CLUSTERING AND FITTING

NAME: BHASKAR CHERUKU
STUDENT ID: 22003064

INTRODUCTION:
ANALYSIS RELATED TO PUBLIC DATA FROM THE WORLD BANK, AND SPECIFICALLY COUNTRY-BY-COUNTRY INDICATORS RELATED TO CLIMATE CHANGE, DATALINK. FOR THE VISUALIZATION PERSPECTIVE THE ADDITIONAL RELEVANT INDICATORS ARE USED FOR GDP PER CAPITA (ADDITIONAL DATALINK). FROM THE DATA RETRIEVED FROM THE INTERNET FIND THE CLUSTER VISUALIZATION FOR THE RELEVANT FEATURES.

DATA EXPLORATION:
DATA WAS DOWNLOADED FROM THE MENTIONED DATALINK, DATASET CONTAINS THE TOTAL 66 FEATURES AND 266 SAMPLES ARE THERE, FROM THE YEAR 1960 TO 2021 DATA WAS SHARED. MAINLY DATA IS HAVING THE COUNTRY NAME, CODE, INDICATOR AND DIFFERENT YEAR CO2 EMISSION DATA. EMISSION DATA HAS BEEN OBSERVED DURING DIFFERENT YEAR AND COUNTRY, IDEA IS TO EXPLORE THE DATA CLUSTERING.

Visualization:

K-means clustering algorithm applied to check the co2 emissions data.
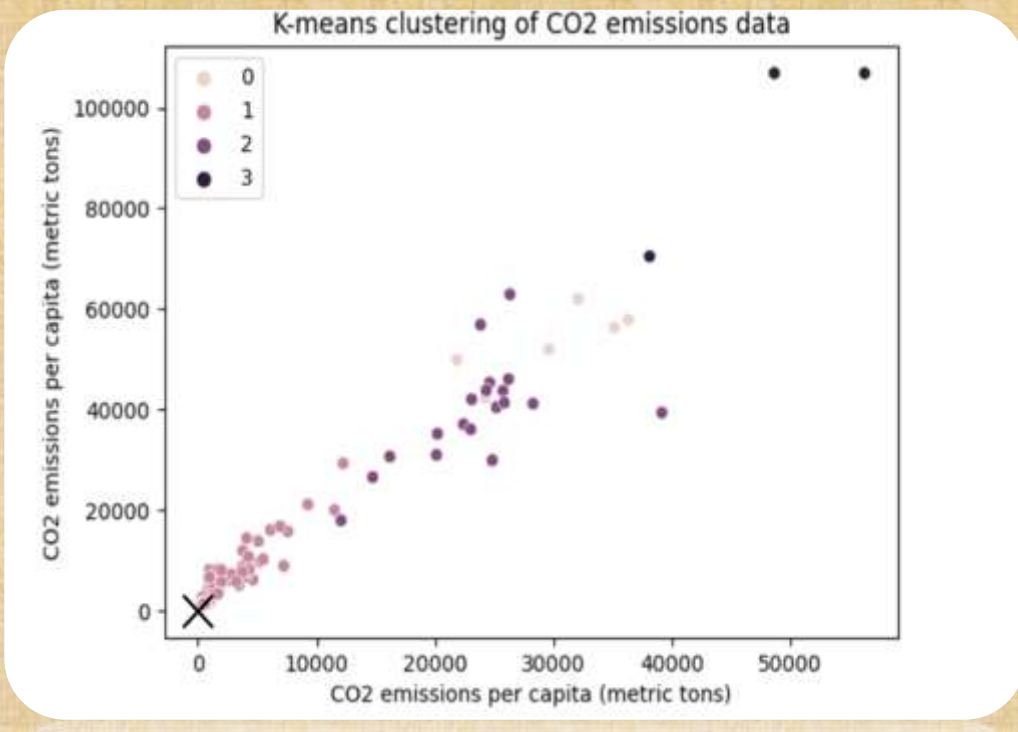Total 4 different clusters are formed.

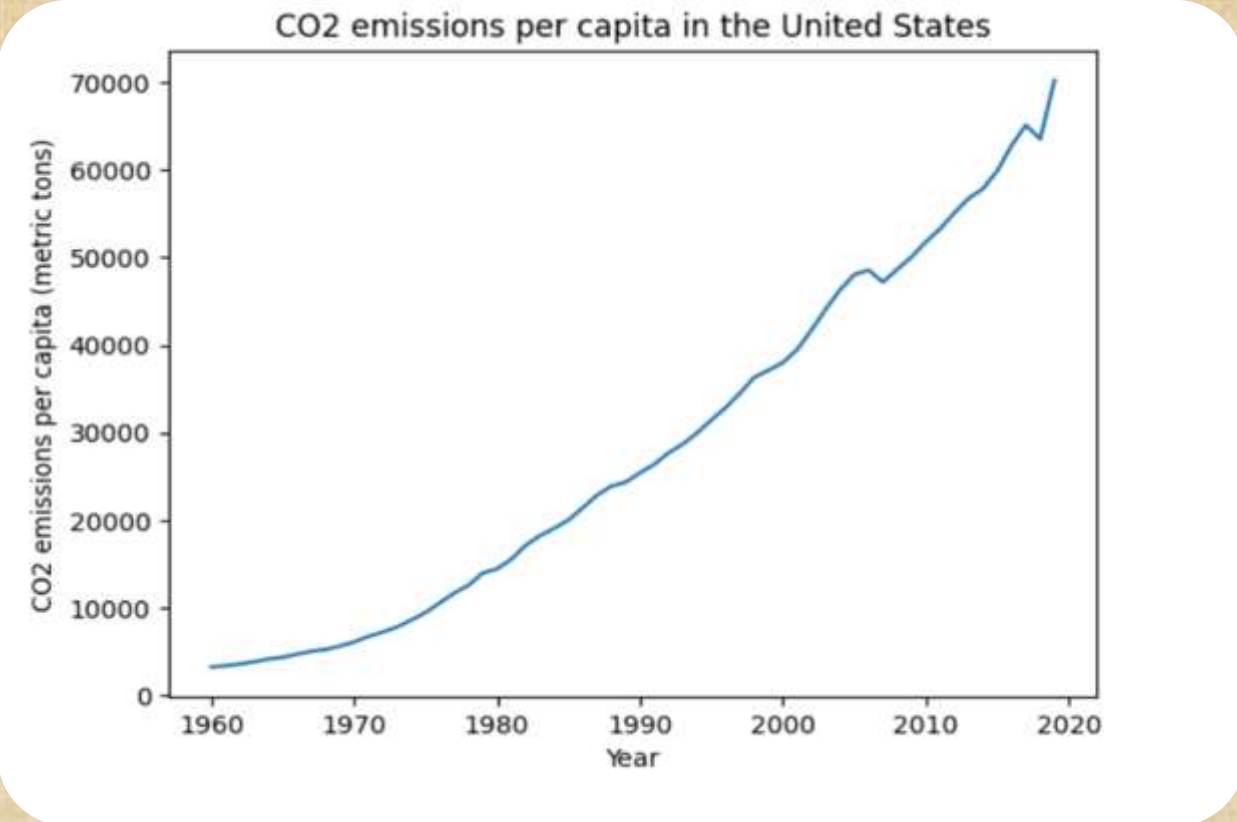Figure 1. Overall Clustering Formation

Figure 2. CO2 emissions per capita during 1960 t0 2020
It has been observed that during each
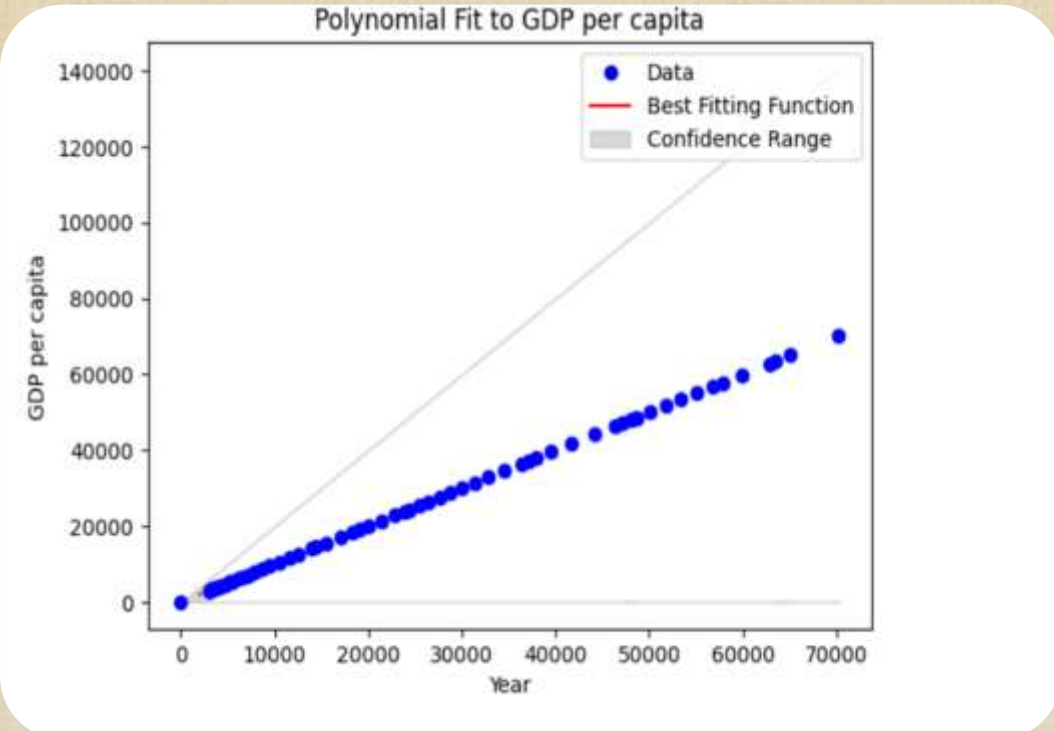and every year the co2 emissions value is increasing.
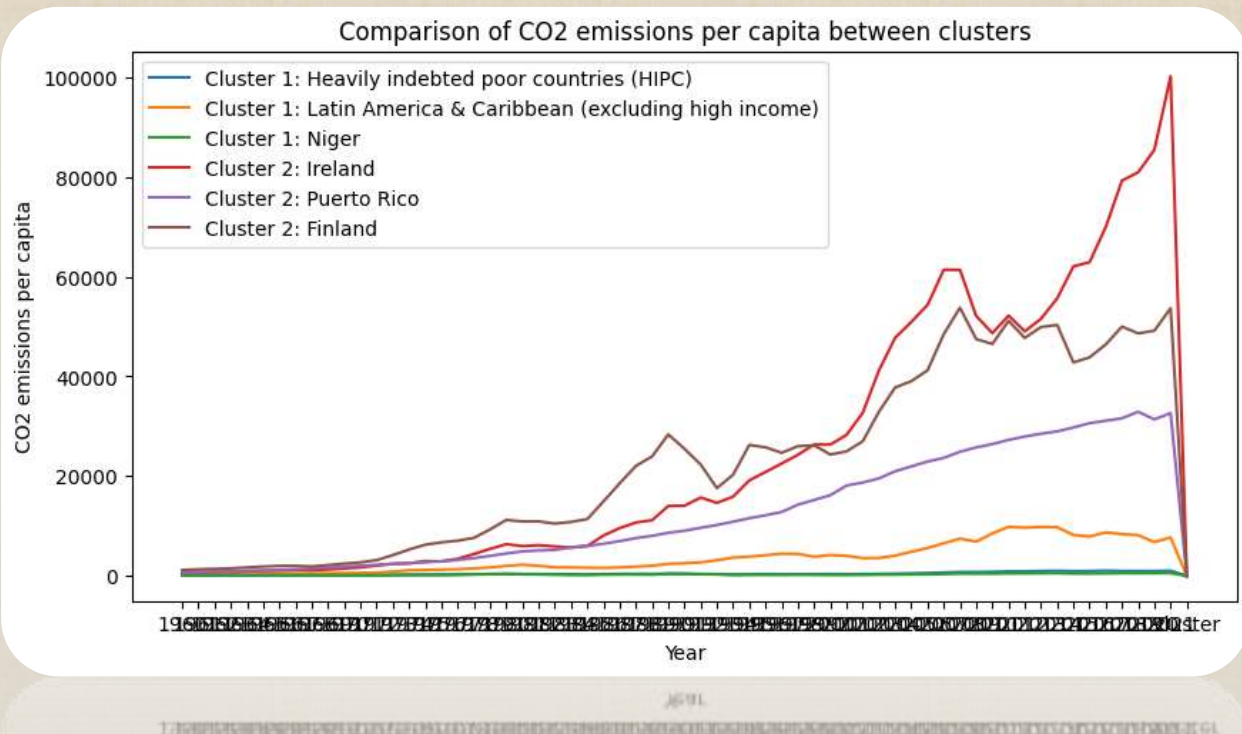

Figure 3. Polynomial Fit to GDP per capita

Figure 4. Different comparison of co2 emissions per capita
between two different clusters country and
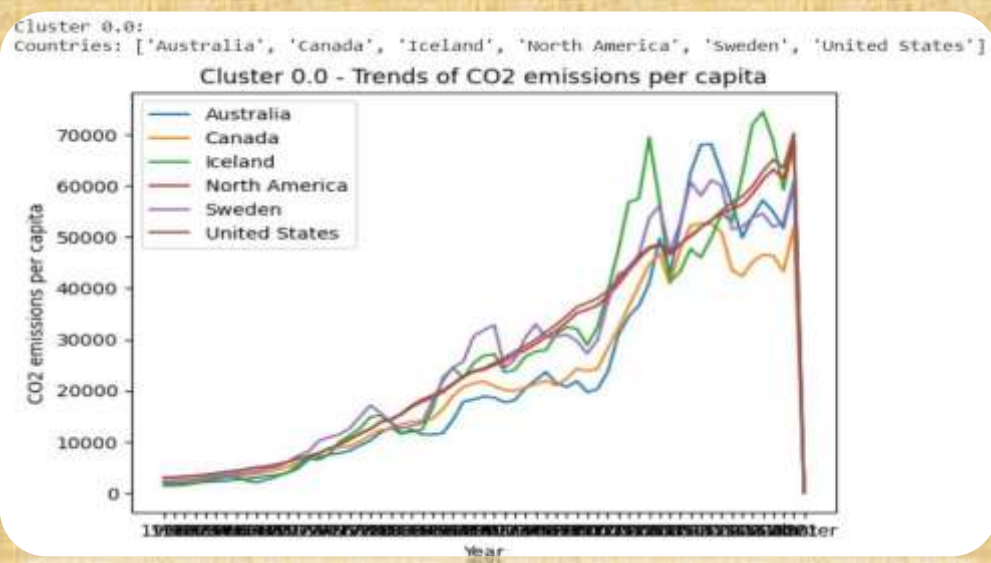based on the country cluster comparisons are made.

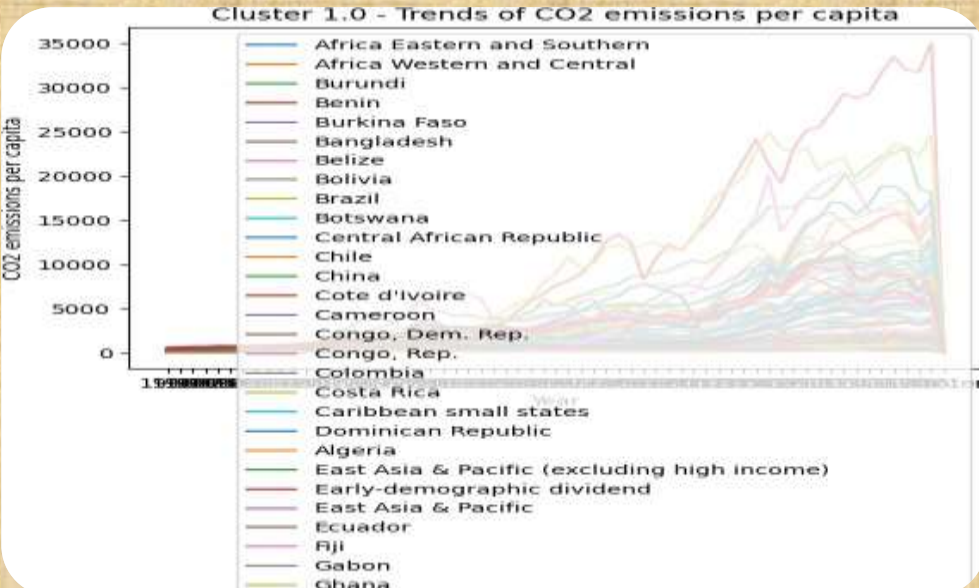Figure 5: Cluster 0 country Trends CO2 emissions per capita

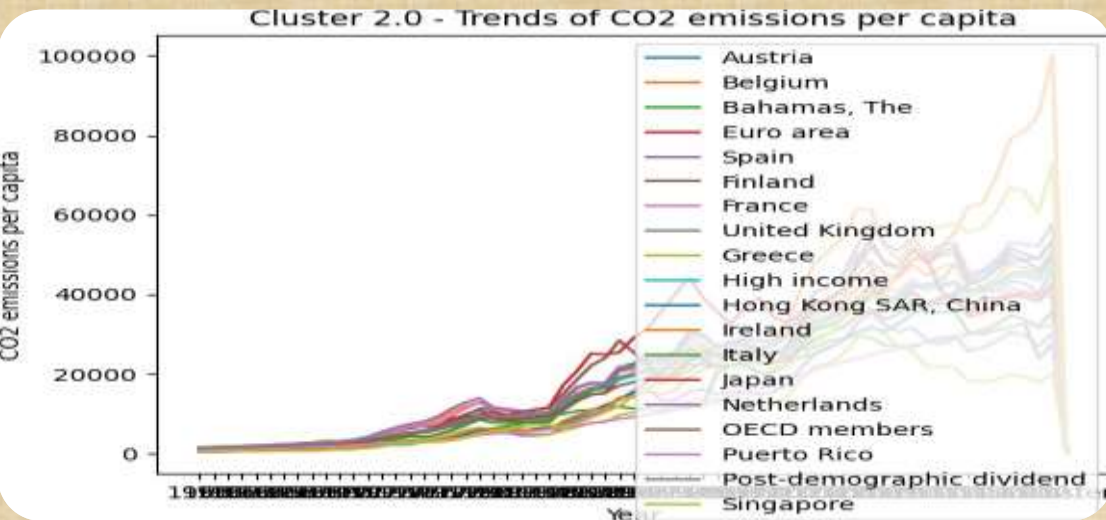Figure 6: Cluster 1 country Trends CO2 emissions per capita

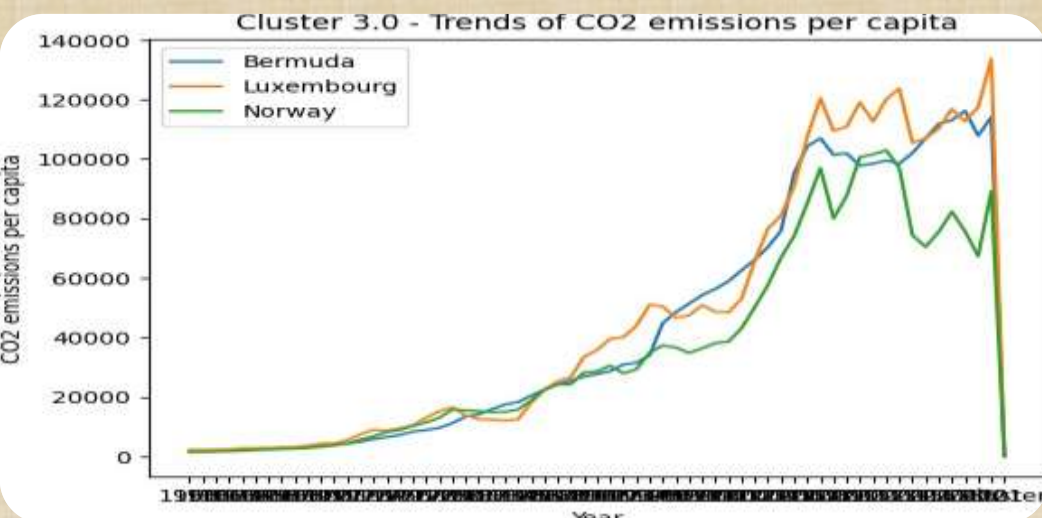Figure 7: Cluster 2 country Trends CO2 emissions per capita

Figure 8: Cluster 3 country Trends CO2 emissions per capita

Cluster Number	Country Name
0	Countries: ['Australia', 'Canada', 'Iceland', 'North America', 'Sweden', 'United States']
1	Countries: ['Africa Eastern and Southern', 'Africa Western and Central', 'Burundi', 'Benin', 'Burkina Faso', 'Bangladesh', 'Belize', 'Bolivia', 'Brazil', 'Botswana', 'Central African Republic', 'Chile', 'China', 'Cote d'Ivoire', 'Cameroon', 'Congo, Dem. Rep.', 'Congo, Rep.', 'Colombia', 'Costa Rica', 'Caribbean small states', 'Dominican Republic', 'Algeria', 'East Asia & Pacific (excluding high income)', 'Early-demographic dividend', 'East Asia & Pacific', 'Ecuador', 'Fiji', 'Gabon', 'Ghana', 'Guatemala', 'Guyana', 'Honduras', 'Heavily indebted poor countries (HIPC)', 'Haiti', 'IBRD only', 'IDA & IBRD total', 'IDA total', 'IDA only', 'India', 'Jamaica', 'Kenya', 'St. Kitts and Nevis', 'Korea, Rep.', 'Latin America & Caribbean (excluding high income)', 'Latin America & Caribbean', 'Low income', 'Sri Lanka', 'Lower middle income', 'Low & middle income', 'Lesotho', 'Late-demographic dividend', 'Morocco', 'Madagascar', 'Mexico', 'Middle income', 'Malawi', 'Malaysia', 'Niger', 'Nigeria', 'Nicaragua', 'Nepal', 'Pakistan', 'Panama', 'Peru', 'Philippines', 'Papua New Guinea', 'Pre-demographic dividend', 'Portugal', 'Rwanda', 'South Asia', 'Sudan', 'Senegal', 'Sierra Leone', 'Sub-Saharan Africa (excluding high income)', 'Sub-Saharan Africa', 'Suriname', 'Eswatini', 'Seychelles', 'Chad', 'East Asia & Pacific (IDA & IBRD countries)', 'Togo', 'Thailand', 'Latin America & the Caribbean (IDA & IBRD countries)', 'South Asia (IDA & IBRD)', 'Sub-Saharan Africa (IDA & IBRD countries)', 'Trinidad and Tobago', 'Turkiye', 'Uganda', 'Upper middle income', 'Uruguay', 'St. Vincent and the Grenadines', 'World', 'South Africa', 'Zambia', 'Zimbabwe']
2	Countries: ['Austria', 'Belgium', 'Bahamas, The', 'Euro area', 'Spain', 'Finland', 'France', 'United Kingdom', 'Greece', 'High income', 'Hong Kong SAR, China', 'Ireland', 'Italy', 'Japan', 'Netherlands', 'OECD members', 'Puerto Rico', 'Post-demographic dividend', 'Singapore']
3	Countries: ['Bermuda', 'Luxembourg', 'Norway']