LIFE EXPECTANCY IN ALL COUNTRIES

1. Introduction

Life expectancy means the estimate of the average number of additional years that a person of a given age can expect to live. There are some differences in life expectancy among countries in the world. While the United Nations estimated that the global average of life expectancy is 72.6 years in 2019.

1.1. Definition of concepts

We would like to see what affects the *Life Expectancy* in terms of health and economic parameters.

1.2. Objectives

The objectives of this project work are as follows:-

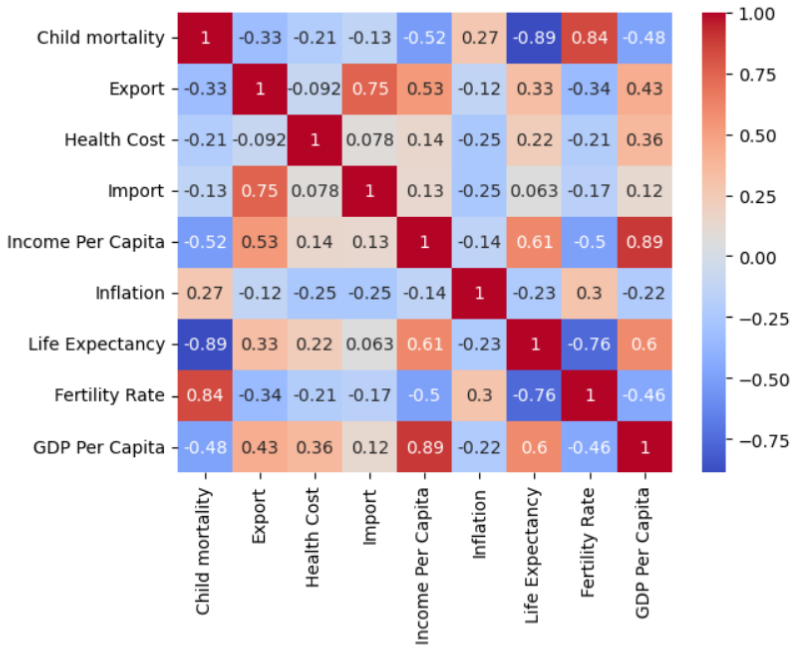
- To see the distribution of life expectancy among countries.
- To relate the life expectancy with increase or decrease of a health and economic parameter.

2. Data Sources

The data used is from an organization and it can not be disclosed.

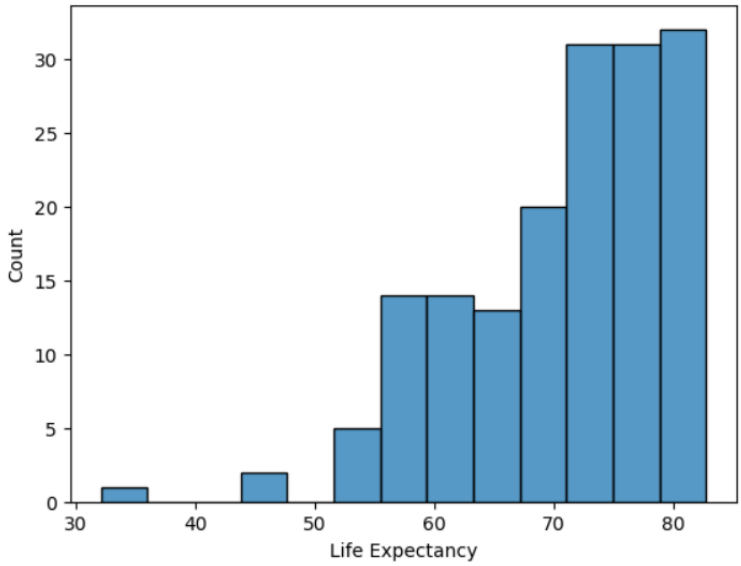
3. Exploratory Data Analysis

3.1. Correlation of All Data



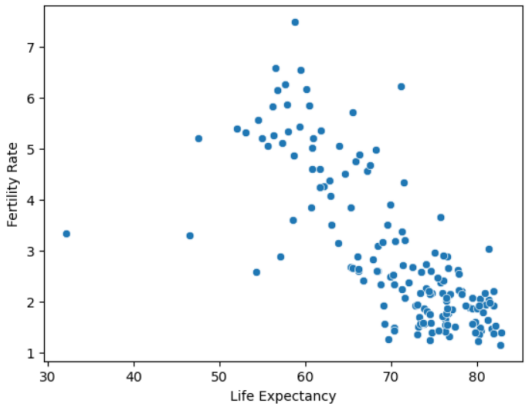
From the heatmap, there is a high correlation between *Life Expectancy* with *Fertility rate* (-0.76) and *Child Mortality* (-0.89). While the *Health Cost* is also considered a health parameter, it has low correlation with *Life Expectancy* (0.22). It has low to zero correlation with *Import* and *Inflation*.

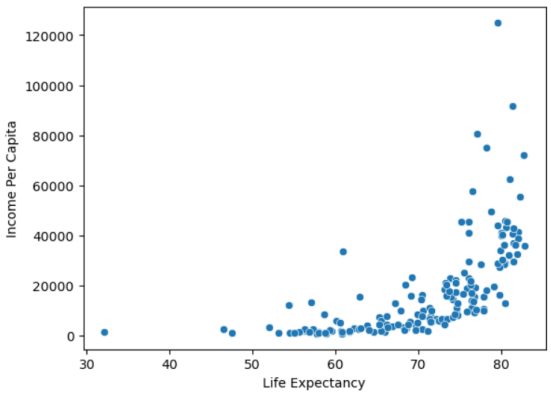
3.2. The Distribution of Life Expectancy

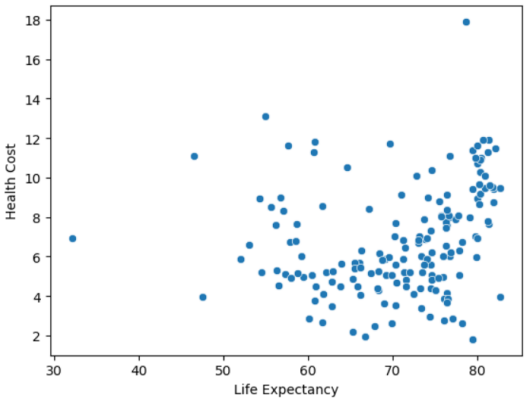


The *Life Expectancy* has left-skewed distribution, with most values being more than 70 years.

3.3. Results from the EDA







Based on the scatterplot, there is a correlation between *Life Expectancy* with *Fertility Rate* (negatively correlated, when the *Life Expectancy* goes higher, *Fertility Rate* gets lower); with the *Income Per Capita* (it is almost positively correlated and more variations for *Income Per Capita* < 20000); while the relation with *Health Cost* is more spread-out.

4. Confirmatory Analysis

We need to establish if the results from the exploratory analysis are actually valid.

4.1. T- test to compare Life Expectancy with Fertility Rate, Income Per Capita, Health Cost

With:

- H0 = Does High Fertility Rate/Income Per Capita/Health Cost Means Higher Life Expectancy?

```
#T-Test

s_mean = df['Life Expectancy'].mean()
nons_mean = df['Fertility Rate'].mean()
print(f'The difference of values: {nons_mean - s_mean}')
```

```
The difference of values: -67.78901840490798
```

```
ttest0 = st.ttest_ind(a = df['Life Expectancy'], b = df['Fertility Rate'])
ttest0.pvalue
```

```
5.146626444374683e-240
```

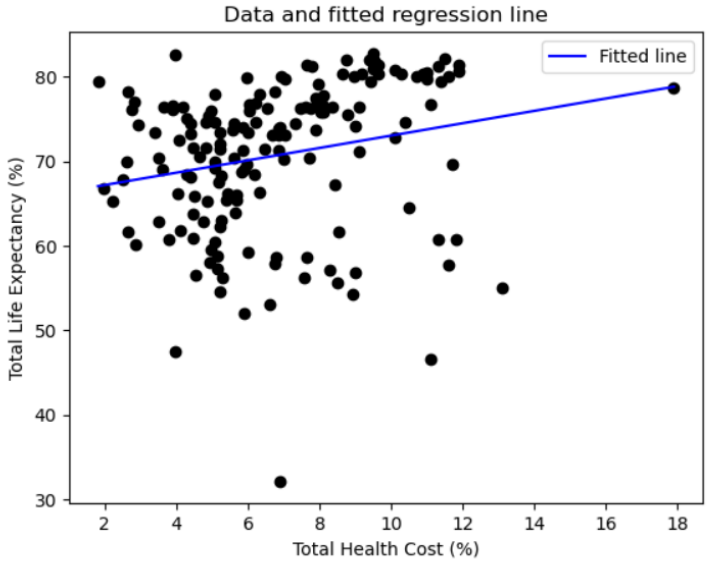
```
p_value0 = ttest0.pvalue
print('P-Value :',p_value0)

if p_value0 >= 0.05:
    print('Countries with high fertility rate have higher life expectancy.')
else:
    print('The fertility rate of a country does not affect its life expectancy.')
```

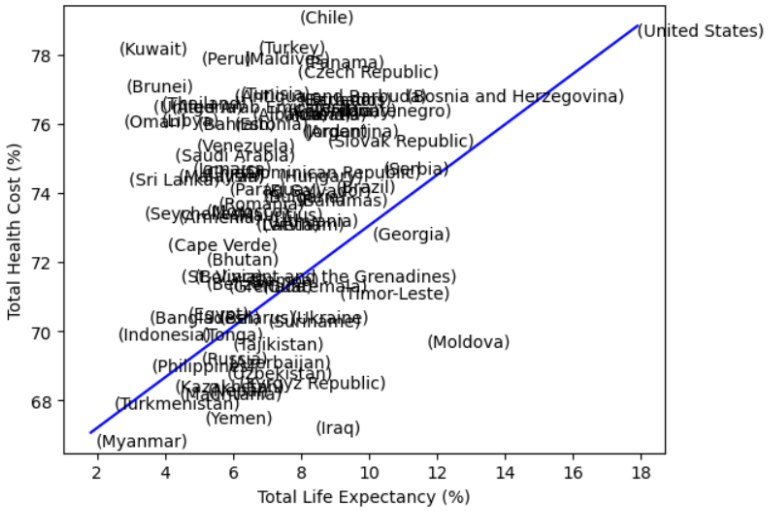
```
P-Value : 5.633547960569634e-228
The fertility rate of a country does not affect its life expectancy.
```

The result of all T-tests, it is concluded that Fertility Rate/Income Per Capita/Health Cost does not affect the Life Expectancy, as the **p-value <= 0.05**.

4.2. Model Fitting



The graphic showed the final model with fitted regression line.



In this graphic, we can see some countries that fit the calculation of *Life Expectancy* and *Health Cost*. As example, The United States has a high percentage of *Life Expectancy* and also *Health Cost*. While Myanmar goes the other way.

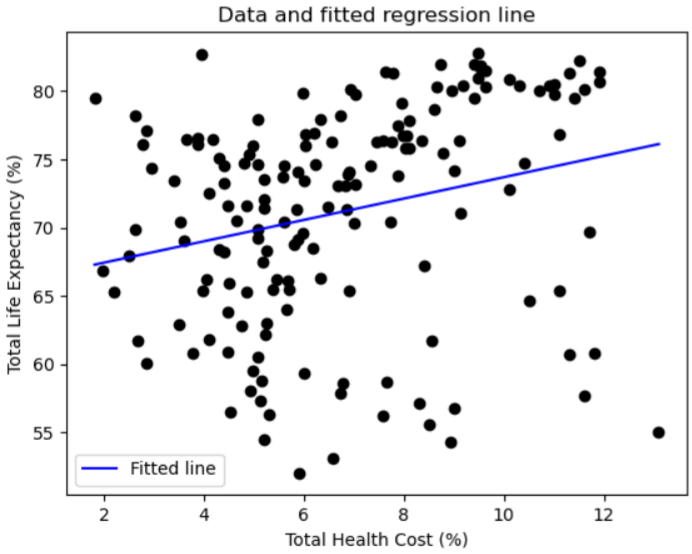
The coefficient of the model.

	coef	std err
Intercept	65.737988	1.850754
Health_cost	0.732186	0.253478

With the interpretation that the difference of 1 point from *Health Cost* will affect the *Life Expectancy* 0.73 point positively (increase *Life Expectancy* by 0.73 point).

The R-squared is only 0.049 which is really low thus meaning that the value has a low effect size.

4.2.1. Comparison After Handling Outliers



As seen in the graphic above for the model after handling outliers, the plot is more spread-out.

	coef	std err
Intercept	65.870287	1.709910
Health_cost	0.782625	0.237581

The interpretation of the coefficient is that the difference of 1 point from *Health Cost* will affect the *Life Expectancy* 0.78 point positively (increase *Life Expectancy* by 0.73 point).

The R-squared is increased to 0.063, however it is still really low thus meaning that the value has a low effect size.

5. Limitations of the research work

The limitations encountered are explained below:

- The data used is only for 1 period, and can not conclude the whole period.
- There are only two variables compared in the model (Life Expectancy and Health Cost).

6. Conclusion

Despite having a correlation in the exploratory analysis, especially in the scatterplot. Through the T-test, it shows that the *Life Expectancy* is not affected by *Income Per Capita*, *Fertility Rate*, nor *Health Cost*. Even looking further, Health Cost only affects the value of Life Expectancy by 0.78.