APPLIED DATA SCIENCE – ASSIGNMENT 3

INTRODUCTION

Nowadays, data has become a major part in our day to day life. To analyze these data's, we use different techniques and methods to obtain inference from it.
In this case, we use the methods of clustering and fitting to gather the data's, to group them and then further analyze to see any trends or patterns.

Clustering is the process where data's that has similar properties are grouped together to form a cluster.

Fitting is a process that aims to make sure that the machine learning model built will have the best suited parameters to in order to solve real life scenarios with high accuracy.

In this research, we compare CO2 emission rate of different countries and a relationship between each year and its corresponding emission rate are established.

The countries selected for this research are Australia, China, Canada, France, Russia, New Zealand, Germany, USA and Argentina.

The indicator that has been used for clustering and fitting is EN.ATM.CO2E.PC which indicates to CO2 emission.

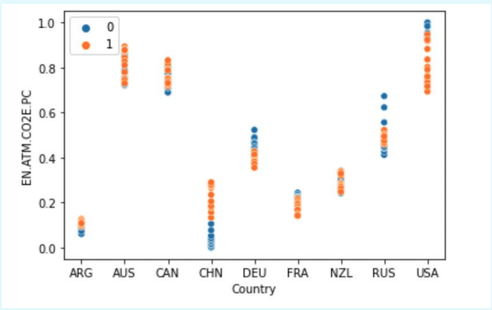
ANALYSIS

Before clustering the data, we first normalize all the data's so that the values are in a uniform value range.

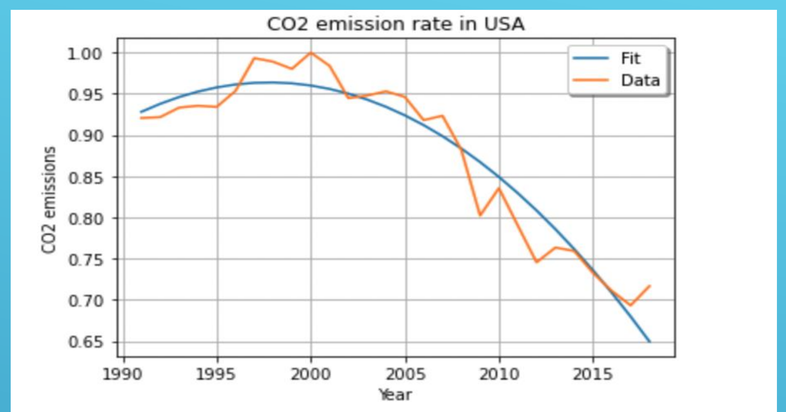
series	Country	Year	EN.ATM.CO2E.PC	EN.ATM.METH.KT.CE
0	ARG	1991	0.064064	0.060993
1	ARG	1992	0.065026	0.062488
2	ARG	1993	0.068642	0.062067
3	ARG	1994	0.070792	0.063281
4	ARG	1995	0.068604	0.063620
...	...	...	...	...
247	USA	2014	0.759328	0.486627
248	USA	2015	0.732744	0.483620
249	USA	2016	0.710894	0.477035
250	USA	2017	0.693253	0.481414
251	USA	2018	0.716802	0.488098

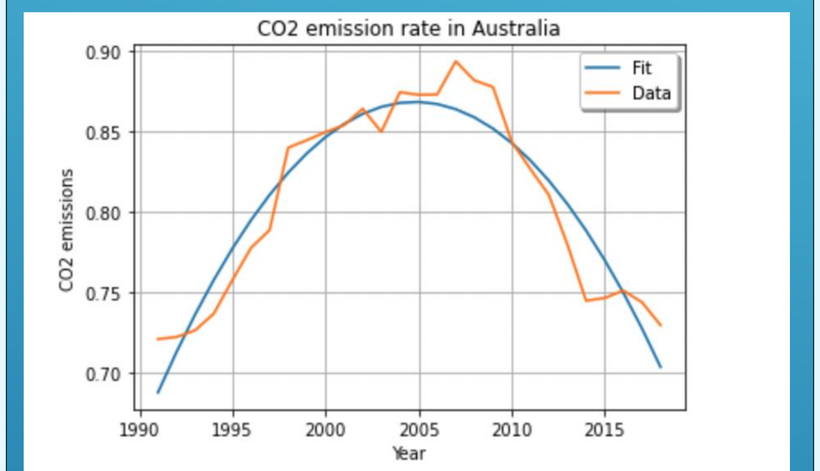
The above table shows the data frame after normalizing all the values to values that range between 0 and 1.

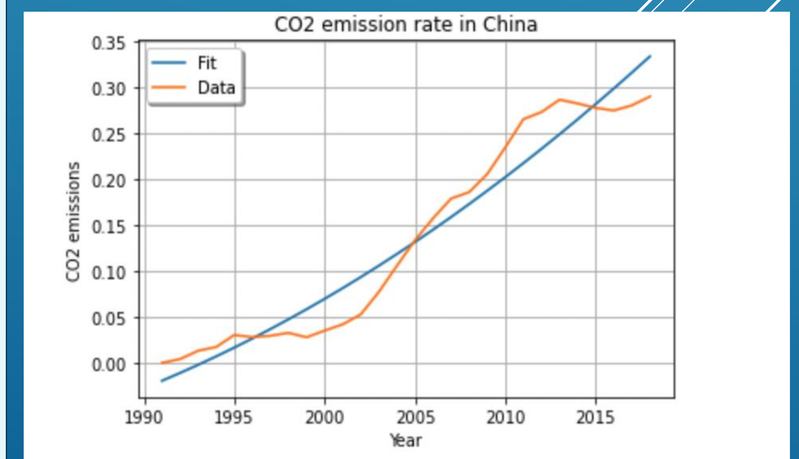
The no. of clusters selected are 2 and the following was plotted after using the clustering technique.



0 being the least emission rate and 1 being the high emission rate, we can see that USA has the highest emission rate, Australia just behind USA and China has the least emission rate.







CONCLUSION

The images shown aside depicts the curve fit used for the country USA, Australia and China which has highest, second highest and least CO2 emission rate respectively.

The analysis were made of the change in rate from 1991 till 2018.

Based on the above plots, we can state that there is a direct relationship between 'year' and the 'CO2 emission rate' even the country is having a highest emission or lowest emission rate.

Based on the analysis we can understand that in USA the emission rate was high in 2000 and then the rate started slowing decreasing, fluctuating between 0.96 and 0.71.
In Australia, the emission rate first increased gradually to a rate of 0.89 in the year 2007 and then decreased to a rate of 0.72 in the year 2018.
In China, whose has shown the least emission rate when compared with Australia and USA, is seen to be increasing from 1991 to 2018 to a rate of 0.34.

REFERENCES:

https://data.worldbank.org/indicator/EN.ATM.CO2E.KT?view=chart