

# **KI-basierte Auswertung kommunaler Daten mit dem GPT Data Analyzer**

## **– Arbeitspapier –**

Dr. Neven Josipovic

Braunschweig, 12.10.2023

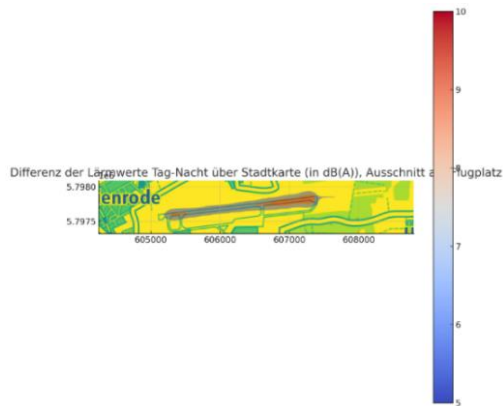
### **Abstract**

Der Einsatz von künstlicher Intelligenz (KI) gewinnt in der öffentlichen Verwaltung zunehmend an Bedeutung. Neben der Ein- und Ausgabe von Texten rückt dabei die KI-basierte Analyse von Daten in den Fokus. Das Arbeitspapier untersucht vor dem Hintergrund, wie kommunale Daten mit dem GPT Data Analyzer ausgewertet und aufbereitet werden können. Dazu wurden öffentlich verfügbare kommunale Daten in ChatGPT hochgeladen und dialogbasiert analysiert. Die ausgewählten Datensätze repräsentieren gängige offene Datenformate und umfassen sowohl Geodaten als auch strukturierte Tabellendaten. Im Ergebnis zeigt sich, dass sich GPT-4 insbesondere durch seine Fähigkeit zur Interaktion in natürlicher Sprache hervorhebt, was weniger Expertise im Hinblick auf die Analyse von Daten erfordert und den Analyseprozess beschleunigt. Als intelligentes System kann es außerdem Unregelmäßigkeiten in den Daten erkennen und beheben sowie Dokumentationen und Interpretationen durchführen. Die Ergebnisse der Datenanalysen mit GPT-4 verdeutlichen die Potenziale des Einsatzes von KI für die Automatisierung von Prozessen. Für eine optimale Nutzung sollten Verwaltungen den Umgang mit Daten standardisieren und ihre Mitarbeiterinnen und Mitarbeiter zur Kommunikation mit KI-Systemen schulen. Da KI-Modelle wahrscheinlich in externen Cloud-Umgebungen gehostet werden müssen, gewinnt der Datenschutz an Bedeutung. Um geeignete Lösungen für die öffentliche Verwaltung zu entwickeln, könnte eine enge Zusammenarbeit zwischen der öffentlichen Verwaltung und privaten IT-Dienstleistern von Vorteil sein.

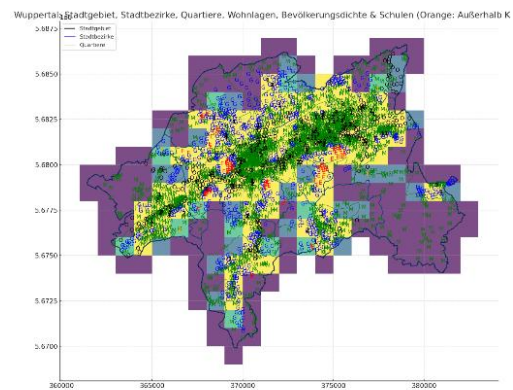
**Inhalt**

Ergebnisbeispiele .....	3
I. Einleitung .....	4
1. Ziele .....	4
2. Methodik.....	4
3. Ausgewählte Datensätze .....	4
II. Datenanalyse in Dialogform .....	6
1. Geodaten (Fluglärm).....	6
2. Geodaten (Bevölkerungsdaten).....	10
3. Haushaltsdaten (Ergebnishaushalt).....	18
4. Bevölkerungsdaten (Demografie) .....	24
5. Sensordaten (Passantenfrequenzen) .....	28
6. Umfragen (Innenstadturnfrage).....	30
7. Verkehrsdaten (Unfallstatistik).....	33
III. Diskussion der Ergebnisse und Potenziale .....	36
1. Dialogbasierte Datenanalyse .....	36
2. Bedeutung der Input-Qualität („Prompts“) .....	36
3. Mensch-Maschine-Interaktion.....	37
4. Fehlerhafte Analysen .....	37
5. Automatisierung von Prozessen .....	37
IV. Anforderungen an die Datenbereitstellung .....	38
V. Fazit.....	39
Literatur.....	40
Anhang: Struktur der eingelesenen CSV-Dateien.....	41

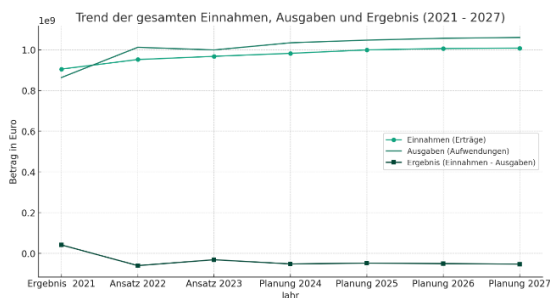
Ergebnisbeispiele



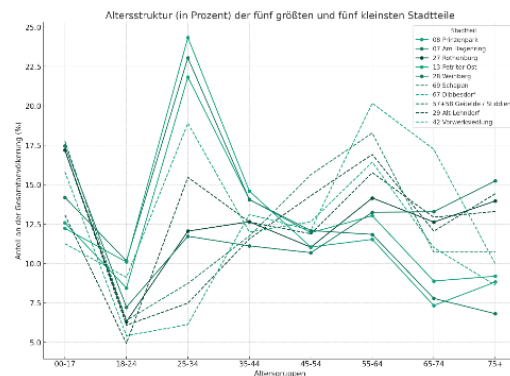
Überlagerung einer aus zwei Geodatenätzen generierten Lärmdifferenzkarte mit einem Stadtplan



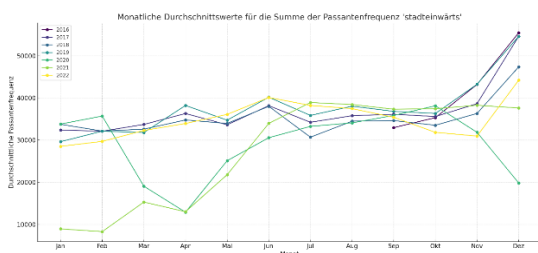
Zusammenführung, Darstellung und Analyse von Geodaten zu den Themen Wohnlage, Bevölkerungsdichte und Schulen



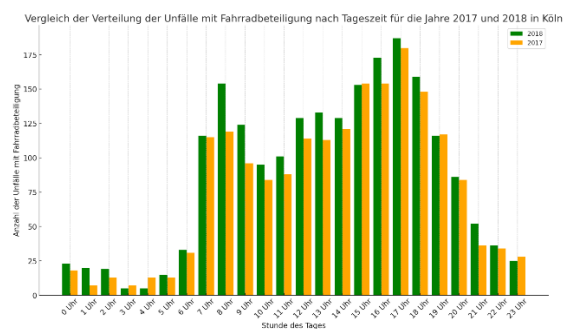
Grafische Darstellung von verschiedenen Auswertungen eines kommunalen Ergebnishaushaltes



Analyse und grafische Darstellung des Anteils bestimmter Altersgruppen an der Gesamtbevölkerung ausgewählter Stadtteile



Darstellung einer Analyse der durchschnittlichen monatlichen Passantenfrequenzen in Richtung Innenstadt



Auswertung und Darstellung einer kommunalen Verkehrsunfallstatistik im Vergleich zum Vorjahr

## **I. Einleitung**

GPT-4 ist ein KI-basiertes Sprachmodell, das in der Lage ist, Texte zu generieren und eine Vielzahl von Anfragen zu verarbeiten. Das Modell wird stetig um neue Funktionen und Plugins erweitert. Eine in der Beta-Phase befindliche Funktion ist der Data Analyzer. Er ermöglicht es dem Nutzer, strukturierte Rohdaten (beispielsweise als CSV-Datei) hochzuladen und durch die KI analysieren zu lassen.

### **1. Ziele**

Das Arbeitspapier geht der Frage nach, wie sich kommunale Daten im GPT-4 Data Analyzer auswerten und aufbereiten lassen. Anhand ausgewählter Anwendungsfälle sollen explorativ (d. h. ohne festgelegte Hypothesen oder Erwartungen) mögliche Potenziale der KI-basierten Datenanalyse für die öffentliche Verwaltung aufgezeigt werden. Außerdem wird untersucht, welche Anforderungen an die Bereitstellung von Daten bestehen, um eine KI-basierte Analyse zu ermöglichen.

### **2. Methodik**

Der methodische Ansatz besteht darin, die jeweiligen Datensätze zunächst einlesen zu lassen, der KI dann den Inhalt der Datensätze entweder anhand von Metadaten oder in Textform zu erläutern und anschließend Visualisierungen der (Roh-)Daten anzufordern. Im nächsten Schritt werden Analysen generiert, die über eine reine Visualisierung der Daten hinausgehen und verschiedene Rechenoperationen erfordern. Zuletzt wird geprüft, welche Einsichten und Schlussfolgerungen die KI aus den Daten ziehen kann.

### **3. Ausgewählte Datensätze**

Für das Arbeitspapier werden ausschließlich öffentlich verfügbare Daten verwendet. Die ausgewählten Datensätze sollen zum einen gängige offene Datenformate abbilden, zum anderen sowohl Geodaten als auch strukturierte Tabellendaten abdecken. Ausgewählt wurden folgende Datensätze (alle Links abgerufen am 12.10.2023) in den Formaten JSON, TIFF und CSV.

*Fall 1: Geodaten (Fluglärm)*

- Lärmkartierung Fluglärm in Braunschweig, 24 Stunden Tag (2022)  
<https://opendata.braunschweig.de/dataset/l%C3%A4rmkartierung-flugl%C3%A4rm-24-stunden-tag-2022>
- Lärmkartierung Fluglärm in Braunschweig, Nacht (2022)  
<https://opendata.braunschweig.de/dataset/l%C3%A4rmkartierung-flugl%C3%A4rm-nacht-2022>
- Regionalkarte Braunschweig 1: 100.000  
<https://opendata.braunschweig.de/dataset/regionalkarte-1-100000>

*Fall 2: Geodaten (Bevölkerung)*

- Bestandsdaten zu den Schulstandorten aller Träger im Wuppertaler Stadtgebiet  
<https://www.offenedaten-wuppertal.de/dataset/schulen-wuppertal>
- Wuppertaler Stadtgebiet als Polygon  
<https://www.offenedaten-wuppertal.de/dataset/stadtgebiet-wuppertal>
- Bevölkerungsdichte in der Stadt Wuppertal anhand eines Kilometerquadratrasters  
<https://www.offenedaten-wuppertal.de/dataset/bev%C3%B6lkerungsdichte-wuppertal>
- Grenzen der 69 Quartiere, die die 10 Wuppertaler Stadtbezirke weiter untergliedern  
<https://www.offenedaten-wuppertal.de/dataset/quartiere-wuppertal>
- Thematische Layer zur Wohnlagenkarte der Stadt Wuppertal  
<https://www.offenedaten-wuppertal.de/dataset/wohnlagen-wuppertal>
- Grenzen der 10 Wuppertaler Stadtbezirke  
<https://www.offenedaten-wuppertal.de/dataset/stadtbezirke-wuppertal>

*Fall 3: Haushaltsdaten (Ergebnishaushalt)*

- Haushaltsplan der Stadt Braunschweig 2023-2024 (nach Umwandlung von PDF zu CSV)  
[https://www.braunschweig.de/politik\\_verwaltung/fb\\_institutionen/fachbereiche\\_referate/fb20/fb20\\_1/Haushaltsplan-2023-2024.pdf](https://www.braunschweig.de/politik_verwaltung/fb_institutionen/fachbereiche_referate/fb20/fb20_1/Haushaltsplan-2023-2024.pdf)

*Fall 4: Bevölkerungsdaten (Demografie)*

- Bevölkerungsdaten der Stadt Braunschweig 2022  
[https://opendata.braunschweig.de/sites/default/files/BS\\_OPDA\\_SBZ\\_60\\_Open\\_Data\\_2022.csv](https://opendata.braunschweig.de/sites/default/files/BS_OPDA_SBZ_60_Open_Data_2022.csv)

*Fall 5: Sensordaten (Passantenfrequenzen)*

Täglichen Passantenfrequenzen für den Zeitraum 19.9.2016 bis 31.12.2022 für Braunschweig

[https://opendata.braunschweig.de/sites/default/files/Passantenfrequenz%20Innenstadt%2019.09.2016%20bis%2031.12.2022\\_all\\_in\\_one.csv](https://opendata.braunschweig.de/sites/default/files/Passantenfrequenz%20Innenstadt%2019.09.2016%20bis%2031.12.2022_all_in_one.csv)

*Fall 6: Umfragen (Innenstadtumfrage)*

- IFH-Kundenbefragung Vitale Innenstädte 2022 für Braunschweig  
<https://opendata.braunschweig.de/dataset/ifh-kundenbefragung-vitale-innenst%C3%A4dte-2022>

*Fall 7: Verkehrsdaten (Unfallstatistik)*

- Unfallstatistik Köln 2017/18  
<https://offenedaten-koeln.de/dataset/unfallstatistik-k%C3%B6ln>

## II. Datenanalyse in Dialogform

### 1. Geodaten (Fluglärm)

Eingabe: Lies bitte die folgende Datei ein. [Datei angefügt]

ChatGPT: The file you've uploaded, AG\_DE\_NI\_2\_Air\_LDEN.shp, appears to be a shapefile. A shapefile is a common geospatial vector data format for geographic information system (GIS) software. It is developed and regulated by Esri, and it can be used to represent geographical features. We can use the geopandas library in Python to read and analyze this file. Let's first load the file and examine its structure. Would that be okay?

Eingabe: Ja, gerne.

ChatGPT: Es scheint, dass beim Versuch, die Shapefile zu öffnen, ein Fehler aufgetreten ist. Shapefiles bestehen normalerweise aus mehreren zusammenhängenden Dateien mit den Erweiterungen .shp, .shx, .dbf und manchmal .prj, .sbn, .sbx, .shp.xml usw. Diese Dateien müssen alle zusammen vorhanden sein, damit die Shapefile korrekt gelesen werden kann. Der spezifische Fehler, den wir hier sehen, deutet darauf hin, dass die .shx-Datei fehlt oder nicht gefunden werden kann. Könnten Sie bitte überprüfen, ob diese Datei vorhanden ist und sie gegebenenfalls erneut hochladen?

Eingabe: Kann ich dir auch einen ganzen Zip-Ordner hochladen, der die genannten Dateien enthält?

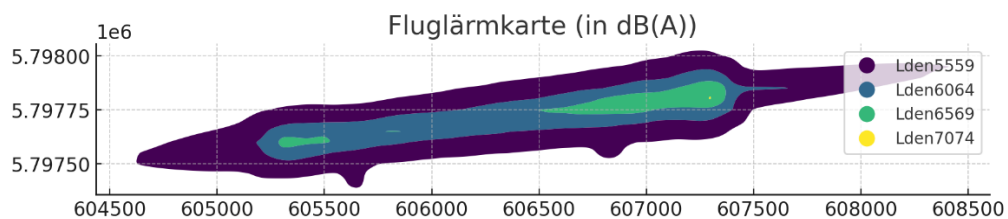
ChatGPT: Ja, absolut! Sie können einen ZIP-Ordner mit allen Dateien hochladen, die zur Shapefile gehören. Sobald Sie das getan haben, kann ich versuchen, die Dateien zu entpacken und die Shapefile erneut zu laden.

Eingabe: [Zip-Ordner angefügt]

ChatGPT: Die Dateien wurden erfolgreich entpackt. Wir haben jetzt alle Dateien, die zur Shapefile gehören, einschließlich der .shx-Datei, die vorher gefehlt hat. Lassen Sie uns nun versuchen, die Shapefile erneut zu laden. Die Shapefile wurde erfolgreich geladen. Hier sind die ersten paar Zeilen der Daten [...] Wie möchten Sie mit diesen Daten weiterfahren?

Eingabe: Kannst du mir die Daten bitte als Karte darstellen? Hinweis: Es handelt sich um eine Fluglärmkarte in der Einheit db(A).

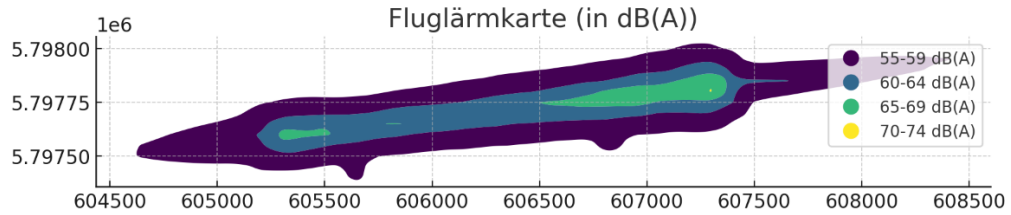
ChatGPT: Hier ist die Fluglärmkarte, die auf den Daten in Ihrer Shapefile basiert. Jede Zone auf der Karte repräsentiert eine bestimmte Lärmkategorie (in dB(A)), die durch unterschiedliche Farben dargestellt wird. Die Farbskala rechts von der Karte zeigt, welche Farben welche Lärmkategorien repräsentieren. Bitte beachten Sie, dass die Karte nur die räumliche Verteilung der Lärmkategorien zeigt und keine spezifischen geografischen Merkmale oder Bezugspunkte enthält. Wenn Sie zusätzliche Informationen auf der Karte darstellen möchten (z.B. Städte, Straßen, Flughäfen usw.), benötigen wir entsprechende Geodaten. Wie möchten Sie fortfahren? Haben Sie spezifische Fragen zu den Daten oder gibt es bestimmte Analysen, die Sie durchführen möchten?.



Eingabe: Kannst du die Legende anpassen, sodass die Werte in der Einheit db(A) dargestellt sind?

ChatGPT: Hier ist die aktualisierte Fluglärmkarte, bei der die Legende nun die Lärmpegel in der Einheit dB(A) anzeigt. Jede Zone auf der Karte repräsentiert einen bestimmten Lärmpegelbereich (in dB(A)), der durch unterschiedliche Farben dargestellt wird. Die

Farbskala rechts von der Karte zeigt, welche Farben welche Lärmpegelbereiche repräsentieren. Wie möchten Sie nun fortfahren? Haben Sie weitere Fragen zu diesen Daten oder gibt es bestimmte Analysen, die Sie durchführen möchten?

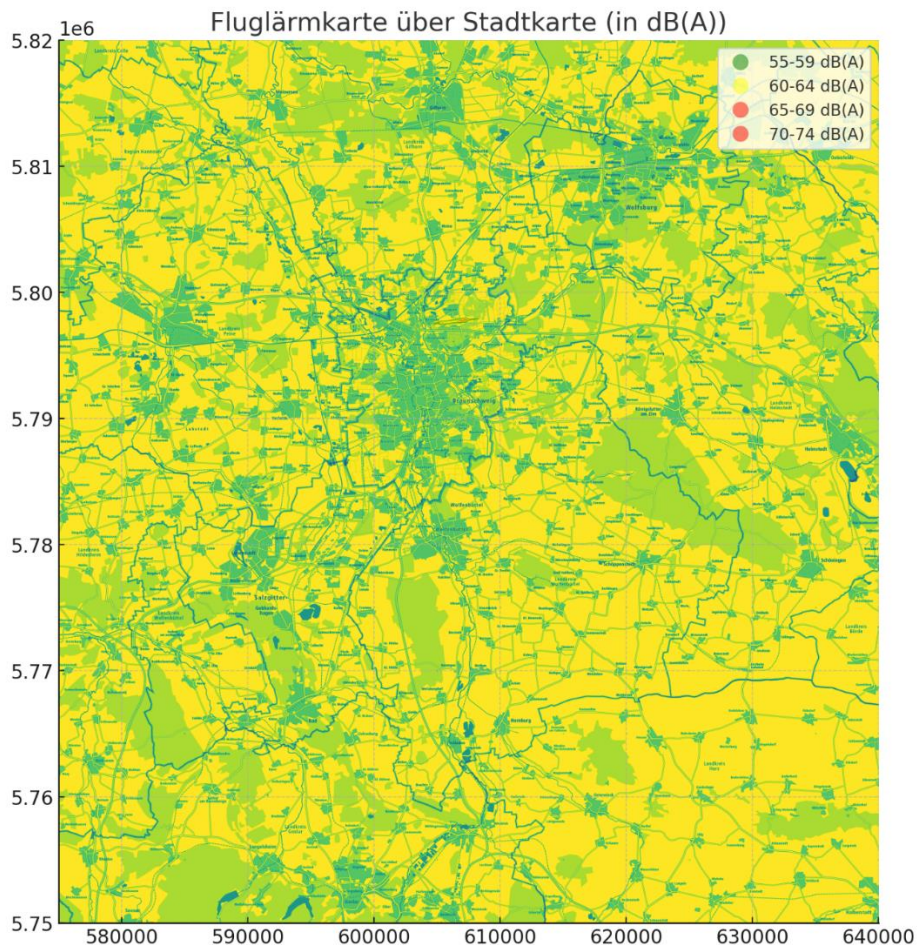


**Eingabe:** Könntest du die Karte als Overlay auf eine Stadtkarte legen, wenn ich dir die Stadtkarte hochlade? Ich habe sie in den Formaten tif und tfw.

**ChatGPT:** Ja, ich kann eine Rasterkarte (z.B. im .tif-Format) mit Ihrer Fluglärnkarte überlagern. Bitte laden Sie die Stadtkarte hoch, und ich werde das Overlay erstellen.

**Eingabe:** [Karte wird hochgeladen]

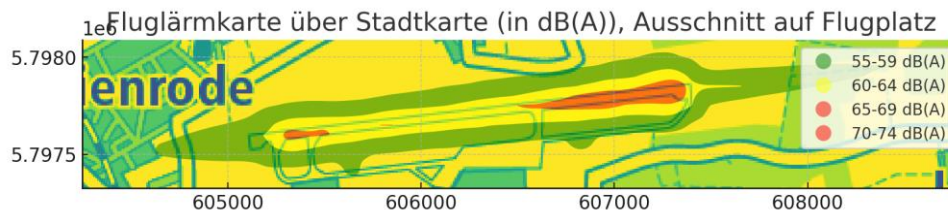
**ChatGPT:** Hier ist die Fluglärnkarte, die auf der Stadtkarte dargestellt wird. Die verschiedenen Farben repräsentieren unterschiedliche Lärmpegelbereiche in dB(A). Die Überlagerung ist halbtransparent, so dass die darunterliegende Stadtkarte sichtbar ist. Bitte beachten Sie, dass die genaue Ausrichtung und Überlagerung der beiden Karten von der Genauigkeit und Konsistenz der geografischen Koordinaten in den beiden Datensätzen abhängt. Wenn die Koordinatensysteme oder Maßstäbe der beiden Karten nicht genau übereinstimmen, kann die Überlagerung ungenau sein. Haben Sie weitere Fragen oder gibt es weitere Analysen, die Sie durchführen möchten?





**Eingabe:** Zeige bitte nur den Ausschnitt der Karte, wo sich der Flugplatz befindet. Du erkennst ihn an den erhöhten Lärmwerten.

**ChatGPT:** Hier ist die Fluglärmkarte, die auf einen Ausschnitt der Stadtkarte fokussiert ist, wo der Flugplatz vermutet wird (basierend auf den erhöhten Lärmwerten). Die verschiedenen Farben repräsentieren unterschiedliche Lärmpegelbereiche in dB(A). Bitte beachten Sie erneut, dass die genaue Ausrichtung und Überlagerung der beiden Karten von der Genauigkeit und Konsistenz der geografischen Koordinaten in den beiden Datensätzen abhängt. Wenn die Koordinatensysteme oder Maßstäbe der beiden Karten nicht genau übereinstimmen, kann die Überlagerung ungenau sein. Haben Sie weitere Fragen oder gibt es weitere Analysen, die Sie durchführen möchten?



**Eingabe:** Die Shape-Datei enthält die Lärmwerte für tagsüber. Wenn ich dir die Shape-Datei mit den Nachtwerten hochlade. Kannst du dann die Karte so anpassen, dass nicht die Lärmbereiche, sondern die db(A)-Differenz Tag-Nacht angezeigt wird?

**ChatGPT:** Ja, das ist möglich. Sie können die Shapefile mit den nächtlichen Lärmwerten hochladen, und ich kann dann eine neue Karte erstellen, die die Differenz zwischen den Lärmwerten tagsüber und nachts zeigt. Bitte laden Sie die entsprechende Datei hoch.

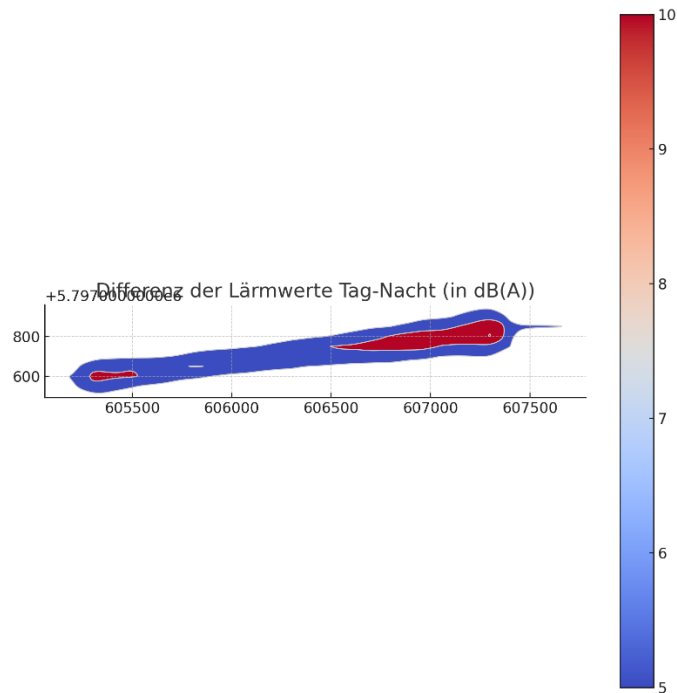
**Eingabe:** [Datei wird hochgeladen]

**ChatGPT:** Die Shapefile mit den nächtlichen Lärmwerten wurde erfolgreich geladen. Wir können nun die Differenz zwischen den Lärmwerten tagsüber und nachts berechnen und diese auf einer Karte darstellen. Bitte beachten Sie, dass dies voraussetzt, dass die Geometrien (d.h. die Orte) in beiden Datensätzen übereinstimmen. Andernfalls könnte die Differenzberechnung ungenau sein. Lassen Sie uns nun fortfahren und die Differenzkarte erstellen.

**Eingabe:** Kannst du als nächstes die Differenzkarte erstellen?

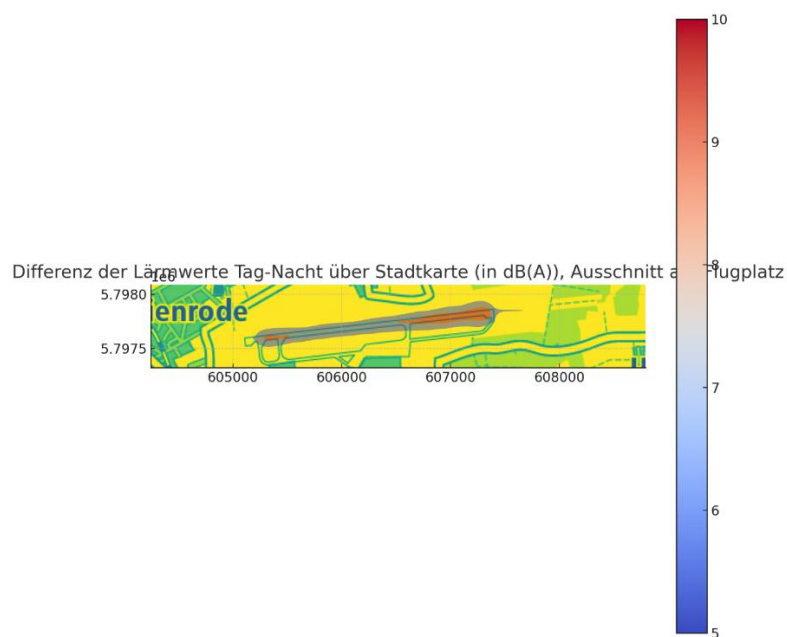
**ChatGPT:** Hier ist die Karte, die die Differenz der Lärmwerte zwischen Tag und Nacht darstellt. Die verschiedenen Farben repräsentieren unterschiedliche Differenzen in dB(A). Positive Werte (in Rot) zeigen Bereiche an, in denen der Lärm tagsüber lauter ist als nachts, während negative Werte (in Blau) Bereiche anzeigen, in denen der Lärm nachts lauter ist als tagsüber. Bitte beachten Sie, dass diese Analyse auf der Annahme basiert, dass die geometrischen Formen (d.h. die Gebiete, die von den Polygonen abgedeckt werden) in den Tag- und Nacht-Datensätzen genau übereinstimmen. Wenn dies nicht der Fall ist, könnte die berechnete Lärmwertdifferenz ungenau sein. Haben Sie weitere Fragen oder gibt es weitere Analysen, die Sie durchführen möchten?





**Eingabe:** Danke. Kannst du die Differenzkarte bitte als Overlay über die tif-Karte legen. Es genügt der Ausschnitt, wo sich der Flugplatz befindet?

**ChatGPT:** Hier ist die Karte, die die Differenz der Lärmwerte zwischen Tag und Nacht darstellt, überlagert auf dem Ausschnitt der Stadtkarte, auf dem wir den Flugplatz vermuten. Die verschiedenen Farben repräsentieren unterschiedliche Differenzen in dB(A). Positive Werte (in Rot) zeigen Bereiche an, in denen der Lärm tagsüber lauter ist als nachts, während negative Werte (in Blau) Bereiche anzeigen, in denen der Lärm nachts lauter ist als tagsüber. Bitte beachten Sie erneut, dass die genaue Ausrichtung und Überlagerung der beiden Karten von der Genauigkeit und Konsistenz der geografischen Koordinaten in den beiden Datensätzen abhängt. Wenn die Koordinatensysteme oder Maßstäbe der beiden Karten nicht genau übereinstimmen, kann die Überlagerung ungenau sein. Haben Sie weitere Fragen oder gibt es weitere Analysen, die Sie durchführen möchten?



## 2. Geodaten (Bevölkerungsdaten)

**Eingabe:** Der Datensatz beinhaltet das Wuppertaler Stadtgebiet als Polygon, wie es in der kleinräumigen Gliederung der Stadt Wuppertal geführt wird. Die kleinräumige Gliederung ist ein Knoten- und Kantenmodell der Stadt, das als Raumbezugsbasis der Wuppertaler Kommunalstatistik dient. Innerhalb dieses Systems wird das Stadtgebiet aus den Stadtbezirken aggregiert. Es handelt sich um keine flurstückscharfe Darstellung, hierfür muss auf Daten des Amtlichen Liegenschaftskatasterinformationssystems ALKIS zurückgegriffen werden. Die Abgrenzung des Wuppertaler Stadtgebiets ist langfristig konstant. Änderungen sind nur im Zusammenhang mit gesetzlich geregelten Gebietsreformen möglich und daher nicht vorhersehbar. Der Datensatz ist unter einer Open-Data-Lizenz (CC BY 4.0) verfügbar.

**ChatGPT:** Es scheint, dass Sie einen Datensatz in Form einer JSON-Datei hochgeladen haben, die die Stadtbezirke von Wuppertal repräsentiert. Wie kann ich Ihnen damit weiterhelfen?

**Eingabe:** Es kommen noch vier Datensätze. Bitte führe alle zusammen und stelle das Ergebnis visuell dar.

**ChatGPT:** Natürlich, bitte laden Sie die weiteren vier Datensätze hoch, und ich werde sie dann zusammenführen und visualisieren.

**Eingabe:** [Datensätze werden hochgeladen]

**ChatGPT:** Es Jetzt haben wir alle fünf Datensätze:

1. Stadtbezirke
2. Bevölkerungsdichte
3. Wohnlagen
4. Quartiere
5. Stadtgebiet

Um die Daten korrekt darzustellen, werde ich folgendermaßen vorgehen:

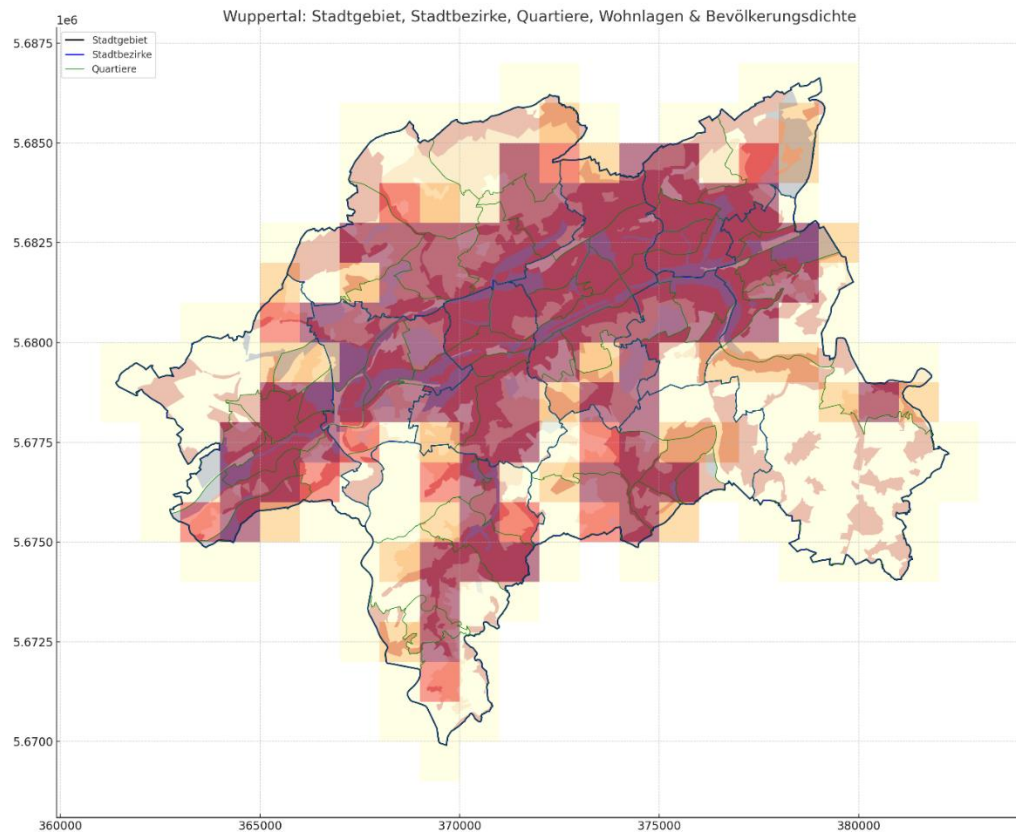
1. Das Stadtgebiet als Grundlage darstellen.
2. Die Stadtbezirke darauf abbilden.
3. Die Quartiere innerhalb der Stadtbezirke visualisieren.
4. Die Wohnlagen und die Bevölkerungsdichte als überlagernde Schichten hinzufügen.

Ich werde jetzt mit der Visualisierung beginnen.

Hier ist die Visualisierung von Wuppertal mit den folgenden Ebenen:

- Stadtgebiet: Dargestellt durch die schwarzen Linien
- Stadtbezirke: Dargestellt durch die blauen Linien
- Quartiere: Dargestellt durch die grünen Linien
- Wohnlagen: Dargestellt durch verschiedene Farbtöne (von hell nach dunkel, basierend auf der Wohnlage)
- Bevölkerungsdichte: Dargestellt durch verschiedene Schattierungen von Gelb bis Rot (je dunkler, desto dichter besiedelt)

Lassen Sie mich wissen, ob Sie weitere Anpassungen oder Informationen zu dieser Visualisierung wünschen!



**Eingabe:** Danke! Die Visualisierungen für Wohnlage und Bevölkerungsdichte sind optisch schwer zu unterscheiden. Kannst das so darstellen, dass man die beiden Informationen weniger leicht verwechseln bzw. leichter unterscheiden kann?

**ChatGPT:** Natürlich, ich verstehe Ihr Anliegen. Eine Möglichkeit, die beiden Informationen besser zu unterscheiden, besteht darin, unterschiedliche Farbpaletten oder sogar unterschiedliche Darstellungsmethoden zu verwenden.

Hier ist mein Vorschlag:

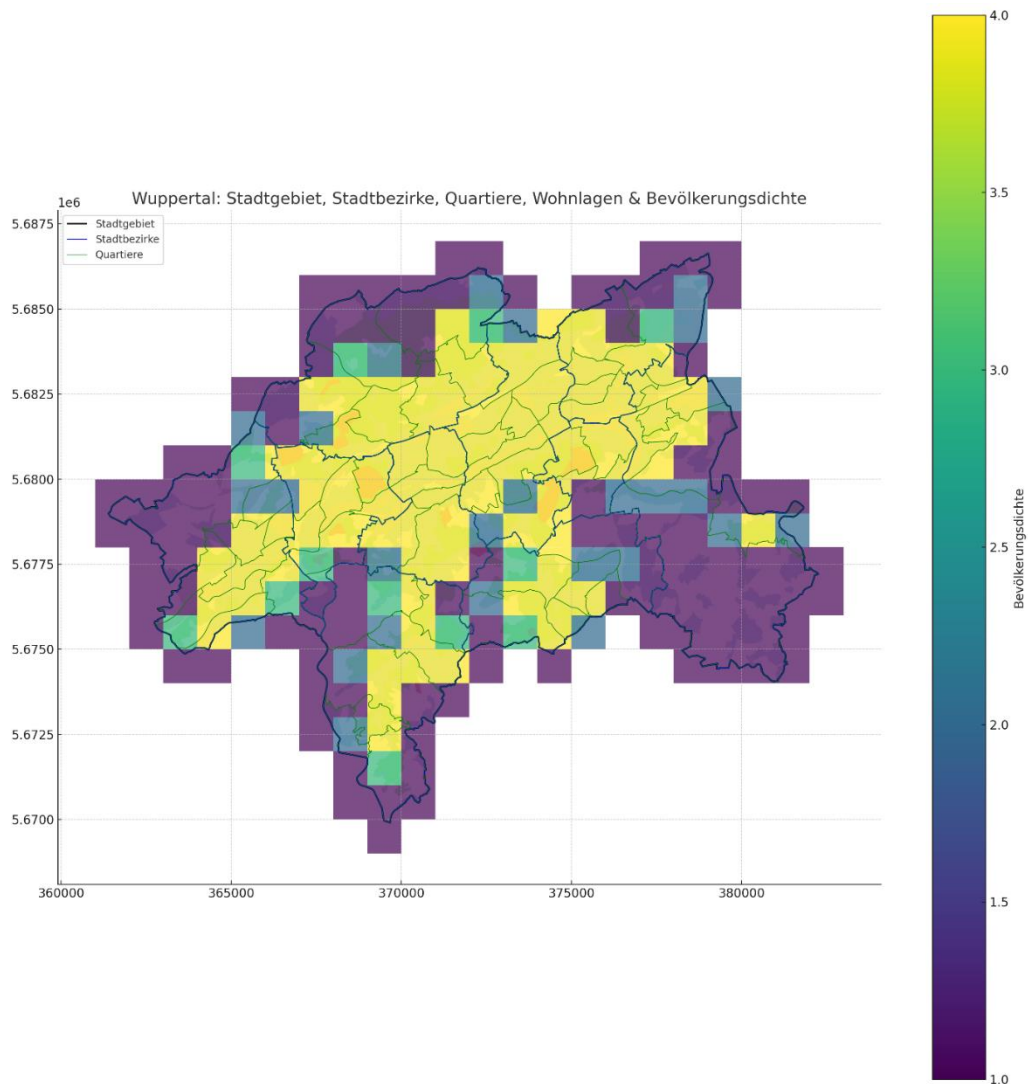
- Wohnlagen: Ich werde eine klare kategorische Farbpalette verwenden, um die verschiedenen Wohnlagen darzustellen.
- Bevölkerungsdichte: Ich werde die Bevölkerungsdichte als kontinuierliches Farbschema darstellen, jedoch mit einer anderen Farbpalette als zuvor.

Ich werde jetzt die Anpassungen vornehmen und Ihnen das aktualisierte Diagramm zeigen.

Hier ist die aktualisierte Visualisierung:

- Stadtgebiet: Dargestellt durch die schwarzen Linien
- Stadtbezirke: Dargestellt durch die blauen Linien
- Quartiere: Dargestellt durch die grünen Linien
- Wohnlagen: Pastellfarbene Schichten repräsentieren die verschiedenen Wohnlagen (z.B. "einfach" in Hellgrau, "mittel" in Hellblau, usw.)
- Bevölkerungsdichte: Die Bevölkerungsdichte wird durch verschiedene Schattierungen von Gelb bis Lila (je dunkler, desto dichter besiedelt) dargestellt.

Diese Farbpalette sollte die Unterscheidung zwischen Wohnlagen und Bevölkerungsdichte erleichtern. Lassen Sie mich wissen, ob dies Ihren Anforderungen entspricht oder ob Sie weitere Anpassungen wünschen!

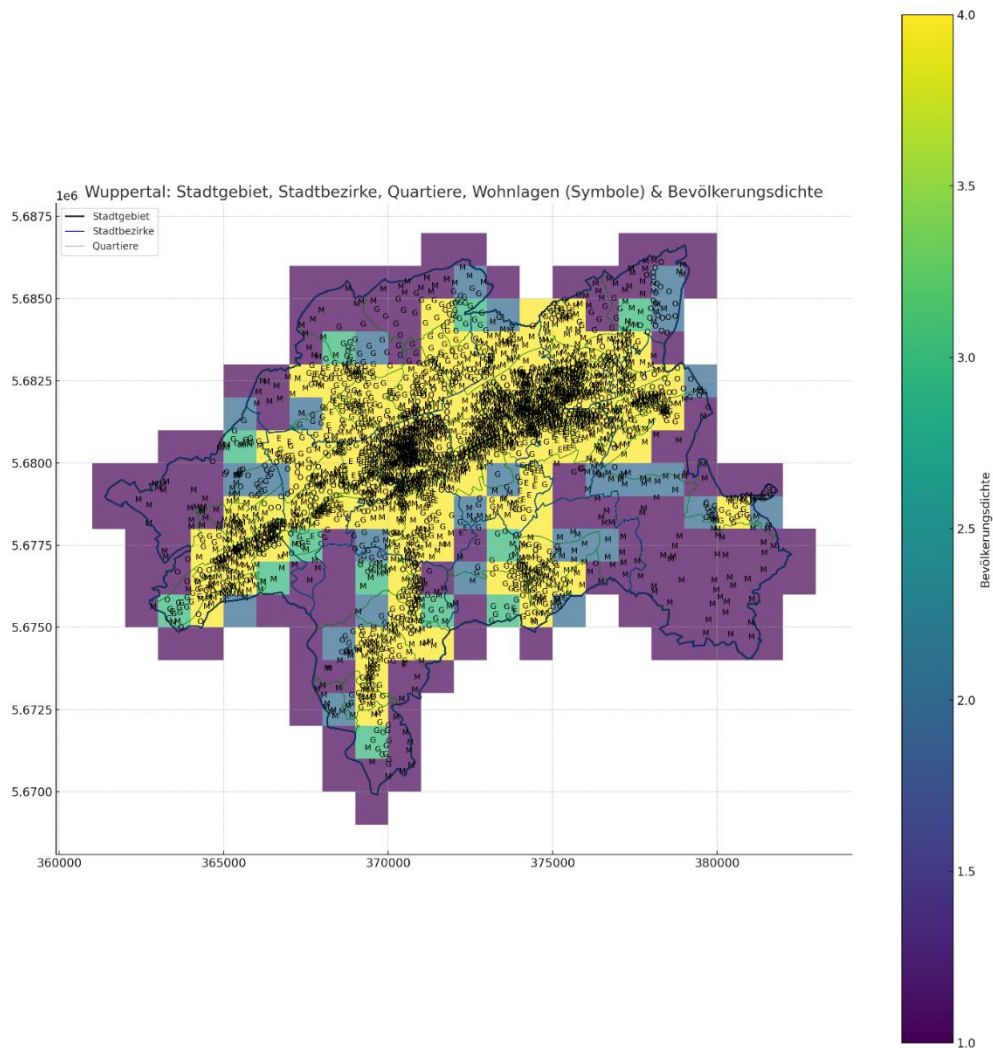


**Eingabe:** Bitte passe die Darstellung so an, dass du die Information über die Wohnlagen nicht mit Farben, sondern mit Symbolen darstellst (O für einfache, M für mittlere, G für gute und E für exklusive Wohnlage). Der Rest kann bleiben, wie er ist.

**ChatGPT:** Hier ist die aktualisierte Visualisierung:

- Stadtgebiet: Dargestellt durch die schwarzen Linien
- Stadtbezirke: Dargestellt durch die blauen Linien
- Quartiere: Dargestellt durch die grünen Linien
- Wohnlagen: Dargestellt durch Textsymbole (O, M, G, E) entsprechend den verschiedenen Wohnlagen (z.B. "O" für einfache Wohnlage, "M" für mittlere Wohnlage usw.)
- Bevölkerungsdichte: Die Bevölkerungsdichte wird durch verschiedene Schattierungen von Gelb bis Lila (je dunkler, desto dichter besiedelt) dargestellt.

Ich hoffe, dass diese Darstellung jetzt Ihren Anforderungen entspricht. Lassen Sie mich wissen, ob Sie weitere Anpassungen oder Informationen wünschen!



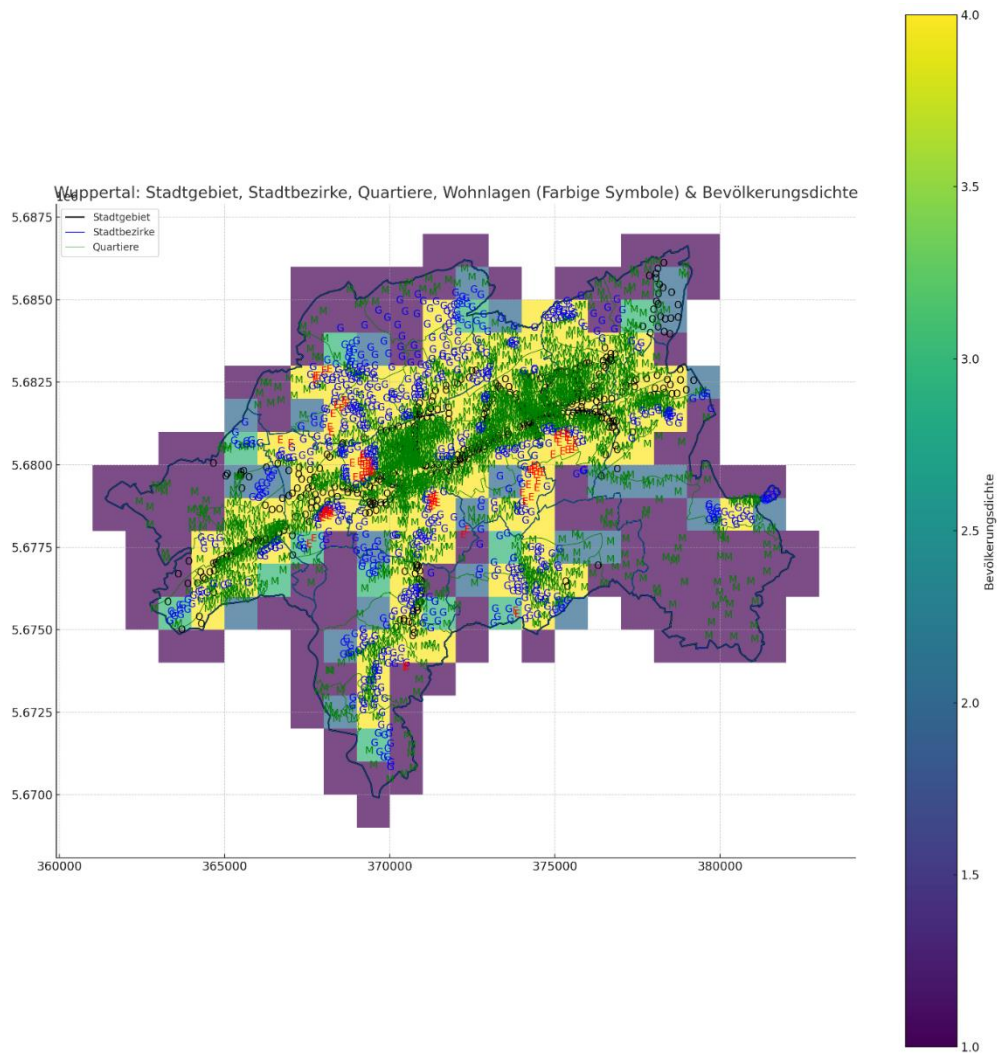
**Eingabe:** Ja, das ist viel besser. Bitte passe die Darstellung der Wohnlagen wie folgt an: schwarzes O für einfache, grünes M für mittlere, blaues G für gute und rotes E für exklusive Wohnlage.

**ChatGPT:** Hier ist die aktualisierte Visualisierung:

- Stadtgebiet: Dargestellt durch die schwarzen Linien
- Stadtbezirke: Dargestellt durch die blauen Linien
- Quartiere: Dargestellt durch die grünen Linien
- Wohnlagen: Dargestellt durch farbige Textsymbole (schwarzes "O" für einfache Wohnlage, grünes "M" für mittlere Wohnlage, blaues "G" für gute Wohnlage und rotes "E" für exklusive Wohnlage)
- Bevölkerungsdichte: Die Bevölkerungsdichte wird durch verschiedene Schattierungen von Gelb bis Lila (je dunkler, desto dichter besiedelt) dargestellt.

Ich hoffe, diese Darstellung entspricht nun genau Ihren Wünschen. Wenn Sie weitere Anpassungen oder Informationen benötigen, lassen Sie es mich bitte wissen!





**Eingabe:** Sehr gut, danke. Bitte ergänze in der Darstellung die Standorte der Schulen aus der beigefügten Datei. Es genügt dabei die Unterscheidung zwischen Grundschule (Standorte mit einer 1 kennzeichnen) und weiterführenden Schulen (Standorte als 2 kennzeichnen).

**ChatGPT:** Danke für die Daten. Ich werde nun die Standorte der Schulen zur bestehenden Visualisierung hinzufügen und dabei die gewünschte Unterscheidung zwischen Grundschulen (mit "1" gekennzeichnet) und weiterführenden Schulen (mit "2" gekennzeichnet) vornehmen.

Hier ist die aktualisierte Visualisierung, die nun auch die Standorte der Schulen zeigt:

- Stadtgebiet: Dargestellt durch die schwarzen Linien
- Stadtbezirke: Dargestellt durch die blauen Linien
- Quartiere: Dargestellt durch die grünen Linien
- Wohnlagen: Dargestellt durch farbige Textsymbole
- Bevölkerungsdichte: Dargestellt durch verschiedene Schattierungen von Gelb bis Lila
- Schulen: Grundschulen sind mit "1" und weiterführende Schulen mit "2" gekennzeichnet.

Ich hoffe, diese Darstellung ist hilfreich. Lassen Sie mich wissen, ob Sie weitere Anpassungen oder Informationen wünschen!



**Eingabe:** Bitte berechne die durchschnittliche Entfernung von der verschiedenen Wohnlagen zu Grundschulen und weiterführenden Schulen.

**ChatGPT:** Um die durchschnittliche Entfernung von den verschiedenen Wohnlagen zu den Schulen zu berechnen, werde ich die folgenden Schritte ausführen:

1. Die Zentroiden (Mittelpunkte) der Polygone für jede Wohnlage bestimmen.
2. Für jeden Zentroiden die Entfernungen zu allen Grundschulen und weiterführenden Schulen berechnen.
3. Die durchschnittliche Entfernung für jeden Wohnlagentyp zu Grundschulen und weiterführenden Schulen bestimmen.

Ich werde jetzt mit der Berechnung beginnen.

Hier sind die durchschnittlichen Entfernungen von den verschiedenen Wohnlagen zu Grundschulen und weiterführenden Schulen:

**Einfache Wohnlage:**

- Durchschnittliche Entfernung zu Grundschulen: ca. 5356 Meter
- Durchschnittliche Entfernung zu weiterführenden Schulen: ca. 4796 Meter

**Exklusive Wohnlage:**

- Durchschnittliche Entfernung zu Grundschulen: ca. 5055 Meter
- Durchschnittliche Entfernung zu weiterführenden Schulen: ca. 4527 Meter

**Gute Wohnlage:**

- Durchschnittliche Entfernung zu Grundschulen: ca. 5986 Meter
- Durchschnittliche Entfernung zu weiterführenden Schulen: ca. 5573 Meter

**Mittlere Wohnlage:**

- Durchschnittliche Entfernung zu Grundschulen: ca. 5619 Meter
- Durchschnittliche Entfernung zu weiterführenden Schulen: ca. 5111 Meter
- Diese Entfernungen geben einen allgemeinen Überblick über die durchschnittliche Nähe der verschiedenen Wohnlagen zu den Schulen in Wuppertal.

Bitte beachten Sie, dass diese Entfernungen auf den Mittelpunkten (Zentroiden) der Wohnlagen basieren und daher nur Näherungswerte sind.

Lassen Sie mich wissen, wenn Sie weitere Analysen oder Informationen wünschen!

**Eingabe:** Wie viele Schulen liegen in Bereichen, die nicht die höchste Bevölkerungsdichte (Klasse 4) aufweisen?

**ChatGPT:** Es gibt insgesamt 129 Schulen in den Bereichen, die nicht die höchste Bevölkerungsdichte (Klasse 4) aufweisen.

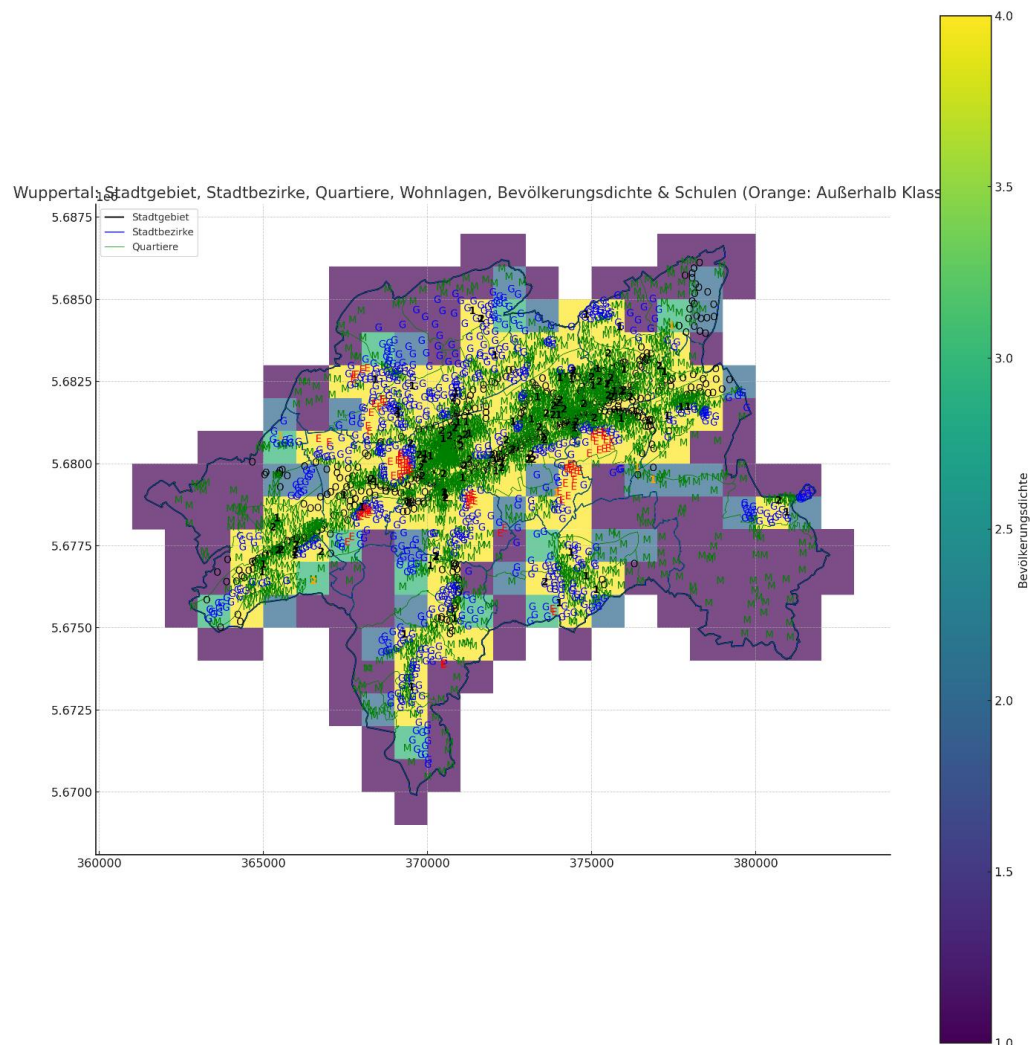
**Eingabe:** Bitte kennzeichne diese 129 Schulen, die nicht in den Gebieten mit der höchsten Bevölkerungsdichte liegen, in der Darstellung orange.

**ChatGPT:** Hier ist die aktualisierte Visualisierung:

- Stadtgebiet: Dargestellt durch die schwarzen Linien
- Stadtbezirke: Dargestellt durch die blauen Linien
- Quartiere: Dargestellt durch die grünen Linien
- Wohnlagen: Dargestellt durch farbige Textsymbole
- Bevölkerungsdichte: Dargestellt durch verschiedene Schattierungen von Gelb bis Lila

- Schulen: Grundschulen und weiterführende Schulen sind mit "1" bzw. "2" gekennzeichnet. Schulen, die nicht in Gebieten mit der höchsten Bevölkerungsdichte (Klasse 4) liegen, sind orange hervorgehoben.

Lassen Sie mich wissen, ob Sie weitere Anpassungen oder Informationen wünschen!



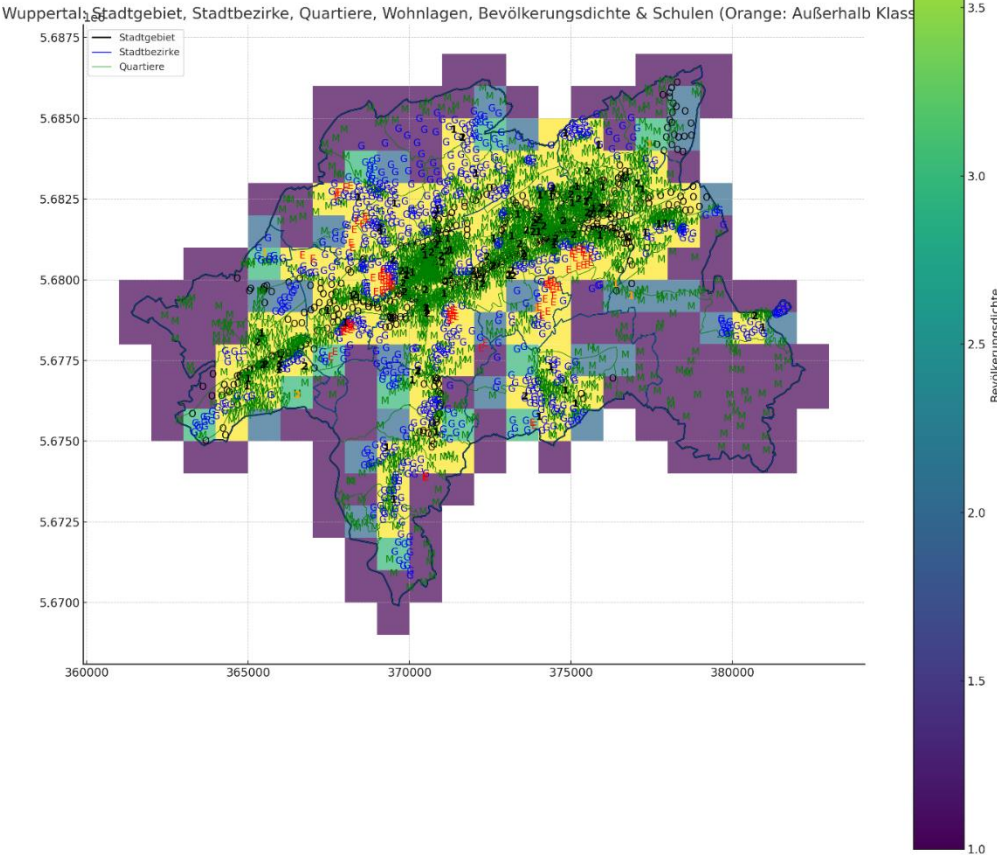
**Eingabe:** Ich zähle in der Darstellung nur etwa 4-5 Schulen, die nicht in den Gebieten mit der höchsten Bevölkerungsdichte liegen. Bist du sicher, dass es 129 sind?

**ChatGPT:** Es tut mir leid, das war ein Fehler meinerseits. Ich werde die Abfrage erneut durchführen und sicherstellen, dass die richtige Anzahl von Schulen, die sich nicht in Gebieten mit der höchsten Bevölkerungsdichte befinden, korrekt in der Darstellung hervorgehoben wird. Lassen Sie mich das korrigieren.

Es tut mir leid für die vorherige Verwirrung. Nach erneuter Überprüfung gibt es tatsächlich 7 Schulen, die sich nicht in Gebieten mit der höchsten Bevölkerungsdichte (Klasse 4) befinden. Ich werde die Darstellung jetzt aktualisieren und diese 7 Schulen orange hervorheben.

Hier ist die korrigierte Visualisierung. Die 7 Schulen, die sich nicht in Gebieten mit der höchsten Bevölkerungsdichte (Klasse 4) befinden, sind nun orange hervorgehoben.

Ich entschuldige mich für das vorherige Missverständnis. Lassen Sie mich wissen, ob Sie weitere Anpassungen oder Informationen wünschen!



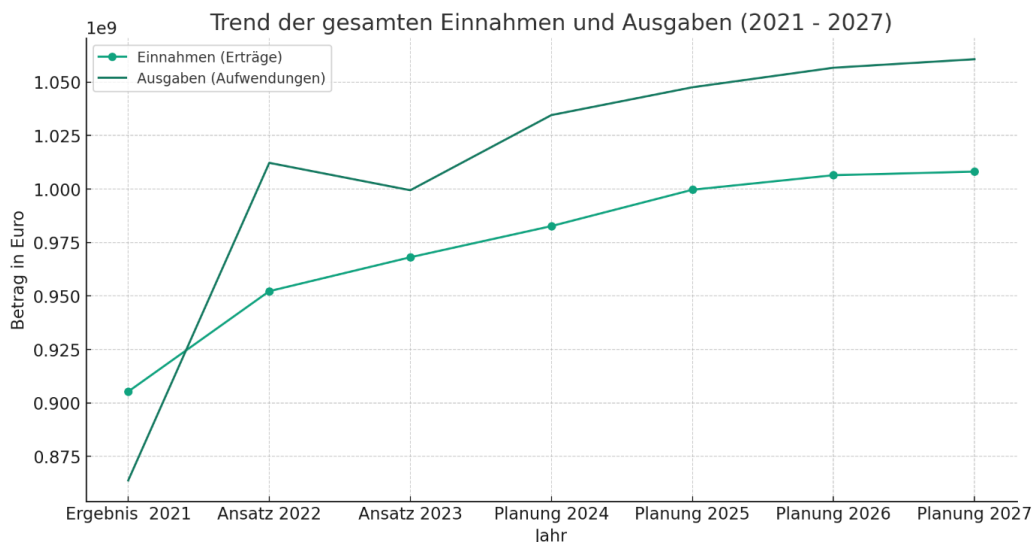
### 3. Haushaltsdaten (Ergebnishaushalt)

Eingabe: Lies bitte die folgende Datei ein. Das Trennzeichen ist ein Semikolon.

ChatGPT: Die Datei Ergebnis3.csv wurde erfolgreich eingelesen. Sie enthält folgende Spalten: [...] Was möchten Sie als Nächstes tun?

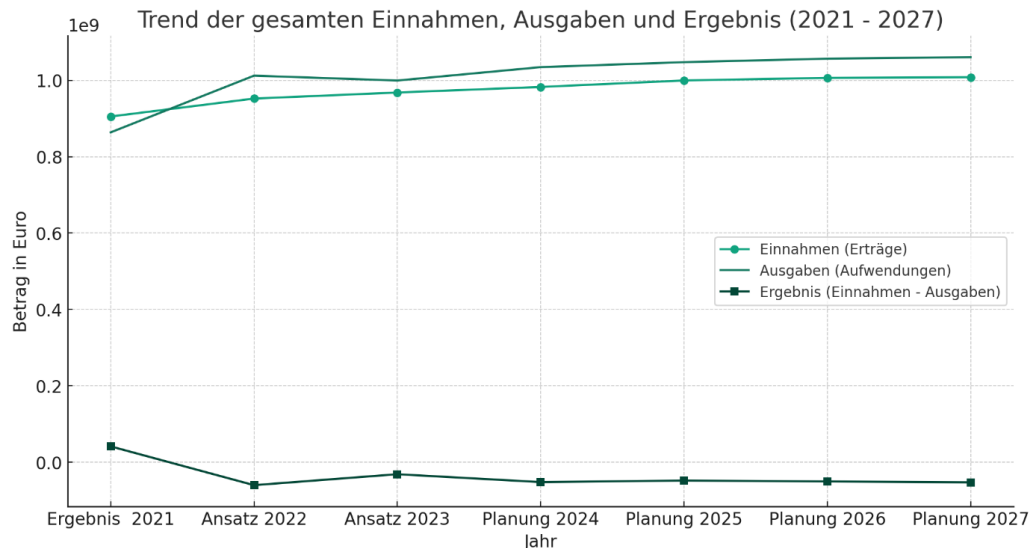
Eingabe: Es handelt sich um den Ergebnishaushalt (Ergebnis 2021 und Plan 2022-2027) einer deutschen Großstadt. Die ordentlichen und außerordentlichen Erträge sind die Einnahmen der Stadt und die ordentlichen und außerordentlichen Aufwendungen sind die Ausgaben der Stadt. Beide sind in der ersten Spalte kenntlich gemacht. Die zweite Spalte in der CSV-Datei zeigt die Unterkategorie an, die die Erträge/Aufwendungen verursachen. Die Spalte Sachkontobezeichnung gibt auf der Sachebene an, was die Erträge/Aufwendungen verursacht. Bitte stelle den Trend der gesamten Einnahmen (ordentliche + außerordentliche Einnahmen) und Ausgaben (ordentliche + außerordentliche Ausgaben) grafisch dar.

ChatGPT: Der Graph zeigt den Trend der gesamten Einnahmen (ordentliche + außerordentliche Erträge) und Ausgaben (ordentliche + außerordentliche Aufwendungen) der Stadt für die Jahre 2021 bis 2027. Was möchten Sie als Nächstes tun?



Eingabe: Bitte ergänze in der Abbildung das Ergebnis (Differenz zwischen Einnahmen und Ausgaben).

ChatGPT: Der aktualisierte Graph zeigt nun den Trend der gesamten Einnahmen (ordentliche + außerordentliche Erträge), Ausgaben (ordentliche + außerordentlichen Aufwendungen) sowie das Ergebnis (Differenz zwischen Einnahmen und Ausgaben) der Stadt für die Jahre 2021 bis 2027. Was möchten Sie als Nächstes tun?



**Eingabe:** Zeig mir, wie du die Differenz zwischen den Einnahmen (Erträge) und den Ausgaben (Aufwendungen) für die Jahre 2021 bis 2027 berechnet hast (mit Werten und Formel), anhand eines Beispieljahres.

**ChatGPT:** Für das Beispieljahr 2021 wurden die Differenz zwischen den Einnahmen (Erträge) und den Ausgaben (Aufwendungen) wie folgt berechnet:

- Gesamte Einnahmen (Erträge) für 2021: 905,321,828.0905,321,828.0 €
- Gesamte Ausgaben (Aufwendungen) für 2021: 863,793,040.0863,793,040.0 €

Die Differenz (Ergebnis) wird dann durch die folgende Formel berechnet:

$$\text{Ergebnis} = \text{Einnahmen (Erträge)} - \text{Ausgaben (Aufwendungen)}$$

Für das Jahr 2021 ergibt das:

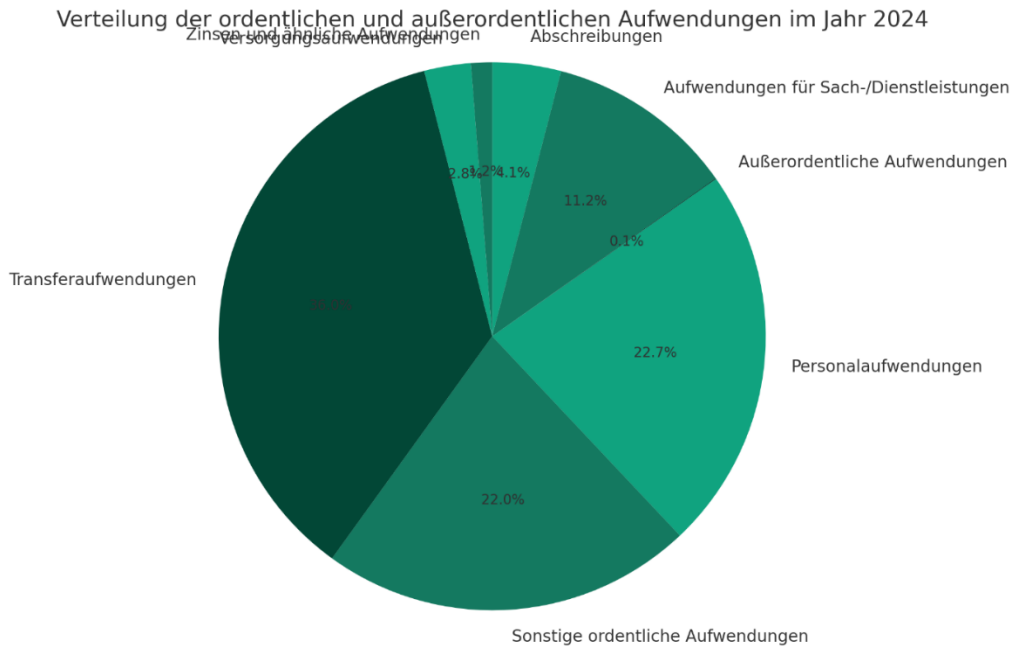
$$\text{Ergebnis für 2021} = 905,321,828.0 \text{ €} - 863,793,040.0 \text{ €} = 41,528,788.0 \text{ €}$$

Das Ergebnis für das Jahr 2021 beträgt also 41,528,788.0 €.

Möchten Sie noch weitere Informationen oder Analysen?

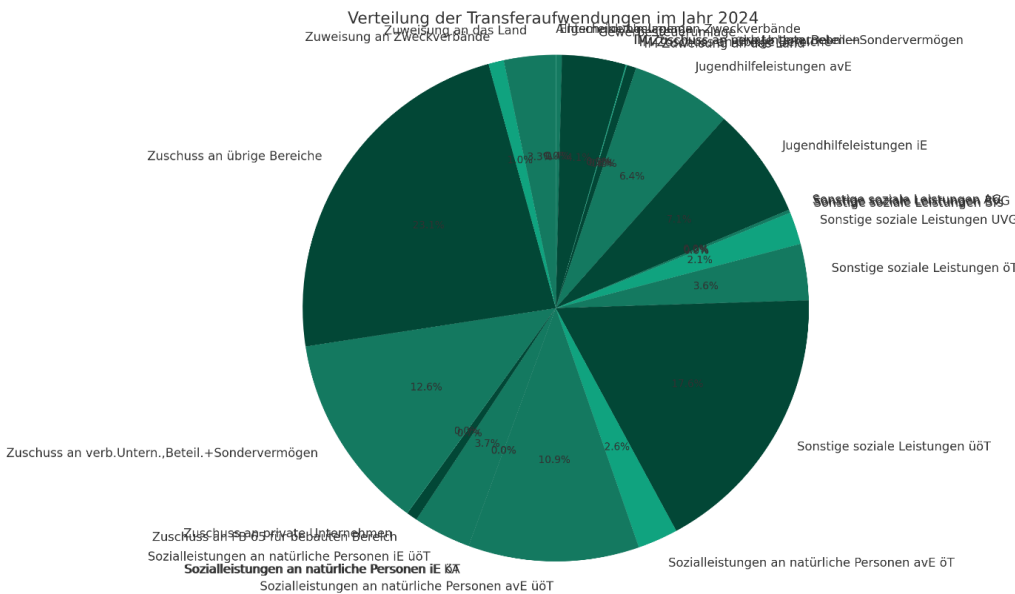
**Eingabe:** Bitte erstelle ein Kuchendiagramm aller Arten von ordentlichen und außerordentlichen Aufwendungen im Jahr 2024.

**ChatGPT:** Das Kuchendiagramm zeigt die Verteilung der verschiedenen Arten von ordentlichen und außerordentlichen Aufwendungen für das Jahr 2024. Jedes Segment repräsentiert einen Anteil der gesamten Aufwendungen, der einer bestimmten Unterkategorie zugeordnet ist. Was möchten Sie als Nächstes tun?



**Eingabe:** Bitte erstelle ein Kuchendiagramm aller Arten von Transferaufwendungen für das Jahr 2024.

**ChatGPT:** Das Kuchendiagramm zeigt die Verteilung der verschiedenen Arten von Transferaufwendungen für das Jahr 2024. Jedes Segment repräsentiert einen Anteil der gesamten Transferaufwendungen, der einer bestimmten Sachkontobezeichnung zugeordnet ist. Möchten Sie weitere Analysen durchführen oder haben Sie andere Fragen?



**Eingabe:** Stelle mir den prozentualen Anteil der "Grundsteuer B" (Spalte "Sachkontobezeichnung") an der Unterkategorie "Steuern und ähnliche Abgaben" und an der Kategorie "Ordentlichen Erträge" pro Jahr über den Zeitraum grafisch dar.

**ChatGPT:** Der Graph zeigt den prozentualen Anteil der "Grundsteuer B" an der Unterkategorie "Steuern und ähnliche Abgaben" sowie an der Kategorie "Ordentliche Erträge" für die Jahre 2021 bis 2027.

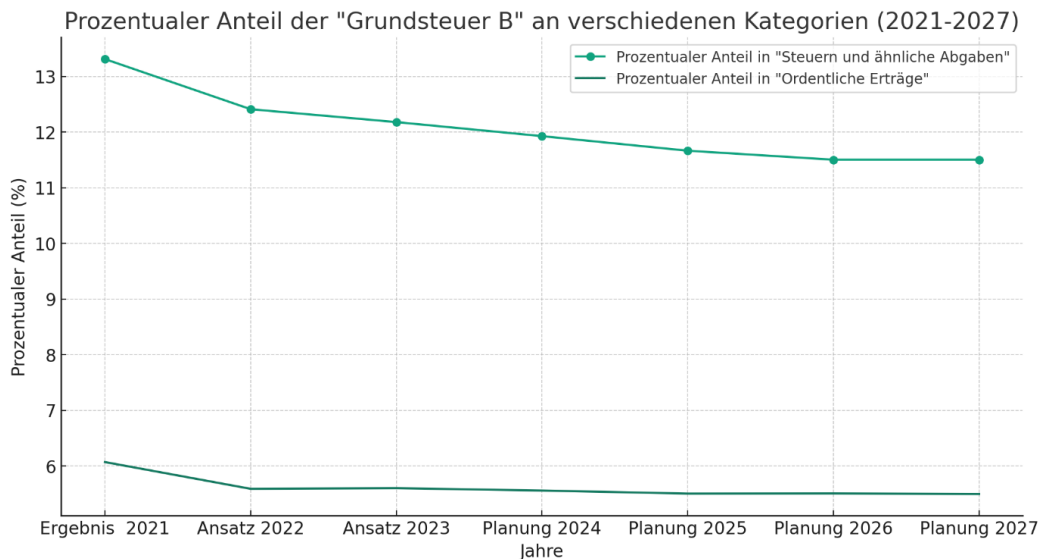
Die Linie mit den Kreismarkierungen zeigt den prozentualen Anteil der "Grundsteuer B" an "Steuern und ähnlichen".



Die Linie mit den Kreuzmarkierungen zeigt den prozentualen Anteil der "Grundsteuer B" an "Ordentliche Erträge".

Wie Sie sehen können, bleibt der prozentuale Anteil der "Grundsteuer B" in beiden Kategorien im betrachteten Zeitraum relativ stabil.

Haben Sie weitere Fragen oder benötigen Sie zusätzliche Informationen?



Eingabe: Wie bewertest du angesichts der Zahlen die finanzielle Lage der Stadt?

ChatGPT: Die finanzielle Differenz zwischen den Einnahmen (Erträgen) und Ausgaben (Aufwendungen) für die Jahre 2021 bis 2027 zeigt folgendes Bild:

- Im Jahr 2021 gibt es einen positiven Überschuss von ca. 41,5 Millionen Euro.
- Für die Jahre 2022 bis 2027 sind jedoch Defizite geplant, wobei das höchste Defizit für das Jahr 2022 mit fast 60 Millionen Euro erwartet wird.

Diese Zahlen deuten darauf hin, dass die Stadt in den kommenden Jahren vor finanziellen Herausforderungen stehen könnte, insbesondere wenn die geplanten Defizite nicht durch zusätzliche Einnahmen oder Kostensenkungen ausgeglichen werden.

Es ist wichtig zu beachten, dass diese Analyse eine grobe Einschätzung der finanziellen Lage darstellt und auf den verfügbaren Daten basiert. Eine detaillierte Analyse, die auch andere Faktoren wie Verschuldung, Investitionen und Liquidität berücksichtigt, wäre für eine vollständige Beurteilung notwendig.

Möchten Sie weitere Analysen oder Erklärungen?

Eingabe: Welche Einspar- bzw. Ertragserhöhungsoptionen würdest du der Stadt vorschlagen? Bitte argumentiere anhand konkreter Sachkonten (Spalte Sachkontobezeichnung) und entwickle zwei mögliche Szenarien. Bitte stelle für jedes Szenario jeweils die Vor- und Nachteile dar.

ChatGPT: Basierend auf den höchsten Ausgaben und Einnahmen für das Jahr 2021 schlage ich zwei Szenarien für Einsparungen bzw. Ertragserhöhungen vor:

**Szenario 1: Reduzierung der Personalkosten**

Reduzierung der Ausgaben für "Tariflich Beschäftigte Entgelt" und "Beamtenbezüge" durch Effizienzsteigerung oder Personalabbau.

**Vorteile:**

1. Schnelle Kostensenkung: Personalkosten sind oft ein großer Posten im Budget, und Einsparungen hier können einen schnellen Einfluss haben.

2. Langfristige Einsparungen: Einsparungen in diesem Bereich könnten sich in den Folgejahren fortsetzen.

**Nachteile:**

1. Moral und Produktivität: Kostensenkungen in diesem Bereich könnten die Mitarbeitermoral und Produktivität negativ beeinflussen.
2. Qualität der Dienstleistungen: Eine Reduzierung des Personals könnte zu einer Verringerung der Qualität öffentlicher Dienstleistungen führen.

**Szenario 2: Steigerung der Steuereinnahmen**

Erhöhung der "Grundsteuer B" und/oder der "Gewerbsteuer" durch Anpassung der Steuersätze.

**Vorteile:**

1. Sofortige Erhöhung der Einnahmen: Eine Steuererhöhung könnte zu einer schnellen Erhöhung der Einnahmen führen.
2. Langfristige Stabilität: Höhere Steuereinnahmen könnten für mehr finanzielle Stabilität in den kommenden Jahren sorgen.

**Nachteile:**

1. Unpopulär: Steuererhöhungen sind oft politisch unpopulär und könnten Widerstand in der Bevölkerung hervorrufen.
2. Wirtschaftliche Auswirkungen: Höhere Gewerbesteuern könnten für Unternehmen abschreckend wirken und langfristig die lokale Wirtschaft beeinträchtigen.

Beide Szenarien haben ihre eigenen Vor- und Nachteile, und es wäre sinnvoll, eine detaillierte Kosten-Nutzen-Analyse durchzuführen, bevor eine Entscheidung getroffen wird. Welches Szenario oder welche Kombination von beiden wäre Ihrer Meinung nach am besten geeignet?

**Eingabe:** Welche Einspar- bzw. Ertragserhöhungsoptionen würdest du der Stadt vorschlagen, wenn du nicht nur Personalkosten und Steuereinnahmen, sondern auch weitere mögliche Sachkonten berücksichtigst? Bitte argumentiere anhand konkreter Sachkonten (Spalte Sachkontobezeichnung) und entwickle zwei mögliche Szenarien. Bitte stelle für jedes Szenario jeweils die Vor- und Nachteile dar.

**ChatGPT:** Basierend auf weiteren hohen Ausgaben- und Einnahmekategorien schlage ich zwei zusätzliche Szenarien vor:

**Szenario 3: Reduzierung der Transferaufwendungen**

Senkung der Ausgaben für "Zuschuss an übrige Bereiche" und "Sonstige soziale Leistungen üöT".

**Vorteile:**

1. Unmittelbare Kostenreduktion: Einsparungen in diesen Bereichen könnten sofortige Auswirkungen auf das Budget haben.
2. Geringere politische Kosten: Im Vergleich zur Reduzierung der Personalkosten könnte diese Option politisch weniger heikel sein.

**Nachteile:**

1. Soziale Auswirkungen: Kürzungen bei sozialen Leistungen könnten die am meisten gefährdeten Bevölkerungsgruppen treffen.
2. Langfristige Kosten: Reduzierte Investitionen in soziale Programme könnten langfristig zu höheren Ausgaben führen (z.B. Gesundheitskosten).

**Szenario 4: Erhöhung der staatlichen Zuwendungen und Kostenerstattungen**

Steigerung der "Schlüsselzuweisungen vom Land" und "Erstattung Land" durch Verhandlungen oder bessere Ausnutzung von Förderprogrammen.

**Vorteile:**

1. Keine direkten Kosten für Bürger: Im Gegensatz zu Steuererhöhungen belastet diese Option die Bürger nicht direkt.
2. Verbesserte finanzielle Stabilität: Höhere staatliche Zuwendungen können zur langfristigen finanziellen Stabilität beitragen.

**Nachteile:**

1. Unsicherheit: Die Erhöhung staatlicher Zuwendungen ist oft unsicher und abhängig von politischen Entscheidungen.
2. Administrative Komplexität: Das Beantragen und Verwalten von Fördermitteln kann komplex und zeitaufwendig sein.

Diese Szenarien könnten einzeln oder in Kombination umgesetzt werden, abhängig von den spezifischen Zielen und Herausforderungen der Stadt. Jede Option sollte sorgfältig analysiert werden, um ihre möglichen Auswirkungen auf die finanzielle Gesundheit der Stadt und das Wohl ihrer Bürger zu verstehen.

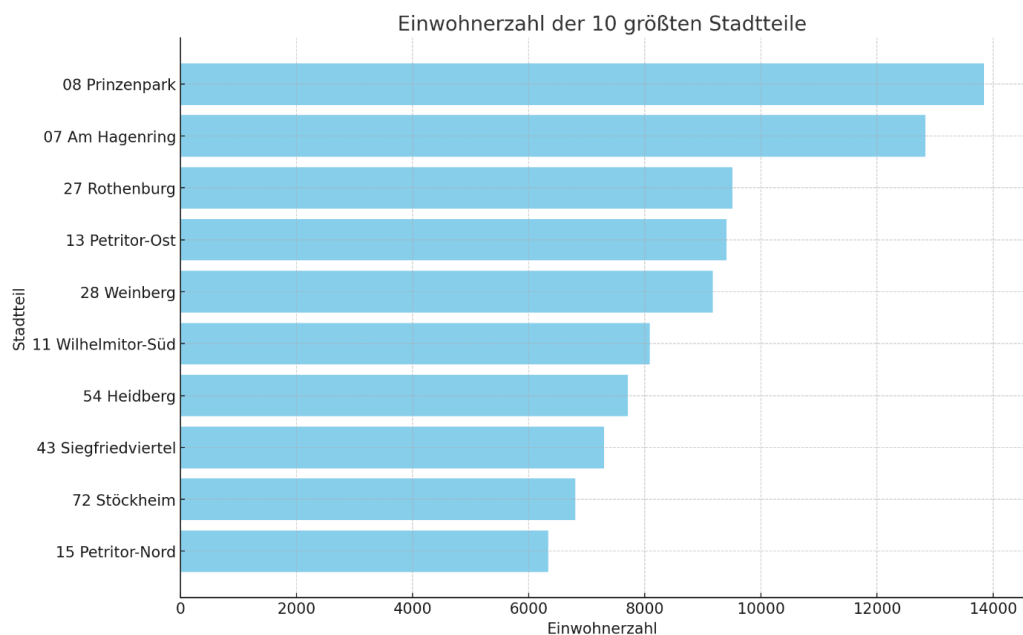
#### 4. Bevölkerungsdaten (Demografie)

**Eingabe:** Bitte lies die CSV-Datei ein. Versuche es mit der Codierung ISO-8859-1 und Semikolon als Trennzeichen.

**ChatGPT:** Es scheint, dass die CSV-Datei erfolgreich eingelesen wurde, aber die Ausgabe ist zu lang für die Darstellung hier. Möchten Sie eine spezielle Analyse durchführen oder bestimmte Informationen aus der Datei extrahieren?

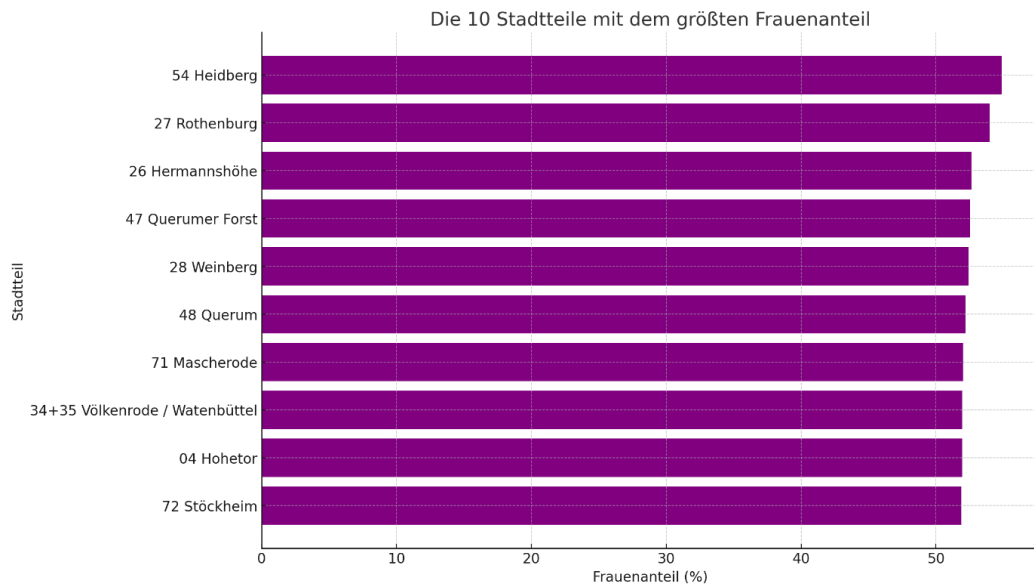
**Eingabe:** Bitte stelle die Einwohnerzahl der 10 größten Stadtteile grafisch dar (Spalte "Einwohnerzahl, insgesamt").

**ChatGPT:** Hier ist die grafische Darstellung der Einwohnerzahl der 10 größten Stadtteile. Wie Sie sehen können, variieren die Einwohnerzahlen deutlich zwischen den verschiedenen Stadtteilen.



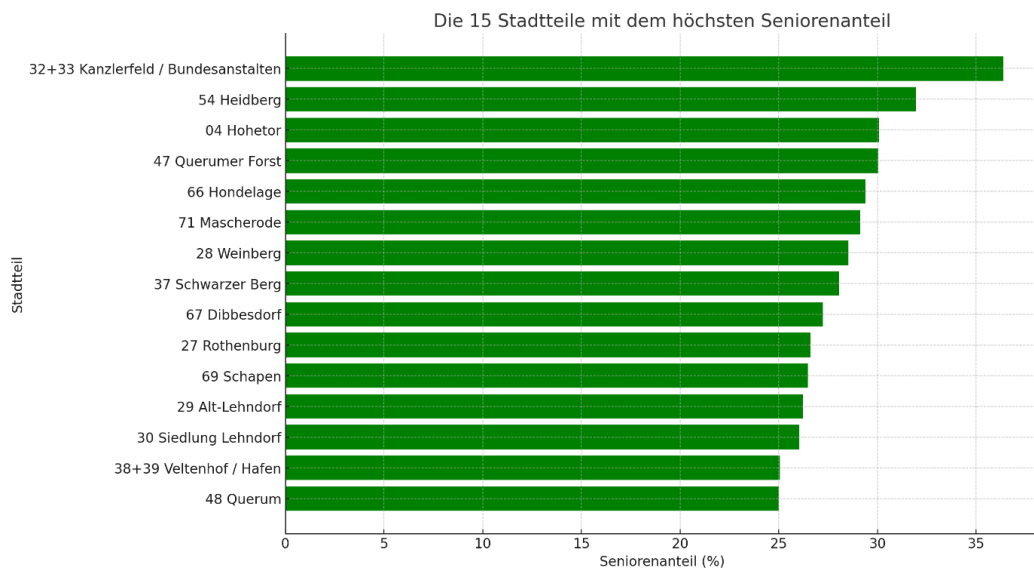
**Eingabe:** Bitte stelle die zehn Stadtteile mit dem größten Frauenanteil (in Prozent) grafisch dar (Spalte "Einwohnerzahl, weiblich" geteilt durch Spalte "Einwohnerzahl, insgesamt").

**ChatGPT:** Hier ist die grafische Darstellung der zehn Stadtteile mit dem größten Frauenanteil. Der Frauenanteil ist in Prozent angegeben und variiert zwischen den verschiedenen Stadtteilen.



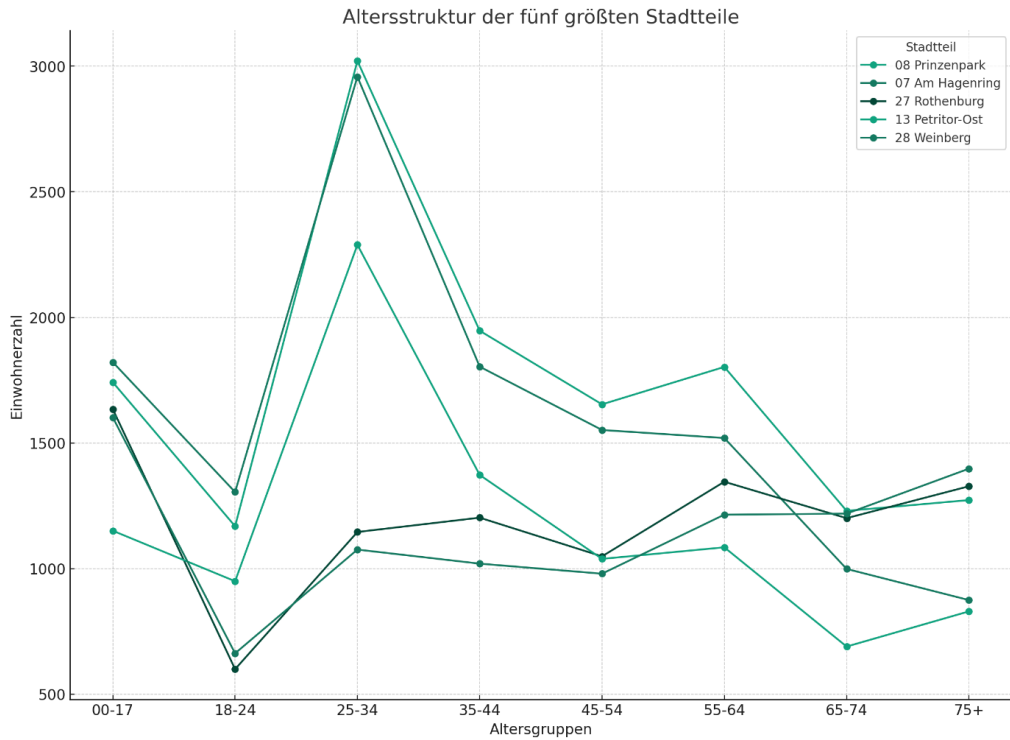
**Eingabe:** Welche Stadtteile haben den größten Anteil an Senioren (Spalte "Einwohnerzahl, insgesamt, 65 < 75 Jahre" + Spalte "Einwohnerzahl, insgesamt, 75 Jahre und älter" geteilt durch Spalte "Einwohnerzahl, insgesamt"). Bitte stelle die fünfzehn Stadtteile mit dem höchsten Seniorenanteil grafisch dar.

**ChatGPT:** Hier ist die grafische Darstellung der fünfzehn Stadtteile mit dem höchsten Seniorenanteil. Der Seniorenanteil ist in Prozent angegeben und zeigt deutliche Unterschiede zwischen den verschiedenen Stadtteilen.



**Eingabe:** Bitte stelle mit einem Liniendiagramm die Altersstruktur der fünf größten Stadtteile dar.

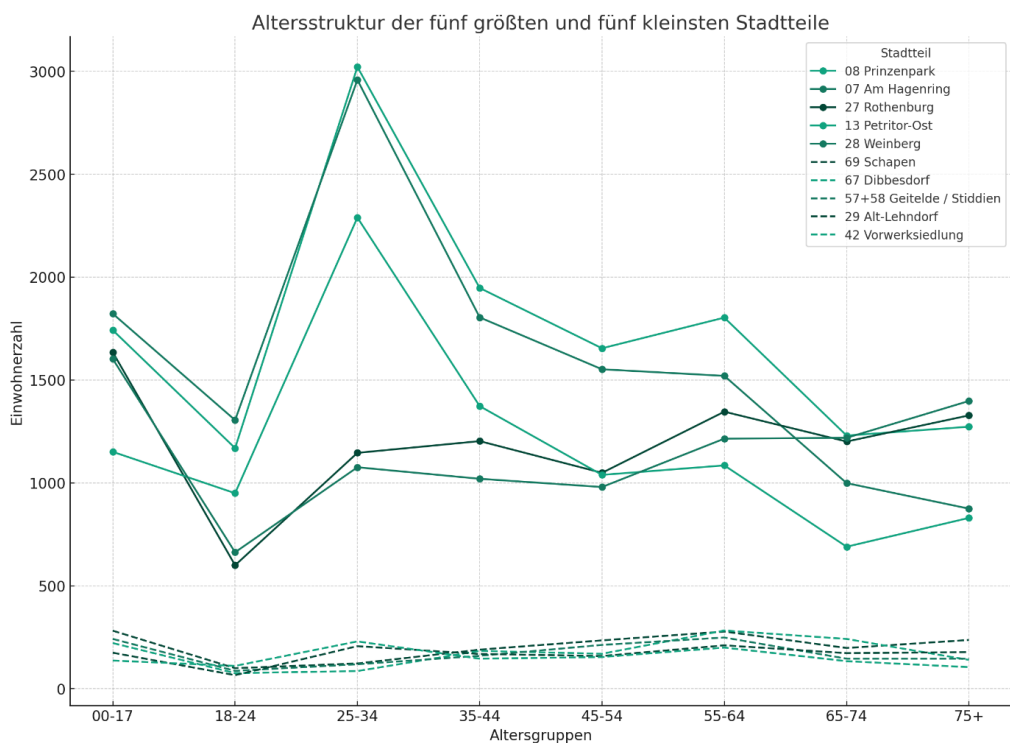
**ChatGPT:** Hier ist die grafische Darstellung der Altersstruktur der fünf größten Stadtteile. Die Altersgruppen sind auf der x-Achse und die Einwohnerzahl für jede Altersgruppe auf der y-Achse dargestellt. Jede Linie repräsentiert einen der fünf größten Stadtteile.



Wie Sie sehen können, gibt es Unterschiede in der Altersstruktur zwischen den verschiedenen Stadtteilen. Manche Stadtteile haben beispielsweise eine höhere Anzahl an jüngeren oder älteren Einwohnern.

**Eingabe:** Bitte stelle in der gleichen Grafik die Altersstruktur der fünf kleinsten Stadtteile dar und verwende zur optischen Unterscheidung einen anderen Farbsatz.

**ChatGPT:** Hier ist die aktualisierte grafische Darstellung der Altersstruktur, die sowohl die fünf größten als auch die fünf kleinsten Stadtteile berücksichtigt. Die Altersgruppen sind auf der x-Achse und die Einwohnerzahl für jede Altersgruppe auf der y-Achse dargestellt.



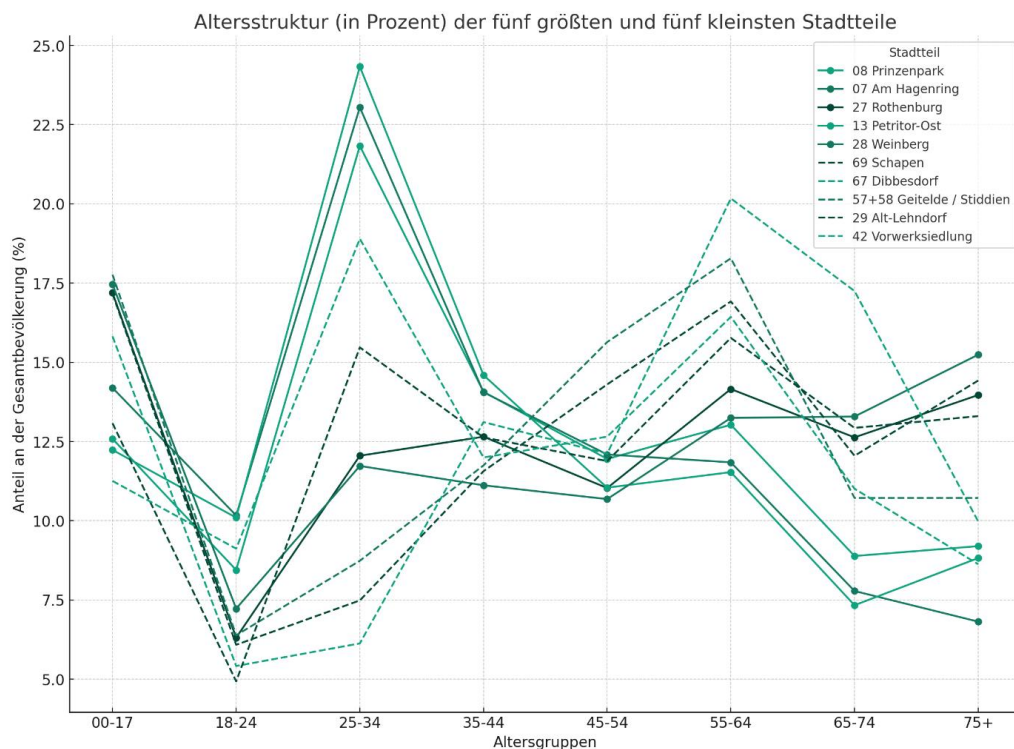


- Die durchgezogenen Linien mit Kreismarkierungen repräsentieren die fünf größten Stadtteile.
- Die gestrichelten Linien mit "X"-Markierungen repräsentieren die fünf kleinsten Stadtteile.

Wie Sie sehen können, zeigt die Grafik deutliche Unterschiede sowohl in der Größe als auch in der Altersstruktur zwischen den verschiedenen Stadtteilen.

**Eingabe:** Danke. Passe die Grafik nun so an, dass nicht die absoluten Einwohnerzahlen, sondern der Anteil der Altersgruppe an der Einwohnerzahl insgesamt im jeweiligen Stadtteil in Prozent dargestellt wird.

**ChatGPT:** Hier ist die aktualisierte Grafik, die den Anteil jeder Altersgruppe an der Gesamtbevölkerung des jeweiligen Stadtteils in Prozent darstellt.



- Die durchgezogenen Linien mit Kreismarkierungen repräsentieren die fünf größten Stadtteile.
- Die gestrichelten Linien mit "X"-Markierungen repräsentieren die fünf kleinsten Stadtteile.

Diese Darstellung ermöglicht es uns, die Altersstrukturen der verschiedenen Stadtteile in Relation zur Gesamtbevölkerung zu betrachten. Dadurch können interessante Muster und Unterschiede in der Altersverteilung erkannt werden.

### 5. Sensordaten (Passantenfrequenzen)

**Eingabe:** Die Datei enthält Informationen über die Passantenfrequenzen in Richtung der Braunschweiger Innenstadt und in Richtung aus der Innenstadt heraus. In den Spalten ist der jeweilige Messpunkt benannt, einmal in Richtung Innenstadt ("stadteinwärts") und aus der Innenstadt heraus ("stadtauswärts") und die Summe von beiden ("Gesamt"). Darüber hinaus gibt es noch Spalten für den Wetter-Status am jeweiligen Tag und die max. Temperatur, die vorgeherrscht hat.

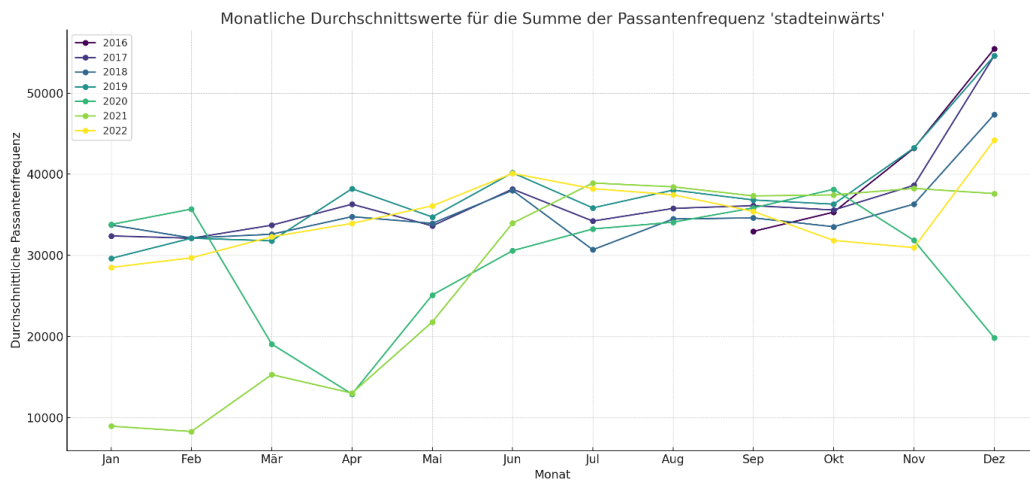
**ChatGPT:** Die Datei wurde erfolgreich mit einem Semikolon als Trennzeichen und der ISO-8859-1-Kodierung geladen. Jetzt können wir die Daten analysieren.

Von den angezeigten Daten können wir erkennen, dass es Spalten für verschiedene Standorte gibt, die die Passantenfrequenz in Richtung Innenstadt ("stadteinwärts"), aus der Innenstadt heraus ("stadtauswärts") und die Gesamtzahl ("Gesamt") anzeigen. Es gibt auch Spalten für das Datum, den Wetterstatus und die maximale Temperatur.

Wie möchten Sie fortfahren? Haben Sie spezielle Fragen oder Analysen, die Sie an diesen Daten durchführen möchten?

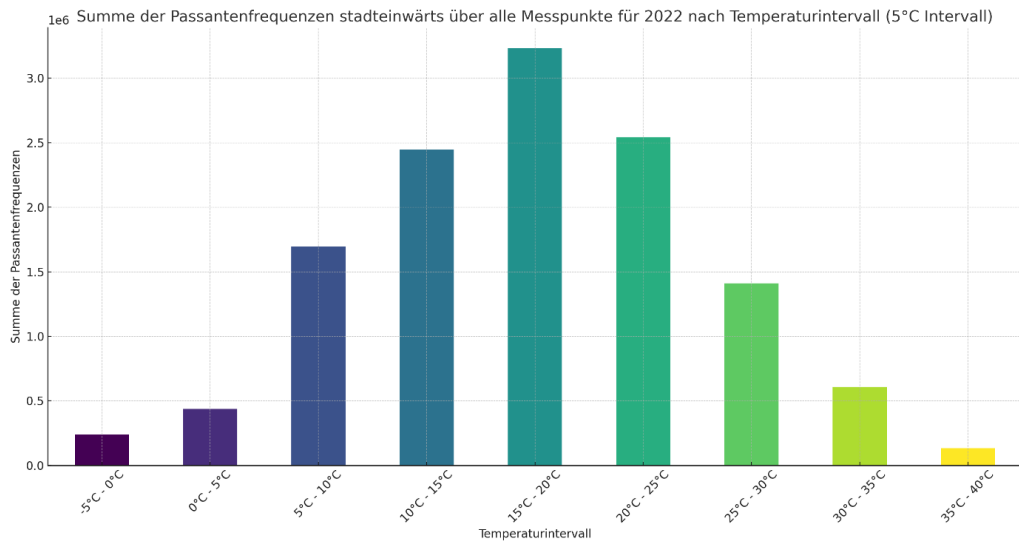
**Eingabe:** Erstelle bitte eine Darstellung, das die monatlichen Durchschnittswerte für die Summe der Passantenfrequenz "stadteinwärts" über alle Messpunkte zeigt. Jedes Jahr soll darin separat dargestellt sein.

**ChatGPT:** Hier ist die Darstellung der monatlichen Durchschnittswerte für die Summe der Passantenfrequenz "stadteinwärts" über alle Messpunkte. Jedes Jahr wird durch eine separate Linie dargestellt.



**Eingabe:** Danke. Erstelle bitte ein Balkendiagramm, aus dem die Summe der Passantenfrequenzen stadteinwärts über alle Messpunkte für das Jahr 2022 für die einzelnen Temperaturintervalle von 5°C dargestellt wird.

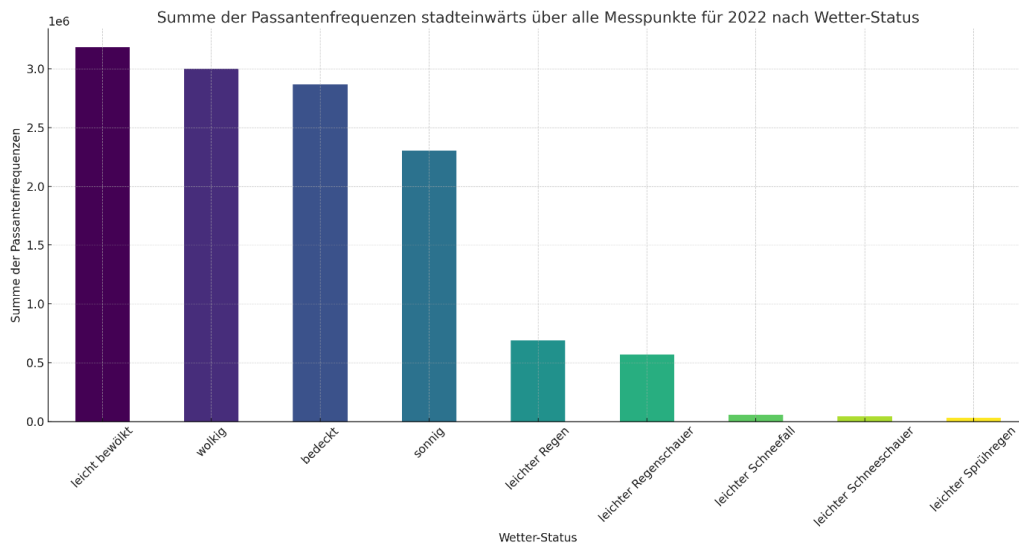
**ChatGPT:** Hier ist das Balkendiagramm, das die Summe der Passantenfrequenzen "stadteinwärts" über alle Messpunkte für das Jahr 2022 in Abhängigkeit von den Temperaturintervallen zeigt, wobei ein Intervall von 5°C verwendet wurde.



Lassen Sie mich wissen, ob dies Ihren Erwartungen entspricht oder ob Sie weitere Anpassungen wünschen!

**Eingabe:** Danke. Erstelle bitte nun ein Balkendiagramm, aus dem die Summe der Passantenfrequenzen stadteinwärts über alle Messpunkte für 2022 in Abhängigkeit des Wetter-Status hervorgeht.

**ChatGPT:** Hier ist das Balkendiagramm, das die Summe der Passantenfrequenzen "stadteinwärts" über alle Messpunkte für das Jahr 2022 in Abhängigkeit vom Wetter-Status zeigt. Die Balken sind nach der Passantenfrequenz absteigend sortiert.



Lassen Sie mich wissen, ob Sie weitere Analysen oder Fragen haben!

## 6. Umfragen (Innenstadtumfrage)

**Eingabe:** Bitte lies die CSV-Datei ein. Versuche es mit der Codierung ISO-8859-1 und Semikolon als Trennzeichen.

**ChatGPT:** Es scheint, dass die Datei erfolgreich eingelesen wurde. Möchten Sie einen bestimmten Teil der Daten sehen oder damit arbeiten?

**Eingabe:** Stell bitte zur Frage 1 dar, aus welchem Gründen männliche bzw. weibliche Besucher die Innenstadt aufsuchen. Verwende bitte ein Liniendiagramm.

**ChatGPT:** Natürlich, um dies zu tun, benötige ich jedoch einige Informationen:

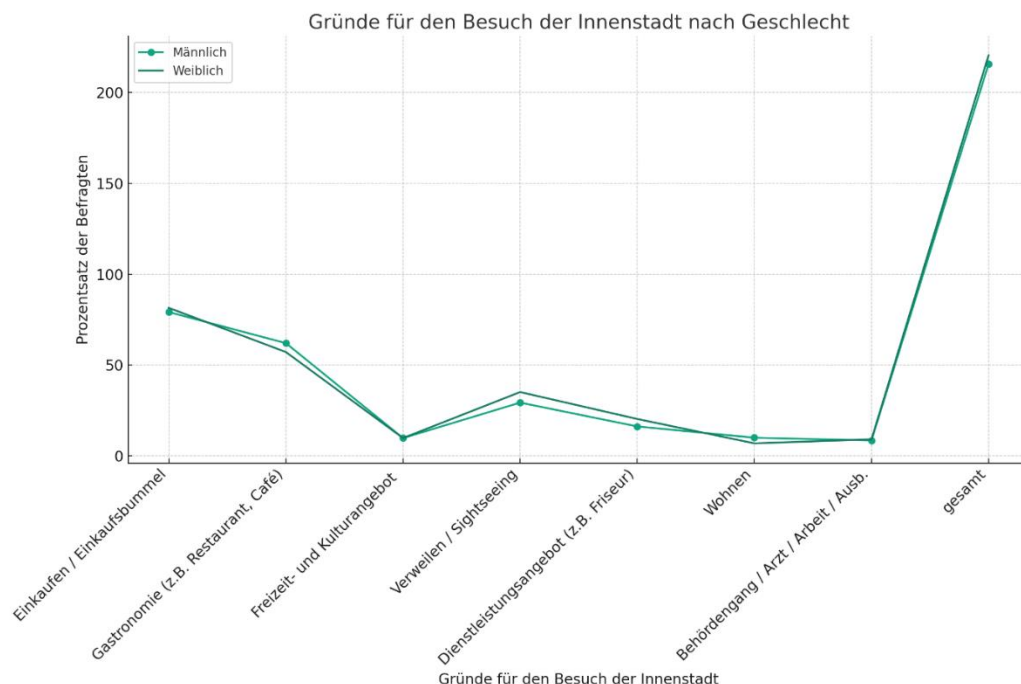
1. Könnten Sie mir die genauen Spaltennamen für die Frage 1 und das Geschlecht der Befragten mitteilen?

2. Sind die Daten in diesen Spalten kategorisch oder numerisch?

Durch diese Informationen kann ich die Daten besser visualisieren.

**Eingabe:** Du findest die Information zur Frage in Spalte 1. Die Daten sind numerisch (Prozent mit Mehrfachnennung).

**ChatGPT:** Vielen Dank für die zusätzlichen Informationen. Um die Frage zu beantworten, werde ich die Daten in der Spalte "Frage" filtern, um nur die Einträge für "Frage 1" zu erhalten. Dann werde ich die Daten für männliche und weibliche Befragte aus den Spalten "Befragte, männlich" und "Befragte, weiblich" verwenden. Lassen Sie uns nun das Liniendiagramm erstellen.

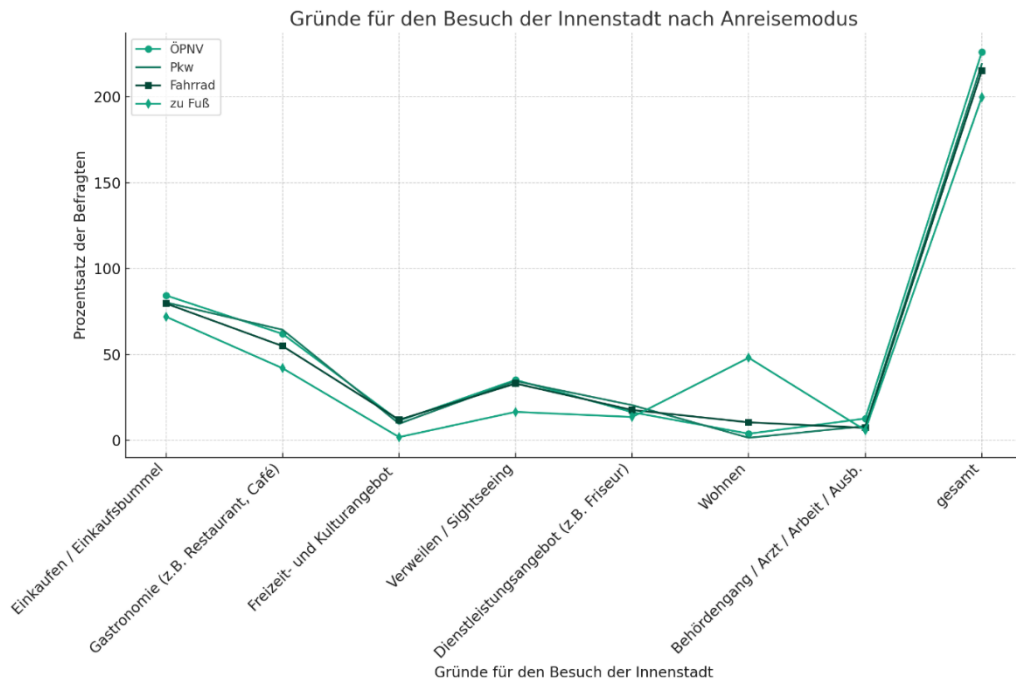


Das Liniendiagramm zeigt die Gründe, aus denen männliche und weibliche Besucher die Innenstadt besuchen, basierend auf der Frage 1 der Umfrage. Die Datenpunkte repräsentieren den Prozentsatz der Befragten für jede Antwortmöglichkeit. Möchten Sie weitere Analysen durchführen oder haben Sie Fragen zum Diagramm?

**Eingabe:** Danke. Erstell die gleiche Auswertung bitte nochmal, nur diesmal für die unterschiedlichen Anreisemodi (ÖPNV, Pkw, Fahrrad, zu Fuß).

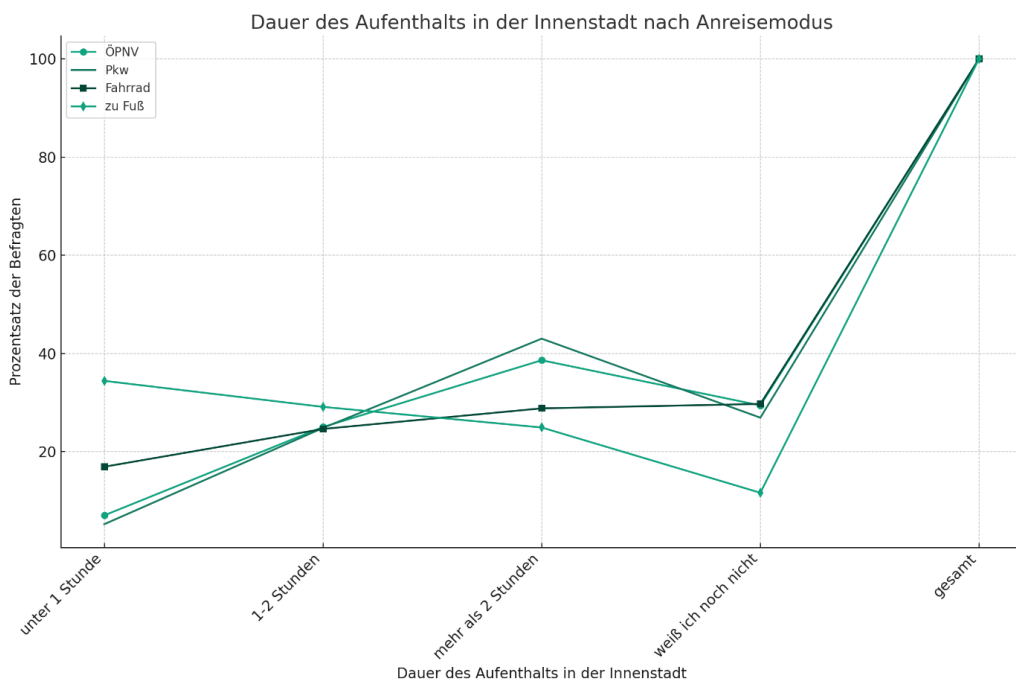
**ChatGPT:** Das Liniendiagramm zeigt die Gründe, aus denen Besucher mit verschiedenen Anreisemodi (ÖPNV, Pkw, Fahrrad, zu Fuß) die Innenstadt besuchen. Die Datenpunkte

repräsentieren den Prozentsatz der Befragten für jede Antwortmöglichkeit. Haben Sie weitere Fragen oder gibt es etwas anderes, bei dem ich Ihnen behilflich sein kann?



**Eingabe:** Danke. Erstell die Auswertung nochmal, nur diesmal nicht für die Gründe des Besuchs, sondern für die Dauer, die sich die Besucher in der Innenstadt aufhalten (s. Frage 3 in Spalte 1).

**ChatGPT:** Das Liniendiagramm zeigt die Dauer des Aufenthalts in der Innenstadt für Besucher, die mit verschiedenen Anreisemodi (ÖPNV, Pkw, Fahrrad, zu Fuß) gekommen sind. Die Datenpunkte repräsentieren den Prozentsatz der Befragten für jede Antwortmöglichkeit. Haben Sie weitere Fragen oder gibt es etwas anderes, bei dem ich Ihnen behilflich sein kann?



**Eingabe:** Schau dir nun bitte die Antworten auf alle vier Fragen an (s. Spalte 1) und erstelle einen kurzen Bericht (500 Wörter) zur Situation der Innenstadt und zu möglichen Entwicklungspotenzialen.

**ChatGPT:** **Bericht zur Situation der Innenstadt**

**Frage 1: Warum besuchen Menschen die Innenstadt?**

Die Hauptgründe für den Besuch der Innenstadt sind Einkaufen (80,4 % der Befragten), Gastronomie (59,6 %), und Verweilen bzw. Sightseeing (32,3 %). Diese Daten zeigen, dass die Innenstadt als Handels- und Freizeitzentrum wahrgenommen wird.

**Frage 2: Wie viele Einzelhandelsgeschäfte werden besucht?**

Leider wurden die Daten hierfür abgeschnitten. Dennoch ist diese Frage entscheidend, um das Einkaufsverhalten und die Attraktivität der Geschäfte in der Innenstadt zu verstehen.

**Frage 3: Wie lange halten sich die Menschen in der Innenstadt auf?**

Die Mehrheit der Menschen hält sich für etwa 1 bis 2 Stunden in der Innenstadt auf. Dies könnte auf ein kurzes, zielorientiertes Einkaufserlebnis hinweisen, das durch weitere Freizeitangebote ausgeweitet werden könnte.

**Frage 4: Schulnoten für die Innenstadt**

Die Innenstadt erhält größtenteils positive Schulnoten, mit einer Mehrheit der Befragten, die die Note "2" geben.

**Entwicklungspotenziale**

**1. Vielfalt der Angebote:** Da Einkaufen und Gastronomie die Hauptgründe für den Besuch sind, könnte eine Diversifizierung des Angebots, etwa durch kulturelle Events oder Freizeitmöglichkeiten, die Attraktivität erhöhen.

**2. Aufenthaltsdauer:** Maßnahmen, die die Aufenthaltsdauer verlängern, könnten das wirtschaftliche Potenzial der Innenstadt steigern. Denkbar wären mehr Sitzgelegenheiten, Unterhaltungsangebote oder sogar kostenlose Wi-Fi-Zonen.

**3. Mobilität:** Die Anreisemodi variieren, was die Notwendigkeit einer verbesserten Infrastruktur für ÖPNV, Fahrradwege und Parkmöglichkeiten unterstreicht.

**4. Kundenerfahrung:** Da die Innenstadt bereits gute Noten erhält, sollten Best Practices identifiziert und weiter ausgebaut werden, um eine noch bessere Kundenerfahrung zu bieten.

Insgesamt zeigt die Umfrage, dass die Innenstadt in vielen Bereichen gut abschneidet, aber durchaus Raum für Verbesserungen und Diversifizierung bietet.



## 7. Verkehrsdaten (Unfallstatistik)

**Eingabe:** Bitte lies die CSV-Datei ein. Versuche es mit der Codierung ISO-8859-1 und Semikolon als Trennzeichen.

**ChatGPT:** Die Datei wurde erfolgreich eingelesen. Sie enthält verschiedene Spalten mit numerischen Daten. Was möchten Sie als nächstes tun? Möchten Sie eine spezielle Analyse durchführen oder bestimmte Informationen aus den Daten extrahieren?

**Eingabe:** Es handelt sich um die Unfallstatistik der Stadt Köln 2018. Hier sind weitere Informationen zu den Spalten: [Metadaten wurden übergeben].

**ChatGPT:** Danke für die detaillierte Erklärung der Spalten. Die Daten scheinen eine umfassende Darstellung der Unfälle in Köln für das Jahr 2018 zu bieten. Welche Art von Analyse oder Information suchen Sie? Hier sind einige mögliche Fragen, die wir untersuchen könnten:

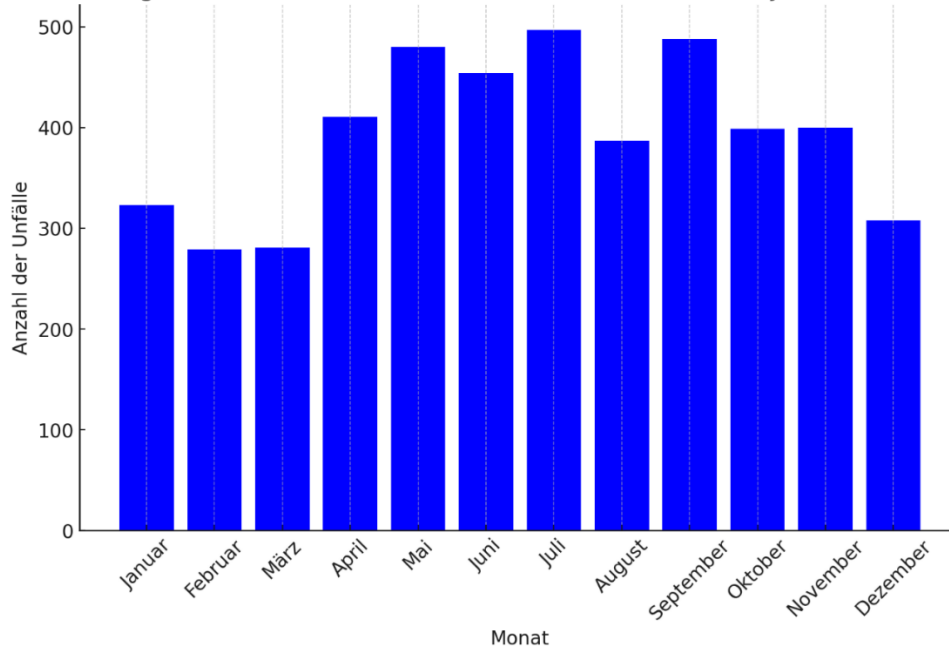
1. Wie ist die Verteilung der Unfälle über die verschiedenen Monate des Jahres?
2. Zu welchen Tageszeiten treten die meisten Unfälle auf?
3. Welche Unfallarten sind am häufigsten?
4. Wie oft sind bestimmte Verkehrsteilnehmer (wie Fahrradfahrer, Fußgänger, PKWs, etc.) an Unfällen beteiligt?
5. Unter welchen Licht- und Straßenbedingungen treten die meisten Unfälle auf?

Fühlen Sie sich frei, Ihre eigene Frage zu stellen oder eine der oben genannten auszuwählen.

**Eingabe:** Wie ist die Verteilung der Unfälle über die verschiedenen Monate des Jahres?

**ChatGPT:** Die Grafik zeigt die Verteilung der Unfälle über die verschiedenen Monate des Jahres 2018 in Köln. Wie man sehen kann, gibt es einige Variationen im Laufe des Jahres, aber es gibt keinen klaren Trend, der auf einen bestimmten Monat als besonders unfallträchtig hinweist.

Verteilung der Unfälle über die verschiedenen Monate des Jahres 2018 in Köln

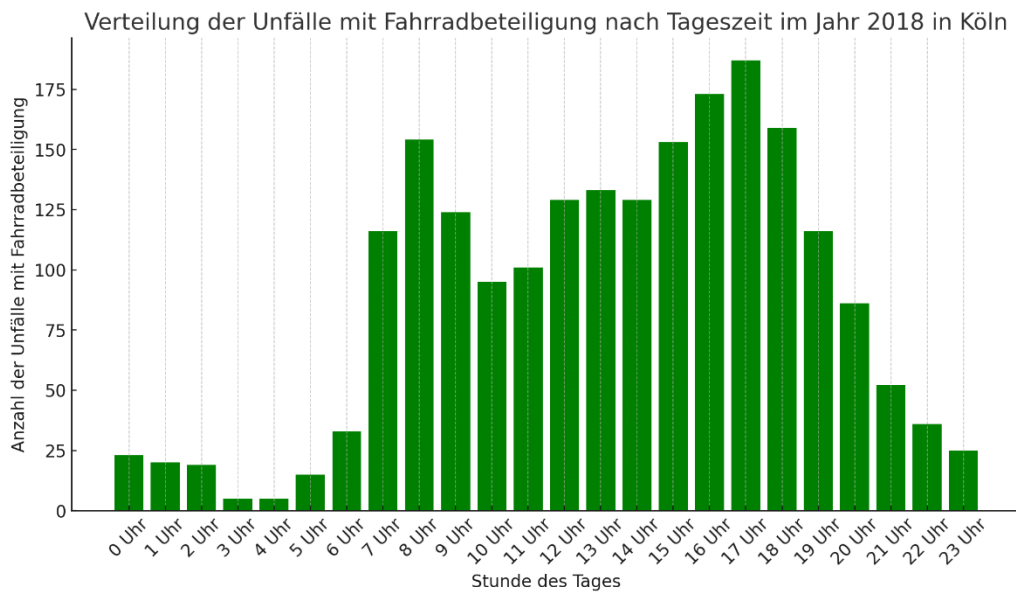


Möchten Sie weitere Analysen durchführen oder haben Sie weitere Fragen zu den Daten?

**Eingabe:** Zu welchen Tageszeiten treten die meisten Unfälle mit Fahrradbeteiligung auf?

**ChatGPT:** Die Grafik zeigt die Verteilung der Unfälle mit Fahrradbeteiligung nach Tageszeit für das Jahr 2018 in Köln. Einige interessante Beobachtungen:

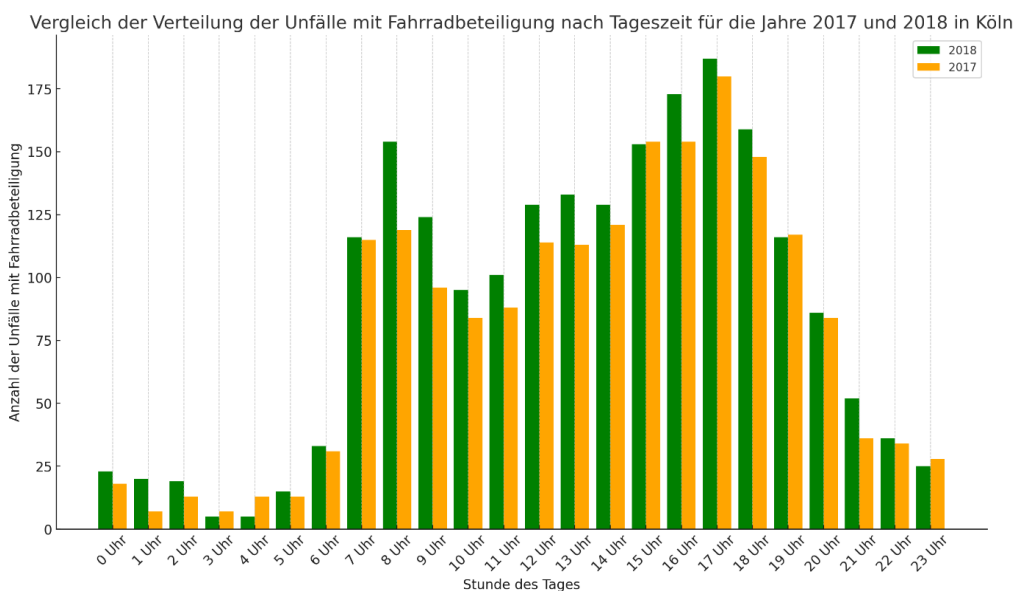
- Die meisten Unfälle mit Fahrradeteiligung treten in den späten Nachmittagsstunden auf, insbesondere zwischen 16 und 19 Uhr.
- In den frühen Morgenstunden gibt es deutlich weniger Unfälle, was wahrscheinlich auf den geringeren Verkehr zu dieser Zeit zurückzuführen ist.



Möchten Sie weitere Analysen durchführen oder haben Sie weitere Fragen?

**Eingabe:** Hier sind noch die Daten für das Jahr 2017. Ich würde die beiden Datensätze gerne vergleichen.

**ChatGPT:** Die Grafik vergleicht die Verteilung der Unfälle mit Fahrradeteiligung nach Tageszeit für die Jahre 2017 und 2018 in Köln.



Hier sind einige Beobachtungen:

- In beiden Jahren treten die meisten Unfälle mit Fahrradeteiligung in den späten Nachmittagsstunden auf, insbesondere zwischen 16 und 19 Uhr.
- Für das Jahr 2018 scheint es mehr Unfälle mit Fahrradeteiligung in den Abendstunden zu geben, verglichen mit dem Jahr 2017.

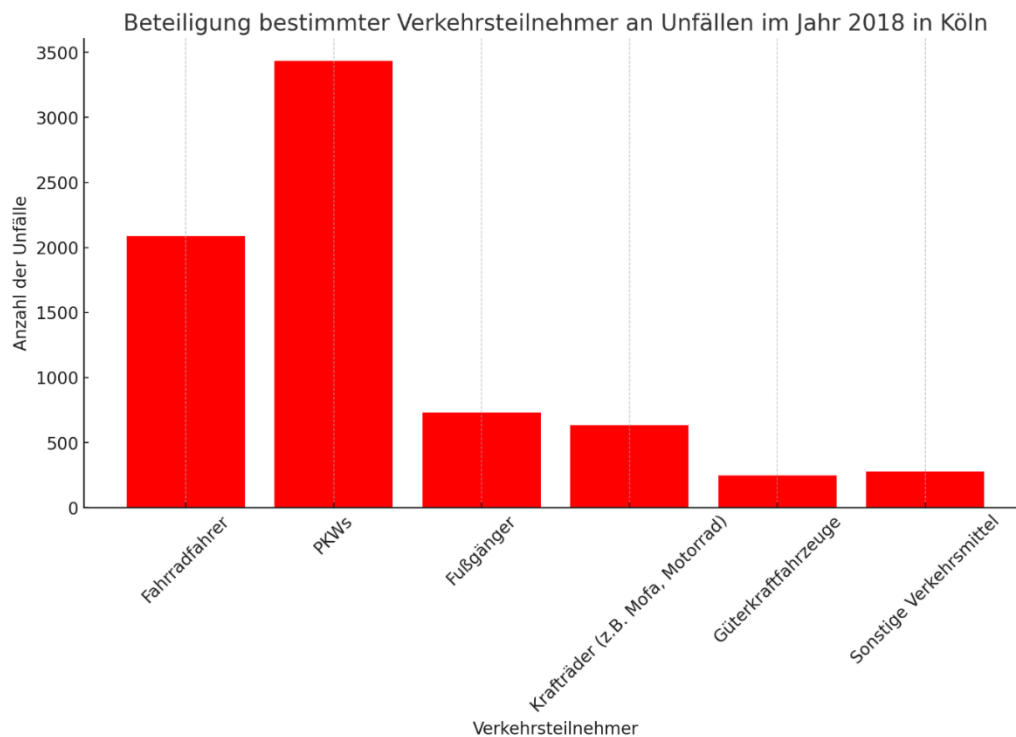
Es ist interessant zu sehen, dass die allgemeine Tendenz über die beiden Jahre ähnlich ist, obwohl es einige Unterschiede in der genauen Verteilung gibt.

Möchten Sie weitere Aspekte der Daten vergleichen oder haben Sie weitere Fragen?

**Eingabe:** Wie oft waren im Jahr 2018 bestimmte Verkehrsteilnehmer (wie Fahrradfahrer, Fußgänger, PKWs, etc.) an Unfällen beteiligt?

**ChatGPT:** Die Grafik zeigt die Beteiligung bestimmter Verkehrsteilnehmer an Unfällen im Jahr 2018 in Köln:

- PKWs waren am häufigsten an Unfällen beteiligt.
- Fahrradfahrer und Fußgänger folgen mit einer signifikanten Anzahl an Unfällen.
- Krafträder (z.B. Mofas, Motorräder) und Güterkraftfahrzeuge sind weniger häufig an Unfällen beteiligt.
- Sonstige Verkehrsmittel, wie Busse oder Straßenbahnen, sind am seltensten an Unfällen beteiligt.



### III. Diskussion der Ergebnisse und Potenziale

#### 1. Dialogbasierte Datenanalyse

Eine Besonderheit an GPT-4 ist die Art und Weise, wie die KI mit ihrem Nutzer interagiert. Computerprogramme auf KI-Basis stehen nicht erst seit der Ankündigung von GPT-4 zur Verfügung.<sup>1</sup> Bislang war es aber oftmals erforderlich, Programmiersprachen zu beherrschen oder Eingaben zu tätigen, für die Fachkenntnisse auf Seiten der Nutzer erforderlich waren. GPT-4 dagegen ist ein multimodales Modell, das über ein umfangreiches Wissen verfügt und als Eingabe natürliche Sprache akzeptiert.

Die Eingabe in natürlicher Sprache ermöglicht es, Daten dialogbasiert auszuwerten und darzustellen, ohne dass Kenntnisse im Umgang mit spezieller Software erforderlich sind. Dadurch sinken nicht nur die Anforderungen an die mit der Datenanalyse befassten Mitarbeiterinnen und Mitarbeiter, sondern Analysen können auch in wesentlich kürzerer Zeit erfolgen. Effizienzsteigerungen durch KI könnten bei standardisierten Verwaltungsaufgaben und Transaktionsprozessen realisiert werden, deren Bewältigung auf regelbasierten Arbeitsschritten in stark formalisierten Umgebungen beruht.<sup>2</sup> Zugleich ist davon auszugehen, dass neuer Bedarf an menschlicher Arbeit bzw. eine Vielfalt neuer Betätigungsfelder für menschliche Mitarbeiterinnen und Mitarbeiter entstehen wird, wenn vermehrt Aufgaben an KI-Systeme abgegeben werden.<sup>3</sup>

Im Alltag der Verwaltung eröffnet sich die Möglichkeit, Datenauswertungen auch an (noch) nicht ausreichend qualifizierte oder erfahrene Mitarbeiterinnen und Mitarbeiter zu übergeben. Die Potenziale werden insbesondere dann deutlich, wenn es um die Analyse von nicht zusammenhängenden Daten in verschiedenen Datenformaten geht, da solche Auswertungen im Regelfall nicht ohne fortgeschrittene IT-Kenntnisse möglich sind. Im Rahmen der Analyse von Fluglärmkarten (II.1) genügte es beispielsweise, dem System zu sagen, was von ihm verlangt wird (z. B. „Könntest du die Karte als Overlay auf eine Stadtkarte legen, wenn ich dir die Stadtkarte hochlade?“). Es waren weder Fachtermini noch Programmierkenntnisse erforderlich.

#### 2. Bedeutung der Input-Qualität („Prompts“)

Gleichzeitig ist festzustellen, dass die Qualität der Eingaben und Anfragen (sogenannte „Prompts“) die Qualität der Ausgaben der KI bestimmt. Damit Nutzer das Ergebnis erhalten, das sie tatsächlich angefordert haben, ist ein Grundverständnis der dialogbasierten Kommunikation mit generativen Sprachmodellen hilfreich. Auch wenn die Anforderungen hier geringer sind als etwa zum Erlernen einer Programmiersprache, handelt es sich doch um Kenntnisse, die wesentlich dazu beitragen, die Chancen der KI-basierten Analyse auszuschöpfen. Für Verwaltungen bedeutet das unter anderem, dass sie die Einführung des KI-unterstützten Arbeitens mit entsprechenden Schulungen und Weiterbildungen begleiten sollten.

Im Hinblick auf die Qualität der Prompts zeigen die Testdialoge beispielsweise, dass die Qualität der Analyse zunimmt, wenn der KI zusätzliche Hinweise über die auszuwertenden Daten mitgeteilt werden, die sich nicht aus dem Datensatz selbst oder aus der Metadatenbeschreibung ergeben. Das kann beispielsweise Hintergrundinformationen zum Kontext der Analyse oder Hinweise auf Unregelmäßigkeiten im Datensatz oder seine Struktur betreffen. Bei der Auswertung von Haushaltsdaten (II.3) erhöhte sich die Qualität der Analyse durch die KI beispielsweise deutlich, nachdem ihm erläutert wurde, um was für Daten es sich handelt und wie die Spalten zu interpretieren sind.

---

<sup>1</sup> Josipovic, Bing Chat/GPT-4 in der öffentlichen Verwaltung, Arbeitspapier, 31.05.2023, S. 2–4.

<sup>2</sup> BT-Drs. 20/3651, S. 23.

<sup>3</sup> Albrecht, ChatGPT und andere Computermodelle zur Sprachverarbeitung, S. 53.

### 3. Mensch-Maschine-Interaktion

Der Aspekt der Benutzerfreundlichkeit wird deutlich, indem das System Unregelmäßigkeiten in den eingelesenen Dateien teilweise selbständig erkennt, behebt und den Datensatz damit in eine für die Analyse geeignete Form bringt. Wenn das System bei der Datenauswertung zusätzliche Informationen berücksichtigt, kann sich der Nutzen der Ergebnisse für die jeweilige Problem- oder Fragestellung erhöhen. Beim Beispiel der Analyse von Umfrageergebnissen (II.6) hat das System von sich aus weitere Informationen angefordert. Es wollte unter anderem wissen, um welche Spalten es sich handelt und ob die Daten kategorisch oder numerisch sind.

Die KI kann Daten, entweder in ihrer Rohfassung oder im Rahmen einer Analyse, auch interpretieren und Handlungsempfehlungen ableiten. Hier bietet sich dem Nutzer die Möglichkeit, je nachdem wie offen die Anfrage formuliert wird, eine Einschätzung des Systems zu erhalten, die nicht nur geringen menschlichen Voreingenommenheiten (Biases) unterliegt, sondern auch den Wissensstand der KI berücksichtigt.<sup>4</sup> Der Nutzer kann die Interpretation mit seinen eigenen Überlegungen und Schlussfolgerungen abgleichen und dadurch sicherstellen, dass seine Schlussfolgerungen logisch, faktenbasiert und unvoreingenommen sind.<sup>5</sup> Die Vorteile des Einsatzes von KI zeigen sich bei der Interaktion von Mensch und Maschine, wenn die jeweiligen Stärken kombiniert werden.<sup>6</sup> Neben einer reinen Datenanalyse kann die KI mit den Handlungsempfehlungen auch Textentwürfe erstellen, die sich in Entwürfe für Berichte, Vorlagen oder Stellungnahmen überführen lassen. In den Testdialogen war sie beispielsweise in der Lage, auf Grundlage der Befragung von Innenstadtbesuchern einen kurzen Bericht über die Lage der Innenstadt zu erstellen (II.6).

### 4. Fehlerhafte Analysen

Dass eine menschliche Prüfung der Ergebnisse teilweise noch erforderlich ist, zeigt sich an offensichtlich fehlerhaften Aussagen des Systems. Im Beispiel der Analyse der Lage von Schulen in Abhängigkeit der Bevölkerungsdichte (II.2) markierte das System in der Abbildung fünf Schulen, sprach in seiner Antwort aber von 129 Schulen. Es erkannte den Fehler, nachdem es auf die Diskrepanz hingewiesen wurde.<sup>7</sup> Der Fehler könnte auf eine frühe Entwicklungsphase des GPT Data Analyzer, missverständliche Prompts, Fehler in den Datensätzen oder eine Kombination davon zurückzuführen sein. Ähnliche Fehler wurden ansonsten bei der Auswertung der Geodaten nicht erkannt. Im Fall der CSV-Tabellendaten wurden nach einer manuellen Kontrolle der durch die KI erstellten Diagramme anhand der Rohdaten keine fehlerhaften Ausgaben gefunden. Insgesamt zeigt GPT-4 damit, eine geringe Fehleranfälligkeit im Hinblick auf die Richtigkeit der Antworten, wenn eine entsprechende Qualität der Rohdaten und der menschlichen Eingaben bzw. Anfragen gegeben ist.

### 5. Automatisierung von Prozessen

Es deutet sich an, dass KI-Systeme wie GPT-4 kurz- bis mittelfristig als Assistenzsysteme durch menschliche Nutzer, etwa zum Zweck der Datenauswertung, verwendet werden. Langfristig wird der Aspekt der Automatisierung immer stärker in den Fokus treten. Hier geht es um KI-Systeme, die direkt auf (Roh-)Daten zugreifen, sie analysieren und anschließend Handlungsempfehlungen aussprechen oder selbständig Entscheidungen treffen. Im Zusammenhang mit der Verwaltungstätigkeit könnte das beispielsweise bedeuten, dass entscheidungsrelevante Unterlagen automatisiert ausgewertet und entweder mit einer Handlungsempfehlung an einen menschlichen Nutzer übergeben werden.

---

<sup>4</sup> Anzumerken ist, dass auch generative Sprachmodelle durch das menschliche Feintuning nicht völlig unvoreingenommen sind, vgl. *Josipovic*, Bing Chat/GPT-4 in der öffentlichen Verwaltung, Arbeitspapier, 31.05.2023, S. 3.

<sup>5</sup> Dies gilt selbstverständlich nur, soweit die Rohdaten frei von menschlicher Voreingenommenheit sind.

<sup>6</sup> *Etscheid/von Lucke/Stroh* (Fraunhofer IAQ), Künstliche Intelligenz in der öffentlichen Verwaltung, S. 33.

<sup>7</sup> Die Ergebnisse der Analysen der CSV-Tabellendaten wurden auf Richtigkeit überprüft.

#### IV. Anforderungen an die Datenbereitstellung

Im Hinblick auf die KI-basierte Auswertung von offenen Verwaltungsdaten gilt, dass die Qualität der Rohdaten die Qualität des Ergebnisses bestimmt. Das System erkennt zwar selbständig Fehler und Unregelmäßigkeiten in den Rohdaten und kann sie beheben (beispielsweise einzelne falsch formatierte Zellen in CSV-Dateien korrigieren), die Daten müssen aber insgesamt den Anforderungen an offene Daten entsprechen. Dazu zählt, dass sie strukturiert, maschinenlesbar und mit den erforderlichen Informationen versehen sein müssen. Beispielsweise sollten in CSV-Dateien die Spalten beschriftet und Formatierungen entfernt sein.<sup>8</sup>

Möchten Verwaltungen das volle Potenzial KI-basierter Datenanalyse in Zukunft ausschöpfen, ist es erforderlich, den Umgang mit Daten zu standardisieren. Dazu müssen sie,

- einen verwaltungsweiten Prozess etablieren, um sich laufend einen Überblick des in der Verwaltung vorhandenen Datenbestandes zu verschaffen,
- den Datenbestand in geeigneter Form (etwa über einen Metadatenkatalog) festhalten,
- innerhalb der Verwaltung verbindliche Datenstandards einführen und ihre Umsetzung gewährleisten,
- die Mitarbeiterinnen und Mitarbeiter über den Umgang mit Daten aufklären,
- die Standardisierung des Umgangs mit Daten als Voraussetzung für die Entfaltung der Möglichkeiten von KI verstehen.

Im Zusammenhang mit der Einführung von KI in der öffentlichen Verwaltung wird auch das Thema Datenschutz an Bedeutung gewinnen. Verwaltungen werden generative Sprachmodelle wie ChatGPT und Bing Chat aufgrund der erforderlichen Rechenleistung sehr wahrscheinlich nicht lokal hosten und betreiben können. Möchten Verwaltungen in Zukunft die Potenziale von KI realisieren, werden sie also Umgebungen und IT-Infrastrukturen schaffen müssen, in denen gewährleistet ist, dass die mit dem System getätigten Ein- und Ausgaben nicht extern verarbeitet werden. Es ist zu erwarten, dass große IT-Dienstleister KI-basierte Produkte und Dienstleistungen im Rahmen von Cloud-Angeboten einführen und damit auch Umgebungen bereitstellen werden, die den Anforderungen des Datenschutzes genügen.

---

<sup>8</sup> Darüber hinaus ist es bei CSV-Dateien hilfreich, Informationen zur Zeichencodierung und zum Trennzeichen zu haben. Im Regelfall hat das System Probleme beim Einlesen von CSV-Dateien aber selbständig erkannt und gelöst.

## V. Fazit

GPT-4 ermöglicht durch die Verwendung natürlicher Sprache als Eingabeform eine effiziente und benutzerfreundliche Datenanalyse. Dadurch sinken zum einen die Anforderungen an die Nutzer, zum anderen können sie Analysen schneller durchführen. Das zeigt sich insbesondere bei komplexen Datensätzen in unterschiedlichen Formaten. Die Qualität der Ergebnisse hängt jedoch von der Qualität der Eingaben ab und erfordert ein gewisses Verständnis für die Kommunikation mit generativen Sprachmodellen.

KI-Systeme wie GPT-4 könnten kurz- bis mittelfristig als Assistenzsysteme in der Verwaltung eingesetzt werden, wobei langfristig eine verstärkte Automatisierung im Vordergrund steht. Die Beschaffenheit der Rohdaten beeinflusst die Ergebnisqualität erheblich. Verwaltungen sollten den Umgang mit Daten standardisieren und Mitarbeiterinnen und Mitarbeiter im Hinblick auf den (künftigen) Einsatz von KI frühzeitig schulen, um das gesamte Potenzial der KI-basierten Datenanalyse auszuschöpfen.

Es ist wahrscheinlich, dass Verwaltungen zum Einsatz von KI auf externe Cloud-Angebote zurückgreifen müssen, was Datenschutzfragen aufwirft und die Notwendigkeit von angemessenen IT-Infrastrukturen betont. Große IT-Dienstleister könnten solche Lösungen anbieten und somit den Anforderungen des Datenschutzes gerecht werden.

Um geeignete Lösungen für die öffentliche Verwaltung zu entwickeln, könnte eine enge Zusammenarbeit zwischen der öffentlichen Verwaltung und privaten IT-Dienstleistern von Vorteil sein. Dies würde es ermöglichen, die spezifischen Anforderungen und Handlungsmöglichkeiten der Verwaltung besser zu verstehen und maßgeschneiderte Lösungen zu entwickeln.



## Literatur

*Alle Links abgerufen am 12.10.2023*

Albrecht, ChatGPT und andere Computermodelle zur Sprachverarbeitung – Grundlagen, Anwendungspotenziale und mögliche Auswirkungen, Deutscher Bundestag, Ausschuss-Drs. 20(18)109, abrufbar unter:  
<https://www.bundestag.de/resource/blob/944148/30b0896f6e49908155fcd01d77f57922/20-18-109-Hintergrundpapier-data.pdf>

Deutscher Bundestag, Bericht des Ausschusses für Bildung, Forschung und Technikfolgenabschätzung gemäß § 56a der Geschäftsordnung, BT-Drs. 20/3651, Technikfolgenabschätzung (TA) - Künstliche Intelligenz und Distributed-Ledger-Technologie in der öffentlichen Verwaltung, abrufbar unter:  
<https://dserver.bundestag.de/btd/20/036/2003651.pdf>

Etscheid, Jan/Jörn von Lucke/Felix Stroh (Fraunhofer IAO), Künstliche Intelligenz in der öffentlichen Verwaltung - Anwendungsfelder und Szenarien, 2020, abrufbar unter:  
<https://publica.fraunhofer.de/bitstreams/d3d9f520-1fd4-4516-98d6-a3370c134155/download>

Josipovic, Bing Chat/GPT-4 in der öffentlichen Verwaltung, Arbeitspapier, 31.05.2023, abrufbar unter:  
[https://nevenjosipovic.de/images/pdf/Bing%20Chat%20GPT-4%20in%20der%20offentlichen%20Verwaltung\\_Arbeitspapier\\_31.05.2023.pdf](https://nevenjosipovic.de/images/pdf/Bing%20Chat%20GPT-4%20in%20der%20offentlichen%20Verwaltung_Arbeitspapier_31.05.2023.pdf)





