
Name: Jann Moises Nyll B. De los Reyes

Section: CPE22S3

Date: March 24, 2024

Submitted to: Engr. Roman M. Richard

Hands-on Activity 11.1 Linear Regression Analysis

Objective(s):

> This activity aims to demonstrate how to apply simple linear regression analysis to solve regression problem

Links to an external site.

Intended Learning Outcomes (ILOs):

> Demonstrate how to solve regression problems using simple linear regression

Use the linear regression model to predict the target value

Resources:

> Jupyter Notebook

Files:

> Life Expectancy Data.csv

Submission Requirements:

- PDF containing initial EDA and Data Wrangling
- PDF showing demonstration of simple linear regression.
- Submit a link to the colab file through the comment section.

📂 Import and Install Libraries

In [111…

```
!pip install hvplot
```

```
Requirement already satisfied: hvplot in /usr/local/lib/python3.10/dist-packages (0.9.2)
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) (2
024.4.0)
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->hvpl
ot) (2.1.5)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.8.2->pandas->hvplo
t) (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->hvplo
t) (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->hvplo
t) (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->hvp
lot) (2.0.7)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->panel>=0.11.0->hvp
lot) (2024.2.2)
```

In [112…]

```
# Import necesarry libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import hvplot.pandas

from sklearn.model_selection import train_test_split

from sklearn import metrics

from sklearn.linear_model import LinearRegression

%matplotlib inline
```

In [113…]

```
# Check the data
df = pd.read_csv('/content/drive/MyDrive/Module 11/Life Expectancy Data.csv')
```

In [114…]

```
df.head()
```

Out[114...

	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	...	Polio	Total expenditure	Diphth
0	Afghanistan	2015	Developing	65.0	263.0	62	0.01	71.279624	65.0	1154	...	6.0	8.16	
1	Afghanistan	2014	Developing	59.9	271.0	64	0.01	73.523582	62.0	492	...	58.0	8.18	
2	Afghanistan	2013	Developing	59.9	268.0	66	0.01	73.219243	64.0	430	...	62.0	8.13	
3	Afghanistan	2012	Developing	59.5	272.0	69	0.01	78.184215	67.0	2787	...	67.0	8.52	
4	Afghanistan	2011	Developing	59.2	275.0	71	0.01	7.097109	68.0	3013	...	68.0	7.87	

5 rows × 22 columns

In [115...
```
df.shape
```

Out[115...
```
(2938, 22)
```

In [116...
```
df.columns
```

Out[116...
```
Index(['Country', 'Year', 'Status', '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'],
      dtype='object')
```

In [117...
```
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
```

✍️Data Wrangling

In [118...
```
df.isnull().sum() #count null values in every column in dataframe
#Notice that we have additional spaces in our dataframe column, We also have null values in dataframe
```

Out[118…

```
Country                             0
Year                                0
Status                              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
```

In [119…

```
df.columns = df.columns.str.strip() #delete some spaces in our columns in dataframe using  .strip()
df.info() # check if theres spaces in our  column
```

```
<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
```

In [120…

```
df = df.rename(columns={'Adult Mortality': 'Adult_Mortality',
                        'Hepatitis B': 'Hepatitis_B',
                        'Total expenditure': 'Total_expenditure',
                        'thinness  1-19 years': 'thinness_1_to_19_yrs',
                        'thinness 5-9 years': 'thinness_5_to_9_yrs',
                        'Income composition of resources' : 'Income_composition_of_resources',
                        'Life expectancy' : 'Life_expectancy'
                       })

#Rename some columns in our dataframe
df
```

Out[120…

	Country	Year	Status	Life_expectancy	Adult_Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis_B	Measles	...	Polio	Tota
0	Afghanistan	2015	Developing	65.0	263.0	62	0.01	71.279624	65.0	1154	...	6.0	
1	Afghanistan	2014	Developing	59.9	271.0	64	0.01	73.523582	62.0	492	...	58.0	
2	Afghanistan	2013	Developing	59.9	268.0	66	0.01	73.219243	64.0	430	...	62.0	
3	Afghanistan	2012	Developing	59.5	272.0	69	0.01	78.184215	67.0	2787	...	67.0	
4	Afghanistan	2011	Developing	59.2	275.0	71	0.01	7.097109	68.0	3013	...	68.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	
2933	Zimbabwe	2004	Developing	44.3	723.0	27	4.36	0.000000	68.0	31	...	67.0	
2934	Zimbabwe	2003	Developing	44.5	715.0	26	4.06	0.000000	7.0	998	...	7.0	
2935	Zimbabwe	2002	Developing	44.8	73.0	25	4.43	0.000000	73.0	304	...	73.0	
2936	Zimbabwe	2001	Developing	45.3	686.0	25	1.72	0.000000	76.0	529	...	76.0	
2937	Zimbabwe	2000	Developing	46.0	665.0	24	1.68	0.000000	79.0	1483	...	78.0	

2938 rows × 22 columns

In [121…

```
#Fill the missing values with mean in Adult_Mortality
df['Adult_Mortality'].fillna(np.mean(df['Adult_Mortality']),inplace=True)
#Fill the missing values with mean in Alcohol
df['Alcohol'].fillna(np.mean(df['Alcohol']),inplace=True)
#Fill the missing values with mean in Hepatitis_B
df['Hepatitis_B'].fillna(np.mean(df['Hepatitis_B']),inplace=True)
#Fill the missing values with mean in BMI
df['BMI'].fillna(np.mean(df['BMI']),inplace=True)
#Fill the missing values with mean in Polio
df['Polio'].fillna(np.mean(df['Polio']),inplace=True)
#Fill the missing values with mean in Total_expenditure
df['Total_expenditure'].fillna(np.mean(df['Total_expenditure']),inplace=True)
#Fill the missing values with mean in Diphtheria
df['Diphtheria'].fillna(np.mean(df['Diphtheria']),inplace=True)
#Fill the missing values with mean in GDP
df['GDP'].fillna(np.mean(df['GDP']),inplace=True)
#Fill the missing values with mean in Population
df['Population'].fillna(np.mean(df['Population']),inplace=True)
#Fill the missing values with mean in thinness_1_to_19_yrs
df['thinness_1_to_19_yrs'].fillna(np.mean(df['thinness_1_to_19_yrs']),inplace=True)
#Fill the missing values with mean in thinness_5_to_9_yrs
df['thinness_5_to_9_yrs'].fillna(np.mean(df['thinness_5_to_9_yrs']),inplace=True)
#Fill the missing values with mean in Income_composition_of_resources
df['Income_composition_of_resources'].fillna(np.mean(df['Income_composition_of_resources']),inplace=True)
#Fill the missing values with mean in Schooling
df['Schooling'].fillna(np.mean(df['Schooling']),inplace=True)
#Fill the missing values with mean in Life_expectancy
df['Life_expectancy'].fillna(np.mean(df['Life_expectancy']),inplace=True)
```

In [ ]:

```
df['Status'].unique()
```

Out[ ]:

```
array(['Developing', 'Developed'], dtype=object)
```

In [ ]:

```
df['Status'] = df['Status'].replace({'Developing':0,'Developed':1}) #Change the multiple values of  a column using .replace()
```

In [ ]:

```
df
```

Out[ ]:

	Country	Year	Status	Life_expectancy	Adult_Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis_B	Measles	...	Polio	Total_exp
0	Afghanistan	2015	0	65.0	263.0	62	0.01	71.279624	65.0	1154	...	6.0	
1	Afghanistan	2014	0	59.9	271.0	64	0.01	73.523582	62.0	492	...	58.0	
2	Afghanistan	2013	0	59.9	268.0	66	0.01	73.219243	64.0	430	...	62.0	
3	Afghanistan	2012	0	59.5	272.0	69	0.01	78.184215	67.0	2787	...	67.0	
4	Afghanistan	2011	0	59.2	275.0	71	0.01	7.097109	68.0	3013	...	68.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	
2933	Zimbabwe	2004	0	44.3	723.0	27	4.36	0.000000	68.0	31	...	67.0	
2934	Zimbabwe	2003	0	44.5	715.0	26	4.06	0.000000	7.0	998	...	7.0	
2935	Zimbabwe	2002	0	44.8	73.0	25	4.43	0.000000	73.0	304	...	73.0	
2936	Zimbabwe	2001	0	45.3	686.0	25	1.72	0.000000	76.0	529	...	76.0	
2937	Zimbabwe	2000	0	46.0	665.0	24	1.68	0.000000	79.0	1483	...	78.0	

2938 rows × 22 columns

In [122…
```
df['Status'].value_counts() #check the Status
```

Out[122…
```
Status
Developing    2426
Developed      512
Name: count, dtype: int64
```

In [123…
```
country = list(df['Country'].unique())
country
```

Out[123…

```
['Afghanistan',
 'Albania',
 'Algeria',
 'Angola',
 'Antigua and Barbuda',
 'Argentina',
 'Armenia',
 'Australia',
 'Austria',
 'Azerbaijan',
 'Bahamas',
 'Bahrain',
 'Bangladesh',
 'Barbados',
 'Belarus',
 'Belgium',
 'Belize',
 'Benin',
 'Bhutan',
 'Bolivia (Plurinational State of)',
 'Bosnia and Herzegovina',
 'Botswana',
 'Brazil',
 'Brunei Darussalam',
 'Bulgaria',
 'Burkina Faso',
 'Burundi',
 "Côte d'Ivoire",
 'Cabo Verde',
 'Cambodia',
 'Cameroon',
 'Canada',
 'Central African Republic',
 'Chad',
 'Chile',
 'China',
 'Colombia',
 'Comoros',
 'Congo',
 'Cook Islands',
 'Costa Rica',
 'Croatia',
 'Cuba',
 'Cyprus',
 'Czechia',
 "Democratic People's Republic of Korea",
 'Democratic Republic of the Congo',
 'Denmark',
 'Djibouti',
 'Dominica',
 'Dominican Republic',
 'Ecuador',
 'Egypt',
 'El Salvador',
 'Equatorial Guinea',
 'Eritrea',
 'Estonia',
 'Ethiopia',
 'Fiji',
 'Finland',
 'France',
 'Gabon',
 'Gambia',
 'Georgia',
 'Germany',
 'Ghana',
 'Greece',
 'Grenada',
 'Guatemala',
 'Guinea',
 'Guinea-Bissau',
 'Guyana',
 'Haiti',
 'Honduras',
 'Hungary',
 'Iceland',
 'India',
 'Indonesia',
 'Iran (Islamic Republic of)',
 'Iraq',
```

```
'Ireland',
'Israel',
'Italy',
'Jamaica',
'Japan',
'Jordan',
'Kazakhstan',
'Kenya',
'Kiribati',
'Kuwait',
'Kyrgyzstan',
"Lao People's Democratic Republic",
'Latvia',
'Lebanon',
'Lesotho',
'Liberia',
'Libya',
'Lithuania',
'Luxembourg',
'Madagascar',
'Malawi',
'Malaysia',
'Maldives',
'Mali',
'Malta',
'Marshall Islands',
'Mauritania',
'Mauritius',
'Mexico',
'Micronesia (Federated States of)',
'Monaco',
'Mongolia',
'Montenegro',
'Morocco',
'Mozambique',
'Myanmar',
'Namibia',
'Nauru',
'Nepal',
'Netherlands',
'New Zealand',
'Nicaragua',
'Niger',
'Nigeria',
'Niue',
'Norway',
'Oman',
'Pakistan',
'Palau',
'Panama',
'Papua New Guinea',
'Paraguay',
'Peru',
'Philippines',
'Poland',
'Portugal',
'Qatar',
'Republic of Korea',
'Republic of Moldova',
'Romania',
'Russian Federation',
'Rwanda',
'Saint Kitts and Nevis',
'Saint Lucia',
'Saint Vincent and the Grenadines',
'Samoa',
'San Marino',
'Sao Tome and Principe',
'Saudi Arabia',
'Senegal',
'Serbia',
'Seychelles',
'Sierra Leone',
'Singapore',
'Slovakia',
'Slovenia',
'Solomon Islands',
'Somalia',
'South Africa',
'South Sudan',
```

```
 'Spain',
 'Sri Lanka',
 'Sudan',
 'Suriname',
 'Swaziland',
 'Sweden',
 'Switzerland',
 'Syrian Arab Republic',
 'Tajikistan',
 'Thailand',
 'The former Yugoslav republic of Macedonia',
 'Timor-Leste',
 'Togo',
 'Tonga',
 'Trinidad and Tobago',
 'Tunisia',
 'Turkey',
 'Turkmenistan',
 'Tuvalu',
 'Uganda',
 'Ukraine',
 'United Arab Emirates',
 'United Kingdom of Great Britain and Northern Ireland',
 'United Republic of Tanzania',
 'United States of America',
 'Uruguay',
 'Uzbekistan',
 'Vanuatu',
 'Venezuela (Bolivarian Republic of)',
 'Viet Nam',
 'Yemen',
 'Zambia',
 'Zimbabwe']
```

In [125…
```
country_dict = {country[i]: i for i in range(len(country))} #make a dictionary where  key = country and value  = number
print(country_dict)
```

```
{'Afghanistan': 0, 'Albania': 1, 'Algeria': 2, 'Angola': 3, 'Antigua and Barbuda': 4, 'Argentina': 5, 'Armenia': 6, 'Australi
a': 7, 'Austria': 8, 'Azerbaijan': 9, 'Bahamas': 10, 'Bahrain': 11, 'Bangladesh': 12, 'Barbados': 13, 'Belarus': 14, 'Belgium':
15, 'Belize': 16, 'Benin': 17, 'Bhutan': 18, 'Bolivia (Plurinational State of)': 19, 'Bosnia and Herzegovina': 20, 'Botswana':
21, 'Brazil': 22, 'Brunei Darussalam': 23, 'Bulgaria': 24, 'Burkina Faso': 25, 'Burundi': 26, "Côte d'Ivoire": 27, 'Cabo Verd
e': 28, 'Cambodia': 29, 'Cameroon': 30, 'Canada': 31, 'Central African Republic': 32, 'Chad': 33, 'Chile': 34, 'China': 35, 'Co
lombia': 36, 'Comoros': 37, 'Congo': 38, 'Cook Islands': 39, 'Costa Rica': 40, 'Croatia': 41, 'Cuba': 42, 'Cyprus': 43, 'Czechi
a': 44, "Democratic People's Republic of Korea": 45, 'Democratic Republic of the Congo': 46, 'Denmark': 47, 'Djibouti': 48, 'Do
minica': 49, 'Dominican Republic': 50, 'Ecuador': 51, 'Egypt': 52, 'El Salvador': 53, 'Equatorial Guinea': 54, 'Eritrea': 55,
'Estonia': 56, 'Ethiopia': 57, 'Fiji': 58, 'Finland': 59, 'France': 60, 'Gabon': 61, 'Gambia': 62, 'Georgia': 63, 'Germany': 6
4, 'Ghana': 65, 'Greece': 66, 'Grenada': 67, 'Guatemala': 68, 'Guinea': 69, 'Guinea-Bissau': 70, 'Guyana': 71, 'Haiti': 72, 'Ho
nduras': 73, 'Hungary': 74, 'Iceland': 75, 'India': 76, 'Indonesia': 77, 'Iran (Islamic Republic of)': 78, 'Iraq': 79, 'Irelan
d': 80, 'Israel': 81, 'Italy': 82, 'Jamaica': 83, 'Japan': 84, 'Jordan': 85, 'Kazakhstan': 86, 'Kenya': 87, 'Kiribati': 88, 'Ku
wait': 89, 'Kyrgyzstan': 90, "Lao People's Democratic Republic": 91, 'Latvia': 92, 'Lebanon': 93, 'Lesotho': 94, 'Liberia': 95,
'Libya': 96, 'Lithuania': 97, 'Luxembourg': 98, 'Madagascar': 99, 'Malawi': 100, 'Malaysia': 101, 'Maldives': 102, 'Mali': 103,
'Malta': 104, 'Marshall Islands': 105, 'Mauritania': 106, 'Mauritius': 107, 'Mexico': 108, 'Micronesia (Federated States of)':
109, 'Monaco': 110, 'Mongolia': 111, 'Montenegro': 112, 'Morocco': 113, 'Mozambique': 114, 'Myanmar': 115, 'Namibia': 116, 'Nau
ru': 117, 'Nepal': 118, 'Netherlands': 119, 'New Zealand': 120, 'Nicaragua': 121, 'Niger': 122, 'Nigeria': 123, 'Niue': 124, 'N
orway': 125, 'Oman': 126, 'Pakistan': 127, 'Palau': 128, 'Panama': 129, 'Papua New Guinea': 130, 'Paraguay': 131, 'Peru': 132,
'Philippines': 133, 'Poland': 134, 'Portugal': 135, 'Qatar': 136, 'Republic of Korea': 137, 'Republic of Moldova': 138, 'Romani
a': 139, 'Russian Federation': 140, 'Rwanda': 141, 'Saint Kitts and Nevis': 142, 'Saint Lucia': 143, 'Saint Vincent and the Gre
nadines': 144, 'Samoa': 145, 'San Marino': 146, 'Sao Tome and Principe': 147, 'Saudi Arabia': 148, 'Senegal': 149, 'Serbia': 15
0, 'Seychelles': 151, 'Sierra Leone': 152, 'Singapore': 153, 'Slovakia': 154, 'Slovenia': 155, 'Solomon Islands': 156, 'Somali
a': 157, 'South Africa': 158, 'South Sudan': 159, 'Spain': 160, 'Sri Lanka': 161, 'Sudan': 162, 'Suriname': 163, 'Swaziland': 1
64, 'Sweden': 165, 'Switzerland': 166, 'Syrian Arab Republic': 167, 'Tajikistan': 168, 'Thailand': 169, 'The former Yugoslav re
public of Macedonia': 170, 'Timor-Leste': 171, 'Togo': 172, 'Tonga': 173, 'Trinidad and Tobago': 174, 'Tunisia': 175, 'Turkey':
176, 'Turkmenistan': 177, 'Tuvalu': 178, 'Uganda': 179, 'Ukraine': 180, 'United Arab Emirates': 181, 'United Kingdom of Great B
ritain and Northern Ireland': 182, 'United Republic of Tanzania': 183, 'United States of America': 184, 'Uruguay': 185, 'Uzbeki
stan': 186, 'Vanuatu': 187, 'Venezuela (Bolivarian Republic of)': 188, 'Viet Nam': 189, 'Yemen': 190, 'Zambia': 191, 'Zimbabw
e': 192}
```

In [126…
```
df['Country'] = df['Country'].replace(country_dict) #replace object  type to int
```

In [127…
```
df
```

Out[127…

	Country	Year	Status	Life_expectancy	Adult_Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis_B	Measles	...	Polio	Total_ex
0	0	2015	Developing	65.0	263.0	62	0.01	71.279624	65.0	1154	...	6.0	
1	0	2014	Developing	59.9	271.0	64	0.01	73.523582	62.0	492	...	58.0	
2	0	2013	Developing	59.9	268.0	66	0.01	73.219243	64.0	430	...	62.0	
3	0	2012	Developing	59.5	272.0	69	0.01	78.184215	67.0	2787	...	67.0	
4	0	2011	Developing	59.2	275.0	71	0.01	7.097109	68.0	3013	...	68.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	
2933	192	2004	Developing	44.3	723.0	27	4.36	0.000000	68.0	31	...	67.0	
2934	192	2003	Developing	44.5	715.0	26	4.06	0.000000	7.0	998	...	7.0	
2935	192	2002	Developing	44.8	73.0	25	4.43	0.000000	73.0	304	...	73.0	
2936	192	2001	Developing	45.3	686.0	25	1.72	0.000000	76.0	529	...	76.0	
2937	192	2000	Developing	46.0	665.0	24	1.68	0.000000	79.0	1483	...	78.0	

2938 rows × 22 columns

📊 Explanatory Data Analysis

In [108…

```
df.describe().T
```

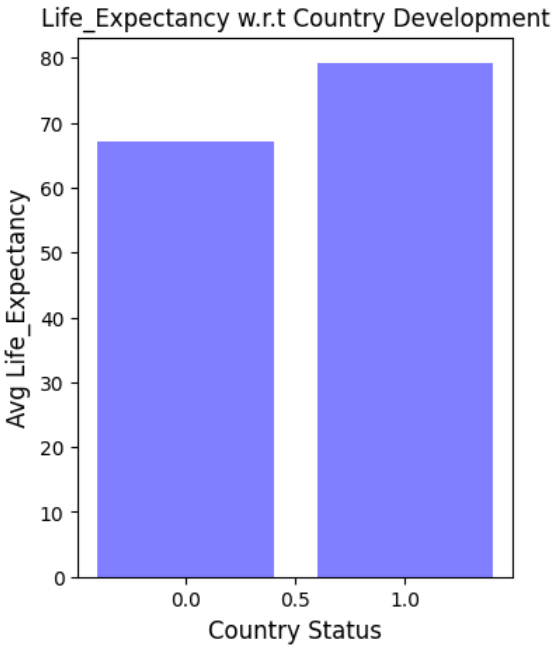
Out[108…

	count	mean	std	min	25%	50%	75%	ma
Country	2938.0	9.509122e+01	5.625004e+01	0.00000	46.000000	9.300000e+01	1.450000e+02	1.920000e+0
Year	2938.0	2.007519e+03	4.613841e+00	2000.00000	2004.000000	2.008000e+03	2.012000e+03	2.015000e+0
Status	2938.0	1.742682e-01	3.794045e-01	0.00000	0.000000	0.000000e+00	0.000000e+00	1.000000e+0
Life_expectancy	2938.0	6.922493e+01	9.507640e+00	36.30000	63.200000	7.200000e+01	7.560000e+01	8.900000e+0
Adult_Mortality	2938.0	1.647964e+02	1.240803e+02	1.00000	74.000000	1.440000e+02	2.270000e+02	7.230000e+0
infant deaths	2938.0	3.030395e+01	1.179265e+02	0.00000	0.000000	3.000000e+00	2.200000e+01	1.800000e+0
Alcohol	2938.0	4.602861e+00	3.916288e+00	0.01000	1.092500	4.160000e+00	7.390000e+00	1.787000e+0
percentage expenditure	2938.0	7.382513e+02	1.987915e+03	0.00000	4.685343	6.491291e+01	4.415341e+02	1.947991e+0
Hepatitis_B	2938.0	8.094046e+01	2.258685e+01	1.00000	80.940461	8.700000e+01	9.600000e+01	9.900000e+0
Measles	2938.0	2.419592e+03	1.146727e+04	0.00000	0.000000	1.700000e+01	3.602500e+02	2.121830e+0
BMI	2938.0	3.832125e+01	1.992768e+01	1.00000	19.400000	4.300000e+01	5.610000e+01	8.730000e+0
under-five deaths	2938.0	4.203574e+01	1.604455e+02	0.00000	0.000000	4.000000e+00	2.800000e+01	2.500000e+0
Polio	2938.0	8.255019e+01	2.335214e+01	3.00000	78.000000	9.300000e+01	9.700000e+01	9.900000e+0
Total_expenditure	2938.0	5.938190e+00	2.400274e+00	0.37000	4.370000	5.938190e+00	7.330000e+00	1.760000e+0
Diphtheria	2938.0	8.232408e+01	2.364007e+01	2.00000	78.000000	9.300000e+01	9.700000e+01	9.900000e+0
HIV/AIDS	2938.0	1.742103e+00	5.077785e+00	0.10000	0.100000	1.000000e-01	8.000000e-01	5.060000e+0
GDP	2938.0	7.483158e+03	1.313680e+04	1.68135	580.486996	3.116562e+03	7.483158e+03	1.191727e+0
Population	2938.0	1.275338e+07	5.381546e+07	34.00000	418917.250000	3.675929e+06	1.275338e+07	1.293859e+0
thinness_1_to_19_yrs	2938.0	4.839704e+00	4.394535e+00	0.10000	1.600000	3.400000e+00	7.100000e+00	2.770000e+0
thinness_5_to_9_yrs	2938.0	4.870317e+00	4.482708e+00	0.10000	1.600000	3.400000e+00	7.200000e+00	2.860000e+0
Income_composition_of_resources	2938.0	6.275511e-01	2.048197e-01	0.00000	0.504250	6.620000e-01	7.720000e-01	9.480000e-0
Schooling	2938.0	1.199279e+01	3.264381e+00	0.00000	10.300000	1.210000e+01	1.410000e+01	2.070000e+0

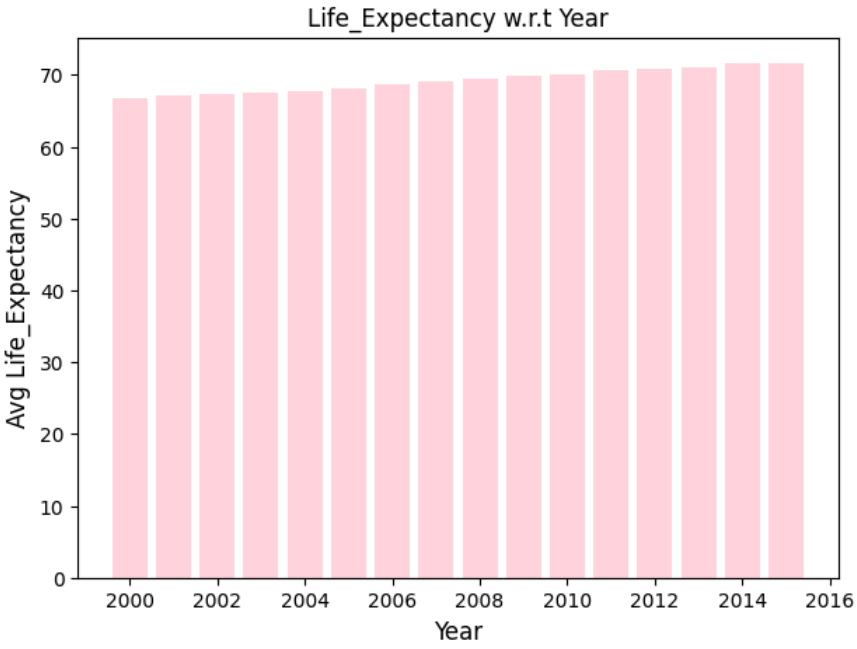
In [105…

```
plt.figure(figsize=(4,5))
plt.bar(df.groupby('Status')['Status'].count().index,df.groupby('Status')['Life_expectancy'].mean(),color='blue',alpha=0.50)
plt.xlabel("Country Status",fontsize=12)
plt.ylabel("Avg Life_Expectancy",fontsize=12)
```

```
plt.title("Life_Expectancy w.r.t Country Development")
plt.show()
```



In [107…
```
plt.figure(figsize=(7,5))
plt.bar(df.groupby('Year')['Year'].count().index,df.groupby('Year')['Life_expectancy'].mean(),color='pink',alpha=0.65)
plt.xlabel("Year",fontsize=12)
plt.ylabel("Avg Life_Expectancy",fontsize=12)
plt.title("Life_Expectancy w.r.t Year")
plt.show()
```



In [ ]:
```
df.corr()
```

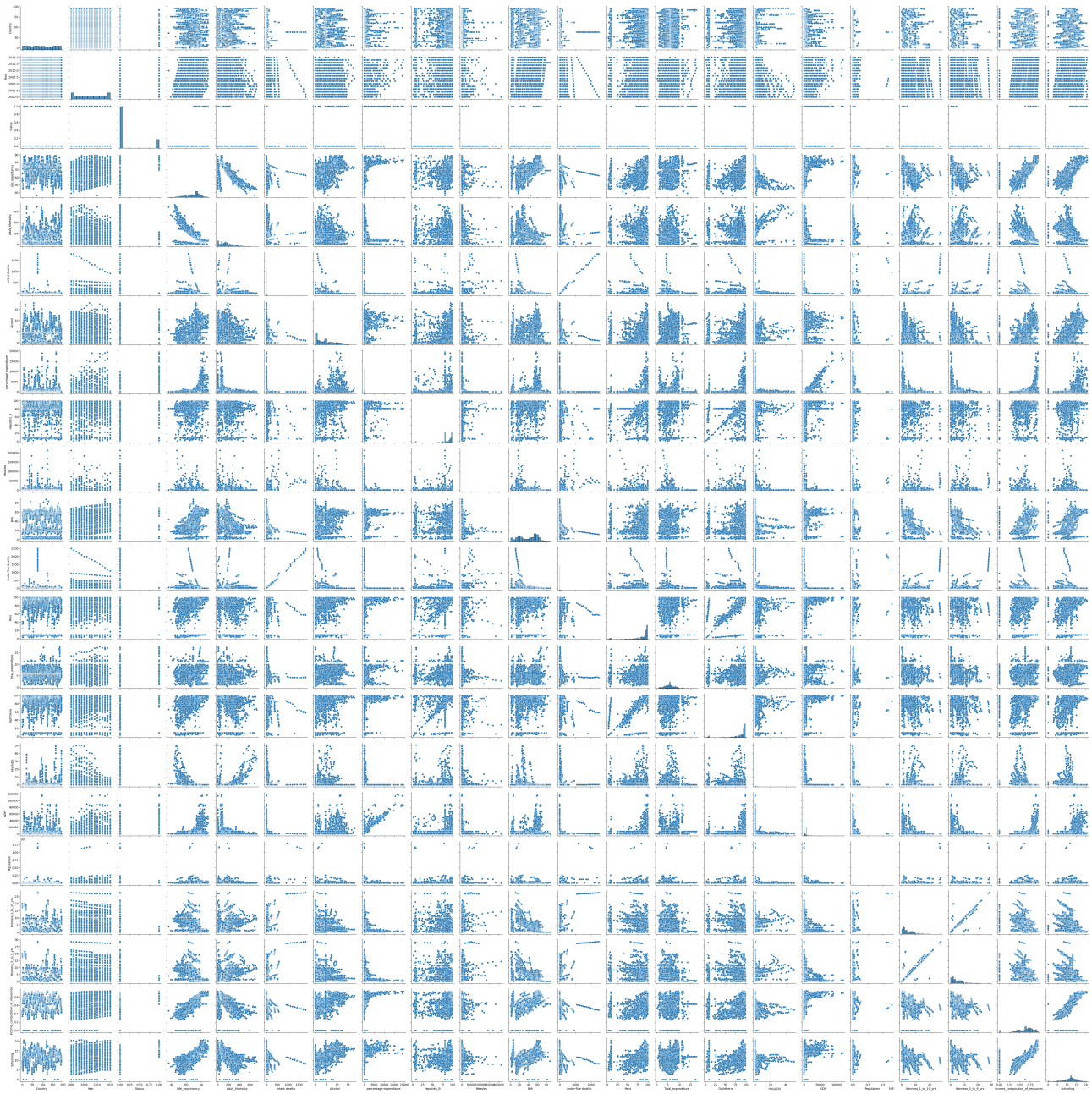
Out[ ]:

	Country	Year	Status	Life_expectancy	Adult_Mortality	infant deaths	Alcohol	percentage expenditure	Hep
Country	1.000000	0.001350	0.032439	-0.013475	0.036221	-0.030985	-0.059199	-0.032465	-0
Year	0.001350	1.000000	-0.001864	0.169623	-0.078861	-0.037415	-0.048168	0.031400	0
Status	0.032439	-0.001864	1.000000	0.481962	-0.315171	-0.112252	0.579371	0.454261	0
Life_expectancy	-0.013475	0.169623	0.481962	1.000000	-0.696359	-0.196535	0.391598	0.381791	0
Adult_Mortality	0.036221	-0.078861	-0.315171	-0.696359	1.000000	0.078747	-0.190408	-0.242814	-0
infant deaths	-0.030985	-0.037415	-0.112252	-0.196535	0.078747	1.000000	-0.113812	-0.085612	-0
Alcohol	-0.059199	-0.048168	0.579371	0.391598	-0.190408	-0.113812	1.000000	0.339634	0
percentage expenditure	-0.032465	0.031400	0.454261	0.381791	-0.242814	-0.085612	0.339634	1.000000	0
Hepatitis_B	-0.018031	0.089398	0.095642	0.203771	-0.138591	-0.178783	0.075447	0.011679	1
Measles	-0.024164	-0.082493	-0.076955	-0.157574	0.031174	0.501128	-0.051055	-0.056596	-0
BMI	0.018960	0.108327	0.310873	0.559255	-0.381449	-0.227220	0.318070	0.228537	0
under-five deaths	-0.027014	-0.042937	-0.115195	-0.222503	0.094135	0.996629	-0.110777	-0.087852	-0
Polio	0.018558	0.093820	0.220098	0.461574	-0.272694	-0.170674	0.213744	0.147203	0
Total_expenditure	0.053273	0.081860	0.289985	0.207981	-0.110875	-0.126564	0.294898	0.173414	0
Diphtheria	-0.005541	0.133853	0.216763	0.475418	-0.273014	-0.175156	0.215242	0.143570	0
HIV/AIDS	0.089229	-0.139741	-0.148590	-0.556457	0.523727	0.025231	-0.048650	-0.097857	-0
GDP	-0.015411	0.093351	0.445911	0.430493	-0.277053	-0.107109	0.318591	0.888140	0
Population	-0.014489	0.014951	-0.041091	-0.019638	-0.012501	0.548522	-0.030765	-0.024648	-0
thinness_1_to_19_yrs	0.007174	-0.047592	-0.367934	-0.472162	0.299863	0.465590	-0.416946	-0.251190	-0
thinness_5_to_9_yrs	0.021713	-0.050627	-0.366297	-0.466629	0.305366	0.471228	-0.405881	-0.252725	-0
Income_composition_of_resources	-0.023711	0.236333	0.457302	0.692483	-0.440062	-0.143663	0.416099	0.380374	0
Schooling	-0.025427	0.203471	0.491444	0.715066	-0.435108	-0.191757	0.497546	0.388105	0

22 rows × 22 columns

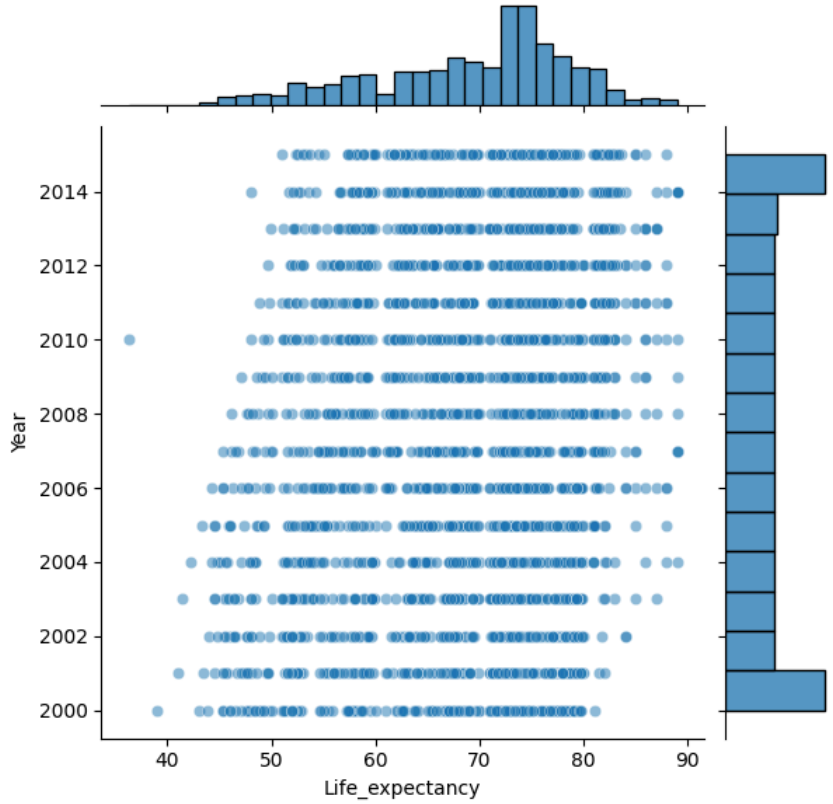
In [101…
```
sns.pairplot(df)
```

Out[101…
```
<seaborn.axisgrid.PairGrid at 0x7dadd8388100>
```



In [ ]:
```
sns.jointplot(x='Life_expectancy', y='Year', data=df, alpha=0.5)
```

Out[ ]:
```
<seaborn.axisgrid.JointGrid at 0x7dae0c87b760>
```

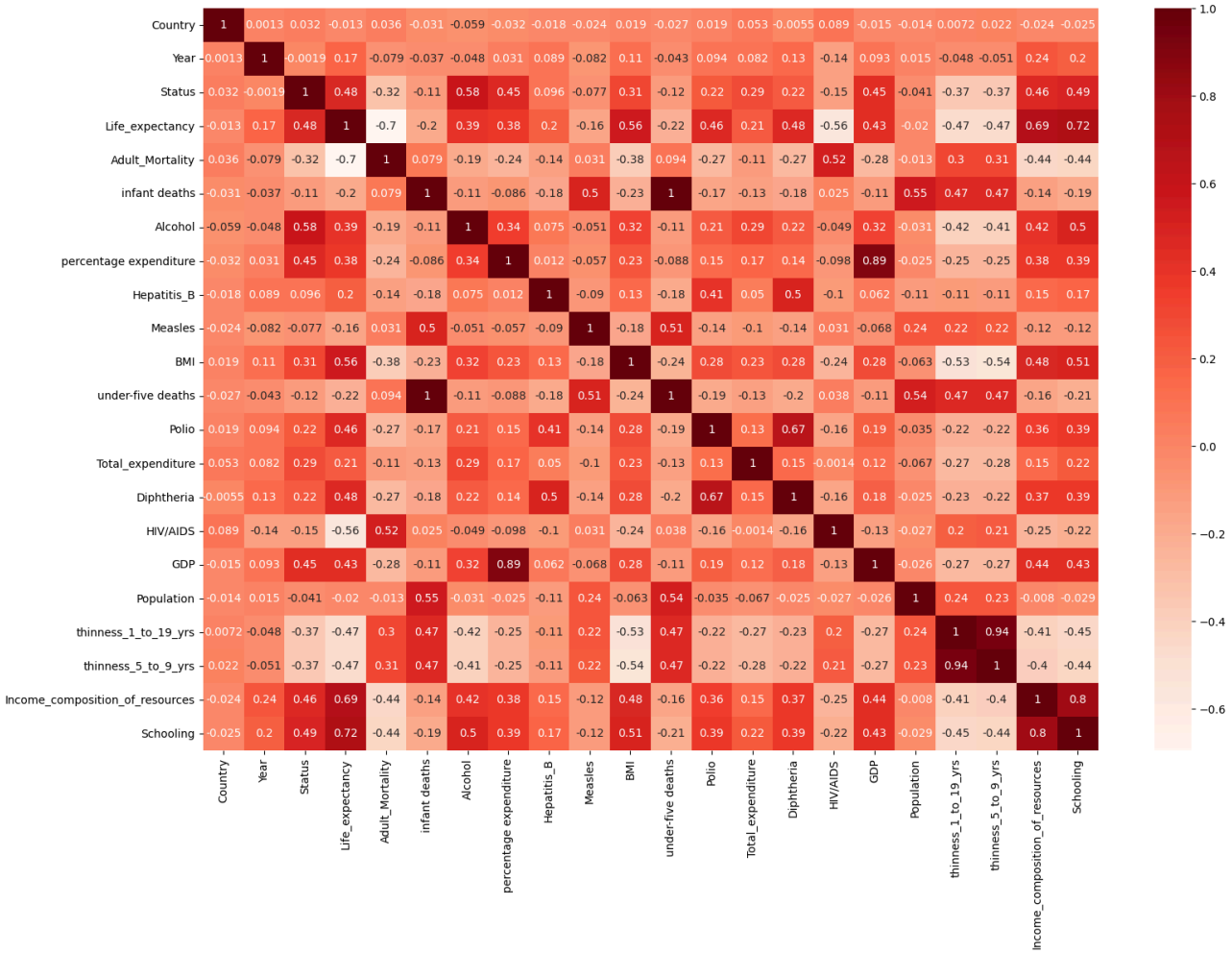


In [100…

```
plt.figure(figsize=(18,12))
sns.heatmap(df.corr(),annot = True, cmap ='Reds')
```

Out[100…

```
<Axes: >
```



Training Linear Regression Model

In [78]:
```
X = df[['Country', 'Year', 'Status', 'Adult_Mortality',
        'Alcohol', 'percentage expenditure', 'Hepatitis_B',
       'Measles', 'BMI', 'Polio', 'Total_expenditure',
       'Diphtheria', 'HIV/AIDS', 'GDP', 'Population', 'thinness_1_to_19_yrs',
       'thinness_5_to_9_yrs', 'Income_composition_of_resources', 'Schooling']]
y= df['Life_expectancy']
```

In [80]:
```
print("X = ",X.shape,"\ny = ",y.shape)
```

```
X =  (2938, 19)
y =  (2938,)
```

In [81]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.3, random_state = 101)
```

In [82]:
```
X_train.shape
```

Out[82]: (2056, 19)

In [83]:
```
X_test.shape
```

Out[83]: (882, 19)

✅ Linear Regression

In [84]:
```
model = LinearRegression()
```

In [85]:
```
model.fit(X_train, y_train)
```

Out[85]:
```
▾ LinearRegression

LinearRegression()
```

In [86]:
```
model.coef_
```

Out[86]:
```
array([ 3.50760662e-03, -1.90935935e-02,  1.09507440e+00, -1.99874815e-02,
        3.84381053e-02,  4.05850770e-05, -1.95725161e-02, -4.02448083e-05,
        4.52961438e-02,  3.50562822e-02,  7.08779811e-02,  4.62104564e-02,
       -4.63249310e-01,  4.59063107e-05,  1.74019171e-09, -1.26224905e-01,
        4.90149338e-02,  5.92039227e+00,  7.33717499e-01])
```

In [87]:
```
pd.DataFrame(model.coef_, X.columns, columns = ['Coefficients'])
```

Out[87]:

	Coefficients
Country	3.507607e-03
Year	-1.909359e-02
Status	1.095074e+00
Adult_Mortality	-1.998748e-02
Alcohol	3.843811e-02
percentage expenditure	4.058508e-05
Hepatitis_B	-1.957252e-02
Measles	-4.024481e-05
BMI	4.529614e-02
Polio	3.505628e-02
Total_expenditure	7.087798e-02
Diphtheria	4.621046e-02
HIV/AIDS	-4.632493e-01
GDP	4.590631e-05
Population	1.740192e-09
thinness_1_to_19_yrs	-1.262249e-01
thinness_5_to_9_yrs	4.901493e-02
Income_composition_of_resources	5.920392e+00
Schooling	7.337175e-01

In [88]:
```
y_pred = model.predict(X_test)
```

In [89]:
```
MAE = metrics.mean_absolute_error(y_test, y_pred)
MSE = metrics.mean_squared_error(y_test, y_pred)
RMSE = np.sqrt(MSE)
```

In [90]:
```
MAE
```

Out[90]:
```
3.0368505447749077
```

In [91]:
```
MSE
```

Out[91]:
```
15.94170767049655
```

In [92]:
```
RMSE
```

Out[92]:
```
3.9927068099844933
```

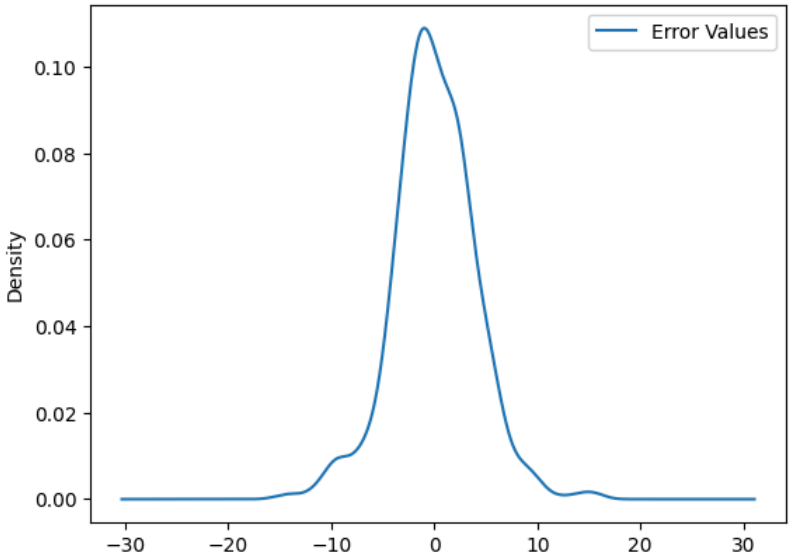
In [93]:
```
df['Life_expectancy'].mean()
```

Out[93]:
```
69.22493169398906
```

In [94]:
```
test_residual = y_test - y_pred
```

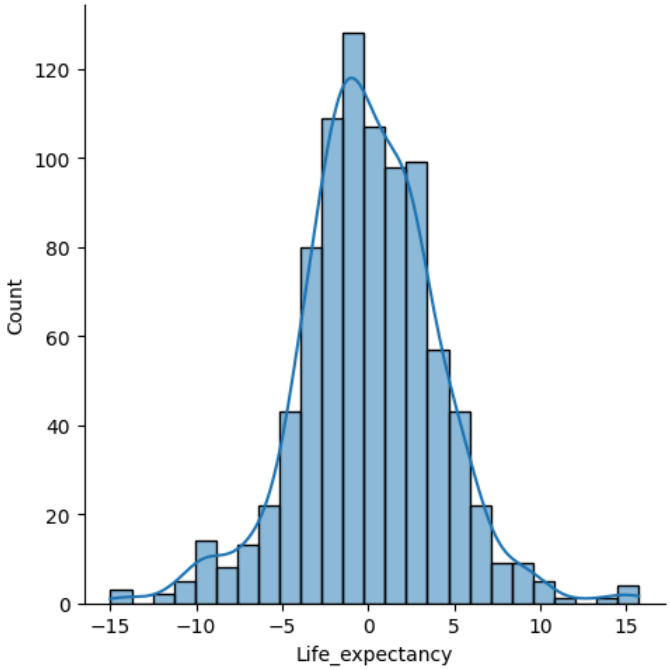
In [96]:
```
pd.DataFrame({'Error Values':(test_residual)}).plot.kde()
```

Out[96]: <Axes: ylabel='Density'>



In [97]:
```
sns.displot(test_residual, bins = 25 , kde =True)
```

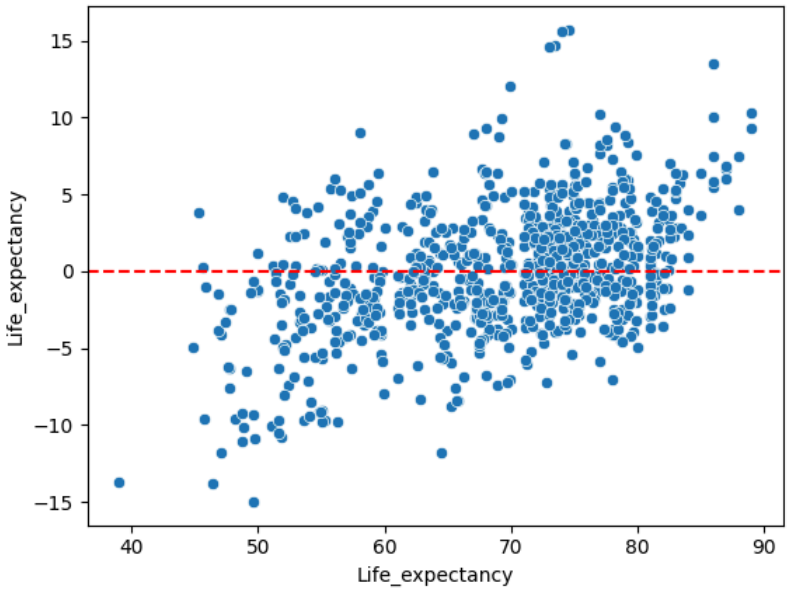
Out[97]: <seaborn.axisgrid.FacetGrid at 0x7dadee8839d0>



In [98]:
```
sns.scatterplot(x=y_test,y =test_residual)

plt.axhline(y=0,color ='r', ls ='--')
```

Out[98]: <matplotlib.lines.Line2D at 0x7dadd8b4fa00>


15
10
5
0
−5
−10
−15
Life_expectancy
40
50
60
70
80
90
Life_expectancy