**Post-join**

After joining the suicides data with the extracted information from the API, we performed some basic data manipulation tasks to deal with missing values. Next, we did some mathematical manipulation as well to include the log values of the suicide rates as well as the GDP per capita and the unemployment rate. This was done primarily because our subsequent analysis would include regressions and we would want to use a model which explains the greatest variation in our dependent variable (number of suicides). Filters were also added to set the year in which the user wants to analyze the relationships as well as the gender category in which the analysis can be done. For this report, all the analysis will be done on the gender category ‘both sexes’ and the year ‘2019’. To find the individual correlations of GDP per capita and unemployment rate with the suicide rate, we filtered out the missing values of only the relevant variable for that correlation to preserve the number of data points that we have.

Diagram

Description automatically generated

Upon filtering out the missing values, we used the group-by node to select the relevant columns on which the correlations will be calculated, and the regressions will be performed.

**Results of the analysis:**

**Impact of GDP per capita on the number of suicides (per 100,000 pop)**

Correlation:

We formed a correlation matrix for the GDP per capita, its log value, the suicide rates (Actual Value) and its log value. The results are summarized below.

Table

Description automatically generated

The output indicates that there is a positive correlation between the number of suicides and the GDP per capita as well as the log of the GDP per capita. The magnitude of the correlation though is weak in both the cases.

Linear Regression:

We also ran a linear regression where we regressed the number of suicides on log\_GDP as this model had a higher value for the correlation coefficient.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficient | Standard Error | t - value | p - value |
| log\_GDP | 0.73 | 0.44 | 1.67 | 0.096 |
| Intercept | 3.21 | 3.85 | 0.83 | 0.406 |

A picture containing chart

Description automatically generated

The regression output indicates that the log\_GDP is a statistically insignificant variable in explaining the variation in the number of suicides if we test at a 5% level of significance. The scatterplot with the regression line shows the positive correlation between the two variables. However, due to the low strength of the correlation, there is not much variation in the number of suicides in a country as the GDP per capita increases.

**Impact of unemployment rate on the number of suicides (per 100,000 pop)**

Correlation:

Like the previous case, a correlation matrix was formed between the unemployment rate, the suicide rates and their respective log values.

Graphical user interface, table

Description automatically generated

The output in this case indicates a positive correlation between the unemployment rate as well its log with the actual number of suicides in a country. The correlation is stronger in the case of the of the level-level correlation.

Linear Regression:

As the previous output indicated a stronger correlation in the case of a level-level model, we therefore ran the linear regression on the same model. The regression output is shown below.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable | Coefficient | Standard Error | t - value | p - value |
| Unemployment | 0.35 | 0.12 | 3.01 | 0.003 |
| Intercept | 6.92 | 0.99 | 6.99 | 6.023581633485264E-11 |

A picture containing text, outdoor

Description automatically generated

The regression output indicates the unemployment is a statistically significant variable in explaining the variation in the number of suicides when tested at a level of significance of 5%. The regression line also reflects the positive correlation between the two variables. As the unemployment rate increases, we see a gradual rise in the number of suicides (per 100,000 population) in a country.

**Conclusion:**