1. No hardware optimizations (except electric vehicle)
2. Cheaper contract results in fewer yearly costs (- 90 €), but higher amortization time (> 3 years)

Summary & Recommendations

Main findings

1. Heating optimization offers limited savings
2. Seasonal > weather influences on solar energy
3. Self-sufficiency achievable in all seasons
4. Amortization time highly driven by import costs

Recommendations to Jürgen

1. Monitor heating patterns during spring / fall
2. Implement data validation checks [NaN values]
3. Evaluate electricity contracts regularly to reduce yearly costs for import

Thank you
for your attention!

Interested in discussing data analytics, or career opportunities in BI?

Connect with me on [LinkedIn]

LinkedIn

GitHub - Repository

Streamlit -App