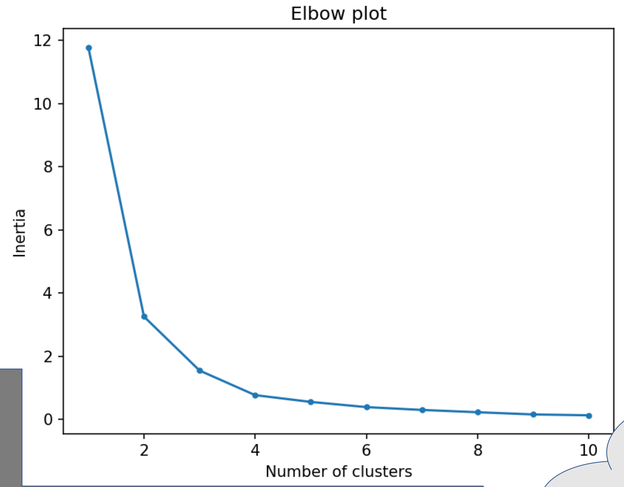
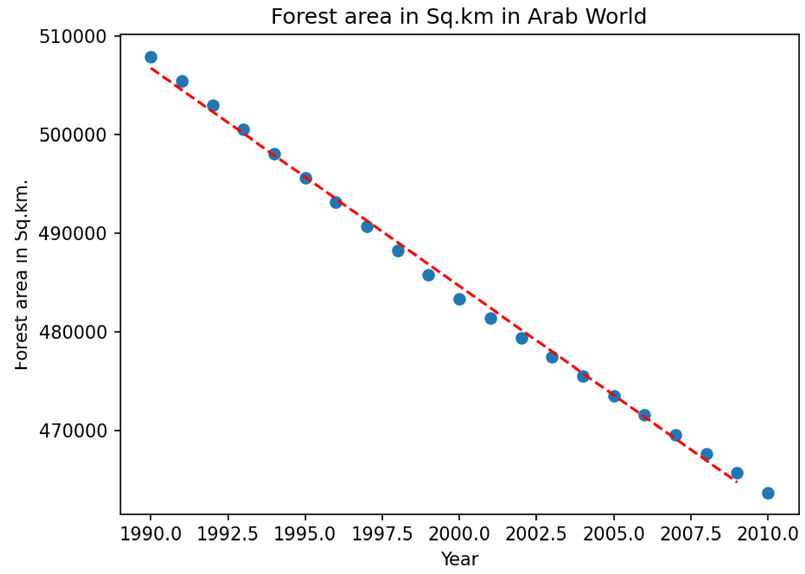
Applied Data Science 1
Clustering and Fitting



From the elbow plot we can see that k=2 in k-means clustering

We all know that human beings and animals take in oxygen and release CO2, where as plants take in CO2 and release oxygen. So, plants are very much needed to control the level of CO2 and play a major role in answering questions on climate change. This project aims to verify if there is a relation between the level of CO2 and the forest area. For this, I have considered the CO2 emission data and forest area data across the countries from the period of 1990 to 2019.

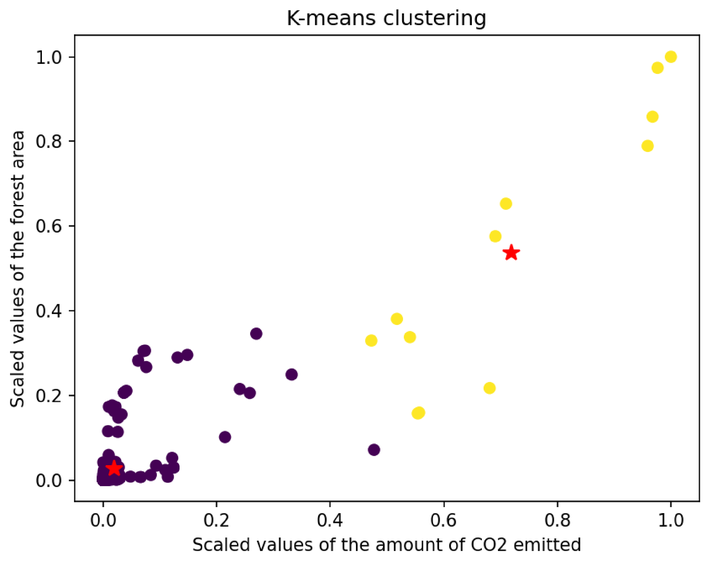


The Silhouette score of the clusters is 0.89409070377 This makes it a good clustering. Isn't it ?

We observe that, in countries belonging to one cluster both the forest area and the amount of CO2 emission levels are low. In another cluster of countries, both values are high. This shows a strong correlation between forest area and CO2 emission levels. So, we can conclude that by increasing the forest area we could control CO2 levels and thereby control climate change

The above curve fits the forest area data.

The forest area values predicted by the above curve



Comparing the data from 2 countries belonging to different clusters

```
The predicted forest area in Arab World in 2055 is 363071.67284694966
The predicted forest area in Arab World in 2060 is 352018.63042178005
The predicted forest area in Arab World in 2070 is 329912.54557144176
The predicted forest area in Arab World in 2080 is 307806.4607211035
```

Country	Amount of CO2 emitted	Forest Area
Africa Eastern and Southern	6.030300e+05	4511676.2
East Asia & Pacific (excluding high income)	1.249182e+07	4867609.8