**Problem Statement**

# **Background**

Suicide is a sad, sombre and dark topic that most people shy away from speaking about. However, the truth is that the more people fail to speak out regarding this occurrence, the deeper the problem lies. To bring the topic into light, we need to be objective and have factual data and statements given the delicate nature of the topic. Suicides largely come from mental health issues and the current generation has been open about this topic and are also actively participating on the suicide and mental health issues activism.

# **The Problem**

This project seeks to shed some light on the extent of suicide rates in the world. Having this data will be particularly helpful in ensuring that as we tackle the issues surrounding mental health, we have a clearer and better understanding of the topic.

# **Research Aims and Objectives**

This project particularly aims at shedding some light on the suicide rates of the people. Notably, the research objectives are as outlined below;

* 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

Other research aims and objectives will be created as the project progresses.

# **Data Exploration**

The data for the project is retrieved from [Kaggle](https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016/tasks). The dataset 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 analysed using python. We will also use sql in some instances.

We will also use the CRISP-DM methodology to guide us throughout the project.