

Home Assignment <3>: Exploratory Data Analysis on COVID-19 Dataset

Learning Objective:

The objective of this assignment is to perform **Exploratory Data Analysis (EDA)** on the COVID-19 dataset using statistical measures and visualization.

Students will apply descriptive statistics, outlier detection (IQR technique), data normalization (StandardScaler), and correlation visualization using Seaborn.

Expected Completion Time:

Best Case: 30 minutes

Average Case: 45 minutes

Assignment Details:

Use the existing **COVID-19 dataset** (`country_wise_latest.csv`) and focus on the following two numerical features:

- **Confirmed Cases**
- **New Cases**

You are required to perform the following tasks step-by-step:

1. **Create a class `CovidEDA`** to perform the below operations.
 - Load the dataset using Pandas.
 - Keep only the columns **Confirmed** and **New cases** for analysis.
2. **Compute Statistical Measures**
 - Calculate and print:
 - Mean
 - Median
 - Variance
 - Standard Deviation
 - Correlation Matrix (between Confirmed and New cases)
3. **Outlier Detection using IQR Technique**
 - Identify outliers in both Confirmed and New cases.
 - Remove the outliers and store the cleaned data in a new DataFrame.
 - Display the cleaned dataset.
4. **Normalization using Standard Scaler**
 - Apply **StandardScaler** from `sklearn.preprocessing` to normalize the Confirmed and New Cases.
 - Display the scaled (normalized) output as a new DataFrame.
5. **Visualization Tasks**
 - Plot **Histograms** for Confirmed and New cases (before and after normalization) using **Seaborn**, to visualize the bell curve.

- Plot a **Heatmap** between Confirmed and New cases to display their correlation visually.

Expected Outcome:

Upon completion of this assignment, you should be able to:

- Apply **EDA techniques** on real-world datasets.
- Calculate **descriptive statistics** (mean, median, variance, std deviation).
- Detect and handle **outliers** using the IQR method.
- Normalize data using **StandardScaler**.
- Visualize statistical distribution using **Seaborn histograms**.
- Create **heatmaps** to represent correlations visually.
- Structure your analysis within a **Python class** for reusability.