**GROUP TASK CW2 (25 Per Cent of total marks)- Report Deadline. Exact date in Canvas.**

**Teams and Tasks overview:**

Your team consist of 5 members. You should have regular weekly meetings with your team members via MS Teams software.

All Team members should discuss first which software to use. The recommended software is R Studio (R is free and can be downloaded from the web), but team members may agree and decide to use Python.

-1 or 2 team members should look at data pre-processing and basic statistics (detecting missing values, outliers, presenting averages, standard deviations, quartiles and box plots).

-Another 1 or 2 team members should look at further statistics (correlation among variables, presenting correlation matrixes, distributions and histograms).

-The rest of the team members should work on building multi regression models (predicting a dependent variable from multiple independent variables). All team members should also have input into this final task and understand multiple regression models. A tutorial on multi regression models can be found in the reference below:

Ref: [2 Linear Regression | The Analytics Edge | Sloan School of Management | MIT OpenCourseWare](https://ocw.mit.edu/courses/sloan-school-of-management/15-071-the-analytics-edge-spring-2017/linear-regression/)

**DATA and TEAM TASK:**

**Data**- You are presented with an excel sheet that contain data collected via a multi-sensor platform (sensor system). The sensor system contains readings on date, PM2.5, NO2, 03, temperature and humidity. The data presented here are 1 hour averages. These 6 columns (A to F) constitute your main raw data set.

**Tasks**-Your team tasks should focus on: 1. providing some descriptive statistics and visualisation (e.g. box plots, correlation matrixes, time series, etc) on the raw data (columns A to F) 2. Providing forecasting models, assessing performance of the models, visualising results. It is recommended you use the reference in the previous section to learn about regression using several variables.

Note: We do not specify which language you use. Suggestions are R Studio and/or Python. I recommend R. We do not specify forecasting methods but we do suggest regression and to follow a procedure such as the one in the link provided in the previous section. We do not specify the dependent variable that you want to predict.

**REPORT: Deadline for the report: Exact date in Canvas**

Only one report per group to be submitted by a chosen team leader.

Your Lab report will have a maximum of 7 pages (including references but excluding appendices). Font size: 11. Anything extra you wish to include can be included in the appendix.

**The structure of the report should include:**

-Title and team member names

-Team member roles: to be written by each team member. Two paragraphs describing individual member tasks and contribution to the team

-Abstract (brief description of this exercise and motivation)

-Introduction: brief intro to data set and purpose of the exercise

-Main body: with 4 sections on each of the main aspects 1. (basic statistics such as box plots, 2. more advanced statistics such as correlations and distributions, 3.multi variable regression forecasting models, visualisation of results, 4. model performance and discussion of results. Most effort should be put in the last 2 sections on forecasting and performance of forecasting.

-Conclusions

-Appendixes do not count towards the 7page limit, so if there is no room for something in the main document you can include it in the appendix.

In the main sections you should describe what your team has achieved during the task. Describe the data and phenomena by making use of your 4 sections on the main body. Include pictures and diagrams as appropriate and include any relevant graphs, especially when visualizing statistics and when presenting results for forecasting models. Include interpretation of results. Include any relevant code in the appendix and any graphs that you feel are important but you could not include in your 7 page limit. Reference any external sources and material used. Do not plagiarize. This is an academic offense. Explain concepts in your own words.

**MARKS BREAKDOWN : Total 25% of module**

-**15 % (Group Mark):**

Task completion: Descriptive statistics, forecasting including methods and visualisation of results and statistics (10 %).

Report: Lab Report including abstract, introduction, basic statistics such as box plots; further statistics such as correlations, distributions and histograms; forecasting models; visualisation of results and conclusions. Follow structure given above. (5 %)

Don’t forget to include team member’s names and two paragraphs on the contribution of each team member.

**-10 % (Individual Mark):**

Individuals to include two paragraphs in the report stating team role and contribution to the team (two paragraphs per member)

Engagement in Group Task (5 %) +

Individual Contribution to Group Task (5 %)

**Important**: **A total mark of zero is a possibility if an individual member does not engage with the task and does not contribute to the Team.**