Figure 1: Number of Countries above life expectancy threshold for given year

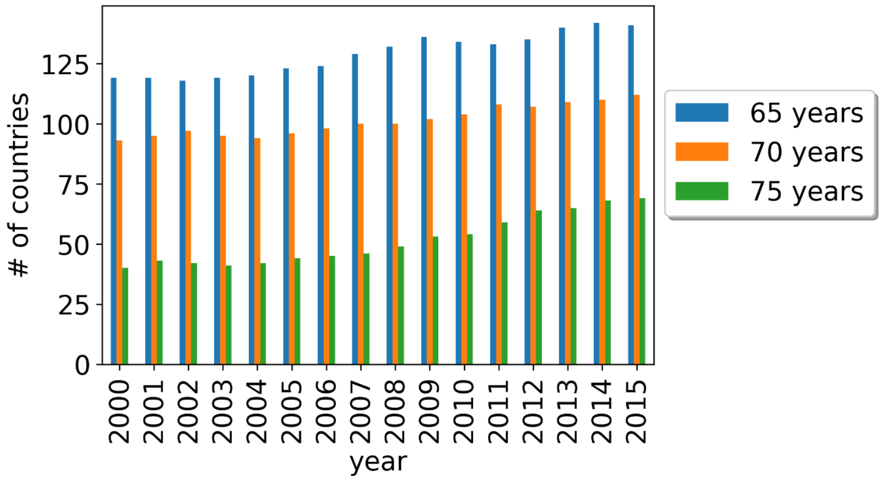
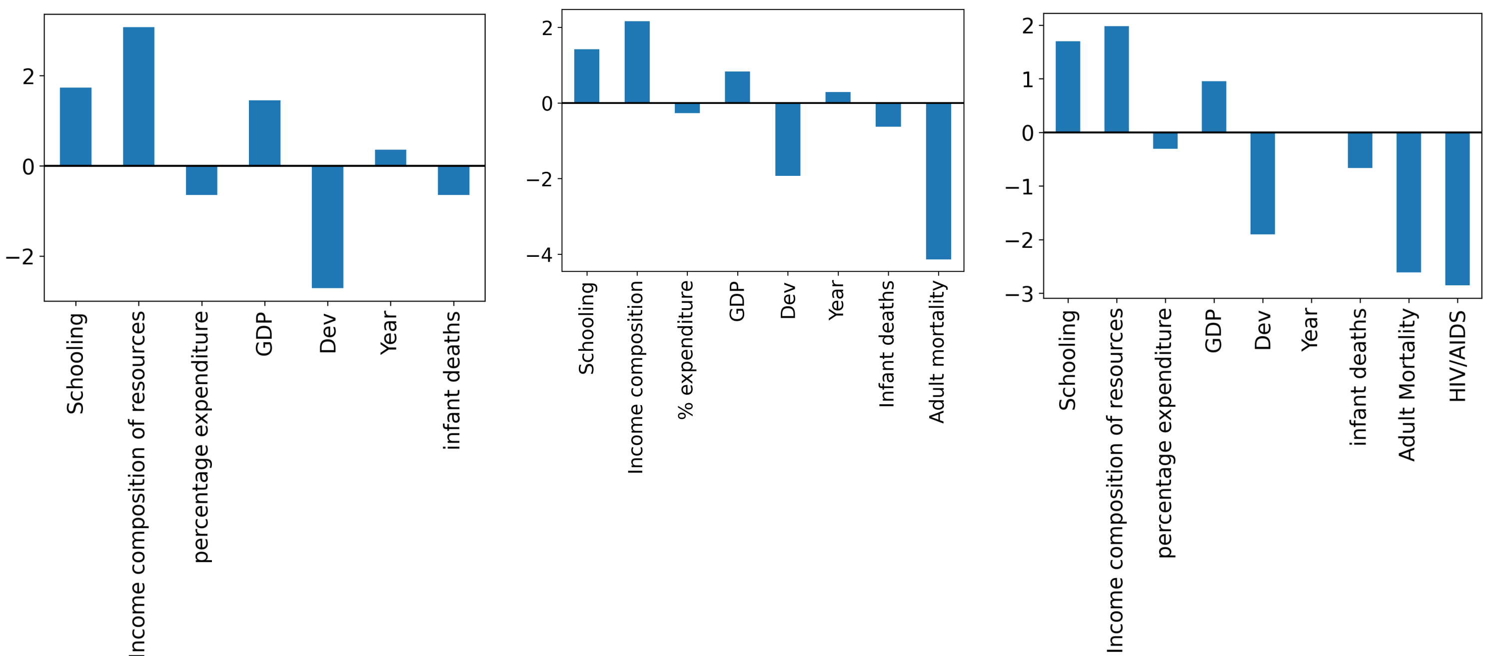


Figure 2.1: Model 1 (7 features)

Figure 2.2: Model 2 (8 features)

Figure 2.3: Model 3 (9 features)



Life Expectancy Predictions

Around the world, life expectancies vary drastically country to country. While it is becoming clearer and clearer what factors play a role in a country's life expectancy, there are some features that can help us better determine a country's life expectancy. The goal of my project today is to see if these features can accurately predict a country's life expectancy.

The dataset was provided on kaggle.com; it was scraped off of the World Health Organization's and United Nations's website. The dataset analyzed 194 countries, with statistics for every year from 2000 to 2015. The dataset contains 23 columns that include gdp per capita, income composition of resources, infant mortality rate, and whether a country was developed or not, among many other features.

For analysis, I selected a total of 9 features and evaluated their effects on life expectancy. These features include: (1) income composition of resources, (2) average years of schooling, (3) percentage expenditure, (4) GDP per capita, (5) year, (6) infant death mortality rate, (7) adult mortality rate, (8) HIV/AIDs rate, (9) status of country: developing or developed.

For **Figure One**, we can see how many countries pass a certain life expectancy threshold. On the y-axis, we have the number of countries that pass a certain life expectancy. On the x-axis, we have years from 2000 to 2015, with three bars on each year representing different life expectancy thresholds. From the graph, we can see that as the years increase, the number of countries that meet every life expectancy threshold increases. We can also see that fewer countries meet higher life expectancy thresholds. In conclusion, we can see that as years increase, life expectancy goes up, most likely due to policy changes and technological innovation.

For **Figure Two,** there are 3 plots considering different features and their associated coefficient values for a pipeline model that connects StandardScaler and LinearRegression. After creating a random 10/90% train-test split, we can compare actual .vs. predicted life expectancies. For **Figure 2.1**, though not pictured, the mean cross validation score with 10 folds was .478, meaning that for 10 models, an average of 47.8% of life expectancy predictions fit the regression model, while considering 7 features. For **Figure 2.2,** the mean cross validation score was .613, a large increase from Figure 2.1 after considering adult mortality as a feature. For **Figure 2.3**, the mean cross validation score was .665, after considering HIV/AIDs rates as a feature. From the figures, every feature but 'year' is similar across all figures. Post-standardization, we can see from the coefficient values of income composition, schooling, adult mortality rates, and HIV/AIDS that these 5 features play a large role in a country's life expectancy.

In conclusion, we discovered that 9 features play a large role in a country's life expectancy, the model with 9 features being fairly accurate. We learned that average schooling years, income composition of resources, and HIV/AIDs rates play a large factor in predicting life expectancy. Overall, these features tell us where a country should allocate resources in order to efficiently raise their life expectancy.