Datengetriebene Validierung der Polizeilichen Kriminalstatistik: Eine Langzeitanalyse von Migration und Delinquenz in Deutschland mittels Apache Spark

Bergmann Justin
Friedrich-Schiller-Universität Jena
Jena, Deutschland
justin.bergmann@uni-jena.de

***Abstract*—Dieses Paper untersucht die statistische Überrepräsentation nichtdeutscher Tatverdächtiger in der Polizeilichen Kriminalstatistik (PKS) im Zeitraum von 1984 bis 2024. Unter Verwendung einer Apache Spark Big-Data-Pipeline wurden Daten des Bundeskriminalamts mit Bevölkerungs-, Justiz- und Strafvollzugsdaten (Destatis) verknüpft und analysiert. Ziel war es, gängige Hypothesen wie „Racial Profiling“ oder sozioökonomische Ursachen datengestützt zu validieren. Die Ergebnisse zeigen eine signifikante strukturelle Ungleichheit, die sich auch in Justizurteilen und im Strafvollzug bestätigt und somit nicht allein durch polizeiliches Kontrollverhalten oder Bagatelldelikte erklärbar ist. Auch die Hypothese der rein ökonomischen Motivation konnte durch die Analyse nicht gewinnorientierter Delikte widerlegt werden.**

***Index Terms*—Polizeiliche Kriminalstatistik, Migration, Apache Spark, Data Analysis, Kriminologie, Racial Profiling**

I. Introduction

Die öffentliche Debatte zum Thema Kriminalität wird zunehmend kontrovers geführt. Insbesondere in Deutschland wird nach der Veröffentlichung der PKS häufig unmittelbar nach systemischen Erklärungen gesucht.

Saskia Esken äußerte dazu 2020 gegenüber den Zeitungen der Funke Mediengruppe, wie die Zeit berichtete [1]: „Auch in Deutschland gibt es latenten Rassismus in den Reihen der Sicherheitskräfte.“ Argumentiert wird hierbei, dass nicht die tatsächliche Delinquenz, sondern die Entdeckungswahrscheinlichkeit ungleich verteilt sei.

Ähnlich argumentiert Amnesty International in einem gemeinsamen offenen Brief mit weiteren Organisationen und verweist auf das Kontrollverhalten: „Die PKS sei in erster Linie ein ‚Tätigkeitsbericht der Polizei‘ [2]. Wer durch Schwerpunktsetzungen öfter kontrolliert wird, tauche folglich öfter in der Statistik auf.“

Als alternativen Erklärungsansatz vertrat Bundesinnenministerin Nancy Faeser die Ansicht, man müsse „bei den sozialen Ursachen ansetzen“, die sich hinter der Gewalt verbergen, wie etwa „fehlende Schulabschlüsse und Perspektivlosigkeit“ [3]. Damit wird die Ursache der statistischen Auffälligkeit primär in sozioökonomischen Faktoren verortet, statt in der Herkunft.

Die genannten Erklärungsansätze, sowohl der Vorwurf des strukturellen Rassismus als auch die reine Armutsthese, bleiben jedoch häufig den Beweis schuldig, ob sie tatsächlich die beobachteten statistischen Phänomene erklären können.

Aus diesem Grund wurden folgende Forschungsfragen formuliert, die im Rahmen dieser Arbeit datengestützt untersucht werden sollen:

RQ1 *Regionale Verteilung:* Welches Delikt dominiert in welchem Bundesland?

RQ2 *Racial Profiling:* Sind Nichtdeutsche statistisch nur deshalb überrepräsentiert, weil sie einer höheren Kontrolldichte unterliegen?

RQ3 *Sozioökonomische Validierung:* Lässt sich die Überrepräsentation allein durch soziale Notlagen erklären, oder besteht sie auch bei Delikten ohne finanziellen Anreiz (z.B. Sexualstraftaten) fort?

RQ4 *Schweregrad-Analyse:* Ist die Statistik primär durch Bagatell-Delikte verzerrt?

II. Methodik und Systemarchitektur

Um die in der Einleitung formulierten Forschungsfragen empirisch zu untersuchen, wurde ein Big-Data-Pipeline implementiert. Der Fokus lag hierbei auf der vereinigung der vier Datenquellen (Polizeiliche Kriminalstatistik, Bevölkerungsdaten, Justizdaten, JVA-Daten)

A. Datenakquise (Extract)

Die Datengrundlage bilden zwei primäre Quellen, die technisch unterschiedlichen angebunden wurden:

1) Statistisches Bundesamt (Genesis): Für die Bevölkerungszahlen, die Justizzahlen sowie die Justizvollzugszahlen wurde die API der Genesis-Datenbank genutzt. Da der Abruf des gesamten Zeitraums (1984–2024) in einer einzelnen Anfrage zu Timeouts und Speicherüberläufen führte, wurde eine *Chunking-Strategie* implementiert. Ein Python-Skript iteriert

hierbei über definierte Zeitfenster, lädt die Daten als CSV-Dateien herunter und überprüft diese auf Integrität vor der Speicherung. Die Daten hätten auch händisch heruntergeladen werden können. Allerdings stellt die Web-Oberfläche der Genesis-Datenbank nur eine bestimmte Anzahl an Ressourcen für alle Nutzer gleichzeitig zu Verfügung. Das führt dazu, dass das Herunterladen der Tabellen abgebrochen wird, wenn zu viele Nutzer gleichzeitig die Daten anfordern. Der Zugriff über die API hat diese Limitierung nicht.

2) Bundeskriminalamt: Die PKS liegt für historische Zeiträume überwiegend in unstrukturierten Formaten (Excel/xlsx) vor. Die Herausforderung bestand hier in inkonsistenten Spaltenbezeichnungen über die Jahrzehnte.

B. Datenverarbeitung (Transform)

Die Transformation der Rohdaten erfolgte mittels **Apache Spark** (PySpark). Gewählt wurde Spark einerseits aufgrund einer Vorgabe, andererseits wird Spark deshalb empfohlen, da es sich durch seine Fähigkeit zur In-Memory-Verarbeitung und horizontalen Skalierbarkeit auszeichnet.

Der ETL-Prozess umfasste folgende Kernschritte:

1) **Schema-Normalisierung:** Vereinheitlichung der Attributnamen z.B. Mapping von kryptischen PKS-Codes zu sprechenden Namen wie "tatverdaechtige_nichtdeutsch").
2) **Data Cleaning:** Bereinigung von Formatierungsartefakten. Ein kritischer Schritt war die Behandlung von Fehlwerten: In der PKS werden Null-Werte oft als Bindestrich (−) dargestellt. Diese wurden in numerische Werte (`0`) umgewandelt, um mathematische Operationen zu ermöglichen.
3) **Aggregation & Berechnung der Überrepräsentation:** Die Datensätze wurden über die Primärschlüssel `Jahr` und `Bundesland` verknüpft.
   Um die Diskrepanz zwischen den Bevölkerungsgruppen messbar zu machen, wurde der *Überrepräsentationsfaktor* berechnet. Dieser ergibt sich aus dem Quotienten der jeweiligen Per-Capita-Raten:

$$\text{Faktor} = \frac{\left(\frac{\text{TV}_{\text{nichtdeutsch}}}{\text{Bev}_{\text{nichtdeutsch}}}\right)}{\left(\frac{\text{TV}_{\text{deutsch}}}{\text{Bev}_{\text{deutsch}}}\right)}$$

   Wobei `TV` für die Anzahl der Tatverdächtigen und `Bev` für die jeweilige Wohnbevölkerung steht. Ein Faktor > 1 indiziert hierbei eine statistische Überrepräsentation.

C. Speicherung der Daten

Die bereinigten und angereicherten Daten wurden im Parquet-Format gespeichert. Im Gegensatz zu zeilenbasierten Formaten wie CSV, ermöglicht Parquet durch spaltenbasierte Speicherung und Kompression signifikant schnellere Lesezugriffe für die anschließende analytische Auswertung der Zeitreihen.

III. Analyse und Ergebnisse

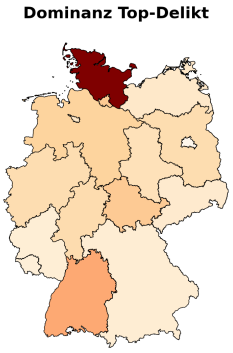
Basierend auf der in Kapitel II beschriebenen Pipeline wurden die Daten der Jahre 1984 bis 2024 ausgewertet.

Eine Einschränkung besteht jedoch bei der Strafvollzugsstatistik (JVA) der Genesis-Online-Datenbank, die erst ab dem Jahr 2017 konsistent bereitgestellt wird. Die Ursache hierfür ist nicht abschließend geklärt, dürfte jedoch in der erst späten Umstellung von händischer auf digitale Dokumentation in den Justizvollzugsanstalten begründet liegen. Im Folgenden werden die Ergebnisse entlang der Forschungsfragen präsentiert.

A. Regionale Deliktstruktur und Dominanzanalyse

In Bezug auf Forschungsfrage **RQ1** wurde untersucht, welches Delikt in den jeweilige Bundesländern dominiert.

Abbildung 1 visualisiert die prozentualen Anteile des häufigsten Delikts pro Bundesland.



Rangliste (Anteil)
Sortiert nach % an Gesamtkriminalität

Schleswig-Holstein	Betrug	**17.5%**
Ba.-Wü.	Betrug	**14.3%**
Hamburg	Betrug	**13.6%**
Thüringen	Betrug	**13.6%**
Niedersachsen	Betrug	**13.4%**
Hessen	Betrug	**13.1%**
Brandenburg	Asyl/Drogen	**13.1%**
NRW	Betrug	**13.0%**
S.-Anhalt	Betrug	**12.9%**
Bremen	Betrug	**12.7%**
Meck.-Pomm.	Sachbeschäd.	**12.6%**
Sachsen	Asyl/Drogen	**12.6%**
Bayern	Betrug	**12.4%**
Rl.-Pfalz	Körperverl.	**12.4%**
Saarland	Körperverl.	**12.3%**
Berlin	Betrug	**11.9%**

Abbildung 1. Dominierendes Delikt pro Bundesland (Anteil der Gesamtdelikte)

In 11 der 16 Bundesländer stellt „Betrug“ die relativ häufigste Deliktart dar, wobei Schleswig-Holstein mit einem Anteil von 17,5 % die höchste Intensität aufweist, gefolgt von Baden-Württemberg (14,3 %).

Signifikante Abweichungen zeigen sich jedoch in spezifischen geografischen Clustern:

- **Grenzregionen (Ost):** In Brandenburg (13,1 %) und Sachsen (12,6 %) dominieren Verstöße gegen „Nebengesetze (Asyl/Drogen)“. Dies lässt sich plausibel auf die geografische Lage als Außengrenze (zu Polen und Tschechien) zurückführen, wo Delikte nach dem Aufenthaltsgesetz oder grenzüberschreitender Schmuggel statistisch stärker ins Gewicht fallen.
- **Südwest-Cluster:** Rheinland-Pfalz und das Saarland bilden ein Ausnahmen, in denen „Körperverletzung“ das häufigste Delikt darstellt (ca. 12,4 %).
- **Mecklenburg-Vorpommern:** Als einziger Ausreißer dominiert hier die „Sachbeschädigung“ (12,6 %).

B. Zeitliche Entwicklung des Überrepräsentationsfaktors

Um zu prüfen, ob die absolute Zunahme nichtdeutscher Tatverdächtiger lediglich eine Folge des Bevölkerungswachstums ist, wurde der in der Methodik definierte Überrepräsentationsfaktor im Zeitverlauf (1987–2024) analysiert (siehe Abbildung 2).

Der Graph zeigt den Quotienten aus der Belastungszahl nichtdeutscher gegenüber deutscher Tatverdächtiger. Eine

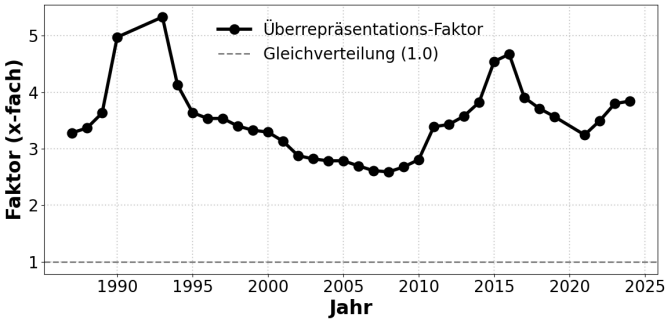


Abbildung 2. Überrepräsentation Nichtdeutscher gegenüber Deutschen per Capita

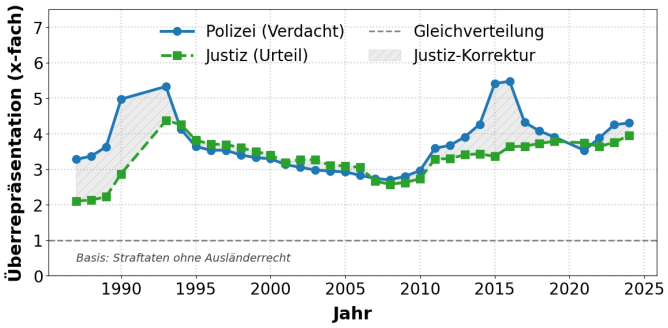


Abbildung 3. Vergleich der Überrepräsentationsfaktoren: Polizeilicher Verdacht gegenüber Gerichtliche Verurteilung (Basis: Straftaten ohne Ausländerrecht)

Gleichverteilung würde einem Faktor von 1,0 entsprechen (gestrichelte Linie). Die Analyse offenbart jedoch eine konsistente, signifikante Ungleichheit:

- **Strukturelles Niveau:** Über den gesamten Betrachtungszeitraum von fast 40 Jahren sinkt der Faktor nie unter den Wert von 2,5. Dies bedeutet, dass Nichtdeutsche selbst im historischen Tiefststand (ca. 2008) mehr als doppelt so häufig als Tatverdächtige registriert wurden wie deutsche Staatsbürger.
- **Historische Peaks:** Es zeigen sich zwei markante Hochphasen. Der erste Peak in den frühen 1990er Jahren (Faktor > 5) korreliert zeitlich mit der damaligen Asyldebatte und den Folgen der Wiedervereinigung. Ein zweiter Anstieg ist im Zuge der Flüchtlingskrise 2015/2016 zu beobachten (Anstieg auf ca. 4,5).
- **Aktueller Trend:** Nach einem Rückgang ab 2017 stabilisiert sich der Faktor aktuell auf einem Niveau von ca. 3,8.

Das Ergebnis widerlegt die Annahme, dass die Kriminalitätsraten proportional zum Bevölkerungswachstum steigen. Die entscheidende Folgefrage (**RQ2**) lautet nun, ob dieser Faktor durch tatsächliche Straffälligkeit oder durch ein verzerrtes Anzeigeverhalten der Polizei zustande kommt.

C. Systemübergreifende Validierung (Justizdaten)

Um diese Hypothese zu prüfen, werden in Abbildung 3 die Daten der PKS (Verdachtsebene) mit der Strafverfolgungsstatistik (Verurteilungsebene) kontrastiert. Wäre die Überrepräsentation lediglich ein Artefakt polizeilicher Vorurteile, müsste die Quote spätestens vor Gericht, wo juristische Beweisstandards gelten, massiv einbrechen (Divergenz der Kurven).

Die Analyse der Zeitreihe (1987–2024) liefert jedoch ein gegenteiliges Bild:

- **Hohe Korrelation:** Die Kurve der Verurteilten (grün) folgt der Kurve der Tatverdächtigen (blau) nahezu synchron. Phasen erhöhter polizeilicher Registrierung (z.B. 1993 oder 2016) spiegeln sich zeitversetzt auch in den Urteilen wider.
- **Die Justiz-Korrektur:** Die graue Fläche zwischen den Kurven visualisiert die Diskrepanz zwischen Verdacht und Verurteilung. Zwar liegt der Faktor bei den Verurteilungen systematisch leicht unter dem der Polizei (was für die Unschuldsvermutung und Filterfunktion der Justiz spricht), jedoch bleibt die *strukturelle Überrepräsentation* bestehen.
- **Niveau der Verurteilungen:** Selbst nach der juristischen Prüfung liegt der Faktor bei verurteilten Nichtdeutschen konstant zwischen 2,5 und 4,0.

Daraus lässt sich ableiten: Zwar existiert ein gewisser Selektionseffekt auf Ebene der Verdachtsschöpfung (graue Fläche), dieser ist jedoch nicht ausreichend, um die statistische Ungleichheit zu erklären. Die unabhängige Instanz der Gerichte bestätigt den Trend der Polizeistatistik im Wesentlichen.

D. Überprüfung der sozioökonomischen Hypothese (Armutsthese)

Zur Beantwortung von Forschungsfrage **RQ3** wird die verbreitete These geprüft, Kriminalität sei primär eine Folge prekärer Lebensverhältnisse. Greift man die eingangs dargelegte Argumentation der Bundesregierung auf, wonach primär soziale Faktoren wie „Armut und Perspektivlosigkeit“ [3] kriminalitätsfördernd wirken, lässt sich hieraus eine klare empirische Erwartungshaltung ableiten: Wäre die ökonomische Not der alleinige Treiber, müsste die statistische Überrepräsentation bei *nicht gewinnorientierten Delikten* (Bereicherungskriminalität) signifikant höher ausfallen als bei Delikten ohne finanziellen Anreiz .

Abbildung 4 zeigt die Entwicklung der Belastungsfaktoren explizit für Delikte gegen die sexuelle Selbstbestimmung.

Die Daten zeigen jedoch, dass die sozioökonomische Erklärung zu kurz greift:

- **Hohes Niveau ohne finanzielle Not:** Auch bei Delikten, die keinerlei materielle Besserung versprechen wie z. B. Vergewaltigung, liegt der Überrepräsentationsfaktor konstant hoch (zwischen 3,5 und 6,0).
- **Widerlegung der reinen Armutsthese:** Wäre Armut der alleinige Treiber, müsste der Faktor in diesem Deliktbereich gegen 1,0 tendieren (Gleichverteilung), da Armut nicht kausal zu Sexualgewalt führt. Das Gegenteil ist der Fall: Die Werte liegen teilweise sogar über denen der reinen Vermögensdelikte.

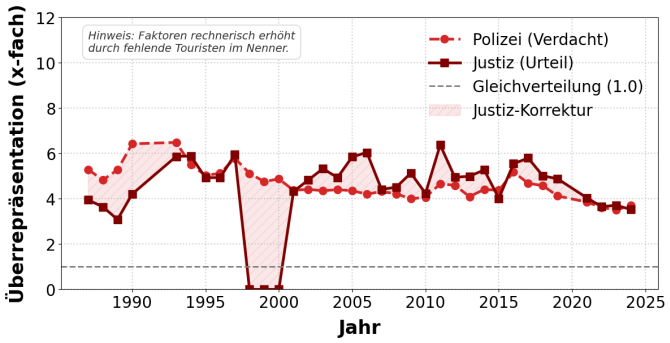


Abbildung 4. Überrepräsentationsfaktor bei nicht gewinnorientierten Delikten (Vergleich Polizei gegenüber Justiz)

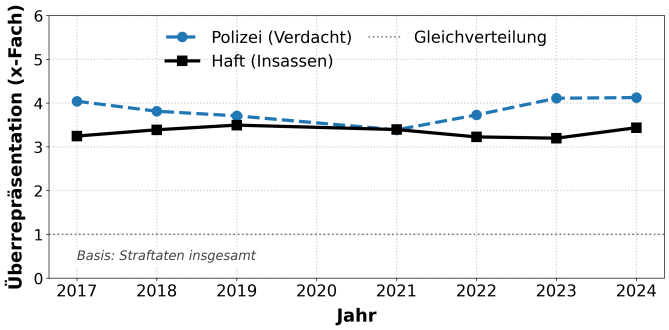


Abbildung 5. Überrepräsentationsfaktor: Polizeilicher Verdacht gegenüber Effektiver Strafvollzug (Insassen)

- **Bestätigung durch die Justiz:** Auch hier zeigt die Kurve der Verurteilten (dunkelrot), dass es sich nicht um ein reines Artefakt polizeilicher Verdächtigung handelt. Zwar gibt es eine sichtbare „Justiz-Korrektur“ (grau-rosa Fläche), doch der Faktor der rechtskräftig Verurteilten pendelt sich stabil auf einem sehr hohen Niveau (ca. 4-fach) ein.

Daraus folgt: Die soziale Lage mag bei Eigentumsdelikten ein Erklärungsfaktor sein, sie taugt jedoch nicht als alleinige Ursache zur Erklärung der gesamtheitlichen statistischen Auffälligkeit, insbesondere im Bereich der Gewaltkriminalität.

E. Einfluss von Bagatell-Delikten (Validierung durch Strafvollzug)

Die letzte Forschungsfrage (**RQ4**) widmet sich dem Einwand, die statistische Verzerrung in der PKS sei primär auf leichte Vergehen (sog. Bagatell-Kriminalität wie Ladendiebstahl oder Beförderungserschleichung) zurückzuführen.

Zur Überprüfung dieser Hypothese dient die Statistik der Strafvollzugsinsassen. Da Freiheitsstrafen in Deutschland als *ultima ratio* nur bei schwerwiegenden Delikten oder hoher Rückfallgeschwindigkeit verhängt werden, fungiert die Inhaftierten-Quote als verlässlicher Indikator für Schwerkriminalität.

Abbildung 5 stellt die Überrepräsentation bei Tatverdächtigen (Polizei) jener der tatsächlich Inhaftierten (JVA) gegenüber. Wie in der Methodik erwähnt, stehen diese Daten konsistent erst ab 2017 zur Verfügung.

Die Ergebnisse widerlegen die Bagatell-Hypothese deutlich:

- **Konvergenz der Kurven:** Die Kurve der Inhaftierten (schwarz) verläuft nahezu deckungsgleich mit der polizeilichen Verdachtskurve (blau). Beide weisen einen Faktor von über 3,0 auf.
- **Stabilität bei Schwerkriminalität:** Würde die Überrepräsentation primär auf Bagatellen beruhen, müsste der Faktor im Gefängnissystem drastisch sinken (in Richtung 1,0), da Ersttäter bei kleinen Delikten in der Regel Geld- oder Bewährungsstrafen erhalten und nicht inhaftiert werden.
- **Schlussfolgerung:** Dass der Faktor im Strafvollzug fast identisch hoch bleibt, beweist, dass die statistische Abweichung nicht durch leichte Delikte verzerrt ist, sondern sich auch bei schweren Straftaten (die zwingend zu Haft führen) in gleichem Maße manifestiert.

Zusammenfassend bestätigen die Daten des Strafvollzugs die Trends der Polizeistatistik und validieren, dass es sich bei der Überrepräsentation um ein phänomenübergreifendes Muster handelt, das unabhängig vom Schweregrad der Tat besteht.

IV. Fazit und Diskussion

Ziel dieser Arbeit war es, die öffentlich geführte Debatte über Kriminalität und Migration durch eine datengestützte Big-Data-Analyse (1984–2024) zu objektivieren. Entgegen der oft emotional geführten Diskussion, die primär methodische Mängel der PKS oder sozioökonomische Benachteiligung als Ursachen für die statistische Überrepräsentation anführt, liefert die vorliegende Untersuchung differenzierte Ergebnisse.

A. Beantwortung der Forschungsfragen

Die Ergebnisse lassen sich entlang der eingangs formulierten Forschungsfragen wie folgt zusammenfassen:

- **Zu RQ1 (Regionale Verteilung):** Kriminalität ist nicht homogen verteilt. Während in Stadtstaaten und Grenzregionen (Ost) spezifische Deliktmuster dominieren (z.B. Verstöße gegen das Aufenthaltsrecht), zeigen wirtschaftsstarke Flächenländer eine Dominanz von Vermögensdelikten.
- **Zu RQ2 (Racial Profiling):** Die Hypothese, die Überrepräsentation sei primär ein Artefakt behördlichen Kontrollverhaltens, konnte durch den Abgleich mit Justizdaten widerlegt werden. Zwar ist eine leichte statistische Bereinigung zwischen Verdacht (Polizei) und Verurteilung (Gericht) messbar, jedoch bestätigt die Judikative die Trends der Exekutive in hohem Maße. Eine rein durch Vorurteile erzeugte Datenlage würde vor Gericht keinen Bestand haben.
- **Zu RQ3 (Armutsthese):** Die Annahme, Kriminalität sei ausschließlich eine Folge prekärer Lebensverhältnisse, erweist sich als unzureichend. Da die Überrepräsentation auch bei nicht gewinnorientierten Delikten wie z. B. Sexualstraftaten ähnlicher oder höherer Intensität auftritt, kann ökonomische Not nicht als alleinige Ursache dienen.

- **Zu RQ4 (Schweregrad):** Die Analyse der Strafvollzugsdaten (Inhaftierte) zeigt, dass die Abweichung nicht durch Bagatell-Delikte verzerrt ist. Im Bereich der Schwerkriminalität, die zu Haftstrafen führt, manifestiert sich das gleiche statistische Bild wie in der PKS.

B. Schlussfolgerung

Die empirischen Daten legen nahe, dass die statistische Überrepräsentation nichtdeutscher Tatverdächtiger ein reales, systemübergreifendes Phänomen ist, das sich weder durch statistische Verzerrungen noch allein durch sozioökonomische Faktoren "wegerklären"lässt. Für die künftige Debatte bedeutet dies, dass der Fokus von der Kritik an der Datenerhebung hin zu einer Analyse der tatsächlichen Ursachen (z.B. kulturelle Faktoren, Integrationsdefizite, Sozialisationserfahrungen) verschoben werden muss, da die Datenbasis als valide bestätigt wurde.

V. Ausblick und Limitationen

Trotz der umfangreichen Zeitreihenanalyse unterliegt diese Arbeit gewissen Einschränkungen. So bildet auch die Justizstatistik nur das Hellfeld ab. Das Dunkelfeld der nicht angezeigten Straftaten bleibt naturgemäß unberücksichtigt. Zudem stand die Strafvollzugsstatistik in digitaler Form erst ab 2017 zur Verfügung, was Langzeitanalysen in diesem spezifischen Sektor erschwerte.

Zukünftige Forschungsarbeiten könnten auf diesem Fundament aufbauen, indem sie:

1) **Differenzierte Untersuchungen** durchführen, die weitere Faktoren (Bildungsgrad, Aufenthaltsstatus, Altersstruktur) einbeziehen, sofern diese Daten verfügbar gemacht werden.
2) **Prognosemodelle** entwickeln, die den Personaleinsatz gezielt an den regionalen Brennpunkten Abbildung 1 ausrichten.

Literatur

[1] S. Esken, „Auch in Deutschland gibt es latenten Rassismus,“ *ZEIT ONLINE*, 8. Juni 2020. [Online]. Verfügbar: https://www.zeit.de/politik/deutschland/2020-06/saskia-esken-spd-polizei-rassismus

[2] Amnesty International et al., „Offener Brief: Die polizeiliche Kriminalstatistik ist als Instrument zur Bewertung der Sicherheitslage ungeeignet,“ April 2025. [Online]. Verfügbar: https://www.amnesty.de/sites/default/files/2025-04/Offener-Brief-Polizeiliche-Kriminalstatistik-PKS-2025.pdf

[3] N. Faeser, „Vorstellung der Polizeilichen Kriminalstatistik 2023,“ Bundespressekonferenz, Berlin, 9. April 2024. Aufzeichnung verfügbar via Jung & Naiv: https://youtu.be/bolPCBt1LOo?t=166 (abgerufen am 1. Februar 2026).

[4] Bundeskriminalamt (BKA), „Polizeiliche Kriminalstatistik (PKS) 1984–2023: Zeitreihen und ausgewählte Zahlen,“ Wiesbaden, 2024. [Online]. Verfügbar: https://www.bka.de/DE/AktuelleInformationen/StatistikenLagebilder/PolizeilicheKriminalstatistik/pks_node.html

[5] Statistisches Bundesamt (Destatis), „Genesis-Online Datenbank: Bevölkerung, Verurteilte und Strafvollzug (Codes 12411, 24311, 24411),“ 2024. [Online]. Verfügbar: https://www-genesis.destatis.de

[6] Apache Software Foundation, „Apache Spark: Unified Engine for Large-Scale Data Analytics,“ 2024. [Online]. Verfügbar: https://spark.apache.org/

[7] Apache Software Foundation, „Apache Parquet: Columnar Storage Format,“ 2024. [Online]. Verfügbar: https://parquet.apache.org/