**Data**

1. **Data Sources:**

The main source of this data set has been Kaggle where individual data sets of different years is available. For this project I’m considering the 2019 year’s data.

1. **Data Insights:**

The data consists of seven different variables each of which are numerical values. The numerical values range from 0 and go all the way above. Based on the values of independent feature sets, a score is derived for each country which decides the rank of the country. There are 155 countries data that is available for modelling this project’s model. The independent feature sets namely GDP, social support, healthy life expectancy, freedom to make life choices, generosity, perceptions of corruption are used for finding the dependent variable which is the score. It is also clearly visible from the table that Country or Region column is filled with country names which is non numeric value. Nevertheless since, we are not utilizing this column for our findings it can remain as it is.

1. **Approach:**

The idea is to utilize 30% of this data to train the model and use the remaining 70% to see how accurate the model is working. Also, when a new country is added to the data we can determine its rank based on the working model available with us.