University of Hertfordshire UH

Clustering and Fitting Analysis

Abstract:
This script investigates global energy and economic variables using data analytic techniques. The CSV file containing the data is imported, cleaned, and transposed in preparation for additional analysis. There are two major analyses carried out: Time series analysis is used to forecast future values of a given characteristic for a subset of nations, while K-Means clustering is used to find patterns in energy-related data. The findings shed light on how nations are grouped according to attributes linked to energy and potential developments in the availability of power.

Introduction:
Finding patterns and trends in the world's energy and economic statistics is the goal of the analysis. Access to electricity, coal rents, mineral rents, and natural gas rentals for a range of nations and years are all included in the dataset. There are two primary analyses carried out: Time series analysis is used to forecast future values and K-Means clustering is used to group nations with comparable energy profiles.

Data loading and cleaning:
The Pandas library is used to load the data from the supplied CSV file. The mean of each column is used to impute missing values once numeric columns have been translated to the proper data types. After that, the data is transposed to make analysis easier.

K means clustering
The script is appropriate. Using K-Means clustering, groups of nations may be found according to certain characteristics, such as coal, natural gas, and power availability. After normalizing the data, clustering is done. The quality of the clustering is assessed using the silhouette score. A scatter plot displaying the clusters and their centers at the original scale is created by examining the properties of each cluster.

Result:
Three unique groupings of nations with various energy characteristics are shown by the K-Means clustering. High coal rents are found in Cluster 1, high mineral rents are found in Cluster 0, and reasonably balanced values across characteristics are found in Cluster 2. The time series analysis offers insights into possible trends by projecting future values of "Access to Electricity" for the chosen nations.
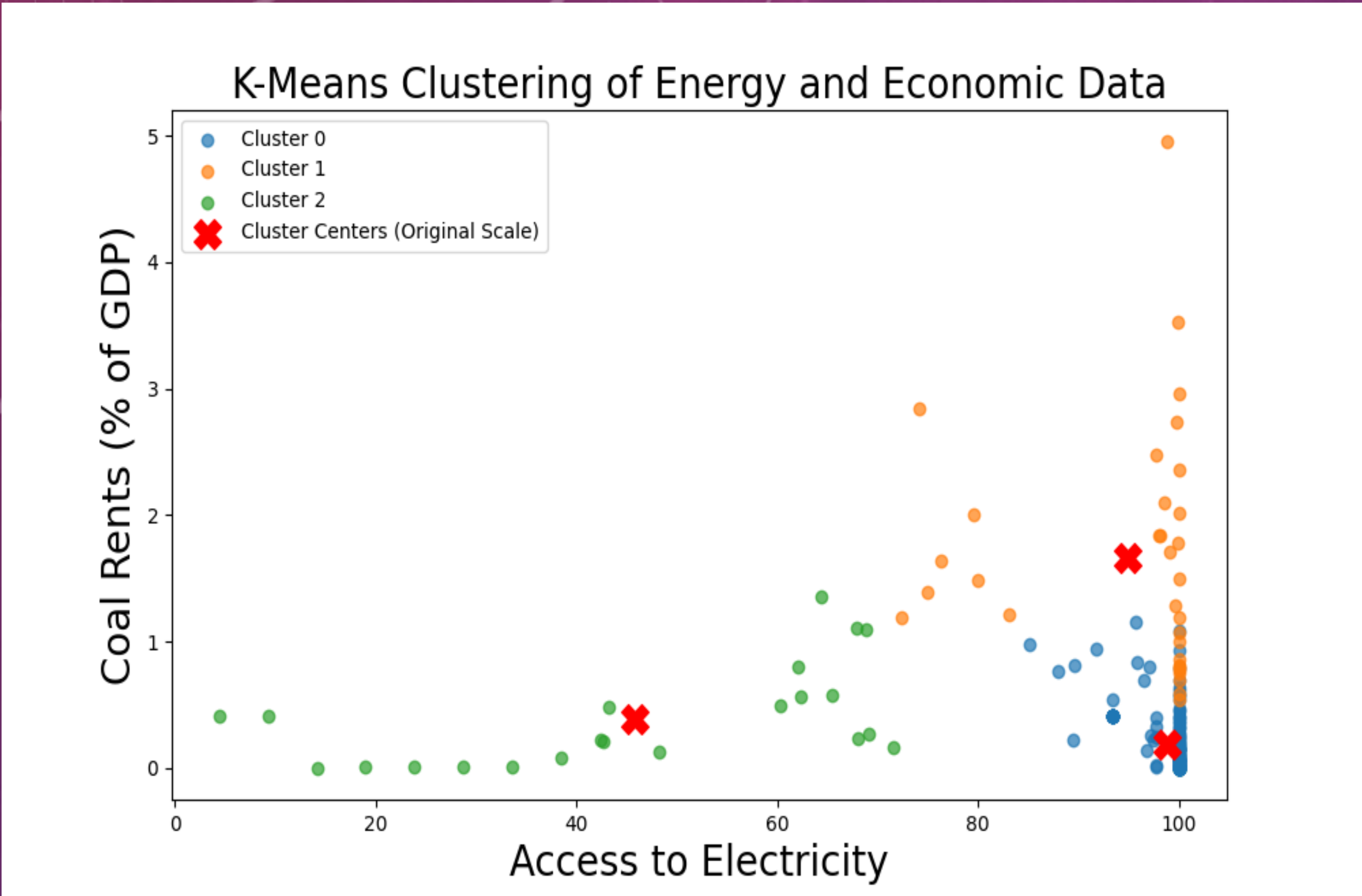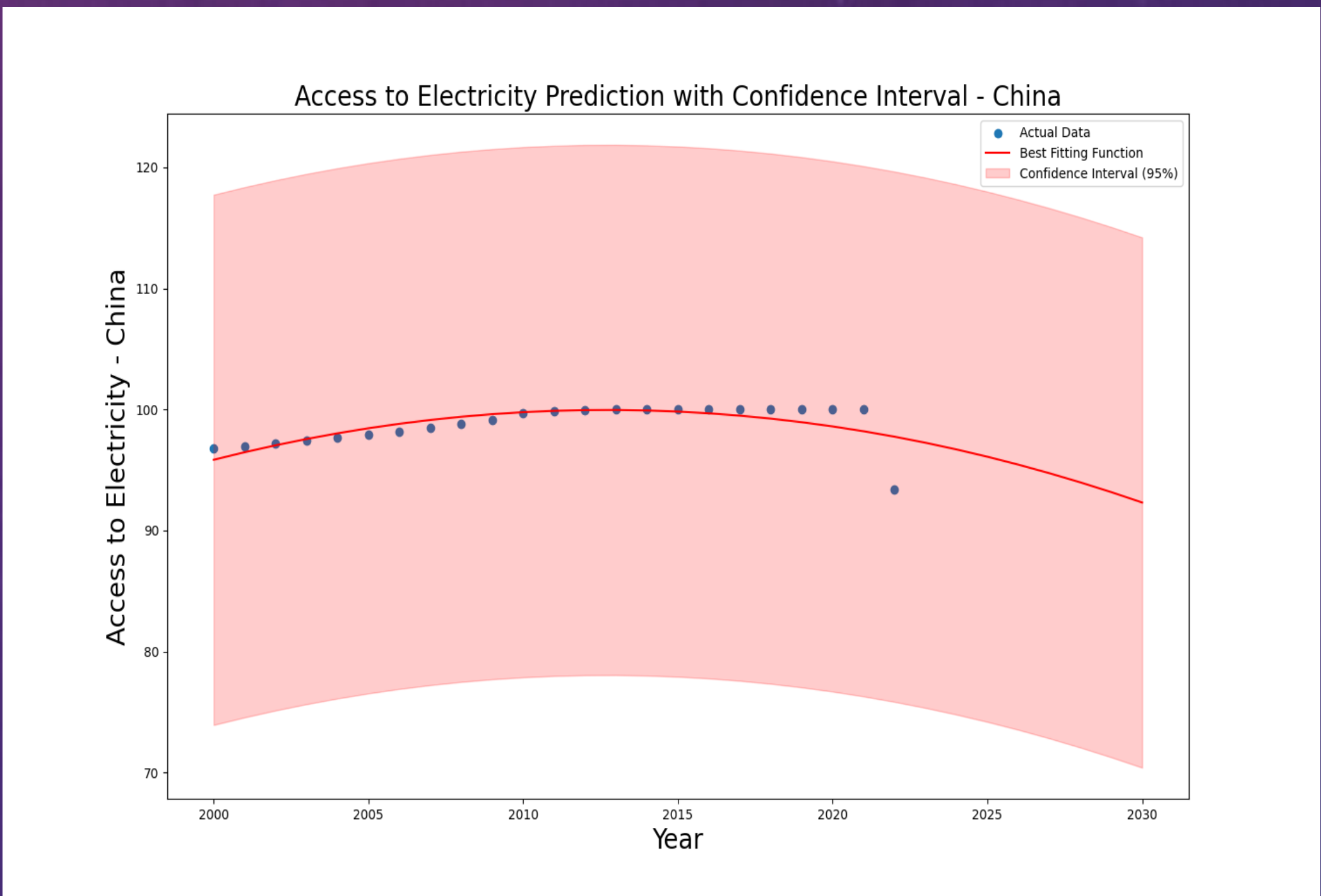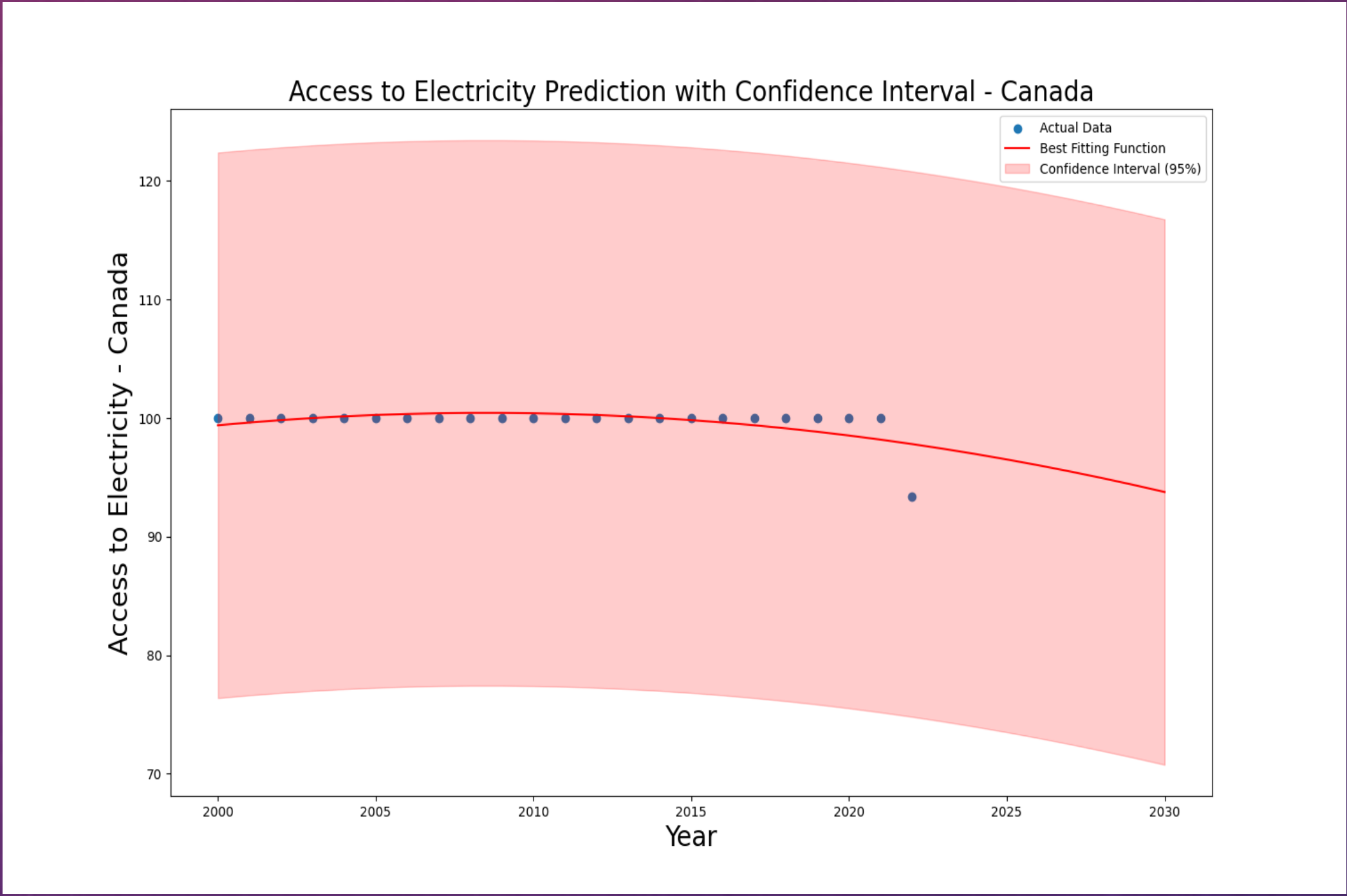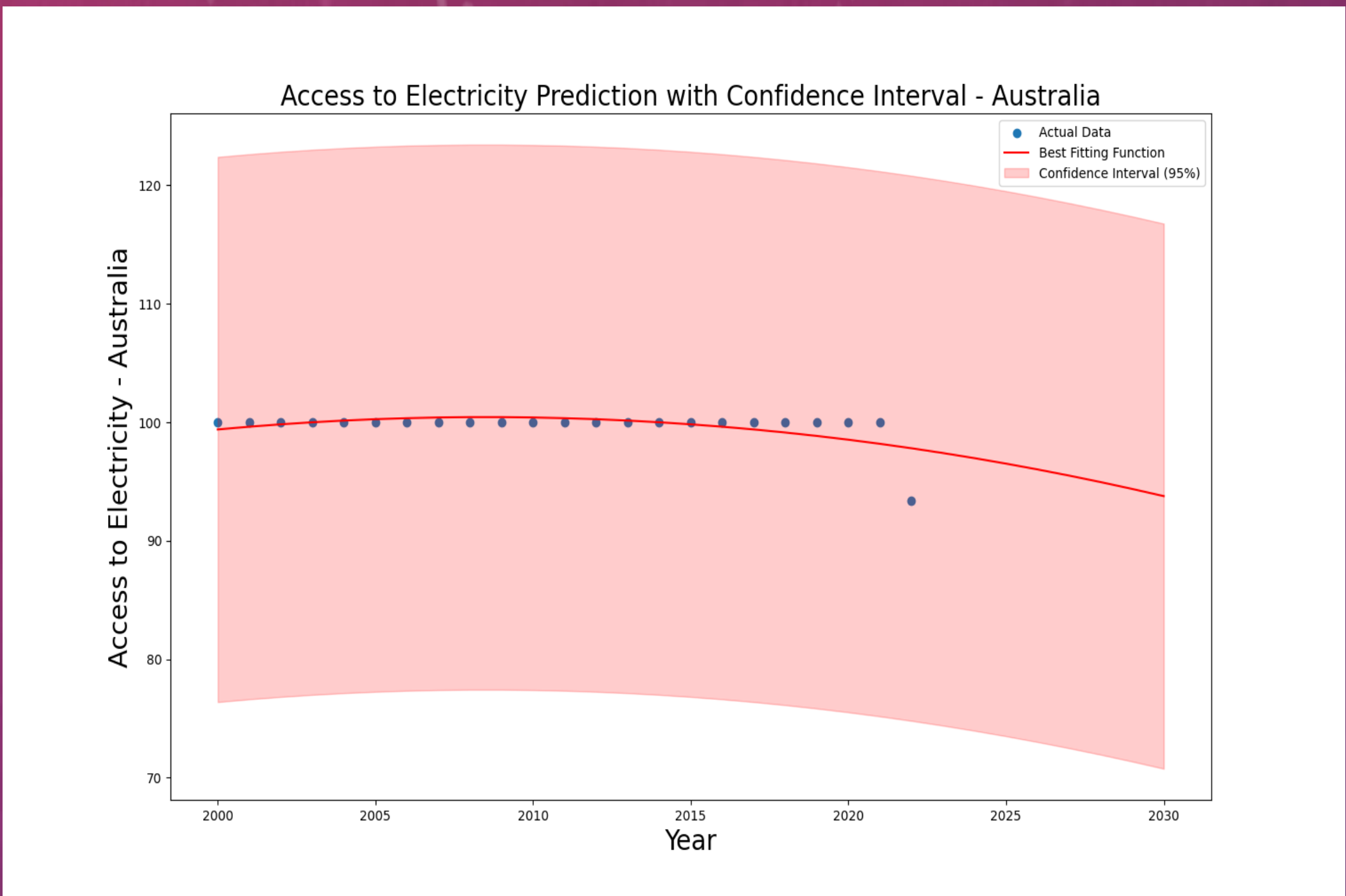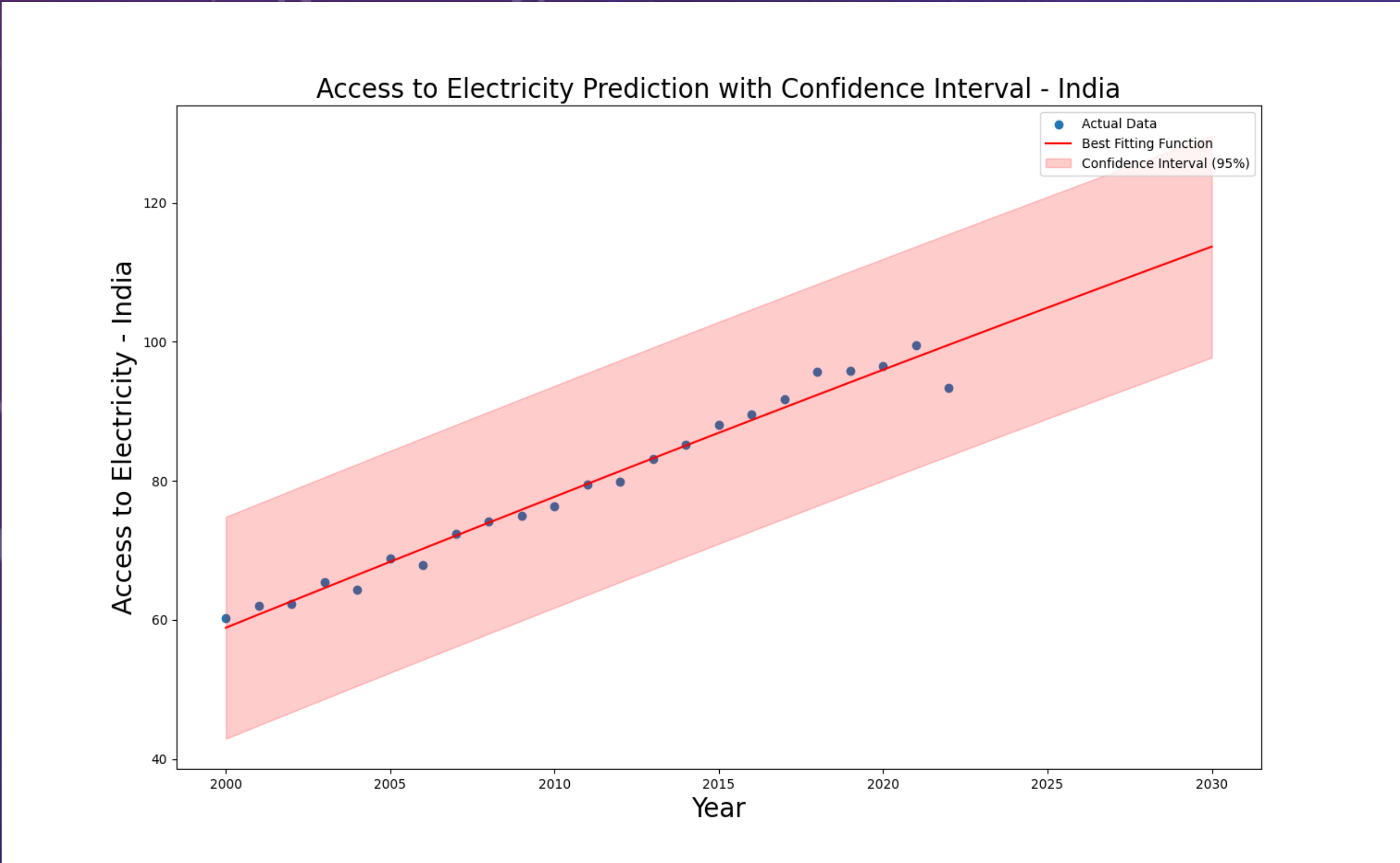

Time series analysis
A time series analysis is done for a few nations (China, Australia, Canada, India), regarding the 'Access to Electricity' characteristic. The script forecasts values by fitting a quadratic model to past data. Together with the real data points and the best-fitting function, a confidence interval (95%) is computed and shown.

Results:
In India, the number of people who can get electricity is rising annually. For China, Canada, and Australia, access to power is declining.

Conclusion:
The screenplay offers a thorough examination of international economic and energy facts. Time series analysis projects future trends in a nation's access to power, whereas K-Means clustering finds country groupings with comparable energy profiles. The comprehension of global energy trends and possible future developments is improved by these investigations.

Datasource: https://databank.worldbank.org/reports.aspx?source=2&series=AG.LND.FRST.ZS&country=

GithubLink : https://github.com/SindhuKavyaAlahari/ClusteringAndFitting.git

Name : Alahari Sindhu Kavya
Student Id : 22070331