

# ANALYSIS OF WORLD BANK CLIMATE CHANGE DATASET

## THE DATASET

The dataset is accessible through the following link:

(<https://api.worldbank.org/v2/en/topic/19?downloadformat=csv>)

```
60] climate_change_df = pd.read_csv("data/API_19_DS2_en_csv_v2_4028487.csv",
                                   skiprows=4).drop(['Unnamed: 66'], axis=1)
pop_growth_ds = climate_change_df[
    climate_change_df['Indicator Code'] == 'SP.POP.GROW'
].drop(['Country Name', 'Indicator Name'],
       axis=1)
pop_growth_ds_filtered = pop_growth_ds.fillna(0)
pop_growth_1990 = pop_growth_ds_filtered['1990'].values
pop_growth_2019 = pop_growth_ds_filtered['2019'].values
pop_growth_ds.head()
```

	Country Code	Indicator Code	1960	1961	1962	1963
4	ABW	SP.POP.GROW	NaN	2.236462	1.432843	0.823502
80	AFE	SP.POP.GROW	NaN	2.539822	2.575182	2.606839
156	AFG	SP.POP.GROW	NaN	1.898499	1.965805	2.029830
232	AFW	SP.POP.GROW	2.034308	2.085972	2.133725	2.173361
308	AGO	SP.POP.GROW	NaN	1.392891	1.383296	1.256405

### K-Means Clustering

Involves Assigning examples to clusters in an effort to minimize the variance of each cluster

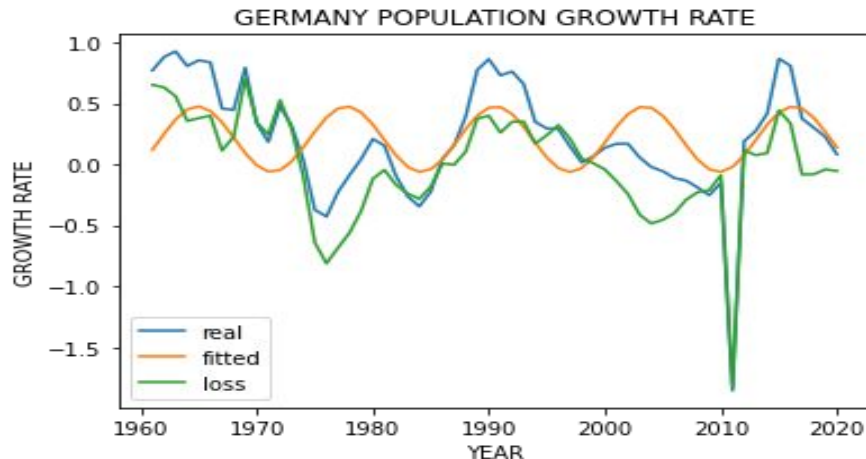
### OPTICS Clustering

Involves creating an augmented ordering of the database representing the density-based clustering of structure of a dataset

### Curve Fitting

is the process of constructing a **curve**, or **mathematical function**, that has the best fit to a series of **data points**

**Curve Fitting** of Population Growth Rate of Germany from 1960 to 2020. It can be seen that the fitted curve is a sinusoidal curve which varies slightly with the original data points as shown by the line plot



**Clusters of CO2** emissions per GDP of all the countries in the world for the year 1990 and 2015. Most countries fall around point (0.5,0.2). This can be an indication of that these countries are in the same scale of industrial development

