Data clustering and Data fitting report by David Adeyemi Runsewe (19046670)
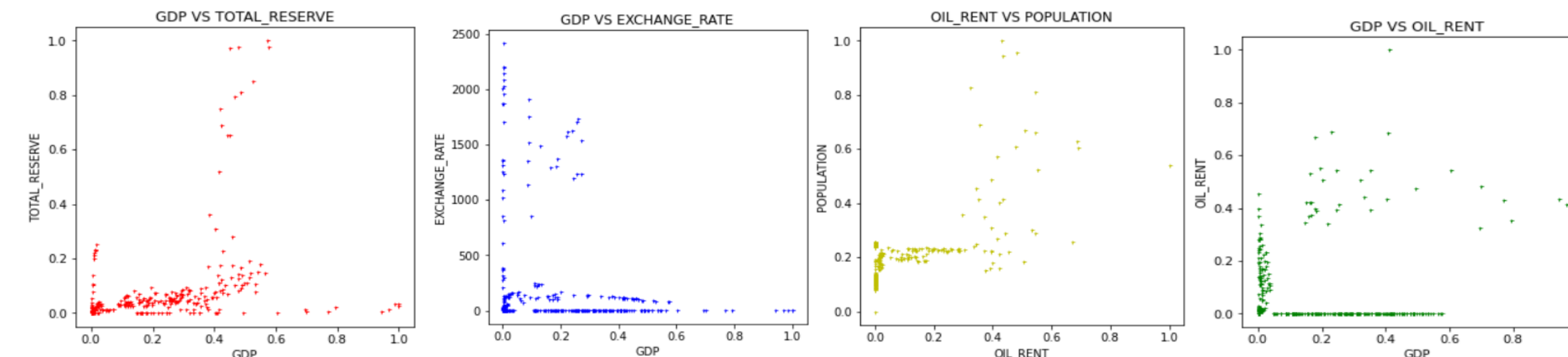
Data Clustering:

Aim : Use a machine language algorithm to find distinct groups or clusters within selected dataset , here we will make use of Kmean Clustering.
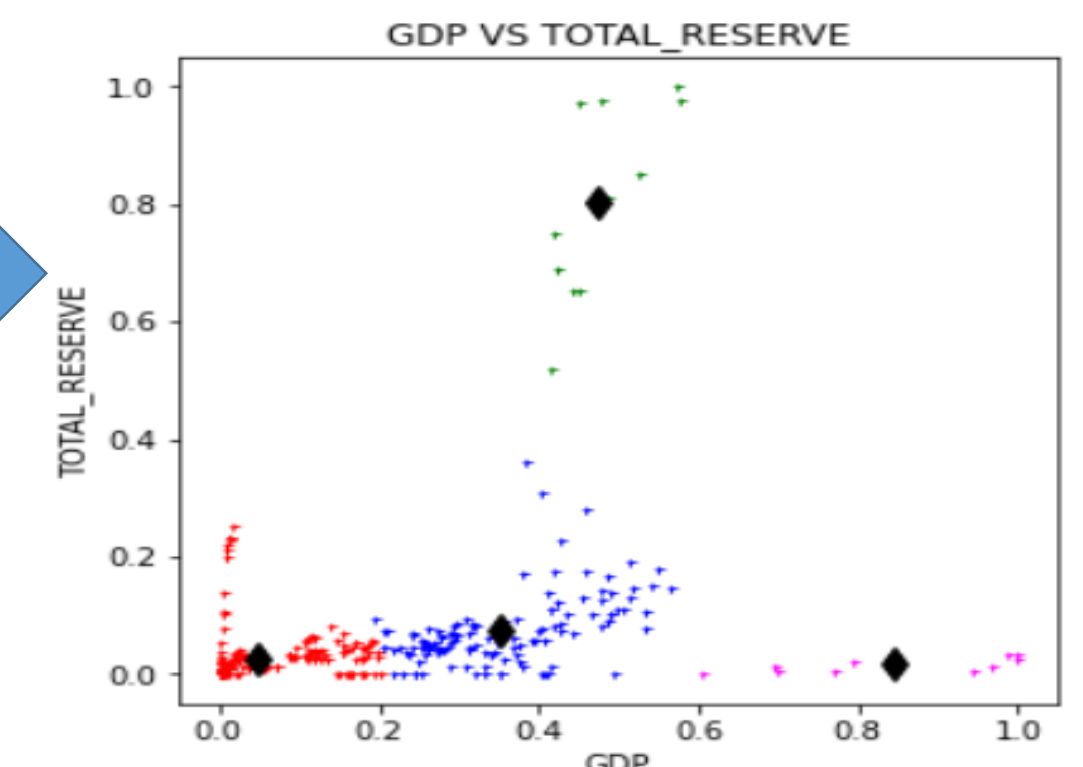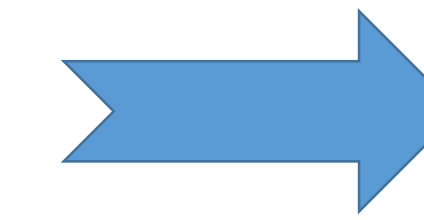
Why Kmean Clustering :
 It scales well for large datasets, and also easily adapts to new examples.

A plot of multiple graphs from the Dataset to see which dataset is most suitable to build a cluster.

Dataset has been neatly divided into 4 clusters using the kmean clustering algorithm.

The dataset used in this analysis is a collection of multiple countries from world bank data and indicators being focused on are GDP per capita, Official exchange rate, Oil rents, Population growth and Total reserve. From year 1980 to 2015.

Data Fitting:

The aim of data fitting by assigning a 'Best fit', function or curve is to capture trend in data that allows prediction of future behavior of data series

Insight on dataset we are fitting: This analysis will try to fit the data from world bank data, we will be looking at fitting Total reserve for the country India over the time range 1960 to 2019. From our the analysis having used done an exponential growth fit and a logistic fit, we were able to make good future predictions with low uncertainty level, below are graphs that says more from the Analysis

Exponential growth plot after conversion has been done, not the best of fit or possible fit outcome expected as our fit line misses out fitting most of the dataset, improvement can be done using logistics function.

Improved starting value from our dataset which will help better fit our data while using logistic function to fit our data.

Fitting in this graph starts with exponential growth and gradually starts to level off, Initial fit start with and exponential growth but as it levels up, this fit improves.

After using our error range function and also the below grow we can see that we do have a very low uncertainty level, hence our data are fitted properly and most of the predictions will likely the correct.
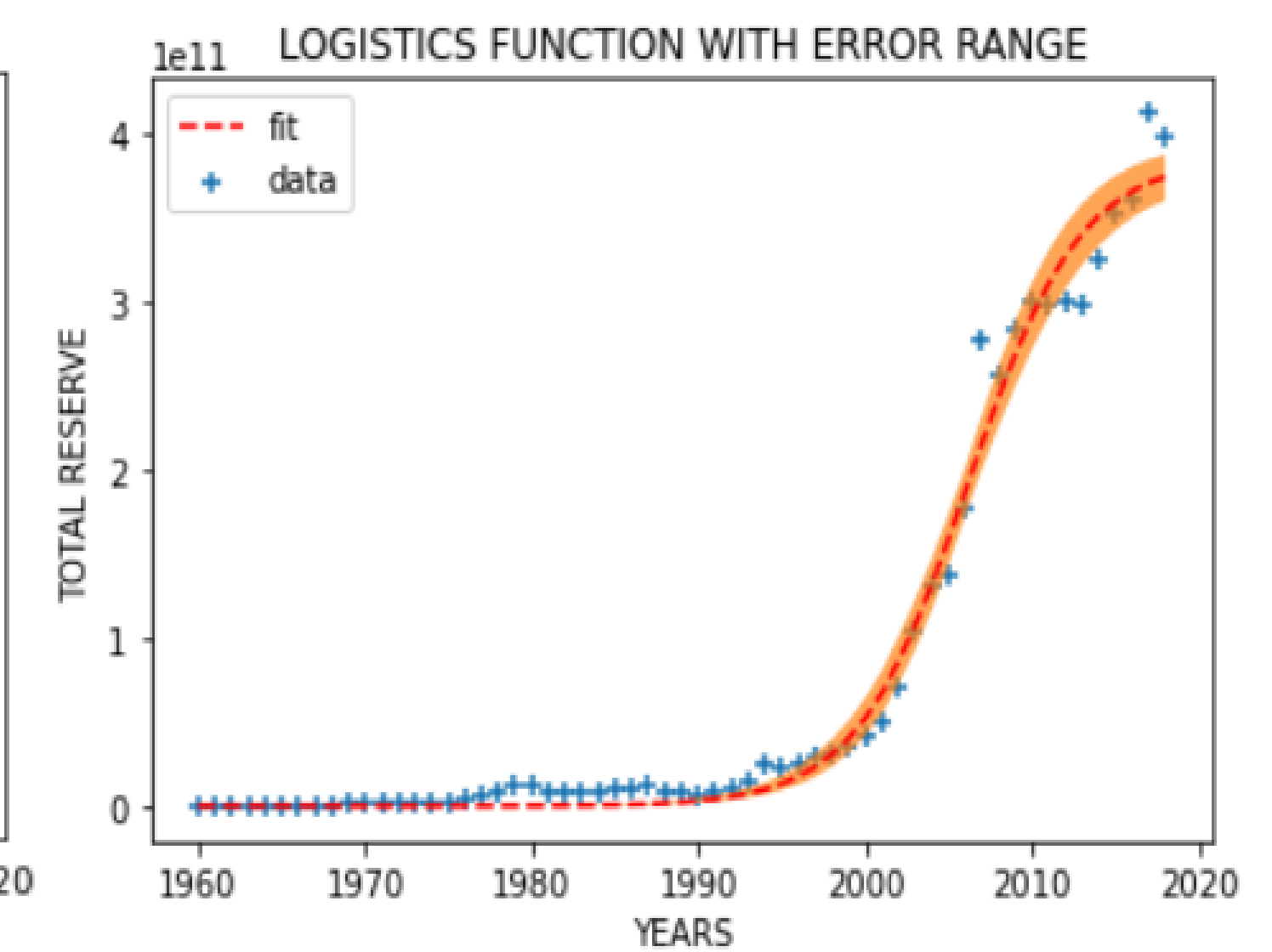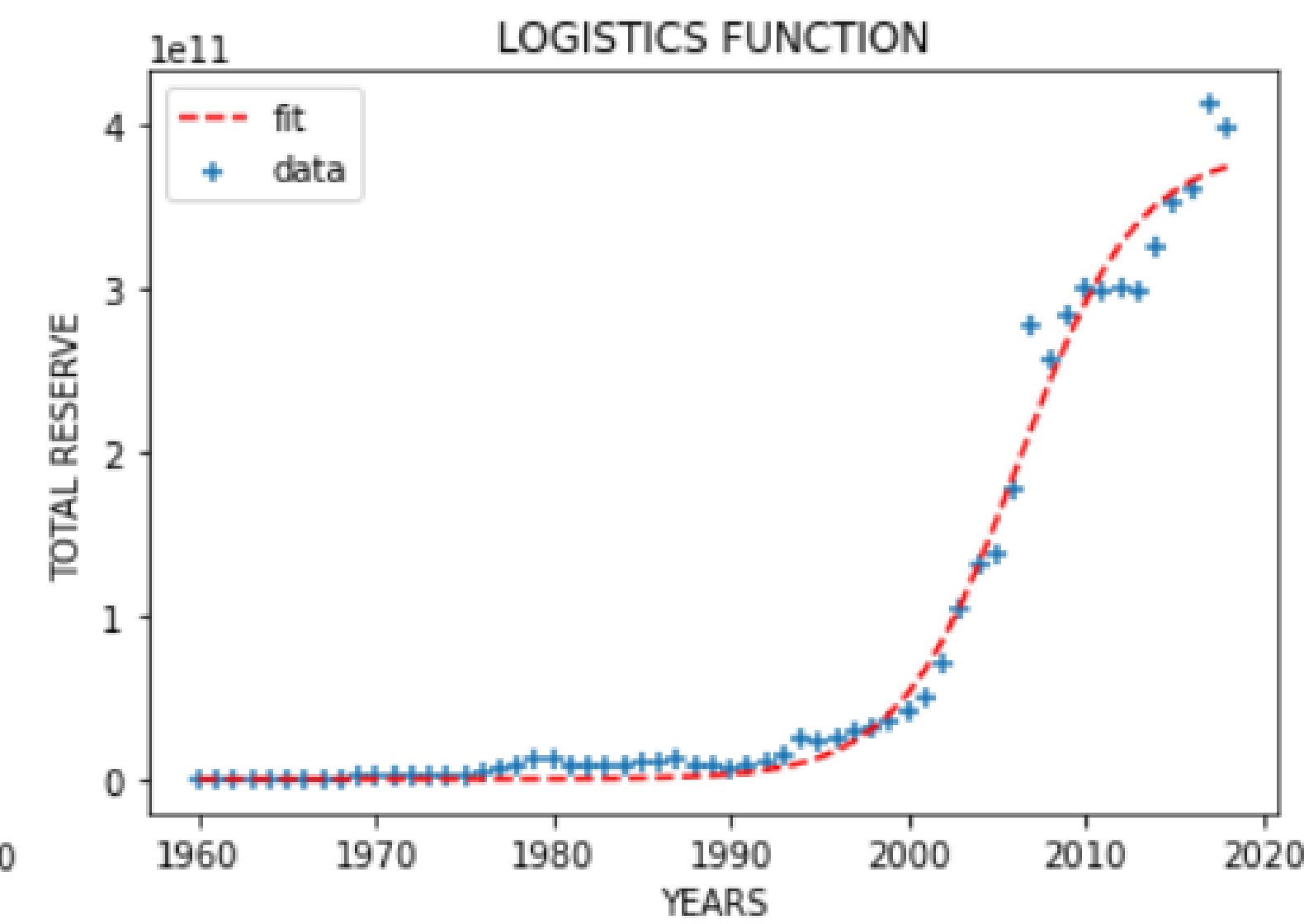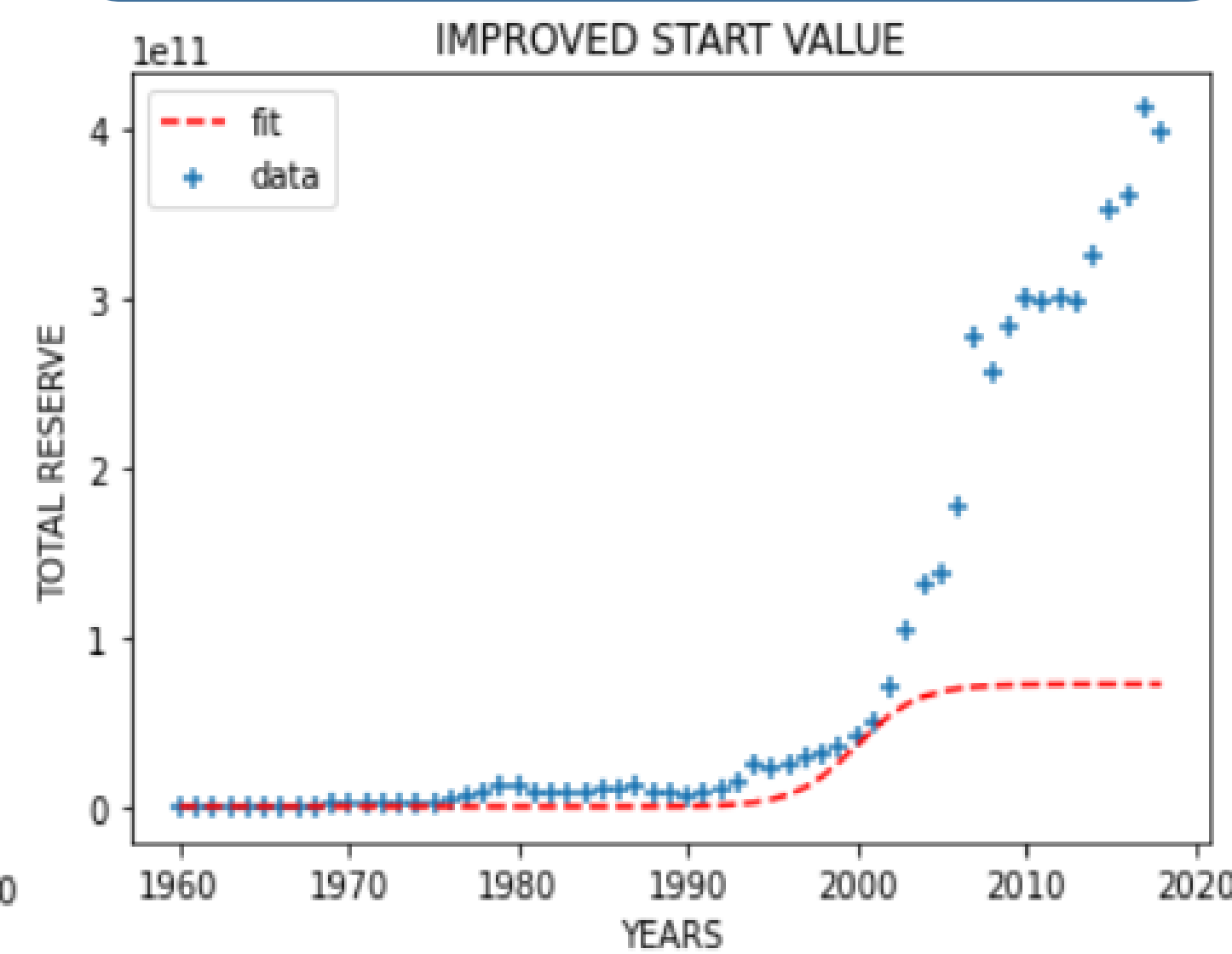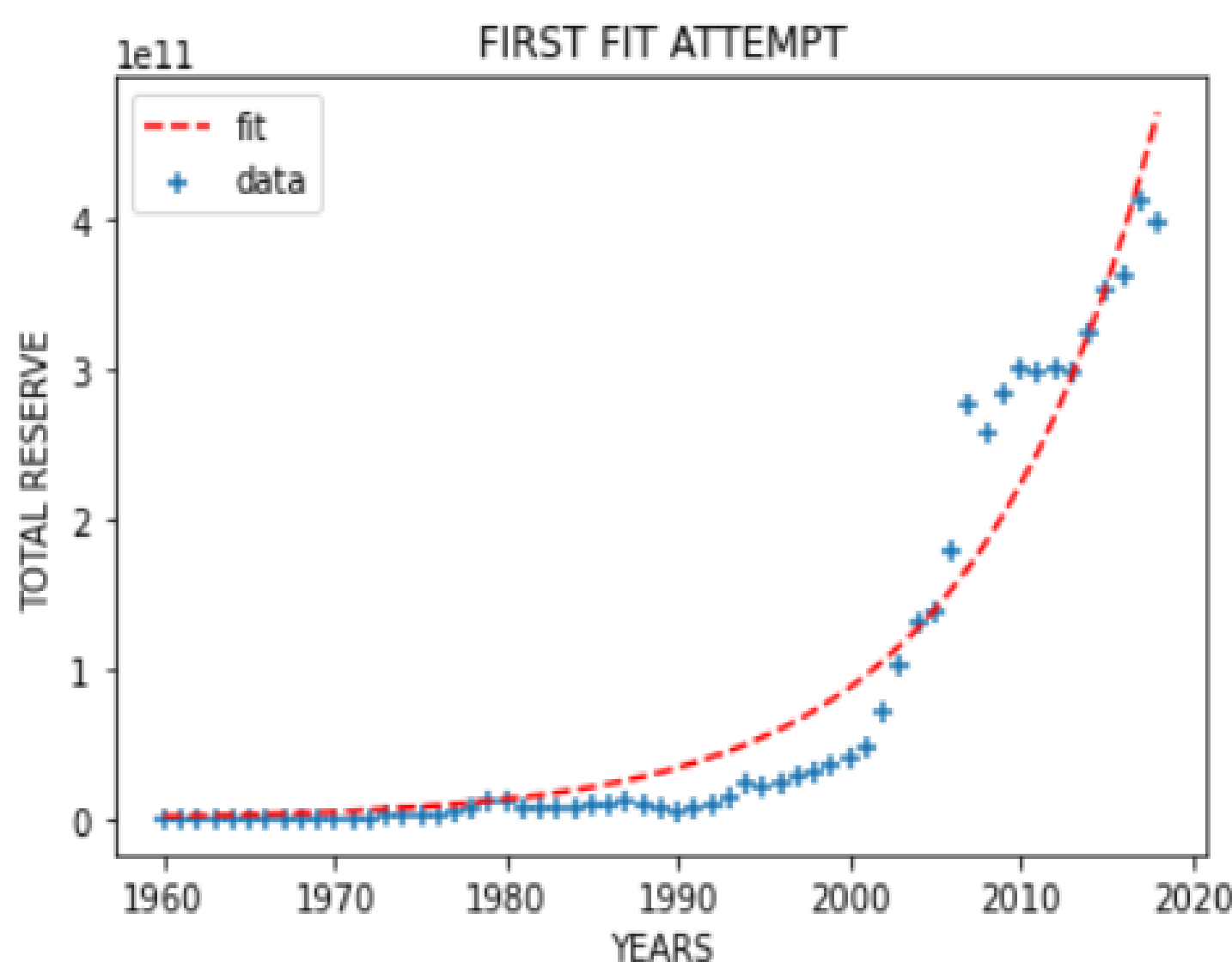
From my analysis, I was able to predict total reserves for India in future years, give that I do have low uncertainty levels from the findings, it is possibly the closest predictions I believe is possible from using a curve fit. Also judging from the trends of the data flow too the predicted year follows a similar pattern with values not far off.