

CLUSTER AND FITTING

NAME : MISHAL MOHAMMED

REG NO: 21031226

CLUSTER

CURVE FITTING

Introduction

The dataset is used from world bank api. The 2 dataset are population and GDP of 2 countries. The following shows the dataset of population of 20 countries.

	YR2009	YR2010	YR2011	YR2012	YR2013	YR2014	YR2015	YR2016	YR2017	YR2018
economy										
AUS	2.169170e+07	2.203175e+07	2.234002e+07	2.273346e+07	2.312813e+07	2.347569e+07	2.381600e+07	2.419091e+07	2.460186e+07	2.498269e+07
BRA	1.938665e+08	1.967136e+08	1.976145e+08	1.992873e+08	2.010359e+08	2.027637e+08	2.044718e+08	2.061631e+08	2.078338e+08	2.094693e+08
CAN	3.362860e+09	3.400489e+09	3.433933e+09	3.471422e+09	3.508295e+09	3.543744e+09	3.5770291e+09	3.610949e+09	3.654524e+09	3.706508e+09
CHN	1.331260e+09	1.337705e+09	1.345035e+09	1.354190e+09	1.363240e+09	1.371860e+09	1.379860e+09	1.387790e+09	1.396215e+09	1.402760e+09
DEU	8.190231e+07	8.177693e+07	8.027498e+07	8.042582e+07	8.064500e+07	8.098250e+07	8.168661e+07	8.234867e+07	8.265700e+07	8.290578e+07
FRA	6.470704e+07	6.502750e+07	6.534279e+07	6.565981e+07	6.599868e+07	6.631207e+07	6.654827e+07	6.672410e+07	6.691802e+07	6.710193e+07
GBR	6.227677e+07	6.276636e+07	6.326881e+07	6.370022e+07	6.412827e+07	6.460230e+07	6.511623e+07	6.561159e+07	6.605886e+07	6.646034e+07
IDN	2.386206e+08	2.418342e+08	2.451160e+08	2.484517e+08	2.518003e+08	2.551281e+08	2.583833e+08	2.615564e+08	2.646510e+08	2.676705e+08
IND	1.211722e+09	1.234281e+09	1.250288e+09	1.265789e+09	1.280842e+09	1.295801e+09	1.310152e+09	1.324517e+09	1.338677e+09	1.352642e+09
IRN	7.292483e+07	7.376252e+07	7.463486e+07	7.553989e+07	7.648196e+07	7.746577e+07	7.849221e+07	7.956399e+07	8.067389e+07	8.180020e+07
ITA	5.909536e+07	5.927742e+07	5.937945e+07	5.953972e+07	6.023395e+07	6.078914e+07	6.073058e+07	6.062750e+07	6.053671e+07	6.042176e+07
JPN	1.280470e+08	1.286700e+08	1.278330e+08	1.276200e+08	1.274450e+08	1.272780e+08	1.271410e+08	1.269945e+08	1.267858e+08	1.265291e+08
KOR	4.930784e+07	4.955411e+07	4.993664e+07	5.019985e+07	5.042889e+07	5.074666e+07	5.101495e+07	5.121780e+07	5.136191e+07	5.156506e+07
MEX	1.124639e+08	1.140630e+08	1.156955e+08	1.172742e+08	1.188272e+08	1.203551e+08	1.218583e+08	1.233334e+08	1.247773e+08	1.261909e+08
POL	3.815160e+07	3.804278e+07	3.806326e+07	3.806316e+07	3.804020e+07	3.801174e+07	3.798641e+07	3.797009e+07	3.797483e+07	3.797475e+07
SAU	1.427853e+08	1.428495e+08	1.429039e+08	1.432017e+08	1.435070e+08	1.438197e+08	1.440969e+08	1.443424e+08	1.444967e+08	1.444779e+08
BAU	2.683030e+07	2.742147e+07	2.826759e+07	2.915491e+07	3.005206e+07	3.091860e+07	3.171768e+07	3.243444e+07	3.310118e+07	3.370278e+07
TUR	7.132141e+07	7.232899e+07	7.344325e+07	7.465105e+07	7.592545e+07	7.722926e+07	7.852941e+07	7.982787e+07	8.111945e+07	8.234009e+07
USA	3.067715e+08	3.093271e+08	3.115355e+08	3.138777e+08	3.160599e+08	3.183863e+08	3.207390e+08	3.230718e+08	3.251221e+08	3.268382e+08
ZAF	5.047701e+07	5.121697e+07	5.200376e+07	5.283266e+07	5.368712e+07	5.454418e+07	5.538637e+07	5.620765e+07	5.700975e+07	5.778252e+07

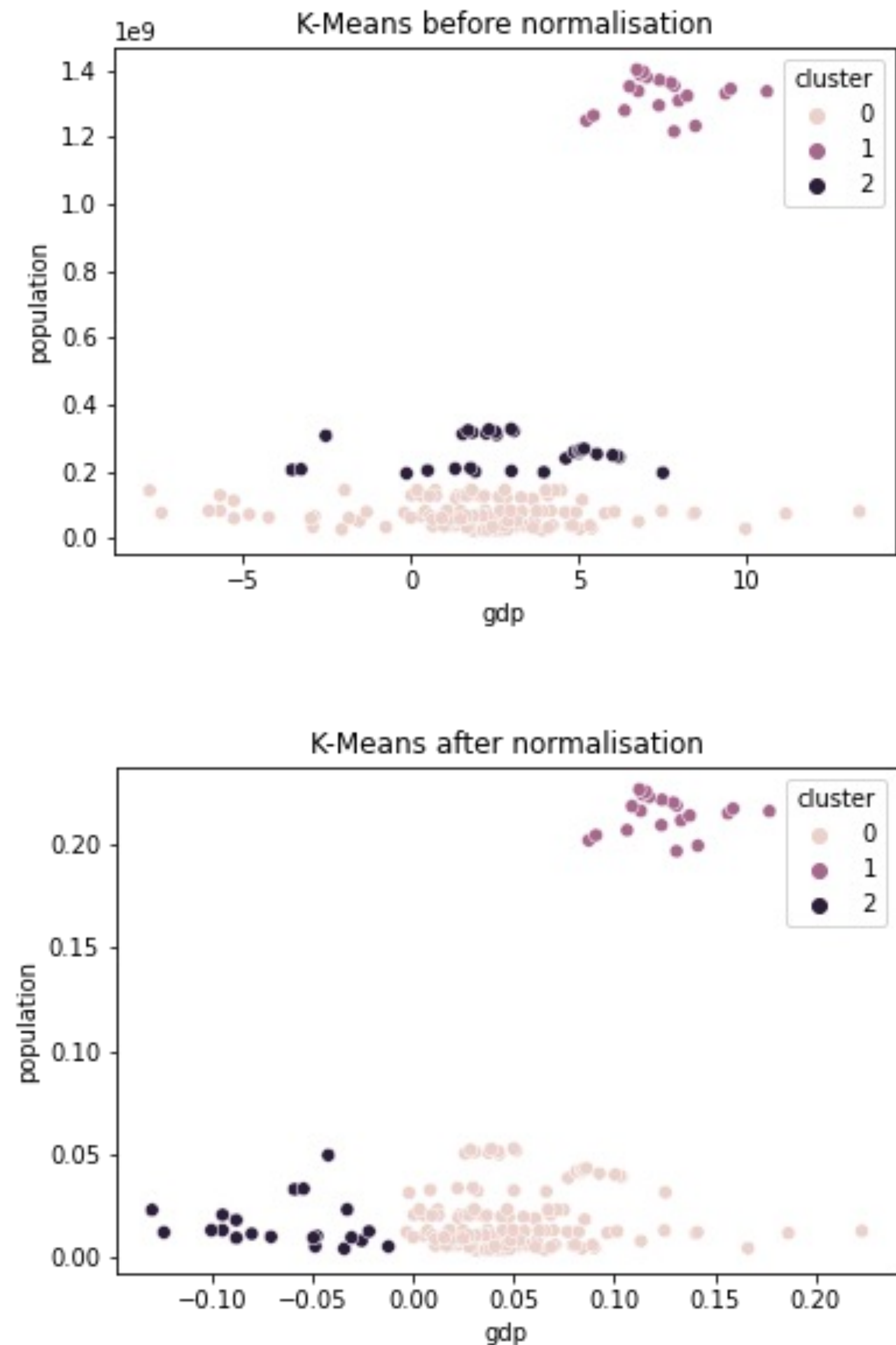
Clustering

To do the clustering we done the following steps

- Combine 2 dataset into a single data frame
- Using libraries sklearn, we imported k-means clustering and done the clustering process.
- After the process we got the ideal number of grouping as 3.
- Next, we normalize the dataset using the library preprocessing also from sklearn to get a better result from the same dataset.
- After normalizing the data, we again done the process of K-Mean clustering.
- This result in having 3 cluster as the ideal case.

Output graph

The output graph of clustering of data set.



Curve Fitting

Curve fitting is the process of finding a mathematical formula that displays the relation between the dependent variable, x and as independent variable y. It also estimate the parameter values with the help of nonlinear regression. For curve fitting we use electricity consumption and access to electricity. The data is from 20 countries in the period of 2004 to 2013, 10 years.

	access	consumtion
0	100.000000	10554.815823
1	96.765106	1956.159381
2	100.000000	17264.736744
3	97.989670	1585.838782
4	100.000000	7108.962524
...
195	100.000000	6539.207375
196	100.000000	8784.756106
197	100.000000	2755.623957
198	100.000000	13004.023569
199	85.199997	4285.645767

Conclusion

The 2 graph with and without curve fit is showing below.

