DATA ANALYSIS ON A CLIMATE CHANGE DATASET

Short introduction:

Year after year, the globe becomes more modernized, and as a result, more polluted. This information was gathered from the US Energy Administration and combined for a simpler examination. It's a collection of important factors that go into C02 Emissions, such as each country's production and consumption of each major energy source, as well as its yearly pollution rating.

Dataset

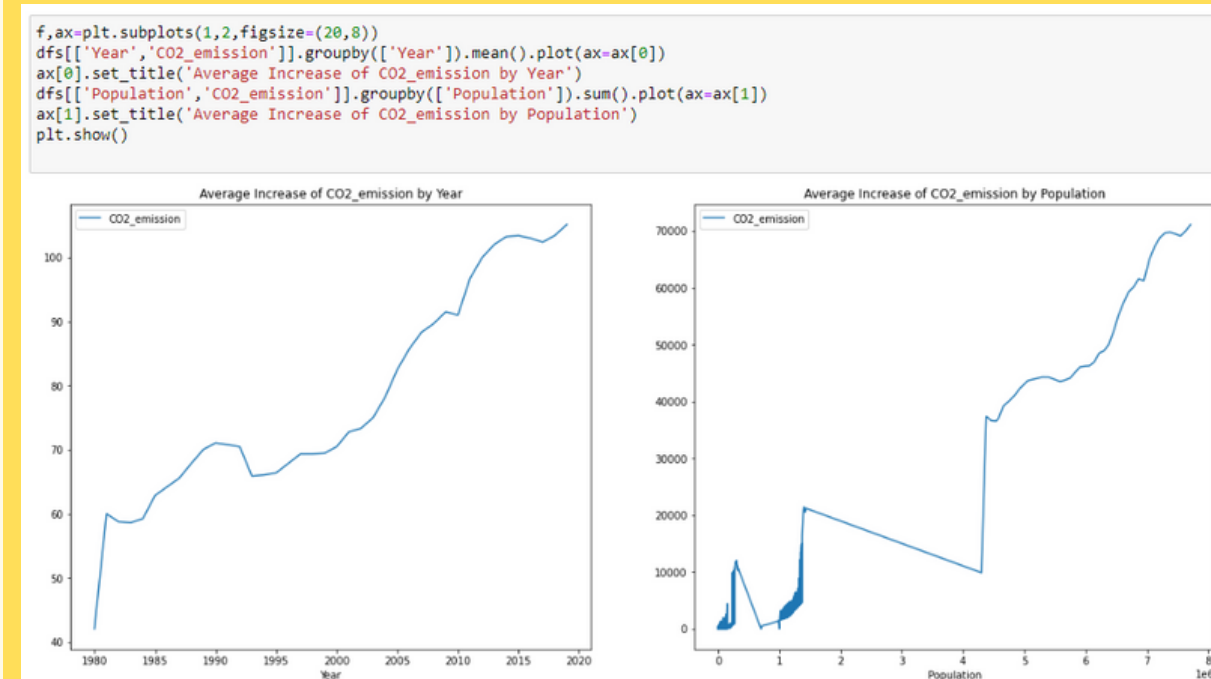|	Country Name	Country Code	Indicator Code	1960	1961	1962	1963	1964	1965	1966
8	Aruba	ABW	SH.DYN.MORT	0.0	0.0	0.0	0.0	0.0	0.0	0.0
84	Africa Eastern and Southern	AFE	SH.DYN.MORT	0.0	0.0	0.0	0.0	0.0	0.0	0.0
160	Afghanistan	AFG	SH.DYN.MORT	358.2	352.2	346.3	340.6	335.1	329.8	324.4
236	Africa Western and Central	AFW	SH.DYN.MORT	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Graph:

Global Annual CO2 level increased in the atmosphere especially in last 30 year up to alarmlying condition.
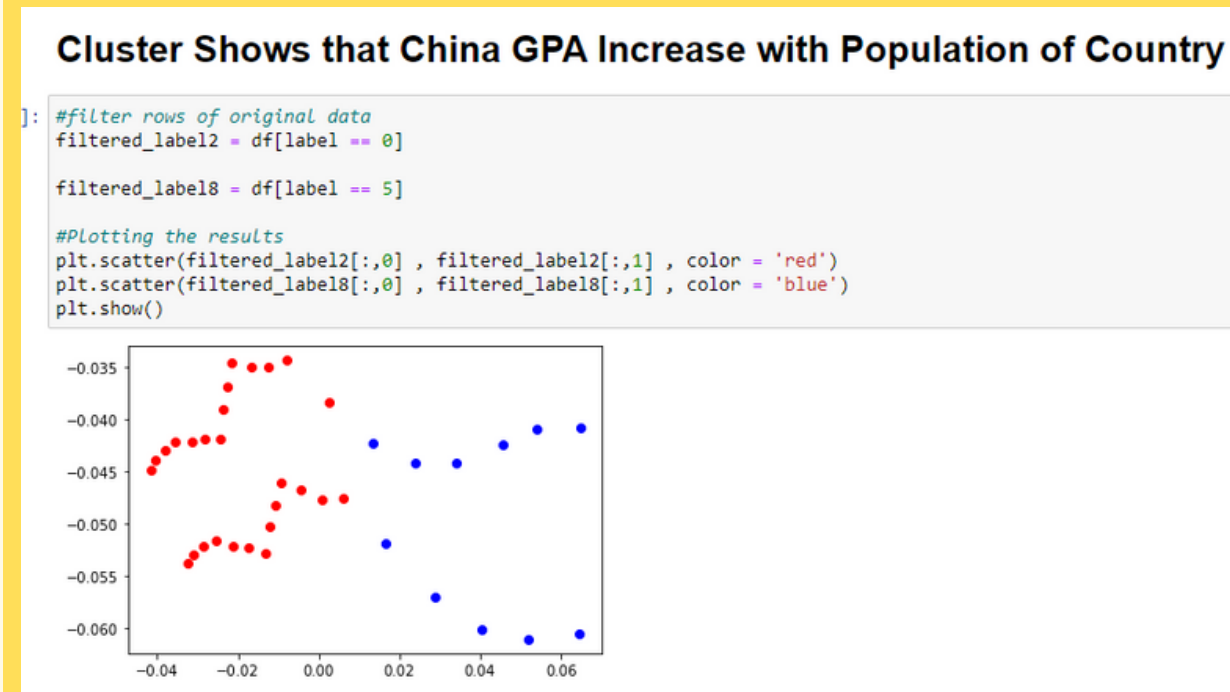

Global annual CO2 levels in atmosphere
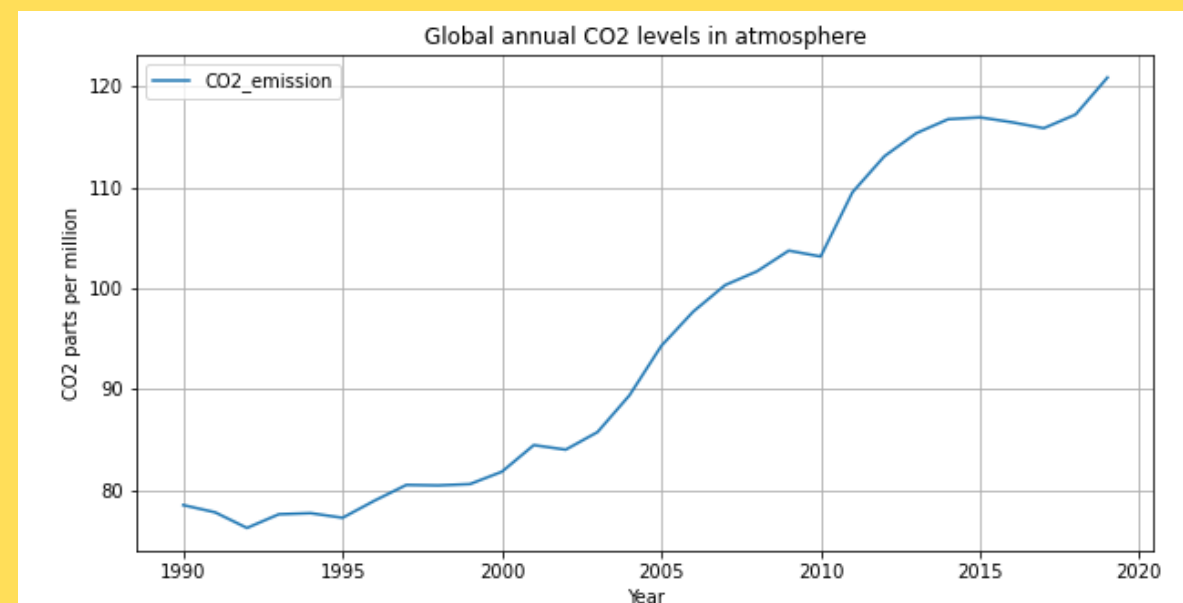
Graph:

Average CO2_emission increase fastly from 1980 to 2020 by population. In the last 10 years, the population increased CO2 emissions rapidly increases
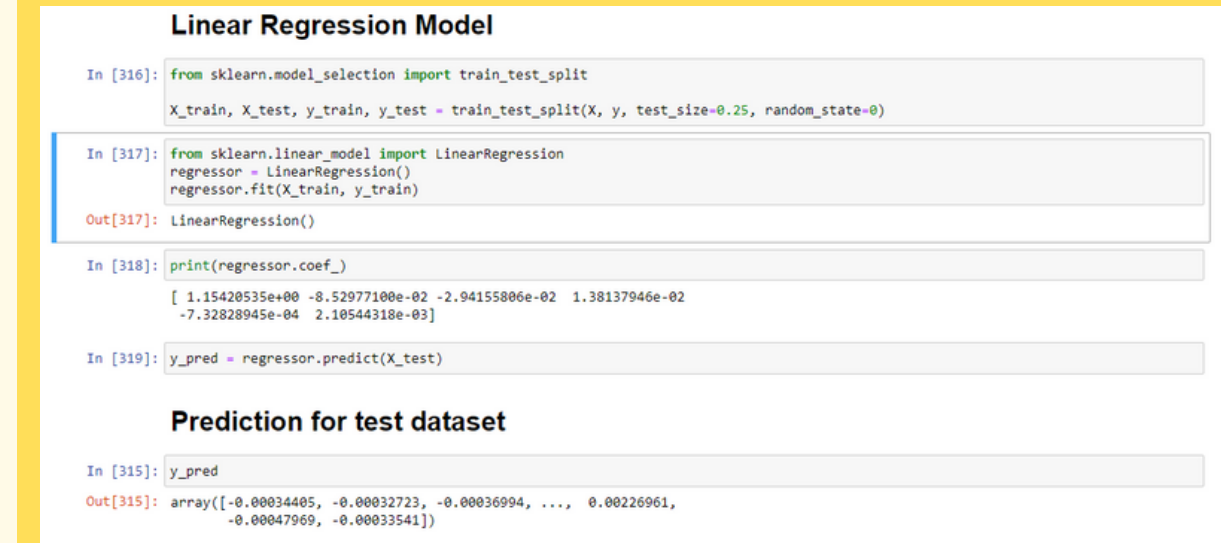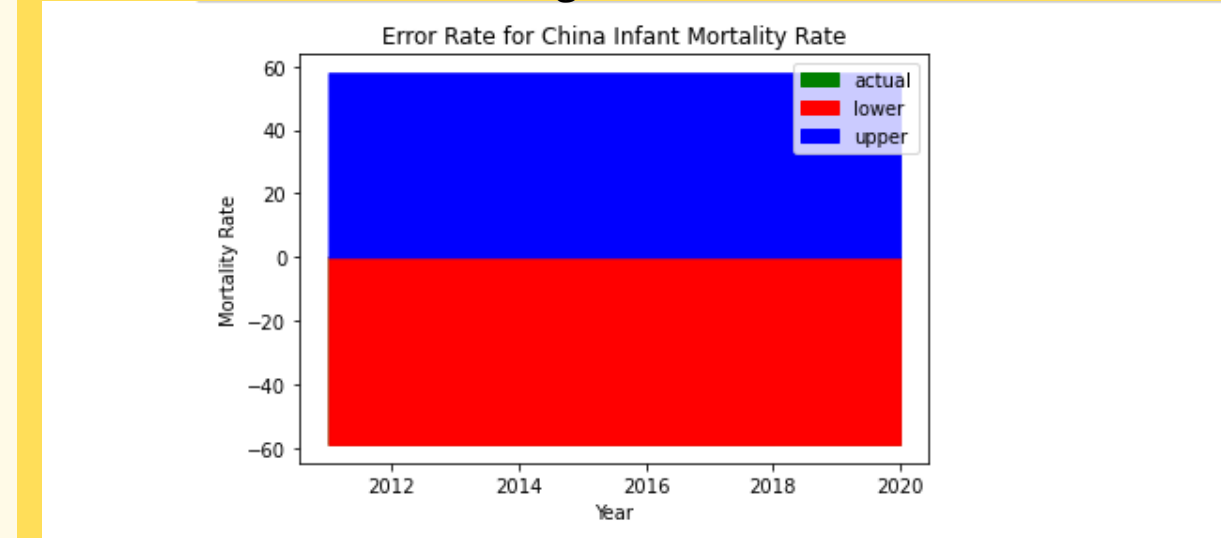
```
f,ax=plt.subplots(1,2,figsize=(20,8))
dfs[['Year','CO2_emission']].groupby(['Year']).mean().plot(ax=ax[0])
ax[0].set_title('Average Increase of CO2_emission by Year')
dfs[['Population','CO2_emission']].groupby(['Population']).sum().plot(ax=ax[1])
ax[1].set_title('Average Increase of CO2_emission by Population')
plt.show()
```


Graph:

K_Mean Clustering shows that GDP of china is increase with Population

Cluster Shows that China GPA Increase with Population of Country

```
]: #filter rows of original data
filtered_label2 = df[label == 0]

filtered_label8 = df[label == 5]

#Plotting the results
plt.scatter(filtered_label2[:,0] , filtered_label2[:,1] , color = 'red')
plt.scatter(filtered_label8[:,0] , filtered_label8[:,1] , color = 'blue')
plt.show()
```


Graph:

Error Rate for China Mortality Rate with lower and upper limits of the confidence range.

Error Rate for China Infant Mortality Rate

Linear Regression Model

```
In [316]: from sklearn.model_selection import train_test_split

          X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=0)

In [317]: from sklearn.linear_model import LinearRegression
          regressor = LinearRegression()
          regressor.fit(X_train, y_train)

Out[317]: LinearRegression()

In [318]: print(regressor.coef_)

          [ 1.15420535e+00 -8.52977100e-02 -2.94155806e-02  1.38137946e-02
           -7.32828945e-04  2.10544318e-03]

In [319]: y_pred = regressor.predict(X_test)
```

Prediction for test dataset

```
In [315]: y_pred

Out[315]: array([-0.00034405, -0.00032723, -0.00036994, ...,  0.00226961,
                 -0.00047969, -0.00033541])
```

Conclusion:

- Most CO2 emissions are caused by renewables and coal, natural gas, and petroleum liquids. Average CO2_emission increase fastly from 1980 to 2020. In especially last 10 years CO2 emissions rapidly increased.
- Average CO2_emission increase fastly from 1980 to 2020 by population. As the population increases CO2 emission rapidly increases.

Reference:

- https://data.worldbank.org/topic/19
- https://ieeexplore.ieee.org/document/9668116