1. **PROBLEM STATEMENT**

**Overview**

The problem question will involve Analysts Consultants. The company wants to establish a way to find out the leading causes of suicide in different countries in the world and how the economic aspects affect the rates of suicide. Through this, a proper model can be devised and recommended proper measures to be taken to reduce the rates of suicide rates globally. Using Data Science techniques, we will be scheming and planning towards a strategy to achieve reliable results.

**Objectives.**

* To show how the economic aspects such as GDP affect the rates of suicide.
* To obtain an overview of suicide rates over a specific period of time.
* To show the distribution of suicide cases across the countries included in our data.

**Business criteria**

The success of the project determined by importing our libraries, loading our datasets, cleaning our data and performing proper analysis to enable us to establish the relation between the rates of suicide and the GDP per year in our dataset.

**Assessing the situation**

**Resources**.

1. Suicide dataset [<https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016/download>]
2. Software (Google Colaboratory, JIRA, Google Docs, Google Slides, GitHub)

**Assumptions.**

I. The data is correct and up to date.

II. The data available is highly attributable.

**Constrains.**

I. The dataset lacks frequency and consistency as some countries do not submit their data.

II. The dataset does not mention the reasons that mostly lead to suicide in each country.

**Data mining goals**.

* Do economic aspects(GDP,per\_capita income) affect the rate of suicide?
* Is population direct proportion to the number of suicides?
* What is the trend of suicides yearly between 1995 to 2015?
* What is the suicide rate by gender and age group?
* The correlation of different economic aspects and suicide rates

**2. Data Understanding**

**Overview.**

Global suicide dataset - contains the rates of suicide as compiled by 101 countries in the world. Its columns consist of the country name, age, the suicide rate per 100,000, gender, year, population, GDP per year($), and GDP per capita for every year.

**Verifying our data quality.**

Our data had no missing errors and no data errors.

**3. Data Preparation**.

The team followed the following steps outlined below to clean their dataset.

* **Loading our dataset**

After importing python and NumPy libraries, the dataset was uploaded to the Python database from the CSV files.

* **Cleaning our dataset.**

1. ***Validity***

This included dropping unnecessary columns and removing any syntax errors from the dataset.

1. ***Accuracy***

This included checking and removing any contradictory rows or columns.

1. ***Completeness***.

This included checking and counting null and missing values, dropping the rows and columns missing values and filling the null values with a random value or the mean or mode or median of the total rows in that column.

1. ***Consistency.***

This involved checking for duplicates in our dataset and dropping them.

1. ***Uniformity***

This stage includes renaming column names, changing the data types of columns and changing the column names to either lowercase or uppercase.

The cleaning processes were done on Google Colaboratory. [<https://github.com/kari-hub/Analysts-Google-Colab>]

**4. Analysis**.

The analysis is the process of breaking a complex project into smaller parts in order to gain a better understanding of it. It is simply the process of identifying trends in the data and utilizing visual presentations about the findings in our identified problem.

**i)**We first wanted to find how age and gender relate to the rates of suicide deaths in our dataset.

1. In terms of age, we realized that most suicides happened for people between the ages of 35-54 years. The least number of suicides happened to those between ages 5-14 years

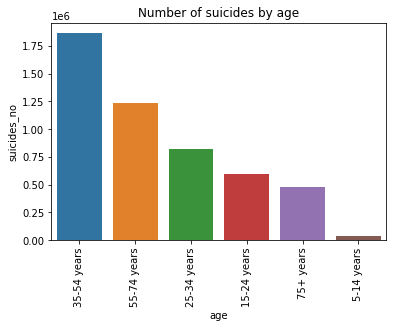
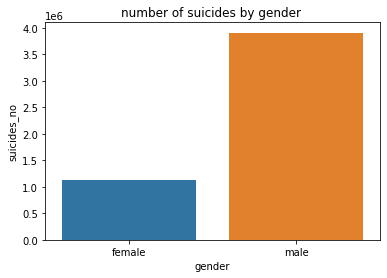


Figure 1.1

A graph showing Suicide numbers against age

1. In the case of gender, males had three times more suicide cases to females

.



**ii)** Having known the ages with the most Suicide cases, we checked for the countries that have the most prevalent causes of suicide

Our findings showed that the Russian Republic and the United States of America had the most cases. This could be assumed maybe because of the high population of about 140 Million and 295 Million at around 2005, a middle point in our data distribution.

Japan is also a hotbed for suicide cases as it comes third with most suicide cases.

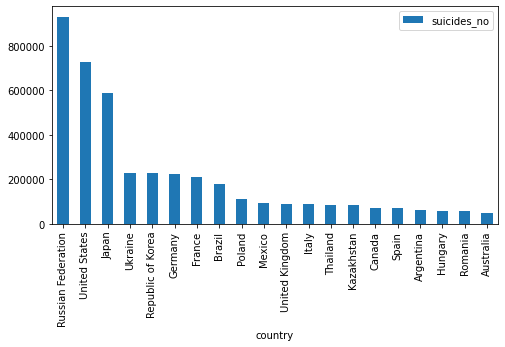


Figure 1.3

A graphical representation of the Number of Suicide against countries.

**iii)**We also checked for the distribution of suicide cases during the years and suicides per 100,000 people. This was to check if there has been a general decrease or increase to 2015.

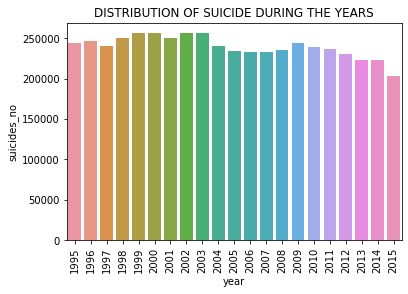


Figure 1.4

A graph representation showing Suicide numbers against time (years)

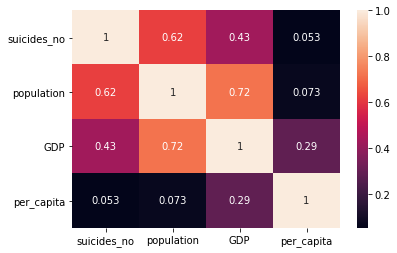
**iv)** We also checked for how GDP relates to the rate of Suicide.

This was to help us have a closer and in-depth examination of how a country's progression on Financial aspects affects its citizens by having either a positive impact or a negative influence that leads to them committing suicide.

The table shows there is a negative relationship between the number of Suicides and the GDP of different countries. This means that there is a high case of suicides in Low GDP Countries.

**v)** We also checked for Correlation between some phenomena to know the relationship between the number of causes of Suicides and population, per capita and GDP. Through this, we see that there is a highly positive correlation of 0.62 between population and no of Suicide cases.

* There is a weakly positive correlation of about 0.43between the GDP and the number of Suicides cases.
* There is a neutral correlation of about 0.053 between the Per Capita income and the number of Suicide cases.



**5.Recommendations**

Our main objectives were to find the main causes of suicides from our data sets then give the best recommendations to help reduce suicidal rates in the world.

* The first and most significant factor was the population. People in more populated countries are more likely to commit suicide than countries that are less populated. This might be because of a shortage of essential resources etc. Another factor that led to suicides was low economic factors. In our case, we used GDP which is the total monetary value of all goods and services produced in a country over a period of one year. The solution to this would be countries and governments to provide employment for the youth in these countries in order to reduce the rates.
* Another thing we realized from the analysis was that the males were more prone to suicide than females in general. The cause of this might be that men are used to bottling up their feelings and they also lack the will to express their issues, hence the government should sensitize men on the importance of mental health and be more open to counseling.
* Another recommendation is that in countries such as US the governments should enforce strict gun laws most suicides recorded are committed using guns hence making guns less available to citizens will significantly reduce the number of suicides in these countries
* Offering follow up community sensitization in areas that have faced war, trauma and abuse because search people are more likely to commit suicide hence these people should be counselled and talked to reduce PTSD they had from their previous experiences