egression-analysis-lifeexpectancy

April 28, 2024

1 Hands-on Activity 11.1 Linear Regression Analysis

Course: CPE 311	Program: BSCpE
Course Title: Computational Thinking with Python	**Date Performed:** April 27 , 2024
Section: BSCPE22S3	**Date Submitted:** April 28, 2024
Student Name: John Louie V. Adornado	**Instructor's Name:** Engr. Roman Richard

[40]:
```
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) (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)
```

[41]:
```
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
```

[42]:
```
df = pd.read_csv('Life Expectancy Data.csv')
df
```

[42]:

	Country	Year	Status	Life expectancy	Adult Mortality	\
0	Afghanistan	2015	Developing	65.0	263.0	
1	Afghanistan	2014	Developing	59.9	271.0	
2	Afghanistan	2013	Developing	59.9	268.0	
3	Afghanistan	2012	Developing	59.5	272.0	
4	Afghanistan	2011	Developing	59.2	275.0	
…	…	…	…	…	…	
2933	Zimbabwe	2004	Developing	44.3	723.0	
2934	Zimbabwe	2003	Developing	44.5	715.0	
2935	Zimbabwe	2002	Developing	44.8	73.0	
2936	Zimbabwe	2001	Developing	45.3	686.0	
2937	Zimbabwe	2000	Developing	46.0	665.0	

	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	\
0	62	0.01	71.279624	65.0	1154	
1	64	0.01	73.523582	62.0	492	
2	66	0.01	73.219243	64.0	430	
3	69	0.01	78.184215	67.0	2787	
4	71	0.01	7.097109	68.0	3013	
…	…	…	…	…	…	
2933	27	4.36	0.000000	68.0	31	
2934	26	4.06	0.000000	7.0	998	
2935	25	4.43	0.000000	73.0	304	
2936	25	1.72	0.000000	76.0	529	
2937	24	1.68	0.000000	79.0	1483	

	…	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  547.358878

      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 22 columns]
```

[43]:
```
df.head(20)
```

[43]:
```
        Country  Year      Status  Life expectancy   Adult Mortality  \
0   Afghanistan  2015  Developing              65.0            263.0
1   Afghanistan  2014  Developing              59.9            271.0
2   Afghanistan  2013  Developing              59.9            268.0
3   Afghanistan  2012  Developing              59.5            272.0
4   Afghanistan  2011  Developing              59.2            275.0
5   Afghanistan  2010  Developing              58.8            279.0
6   Afghanistan  2009  Developing              58.6            281.0
7   Afghanistan  2008  Developing              58.1            287.0
```

```
8   Afghanistan  2007  Developing                57.5            295.0
9   Afghanistan  2006  Developing                57.3            295.0
10  Afghanistan  2005  Developing                57.3            291.0
11  Afghanistan  2004  Developing                57.0            293.0
12  Afghanistan  2003  Developing                56.7            295.0
13  Afghanistan  2002  Developing                56.2              3.0
14  Afghanistan  2001  Developing                55.3            316.0
15  Afghanistan  2000  Developing                54.8            321.0
16      Albania  2015  Developing                77.8             74.0
17      Albania  2014  Developing                77.5              8.0
18      Albania  2013  Developing                77.2             84.0
19      Albania  2012  Developing                76.9             86.0

    infant deaths  Alcohol  percentage expenditure  Hepatitis B  Measles   \
0              62     0.01                71.279624         65.0      1154
1              64     0.01                73.523582         62.0       492
2              66     0.01                73.219243         64.0       430
3              69     0.01                78.184215         67.0      2787
4              71     0.01                 7.097109         68.0      3013
5              74     0.01                79.679367         66.0      1989
6              77     0.01                56.762217         63.0      2861
7              80     0.03                25.873925         64.0      1599
8              82     0.02                10.910156         63.0      1141
9              84     0.03                17.171518         64.0      1990
10             85     0.02                 1.388648         66.0      1296
11             87     0.02                15.296066         67.0       466
12             87     0.01                11.089053         65.0       798
13             88     0.01                16.887351         64.0      2486
14             88     0.01                10.574728         63.0      8762
15             88     0.01                10.424960         62.0      6532
16              0     4.60               364.975229         99.0         0
17              0     4.51               428.749067         98.0         0
18              0     4.76               430.876979         99.0         0
19              0     5.14               412.443356         99.0         9

    …  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
5   …   66.0               9.20         66.0        0.1  553.328940
6   …   63.0               9.42         63.0        0.1  445.893298
7   …   64.0               8.33         64.0        0.1  373.361116
8   …   63.0               6.73         63.0        0.1  369.835796
9   …   58.0               7.43         58.0        0.1  272.563770
10  …   58.0               8.70         58.0        0.1   25.294130
```

```
11  …    5.0              8.79          5.0        0.1   219.141353
12  …   41.0              8.82         41.0        0.1   198.728544
13  …   36.0              7.76         36.0        0.1   187.845950
14  …   35.0              7.80         33.0        0.1   117.496980
15  …   24.0              8.20         24.0        0.1   114.560000
16  …   99.0              6.00         99.0        0.1  3954.227830
17  …   98.0              5.88         98.0        0.1  4575.763787
18  …   99.0              5.66         99.0        0.1  4414.723140
19  …   99.0              5.59         99.0        0.1  4247.614380

    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
5    2883167.0                   18.4                18.4
6     284331.0                   18.6                18.7
7    2729431.0                   18.8                18.9
8   26616792.0                   19.0                19.1
9    2589345.0                   19.2                19.3
10    257798.0                   19.3                19.5
11  24118979.0                   19.5                19.7
12   2364851.0                   19.7                19.9
13  21979923.0                   19.9                 2.2
14   2966463.0                    2.1                 2.4
15    293756.0                    2.3                 2.5
16     28873.0                    1.2                 1.3
17    288914.0                    1.2                 1.3
18    289592.0                    1.3                 1.4
19      2941.0                    1.3                 1.4

    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                             0.433        8.7
8                             0.415        8.4
9                             0.405        8.1
10                            0.396        7.9
11                            0.381        6.8
12                            0.373        6.5
13                            0.341        6.2
```

```
14                              0.340        5.9
15                              0.338        5.5
16                              0.762       14.2
17                              0.761       14.2
18                              0.759       14.2
19                              0.752       14.2

[20 rows x 22 columns]
```

[44]:
```
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
```

[45]:
```
df.describe()
```

[45]:
```
              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
```

[46]:
```
df.isnull().sum()
```

[46]:
```
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
```

[47]:
```
df.dtypes
```

[47]:
```
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
```

[48]:
```
nullv = ['Life expectancy ', 'Adult Mortality', 'Alcohol', 'Hepatitis B', ' BMI␣
 ↪', 'Polio', 'Total expenditure',
       'Diphtheria ', 'GDP', 'Population',' thinness  1-19 years', ' thinness␣
 ↪5-9 years','Income composition of resources', 'Schooling']
for i in nullv:
    mean = df[i].mean()
    df[i].fillna(value=mean, inplace = True)
```

[50]:
```
print(df.columns)
```

```
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')
```

[51]:
```
print(df.isnull().sum())
```

```
Country                          0
Year                             0
Status                           0
Life expectancy                  0
Adult Mortality                  0
infant deaths                    0
Alcohol                          0
percentage expenditure           0
Hepatitis B                      0
Measles                          0
 BMI                             0
under-five deaths                0
Polio                            0
```

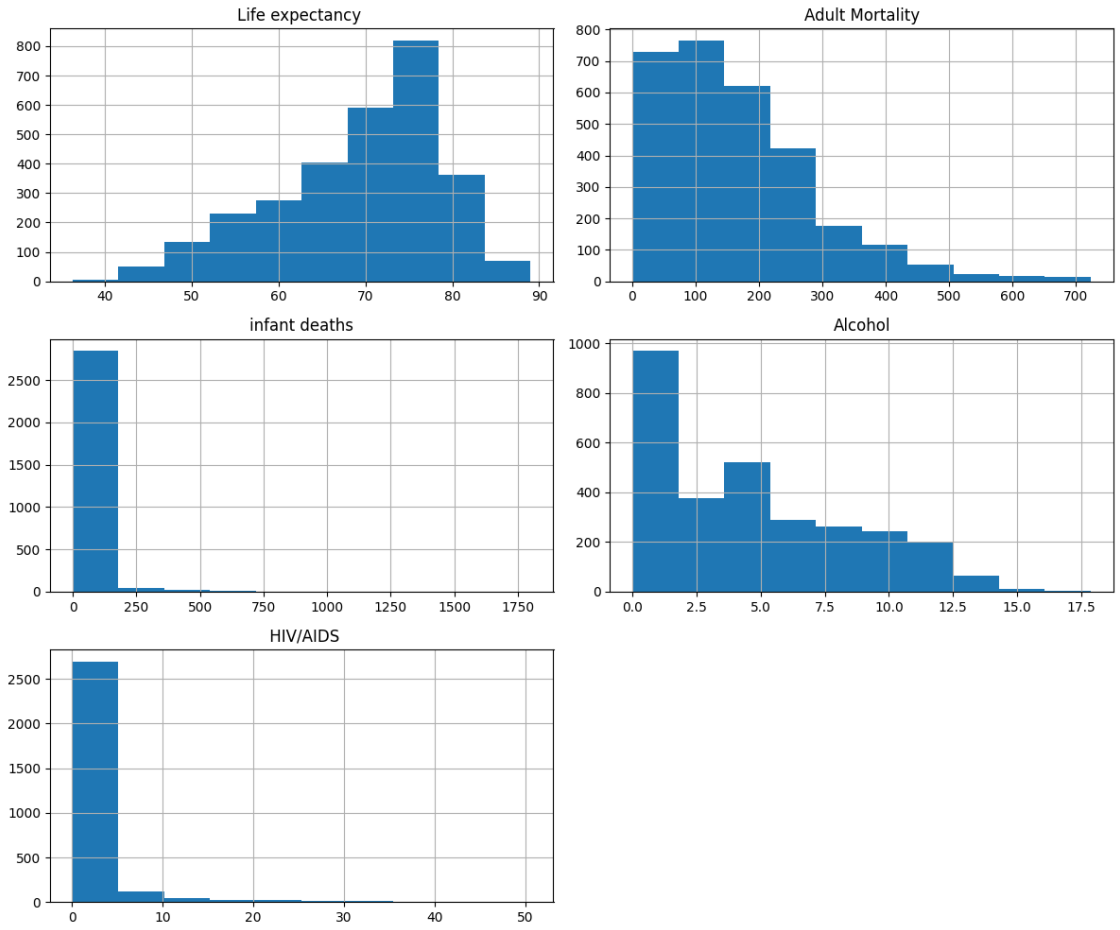
```
Total expenditure                  0
Diphtheria                         0
 HIV/AIDS                          0
GDP                                0
Population                         0
 thinness  1-19 years              0
 thinness 5-9 years                0
Income composition of resources    0
Schooling                          0
dtype: int64
```

2 Exploratory Data Analysis(EDA)

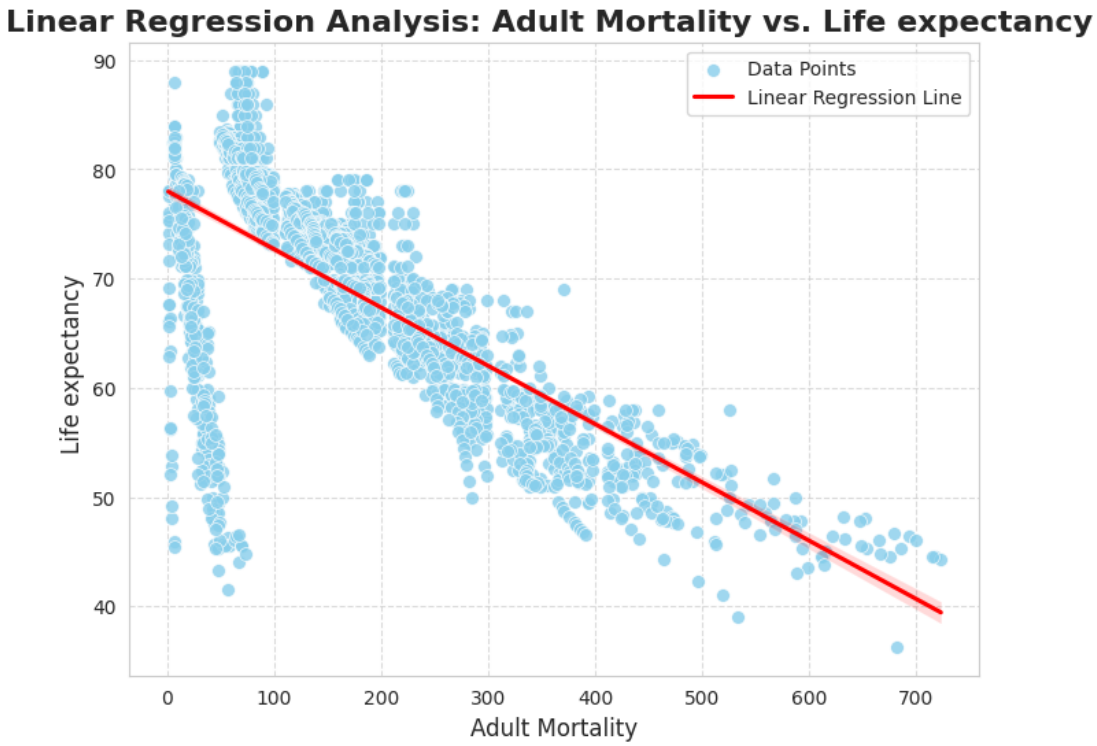
[59]:
```
columns_of_interest = ['Life expectancy ', 'Adult Mortality', 'infant deaths',␣
 ↪'Alcohol', ' HIV/AIDS']
df_subset = df[columns_of_interest]

summary_stats = df_subset.describe()
print(summary_stats)
```

```
       Life expectancy   Adult Mortality  infant deaths      Alcohol  \
count       2938.000000      2938.000000    2938.000000  2938.000000
mean          69.224932       164.796448      30.303948     4.602861
std            9.507640       124.080302     117.926501     3.916288
min           36.300000         1.000000       0.000000     0.010000
25%           63.200000        74.000000       0.000000     1.092500
50%           72.000000       144.000000       3.000000     4.160000
75%           75.600000       227.000000      22.000000     7.390000
max           89.000000       723.000000    1800.000000    17.870000

          HIV/AIDS
count  2938.000000
mean      1.742103
std       5.077785
min       0.100000
25%       0.100000
50%       0.100000
75%       0.800000
max      50.600000
```

[60]:
```
df_subset.hist(figsize=(12, 10))
plt.tight_layout()
plt.show()
```

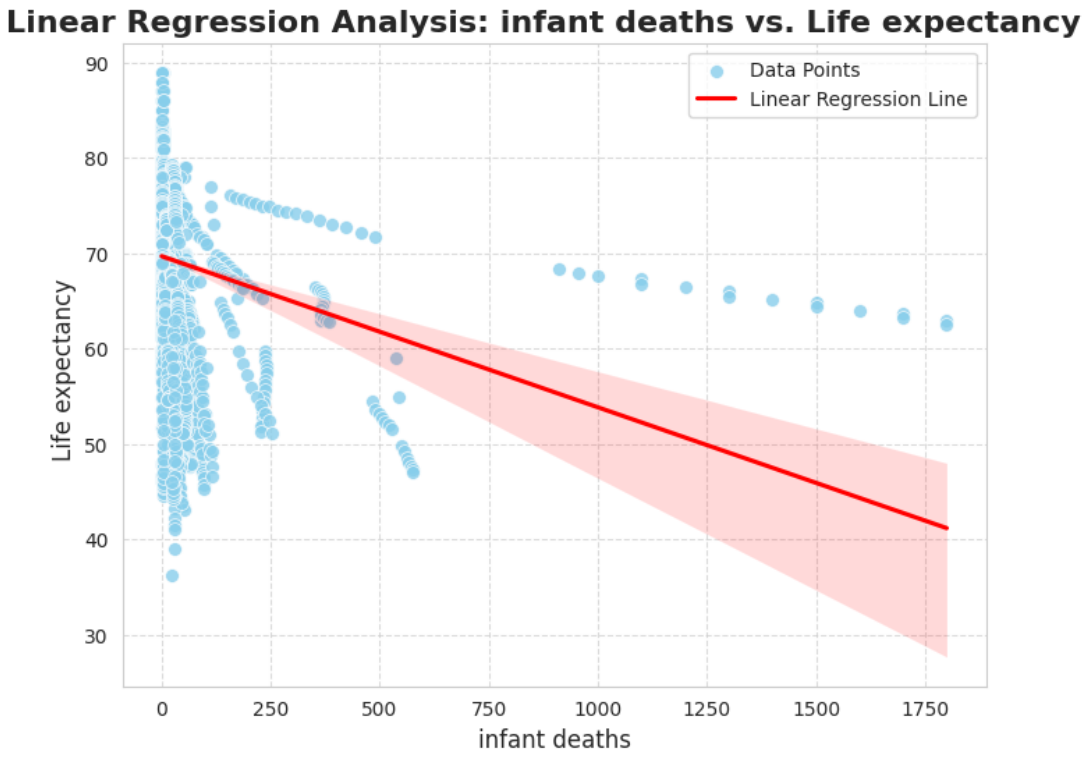


[82]:
```
import seaborn as sns
from matplotlib import pyplot as plt
sns.set_style("whitegrid")
fig, ax = plt.subplots(figsize=(8, 6))
sns.scatterplot(data=df, x='Adult Mortality', y='Life expectancy ', s=50,␣
 ↪alpha=0.8, color='skyblue', ax=ax)
sns.regplot(data=df, x='Adult Mortality', y='Life expectancy ', scatter=False,␣
 ↪color='red', ax=ax)

plt.title('Linear Regression Analysis: Adult Mortality vs. Life expectancy ',␣
 ↪fontsize=16, weight='bold')
plt.xlabel('Adult Mortality', fontsize=12)
plt.ylabel('Life expectancy ', fontsize=12)

plt.grid(True, linestyle='--', alpha=0.7)
plt.legend(['Data Points', 'Linear Regression Line'], loc='upper right')
plt.show()
```



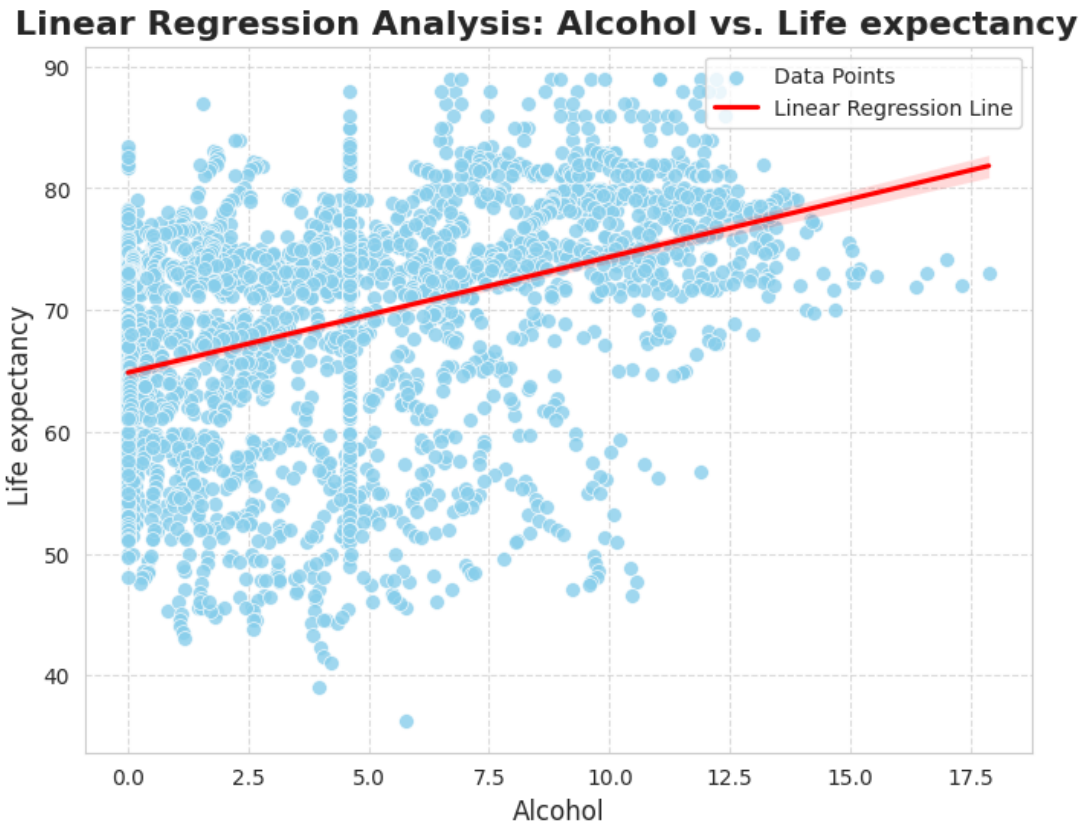
Meaning that with a higher adult mortality rates have a lower life expectancy.

[84]:
```
import seaborn as sns
from matplotlib import pyplot as plt
sns.set_style("whitegrid")
fig, ax = plt.subplots(figsize=(8, 6))
sns.scatterplot(data=df, x='infant deaths', y='Life expectancy ', s=50, alpha=0.
 ↪8, color='skyblue', ax=ax)
sns.regplot(data=df, x='infant deaths', y='Life expectancy ', scatter=False,␣
 ↪color='red', ax=ax)

plt.title('Linear Regression Analysis: infant deaths vs. Life expectancy  ',␣
 ↪fontsize=16, weight='bold')
plt.xlabel('infant deaths', fontsize=12)
plt.ylabel('Life expectancy ', fontsize=12)

plt.grid(True, linestyle='--', alpha=0.7)
plt.legend(['Data Points', 'Linear Regression Line'], loc='upper right')
plt.show()
```

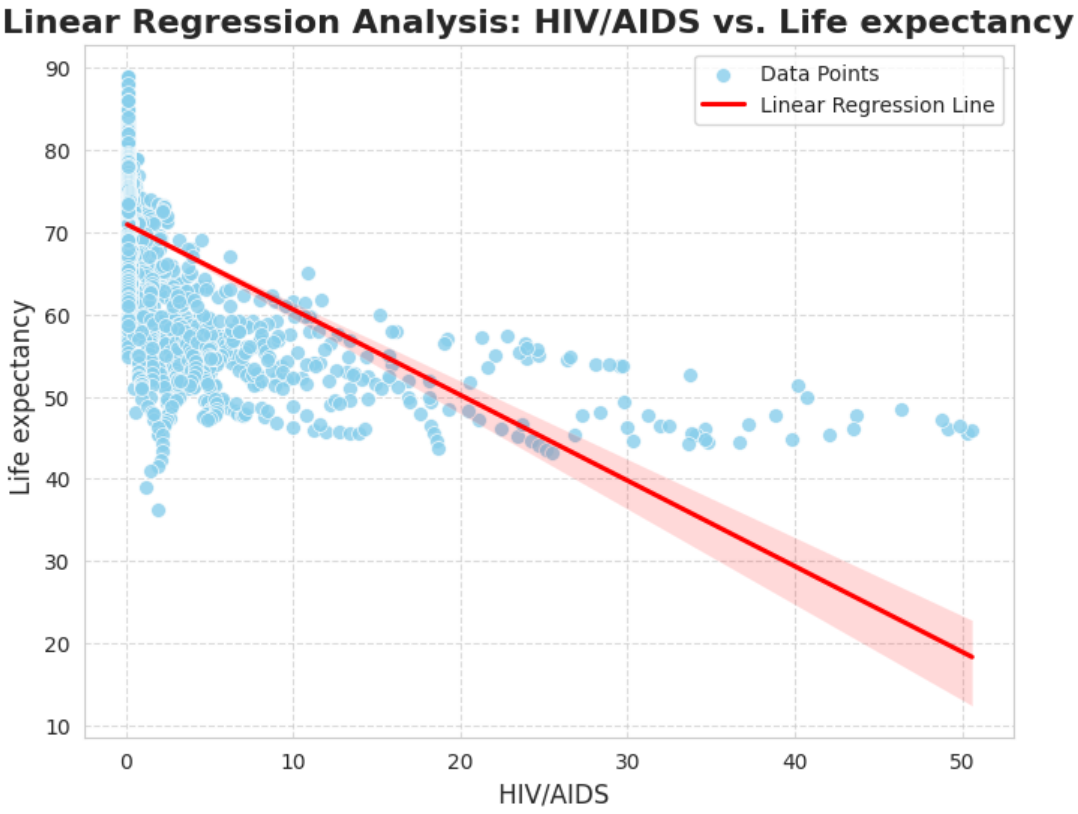


Indicating that higher infant mortality rates have a lower life expectancy.

[85]:
```
import seaborn as sns
from matplotlib import pyplot as plt
sns.set_style("whitegrid")
fig, ax = plt.subplots(figsize=(8, 6))
sns.scatterplot(data=df, x='Alcohol', y='Life expectancy ', s=50, alpha=0.8,␣
 ↪color='skyblue', ax=ax)
sns.regplot(data=df, x='Alcohol', y='Life expectancy ', scatter=False,␣
 ↪color='red', ax=ax)

plt.title('Linear Regression Analysis: Alcohol vs. Life expectancy  ',␣
 ↪fontsize=16, weight='bold')
plt.xlabel('Alcohol', fontsize=12)
plt.ylabel('Life expectancy ', fontsize=12)

plt.grid(True, linestyle='--', alpha=0.7)
plt.legend(['Data Points', 'Linear Regression Line'], loc='upper right')
plt.show()
```

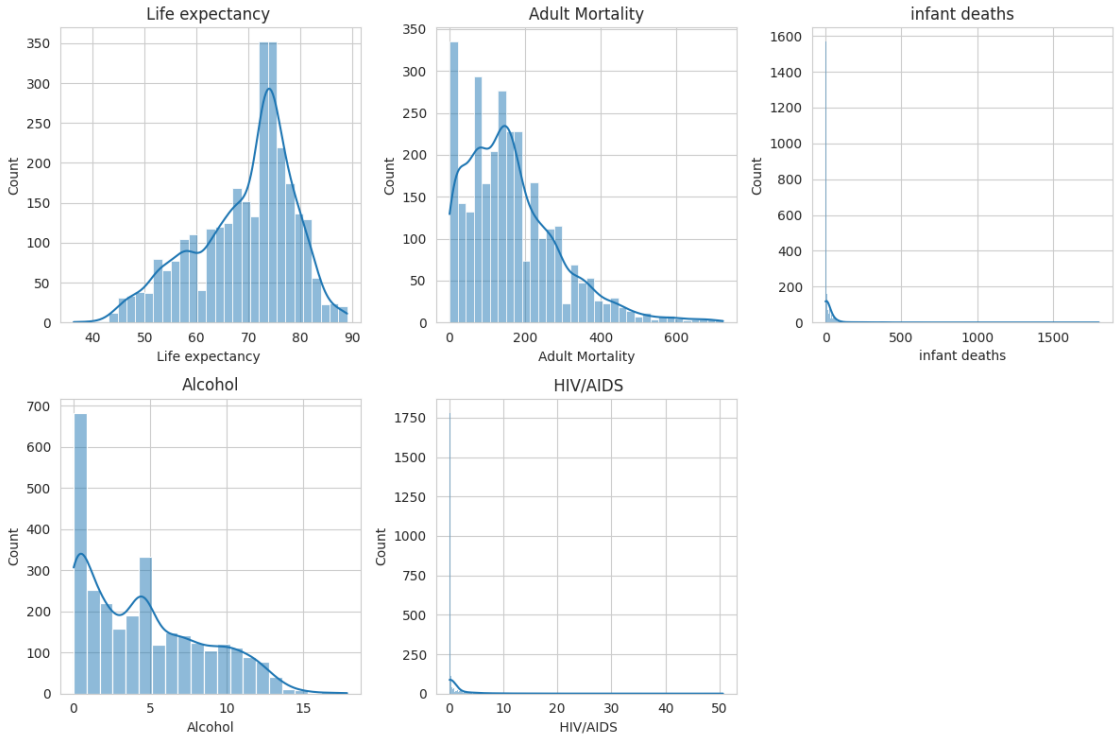


[87]:
```
import seaborn as sns
from matplotlib import pyplot as plt
sns.set_style("whitegrid")
fig, ax = plt.subplots(figsize=(8, 6))
sns.scatterplot(data=df, x=' HIV/AIDS', y='Life expectancy ', s=50, alpha=0.8,␣
 ↪color='skyblue', ax=ax)
sns.regplot(data=df, x=' HIV/AIDS', y='Life expectancy ', scatter=False,␣
 ↪color='red', ax=ax)

plt.title('Linear Regression Analysis: HIV/AIDS vs. Life expectancy  ',␣
 ↪fontsize=16, weight='bold')
plt.xlabel(' HIV/AIDS', fontsize=12)
plt.ylabel('Life expectancy ', fontsize=12)

plt.grid(True, linestyle='--', alpha=0.7)
plt.legend(['Data Points', 'Linear Regression Line'], loc='upper right')
plt.show()
```

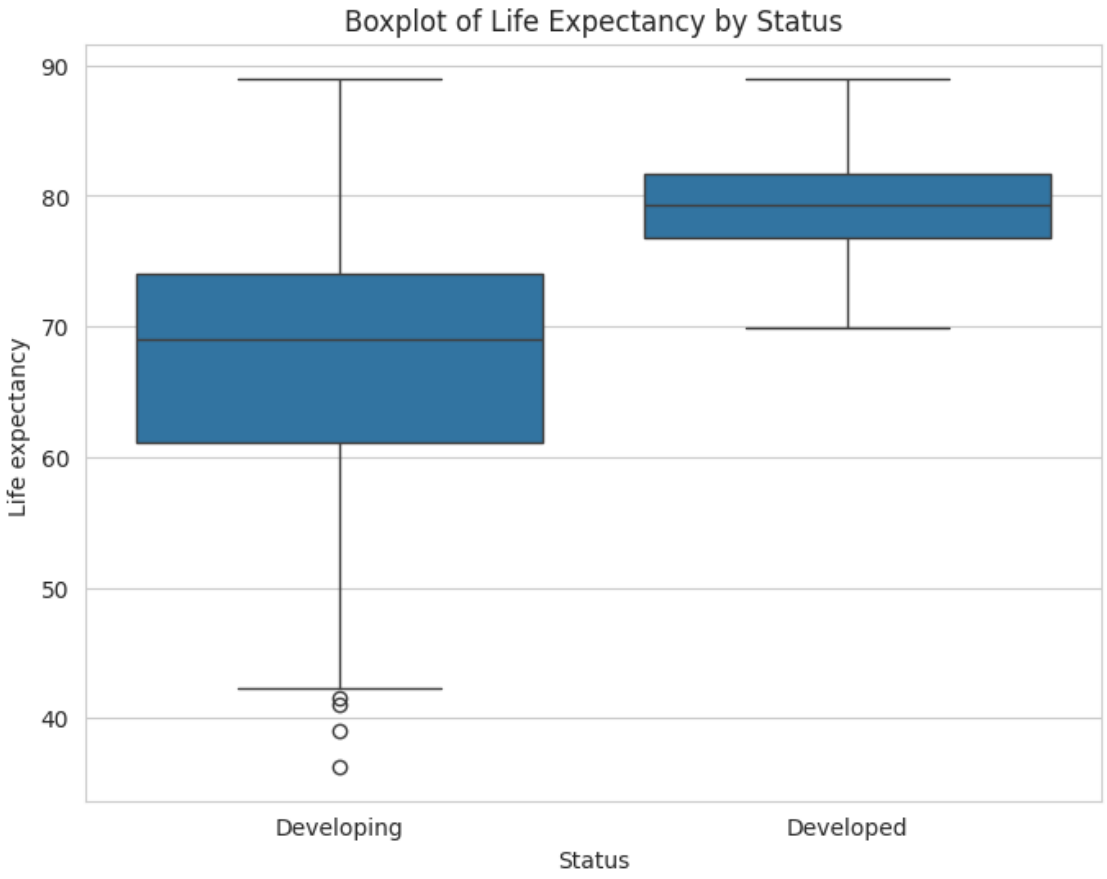


as higher HIV/AIDS rates are typically have a lower life expectancy due to the impact of the disease on health and mortality rates.

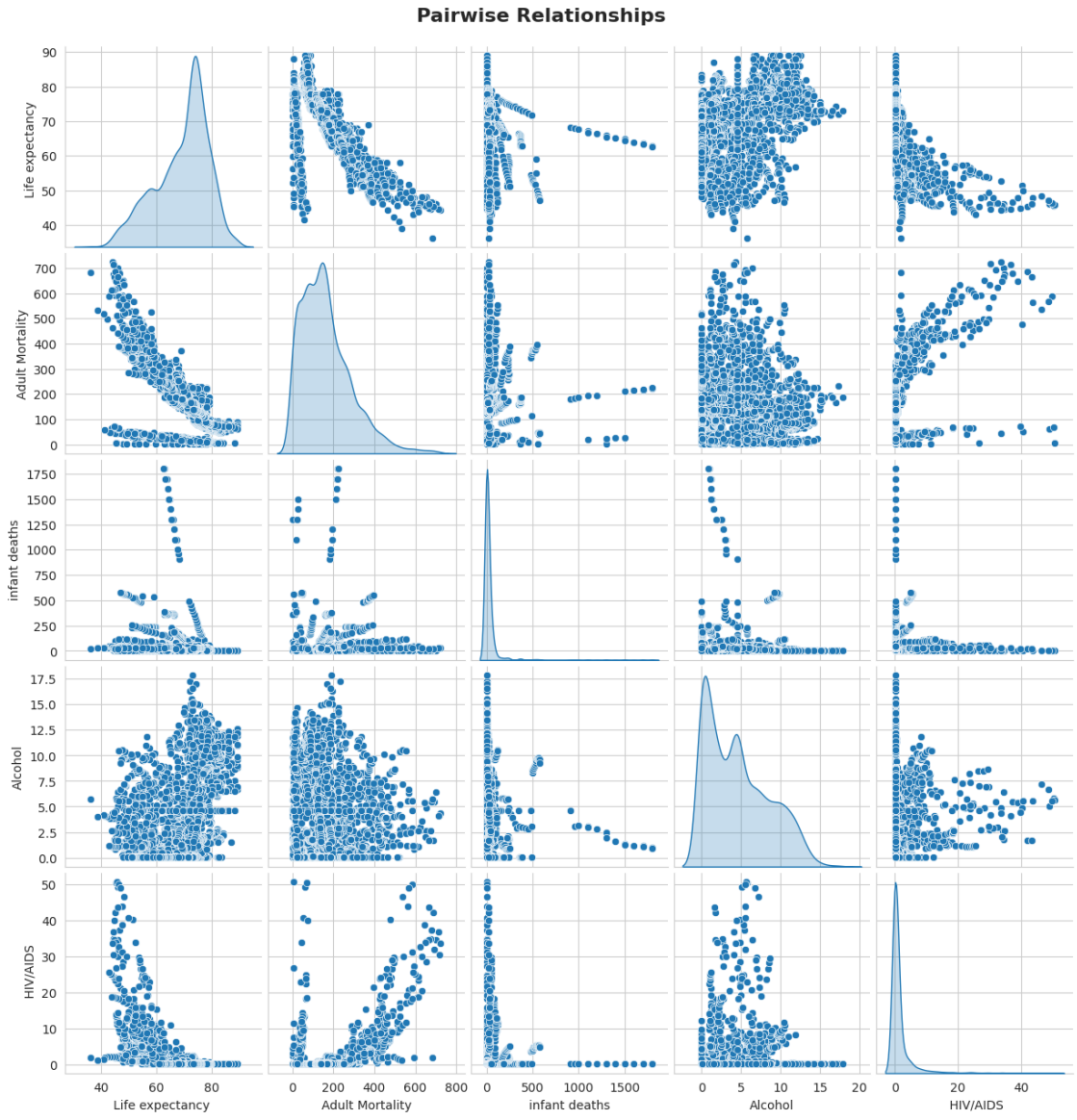
[89]:
```
important_variables = ['Life expectancy ', 'Adult Mortality', 'infant deaths',␣
 ↪'Alcohol', ' HIV/AIDS']

# Subset the dataframe
selected_df = df[important_variables]

# Plot histograms for each variable
plt.figure(figsize=(12, 8))
for i, variable in enumerate(selected_df.columns):
    plt.subplot(2, 3, i+1)
    sns.histplot(selected_df[variable], kde=True)
    plt.title(variable)
plt.tight_layout()
plt.show()
```



[93]:
```
plt.figure(figsize=(8, 6))
sns.boxplot(x='Status', y='Life expectancy ', data=df)
plt.title('Boxplot of Life Expectancy by Status')
plt.show()
```

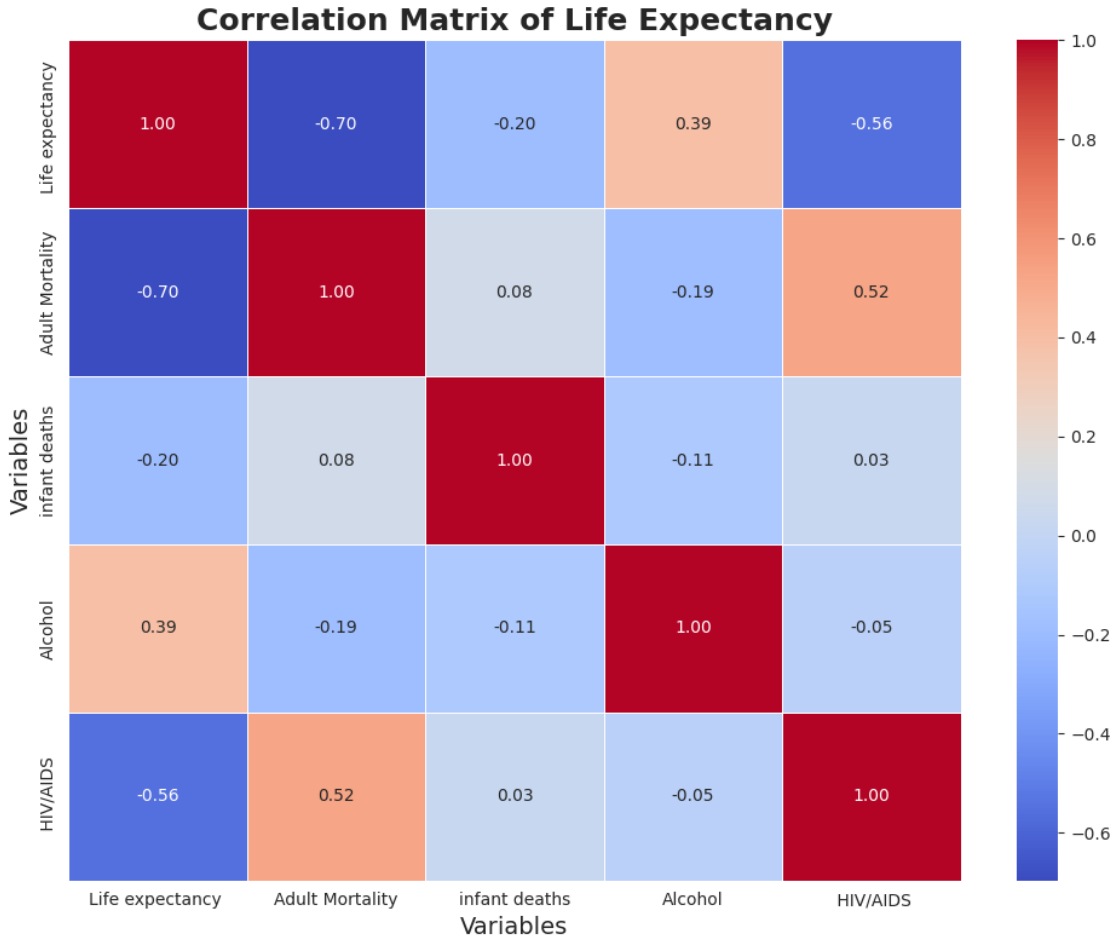


[90]:
```
sns.pairplot(selected_df, diag_kind='kde')
plt.suptitle('Pairwise Relationships', y=1.02, fontsize=16, weight='bold')
plt.show()
```



[91]:
```
import seaborn as sns
import matplotlib.pyplot as plt

correlation_matrix = selected_df.corr()
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f",␣
↪linewidths=0.5)
plt.title('Correlation Matrix of Life Expectancy', fontsize=18, weight='bold')

plt.xlabel('Variables', fontsize=14)
plt.ylabel('Variables', fontsize=14)
```

```
plt.tight_layout()
plt.show()
```



[94]:
```
print("Summary of Correlations:")
for i in range(len(correlation_matrix)):
    for j in range(i+1, len(correlation_matrix)):
        variable1 = correlation_matrix.index[i]
        variable2 = correlation_matrix.columns[j]
        correlation = correlation_matrix.iloc[i, j]
        if correlation > 0:
            print(f"The correlation between '{variable1}' and '{variable2}' is␣
 ↪positive: {correlation:.2f}")
        elif correlation < 0:
            print(f"The correlation between '{variable1}' and '{variable2}' is␣
 ↪negative: {correlation:.2f}")
```

```
Summary of Correlations:
The correlation between 'Life expectancy ' and 'Adult Mortality' is negative:
```

```
-0.70
The correlation between 'Life expectancy ' and 'infant deaths' is negative:
-0.20
The correlation between 'Life expectancy ' and 'Alcohol' is positive: 0.39
The correlation between 'Life expectancy ' and ' HIV/AIDS' is negative: -0.56
The correlation between 'Adult Mortality' and 'infant deaths' is positive: 0.08
The correlation between 'Adult Mortality' and 'Alcohol' is negative: -0.19
The correlation between 'Adult Mortality' and ' HIV/AIDS' is positive: 0.52
The correlation between 'infant deaths' and 'Alcohol' is negative: -0.11
The correlation between 'infant deaths' and ' HIV/AIDS' is positive: 0.03
The correlation between 'Alcohol' and ' HIV/AIDS' is negative: -0.05
```

[95]:
```
from sklearn.datasets import load_iris
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
```

[96]:
```
iris = load_iris()
X = iris.data[:, 2].reshape(-1, 1)
y = iris.data[:, 3]
```

[97]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
↪random_state=42)
```

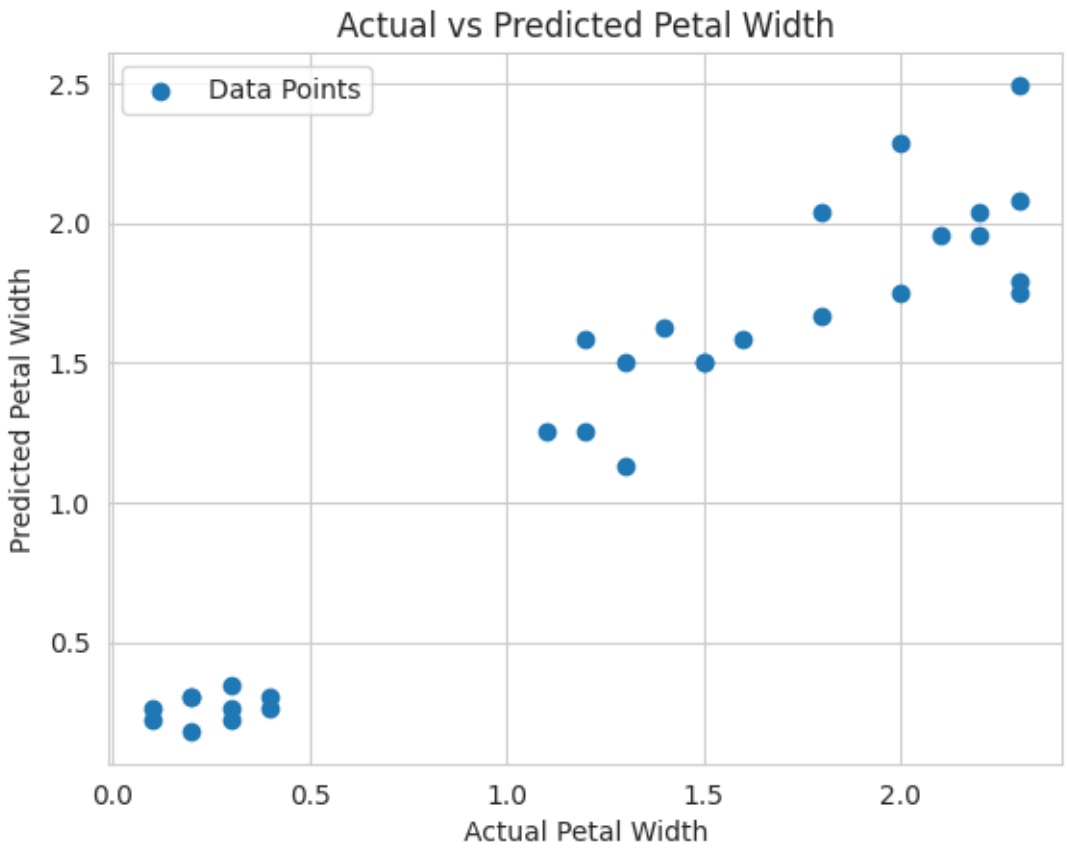
[98]:
```
model = LinearRegression()
```

[99]:
```
model.fit(X_train, y_train)
```

[99]:
```
LinearRegression()
```

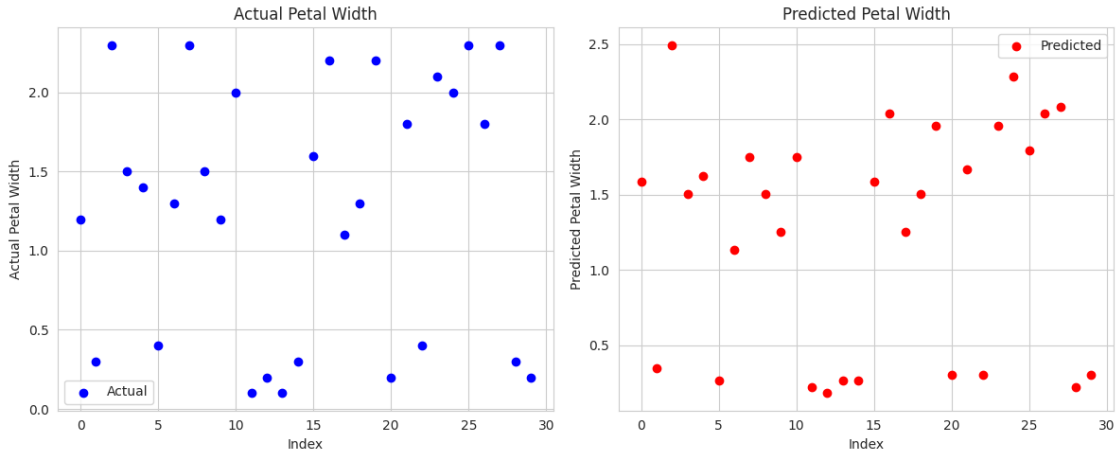
[100]:
```
y_pred = model.predict(X_test)
```

[101]:
```
import matplotlib.pyplot as plt


plt.scatter(y_test, y_pred, label='Data Points')
plt.xlabel('Actual Petal Width')
plt.ylabel('Predicted Petal Width')
plt.title('Actual vs Predicted Petal Width')


plt.legend()

plt.show()
```

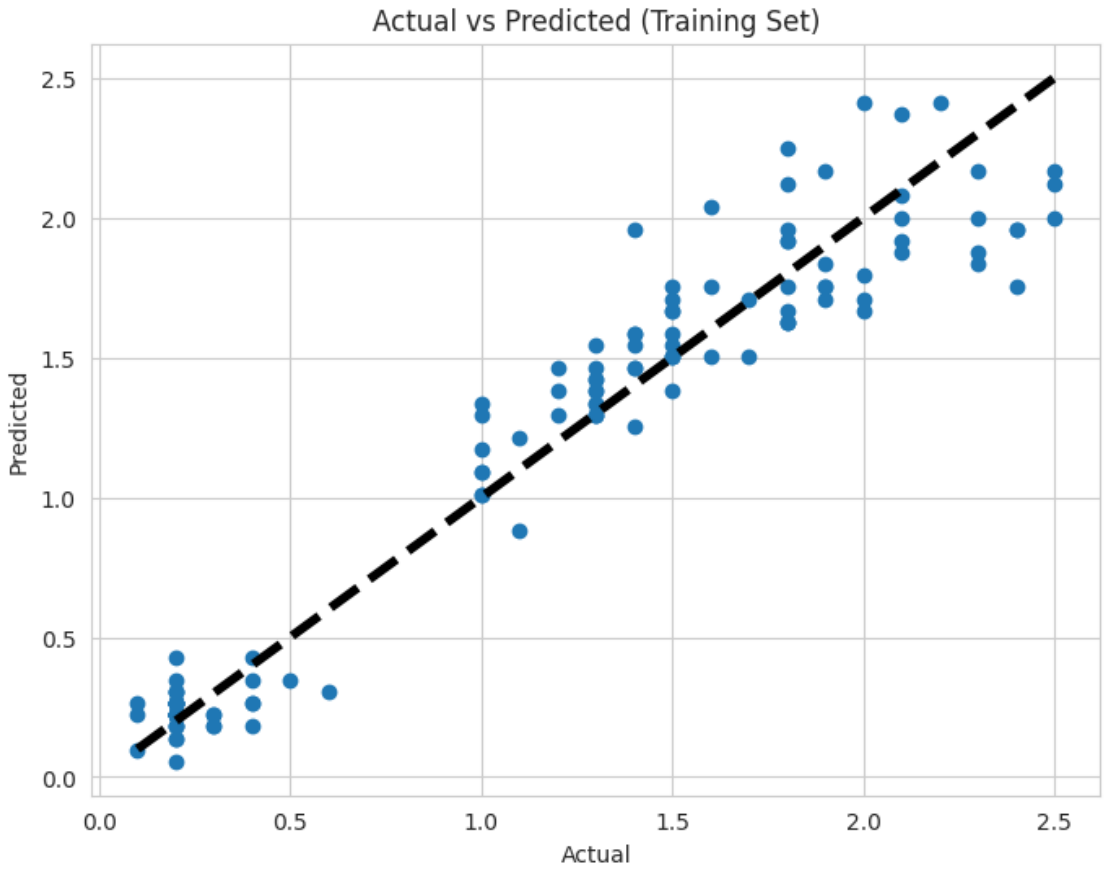


[102]:
```
import matplotlib.pyplot as plt

# Create subplots
fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(12, 5))

# Plot for Actual Petal Width
ax1.scatter(range(len(y_test)), y_test, color='blue', label='Actual')
ax1.set_xlabel('Index')
ax1.set_ylabel('Actual Petal Width')
ax1.set_title('Actual Petal Width')
ax1.legend()

# Plot for Predicted Petal Width
ax2.scatter(range(len(y_pred)), y_pred, color='red', label='Predicted')
ax2.set_xlabel('Index')
ax2.set_ylabel('Predicted Petal Width')
ax2.set_title('Predicted Petal Width')
ax2.legend()

plt.tight_layout()
```

```
plt.show()
```



[105]:
```
y_train_pred = model.predict(X_train)
plt.figure(figsize=(8, 6))
plt.scatter(y_train, y_train_pred)
plt.plot([y_train.min(), y_train.max()], [y_train.min(), y_train.max()], 'k--',␣
↪lw=4)
plt.xlabel('Actual')
plt.ylabel('Predicted')
plt.title('Actual vs Predicted (Training Set)')
plt.show()
```



[107]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
↪random_state=42)

# Training the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Predicting on the test set
y_pred = model.predict(X_test)

# Calculating evaluation metrics
MAE = metrics.mean_absolute_error(y_test, y_pred)
MSE = metrics.mean_squared_error(y_test, y_pred)
RMSE = np.sqrt(MSE)

# Displaying evaluation metrics
print("Mean Absolute Error (MAE):", MAE)
print("Mean Squared Error (MSE):", MSE)
print("Root Mean Squared Error (RMSE):", RMSE)
```

```
Mean Absolute Error (MAE): 0.16818126256563326
Mean Squared Error (MSE): 0.045604284097661846
Root Mean Squared Error (RMSE): 0.21355159586774772
```

By these insights, healthcare professionals can develop to improve public health outcomes and enhance life expectancy worldwide.