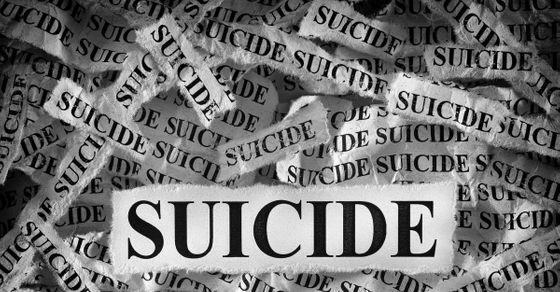
**DATA ANALYSIS REPORT ON SUICIDE RATES GLOBALLY.**



**BY**

**ESTRELLA.**

* ***ALLAN MUTISYA***
* ***HOSEA NJUGUNA***
* ***RUTH NDUTA***

# 

# 

# 

# **Business Understanding**

## **Business Overview**

Globally, close to 800,000 people die from suicide every year. That’s one person every 40 seconds. Due to the stigma associated with suicide – and the fact that it is illegal in some countries – this figure is also likely to be an underestimate, with some suicides being classified as unintentional injuries. This data analysis is done on data collected all over the world. Here we will see why and how suicide attempts are made and try to make a prediction for the same.

## **Business Objective**

The main objective of this project is to highlight the suicide rates across different generations ie age groups and investigate the pattern over the given time period between 1985 - 2016.

## **Assessing the situation**

1. Resource Inventory: Datasets: Suicide Rates Overview 1985 to 2016 -[[Link]](https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016/download%5D)
2. Software: Github, Google Collaboratory, Trello)
3. Assumptions: The data provided is correct and up to date
4. Constraints: There are no constraints

## **Data Mining Goals**

Our data mining goals for this project are as follows:

* Finding the country with the highest suicide rates
* Understanding which generation has the highest suicide rates
* Investigating which gender has more suicide rates compared to the other one
* Knowing the age which most people commit suicide
* Finding out the year where people committed suicide the most

# 

## **Data Mining Success Criteria**

Our success criteria will be measured by the following criteria;

* We give an objective account of the data analysis, with insights majorly coming from the dataset.

# 

# **Data Understanding**

## **Data Understanding Overview**

* The data for the project is retrieved from [Kaggle](https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016/tasks).
  + Master dataset - this dataset compares socio-economic info with suicide rates by year and country

## **Data Description**

For the dataset provided, here is a brief description.

* The original dataset - Master - consists of 12 columns and 27,820 rows. This is an extensive dataset with a lot of information on the research interest.
* The data is from 1985 to 2016.
* The data will be majorly cleaned and analyzed using python and SQL.

.

## **Verifying Data Quality**

The dataset has missing values on HDI for the year column and consequently it was dropped; leaving us with 11 columns.

# **Data Preparation**

These are the steps followed in preparing the data

## Loading Data

Loaded the datasets from the CSV and then created a python collab notebook from them.

## Cleaning Data

While doing data exploration, we noticed that the country-year column is redundant and irrelevant and had to be dropped. There was no evidence of any duplicated data. The ‘gdp\_per\_year ($)’ column was renamed to gdp\_per\_year and its datatype converted from string to float. The syntax errors including the commas in the numeric values were removed. The gdp\_per\_capita ($) column was renamed to gdp\_per\_capita.

There were no null values since we’d dropped the HDI year column. For uniformity purposes, all columns were converted to the same case. Outliers were removed using the interquartile range and a total of 8,012 records were dropped from 3 columns namely: Gdp\_for\_year, Suicides\_no, and Gdp\_per\_capita.

We added another column, to calculate the suicide rate by dividing the number of suicides by the total population of the specific countries; and labeled it as Suicide\_rate.

## Exporting the Cleaned Dataset

After cleaning the data we created a new dataset that has 19808 rows and 11 columns that were exported for further analysis.

# **Analysis**

To understand how the generations are grouped the table below explains the different age groups used in the dataset.

|  |  |
| --- | --- |
| G.I. Generation | 75+ years |
| The Silent Generation | 55-74 years |
| Baby Boomer Generation | 35-54 years |
| Generation X (Baby Bust) | 25-34 years |
| Millennials | 15-24 years |
| iGen / Gen Z | 5-14 years |

### *Finding the country with the highest suicide rates*

The country with the highest suicide rate is Bulgaria with 31114 cases and a scatter plot of the cases over the years is plotted on the python notebook.

Dominica, Saint Kitts, and Nevis have 0 cases hence the countries with the lowest number of suicide cases.

### *Understanding which generation has the highest and lowest suicide rates*

The Silent generation leads with the highest rates at 68.9% while Gen Z is the lowest at 0.69% on an overall scale.

Before 2000, the G.I generation leads with 45.04% while Millenials are at 0.74% being the lowest.

After 2000, the Silent generation led with 51.51% and the bar graph plotted to visualize the results. Gen Z had the lowest with 0.69%.

### *Investigating which gender has more suicide rates compared to the other one*

Upon investigating which gender has more suicide rates compared to the other one, we realized the male gender leads with 498619 cases while the female gender has 329940 cases. bar plots visualizing the suicide rate for both genders male and females were plotted to that effect

### *Knowing the age which most people commit suicide*

The age group leading in the rate of suicides was found to be 75+ years with the number of deaths at 77.86% overall and 39.00% before 2000, 38.86% after 2000.

### *Finding out the year where people committed suicide the most*

Most suicides were committed in the year 1999 with the death toll at 35636 with the number of females highest during this period.

For every 100,000 people in the countries, Lithuania leads in the rate of suicides by 8091.18 while Dominica, Saint Kitts, and Nevis were at 0. Consequently, for the same group, the age of 75+ years led at 77855.35.

# **Evaluation**

Based on our analysis, it is evident that European and Latin American countries e.g Bulgaria and Chile respectively lead in the rate of suicides. The common aspect among the states with lower suicide rates is that they have smaller economies and most of them being smaller island states.

In analyzing the rate of suicides committed among the 6 generations, we found that the older generations; the Silent and G.I generations are on the lead. This may be due to the reason that most of the population in these European countries is the older demographic.

The male demographic leads in the rate of suicide through the years across all generations. The reason is not well defined in our analysis but external research proves that the male gender is more susceptible to suicide than the female within the same age group or even older.

Countries with a higher population have a higher suicide rate for every 100,000 people compared to countries with a lower population which have a lower suicide rate for every 100,000 people within its population.

Countries with lower GDP have higher suicide rates and this may be due to low unemployment rates, high crime rates, and low quality of life.

# **Recommendations**

We made the following recommendations after the analysis and evaluation.

# 

1. The analysis revealed that European and Latin American countries are the most affected by the high rates of suicides. These governments need to put measures in place to ensure that there is a widespread sensitization on matters pertaining mental wellness and health and some of the reasons as to why people resort to suicide. Doing this will ensure that the population is conversant with this topic and can navigate through dealing with the causes that eventually drive someone to contemplate on suicide as an option.
2. The second recommendation would mainly focus on the older generations. From the analysis, it is evident that the older generation is the most affected by the suicide rates. Therefore, investing in healthcare, wellness facilities and nursing plans that specifically cater for the older generation would mitigate this negative occurrence. Furthermore, the family and caregivers of these individuals need to be active participants in their lives.
3. The male demographic tends to be on the receiving end of the suicide rates as they report high figures than that of their female counterparts. From external research, it is found that men resort to more dangerous means to commit suicide and they also tend to stay aloof with their suffering. There should be a sensitization across the board (inclusive of men and women) on the importance of sharing and opening up to others. Instead of using reactive measures to curb suicide, it would be prudent if we used the proactive approach that stops the acts from happening in the first place.
4. Furthermore, the evaluation showed us that countries with a higher population had lower suicide rates and vice versa. Bigger populations tend to have better welfare groups and also become more social than the smaller populations. In the case of a smaller population, we need to encourage the populace to interact more and enjoy the social life. In the case of a suicide attempt, one ought to reach out to a friendly person who would be more than willing to lend a hand. Also, these respective governments need to have special teams in place to assist avoid incidences that would prove to be disastrous to the person’s own life.
5. People usually derive their meaning from their purpose. Most people go to work to have a sense of direction and meaning in their lives. Countries with higher GDP have better economic opportunities and therefore, people have better employment opportunities. However, the same cannot be said for countries with lower GDPs. These countries are characterized by low employment rates, poor living standards, poor health conditions, increased drug abuse and an overall sense of despair within the community. The countries also lack the incentive and the framework to combat suicide given the lack of infrastructure and know-how to do so. These countries need to empower their populace and ensure that they can also derive meaning in their lives. In doing so, the population becomes more content and the overall sense of the community does not emanate from despair; but rather from happiness and meaning.

The above analysis was done using a python colab notebook and the results verified in the SQL notebook. For project management, we used Trello and hosted on Github. The links are as follows:

* Python notebook [[Link]](https://colab.research.google.com/drive/1pXi8GO67vsurAd-eqPWFip-mO6g5pM2T?usp=sharing)
* SQL notebook [[Link]](https://colab.research.google.com/drive/1lWUUwPOPGzdI3JDvXbSt1Ov6wsu9DONy?usp=sharing)
* GitHub repo [[Link]](https://github.com/allanstar-byte/ESTRELLA.git)
* Trello [[Link]](https://trello.com/b/JynZFMbI/estrella)