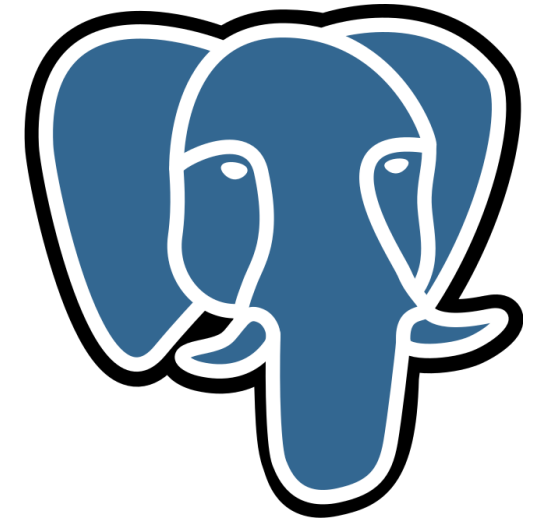# BI-MINI-PROJEKT
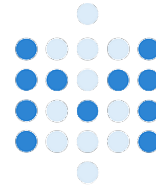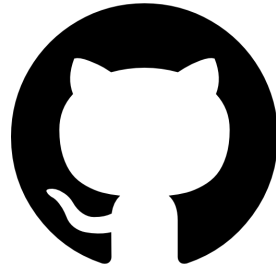# THESIS TOPIC EVALUATION
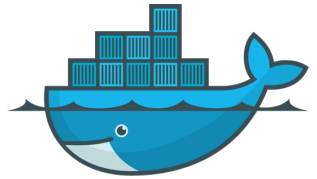
Keno Coldewey
30.01.2024

AGENDA

- Overview

- Results

- Lessons learned

TECHNOLOGY STACK

ETL PROCESS

- Dataset
  - Export from Studi.Ip: thesis topics
  - 2021 February → 2023 December
  - Nested dicts for every day, sometimes

- ETL Import to PostgreSQL
  → Python

```
∨ 20210721_e1ca
  > additional
  <> 20210721-191704-9bc8.html
  <> 20210721-191706-b2e0.html
  <> 20210721-191708-9aa8.html
  <> 20210721-191710-0666.html
  ▦ db-topics-additional.csv
  {} db-topics-additional.json
  ▦ db-topics.csv
  {} db-topics.json
  > 20210722_31aa
  > 20210723_de4b
  > 20210724_c9be
```
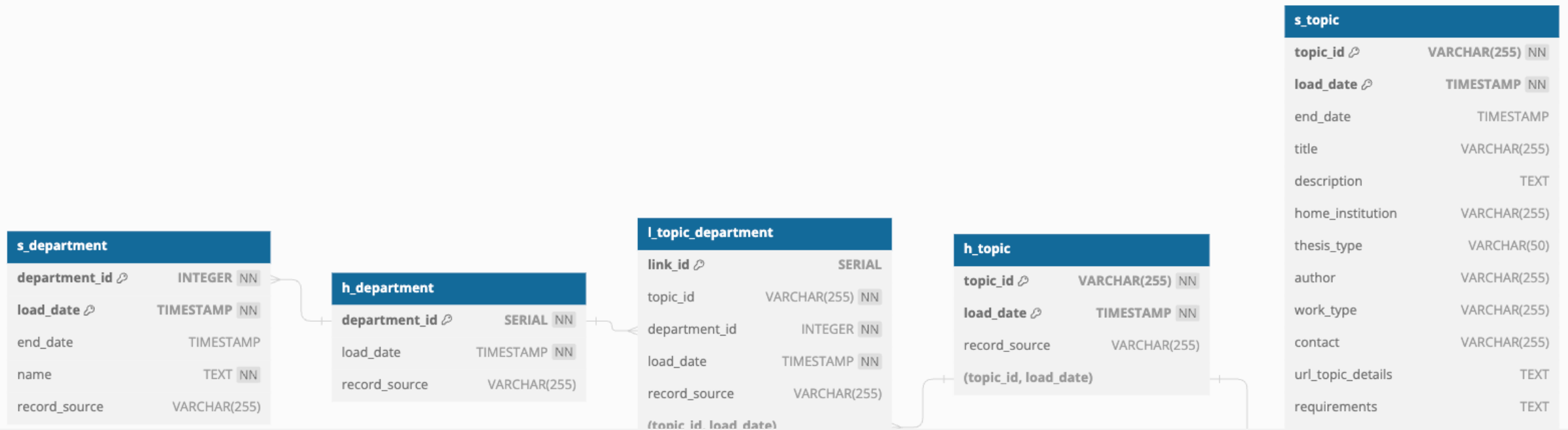
DATABASE SCHEMA as DATA VAULT

- 2 hubs

- 2 satellite

- 1 link

Title	Level		Type	Status	Date	
Bodenmegafauna-Grabgänge: Highways der biologischen Aktivität?	Bachelor	(i)	praktisch / anwendungsbezogen	vergeben	20.02.2023	
Masterarbeit in Anlehnung an das Projektband im Sachunterricht	Master	(i)	nicht spezifiziert	verfügbar	15.02.2023	
Hey ChatGPT! Please verify my program.	Bachelor	(i)	praktisch / anwendungsbezogen	reserviert	13.02.2023	
Peer-Interaktionen in der Kita - Fragestrategien	Master	(i)	praktisch / anwendungsbezogen	verfügbar	10.02.2023	
Entwicklung einer Testautomatisierung für das Testen in operativen Umgebungen	Bachelor oder Master	(i)	praktisch / anwendungsbezogen	verfügbar	09.02.2023	
Fehlerinjektion in NMEA0183 Netzwerke mithilfe einer ROS2 basierten Architektur	Bachelor oder Master	(i)	praktisch / anwendungsbezogen	vergeben	09.02.2023	
From incorrectness logic proofs to violation witnesses	Bachelor	(i)	konzeptuell / theoretisch	verfügbar	06.02.2023	

< | 1 | **2** | 3 | 4 | … | 7 | >

Title	Level		Type	Status	Date	
Hey ChatGPT! Please verify my program.	Bachelor	(i)	praktisch / anwendungsbezogen	reserviert	13.02.2023	
Peer-Interaktionen in der Kita - Fragestrategien	Master	(i)	praktisch / anwendungsbezogen	verfügbar	10.02.2023	
Entwicklung einer Testautomatisierung für das Testen in operativen Umgebungen	Bachelor oder Master	(i)	praktisch / anwendungsbezogen	verfügbar	09.02.2023	
Fehlerinjektion in NMEA0183 Netzwerke mithilfe einer ROS2 basierten Architektur	Bachelor oder Master	(i)	praktisch / anwendungsbezogen	vergeben	09.02.2023	
From incorrectness logic proofs to violation witnesses	Bachelor	(i)	konzeptuell / theoretisch	verfügbar	06.02.2023	
Urbane Gewässer und Renaturierung - Unauflösbarer Gegensatz?	Bachelor	(i)	praktisch / anwendungsbezogen	verfügbar	06.02.2023	
Producing an automated tool to extract colour from research imagery	Master	(i)	praktisch / anwendungsbezogen	verfügbar	03.02.2023	
Paradigm shifts in dissolved organic matter understanding: The persistence of the overturned	Master	(i)	konzeptuell / theoretisch	verfügbar	03.02.2023	

< | 1 | … | 3 | 4 | **5** | 6 | 7 | >

Year 2021
→ Crawler issue, not all pages are within the csv/json

24.05.2023
- Only 54 titles

- Same data items on different pages
  → based on crawler issuer or stud.Ip problem

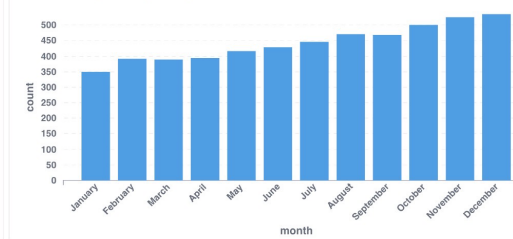- Overwrite the date 24.05.2023 with the data from date 23.05.2023

6

RESULTS

TASKS
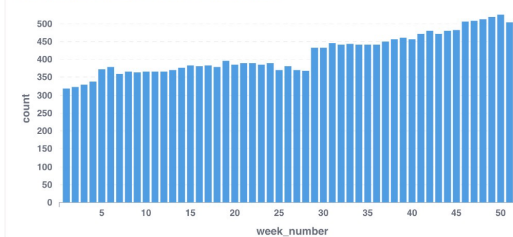
- T1 – How many thesis topics are published in a week, in a month, in a year?
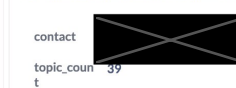
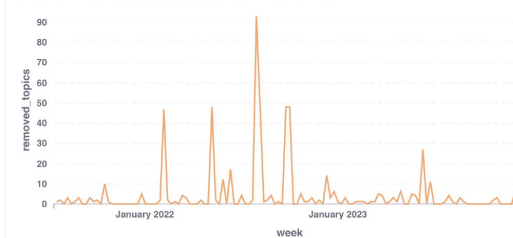- T2 – Which supervisor has the most thesis topics to offer?

- T3 – Which department has the most thesis topics to offer?

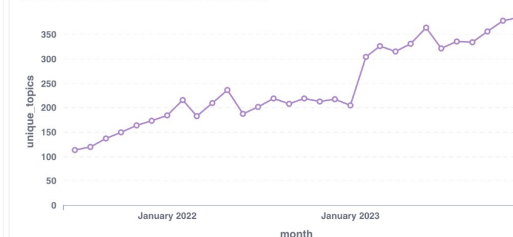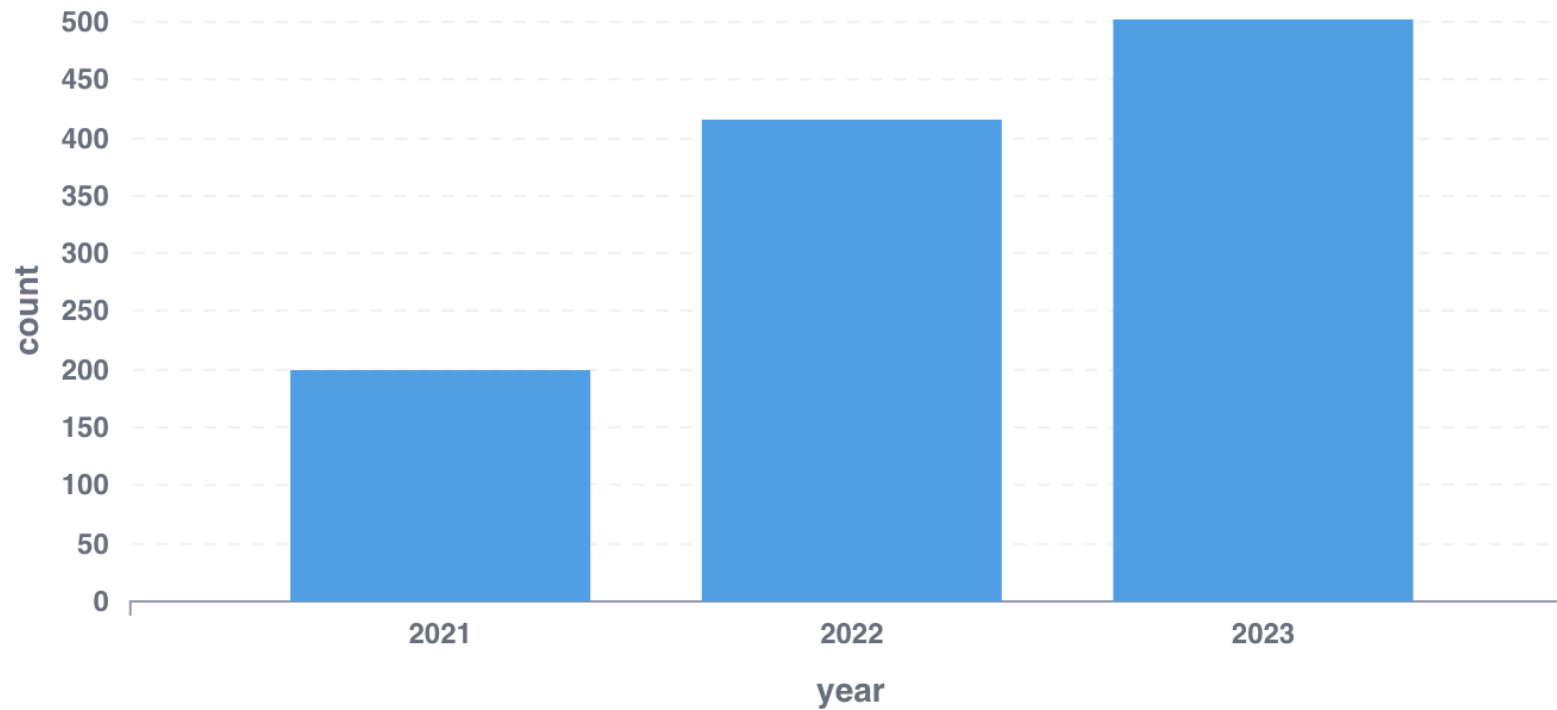- T4 – How many thesis topics are "removed from the list" in a week, in a month, in a year?

- T5 – Create 1 task/business question of your own and answer this question:
  Which Thesis Type is Most Common?

- KPI 1 – Unique thesis topics published every month

- KPI 2 – Average thesis topics for each department

- KPI 3 – Create 1 KPI of your own, describe how to calculate it and put it on the dashboard:
  Overview of the distribution of thesis topics across different thesis types.
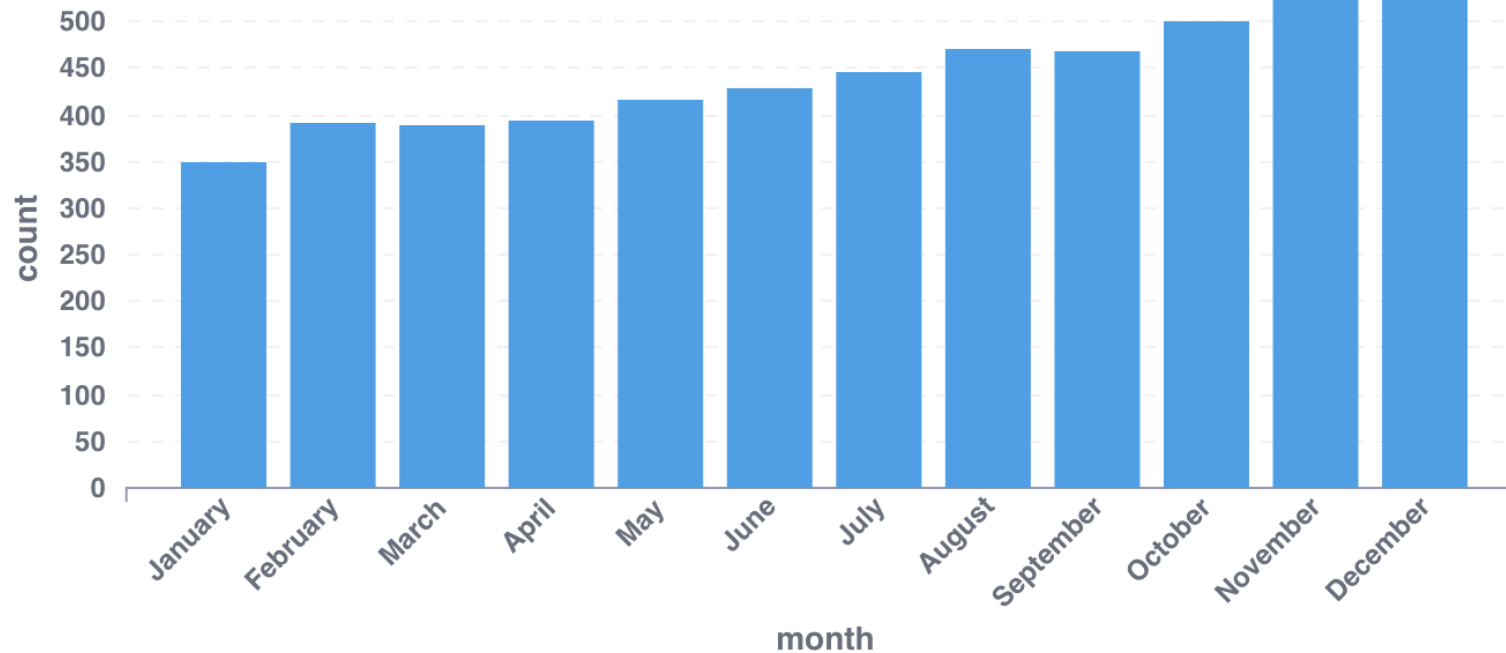
T1 - HOW MANY THESIS TOPICS ARE PUBLISHED IN A WEEK, IN A MONTH, IN A YEAR?

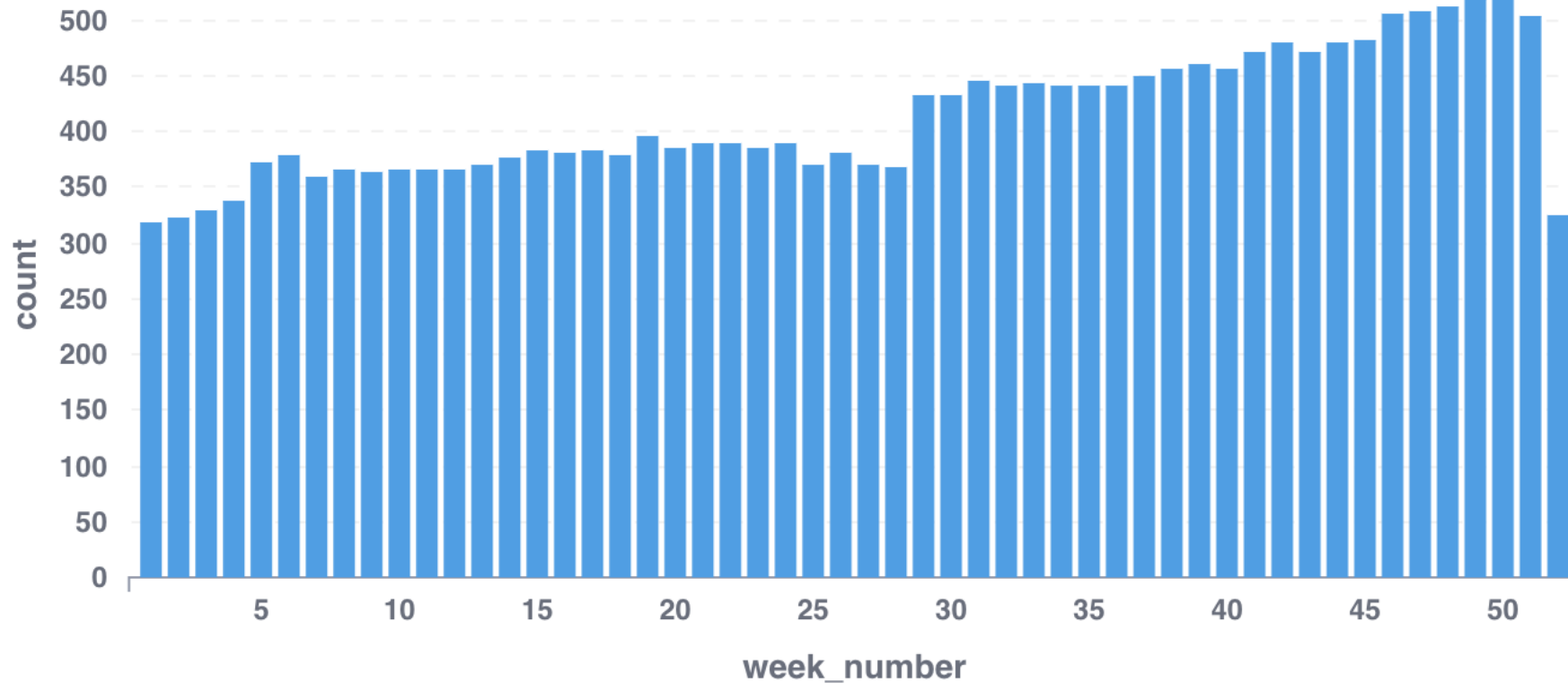**T1.1 - How many thesis topics are published in a year?**

T1 - HOW MANY THESIS TOPICS ARE PUBLISHED IN A WEEK, IN A MONTH, IN A YEAR?
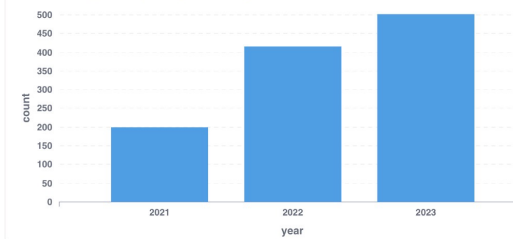
T1.2 - How many thesis topics are published in a month?   ...

T1 - HOW MANY THESIS TOPICS ARE PUBLISHED IN A WEEK, IN A MONTH, IN A YEAR?

T1.3 - How many thesis topics are published in a week?

T2, T3 AND T5

T2 - Which supervisor has the most thesis top...

contact   ▨

topic_coun t   39

T3 - Which department has the most thesis t...

departmen t_name   **Fach-Bachelor Wirtschaftsinformatik**
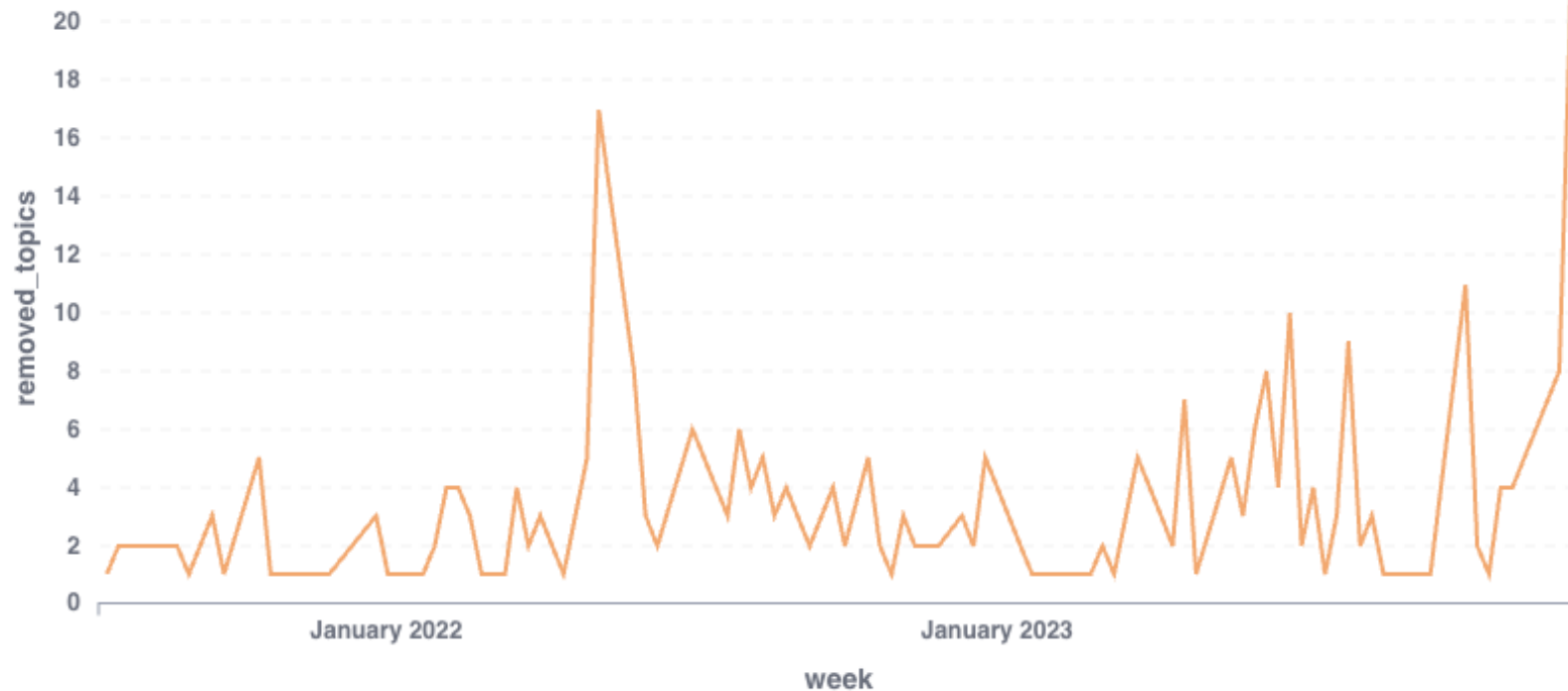
topic_coun t   113

T5 - Which Thesis Type is Most Com...

thesis_t ype   **Bachelor oder Master**

Count   323

T4 - HOW MANY THESIS TOPICS ARE "REMOVED FROM THE LIST" IN A WEEK, IN A MONTH, IN A YEAR?

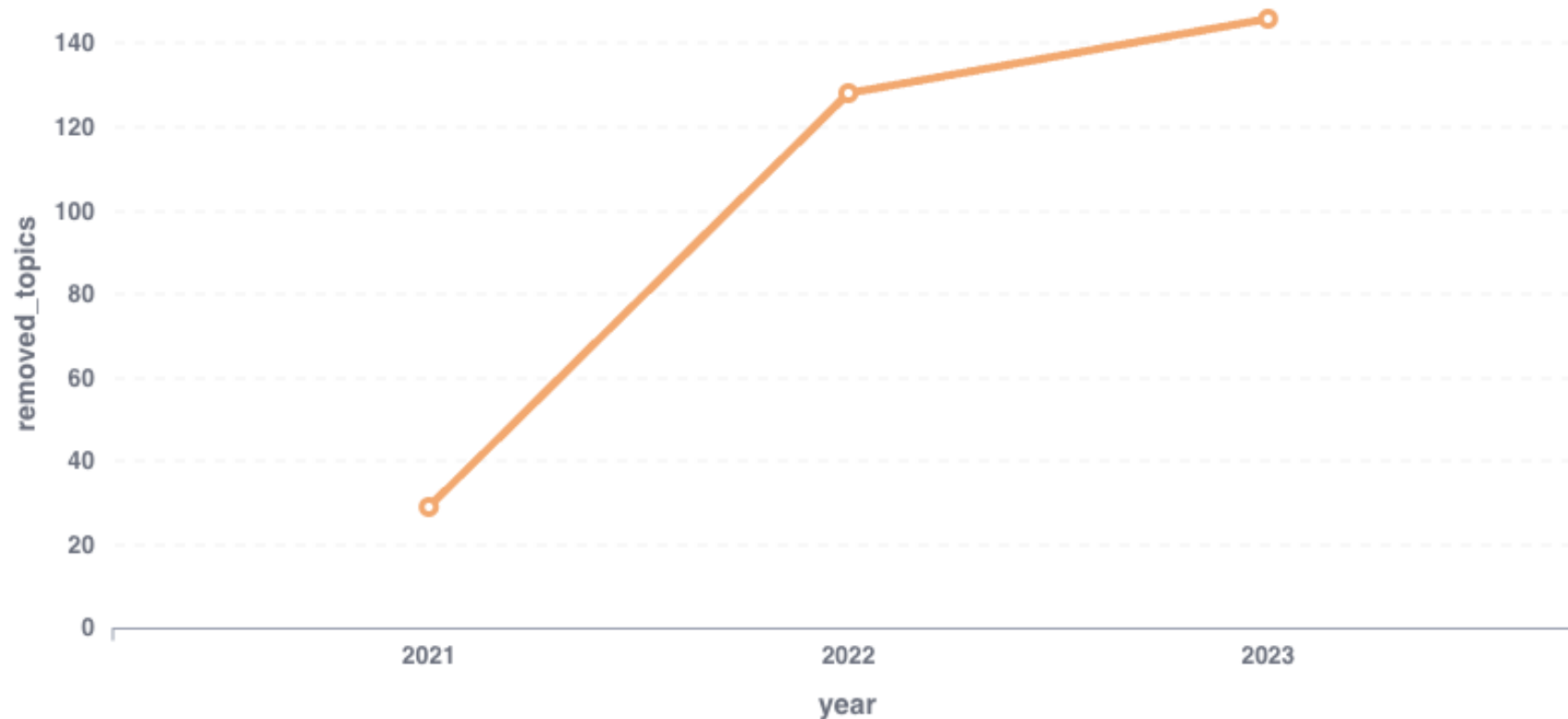T4.1 - How many thesis topics are "removed from the list" in a week?

T4 - HOW MANY THESIS TOPICS ARE "REMOVED FROM THE LIST" IN A WEEK, IN A MONTH, IN A YEAR?

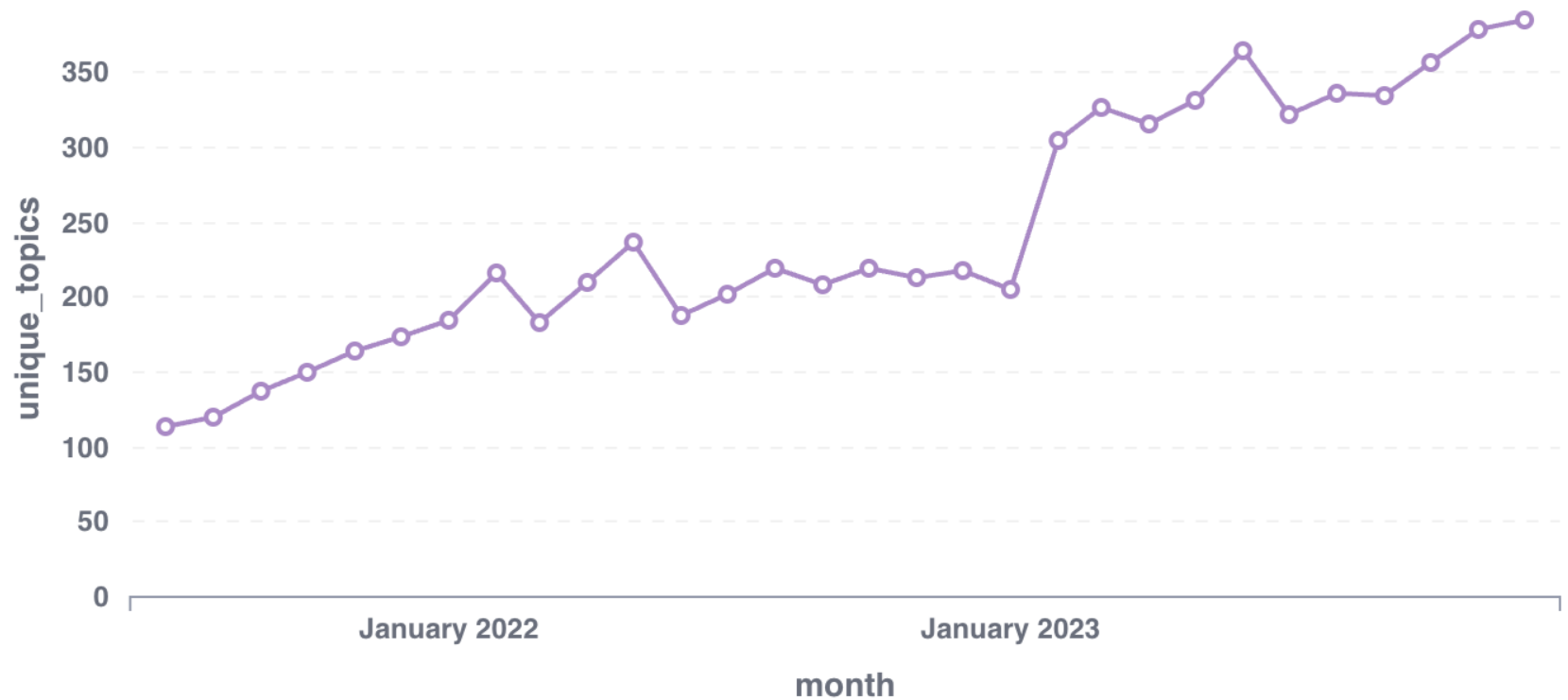T4.2 - How many thesis topics are "removed from the list" in a month?

T4 - HOW MANY THESIS TOPICS ARE "REMOVED FROM THE LIST" IN A WEEK, IN A MONTH, IN A YEAR?

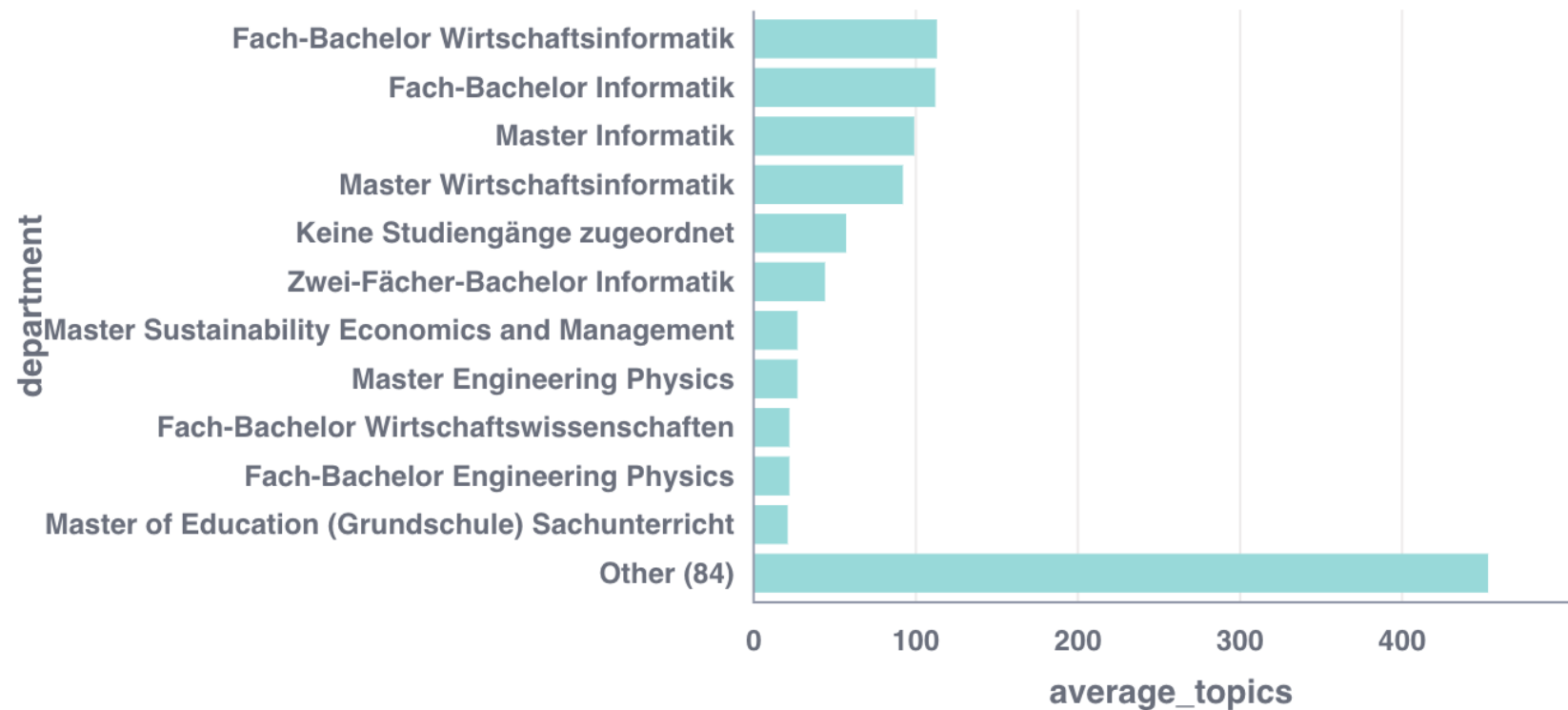T4.3 - How many thesis topics are "removed from the list" in a year?

KPI 1 - UNIQUE THESIS TOPICS PUBLISHED EVERY MONTH

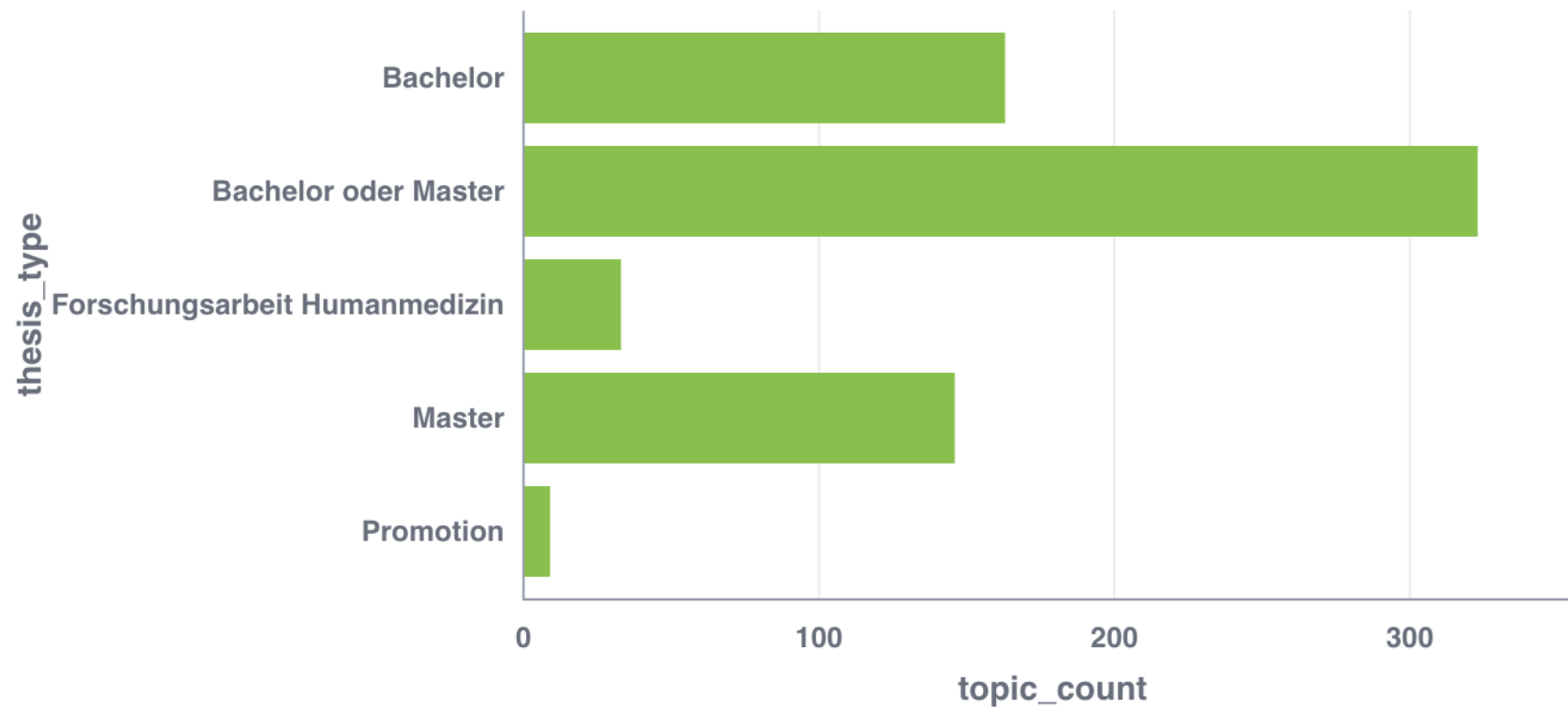
KPI 1 - Unique thesis topics published every month

KPI 2 - AVERAGE THESIS TOPICS FOR EACH DEPARTMENT

KPI 2 - Average thesis topics for each department

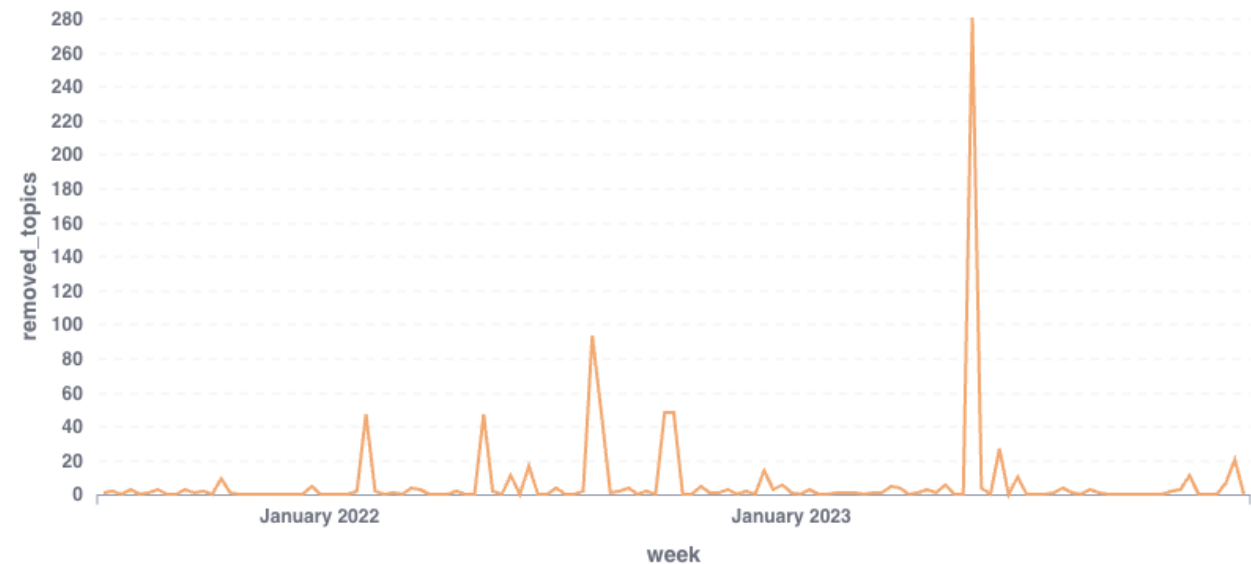# KPI 3 - DISTRIBUTION OF THESIS TOPICS ACROSS DIFFERENT THESIS TYPES

KPI 3 - distribution of thesis topics across different thesis types

LESSONS LEARNED: DATA CLEANING

- Extremlly highs in the visualisation

- Comparison with LevinSthein for import-process of the ETL process was helpful

T4.1 - How many thesis topics are "removed from the list" in a week?

LESSONS LEARNED / FEEDBACK

- Data Vault have good expansion options

- Aggregated json/csv files in the dataset was very helpful

- Dataset has a issue with year 2021

- ETL process was time-consuming

- Adjustment for the database schema in the process of evaluation needed (departments table)

- Use host.docker.internal instead of localhost for the host

- Usage of natural keys is not optimal

THANKS FOR THE ATTENTION!

Keno Coldewey