







# Motivation



The screenshot shows the 'socialnet.' website interface. At the top right, there is a language selector 'DE / EN'. Below it are icons for search, portals, and a menu. A navigation bar contains the following categories: Materialien, Lexikon, Rezensionen (highlighted), International, Studium, Branchenbuch, Kalender, and Stellenmarkt. The main content area features a search box with the text 'In den Rezensionen suchen' and a magnifying glass icon. Below the search box, there are several filters: 'Suche nach Titel, Autorin, Rezensentin, Verlag, ISBN/EAN, Schlagwort', 'Suche einschränken', 'Alle Themenbereiche', 'Erscheinungsjahr', 'auch geplante Rezensionen', 'Rezensionen der letzten 3 Jahre', and 'Alle Sprachen'. A trash icon is also present.

socialnet > Rezensionen

## socialnet Rezensionen

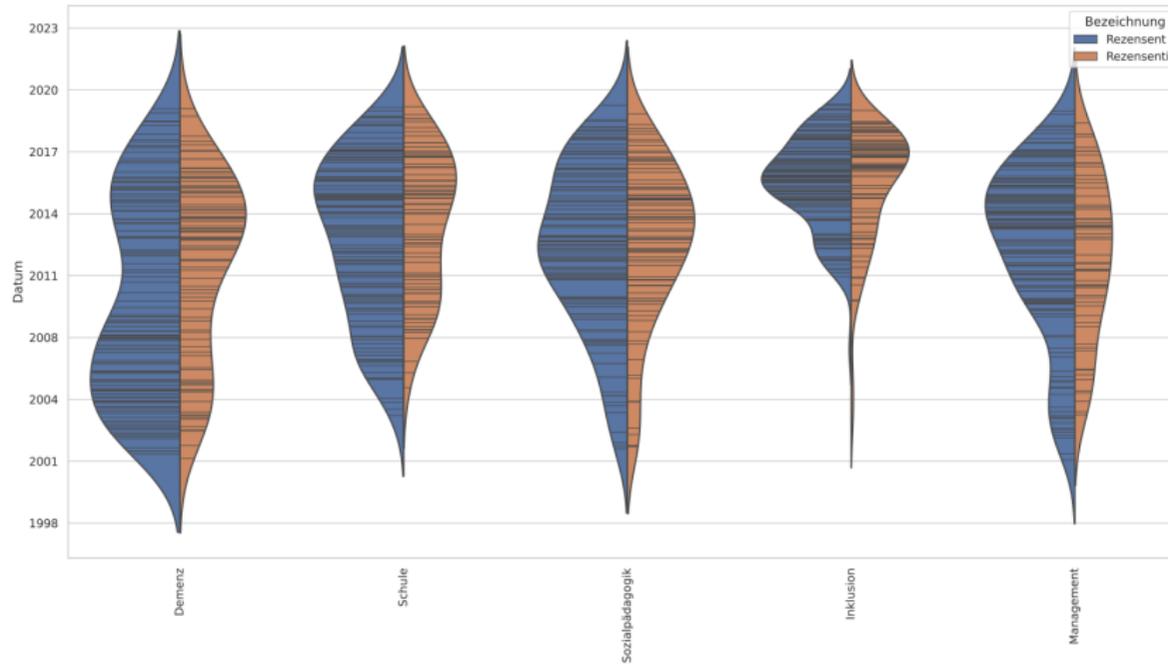
25.084 Rezensionen | Anzeigen nach Schlagwort Verlag Rezensentin Jahr | Empfehlungen der Redaktion

### Themenfelder

Altenhilfe | Behinderung & Inklusion | Gesundheitswesen | Frühpädagogik | Kinder- und Jugendhilfe | Soziale Arbeit und weitere Arbeitsfelder | Methoden & Grundlagen | Sozialpolitik und -verwaltung | Sozialmanagement



[www.extra-mining.de](http://www.extra-mining.de)







## Auswertung “großer” Datenmengen in der Sozialen Arbeit mit KI?

- ▶ “Groß”: Auswertung durch Menschen schwierig
- ▶ Könnten statistisch ausgewertet werden
- ▶ Aber: evtl. fehlen Programmierkenntnisse
- ▶ Auftragsarbeit: ggf. zu kostenintensiv

### KIs (genauer ChatBots bzw. LLMs [1]) können helfen

Im Folgenden:

- ▶ Möglichkeiten und
- ▶ Limitationen am Beispiel eines Datensatzes





# Datensatz

Journalzusammenzug\_aller\_Traeger1.xlsx — LibreOffice

File Edit View Insert Format Styles Sheet Data Tools Window Help

Calibri 11 pt B

A1	Träger							
1	Träger		Journaltyp		Erstelldatum	Verfasser:!	Bereiche	Journaltext
2	Träger 4	4.1	Teilhabe	Sclla	11.07.2013 16:58:11	Ho M		Habe mit Ella den Wäschewa
3	Träger 4	4.1	Pfleg. mediz. Massnahm	Sclla	10.07.2013 07:38:47	Sc R	Gesundheit	Wär gestern mit E.Sch. im RC
4	Träger 4	4.1	Pfleg. mediz. Massnahm	Sclla	20.06.2013 09:32:16	Sc R	Gesundheit	Knie links ist immer noch star
5	Träger 4	4.1	Pfleg. mediz. Massnahm	Eirix	06.03.2013 17:48:19	Sc R	Gesundheit	War heute mit Beatrix beim H
6	Träger 4	4.1	Pfleg. mediz. Massnahm	Eirix	24.08.2013 11:50:20	Sc R	Gesundheit	B.E hat am Knie links eine 5 I
7	Träger 4	4.1	Pfleg. mediz. Massnahm	Sclla	18.04.2014 10:41:04	Cozi		Ella hatte heute Morgen stark
8	Träger 4	4.1	Pfleg. mediz. Massnahm	Hamil	14.11.2012 21:15:21	Sc R	Gesundheit	War mit Emil beim Hausarzt [
9	Träger 4	4.1	Pfleg. mediz. Massnahm	MüMax	17.03.2014 06:48:09	Cozi	Gesundheit	Max möchte heute noch zuha
10	Träger 4	4.1	Teilhabe	Hamil	21.05.2013 13:55:40	Cozi	Freizeit	Nachtrag zum Sonntag: Emil
11	Träger 4	4.1	Verlaufsprotokoll	Meter	10.11.2012 11:21:21	Po L	Wohnen	Wädi hat den Frühstück teils r
12	Träger 4	4.1	Teilhabe	Eirix	14.02.2014 20:26:27	Stle	Bildung, Wohnen	B.E. wollte mir heute Abend n
13	Träger 4	4.1	Pfleg. mediz. Massnahm	Meter	12.11.2012 14:44:18	Po L	Gesundheit	Wädi hat am Oberkopf eine S
14	Träger 4	4.1	Teilhabe	Hamil	10.01.2014 19:14:29	Po L	Wohnen	Vor den Festtagen fragte MM
15								Teilhaberückblick der letzten Das Leben und die Teilhabe Die letzten Monate war die Zu Die Wohnungsferien in Interla Ein Höhepunkt waren die Fes Emil sein Alltag ist immer wie Emil genießt es wenn sich di Emil wird Ende Monat pensio Für die Zeit nach der Pension
	Träger 4	4.1	Teilhabe	Hamil	04.03.2013 15:19:22	Cozi	Wohnen, Freizeit,	
16	Träger 4	4.1	Verlaufsprotokoll	Meter	10.10.2013 13:51:21	Po L	Wohnen	Vor dem Mittagessen hat BE
17	Träger 4	4.1	Teilhabe	Eirix	14.11.2013 08:26:54	Sur	Wohnen	Da Ella heute frei hat und aus



# Datenschutz

- ▶ Daten aus Bereichen Gesundheit etc. oft kritisch
- ▶ Dürfen ggf. nicht durch Dritte verarbeitet werden
- ▶ Ist KI-gestützte Auswertung trotzdem möglich?



# ChatGPT

- ▶ Chatbot, der darauf trainiert wurde, Text-Instruktionen zu befolgen
- ▶ GPT-4 (25€/Monat)

## Randbedingungen:

- ▶ ChatGPT bekommt niemals direkten Zugriff auf die Daten
- ▶ Ich schreibe keine einzige Zeile Code selbst



# Explorative Datenanalyse



# Explorative Datenanalyse

Wie viele Einträge hat unser Datensatz überhaupt?

## Nachricht

Ich habe eine xlsx-Datei mit einem Datensatz. Die Datei hat folgende Spalten: Träger, Unbenannt, Journaltyp, Klient, Erstelldatum, Verfasser:in, Bereiche, Journaltext. Bitte schreibe mir Code, um die Anzahl der Einträge pro Monat zu visualisieren.



Reasoned about data visualization in Excel for 4 seconds ▾

Hier ein Beispielcode in Python, der die Daten aus der XLSX-Datei lädt, das Erstelldatum in einen Datumswert umwandelt, nach Monaten gruppiert und das Ergebnis als Balkendiagramm anzeigt:

```
python Copy Edit

import pandas as pd
import matplotlib.pyplot as plt

# 1) XLSX einlesen
df = pd.read_excel("daten.xlsx")

# 2) Erstelldatum in Datumsformat umwandeln (ggf. dayfirst=True setzen, falls das Datum
df["Erstelldatum"] = pd.to_datetime(df["Erstelldatum"], dayfirst=True)

# 3) Monatsspalte hinzufügen (Jahr-Monat)
df["Monat"] = df["Erstelldatum"].dt.to_period("M")

# 4) Anzahl Einträge je Monat zählen
eintraege_pro_monat = df.groupby("Monat").size()

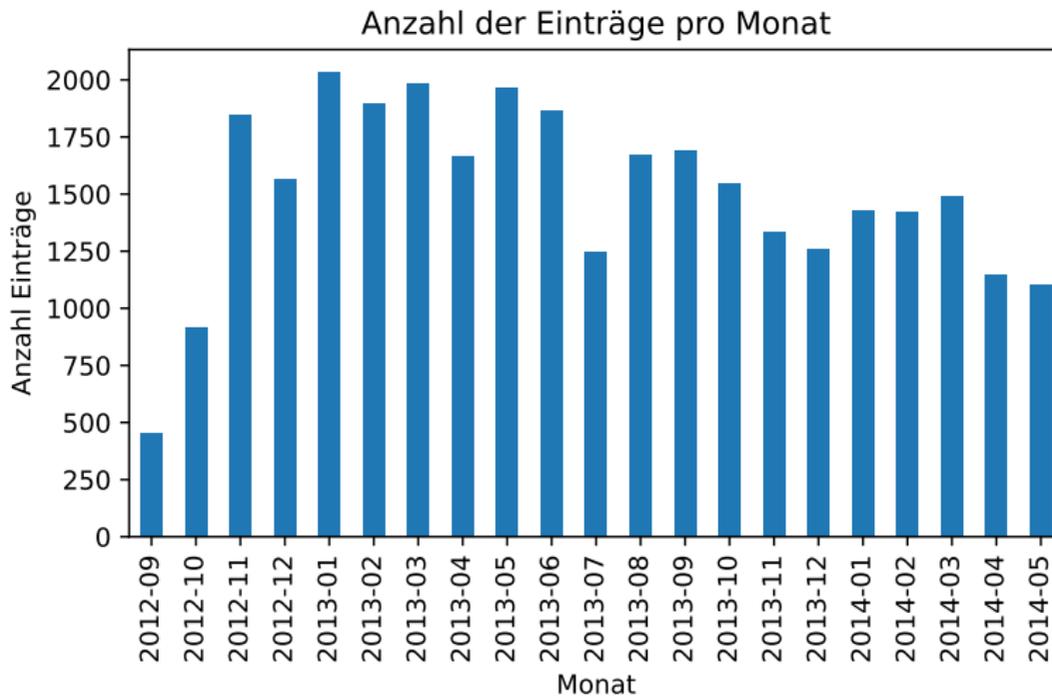
# 5) Balkendiagramm erstellen
eintraege_pro_monat.plot(kind="bar")
plt.title("Anzahl der Einträge pro Monat")
plt.xlabel("Monat")
plt.ylabel("Anzahl Einträge")
plt.show()
```

#### Hinweise:

- Passe den Dateinamen (`daten.xlsx`) an deinen tatsächlichen Dateinamen an.
- Falls das Datumsformat in der Tabelle anders ist, musst du eventuell weitere Parameter bei `to_datetime` angeben (z.B. `format="%d.%m.%Y"`).



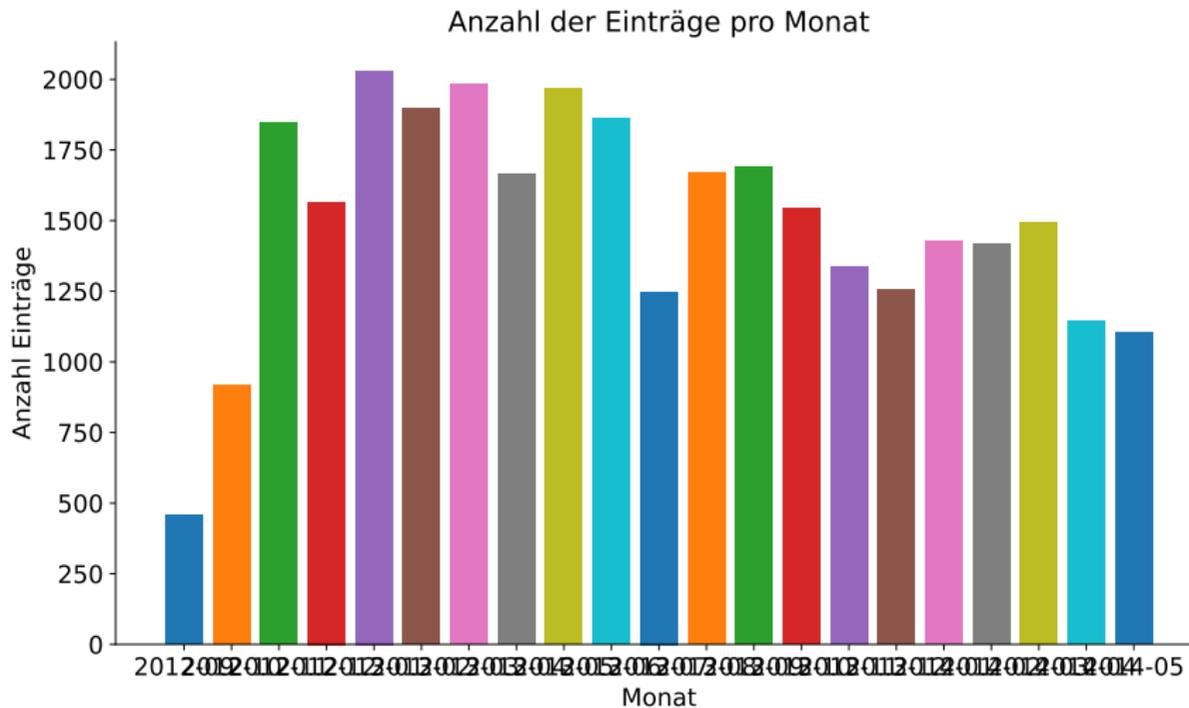
# Ergebnis







# Ergebnis





# Explorative Datenanalyse

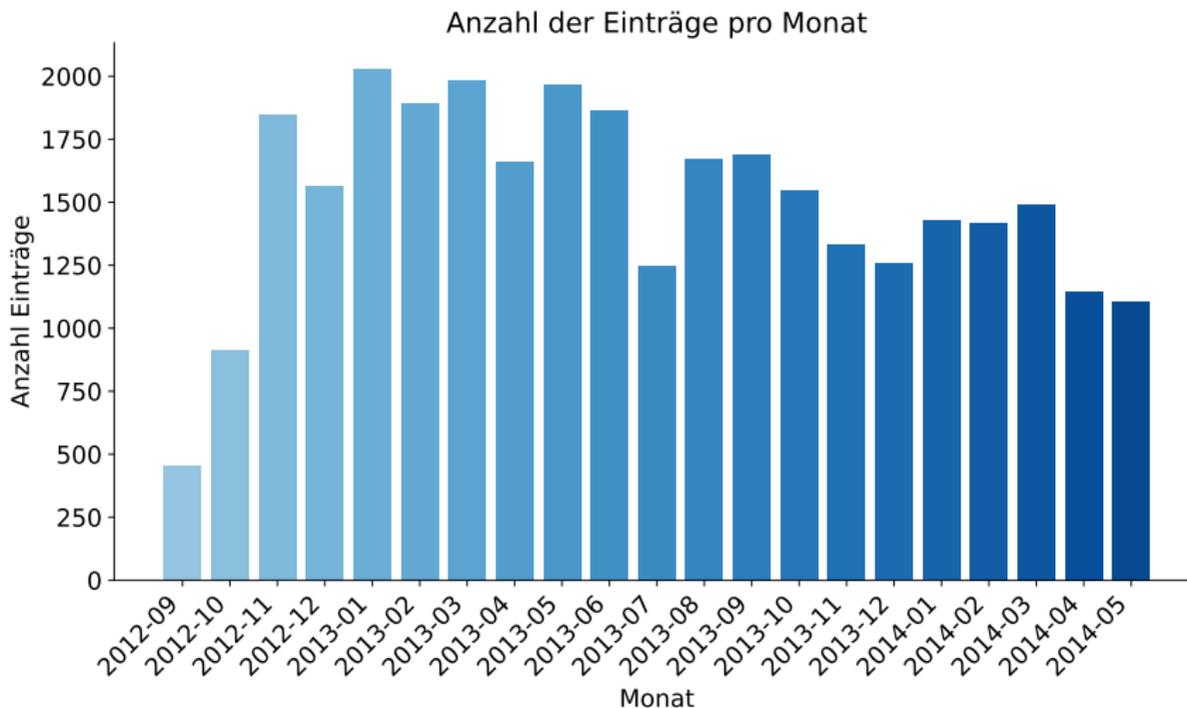
Geht das schöner?

## Nachricht

Die Zahlen unten überlappen sich. Außerdem ist das Farbschema zu bunt, bitte benutze etwas Einheitliches, am besten mit unterschiedlichen Blautönen.



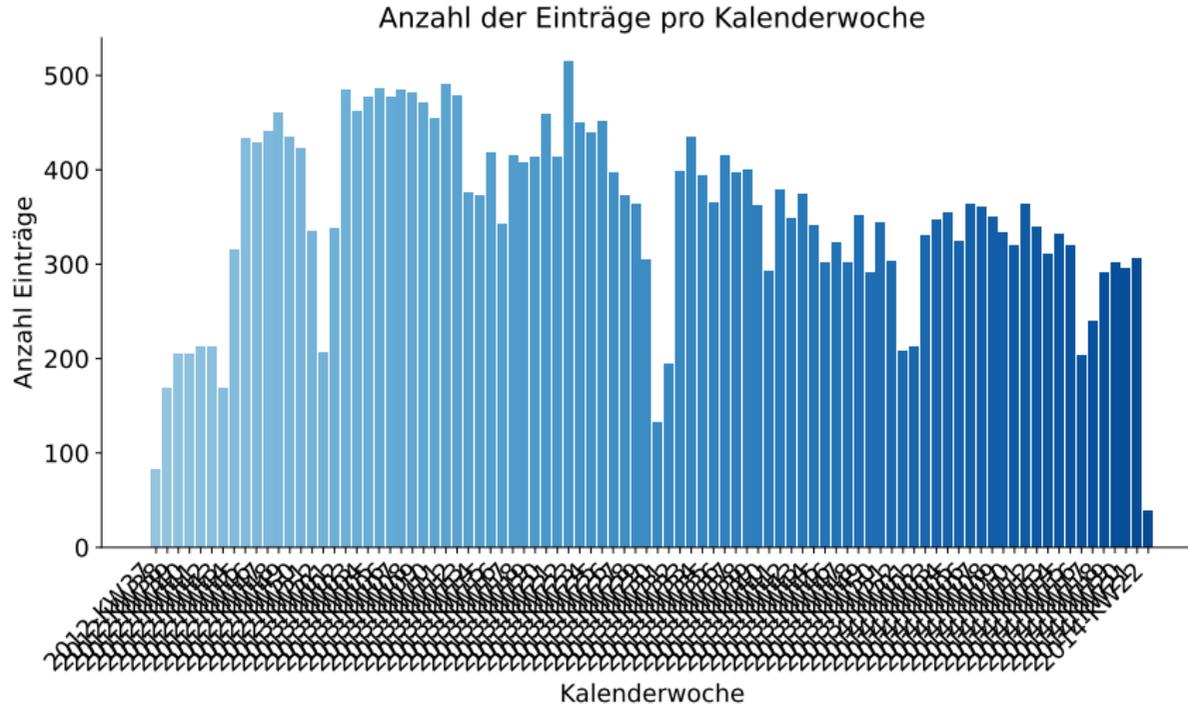
# Ergebnis







# Ergebnis





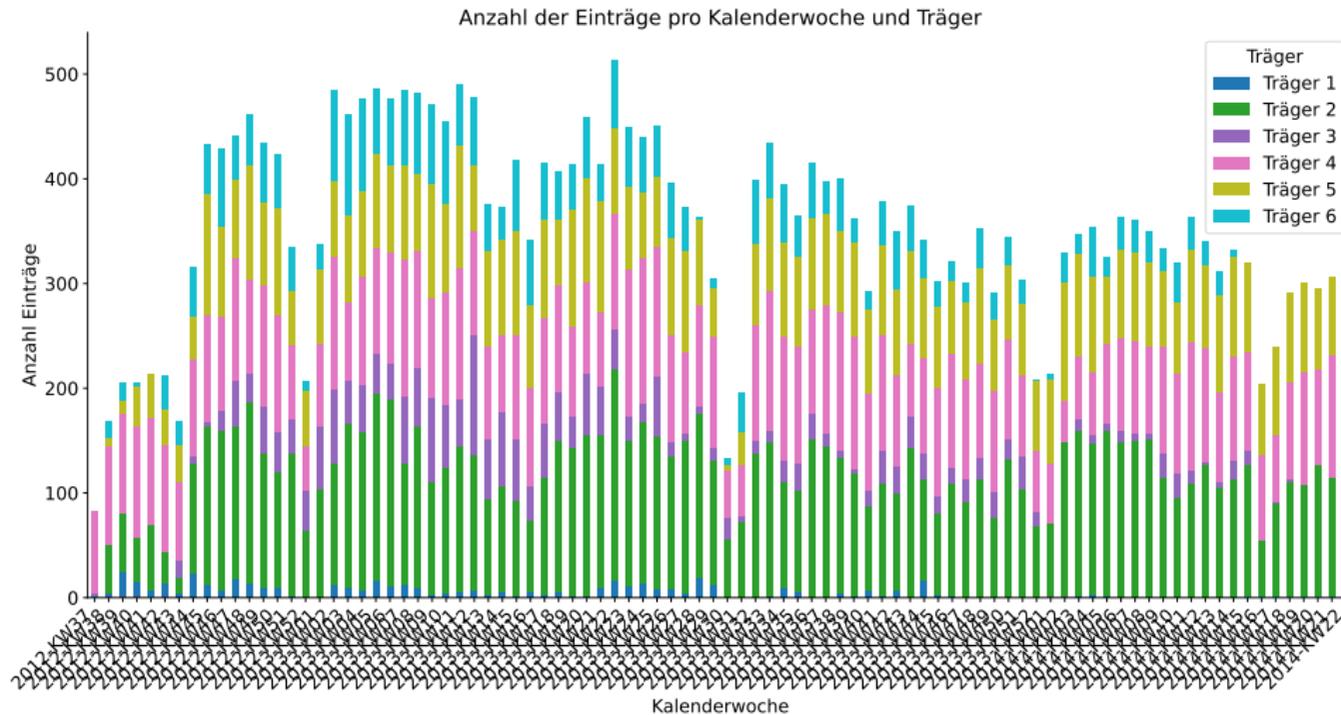
# Explorative Datenanalyse

## Nachricht

Die Zahlen unten überlappen sich wieder, weil zu wenig Platz ist. Ich möchte mir außerdem einzelne Träger anschauen. Kannst du mir die Anzahl der Einträge pro Woche **pro Träger** visualisieren? Vielleicht kann man den unterschiedlichen Trägern unterschiedliche Farben geben?



# Ergebnis





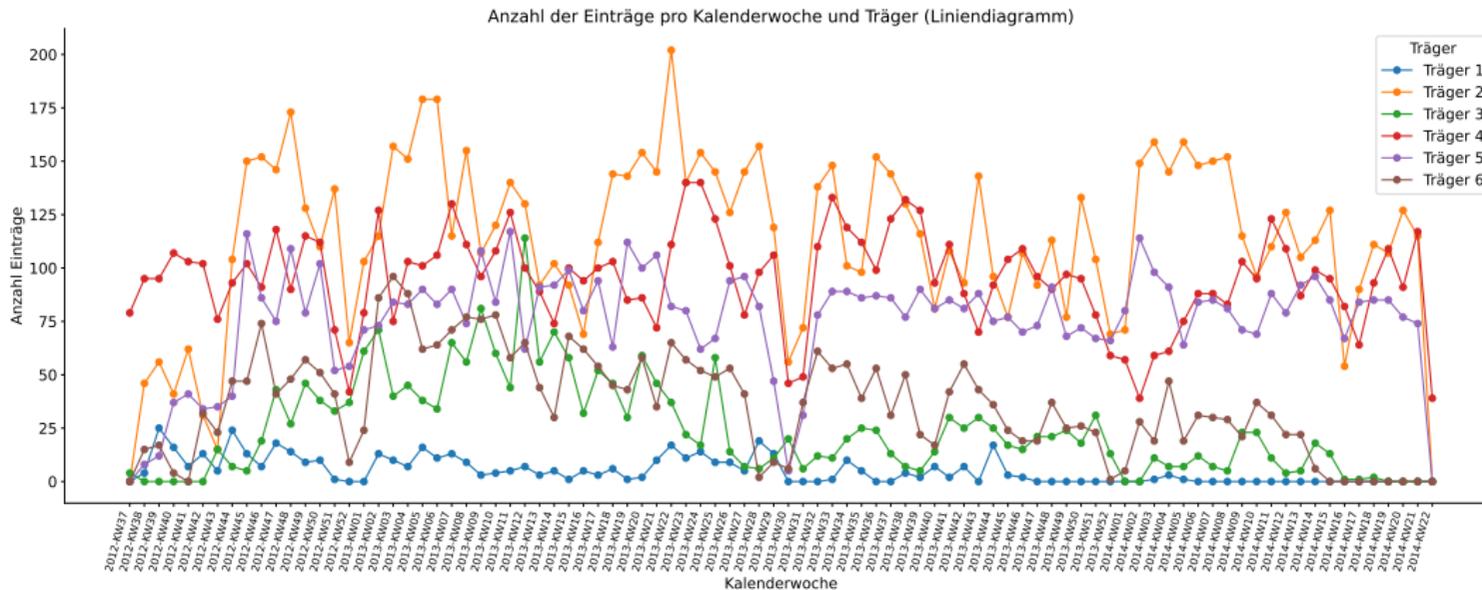
# Explorative Datenanalyse

## Nachricht

Die Labels unten überlagern sich noch immer. Vielleicht kann man sie verkleinern oder stärker rotieren? Kannst du das Ganze bitte als Liniendiagramm darstellen?



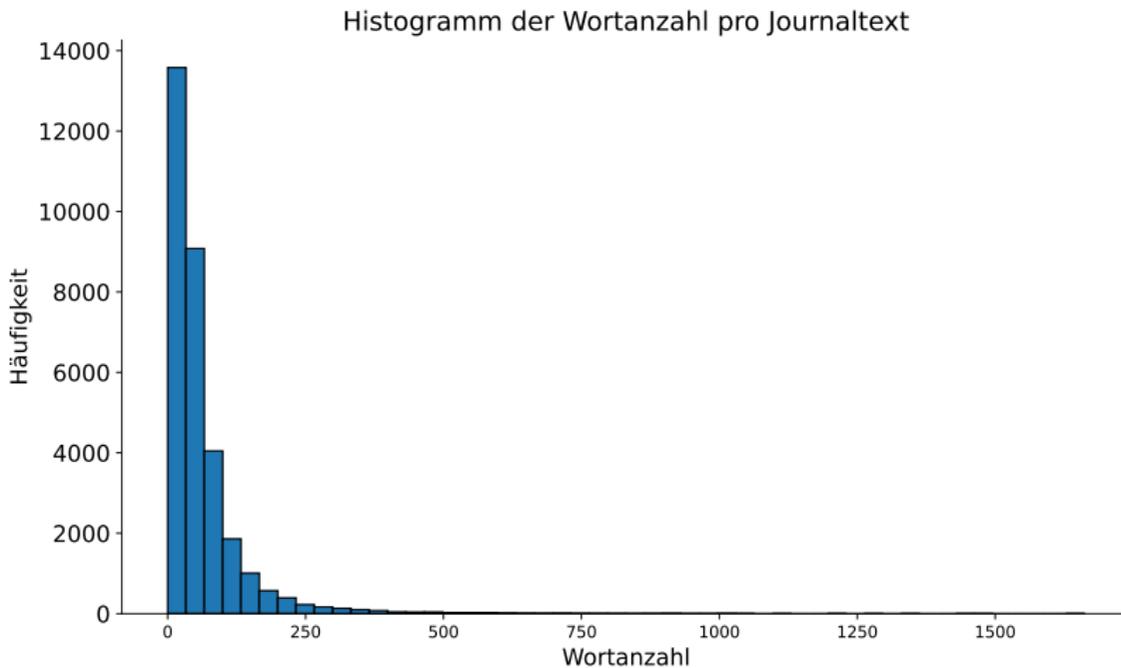
# Ergebnis







# Ergebnis





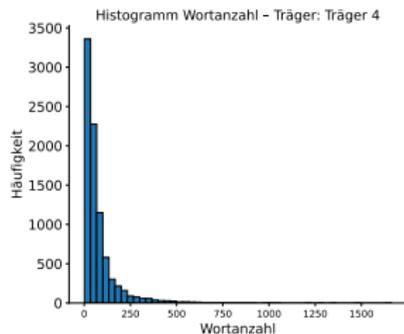
# Explorative Datenanalyse

## Nachricht

Bitte zeichne ein solches Histogramm für jeden der Träger. Die Diagramme sollen in einem Gitter mit je drei Trägern pro Zeile angeordnet sein. Es gibt 6 Träger insgesamt.



# Ergebnis





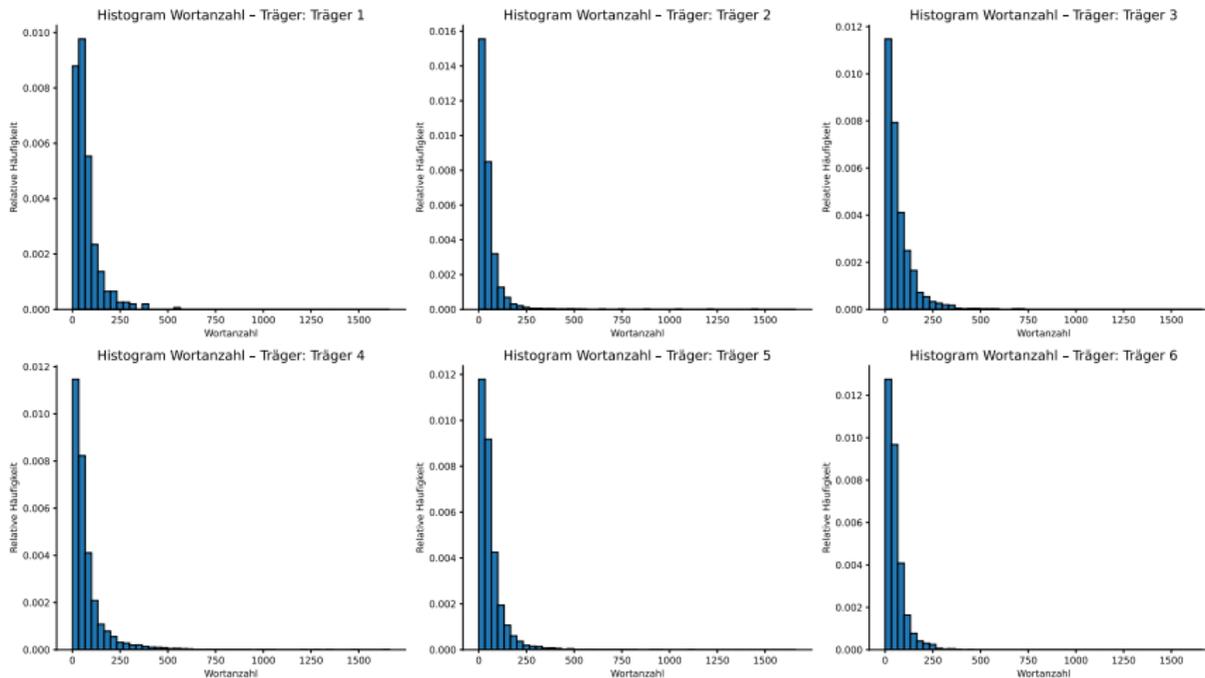
# Explorative Datenanalyse

## Nachricht

Die Skalierung der x-Achsen ist unterschiedlich. Außerdem sind die y-Achsen nicht normalisiert. Ordne die Träger bitte alphabetisch.



# Ergebnis







## Vorschläge

- ▶ Häufigkeitsverteilung wichtiger Worte (Wortwolken, Keyword-Analysen)
- ▶ Journaltyp vs. Wortanzahl: Unterscheidet sich die durchschnittliche Wortanzahl zwischen verschiedenen Journaltypen?
- ▶ Bereiche vs. Wortanzahl: Welche Bereiche haben tendenziell längere Journale?
- ▶ Bereiche vs. Zeitraum: Sind bestimmte Bereiche nur zu bestimmten Zeitpunkten aktiv?
- ▶ Stimmungsmessung (Sentiment Analysis): Falls der Journaltext etwas über Stimmungslagen oder Bewertungen aussagt (benötigt eventuell eine vordefinierte Sentiment-Pipeline).
- ▶ Für die schnelle Exploration könnte man ein kleines Dashboard (z.B. mit Plotly, Dash oder Streamlit) erstellen, um Filter (Träger, Journaltyp, Zeitraum) direkt anzuwenden und die Daten nach Bedarf live zu visualisieren.



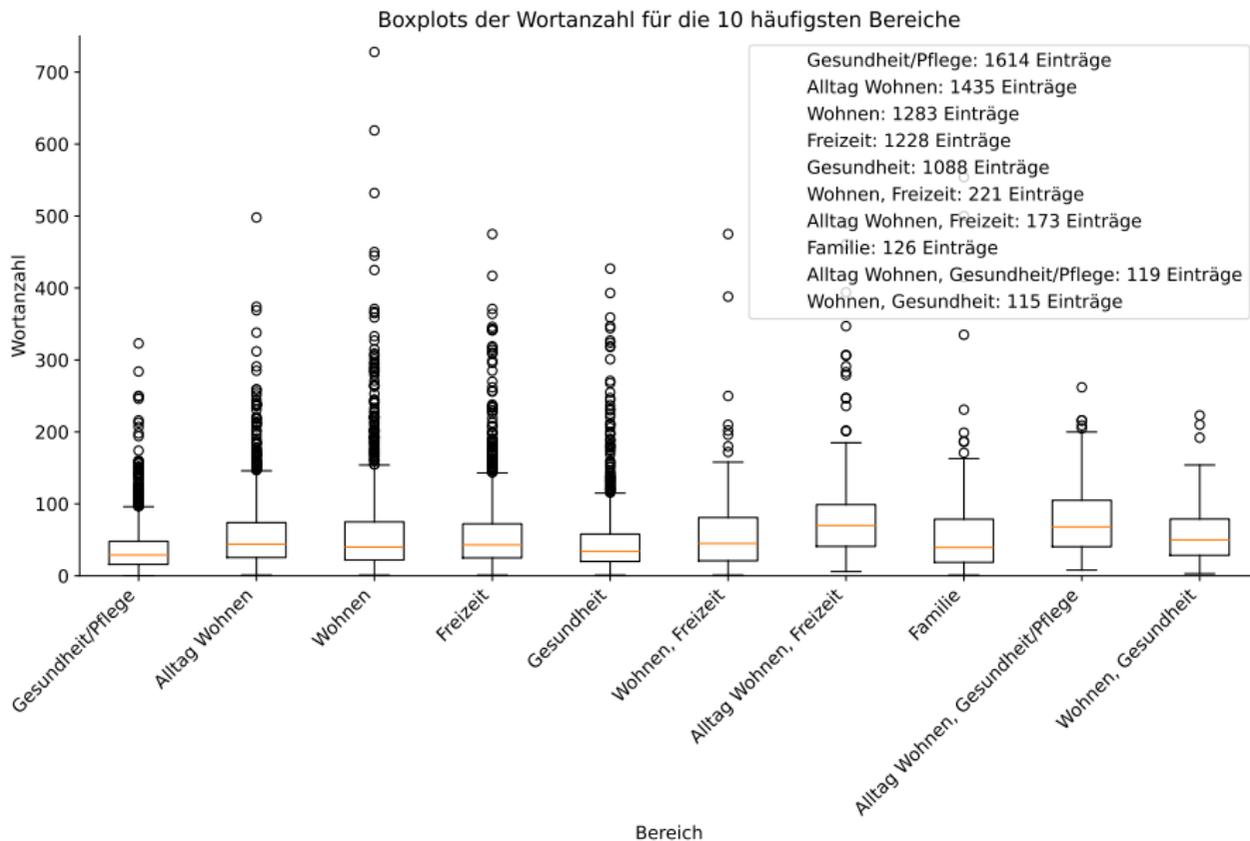
# Bereiche

## Nachricht

Bitte erstelle Boxplots für die Anzahl der Wörter für jeden Bereich.

[...]

Der Plot ist überfüllt. Bitte nutze nur die 10 Bereiche mit den meisten Einträgen. Sortiere nach Anzahl. Schreibe die Anzahl der Einträge pro Bereich in die Legende.



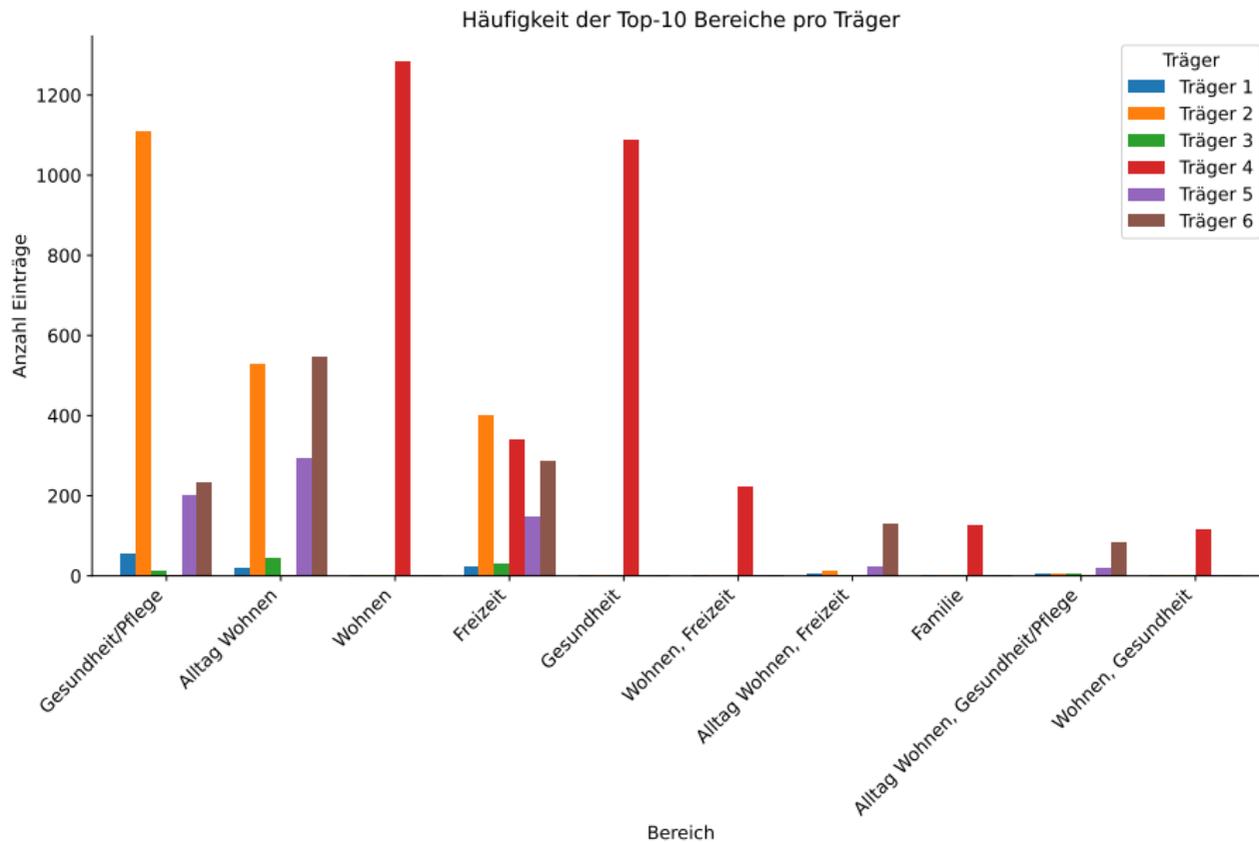


# Bereiche

Einige Bereiche überlappen sich? Nutzen unterschiedliche Träger unterschiedliche Bereiche?

## Nachricht

Bitte zeichne mir ein Balkendiagramm, das für jeden der Top-10-Bereiche zeigt, wie oft ein jeweiliger Träger sie verwendet hat. Bitte gib jedem der 6 Träger eine eigene Farbe. Ordne die Träger alphabetisch.

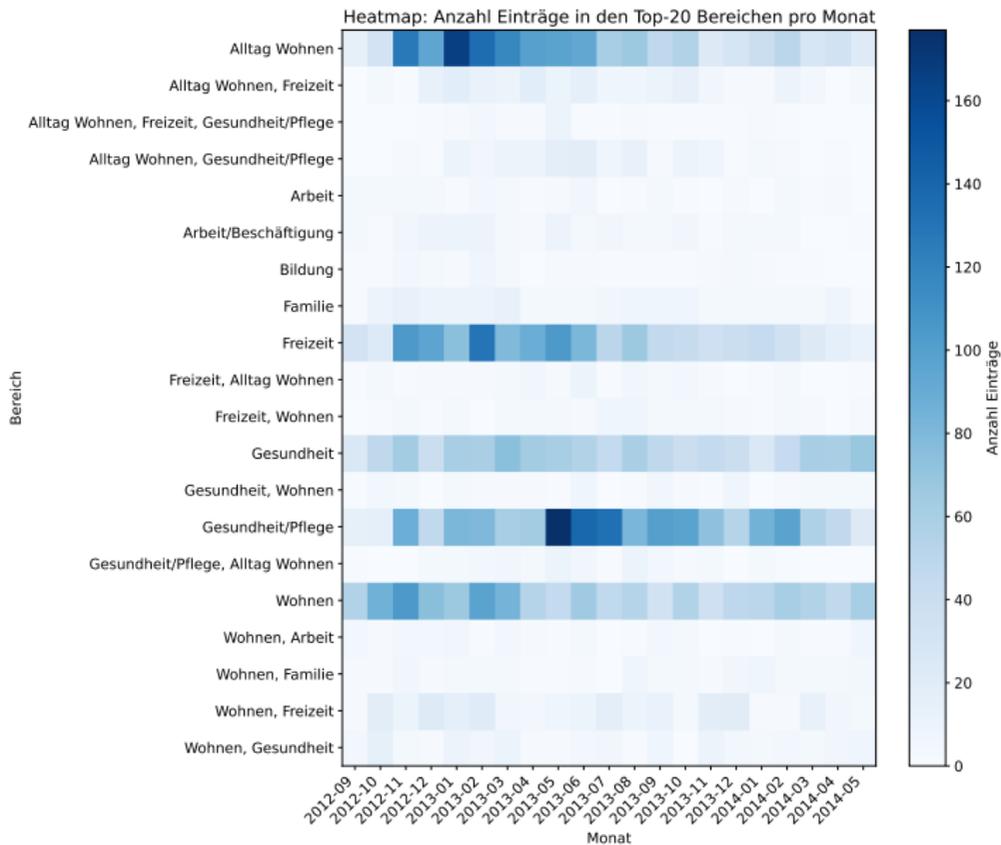


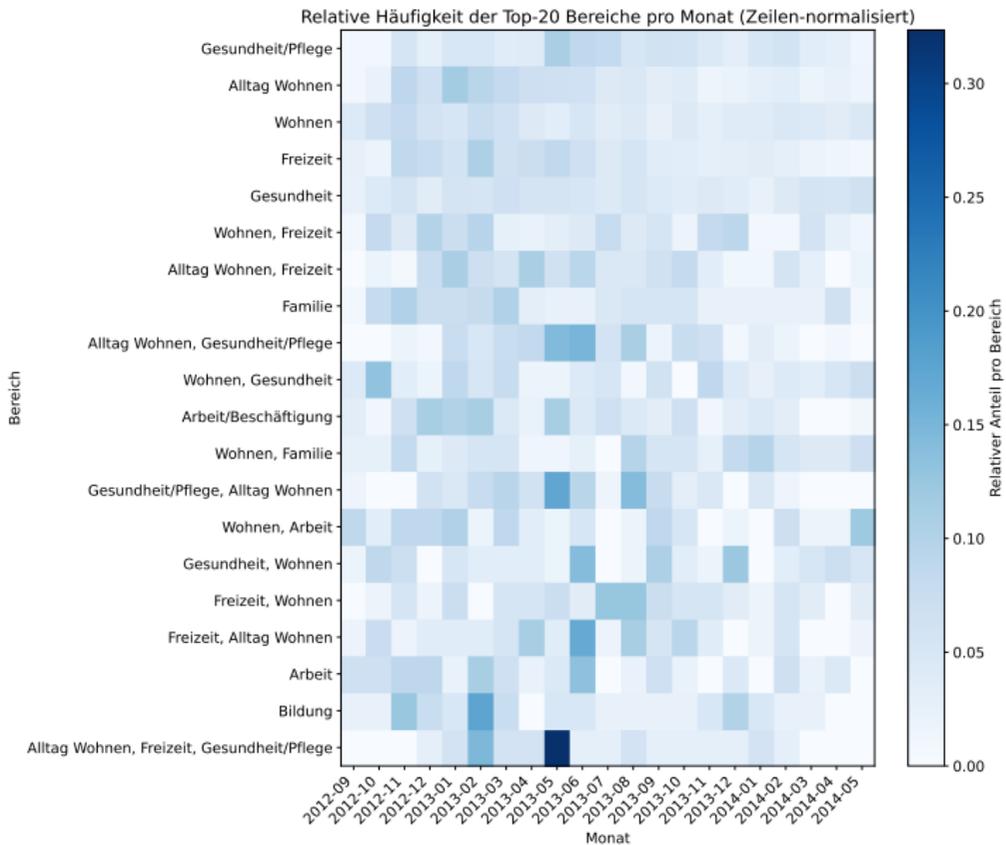


## Bereiche + Zeit

### Nachricht

Du hast vorgeschlagen: Bereiche vs. Zeitraum: Sind bestimmte Bereiche nur zu bestimmten Zeitpunkten aktiv? Wie kann ich das herausfinden?  
Bitte zeige nur die Top-20-Bereiche.





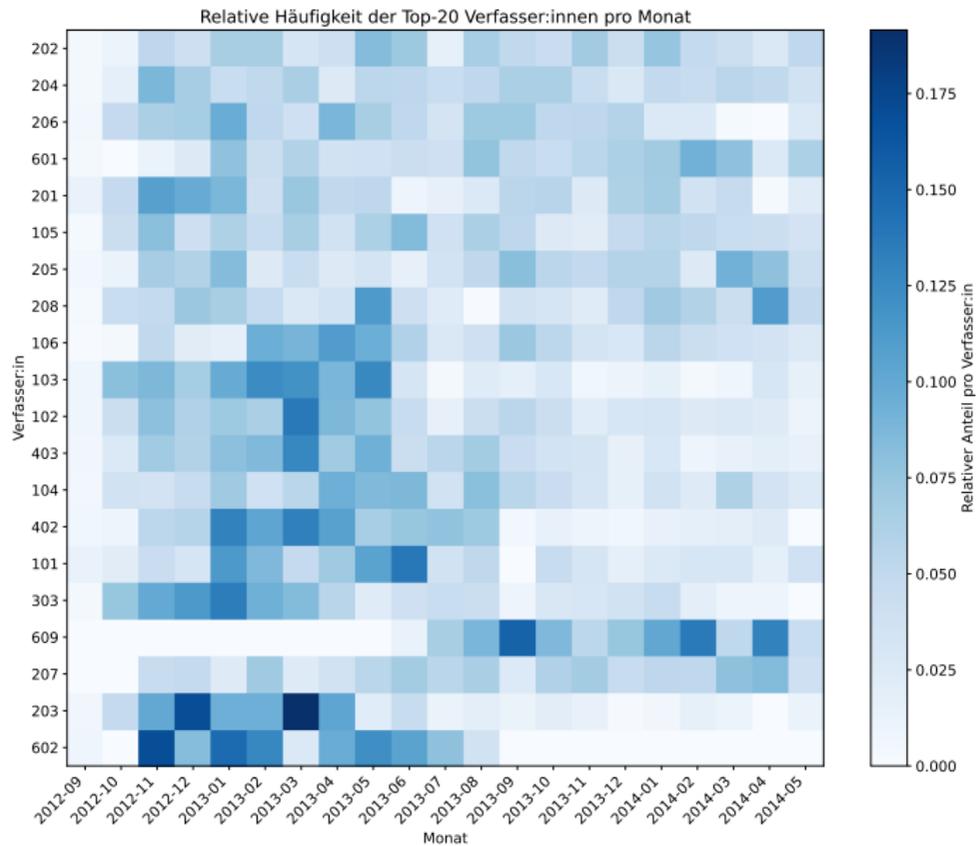


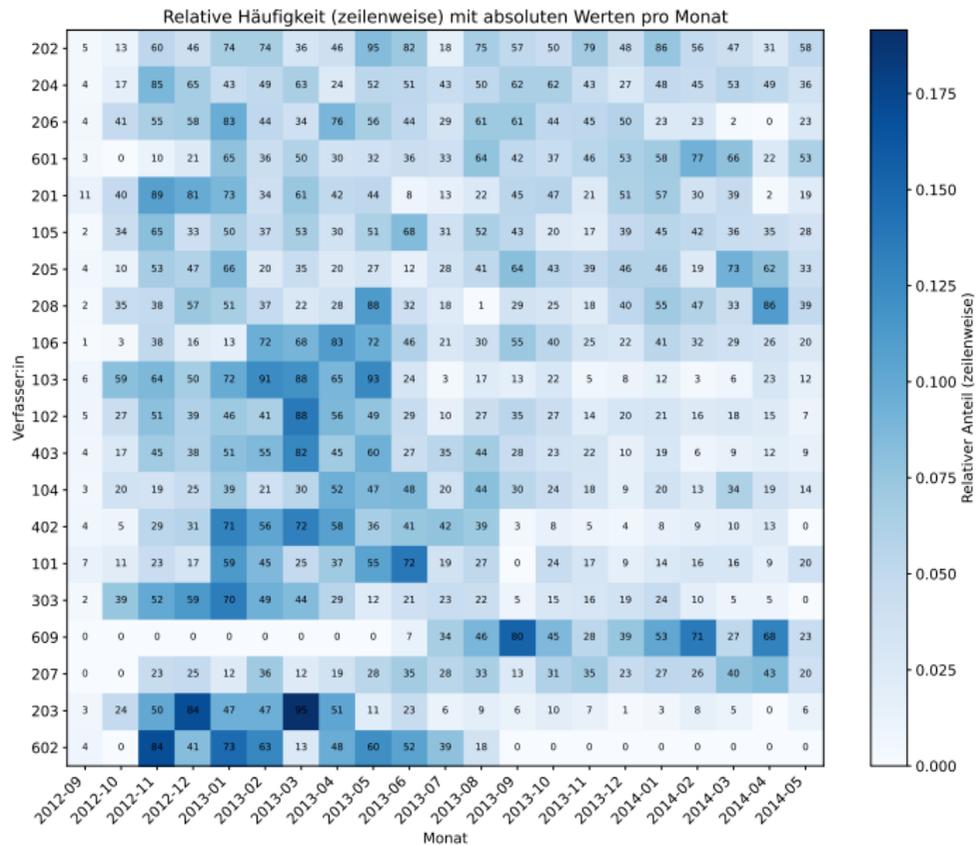
# Verfasser + Zeit

Können wir uns das für die Verfasser:innen anschauen?

## Nachricht

Bitte zeichne mir ein solches Diagramm für die Spalte Verfasser:in.







# Textanalyse

## Häufigkeitsverteilung wichtiger Worte

### Nachricht

Bitte erstelle für jeden Träger eine Wortwolke.

- ▶ Gleiches könnte für Bereiche, Klienten, Verfasser:Innen etc. durchgeführt werden.



Wortwolke für Träger: Träger 1



Wortwolke für Träger: Träger 2



Wortwolke für Träger: Träger 3



Wortwolke für Träger: Träger 4



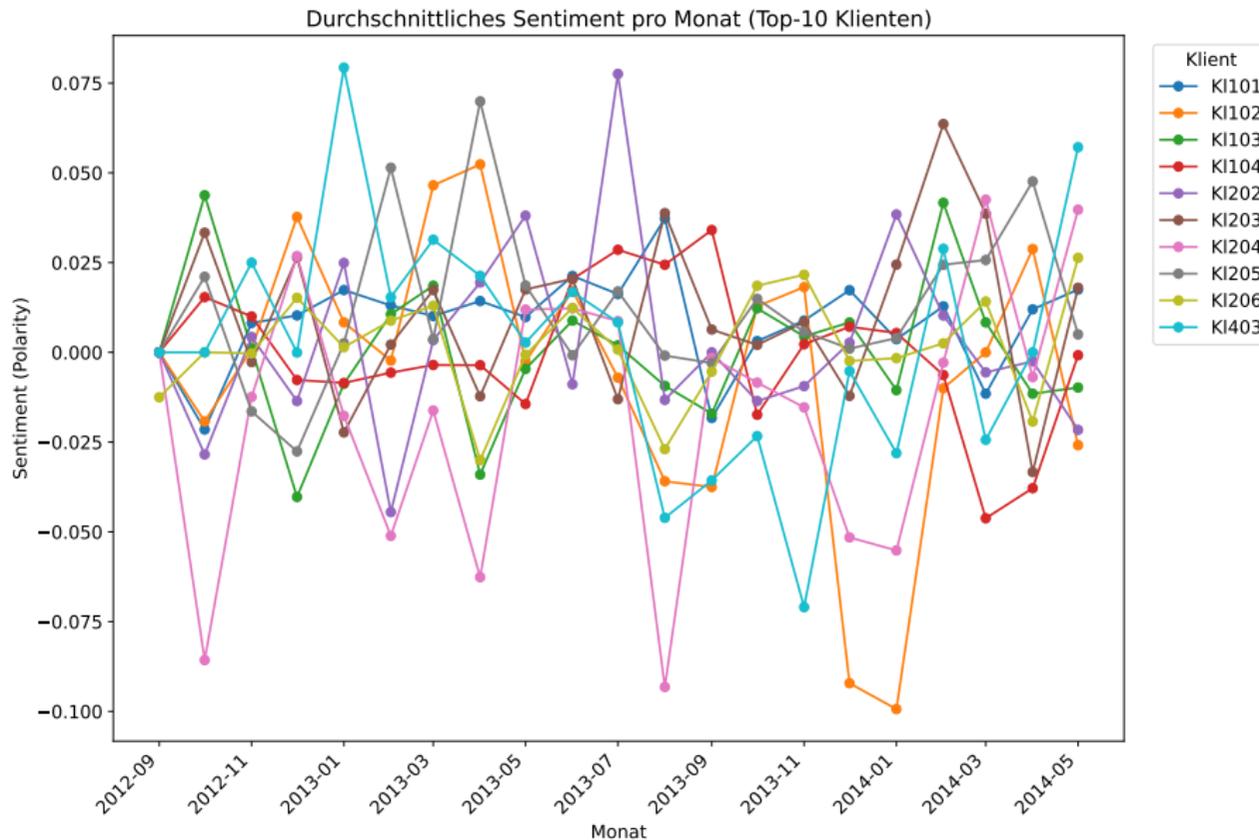
Wortwolke für Träger: Träger 5

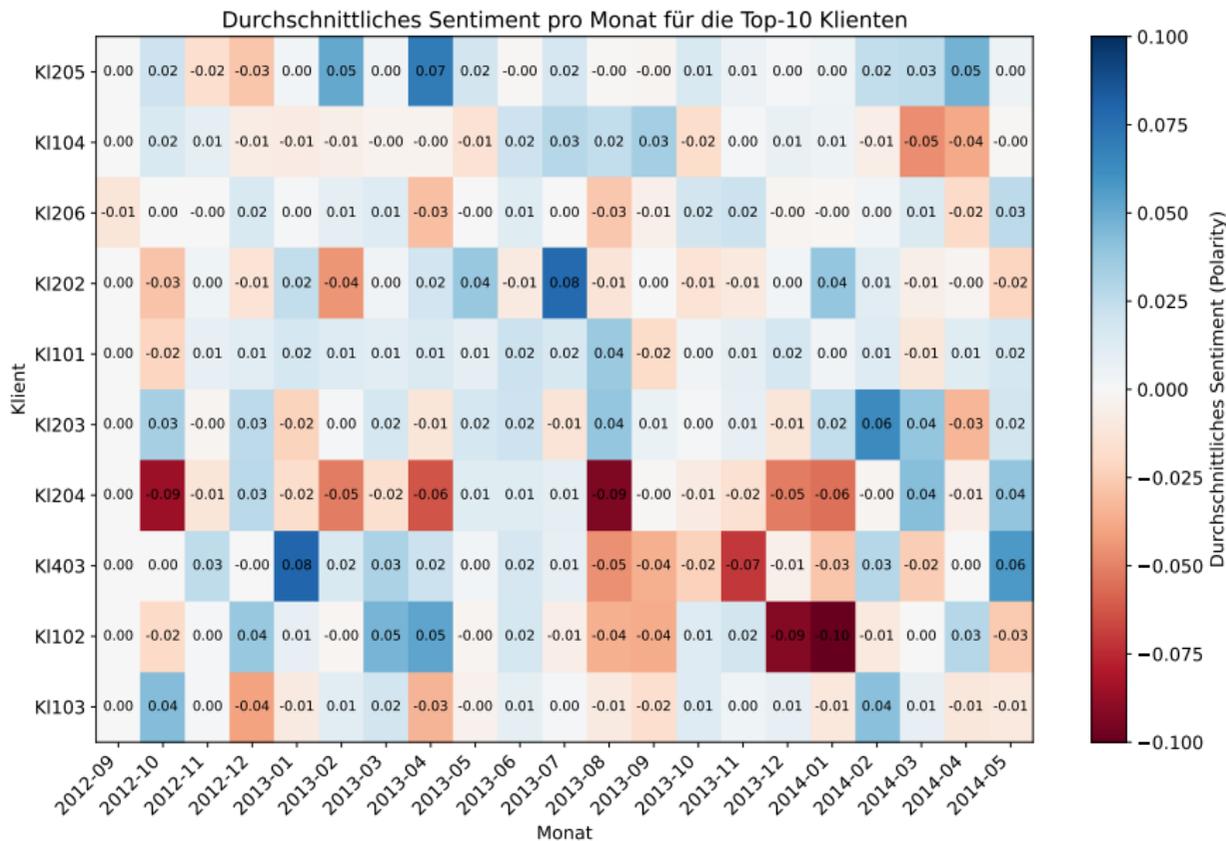


Wortwolke für Träger: Träger 6











# Interaktives Dashboard







Fazit







# References I

- [1] OpenAI, “Gpt-4 technical report,” *ArXiv*, vol. abs/2303.08774, 2023.