

# **REPORT: STUDY OF THE AVERAGE NUMBER OF SUICIDES AND HOMICIDES IN THE WORLD DEPENDING ON THE INCOME OF THE COUNTRY**

**By: Amélie Meyer**  
25/03/2022

Date *last edit*:

## **INTRODUCTION:**

The purpose of this report is to give a **highly detailed explanation** of the step-by-step process of this analysis.

It will give **insights** as to the correlation (if one exists) between the number of **suicides and homicides** in the world depending of the **income of the country**.

Note: My analysis will be completed in **Microsoft Excel, Microsoft SQL Server Management Studio 18 and Tableau Public**.

## **PHASE 1: ASK**

The purpose of this analysis is to discover if there is any kind of **correlation, link or relationship** between the number of suicides and the number of homicides in the world depending on the **income** of the country.

## **PHASE 2: PREPARE**

In this analysis, the public data set "Homicide rate, Suicide rate and GPD" ([Homicide rate, Suicide rate and GDP | Kaggle](#)) will be used.

The dataset contains 50 years of data of the number of homicides, suicides and the GPD of most country around the world. This data is originally from World Bank Open Data.

- **Bias and Credibility**

The **ROCCC system can be used in order to determine** the credibility and extent to which the 'Homicide rate, Suicide rate and GCP data' is biased.

- Reliable: This dataset **is** reliable. Even if it contains secondary data collected by different organisations then posted on World Bank Open Data. As World Bank Open Data is a reliable source, so should be this data.
- Original: This dataset **is not** original as it was collected by different organisations.
- Comprehensive: This dataset **is** comprehensive as each number is linked to a year and country; we have enough information to understand the data.
- Current: The dataset **is not** current, the last year reported is 2020.
- Cited: This data set **is** cited and originally from World Bank Open Data, making it reliable.

The **data integrity and credibility** is sufficient to make the insight of this analysis reliable and comprehensive. Nevertheless, it is important to state again that this data is **slightly outdated**, and so can only represents the relationship between the homicide rate, suicide rate and the income of the country from 1970 to 2020.

## PHASE 3: PROCESS

Before analysing the data, the data's integrity has to be checked and the data has to be cleaned. To do so, the data will be cleaned in **Microsoft Excel**.

### - **Summary of the step taken:**

-Removing duplicates: using the action *remove duplicate*

-Renaming the columns *for example*: 'Intentional homicides (per 100,000 people)' to 'homicides'

-Checking the data for errors *for example*: finding region's names in the country's column and their income was classified as "Aggregates". The decision to remove that data was made.

-The data in the column "adminregion" was very incomplete, a lot of data was missing on the region the country was in. So, a new table was created using the data from this website: [List of Countries by Continent - StatisticsTimes.com](#). This table is to be joined with the first table on SQL to easily fix the problem of this missing data.

Note: Please note that the two tables that can be found in the GitHub repositories are the tables modified on excel for this analysis.

## PHASE 4: ANALYSE

The analysis itself was done using **Microsoft SQL Server Management Studio 18**. And here is the code written.

```
-- TABLES
```

```
SELECT *  
  
FROM PortfolioProject.dbo.SuicideHomicideGDP
```

```
SELECT *  
  
FROM PortfolioProject.dbo.CountryRegion
```

```
SELECT *  
  
FROM PortfolioProject.dbo.SuicideHomicideGDP AS SHG  
JOIN PortfolioProject.dbo.CountryRegion AS CoRe  
    ON SHG.country = CoRe.Country  
    AND SHG.iso3c = CoRe.iso3c
```

```
-- AVERAGE PER COUNTRY
```

```
SELECT country, AVG(CAST (homicides AS DECIMAL)) AS avg_homicides, AVG  
(CAST(suicides AS DECIMAL)) AS avg_suicides  
  
FROM PortfolioProject.dbo.SuicideHomicideGDP  
  
WHERE homicides IS NOT NULL AND suicides IS NOT NULL  
  
GROUP BY country  
  
ORDER BY avg_homicides DESC  
  
--ORDER BY avg_suicides DESC
```

-- AVERAGE PER REGION

```
SELECT CoRe.region, AVG(CAST (homicides AS DECIMAL)) AS avg_homicides, AVG  
(CAST(suicides AS DECIMAL)) AS avg_suicides
```

```
FROM PortfolioProject.dbo.SuicideHomicideGDP AS SHG
```

```
JOIN PortfolioProject.dbo.CountryRegion AS CoRe
```

```
    ON SHG.country = CoRe.Country
```

```
    AND SHG.iso3c = CoRe.iso3c
```

```
GROUP BY CoRe.region
```

```
ORDER BY avg_homicides DESC
```

--ORDER BY avg\_suicides DESC

-- AVERAGE PER CONTINENT

```
SELECT CoRe.continent, AVG(CAST (homicides AS DECIMAL)) AS avg_homicides, AVG  
(CAST(suicides AS DECIMAL)) AS avg_suicides
```

```
FROM PortfolioProject.dbo.SuicideHomicideGDP AS SHG
```

```
JOIN PortfolioProject.dbo.CountryRegion AS CoRe
```

```
    ON SHG.country = CoRe.Country
```

```
    AND SHG.iso3c = CoRe.iso3c
```

```
GROUP BY CoRe.continent
```

```
ORDER BY avg_homicides DESC
```

--ORDER BY avg\_suicides DESC

--AVERAGE GLOBAL PER YEAR

```
SELECT year, AVG(CAST (homicides AS DECIMAL)) AS avg_homicides, AVG (CAST(suicides  
AS DECIMAL)) AS avg_suicides
```

```
FROM PortfolioProject.dbo.SuicideHomicideGDP
```

```
WHERE homicides IS NOT NULL AND suicides IS NOT NULL
```

```
GROUP BY year
```

```
ORDER BY 1
```

--ORDER BY avg\_homicides DESC

```
--ORDER BY avg_suicides DESC
```

```
--INCOMES
```

```
SELECT DISTINCT country, income  
FROM PortfolioProject.dbo.SuicideHomicideGDP  
ORDER BY 2
```

```
--AVERAGE PER COUNTRY VS INCOMES
```

```
SELECT country, AVG(CAST (homicides AS DECIMAL)) AS avg_homicides, AVG  
(CAST(suicides AS DECIMAL)) AS avg_suicides, income  
FROM PortfolioProject.dbo.SuicideHomicideGDP  
WHERE homicides IS NOT NULL AND suicides IS NOT NULL  
GROUP BY country, income  
ORDER BY avg_homicides DESC  
--ORDER BY avg_suicides DESC
```

```
-- AVERAGE PER CONTINENT VS INCOMES
```

```
SELECT CoRe.continent, AVG(CAST (homicides AS DECIMAL)) AS avg_homicides, AVG  
(CAST(suicides AS DECIMAL)) AS avg_suicides, SHG.income  
FROM PortfolioProject.dbo.SuicideHomicideGDP AS SHG  
JOIN PortfolioProject.dbo.CountryRegion AS CoRe  
ON SHG.country = CoRe.Country  
AND SHG.iso3c = CoRe.iso3c  
--WHERE CoRe.continent = 'Europe'  
--WHERE CoRe.continent = 'Asia'  
--WHERE SHG.income = 'high income'  
GROUP BY CoRe.continent, SHG.income  
ORDER BY avg_homicides DESC  
--ORDER BY avg_suicides DESC
```

```
--AVERAGE PER REGION VS INCOMES
```

```

SELECT CoRe.region, AVG(CAST (homicides AS DECIMAL)) AS avg_homicides, AVG
(CAST(suicides AS DECIMAL)) AS avg_suicides, SHG.income

FROM PortfolioProject.dbo.SuicideHomicideGDP AS SHG

JOIN PortfolioProject.dbo.CountryRegion AS CoRe

    ON SHG.country = CoRe.Country

    AND SHG.iso3c = CoRe.iso3c

GROUP BY CoRe.region, SHG.income

ORDER BY avg_homicides DESC

--ORDER BY avg_suicides DESC

```

## PHASE 5: SHARE

Key points of the conclusion of this analysis:

*Note: the visuals for this study are available here: [Dashboard - Study Rate Homicide-Suicide/Income](#)*

-**El Salvador** is the country with the highest average of number of homicides, **Lesotho** is the country with the highest average of number of suicides

- **Central America** is the region with the highest average of number of homicides, **Southern Africa** is the region with the highest average of number of suicides.

-**North America** is the continent with the highest average of number of homicides, **Europe** is the continent with the highest average of number of suicides.

-**2002** was the year with the highest rate of homicide across the world, **2000** was the year with the highest rate of suicide across the world.

-There is a **clear correlation** between the suicide rate and the income of the country, as **the higher the income of the country, the higher the suicide rate is** as well. Nonetheless, the average amount of suicides around the world has been **decreasing** over the past 50 years.

-There is **no clear correlation** between the homicide rate and the income of the country, but it is clear to say that the country with **middle incomes** is where the average in homicides is the **highest**.

In conclusion, there is a **correlation** between a certain level of income and the average number of suicides and homicides. We can see a **trend** in the numbers around the world, but it is important to highlight that it is **not a causality**. So there are **many factors** that could explain the results of this analysis, income of the country being one of them. Based on the numbers found for the suicides rate across the world, a **new question** appear, is having no money in a high income society makes you more likely to resort to suicide ? Or is it the other way around ?

## PERSONAL CONCLUSION ON THE PROJECT

This project allowed me to **think outside of the box** in my opinion. I faced **many challenges** when trying to pursue this analysis. The first one was when I noticed that the region in my original dataset was for most countries empty, and no information on continents was present. I knew that I wanted to be able to explore my dataset looking at regions and continents as they are easier to look through than countries individually. Creating a new dataset and joining them using SQL was **easy and efficient**, it prevented me from looking through each country's location so I could manually fill my original dataset.

I also came across a few problems when writing my queries, but the challenge is the reason why I'm **passionate** about data analysis. With the help of the online community, or just trying to **change perspective** when you look at your problem, the reason why your query is not working can always be found. And there is no better feeling than having your query do what you wanted it to do. So I was able to **test and develop my knowledge** of the SQL languages and managed to **pursue my analysis** despite facing many challenges.