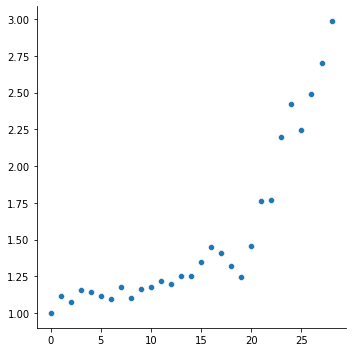
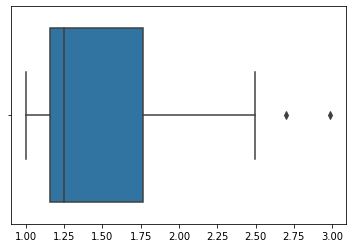
***Exploratory Data Analysis***

1. **Scatter Plot**



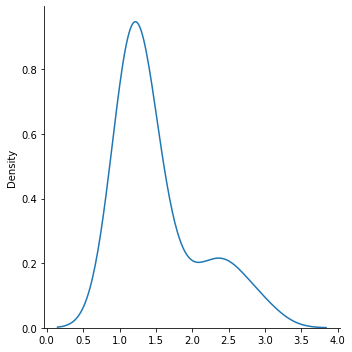
* Year values on X-axis and Index values on Y-axis are plotted in the graph here.
* We can see two different trends in the data here. The index values up to year 20 follow a line with some fluctuations. But after year 20 they seem to be increasing exponentially.

1. **Box and whisker plot**



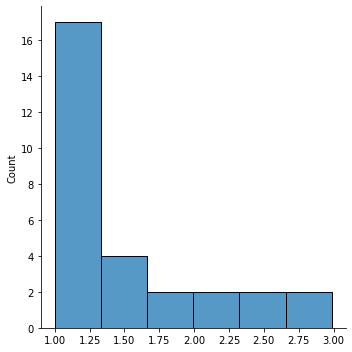
* The box plot helps to identify outliers in the data.
* We see two values after the right whisker line here. We should decide if they’re the outliers with further analysis.

1. **KDE plot**



* The X-axis represents the values and the Y-axis shows their density.
* So we see most of the values in the data are between 0.5 to 2.0

1. **Histogram**



* The first interval consists of 17 values, the second consists of 4 values, and the remaining 4 intervals consist of 2 values each.

***Curve Fitting***

* As noticed in the scatter plot, there’s an exponential rise in values at the end. So we can already imagine that a straight line won’t be a good fit for the data.
* In the excel sheet, I have tried different curves such as Linear, Exponential, Quadratic, Logarithmic are tried on the data.
* The R-square value shown for each graph decides which is the best fit for our data.

It decides how much the Y variable (dependent) depends upon the X (variable).

* The **quadratic curve** gives us about 93.7 % accuracy, higher than all others. Hence we can say that this curve would be the best fit for data.
* The R-square value tells us what percentage of prediction error is eliminated when we use one curve over other one.

Please find the plots of different curves with their respective R-square value below.

