****

**BUCHAREST UNIVERSITY OF ECONOMIC STUDIES**

**CYBERNETICS, STATISTICS AND ECONOMIC INFORMATICS FACULTY**

SOFTWARE DEVELOPMENT FOR DATA ANALYSIS PROJECT

Criminality in the European Union: A Comprehensive Study

**Bucharest**

**2024**

**CONTENTS**

[1. Data source 3](#_Toc156084665)

[2. Time reference on dates 3](#_Toc156084666)

[3. Description of the variables 3](#_Toc156084667)

[4. Description of observations 4](#_Toc156084668)

[5. Data analysis approach 4](#_Toc156084669)

[6. The motivation to opt for a particular data analysis method 4](#_Toc156084670)

[7. Presentation of the results 4](#_Toc156084671)

[8. Interpretation 5](#_Toc156084672)

[9. Discussions and conclusions 7](#_Toc156084673)

[References 8](#_Toc156084674)

# Data source

Our goal is to examine crime within the European Union (EU) countries choosing a dataset that includes various criminal offenses. The data for this study was obtained from Eurostat [1]. Our model has 20 variables and 41 observations.

# Time reference on dates

The data selected represents values at the end of 2020. Choosing a recent year allowed us to make assumptions about the results easier, since we have some knowledge of the circumstances in different nations.

# Description of the variables

Our model includes 20 variables, chosen for studying criminality in various countries from the EU. Each variable represents a criminal offense, their abbreviations being displayed in table 3.1. All of the variables are referring to the number of convictions.

|  |  |  |
| --- | --- | --- |
| **No.** | **Label of the variable** | **Description of the variable** |
| 1 | IH | Intentional Homicide |
| 2 | AIH | Attempted Intentional Homicide |
| 3 | A | Serious Assault |
| 4 | Kd | Kidnapping |
| 5 | SV | Sexual Violence |
| 6 | R | Rape |
| 7 | SA | Sexual Assault |
| 8 | SE | Sexual Exploitation |
| 9 | Rb | Robbery |
| 10 | Brg | Burglary |
| 11 | BrgPRP | Burglary of Private Residential Premises |
| 12 | Th | Theft |
| 13 | ThMV | Theft of a Motorized Vehicle or Parts Thereof |
| 14 | Drg | Unlawful Acts Involving Controlled Drugs or Precursors |
| 15 | Frd | Fraud |
| 16 | Crp | Corruption |
| 17 | Br | Bribery |
| 18 | ML | Money Laundering |
| 19 | Acomp | Acts Against Computer Systems |
| 20 | OCGParticip | Participation in an Organized Criminal Group |

Table 3.1. Variable labels and descriptions

# Description of observations

We chose a sample of 41 observations for our study, aiming to encompass as many countries from the EU as possible in order to have a better understanding of criminality across our country, neighbors and possible future destinations. The values of each country from the EU are provided in the *Offences.csv* or *Offences.xlsx* file, both located in the folder *dataIN*. We used the country names as the observations labels.

# Data analysis approach

The data analysis approach involves using statistical methods such as principal component analysis (PCA) and exploratory factor analysis (EFA) to explore relationships and patterns in criminal offenses across EU countries. PCA allows us to highlight the significant information regarding the overall data set, whilst EFA allows us to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called factors.

# The motivation to opt for a particular data analysis method

We chose these data analysis methods to look at the data because we want to find out more than just basic facts. PCA enables us to observe the information much more compactly, allowing us to have a detailed analysis on a large number of variables, whilst EFA helps us identify the hidden causes of the connections we observe. We are using these two methods to dig deeper and understand not only how crimes are connected, but also why they happen. These two methods allowed us to get a complete picture of criminal activity throughout the EU.

# Presentation of the results

First, we ran the PCA to learn more about the issues regarding crime in the EU.

On the correlogram of correlation most of our variables present strong and very strong relationships between one another.

The correlogram of scores suggests that France and Germany differ the most from the rest of the countries in the EU, having the highest number of convicts.

Moving on to the eigenvalues graphic, which displays the principal components, we can observe that we have C1, C2, C3 and C4 situated above the red line, emphasizing their importance in capturing relevant information.

In our study, by using a dataset which contains observations about various criminal offenses from the EU, we were able to draw the following conclusions about the principal components.

The first component represents the criminalities from two popular tourist destinations, known for their varying landscapes, historical sites, and rich cultural legacy, both in geographical proximity: Germany, and France.

The second component mainly characterizes France, a Western European country with a rich history of security challenges, highlighting the criminality from that country.

The third component represents the criminalities from a country with iconic landmarks, infamous for its tourist pickpocketing: Italy.

The last component, the fourth one, characterizes criminality in a Nordic country, known for its safety and low crime rate: Sweden.

For further analysis, we examined the observation contributions to the axes variance, correlogram of quality of observations, and correlogram of factor loadings in order to draw assumptions regarding our model.

* For the first principal component:

Physical crimes seem to be the main problem of Germany and France. This may be due to light punishments or due to law enforcement not doing its job properly. An additional cause regarding violence could be the way people were raised in these countries, preferring not to take action against their abuser or not to help victims of violence.

* For the second principal component:

Participation in an organized criminal group and rape seem to represent most of the issues related to criminal activity in France. Being more focused on terrorism could have led to this, as they were paying more attention to possible terrorist attacks from outside and ignored offenses that take place inside their country. Moreover, mass immigration can bring in different kinds of people, and if the French government does not properly confront these problems, people will continue to do these offenses.

* For the third principal component:

Italy seems to have a problem with burglary of private residential premises and theft of motorized vehicles or parts of them. This may be due to the fact that a lot of tourists come to Italy, and they are an easier target for thieves. Another reason may be that a lot of foreigners are choosing Italy as their new place of residence, and the country takes in all kinds of people, some of them choosing to steal for securing an additional income.

* For the fourth principal component:

Sweden seems to be tied to corruption and money laundering. Maybe the country has had or still has payment inequality problems, favorizing some citizens over others.

The correlogram of communalities displays the quantity of variance explained in common by a group of principal components. Variables with high communalities contribute significantly to the factors identified by PCA. They play a substantial role in representing the underlying structure of the data.

Moving on to EFA, we used in our study the eigenvalues, the Bartlett sphericity test, we computed the Kaiser-Meyer-Olkin (KMO) indices and we made interpretations on the significant factor by using the correlogram of factor loadings. All of these are detailed in the following chapter. The correlogram of quality of observations and the observation contributions to the axes variance have the same interpretation as the ones displayed in the PCA, having used the same dataset.

# Interpretation

The PCA results provide insight into the main variables influencing the trends in criminal offenses throughout the EU.

By looking at the correlation correlogram, we were able to identify strong and very strong relationships between various criminal offenses.

Sexual violence, rape, and implicitly sexual assault are strongly related to attempted intentional homicide, meaning that certain people are more likely to have their lives endangered in relationships where there is an imbalance of power.

The strong relationship between serious assault and theft, rape, and sexual violence indicates that these three actions often lead to serious assaults.

The relationship between fraud and money laundering is strong, showing that fraud-related criminal activity generates money that needs to be washed. Moreover, the very strong relationship between fraud and kidnapping suggests that individuals frequently mislead others, generating criminal activity. Fraud can include unexpected money scams and romance scams.

The correlogram of scores suggests that France and Germany differ the most from the rest of the countries in the EU, having the highest number of convicts.

Moving further to the eigenvalues, four principal components stand out, having an explained variance of over 1. We will further continue with highlighting their meaning, delving deeper into analyzing criminality in the EU.

By studying the principal components in depth, we were able to explore the problems related to criminality from other countries, along with what may cause them or why these problems persist.

* For the first principal component:

We found out that Germany and France are mainly connected to physical crimes. Further studying Germany’s penal systems and sentencing practices, we discovered that Germany has light punishments for physical crimes, such as imprisonment not exceeding five years or a fine for physical assaults. Another example consists of theft and burglary, which presume imprisonment not exceeding one year or a fine.

* For the second principal component:

As a significant number of people have been caught in France for participation in an organized criminal group [2], France’s security seems to keep up to its former appearances, the country being not only threatened by potential terrorist attacks, but also from inside, by allowing all kinds of immigrants in and letting serious assaults to take place in their country. This problem related to serious attacks is due to the fact that the penalty consists of imprisonment only if the victim is incapacitated for over 8 days or worse things happen, other abusers getting a fine. Addressing the issue related to rape in France, the country has a problem with this offence. Not only rape victims do not get any justice because only a few cases result in conviction, but they also have to deal with rigid administrative system and prejudices, which frequently influence processes and decisions [3].

* For the third principal component:

Italy is strongly linked to burglary of private residential premises and theft of motorized vehicles or parts of them. We conducted further investigations. Regarding vehicle theft, when you rent a vehicle in Italy, they don't insure you against theft if you visit certain cities, like Gaeta, Puglia, Cerignola, and Naples. The thieves remove all the car parts without the factory number to further sell them at higher or at their real price, then they sell the parts with the factory number at the lowest price. Moreover, most tourists who got robbed either of their car or personal belongings from their homes chose to stay at villas during their trip to Italy. Villas are more prone to robberies than apartments, having more access roads and being less guarded.

* For the fourth principal component:

Sweden is indeed tied to corruption and money laundering, which are both having a strong relationship of correlation. The country has had in the year prior to the one from our study (2019) an extremely high level of income inequality, displayed in the figure 8.1 with a Gini coefficient between 85.0 and 89.9 [4]. A Gini coefficient in this range indicates that while the majority of people have much lower incomes, a significant amount of the total income is distributed among a small fraction of the population. Corruption can disappear very hard in a nation, as demonstrated by the fact that it persisted in Sweden in 2020. All in one, corruption is fueled by wealth inequality, and capitalism is a system that encourages both of these things.

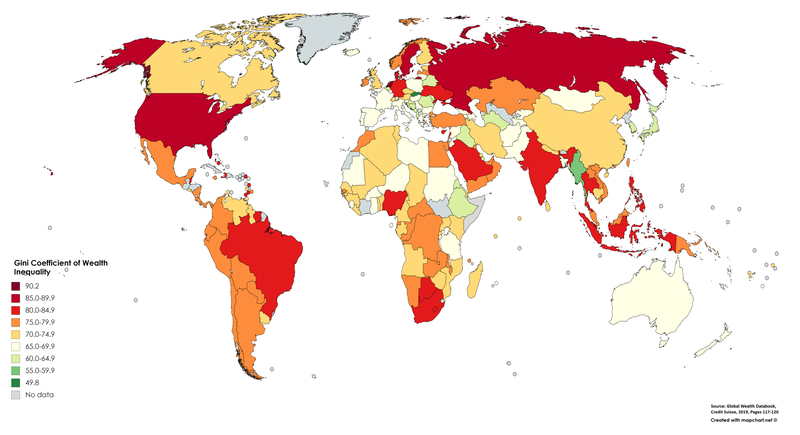


Figure 8.1. Gini Coefficient of Wealth Inequality (2019)

As high communality variables make a substantial contribution to the PCA-identified factors, we can observe in the correlogram of communalities that mainly sexual violence, sexual assault, and stealing-related criminalities, such as burglary and theft, contribute to the first four components.

Transitioning to EFA, the eigenvalues of the initial model and the eigenvalues of the model based on the extracted display the principal components relevant in explaining variation. In both cases, three components proved to be significant in variation explaining.

The Bartlett sphericity test indicates that there is at least one common factor who explains most of the model’s variation.

The Kaiser-Meyer-Olkin (KMO) indices specify the factorability of each variable of the model, helping in determining the suitability for factor analysis. The global KMO index has the value 0.7872562394354212, indicating an average potential of factorability.

After computing the value of the Bartlett statistic, we have identified one significant factor, compared to the 20 variables observed initially.

From the correlogram of factor loadings we can suppose that F1 might represent the level of poverty of a country. Criminal activity therefore increases with a nation's poverty.

This might explain why corruption, bribery, and money laundering yield lower values. If everybody is poor in the country, there are not that many opportunities for money-related crimes to occur.

Sexual assault, sexual violence and drug consumption are quite high in value: less money is allocated to the police in poorer nations, which means that drug prevention will receive less attention. Moreover, the victims of SA or SV might be under the influence of drugs, being unconscious when the act takes place, thus not being able to further take action and defend themselves.

# Discussions and conclusions

Through this study, our team wanted to gain insights into the criminal side of each country, by analyzing various criminal offenses. We discovered the depressing reality about many popular tourist spots, where laws are drafted only for paper use and are not actually implemented. Moreover, we noticed that the countries we analyzed have had these problems and still have these problems regarding criminality inside their borders.

Through our research we were able to observe what drives people to commit crimes. The discrepancies between people lead to misunderstandings between them, some choosing to fight for power through methods that are against the law

In conclusion, criminality is often found in highly populated countries, where there is inequality between people in means of income. With money being the principal reason for crime, convicts often find themselves in the category of individuals who profit from the weaknesses in laws and regulations. The victims, because of this money problem, are also put in a situation where justice cannot be done to them, the law enforcement officers not caring due to the income inequality present in their country.

# References

[1]

<https://ec.europa.eu/eurostat/databrowser/view/crim_off_cat__custom>

[2] <https://www.spectator.co.uk/article/the-gang-wars-of-paris/>

[3] <https://www.france24.com/en/tv-shows/france-in-focus/20230508-is-france-failing-its-rape-victims>

[4] <https://en.wikipedia.org/wiki/File:Gini_Coefficient_of_Wealth_Inequality_source_(2019).png>