Creating clustering and fitting on world indicators (GDP, year, country, impact rating)
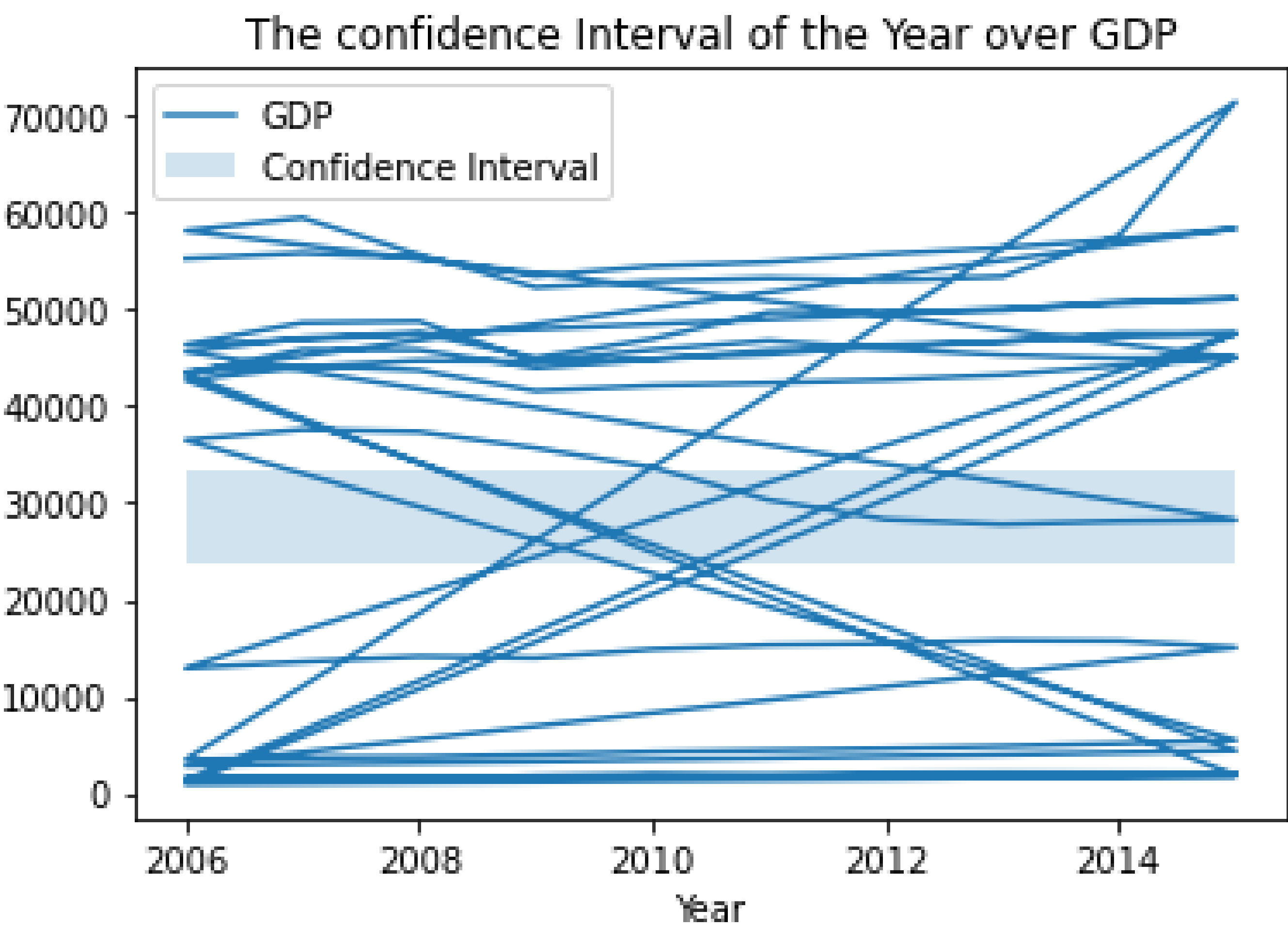
Nwadishi Chukwuebuke

University of Hertfordshire UH

INTRODUCTION TO THE PROJECT

Clustering and fitting are one key techniques I worked on in the project by creating a curve fit on a linear model on the Year and GDP of the selected countries and also finding the confidence interval of GDP over the years, while I finally created a cluster model using Kmeans.
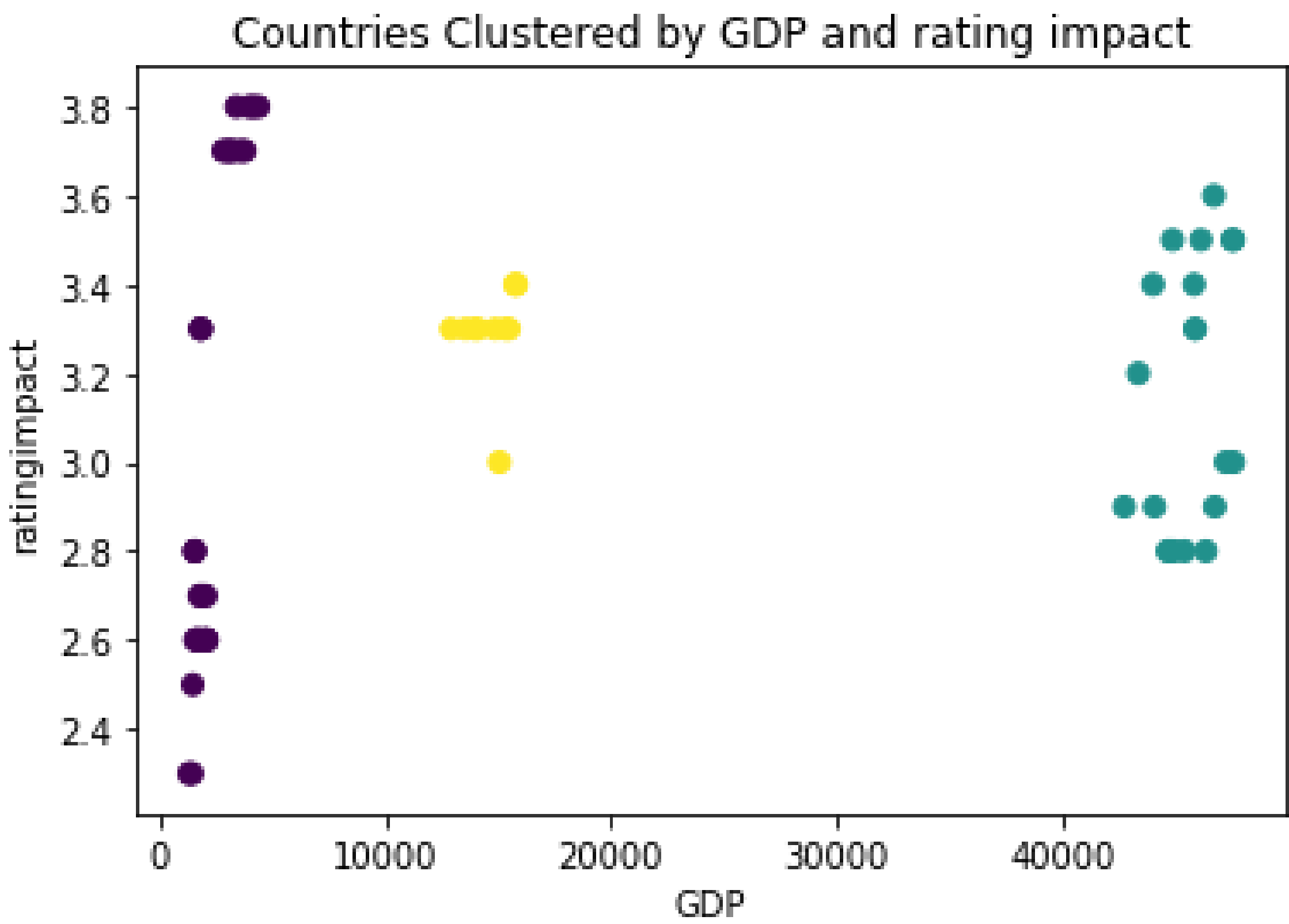

GDP Fitted with a Linear Model
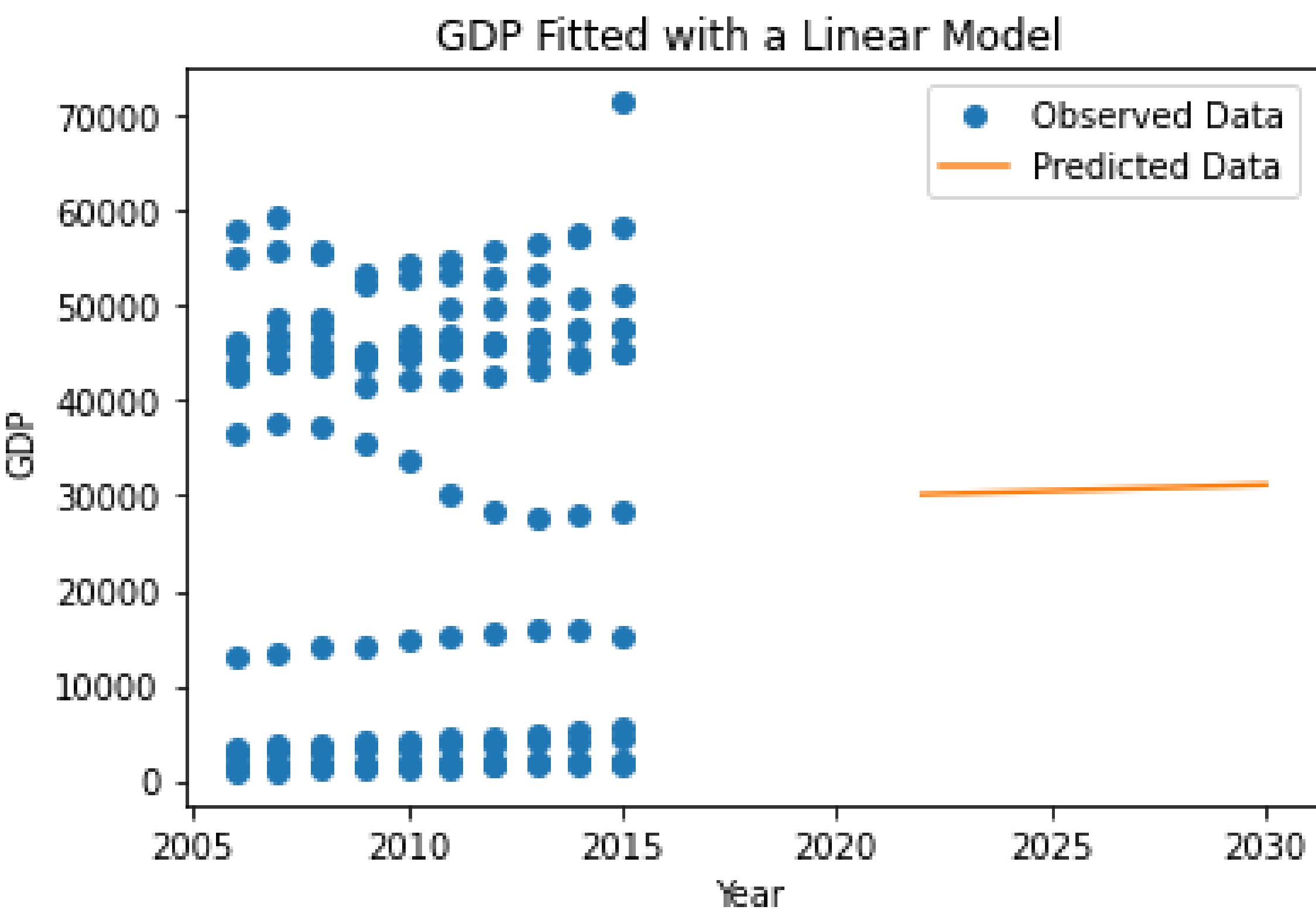
I performed fitting with curve_fit on GDP data, you will need a dataset that contains GDP values for different countries or regions over time. You can then use curve_fit to fit a mathematical model to the data and predict future GDP values.

Countries Clustered by GDP and rating impact

The resulting plot shows two clusters, one with lower GDP and population values and another with higher GDP and population values. This suggests that there is a relationship between GDP and population, where countries with higher GDP tend to have larger populations.

I used the KMeans clustering algorithm to group countries into clusters based on certain features. For instance, we can use data on the gross domestic product (GDP), population, and life expectancy of different countries to group them into clusters.

I compared countries from different clusters and see how their GDP, population, and life expectancy differ. For example, we can pick a few countries from one cluster and compare them with countries from another cluster or different regions.

The confidence Interval of the Year over GDP

The confidence interval plotted on the line chart provides a range of values that are likely to contain the true mean of the sample (in this case, the average GDP over the given years). The confidence level indicates the probability that the true mean falls within this range. For example, if the confidence level is 95%, there is a 95% probability that the true mean is within the confidence interval.

The confidence interval plot above is fairly narrow, which suggests that the sample data is relatively consistent and the true mean (the average GDP over the given years) is likely to be close to the sample mean. However, you should always keep in mind that the confidence interval is just a statistical estimate, and there is always some uncertainty about the true mean of the data.