Applied Data Science – Assignment 3

Introduction

Clustering in simple terms can be understood as grouping of similar data that has been gathered for the purpose of further analysis. It depends on the unsupervised machine learning process. It helps in undertaking better comparison (Google Developers, 2022)

Fitting is mainly used for enhancing the accuracy of the data that has been gathered so that the relationship between the elements can be identified. Both these techniques have been used in this research (Data Robot, 2022)

Method

Countries selected includes Japan, Australia, Jamaica, Pakistan, Switzerland, India, Chile, Great Britain, Luxemburg, and Bulgaria.

Indicators selected are as follows:

NE.IMP.GNFS.ZS: Import

NY.GDP.MKTP.PP.CD: GDP, PPP basis

EN.ATM.CO2E.PC: CO2 emissions calculated in metric tons per capita

EN.ATM.GHGT.KT.CE: Greenhouse gas emission

Results

series	Country_Code	Year	NE.IMP.GNFS.ZS	NY.GDP.MKTP.PP.CD	EN.ATM.CO2E.PC	EN.ATM.GHGT.KT.CE
0	AUS	2015	21.556339	1.101457e+12	15.786449	594580.0
1	AUS	2016	21.547899	1.143149e+12	15.872080	573390.0
2	AUS	2017	20.714438	1.190694e+12	15.738647	619790.0
3	AUS	2018	21.512513	1.253361e+12	15.475516	615380.0
4	BGR	2015	62.900855	1.320171e+11	6.225976	57310.0
5	BGR	2016	58.963344	1.430866e+11	5.855926	54270.0
6	BGR	2017	62.682001	1.519202e+11	6.223902	56450.0
7	BGR	2018	63.155915	1.616873e+11	5.854773	53330.0

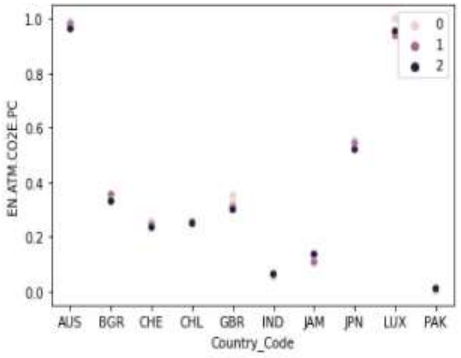
Above image depicts that data that has been achieved after merging the data frames.

,

series	Country_Code	Year	NE.IMP.GNFS.ZS	NY.GDP.MKTP.PP.CD	EN.ATM.CO2E.PC	EN.ATM.GHGT.KT.CE
0	AUS	2015	0.042463	0.119507	0.984073	0.174046
1	AUS	2016	0.042406	0.124138	0.989703	0.167751
2	AUS	2017	0.036792	0.129418	0.980929	0.181535

series	Country_Code	Year	NE.IMP.GNFS.ZS	NY.GDP.MKTP.PP.CD	EN.ATM.CO2E.PC	EN.ATM.GHGT.KT.CE
3	AUS	2018	0.042168	0.136378	0.963628	0.180225
4	BGR	2015	0.320952	0.011839	0.355441	0.014443
5	BGR	2016	0.294430	0.013069	0.331109	0.013540
6	BGR	2017	0.319478	0.014050	0.355304	0.014188
7	BGR	2018	0.322670	0.015134	0.331033	0.013261
8	CHE	2015	0.238941	0.057906	0.255918	0.011591
9	CHE	2016	0.261364	0.060512	0.257108	0.011734

Data achieved after normalization of the data set used in this project.

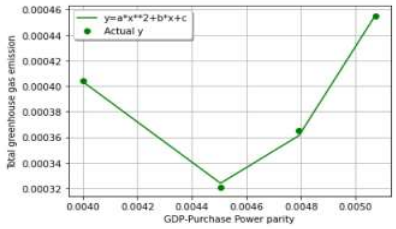


Above image depicts the clustering of all the countries related to carbon dioxide emission. It has highest in the countries Australia and Luxemburg as compared to the other countries selected.

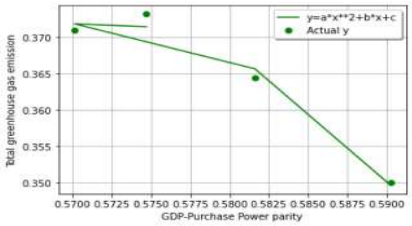
References

Data Robot, 2022. *Model Fitting*. [Online] Available at: https://www.datarobot.com/wiki/fitting/ [Accessed 14 May 2022].

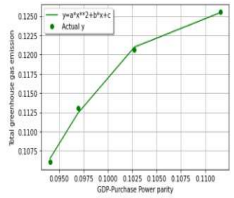
Google Developers, 2022. *What is Clustering?* [Online] Available at: https://developers.google.com/machine-learning/clustering/overview [Accessed 14 May 2022].



Luxemburg has high carbon dioxide emission; hence curve fit has been implemented for further analysis



Curve fit function has been implemented for the country Japan which has medium carbon dioxide emission



Curve fit has been used for the country Pakistan which has low carbon dioxide emission

Conclusion

Countries with high carbon dioxide emission reflects an indirect relationship between total greenhouse emission and GDP, PPP but after a certain value of GDP, PPP the relationship becomes direct.

Country with medium carbon dioxide emission reflect direct relationship between indicators of greenhouse emission and GDP, PPP

Countries with low carbon dioxide emission reflect direct relationship between greenhouse emission and GDP, PPP.