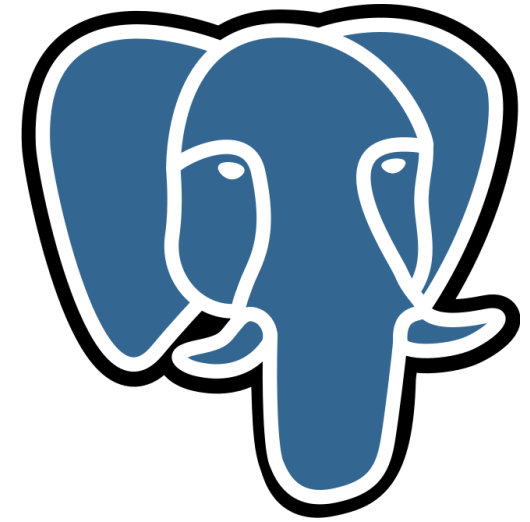
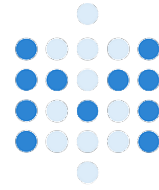


# BI-MINI-PROJEKT THESIS TOPIC EVALUATION

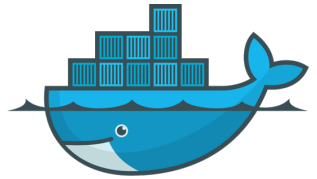
Keno Coldewey  
30.01.2024

# AGENDA

- Overview
- Results
- Lessons learned



# Metabase



docker

TECHNOLOGY STACK

# ETL PROCESS

- Dataset
  - Export from studi.ip: thesis topics
  - 2021 February → 2023 December
  - Nested dicts for every day
- 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
```

s_department	
department_id	INTEGER NN
load_date	TIMESTAMP NN
end_date	TIMESTAMP
name	TEXT NN
record_source	VARCHAR(255)

h_department	
department_id	SERIAL NN
load_date	TIMESTAMP NN
record_source	VARCHAR(255)

l_topic_department	
link_id	SERIAL
topic_id	VARCHAR(255) NN
department_id	INTEGER NN
load_date	TIMESTAMP NN
record_source	VARCHAR(255)
(topic_id, load_date)	

h_topic	
topic_id	VARCHAR(255) NN
load_date	TIMESTAMP NN
record_source	VARCHAR(255)
(topic_id, load_date)	

s_topic	
topic_id	VARCHAR(255) NN
load_date	TIMESTAMP NN
end_date	TIMESTAMP
title	VARCHAR(255)
description	TEXT
home_institution	VARCHAR(255)
thesis_type	VARCHAR(50)
author	VARCHAR(255)
work_type	VARCHAR(255)
contact	VARCHAR(255)
url_topic_details	TEXT
requirements	TEXT

## DATABASE SCHEMA

- 2 hubs
- 2 satellite
- 1 link

Bodenmegafauna-Grabgänge: Highways der biologischen Aktivität?	Bachelor	<a href="#">i</a>	praktisch / anwendungsbezogen	Dr. Philipp Maurischat	vergeben	20.02.2023	<a href="#">📄</a>
Masterarbeit in Anlehnung an das Projektband im Sachunterricht	Master	<a href="#">i</a>	nicht spezifiziert	Dr. Mark-Tilo Schmitt	verfügbar	15.02.2023	<a href="#">📄</a>
Hey ChatGPT! Please verify my program.	Bachelor	<a href="#">i</a>	praktisch / anwendungsbezogen	Cedric Richter	reserviert	13.02.2023	<a href="#">📄</a>
Peer-Interaktionen in der Kita - Fragestrategien	Master	<a href="#">i</a>	praktisch / anwendungsbezogen	Aleksandra Anna Kappenberg Prof. Dr. Ulla Licandro	verfügbar	10.02.2023	<a href="#">📄</a>
Entwicklung einer Testautomatisierung für das Testen in operativen Umgebungen	Bachelor oder Master	<a href="#">i</a>	praktisch / anwendungsbezogen	Janusz Andrzej Piotrowski, M. Sc.	verfügbar	09.02.2023	<a href="#">📄</a>
Fehlerinjektion in NMEA0183 Netzwerke mithilfe einer ROS2 basierten Architektur	Bachelor oder Master	<a href="#">i</a>	praktisch / anwendungsbezogen	Janusz Andrzej Piotrowski, M. Sc.	vergeben	09.02.2023	<a href="#">📄</a>
From incorrectness logic proofs to violation witnesses	Bachelor	<a href="#">i</a>	konzeptuell / theoretisch	Prof. Dr. Heike Wehrheim	verfügbar	06.02.2023	<a href="#">📄</a>

< | 1 | 2 | 3 | 4 | ... | 7 | >

Hey ChatGPT! Please verify my program.	Bachelor	<a href="#">i</a>	praktisch / anwendungsbezogen	Cedric Richter	reserviert	13.02.2023	<a href="#">📄</a>
Peer-Interaktionen in der Kita - Fragestrategien	Master	<a href="#">i</a>	praktisch / anwendungsbezogen	Aleksandra Anna Kappenberg Prof. Dr. Ulla Licandro	verfügbar	10.02.2023	<a href="#">📄</a>
Entwicklung einer Testautomatisierung für das Testen in operativen Umgebungen	Bachelor oder Master	<a href="#">i</a>	praktisch / anwendungsbezogen	Janusz Andrzej Piotrowski, M. Sc.	verfügbar	09.02.2023	<a href="#">📄</a>
Fehlerinjektion in NMEA0183 Netzwerke mithilfe einer ROS2 basierten Architektur	Bachelor oder Master	<a href="#">i</a>	praktisch / anwendungsbezogen	Janusz Andrzej Piotrowski, M. Sc.	vergeben	09.02.2023	<a href="#">📄</a>
From incorrectness logic proofs to violation witnesses	Bachelor	<a href="#">i</a>	konzeptuell / theoretisch	Prof. Dr. Heike Wehrheim	verfügbar	06.02.2023	<a href="#">📄</a>
Urbane Gewässer und Renaturierung - Unauflösbarer Gegensatz?	Bachelor	<a href="#">i</a>	praktisch / anwendungsbezogen	Dr. Philipp Maurischat	verfügbar	06.02.2023	<a href="#">📄</a>
Producing an automated tool to extract colour from research imagery	Master	<a href="#">i</a>	praktisch / anwendungsbezogen	Dr. Philipp Maurischat	verfügbar	03.02.2023	<a href="#">📄</a>
Paradigm shifts in dissolved organic matter understanding: The persistence of the overturned	Master	<a href="#">i</a>	konzeptuell / theoretisch	Dr. Philipp Maurischat	verfügbar	03.02.2023	<a href="#">📄</a>

< | 1 | ... | 3 | 4 | 5 | 6 | 7 | >

Year 2021  
→ Crawler issue

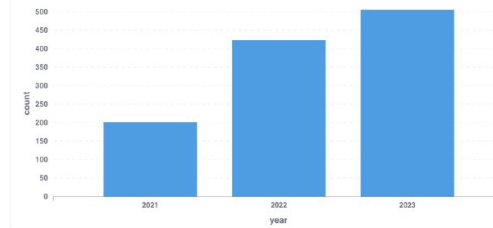
24.05.2023

- Only 54 titles
- Same data items on different pages  
→ based on crawler issuer or stud.ip problem
- Deleted this date

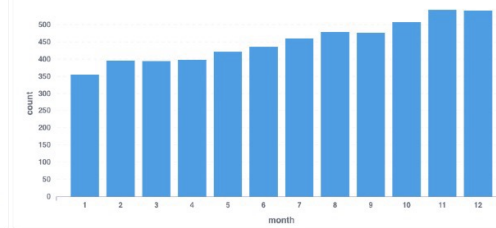
# RESULTS

## thesis topics

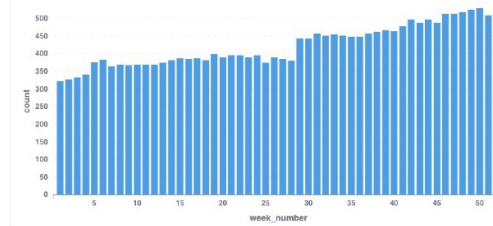
T1.1 - How many thesis topics are published in a year?



T1.2 - How many thesis topics are published in a month?



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



T2 - Which supervisor has the most thesis topics to offer?

contact Dr. Marco Grawunder  
topic\_count 39

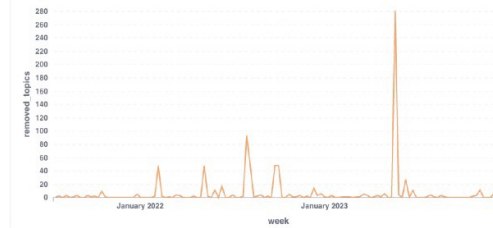
T3 - Which department has the most thesis topics to offer?

department\_name Fach-Bachelor Informatik  
topic\_count 224

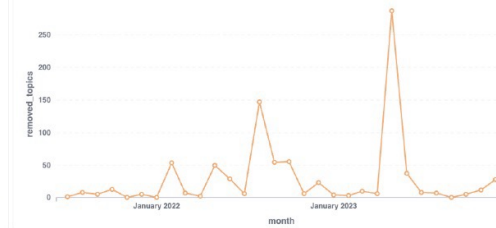
T5 - Which Thesis Type is Most Common?

thesis\_type Bachelor oder Master  
Count 100,094

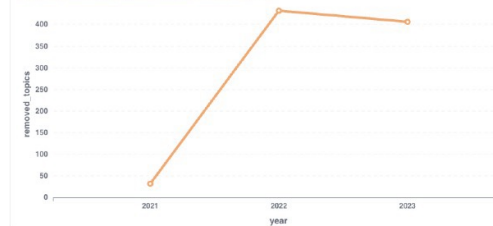
T4.1 - How many thesis topics are "removed from the list" in a week?



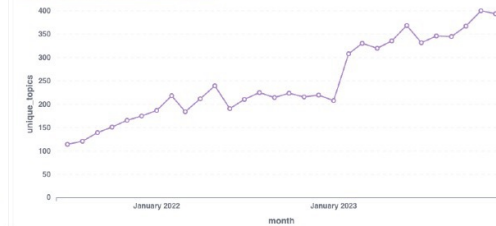
T4.2 - How many thesis topics are "removed from the list" in a month?



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



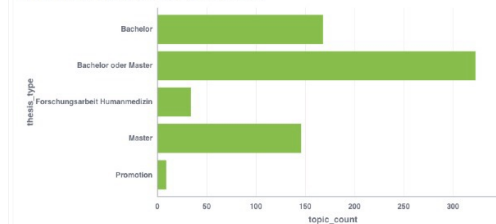
KPI 1 - Unique thesis topics published every month



KPI 2 - Average thesis topics for each department



KPI 3 - distribution of theses topics across different thesis types



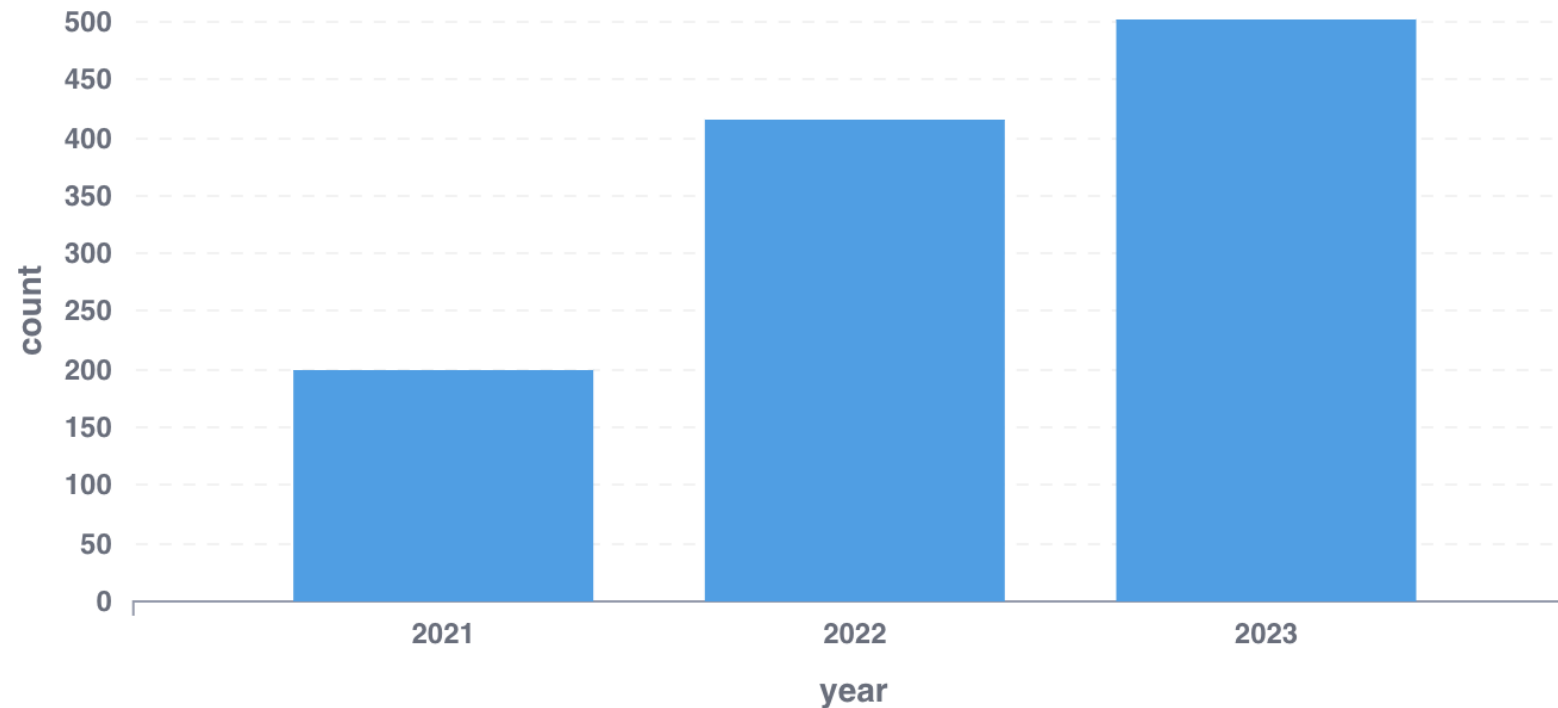
# TASKS

- 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.



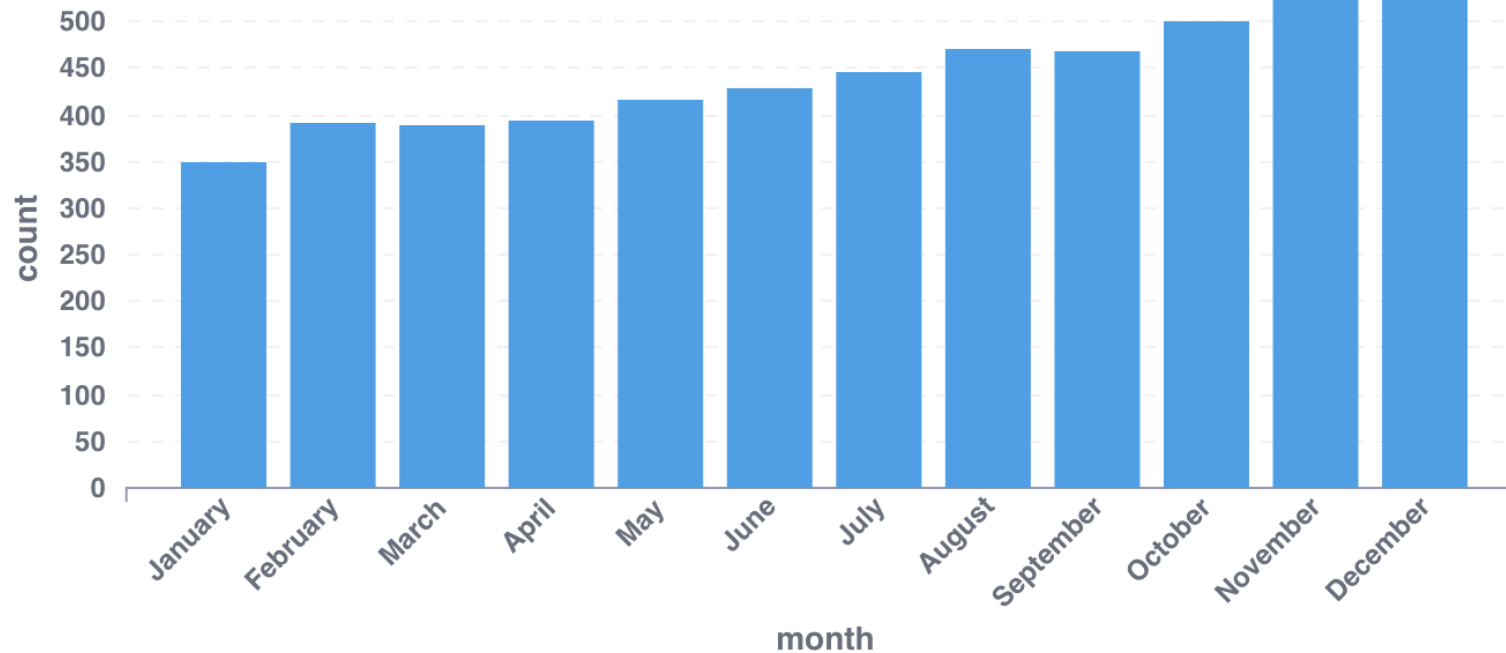
# TI - HOW MANY THESIS TOPICS ARE PUBLISHED IN A WEEK, IN A MONTH, IN A YEAR?

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



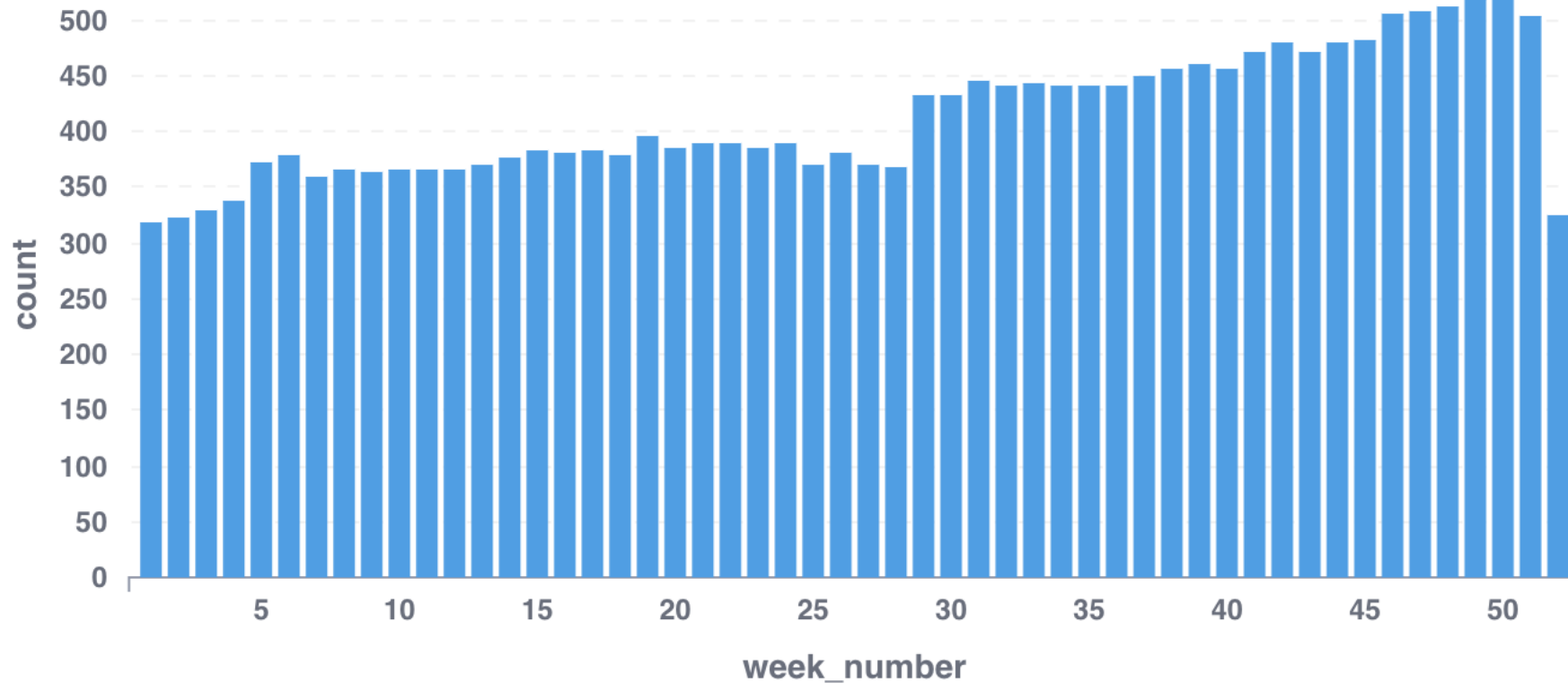
# TI - HOW MANY THESIS TOPICS ARE PUBLISHED IN A WEEK, IN A MONTH, IN A YEAR?

T1.2 - How many thesis topics are published in a month?



# TI - 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 Dr. Marco Grawunder

topic\_coun 39  
t

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

departmen Fach-Bachelor  
t\_name Wirtschaftsinformatik

topic\_coun 113  
t

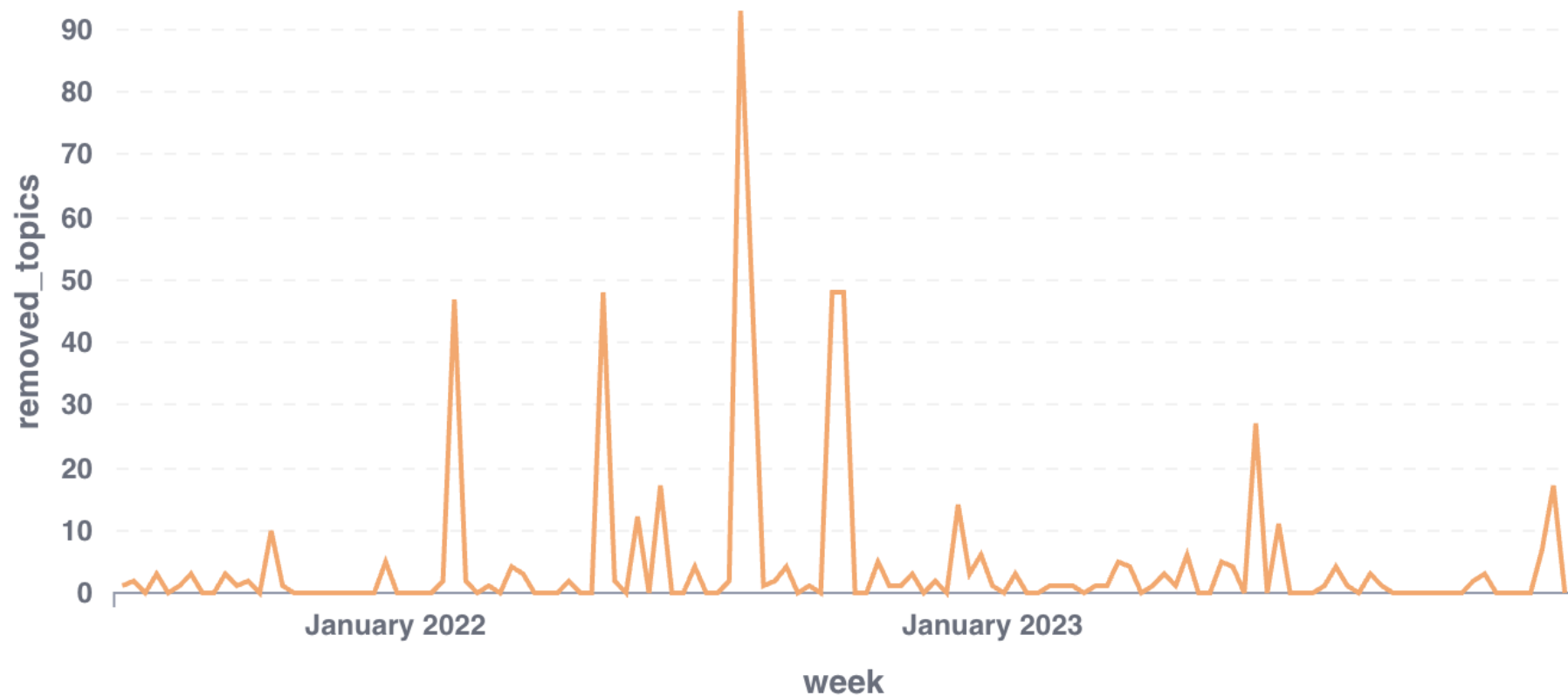
T5 - Which Thesis Type is Most Com...

thesis\_t Bachelor oder  
ype Master

Count 323

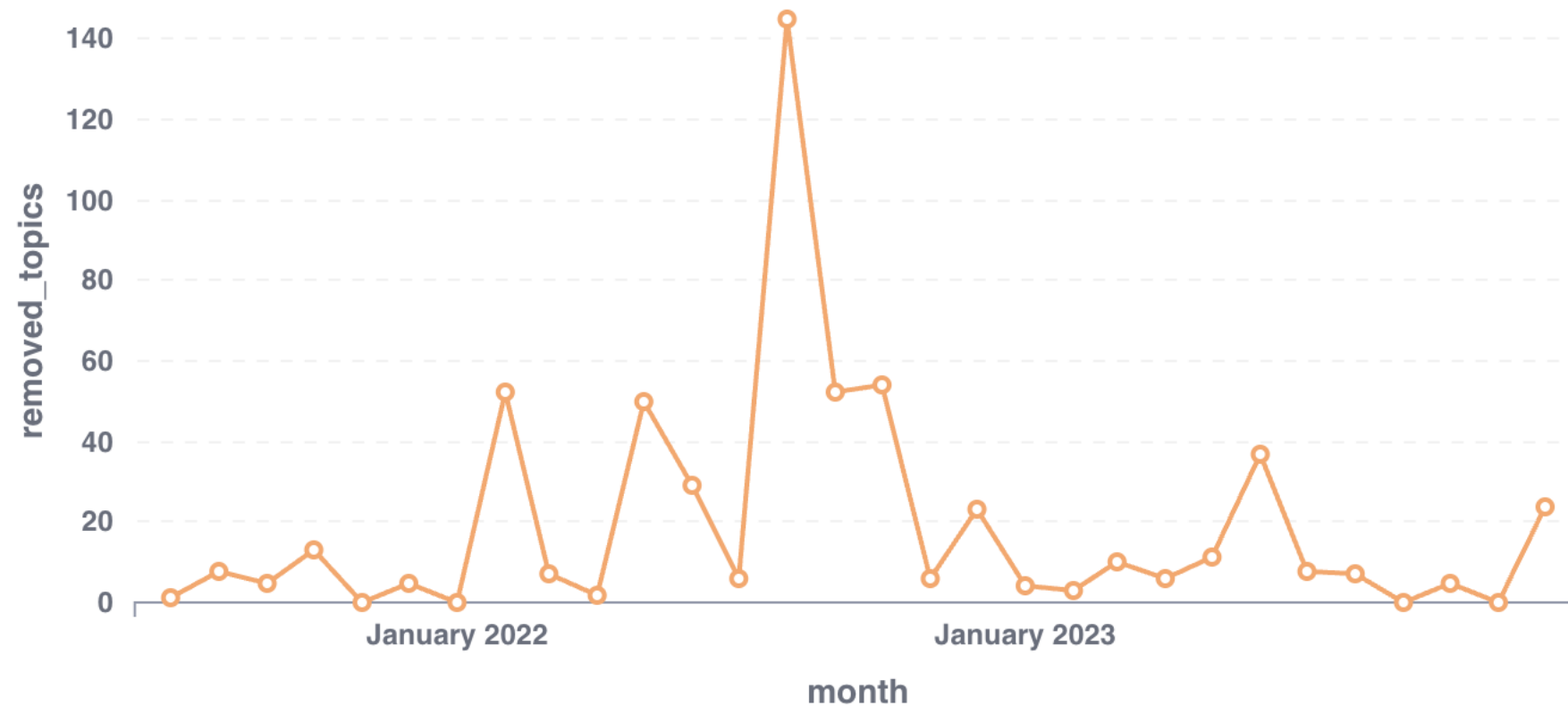
## T4 - HOW MANY THESIS TOPICS ARE "REMOVED FROM THE LIST" IN A WEEK, IN A MONTH, IN A YEAR?

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



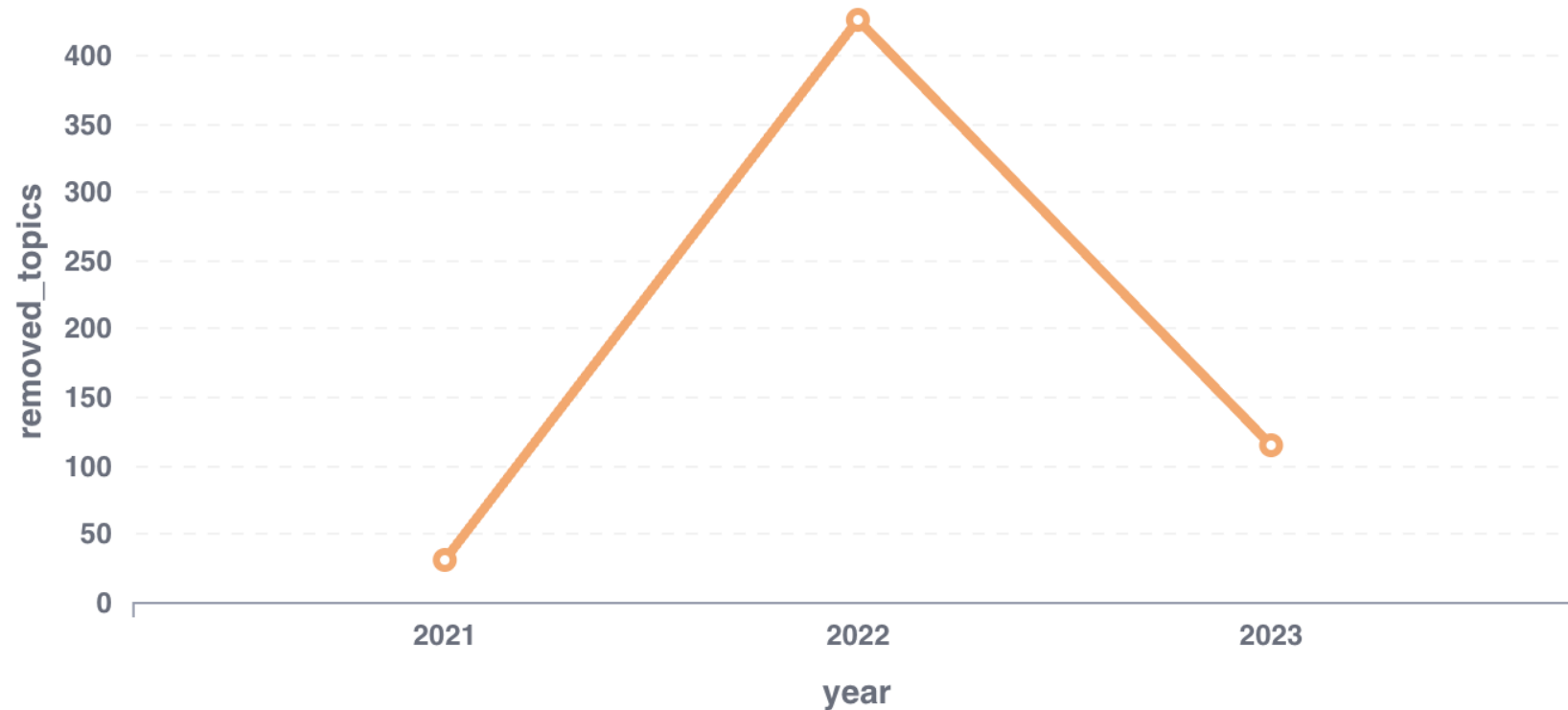
## T4 - HOW MANY THESIS TOPICS ARE "REMOVED FROM THE LIST" IN A WEEK, IN A MONTH, IN A YEAR?

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



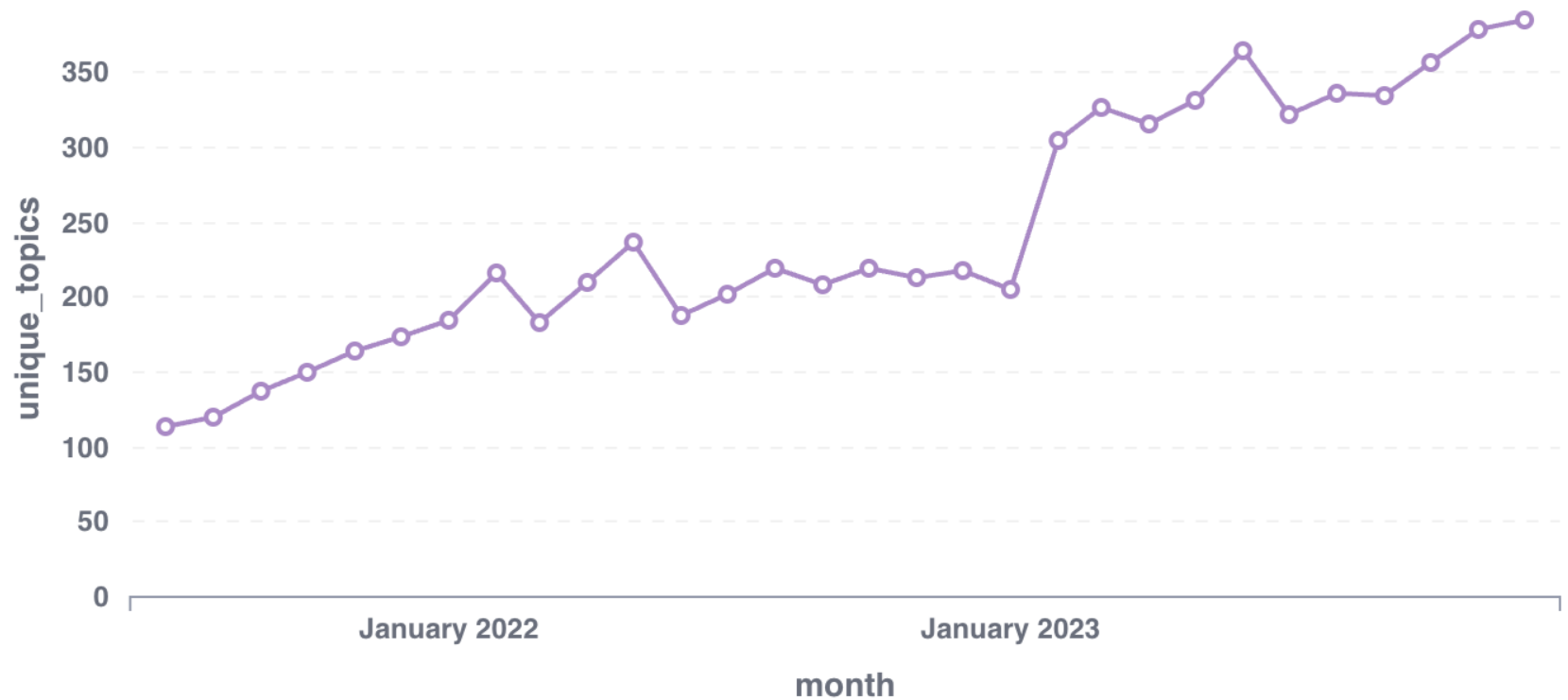
## T4 - HOW MANY THESIS TOPICS ARE "REMOVED FROM THE LIST" IN A WEEK, IN A MONTH, IN A YEAR?

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



## KPI 1 - UNIQUE THESIS TOPICS PUBLISHED EVERY MONTH

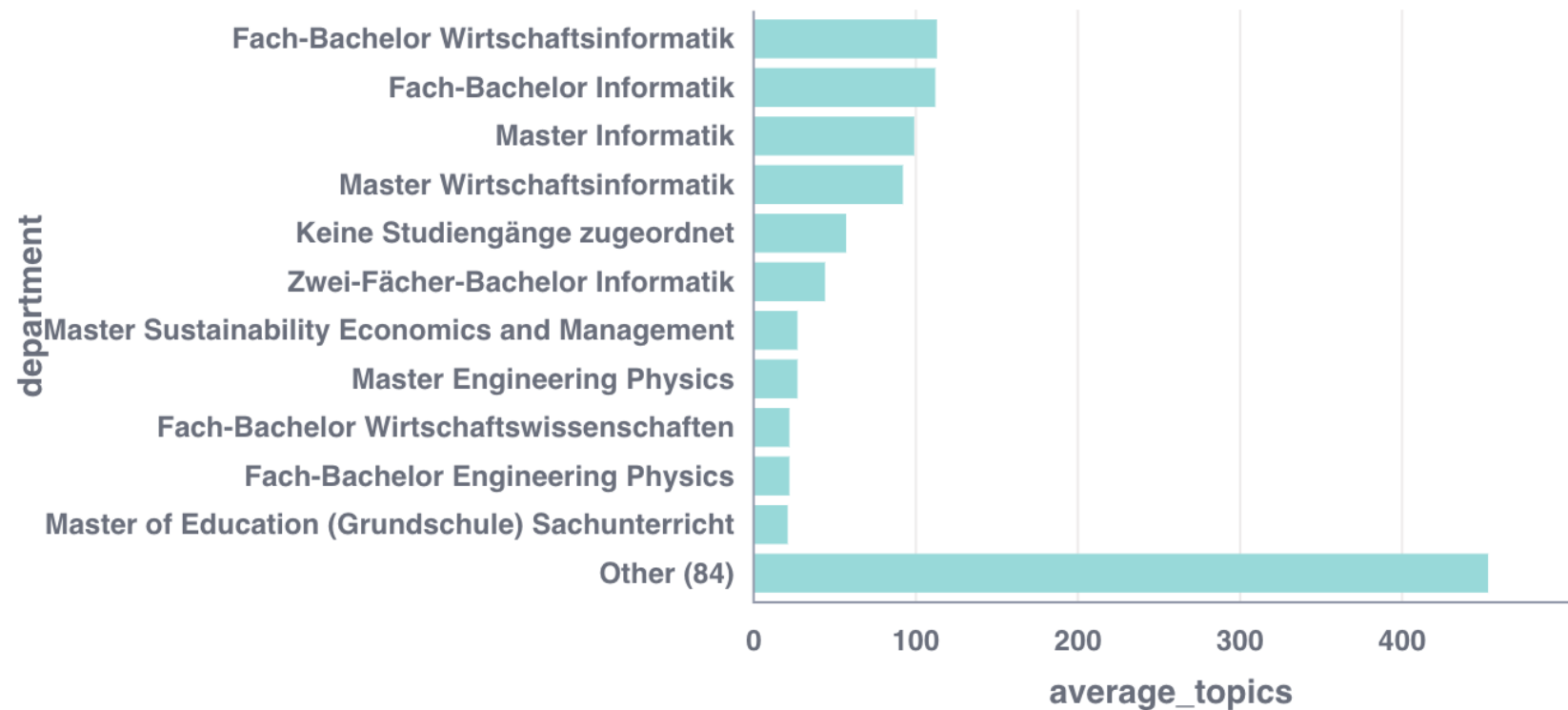
KPI 1 - Unique thesis topics published every month





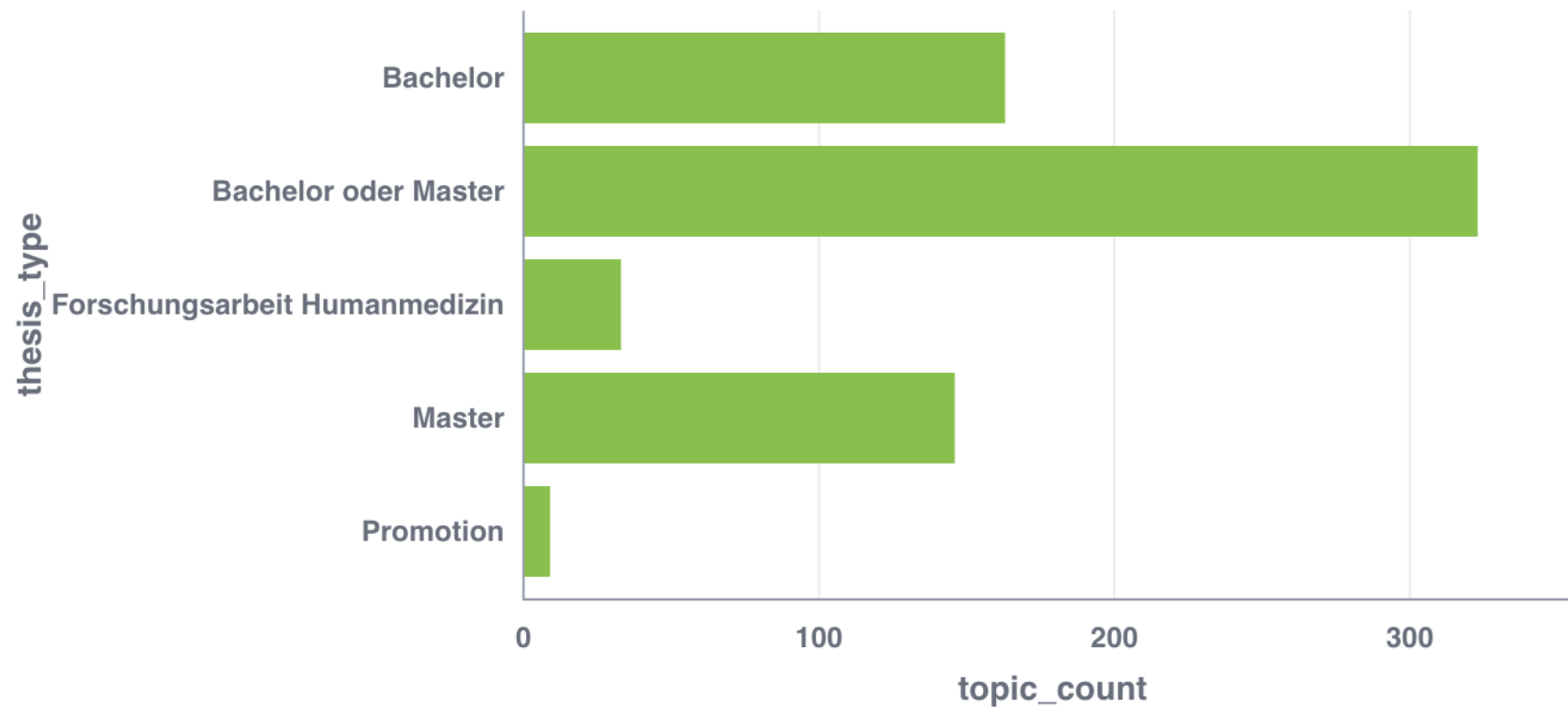
## KPI 2 - AVERAGE THESIS TOPICS FOR EACH DEPARTMENT

KPI 2 - Average thesis topics for each department



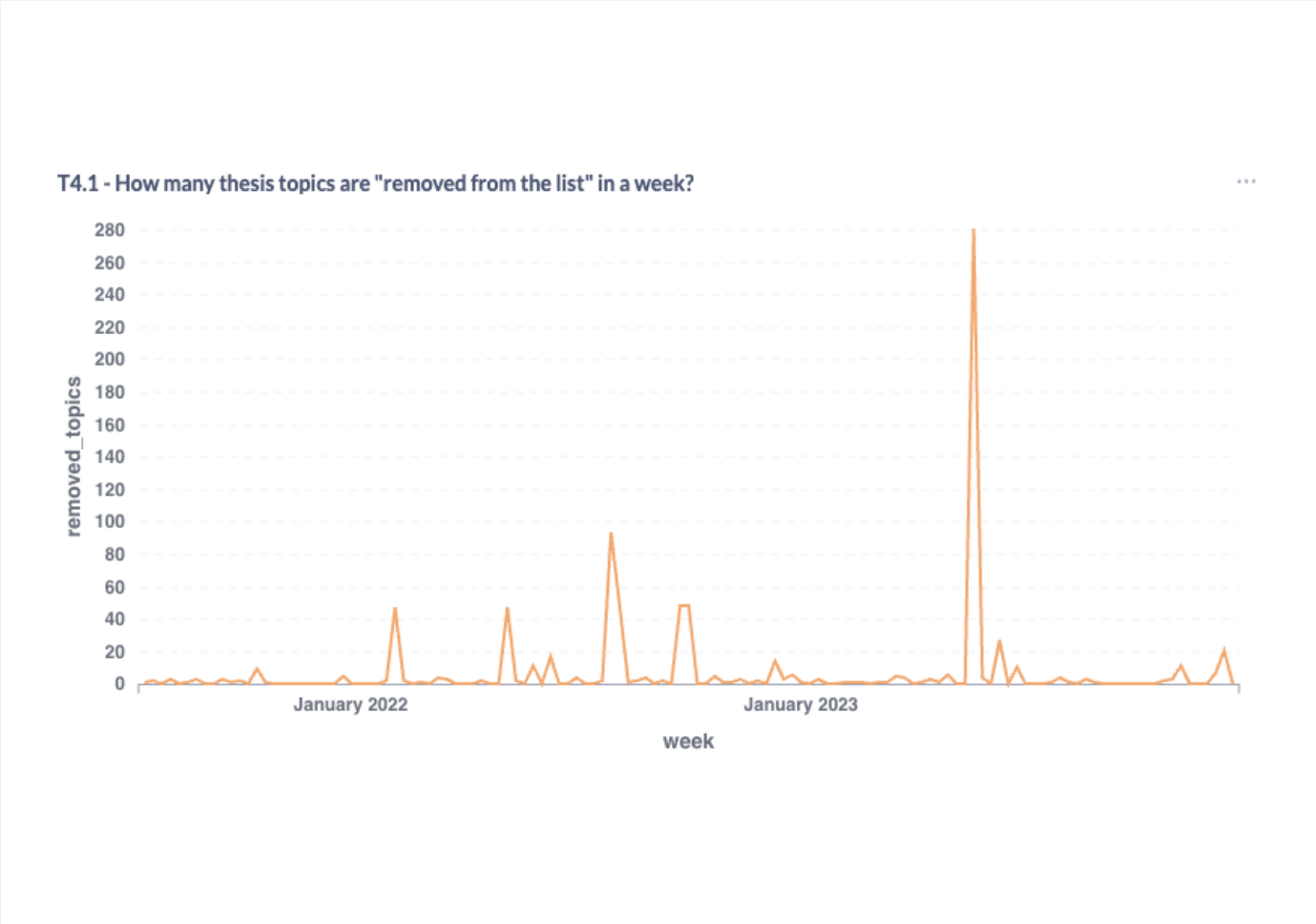
## KPI 3 - DISTRIBUTION OF THESIS TOPICS ACROSS DIFFERENT THESIS TYPES

KPI 3 - distribution of thesis topics across different thesis types



## LESSONS LEARNED: DATA CLEANING

- Extremely high
- Comparison with Levinstein for import-process of the ETL process



## LESSONS LEARNED / FEEDBACK

- Dataset has a issue with year 2021
- Aggregated json/csv files in the dataset was very helpful
- 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