

ADT Portfolio Teil 4

Visualisierung des Lobbyregisters



Noah Raupold (5022097),
David Gläsle (5022114)

Eingereicht am: 12. Dezember 2025

Inhaltsverzeichnis

1 Einleitung: Vom Datengrab zum Dashboard	1
2 Visualisierungsstrategie mit Grafana	2
2.1 Architektur der Dashboards	2
2.2 Technische Herausforderungen	7
3 Implementierung komplexer Analysen	8
3.1 Network of Influence: Die unsichtbaren Verbindungen	8
3.2 Finanzielle Rankings: Window Functions im Einsatz	9
4 Data Stories: Erkenntnisse aus den Daten	11
4.1 Die „Drehtür“-Ministerien	11
4.2 Die teuersten Themenfelder	11
4.3 Der Gesetzgebungs-Fußabdruck	13
4.4 Effizienz-Analyse: Wer hat den größten Hebel?	13
4.5 Lobbyregister Fun Facts	14
5 Geografische Deep-Dives	15
5.1 Globale Akteure: Lobbying aus dem Ausland	15
5.2 Die PLZ-Heatmap	16
5.3 München vs. Berlin: Wo sitzt das Geld?	18
6 Transparenz & Compliance	20
6.1 Die Verweigerer	20
6.2 Heuristische Parteinähe-Erkennung	20
6.3 Strukturanalyse: Rechtsformen und ihre Agenda	21
7 Gesamtfazit	23

1 Einleitung: Vom Datengrab zum Dashboard

Nach dem Aufbau der ETL-Strecke und der optimierten Datenbankarchitektur (Indexing, Materialized Views) richtet sich der Abschluss dieses Portfolios auf die Visualisierungsschicht. Tausende Tabellenzeilen sind für Entscheidungen wertlos, solange sie nicht verdichtet, in Kontext gesetzt und visuell lesbar gemacht werden.

Ziel dieses Teils ist es, die technische Tiefe der PostgreSQL-Datenbank (Window Functions, CTEs¹, rekursive Abfragen) in prägnante Dashboards zu übersetzen. Grafana passt hier, weil es sich reibungslos in den bestehenden Docker-Stack einfügt und SQL nicht hinter einer Oberfläche versteckt. Sichtbar werden sollen verdeckte Netzwerke („Wer kennt wen?“), finanzielle Ausreißer und thematische Schwerpunkte im deutschen Lobbyregister.

Die Visualisierung steht damit am Ende einer Kette von Entscheidungen, die in den vorangegangenen Portfolios getroffen wurden: Datenmodellierung nach strengem Normalisierungsgrad, robuste ETL-Pipeline mit `asyncio` und gezielte Performance-Tuning-Maßnahmen. Erst auf dieser Basis können Dashboards belastbare Aussagen treffen, ohne sich auf Voraggregationen oder manuelle Datenaufbereitung verlassen zu müssen.

Adressaten der Visualisierungen sind unterschiedliche Gruppen: politische Entscheidungsträger, die einen schnellen Überblick benötigen; Analystinnen, die explorative Drill-downs durchführen; und Öffentlichkeit beziehungsweise Medien, die Transparenz über Verflechtungen erwarten. Alle drei profitieren von reproduzierbaren Panels, die unmittelbar aus der Datenbank gespeist werden und so eine konsistente „Single Source of Truth“ wahren.

¹Common Table Expressions.

2 Visualisierungsstrategie mit Grafana

Grafana bietet – anders als klassische BI-Tools wie Tableau oder PowerBI – einen direkten, code-basierten Zugriff auf die Datenbank. So bleibt die volle SQL-Funktionalität erhalten, statt auf generische Drag-and-Drop-Aggregationen reduziert zu werden. Der Grafana-Container ist in den bestehenden Docker-Stack eingebunden und greift über eine dedizierte PostgreSQL-Datasource auf die Materialized Views und Indizes zu, die in den vorherigen Phasen erarbeitet wurden.

Im Fokus stehen reproduzierbare Panels mit klar definierten SQL-Queries, die ohne proprietäre Berechnungen auskommen. Durch diesen Ansatz bleibt das gesamte Datenmodell transparent; jede Kennzahl lässt sich auf eine Query zurückführen, die versioniert und nachvollziehbar dokumentiert ist.

2.1 Architektur der Dashboards

Unsere Visualisierungsstrategie folgt dem „Schneidenbohrer-Prinzip“ (Drill-Down):

1. **High-Level Overview:** Der Einstieg erfolgt über ein globales Dashboard, das die wichtigsten KPIs aggregiert. Entscheider sehen dort sofort die Gesamtzahl der aktiven Lobbyisten, das kumulierte Finanzvolumen sowie die personelle Schlagkraft (FTE¹).

¹Full Time Equivalent.

2 Visualisierungsstrategie mit Grafana

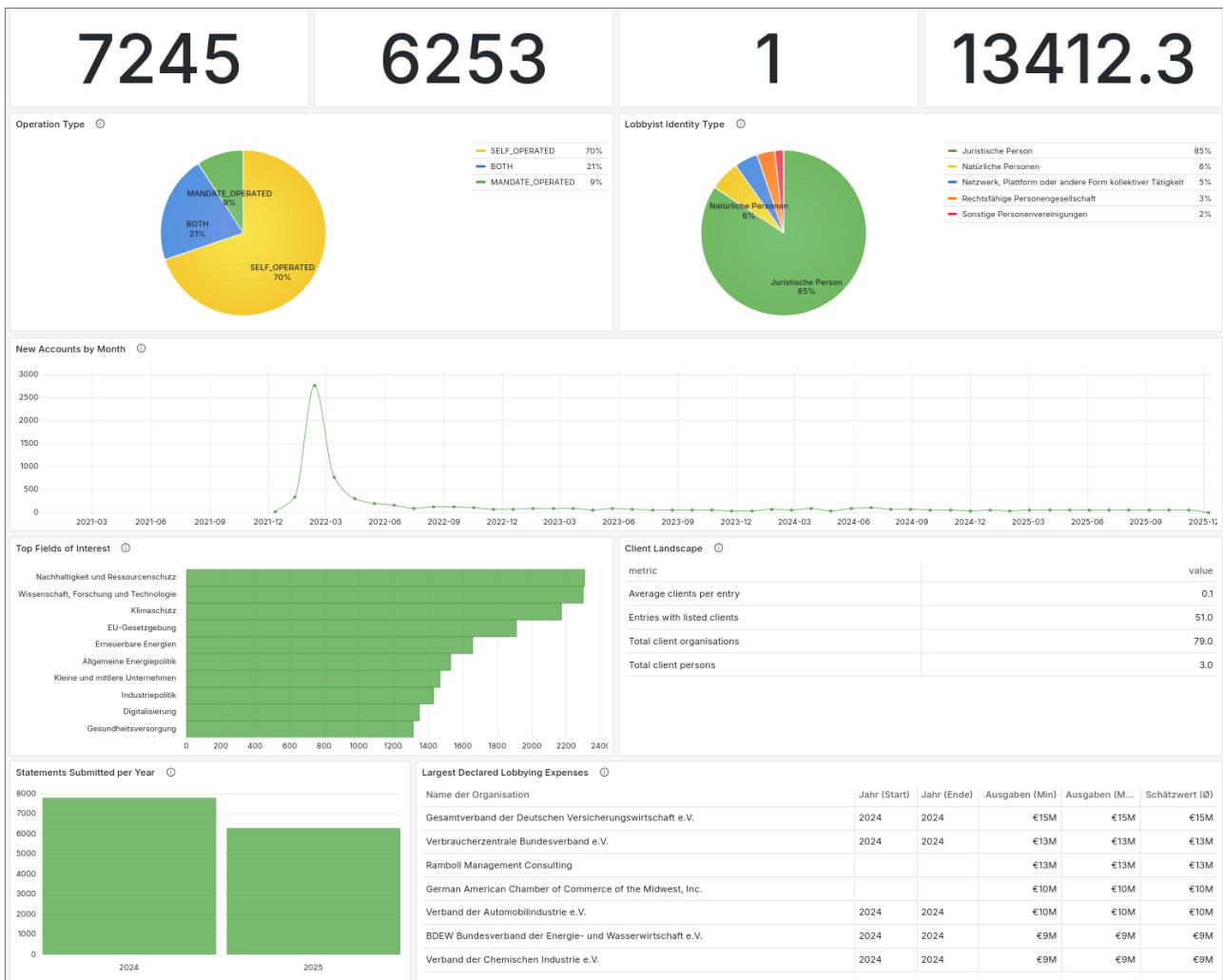


Abbildung 2.1: Das Overview-Dashboard: Zentrale KPIs, FTE-Analyse und Themenfelder auf einen Blick.

2. **Analytical Deep-Dive (Geo & Organisation):** Für detaillierte Analysen bieten spezialisierte Dashboards tiefe Einblicke. Das Geo-Dashboard visualisiert die Verteilung der Akteure auf einer Karte.

2 Visualisierungsstrategie mit Grafana

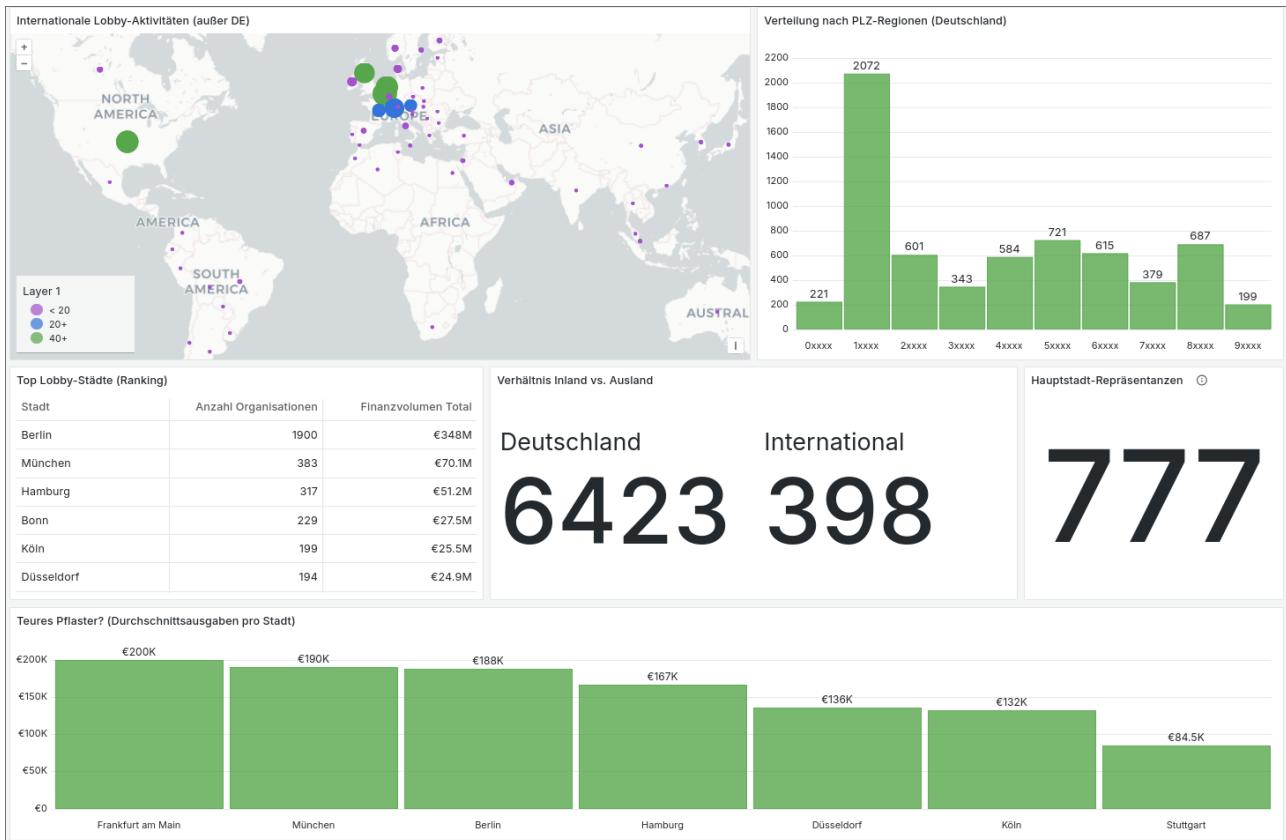


Abbildung 2.2: Geografische Verteilung der Lobby-Akteure (Inland & Ausland).

Besonders wertvoll ist der Drill-Down auf Städte-Ebene. Im Beispiel **Würzburg** (Abb. 2.3) lässt sich gezielt untersuchen, welche Akteure in einer spezifischen Region ansässig sind und wie hoch deren lokales Finanzvolumen ist.

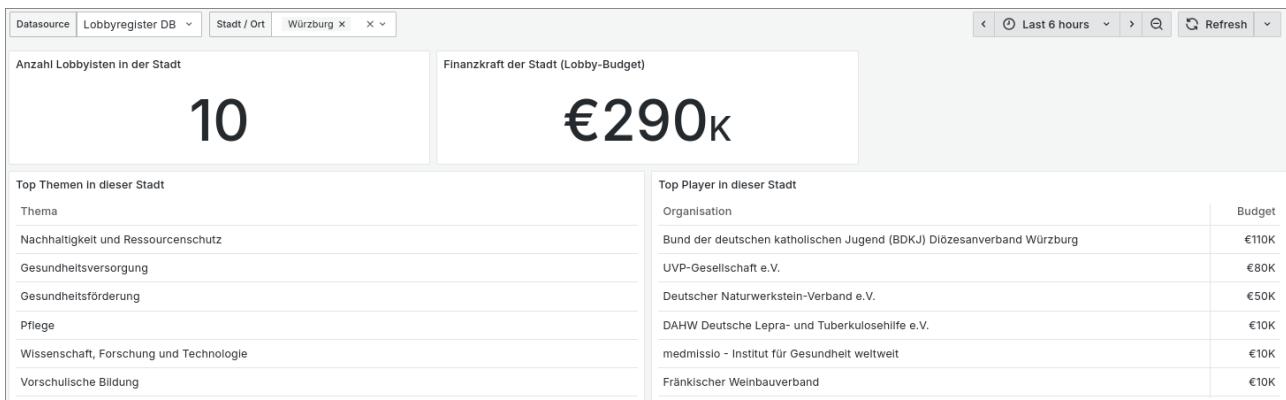


Abbildung 2.3: Detailansicht "City: Filterung am Beispiel der Stadt Würzburg."

Ebenso erlaubt der Organization-Profiler die Durchleuchtung einzelner Akteure. Am Beispiel der **Fraunhofer-Gesellschaft zur Förderung der angewandten Forschung e. V.** (Abb. 2.4) werden Mitarbeiterzahlen, Finanzhistorie, Mitgliedschaften und spezifische Gesetzesvorhaben transparent gemacht.

2 Visualisierungsstrategie mit Grafana

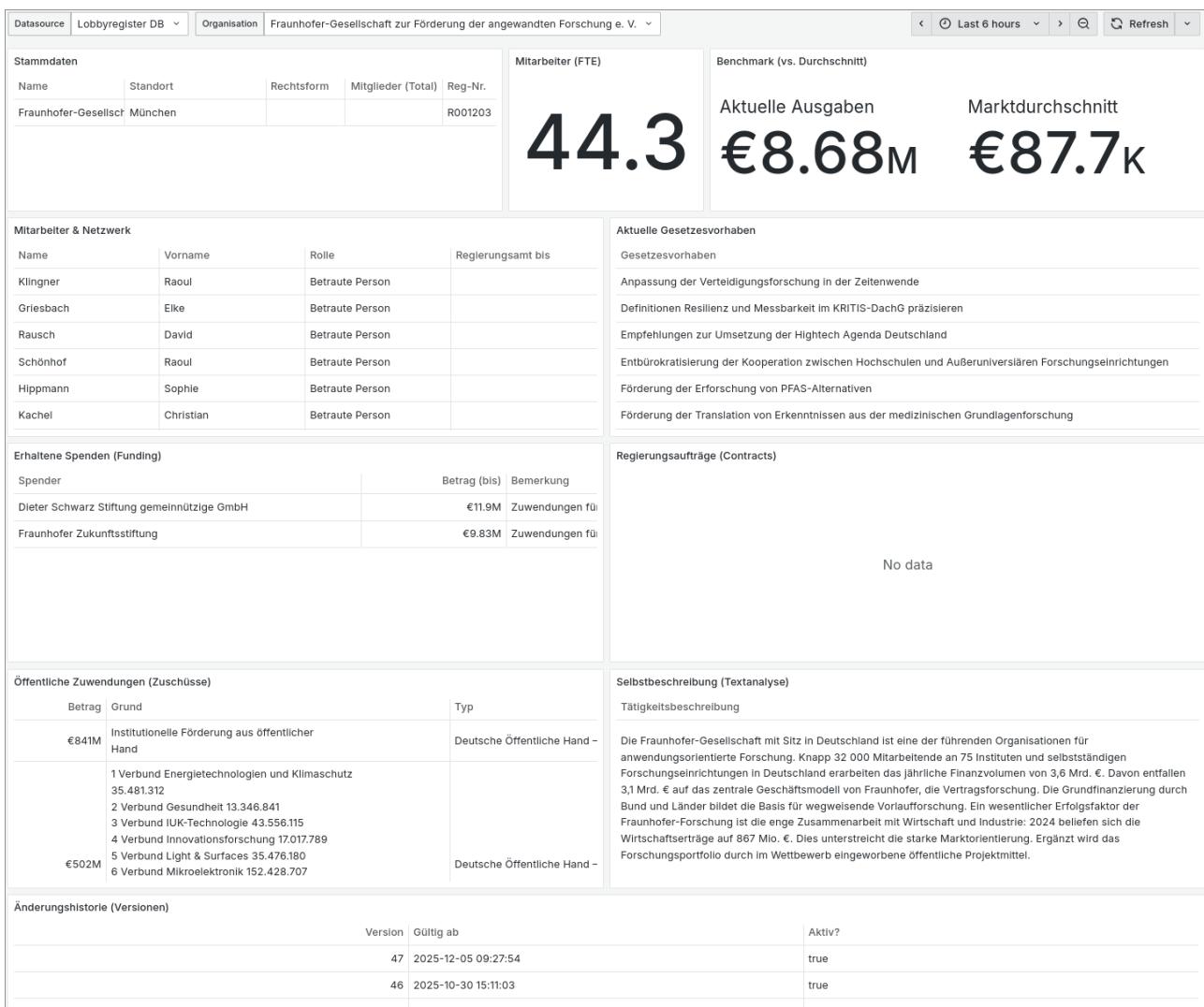


Abbildung 2.4: Organization-Profiler: 360°-Sicht am Beispiel der Fraunhofer-Gesellschaft.

3. Forensische Detailansicht & Netzwerke: Abschließend ermöglichen Spezial-Dashboards die Prüfung auf Compliance-Verstöße und Netzwerke. Das Transparenz-Dashboard listet Akteure, die Finanzangaben verweigern oder Jahresberichte schuldig bleiben.

2 Visualisierungsstrategie mit Grafana

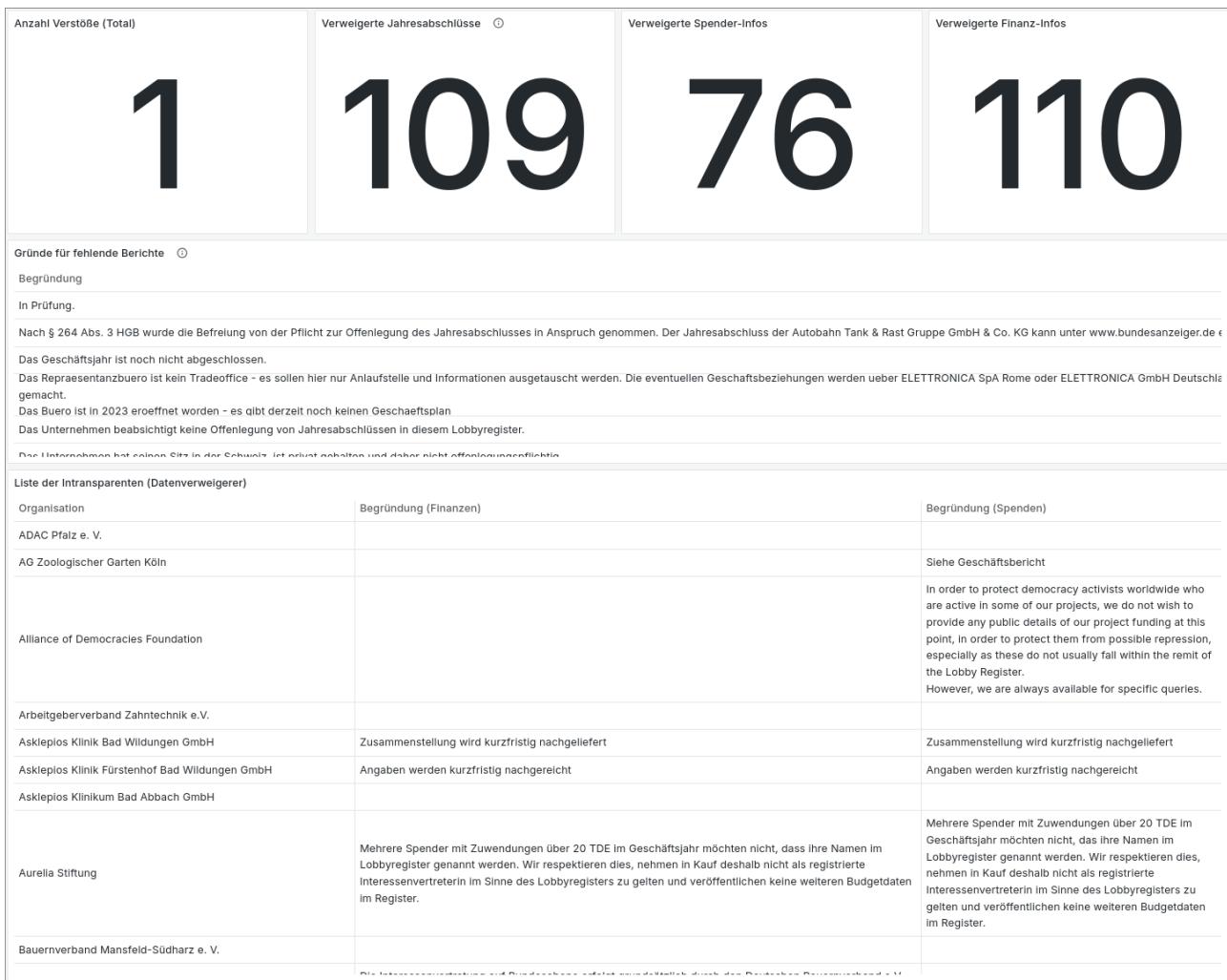


Abbildung 2.5: Transparenz-Dashboard: Identifikation von Akteuren mit verweigerten Angaben.

Das Advanced-Analytics Dashboard (Abb. 2.6) visualisiert komplexe Zusammenhänge wie legislative Fußabdrücke und Netzwerke.

2 Visualisierungsstrategie mit Grafana

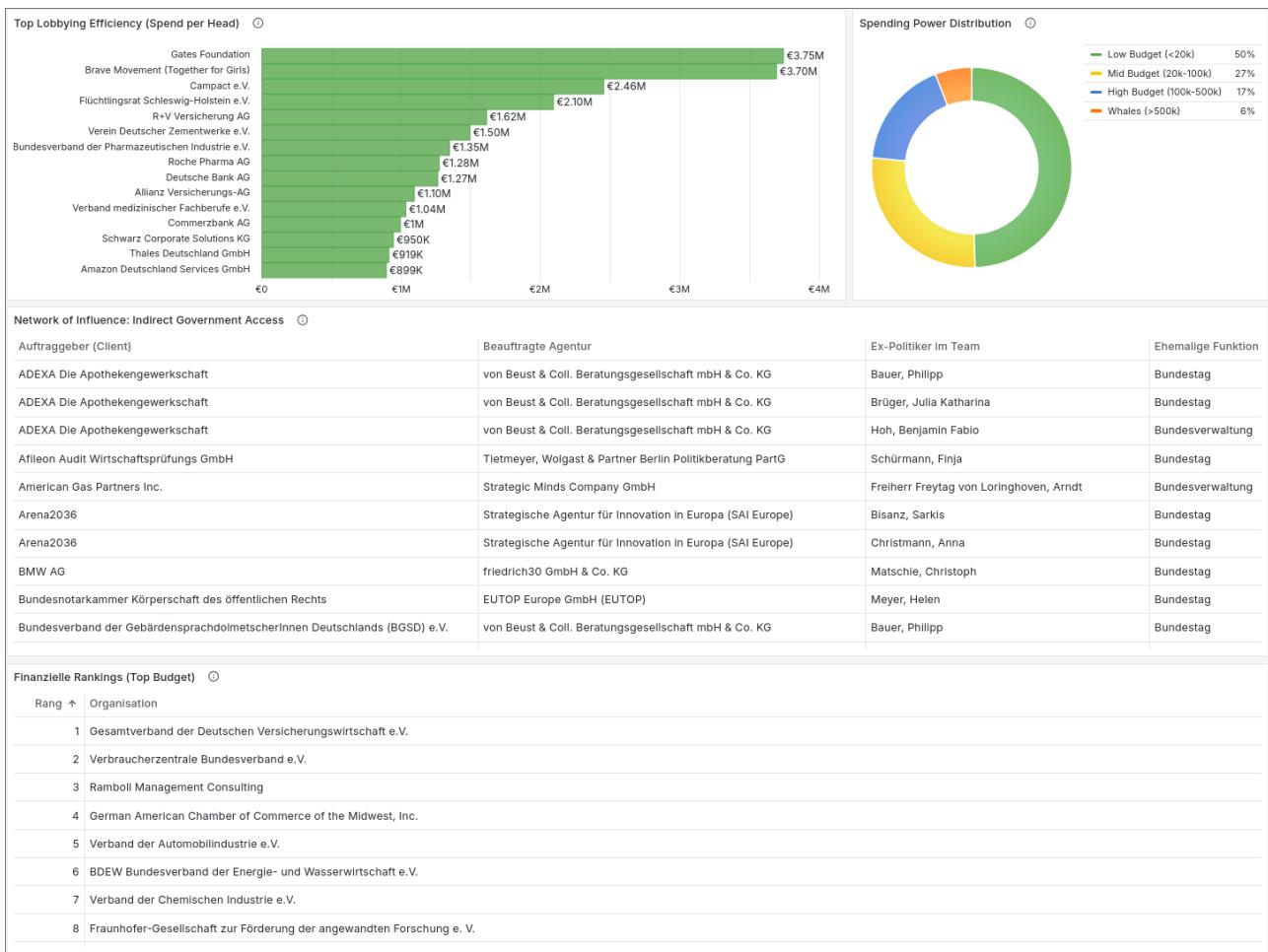


Abbildung 2.6: Advanced Analytics: Netzwerkanalysen und legislative Zusammenhänge.

2.2 Technische Herausforderungen

Knackpunkte waren die Visualisierung von Netzwerken und Zeitreihen, weil das Lobbyregister häufig nur Monats- und Jahresangaben liefert. Durch SQL-Casts (`TO_DATE`) und Fallback-Logiken haben wir dennoch stabile Zeitachsen für Grafana aufgebaut.

3 Implementierung komplexer Analysen

Im Mittelpunkt stehen nicht die Charts, sondern die SQL-Abfragen, die sie speisen. Grafana fungiert hier als Visualisierungsschicht, bleibt aber eng an der Datenbank: Alle Metriken entstehen aus versionierten Queries, die direkt auf den Materialized Views und Indizes aufsetzen. Damit sind die Analysen reproduzierbar und können sowohl im Dashboard als auch in isolierten SQL-Umgebungen ausgeführt werden.

Die folgenden Beispiele zeigen drei Abfrageklassen, die über einfache Aggregationen hinausgehen: mehrstufige Joins zur Rekonstruktion von Personen- und Organisationsnetzwerken, Window Functions für Rankings sowie Text- und Datumsnormalisierung, um unvollständige Eingaben auszuwerten. Sie illustrieren, wie wir Transparenz, Performanz und fachliche Aussagekraft miteinander verbinden.

3.1 Network of Influence: Die unsichtbaren Verbindungen

Transparenz scheitert oft an indirekten Ketten: Ein Unternehmen (Client) beauftragt eine Agentur, die wiederum einen ehemaligen Minister beschäftigt. Sichtbar wird diese Spur mit einem 5-Wege-Join.

Abbildung 3.1: SQL-Query für das 'Network of Influence' Panel

```

1  SELECT
2      cco.name AS "Auftraggeber (Client)",
3      li.name_text AS "Beauftragte Agentur",
4      -- Fallback: Name aus Personentabelle oder Organisationstabellle
5      COALESCE(ep.last_name || ' , ' || ep.first_name, li2.name_text)
6          AS "Ex-Politiker im Team",
7      cl.de AS "Ehemalige Funktion"
8  FROM contract_client_org cco
9  -- Der Weg des Geldes: Client -> Vertrag -> Agentur
10 JOIN contract_clients cc ON cco.clients_id = cc.id
11 JOIN contract_item c_item ON cc.contract_item_id = c_item.id
12 JOIN contracts c ON c_item.parent_id = c.id
13 JOIN register_entry re ON c.entry_id = re.id
14 JOIN lobbyist_identity li ON re.id = li.entry_id
15 -- Der personelle Link: Agentur -> Ex-Politiker
16 LEFT JOIN entrusted_person ep ON li.id = ep.identity_id
17 LEFT JOIN legalRepresentative lr ON li.id = lr.identity_id
18 JOIN recent_government_function rgf
19     ON (ep.recent_gov_function_id = rgf.id
20         OR lr.recent_gov_function_id = rgf.id)
21 LEFT JOIN code_label cl ON rgf.type_label_id = cl.id
22 WHERE rgf.id IS NOT NULL;

```

Ein konkreter Lauf der Abfrage zeigte beispielsweise:

- **Auftraggeber:** Arena2036
- **Beauftragte Agentur:** Strategische Agentur für Innovation in Europa (SAI Europe)
- **Verbindung:** Bundestag (via ehemaligem Funktionsträger)

Dies zeigt exemplarisch, wie Forschungscampus-Projekte (Arena2036) über spezialisierte Agenturen Zugang zu politischen Entscheidern suchen.

3.2 Finanzielle Rankings: Window Functions im Einsatz

Um die finanzkräftigsten Akteure zu identifizieren, setzen wir SQL Window Functions ein. RANK() erzeugt eine dynamische Bestenliste direkt in der Datenbank; ein Umweg über die Applikation entfällt.

Abbildung 3.2: Berechnung des Budget-Rankings mittels RANK()

```

1  SELECT
2      name,
3      expenses,
4      rank
5 FROM (
6     SELECT
7         COALESCE(li.name_text, li.company_name) AS name,
8         fe.expenses_to_eur AS expenses,
9         -- Ranking basierend auf Ausgaben
10        RANK() OVER (
11            ORDER BY fe.expenses_to_eur DESC NULLS LAST
12        ) AS rank
13    FROM financial_expenses fe
14    JOIN lobbyist_identity li ON fe.entry_id = li.entry_id
15 ) AS ranked_stats
16 WHERE rank <= 3;

```

Das Ergebnis der Analyse zeigt die "Heavy Hitter" des deutschen Lobbyismus (Stand: aktueller Datenabzug):

Rang	Organisation	Budget (EUR)
1	Gesamtverb. d. Dt. Versicherungswirtschaft	15.300.000
2	Verbraucherzentrale Bundesverband e.V.	12.740.000
3	Ramboll Management Consulting	12.620.000

Tabelle 3.1: Top 3 Lobbyisten nach gemeldeten Finanzaufwendungen.

4 Data Stories: Erkenntnisse aus den Daten

Neben der Implementierung haben wir die Datenbank explorativ („Data Mining“) genutzt, um Muster im deutschen Lobbyismus freizulegen. Die folgenden Abschnitte fassen ausgewählte Befunde zusammen.

4.1 Die „Drehtür“-Ministerien

Wir haben analysiert, welche Bundesministerien oder Behörden am häufigsten in den Lebensläufen von Lobbyisten auftauchen (Tabelle 4.1).

Kürzel	Ministerium / Behörde	Anzahl Treffer
BMVg	Bundesministerium der Verteidigung	6
BMU	Bundesministerium für Umwelt	3
BMAS	Bundesministerium für Arbeit	2
BMJ	Bundesministerium der Justiz	2
BMBFSFJ	Bundesministerium für Familie	1

Tabelle 4.1: Top 5 Herkunftsorte von Lobbyisten mit Regierungshintergrund. Das Verteidigungsministerium führt die Liste an.

Es ist auffällig, dass das Verteidigungsressort dominiert, gefolgt von Umwelt- und Arbeitsthemen.

4.2 Die teuersten Themenfelder

Welche politischen Themen sind den Lobbyisten das meiste Geld wert? Wir berechnen das Durchschnittsbudget aller Organisationen, die ein bestimmtes Interessenfeld angeben. Das gesamte Finanzvolumen im Register liegt bei **909 Millionen Euro**.

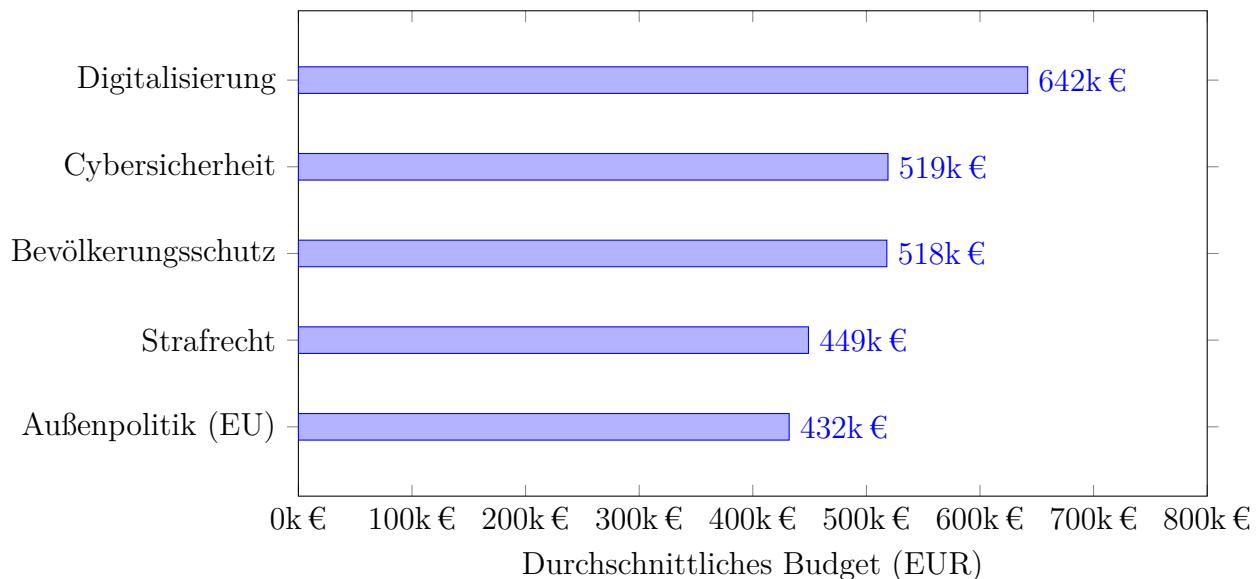


Abbildung 4.1: Top 5 Themenfelder mit dem höchsten Durchschnittsbudget pro Lobby-Akteur.

- **Platz 1: Digitalisierung ($\varnothing 641.667$ €)** – Das Megathema unserer Zeit zieht massives Kapital an.
- **Platz 2: Cybersicherheit ($\varnothing 519.129$ €)** – Angesichts wachsender Bedrohungen ist die digitale Verteidigung das teuerste Pflaster des Lobbyismus (495 Akteure).
- **Platz 3: Bevölkerungsschutz und Katastrophenhilfe ($\varnothing 517.532$ €)** – Ein Bereich, der oft staatliche Großaufträge impliziert und entsprechend finanzstarke Akteure anzieht (246 Akteure).

Klassische Industriethemen (Auto, Chemie) sind in Summe groß, werden im *Durchschnitt* aber durch viele kleine Zulieferer verwässert. Die oben genannten Felder werden dagegen fast ausschließlich von „Big Playern“ bespielt.

Die Daten zeigen, dass diese teuren Themenfelder von finanzstarken Verbänden dominiert werden. Oft streuen sie ihre hohen Budgets breit:

- **Gesamtverband der Deutschen Versicherungswirtschaft e.V.** (15,3 Mio. € Jahresbudget) treibt den Durchschnitt im Bereich Strafrecht nach oben, verteilt seine Ressourcen jedoch auf insgesamt **37 Themenfelder**.
- **Verbraucherzentrale Bundesverband e.V.** (12,7 Mio. €) ist federführend im Bereich Cybersicherheit, bespielt aber gleichzeitig **46 weitere Themen**.
- **Verband der Automobilindustrie e.V.** (9,9 Mio. €) ist in allen drei Top-Kategorien aktiv, streut sein Budget jedoch am breitesten über **60 verschiedene Interessenfelder**.

4.3 Der Gesetzgebungs-Fußabdruck

Das Lobbyregister erfasst nicht nur Interessenfelder, sondern auch konkrete Gesetzesvorhaben (*Regulatory Projects*). Damit lässt sich messen, welche Gesetze die meisten Lobbyisten mobilisieren.

Abbildung 4.2: Identifikation der umkämpftesten Gesetzesvorhaben

```

1 SELECT
2   rpi.title as "Gesetzesvorhaben",
3   COUNT(DISTINCT re.id) as "Lobbyisten-Anzahl"
4 FROM register_entry re
5 JOIN regulatory_projects rp ON re.id = rp.entry_id
6 JOIN regulatory_project_item rpi ON rp.id = rpi.parent_id
7 -- Filterung nach Relevanz (z.B. Digitalisierung)
8 JOIN activities_interests ai ON re.id = ai.entry_id
9 JOIN field_of_interest foi ON ai.id = foi.activities_id
10 JOIN code_label cl ON foi.label_id = cl.id
11 WHERE cl.de = 'Digitalisierung'
12 GROUP BY rpi.title
13 ORDER BY 2 DESC
14 LIMIT 5;

```

So werden politische Konfliktlinien sichtbar, bevor sie breit diskutiert werden. Ein Lauf der Abfrage für den Sektor „Digitalisierung“ zeigte überraschend: Nicht ein IT-Sicherheitsgesetz, sondern die „**Praxisorientierte Anpassung des Energieeffizienzgesetzes**“ (10 betroffene Lobbyisten) führte. Das unterstreicht die Verzahnung von Digitalisierung und Energiepolitik. Auch die **EU-Plastikabgabe** und die **Digitalisierbarkeit steuerlicher Prozesse** (je 9 Lobbyisten) stehen weit oben.

4.4 Effizienz-Analyse: Wer hat den größten Hebel?

Die Metrik „Ausgaben pro Mitarbeiter“ (Spending per Head) brachte unerwartete Spitzenreiter hervor. Großkonzerne wie *Deutsche Bank* oder *Allianz* haben erwartbar hohe Budgets, doch spezialisierte Kammern und Tech-Konzerne führen das Effizienz-Ranking. Die **German American Chamber of Commerce** (147.000 € pro Kopf) und **Meta Platforms** (129.000 € pro Kopf) investieren am meisten pro registriertem Lobbyisten – ein Hinweis auf wenige, aber hochspezialisierte Experten mit großen Budgets.

Zum Vergleich: Etablierte Großkonzerne agieren oft mit anderen Relationen. Die **Deutsche Bank AG** meldet beispielsweise ein Budget von 1,84 Mio. € bei nur 1,45 angegebenen Vollzeitäquivalenten (FTE), was auf den Einsatz hochbezahlter externer Berater oder weniger, aber extrem teurer Spitzenlobbyisten hindeutet. Die **Allianz SE** operiert mit 980.000 € Budget bei 3,73 FTE etwas „konservativer“.

4.5 Lobbyregister Fun Facts

Zum Abschluss noch einige kuriose Statistiken, die wir direkt aus der Datenbank extrahiert haben:

- **Der Spendenkönig:** Bill Gates (Privatperson) führt die Liste der Spender mit 97,9 Mio. € unangefochten an.
- **Die Lobby-Hauptstädte:** Außerhalb von Berlin sitzen die meisten Lobbyisten in **München (383)**, gefolgt von Hamburg (318) und der ehemaligen Hauptstadt Bonn (230).
- **Die „Vielseitigsten“:** Das „Aktionsbündnis Für die Würde unserer Städte“ vertritt stolze 62 verschiedene Auftraggeber (Clients) – ein Rekordwert für Bündelung von Interessen.
- **Namensvettern:** Wer in Deutschland Lobbyist werden will, hat mit dem Vornamen **Andreas** (8 Treffer) statistisch die besten Chancen, dicht gefolgt von Peter und Michael.
- **Bestvernetzt:** Die *Deutsch-Taiwanische Gesellschaft e. V.* pflegt mit 12 direkten Verbindungen zu ehemaligen Regierungsmitgliedern das dichteste Kontaktnetzwerk im Datensatz.

5 Geografische Deep-Dives

Neben der inhaltlichen Dimension ("Wer spricht worüber?") ermöglicht das Datenmodell präzise geografische Analysen. Wir wollten wissen, wo die Lobbyisten tatsächlich sitzen – und ob "Berlin" wirklich der einzige Ort ist, an dem Entscheidungen beeinflusst werden.

5.1 Globale Akteure: Lobbying aus dem Ausland

Das Lobbyregistergesetz gilt nicht nur für deutsche Unternehmen. Viele internationale Akteure versuchen, Einfluss auf die Bundespolitik zu nehmen. Unsere "International-Query" filtert alle Einträge mit Hauptsitz außerhalb Deutschlands.

Abbildung 5.1: SQL-Query zur Analyse internationaler Herkunftsländer

```
1 SELECT
2   cl.code as "Land",
3   COUNT(DISTINCT re.id) as "Anzahl Akteure"
4 FROM register_entry re
5 JOIN lobbyist_identity li ON re.id = li.entry_id
6 JOIN address a ON li.address_id = a.id
7 JOIN country_label cl ON a.country_id = cl.id
8 WHERE cl.code != 'DE' -- Ausschluss von Deutschland
9 GROUP BY cl.code
10 ORDER BY 2 DESC
11 LIMIT 10;
```

Das Ergebnis: Neben erwartbaren Nachbarn (Belgien/Brüssel, Österreich) tauchen auch USA und Großbritannien weit oben auf. Lobbyismus in Berlin ist damit klar international. Spitzenspieler sind **Belgien (49 Akteure)** und die **USA (47)**, gefolgt von den Niederlanden (43), Großbritannien (41) und der Schweiz (38).

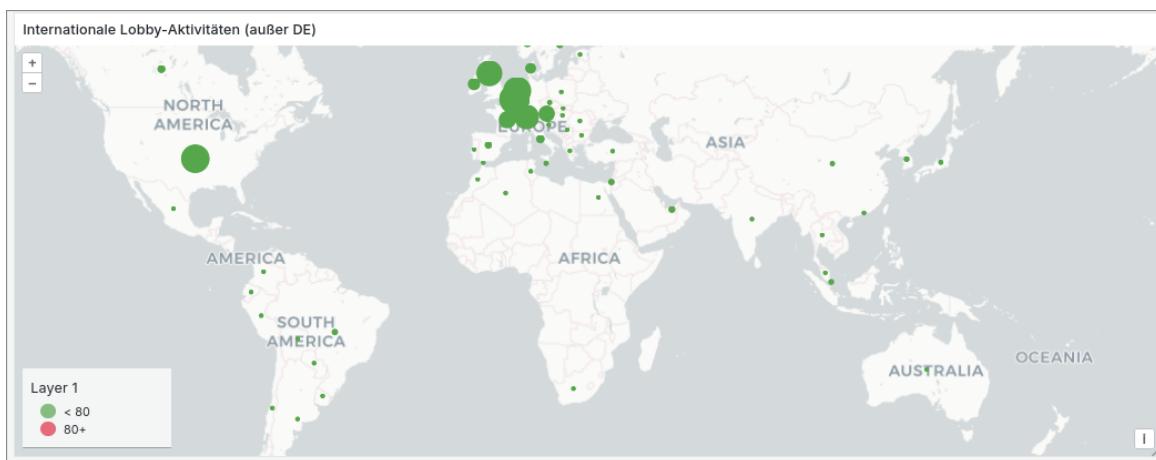


Abbildung 5.2: Geomap-Visualisierung der Herkunftsländer internationaler Lobbyisten (außer Deutschland).

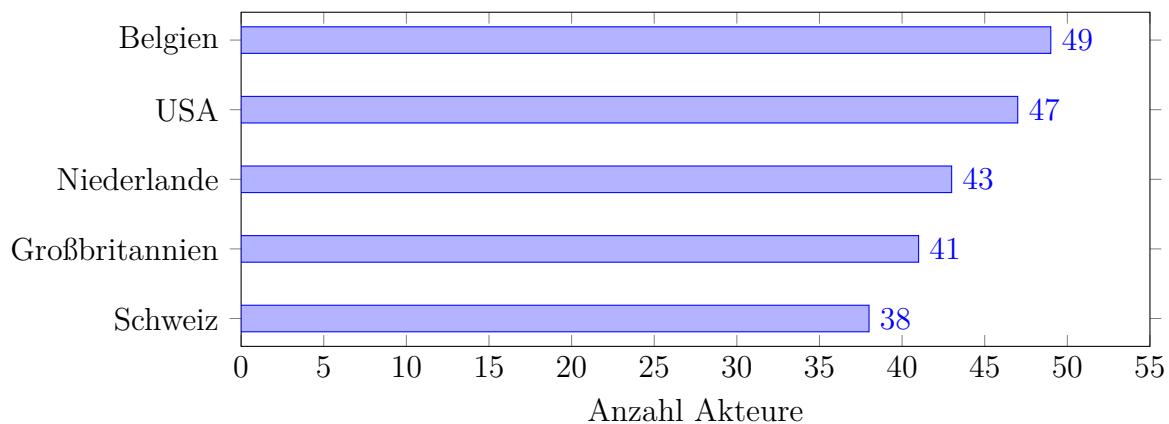


Abbildung 5.3: Top 5 Herkunftsländer internationaler Lobbyisten (ohne Deutschland).

5.2 Die PLZ-Heatmap

Um eine geografische Verteilung innerhalb Deutschlands zu erhalten, ohne auf komplexe GIS-Systeme¹ zurückgreifen zu müssen, haben wir eine Analyse auf Basis der ersten Ziffer der Postleitzahl durchgeführt. Dies liefert eine robuste „Heatmap“ der Lobby-Dichte.

¹Geographic Information Systems.

Abbildung 5.4: Verteilung nach Postleitzahl-Regionen

```

1 SELECT
2   SUBSTRING(a.zip_code, 1, 1) || 'xxxx' as "PLZ-Region",
3   COUNT(DISTINCT re.id) as "Anzahl Organisationen"
4 FROM register_entry re
5 JOIN lobbyist_identity li ON re.id = li.entry_id
6 JOIN address a ON li.address_id = a.id
7 JOIN country_label cl ON a.country_id = cl.id
8 WHERE cl.code = 'DE' AND a.zip_code IS NOT NULL
9 GROUP BY SUBSTRING(a.zip_code, 1, 1)
10 ORDER BY 1 ASC;

```

Die Ergebnisse zeigen erwartungsgemäß eine massive Konzentration im Bereich „1xxxx“ (Berlin/Brandenburg) mit **2.072 Organisationen**. Spannend sind die Cluster außerhalb der Hauptstadt:

- **Region 5xxxx (Köln/Bonn):** Mit 721 Organisationen immer noch ein Schwergewicht – ein Erbe der Bonner Republik.
- **Region 8xxxx (München/Bayern):** 687 Akteure, was die wirtschaftliche Stärke des Südens widerspiegelt.
- **Region 6xxxx (Frankfurt/Rhein-Main):** 615 Akteure, getrieben durch den Finanzsektor.

Im Gegensatz dazu ist der Osten (PLZ 0xxxx) mit nur 221 Einträgen (außerhalb Berlins) lobbyistisch kaum repräsentiert.

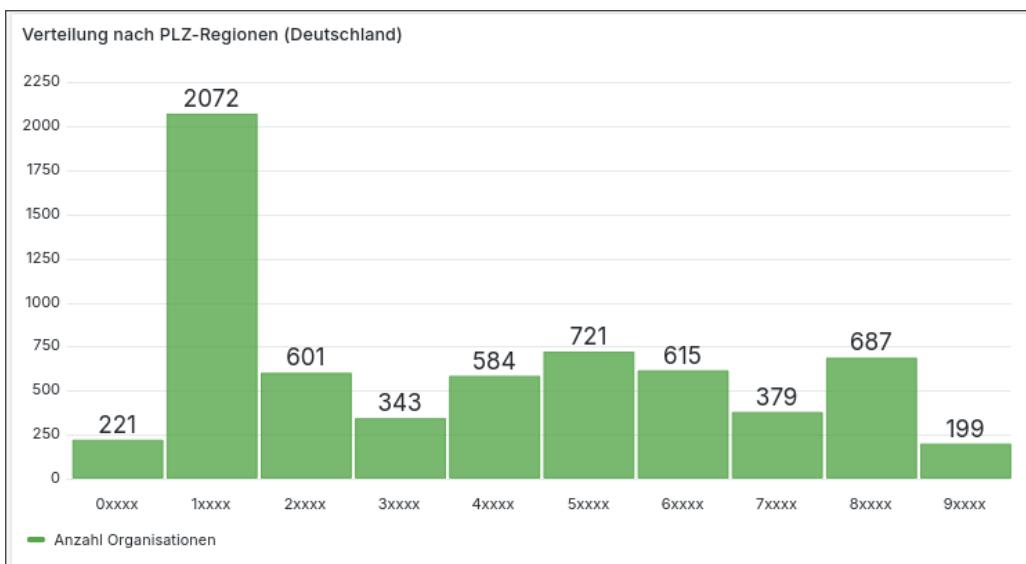


Abbildung 5.5: Verteilung der Lobby-Akteure nach Postleitzahl-Regionen (Grafana).

5.3 München vs. Berlin: Wo sitzt das Geld?

Ein besonders spannendes "Data Storytelling Element ist der Vergleich der durchschnittlichen Finanzausstattung pro Stadt. Berlin hat die meisten Büros, doch das "große Geld" sitzt oft in den Konzernzentralen.

Wir haben dazu die Tabelle `financial_expenses` mit der Adress-Tabelle `address` verknüpft und nach Städten gruppiert.

Abbildung 5.6: Vergleich der durchschnittlichen Budgets nach Städten

```

1 SELECT
2   a.city as "Stadt",
3   ROUND(AVG(fe.expenses_to_eur), 0) as "Avg Budget (EUR)"
4 FROM register_entry re
5 JOIN lobbyist_identity li ON re.id = li.entry_id
6 JOIN address a ON li.address_id = a.id
7 JOIN financial_expenses fe ON re.id = fe.entry_id
8 WHERE a.city IN ('Berlin', 'München', 'Hamburg', 'Frankfurt am Main')
9 GROUP BY a.city
10 ORDER BY 2 DESC;

```

Das Ergebnis überraschte uns: Nicht die Politik-Hauptstadt Berlin, sondern die Finanzmetropole **Frankfurt am Main** führt das Ranking mit einem Durchschnittsbudget von ca. **199.679 €**

an. Dicht darauf folgt **München** (189.133 €). Berlin landet mit ca. **183.790 €** nur auf dem dritten Platz, noch vor Hamburg (166.201 €). Dies zeigt deutlich, dass hohe Budgets vor allem im Finanz- und Industriesektor (Frankfurt/München) zu finden sind.

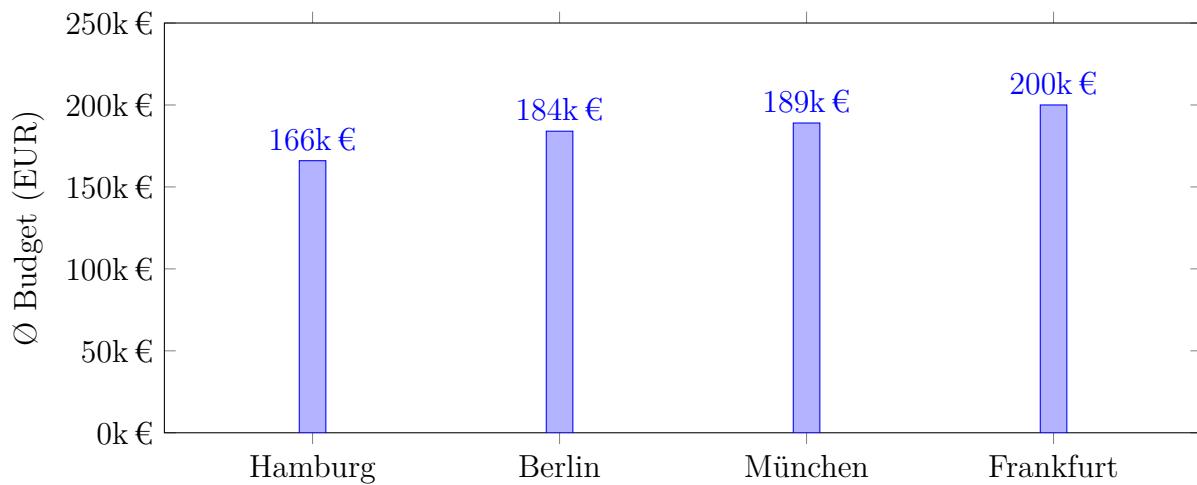


Abbildung 5.7: Vergleich der durchschnittlichen Lobby-Budgets deutscher Großstädte. Finanzmetropolen liegen vor der politischen Hauptstadt.

6 Transparenz & Compliance

Ein Register steht und fällt mit der Datenqualität. Ein Teil unseres Dashboards widmet sich daher der "Meta-Analyse: Wie gut halten sich die Akteure an die Regeln?

6.1 Die Verweigerer

Das Gesetz erlaubt in bestimmten Härtefällen, Finanzdaten zu verweigern (Feld `refuse_info`). Wir wollten wissen, wie oft von dieser Klausel Gebrauch gemacht wird.

Abbildung 6.1: Ermittlung von Transparenz-Lücken

```
1 SELECT
2     COUNT(*) as "Anzahl Verweigerer",
3     SUM(CASE WHEN refuse_reason IS NOT NULL THEN 1 ELSE 0 END)
4         as "Mit Begründung"
5 FROM financial_expenses
6 WHERE refuse_info = true;
```

Diese Query speist ein Alarm-Panel im Dashboard. Aktuell haben **113 Akteure** die Finanzangaben verweigert, wovon nur 33 eine explizite Begründung hinterlegt haben. Dies ist zwar eine kleine Minderheit, aber für Journalisten ("Watchdogs") eine extrem relevante Zielgruppe.

6.2 Heuristische Parteinähe-Erkennung

Das Datenmodell enthält kein explizites Feld für "Parteizugehörigkeit". Dennoch möchten Nutzer wissen, welchem Spektrum ein Ex-Politiker zuzuordnen ist. Eine Text-Mining-Lösung in SQL analysiert deshalb das Feld `function_position` (z.B. "Ehemaliger MdB (SPD)").

Abbildung 6.2: SQL-basierte Textklassifikation von Parteizugehörigkeiten

```

1  SELECT
2    CASE
3      WHEN function_position ILIKE '%CDU%'
4          OR function_position ILIKE '%CSU%', THEN 'Union'
5      WHEN function_position ILIKE '%SPD%', THEN 'SPD'
6      WHEN function_position ILIKE '%Grüne%', THEN 'Grüne'
7      WHEN function_position ILIKE '%FDP%', THEN 'FDP'
8      ELSE 'Sonstige / Unbekannt'
9    END as "Partei",
10   COUNT(*) as "Treffer"
11  FROM recent_gov_house_reps
12 WHERE function_position IS NOT NULL
13 GROUP BY 1
14 ORDER BY 2 DESC;

```

Dieser Ansatz demonstriert die Mächtigkeit von CASE WHEN Statements. Die Analyse ergab zwar eine Dominanz der Kategorie "Sonstige" (664 Treffer), was auf unstrukturierte Eingaben hindeutet. Unter den erkennbaren Parteien führte jedoch die **FDP** (9) vor Union (4) und SPD (4).

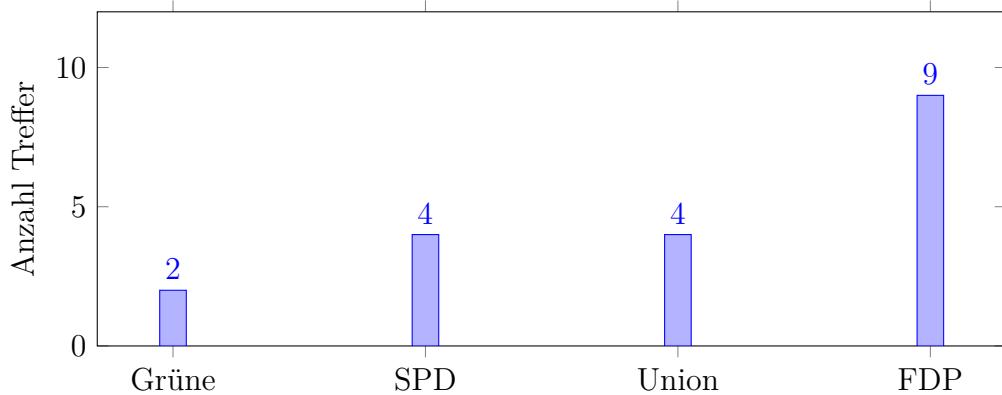


Abbildung 6.3: Erkannte Parteizugehörigkeiten bei Ex-Politikern (Text-Mining).

6.3 Strukturanalyse: Rechtsformen und ihre Agenda

Ein tieferer Blick in die Daten zeigt eine Arbeitsteilung zwischen den Rechtsformen. Untersucht wurde, welches Themenfeld bei welcher Rechtsform am häufigsten genannt wird und wie viel Budget durchschnittlich dahintersteht.

Die Analyse bestätigt das klassische Bild der Interessensvertretung:

- **Zivilgesellschaft (e.V. und Stiftungen):** Hier dominieren ökologische Themen. Sowohl bei eingetragenen Vereinen als auch bei Stiftungen steht „Nachhaltigkeit und Ressourcenschutz“ unangefochten auf Platz 1.
- **Wirtschaft (GmbH und AG):** Kapitalgesellschaften setzen andere Prioritäten. Ihr Top-Thema ist konsistent „Wissenschaft, Forschung und Technologie“.

Dies deutet darauf hin, dass Unternehmen Lobbying primär als Hebel für Innovationsförderung und Standortpolitik nutzen, während NGOs die normative Debatte um Nachhaltigkeit treiben.

Rechtsform	Top-Thema	Ø Budget (EUR)
Stiftung	Nachhaltigkeit und Ressourcenschutz	269.333
Eingetragener Verein (e.V.)	Nachhaltigkeit und Ressourcenschutz	262.830
GmbH	Wissenschaft, Forschung und Technologie	177.708
Aktiengesellschaft (AG)	Wissenschaft, Forschung und Technologie	94.242

Tabelle 6.1: Dominante Themenfelder je nach Rechtsform des Lobby-Akteurs.

7 Gesamtfazit

Dieses Portfolio markiert den Abschluss eines umfassenden Data-Engineering-Projekts. Das Ziel, die Strukturen des Lobbyregisters sichtbar zu machen, wurde zu einer konsistenten Analytics-Plattform weiterentwickelt. Die Arbeit lässt sich in vier Phasen gliedern, die jeweils prägende Erkenntnisse hervorbrachten.

Phase 1: Modellierung

Die Modellierung zeigte früh, wie komplex die Abbildung der Registerdaten ist. Die tief verschachtelten JSON-Strukturen der Bundestags-API erforderten ein streng normalisiertes Schema in der 3. Normalform mit über 80 Tabellen.

Zentrale Erkenntnis:

Eine hohe Normalisierung sichert Datenintegrität und vermeidet Redundanzen, verschiebt aber Komplexität in spätere Performance-Optimierungen. Jedes Array in der Quelle benötigt eine eigene 1:N-Struktur – ein Aufwand, der sich zugunsten der Datenqualität auszahlt.

Phase 2: ETL Prozess

In der Implementierungsphase stand Prozessstabilität im Mittelpunkt. Eine asynchrone Python-Pipeline entkoppelte Producer (API) und Consumer (Datenbank) und wurde komplett containerisiert.

Zentrale Erkenntnis:

Resiliente ETL-Strecken entstehen durch saubere Entkopplung und reproduzierbare Umgebungen. `asyncio`, Queues und Docker sorgten dafür, dass die Pipeline auch unter Last kontrolliert arbeitet und teamfähig bleibt.

Phase 3: Optimierung

Die anschließende Optimierung adressierte die unvermeidbare Latenz einer stark normalisierten Datenbank. Analysen mit vielen Joins wurden zu langsam, um sie direkt in Dashboards zu nutzen.

Zentrale Erkenntnis:

Performanz verlangt gezielte Redundanz. **Materialized Views** und spezialisierte **GIN-Indizes**¹ reduzierten Antwortzeiten von Sekunden auf Millisekunden und machten die Daten für interaktive Analysen nutzbar.

Phase 4: Visualisierung

Die Visualisierung überführte die technische Basis in auswertbare Dashboards. Auf dieser Grundlage ließen sich Muster wie der Fokus auf das „Energieeffizienzgesetz“ oder PLZ-basierte Cluster transparent machen.

Zentrale Erkenntnis:

SQL bleibt auch im BI-Kontext ein zentrales Werkzeug. Erst die Kombination aus belastbarer Datenbasis und klaren Visualisierungen ermöglicht fundierte Aussagen zu Netzwerken, Budgets und Themenfeldern.

Schlusswort

Das Modul „Advanced Database Techniques“ machte sichtbar, wie Modellierung, ETL, Optimierung und Visualisierung ineinander greifen. Das Ergebnis ist eine nachvollziehbare Plattform, die mit modernen IT-Methoden Transparenz im politischen Umfeld fördert. Sie übergibt nicht nur Quellcode, sondern ein belastbares Werkzeug zur Analyse des Lobbyregisters.

¹Generalized Inverted Index.

Literatur

- [1] Deutscher Bundestag. *Lobbyregister des Deutschen Bundestages*. <https://www.lobbyregister.bundestag.de/>. Datenquelle. 2025.
- [2] Noah Raupold und David Gläsle. *Lobbyregister Ingestor*. <https://github.com/dav354/adt>. Git-Repository. 2025.