In [388]:

```
import pandas as pd
import scipy
import matplotlib.pyplot as plt
import seaborn as sb
import numpy as np
import pprint
```

In [389]:

```
# Processed CSV File
TEST_CSV = "/mnt/c/Users/User/Documents/GitHub/Last-Mile-206/datasets/consolid
ated.csv"
```

In [390]:

```
# Reading as dataframe
df = pd.read_csv(TEST_CSV)
df.head()
```

Out[390]:

	Unnamed: 0	Unnamed: 0_x	Country	Country Code	Indicator Name	Indicator Code	1980	1981
0	0	1.0	Afghanistan	AFG	CO2 emissions (metric tons per capita)	EN.ATM.CO2E.PC	0.131783	0.150615
1	1	3.0	Albania	ALB	CO2 emissions (metric tons per capita)	EN.ATM.CO2E.PC	1.935058	2.693024
2	2	58.0	Algeria	DZA	CO2 emissions (metric tons per capita)	EN.ATM.CO2E.PC	3.460646	2.342523
3	3	2.0	Angola	AGO	CO2 emissions (metric tons per capita)	EN.ATM.CO2E.PC	0.640966	0.611135
4	4	10.0	Antigua and Barbuda	ATG	CO2 emissions (metric tons per capita)	EN.ATM.CO2E.PC	2.311695	1.720761

In [391]:

```
# Filling NAN values and other unsupported types
df = df.fillna(0)
df = df.replace(to_replace = '--', value = 0)
```

In [392]:

```
# Using the Carbon Emissions by country (kilo tonnes per capita), the top 16 offenders were chosen
country_codes = [137, 88, 11, 183, 144, 97, 185, 30, 8, 126, 83, 179, 140, 86, 149, 119]
print(len(country_codes))
```

```
16
```

In [393]:

```
# Building a list of facotrs (average) that might be a direct effect of CO2 emissions
co2_emissions = []
for itr in country_codes:
    # CO2 Emissions by country
    co2_country = df.loc[itr, '1980':'2016']
    sum_co2_country = co2_country.sum()
    avg_co2_country = sum_co2_country/len(co2_country)
    # GDP Const
    sum_gdp_const_country = 0.0
    gdp_const_country = df.loc[itr, 'gdp_const_1980':'gdp_const_2016']
    for seriesData in gdp_const_country:
        sum_gdp_const_country += float(str(seriesData).replace(',',''))
    avg_gdp_const_country = sum_gdp_const_country/len(gdp_const_country)
    # GDP Curr
    sum_gdp_curr_country = 0.0
    gdp_curr_country = df.loc[itr, 'gdp_curr_1980':'gdp_curr_2016']
    for seriesData in gdp_curr_country:
        sum_gdp_curr_country += float(str(seriesData).replace(',',''))
    avg_gdp_curr_country = sum_gdp_curr_country/len(gdp_curr_country)
    # Average Inflation
    sum_infl_avg_prices_country = 0.0
    infl_avg_prices_country = df.loc[itr, 'infl_avg_prices_1980':'infl_avg_prices_2016']
    for seriesData in infl_avg_prices_country:
        sum_infl_avg_prices_country += float(str(seriesData).replace(',',''))
    avg_infl_avg_prices_country = sum_infl_avg_prices_country/len(infl_avg_prices_country)
    # Unemployment
    sum_unemployment_country = 0.0
    unemployment_country = df.loc[itr, 'unemployment_1980':'unemployment_2016']
    for seriesData in unemployment_country:
        sum_unemployment_country += float(str(seriesData).replace(',',''))
    avg_unemployment_country = sum_unemployment_country/len(unemployment_country)
    # Acc Balance
    sum_acc_bal_country = 0.0
    acc_bal_country = df.loc[itr, 'acc_bal_1980':'acc_bal_2016']
    for seriesData in acc_bal_country:
        sum_acc_bal_country += float(str(seriesData).replace(',',''))
    avg_acc_bal_country = sum_acc_bal_country/len(acc_bal_country)
    # Consolidating the results
    co2_emissions.append([itr, avg_co2_country, avg_gdp_const_country, avg_gdp_curr_country, avg_infl_avg_prices_country, avg_unemployment_country, avg_acc_bal_country])
```

In [394]:
```
# Converting the list to a Dataframe
co2_emissions_df = pd.DataFrame(co2_emissions, columns = ['country_code', 'avg
_co2', 'avg_gdp_const', 'avg_gdp_curr', 'avg_infl_prices', 'avg_unemployment',
'avg_acc_bal'])
```

In [395]:
```
print(co2_emissions_df)
```

```
    country_code     avg_co2  avg_gdp_const  avg_gdp_curr  avg_infl_prices  \
0            137   47.969597       5.951595     93.792135         3.749595   
1             88   22.272511       3.513514    134.247108         3.453108   
2             11   24.237600       4.265514     27.222054         1.810270   
3            183   27.898241       3.701000    277.183108         4.161865   
4            144   15.422268       2.258973    788.697189         1.234973   
5             97   22.413107       4.108216     27.218973         2.939081   
6            185   18.776334       2.644243   9767.411351         3.313054   
7             30   16.327194       2.383108    890.925946         3.239243   
8              8   16.297615       3.172973    568.144324         4.227135   
9            126    9.823519       5.409649     78.568811         2.006081   
10            83    8.179348       2.454595    154.352568        93.487595   
11           179    6.631358       4.899162     23.141486       191.570459   
12           140    0.000000       1.142649   1668.583243        47.359459   
13            86    0.000000       6.378676    822.673676         5.025946   
14           149   11.441580       6.608108    187.938459         2.080243   
15           119   10.816186       2.068676    494.419757         2.024649   

    avg_unemployment  avg_acc_bal  
0           0.000000    21.260892  
1           1.308270    19.591919  
2           1.076135     2.001243  
3           0.000000    10.251324  
4           2.648892     3.521514  
5           3.221676     5.015514  
6           6.379730    -2.536541  
7           8.380649    -1.415784  
8           6.958324    -4.243622  
9           0.000000     2.744757  
10          5.239405    -2.728189  
11          0.000000    -2.074108  
12          5.212324     3.389595  
13          3.549054     0.816378  
14          2.882811    11.981324  
15          5.591027     4.714838  
```

In [396]:

```
# Finding the Pearson Correlation
pearsoncorr = co2_emissions_df.corr(method='pearson')
print(pearsoncorr)
pearsoncorr.to_csv('/mnt/c/Users/User/Documents/GitHub/Last-Mile-206/datasets/
pearsoncorr.csv')
```

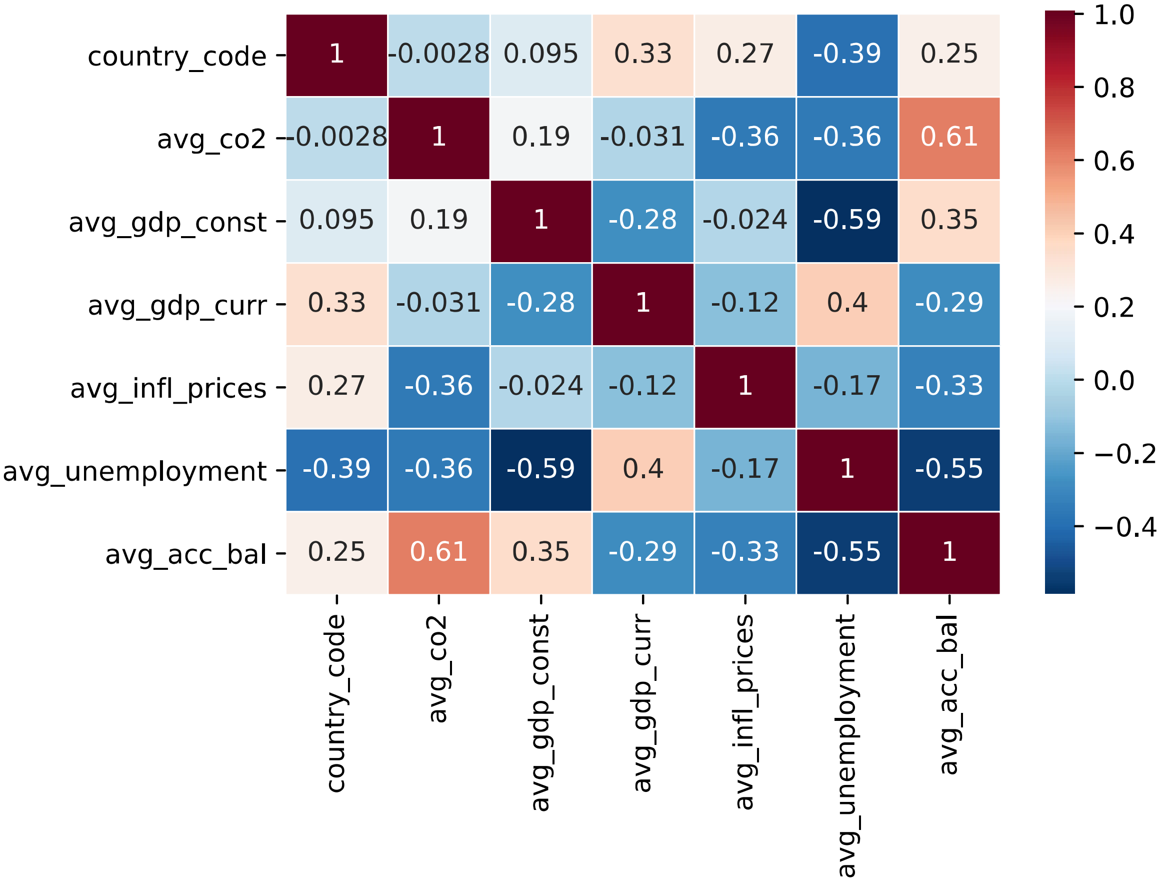
```
                  country_code   avg_co2  avg_gdp_const  avg_gdp_curr  \
country_code          1.000000 -0.002825       0.095280      0.334774   
avg_co2              -0.002825  1.000000       0.192756     -0.031103   
avg_gdp_const         0.095280  0.192756       1.000000     -0.283765   
avg_gdp_curr          0.334774 -0.031103      -0.283765      1.000000   
avg_infl_prices       0.267732 -0.357737      -0.024219     -0.124653   
avg_unemployment     -0.385708 -0.355211      -0.587706      0.399230   
avg_acc_bal           0.251134  0.614562       0.353589     -0.291355   

                  avg_infl_prices  avg_unemployment  avg_acc_bal  
country_code             0.267732         -0.385708     0.251134  
avg_co2                 -0.357737         -0.355211     0.614562  
avg_gdp_const           -0.024219         -0.587706     0.353589  
avg_gdp_curr            -0.124653          0.399230    -0.291355  
avg_infl_prices          1.000000         -0.165095    -0.331369  
avg_unemployment        -0.165095          1.000000    -0.548674  
avg_acc_bal             -0.331369         -0.548674     1.000000  
```

In [397]:

```
# Plotting the heatmap using Seaborn
sb.heatmap(pearsoncorr,
            xticklabels=pearsoncorr.columns,
            yticklabels=pearsoncorr.columns,
            cmap='RdBu_r',
            annot=True,
            linewidth=0.5)
```

Out[397]: <AxesSubplot:>



	country_code	avg_co2	avg_gdp_const	avg_gdp_curr	avg_infl_prices	avg_unemployment	avg_acc_bal
country_code	1	-0.0028	0.095	0.33	0.27	-0.39	0.25
avg_co2	-0.0028	1	0.19	-0.031	-0.36	-0.36	0.61
avg_gdp_const	0.095	0.19	1	-0.28	-0.024	-0.59	0.35
avg_gdp_curr	0.33	-0.031	-0.28	1	-0.12	0.4	-0.29
avg_infl_prices	0.27	-0.36	-0.024	-0.12	1	-0.17	-0.33
avg_unemployment	-0.39	-0.36	-0.59	0.4	-0.17	1	-0.55
avg_acc_bal	0.25	0.61	0.35	-0.29	-0.33	-0.55	1

In [ ]: