Suicide Rate Analysis

Source Dataset: https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016

1. Introduction

This tableau visualization is the analysis of suicide-rates over the last three decades among various countries. This project visualizes the following;

- statistical insights in the dataset
- correlation between attributes
- linear model

2. Dataset Discussion

This is the dataset taken from Kaggle, and it has all the data from the year 1985 to 2016. This is a small snippet of the dataset;

master.csv Country	master.csv Year	master.csv Sex	master.csv Age	master.csv Suicides No	master.csv Population	master.csv Suicides/100...	master.csv Country-Year	master.csv HDI for year	master.csv Gdp Per Year (...	master.csv Gdp Per Capit...	master.csv Generation
Albania	1987	male	15-24 years	21	312,900	6.7100	Albania1987	*null*	2156624900	796	Generation X
Albania	1987	male	35-54 years	16	308,000	5.1900	Albania1987	*null*	2156624900	796	Silent

This dataset has two redundant columns and null values for HDI (Human Development Index). This analysis focuses on comparing the suicide rates based on the attributes; *country, year, sex, age* and finding the correlation between attributes.

3. Procedure

3.1. Statistical Insights

- **Number of Suicides Vs Age Group**
  Plotted a bar graph to illustrate the number of suicides in each of the age category. The age group 35-54 has the maximum number of suicides which is 2,452,141.
- **Number of Suicides Vs Sex**
  Plotted a bar graph to show the suicide rate among male and female, and it is noticed that male have the maximum number of suicides.
- **Number of Suicides Vs Generation**
  Plotted a packed bubbles plot to find the generation with the maximum number of suicides and found that Boomers is the one with maximum suicide rate.
- **Average No. of Suicides Vs Year**
  Plotted a trend line to visualize the average number of suicides over the years, and it is seen that, it has been increasing until 2015, and it has decreased tremendously for 2016.
- **Comparison of GDP per Capita for countries**
  Developed a bar graph to compare average of the gdp_per_capita ($) attribute for various countries. It is seen that Luxembourg has the maximum gdp per capita over the last three decades.

- **Population Vs Suicides among countries**
  It is clear from the graph that, United States has the maximum suicide rate.

3.2. Correlation Plot

Since this is a regression problem, plotted the graphs to find the type of correlation between the attributes; *suicides_no, population, gdp_per_capita ($)* (*Note: HDI is ignored, since it has null values, and we cannot assume any default value to replace them).* It is observed that *suicides_no* and *population* have the positive correlation.

3.3. Linear Model

Developed a linear model for suicides_no and population, this model has r-squared value of 70% and a low Mean Squared Error.

Analysis Results

To answer the question whether the suicide rate has been decreasing over the years, it would be yes, we can see a gradual decrease over the years. Hopefully, this continues to decrease over the future. It has been observed that, population is one of the factors which is highly influencing the suicide rate. It is noticed that males have more tendency to commit suicides, which is another problem to dig into and find the reasons.

Conclusion

The future steps of this analysis would be to perform predictive analysis by combining this data with other similar datasets with many more attributes which support predictions.