Exploratory Data Analysis (EDA) is used to detect any errors, outliers as well as to understand different patterns in the data using some statistical graphs and other visualization techniques. It allows us to understand the data better before making any assumptions. In EDA we get maximum insights from a data set, uncover underlying structure, extract important variables from the dataset, detect outliers and anomalies, test underlying assumptions and determine the optimal factor settings.

Here we used covid19 world cases death dataset for analysis. In Jupyter notebook, we import libraries such as numpy, matplotlib and loaded our dataset. We checked for null values, missing values, data types, asking some analytical questions and have done visualizations.

Before using a dataset to train a model, the data needs to be clean and in right format. The dataset will have many issues like imbalanced dataset, inconsistent format, missing values and outliers.

Dataset: <https://dj2taa9i652rf.cloudfront.net/> [rearc-covid-19-world-cases-deaths-testing/](https://covid19-lake.s3.us-east-2.amazonaws.com/rearc-covid-19-world-cases-deaths-testing/)

Dataset Features: ['iso\_code', 'continent', 'location', 'date', 'total\_cases', 'new\_cases',

'new\_cases\_smoothed', 'total\_deaths', 'new\_deaths',

'new\_deaths\_smoothed', 'total\_cases\_per\_million',

'new\_cases\_per\_million', 'new\_cases\_smoothed\_per\_million',

'total\_deaths\_per\_million', 'new\_deaths\_per\_million',

'new\_deaths\_smoothed\_per\_million', 'reproduction\_rate', 'icu\_patients',

'icu\_patients\_per\_million', 'hosp\_patients',

'hosp\_patients\_per\_million', 'weekly\_icu\_admissions',

'weekly\_icu\_admissions\_per\_million', 'weekly\_hosp\_admissions',

'weekly\_hosp\_admissions\_per\_million', 'new\_tests', 'total\_tests',

'total\_tests\_per\_thousand', 'new\_tests\_per\_thousand',

'new\_tests\_smoothed', 'new\_tests\_smoothed\_per\_thousand',

'positive\_rate', 'tests\_per\_case', 'tests\_units', 'total\_vaccinations',

'people\_vaccinated', 'people\_fully\_vaccinated', 'total\_boosters',

'new\_vaccinations', 'new\_vaccinations\_smoothed',

'total\_vaccinations\_per\_hundred', 'people\_vaccinated\_per\_hundred',

'people\_fully\_vaccinated\_per\_hundred', 'total\_boosters\_per\_hundred',

'new\_vaccinations\_smoothed\_per\_million', 'stringency\_index',

'population', 'population\_density', 'median\_age', 'aged\_65\_older',

'aged\_70\_older', 'gdp\_per\_capita', 'extreme\_poverty',

'cardiovasc\_death\_rate', 'diabetes\_prevalence', 'female\_smokers',

'male\_smokers', 'handwashing\_facilities', 'hospital\_beds\_per\_thousand',

'life\_expectancy', 'human\_development\_index',

'excess\_mortality\_cumulative\_absolute', 'excess\_mortality\_cumulative',

'excess\_mortality', 'excess\_mortality\_cumulative\_per\_million'],