

Abschlussarbeit zum Zertifikatslehrgang Data Analyst (IHK)

BEISPIEL EINES DATENANALYSE-PROZESSES
MIT DATENSÄTZEN DER POLIZEILICHEN
KRIMINALSTATISTIK 2022 DES BKA

PATRIK PAUL

Inhaltsverzeichnis

1. Einleitung	2
1.1. Die Schritte der Analyse	2
1.2. Die Datensätze	2
1.3. Ziele der Analyse	2
2. Verarbeitung und Darstellung	3
2.1. Datenvorverarbeitung	3
2.1.1. Datenqualität	3
2.1.2. Datenbereinigung & -transformation	4
2.1.3. Datenmodell und Beziehungen	5
2.1.4. Gruppierung und Aggregation	6
2.1.5. Tabellen zusammenführen – „Join“	7
2.2. Tabellen und Visualisierungen	8
2.2.1. Die einfache Tabelle	8
2.2.2. Säulendiagramm	9
2.2.3. Landkarte & Drilldown	9
2.2.4. Liniendiagramm bzw. Zeitreihe	11
3. Analyse mit Kennzahlen	12
3.1. Benutzerdefinierte Spalten & Measures	12
3.2. Werte der erstellten Kennzahlen	12
3.3. Aufklärung in Abhängigkeit des Verhältnisses „TV zu Fälle“	12
3.4. Weitere Auswertung	13
3.5. Ergebnisse	14
4. Fazit und Ausblick	14
5. Abbildungsverzeichnis	15
6. Abkürzungsverzeichnis	15
7. Quellenverzeichnis	15

1. Einleitung

1.1. Die Schritte der Analyse

In dieser Abschlussarbeit soll beispielhaft die Vorgehensweise in einem typischen Datenanalyse-Prozess beschrieben werden. Mit ausgewählten Datensätzen sollen Schritte wie die Vorverarbeitung von Daten, das Erstellen eines Berichts mit Visualisierungen und gewissen Funktionen der Software, sowie eine Auswertung mit generierten Werten durchgeführt werden. Diese Werte kann man als Metriken oder Kennzahlen bezeichnen, in einer Geschäftsumgebung spricht man auch von KPIs.

1.2. Die Datensätze

Als Grundlage für dieses Projekt wurden aus Interesse Datensätze der Polizeilichen Kriminalstatistiken (PKS) des Bundeskriminalamtes (BKA) für das Jahr 2022 gewählt. Diese stehen über die Homepage www.bka.de in der Rubrik „Aktuelle Publikationen“ zur Verfügung. Dort gibt es verschiedene Statistiken der vergangenen rund 10 Jahre. Für dieses Projekt werden aus einer Vielzahl an Tabellen die folgenden verwendet:

- ‚Grundtabelle Fälle, auf Bundesebene‘
- ‚Tatverdächtige insgesamt nach Alter und Geschlecht‘
- ‚Opfer insgesamt nach Alter und Geschlecht‘
- ‚Grundtabelle - Fälle mit Häufigkeitszahl - Städte‘
- ‚Zeitreihen - Grundtabelle - Fälle ab 1987‘

1.3. Ziele der Analyse

Wie Ursprung und Bezeichnung der Datensätze erkennen lassen, beschäftigt sich diese Arbeit mit Straftaten, deren Häufigkeitszahlen und räumlichen Verteilungen, Aufklärungsquoten (AQ), sowie mit den dazugehörigen Tatverdächtigen (TV). Die Verdächtigen werden pro Straftat nach Alter und Geschlecht, sowie nach Deutsch oder nicht-Deutsch aufgelistet. Ähnlich sieht die Tabelle über die Opfer-Statistik aus. Diese Inhalte werden teilweise in Visualisierungen dargestellt. Weiter werden auf Spalten Gruppierung und Aggregation angewendet, sowie Kennwerte durch benutzerdefinierte Operationen ermittelt. Konkret soll das Verhältnis zwischen den Fallzahlen und Anzahl an TV untersucht werden, sowie die Verteilung der höchsten TV-Zahlen nach Altersgruppen. Diese beiden Werten werden abschließend mit den entsprechenden Aufklärungsraten verglichen, um Korrelationen zu erkennen.

2. Verarbeitung und Darstellung

2.1. Datenvorverarbeitung

2.1.1. Datenqualität

Ein wichtiger Schritt der Analyse ist die Beurteilung der Qualität der vorliegenden Datensätze. Je besser die Qualität ist, umso besser kann eine Analyse durchgeführt werden bzw. umso zuverlässiger sind die Ergebnisse. Auch nachgelagerte Schritte wie das Ableiten eines Handlungsbedarfs sind letztendlich zutreffender. Zur Beurteilung der Qualität eines Datensatzes werden im Wesentlichen folgende Eigenschaften betrachtet: Vollständigkeit, Eindeutigkeit, Korrektheit, Zuverlässigkeit, Relevanz, Konsistenz, Einheitlichkeit, Verständlichkeit.

Um diese Punkte beurteilen zu können, muss man sich jeden Datensatz genauer anschauen bevor man anfängt, mit ihm zu arbeiten. Dies ist der erste Schritt nachdem man die Daten in ein Analysetool importiert hat. In diesem Bericht wird „Power BI Desktop“ von Microsoft verwendet. Hier hat man beim Import jedes Datensatzes die Möglichkeit, den Schritt „Daten transformieren“ zu wählen. Dann öffnet sich der sogenannte „Power Query-Editor“, mit dem man den Datensatz betrachten und bearbeiten kann. Alternativ können Datensätze, die hier als csv und xls vorliegen, auch zuvor mit Excel betrachtet werden. Da es sich hier um nicht sehr große Datensätze handelt (20-30 Spalten und 1000-3500 Zeilen), stellt die Betrachtung in Excel kein Problem dar.

Zu den Beurteilungen nach einigen genannten Kriterien, stichpunktartig:

- Vollständigkeit: gegeben, alle Spalten und Zeilen befüllt
- Eindeutigkeit: gegeben, Einträge sind sauber getrennt (nach Straftaten)
- Korrektheit: vermutlich gegeben, da aus erster Hand (Quelle: BKA)
- Zuverlässigkeit: gut, Statistik wird jedes Jahr veröffentlicht
- Einheitlichkeit: gut, Spalten mit Text, Ganz- / Kommazahlen sind durchgängig
- Verständlichkeit: gut, klare Spaltenbezeichnungen

Zu dem Punkt Vollständigkeit ist noch folgendes zu sagen: Die Datensätze des BKA sind in sich vollständig und lückenlos. Allerdings sind in diesen Tabellen nicht sämtliche Straftaten enthalten. Es wird darauf hingewiesen, dass nur „ausgewählte Delikte“ aufgeführt werden und gewisse Straftatengruppen außen vor sind.

2.1.2. Datenbereinigung & -transformation

Nachdem die Datensätze auf Ihre Qualität hin beurteilt wurden und eine Verwendbarkeit gegeben ist, sollten und müssen weitere Schritte vorgenommen werden. Da die Spaltenüberschriften mehrzeilig angeordnet sind, ist es nötig, diese auf eine Zeile zu reduzieren. Bedeutet die Beschriftungen sind zusammenzufassen und leere oder überflüssige Zeilen zu löschen. Danach ist häufig die Funktion „Erste Zeile als Überschriften verwenden“ anzuwenden.

Ein in Zusammenhang mit Datenqualität genannter Punkt ist die Relevanz. Dieser kommt auch bei der Datenbereinigung zum Tragen. Denn in den vorliegenden Tabellen sind wie in vielen Fällen Spalten enthalten, die für die beabsichtigte Analyse nicht relevant sind. In der Grundtabelle zu den erfassten Straftaten des Jahres 2022 sind es die Spalten zur Tatortverteilung. Da das Thema Örtlichkeit anhand einer anderen Tabelle behandelt werden soll, werden diese Spalten hier gelöscht.

	A	B	C	D	E	F	G	H	I	J
1	Polizeiliche Kriminalstatistik		Tabelle 01							
2	Grundtabelle		Bereich: Bundesrepublik Deutschland (70)							
3	V1.0 erstellt am: 14.02.2023		Berichtszeitraum: 01.01.2022 bis 31.12.2022							
4	Schlüssel	Straftat	Anzahl erfasste Fälle	% -Anteil an allen Fällen	erfasste Fälle		mit Schusswaffe		Aufklärung	
davon: Versuche					gedroht	ge-schossen	Anzahl Fälle	in % (AQ)		
									Anzahl	in %
5	1	2	3	4	5	6	12	13	14	15
9	-----	Straftaten insgesamt	5.628.584	100,0	414.197	7,4	4.092	4.442	3.226.935	57,3
10	000000	Straftaten gegen das Leben	3.077	0,1	1.770	57,5	8	133	2.732	88,8
11	010000	Mord § 211 StGB	662	0,0	451	68,1	3	59	604	91,2
12	010079	Sonstiger Mord	625	0,0	425	68,0	3	55	570	91,2
13	011000	Mord im Zusammenhang mit Raubdelikten	29	0,0	23	79,3	0	4	26	89,7
14	012000	Mord im Zusammenhang mit Sexualdelikten	8	0,0	3	37,5	0	0	8	100,0
15	020000	Totschlag und Tötung auf Verlangen §§ 212, 213 StGB	1.574	0,0	1.272	80,8	5	73	1.494	94,9
16	020010	Totschlag § 212 StGB	1.549	0,0	1.267	81,8	5	69	1.469	94,8
17	020020	Minder schwerer Totschlag § 213 StGB	1	0,0	1	100,0	0	0	1	100,0
18	020030	Tötung auf Verlangen § 216 StGB	24	0,0	4	16,7	0	4	24	100,0
19	030000	Fahrlässige Tötung § 222 StGB - nicht i.V.m. V	748	0,0	0	0,0	0	1	543	72,6
20	040000	Abbruch der Schwangerschaft §§ 218, 218b, 218c StGB	93	0,0	47	50,5	0	0	91	97,8
21	040010	Schwangerschaftsabbruch § 218 StGB	92	0,0	47	51,1	0	0	90	97,8
22	040020	Schwangerschaftsabbruch ohne ärztliche Festst	1	0,0	0	0,0	0	0	1	100,0
23	040030	Ärztliche Pflichtverletzung § 218c StGB	0	0,0	0	0,0	0	0	0	0,0
24	040040	Werbung für den Abbruch der Schwangerschaft	0	0,0	0	0,0	0	0	0	0,0
25	040050	Inverkehrbringen von Mitteln Schwangerschafts-	0	0,0	0	0,0	0	0	0	0,0
26	100000	Straftaten gegen die sexuelle Selbstbestimmur	118.196	2,1	2.902	2,5	27	2	98.376	83,2
27	110000	Straftaten gegen die sexuelle Selbstbestimmur	37.076	0,7	1.855	5,0	26	2	29.710	80,1

Abbildung 1: Grundtabelle in Excel, Auszug

Wichtig ist auch das Betrachten der Datentypen. Teilweise werden diese beim Import nicht korrekt erkannt. So auch hier: Die Spalte „Straftatschlüssel“ wurde als Text erkannt, obwohl bis auf den ersten Eintrag alle Felder mit Ganzzahlen befüllt sind. Daneben wurde „Straftat“ als gemischter Datentyp erkannt, obwohl alle Felder Texte enthalten. Üblicherweise ändert man diese Datentypen entsprechend. Allerdings liegt bei dem Schlüssel der spezielle Fall vor, das es zwar eine Zahl ist, aber dies ein Wert ist, mit dem man keine Rechenoperationen vornimmt. In diesem Fall belässt

man den Datentyp bei Text. Bei weiteren Spalten ist ebenfalls der Typ von gemischt auf Ganz- bzw. Dezimalzahl zu ändern. Der Grund für diese falsche Erkennung liegt in dem mehrzeiligen Header. Durch vorheriges Entfernen dieser Zeilen wäre die Erkennung vermutlich besser ausgefallen.

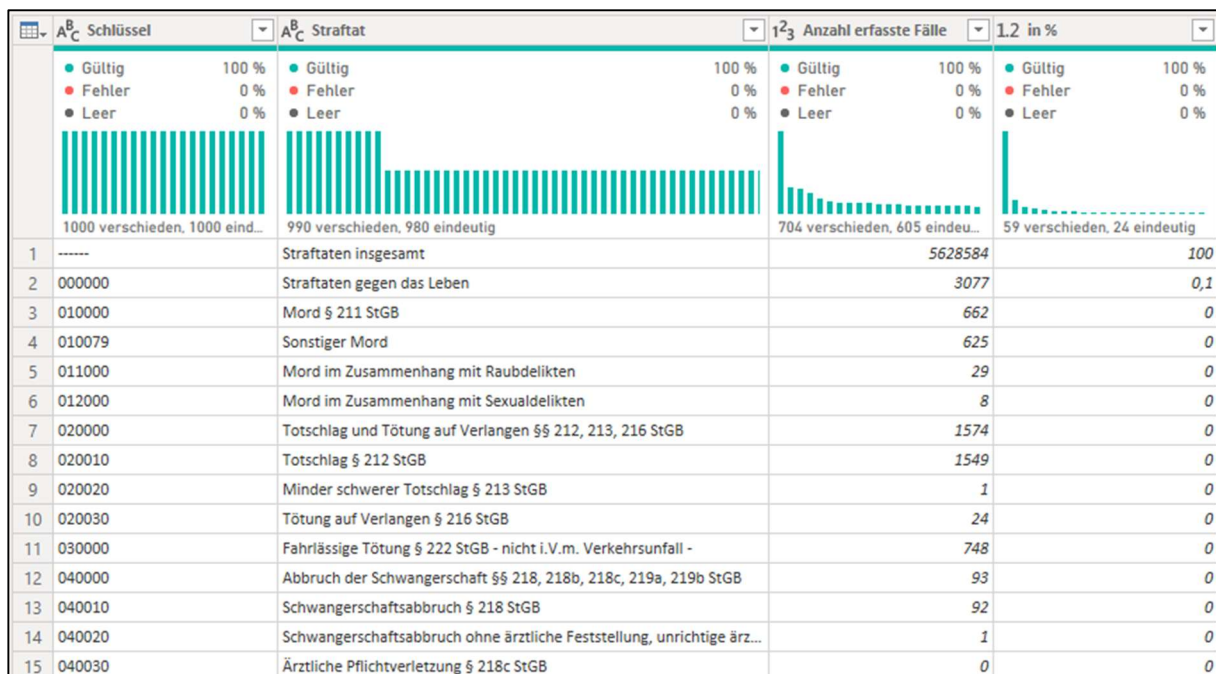


Abbildung 2: Grundtabelle im Power Query-Editor, Auszug

Es gibt noch eine Menge weiterer Schritte, die man falls nötig hier ausführen kann. Wie das Entfernen von Zeilen, die nur null-Werte enthalten, oder das Auftrennen von Spalten, wenn neben Zahlenwerte unerwünschte Zeichen enthalten sind. Abschließend sollte man die Abfrage noch sinnvoll benennen.

2.1.3. Datenmodell und Beziehungen

In ‚Power BI Desktop‘ gibt es die Modellansicht, in der das Datenmodell graphisch dargestellt wird. Hier werden die importierten Datensätze mit ihren Spalten jeweils als eine Art Kachel dargestellt und Beziehungen zwischen den Tabellen als Pfade. Mit einer Beziehung ist gemeint, dass es in den jeweiligen Tabellen Spalten mit gleichem Inhalt gibt, worüber die Tabellen verknüpft werden können. Was bedeutet, dass die Inhalte der entsprechenden Zeilen einander zugeordnet sind. Die entsprechenden Spalten bezeichnet man als Schlüssel. In unserem Fall hat die Spalte zufällig auch den Name Schlüssel, also der Straftaten-Schlüssel. In den drei hier verwendeten Datensätzen sind so die Einträge zu Fällen, den Tatverdächtigen und den Opfern pro Straftat miteinander verknüpft. Eine Tabelle mit verknüpften Einträgen zeigt Abb.7 in Kapitel 2.2.1.

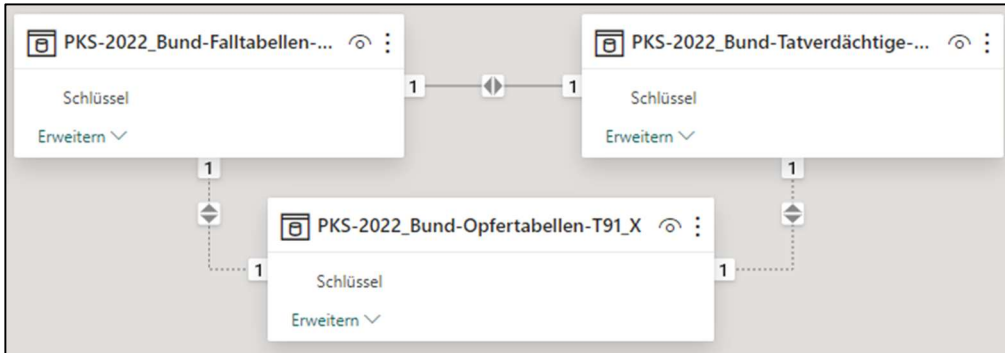


Abbildung 3: Modellansicht in Power BI Desktop

Generell gibt es verschiedene Typen von Beziehungen. Dieser Typ ist abhängig davon, ob die Einträge der einen Tabelle nur einmal oder mehrfach in der anderen Tabelle zugeordnet werden können. Man bezeichnet dies als 1:1- oder als 1:n-Beziehung. Ist im zweiten Fall die Zuordnung auch in der Gegenrichtung mehrfach möglich, ist es eine m:n-Beziehung. Dies sind die Grundtypen, weitere Varianten werden hier außen vor gelassen.

2.1.4. Gruppierung und Aggregation

Wenn Felder einer Dimension mehrfach vorkommen, weil in einer anderen Spalte mehrere Werte zugeordnet sind, können diese Werte gruppiert und untereinander angezeigt werden. Dann können diese Gruppen noch zusammengefasst werden. Die Art der Zusammenfassung bezeichnet man als Aggregation. In einer der genutzten Tabellen werden unter anderem die Anzahl Fälle pro Straftat und Stadt aufgeführt. Somit kann man z.B. nach den Städten gruppieren und die Fallzahlen aufsummieren.

A ^B C Stadt	1 ² 3 Summe aller Fälle
<ul style="list-style-type: none"> ● Gültig 100 % ● Fehler 0 % ● Leer 0 % <p>84 verschieden, 84 eindeutig</p>	<ul style="list-style-type: none"> ● Gültig 100 % ● Fehler 0 % ● Leer 0 % <p>84 verschieden, 84 eindeutig</p>
Aachen	83260
Augsburg	53014
Bergisch Gladbach	17435
Berlin	1504690
Bielefeld	67979
Bochum	81612
Bonn	102031
Bottrop	24294
Braunschweig	54433
Bremen	193419


Abbildung 4: Tabelle, gruppiert und aggregiert, Auszug

2.1.5. Tabellen zusammenführen – „Join“


Wenn es wie oben beschrieben einen „Schlüssel“ gibt, über den Einträge verschiedener Tabellen verknüpft sind, kann man somit diese Tabellen paarweise zusammenführen. Dies wird auch als ‚Join‘ bezeichnet. Alle Spalten werden somit in eine gemeinsame Tabelle geschrieben. Hierbei gibt es verschiedene Arten von Joins. Entweder werden nur die zusammenpassenden Einträge übernommen – „Inner Join“. Oder es werden die jeweils überhängenden Werte der einen bzw. der anderen Tabelle auch übernommen und die fehlenden Werte mit „null“ gefüllt – „Left / Right Outer Join“. Oder es werden beide überhängenden Werte übernommen, dann ist es ein „Full Outer Join“. Als letzte Möglichkeit gibt es den „Cross Join“, bei dem alle möglichen Kombinationen der Spalten aus beiden Tabellen erzeugt werden. Da es in den vorliegenden Tabellen „Fälle“ und „Tatverdächtige“ jeweils 1139 zueinander passende Einträge gibt, wählt man den „Inner Join“.

Zusammenführen

Wählen Sie eine Tabelle und übereinstimmende Spalten aus, um eine zusammengeführte Tabelle zu erstellen.

PKS-2022_Bund-Falltabellen-T01 (2) 

Nr.	Schlüssel	Straftat	Anzahl erfasste Fälle	in %	Versuche	Versuche in
1	-----	Straftaten insgesamt	5628584	100	414197	
2	000000	Straftaten gegen das Leben	3077	0,1	1770	
3	010000	Mord § 211 StGB	662	0	451	
4	010079	Sonstiger Mord	625	0	425	

PKS-2022_Bund-Tatverdächtige-T20... 

Nr.	Schlüssel	Straftat	TV insgesamt	unter 6	6 - 7	8 - 9	10 - 11
1	-----	Straftaten insgesamt	2093782	4283	4109	8600	21559
2	000000	Straftaten gegen das Leben	3539	2	0	1	2
3	010000	Mord § 211 StGB	763	0	0	1	0
4	010079	Sonstiger Mord	711	0	0	1	0

Join-Art

Fuzzyübereinstimmungen zum Zusammenführen verwenden

> Optionen für Fuzzyübereinstimmung

✓ Die Auswahl stimmt mit 1139 von 1139 Zeilen der ersten Tabelle und mit...

Abbildung 5: Tabellen zusammenführen

Da hier sogar mehrere Spalten identisch sind, lässt man diese beim Ausklappen der angefügten Tabelle nicht anzeigen. Solche Redundanzen sind auszuschließen. Genauso kann man diese Spalten vor oder nach dem Zusammenführen löschen.

2.2. Tabellen und Visualisierungen

2.2.1. Die einfache Tabelle

Um einen Eindruck über die wichtigsten Zahlen des Datensatzes zu geben, kann man diese in Form einer Tabelle darstellen. Über einen Datenschnitt können einzelne oder mehrere Straftaten ausgewählt werden, um die Tabelle auf die entsprechenden Werte zu beschränken.

Polizeiliche Kriminalstatistik 2022 Bundesebene Grundtabelle							
Straftat	Straftat	Anzahl erfasste Fälle	Tatverdächtige insgesamt	männlich	weiblich	Anzahl Aufklärung	Aufklärung in %
<input checked="" type="checkbox"/> Bodenverunreinigung	Bodenverunreinigung	828	554	502	52	469	56,60
<input type="checkbox"/> Bodenverunreinigung - ...	Bodenverunreinigung 3...	71	58	55	3	55	77,50
<input type="checkbox"/> Börsengesetz	Bundesnaturschutzgesetz	512	387	318	69	339	66,20
<input type="checkbox"/> Brandstiftung mit Todes...	Computersabotage	546	112	100	12	113	20,70
<input type="checkbox"/> Brandstiftung und Herb...							
<input type="checkbox"/> Bundesbankgesetz							
<input checked="" type="checkbox"/> Bundesjagdgesetz							
<input checked="" type="checkbox"/> Bundesnaturschutzgesetz							
<input type="checkbox"/> Cannabis und Zubereitu...							
<input type="checkbox"/> Computerbetrug (sonsti...							
<input type="checkbox"/> Computerbetrug (sonsti...							
<input type="checkbox"/> Computerbetrug § 263a...							
<input type="checkbox"/> Computerbetrug mittels...							
<input type="checkbox"/> Computerbetrug mittels...							
<input type="checkbox"/> Computerbetrug mittels...							
<input checked="" type="checkbox"/> Computersabotage							

Abbildung 6: Tabelle mit Datenschnitt, Auszug

Nr.	Schlüssel	Straftat	Anzahl erfasste Fälle	Anzahl Aufklärung	TV insgesamt	Opfer insgesamt
1	-----	Straftaten insgesamt	5628584	3226935	2093782	1151908
2	000000	Straftaten gegen das Leben	3077	2732	3539	967
3	010000	Mord § 211 StGB	662	604	763	924
4	010079	Sonstiger Mord	625	570	711	35
5	011000	Mord im Zusammenhang mit Raubdelikten	29	26	46	8
6	012000	Mord im Zusammenhang mit Sexualdelikten	8	8	8	1834
7	020000	Totschlag und Tötung auf Verlangen §§ 212, ...	1574	1494	1948	1809
8	020010	Totschlag § 212 StGB	1549	1469	1926	1
9	020020	Minder schwerer Totschlag § 213 StGB	1	1	2	24
10	020030	Tötung auf Verlangen § 216 StGB	24	24	21	777
11	030000	Fahrlässige Tötung § 222 StGB - nicht i.V.m. V...	748	543	745	38621

Abbildung 7: Tabelle mit verknüpften Einträgen, Auszug

2.2.2. Säulendiagramm

Eine weniger zahlen-lastige Darstellung ist natürlich eine graphische. Um beispielsweise die unterschiedlichen Attribute verschiedener Straftaten vergleichbar darzustellen, bietet sich das beliebte Säulendiagramm an.

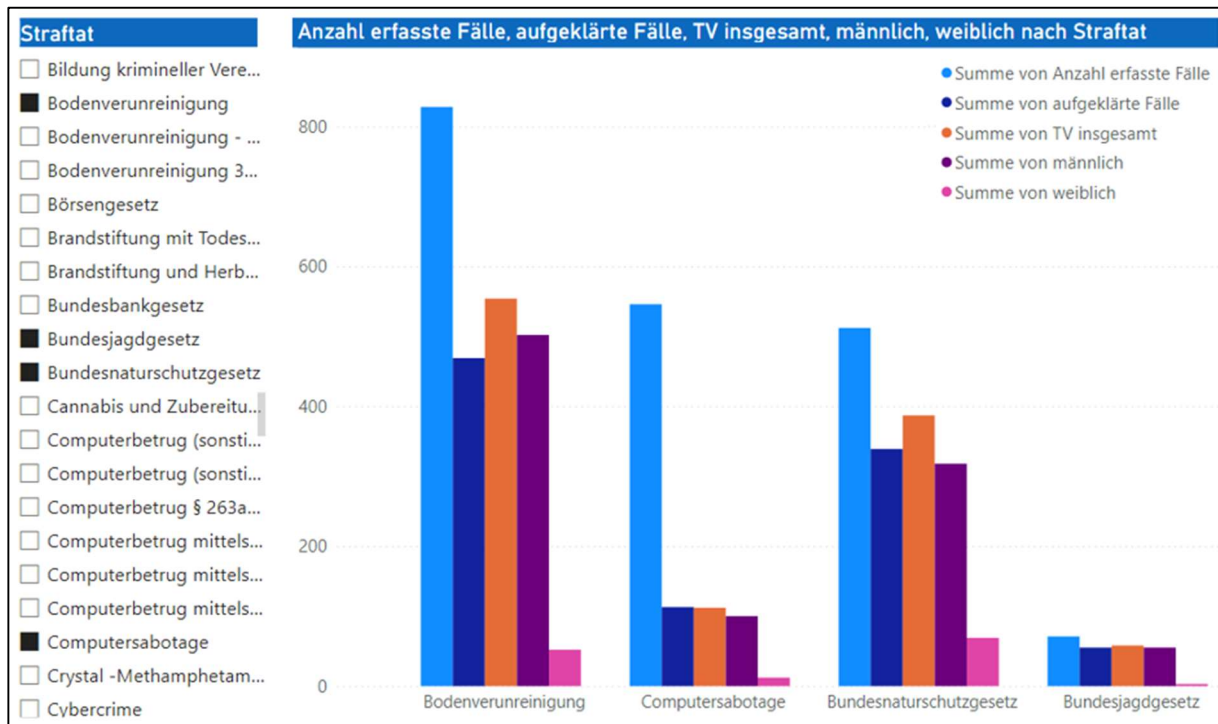


Abbildung 8: Säulendiagramm mit Datenschnitt

2.2.3. Landkarte & Drilldown

Mit der Visualisierung „Landkarte“ lassen sich Verhältnisse zwischen verschiedenen Orten darstellen. In der Karte unten wird durch Blasen unterschiedlicher Größe die jeweilige Anzahl von Straftaten in verschiedenen Städten gezeigt.

In der zunächst zugrundeliegenden Tabelle sind die Straftaten nur nach Städten aufgeschlüsselt. In einer anderen Tabelle sind die Städte ihren Bundesländern zugeordnet. Führt man diese beiden Tabellen zusammen (siehe Kapitel 2.5) erhält man den Zusammenhang zwischen den Straftaten-Zahlen und den Bundesländern. Dadurch ergibt sich die Möglichkeit in der Landkarte einen sogenannten Drilldown durchzuführen, also zwischen den Ebenen Bundesländer und Städte umzuschalten. In der anderen Richtung heißt es Drillup, bedeutet von Städten nach Bundesländern - also in der Hierarchie aufwärts - umzuschalten.

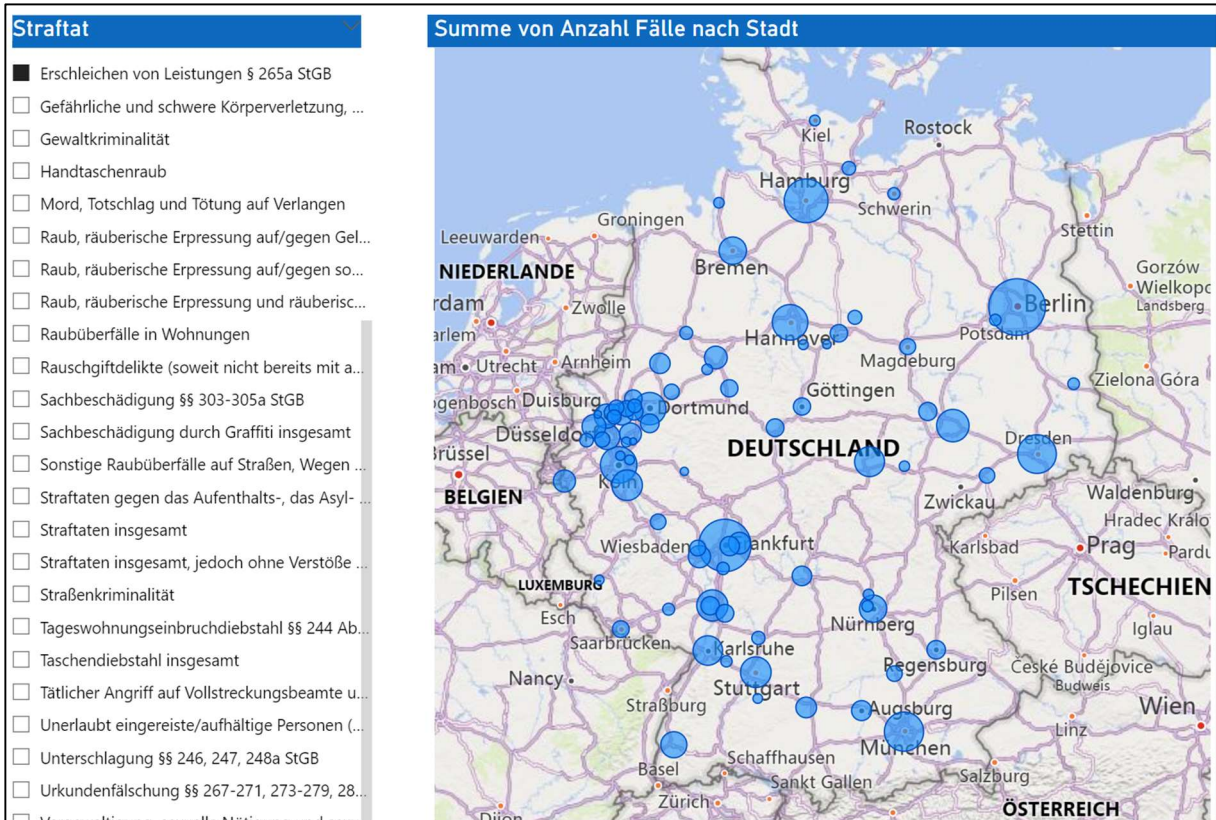


Abbildung 9: Landkarte mit Datenschnitt

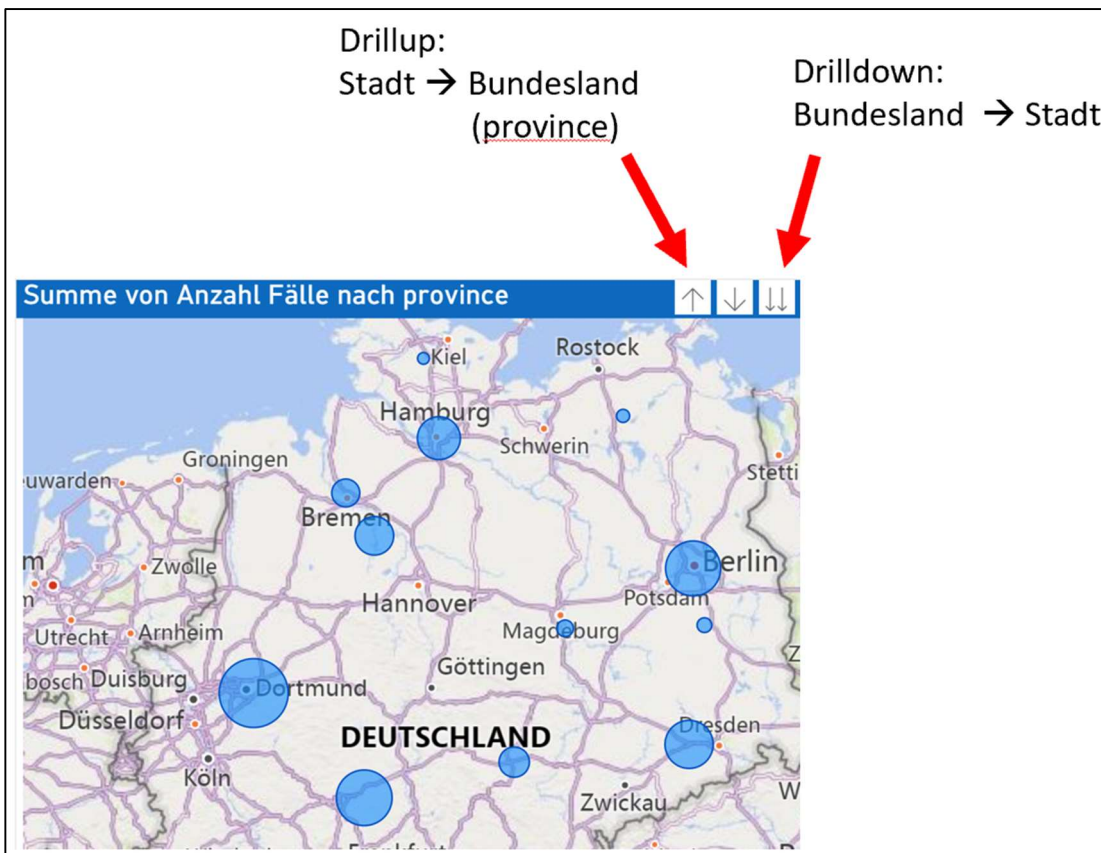


Abbildung 10: Drillup bzw. Drilldown

2.2.4. Liniendiagramm bzw. Zeitreihe

Ein weiteres klassisches Diagramm ist das Liniendiagramm, mit dem bevorzugt auch zeitliche Verläufe - so genannte Zeitreihen - dargestellt werden. Hier die Fallzahlen einer Straftat der Jahre 1987 bis 2020.

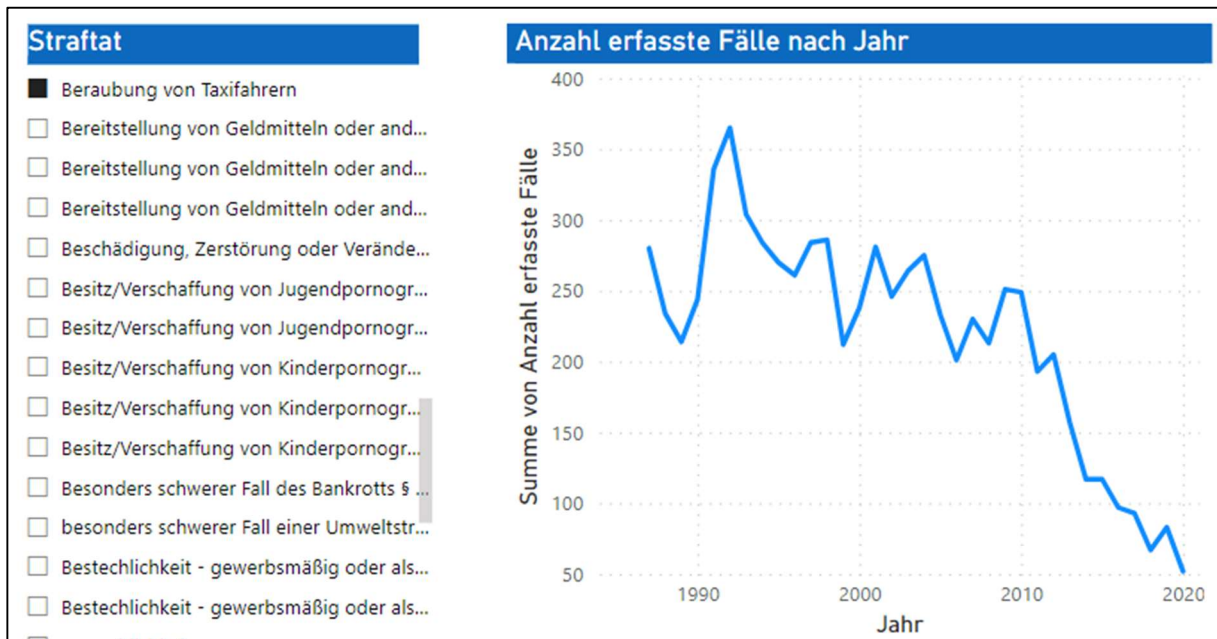


Abbildung 11: Zeitreihe mit Datenschnitt

3. Analyse mit Kennzahlen

3.1. Benutzerdefinierte Spalten & Measures

Im Power Query-Editor können auch eigene Spalten hinzugefügt werden. Dabei kann über eine Skriptsprache eine sehr große Anzahl an Vorgängen ausgeführt werden. Hier wird eine Fallunterscheidung mit einer if-else-Verzweigung angewendet. Ein Feld soll mit „gleich“, „mehr“ oder „weniger“ befüllt werden, abhängig davon, ob die Zahl der Tatverdächtigen (TV) gleich, größer oder kleiner als die Fallzahl ist.

```
#"Hinzugefügte benutzerdefinierte Spalte" = Table.AddColumn(#"Entfernte untere Zeilen",  
"TV zu Fälle",  
each if [Anzahl erfasste Fälle] > ([Tatverdächtige insgesamt]*0.95)  
and [Anzahl erfasste Fälle] < ([Tatverdächtige insgesamt]*1.05) then "gleich +/-5%"  
else if [Anzahl erfasste Fälle] < [Tatverdächtige insgesamt] then "mehr"  
else if [Anzahl erfasste Fälle] > [Tatverdächtige insgesamt] then "weniger" else "0 Fälle")
```

Abbildung 12: if-else-Verzweigung

Als weitere Funktion kann man in Power Bi Desktop sogenannte ‚Measures‘ erstellen – also Messwerte oder Metriken. Diese können mittels der DAX-Formelsprache geschrieben werden. Hier wird die hinzugefügte Spalte „TV zu Fälle“ verwendet und in drei Measures die Zeilen mit „mehr“, „weniger“ und „gleich“ gezählt.

```
TV mehr als Fälle = CALCULATE(COUNTROWS('PKS-2022_Bund-Falltabellen-T01'),  
'PKS-2022_Bund-Falltabellen-T01'[TV zu Fälle] = "mehr")  
  
TV weniger als Fälle = CALCULATE(COUNTROWS('PKS-2022_Bund-Falltabellen-T01'),  
'PKS-2022_Bund-Falltabellen-T01'[TV zu Fälle] = "weniger")  
  
TV gleich Fälle = CALCULATE(COUNTROWS('PKS-2022_Bund-Falltabellen-T01'),  
'PKS-2022_Bund-Falltabellen-T01'[TV zu Fälle] = "gleich +/-5%")  
  
keine Fälle = CALCULATE(COUNTROWS('PKS-2022_Bund-Falltabellen-T01'),  
'PKS-2022_Bund-Falltabellen-T01'[TV zu Fälle] = "0 Fälle")
```

Abbildung 13: Measures

3.2. Werte der erstellten Kennzahlen

Die Zahl der gelisteten Straftaten - also die reinen Erfassungsstraftaten ohne Summenstraftaten - beträgt 782. Diese verteilen sich wie folgt auf die oben erstellten Measures:

147	467	112	56
TV mehr als Fälle	TV weniger als Fälle	TV gleich Fälle	keine Fälle

Abbildung 14: Kennzahlen, Vergleich „TV zu Fälle“

3.3. Aufklärung in Abhängigkeit des Verhältnisses „TV zu Fälle“

Als weitere Kennzahl wird die durchschnittliche Aufklärungsquote erhoben, zunächst über alle Straftaten. Anschließend wird mittels der Spalte, die das Verhältnis Zahl der Tatverdächtigen zu Zahl der Fälle kategorisiert, gefiltert.

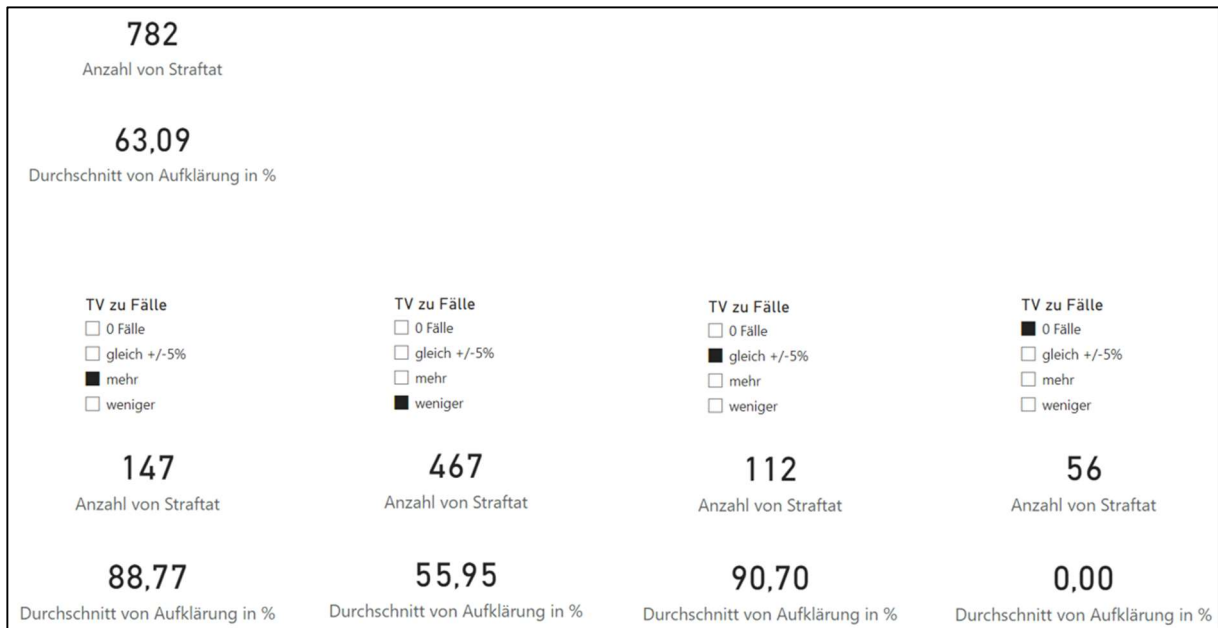


Abbildung 15: Kennzahlen, Straften und Aufklärungsquoten

3.4. Weitere Auswertung

Vergleich der Anzahl an maximalen Tatverdächtigen mit der Aufklärungsrate nach Altersgruppen über alle Straftaten.

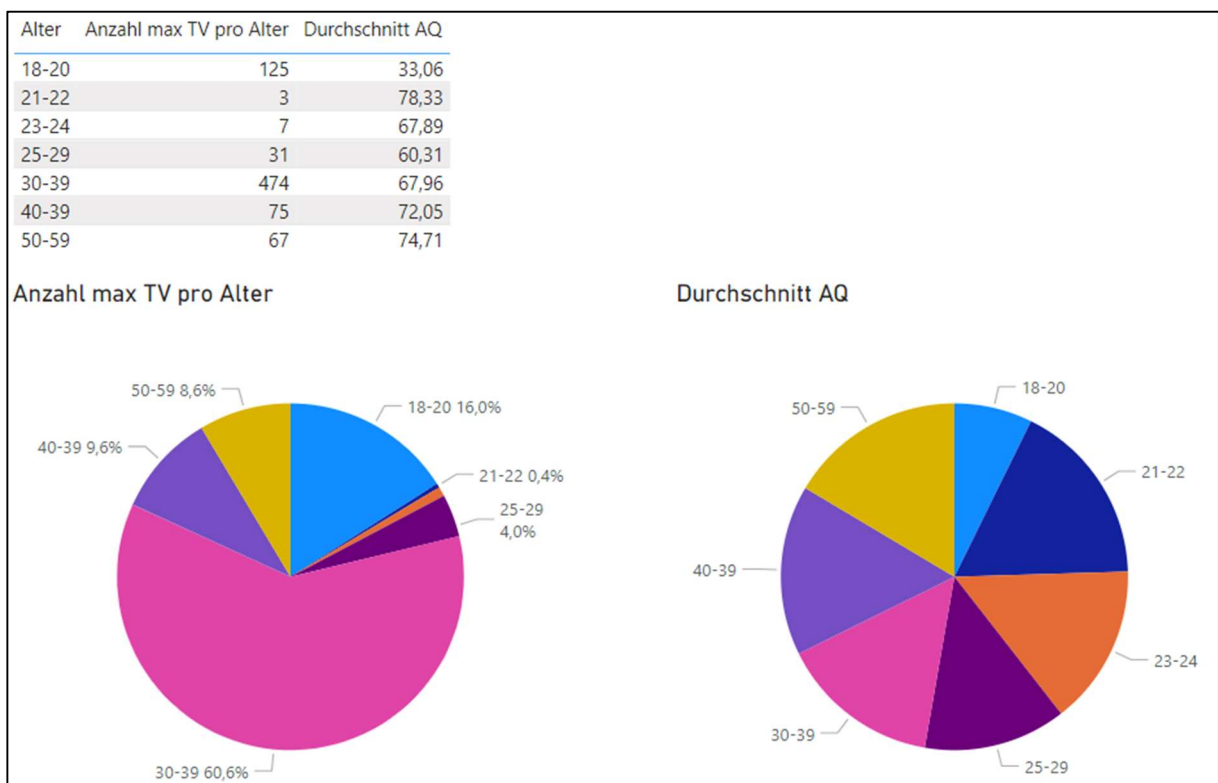


Abbildung 16: Auswertung, TV und Aufklärungsquoten

3.5. Ergebnisse

Die erhobenen Kennzahlen verraten also, dass es bei den meisten Straftaten weniger Tatverdächtige als erfasste Fälle gibt. Weiter zeigt sich, dass bei den Straftaten mit mehr Tatverdächtigen die Aufklärungsrate höher ist. Dies war auch zu erwarten. Die Aufklärung ist ebenfalls hoch, liegen gleich viele Tatverdächtige wie Fälle vor. Wobei hier eine Toleranz von $\pm 5\%$ berücksichtigt wurde, da sich bei exakter Übereinstimmung eine deutlich niedrigere durchschnittliche Aufklärungsrate zeigte, was verfälscht erschien.

Bei dem Vergleich der Anzahl der maximalen Tatverdächtigen nach Altersgruppe mit der entsprechenden Aufklärungsquoten hätte man erwarten können, dass dies in irgendeiner Weise korreliert. Also das es - kurz gesagt - bei der Altersgruppe mit den meisten Verdächtigen auch die meisten Aufklärungen gibt. Dem ist aber nicht so. Die Quoten sind annähernd gleichmäßig verteilt, lässt man die Gruppe der 18- bis 20-jährigen außen vor.

4. Fazit und Ausblick

Was hier im kleinen Stil gemacht wurde, ist natürlich auch mit wesentlich größeren Datenmengen und Strukturen möglich. Ebenso können komplexere Auswertungen durchgeführt werden. Oft werden Analysen regelmäßig wiederholt, um beispielsweise Kennzahlen permanent zu beobachten. Da die Statistiken vom BKA jährlich veröffentlicht werden, könnte man diese über Jahre analysieren und auswerten welche Veränderungen es gibt.

5. Abbildungsverzeichnis

Abbildung 1: Grundtabelle in Excel, Auszug	4
Abbildung 2: Grundtabelle im Power Query-Editor, Auszug	5
Abbildung 3: Modellansicht in Power BI Desktop	6
Abbildung 4: Tabelle, gruppiert und aggregiert, Auszug	6
Abbildung 5: Tabellen zusammenführen	7
Abbildung 6: Tabelle mit Datenschnitt, Auszug	8
Abbildung 7: Tabelle mit verknüpften Einträgen, Auszug	8
Abbildung 8: Säulendiagramm mit Datenschnitt	9
Abbildung 9: Landkarte mit Datenschnitt	10
Abbildung 10: Drillup bzw. Drilldown	10
Abbildung 11: Zeitreihe mit Datenschnitt	11
Abbildung 12: if-else-Verzweigung	12
Abbildung 13: Measures	12
Abbildung 14: Kennzahlen, Vergleich „TV zu Fälle“	12
Abbildung 15: Kennzahlen, Straften und Aufklärungsquoten	13
Abbildung 16: Auswertung, TV und Aufklärungsquoten	13

6. Abkürzungsverzeichnis

AQ:	Aufklärungsquote
BKA:	Bundekriminalamt
CSV:	Comma-separated Values
DAX:	Data Analysis Expressions
KPI:	Key Performance Indicator
PKS:	Polizeiliche Kriminalstatistik
TV:	Tatverdächtige
XLS:	Excel Spreadsheet

7. Quellenverzeichnis

Daten der Polizeilichen Kriminalstatistik, 2022

- T01 Grundtabelle – Fälle (V1.0)
- T20 Tatverdächtige insgesamt nach Alter und Geschlecht (V1.0)
- T91 Opfer insgesamt nach Alter und Geschlecht (V1.0)
- T01 Grundtabelle – Fälle mit Häufigkeitszahl (HZ) – Städte (V1.0)
- T01 Grundtabelle - Fälle ab 1987 (V1.0)