In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
```

In [2]:

```
df = pd.read_csv("world.csv").dropna()

df
```

Out[2]:

	Country	Density\n(P/Km2)	Abbreviation	Agricultural Land(%)	Land Area(Km2)	Armed Forces size	Birth Rate	Calling Code	Ca
0	Afghanistan	60	AF	58.10%	652,230	323,000	32.49	93.0	
1	Albania	105	AL	43.10%	28,748	9,000	11.78	355.0	
2	Algeria	18	DZ	17.40%	2,381,741	317,000	24.28	213.0	
4	Angola	26	AO	47.50%	1,246,700	117,000	40.73	244.0	
6	Argentina	17	AR	54.30%	2,780,400	105,000	17.02	54.0	E
...	...	...	...	...	...	...	...	...	
185	United Kingdom	281	GB	71.70%	243,610	148,000	11.00	44.0	
186	United States	36	US	44.40%	9,833,517	1,359,000	11.60	1.0	\
187	Uruguay	20	UY	82.60%	176,215	22,000	13.86	598.0	
191	Vietnam	314	VN	39.30%	331,210	522,000	16.75	84.0	
193	Zambia	25	ZM	32.10%	752,618	16,000	36.19	260.0	

110 rows × 35 columns

In [3]:

```
df.head()
```

Out[3]:

	Country	Density\n(P/Km2)	Abbreviation	Agricultural Land(%)	Land Area(Km2)	Armed Forces size	Birth Rate	Calling Code	Capital
0	Afghanistan	60	AF	58.10%	652,230	323,000	32.49	93.0	
1	Albania	105	AL	43.10%	28,748	9,000	11.78	355.0	
2	Algeria	18	DZ	17.40%	2,381,741	317,000	24.28	213.0	
4	Angola	26	AO	47.50%	1,246,700	117,000	40.73	244.0	

	Country	Density\n(P/Km2)	Abbreviation	Agricultural Land(%)	Land Area(Km2)	Armed Forces size	Birth Rate	Calling Code	Capital
6	Argentina	17	AR	54.30%	2,780,400	105,000	17.02	54.0	Buen

5 rows × 35 columns

Data cleaning and pre processing

In [4]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 110 entries, 0 to 193
Data columns (total 35 columns):
 #   Column                                     Non-Null Count  Dtype  
---  ------                                     --------------  -----  
 0   Country                                    110 non-null    object 
 1   Density
(P/Km2)                          110 non-null    object 
 2   Abbreviation                               110 non-null    object 
 3   Agricultural Land( %)                      110 non-null    object 
 4   Land Area(Km2)                             110 non-null    object 
 5   Armed Forces size                          110 non-null    object 
 6   Birth Rate                                 110 non-null    float64
 7   Calling Code                               110 non-null    float64
 8   Capital/Major City                         110 non-null    object 
 9   Co2-Emissions                              110 non-null    object 
 10  CPI                                        110 non-null    object 
 11  CPI Change (%)                             110 non-null    object 
 12  Currency-Code                              110 non-null    object 
 13  Fertility Rate                             110 non-null    float64
 14  Forested Area (%)                          110 non-null    object 
 15  Gasoline Price                             110 non-null    object 
 16  GDP                                        110 non-null    object 
 17  Gross primary education enrollment (%)     110 non-null    object 
 18  Gross tertiary education enrollment (%)    110 non-null    object 
 19  Infant mortality                           110 non-null    float64
 20  Largest city                               110 non-null    object 
 21  Life expectancy                            110 non-null    float64
 22  Maternal mortality ratio                   110 non-null    float64
 23  Minimum wage                               110 non-null    object 
 24  Official language                          110 non-null    object 
 25  Out of pocket health expenditure           110 non-null    object 
 26  Physicians per thousand                    110 non-null    float64
 27  Population                                 110 non-null    object 
 28  Population: Labor force participation (%)  110 non-null    object 
 29  Tax revenue (%)                            110 non-null    object 
 30  Total tax rate                             110 non-null    object 
 31  Unemployment rate                          110 non-null    object 
 32  Urban_population                           110 non-null    object 
 33  Latitude                                   110 non-null    float64
 34  Longitude                                  110 non-null    float64
dtypes: float64(9), object(26)
memory usage: 30.9+ KB
```

In [5]:

```
df.describe()
```

Out[5]:

	Birth Rate	Calling Code	Fertility Rate	Infant mortality	Life expectancy	Maternal mortality ratio	Physicians per thousand	Latitude
count	110.000000	110.000000	110.000000	110.000000	110.000000	110.000000	110.000000	110.00000
mean	20.196455	344.290909	2.672182	20.271818	72.671818	137.227273	1.919182	20.36267
std	10.039056	341.231562	1.308142	18.453214	7.000788	201.171462	1.598116	24.43214
min	6.400000	1.000000	0.980000	1.700000	54.300000	2.000000	0.010000	-40.90055
25%	11.075000	70.000000	1.682500	6.100000	67.625000	15.250000	0.467500	7.62325
50%	17.830000	239.500000	2.200000	13.600000	74.400000	41.000000	1.640000	21.03360
75%	27.962500	420.750000	3.505000	31.500000	77.350000	176.000000	3.007500	40.12460
max	46.080000	1876.000000	6.910000	78.500000	83.300000	1120.000000	7.120000	61.52401

In [6]:

```
df.columns
```

Out[6]:

```
Index(['Country', 'Density\n(P/Km2)', 'Abbreviation', 'Agricultural Land( %)',
       'Land Area(Km2)', 'Armed Forces size', 'Birth Rate', 'Calling Code',
       'Capital/Major City', 'Co2-Emissions', 'CPI', 'CPI Change (%)',
       'Currency-Code', 'Fertility Rate', 'Forested Area (%)',
       'Gasoline Price', 'GDP', 'Gross primary education enrollment (%)',
       'Gross tertiary education enrollment (%)', 'Infant mortality',
       'Largest city', 'Life expectancy', 'Maternal mortality ratio',
       'Minimum wage', 'Official language', 'Out of pocket health expenditure',
       'Physicians per thousand', 'Population',
       'Population: Labor force participation (%)', 'Tax revenue (%)',
       'Total tax rate', 'Unemployment rate', 'Urban_population', 'Latitude',
       'Longitude'],
      dtype='object')
```

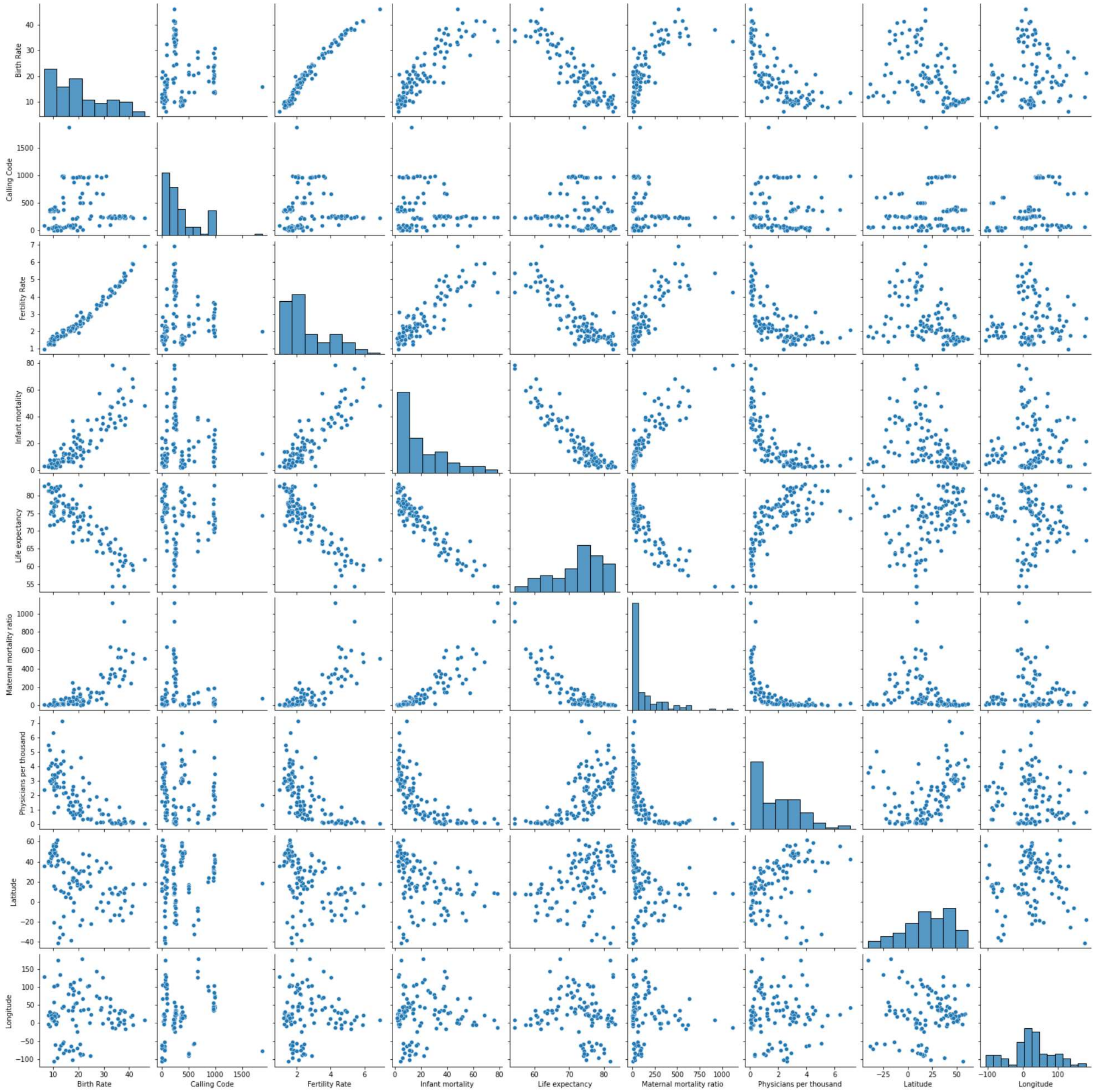
EDA and VISUALIZATION

In [7]:

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

Out[7]:

```
<seaborn.axisgrid.PairGrid at 0x1fadabb5970>
```

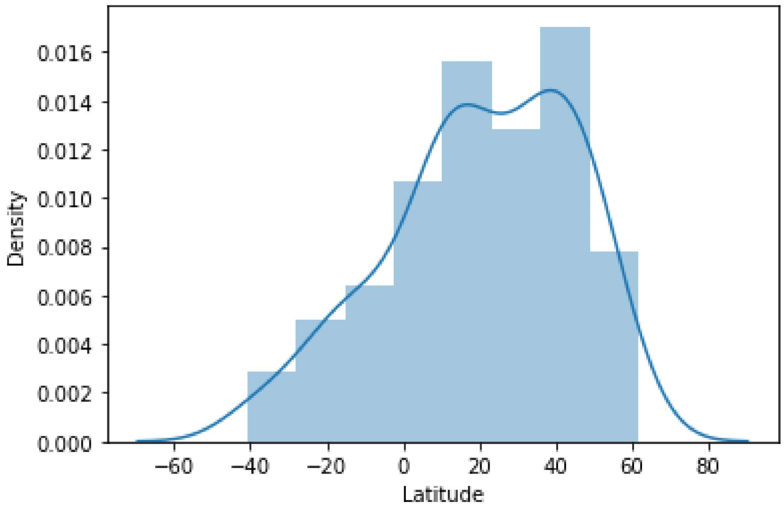


In [8]:

```
sns.distplot(df['Latitude'])
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning:
`distplot` is a deprecated function and will be removed in a future version. Please adap
t your code to use either `displot` (a figure-level function with similar flexibility) o
r `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```

Out[8]: <AxesSubplot:xlabel='Latitude', ylabel='Density'>



In [9]:
```
df1 = df[['Country', 'Density\n(P/Km2)', 'Abbreviation', 'Agricultural Land( %)',
        'Land Area(Km2)', 'Armed Forces size', 'Birth Rate', 'Calling Code',
        'Capital/Major City', 'Co2-Emissions', 'CPI', 'CPI Change (%)',
        'Currency-Code', 'Fertility Rate', 'Forested Area (%)',
        'Gasoline Price', 'GDP', 'Gross primary education enrollment (%)',
        'Gross tertiary education enrollment (%)', 'Infant mortality',
        'Largest city', 'Life expectancy', 'Maternal mortality ratio',
        'Minimum wage', 'Official language', 'Out of pocket health expenditure',
        'Physicians per thousand', 'Population',
        'Population: Labor force participation (%)', 'Tax revenue (%)',
        'Total tax rate', 'Unemployment rate', 'Urban_population', 'Latitude',
        'Longitude']]
```

In [10]:
```
sns.heatmap(df1.corr())
```

Out[10]: <AxesSubplot:>

In [11]:
```
x = df1[[ 'Birth Rate','Calling Code','Fertility Rate','Infant mortality','Life expecta
y = df1[ 'Physicians per thousand']
```

split the data into training and test data

In [12]:
```
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.3)
```

In [13]:
```
lr = LinearRegression()
lr.fit(x_train, y_train)
```

Out[13]:
```
LinearRegression()
```

In [14]:
```
lr.intercept_
```

Out[14]:
```
-1.5987211554602254e-14
```

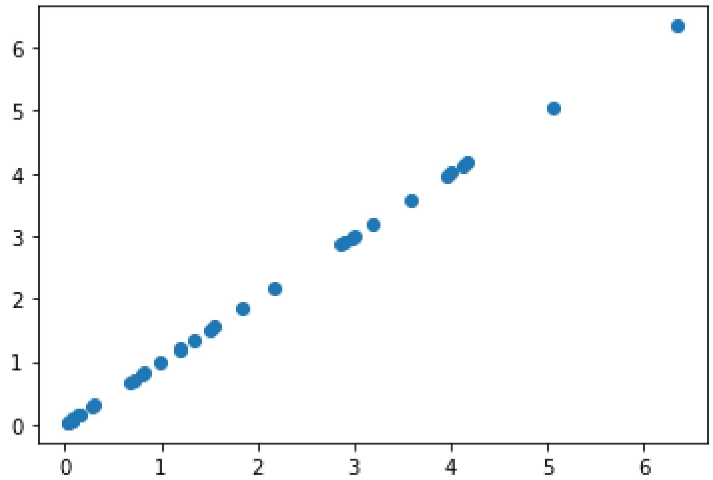
In [15]:
```
coeff = pd.DataFrame(lr.coef_, x.columns, columns =['Fertility Rate'])
coeff
```

Out[15]:

	Fertility Rate
Birth Rate	1.399787e-16
Calling Code	4.654202e-18
Fertility Rate	-1.709987e-15
Infant mortality	1.292316e-16
Life expectancy	2.058709e-16
Maternal mortality ratio	7.166756e-18
Physicians per thousand	1.000000e+00
Latitude	-5.380168e-17
Longitude	-1.795956e-17

In [16]:
```
prediction = lr.predict(x_test)
plt.scatter(y_test, prediction)
```

Out[16]:
```
<matplotlib.collections.PathCollection at 0x1fadf32e160>
```



In [17]:
```
lr.score(x_test,y_test)
```
Out[17]: 1.0

ACURACY

In [18]:
```
from sklearn.linear_model import Ridge,Lasso
```

In [19]:
```
rr=Ridge(alpha=10)
rr.fit(x_train,y_train)
rr.score(x_test,y_test)
rr.score(x_train,y_train)
```
Out[19]: 0.9939179744721426

In [20]:
```
rr.score(x_test,y_test)
```
Out[20]: 0.9896887417510037

In [21]:
```
la = Lasso(alpha=10)
la.fit(x_train,y_train)
```
Out[21]: Lasso(alpha=10)

In [22]:
```
la.score(x_test,y_test)
```
Out[22]: 0.314710323530639