1my_edit_Life Expectency Prediction Using Linear Regression

July 23, 2021

1 Life Expectancy Prediction Model

Life expectancy prediction model created using data from WHO

[2]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

[3]:
```
from jupyterthemes import jtplot
jtplot.style(theme = 'monokai', context = 'notebook', ticks = True, grid =␣
↪False)
# setting the style of the notebook to be monokai theme
# this line of code is important to ensure that we are able to see the x and y␣
↪axes clearly
# If you don't run this code line, you will notice that the xlabel and ylabel␣
↪on any plot is black on black and it will be hard to see them.
```

[4]:
```
# read the csv file
life_expectancy_df = pd.read_csv('Life_Expectancy_Data.csv')
```

[5]:
```
life_expectancy_df
```

[5]:

	Year	Status	Life expectancy	Adult Mortality	infant deaths	\
0	2015	Developing	65.0	263.0	62	
1	2014	Developing	59.9	271.0	64	
2	2013	Developing	59.9	268.0	66	
3	2012	Developing	59.5	272.0	69	
4	2011	Developing	59.2	275.0	71	
…	…	…	…	…	…	
2933	2004	Developing	44.3	723.0	27	
2934	2003	Developing	44.5	715.0	26	
2935	2002	Developing	44.8	73.0	25	
2936	2001	Developing	45.3	686.0	25	
2937	2000	Developing	46.0	665.0	24	

	Alcohol	percentage expenditure	Hepatitis B	Measles	BMI	…	\
0	0.01	71.279624	65.0	1154	19.1	…	
1	0.01	73.523582	62.0	492	18.6	…	

```
2        0.01               73.219243         64.0       430   18.1  …
3        0.01               78.184215         67.0      2787   17.6  …
4        0.01                7.097109         68.0      3013   17.2  …
…        …                     …          …       …    …  …
2933     4.36                0.000000         68.0        31   27.1  …
2934     4.06                0.000000          7.0       998   26.7  …
2935     4.43                0.000000         73.0       304   26.3  …
2936     1.72                0.000000         76.0       529   25.9  …
2937     1.68                0.000000         79.0      1483   25.5  …

      Polio  Total expenditure  Diphtheria    HIV/AIDS         GDP  \
0       6.0                8.16         65.0       0.1  584.259210
1      58.0                8.18         62.0       0.1  612.696514
2      62.0                8.13         64.0       0.1  631.744976
3      67.0                8.52         67.0       0.1  669.959000
4      68.0                7.87         68.0       0.1   63.537231
…     …                 …           …        …          …
2933   67.0                7.13         65.0      33.6  454.366654
2934    7.0                6.52         68.0      36.7  453.351155
2935   73.0                6.53         71.0      39.8   57.348340
2936   76.0                6.16         75.0      42.1  548.587312
2937   78.0                7.10         78.0      43.5  547.358879

      Population   thinness  1-19 years   thinness 5-9 years  \
0     33736494.0                   17.2                17.3
1       327582.0                   17.5                17.5
2     31731688.0                   17.7                17.7
3      3696958.0                   17.9                18.0
4      2978599.0                   18.2                18.2
…          …                     …                 …
2933  12777511.0                    9.4                 9.4
2934  12633897.0                    9.8                 9.9
2935    125525.0                    1.2                 1.3
2936  12366165.0                    1.6                 1.7
2937  12222251.0                   11.0                11.2

      Income composition of resources  Schooling
0                               0.479       10.1
1                               0.476       10.0
2                               0.470        9.9
3                               0.463        9.8
4                               0.454        9.5
…                               …        …
2933                            0.407        9.2
2934                            0.418        9.5
2935                            0.427       10.0
2936                            0.427        9.8
```

```
2937                            0.434        9.8

[2938 rows x 21 columns]
```

[6]:
```
life_expectancy_df.head(7)
```

[6]:

	Year	Status	Life expectancy	Adult Mortality	infant deaths	\
0	2015	Developing	65.0	263.0	62	
1	2014	Developing	59.9	271.0	64	
2	2013	Developing	59.9	268.0	66	
3	2012	Developing	59.5	272.0	69	
4	2011	Developing	59.2	275.0	71	
5	2010	Developing	58.8	279.0	74	
6	2009	Developing	58.6	281.0	77	

	Alcohol	percentage expenditure	Hepatitis B	Measles	BMI	…	Polio	\
0	0.01	71.279624	65.0	1154	19.1	…	6.0	
1	0.01	73.523582	62.0	492	18.6	…	58.0	
2	0.01	73.219243	64.0	430	18.1	…	62.0	
3	0.01	78.184215	67.0	2787	17.6	…	67.0	
4	0.01	7.097109	68.0	3013	17.2	…	68.0	
5	0.01	79.679367	66.0	1989	16.7	…	66.0	
6	0.01	56.762217	63.0	2861	16.2	…	63.0	

	Total expenditure	Diphtheria	HIV/AIDS	GDP	Population	\
0	8.16	65.0	0.1	584.259210	33736494.0	
1	8.18	62.0	0.1	612.696514	327582.0	
2	8.13	64.0	0.1	631.744976	31731688.0	
3	8.52	67.0	0.1	669.959000	3696958.0	
4	7.87	68.0	0.1	63.537231	2978599.0	
5	9.20	66.0	0.1	553.328940	2883167.0	
6	9.42	63.0	0.1	445.893298	284331.0	

	thinness 1-19 years	thinness 5-9 years	\
0	17.2	17.3	
1	17.5	17.5	
2	17.7	17.7	
3	17.9	18.0	
4	18.2	18.2	
5	18.4	18.4	
6	18.6	18.7	

	Income composition of resources	Schooling
0	0.479	10.1
1	0.476	10.0
2	0.470	9.9
3	0.463	9.8

```
4                              0.454        9.5
5                              0.448        9.2
6                              0.434        8.9

[7 rows x 21 columns]
```

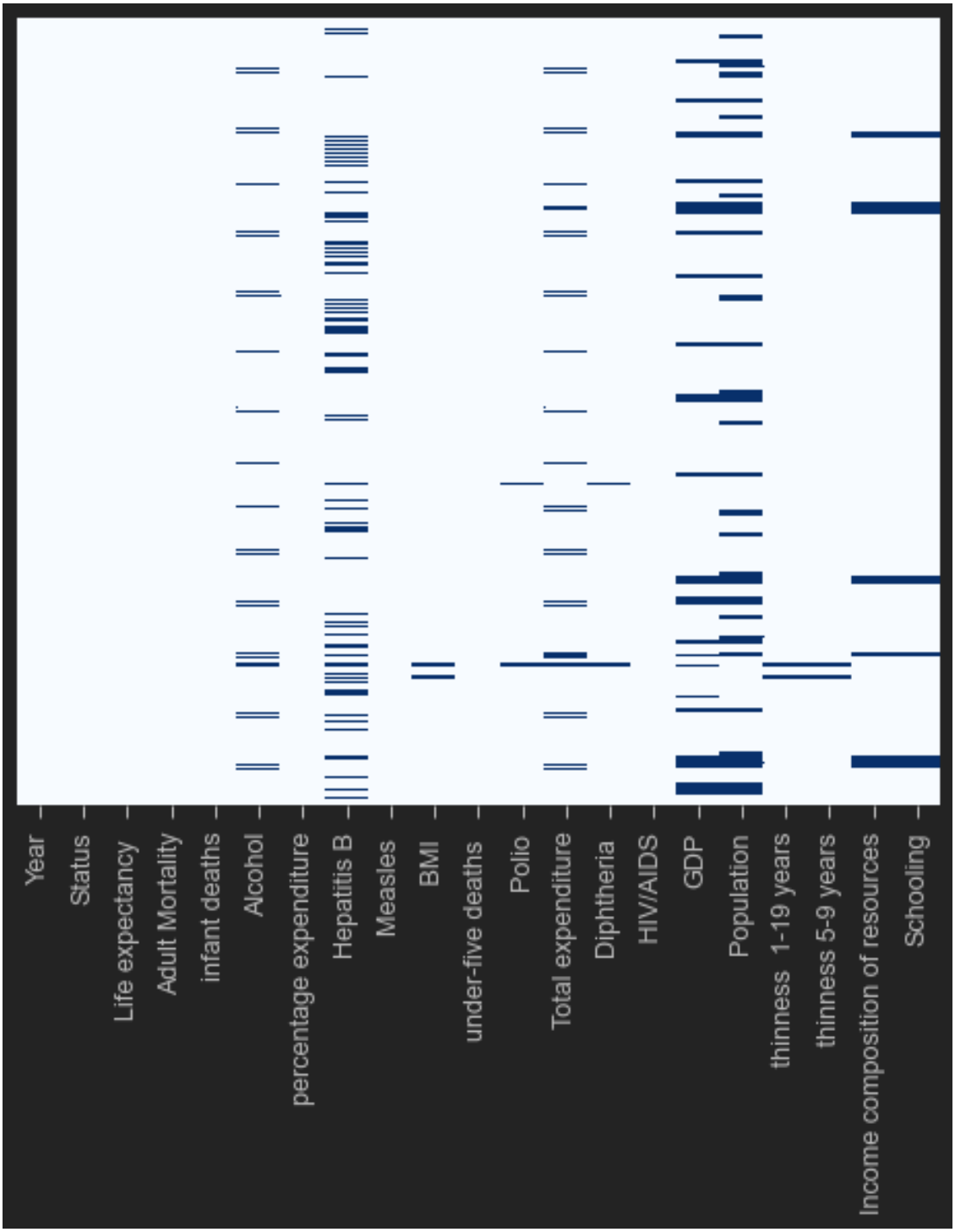
[7]:
```
life_expectancy_df.tail(7)
```

[7]:
```
      Year      Status  Life expectancy   Adult Mortality  infant deaths  \
2931  2006  Developing              45.4              7.0             28
2932  2005  Developing              44.6            717.0             28
2933  2004  Developing              44.3            723.0             27
2934  2003  Developing              44.5            715.0             26
2935  2002  Developing              44.8             73.0             25
2936  2001  Developing              45.3            686.0             25
2937  2000  Developing              46.0            665.0             24

      Alcohol  percentage expenditure  Hepatitis B  Measles    BMI   …  \
2931     4.57               34.262169         68.0       212   27.9  …
2932     4.14                8.717409         65.0       420   27.5  …
2933     4.36                0.000000         68.0        31   27.1  …
2934     4.06                0.000000          7.0       998   26.7  …
2935     4.43                0.000000         73.0       304   26.3  …
2936     1.72                0.000000         76.0       529   25.9  …
2937     1.68                0.000000         79.0      1483   25.5  …

      Polio  Total expenditure  Diphtheria    HIV/AIDS         GDP  \
2931   71.0               5.12          7.0       26.8  414.796232
2932   69.0               6.44         68.0       30.3  444.765750
2933   67.0               7.13         65.0       33.6  454.366654
2934    7.0               6.52         68.0       36.7  453.351155
2935   73.0               6.53         71.0       39.8   57.348340
2936   76.0               6.16         75.0       42.1  548.587312
2937   78.0               7.10         78.0       43.5  547.358879

      Population   thinness  1-19 years   thinness 5-9 years  \
2931  13124267.0                    8.6                  8.6
2932    129432.0                    9.0                  9.0
2933  12777511.0                    9.4                  9.4
2934  12633897.0                    9.8                  9.9
2935    125525.0                    1.2                  1.3
2936  12366165.0                    1.6                  1.7
2937  12222251.0                   11.0                 11.2

      Income composition of resources  Schooling
2931                            0.408        9.5
2932                            0.406        9.3
```

```
2933                             0.407        9.2
2934                             0.418        9.5
2935                             0.427       10.0
2936                             0.427        9.8
2937                             0.434        9.8

[7 rows x 21 columns]
```

[8]:
```
# check if there are any Null values
sns.heatmap(life_expectancy_df.isnull(), yticklabels = False, cbar = False,␣
↪cmap="Blues")
```

[8]:
```
<AxesSubplot:>
```



[9]:
```
# Check the dataframe info
life_expectancy_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 2938 entries, 0 to 2937
Data columns (total 21 columns):
 #   Column                           Non-Null Count  Dtype
---  ------                           --------------  -----
 0   Year                             2938 non-null   int64
 1   Status                           2938 non-null   object
 2   Life expectancy                  2928 non-null   float64
 3   Adult Mortality                  2928 non-null   float64
 4   infant deaths                    2938 non-null   int64
 5   Alcohol                          2744 non-null   float64
 6   percentage expenditure           2938 non-null   float64
 7   Hepatitis B                      2385 non-null   float64
 8   Measles                          2938 non-null   int64
 9    BMI                             2904 non-null   float64
 10  under-five deaths                2938 non-null   int64
 11  Polio                            2919 non-null   float64
 12  Total expenditure                2712 non-null   float64
 13  Diphtheria                       2919 non-null   float64
 14   HIV/AIDS                        2938 non-null   float64
 15  GDP                              2490 non-null   float64
 16  Population                       2286 non-null   float64
 17   thinness  1-19 years            2904 non-null   float64
 18   thinness 5-9 years              2904 non-null   float64
 19  Income composition of resources  2771 non-null   float64
 20  Schooling                        2775 non-null   float64
dtypes: float64(16), int64(4), object(1)
memory usage: 482.1+ KB
```

[10]:
```
# Statistical summary of the dataframe

life_expectancy_df.describe()
```

[10]:

	Year	Life expectancy	Adult Mortality	infant deaths	\
count	2938.000000	2928.000000	2928.000000	2938.000000	
mean	2007.518720	69.224932	164.796448	30.303948	
std	4.613841	9.523867	124.292079	117.926501	
min	2000.000000	36.300000	1.000000	0.000000	
25%	2004.000000	63.100000	74.000000	0.000000	
50%	2008.000000	72.100000	144.000000	3.000000	
75%	2012.000000	75.700000	228.000000	22.000000	
max	2015.000000	89.000000	723.000000	1800.000000	

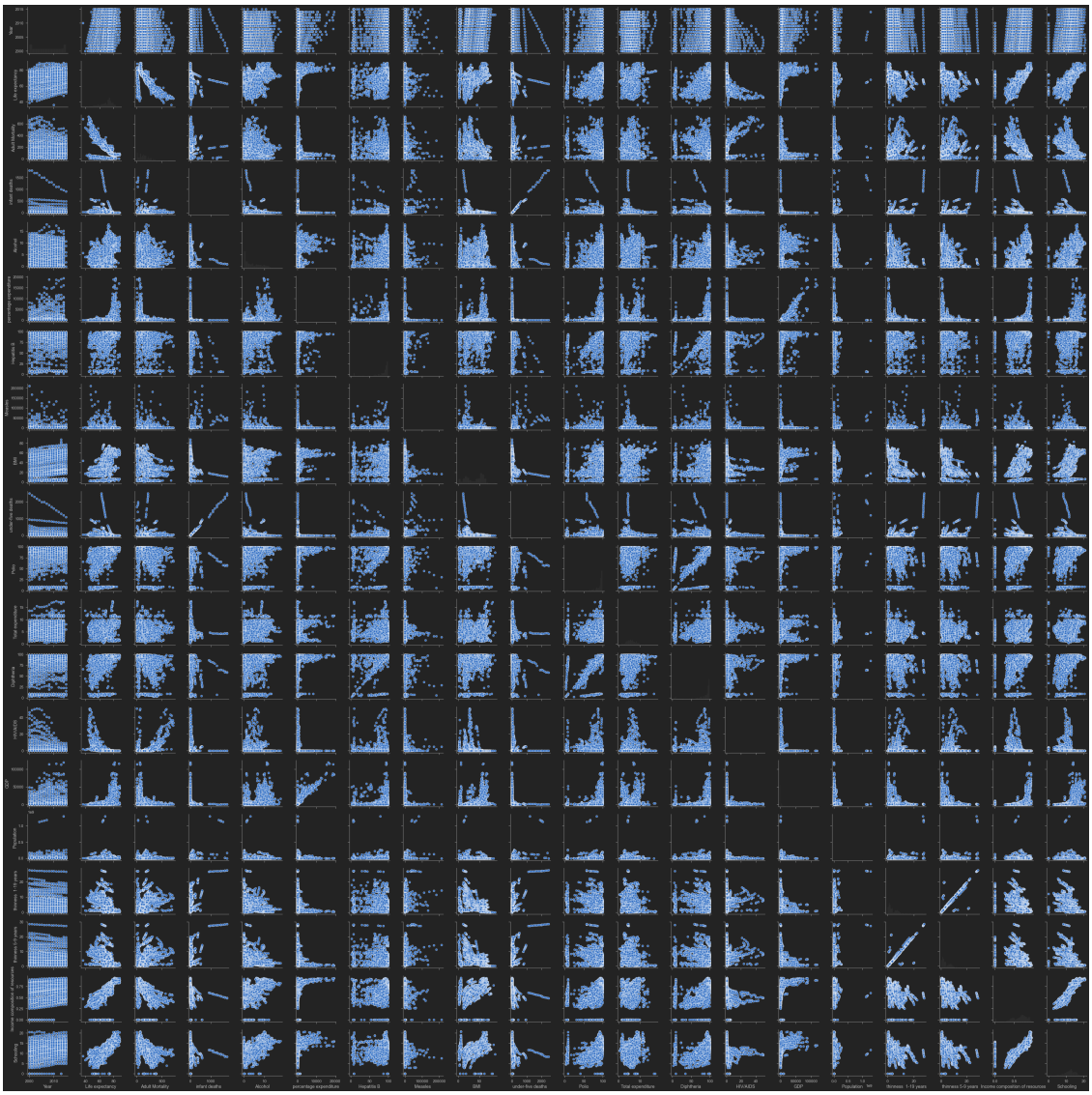
	Alcohol	percentage expenditure	Hepatitis B	Measles	\
count	2744.000000	2938.000000	2385.000000	2938.000000	
mean	4.602861	738.251295	80.940461	2419.592240	
std	4.052413	1987.914858	25.070016	11467.272489	
min	0.010000	0.000000	1.000000	0.000000	

```
25%       0.877500               4.685343    77.000000       0.000000
50%       3.755000              64.912906    92.000000      17.000000
75%       7.702500             441.534144    97.000000     360.250000
max      17.870000           19479.911610    99.000000  212183.000000

              BMI   under-five deaths         Polio  Total expenditure  \
count  2904.000000         2938.000000  2919.000000         2712.00000
mean     38.321247           42.035739    82.550188            5.93819
std      20.044034          160.445548    23.428046            2.49832
min       1.000000            0.000000     3.000000            0.37000
25%      19.300000            0.000000    78.000000            4.26000
50%      43.500000            4.000000    93.000000            5.75500
75%      56.200000           28.000000    97.000000            7.49250
max      87.300000         2500.000000    99.000000           17.60000

       Diphtheria      HIV/AIDS             GDP    Population  \
count  2919.000000  2938.000000     2490.000000  2.286000e+03
mean     82.324084     1.742103     7483.158469  1.275338e+07
std      23.716912     5.077785    14270.169342  6.101210e+07
min       2.000000     0.100000        1.681350  3.400000e+01
25%      78.000000     0.100000      463.935626  1.957932e+05
50%      93.000000     0.100000     1766.947595  1.386542e+06
75%      97.000000     0.800000     5910.806335  7.420359e+06
max      99.000000    50.600000   119172.741800  1.293859e+09

        thinness  1-19 years   thinness 5-9 years  \
count            2904.000000          2904.000000
mean                4.839704             4.870317
std                 4.420195             4.508882
min                 0.100000             0.100000
25%                 1.600000             1.500000
50%                 3.300000             3.300000
75%                 7.200000             7.200000
max                27.700000            28.600000

       Income composition of resources    Schooling
count                      2771.000000  2775.000000
mean                          0.627551    11.992793
std                           0.210904     3.358920
min                           0.000000     0.000000
25%                           0.493000    10.100000
50%                           0.677000    12.300000
75%                           0.779000    14.300000
max                           0.948000    20.700000
```

[11]:
```
# Plot the histogram
life_expectancy_df.hist(bins = 30, figsize = (20, 20), color = 'r');
```

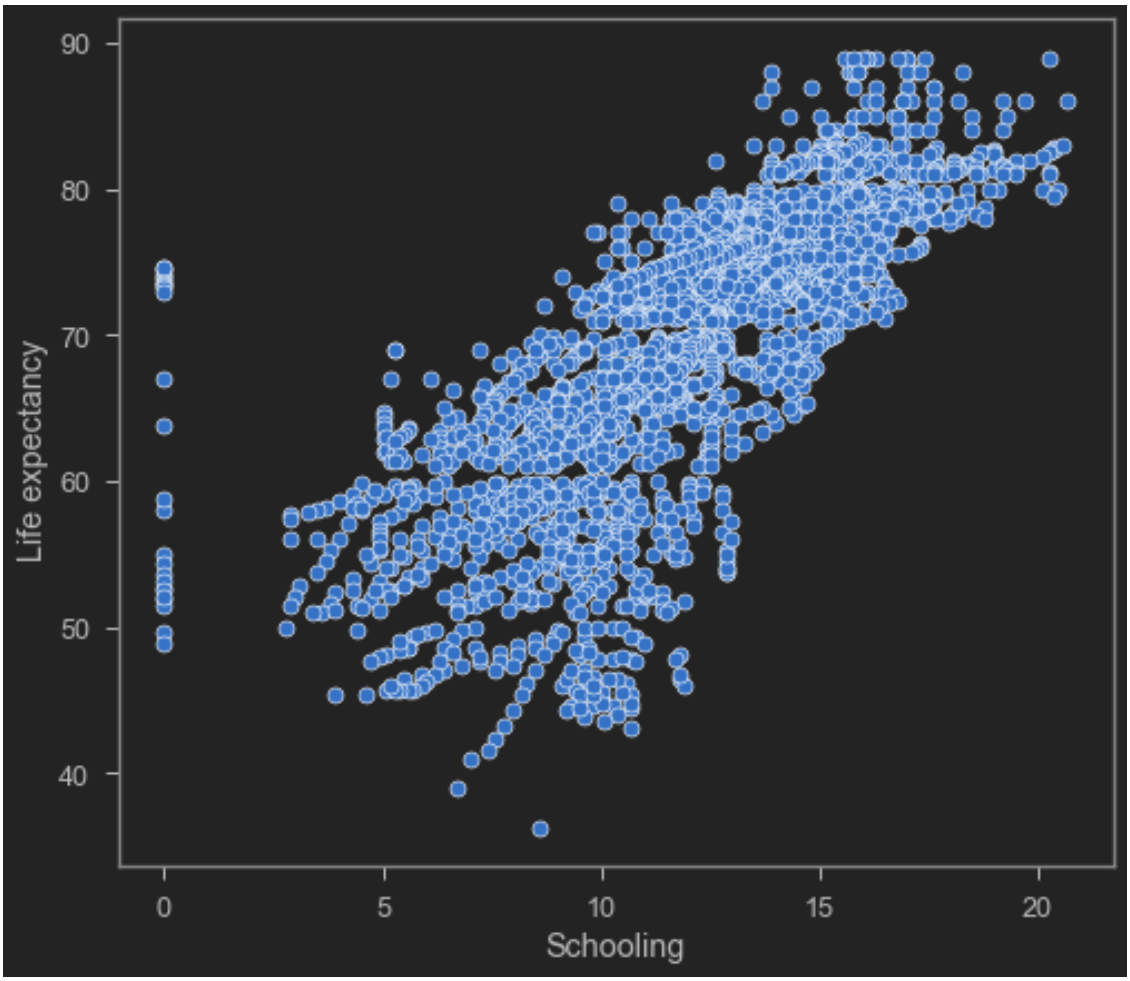
[12]:
```
# Plot pairplot
plt.figure(figsize = (20,20))
sns.pairplot(life_expectancy_df)
```

[12]:
```
<seaborn.axisgrid.PairGrid at 0x24b2471c8e0>

<Figure size 1440x1440 with 0 Axes>
```



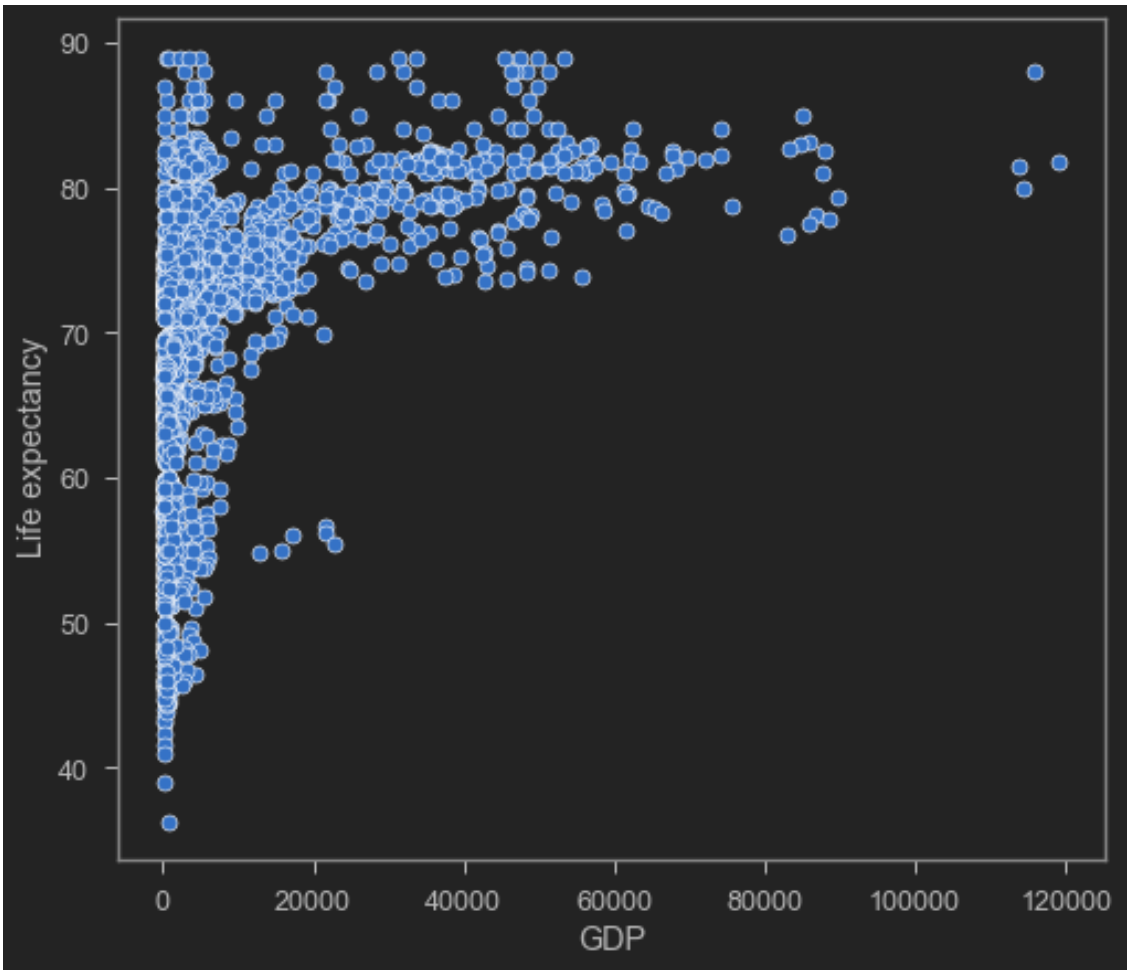
[13]:
```
sns.scatterplot(data = life_expectancy_df, x = 'Schooling', y = 'Life␣
↪expectancy ')
```

[13]:
```
<AxesSubplot:xlabel='Schooling', ylabel='Life expectancy '>
```



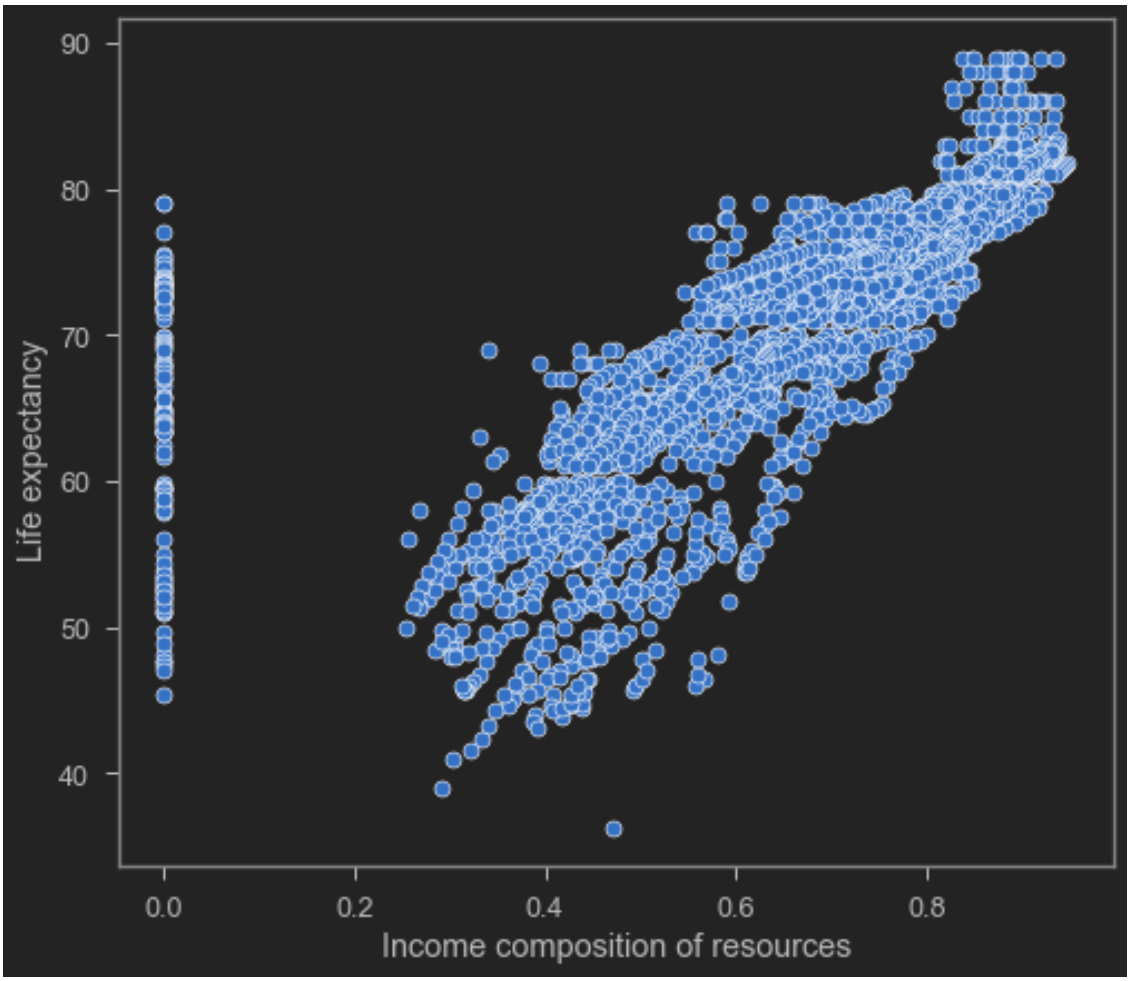
[14]:
```
sns.scatterplot(data = life_expectancy_df, x = 'GDP', y = 'Life expectancy ')
```

[14]: <AxesSubplot:xlabel='GDP', ylabel='Life expectancy '>



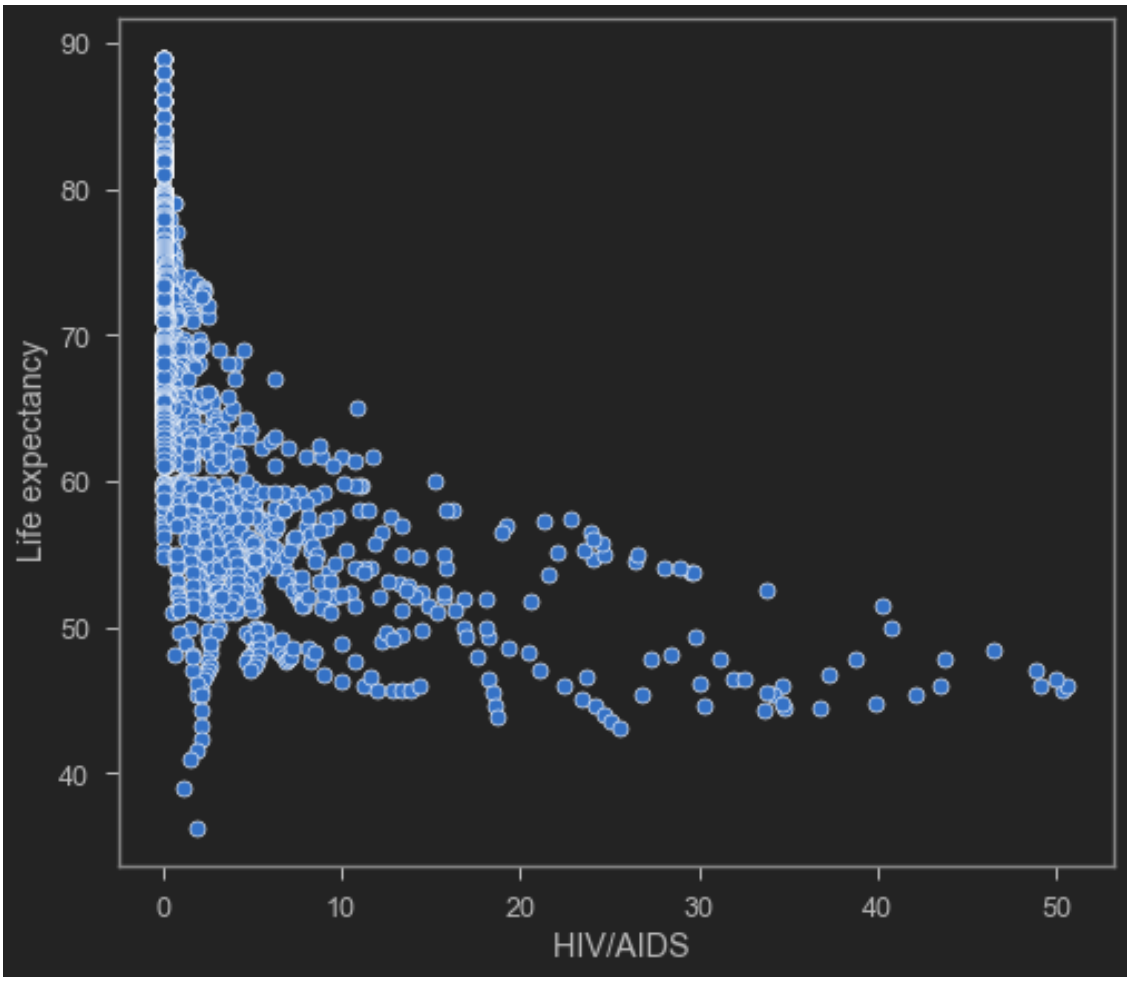
[15]:
```
sns.scatterplot(data = life_expectancy_df, x = 'Income composition of␣
↪resources', y = 'Life expectancy ')
# How productive resources are used
```

[15]:
```
<AxesSubplot:xlabel='Income composition of resources', ylabel='Life expectancy
'>
```



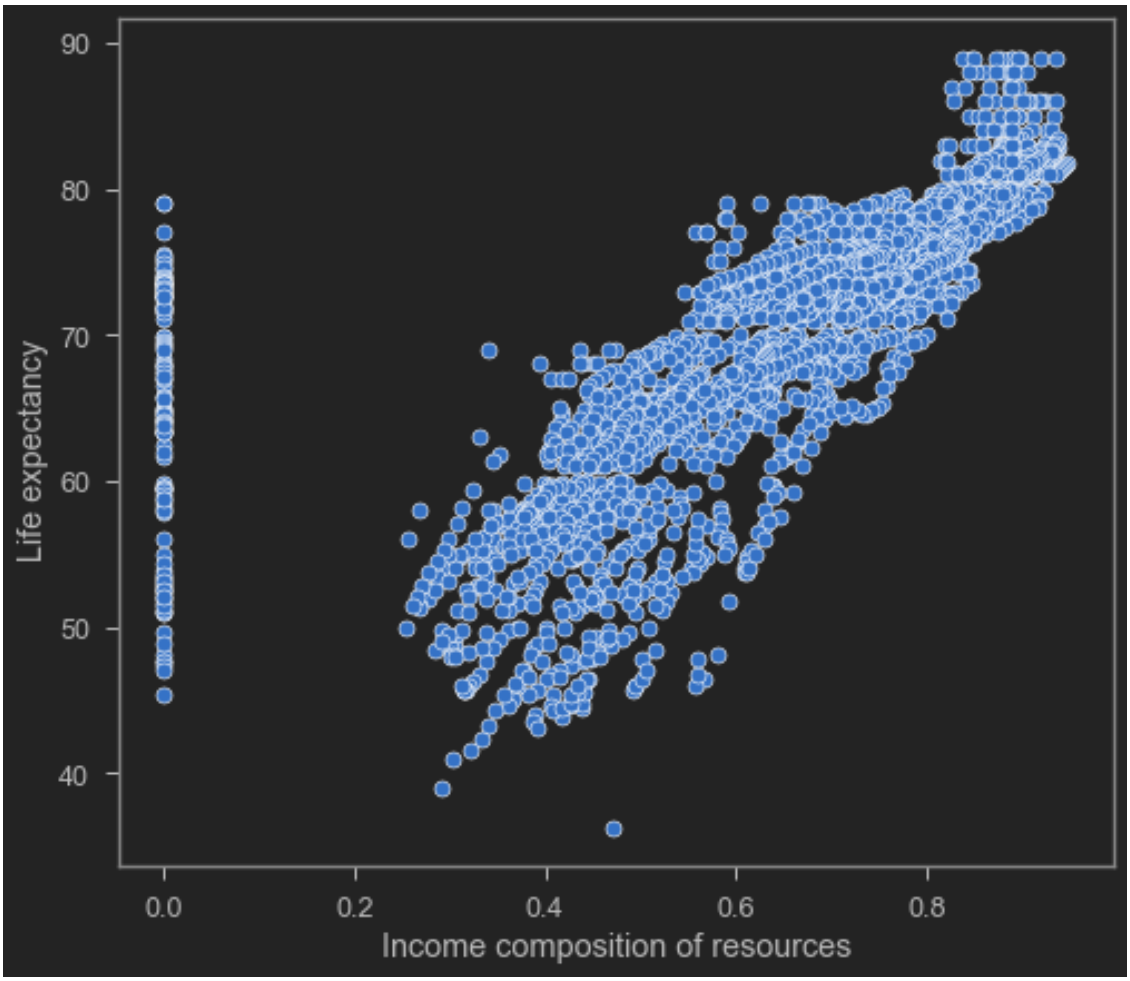
[16]:
```
sns.scatterplot(data = life_expectancy_df, x = ' HIV/AIDS', y = 'Life␣
↪expectancy ')
# HIV/AIDs death rate increases, life expectancy is reduced
```

[16]:
```
<AxesSubplot:xlabel=' HIV/AIDS', ylabel='Life expectancy '>
```

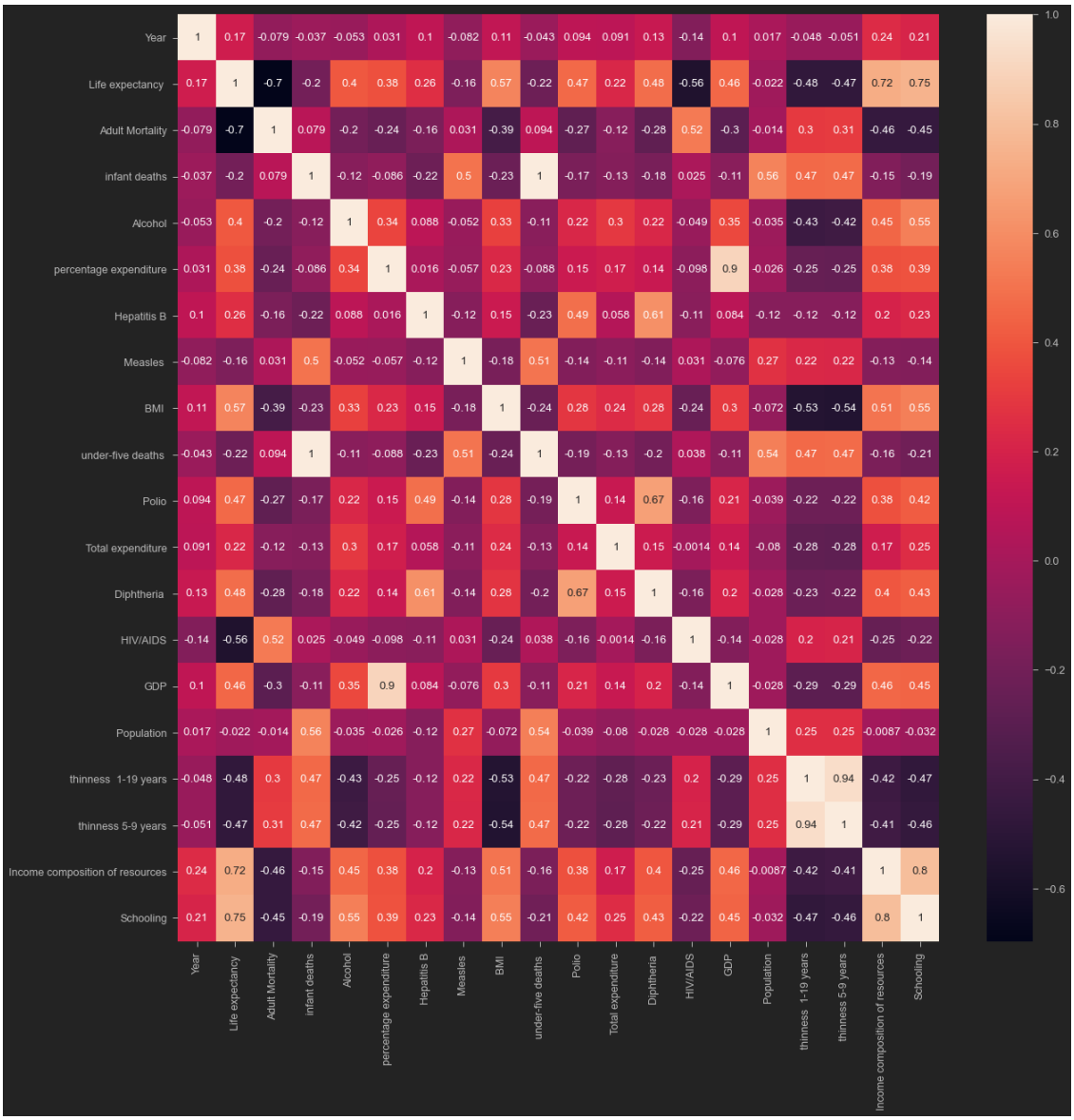


[17]:
```
sns.scatterplot(data = life_expectancy_df, x = 'Income composition of␣
↪resources', y = 'Life expectancy ')
# How productive resources are used
```

[17]:
```
<AxesSubplot:xlabel='Income composition of resources', ylabel='Life expectancy
'>
```



[18]:
```
# Plot the correlation matrix

plt.figure(figsize = (20,20))
corr_matrix = life_expectancy_df.corr()
sns.heatmap(corr_matrix, annot = True)
plt.show()
```



[19]:
```
life_expectancy_df
```

[19]:

	Year	Status	Life expectancy	Adult Mortality	infant deaths	\
0	2015	Developing	65.0	263.0	62	
1	2014	Developing	59.9	271.0	64	
2	2013	Developing	59.9	268.0	66	
3	2012	Developing	59.5	272.0	69	
4	2011	Developing	59.2	275.0	71	
…	…	…	…	…	…	
2933	2004	Developing	44.3	723.0	27	
2934	2003	Developing	44.5	715.0	26	

```
2935  2002  Developing              44.8             73.0             25
2936  2001  Developing              45.3            686.0             25
2937  2000  Developing              46.0            665.0             24

      Alcohol  percentage expenditure  Hepatitis B  Measles    BMI   …  \
0        0.01               71.279624         65.0      1154   19.1  …
1        0.01               73.523582         62.0       492   18.6  …
2        0.01               73.219243         64.0       430   18.1  …
3        0.01               78.184215         67.0      2787   17.6  …
4        0.01                7.097109         68.0      3013   17.2  …
…         …                      …           …         …     …  …
2933     4.36               0.000000         68.0        31   27.1  …
2934     4.06               0.000000          7.0       998   26.7  …
2935     4.43               0.000000         73.0       304   26.3  …
2936     1.72               0.000000         76.0       529   25.9  …
2937     1.68               0.000000         79.0      1483   25.5  …

      Polio  Total expenditure  Diphtheria    HIV/AIDS         GDP  \
0       6.0               8.16         65.0        0.1  584.259210
1      58.0               8.18         62.0        0.1  612.696514
2      62.0               8.13         64.0        0.1  631.744976
3      67.0               8.52         67.0        0.1  669.959000
4      68.0               7.87         68.0        0.1   63.537231
…      …                  …            …           …          …
2933   67.0               7.13         65.0       33.6  454.366654
2934    7.0               6.52         68.0       36.7  453.351155
2935   73.0               6.53         71.0       39.8   57.348340
2936   76.0               6.16         75.0       42.1  548.587312
2937   78.0               7.10         78.0       43.5  547.358879

      Population   thinness  1-19 years   thinness 5-9 years  \
0     33736494.0                   17.2                 17.3
1       327582.0                   17.5                 17.5
2     31731688.0                   17.7                 17.7
3      3696958.0                   17.9                 18.0
4      2978599.0                   18.2                 18.2
…          …                       …                    …
2933  12777511.0                    9.4                  9.4
2934  12633897.0                    9.8                  9.9
2935    125525.0                    1.2                  1.3
2936  12366165.0                    1.6                  1.7
2937  12222251.0                   11.0                 11.2

      Income composition of resources  Schooling
0                               0.479       10.1
1                               0.476       10.0
2                               0.470        9.9
```

```
3                                0.463        9.8
4                                0.454        9.5
…                                …          …
2933                             0.407        9.2
2934                             0.418        9.5
2935                             0.427       10.0
2936                             0.427        9.8
2937                             0.434        9.8

[2938 rows x 21 columns]
```

[20]:
```
# Checking the unique values in country to consider it as a categorical variable
life_expectancy_df['Status'].nunique()
```

[20]:
```
2
```

[21]:
```
# Perform one-hot encoding
life_expectancy_df = pd.get_dummies(life_expectancy_df, columns = ['Status'])
```

[22]:
```
life_expectancy_df
```

[22]:
```
      Year  Life expectancy   Adult Mortality  infant deaths  Alcohol  \
0     2015              65.0            263.0             62     0.01
1     2014              59.9            271.0             64     0.01
2     2013              59.9            268.0             66     0.01
3     2012              59.5            272.0             69     0.01
4     2011              59.2            275.0             71     0.01
…    …                 …               …                 …       …
2933  2004              44.3            723.0             27     4.36
2934  2003              44.5            715.0             26     4.06
2935  2002              44.8             73.0             25     4.43
2936  2001              45.3            686.0             25     1.72
2937  2000              46.0            665.0             24     1.68

      percentage expenditure  Hepatitis B  Measles    BMI   \
0                  71.279624         65.0      1154   19.1
1                  73.523582         62.0       492   18.6
2                  73.219243         64.0       430   18.1
3                  78.184215         67.0      2787   17.6
4                   7.097109         68.0      3013   17.2
…                        …          …         …     …
2933                0.000000         68.0        31   27.1
2934                0.000000          7.0       998   26.7
2935                0.000000         73.0       304   26.3
2936                0.000000         76.0       529   25.9
2937                0.000000         79.0      1483   25.5
```

```
      under-five deaths    …  Diphtheria    HIV/AIDS         GDP  Population  \
0                     83   …         65.0         0.1  584.259210  33736494.0
1                     86   …         62.0         0.1  612.696514    327582.0
2                     89   …         64.0         0.1  631.744976  31731688.0
3                     93   …         67.0         0.1  669.959000   3696958.0
4                     97   …         68.0         0.1   63.537231   2978599.0
…                   …  …           …          …          …           …
2933                  42   …         65.0        33.6  454.366654  12777511.0
2934                  41   …         68.0        36.7  453.351155  12633897.0
2935                  40   …         71.0        39.8   57.348340    125525.0
2936                  39   …         75.0        42.1  548.587312  12366165.0
2937                  39   …         78.0        43.5  547.358879  12222251.0

       thinness  1-19 years   thinness 5-9 years  \
0                      17.2                 17.3
1                      17.5                 17.5
2                      17.7                 17.7
3                      17.9                 18.0
4                      18.2                 18.2
…                      …                  …
2933                    9.4                  9.4
2934                    9.8                  9.9
2935                    1.2                  1.3
2936                    1.6                  1.7
2937                   11.0                 11.2

      Income composition of resources  Schooling  Status_Developed  \
0                               0.479       10.1                 0
1                               0.476       10.0                 0
2                               0.470        9.9                 0
3                               0.463        9.8                 0
4                               0.454        9.5                 0
…                               …          …                …
2933                            0.407        9.2                 0
2934                            0.418        9.5                 0
2935                            0.427       10.0                 0
2936                            0.427        9.8                 0
2937                            0.434        9.8                 0

      Status_Developing
0                     1
1                     1
2                     1
3                     1
4                     1
…                   …
2933                  1
```

```
2934                 1
2935                 1
2936                 1
2937                 1

[2938 rows x 22 columns]
```

[23]:
```
# Check the number of null values for the columns having null values
life_expectancy_df.isnull().sum()[np.where(life_expectancy_df.isnull().sum() !=␣
 ↪0)[0]]
```

[23]:
```
Life expectancy                  10
Adult Mortality                  10
Alcohol                         194
Hepatitis B                     553
 BMI                             34
Polio                            19
Total expenditure               226
Diphtheria                       19
GDP                             448
Population                      652
 thinness  1-19 years            34
 thinness 5-9 years              34
Income composition of resources  167
Schooling                       163
dtype: int64
```

[24]:
```
# Since most of the are continous values we fill them with mean
life_expectancy_df = life_expectancy_df.apply(lambda x: x.fillna(x.
 ↪mean()),axis=0)
```

[25]:
```
life_expectancy_df.isnull().sum()[np.where(life_expectancy_df.isnull().sum() !=␣
 ↪0)[0]]
```

[25]:
```
Series([], dtype: int64)
```

[26]:
```
life_expectancy_df['Life expectancy '].max()
```

[26]:
```
89.0
```

[27]:
```
life_expectancy_df.describe()
```

[27]:
```
              Year  Life expectancy   Adult Mortality  infant deaths  \
count  2938.000000       2938.000000      2938.000000    2938.000000
mean   2007.518720         69.224932       164.796448      30.303948
std       4.613841          9.507640       124.080302     117.926501
min    2000.000000         36.300000         1.000000       0.000000
25%    2004.000000         63.200000        74.000000       0.000000
```

```
50%    2008.000000         72.000000       144.000000       3.000000
75%    2012.000000         75.600000       227.000000      22.000000
max    2015.000000         89.000000       723.000000    1800.000000

           Alcohol  percentage expenditure  Hepatitis B       Measles   \
count  2938.000000             2938.000000  2938.000000    2938.000000
mean      4.602861              738.251295    80.940461    2419.592240
std       3.916288             1987.914858    22.586855   11467.272489
min       0.010000                0.000000     1.000000       0.000000
25%       1.092500                4.685343    80.940461       0.000000
50%       4.160000               64.912906    87.000000      17.000000
75%       7.390000              441.534144    96.000000     360.250000
max      17.870000            19479.911610    99.000000  212183.000000

              BMI   under-five deaths   …  Diphtheria      HIV/AIDS  \
count  2938.000000         2938.000000  …  2938.000000  2938.000000
mean     38.321247           42.035739  …    82.324084     1.742103
std      19.927677          160.445548  …    23.640073     5.077785
min       1.000000            0.000000  …     2.000000     0.100000
25%      19.400000            0.000000  …    78.000000     0.100000
50%      43.000000            4.000000  …    93.000000     0.100000
75%      56.100000           28.000000  …    97.000000     0.800000
max      87.300000         2500.000000  …    99.000000    50.600000

                GDP    Population   thinness  1-19 years  \
count    2938.000000  2.938000e+03             2938.000000
mean     7483.158469  1.275338e+07                4.839704
std     13136.800417  5.381546e+07                4.394535
min         1.681350  3.400000e+01                0.100000
25%       580.486996  4.189172e+05                1.600000
50%      3116.561755  3.675929e+06                3.400000
75%      7483.158469  1.275338e+07                7.100000
max    119172.741800  1.293859e+09               27.700000

        thinness 5-9 years  Income composition of resources    Schooling  \
count          2938.000000                      2938.000000  2938.000000
mean              4.870317                         0.627551    11.992793
std               4.482708                         0.204820     3.264381
min               0.100000                         0.000000     0.000000
25%               1.600000                         0.504250    10.300000
50%               3.400000                         0.662000    12.100000
75%               7.200000                         0.772000    14.100000
max              28.600000                         0.948000    20.700000

       Status_Developed  Status_Developing
count       2938.000000        2938.000000
mean           0.174268           0.825732
```

```
std            0.379405          0.379405
min            0.000000          0.000000
25%            0.000000          1.000000
50%            0.000000          1.000000
75%            0.000000          1.000000
max            1.000000          1.000000

[8 rows x 22 columns]
```

[28]:
```
life_expectancy_df
```

[28]:
```
      Year  Life expectancy   Adult Mortality  infant deaths  Alcohol  \
0     2015              65.0            263.0             62     0.01
1     2014              59.9            271.0             64     0.01
2     2013              59.9            268.0             66     0.01
3     2012              59.5            272.0             69     0.01
4     2011              59.2            275.0             71     0.01
…     …                 …               …                 …      …
2933  2004              44.3            723.0             27     4.36
2934  2003              44.5            715.0             26     4.06
2935  2002              44.8             73.0             25     4.43
2936  2001              45.3            686.0             25     1.72
2937  2000              46.0            665.0             24     1.68

      percentage expenditure  Hepatitis B  Measles    BMI   \
0                  71.279624         65.0      1154   19.1
1                  73.523582         62.0       492   18.6
2                  73.219243         64.0       430   18.1
3                  78.184215         67.0      2787   17.6
4                   7.097109         68.0      3013   17.2
…                        …            …         …     …
2933                0.000000         68.0        31   27.1
2934                0.000000          7.0       998   26.7
2935                0.000000         73.0       304   26.3
2936                0.000000         76.0       529   25.9
2937                0.000000         79.0      1483   25.5

      under-five deaths   …  Diphtheria    HIV/AIDS         GDP  Population  \
0                    83   …         65.0         0.1  584.259210  33736494.0
1                    86   …         62.0         0.1  612.696514    327582.0
2                    89   …         64.0         0.1  631.744976  31731688.0
3                    93   …         67.0         0.1  669.959000   3696958.0
4                    97   …         68.0         0.1   63.537231   2978599.0
…                   …     …            …         …          …           …
2933                 42   …         65.0        33.6  454.366654  12777511.0
2934                 41   …         68.0        36.7  453.351155  12633897.0
2935                 40   …         71.0        39.8   57.348340    125525.0
```

```
2936                 39  …         75.0       42.1  548.587312  12366165.0
2937                 39  …         78.0       43.5  547.358879  12222251.0

       thinness  1-19 years   thinness 5-9 years  \
0                      17.2                 17.3
1                      17.5                 17.5
2                      17.7                 17.7
3                      17.9                 18.0
4                      18.2                 18.2
…                      …                    …
2933                    9.4                  9.4
2934                    9.8                  9.9
2935                    1.2                  1.3
2936                    1.6                  1.7
2937                   11.0                 11.2

      Income composition of resources  Schooling  Status_Developed  \
0                               0.479       10.1                 0
1                               0.476       10.0                 0
2                               0.470        9.9                 0
3                               0.463        9.8                 0
4                               0.454        9.5                 0
…                               …        …                   …
2933                            0.407        9.2                 0
2934                            0.418        9.5                 0
2935                            0.427       10.0                 0
2936                            0.427        9.8                 0
2937                            0.434        9.8                 0

      Status_Developing
0                     1
1                     1
2                     1
3                     1
4                     1
…                   …
2933                  1
2934                  1
2935                  1
2936                  1
2937                  1

[2938 rows x 22 columns]
```

[29]:
```
# Create train and test data

X = life_expectancy_df.drop(columns = ['Life expectancy '])
```

```
y = life_expectancy_df[['Life expectancy ']]
```

[30]:
```
X
```

[30]:
```
      Year  Adult Mortality  infant deaths  Alcohol  percentage expenditure  \
0     2015            263.0             62     0.01               71.279624
1     2014            271.0             64     0.01               73.523582
2     2013            268.0             66     0.01               73.219243
3     2012            272.0             69     0.01               78.184215
4     2011            275.0             71     0.01                7.097109
…      …                …              …        …                       …
2933  2004            723.0             27     4.36                0.000000
2934  2003            715.0             26     4.06                0.000000
2935  2002             73.0             25     4.43                0.000000
2936  2001            686.0             25     1.72                0.000000
2937  2000            665.0             24     1.68                0.000000

      Hepatitis B  Measles    BMI   under-five deaths   Polio  …  \
0            65.0      1154   19.1                  83    6.0  …
1            62.0       492   18.6                  86   58.0  …
2            64.0       430   18.1                  89   62.0  …
3            67.0      2787   17.6                  93   67.0  …
4            68.0      3013   17.2                  97   68.0  …
…             …         …     …                     …    …  …
2933         68.0        31   27.1                  42   67.0  …
2934          7.0       998   26.7                  41    7.0  …
2935         73.0       304   26.3                  40   73.0  …
2936         76.0       529   25.9                  39   76.0  …
2937         79.0      1483   25.5                  39   78.0  …

      Diphtheria    HIV/AIDS         GDP  Population   thinness  1-19 years  \
0            65.0        0.1  584.259210  33736494.0                   17.2
1            62.0        0.1  612.696514    327582.0                   17.5
2            64.0        0.1  631.744976  31731688.0                   17.7
3            67.0        0.1  669.959000   3696958.0                   17.9
4            68.0        0.1   63.537231   2978599.0                   18.2
…             …          …          …           …                       …
2933         65.0       33.6  454.366654  12777511.0                    9.4
2934         68.0       36.7  453.351155  12633897.0                    9.8
2935         71.0       39.8   57.348340    125525.0                    1.2
2936         75.0       42.1  548.587312  12366165.0                    1.6
2937         78.0       43.5  547.358879  12222251.0                   11.0

       thinness 5-9 years  Income composition of resources  Schooling  \
0                    17.3                            0.479       10.1
1                    17.5                            0.476       10.0
2                    17.7                            0.470        9.9
```

```
3                    18.0                              0.463        9.8
4                    18.2                              0.454        9.5
…                     …                                …          …
2933                  9.4                              0.407        9.2
2934                  9.9                              0.418        9.5
2935                  1.3                              0.427       10.0
2936                  1.7                              0.427        9.8
2937                 11.2                              0.434        9.8

      Status_Developed  Status_Developing
0                    0                  1
1                    0                  1
2                    0                  1
3                    0                  1
4                    0                  1
…                   …                  …
2933                 0                  1
2934                 0                  1
2935                 0                  1
2936                 0                  1
2937                 0                  1

[2938 rows x 21 columns]
```

[31]:
```
y
```

[31]:
```
      Life expectancy
0                65.0
1                59.9
2                59.9
3                59.5
4                59.2
…                …
2933             44.3
2934             44.5
2935             44.8
2936             45.3
2937             46.0

[2938 rows x 1 columns]
```

[32]:
```
X.shape
```

[32]:
```
(2938, 21)
```

[33]:
```
y.shape
```

[33]: (2938, 1)

[34]:
```
# Convert the data type to float32

X = np.array(X).astype('float32')
y = np.array(y).astype('float32')
```

[35]:
```
# Only take the numerical variables and scale them
X
```

[35]:
```
array([[2.015e+03, 2.630e+02, 6.200e+01, …, 1.010e+01, 0.000e+00,
        1.000e+00],
       [2.014e+03, 2.710e+02, 6.400e+01, …, 1.000e+01, 0.000e+00,
        1.000e+00],
       [2.013e+03, 2.680e+02, 6.600e+01, …, 9.900e+00, 0.000e+00,
        1.000e+00],
       …,
       [2.002e+03, 7.300e+01, 2.500e+01, …, 1.000e+01, 0.000e+00,
        1.000e+00],
       [2.001e+03, 6.860e+02, 2.500e+01, …, 9.800e+00, 0.000e+00,
        1.000e+00],
       [2.000e+03, 6.650e+02, 2.400e+01, …, 9.800e+00, 0.000e+00,
        1.000e+00]], dtype=float32)
```

[36]:
```
# split the data into test and train sets
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2)
```

[37]:
```
# Scale the data
from sklearn.preprocessing import StandardScaler

scaler_X = StandardScaler()
X_train = scaler_X.fit_transform(X_train)
X_test = scaler_X.transform(X_test)

scaler_y = StandardScaler()
y_train = scaler_y.fit_transform(y_train)
y_test = scaler_y.transform(y_test)
```

MINI CHALLENGE #3: - Try splitting the data into 75% for training and the rest for testing

[ ]:

[38]:
```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test=train_test_split(X,y,test_size=0.20)
```

[39]:
```
X_train.shape
```

[39]:
```
(2350, 21)
```

[40]:
```
X_test.shape
```

[40]:
```
(588, 21)
```

[41]:
```
# using linear regression model
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, accuracy_score

regresssion_model_sklearn = LinearRegression(fit_intercept = True)
regresssion_model_sklearn.fit(X_train, y_train)
```

[41]:
```
LinearRegression()
```

[42]:
```
regresssion_model_sklearn_accuracy = regresssion_model_sklearn.score(X_test,␣
 ↪y_test)
regresssion_model_sklearn_accuracy
```

[42]:
```
0.6673040770349055
```

[43]:
```
print('Linear Model Coefficient (m): ', regresssion_model_sklearn.coef_)
print('Linear Model Coefficient (b): ', regresssion_model_sklearn.intercept_)
```

```
Linear Model Coefficient (m):  [[ 5.71493059e-03 -3.36682610e-02  1.18380375e-01
-4.37989132e-03
   2.17831926e-04 -1.80292260e-02 -1.88034519e-05  1.22053288e-01
  -8.97760913e-02  4.26209085e-02  1.81469857e-03  6.73580617e-02
  -9.91011225e-03  9.72555717e-05  3.47874973e-09 -5.43040456e-03
  -3.60582490e-03  7.72881263e-04  1.08329700e-02 -4.54001292e-06
   4.54001292e-06]]
Linear Model Coefficient (b):  [50.18261]
```

MINI CHALLENGE #4: - Retrain the model while setting the fit_intercept = False, what do you notice?

[44]:
```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, accuracy_score

regresssion_model_sklearn = LinearRegression(fit_intercept = True)
regresssion_model_sklearn.fit(X_train, y_train)
regresssion_model_sklearn_accuracy = regresssion_model_sklearn.score(X_test,␣
 ↪y_test)
print('Linear Model Coefficient (m): ', regresssion_model_sklearn.coef_)
print('Linear Model Coefficient (b): ', regresssion_model_sklearn.intercept_)
```

```
Linear Model Coefficient (m):  [[ 5.71493059e-03 -3.36682610e-02  1.18380375e-01
-4.37989132e-03
   2.17831926e-04 -1.80292260e-02 -1.88034519e-05  1.22053288e-01
```

```
  -8.97760913e-02  4.26209085e-02  1.81469857e-03  6.73580617e-02
  -9.91011225e-03  9.72555717e-05  3.47874973e-09 -5.43040456e-03
  -3.60582490e-03  7.72881263e-04  1.08329700e-02 -4.54001292e-06
   4.54001292e-06]]
Linear Model Coefficient (b):  [50.18261]
```

[45]:
```
# Make prediction

y_predict = regresssion_model_sklearn.predict(X_test)
```

[46]:
```
y_predict
```

[46]:
```
array([[80.58934 ],
       [55.917507],
       [80.084274],
       [65.14007 ],
       [76.79739 ],
       [70.81549 ],
       [76.14473 ],
       [63.41815 ],
       [69.74739 ],
       [55.325638],
       [72.85785 ],
       [58.103493],
       [58.017475],
       [73.225655],
       [76.11931 ],
       [76.91576 ],
       [70.85108 ],
       [72.352005],
       [66.4031  ],
       [69.21193 ],
       [66.92422 ],
       [66.799194],
       [77.70748 ],
       [62.547447],
       [64.309074],
       [71.10952 ],
       [50.672176],
       [76.422485],
       [70.93765 ],
       [66.81561 ],
       [80.53246 ],
       [73.57443 ],
       [73.54893 ],
       [72.632034],
       [66.1771  ],
```

```
[55.86878 ],
[70.221085],
[70.29849 ],
[64.78299 ],
[69.54752 ],
[66.99455 ],
[56.893875],
[80.68474 ],
[72.41958 ],
[61.393612],
[63.746365],
[53.453857],
[69.880554],
[69.08386 ],
[65.235245],
[56.010513],
[71.647644],
[71.566246],
[56.873222],
[67.54059 ],
[79.75193 ],
[77.67082 ],
[70.78409 ],
[52.648876],
[79.51733 ],
[76.58489 ],
[59.491127],
[68.83852 ],
[75.58244 ],
[72.43326 ],
[61.995358],
[63.01261 ],
[60.735596],
[73.913574],
[71.199844],
[62.40103 ],
[82.45397 ],
[70.40784 ],
[69.42379 ],
[66.23884 ],
[69.87781 ],
[69.33186 ],
[69.56039 ],
[76.070724],
[53.124786],
[47.431965],
[53.8238  ],
```

```
[53.154556],
[71.37909 ],
[58.846146],
[74.598465],
[70.62909 ],
[57.54695 ],
[76.62864 ],
[43.428856],
[73.868454],
[67.22849 ],
[61.02767 ],
[63.378532],
[57.631023],
[55.58265 ],
[81.99772 ],
[79.53713 ],
[67.976036],
[78.27576 ],
[55.383293],
[72.97229 ],
[77.91264 ],
[66.910385],
[47.418213],
[83.10025 ],
[71.327255],
[58.125973],
[68.76688 ],
[73.653305],
[58.209476],
[72.55483 ],
[69.86521 ],
[70.20803 ],
[53.61019 ],
[76.733406],
[68.8277  ],
[73.00745 ],
[67.53569 ],
[79.262405],
[77.49602 ],
[72.62472 ],
[61.15097 ],
[69.5208  ],
[70.128555],
[68.91522 ],
[60.21156 ],
[69.24339 ],
[73.21512 ],
```

```
[72.76382 ],
[72.68309 ],
[70.422455],
[76.501236],
[64.65337 ],
[52.42755 ],
[73.05197 ],
[70.29227 ],
[77.11867 ],
[72.21271 ],
[81.9994  ],
[74.845474],
[76.16255 ],
[70.56378 ],
[69.17212 ],
[62.11275 ],
[63.7269  ],
[65.22955 ],
[57.16826 ],
[66.99128 ],
[51.023575],
[72.795296],
[73.29509 ],
[74.54265 ],
[65.76767 ],
[79.666435],
[68.263   ],
[59.417057],
[76.88797 ],
[63.162704],
[56.25554 ],
[72.80091 ],
[60.078316],
[61.07898 ],
[75.6492  ],
[73.29383 ],
[75.547104],
[71.7791  ],
[77.10523 ],
[75.67863 ],
[69.393875],
[65.47504 ],
[53.615837],
[76.6653  ],
[80.05727 ],
[75.95909 ],
[70.25573 ],
```

```
[73.76813 ],
[73.23988 ],
[66.32898 ],
[69.13969 ],
[64.53602 ],
[57.240715],
[76.30483 ],
[67.27693 ],
[71.95332 ],
[65.7961  ],
[75.18176 ],
[71.59563 ],
[77.3637  ],
[76.20502 ],
[80.83119 ],
[68.32204 ],
[65.92694 ],
[56.720505],
[76.440994],
[73.18414 ],
[54.010784],
[73.54417 ],
[53.06186 ],
[64.46517 ],
[71.87745 ],
[61.32178 ],
[81.705246],
[76.644   ],
[66.220055],
[50.66408 ],
[73.4028  ],
[79.5016  ],
[66.752785],
[74.26907 ],
[64.35736 ],
[74.56084 ],
[71.10534 ],
[79.358574],
[51.90413 ],
[54.039375],
[76.167175],
[62.741417],
[76.431305],
[70.79687 ],
[76.2809  ],
[74.598366],
[68.05918 ],
```

[69.30491],
[61.911186],
[67.58536],
[60.64859],
[60.05945],
[69.377686],
[71.06184],
[70.44601],
[57.577034],
[73.95355],
[73.066414],
[69.012985],
[68.58326],
[68.511856],
[67.95749],
[70.00306],
[76.391975],
[66.71825],
[70.22889],
[39.97853],
[65.20597],
[69.364075],
[64.019966],
[70.55681],
[70.74203],
[64.16548],
[73.35677],
[60.85255],
[67.962326],
[75.024185],
[71.22139],
[67.51558],
[55.938545],
[44.679142],
[69.079666],
[55.9115],
[76.00175],
[54.690758],
[63.830242],
[70.77961],
[56.867435],
[76.86063],
[71.61151],
[84.70001],
[50.953785],
[70.193344],
[65.81881],

```
[78.61712 ],
[75.84401 ],
[57.242443],
[86.01918 ],
[71.31578 ],
[56.18623 ],
[63.37242 ],
[64.24695 ],
[67.458694],
[63.894714],
[73.250626],
[68.63004 ],
[75.95959 ],
[72.99304 ],
[60.303543],
[74.922295],
[68.02803 ],
[72.62897 ],
[64.31707 ],
[73.93691 ],
[70.956894],
[62.756355],
[66.69403 ],
[66.974075],
[83.8615  ],
[78.270065],
[68.61348 ],
[77.59049 ],
[55.720238],
[79.14634 ],
[58.84308 ],
[67.81693 ],
[60.594143],
[73.80217 ],
[79.190094],
[79.368095],
[69.256256],
[73.09564 ],
[78.39492 ],
[70.59516 ],
[65.545135],
[76.36004 ],
[61.488323],
[71.53942 ],
[70.42683 ],
[57.26667 ],
[68.6906  ],
```

```
[58.115944],
[73.50658 ],
[58.641193],
[79.34309 ],
[78.01307 ],
[72.2733  ],
[71.21321 ],
[73.29703 ],
[75.986   ],
[69.12429 ],
[62.59523 ],
[76.25482 ],
[76.37598 ],
[68.14957 ],
[72.1156  ],
[62.45862 ],
[75.019714],
[64.03378 ],
[73.12152 ],
[69.07567 ],
[77.39453 ],
[58.820183],
[68.64554 ],
[72.96172 ],
[71.93889 ],
[68.536934],
[78.61095 ],
[65.802055],
[72.443596],
[76.653336],
[59.979343],
[75.47627 ],
[56.086803],
[73.38674 ],
[77.6638  ],
[68.13034 ],
[68.003494],
[57.28502 ],
[66.144775],
[65.348816],
[75.96449 ],
[68.12697 ],
[74.7174  ],
[57.30956 ],
[67.716194],
[82.979744],
[68.33775 ],
```

```
[73.56966 ],
[64.88919 ],
[72.76103 ],
[62.434437],
[77.07836 ],
[57.90441 ],
[69.0654  ],
[70.35305 ],
[81.07982 ],
[73.00481 ],
[76.831474],
[73.81197 ],
[71.605064],
[63.16722 ],
[66.88185 ],
[70.30853 ],
[72.72699 ],
[70.08819 ],
[67.04757 ],
[64.53994 ],
[69.03232 ],
[59.978695],
[47.554256],
[72.66286 ],
[77.35136 ],
[68.06872 ],
[75.61194 ],
[63.18627 ],
[64.07055 ],
[72.24607 ],
[62.056885],
[66.90651 ],
[73.25249 ],
[75.91372 ],
[61.85061 ],
[66.96493 ],
[78.65423 ],
[73.68138 ],
[75.296715],
[77.8911  ],
[77.97929 ],
[64.63207 ],
[82.94056 ],
[80.4916  ],
[71.13336 ],
[71.20915 ],
[68.9114  ],
```

```
[62.47932 ],
[66.14798 ],
[78.40641 ],
[75.41904 ],
[76.64599 ],
[78.755714],
[77.04816 ],
[70.96801 ],
[58.192047],
[69.38227 ],
[71.04652 ],
[73.47714 ],
[75.956   ],
[73.51676 ],
[60.403122],
[67.14332 ],
[78.82875 ],
[75.532104],
[63.640774],
[65.83181 ],
[76.06582 ],
[64.266846],
[67.55332 ],
[75.36337 ],
[58.6632  ],
[73.356674],
[71.35309 ],
[73.91118 ],
[62.913445],
[73.10293 ],
[74.43396 ],
[71.46292 ],
[67.80356 ],
[68.03362 ],
[67.27316 ],
[63.48135 ],
[70.809494],
[81.03255 ],
[71.86958 ],
[66.30582 ],
[72.91072 ],
[64.977974],
[56.848736],
[75.562706],
[69.45861 ],
[70.69133 ],
[72.15283 ],
```

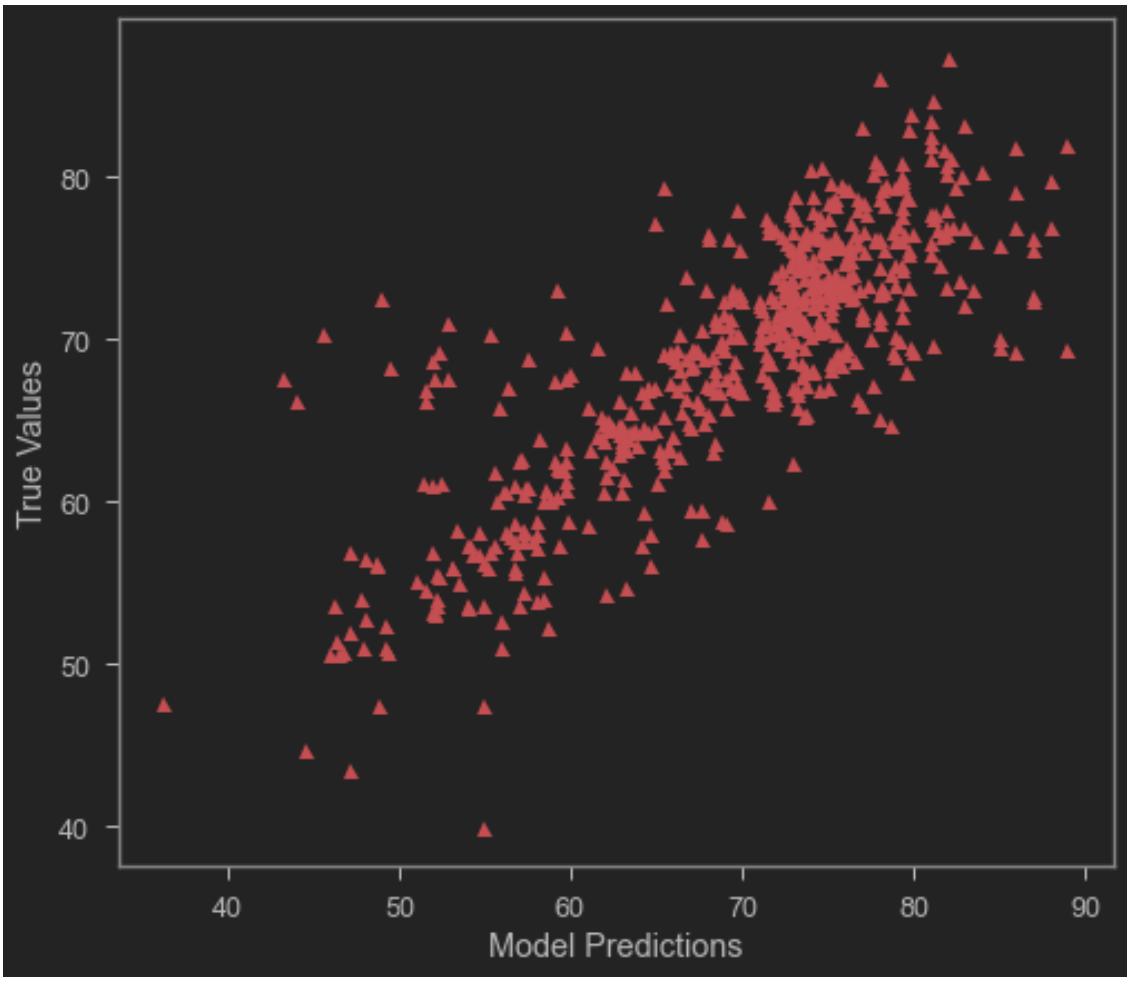
```
[66.842255],
[53.610695],
[70.310905],
[69.35025 ],
[67.03751 ],
[70.725334],
[71.28952 ],
[67.112854],
[77.06624 ],
[75.509094],
[79.73018 ],
[70.03879 ],
[70.80191 ],
[60.95331 ],
[78.6263  ],
[74.82854 ],
[66.33368 ],
[73.444496],
[72.60324 ],
[59.46156 ],
[50.973194],
[69.49855 ],
[64.4733  ],
[52.283978],
[54.03982 ],
[64.72506 ],
[67.37655 ],
[68.270294],
[57.66253 ],
[69.2416  ],
[81.73343 ],
[75.4781  ],
[72.51446 ],
[72.72553 ],
[76.17113 ],
[50.554184],
[68.02292 ],
[74.43485 ],
[72.38485 ],
[73.888016],
[72.57619 ],
[61.10907 ],
[69.35773 ],
[66.11013 ],
[60.887875],
[77.6472  ],
[68.84299 ],
```

```
[64.65614 ],
[66.88781 ],
[55.10727 ],
[75.98692 ],
[75.877884],
[74.98157 ],
[74.294304],
[75.70006 ],
[74.40963 ],
[80.13984 ],
[78.550255],
[71.01561 ],
[60.549286],
[72.38141 ],
[75.68578 ],
[80.10564 ],
[69.67556 ],
[76.159454],
[66.305824],
[50.52682 ],
[66.984634],
[64.48521 ],
[57.991676],
[83.50929 ],
[73.142   ],
[74.710594],
[78.08568 ],
[54.938408],
[71.77594 ],
[72.41804 ],
[81.17804 ],
[72.09921 ],
[52.810116],
[74.02733 ],
[75.54865 ],
[67.15611 ],
[58.514805],
[68.97707 ],
[54.338833],
[56.6862  ],
[74.29269 ],
[69.72372 ],
[69.50523 ],
[54.597168],
[68.495735],
[76.92918 ],
[71.43854 ],
```

```
[75.76344 ],
[87.33328 ],
[53.622078],
[51.368023],
[72.399635],
[71.51123 ],
[54.359833],
[61.95117 ],
[80.27665 ],
[73.5398  ],
[68.608795],
[72.649734],
[53.638172],
[63.150116],
[76.41124 ],
[60.59329 ],
[75.42286 ],
[74.23348 ],
[77.45598 ],
[72.81793 ],
[73.622246],
[78.73018 ],
[70.19907 ],
[65.3129  ],
[62.45713 ],
[72.27916 ],
[73.02316 ],
[64.835304],
[66.596565],
[79.07062 ],
[78.02291 ],
[74.57812 ],
[60.71723 ],
[56.533863],
[67.59474 ],
[73.35457 ]], dtype=float32)
```

[47]:
```
# Plot the scaled result

plt.plot(y_test, y_predict, "^", color = 'r')
plt.xlabel('Model Predictions')
plt.ylabel('True Values')
```

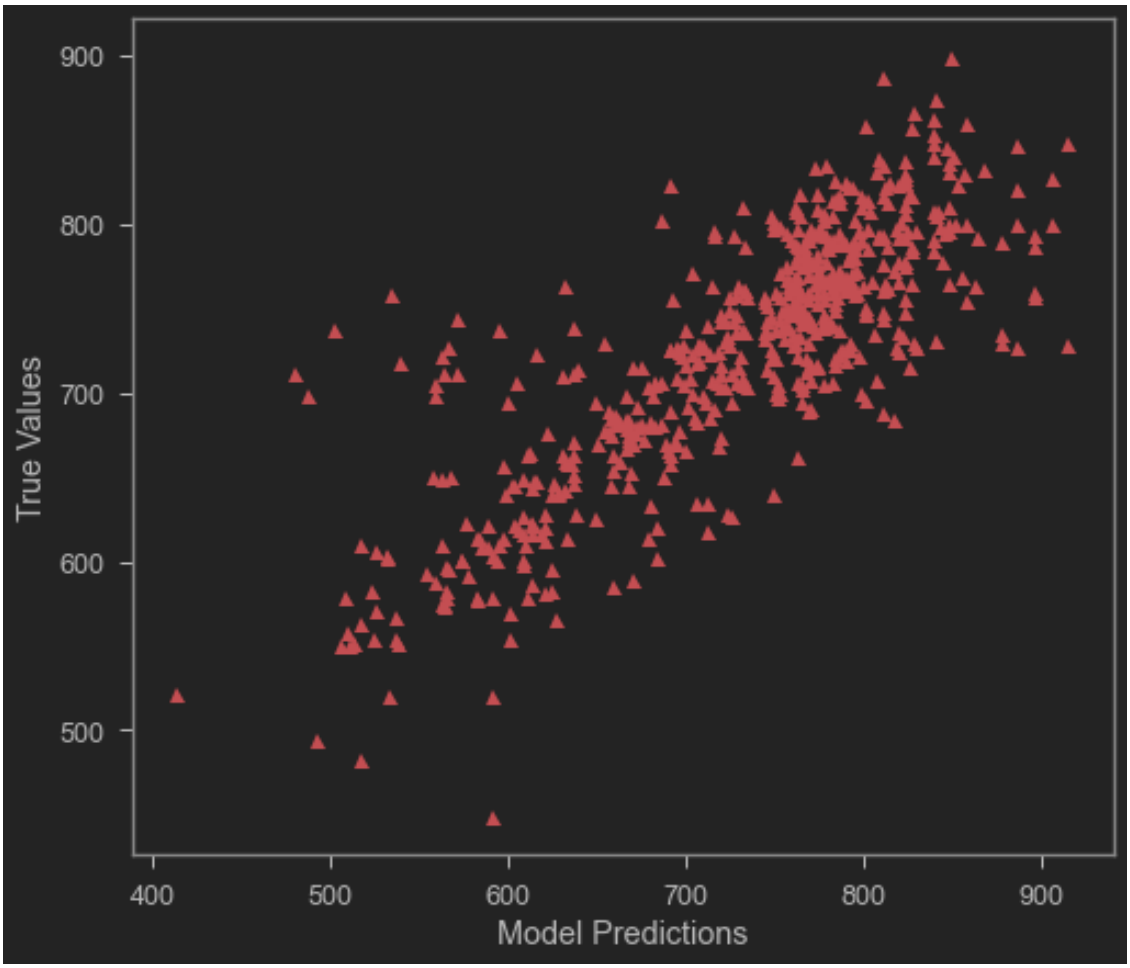
[47]:
```
Text(0, 0.5, 'True Values')
```



[48]:
```
y_predict_orig = scaler_y.inverse_transform(y_predict)
y_test_orig = scaler_y.inverse_transform(y_test)
```

[49]:
```
# Plot the original values

plt.plot(y_test_orig, y_predict_orig, "^", color = 'r')
plt.xlabel('Model Predictions')
plt.ylabel('True Values')
```

[49]:
```
Text(0, 0.5, 'True Values')
```



[50]:
```
# Plot the KPIs

from sklearn.metrics import r2_score, mean_squared_error, mean_absolute_error
from math import sqrt

k = X_test.shape[1]
n = len(X_test)
RMSE = float(format(np.sqrt(mean_squared_error(y_test_orig, y_predict_orig)),'.
 ↪3f'))
MSE = mean_squared_error(y_test_orig, y_predict_orig)
MAE = mean_absolute_error(y_test_orig, y_predict_orig)
r2 = r2_score(y_test_orig, y_predict_orig)
adj_r2 = 1-(1-r2)*(n-1)/(n-k-1)

print('RMSE =',RMSE, '\nMSE =',MSE, '\nMAE =',MAE, '\nR2 =', r2, '\nAdjusted R2␣
 ↪=', adj_r2)
```

```
RMSE = 52.287
```

```
MSE = 2733.9524
MAE = 37.247715
R2 = 0.6673040595597917
Adjusted R2 = 0.6549602172466391
```

[51]:
```
regresssion_model_sklearn = LinearRegression(fit_intercept = False)
regresssion_model_sklearn.fit(X_train, y_train)
from sklearn.metrics import r2_score, mean_squared_error, mean_absolute_error
from math import sqrt

k = X_test.shape[1]
n = len(X_test)
RMSE = float(format(np.sqrt(mean_squared_error(y_test_orig, y_predict_orig)),'.
↪3f'))
MSE = mean_squared_error(y_test_orig, y_predict_orig)
MAE = mean_absolute_error(y_test_orig, y_predict_orig)
r2 = r2_score(y_test_orig, y_predict_orig)
adj_r2 = 1-(1-r2)*(n-1)/(n-k-1)

print('RMSE =',RMSE, '\nMSE =',MSE, '\nMAE =',MAE, '\nR2 =', r2, '\nAdjusted R2␣
↪=', adj_r2)
```

```
RMSE = 52.287
MSE = 2733.9524
MAE = 37.247715
R2 = 0.6673040595597917
Adjusted R2 = 0.6549602172466391
```

[52]:
```
life_expectancy_df.head(7)
```

[52]:

	Year	Life expectancy	Adult Mortality	infant deaths	Alcohol	\
0	2015	65.0	263.0	62	0.01	
1	2014	59.9	271.0	64	0.01	
2	2013	59.9	268.0	66	0.01	
3	2012	59.5	272.0	69	0.01	
4	2011	59.2	275.0	71	0.01	
5	2010	58.8	279.0	74	0.01	
6	2009	58.6	281.0	77	0.01	

	percentage expenditure	Hepatitis B	Measles	BMI	under-five deaths	\
0	71.279624	65.0	1154	19.1	83	
1	73.523582	62.0	492	18.6	86	
2	73.219243	64.0	430	18.1	89	
3	78.184215	67.0	2787	17.6	93	
4	7.097109	68.0	3013	17.2	97	
5	79.679367	66.0	1989	16.7	102	
6	56.762217	63.0	2861	16.2	106	

```
   …  Diphtheria    HIV/AIDS         GDP  Population   thinness  1-19 years  \
0  …         65.0         0.1  584.259210  33736494.0                  17.2
1  …         62.0         0.1  612.696514    327582.0                  17.5
2  …         64.0         0.1  631.744976  31731688.0                  17.7
3  …         67.0         0.1  669.959000   3696958.0                  17.9
4  …         68.0         0.1   63.537231   2978599.0                  18.2
5  …         66.0         0.1  553.328940   2883167.0                  18.4
6  …         63.0         0.1  445.893298    284331.0                  18.6

    thinness 5-9 years  Income composition of resources  Schooling  \
0                 17.3                            0.479       10.1
1                 17.5                            0.476       10.0
2                 17.7                            0.470        9.9
3                 18.0                            0.463        9.8
4                 18.2                            0.454        9.5
5                 18.4                            0.448        9.2
6                 18.7                            0.434        8.9

   Status_Developed  Status_Developing
0                 0                  1
1                 0                  1
2                 0                  1
3                 0                  1
4                 0                  1
5                 0                  1
6                 0                  1

[7 rows x 22 columns]
```

[53]:
```
life_expectancy_df['Life expectancy '].max()
```

[53]:
```
89.0
```