Impacts of COVID-19 on criminality in Germany

Project Report

Student
Can Aktaran
0085356

Instructor
Mehmet Fuat Kina

CSSM502 – Data Analysis for Social Sciences (Python)
Koç University

20.01.2022

Contents

List of figures

List of tables

1 Introduction

The COVID-19 pandemic has triggered unprecedented changes in our daily lives, influencing various aspects of society. With widespread lockdowns and isolation measures, the way we work and interact has transformed significantly. While the surge in digitalization brought about remote work as the new normal, it's essential to acknowledge the adverse effects, particularly on the mental health of the younger population, who experienced an upswing in mental health issues like depression.

In the midst of these transformations, a pivotal question emerges: How did the COVID-19 pandemic impact criminality in Germany? This project aims to explore the intricacies of this question, examining shifts in criminal behavior before and after the pandemic. The objectives include analyzing crime rates, identifying patterns in specific crime types, and exploring the characteristics of individuals involved in criminal activities.

The changes in societal norms and routines during the pandemic may have contributed to shifts in criminal behavior. Prolonged isolation and the psychological toll, especially on the younger population, prompt questions about potential correlations between the pandemic and criminal activities. Did domestic violence escalate due to extended periods of isolation? Did the rise in digital interactions lead to an increase in internet-related crimes?

Therefore, this study will mainly focus on

- examining the criminality in Germany before and after the pandemic.
- analyzing what kind of crimes have seen an increase and which ones were declining.
- explore patterns regarding the characteristics of delinquents (e.g., age, gender, location, crime and socio demographics)

2 Data Pre-Processing

2.1 Data Sources

To comprehensively investigate the impact of the COVID-19 pandemic on criminality in Germany, the goal for this study was to draw from a diverse array of data sources.

2.1.1 Bundeskriminalamt

The Bundeskriminalamt (BKA) is Germany's national investigative police agency, operating under the Federal Ministry of the Interior. Tasked with addressing organized crime, terrorism, and cybercrime at the federal level, the BKA collaborates internationally and plays a vital role in counterterrorism efforts. The Police Crime Statistics (Polizeiliche Kriminalstatistik or PKS) published annually by the BKA provide detailed insights into various criminal activities, offering a comprehensive breakdown of crime data by type, region, and demographics (*Das BKA*, 2024).

2.1.2 Statistisches Bundesamt

The Statistisches Bundesamt (Federal Statistical Office) is the national statistical agency of Germany, operating under the Federal Ministry of the Interior. Its primary role is to collect, process, and disseminate statistical information to support evidence-based policymaking and public discourse. The agency provides comprehensive data on various aspects of German society, including demographics, economics, and social indicators. Whereas the Statistisches Bundesamt offers loads of data concerning the penitentiary system, none of the data sets could provide additional valuable insights, at least not in the scope of this study (*Statistisches Bundesamt*, 2024).

2.1.3 Local Police Departments

As another source regarding more in-depth regional data, the goal was to contact a number of police departments in and around Munich, Germany. However, after talking to a handful of police station representatives, it became clear that the data they accumulate is not intended for the public and therefore they sadly could not provide this study with any additional data.

2.1.4 Web Scraping

In the context of criminality statistics in Germany, web scraping involves the automated extraction of information from online sources related to publicly accessible sources, such as law enforcement reports, or government databases. This technique allows to gather real-time and detailed insights into criminal activities, helping to complement official reports and enhance the understanding of evolving crime trends and patterns in the country. In this case, the available data mainly originating from the BKA did not need to be scraped but was ready to download as .xlsx or .csv files. Other suitable sources of data were either not publicly attainable (e.g. local police departments) or supply too little data points to complement the level of detail found in the BKA's data (e.g. Kaggle or newspaper reports). As a result, web scraping was not used in the course of this research.

2.2 Additional factors

I chose to analyze crime data from 2017 to 2022, covering three years before and during the pandemic. This time frame provides a balanced view of historical crime trends and potential pandemic-related influences. Examining pre-pandemic years (2017-2019) establishes a baseline, while post-2019 data helps identify any changes during the pandemic (2020-2022). This approach offers insights into both consistent patterns and pandemic impacts on crime rates in Germany.

Moreover, due to availability reasons, all of the data mentioned and used in the course of this study, related to criminal suspects and not convicted crimes.

2.3 Preparing the data

Data preprocessing is a crucial phase in any research project, serving as the foundation for accurate analysis and meaningful insights. This chapter outlines the steps undertaken to prepare the raw datasets for analysis in the context of this project.

This study leveraged a series of meticulously curated Excel tables to house the diverse and intricate datasets collected from mainly the BKA database. Each of their annually published reports consist of over 20 tables, clustered in different categories such as thematic focus, spatial focus as well as time-series and population data. In each of these clusters there is data available for specific characteristics, e.g., tables for German citizens or foreigners. However, for the scope of this study, the center of interest remained on the holistic view of criminality in Germany. As a result, the following tables were selected, which served as the foundational repositories for further analysis.

- Suspects per age group [3276x24] (*PKS 2017–2022*)
- Suspects per state per age group [10342x26] (*PKS 2017-2022*)
- Suspects per district per age group [50536x27] (*PKS 2017-2022*)
- Suspects per city per age group [50364x25] (*PKS 2017-2022*)
- Population per state [24x11] (*PKS 2018-2022*)

To start things with all of the Excel files were read from the directory and saved in their respective pandas df for each year named *age2018*, *state2018*, *district2018*, *city2018* and *pop2018.* Moreover, in order to optimize further utilization and batch processing of the data frames, each of them was added to a list, such as *age_list* or *pop_list*.

As all of these tables came from the same source, their structure, as seen in figure 1, was exactly the same, enabling a seamless data manipulation and preprocessing. This process involved a series of strategic measures to refine and distill the Excel tables for optimal analytical utility. To begin, the removal of the initial eight rows from each table was a crucial step, as these rows contained redundant information and headers that were extraneous to the analytical objectives. Subsequently, a meticulous column pruning was undertaken to eliminate data for age groups that were just too granular for the scope of this study, these included the columns *[key]*, *[6 up to 8]*, *[8 up to 10], [10 up to 12], [12 up to 14], [14 up to*

16], [16 up to 18], [juveniles 14<18 (c 11+12)], [21 up to 23] and *[23 up to 25]*. By doing so, a streamlined dataset with variables essential for the research questions at hand was ensured.

A10 fx ------

Police crime statistics
Suspect classification ba age and sex
V1.0 produced:11.03.2022

key	offence or offence category	sex	suspects number	children				
				up to 6	6 up to 8	8 up to 10	10 up to 12	12 up to 14
1	2	3	4	5	6	7	8	9
------	total offences	M	1,419,594	1,983	2,152	5,085	11,665	26,116
------	total offences	W	472,409	1,639	837	1,590	4,613	13,045
------	total offences	X	1,892,003	3,622	2,989	6,675	16,278	39,161
000000	offences against life	M	2,851	0	1	1	2	10
000000	offences against life	W	668	0	0	0	0	5
000000	offences against life	X	3,519	0	1	1	2	15
010000	murder (sect. 211 PC)	M	635	0	0	1	2	2
010000	murder (sect. 211 PC)	W	89	0	0	0	0	2
010000	murder (sect. 211 PC)	X	724	0	0	1	2	4
010079	other types of murder	M	584	0	0	1	2	2
010079	other types of murder	W	83	0	0	0	0	2
010079	other types of murder	X	667	0	0	1	2	4
011000	robbery attended with murder	M	43	0	0	0	0	0
011000	robbery attended with murder	W	5	0	0	0	0	0

Figure 1: Raw Excel data set from the BKA

Moreover, the enhancement of table readability and interpretability was accomplished through the judicious replacement of numerical headers (currently numbers from 1 to 24 as seen in row 9 in figure 1) with semantically meaningful labels. To succinctly encapsulate the nature of the data they represent, the following names were attached to the columns: *crime, gender, total_suspects, children_u_14, teenager_14_18, young_adult_18_21, total_u_21, total_21_25, adult_25_30, adult_30_40, adult_40_50, adult_50_60, adult_o_60, total_o_21.*

Furthermore, to ensure consistency and eliminate potential language-related discrepancies, a process of standardizing categorical values was undertaken. Specifically, gender notations within the dataset were homogenized, with the substitution of *W* and *w* by *f*, respectively. This harmonization simplifies the interpretation of gender-related variables, contributing to the overall coherence of the dataset. Additionally, variations in the representation of male and total numbers, denoted by *M* and *X*, were uniformly replaced with *m* and *x*, respectively. Additionally, instances of the German phrase *Straftaten insgesamt* were uniformly replaced with its English equivalent, *total offences*. After that, rows wherein the *crime* column denoted *Bund echte Zählung der Tatverdächtigen* or *Bundesrepublik Deutschland* were systematically removed, as they represented summarized values. Lastly, to conclude the measurements taken to streamline the raw data, a new column *year* was added to each data frame (excluding data frames regarding the population) containing the respective year. The final structure of each data frame, in this case for the year 2020, can be seen in figure 2.

	crime	gender	total_suspects	children_u_14	teenager_14_18	young_adult_18_21	total_u_21	total_21_25	adult_25_30	adult_30_40	adult_40_50	adult_50_60	adult_o_60	total_o_21	year
0	total offences	m	1481252	44010	119257	130224	293491	167521	187118	332705	222189	164136	114092	1187761	2020
1	total offences	f	488365	18658	43707	35809	98174	45257	55089	105778	74904	60523	48640	390191	2020
2	total offences	x	1969617	62668	162964	166033	391665	212778	242207	438483	297093	224659	162732	1577952	2020
3	offences against life	m	3037	10	127	290	427	397	440	683	442	350	298	2610	2020
4	offences against life	f	612	1	16	24	41	36	75	166	98	118	78	571	2020
5	offences against life	x	3649	11	143	314	468	433	515	849	540	468	376	3181	2020
6	murder (sect. 211 PC)	m	664	3	32	59	94	92	107	157	93	76	45	570	2020
7	murder (sect. 211 PC)	f	99	1	6	3	10	4	6	28	24	13	14	89	2020
8	murder (sect. 211 PC)	x	763	4	38	62	104	96	113	185	117	89	59	659	2020

Figure 2: Structure of data frame after preprocessing

To consolidate the disparate datasets into comprehensive entities, four distinct dataframes, namely *crimes_df*, *state_df*, *district_df* and *city_df* were concatenated. Moreover, to prepare further calculations, for each of the previously mentioned data frames, a filtered version was created to include only the total numbers per crime, hence only rows where *crime* equals *x*, namely *total_crimes*, *total_state_crimes*, *total_district_crimes* and *total_city_crimes*.

3 Data Analysis

3.1 Introduction

In this chapter, the exploration deepens as a dual approach of Exploratory and Comparative Data Analysis (EDA and CDA) is employed. Guided by EDA, the dataset is navigated, unveiling patterns, trends, and anomalies that reside beneath the surface. Complementing this, a CDA was conducted, systematically comparing values across different years, states, districts, and cities. As a result, a wide variety of different Key Performance Indicators (KPIs) were calculated, all of which aim to contribute to the research focus of this study, the impacts of COVID-19 on the criminality in Germany.

3.2 Calculations

3.2.1 Number of total crimes by men per year (EDA)

The total number of crimes that men were accused of doing for each respective year. To save the relevant data, a dictionary named *crimes_m_dict* was formulated. This dictionary features a key-value pair for each year, with the year serving as the key and the corresponding number as the value. Based on *crimes_df*, a new data frame containing only rows where *crime* equals *total offences*, *gender* equals *m* and *year* equals the particular year, was deducted. The number of total_suspects was then saved as value for its year in the *crimes_m_dict.*

3.2.2 Number of total crimes per year (EDA)

The total number of crimes that men and women were accused of doing for each respective year. Analogous to the number of crimes by men, the total number of crimes per year was derived. The specific values were saved in the *crimes_t_dict* dictionary.

3.2.3 Number of total crimes by women per year (CDA)

The total number of crimes that women were accused of for each respective year. Having extracted the number of total crimes and crimes by men, the number of crimes by women per year could be simply determined by subtraction. The resulting values were stored in the *crimes_f_dict* dictionary.

3.2.4 Percentage of total crimes by men per year (CDA)

The percentage of total crimes men were accused across the years between 2017 and 2022. As in prior steps, the absolute numbers were already extracted, calculating the ratio per year was straight forward. The fraction of crimes by men over the total number of crimes was saved into yet another new dictionary, called *crime_dtr_m_dict*. Each value was multiplied by 100 and rounded to two decimals.

3.2.5 Percentage of total crimes by women per year (CDA)

The percentage of total crimes women were accused of across the years. This value was simply derived from subtracting the percentage of male crimes from 1 before multiplying by 100 and rounding to two decimals. The derived values were stored in yet another dictionary, *crime_dtr_f_dict*.

3.2.6 Total number of crimes in all years (EDA)

The sum of total crimes per year. By iterating through *crimes_t_dict* and adding each year's value *the* respective number of total crimes overall years was stored in *all_crimes.*

3.2.7 Top ten most committed crimes per year (EDA)

The list of the ten most committed crimes per year. To derive the name of these ten most performed crimes per year, a new data frame *most_crimes* based on *total_crimes* was derived, which got the rows where *crime* equals *total offences* removed. In the next step, *most_crimes* was sorted descending per values in the *total_suspects* column. Finally, to extract the name of these ten crimes, *most_crimes* was filtered for each year, also dropping all other columns than *crime.* The resulting lists were then stored into the *most_crimes_yearly* dictionary.

3.2.8 Top ten most committed crimes per age group per year (EDA)

The list of the ten most committed crimes of each age group per year. Similarly to the previous extraction, a copy named *most_crimes_age* of *total_crimes* was derived that excludes rows in which the value for *crime* equals *total_suspects*. After that, *most_crimes_age* is iteratively filtered by year and age group, resulting in a data frame for every particular year-age-group pair which were stored in yet another dictionary: *most_crimes_age_yearly.* This time however, the key not only consists of the year, but also the age group, so that each data frame is individually saved (e.g., *total_o_21_2019*).

3.2.9 Number of crimes <100 and >10,000 per year (EDA)

The number of crimes people have been accused of more than a hundred times and more than 10,000 times per year. To do so, the *total_crimes* was filtered twice per year-iteration, once for *total_suspects* smaller than 100 and a second time for it being greater than 10,000. The emerged data frames were temporarily stored and further filtered (*year* and *crime* unequal to *total offences*) before the final number of rows was extracted. Finally, the values were stored in their respective dictionaries, *crimes_lt100_dict* and *crimes_gt10000_dict*.

3.2.10 Total crimes per state per year (EDA)

The total number of crimes in each of the 16 German states per year. To extract the individual crime numbers per state, this time the *total_state_crimes* was used as a starting point. This concatenated data frame was filtered for each of the years and for rows where *crime* equals *total_suspects*. Next, the remaining data frame was again purified by only keeping the *state* and *total_suspects* columns. Lastly, the data frame was stored in *crimes_state_yearly*.

3.2.11 **Total crimes per district per year (EDA)**

The total number of crimes in each of Germany's districts per year. The calculation for these numbers is very much the same as for the previous KPI but using *total_district_crimes* as the foundation and filtering for *district* instead. The final result was being stored in *crimes_district_yearly*.

3.2.12 **Ten districts with highest and lowest number of crimes per year (EDA)**

The ten districts with each the highest and lowest number of crimes per year. After sorting the *total_district_crimes* descending for their *total_suspects* values, both the first and last ten rows were extracted and stored into their respective dictionaries, *top10_crimes_district_yearly* and *least10_crimes_district_yearly*.

3.2.13 **Total crimes per city per year (EDA)**

The total number of crimes in each of Germany's cities with 100,000 inhabitants or more per year. Based on the *total_city_crimes* data frame, the values for crimes across cities were developed analogously to the ones for states and districts. Eventually, the derived numbers were stored in the *crimes_city_yearly* dictionary.

3.2.14 **Ten districts with highest and lowest number of crimes per year (EDA)**

The ten cities with each the highest and lowest number of crimes per year. As for the previous KPI, *total_city_crimes* were firstly sorted descending for their *total_suspects* values. Next both the first and last ten rows were extracted yet again and stored into their according dictionaries, *top10_crimes_city_yearly* and *least10_crimes_city_yearly*.

3.2.15 Average crime rate per state per year (CDA)

The ratio between crimes and population per state per year. This calculation involves two data frames for each year, one for the crimes per state and another one for the population numbers. The former was already extracted and stored in *crimes_state_yearly*, for the latter the *poplist* was used. To begin, each of the *poplist*'s data frames was filtered for their columns *state*, *total_suspects* and *x*. Next, these data frames were merged with *crimes_state_yearly* before another column called *ratio* was added. That column contained the fraction of the number of suspects over the population. Lastly, the derived data frames were stored into the *crimes_state_pop_yearly* dictionary.

3.2.16 Average crime rate per year (CDA)

The average crime rate per year. The overall crime rate per year could be easily derived by calculating the mean of *crimes_state_pop_yearly*'s values. The results were stored into the *crime_rate_yearly* dictionary.

3.2.17 Crime rate per state per year (CDA)

The ratio between crimes and population per state per year. Similarly to the previous KPI, the average crime rate per state per year was calculated. However, in order to prepare the data for further computation, the output differs. The final result consists of a data frame, namely *total_crimes_state_rate*, including the following columns: *crime*, *state*, *total_suspects*, *year*, *crime_rate* and *population.*

3.2.18 Top 50 yearly crimes (EDA)

The 50 most committed crimes per year. To begin, all rows in which *crime* equals *total offences* were removed from *total_crimes*. Next, the resulting data frame was sorted descending for their *total_suspects* values. After that, the data frame was filtered for each of the respective years before the first 50 rows were extracted and saved in the *top50_crimes* dictionary.

3.2.19 Year-over-Year change in total number of crimes (CDA)

The percentage change in total numbers of crimes for the 50 most committed crimes per year. The percentage change in growth is simply calculated by dividing the number of *total_suspects* of the previous year, by its number for the current year. However, in any case where the number of suspects has seen a negative growth, the fraction needs to be inverted and subtracted from 100 and negated, to indicate the percentage change was actually a negative growth compared to the previous year. The derived values were stored in the *top50_change_yearly* dictionary.

3.2.20 Cybercrime related crimes per year (EDA)

The total number of crimes related to cybercrimes per year. To start, *total_crimes* was filtered for rows in which *crime* included any given cybercrime related criminal offence. Next, the sum of values in the *total_suspects* column was extracted. Lastly, the remaining values were stored in the *cybercrimes_yearly* dictionary.

3.2.21 Year-over-year change in cybercrime related crimes (CDA)

The percentage change in total number of cybercrime related offences per year. Using *cybercrimes_yearly,* the year-over-year change is simply calculated by taking the fraction of total suspects of two consecutive years. The derived values are stored in yet another new dictionary: *cybercrime_change_yearly.*

3.3 Fact sheet

Finally, a comprehensive fact sheet was generated, presenting all calculated values for in-depth examination. This fact sheet serves as a tangible reference, allowing for a detailed and transparent view of the derived metrics. The full fact sheet can be found in the appendix.

3.4 Regression Analysis

Engaging in a Regression Analysis with *crime_rate* and *population* as independent factors and *total_suspects* as the dependent one was driven by a desire to dig into potential connections in the available data. While recognizing that we might not accurately predict the future suspect count with our current dataset, the goal here is to bring a scientific lens to the analysis. By using regression methods, we're not just looking for immediate predictions but creating a numeric framework to understand how crime rate, population changes, and their interaction might affect the total number of suspects. This analysis adds a layer of understanding about what influences suspect numbers, even if it acknowledges the challenge of precisely forecasting future outcomes. The interpretation of the Regression Analysis can be found in the next chapter, the full results in the appendix.

4 Interpretation

4.1 No significant increases in crime numbers due to the pandemic and isolation

The examination of the total number of crimes over the years reveals a lack of significant increase since the onset of the pandemic. However, it is crucial to acknowledge potential distortions in these values. Ongoing regulations and pandemic-related circumstances may have created an environment less conducive to criminal activities. Despite this observation, it is important to note that investigating the specific distortions in detail falls outside the scope of the current study. The total number of crimes for each year is as follows:

Table 1: Total suspects per year

Year	2017	2018	2019	2020	2021	2022
Total suspects	2,112,715	2,051,266	2,019,211	1,969,617	1,892,003	2,093,782

This table provides an overview of the general trend in total crimes, highlighting the need for a more in-depth exploration to fully grasp the potential impact of external factors on the observed patterns.

4.2 No significant shifts in spatial distribution of crimes across Germany

The analysis of total crimes per state over the six-year period from 2017 to 2022 reveals a relatively stable pattern without any significant peaks or distortions. Nordrhein-Westfalen consistently holds the highest number of total suspects across all years, reaching its peak in 2017 with 475,452 suspects and maintaining a similar range in subsequent years. Bayern and Baden-Württemberg also consistently feature among the states with higher total suspects, reflecting a certain degree of stability in crime rates. While there are fluctuations in the number of total suspects for individual states from year to year, these variations do not appear to follow a distinct trend or indicate any noteworthy spikes. Notably, states like Bremen consistently exhibit lower total suspects compared to others, contributing to the overall consistency in the ranking of states. This analysis suggests that, on the whole, the distribution of total crime suspects across German states remains relatively steady over the examined period, without any pronounced anomalies or extraordinary deviations.

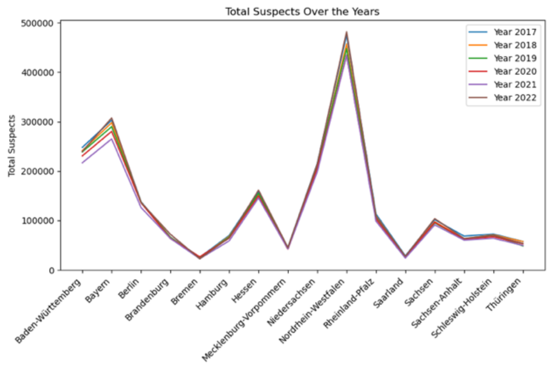

Figure 3: Annual crimes per state

4.3 Decrease in number of crimes of under 14-year-olds

The analysis uncovered a noteworthy trend indicating a significant decrease in crimes committed by individuals under 14 years old. For instance, the number of thefts dropped from 28,805 in 2019 to 21,354 in 2020. Similarly, the total number of crimes decreased from 66,907 in 2019 to 57,678 in 2020. While these figures might initially suggest compliance with COVID-19 related measures, it's essential to consider the broader context. The prolonged closures of stores and restaurants during the ongoing pandemic likely resulted in reduced opportunities for individuals under 14 to engage in criminal activities. When adjusting the numbers to account for these circumstances, the apparent decrease may be influenced by the limited time frames available for such activities. Therefore, further investigations are warranted to thoroughly examine and contextualize these findings, ensuring a nuanced understanding of the impact of pandemic-related measures on the criminal behavior of this specific age group.

4.4 Highest crime rate in the first year with COVID-19 in Bremen

An intriguing revelation surfaced during the analysis, identifying the state of Bremen as a notable outlier in the crime rate trends. In the year 2020, which marked the onset of the ongoing pandemic, Bremen exhibited a significantly higher crime rate compared to other years, surpassing even the leading state, Berlin. The following table illustrate this distinctive pattern:

Table 2: Annual crime rate for selected states

State	**2018**	**2019**	**2020**	**2021**	**2022**
Berlin	**3.76 %**	**3.73 %**	3.71 %	**3.45 %**	**3.71 %**
Bremen	3.53 %	3.54 %	**3.94 %**	3.37 %	3.37 %
Hamburg	3.70 %	3.61 %	3.49 %	3.18 %	3.52 %
Sachsen-Anhalt	2.87 %	2.83 %	2.83 %	2.78 %	2.94 %
Saarland	2.73 %	2.71 %	2.63 %	2.46 %	2.82 %

The values indicate that in 2020, Bremen experienced a distinctive surge in crime rate, surpassing all other states, including Berlin, which held the highest crime rate in other years. This unusual spike prompts further investigation into the specific socio-economic and regulatory dynamics within Bremen during the initial stages of the pandemic, as these factors may have contributed to the observed anomaly in crime rates. Understanding the underlying factors behind Bremen's exceptional crime rate in 2020 can provide valuable insights into the impact of external events on regional crime patterns.

4.5 Biggest increase in cybercrime in the first year with COVID-19

In the analysis of the Year-over-Year change in cybercrime related crimes (CDA), the goal was to examine the percentage variation in the total number of cybercrime related incidents annually.

The results underscored a substantial 7.02% surge in cybercrime-related offenses in 2020, marking this year as an outlier when compared to the relatively lower YoY changes in 2019 (4.69%) and 2021 (3.99%), as can be seen in table 3. This stark contrast in percentages emphasizes the exceptional nature of the cybercrime landscape in 2020, hinting at a pronounced increase during the initial phase of the ongoing pandemic. The disproportionately higher YoY change in 2020 prompts further exploration into the potential impact of regulatory measures, such as isolation protocols and restrictions on outdoor activities, on the surge in cybercrimes in Germany.

Table 3: Annual growth in cybercrime related suspects

Year	YoY-Change
2017	-
2018	2.95 %
2019	4.69 %
2020	**7.02 %**
2021	3.99 %
2022	-8.81 %

4.6 Regression Analysis

The R-squared value, a measure of how well the independent variables explain the variability of the dependent variable, is very low at 0.001. This indicates that only a minimal proportion of the total variation in *total_suspects* can be explained by the *crime_rate_lag1* and population_lag1.

The coefficients of the independent variables provide insights into their impact on the dependent variable. The constant (const) has a coefficient of 28.3103, though it is not statistically significant with a high p-value of 0.975. The *crime_rate_lag1* variable has a coefficient of 270.8980, and *population_lag1* has a coefficient of 4.499e-05. The p-values for these coefficients suggest that *population_lag1* is statistically significant, while *crime_rate_lag1* is not (p-values of 0.000 and 0.414, respectively).

The F-statistic of 30.96 tests the overall significance of the model, and its associated p-value is extremely low (3.64e-14), indicating that at least one of the independent variables has a significant effect on the dependent variable.

However, it's crucial to note that the R-squared is very low, suggesting that the model does not fit the data well. Additionally, the large condition number (2.45e+08) raises concerns about potential multicollinearity issues or numerical instability in the model. Further diagnostics and refinement of the model may be necessary for more reliable results. The entirety of results can be found in the appendix.

5 Conclusion

In conclusion, this study delved into the analysis of crime-related data in Germany from 2017 to 2022, offering valuable insights into trends and patterns. Key findings include variations in crime rates across states and notable changes during the ongoing pandemic, particularly in cybercrime-related offenses. However, it is essential to acknowledge the limitations of this study. The focus solely on suspects, not convicted individuals, and the reliance on reported data underline the need for caution in drawing definitive conclusions. Additionally, the study's scope is bounded by available data and may not capture the complete landscape of criminal activities. For future research, exploring the nuanced relationship between pandemic-related measures and crime rates, coupled with a more extensive dataset, could yield more comprehensive insights into the complex dynamics of criminal behaviors in Germany.

6 References

Das BKA. (2024). [Government Insitution]. Das Bundeskriminalamt. https://www.bka.de/DE/DasBKA/dasBKA_node.html

PKS 2017 Table 20: Suspect classification ba age and sex. (2020). Federal Republic of Germany.

PKS 2017 Table 20: Tatverdächtige nach Alter und Geschlecht—Kreise. (2018). Federal Republic of Germany.

PKS 2017 Table 20: Tatverdächtige nach Alter und Geschlecht—Länder. (2018). Federal Republic of Germany.

PKS 2017 Table 20_BKA: Tatverdächtige nach Alter und Geschlecht – Städte. (2018). Federal Republic of Germany.

PKS 2018: Bevölkerungsdaten insgesamt, deutsch, nichtdeutsch. (2020). Federal Republic of Germany.

PKS 2018 Table 20: Suspect classification ba age and sex. (2020). Federal Republic of Germany.

PKS 2018 Table 20: Tatverdächtige nach Alter und Geschlecht—Kreise. (2019). Federal Republic of Germany.

PKS 2018 Table 20: Tatverdächtige nach Alter und Geschlecht—Länder. (2019). Federal Republic of Germany.

PKS 2018 Table 20_BKA: Tatverdächtige nach Alter und Geschlecht – Städte. (2019). Federal Republic of Germany.

PKS 2019: Bevölkerungsdaten insgesamt, deutsch, nichtdeutsch. (2021). Federal Republic of Germany.

PKS 2019 Table 20: Suspect classification ba age and sex. (2020). Federal Republic of Germany.

PKS 2019 Table 20: Tatverdächtige nach Alter und Geschlecht—Kreise. (2020). Federal Republic of Germany.

PKS 2019 Table 20: Tatverdächtige nach Alter und Geschlecht—Länder. (2020). Federal Republic of Germany.

PKS 2019 Table 20_BKA: Tatverdächtige nach Alter und Geschlecht – Städte. (2019). Federal Republic of Germany.

PKS 2020: Bevölkerungsdaten insgesamt, deutsch, nichtdeutsch. (2021). Federal Republic of Germany.

PKS 2020 Table 20: Suspect classification ba age and sex. (2021). Federal Republic of Germany.

PKS 2020 Table 20: Tatverdächtige nach Alter und Geschlecht—Kreise. (2021). Federal Republic of Germany.

PKS 2020 Table 20: Tatverdächtige nach Alter und Geschlecht—Länder. (2021). Federal Republic of Germany.

PKS 2020 Table 20_BKA: Tatverdächtige nach Alter und Geschlecht – Städte. (2021). Federal Republic of Germany.

PKS 2021: Bevölkerungsdaten insgesamt, deutsch, nichtdeutsch. (2022). Federal Republic of Germany.

PKS 2021 Table 20: Suspect classification ba age and sex. (2022). Federal Republic of Germany.

PKS 2021 Table 20: Tatverdächtige nach Alter und Geschlecht—Kreise. (2022). Federal Republic of Germany.

PKS 2021 Table 20: Tatverdächtige nach Alter und Geschlecht—Länder. (2022). Federal Republic of Germany.

PKS 2021 Table 20_BKA: Tatverdächtige nach Alter und Geschlecht – Städte. (2022). Federal Republic of Germany.

PKS 2022: Bevölkerungsdaten insgesamt, deutsch, nichtdeutsch. (2023). Federal Republic of Germany.

PKS 2022 Table 20: Suspect classification ba age and sex. (2023). Federal Republic of Germany.

PKS 2022 Table 20: Tatverdächtige nach Alter und Geschlecht—Kreise. (2023). Federal Republic of Germany.

PKS 2022 Table 20: Tatverdächtige nach Alter und Geschlecht—Länder. (2023). Federal Republic of Germany.

PKS 2022 Table 20_BKA: Tatverdächtige nach Alter und Geschlecht – Städte. (2023). Federal Republic of Germany.

Statistisches Bundesamt. (2024). Obere Bundesbehörde Im Geschäftsbereich Des Bundesministeriums Des Innern Und Für Heimat. https://www.destatis.de/DE/Home/_inhalt.html

Appendix

Fact Sheet

Please find the fact sheet in its respective folder within the repository

Regression Analysis

```
22. Regression Analysis
                            OLS Regression Results
==============================================================================
Dep. Variable:         total_suspects   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                  0.001
Method:                 Least Squares   F-statistic:                     30.96
Date:                Sat, 20 Jan 2024   Prob (F-statistic):           3.64e-14
Time:                        16:09:40   Log-Likelihood:            -9.2700e+05
No. Observations:               88879   AIC:                         1.854e+06
Df Residuals:                   88876   BIC:                         1.854e+06
Df Model:                           2
Covariance Type:            nonrobust
===================================================================================
                      coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------
const              28.3103    895.712      0.032      0.975   -1727.277    1783.897
crime_rate_lag1   270.8980    331.329      0.818      0.414    -378.503     920.299
population_lag1  4.499e-05   5.75e-06      7.826      0.000    3.37e-05    5.63e-05
==============================================================================
Omnibus:                   228191.338   Durbin-Watson:                   0.717
Prob(Omnibus):                  0.000   Jarque-Bera (JB):       5771573768.660
Skew:                          29.286   Prob(JB):                         0.00
Kurtosis:                    1250.024   Cond. No.                     2.45e+08
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 2.45e+08. This might indicate that there are
strong multicollinearity or other numerical problems.
```