Foreign Aid Effect

An Analysis

Hima Gharat, Jessica Johnson, Rejane Beringer, Salita Epifania

- This project aims to explore if there is some significant effect of financial aid given to some countries in their Sustainability Indexes, such as Carbon dioxide average emissions, Employment Rate and Gross Domestic Product(GDP).

- Data: Countries that received Foreign Aid Assistance from 2011 to 2016.

  databank.worldbank.org

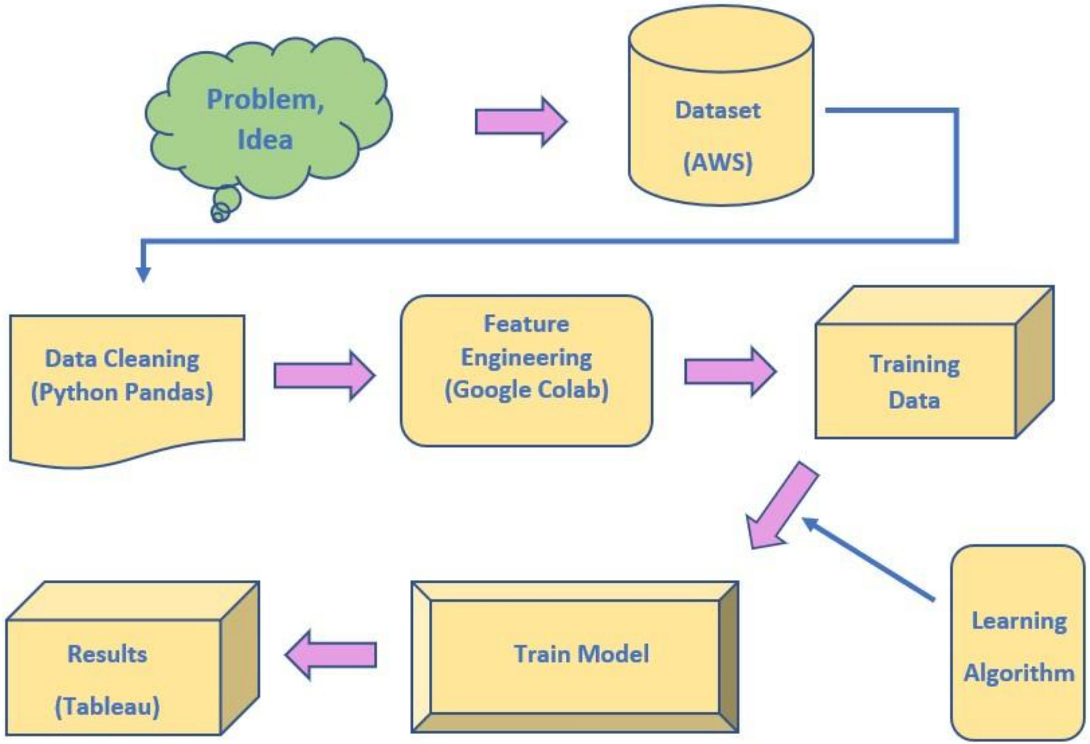
Questions:

- Is there a relationship between Assistance provided and CO2 reduction?

- Is there a relationship between Assistance provided and Employment increase?

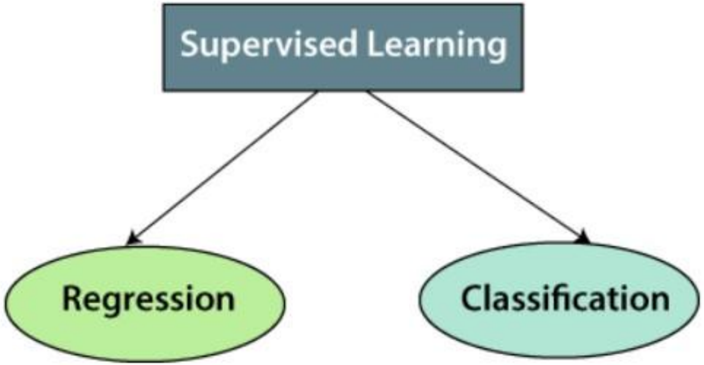
- Is there a relationship between Assistance provided and GDP increase?

Machine Learning



Regression Analysis

- Supervised Learning
- Continuous data set
- Prediction model
- Utilizing historic data
- Trends
- Numeric Results



Linear Regression

- Statistical technique
- Model the relationship between two sets of variables
- Problem
- Sklearn

```
]  # importing the libraries
   import pandas as pd
   import numpy as np
   import matplotlib.pyplot as plt
   import seaborn as sns
   import os
   import warnings
   from sklearn.linear_model import LinearRegression
   from sklearn.preprocessing import StandardScaler
   from sklearn.model_selection import train_test_split
   from sklearn.metrics import r2_score, mean_squared_error

   warnings.filterwarnings('ignore')
   RANDOM_SEED = 42
```

Preparation

```
]: assistance_csv = "Resources/Assistance Received by Country.csv"
   emission_csv = "Resources/CO2 Emission by Country.csv"
   employment_csv = "Resources/Employment by Country.csv"
   gdp_csv = "Resources/GDP by Country.csv"
```

```
: merge3_df.columns

: Index(['Country', 'Assitance 2011', 'Assitance 2012', 'Assitance 2013',
         'Assitance 2014', 'Assitance 2015', 'Assitance 2016', 'CO2 2011',
         'CO2  2012', 'CO2  2013', 'CO2  2014', 'CO2  2015', 'CO2 2016',
         'Employment 2011', 'Employment 2012', 'Employment 2013',
         'Employment 2014', 'Employment 2015', 'Employment 2016', 'GDP 2011',
         'GDP 2012', 'GDP 2013', 'GDP 2014', 'GDP 2015', 'GDP 2016'],
        dtype='object')
```

```
df=df.iloc[:, 1:]
df=df.set_index('Country')
df = df.apply(pd.to_numeric, errors='coerce')
df.columns = [i.replace('  ', ' ') for i in df.columns]
```

```
]: attributes=['Assistance', 'CO2', 'Employment', 'GDP']

   countries=[]

   for attribute in attributes:
       countries.append(df[[f'{attribute} 2011', f'{attribute} 2012', f'{attribute} 2013',
          f'{attribute} 2014', f'{attribute} 2015', f'{attribute} 2016']].mean(axis=1))
```

```
]: result = pd.DataFrame([countries[0], countries[1],countries[2], countries[3]]).T
   result.columns=['Mean Assistance (2011-2016)', 'Mean CO2 (2011-2016)', 'Mean Employment (2011-2016)', 'Mean GDP (2011-2016)']
```

```
]: result
```

]:

	Mean Assistance (2011-2016)	Mean CO2 (2011-2016)	Mean Employment (2011-2016)	Mean GDP (2011-2016)
Country				
Afghanistan	5.308612e+09	9703.495000	16.351667	1961.268333

Cleaning

```
# checking for missing data
data.isnull().sum()
```

```
Unnamed: 0                    0
Country                       0
Mean Assistance (2011-2016)   0
Mean CO2 (2011-2016)          0
Mean Employment (2011-2016)   0
Mean GDP (2011-2016)          0
dtype: int64
```

```
# checking data type
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 195 entries, 0 to 136
Data columns (total 6 columns):
 #   Column                       Non-Null Count  Dtype
---  ------                       --------------  -----
 0   Unnamed: 0                   195 non-null    int64
 1   Country                      195 non-null    object
 2   Mean Assistance (2011-2016)  195 non-null    float64
 3   Mean CO2 (2011-2016)         195 non-null    float64
 4   Mean Employment (2011-2016)  195 non-null    float64
 5   Mean GDP (2011-2016)         195 non-null    float64
dtypes: float64(4), int64(1), object(1)
memory usage: 10.7+ KB
```

```
# removing unused columns
data = data.drop(['Unnamed: 0'], axis = 1)
data.head(1)
```

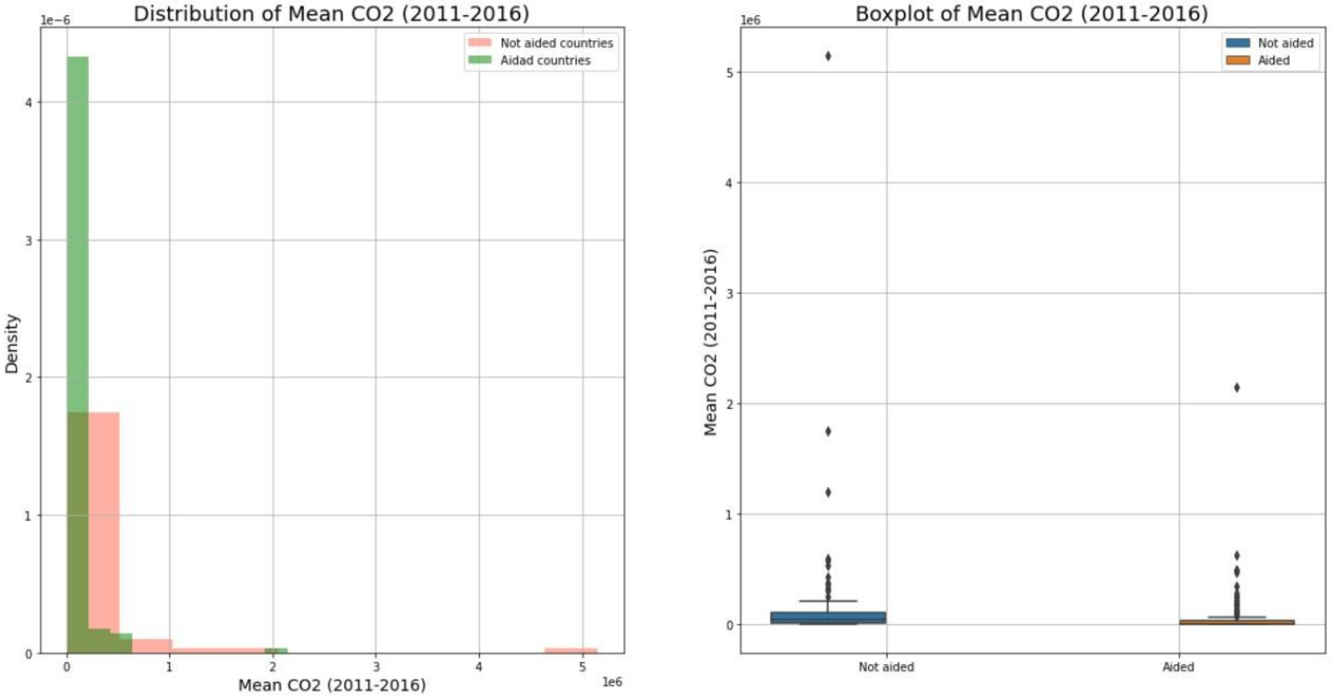
	Country	Mean Assistance (2011-2016)	Mean CO2 (2011-2016)	Mean Employment (2011-2016)	Mean GDP (2011-2016)
0	Aruba	0.0	1227.836667	0.0	35823.64333

```
# creating categorical column indicating the presence of financial aid
data['Aided'] = 'Not aided'
data.loc[
    data['Mean Assistance (2011-2016)'] > 0,
    'Aided'
] = 'Aided'

# checking relative distribution of aided countries
data['Aided'].value_counts(normalize = True)
```

```
Aided        0.702564
Not aided    0.297436
Name: Aided, dtype: float64
```

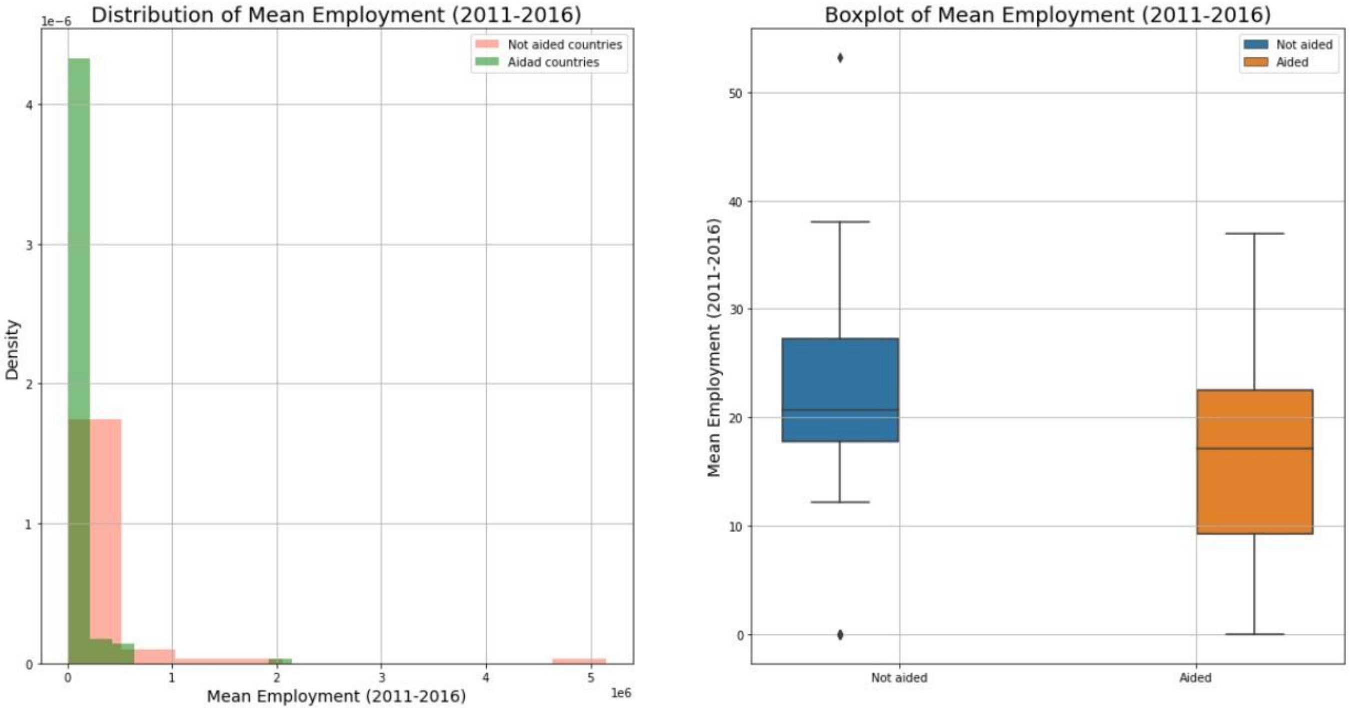
Descriptive Analysis

```
]: # EDA for CO2
hist_box_plot(data, 'Mean CO2 (2011-2016)')
```



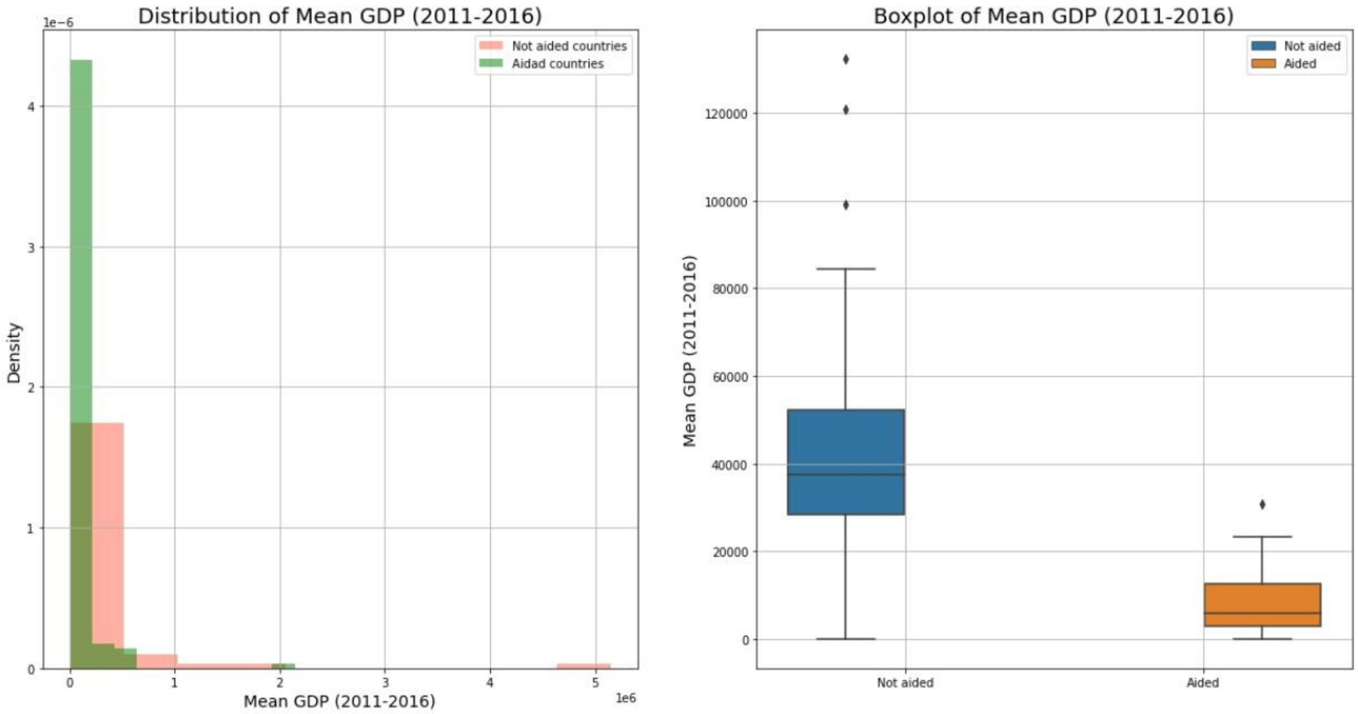
Descriptive Analysis

```
# EDA for Employment
hist_box_plot(data, 'Mean Employment (2011-2016)')
```



Descriptive Analysis

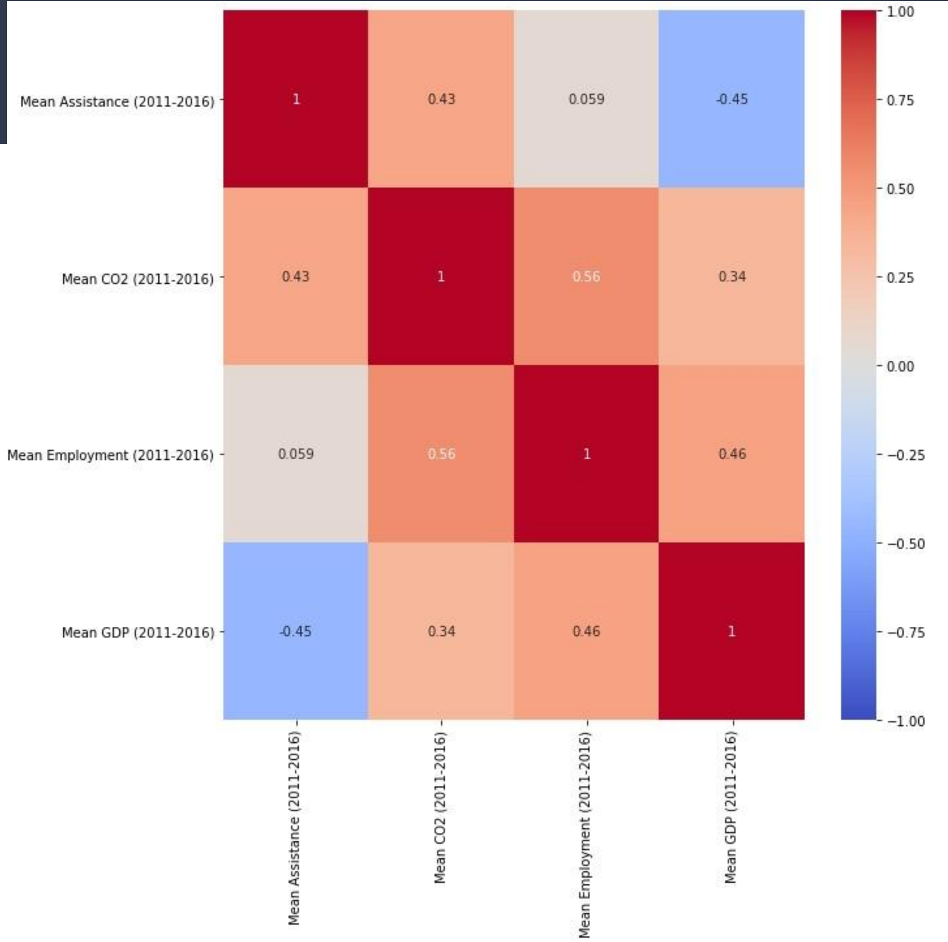
```
]: # EDA for Mean GDP
hist_box_plot(data, 'Mean GDP (2011-2016)')
```



Deriving the variables

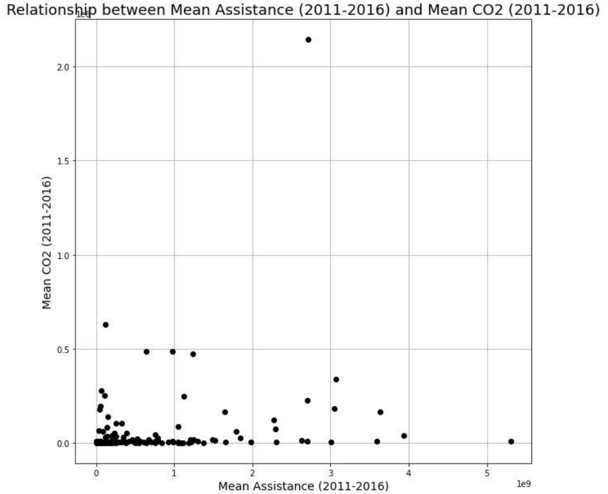
 # Spearman's correlation coefficient

```
# data for correlation
data_corr = data.loc[
    data['Aided'] == 'Aided',
    ['Mean Assistance (2011-2016)', 'Mean CO2 (2011-2016)',
     'Mean Employment (2011-2016)', 'Mean GDP (2011-2016)']
]

# calculating the Spearman's correlation coefficient
plt.figure(figsize = (10,10))
sns.heatmap(data_corr.corr(method = 'spearman'),
            vmin = -1, vmax = 1,
            cmap = 'coolwarm',
            annot = True)
```

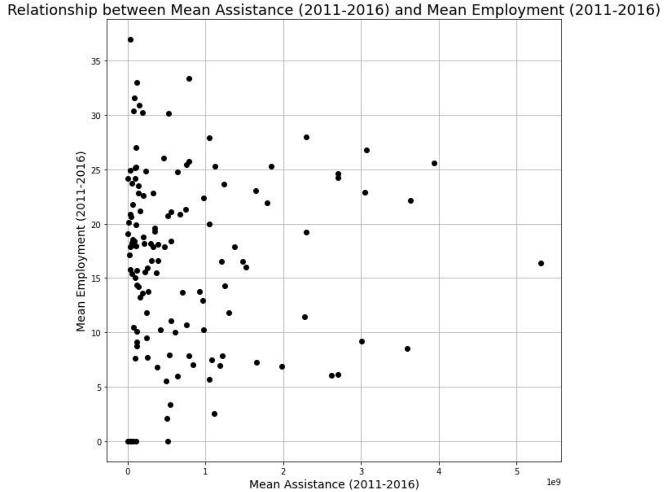


Scatter Plots

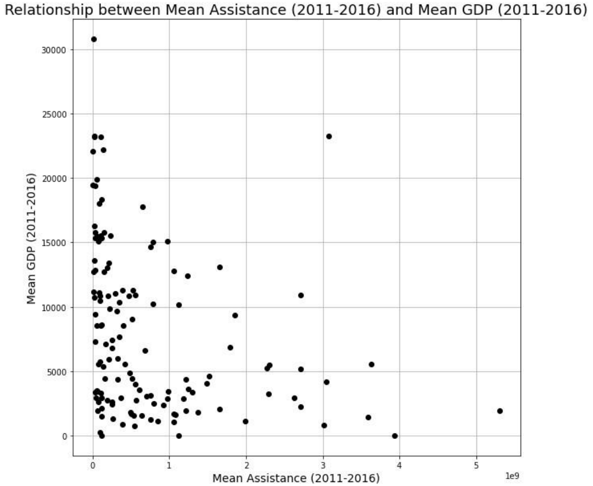
```
# scatteplot for mean CO2
scatter_analysis(data, 'Mean CO2 (2011-2016)')
```




```
# scatteplot for mean Employment
scatter_analysis(data, 'Mean Employment (2011-2016)')
```



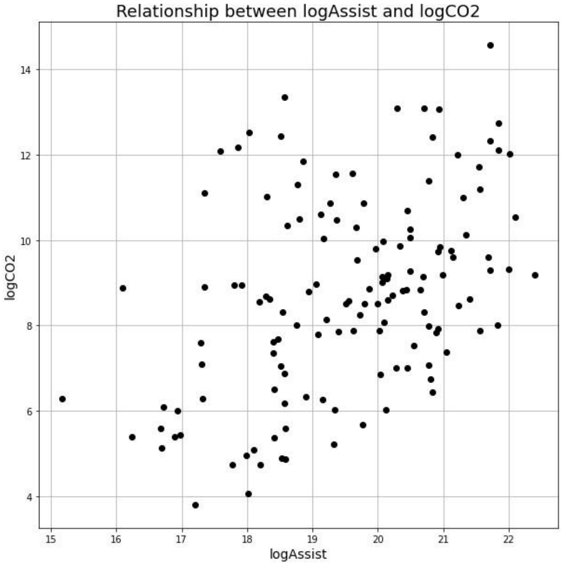

```
# scatteplot for mean GDP
scatter_analysis(data, 'Mean GDP (2011-2016)')
```

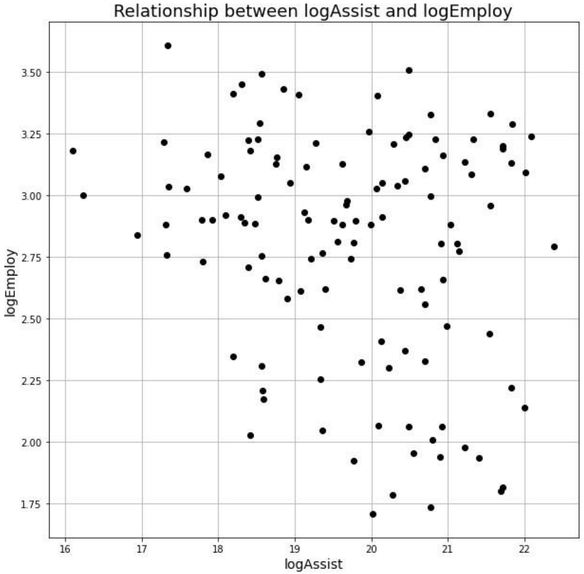


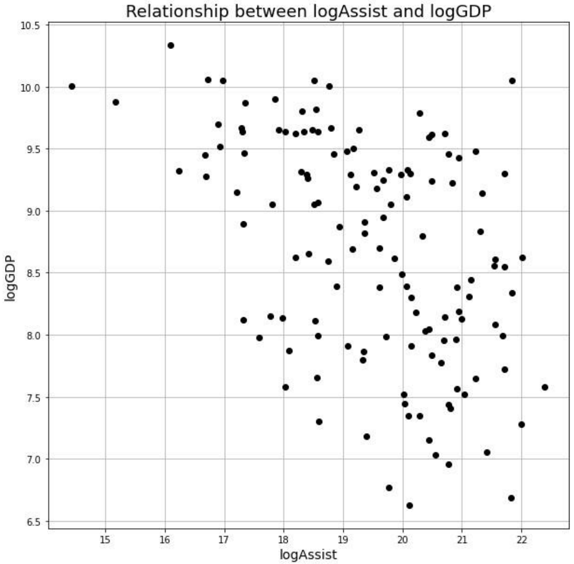
Feature Engineering

```
# creating new features by applying the log function
data['logAssist'] = np.log(data['Mean Assistance (2011-2016)'])
data['logCO2'] = np.log(data['Mean CO2 (2011-2016)'])
data['logEmploy'] = np.log(data['Mean Employment (2011-2016)'])
data['logGDP'] = np.log(data['Mean GDP (2011-2016)'])
```

Clean from outliers







Model training: train/test split

Splitting the training and test sets

```
]: # extracting predictors and targets
x_CO2 = dataCO2['logAssist'].values
y_CO2 = dataCO2['logCO2'].values

# train-test split
x_trainCO2, x_testCO2, y_trainCO2, y_testCO2 = train_test_split(x_CO2, y_CO2, test_size = 0.2,
                                                                  random_state = RANDOM_SEED)
```

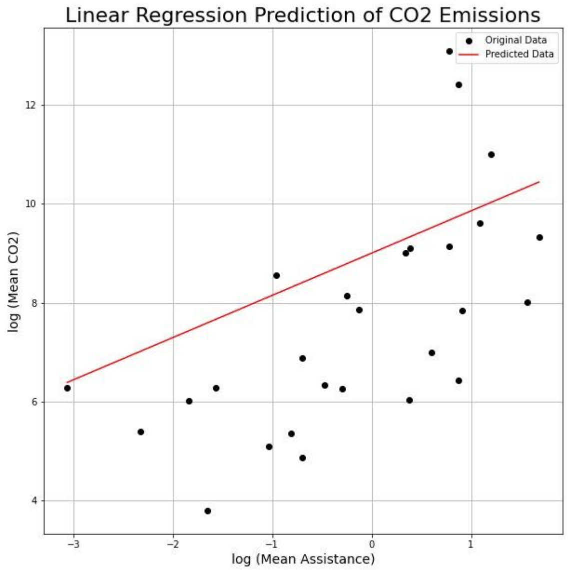
Feature Scaling

```
]: scaler = StandardScaler()
scaler = scaler.fit(x_trainCO2.reshape(-1,1))

# applying the scaler to training and test sets
x_trainCO2 = scaler.transform(x_trainCO2.reshape(-1,1))
x_testCO2 = scaler.transform(x_testCO2.reshape(-1,1))
```

Model Training and Evaluation - CO2 Emissions

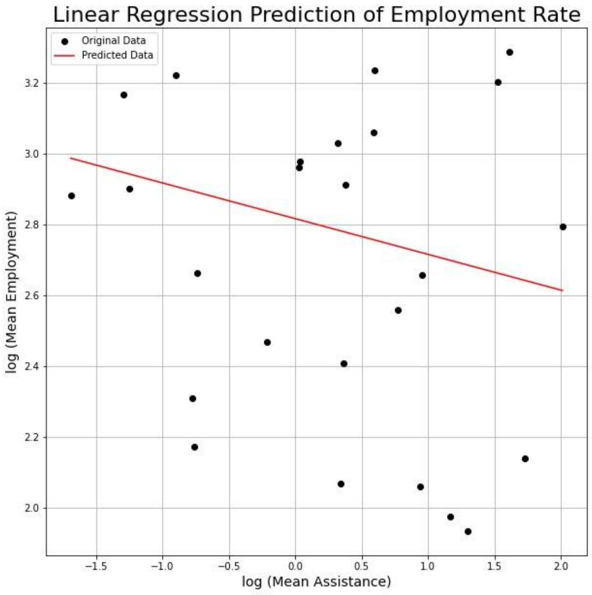
```
]: modelCO2 = LinearRegression()
modelCO2 = modelCO2.fit(x_trainCO2.reshape(-1,1), y_trainCO2)

# predicting values
ypredCO2 = modelCO2.predict(x_testCO2.reshape(-1,1))
```

Model Results



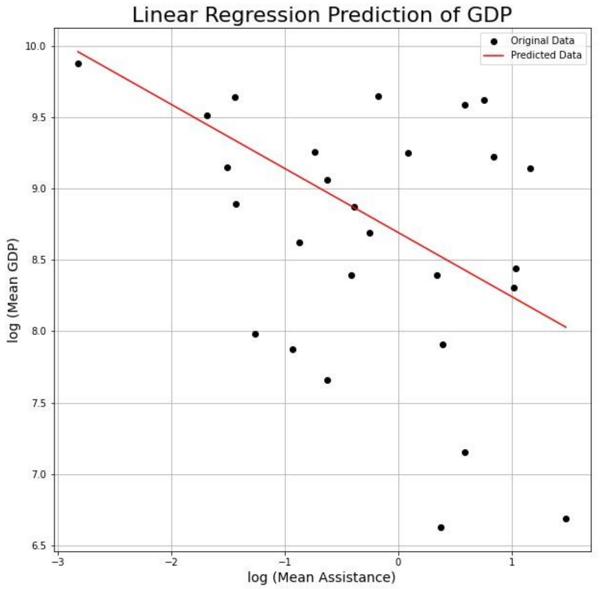

```
# performance metrics
print('R2 Score: %.4f'%(r2_score(y_testCO2, ypredCO2)))
print('MSE: %.4f'%(mean_squared_error(y_testCO2, ypredCO2)))
print('RMSE: %.4f'%(np.sqrt(mean_squared_error(y_testCO2, ypredCO2))))
```

```
R2 Score: 0.0508
MSE: 4.6422
RMSE: 2.1546
```




```
# performance metrics
print('R2 Score: %.4f'%(r2_score(y_testEmp, ypredEmp)))
print('MSE: %.4f'%(mean_squared_error(y_testEmp, ypredEmp)))
print('RMSE: %.4f'%(np.sqrt(mean_squared_error(y_testEmp, ypredEmp))))
```

```
R2 Score: -0.0304
MSE: 0.1937
RMSE: 0.4401
```




```
# performance metrics
print('R2 Score: %.4f'%(r2_score(y_testGDP, ypredGDP)))
print('MSE: %.4f'%(mean_squared_error(y_testGDP, ypredGDP)))
print('RMSE: %.4f'%(np.sqrt(mean_squared_error(y_testGDP, ypredGDP)))
```

```
R2 Score: 0.0885
MSE: 0.7076
RMSE: 0.8412
```

Test Predictions

Predictions - CO2

```
x_CO2 = dataCO2['logAssist'].values
y_CO2 = dataCO2['logCO2'].values
```

```
scaler = StandardScaler()
scaler = scaler.fit(x_CO2.reshape(-1,1))
x_CO2 = scaler.transform(x_CO2.reshape(-1,1))
modelCO2 = LinearRegression()
modelCO2.fit(x_CO2,y_CO2)
```

```
LinearRegression()
```

```
dataCO2['CO2 prediction']= modelCO2.predict(x_CO2)
```

	Country	Mean Assistance (2011-2016)	Mean CO2 (2011-2016)	Mean Employment (2011-2016)	Mean GDP (2011-2016)	Aided	logAssist	logCO2	logEmploy	logGDP	CO2 prediction	Employment prediction	GDP prediction
0	Afghanistan	5.308612e+09	9703.495000	16.351667	1961.268333	Aided	22.392596	9.180241	2.794330	7.581347	10.650099	2.591865	7.894904
1	Albania	2.976650e+08	4953.505000	18.136667	11050.331670	Aided	19.511479	8.507851	2.897936	9.310216	8.708581	2.803696	8.670029
2	Algeria	1.533983e+08	139031.861700	30.910000	12750.568330	Aided	18.848549	11.842458	3.431080	9.453331	8.261847	2.852437	8.848382
3	Angola	2.573450e+08	35430.556670	7.741667	7404.383333	Aided	19.365928	10.475330	2.046617	8.909827	8.610497	2.814397	8.709188
5	Argentina	5.701333e+07	196107.491700	23.713333	19865.350000	Aided	17.858796	12.186418	3.166037	9.896732	7.594876	2.925208	9.114661
...	...	...	...	...	...	...	...	...	...	...	...	...	...
30	Vietnam	3.636665e+09	165060.838300	22.100000	5557.580000	Aided	22.014333	12.014069	3.095578	8.622918	10.395196	2.619676	7.996670
31	West Bank and Gaza	2.300815e+09	2661.020000	27.946667	5482.028333	Aided	21.556529	7.886465	3.330298	8.609230	10.086693	2.653336	8.119836
32	Yemen, Rep.	1.245327e+09	18875.270000	14.271666	3595.601667	Aided	20.942664	9.845608	2.658276	8.187467	9.673023	2.698469	8.284988
33	Zambia	9.828650e+08	4122.318333	10.260000	3449.990000	Aided	20.705982	8.324171	2.328253	8.146127	9.513529	2.715871	8.348664
34	Zimbabwe	7.924700e+08	10780.370000	7.871667	2522.828333	Aided	20.490665	9.285482	2.063270	7.833136	9.368432	2.731702	8.406592

Model Implementation

- Forecast real GDP growth, which is a development indicator.
- Forecast real CO2 in order to prevent and/or reduce emissions and its impacts.
- Forecast real Employment Rate as contribution to the economy, production of goods and services.

Conclusions

- There is a **positive** correlation between Mean Assistance and Mean CO_2 emissions: countries that receive higher foreign aid tend to generate more CO_2. This probably can be due to older technologies and lower educational levels, leading to more pollution and inadequate disposal.

- There is a **negative** correlation between the Mean Assistance and Employment Rate: as an economic indicator, the higher the Employment Rate, the lower will be the need for foreign aid.

- There is a **negative** correlation between the Mean Assistance and GDP: countries that do not produce significant wealth have a higher need for financial aid.

Questions?

Thank you!