Life Expectancy Report (WHO Dataset)

The purpose of this report is to review the life expectancy statistics released by WHO. The data contains information on 193 countries from the year 2000 until 2015. Due to the limited data available on this topic, this report will be written under the assumtion the year is currently 2016.

Chapter 1: Lay of the Land

We will first begin by exploring the global life expectancy, this will be generated as an average from 2000 to 2015 for each country in the dataset.

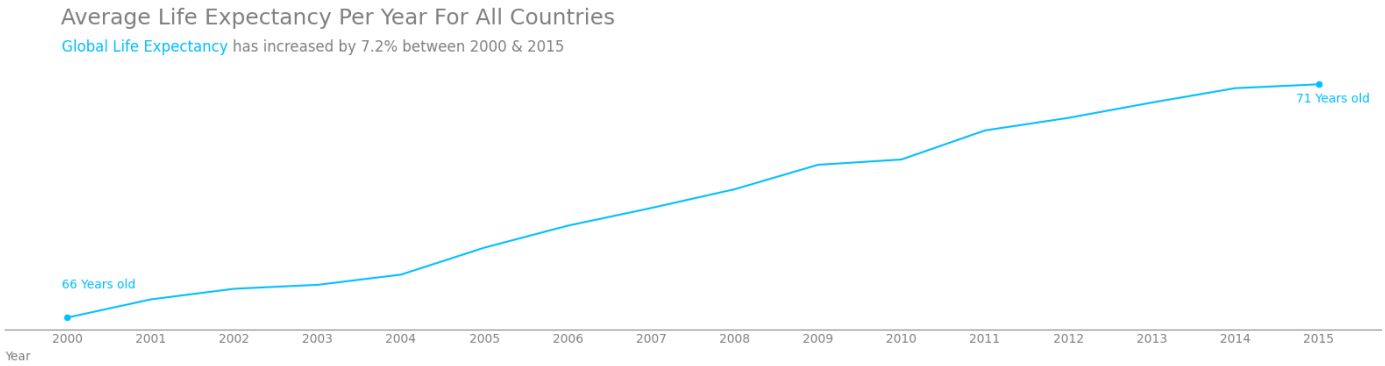


Figure 1.1

From figure 1.1 we can see that global life expectancy has risen from 66 years old in the year 2000 up to 71 years old in the year 2015, this is an increase of 7.2%. This increase can be explained by many factors, the first and main one being, the advances in medical science over the last years. The improvement in vaccines, antibiotics and better surgery techniques have been a major factor in increasing life expectancy. This can also be explained by public health improvements, a decline of child mortality and better living standards in third world countries.

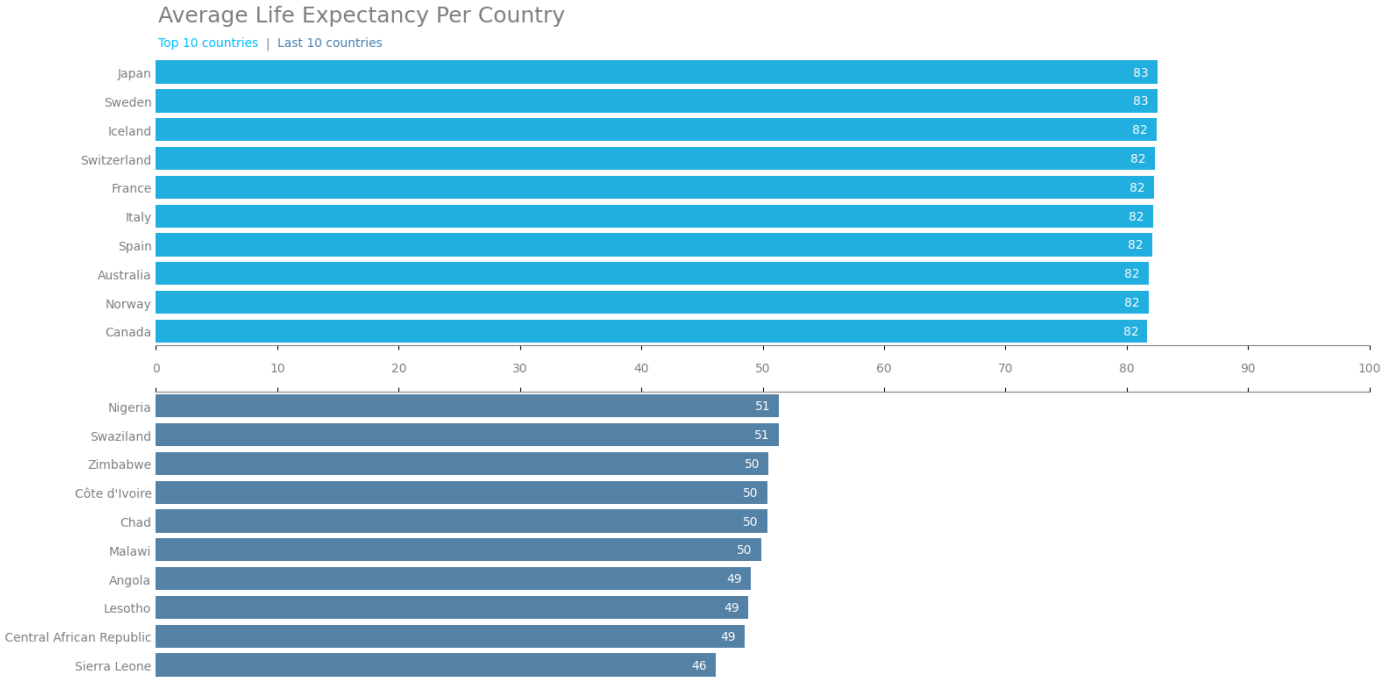


Figure 1.2

Moving on to figure 1.2 we can see the top 10 countries for average life expectancy vs the bottom 10 countries. Although life expectancy has been generally increasing over the years and is currently at 71 years old, there are still some countries that have a life expectancy that does not even reach 50 years old. During the course of this report we will be investing into why there is such a discrepancy between the top and bottom countries in order to decide if any changes can be made.

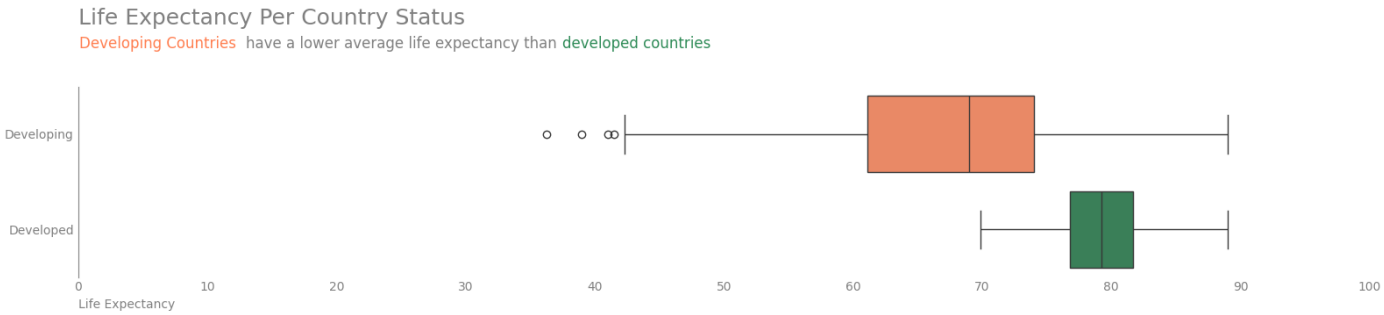


Figure 1.3

Figure 1.3 highlights the point which has been made before about the differences in life expectancy between some of the top and bottom countries, here we can see that developing countries have only an average of about 69 while developed countries have an average of 79. With the global average only being at 71, we can see how the developing countries have more of a dominance when it comes to controlling the global life expectancy rate. As a measure, there are currently 512 developed countries vs 2426 developing countries as per the dataset.

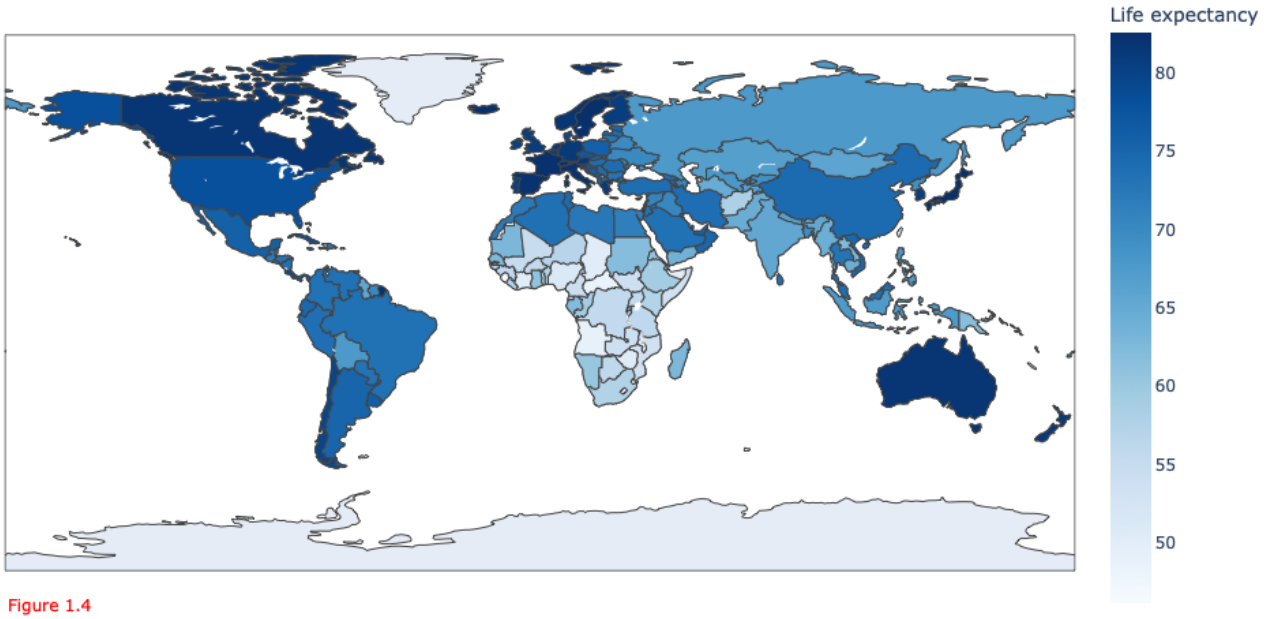


Figure 1.4

We will finish off this chapter by reviewing figure 4 which displays a map of the world with each country colored with a shade of blue, the darker the blue the higher the life expectancy. From this figure We can see that majority of the darker countries seem to fall on the westen side of the map however there are a few anomalies such as china and Australia. Digging into this a little deeper we can see that the darkest shades of blues are located at Canada, most European countries and Australia.

Chapter 2: Can we explain it?

In this next section we will be digging deeper into why the some countries have better life expectancy than others, we will review various features, how these features effect life expectancy and how they compare with each other.

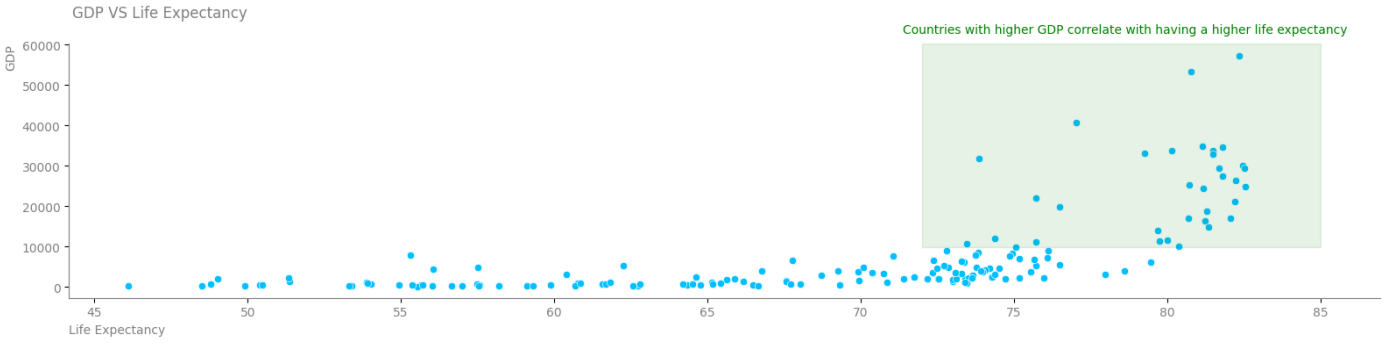


Figure 2.1

We start off by reviewing the average GDP and life expectancy rate of each country. If you look at figure 2.1 you will see each country represented by a blue dot, the further along the dot on the X axis, the higher the life expectancy and the higher the dot on the y axis, the more GDP the country has. As can be seen, countries that have a life expectancy up to 73 years tend to have a much lower GDP when compared to countries that have a life expectancy higher than 73. The green squre on the graph shows the richer countries and as explained they have the highest life expectancy ratings.

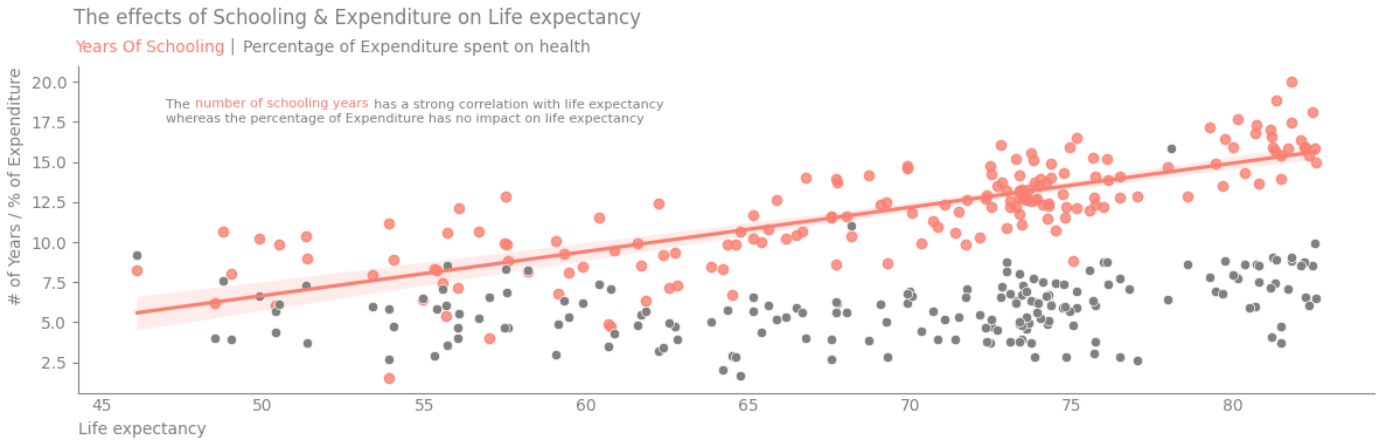


Figure 2.2

In this next graph we will be comparing two different features at once in order to review the difference of importance against life expectancy. Once again life expectancy can be seen on the x axis and on the y axis you can see number of schooling years which matches against the red dots and % of expenditure spent on health which matches against the grey dots.

If we first review the expenditure spent on health (grey dots) we can see that as life expectancy increases there seems to be no difference with the percentage of expenditure spent, in fact the country with the lowest life expectancy seems to have the highest rate of expenditure on health.

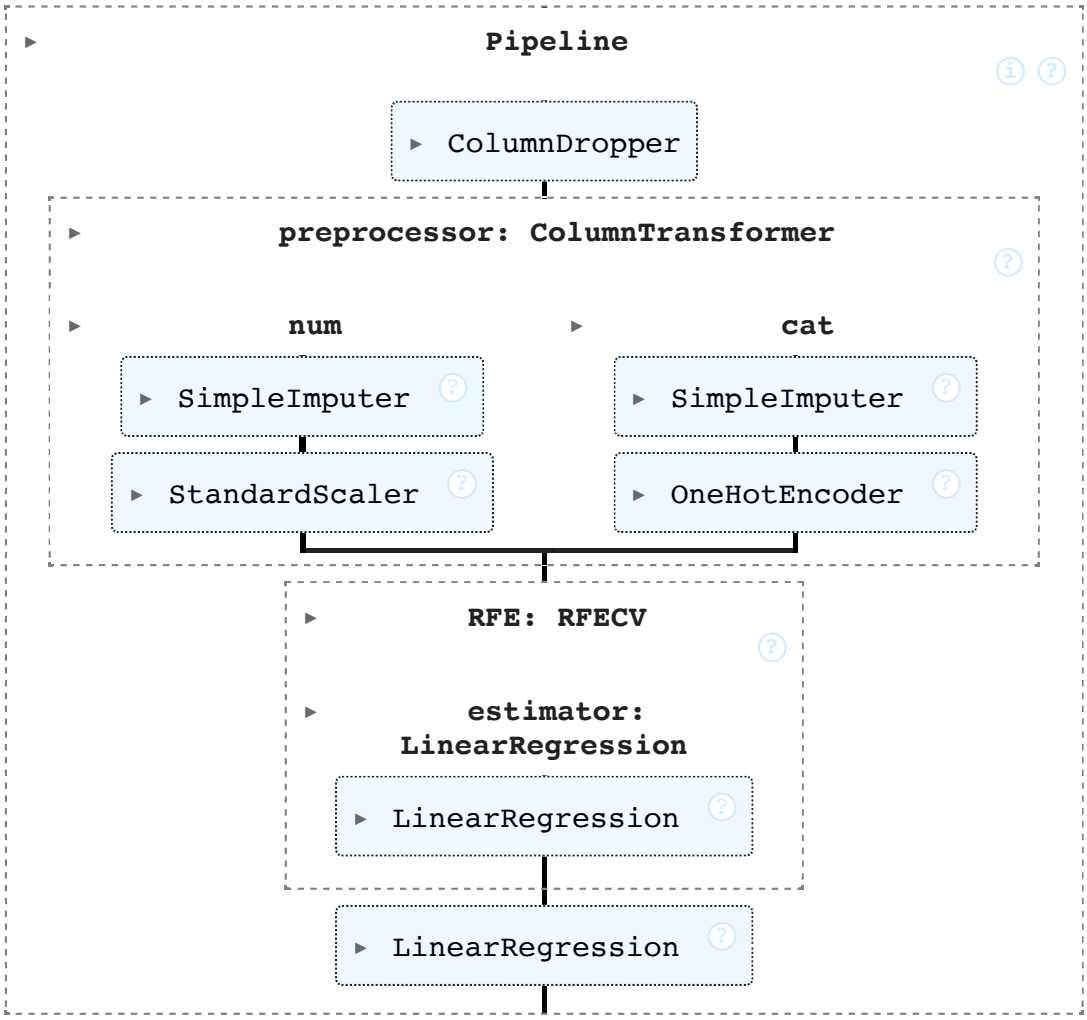
Moving onto the years of schooling for each country, we can see a positive correlation, the more schooling years a country has, the higher the life expectancy for that country.

What does this mean? well as seen in figure 2.1, having more GDP meant having higher life expectancy but this graph shows us that the GDP being spent on health does not make as much of an effect as other important features so perhaps this money is better spent on other aspects such as educating the population.

Chapter 3: Can we predict it

In the final chapter we will be using machine learning to make predictions based on life expectancy and using those predictions to review the most important features for increasing or decreasing the life expectancy.

Creating a Linear model for prediction



I have first created a pipeline which ends with a linear regression model to attempt to predict the life expectancy for each country. The model works by first dropping uneccessary columns such as Year and Country Name as they are of no use to the predictions. We then seperate the numerical and categorical columns for preprocessing, in this instance I have decided to opt for imputation for missing values and the columns are then scaled so that we are able to review the importance without the issue of scale causing us to misinterpret the results.

```
Root Mean Squared Error:  4.176900093097624
Accuracy Score: 81%
```

After testing I have managed to achieve an accuracy level of 81% and a root mean square error of 4.1 years. These results mean that our model is able to give highly accurate predictions with the features available, as a note, due to this being a reseach report the model is not needed for real life applications so a accuracy level of 81% will suffice in this instance.

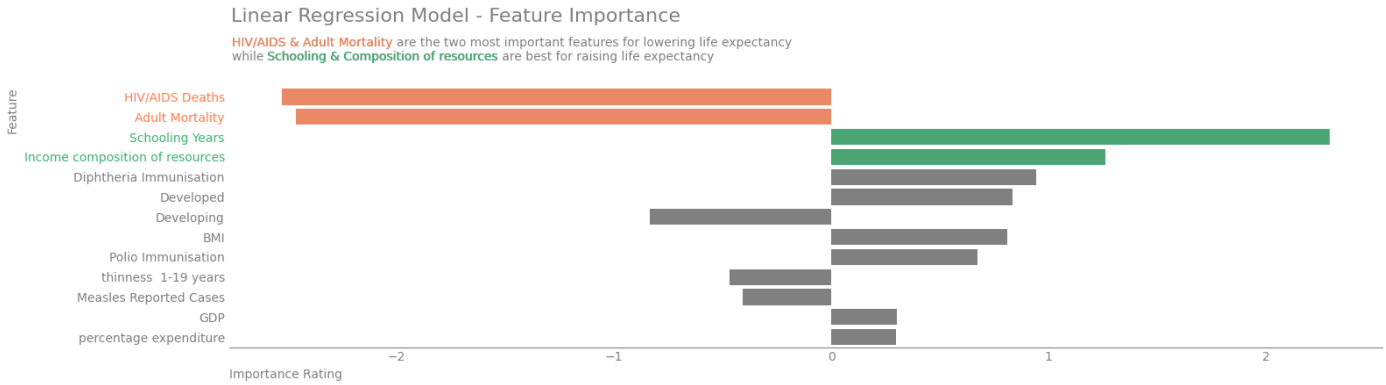


Figure 3.1

From running the linear model for predictions we are able to view the most and least important features when determaining life expectancy. This information is displayed in figure 3.1. Along the x axis of the graph you can see importance rating, this shows how much of an impact that particual feature had when deciding if life expectancy would increase or decrease.

First we will review the two most important features for lowering life expectancy (orange bars), this is the number of HIV/AID deaths per 1000 people and adult mortality which measures deaths between 15 year olds and 60 year olds.

On the other end of the spectrum we can see that the two most important features (green) for increasing life expectancy were the number of schooling years and income composition of resources which displays how efficient a country is at utilising its resources to generate income.

Using ML for clustering

We will now move on from using the linear regression model for prediction and create a new K-means model to cluster each country into groups.

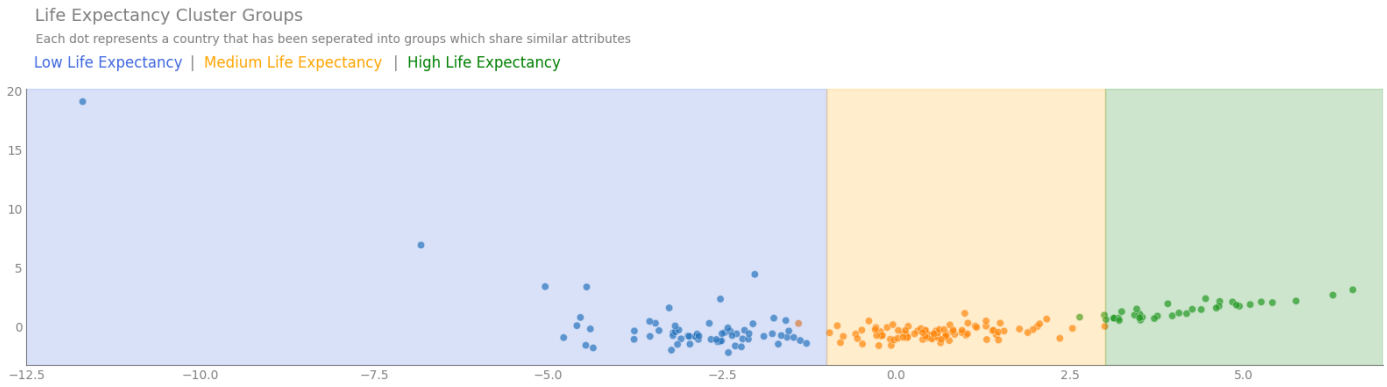


Figure 3.2

Using the K-means model, each country was seperated into one of three groups depending on the values of their features, once they had been clustered I then used principle component analysis to combine the multiple features into two main features so that it would be displayed on a graph (figure 3.2). From figure 3.2 you can see three groups seperated by color and each dot within the groups represents a country. The countries in the blue section have a lower life expectancy, they generally have the lowest number of schooling years, highest death from diseases, and lower GDP when compared to the other two groups. The orange group have medium life expectancy and generally fair pretty well with important predictive features and the final group which can be seen in the green shaded area are the countries which have higher GDP, higher levels of immunisation for diseases and the best schooling.

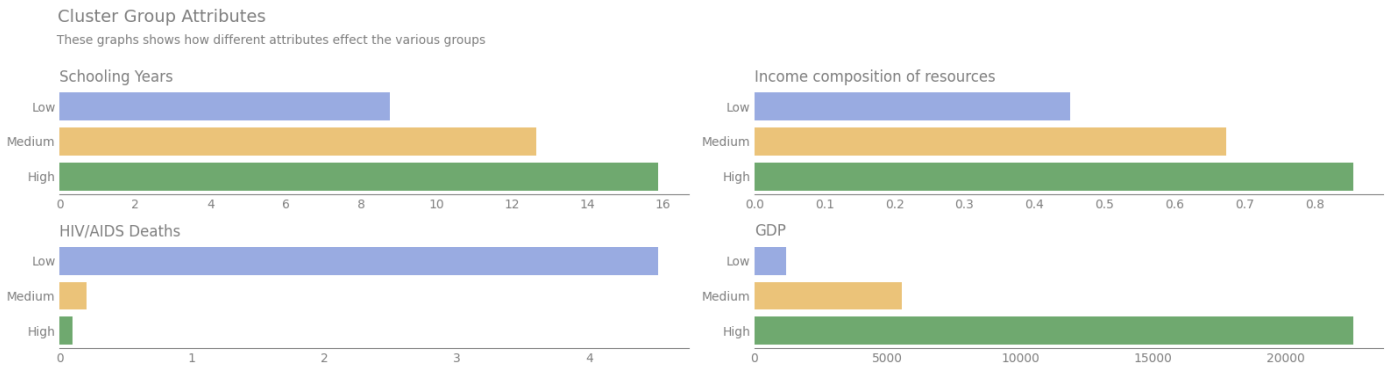


Figure 3.3

Figure 3.3 shows multiple graphs which compare various features for each cluster group, by these graphs we can see the differences which make up each group as highlighted previously.

Forecasting the future

I will now end with one final machine learning model which will predict the global life expectancy rate from 2016 until 2030.

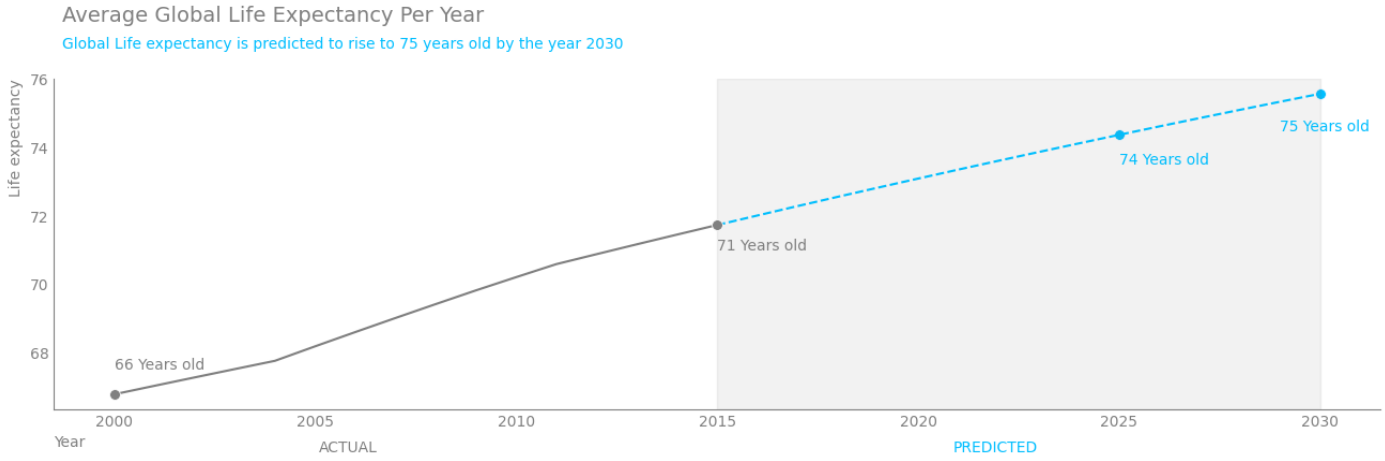


Figure 3.4

Figure 3.4 is seperated into two components, the first one being the graph you viewed at the very begining of this report displaying average global life expectancy and the second component being a prediction of the average global life expectancy until 2030. Using the information we have available, I can predict that by 2030 the average life expectancy will be 75 years old, now generally it is quite difficult to judge the accuracy of these models as they would require us to wait for future events but in this case we can view the accuracy of this model since we are already in the year 2025. The current global life expectancy for 2025 is around 73.5 and compared to the predicted value of the model, 74 years old, this shows the accuracy of this model to be very high.

Conclusion

In conclusion we were able to review the life expectancy for each country and how that looked on a global scale, we then dug deeper to try and figure out why some countries were doing better than others and then we used the support of AI to make some prediction on life expectancy and reviewed the results of the predictions to help fill in some of the gaps of this research.