

ASSIGNMENT 1

Due date: By 11.59PM on Sunday, 24 January 2021

Part 1 :

Consider real-world datasets of different scale (at least 3 different standard datasets with varied sizes in the order of thousands to millions). For the datasets chosen by you, perform descriptive and exploratory analysis using R/Python/Scikit-learn etc, to generate insights into the given data in the form of -

- (i) Frequency distributions
- (ii) Graphical representations
- (iii) Summary Statistics

Submit a detailed report on the findings of your analysis, supported by the necessary plots and visualizations. Also comment on the nature of the datasets analyzed by you w.r.t their properties like central tendency, dispersion and shape, based on your formal analysis.

You may choose appropriate data resources available at -

<https://datascience.nih.gov/covid-19-open-access-resources>

<https://data.world/datasets/covid-19>

<https://ourworldindata.org/coronavirus-source-data>

<https://ieee-dataport.org/datasets>