Analysis of Crime Offences in European Countries (2022)

Aakriti Abhay Singh 24215296

Introduction

This report analyzes the dataset `crim_off_cat_2022.xlsx`, which records various offenses per hundred thousand inhabitants in 41 European countries for the year 2022. The objective of this report is to perform data manipulation and cleanup, and to derive insights regarding crime rates across different countries.

Task 1: Data Manipulation

1. Loading the Dataset

First, we will load the necessary library and import the dataset from the provided Excel file. We will skip the first eight rows to obtain the relevant data.

```
library(readxl)

dataset <- read_excel("crim_off_cat_2022.xlsx",
                      skip = 8,
                      n_max = 41)


head(dataset, 5)
```

```
# A tibble: 5 x 22
  `ICCS (Labels)` `Intentional homicide` `Attempted intentional homicide`
  <chr>           <chr>                  <chr>
1 Belgium         1.54                   10.29
2 Bulgaria        1.1100000000000001     0.67
3 Czechia         0.75                   0.6
```

```
4 Denmark            1                   2.38
5 Germany            0.74                2.0699999999999998
# i 19 more variables: `Serious assault` <chr>, Kidnapping <chr>,
#   `Sexual violence` <chr>, Rape <chr>, `Sexual assault` <chr>,
#   `Sexual exploitation` <chr>, `Child pornography` <chr>, Robbery <chr>,
#   Burglary <chr>, `Burglary of private residential premises` <chr>,
#   Theft <chr>, `Theft of a motorized vehicle or parts thereof` <chr>,
#   `Unlawful acts involving controlled drugs or precursors` <chr>,
#   Fraud <chr>, Corruption <chr>, Bribery <chr>, `Money laundering` <chr>, ...
```

2. Getting the size and structure of this dataset

Next, we will examine the size (number of observations and variables) and the structure of the
dataset to understand its composition.

```
dimensions <- dim(dataset)

cat("Number of observations (rows):", dimensions[1], "\n")
```

```
Number of observations (rows): 41
```

```
cat("Number of variables (columns):", dimensions[2], "\n\n\n")
```

```
Number of variables (columns): 22
```

```
cat("Structure of dataset: \n\n")
```

```
Structure of dataset:
```

```
str(dataset)
```

```
tibble [41 x 22] (S3: tbl_df/tbl/data.frame)
 $ ICCS (Labels)                 : chr [1:41] "Belgium" "Bulgaria" "(
 $ Intentional homicide          : chr [1:41] "1.54" "1.1100000000000(
 $ Attempted intentional homicide: chr [1:41] "10.29" "0.67" "0.6" "2
 $ Serious assault               : chr [1:41] "560.21" "45.78" "39.65
 $ Kidnapping                    : chr [1:41] ":" "1.21" "0.02" ":" '
 $ Sexual violence               : chr [1:41] "92.04" "8.289999999999
 $ Rape                          : chr [1:41] "37.03" "1.4" "16.29" '
```

```
 $ Sexual assault                                        : chr [1:41] "55.01" "6.89" "0.86" "
 $ Sexual exploitation                                   : chr [1:41] "18.18" "0.95" "9.01" "
 $ Child pornography                                     : chr [1:41] ":" ":" ":" ":" ...
 $ Robbery                                               : chr [1:41] "104.38" "13.19" "13.5:
 $ Burglary                                              : chr [1:41] "445" "63.46" "320.92" "
 $ Burglary of private residential premises              : chr [1:41] "336.8" "63.61" "56.03"
 $ Theft                                                 : chr [1:41] "1686" "413.08" "330.6!
 $ Theft of a motorized vehicle or parts thereof         : chr [1:41] "92.17" "21.99" "32.979
 $ Unlawful acts involving controlled drugs or precursors: chr [1:41] "484.04" "78.3199999999
 $ Fraud                                                 : chr [1:41] "977.18" "33.6199999999
 $ Corruption                                            : chr [1:41] "43.21" "7.81" "15.31" "
 $ Bribery                                               : chr [1:41] "0.59" "1.54" "1.71" "(
 $ Money laundering                                      : chr [1:41] "35.200000000000003" "(
 $ Acts against computer systems                         : chr [1:41] "82.15" "0.88" "27.08" "
 $ Participation in an organized criminal group          : chr [1:41] "2.94" "0.72" "0.21" "
```

3. Renaming the First Column

We will rename the first column to "Country" for clarity

```
colnames(dataset)[1] <- "Country"
colnames(dataset)
```

```
 [1] "Country"
 [2] "Intentional homicide"
 [3] "Attempted intentional homicide"
 [4] "Serious assault"
 [5] "Kidnapping"
 [6] "Sexual violence"
 [7] "Rape"
 [8] "Sexual assault"
 [9] "Sexual exploitation"
[10] "Child pornography"
[11] "Robbery"
[12] "Burglary"
[13] "Burglary of private residential premises"
[14] "Theft"
[15] "Theft of a motorized vehicle or parts thereof"
[16] "Unlawful acts involving controlled drugs or precursors"
[17] "Fraud"
[18] "Corruption"
```

```
[19] "Bribery"
[20] "Money laundering"
[21] "Acts against computer systems"
[22] "Participation in an organized criminal group"
```

4. Remove the following columns:

- Child pornography
- Rape
- Sexual assault
- Theft
- Theft of a motorized vehicle or parts thereof
- Burglary
- Burglary of private residential premises

we will install dplyr package

```r
library(dplyr)
```

```
Attaching package: 'dplyr'


The following objects are masked from 'package:stats':

    filter, lag


The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union
```

```r
dataset_clean <- dataset %>%
  select(-`Child pornography`,
         -Rape,
         -`Sexual assault`,
         -Theft,
         -`Theft of a motorized vehicle or parts thereof`,
         -Burglary,
         -`Burglary of private residential premises`)

colnames(dataset_clean)
```

```
 [1] "Country"
 [2] "Intentional homicide"
 [3] "Attempted intentional homicide"
 [4] "Serious assault"
 [5] "Kidnapping"
 [6] "Sexual violence"
 [7] "Sexual exploitation"
 [8] "Robbery"
 [9] "Unlawful acts involving controlled drugs or precursors"
[10] "Fraud"
[11] "Corruption"
[12] "Bribery"
[13] "Money laundering"
[14] "Acts against computer systems"
[15] "Participation in an organized criminal group"
```

5. Identifying Countries with Missing Data

we will identify countries that contain any missing data. We will replace any occurrences of
":" with NA and then count the number of missing values for each country.

```r
dataset_clean[dataset_clean == ":"] <- NA

dataset_clean$missing_count <- rowSums(is.na(dataset_clean))

countries_with_missing_data <- dataset_clean %>%
  filter(missing_count > 0) %>%
  select(Country, missing_count)


cat("Countries with missing data and their missing value counts:\n")
```

Countries with missing data and their missing value counts:

```r
print(countries_with_missing_data)
```

```
# A tibble: 27 x 2
   Country     missing_count
   <chr>               <dbl>
 1 Belgium                 1
```

```
 2 Denmark              3
 3 Estonia              1
 4 Ireland              2
 5 France               1
 6 Cyprus               1
 7 Latvia               1
 8 Luxembourg           2
 9 Hungary              2
10 Netherlands          6
# i 17 more rows
```

```r
dataset_clean <- dataset_clean %>%
  select(-missing_count)
```

6. Removing Countries with Missing Data

```r
dataset_final <- dataset_clean %>%
  filter(complete.cases(.))

cat("Dataset after removing countries with missing data:\n")
```

Dataset after removing countries with missing data:

```r
head(dataset_final , 5)
```

```
# A tibble: 5 x 15
  Country  `Intentional homicide` Attempted intentional homi~1 `Serious assault`
  <chr>    <chr>                  <chr>                        <chr>
1 Bulgaria 1.1100000000000001     0.67                         45.78
2 Czechia  0.75                   0.6                          39.6599999999999~
3 Germany  0.74                   2.0699999999999998           173.8
4 Greece   0.76                   1.6                          12.74
5 Spain    0.69                   2.56                         55.75
# i abbreviated name: 1: `Attempted intentional homicide`
# i 11 more variables: Kidnapping <chr>, `Sexual violence` <chr>,
#   `Sexual exploitation` <chr>, Robbery <chr>,
#   `Unlawful acts involving controlled drugs or precursors` <chr>,
#   Fraud <chr>, Corruption <chr>, Bribery <chr>, `Money laundering` <chr>,
#   `Acts against computer systems` <chr>,
#   `Participation in an organized criminal group` <chr>
```

7. Adding a Column for Total Offences

We will add a new column to our dataset that contains the overall record of offences for each country, calculated as the sum of the other offence-related columns.

```r
offence_columns <- dataset_final[ , -which(names(dataset_final) == "Country")]
offence_columns_numeric <- offence_columns %>%
  mutate_all(~as.numeric(.))

total_offences <- rowSums(offence_columns_numeric, na.rm = TRUE)

dataset_final$Total_Offences <- total_offences

cat("Updated dataset with total offences:\n")
```

```
Updated dataset with total offences:
```

```r
head(dataset_final)
```

```
# A tibble: 6 x 16
  Country  `Intentional homicide` Attempted intentional homi~1 `Serious assault`
  <chr>    <chr>                  <chr>                        <chr>
1 Bulgaria 1.1100000000000001     0.67                         45.78
2 Czechia  0.75                   0.6                          39.6599999999999~
3 Germany  0.74                   2.0699999999999998           173.8
4 Greece   0.76                   1.6                          12.74
5 Spain    0.69                   2.56                         55.75
6 Croatia  0.8                    2.8                          20.14
# i abbreviated name: 1: `Attempted intentional homicide`
# i 12 more variables: Kidnapping <chr>, `Sexual violence` <chr>,
#   `Sexual exploitation` <chr>, Robbery <chr>,
#   `Unlawful acts involving controlled drugs or precursors` <chr>,
#   Fraud <chr>, Corruption <chr>, Bribery <chr>, `Money laundering` <chr>,
#   `Acts against computer systems` <chr>,
#   `Participation in an organized criminal group` <chr>, ...
```

8. Final Size of the Cleaned Dataset

```
dimensions <- dim(dataset_final)

cat("Number of observations (rows):", dimensions[1], "\n")
```

```
Number of observations (rows): 14
```

```
cat("Number of variables (columns):", dimensions[2], "\n")
```

```
Number of variables (columns): 16
```

Data Summary

After cleaning the dataset, we have a total of 14 countries with no missing data and 16 variables (including the new total offenses column).

Key Observations:

- The country with the highest recorded offenses is **Germany** at **3553.38**, primarily driven by serious assault and fraud rates.

- **Spain** follows with **2440.44** offenses, while **Slovenia** has the lowest total at **665.64**.

- Notably, countries like **Italy** and **Finland** exhibit significant rates in various categories, warranting further investigation.

Task 2: Analysis

1. Table of Organized Criminal Group Participation

We will create a table displaying the countries along with their records of participation in an organized criminal group, sorted from highest to lowest values.

the columns (including **"Participation in an organized criminal group"**) are stored as character strings instead of numeric values. So firstly we need to convert these character strings to numeric values.

A.

Convert Character Columns to Numeric

Converting relevant columns to numeric

```
dataset_final <- dataset_final %>%
  mutate(across(c(`Participation in an organized criminal group`,
                  `Intentional homicide`,
                  `Attempted intentional homicide`,
                  `Serious assault`,
                  `Kidnapping`,
                  `Sexual violence`,
                  `Sexual exploitation`,
                  `Robbery`,
                  `Unlawful acts involving controlled drugs or precursors`,
                  `Fraud`,
                  `Corruption`,
                  `Bribery`,
                  `Money laundering`,
                  `Acts against computer systems`),
                as.numeric))
```

B.

Once the conversion is done, we can create the table and identify the country with the highest record as follows:

```
library(kableExtra)
```

```
Attaching package: 'kableExtra'
```

```
The following object is masked from 'package:dplyr':

    group_rows
```

```
organized_crime_participation <- dataset_final %>%
  select(Country, `Participation in an organized criminal group`) %>%
  arrange(desc(`Participation in an organized criminal group`)) %>%
  mutate(`Participation in an organized criminal group` =
           round(`Participation in an organized criminal group`, 1))
```

```r
kable(organized_crime_participation, caption =
        "Organized Criminal Group Participation by Country (2022)") %>%
  kable_styling(full_width = F, position = "left")
```

Table 1: Organized Criminal Group Participation by Country (2022)

Country	Participation in an organized criminal group
Spain	4.3
Greece	3.3
Romania	2.7
Albania	1.4
Austria	1.0
Croatia	0.9
Germany	0.8
Italy	0.8
Bulgaria	0.7
Czechia	0.2
Lithuania	0.2
Finland	0.0
Malta	0.0
Slovenia	0.0

2. Country with the Highest Participation

```r
highest_participation_country <- dataset_final %>%
  filter(`Participation in an organized criminal group` ==
          max(`Participation in an organized criminal group`, na.rm = TRUE)) %>%
  pull(Country)

cat("Country with the highest record of participation in
    an organized criminal group in 2022:", highest_participation_country, "\n")
```

```
Country with the highest record of participation in
    an organized criminal group in 2022: Spain
```

3. plot that visualizes the relationship between Robbery and Unlawful acts involving controlled drugs or precursors.

```
library(ggplot2)

ggplot(dataset_final, aes(x = `Robbery`,
                          y = `Unlawful acts involving controlled drugs or precursors`,
                          label = Country)) +
  geom_point(color = "blue", size = 2) +
  geom_text(vjust = -0.5, size = 3.5) +

  labs(title = "Relationship between Robbery and Unlawful Acts ",
       x = "Robbery (per 100,000 inhabitants)",
       y = "Unlawful Acts (per 100,000 inhabitants)") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5))  #
```

Relationship between Robbery and Unlawful Acts
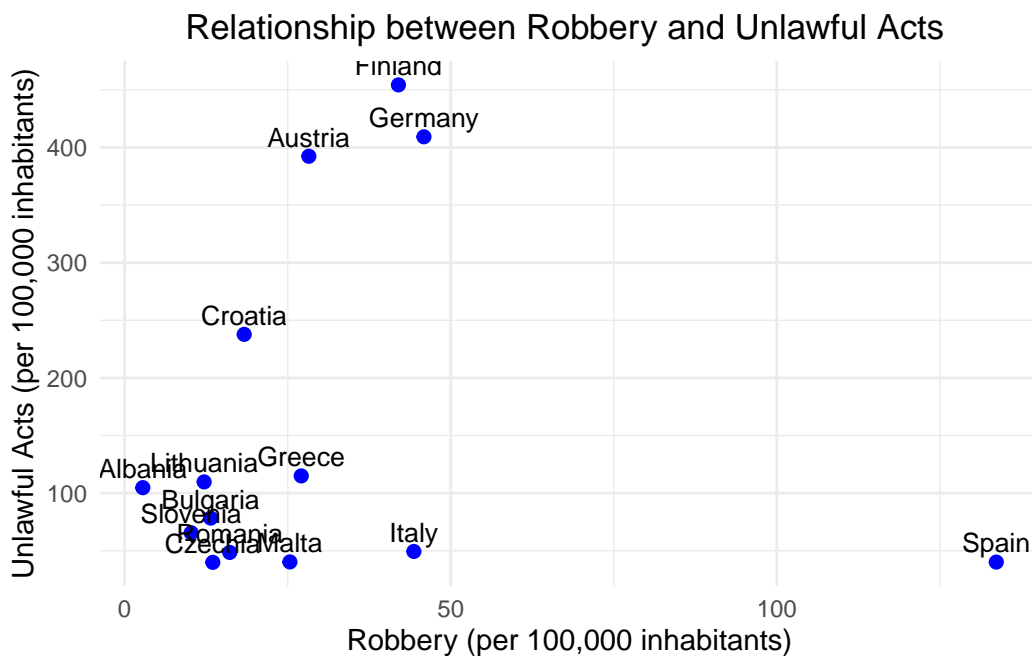
Conclusion

This analysis provides a detailed overview of crime offenses in European countries, highlighting critical insights and visualizations that can aid in understanding the dynamics of crime in the

11

region.

Task 3: Creativity

PLot1: Heatmap of Serious Assault, Robbery, and Fraud

The heatmap will visualize the relationships between serious assault, robbery, and fraud across countries, allowing us to see how these crime categories correlate with one another.

```
library(ggplot2)
library(reshape2)

crime_data <- dataset_final %>%
  select(Country, `Serious assault`, Robbery, Fraud)


correlation_matrix <- cor(crime_data[, -1], use = "pairwise.complete.obs")

melted_cor <- melt(correlation_matrix)

ggplot(melted_cor, aes(Var1, Var2, fill = value)) +
  geom_tile() +
  scale_fill_gradient2(low = "darkgreen", high = "pink", mid = "white",
                       midpoint = 0, limit = c(-1, 1),
                       name = "Correlation") +
  geom_text(aes(label = round(value, 2)), color = "black", size = 4) +
  labs(title = "Correlation Heatmap with Values: Serious Assault, Robbery, Fraud",
       x = "Crime Categories",
       y = "Crime Categories") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Correlation Heatmap with Values: Serious Assault, Ro
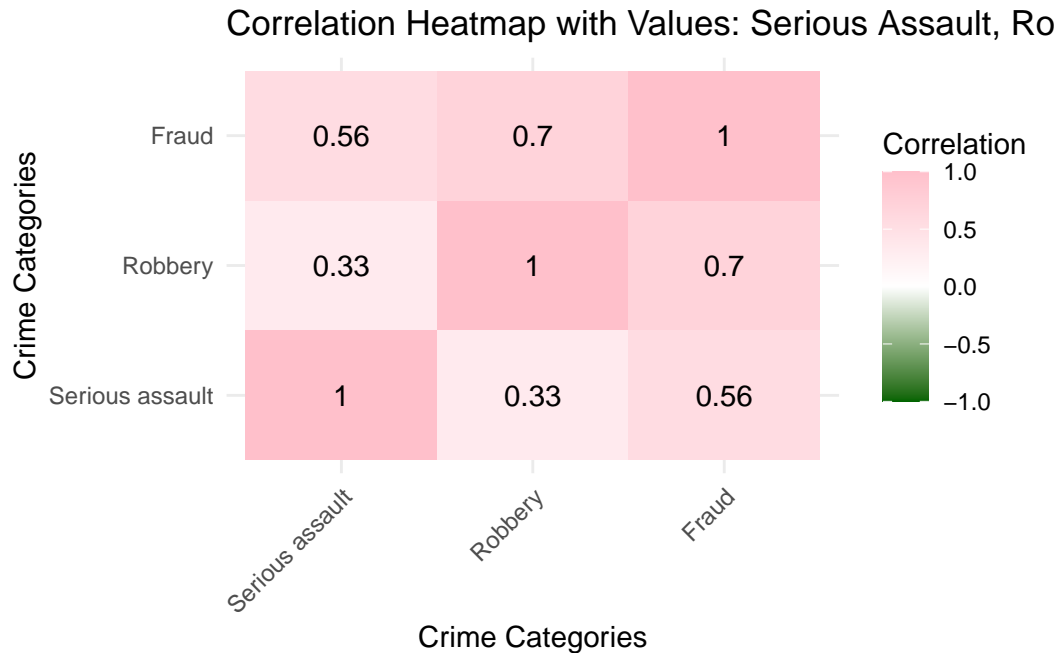
This Heatmap summarizes the pairwise correlations:

- The correlation between Serious Assault and Fraud is 0.56, which indicates a moderate positive relationship.
- Robbery and Fraud have a stronger positive correlation of 0.70.
- Serious Assault and Robbery have a lower positive correlation of 0.33.

Interpretation

- `Serious Assault and Fraud`: Moderate correlation suggests that countries with high levels of serious assault may also experience higher fraud rates.

- `Robbery and Fraud`: The strong correlation between robbery and fraud implies that these two crime types tend to occur together more frequently.

- `Serious Assault and Robbery`: The weaker correlation of 0.33 suggests that while they are positively related, the connection between these two crimes is not as strong as with fraud.

Plot 2: Top 5 Countries for Cybercrime and Drug-Related Offenses

```r
library(ggplot2)
library(reshape2)

top_5_cyber_drug <- dataset_final %>%
  arrange(desc(`Acts against computer systems` +
                `Unlawful acts involving controlled drugs or precursors`)) %>%
  select(Country, `Acts against computer systems`,
        `Unlawful acts involving controlled drugs or precursors`) %>%
  slice(1:5)

top_5_long <- melt(top_5_cyber_drug, id.vars = "Country",
                    variable.name = "Crime_Type", value.name = "Offenses")

ggplot(top_5_long, aes(x = reorder(Country, Offenses),
                        y = Offenses,
                        fill = Crime_Type)) +
  geom_bar(stat = "identity",
          position = position_dodge(width = 0.8), width = 0.7) +
  geom_text(aes(label = round(Offenses, 1)),
            position = position_dodge(width = 0.8),
            vjust = -0.5, size = 2.5) +
  scale_fill_manual(values = c("dodgerblue", "darkorange")) +
  labs(title = "Top 5 Countries for Cybercrime and Drug-Related Offenses",
      x = "Country",
      y = "Offenses (per 100,000 inhabitants)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

Top 5 Countries for Cybercrime and Drug–Related Offenses
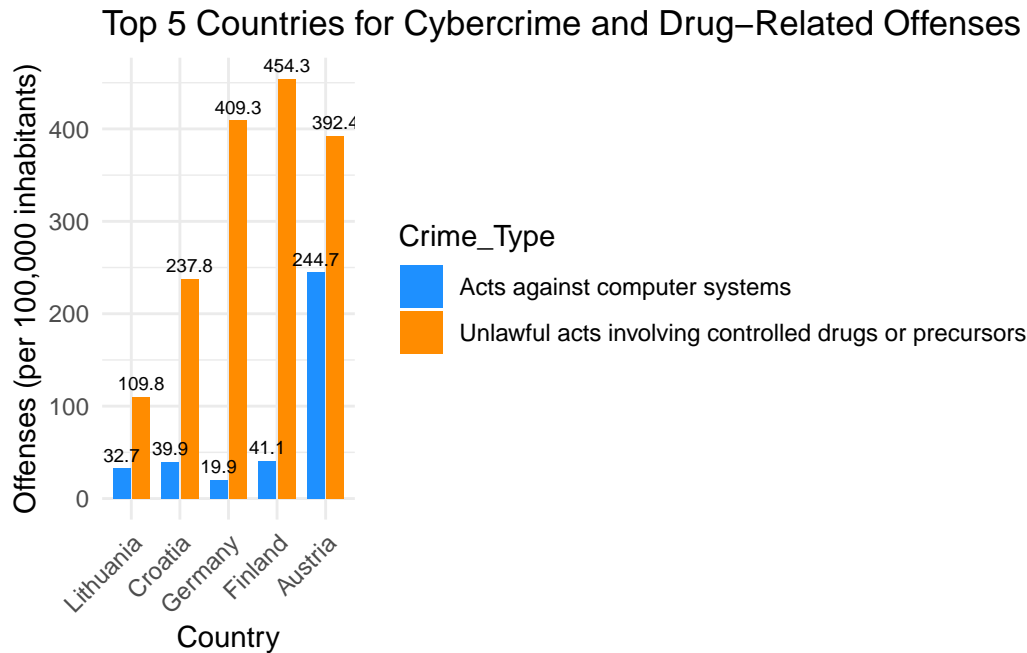
Insights

The bar plot highlights the top 5 countries with the highest recorded offenses in **cyber crime (Acts against computer systems)** and **drug-related offenses (Unlawful acts involving controlled drugs or precursors)**. These countries are **Lithuania**, **Croatia**, **Germany**, **Finland**, and **Austria** .

1. **Austria's Cyber crime Surge**: Austria is a major outlier in terms of cybercrime, with **244.7 offenses per 100,000 inhabitants**, significantly higher than the other countries. This suggests that Austria faces a unique challenge in the realm of cybercrime, perhaps due to higher internet penetration or sophisticated criminal networks targeting digital systems.

2. **Finland's Drug-Related Offenses**: Finland records the highest drug-related offenses at **454.3**. This could indicate higher drug consumption or stricter law enforcement practices targeting drug-related crimes.

3. **Germany's Balance**: Germany shows a moderate level of both cybercrime and drug-related offenses, indicating that it faces issues in both categories but is not an extreme outlier in either.

4. **Lithuania's Cyber Focus**: Lithuania has lower drug-related offenses (**109.8**) but still reports a moderate rate of cybercrime (**32.7**). This could suggest a focus on digital

crime prevention or that cybercrime is a growing threat in the country compared to drug offenses.

5. **Dual Crime Threat in Croatia and Finland**: Both **Croatia** and **Finland** report relatively high levels of both cybercrime and drug-related offenses. This suggests that these countries may face dual crime challenges and require comprehensive strategies to address both the digital and narcotics-related criminal activities.