

Jonathan Ahern

Data Science

March 1, 2022

HW3

Report:

For the first part in MySQL, I cleaned the data by deleting all the rows where population. Next, I took all the averages for Adult Mortality, GDP, Schooling, and Alcohol and made all zero values into the average. I then selected the count of distinct countries after I cleaned the data. I then went through and found the highest and lowest averages for mortality, population, GDP, years of schooling, and alcohol consumption. I took the averages and country and ordered them by the average first in ascending order to find the lowest of each average and then descending order to find the highest. I decided to exclude values of zero as outliers and took the lowest values above 0.

My findings are as follows:

Count of distinct Countries: 145

Adult Mortality(High): 436.00 – Lesotho

Adult Mortality(Low):10.667-Tunisia

Population(High) 281,099,848- India

Population(Low): 292 – Palau

GDP(High): 70817.14495 - Switzerland

GDP(Low): 94.403 – Senegal

Schooling(High): 20.083- Australia

Schooling(Low): 5.05 – Eritrea

When posed with the question, do densely populated countries tend to have lower life expectancy, I found no real significant correlation between the values of population and life expectancy. I believe that life expectancy has to do with the other attributes more than the population density.

The information regarding the google colab is in the google colab but I will copy and paste the text here.

Adult Mortality correlation: -0.69913

Eating Habits correlation: 0.47074

Alcohol correlation: 0.41723

Population correlation: -0.02695

Total Expenditures correlation: 0.10357

GDP correlation: 0.44076

Looking at the correlations between the different attributes, we can see that the correlations were all fairly weak. The best correlations were Adult mortality with the highest negative correlation and Eating Habits with the highest positive correlation followed closely by Total Expenditures.

Answer the following questions by performing correlation analysis:

How does Adult mortality rates affect life expectancy?

The Life Expectancy increases as the Adult Mortality Rate decreases. This can be seen by the negative correlation as well as seen in the Linear Regression model below.

Does life expectancy have positive or negative correlation with eating habits, drinking alcohol, social factors, and economic factors?

Eating Habits:Positive

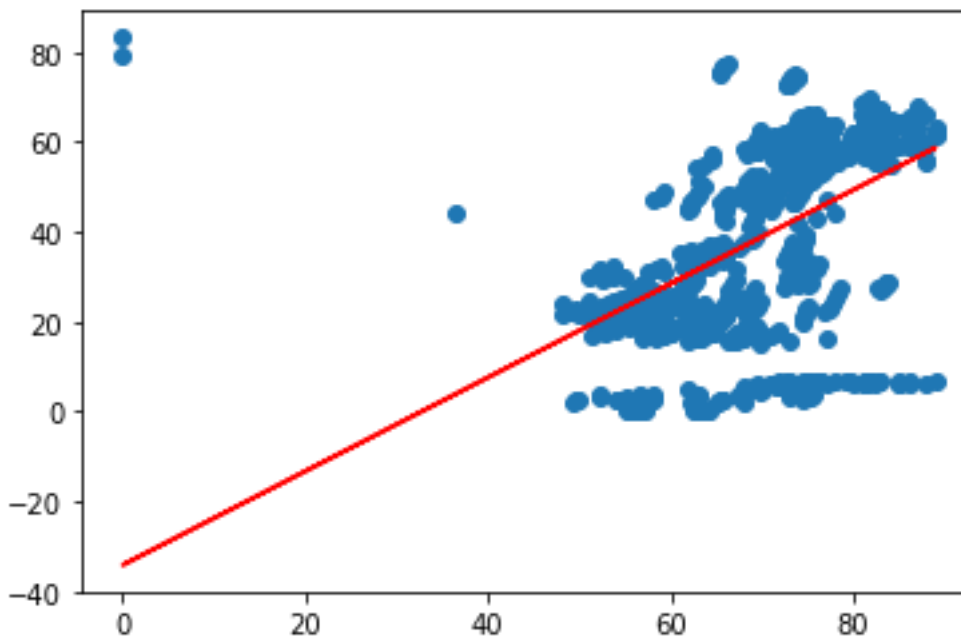
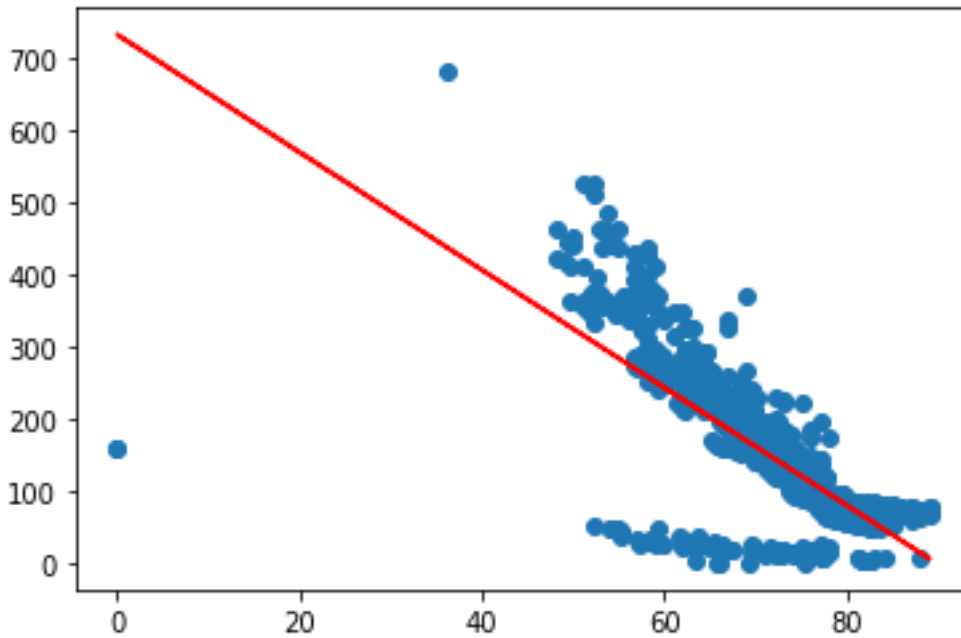
Drinking Alcohol:Positive

Social Factors:Negative

Economic factors:Positive

What is the impact of schooling on the lifespan of humans?

The more years of education, the longer the life expectancy is for an individual.



y= Life_Expectancy

x1= Adult_Mortality

x2= BMI

$y = (-7.87)x_1 + 712.209$

MSE: 6774.677574449944

R2: 0.4526019489966343

$$y = (1.044)x^2 - 34.167$$

MSE: 346.5798258507431

R2: 0.2215920118929937

When looking at the highest negative and positive correlation between the Attributes and Life Expectancy, the graphs seem to have several zeros that throw off the percent correlation. When looking at the Adult Mortality rate in comparison to Life Expectancy, as one would assume, as Mortality increased, life expectancy decreases. The more deaths there are, the less likely that a country would have a higher mortality rate. However, there were several countries with low adult mortality numbers but still low life expectancy. I am not exactly sure why those numbers produce those results but I assume it is based off of a strong correlation between one of the other attributes, such as BMI. Speaking of BMI, it had the highest positive correlation to Life expectancy. As the average BMI for a country increases, the life expectancy increases. Between the two models, the Adult Mortality model performed the best. It had a higher correlation value, although negative. It was closer to -1 than the BMI positive correlation was close to 1. It means that it was more strongly negatively correlated with the Adult Mortality correlation.