

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?	14
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 und Lessons Learned	23

1 Einleitung: Vom Datengrab zum Dashboard

Nachdem in den vorangegangenen Projektphasen eine robuste Datenpipeline (ETL) und eine hochperformante Datenbankarchitektur (Indexing, Materialized Views) geschaffen wurden, widmet sich dieser letzte Portfolioteil der wohl wichtigsten Ebene: der *Visualisierung*. Daten, die nur als Millionen von Tabellenzeilen existieren, sind für den menschlichen Entscheider wertlos. Erst durch Aggregation, Kontextualisierung und grafische Aufbereitung werden aus abstrakten Bytes greifbare Informationen.

Ziel dieses Teils ist es, die technische Tiefe der PostgreSQL-Datenbank (Window Functions, CTEs, rekursive Abfragen) in intuitive Dashboards zu übersetzen. Dabei setzen wir auf Grafana als Visualisierungs-Layer, da es sich nahtlos in den bestehenden Docker-Stack integriert und SQL als First-Class-Citizen behandelt. Wir zeigen, wie wir versteckte Netzwerke („Wer kennt wen?“), finanzielle Ausreißer und inhaltliche Schwerpunkte des deutschen Lobbyismus sichtbar machen.

2 Visualisierungsstrategie mit Grafana

Die Wahl fiel auf Grafana, da es – im Gegensatz zu klassischen BI-Tools wie Tableau oder PowerBI – eine direkte, code-basierte Verbindung zur Datenbank erlaubt. Dies ermöglicht uns, die volle Mächtigkeit von SQL zu nutzen, anstatt uns auf Drag-and-Drop-Aggregationen zu beschränken.

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. Hier sehen Entscheider sofort die Gesamtzahl der aktiven Lobbyisten, das kumulierte Finanzvolumen sowie die personelle Schlagkraft (FTE).

2 Visualisierungsstrategie mit Grafana

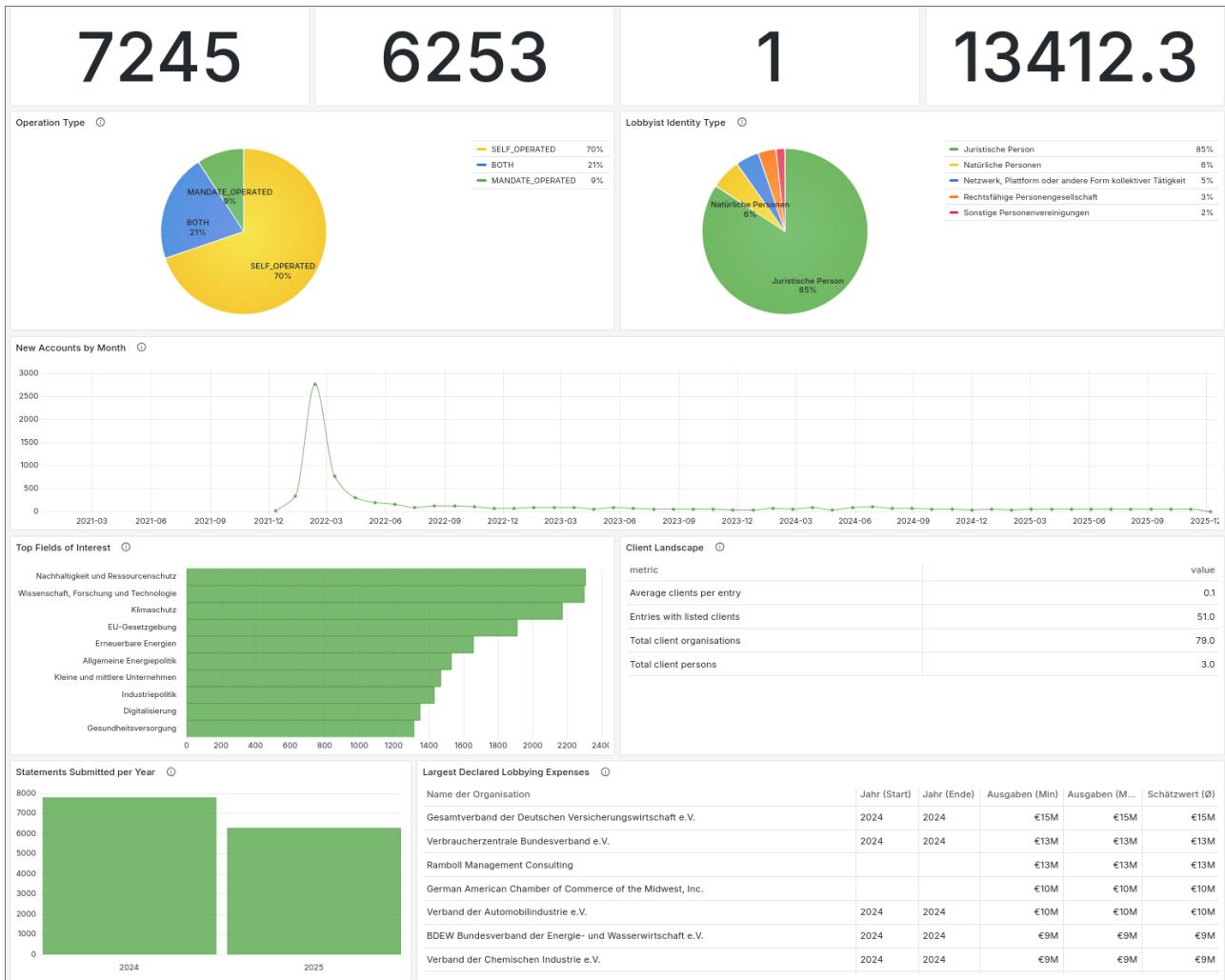


Abbildung 2.1: Das Overview-Dashboard: Zentrale KPIs, FTE-Analyse und Themenfelder auf einen Blick.

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

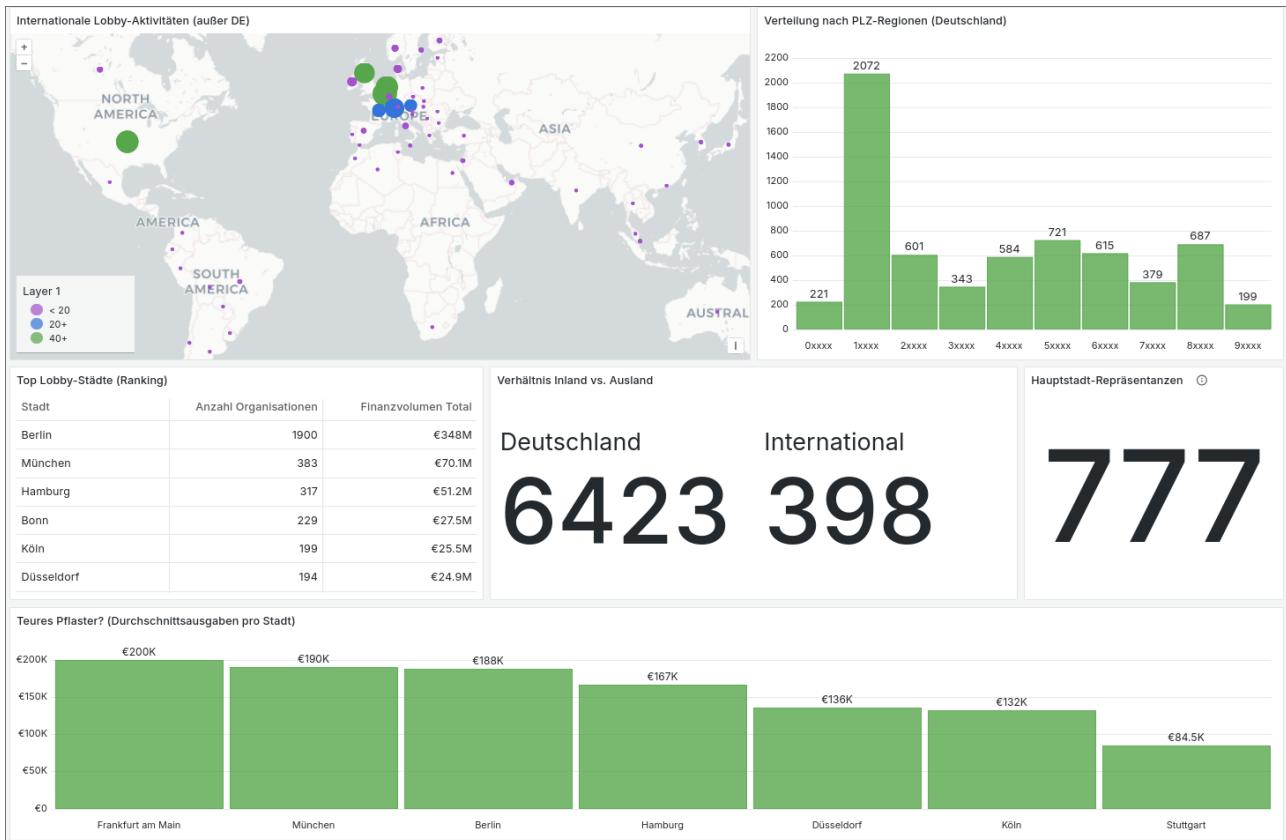


Abbildung 2.2: Geografische Verteilung der Lobby-Akteure (Inland & Ausland).

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.

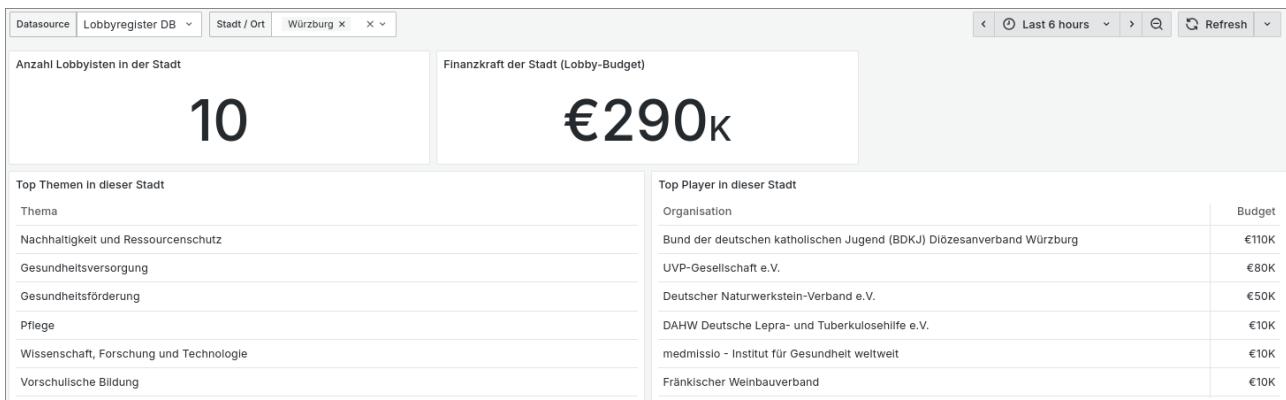


Abbildung 2.3: Detailansicht "City: Filterung am Beispiel der Stadt Würzburg.

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

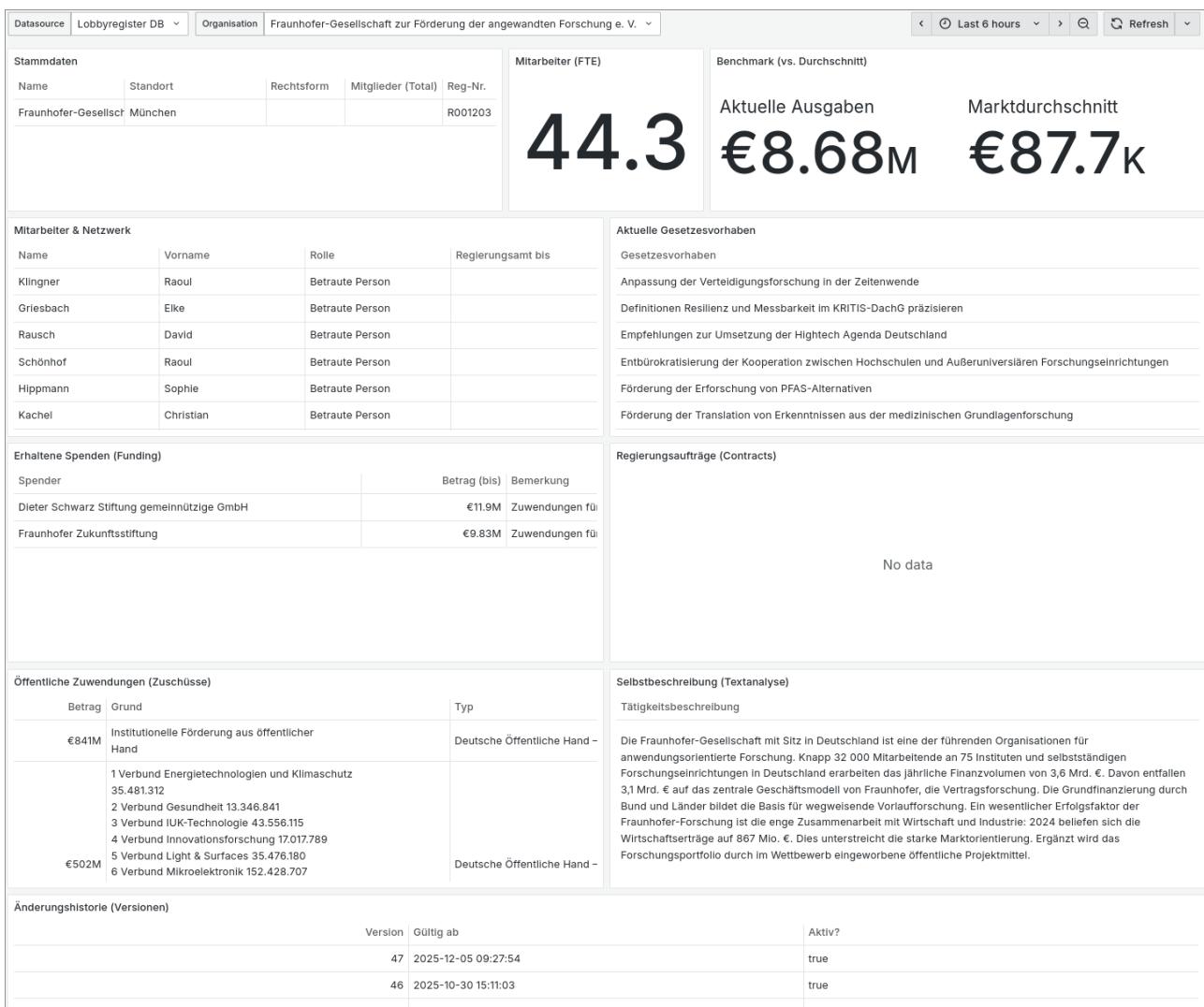


Abbildung 2.4: Organization-Profiler: 360°-Sicht am Beispiel der Fraunhofer-Gesellschaft.

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

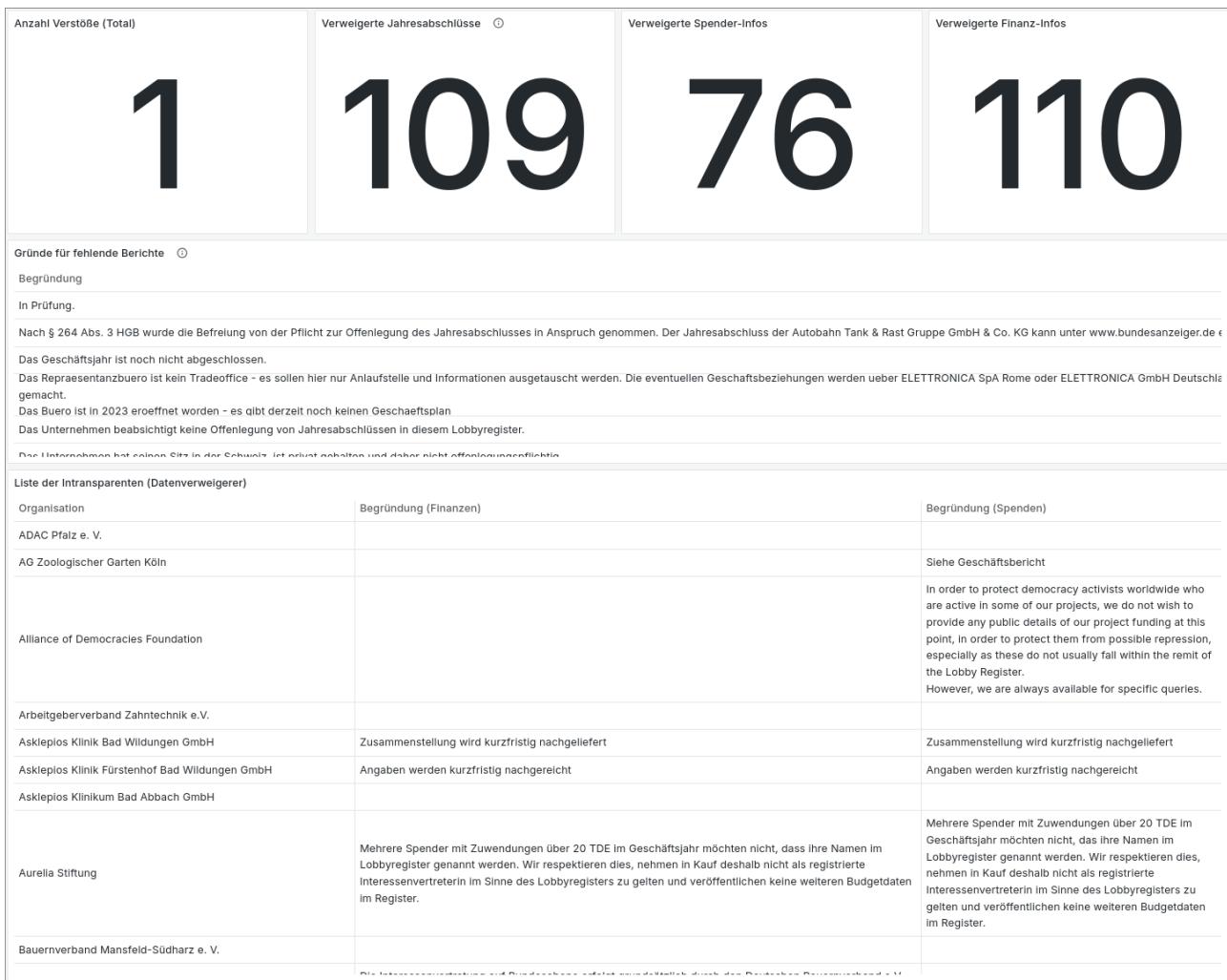


Abbildung 2.5: Transparenz-Dashboard: Identifikation von Akteuren mit verweigerten Angaben.

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

2 Visualisierungsstrategie mit Grafana

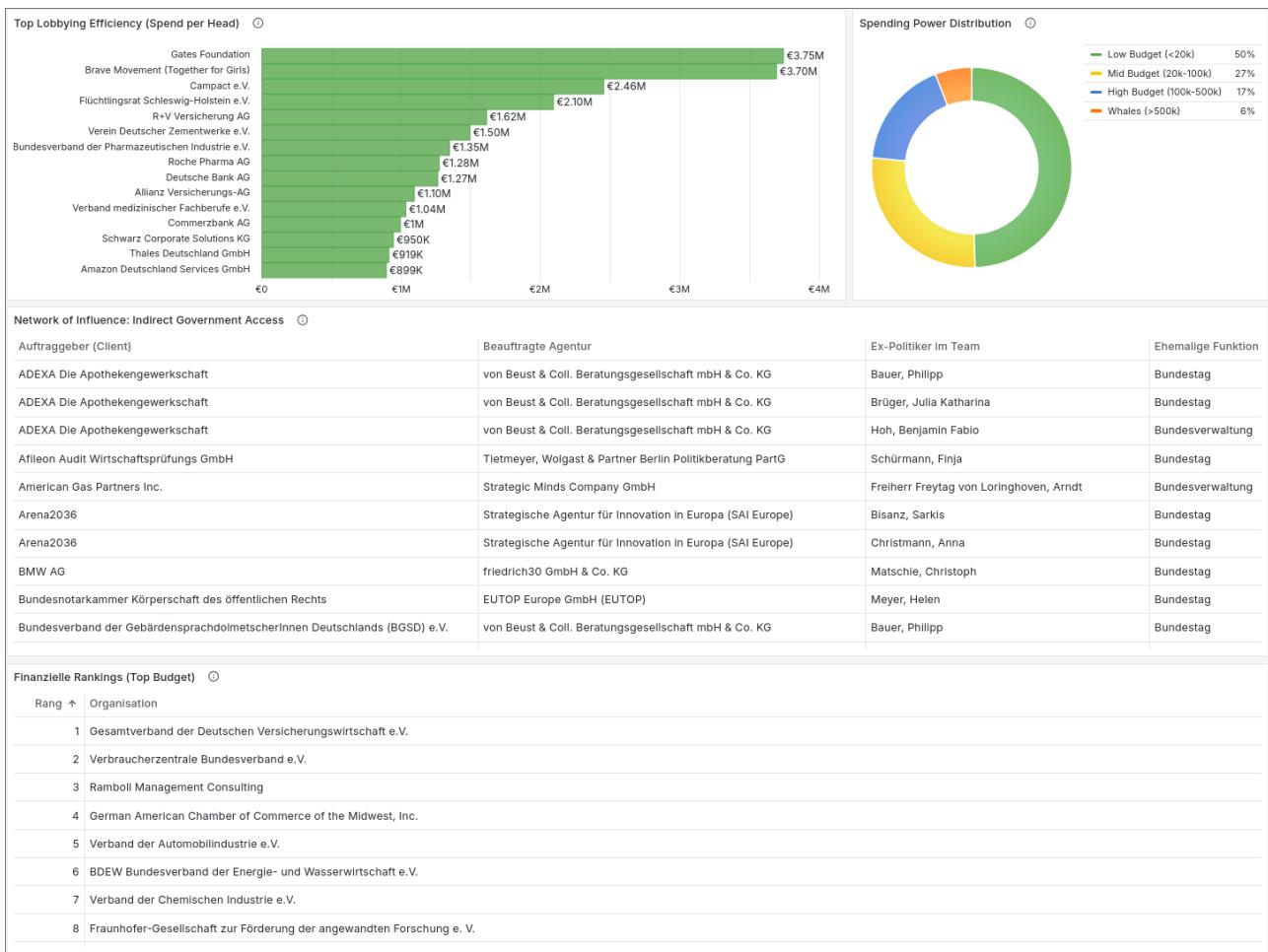


Abbildung 2.6: Advanced Analytics: Netzwerkanalysen und legislative Zusammenhänge.

2.2 Technische Herausforderungen

Eine besondere Herausforderung stellte die Visualisierung von Graphen (Netzwerken) und Zeitreihen dar, da das Lobbyregister oft unvollständige Zeitstempel (nur Monat/Jahr) liefert. Hier mussten wir mittels SQL-Casting (`TO_DATE`) und Fallback-Logiken robuste Zeitachsen für Grafana konstruieren.

3 Implementierung komplexer Analysen

Das Herzstück dieses Portfolios sind nicht die bunten Balken, sondern die SQL-Abfragen, die sie generieren. Wir präsentieren hier drei fortgeschrittene Analysen, die weit über einfaches SELECT * FROM hinausgehen.

3.1 Network of Influence: Die unsichtbaren Verbindungen

Ein zentrales Anliegen des Lobbyregisters ist Transparenz. Doch oft sind Verbindungen indirekt: Ein Unternehmen (Client) beauftragt eine Agentur, und diese Agentur beschäftigt einen ehemaligen Minister. Um diese Kette sichtbar zu machen, nutzen wir einen 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 Organisationstabellen
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 Testlauf dieser Abfrage förderte folgende interessante Verbindung zutage:

- **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 zu erkennen, welche Lobbyisten über die größten finanziellen Ressourcen verfügen, nutzen wir SQL Window Functions. Da der aktuelle Datensatz eine Momentaufnahme darstellt, verwenden

3 Implementierung komplexer Analysen

wir RANK(), um eine dynamische Bestenliste direkt in der Datenbank zu erzeugen, ohne die Daten erst in die Applikation laden zu müssen.

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

Jenseits der technischen Implementierung haben wir die Datenbank explorativ befragt („Data Mining“), um interessante Muster im deutschen Lobbyismus aufzudecken. Hier sind die Ergebnisse unserer „Fun Facts“-Analyse.

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 haben das Durchschnittsbudget aller Organisationen berechnet, die ein bestimmtes Interessenfeld angeben. Das

gesamte im Register erfasste Finanzvolumen beläuft sich auf beeindruckende **909 Millionen Euro**.

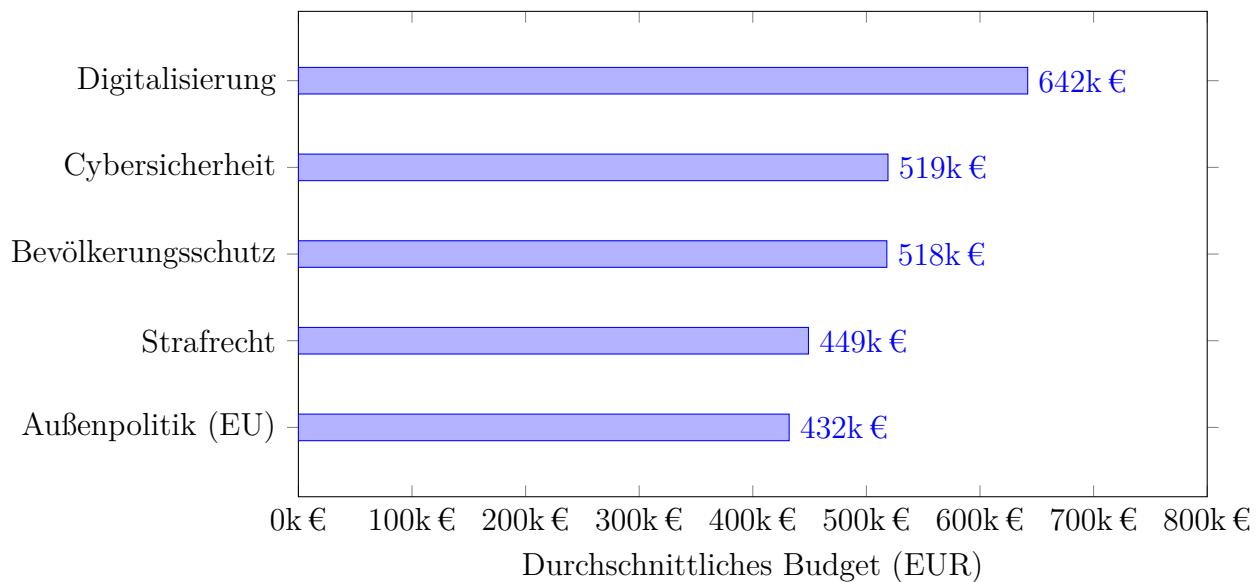


Abbildung 4.1: Top 5 Themenfelder mit dem höchsten Durchschnittbudget pro Lobby-Akteur.

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

Interessant ist hierbei, dass klassische Industriethemen (Auto, Chemie) zwar in der Summe riesig sind, aber durch viele kleine Zulieferer im *Durchschnitt* verwässert werden. Die oben genannten Themen werden hingegen fast ausschließlich von „Big Playern“ bespielt.

Konkret zeigen die Daten, dass diese teuren Themenfelder von den finanzstärksten Verbänden Deutschlands dominiert werden. Dabei dienen die enormen Budgets oft einer breiten Themenabdeckung:

- **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 abstrakte Interessenfelder, sondern auch konkrete Gesetzesvorhaben (*Regulatory Projects*), auf die Einfluss genommen werden soll. Dies ermöglicht uns, den „legislativen Fußabdruck“ zu messen: Welche Gesetze mobilisieren die meisten Lobbyisten?

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;

```

Diese Analyse erlaubt es, frühzeitig politische Konfliktlinien zu erkennen, noch bevor sie in den Medien breit diskutiert werden. Ein konkreter Lauf dieser Abfrage für den Sektor „Digitalisierung“ zeigte überraschende Ergebnisse: Auf Platz 1 landete nicht etwa ein klassisches IT-Sicherheitsgesetz, sondern die „**Praxisorientierte Anpassung des Energieeffizienzgesetzes**“ (10 betroffene Lobbyisten). Dies verdeutlicht, wie stark Digitalisierung und Energiepolitik mittlerweile verzahnt sind (Stichwort: Rechenzentren). Auch Themen wie die **EU-Plastikabgabe** oder die **Digitalisierbarkeit steuerlicher Prozesse** (je 9 Lobbyisten) stehen weit oben auf der Agenda.

4.4 Effizienz-Analyse: Wer hat den größten Hebel?

Ein überraschendes Finding lieferte die Metrik „Ausgaben pro Mitarbeiter“ (Spending per Head). Während Großkonzerne wie die *Deutsche Bank* oder *Allianz* erwartungsgemäß hohe Budgets haben, führen spezialisierte Kammern und Tech-Konzerne das Effizienz-Ranking an. Die **German American Chamber of Commerce** (147.000 € pro Kopf) und **Meta Platforms** (129.000 € pro Kopf) investieren am meisten pro registriertem Lobbyisten. Dies deutet auf eine Strategie hin, bei der hochspezialisierte Experten mit großen Budgets ausgestattet werden.

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)¹, 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.

¹FTE (Full Time Equivalent) bezeichnet das Vollzeitäquivalent. Es ist eine Kennzahl, die Teilzeitstellen in vergleichbare Vollzeitstellen umrechnet (z.B. zwei 50%-Stellen = 1,0 FTE), um die tatsächliche Personalkapazität vergleichbar zu machen.

5 Geografische Deep-Dives

Neben der inhaltlichen Dimension ("Wer spricht worüber?") bietet das Datenmodell auch exzellente Möglichkeiten für geografische Analysen. Wir haben untersucht, 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, deren Hauptsitz nicht in Deutschland liegt.

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 ist aufschlussreich: Neben den erwartbaren Nachbarn (Belgien/Brüssel, Österreich) tauchen auch die USA und Großbritannien weit oben auf. Dies verdeutlicht, dass Lobbyismus in Berlin ein globales Geschäft ist. Konkret führen **Belgien (49 Akteure)** und die **USA (47)** die Liste an, gefolgt von den Niederlanden (43), Großbritannien (41) und der Schweiz (38).

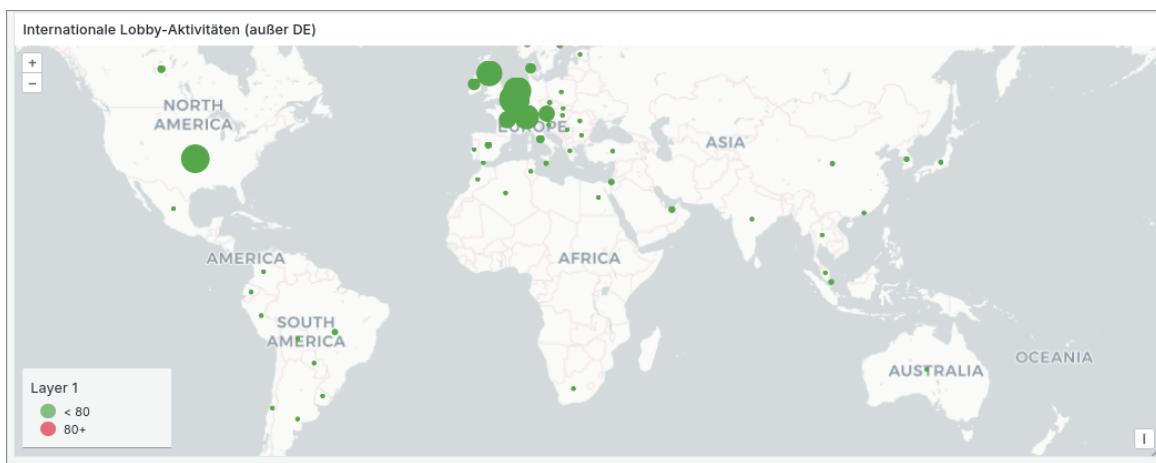


Abbildung 5.2: Geomap-Visualisierung der Herkunftsländer internationaler Lobbyisten (außer Deutschland).

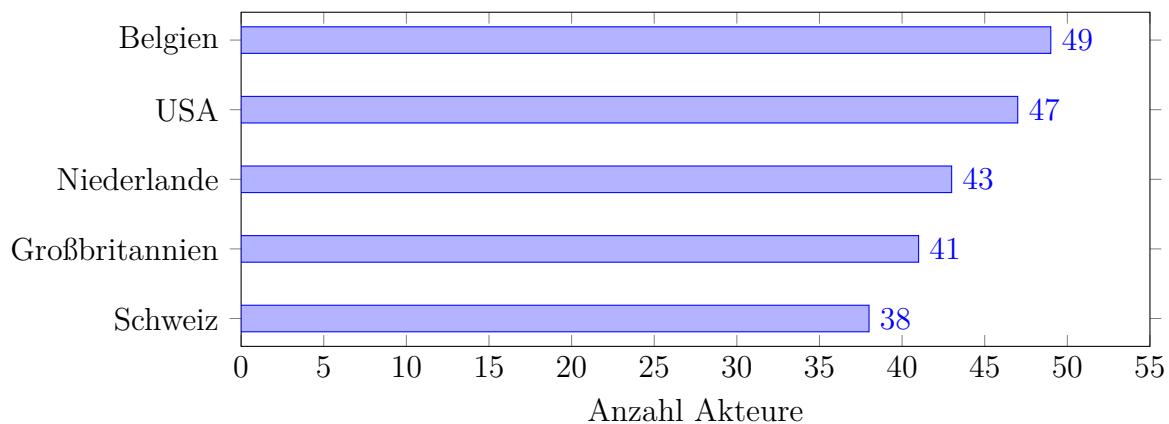


Abbildung 5.3: Top 5 Herkunftsländer internationaler Lobbyisten (ohne Deutschland).

5.2 Die PLZ-Heatmap

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

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**. Interessant sind jedoch die Cluster in der „Provinz“:

- **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.

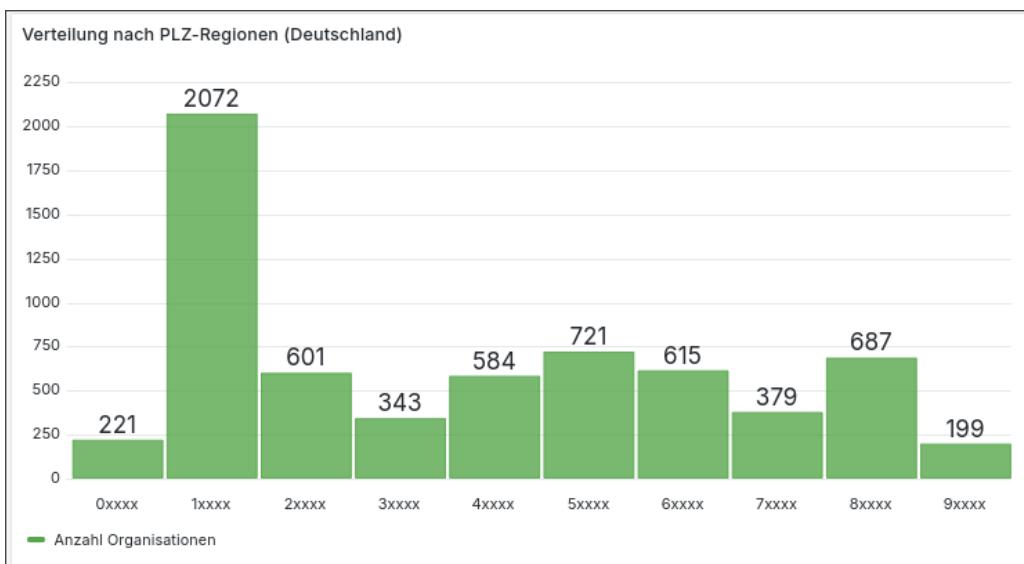


Abbildung 5.5: Verteilung der Lobby-Akteure nach Postleitzahl-Regionen (Grafana).

5.3 München vs. Berlin: Wo sitzt das Geld?

Ein besonders spannendes "Data Storytelling Element ergab sich aus dem Vergleich der durchschnittlichen Finanzausstattung pro Stadt. Zwar hat Berlin die absolute Mehrheit an Lobbybüros, 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.

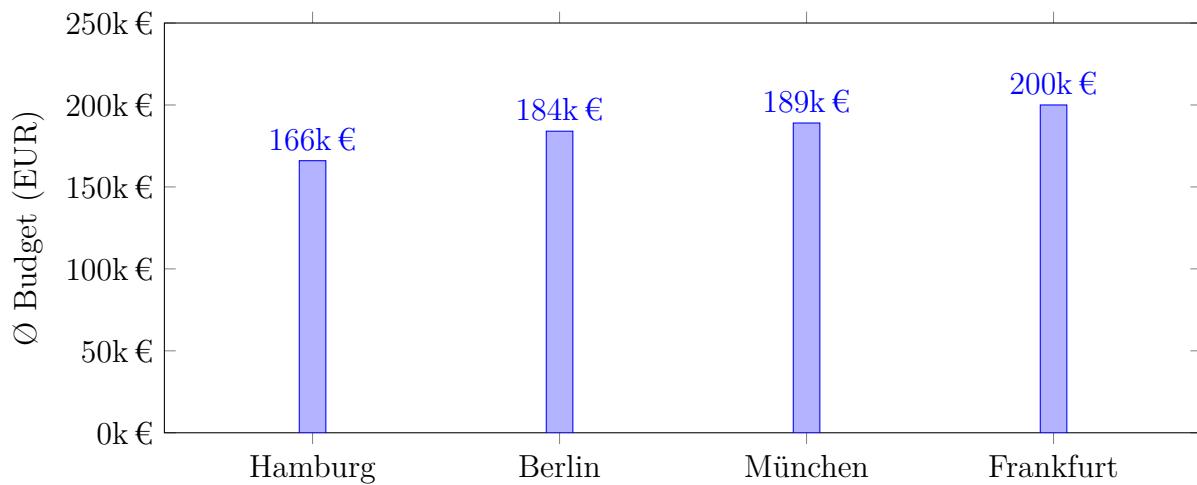


Abbildung 5.7: Vergleich der durchschnittlichen Lobby-Budgets deutscher Großstädte. Finanzmetropolen liegen vor der politischen Hauptstadt.

6 Transparenz & Compliance

Ein Register ist nur so gut wie die Datenqualität, die eingepflegt wird. 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 ist für Nutzer interessant, welchem politischen Spektrum ein Ex-Politiker zuzuordnen ist. Wir haben hierfür eine Text-Mining-Lösung direkt in SQL implementiert, die das Feld `function_position` (z.B. "Ehemaliger MdB (SPD)") analysiert.

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).

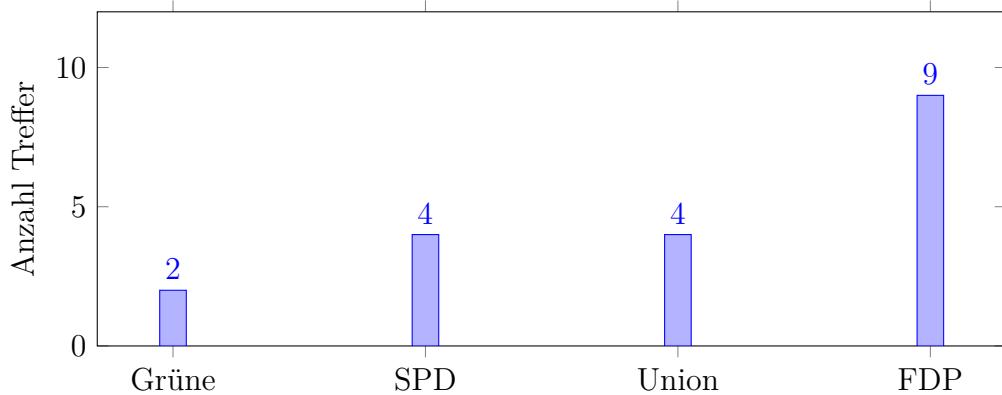


Abbildung 6.3: Erkannte Parteizugehörigkeiten bei Ex-Politikern (Text-Mining).

6.3 Strukturanalyse: Rechtsformen und ihre Agenda

Ein weiterer tieferer Blick in die Daten offenbart eine klare Arbeitsteilung zwischen den verschiedenen Rechtsformen der Akteure. Wir haben untersucht, 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 ganz klar 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 und Lessons Learned

Dieses Portfolio markiert den Abschluss eines umfassenden Data-Engineering-Projekts, das uns über das gesamte Semester begleitet hat. Was mit der abstrakten Idee begann, „Lobbyismus sichtbar zu machen“, entwickelte sich zu einer ausgewachsenen Analytics-Plattform. Rückblickend lässt sich dieser Prozess als Evolution in vier Phasen beschreiben, die jede für sich essenzielle Erkenntnisse lieferte.

Phase 1: Die Komplexität der Realität (Modellierung)

Der erste „Realitätsschock“ traf uns bereits beim Entwurf des ER-Modells. Die Annahme, Lobbyismus ließe sich in wenigen Tabellen abbilden, erwies sich als Trugschluss. Die API des Bundestags lieferte tief verschachtelte JSON-Strukturen, die wir in ein strenges Schema der 3. Normalform überführten. **Learning:** Datenmodellierung ist kein akademischer Selbstzweck. Die Entscheidung für ein hochgradig normalisiertes Schema (über 80 Tabellen) garantierte uns zwar maximale Datenintegrität und Vermeidung von Redundanzen, legte aber bereits hier den Grundstein für die späteren Performance-Herausforderungen. Wir lernten, dass jedes ‘ARRAY‘ im JSON eine eigene 1:N-Tabelle fordert – ein massiver Aufwand, der sich jedoch für die Datenqualität auszahlte.

Phase 2: Robustheit durch Engineering (ETL)

In der Implementierungsphase verschob sich der Fokus von der Struktur auf den Prozess. Wir bauten eine asynchrone Python-Pipeline, die nicht nur Daten „schaufelt“, sondern resilient gegenüber Fehlern ist. **Learning:** Eine gute Pipeline ist langweilig – sie bricht nicht ab. Der Einsatz von `asyncio` und Queues lehrte uns, wie man Producer (API) und Consumer (DB) entkoppelt. Besonders wertvoll war die Erfahrung mit Docker: Die Containerisierung der gesamten Umgebung (DB, Ingestor, Adminer, Grafana) machte das Projekt reproduzierbar und „teamfähig“.

Phase 3: Der Preis der Normalisierung (Optimierung)

Dies war der technische Wendepunkt. Unsere normalisierte Datenbank war zwar sauber, aber für Analysen unerträglich langsam. Einfache Fragen wie „Wer kennt wen?“ erforderten Joins über fünf Tabellen. **Learning:** „Korrektheit“ ist nicht gleich „Performance“. Wir mussten lernen, unsere Prinzipien (Redundanzfreiheit) pragmatisch zu opfern. Durch den Einsatz von **MaterIALIZED VIEWS** und spezialisierten **GIN-INDIZES** (für Textsuche) konnten wir Antwortzeiten von Sekunden auf Millisekunden drücken. Die Erkenntnis: Ein Data Engineer muss nicht nur wissen, wie man Daten speichert, sondern auch, wie die Disk-I/O funktioniert.

Phase 4: Von Daten zu Antworten (Visualisierung)

Im finalen Schritt schloss sich der Kreis. Die technischen Vorarbeiten mündeten in Grafana-Dashboards, die nun auch inhaltliche Fragen beantworten. Wir fanden heraus, dass das „Energieeffizienzgesetz“ ein Lobby-Hotspot ist und dass PLZ-Regionen klare Cluster bilden. **Learning:** SQL ist ein mächtiges BI-Tool. Die Fähigkeit, Daten nicht nur zu besitzen, sondern sie visuell explorierbar zu machen, ist der eigentliche Mehrwert des Projekts.

Schlusswort

Zusammenfassend hat das Modul „Advanced Database Techniques“ weit mehr vermittelt als nur SQL-Syntax. Wir haben gelernt, ein komplexes System von der ersten Skizze bis zum produktiven Dashboard zu denken. Das Ergebnis ist eine Plattform, die beweist, wie moderne IT-Methoden (Container, Indizes, Analytics) einen Beitrag zur demokratischen Transparenz leisten können. Wir hinterlassen nicht nur Code, sondern ein Werkzeug, das Licht ins Dunkel politischer Einflussnahme bringt.

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.