```
1 !pip install hvplot
```

```
Collecting hvplot
  Downloading hvplot-0.9.2-py2.py3-none-any.whl (1.8 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.8/1.8 MB 9.8 MB/s eta 0:00:00
Requirement already satisfied: bokeh>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from hvplot) (3.3.4)
Requirement already satisfied: colorcet>=2 in /usr/local/lib/python3.10/dist-packages (from hvplot) (3.1.0)
Requirement already satisfied: holoviews>=1.11.0 in /usr/local/lib/python3.10/dist-packages (from hvplot) (1.17.1)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from hvplot) (2.0.3)
Requirement already satisfied: numpy>=1.15 in /usr/local/lib/python3.10/dist-packages (from hvplot) (1.25.2)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from hvplot) (24.0)
Requirement already satisfied: panel>=0.11.0 in /usr/local/lib/python3.10/dist-packages (from hvplot) (1.3.8)
Requirement already satisfied: param<3.0,>=1.12.0 in /usr/local/lib/python3.10/dist-packages (from hvplot) (2.1.0)
Requirement already satisfied: Jinja2>=2.9 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (3.1.3)
Requirement already satisfied: contourpy>=1 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (1.2.1)
Requirement already satisfied: pillow>=7.1.0 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (9.4.0)
Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (6.0.1)
Requirement already satisfied: tornado>=5.1 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (6.3.3)
Requirement already satisfied: xyzservices>=2021.09.1 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (2024.4.0)
Requirement already satisfied: pyviz-comms>=0.7.4 in /usr/local/lib/python3.10/dist-packages (from holoviews>=1.11.0->hvplot) (3.0.2)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas->hvplot) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->hvplot) (2023.4)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages (from pandas->hvplot) (2024.1)
Requirement already satisfied: markdown in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (3.6)
Requirement already satisfied: markdown-it-py in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (3.0.0)
Requirement already satisfied: linkify-it-py in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (2.0.3)
Requirement already satisfied: mdit-py-plugins in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (0.4.0)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (2.31.0)
Requirement already satisfied: tqdm>=4.48.0 in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (4.66.2)
Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (6.1.0)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (4.11.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2>=2.9->bokeh>=1.0.0->hvplot) (2.1.5)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2->pandas->hvplot) (1.16.0)
Requirement already satisfied: webencodings in /usr/local/lib/python3.10/dist-packages (from bleach->panel>=0.11.0->hvplot) (0.5.1)
Requirement already satisfied: uc-micro-py in /usr/local/lib/python3.10/dist-packages (from linkify-it-py->panel>=0.11.0->hvplot) (1.0.3)
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.10/dist-packages (from markdown-it-py->panel>=0.11.0->hvplot) (0.1.2)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages (from requests->panel>=0.11.0->hvplot) (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages (from requests->panel>=0.11.0->hvplot) (3.7)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages (from requests->panel>=0.11.0->hvplot) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->panel>=0.11.0->hvplot) (2024.2.2)
Installing collected packages: hvplot
Successfully installed hvplot-0.9.2
```

```
 1 import pandas as pd
 2 import numpy as np
 3 import matplotlib.pyplot as plt
 4 import seaborn as sns
 5 import hvplot.pandas
 6
 7 from sklearn.model_selection import train_test_split
 8
 9 from sklearn import metrics
10
11 from sklearn.linear_model import LinearRegression
12
13 %matplotlib inline
```

Check Out Data

```
df=pd.read_csv('/content/Life Expectancy Data.csv')
```

```
df.head()
```

	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	BMI	under-five deaths	P
0	Afghanistan	2015	Developing	65.0	263.0	62	0.01	71.279624	65.0	1154	19.1	83	
1	Afghanistan	2014	Developing	59.9	271.0	64	0.01	73.523582	62.0	492	18.6	86	
2	Afghanistan	2013	Developing	59.9	268.0	66	0.01	73.219243	64.0	430	18.1	89	
3	Afghanistan	2012	Developing	59.5	272.0	69	0.01	78.184215	67.0	2787	17.6	93	
4	Afghanistan	2011	Developing	59.2	275.0	71	0.01	7.097109	68.0	3013	17.2	97	

```
df.shape
```

```
(2938, 22)
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2938 entries, 0 to 2937
Data columns (total 22 columns):
 #   Column                           Non-Null Count  Dtype  
---  ------                           --------------  -----  
 0   Country                          2938 non-null   object 
 1   Year                             2938 non-null   int64  
 2   Status                           2938 non-null   object 
 3   Life expectancy                  2928 non-null   float64
 4   Adult Mortality                  2928 non-null   float64
 5   infant deaths                    2938 non-null   int64  
 6   Alcohol                          2744 non-null   float64
 7   percentage expenditure           2938 non-null   float64
 8   Hepatitis B                      2385 non-null   float64
 9   Measles                          2938 non-null   int64  
 10   BMI                             2904 non-null   float64
 11  under-five deaths                2938 non-null   int64  
 12  Polio                            2919 non-null   float64
 13  Total expenditure                2712 non-null   float64
 14  Diphtheria                       2919 non-null   float64
 15   HIV/AIDS                        2938 non-null   float64
 16  GDP                              2490 non-null   float64
 17  Population                       2286 non-null   float64
 18   thinness  1-19 years            2904 non-null   float64
 19   thinness 5-9 years             2904 non-null   float64
 20  Income composition of resources  2771 non-null   float64
 21  Schooling                        2775 non-null   float64
dtypes: float64(16), int64(4), object(2)
memory usage: 505.1+ KB
```

```
print(df.dtypes)
```

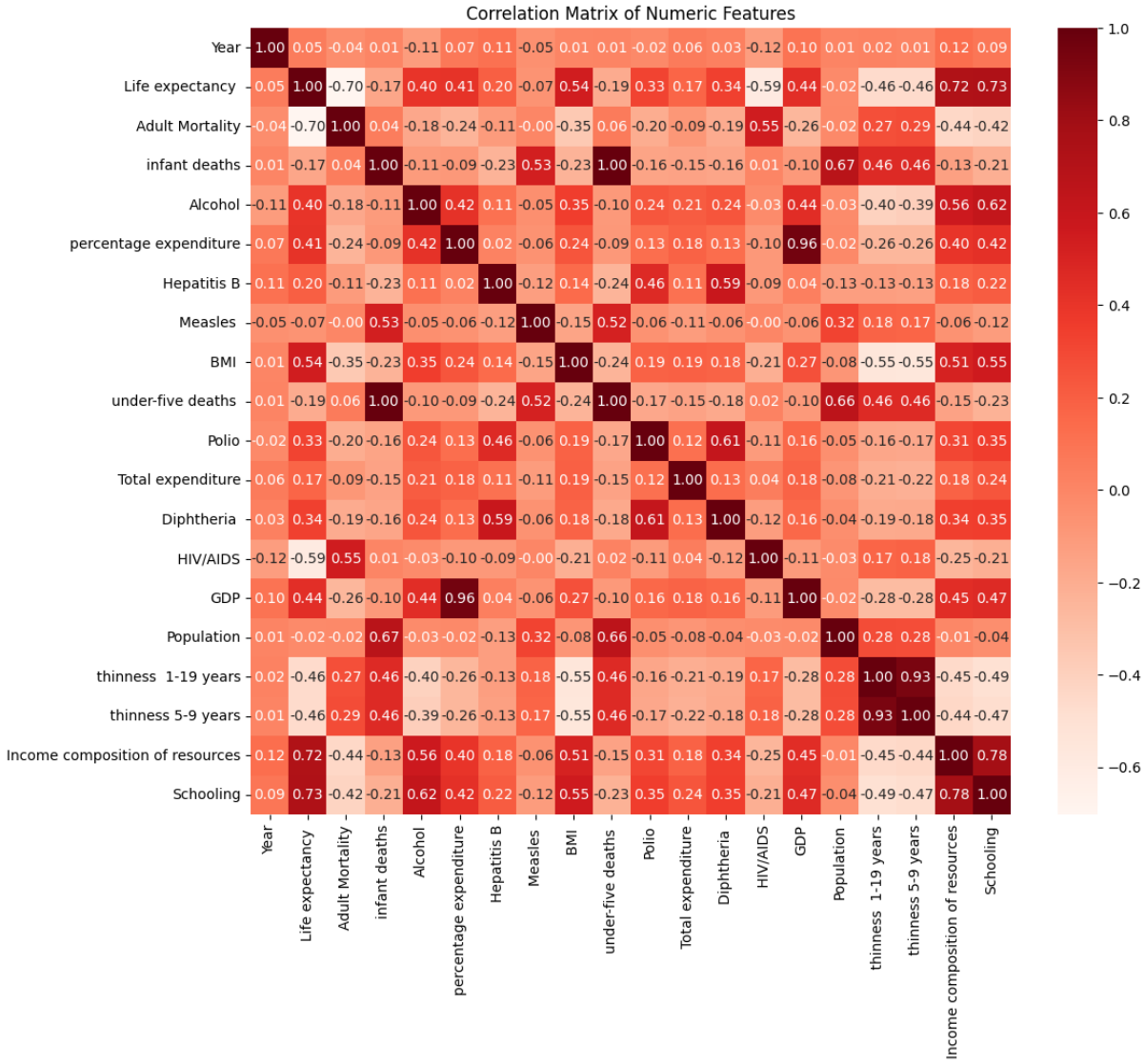
```
Country                           object
Year                               int64
Status                            object
Life expectancy                  float64
Adult Mortality                  float64
infant deaths                      int64
Alcohol                          float64
percentage expenditure           float64
Hepatitis B                      float64
Measles                            int64
 BMI                             float64
under-five deaths                  int64
Polio                            float64
Total expenditure                float64
Diphtheria                       float64
 HIV/AIDS                        float64
GDP                              float64
Population                       float64
 thinness  1-19 years            float64
 thinness 5-9 years              float64
Income composition of resources  float64
Schooling                        float64
dtype: object
```

```
numeric_df = df.select_dtypes(include=[np.number])

print(numeric_df.isnull().sum())

numeric_df = numeric_df.dropna()

correlation_matrix = numeric_df.corr()

print(correlation_matrix)
```

```
Year                               0
Life expectancy                   10
Adult Mortality                   10
infant deaths                      0
Alcohol                          194
percentage expenditure             0
Hepatitis B                      553
Measles                            0
 BMI                              34
under-five deaths                  0
Polio                             19
Total expenditure                226
Diphtheria                        19
 HIV/AIDS                          0
GDP                              448
Population                       652
 thinness  1-19 years             34
 thinness 5-9 years               34
Income composition of resources  167
Schooling                        163
dtype: int64
                                   Year  ...  Schooling
Year                           1.000000  ...   0.088732
Life expectancy                0.050771  ...   0.727630
```

```
Adult Mortality                -0.037092  ...  -0.421171
infant deaths                   0.008029  ...  -0.214372
Alcohol                        -0.113365  ...   0.616975
percentage expenditure          0.069553  ...   0.422088
Hepatitis B                     0.114897  ...   0.215182
Measles                        -0.053822  ...  -0.115660
 BMI                            0.005739  ...   0.554844
under-five deaths               0.010479  ...  -0.226013
Polio                          -0.016699  ...   0.350147
Total expenditure               0.059493  ...   0.243783
Diphtheria                      0.029641  ...   0.350398
 HIV/AIDS                      -0.123405  ...  -0.211840
GDP                             0.096421  ...   0.467947
Population                      0.012567  ...  -0.040312
 thinness  1-19 years           0.019757  ...  -0.491199
 thinness 5-9 years             0.014122  ...  -0.472482
Income composition of resources  0.122892  ...   0.784741
Schooling                       0.088732  ...   1.000000

[20 rows x 20 columns]
```

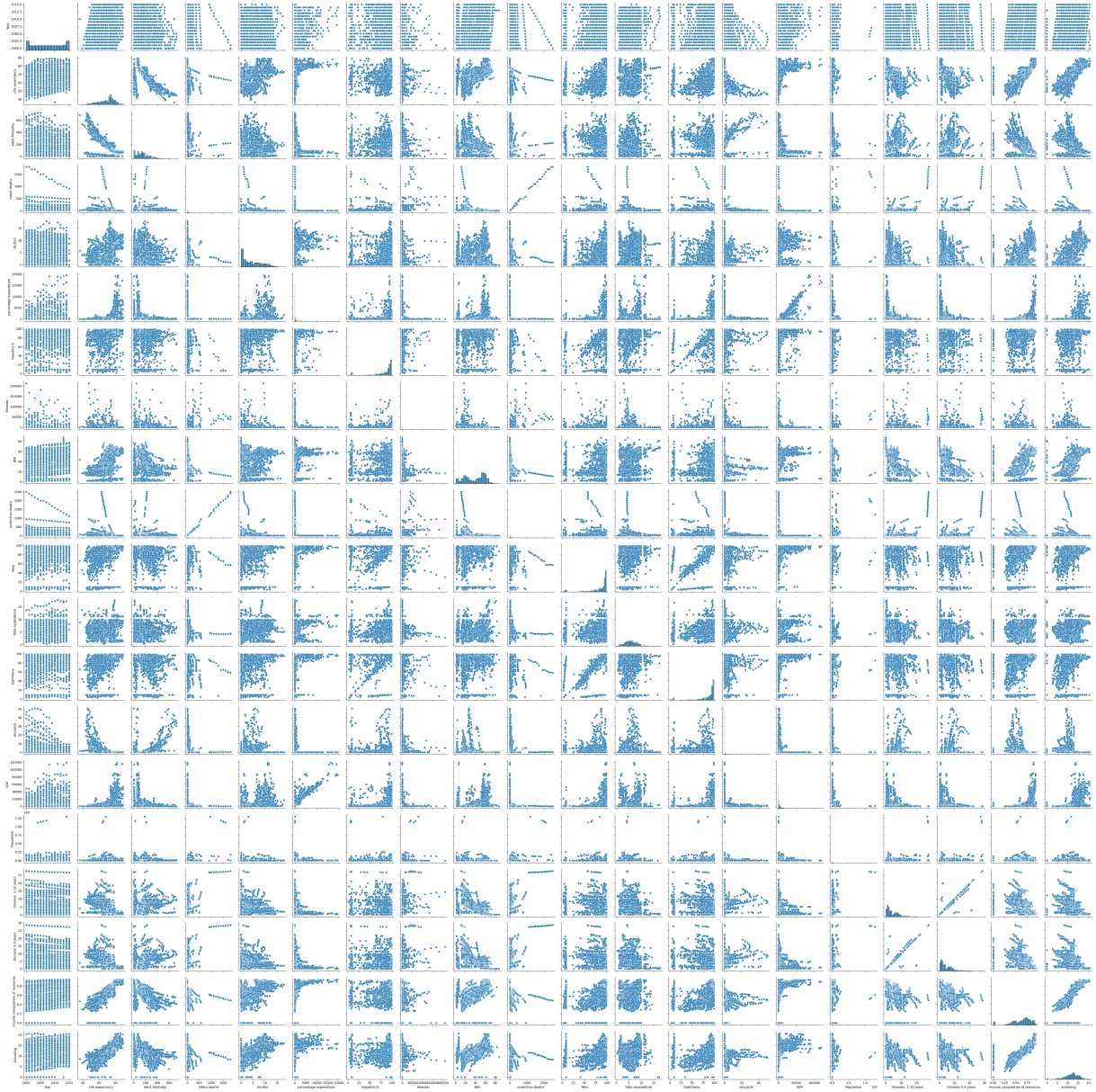
```
numeric_df = df.select_dtypes(include=[np.number])

numeric_df = numeric_df.dropna()

correlation_matrix = numeric_df.corr()

plt.figure(figsize=(12, 10))
sns.heatmap(correlation_matrix, annot=True, cmap='Reds', fmt=".2f")
plt.title('Correlation Matrix of Numeric Features')
plt.show()
```


Correlation Matrix of Numeric Features
Year 1.00 0.05 -0.04 0.01 -0.11 0.07 0.11 -0.05 0.01 0.01 -0.02 0.06 0.03 -0.12 0.10 0.01 0.02 0.01 0.12 0.09
Life expectancy 0.05 1.00 -0.70 -0.17 0.40 0.41 0.20 -0.07 0.54 -0.19 0.33 0.17 0.34 -0.59 0.44 -0.02 -0.46 -0.46 0.72 0.73
Adult Mortality -0.04 -0.70 1.00 0.04 -0.18 -0.24 -0.11 -0.00 -0.35 0.06 -0.20 -0.09 -0.19 0.55 -0.26 -0.02 0.27 0.29 -0.44 -0.42
infant deaths 0.01 -0.17 0.04 1.00 -0.11 -0.09 -0.23 0.53 -0.23 1.00 -0.16 -0.15 -0.16 0.01 -0.10 0.67 0.46 0.46 -0.13 -0.21
Alcohol -0.11 0.40 -0.18 -0.11 1.00 0.42 0.11 -0.05 0.35 -0.10 0.24 0.21 0.24 -0.03 0.44 -0.03 -0.40 -0.39 0.56 0.62
percentage expenditure 0.07 0.41 -0.24 -0.09 0.42 1.00 0.02 -0.06 0.24 -0.09 0.13 0.18 0.13 -0.10 0.96 -0.02 -0.26 -0.26 0.40 0.42
Hepatitis B 0.11 0.20 -0.11 -0.23 0.11 0.02 1.00 -0.12 0.14 -0.24 0.46 0.11 0.59 -0.09 0.04 -0.13 -0.13 -0.13 0.18 0.22
Measles -0.05 -0.07 -0.00 0.53 -0.05 -0.06 -0.12 1.00 -0.15 0.52 -0.06 -0.11 -0.06 -0.00 -0.06 0.32 0.18 0.17 -0.06 -0.12
BMI 0.01 0.54 -0.35 -0.23 0.35 0.24 0.14 -0.15 1.00 -0.24 0.19 0.19 0.18 -0.21 0.27 -0.08 -0.55 -0.55 0.51 0.55
under-five deaths 0.01 -0.19 0.06 1.00 -0.10 -0.09 -0.24 0.52 -0.24 1.00 -0.17 -0.15 -0.18 0.02 -0.10 0.66 0.46 0.46 -0.15 -0.23
Polio -0.02 0.33 -0.20 -0.16 0.24 0.13 0.46 -0.06 0.19 -0.17 1.00 0.12 0.61 -0.11 0.16 -0.05 -0.16 -0.17 0.31 0.35
Total expenditure 0.06 0.17 -0.09 -0.15 0.21 0.18 0.11 -0.11 0.19 -0.15 0.12 1.00 0.13 0.04 0.18 -0.08 -0.21 -0.22 0.18 0.24
Diphtheria 0.03 0.34 -0.19 -0.16 0.24 0.13 0.59 -0.06 0.18 -0.18 0.61 0.13 1.00 -0.12 0.16 -0.04 -0.19 -0.18 0.34 0.35
HIV/AIDS -0.12 -0.59 0.55 0.01 -0.03 -0.10 -0.09 -0.00 -0.21 0.02 -0.11 0.04 -0.12 1.00 -0.11 -0.03 0.17 0.18 -0.25 -0.21
GDP 0.10 0.44 -0.26 -0.10 0.44 0.96 0.04 -0.06 0.27 -0.10 0.16 0.18 0.16 -0.11 1.00 -0.02 -0.28 -0.28 0.45 0.47
Population 0.01 -0.02 -0.02 0.67 -0.03 -0.02 -0.13 0.32 -0.08 0.66 -0.05 -0.08 -0.04 -0.03 -0.02 1.00 0.28 0.28 -0.01 -0.04
thinness 1-19 years 0.02 -0.46 0.27 0.46 -0.40 -0.26 -0.13 0.18 -0.55 0.46 -0.16 -0.21 -0.19 0.17 -0.28 0.28 1.00 0.93 -0.45 -0.49
thinness 5-9 years 0.01 -0.46 0.29 0.46 -0.39 -0.26 -0.13 0.17 -0.55 0.46 -0.17 -0.22 -0.18 0.18 -0.28 0.28 0.93 1.00 -0.44 -0.47
Income composition of resources 0.12 0.72 -0.44 -0.13 0.56 0.40 0.18 -0.06 0.51 -0.15 0.31 0.18 0.34 -0.25 0.45 -0.01 -0.45 -0.44 1.00 0.78
Schooling 0.09 0.73 -0.42 -0.21 0.62 0.42 0.22 -0.12 0.55 -0.23 0.35 0.24 0.35 -0.21 0.47 -0.04 -0.49 -0.47 0.78 1.00
Year
Life expectancy
Adult Mortality
infant deaths
Alcohol
percentage expenditure
Hepatitis B
Measles
BMI
under-five deaths
Polio
Total expenditure
Diphtheria
HIV/AIDS
GDP
Population
thinness 1-19 years
thinness 5-9 years
Income composition of resources
Schooling
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6

Exploratory Data Analysis (EDA)

```
1 sns.pairplot(df)
```

<seaborn.axisgrid.PairGrid at 0x7ad6ddc71450>



Training a Linear Regression Moodel

X and y Arrays

```
df.columns = df.columns.str.strip()
```

```
X = df.drop('Life expectancy', axis=1)
y = df['Life expectancy']
```

Train Test Split

```
X_train, X_test, y_train, y_test = train_test_split(features, target, test_size=0.3, random_state=101)
```

```
X_train.shape
```

```
(2056, 214)
```

```
X_test.shape
```

```
(882, 214)
```

Linear Regression

```
1 model = LinearRegression()
```

```
1 model.fit(X_train, y_train)
```

▾ LinearRegression

LinearRegression()

```
1 model.coef_
```

```
array([ 2.63644186e-01, -2.40133942e-03,  1.08407317e-01, -7.83409090e-02,
        1.30738853e-04, -2.95212685e-03, -8.22343021e-06, -2.12118090e-03,
       -8.20382163e-02,  3.74677532e-03, -8.96052195e-03,  7.06028674e-03,
       -3.11921426e-01, -1.09325914e-05,  1.61644920e-10, -1.20523567e-02,
        5.69775053e-02, -6.17774814e-01,  1.54759935e-01, -9.19236963e+00,
        6.84769665e+00,  4.69103310e+00, -1.56854428e+01,  7.30302397e+00,
        6.53967147e+00,  5.16622440e+00,  2.18934862e+00,  3.22964390e+00,
        2.14940857e+00,  6.76677706e+00,  6.95652980e+00,  7.26011021e-01,
        6.36158412e+00,  2.52950580e+00,  2.34057845e+00,  1.63617795e+00,
       -9.15499070e+00, -2.90262876e+00, -6.39344610e-01,  7.84883228e+00,
       -6.96377887e+00,  4.19736970e+00,  7.60374996e+00, -5.65427026e+00,
       -9.20182925e+00, -1.11026005e+01,  4.24190169e+00, -3.50378887e+00,
       -1.05605207e+01,  1.29552172e+01, -1.59255036e+01, -1.41953143e+01,
        1.12294437e+01,  3.33281363e+00,  5.05269237e+00, -6.40929171e+00,
       -6.69514427e+00, -3.81781754e-01,  1.01604014e+01, -2.36603017e+00,
        9.32509045e+00,  1.16040157e+00, -1.48942418e+00, -1.51676393e+01,
        9.76401254e-01, -9.89303407e+00,  2.34167936e-01, -5.16308489e+00,
       -6.38134354e-01,  5.48409809e+00,  6.56050194e+00,  3.39306828e+00,
        3.39915717e+00, -9.71743039e+00, -6.20375775e+00,  6.58188318e+00,
       -4.53384527e+00,  2.09661601e-02,  1.26087171e+01,  1.35596993e+01,
       -2.73694706e+00, -7.08322193e+00,  5.38412343e+00,  2.44290770e+00,
       -6.08236038e+00,  1.30556636e+01,  5.24386382e+00,  4.09747147e+00,
       -1.01341676e+01, -1.05853184e+01, -1.41015567e+00, -6.77998127e+00,
        5.14785781e+00, -4.80169690e+00,  1.41398402e+00, -3.57898247e+00,
       -1.57804065e+00,  4.78632246e+00,  2.03167027e+00,  9.96404504e-01,
        1.26263005e+01,  3.12129376e+00,  6.56443485e+00,  3.17462693e+00,
        4.10005159e+00, -1.13831824e+00, -6.64117443e+00, -2.95050759e+00,
        5.40916443e+00,  1.01870392e+00, -5.12234561e+00, -4.75792212e+00,
        5.51563580e+00, -1.11787330e+01, -1.04491170e+01,  3.23765381e+00,
       -5.87787987e+00,  1.48602681e+00, -4.78306951e+00, -1.22212749e+01,
        5.00356828e+00,  6.93450872e+00, -1.04394447e+01,  9.98709909e-01,
        1.37746955e+00, -4.24688940e+00,  4.18629021e+00,  7.05505037e+00,
        7.29305385e-01, -6.92209177e-01, -2.28652413e+00,  6.25329573e+00,
        3.03963553e+00, -9.51724006e+00, -3.83351510e+00, -3.32208676e+00,
        1.17019408e-02, -1.95405865e+00,  8.66653552e-01,  1.99515495e+00,
        5.69891311e+00, -6.24915423e+00, -2.42975528e+00, -3.68866852e-01,
        2.39165160e+00,  6.02045198e+00, -5.64869851e+00, -3.29499283e-01,
        8.58386805e+00, -5.12031167e+00,  4.52674127e+00,  5.51513999e+00,
       -8.78902025e-01, -3.14959549e+00,  1.05871413e+00,  8.67882482e+00,
        1.24566229e+01,  3.22265575e+00, -5.33286845e+00, -1.50496598e-01,
       -7.07353054e+00,  8.24229573e-13,  6.11434659e+00,  5.82285591e+00,
        6.05944355e+00, -9.17719519e-01, -2.63321128e+00,  4.48468835e+00,
       -4.55361061e+00,  5.99894531e+00,  4.00075634e+00, -2.01447082e+01,
        2.32442180e+00, -3.77286218e+00,  8.51494966e-01, -1.10664859e-01,
       -1.28728570e+01, -3.27556478e+00, -1.01984139e+01,  2.69385000e+00,
        4.35477362e+00, -4.01099114e+00,  1.82097273e+00, -5.63850273e+00,
        3.31601326e+00,  2.70136815e+00,  3.75321132e+00, -1.55046256e+00,
        4.61516142e+00,  5.53381643e+00, -3.46372550e+00, -9.96225176e+00,
        4.04835946e+00,  3.61459631e+00,  5.36095257e+00,  5.29857155e+00,
       -3.67613801e+00, -1.89537214e-07, -7.22960068e+00,  3.01355678e+00,
```

```
       7.11759563e+00,  2.29954940e+00, -7.74457607e+00, -7.30631570e-01,
       7.76850750e+00, -5.65020584e-01,  3.38159937e+00,  6.07821567e+00,
       6.04426543e+00, -4.10937147e+00, -8.99853937e+00, -8.87538376e+00,
       5.35379703e+00, -5.35377051e+00])
```

```
1 pd.DataFrame(model.coef_, X.columns, columns=['Coedicients'])
```

	Coedicients
Year	0.263644
Adult Mortality	-0.002401
infant deaths	0.108407
Alcohol	-0.078341
percentage expenditure	0.000131
...	...
Country_Yemen	-4.109371
Country_Zambia	-8.998539
Country_Zimbabwe	-8.875384
Status_Developed	5.353797
Status_Developing	-5.353771

214 rows × 1 columns

Prediction from our Model

```
1 y_pred = model.predict(X_test)
```

Registration Evaluation Metrics

- **Mean Absolute Error** (MAE) is the mean of the absolute value of the errors:

$$\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

- **Mean Squared Error** (MSE) is the mean of the squared errors:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

```
1 MAE = metrics.mean_absolute_error(y_test, y_pred)
2 MSE = metrics.mean_squared_error(y_test, y_pred)
3 RMSE = np.sqrt(MSE)
```

```
1 MAE
```

```
    1.2139823762077666
```

```
1 MSE
```

```
    3.683311185459279
```

```
1 RMSE
```

```
    1.9191954526465715
```

```
1 df['Life expectancy'].mean()
```

```
    69.22493169398906
```

Residual Histogram

```
1 test_residual = y_test - y_pred
```