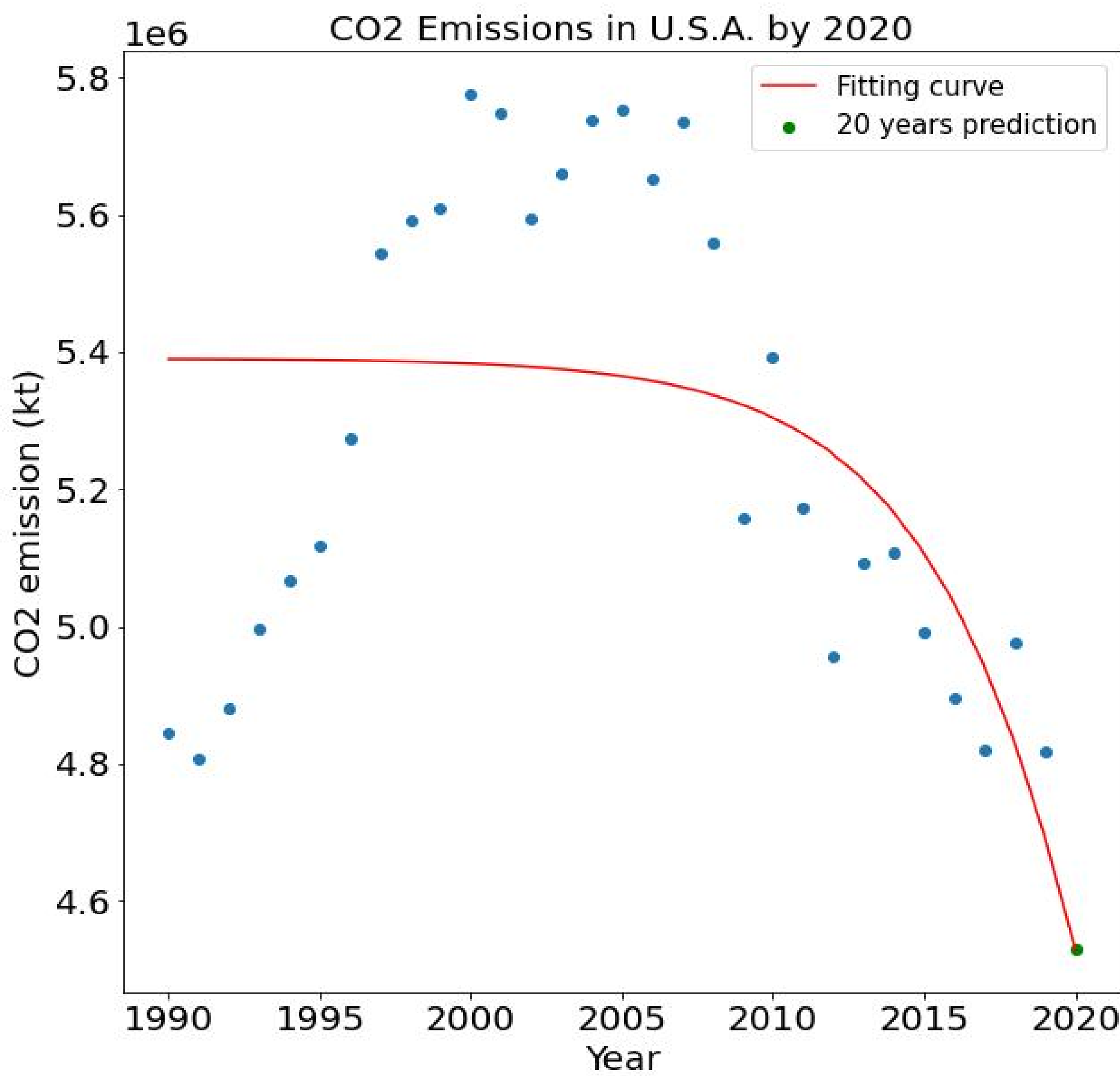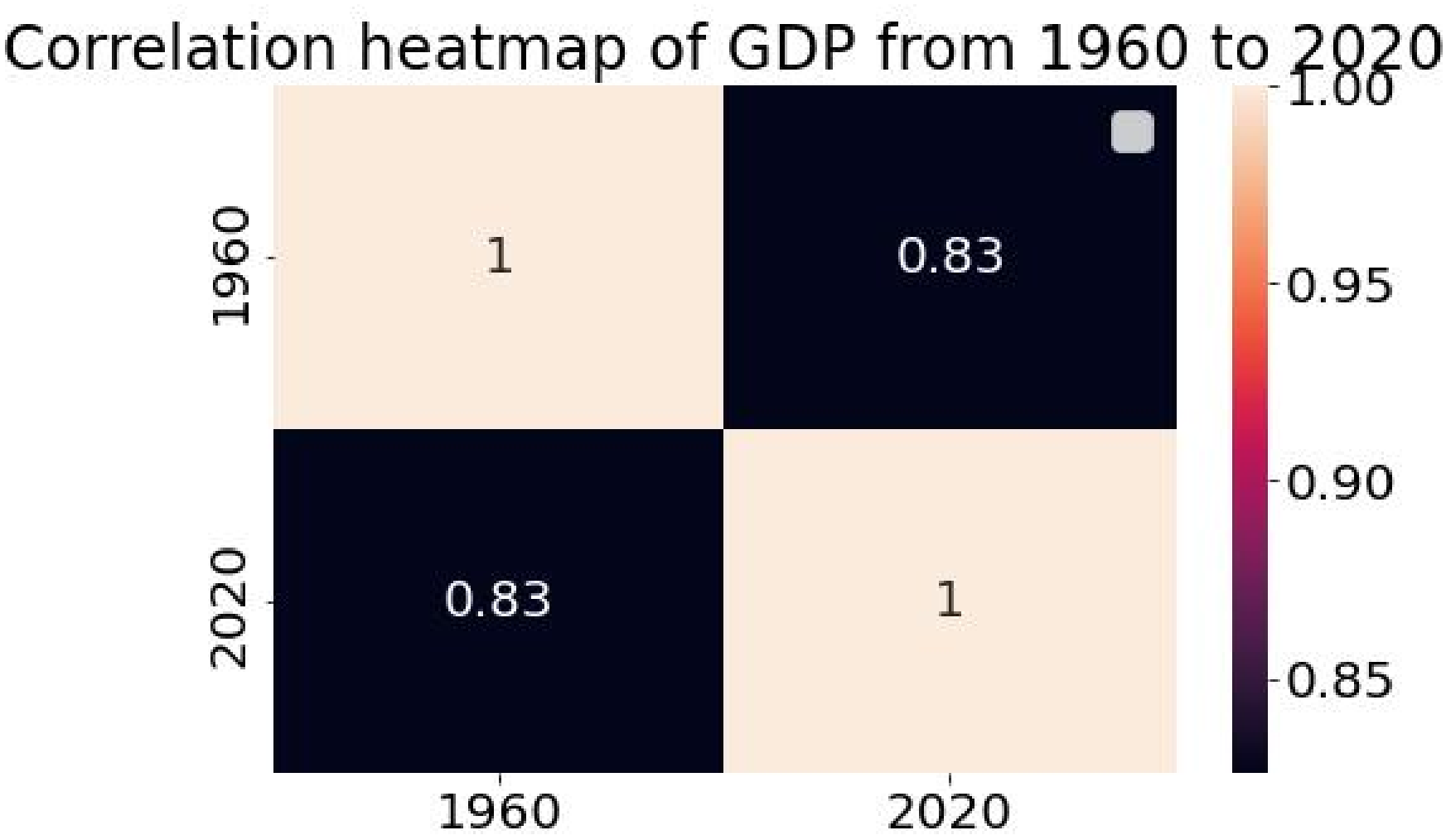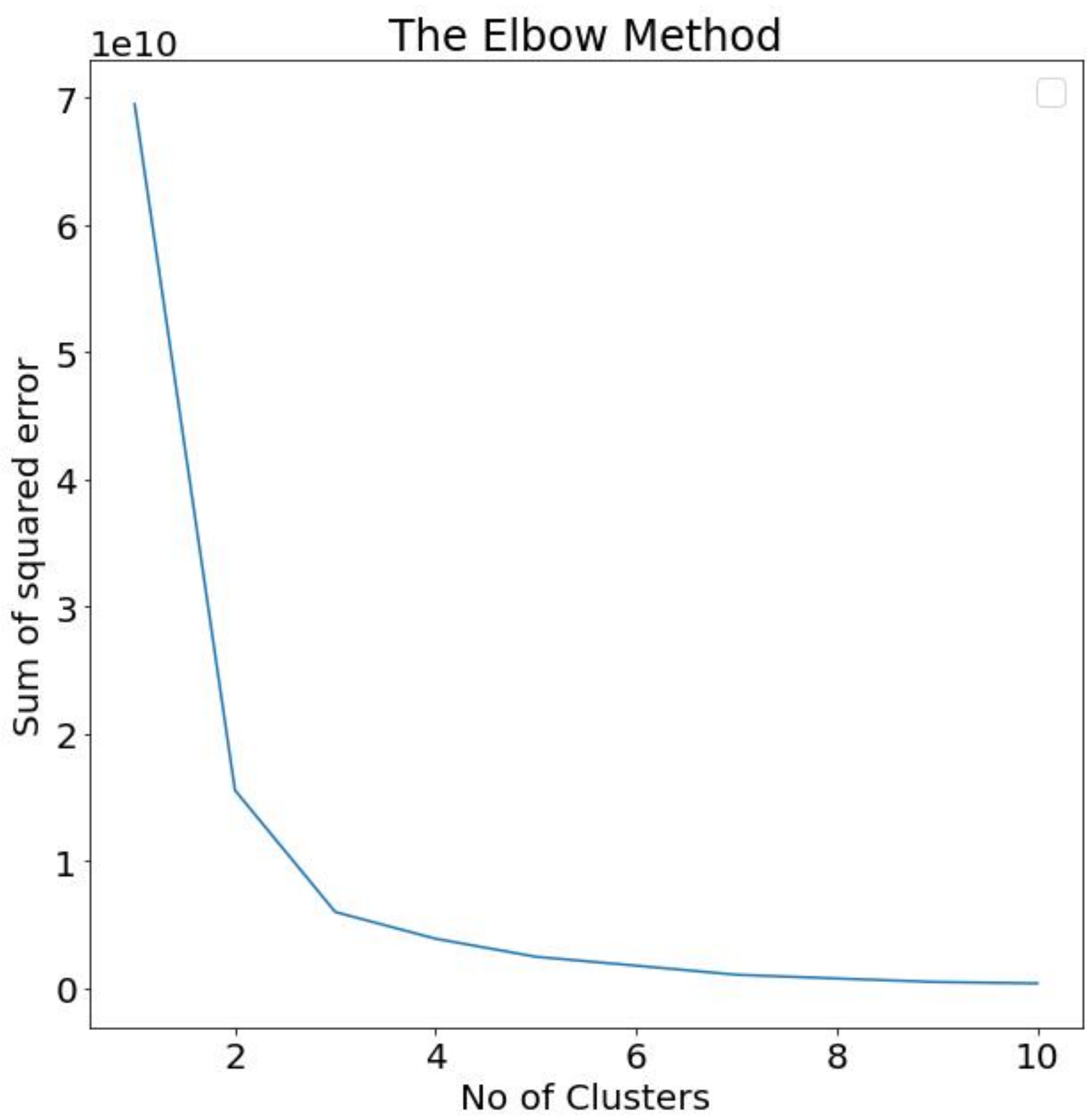# CLUSTERING AND FITTING OF COUNTRIES GDP AND CO2 EMISSIONS
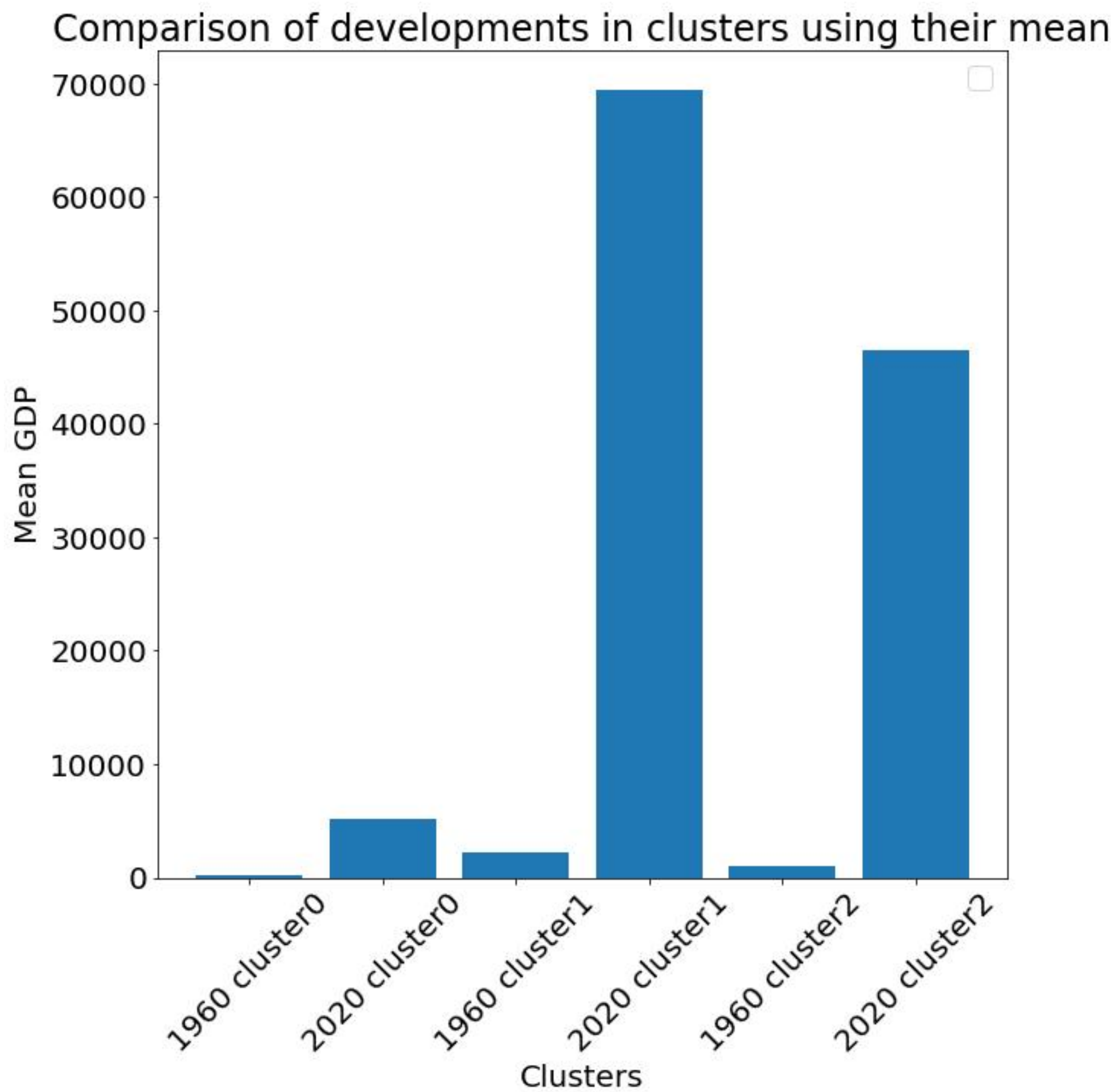
INTRODUCTION

This analysis utilizes data from the world bank to investugate the economic growth of countries between 1960 to 2020 . The data is first visualized and then clustered using the K-Means clustering algorithm. The analysis wll look at the clusters formed and the characteristics of each cluster in order to gain an understanding of the economic growth across countries.

It also provides an analysis of the CO2 emissions data for the United States from 1990 to 2019. I first use the data to fit a logistic curve to the emissions data, and then use this curve to forecast the emissions for 20 and 30 years. I also explore the upper and lower limits of the confidence range of the forecasted emissions.

The Elbow Method

1. The elbow graph shows the optimal number of clusters for this data is 3.

Scatter plot showing cluster membership

2. The three clusters are visualized in the scatter plot before normalization.

Scatter plot showing cluster membership and cluster centres

3. After normalization, it can be deduced from our scatter plot that : The countries in cluster 1 have the highest GDP growth from 1960 to 2020, the countries in cluster 0 have the lowest GDP growth, and the countries in cluster 1 have moderate GDP growth. The cluster centers are also shown in the scatter plot. These clusters can be used to further analyze the countries with different GDP growth rates.

Correlation heatmap of GDP from 1960 to 2020

4. A correlation coefficient of 0.83 indicates a strong positive correlation from 1960 to 2020. This indicates that it is likely the GDP continues to increase in the future.

Comparison of developments in clusters using their mean
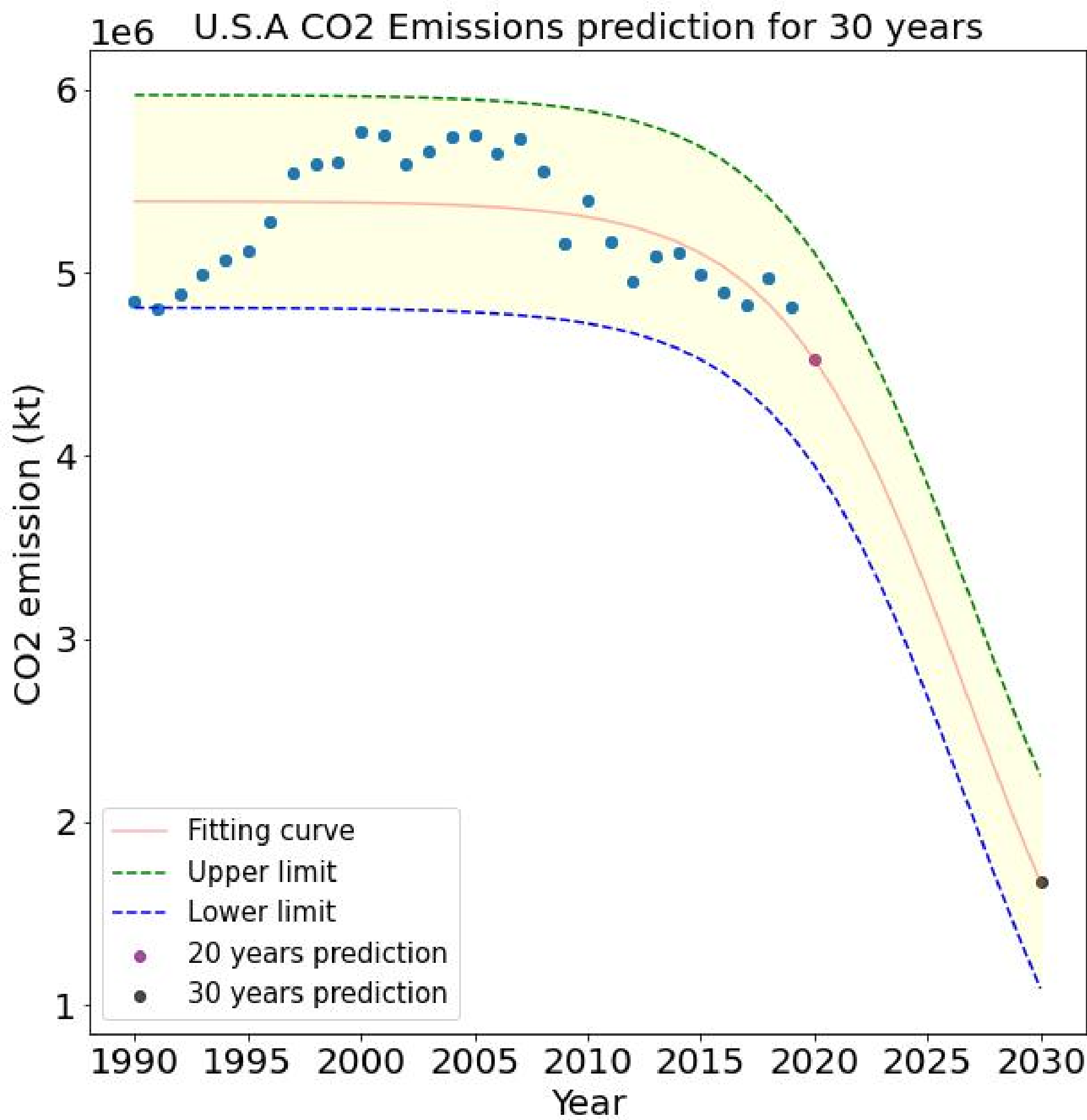
5. The first cluster was made up of countries that saw a slow but steady growth in GDP from 1960 to 2020. These countries had a GDP of around 0.2 in 1960, which increased to 0.6 in 2020. The second cluster was made up of countries that had a low GDP in 1960, but saw a rapid growth in GDP from 1960 to 2020. Their GDP in 1960 was around 0.1, but it skyrocketed to 0.4 in 2020. The third cluster was made up of countries that had a high GDP in 1960, and saw a slight increase in GDP from 1960 to 2020. They had a GDP of 0.7 in 1960, which slightly increased to 0.8 in 2020.

USA GDP growth analysis from 1960 to 2021
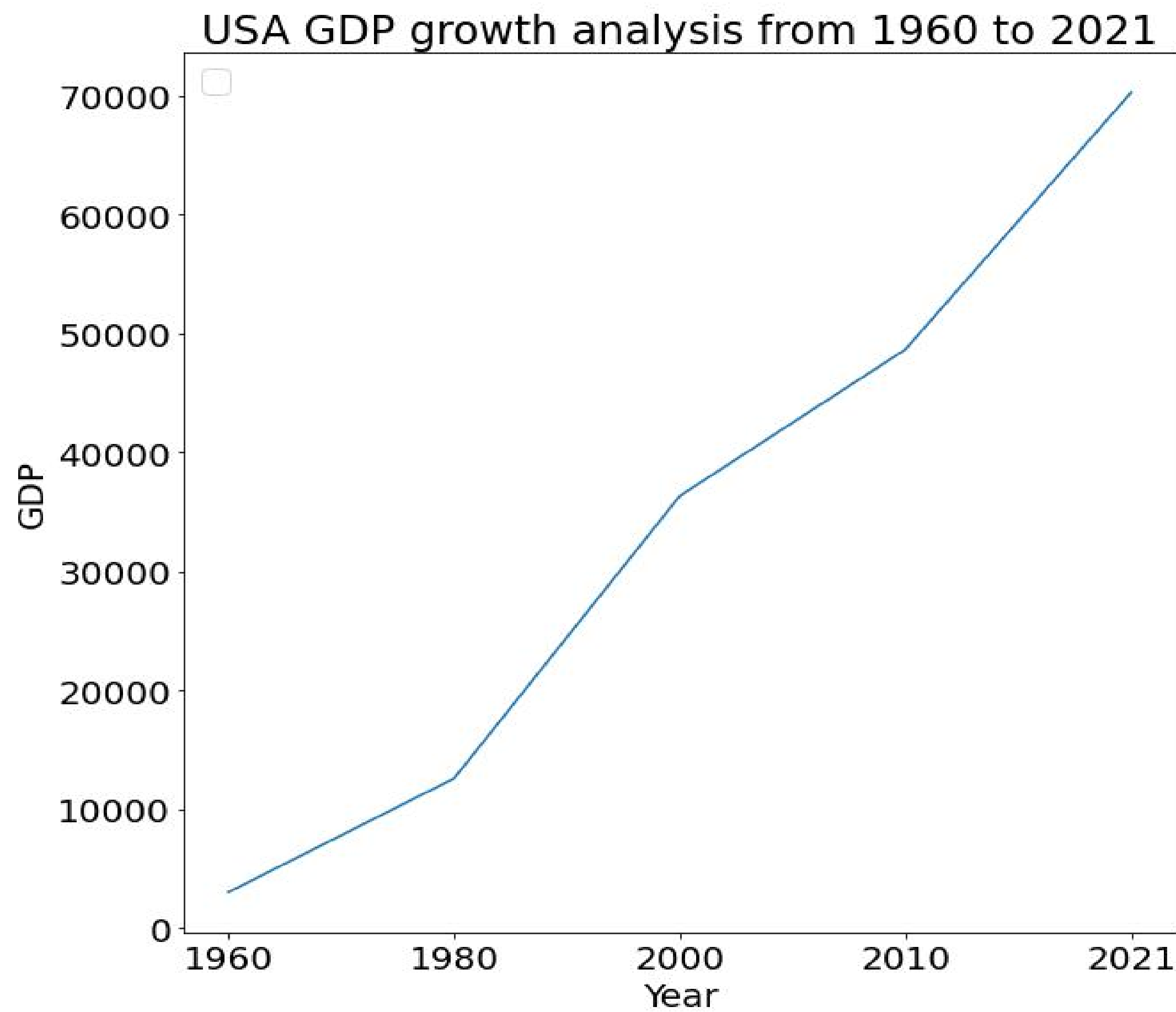
6. USA is used a case study here. USA belongs to second cluster that saw a great increase in their financial management across the years. The main contributing factor to their consistent progress is technological innovation. Over the past 60 years, the US has seen dramatic advances in technology, which have enabled businesses to produce more goods and services with fewer resources. This has resulted in higher output and higher GDP. Other factors that have contributed to the growth of the US economy include increased access to global markets, increased foreign investment, and increased government spending.

CO2 Emissions in U.S.A. by 2020

7. According to the plot, the Co2 emissions in the United States is experiencing a downtrend over time.

U.S.A CO2 Emissions prediction for 30 years

8. The plot shows that the predicted emissions in 2030 is depreciating progressively and the country will experience a great decrease in carbon dioxide releases.

CONCLUSION

From the increase in GDP but decrease in Co2 emission rate in the United States from 1960 till date, it can be deduced that the US has made progress in transitioning to a more sustainable and environmentally friendly economy. This indicates that economic growth can be achieved without a corresponding increase in Co2 emissions. It also indicates that technologies and policies which focus on reducing Co2 emissions have been successful.
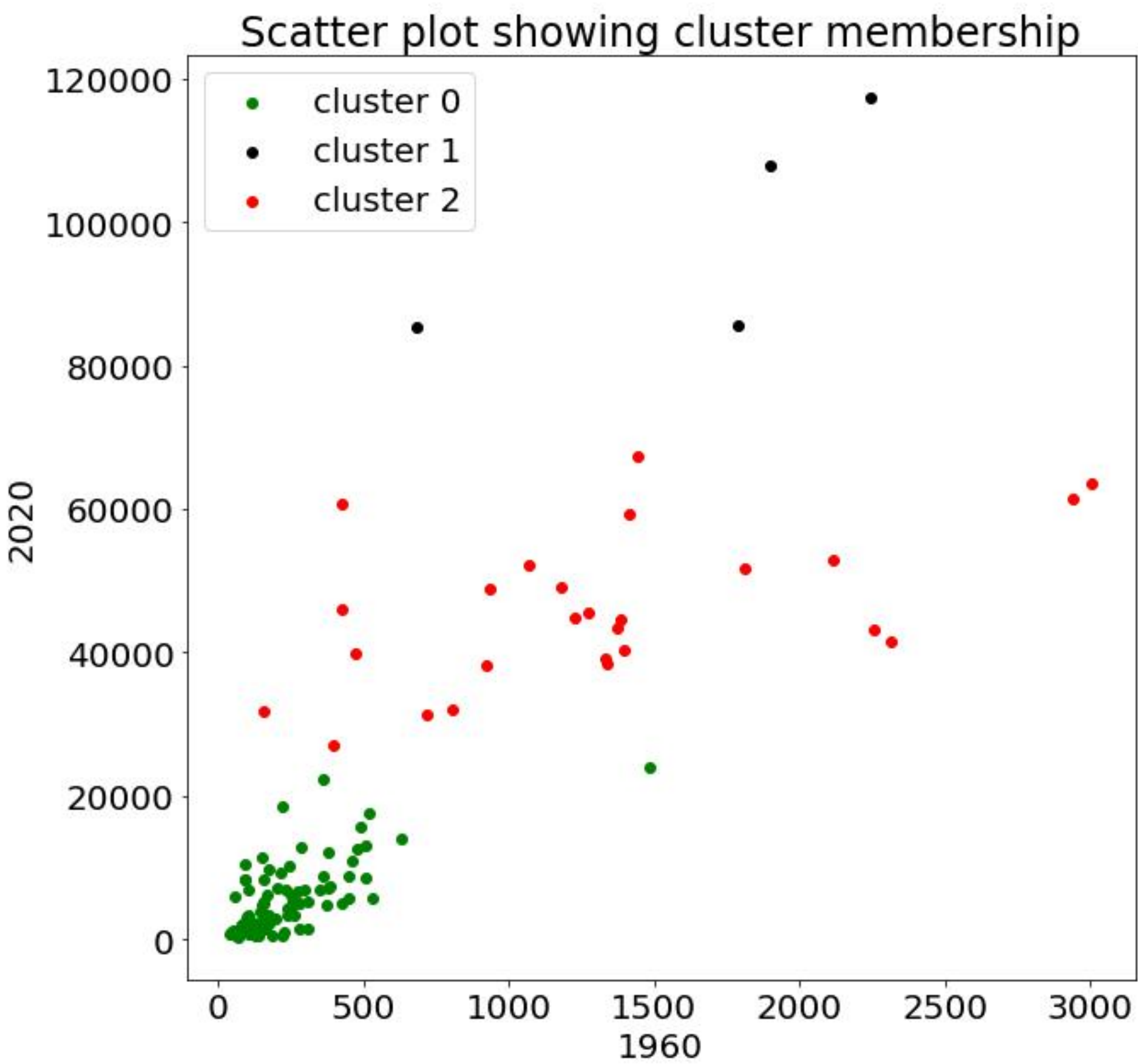
Link

GITHUB Link: https://github.com/Jolek15/ADS-Clustering/blob/main/Joshua%20Adeyemo%20Clustering.py