CLUSTERING AND FITTING CO2 EMMISSIONS VS GDP

NAME: AYISHA FIDHA MANIYODAN
STUDENT ID: 22091853

Dataset link: https://ourworldindata.org/grapher/co2-emissions-vs-gdp

Introduction

•In this dataset, it take a closer look at the correlation between per-capita annual CO_2 emissions, GDP per capita, and estimated historical population for different entities including countries and regions.
•The data covers various years, allowing for a detailed analysis of the relationship between economic growth and carbon emissions.
•The aim of our study is to conduct clustering and fitting analyses to analyse patterns and trends within the given dataset.
•Through the use of these methods, we hope to gain insights into how various entities have changed over time with regard to their GDP, ecological costs and population.
•The dataset has a wide variation from well-developed nations to places like Abkhazia that allows to study the differences in environmental damage and economic progress.

```
# Print column names to identify the correct names
print(df.columns)
```

```
Index(['Entity', 'Code', 'Year', 'Annual CO₂ emissions (per capita)',
       'GDP per capita', '417485-annotations',
       'Population (historical estimates)', 'Continent'],
      dtype='object')
```

```
df.head()
```

	Entity	Code	Year	Annual CO₂ emissions (per capita)	GDP per capita	417485-annotations	Population (historical estimates)	Continent
0	Abkhazia	OWID_ABK	2015	NaN	NaN	NaN	NaN	Asia
1	Afghanistan	AFG	1949	0.001992	NaN	NaN	7356890.0	NaN
2	Afghanistan	AFG	1950	0.011266	1156.0	NaN	7480464.0	NaN
3	Afghanistan	AFG	1951	0.012098	1170.0	NaN	7571542.0	NaN
4	Afghanistan	AFG	1952	0.011946	1189.0	NaN	7667534.0	NaN

Aims and Objectives

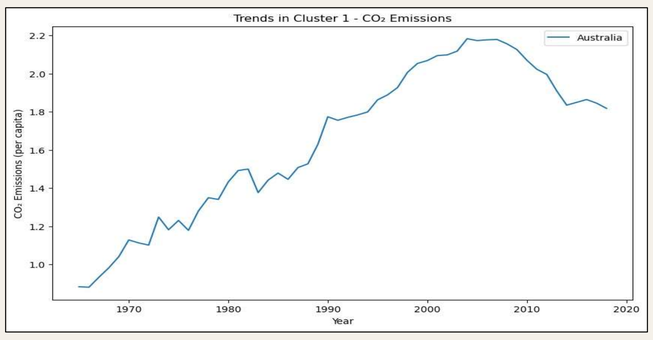
Aim
•Clustering and fitting examinations point to revealing characteristic structures in datasets and show basic designs.
Objectives
•To reveal inherent structures in datasets and demonstrate fundamental designs.
•To recognize normal groupings or clusters of interior datasets.
•To assist in recognizing likenesses and contrasts among information focuses.
•To assist in categorizing and labelling data based on shared characteristics.
•To make exact models that capture connections and plans interior the Information.

	Entity	Code	Year	Annual CO₂ emissions (per capita)	GDP per capita
2	Afghanistan	AFG	1950	-0.524823	-0.644451
3	Afghanistan	AFG	1951	-0.524706	-0.643225
4	Afghanistan	AFG	1952	-0.524727	-0.641561
5	Afghanistan	AFG	1953	-0.524482	-0.637096
6	Afghanistan	AFG	1954	-0.524507	-0.636658

	417485-annotations	Population (historical estimates)	Continent	cluster
2	NaN	-0.140209	NaN	0
3	NaN	-0.139727	NaN	0
4	NaN	-0.139219	NaN	0
5	NaN	-0.138706	NaN	0
6	NaN	-0.138178	NaN	0

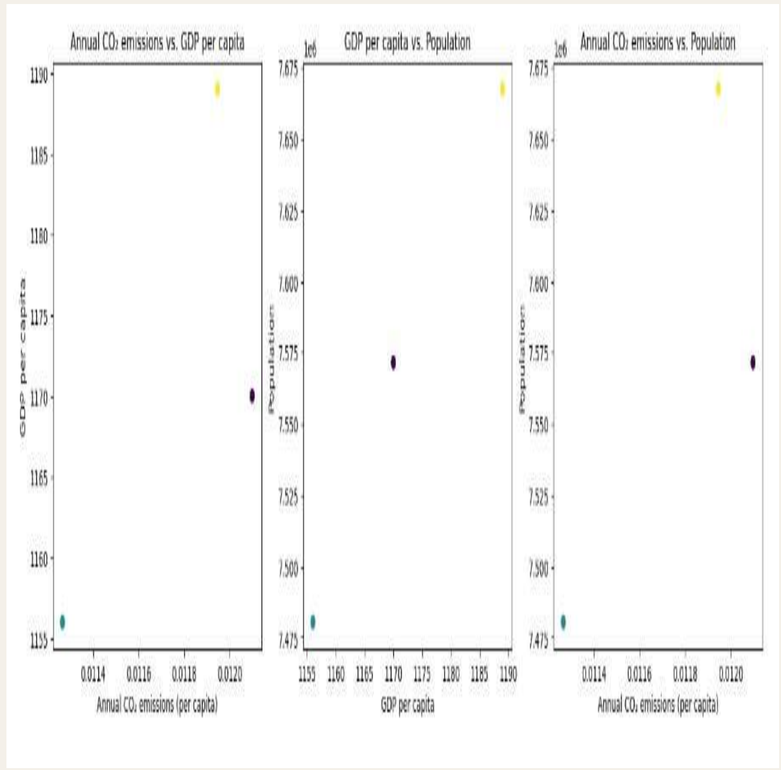
Background

•The background is driven by information science procedures such as clustering and fitting models to determine valuable data from multi-dimensional datasets. The utilize of clustering algorithms permits the discovery of basic structures within the information, thus encouraging design distinguishing proof and classification. Fitting the models permits the modelling of covered-up patterns, increasing prescient control. In this way, information disclosure is achieved giving informed decision- making forms and data extraction from heterogeneous information.



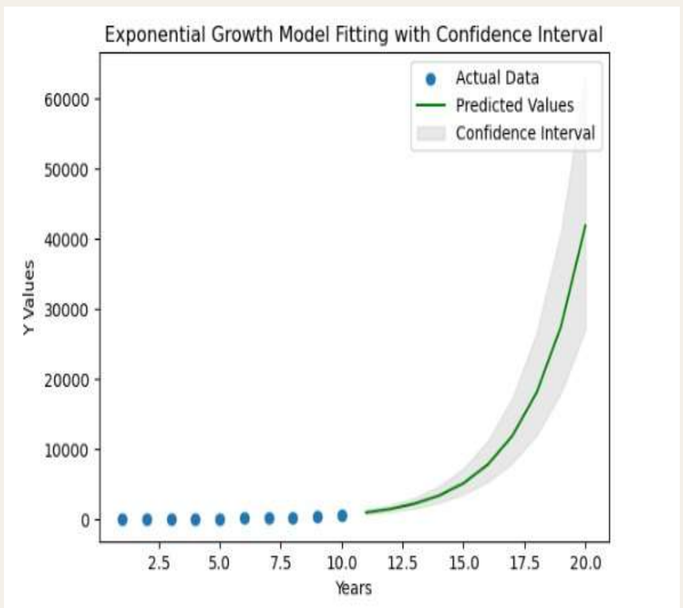
Results Goal 1

•The image shows how to present the libraries pandas and sci-kit-learn utilizing the pip command in a terminal. The output appears that both libraries have as of now been presented.
•The above image shows the yield of printing the column names of a Pandas Data Frame. The Data Frame incorporates data about distinctive nations, such as their code, CO2 emissions, GDP, population, and continent.
These visualizations offer experiences into the connections between CO2 emanations, GDP per capita, and the population.



Goal 2

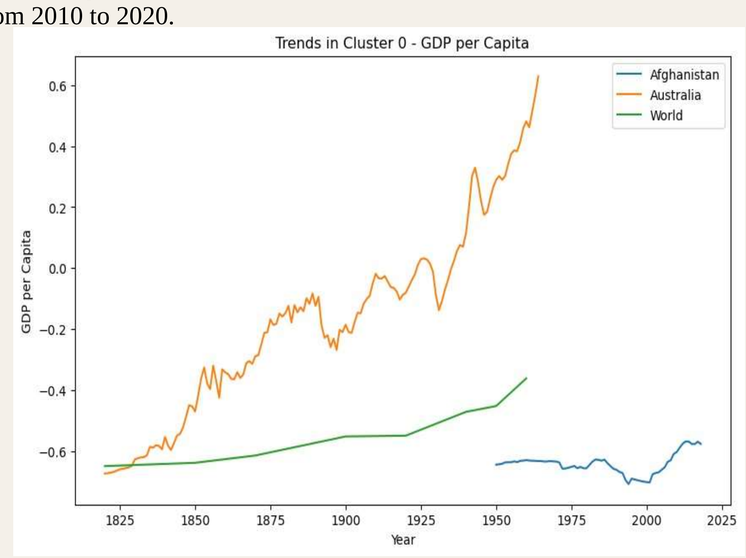
The above code image brings in three libraries commonly utilized in information science: NumPy for effective scientific operations on clusters, matplotlib for information visualization, and SciPy. Optimize for optimization tasks (Migkas *et al.* 2020).
The above image displays the code which is written for plotting the exponential growth model fitting with confidence interval.
The above graph shows up to show the results of fitting an exponential development show to a few data, besides a certainty interval (Donor *et al.* 2020). The curve of the blue line suggests an expanding trend over time, with the shaded area around it representing the run of values inside which the true values are likely to drop 95% of the time.
This suggests that another step involves numerical computations, making plots or charts, and possibly fitting models to optimize certain criteria.
The code characterizes the model, fits it to test information, extricates the fitted parameters, and calculates the standard blunders utilizing the diagonal of the covariance matrix.
The above image displays the code which is written for plotting the exponential growth model fitting with confidence interval.

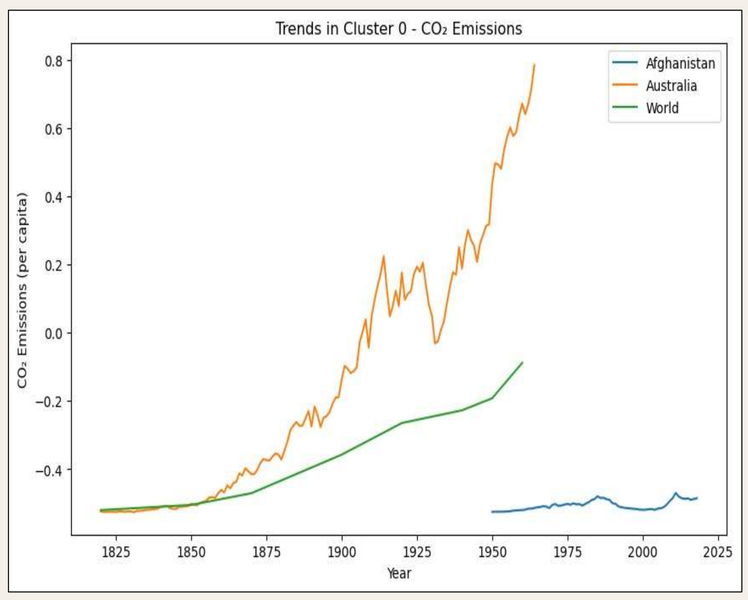


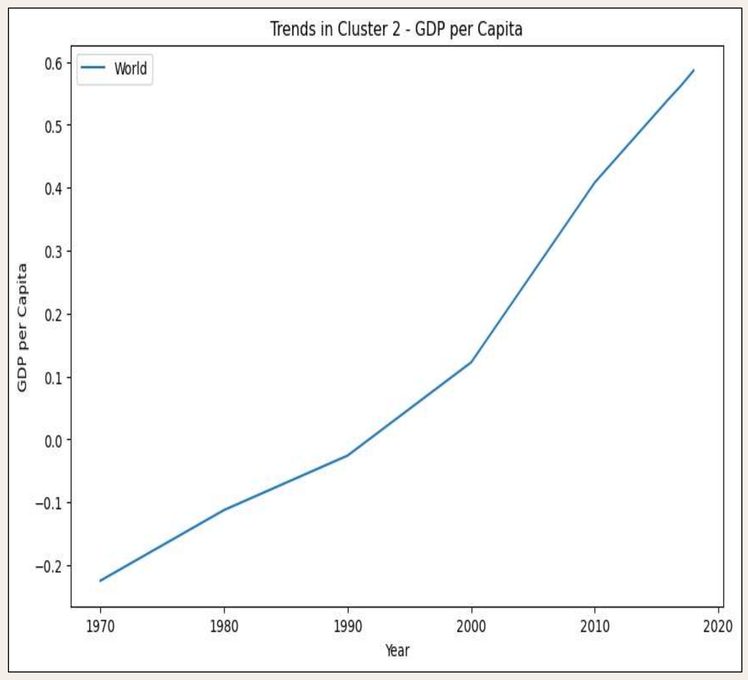
Goal 3

The picture shows the output of printing the primary five lines of a Pandas Data Frame utilizing the head() strategy.
Each push represents a country, distinguished by its code and year (Mach, 2020).
The columns appear in different details like annual CO2 emissions per capita, population, and GDP per capita.
It to begin with chooses these highlights, normalizes them, and checks for an existing "cluster" column before dropping it.
The above image displays the full code for plotting the cluster graphs and the data frame headings.

The code applies K-means clustering to group nations based on three highlights: annual CO2 emissions per capita, GDP per capita, and population (Heymans *et al.* 2021). It to begin with chooses these highlights, normalizes them, and checks for an existing "cluster" column before dropping it.
The data frame shows annual CO2 emanations per capita, GDP per capita, and populace gauges for each nation over a year. The beat rows include sections for Afghanistan from 1950 to 1954.
The graph shows the patterns in CO2 emissions for a cluster of the nations analysed. There appears to be a common downward trend in CO2 emissions for clusters over the long time that appeared in the graph, from around 0.8 metric tons per capita in 1825 to around 0.2 metric tons per capita in 2020.
The graph shows a mixed trend in CO2 emissions for cluster 2 over the long time 1970 to 2020. There is a slight diminish from 1970 to 1990, followed by a slow increase until 2010, and then a more extreme increase from 2010 to 2020.







Conclusion

•Overall, clustering and fitting results shed light on the complex dynamics between economic drivers and carbon emissions at various levels across entities and periods.
•The detected clusters revealed groups of entities demonstrating similar trends in GDP per capita, population, and CO_2 emissions.
•This information can guide specific policy intervention and sustainable development paths.
•Further, the incorporation of Abkhazia also shows a need for taking into consideration different entities while analysing the whole world. However, this is how contribute to a deeper understanding of the subtle relationships between economic development and environmental impact toward informed decision- making aimed at sustainable future.