



VIT[®]
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

School of Advanced Sciences
Department of Mathematics
M.Sc. Data Science
PMDS508P - Python Programming Lab
Slot: L19+L20+L29+L30
Lab Assignment #06

Assignment #06 - Python Pandas

Due-date to Upload of the Assignment Records to VTOP is 16-Nov-2024¹

Dear Students to carry out this assignment, you need to download the following three data files

- DOData.csv (<https://bsrvp.github.io/data/DOData.csv>)
- NutAverage.xlsx (<https://bsrvp.github.io/data/NutAverage.xlsx>)
- PythoBiomass.xlsx (<https://bsrvp.github.io/data/PythoBiomass.xlsx>)

The first data file **DOData.csv** contains the monthly mean values of Dissolved Oxygen in water for a year.

The second data file **NutAverage.xlsx** lists the monthly mean values of NH₄-N (Ammonia), NO₂-N (Nitrite), NO₃-N (Nitrate) and TN (Total Nitrogen) content in the water for a year.

The final data file **PythoBiomass.xlsx** lists the monthly mean biomass of the two Phytoplankton, namely Cyanophyceae and Chlorophyceae, along with the Total Biomass.

You now need to perform the following tasks on these three data sets.

Q1) Read the data file **NutAverage.xlsx** into a DataFrame and perform the following tasks:

- Find the null values, if any, in the data set and fill the null values with the method of your choice.
- Add a column **DIN** (stands for Dissolved Inorganic Nitrogen) to this DataFrame, where **DIN = NH₄-N+NO₂-N+NO₃-N**.
- Add another column **DON** (Dissolved Organic Nitrogen) to this DataFrame, where **DON = TN - DIN**.
- Add a row named **Averages** which contains the averages of each of **NH₄-N, NO₂-N, NO₃-N, TN, DIN, DON**.
- Describe characteristics of the DataFrame.
- Plot all the data (except the **Day Count** column) using the area plot, line plot and box plot of DataFrame. (use Subplots where ever appropriate for better visualisation of the data)
- Compare the **NH₄-N** vs **DIN** composition graphically using a scatter plot.

Q2) Read the data file **PhytoBiomass.xlsx** into a DataFrame and perform the following tasks:

- Find the null values, if any, in the data set and fill these null values with the method of your choice.
- Add a column **Others** which list the biomass of other phytoplankton groups obtained by subtracting Total Biomass with the sum of the biomass of Cyanophyceae and Chlorophyceae.
- Describe the characteristics of the DataFrame.
- Plot the biomass composition of each group using a barh and kde plot.

¹Upload the Assignments in PDF format only

Q3) Read the data file **DODData.csv** into a DataFrame and perform the following tasks:

- Plot the monthly **DO** data using a bar plot.
- Plot the monthly **DO** vs **NH4** (from NutAverage.xlsx) as a scatter plot.
- Construct an interpolating polynomial to estimate the **DO** for the entire year starting from day 1 to day 365. Visualise the interpolated and monthly data (with monthly data plotted as points).