Importing Libraries For Performing Exploratory Data Analysis

In [151…
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
%matplotlib inline
import warnings
warnings.filterwarnings('ignore')
```

In [152…
```
Q=pd.read_excel('Life Expectancy Data_HV22.xlsx')
```

In [153…
```
Q.head()
```

Out[153]:

	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B
0	Afghanistan	2015	Developing	65.0	263.0	62	0.01	71.279624	65.0
1	Afghanistan	2014	Developing	59.9	271.0	64	0.01	73.523582	62.0
2	Afghanistan	2013	Developing	59.9	268.0	66	0.01	73.219243	64.0
3	Afghanistan	2012	Developing	59.5	272.0	69	0.01	78.184215	67.0
4	Afghanistan	2011	Developing	59.2	275.0	71	0.01	7.097109	68.0

5 rows × 22 columns

Now we have successfully loaded our dataset.

In [154…
```
Q.shape
```

Out[154]:
```
(2938, 22)
```

As it is showing there are 2938 records and 22 features.

In [155…
```
Q.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
```

Using Q.info() we get detailed information

In [156…:

```
Q.isna().sum()
```

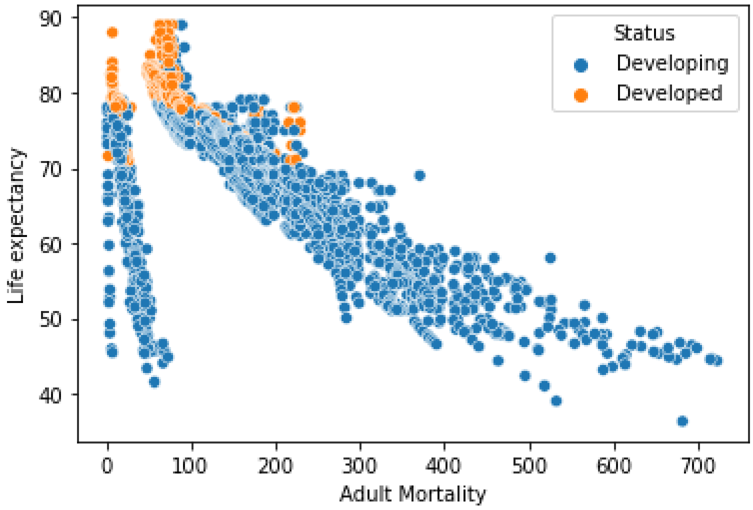
Out[156]:

```
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
```

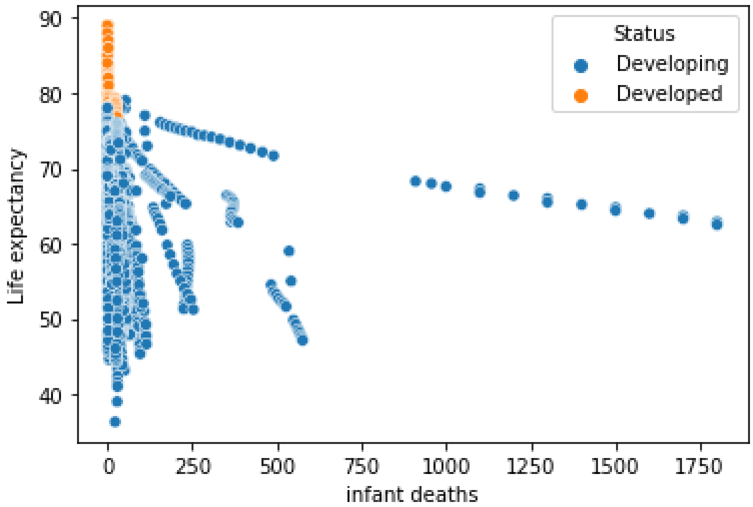
It shows how many null values are there and in which feature.

Data Visualisation:Graphs

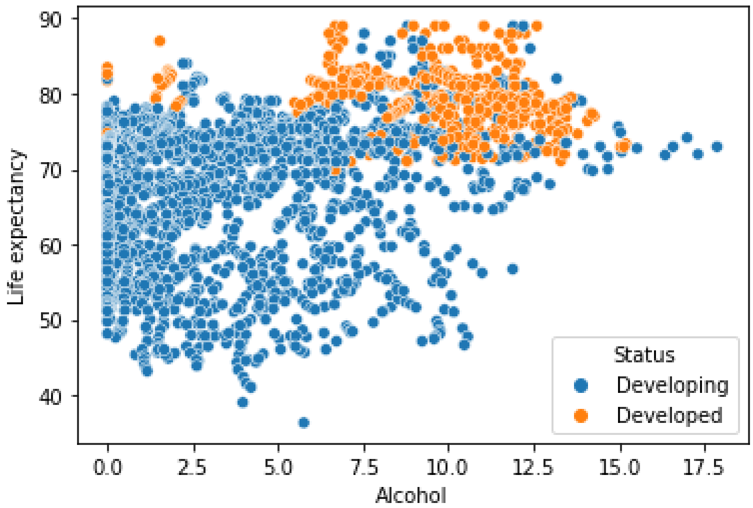
In [157…
```
sns.scatterplot(data=Q, x="Adult Mortality", y="Life expectancy", hue="Status")
plt.show()
```



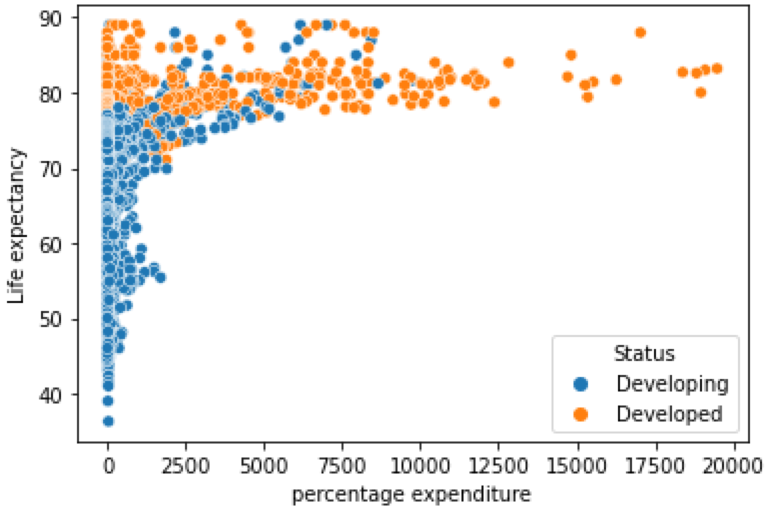
In [158…
```
sns.scatterplot(data=Q, x="infant deaths", y="Life expectancy", hue="Status")
plt.show()
```



In [159…
```
sns.scatterplot(data=Q, x="Alcohol", y="Life expectancy", hue="Status")
plt.show()
```

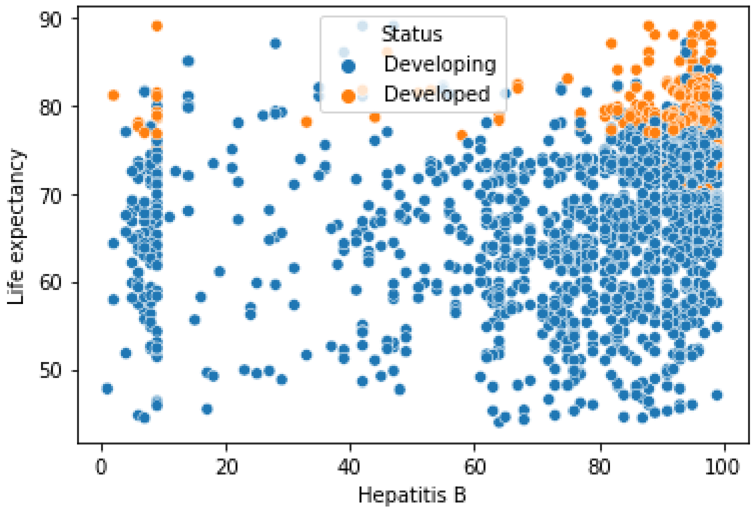


In [160…
```
sns.scatterplot(data=Q, x="percentage expenditure", y="Life expectancy", hue="Statu
plt.show()
```



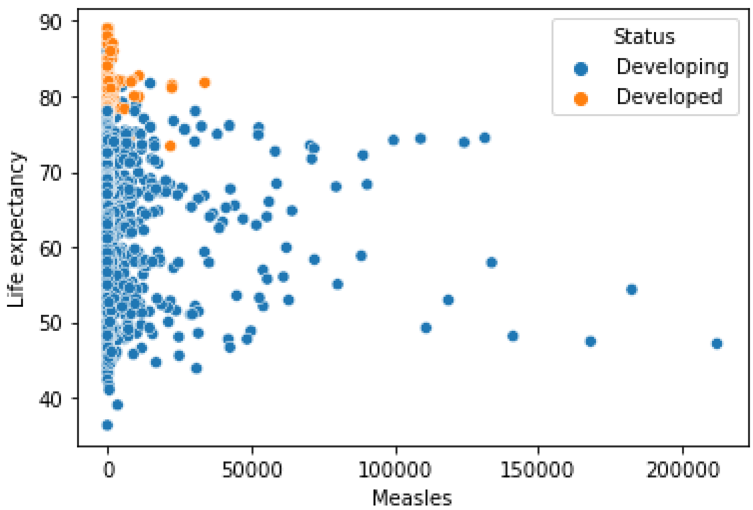
In [161…

```
sns.scatterplot(data=Q, x="Hepatitis B", y="Life expectancy", hue="Status")
plt.show()
```



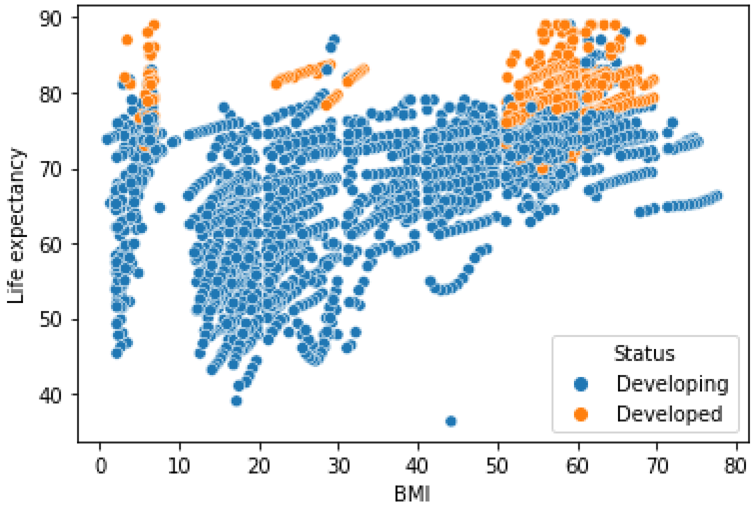
In [162…

```
sns.scatterplot(data=Q,x="Measles",y="Life expectancy",hue="Status")
plt.show()
```

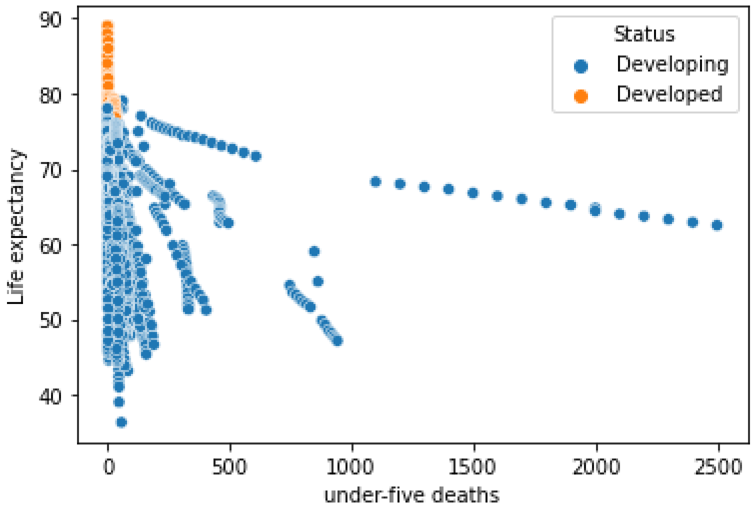


In [163…

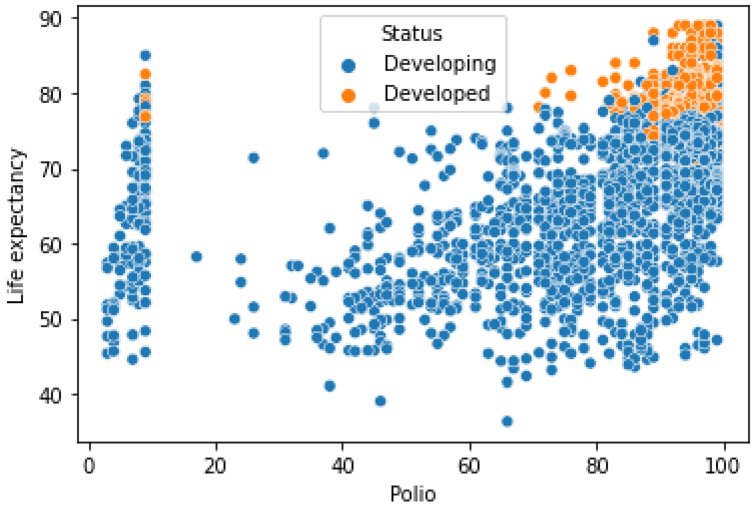
```
sns.scatterplot(data=Q,x="BMI",y="Life expectancy",hue="Status")
plt.show()
```



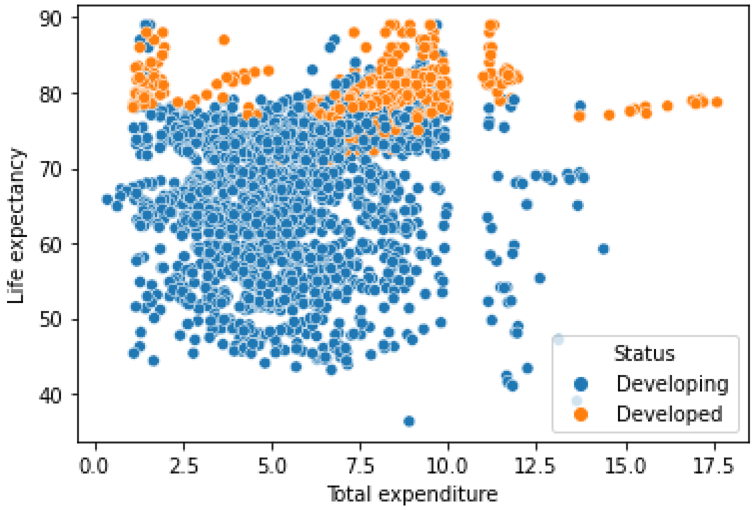
In [164…
```
sns.scatterplot(data=Q,x="under-five deaths",y="Life expectancy",hue="Status")
plt.show()
```



In [165…
```
sns.scatterplot(data=Q,x="Polio",y="Life expectancy",hue="Status")
plt.show()
```

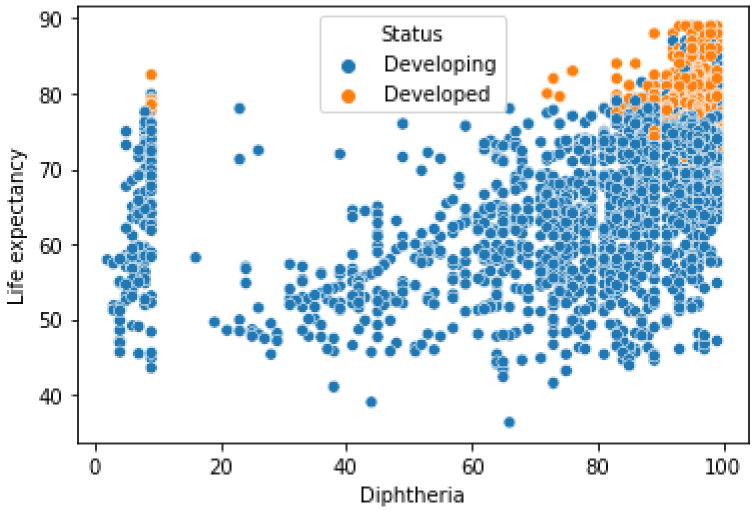


In [166…
```
sns.scatterplot(data=Q,x="Total expenditure",y="Life expectancy",hue="Status")
plt.show()
```



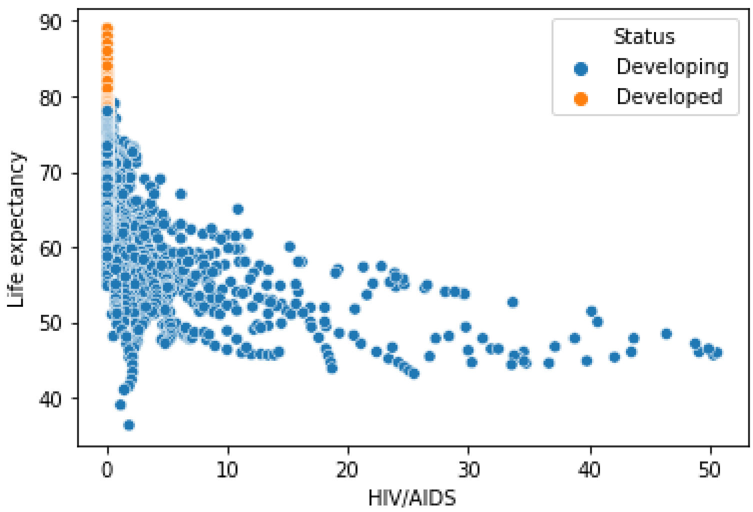
In [167…

```
sns.scatterplot(data=Q,x="Diphtheria",y="Life expectancy",hue="Status")
plt.show()
```



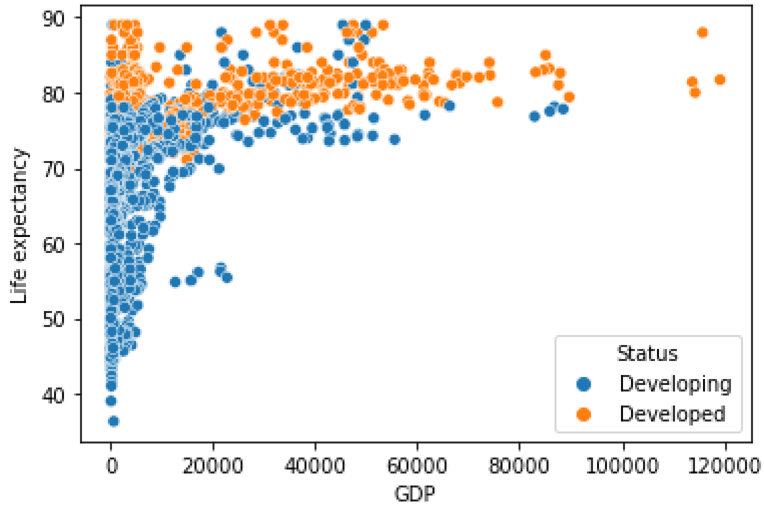
In [168…

```
sns.scatterplot(data=Q,x="HIV/AIDS",y="Life expectancy",hue="Status")
plt.show()
```



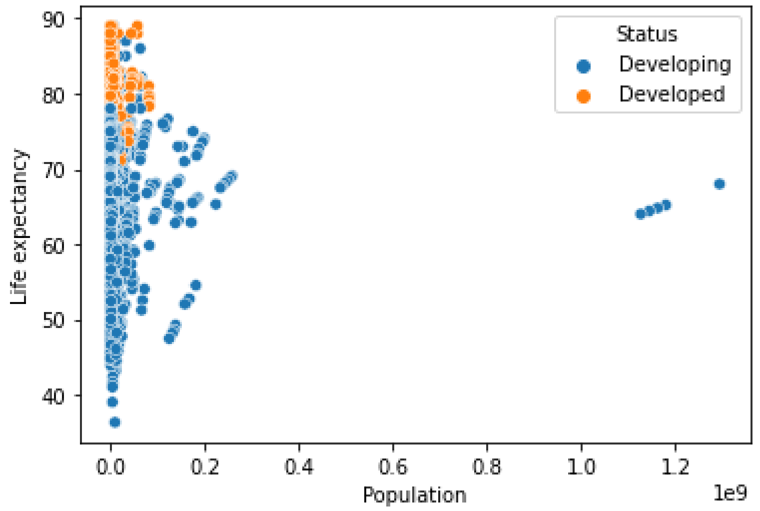
In [169…

```
sns.scatterplot(data=Q,x="GDP",y="Life expectancy",hue="Status")
plt.show()
```



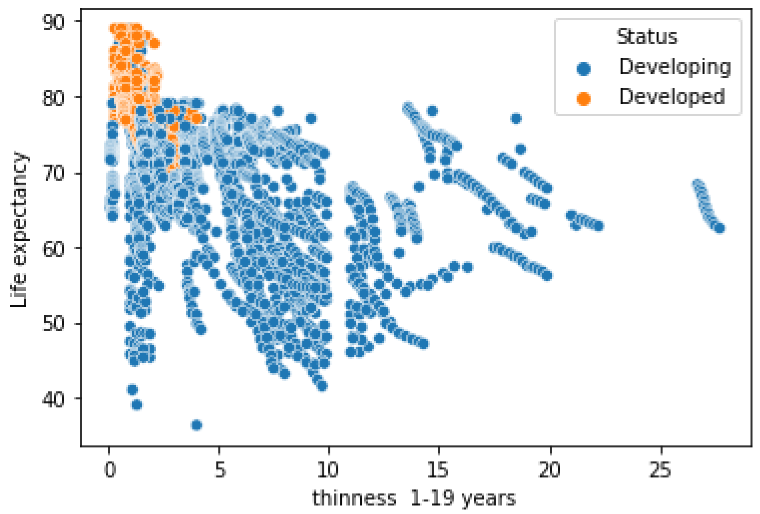
In [170…

```
sns.scatterplot(data=Q,x="Population",y="Life expectancy",hue="Status")
plt.show()
```



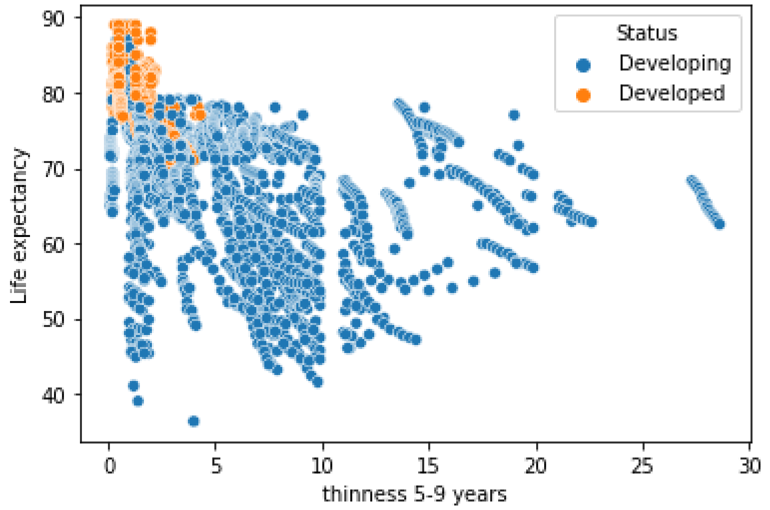
In [171…

```
sns.scatterplot(data=Q,x="thinness  1-19 years",y="Life expectancy",hue="Status")
plt.show()
```



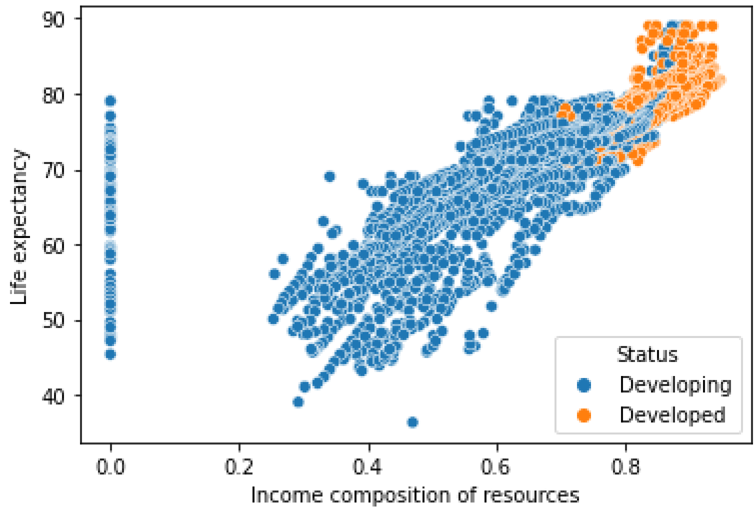
In [172…

```
sns.scatterplot(data=Q,x="thinness 5-9 years",y="Life expectancy",hue="Status")
plt.show()
```



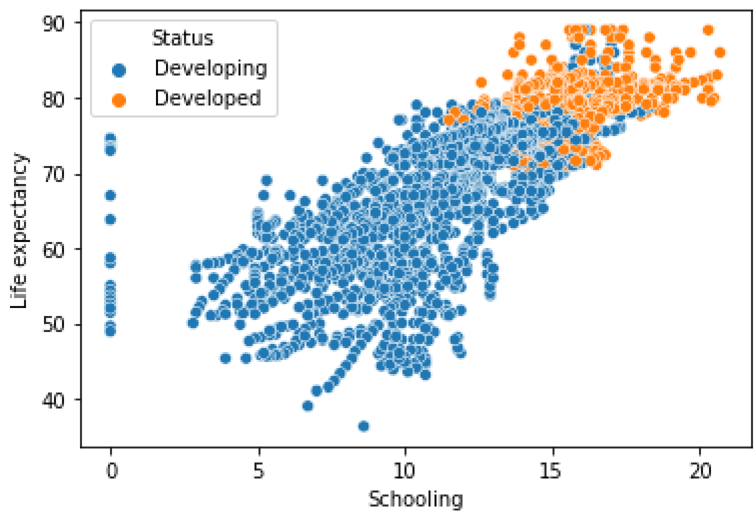
In [173…

```
sns.scatterplot(data=Q,x="Income composition of resources",y="Life expectancy",hue=
plt.show()
```



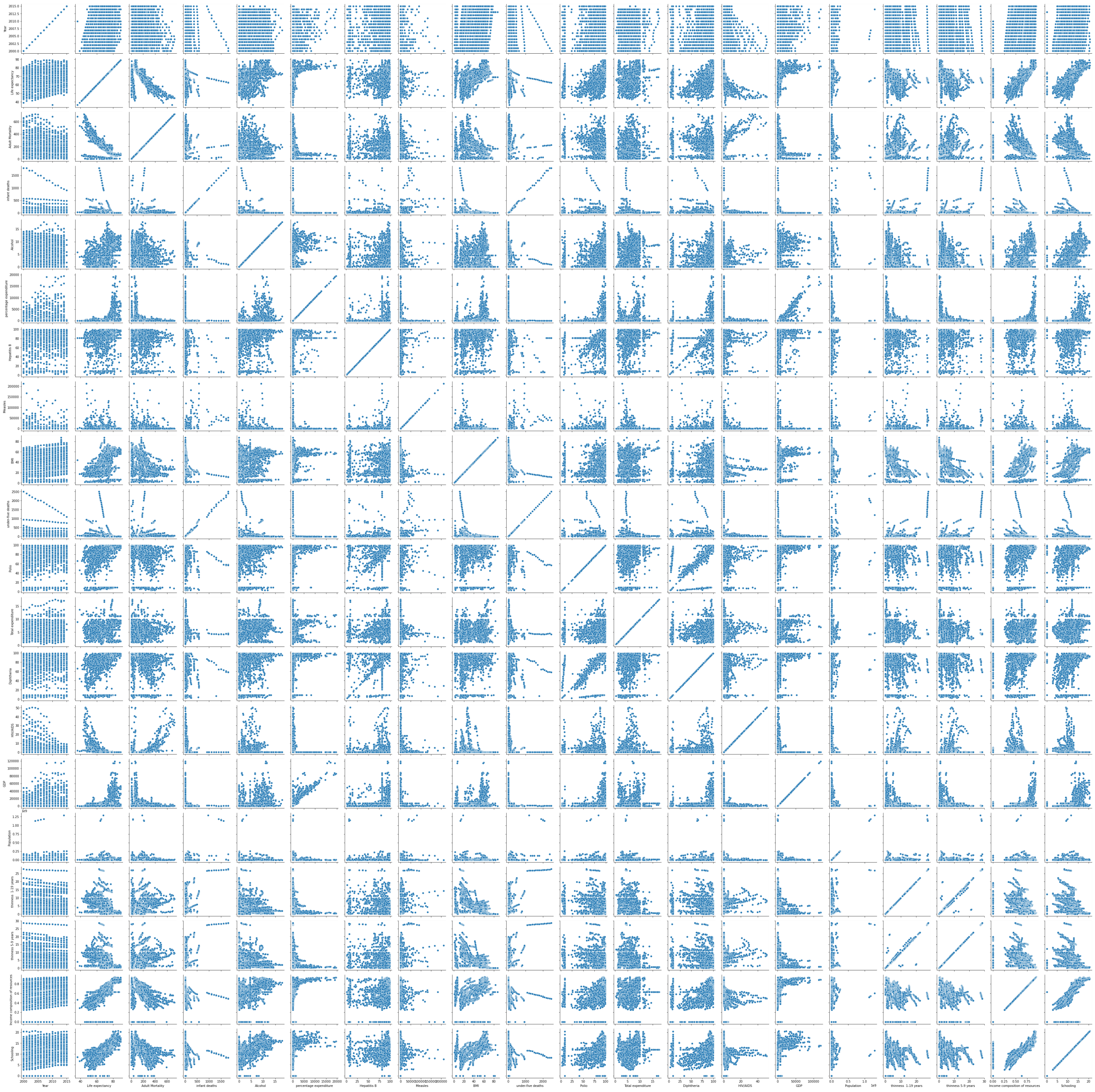
In [174…

```
sns.scatterplot(data=Q,x="Schooling",y="Life expectancy",hue="Status")
plt.show()
```



In [218…

```
g = sns.PairGrid(Q)
g.map(sns.scatterplot)
plt.show()
```



In [176…

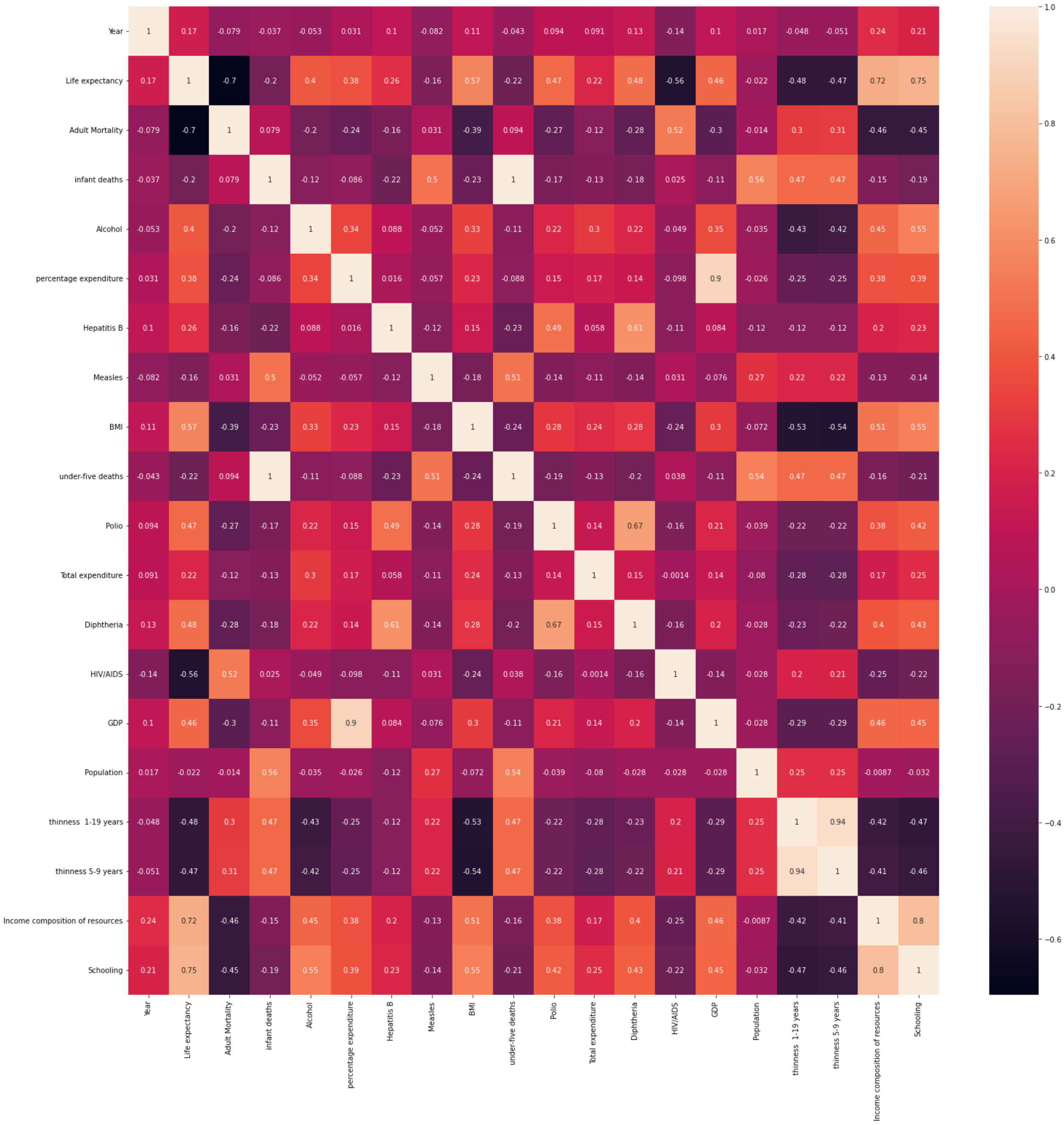
```
Q.corr()
```

Out[176]:

	Year	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Me
Year	1.000000	0.170033	-0.079052	-0.037415	-0.052990	0.031400	0.104333	-0.08
Life expectancy	0.170033	1.000000	-0.696359	-0.196557	0.404877	0.381864	0.256762	-0.1
Adult Mortality	-0.079052	-0.696359	1.000000	0.078756	-0.195848	-0.242860	-0.162476	0.0
infant deaths	-0.037415	-0.196557	0.078756	1.000000	-0.115638	-0.085612	-0.223566	0.5
Alcohol	-0.052990	0.404877	-0.195848	-0.115638	1.000000	0.341285	0.087549	-0.0
percentage expenditure	0.031400	0.381864	-0.242860	-0.085612	0.341285	1.000000	0.016274	-0.0
Hepatitis B	0.104333	0.256762	-0.162476	-0.223566	0.087549	0.016274	1.000000	-0.1
Measles	-0.082493	-0.157586	0.031176	0.501128	-0.051827	-0.056596	-0.120529	1.0
BMI	0.108974	0.567694	-0.387017	-0.227279	0.330408	0.228700	0.150380	-0.1
under-five deaths	-0.042937	-0.222529	0.094146	0.996629	-0.112370	-0.087852	-0.233126	0.5
Polio	0.094158	0.465556	-0.274823	-0.170689	0.221734	0.147259	0.486171	-0.1
Total expenditure	0.090740	0.218086	-0.115281	-0.128616	0.296942	0.174420	0.058280	-0.1
Diphtheria	0.134337	0.479495	-0.275131	-0.175171	0.222020	0.143624	0.611495	-0.1
HIV/AIDS	-0.139741	-0.556556	0.523821	0.025231	-0.048845	-0.097857	-0.112675	0.0
GDP	0.101620	0.461455	-0.296049	-0.108427	0.354712	0.899373	0.083903	-0.0
Population	0.016969	-0.021538	-0.013647	0.556801	-0.035252	-0.025662	-0.123321	0.2
thinness 1-19 years	-0.047876	-0.477183	0.302904	0.465711	-0.428795	-0.251369	-0.120429	0.2
thinness 5-9 years	-0.050929	-0.471584	0.308457	0.471350	-0.417414	-0.252905	-0.124960	0.2
Income composition of resources	0.243468	0.724776	-0.457626	-0.145139	0.450040	0.381952	0.199549	-0.1
Schooling	0.209400	0.751975	-0.454612	-0.193720	0.547378	0.389687	0.231117	-0.1

In [177…

```
plt.figure(figsize=(25,25))
sns.heatmap(Q.corr(),annot=True)
plt.show()
```



Graphs and it's Conclusion

1) As adult mortality increased life expectancy decreases.

2) As infant death increases Life expectancy decreases.

3) Alcohol does not show any significant impact on life expectancy, but it's greater consumption shows low life expectancy especially in developing countries.

4) Percentage expenditure also does not show any significant impact on life expectancy, but it's increase does not decreases the life expectancy.

5) Hepatitis B also not show impact on life expectancy probably due to increase of vaccination and awarenss about it.

6) If measles is more then life expectancy is less.

7) If polio is more then life expectancy is less mostly in developing countries, but is also does not show negative impact all the time,especially in developed countries probably due

to increase of vaccination drives about polio and it's awareness.

8) If HIV/AIDS is more then life expectancy is less.

9) As we ca see from the graph that if population is more then life expectancy is low.

10) If thinness is more then life expectancy is expected to be less.

11) If income composition of resources is more then life expectancy is more mostly in developed countries.

12) If Schooling is more then life expectancy is more, here also this case is more in developed countries.

13) Also the heatmap shows the positive relation between Life expectancy and Alcohol,percentage expenditure,Hepatitis B, BMI, Polio, Total expenditure, Diphtheria, GDP, Income composition of resources and Schooling. 14) It shows positive relation to certain diseases it means healthcare system work probably good in those cases and awareness campaigns and vaccination drives should have helped a lot in this.

15) It shows negative relation with rest of the features means inverse relation to life expectancy. 16) Though we have been seen as successful in combating Hepatitis B,Polio,Diphtheria and health issues due to alcohol, we need an improvement in the areas like Measles, under five deaths, HIV/AIDS, thinness and population contro

Filling the missing values

In [178…
```
Q.isna().sum()
```

Out[178]:
```
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 [179…
```
Q['Life expectancy'].fillna(value=Q['Life expectancy'].mean(),inplace=True)
```

In [180…
```
Q['Adult Mortality'].fillna(value=Q['Adult Mortality'].mean(),inplace=True)
```

In [181…
```
Q['Alcohol'].fillna(value=Q['Alcohol'].mean(),inplace=True)
```

In [182…
```
Q['Hepatitis B'].fillna(value=Q['Hepatitis B'].mean(),inplace=True)
```

In [183…
```
Q['BMI'].fillna(value=Q['BMI'].mean(),inplace=True)
```

In [184…
```
Q['Polio'].fillna(value=Q['Polio'].mean(),inplace=True)
```

In [185…
```
Q['Total expenditure'].fillna(value=Q['Total expenditure'].mean(),inplace=True)
```

In [186…
```
Q['Diphtheria'].fillna(value=Q['Diphtheria'].mean(),inplace=True)
```

In [187…
```
Q['GDP'].fillna(value=Q['GDP'].mean(),inplace=True)
```

In [188…
```
Q['Population'].fillna(value=Q['Population'].mean(),inplace=True)
```

In [189…
```
Q['thinness  1-19 years'].fillna(value=Q['thinness  1-19 years'].mean(),inplace=Tru
```

In [190…
```
Q['thinness 5-9 years'].fillna(value=Q['thinness 5-9 years'].mean(),inplace=True)
```

In [191…
```
Q['Income composition of resources'].fillna(value=Q['Income composition of resourc
```

In [192…
```
Q['Schooling'].fillna(value=Q['Schooling'].mean(),inplace=True)
```

In [193…
```
Q.isna().sum()
```

Out[193]:
```
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
```

We have successfully filled all the null values

ML model deployment

In [194…
```
Q.head()
```

Out[194]:

	Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B
0	Afghanistan	2015	Developing	65.0	263.0	62	0.01	71.279624	65.0
1	Afghanistan	2014	Developing	59.9	271.0	64	0.01	73.523582	62.0
2	Afghanistan	2013	Developing	59.9	268.0	66	0.01	73.219243	64.0
3	Afghanistan	2012	Developing	59.5	272.0	69	0.01	78.184215	67.0
4	Afghanistan	2011	Developing	59.2	275.0	71	0.01	7.097109	68.0

5 rows × 22 columns

Assigning the dependent and independent variable

In [196…

```
X=Q.iloc[:,[4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19]].values
y=Q.iloc[:,[3]].values
```

Splitting the dataset

In [197…

```
from sklearn.model_selection import train_test_split
X_train,X_test,y_train,y_test=train_test_split(X,y,random_state=0)
```

Using Linear Regression model

In [198…

```
from sklearn.linear_model import LinearRegression
lr=LinearRegression()
```

In [199…

```
lr.fit(X_train,y_train)
```

Out[199]:

```
LinearRegression()
```

In [200…

```
y_pred=lr.predict(X_test)
```

In [201…

```
from sklearn import metrics
print("RMSE=",np.sqrt(metrics.mean_squared_error(y_test,y_pred)))
print("R^2=",metrics.r2_score(y_test,y_pred))
```

```
RMSE= 4.780029511651427
R^2= 0.7562021964220327
```

Using Random Forest Regressor model

In [202…

```
from sklearn.ensemble import RandomForestRegressor
rgs=RandomForestRegressor(n_estimators=300,criterion='squared_error',random_state=(
rgs.fit(X_train,y_train)
```

Out[202]:
```
RandomForestRegressor(n_estimators=300, random_state=0)
```

In [205…
```
y_pre=rgs.predict(X_test)
```

In [206…
```
print("RMSE=",np.sqrt(metrics.mean_squared_error(y_test,y_pre)))
print("R^2=",metrics.r2_score(y_test,y_pre))
```

```
RMSE= 2.136543091412308
R^2= 0.9512929366737587
```

Conclusion

Linear Regression model gives 75.62% accuracy whereas Random Forest Regressor model gives us 95.14% accuracy which i consider excellent and fit to use for further predictions

In [233…
```
0.7562021964220327+0.9512929366737587
```

Out[233]:
```
1.7074951330957915
```

In [234…
```
1.7074951330957915/2
```

Out[234]:
```
0.8537475665478957
```

In [235…
```

```

In [ ]:
```

```