

## ✓ Hands On Activity 11.1 Linear Regression Analysis

**Name:** Calingo, Christian Lei S.

**Section:** CPE22S3

**Course:** Computational Thinking with Python

**Course Code:** CPE311

```
!pip install hvplot
```

```
Collecting hvplot
  Downloading hvplot-0.9.2-py2.py3-none-any.whl (1.8 MB)
    1.8/1.8 MB 10.0 MB/s eta 0:00:00
Requirement already satisfied: bokeh>=1.0.0 in /usr/local/lib/python3.10/dist-packages (from hvplot) (3.3.4)
Requirement already satisfied: colorcet>=2 in /usr/local/lib/python3.10/dist-packages (from hvplot) (3.1.0)
Requirement already satisfied: holoviews>=1.11.0 in /usr/local/lib/python3.10/dist-packages (from hvplot) (1.17.1)
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages (from hvplot) (2.0.3)
Requirement already satisfied: numpy>=1.15 in /usr/local/lib/python3.10/dist-packages (from hvplot) (1.25.2)
Requirement already satisfied: packaging in /usr/local/lib/python3.10/dist-packages (from hvplot) (24.0)
Requirement already satisfied: panel>=0.11.0 in /usr/local/lib/python3.10/dist-packages (from hvplot) (1.3.8)
Requirement already satisfied: param<3.0,>=1.12.0 in /usr/local/lib/python3.10/dist-packages (from hvplot) (2.1.0)
Requirement already satisfied: Jinja2>=2.9 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (3.1.3)
Requirement already satisfied: contourpy>=1 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (1.2.1)
Requirement already satisfied: pillow>=7.1.0 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (9.4.0)
Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (6.0.1)
Requirement already satisfied: tornado>=5.1 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (6.3.3)
Requirement already satisfied: xyzservices>=2021.09.1 in /usr/local/lib/python3.10/dist-packages (from bokeh>=1.0.0->hvplot) (2024.4.0)
Requirement already satisfied: pyviz-comms>=0.7.4 in /usr/local/lib/python3.10/dist-packages (from holoviews>=1.11.0->hvplot) (3.0.2)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.10/dist-packages (from pandas->hvplot) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages (from pandas->hvplot) (2023.4)
Requirement already satisfied: tzdata>=2022.1 in /usr/local/lib/python3.10/dist-packages (from pandas->hvplot) (2024.1)
Requirement already satisfied: markdown in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (3.6)
Requirement already satisfied: markdown-it-py in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (3.0.0)
Requirement already satisfied: linkify-it-py in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (2.0.3)
Requirement already satisfied: mdit-py-plugins in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (0.4.0)
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (2.31.0)
Requirement already satisfied: tqdm>=4.48.0 in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (4.66.2)
Requirement already satisfied: bleach in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (6.1.0)
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.10/dist-packages (from panel>=0.11.0->hvplot) (4.11.0)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages (from Jinja2>=2.9->bokeh>=1.0.0->hvplot) (2.1.
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) (2.0.1)
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
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages (from requests->panel>=0.11.0->hvplot) (202
Installing collected packages: hvplot
Successfully installed hvplot-0.9.2
```

```
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
```

### Importing the dataset

```
data = pd.read_csv('data/Life Expectancy Data.csv')
#data.set_index(['Country', 'Year', 'Status'], inplace = True)
data.drop(columns = ['Status'], inplace = True)
data
```

	Country	Year	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	BMI	under-five deaths	Polio	Total expenditure	Diphtheria
0	Afghanistan	2015	65.0	263.0	62	0.01	71.279624	65.0	1154	19.1	83	6.0	8.16	6.0
1	Afghanistan	2014	59.9	271.0	64	0.01	73.523582	62.0	492	18.6	86	58.0	8.18	6.0
2	Afghanistan	2013	59.9	268.0	66	0.01	73.219243	64.0	430	18.1	89	62.0	8.13	6.0
3	Afghanistan	2012	59.5	272.0	69	0.01	78.184215	67.0	2787	17.6	93	67.0	8.52	6.0
4	Afghanistan	2011	59.2	275.0	71	0.01	7.097109	68.0	3013	17.2	97	68.0	7.87	6.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
2933	Zimbabwe	2004	44.3	723.0	27	4.36	0.000000	68.0	31	27.1	42	67.0	7.13	6.0
2934	Zimbabwe	2003	44.5	715.0	26	4.06	0.000000	7.0	998	26.7	41	7.0	6.52	6.0
2935	Zimbabwe	2002	44.8	73.0	25	4.43	0.000000	73.0	304	26.3	40	73.0	6.53	7.0
2936	Zimbabwe	2001	45.3	686.0	25	1.72	0.000000	76.0	529	25.9	39	76.0	6.16	7.0
2937	Zimbabwe	2000	46.0	665.0	24	1.68	0.000000	79.0	1483	25.5	39	78.0	7.10	7.0

2938 rows × 21 columns

Average of each column per country

```
data1 = data.drop(columns = ['Year'], inplace = True)
data1 = data.groupby('Country').mean()
data1
```

	Country	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	BMI	under-five deaths	Polio	Total expenditure	Diphtheria
	Afghanistan	58.19375	269.0625	78.2500	0.014375	34.960110	64.562500	2362.2500	15.51875	107.5625	48.3750	8.252500	52.31
	Albania	75.15625	45.0625	0.6875	4.848750	193.259091	98.000000	53.3750	49.06875	0.9375	98.1250	5.945625	98.06
	Algeria	73.61875	108.1875	20.3125	0.406667	236.185241	78.000000	1943.8750	48.74375	23.5000	91.7500	4.604000	91.87
	Angola	49.01875	328.5625	83.7500	5.740667	102.100268	70.222222	3561.3125	18.01875	132.6250	46.1250	3.919333	47.68
	Antigua and Barbuda	75.05625	127.5000	0.0000	7.949333	1001.585226	98.266667	0.0000	38.42500	0.0000	96.9375	4.791333	98.31
	...	...	...	...	...	...	...	...	...	...	...	...	...
	Venezuela (Bolivarian Republic of)	73.38750	163.0000	9.3750	7.420000	0.000000	66.250000	165.0000	54.48750	10.7500	74.6875	4.998667	68.50
	Viet Nam	74.77500	126.5625	29.1875	3.087333	0.000000	87.538462	4232.9375	11.18750	36.5000	94.9375	5.977333	91.75
	Yemen	63.86250	211.8125	39.3750	0.047333	0.000000	55.687500	2761.1875	33.48750	51.6250	67.1250	5.005333	72.62
	Zambia	53.90625	354.3125	33.4375	2.239333	89.650407	69.818182	6563.8125	17.45000	52.3750	64.3750	5.824000	74.25
	Zimbabwe	50.48750	462.3750	26.5625	4.482000	20.364271	70.562500	923.0000	25.13750	40.8125	75.6250	6.158667	75.18

193 rows × 19 columns

Next steps: [View recommended plots](#)

Checking for NaN Values in DataFrame

```
new_data = data.groupby('Country').apply(lambda group: group.fillna(group.mean()))
new_data.info()

<class 'pandas.core.frame.DataFrame'>
MultiIndex: 2938 entries, ('Afghanistan', 0) to ('Zimbabwe', 2937)
Data columns (total 19 columns):
#   Column                                Non-Null Count  Dtype
#   ...
```

```

---  -----
0   Life expectancy          2928 non-null  float64
1   Adult Mortality          2928 non-null  float64
2   infant deaths            2938 non-null  int64
3   Alcohol                  2921 non-null  float64
4   percentage expenditure    2938 non-null  float64
5   Hepatitis B              2794 non-null  float64
6   Measles                  2938 non-null  int64
7   BMI                      2904 non-null  float64
8   under-five deaths        2938 non-null  int64
9   Polio                    2938 non-null  float64
10  Total expenditure        2906 non-null  float64
11  Diphtheria               2938 non-null  float64
12  HIV/AIDS                 2938 non-null  float64
13  GDP                      2533 non-null  float64
14  Population               2290 non-null  float64
15  thinness 1-19 years      2904 non-null  float64
16  thinness 5-9 years       2904 non-null  float64
17  Income composition of resources 2771 non-null  float64
18  Schooling                2775 non-null  float64
dtypes: float64(16), int64(3)
memory usage: 540.7+ KB

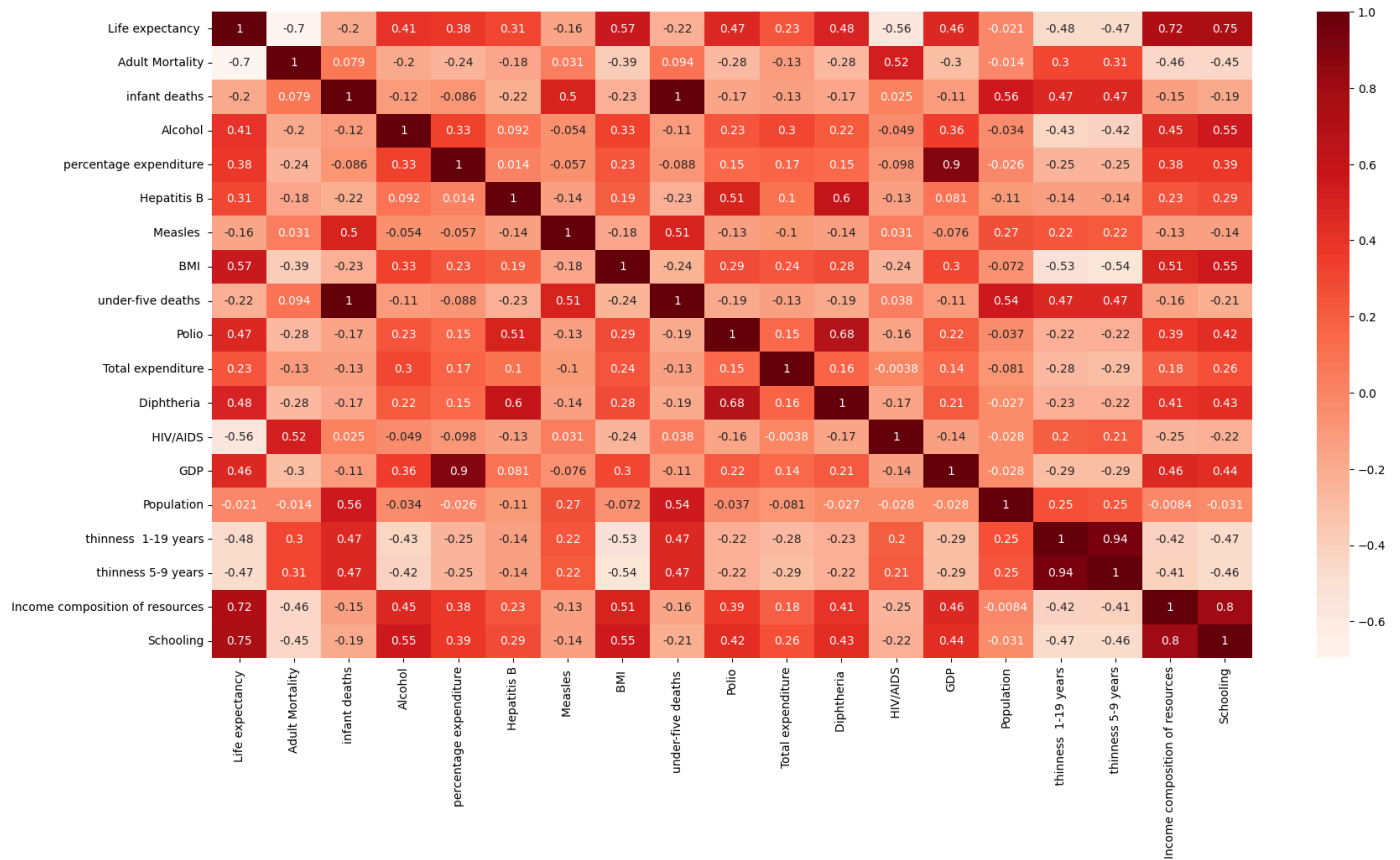
```

### Checking for the correlation of each columns

```

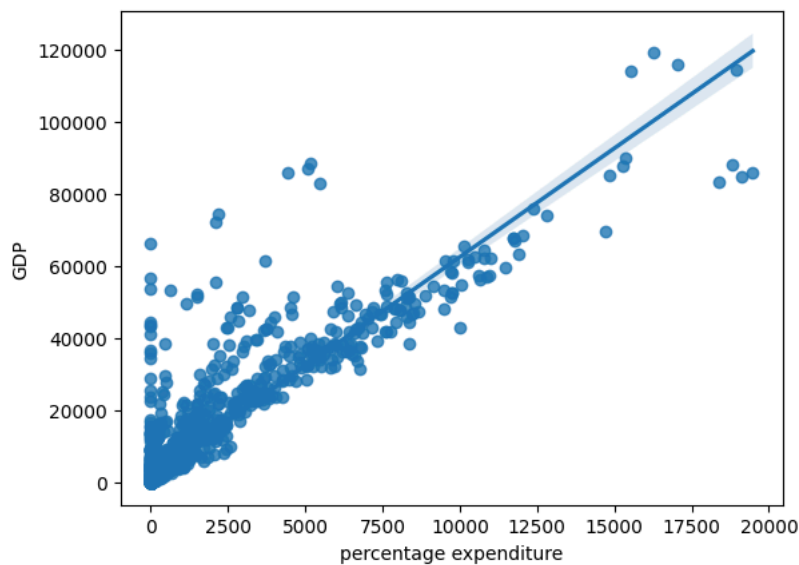
plt.figure(figsize=(20,10))
ax = sns.heatmap(new_data.corr(), annot = True, cmap = 'Reds')

```



```
sns.regplot(data = new_data, x = 'percentage expenditure', y = 'GDP')
```

<Axes: xlabel='percentage expenditure', ylabel='GDP'>



```
sns.regplot(data = new_data, x = 'Income composition of resources', y = 'Schooling')
```