Cause of Death

Table of Contents

. Introduction

Scenario:

You have just find out number of suicide **clusters** being reported across the world. A data collector hands you this data to **perform Data Analysis** and wants you to examine **trends** & **correlations** within our data. We would like to make a **Machine Learning algorithm** where we can train our **AI** to **learn** & improve from experience. Thus, we would want to **predict** the amount of suicides numbers in a certain demographic.

2. Data Wrangling

Our data set has 5 **Features** (Country, Year, Gender, Age, Population). We will explore all of these in detail. While the suicide_no is what we would like to **predict**.

3. Exploratory Data Analysis

From observing our **Time Series Line Plot**, we can see a **sharp drop** in suicides in 1985. This **decrease** could be due to **awareness** of suicide & **mental health** in the 80s, as well as **improved recognition** of those at risk. This is indeed **accurate**, as the research, "Suicide in the elderly" **supports** this claim.

The graph & find_minmax function above, **confirm** that Albania had the **lowest** suicide count, while Zimbabwe & Russian Federation, had the largest suicide count. A **reason** the Russian Federations may have a **large** suicide count may be that they have a very large population (144.5 million, while Albania only has about 3 million). It has been reported that Russian levels of alcohol consumption plays an immense role in it's **large suicide count**, but their is a **lack** of data to **support** this due to Soviet secrecy.

The data illustrates that middle aged adults, between the ages of 30 through 60, have the **highest** suicide count. While elderly and adolescents have about **half** the amount as middle aged adults.

Suicide is one of the **leading** causes of death among all Americans adults. Data show **heightened differences** in suicide for different sexes. It's evident that **males** are **more inclined** to suicide. For Females, the 4 age categories seem to **level off** at 150. We can't say the same for males. Male adults & male middle aged adults are at very high risk of suicide. Both genders show middle aged adults as the leading age group of suicide.

4. Machine Learning + Predictive Analytics

Our **goal** in this section is to build a **multiple linear regression** model that will be **trained** to understand **correlation** between our **features** and our **predictor**. We want to **predict** Y (suicides count), given a a specific year, pertaining to a specific age group & gender.

Conclusions

1. There was a **decrease** in suicides toward the 80's. This could be due to awareness of suicide & mental health in the 80s, as well as **improved** recognition of those at risk. But shortly after that their is a **rise** suicides that we are seeing.

2.Russian levels of alcohol consumption plays an immense role in it's large suicide count, but their is a **lack of data to support** this due to Soviet secrecy.

3.The data illustrates that middle aged adults, between the ages of 30 through 60, have the **highest suicide** count. While elderly and adolescents have about **half** the amount as middle aged adults.

4.Suicide is one of the **leading** causes of death among all Americans adults. Data show **alarming differences** in suicide for different sexes. It's evident that males are more inclined to suicide, than females. In addition, Mental health is a major predictor for suicide.