

David Porochyn, Benjamin Leucht

Landeshauptstadt München |
21/03/2026



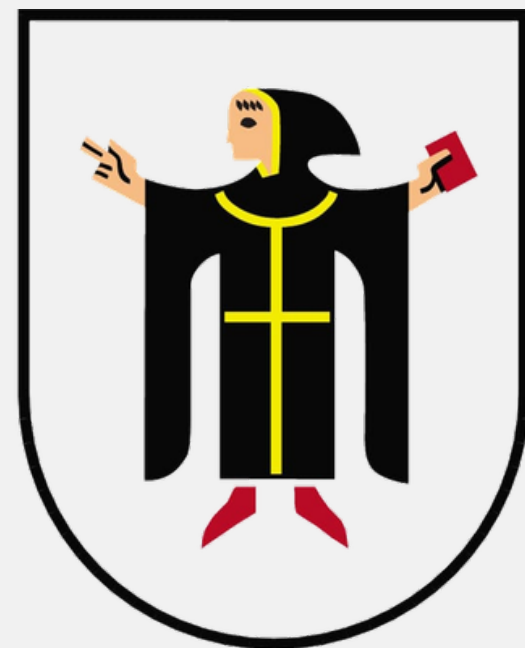
München Open Atlas

Open Data zugänglich machen

Hintergrund



x



Wir haben im Rahmen eines interdisziplinären TUM-Seminars zusammen mit dem IT-Referat an innovativen Lösungen für mehr digitale Inklusion gearbeitet.

IT-Tutorplattformen

**Ereignisbasierte
Verwaltung**

Eligibility Checker

Digitale Teilhabe

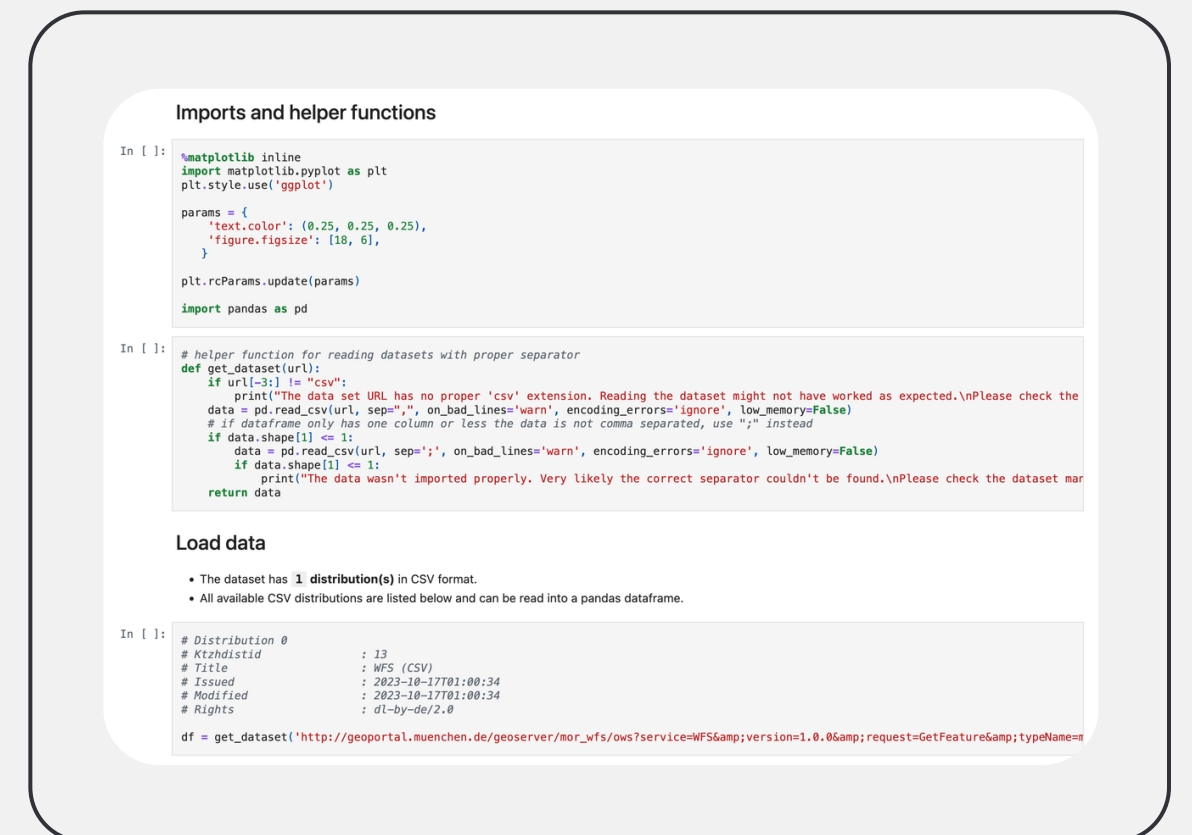
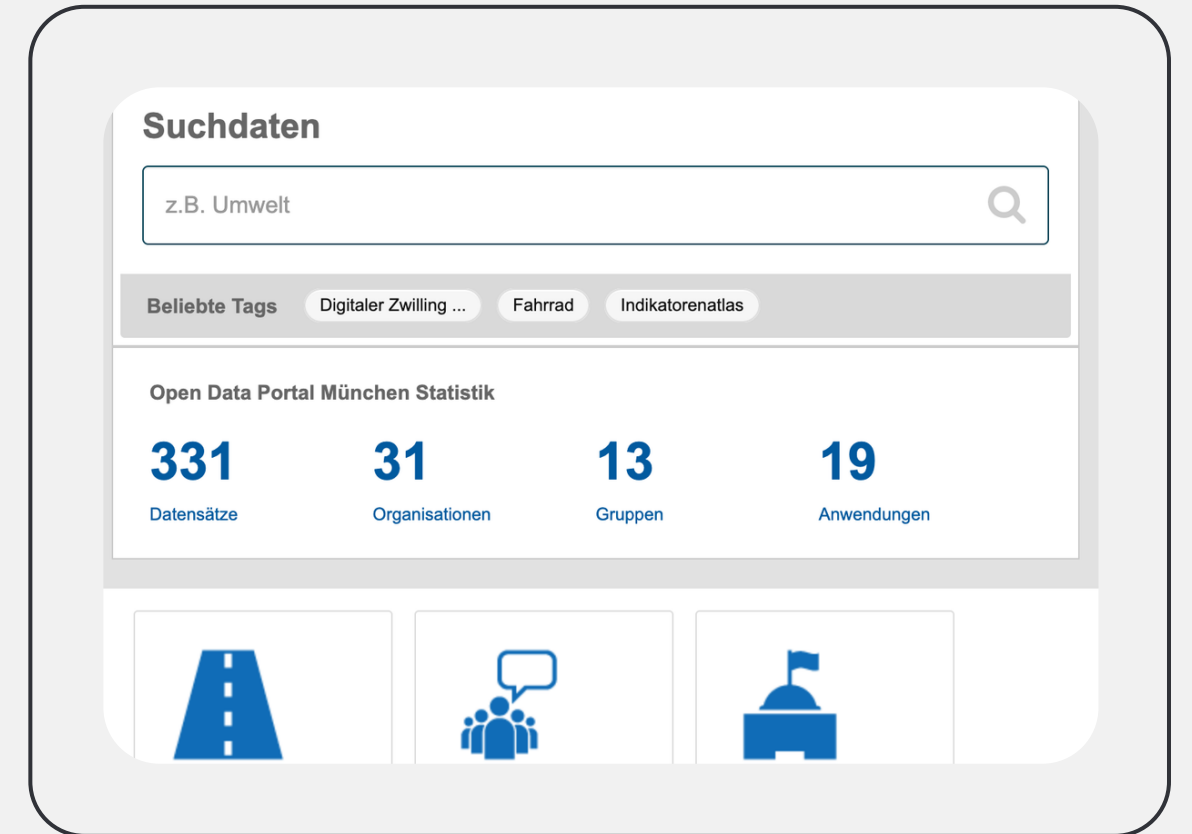
Open Atlas

...

...

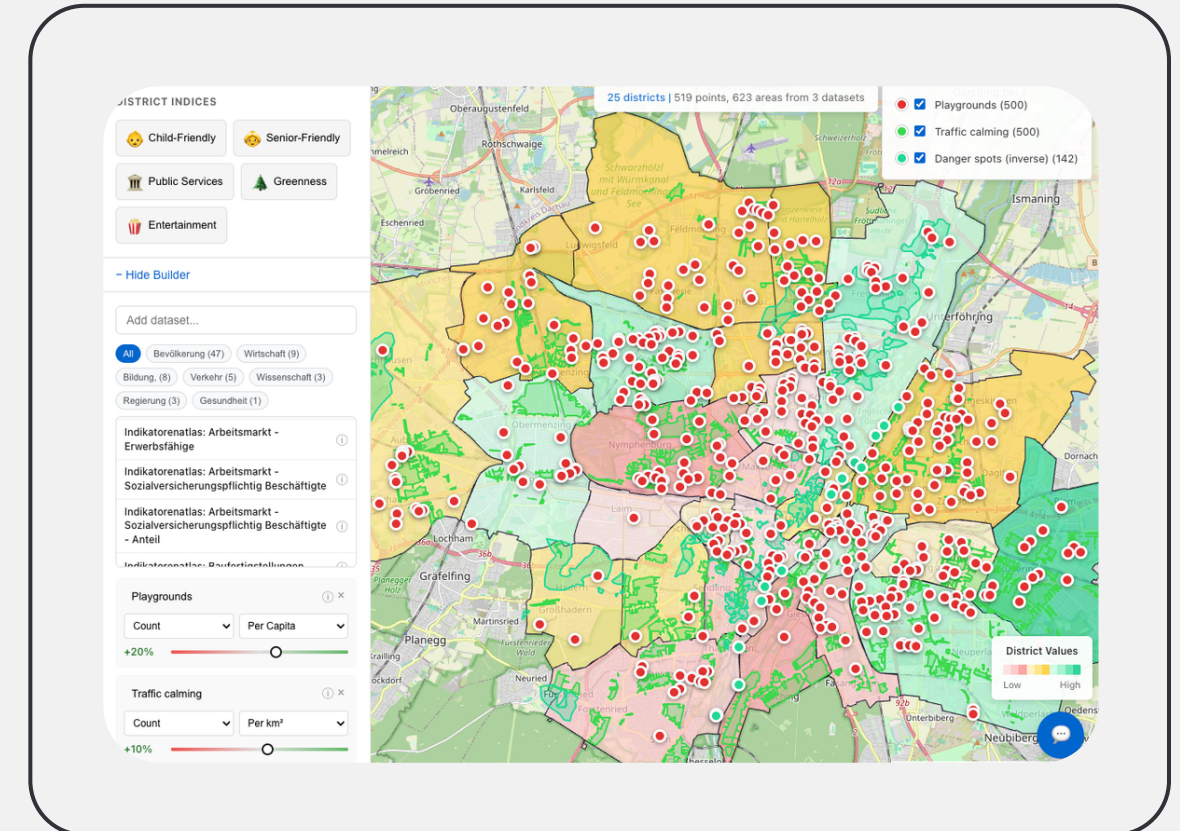
Das Problem

- Die Stadt veröffentlicht sehr viele offene Daten
- Doch es braucht technisches und fachspezifisches Wissen, um sie zu nutzen
 - Schlagwortsuche
 - Viele verschiedene Datenformate



Unsere Lösung

- Eine interaktive Karte auf der 154 Datensätze der Stadt visualisiert werden können
- Ein Indexbuilder mit dem man Stadtbezirksweite Rankings erstellen kann
- Ein Chatbot der verschiedene Anfragen datenbasiert beantworten kann

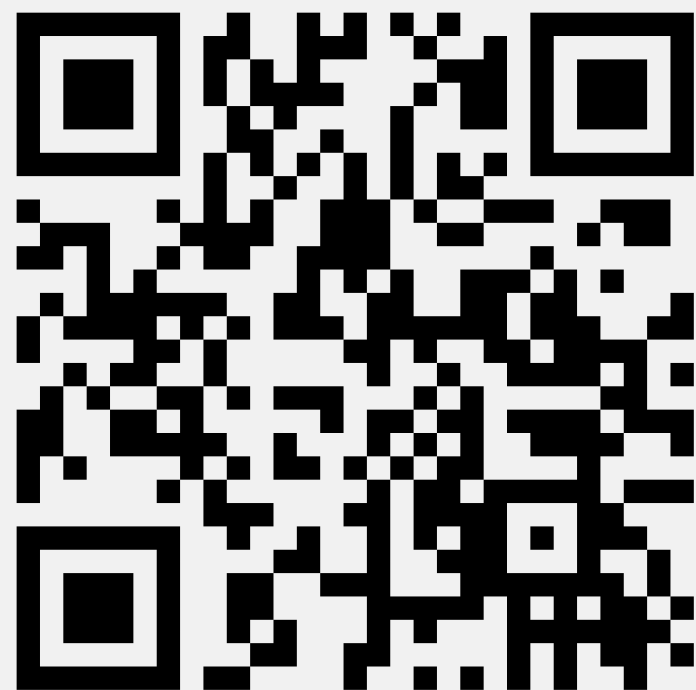


The screenshot shows a chatbot interface titled 'Data Assistant' with a 'Connected' status. The chat history shows a user asking 'What social centers are there in Sendling'. The chatbot responds with the title 'Alten- und Servicezentrum' and a description: 'The dataset lists one Alten- und Service-Zentrum in Sendling: Alten- und Service-Zentrum Sendling, located on Daiserstr. This is the only Sendling entry found in the provided data.' Below the text, it shows a table with 1 row of data:

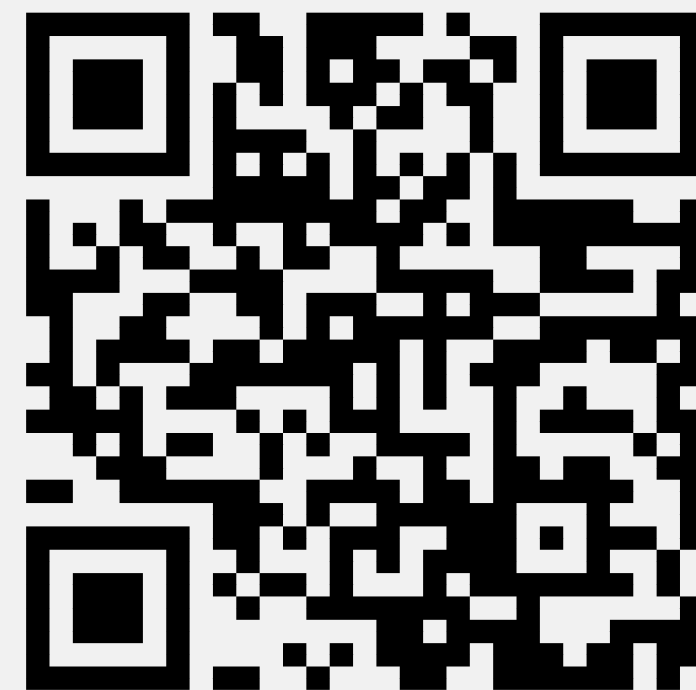
name	st2	za2	zk1	zk2	zs03	objectid
Alten- und Service-Zentrum Sendling	Daiserstr.	806	689328	5.3326e+06	6	3

At the bottom, there is a text input field with the placeholder 'Ask about Munich data...' and a 'Send' button.

Demo



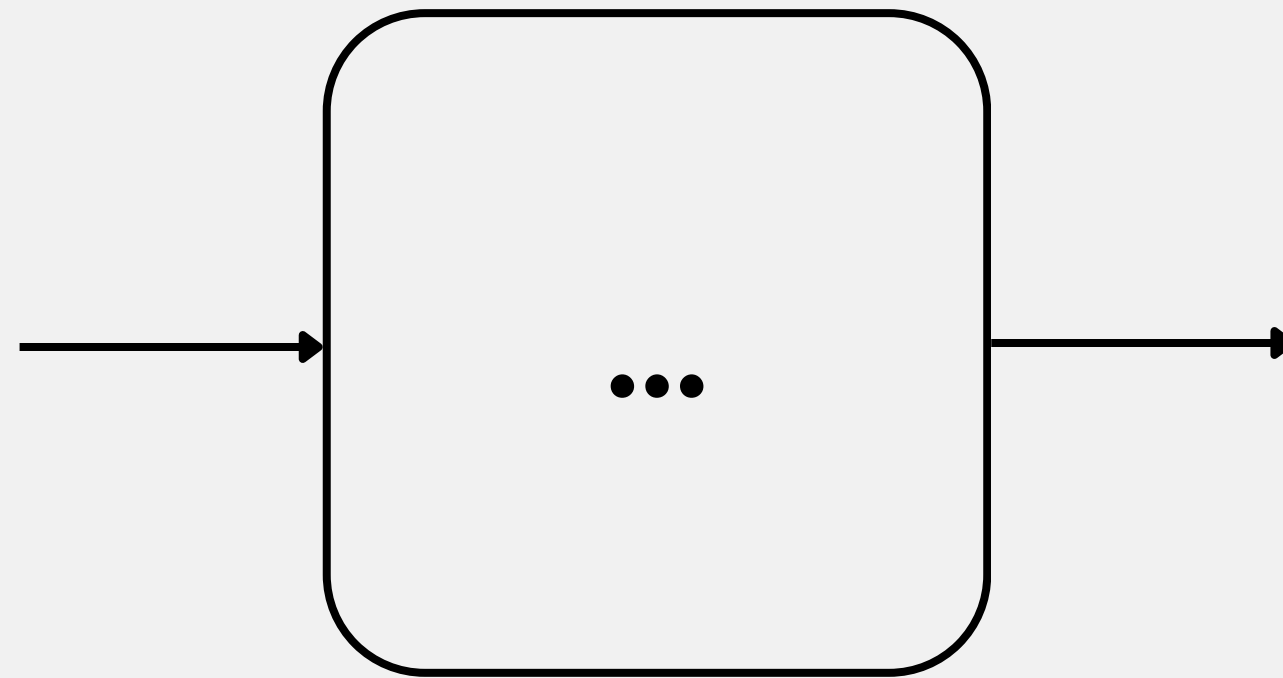
Code



**Open Data
Portal**



(321 Datensätze, 7+ Formate, 2
CRS, diverse Encodings)



**Einheitliche
Datenbank**

(154 Datensätze, 285.898
Features, Bezirkszuordnung)

Format-Filterung

- Nur nutzbare Formate werden berücksichtigt: CSV, JSON, GeoJSON, WFS, GML
- PDFs, HTML, XLSX, WMS → nicht sinnvoll maschinenlesbar

...

...

Parsing

- Jedes Format braucht einen eigenen Parser/eine andere Heuristik um relevante Daten zu erkennen.
- **Beispiel CSV:**
 - Delimiter erraten
 - das Encoding durch 5 Fallbacks reparieren
 - die Spalten mit Koordinaten finden

...

...

Koordinaten

- Verschiedene Formate: WGS84 oder EPSG:25832, oft ohne explizite Angabe
- Also Heuristik:
 - Koordinaten gegen München-Grenzen prüfen und, falls sie außerhalb sind, in WGS84 umrechnen

...

...

Bezirkszuordnung

- Datensätze benennen relevante Spalte unterschiedlich:
 - "sb_nummer", "stadtbezirk", "raumbezug" ...
- Falls keine Bezirksinfo vorhanden, wird der Datenpunkt bzw. sein Centroid über Bezirksgrenzen manuell zugeordnet

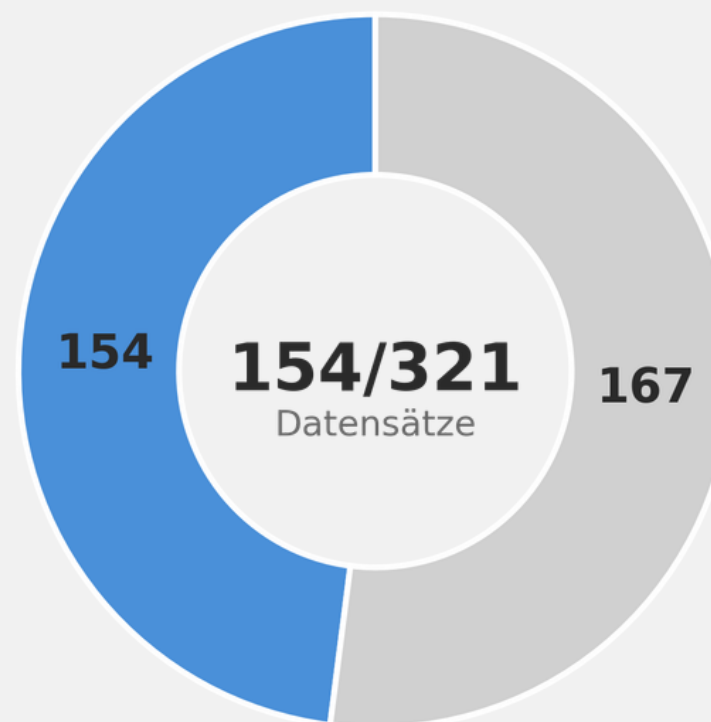
...

...

...



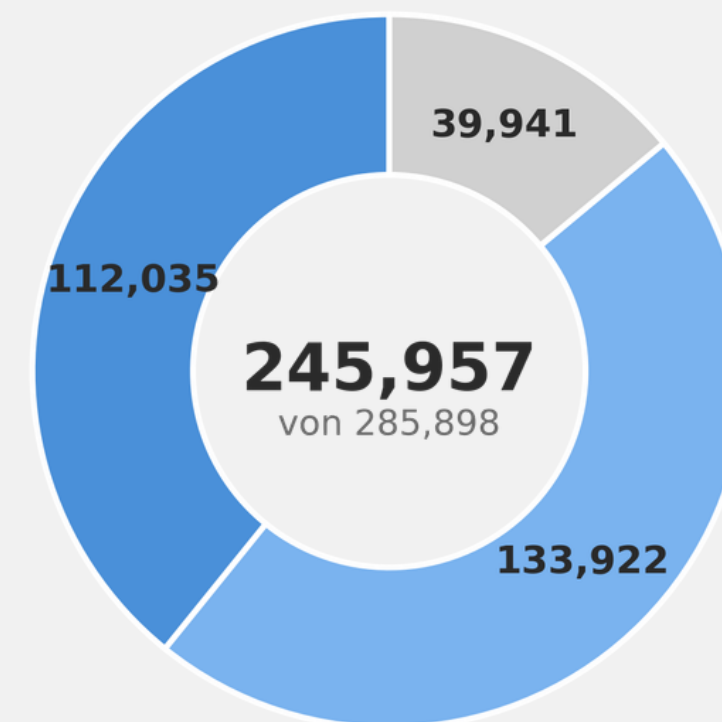
Datensätze



nutzbar

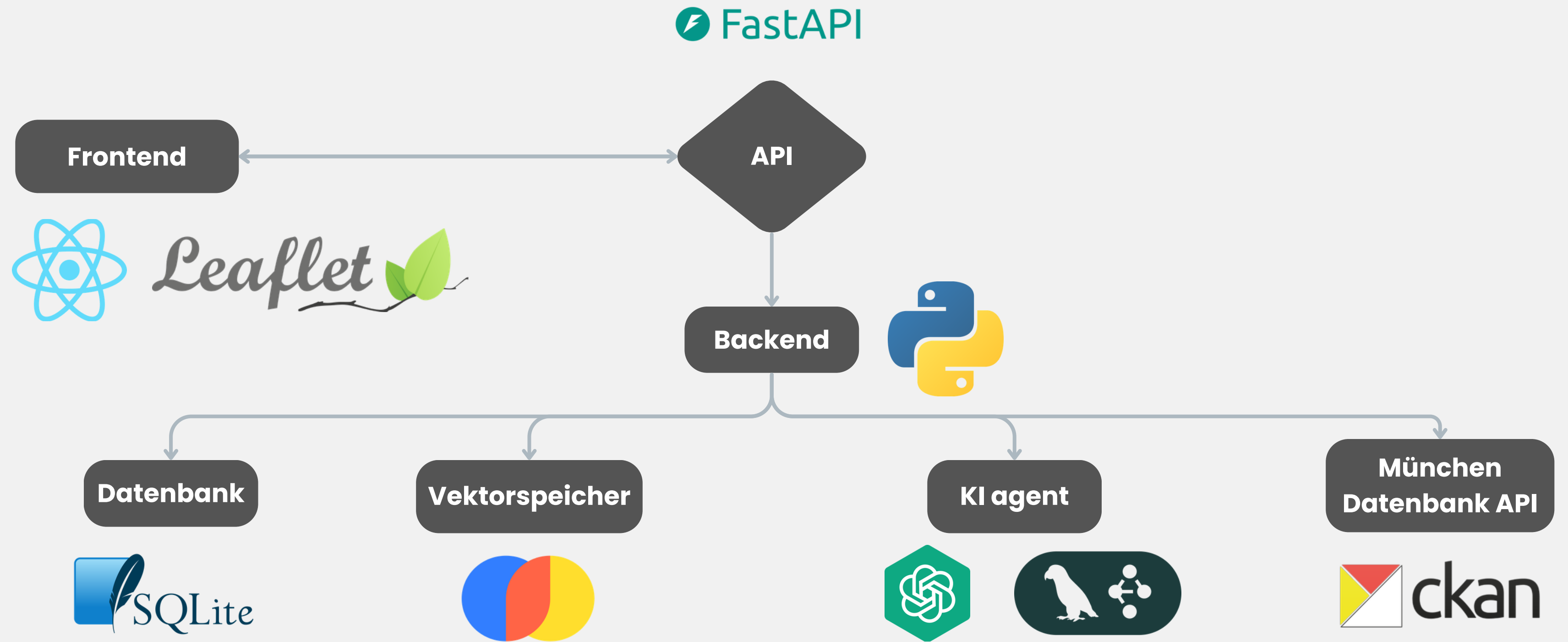
(154 mit Features, 167 ohne)

Features



- Geo + Bezirk (112,035)
- Nur Bezirk (133,922)
- Nicht zugeordnet (39,941)

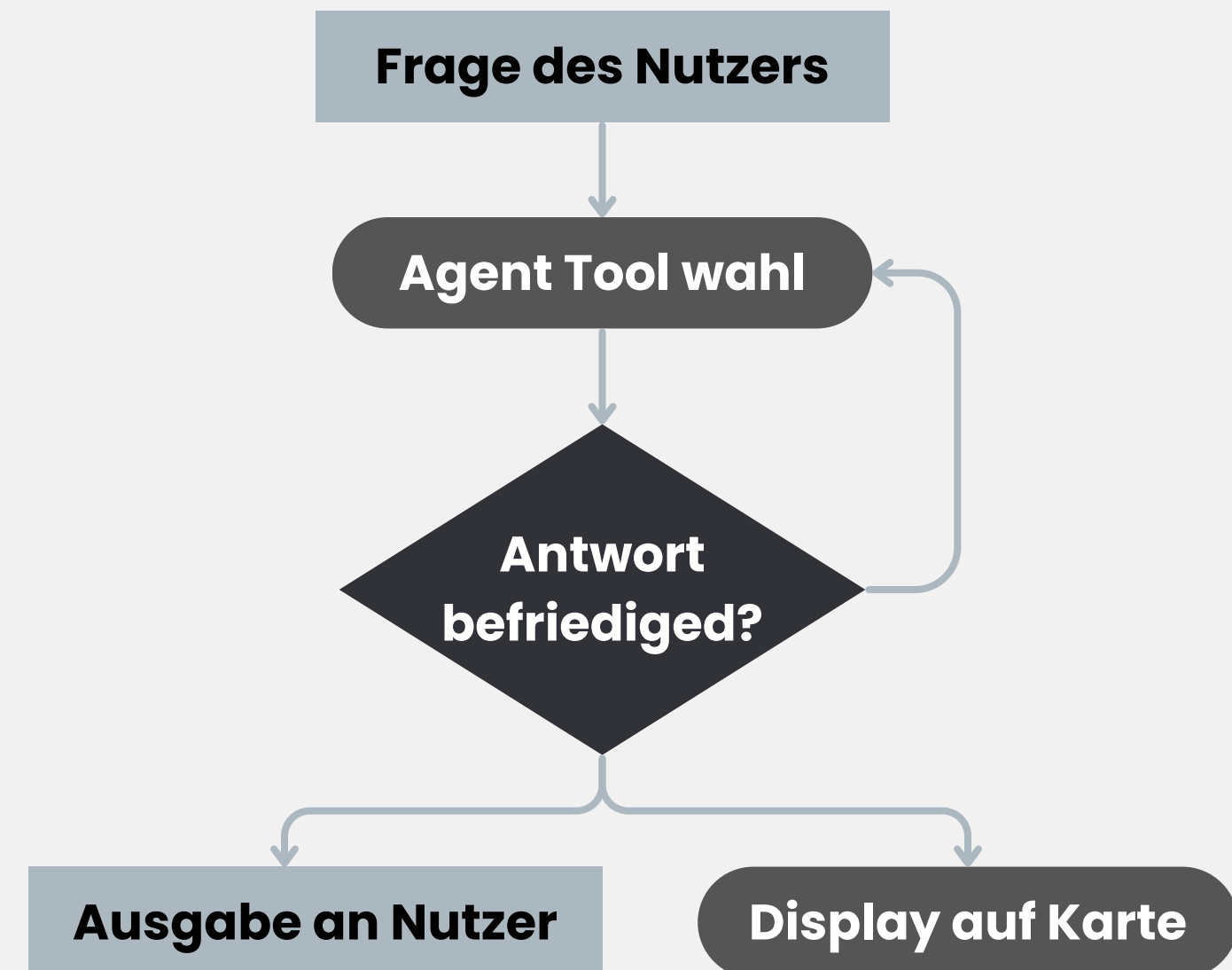
Architekturübersicht



Architektur

- **LangGraph Agent**
- **Tool Aufruf Schleife:**
 - Semantische Suche
 - Bedingte generierung von SQL-Anfragen
 - Index-Berechnung
 - Näherungssuche
- **Vektorspeicher:**
 - ChromaDB
 - OpenAI embeddings
- **DuckDB** für KI-generierten SQL-Analysen

Ablauf einer Anfrage



Unterstützte Fragetypen

“Welcher Bezirk ist am besten für Familien?”

↳ **prüft existierende Indizes** → **Gibt Rangliste + Index-Karte aus**

“Wie viele Spielplätze gibt es in Schwabing?”

↳ **Semantische Suche mit ChromaDB** → **SQL-Anfrage wird generiert** → **LLM Antwort mit Zahl**

“Welche Krankenhäuser sind in der Nähe des Marienplatzes?”

↳ **Geokodierung** → **räumliches SQL auf den passenden Datensatz** → **Auf der Karte angezeigt**

“Vergleiche Radwege vs. ÖPNV pro Bezirk”

↳ **Multi-Datensatz Synthese** → **Tabelle**

“Zeig mir die Fußgängerfreundlichste Bezirke”

↳ **Datensatz-Wahl** → **Index-berechnung** → **Index in der App angezeigt**

Munich Data

DE

DISTRICT INDICES



Child-Friendly

Senior-Friendly



Public Services

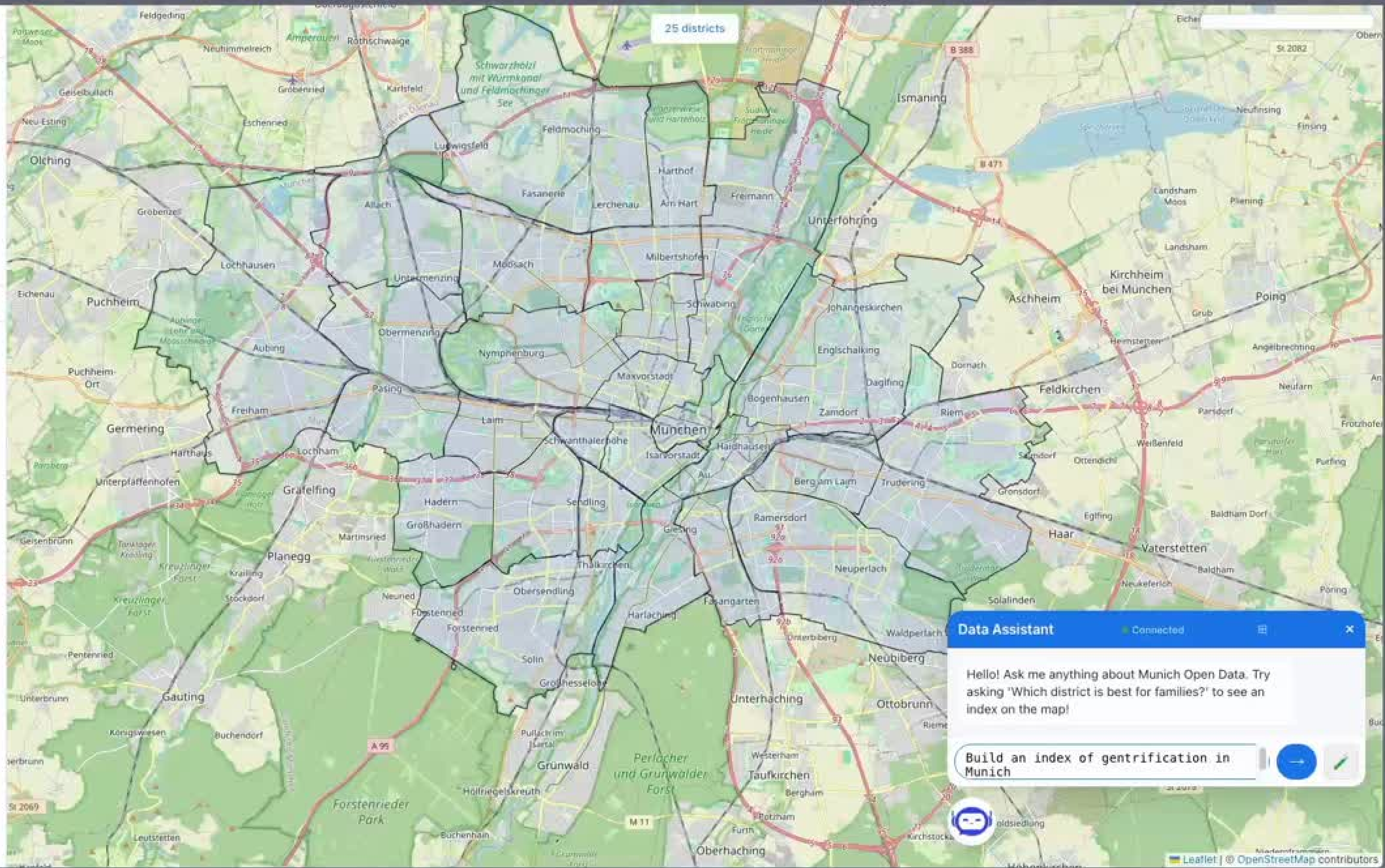


Sustainable Mobility



Entertainment

+ Custom Index



Data Assistant

● Connected

Hello! Ask me anything about Munich Open Data. Try asking "Which district is best for families?" to see an index on the map!

Build an index of gentrification in Munich



oldiedlung

Herausforderungen und Überraschungen

Datenheterogenität

- Derselbe Inhalt hat viele **verschiedene Spaltennamen**
- Für Bezirk:
 - sb_nummer
 - stadbezirk
 - bezirk
 - Stadtbezirk-NR
- Erkennung durch **Parsing** und **Mustererkennung**

Koordinatensysteme

- München speichert Koordinaten in **verschiedenen Formaten**
- Datenformate könnten **vereinheitlicht** werden

LLM-Halluzinationen

- Bei **SQL Generierung**:
 - LLM erfindet Spaltennamen
 - Generiert fehlerhaftes SQL
- Lösung:
 - **Genaues Schema** in jeden Prompt injizieren
 - **Format** von SQL Anfragen genau bedingen
 - **Sicherheitsvalidierung**

Was haben wir gelernt?

1.

Open data ≠ zugängliche Daten

Das Sammeln von Opendata hat nur Wert, wenn die Daten informativ und verfügbar für alle sind.

2.

Datenqualität ist essenziell

Kein Algorithmus kann einen Datensatz retten, der nur 12 von 25 Bezirken abdeckt. Lücken transparent zu machen ist wichtiger als sie zu verstecken.

3.

LLMs brauchen Kontext, nicht Freiheit

Ein KI-System ist nur so gut wie der Kontext, den man ihm gibt. RAG ist keine Optimierung, sondern Voraussetzung.

Was als nächstes kommt



Datenbasis erweitern

Andere Datenquellen

Teilbare Nutzer-Indizes



Zeitreihen:

Wie verändert sich
ein Bezirk über die Jahre

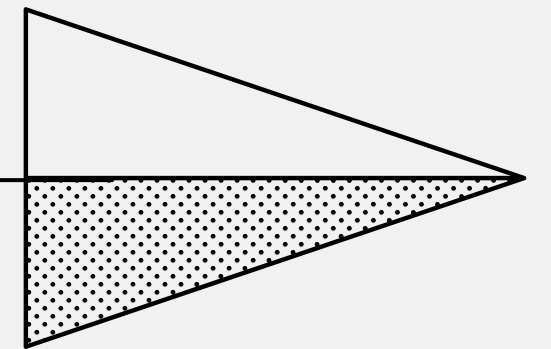
Integration in den Webstack der Stadt



Mehr Städte

Live Datenintegration:

Verkehr, Luftqualität...



Danke, Fragen?