Part 1

In this week, you will be getting familiar with the data. At the end of the week, the goal is to have clean data, efficiently imported into your workspace, and visualized, and to understand what the data is used for. The PDF you will return will have max. 2 pages.

Breakdown of the points:

- 0.25p – an established communication channel and appropriate strategy for code sharing.
- 0.25p – data correctly imported into appropriate matrices completely: observations as rows, variables (predictors) as columns.
- 0.5p – identification of challenges of the data: for example: time series not synchronized, missing values in data, extra variables, variables with unknown physical meanings, etc.
- 0.5p – a visualization and comment on the dataset: variable distribution, number of observations, type of measurements (time series or not time series)
- 3p - exploratory data analysis with PCA: explain variable correlations and visualize the PCs using biplots, loading plots; (! only on the X matrix - we are not looking at the response variable now)
- 0.5p – identification of pretreatment steps, and a plan on how to do data pretreatment

Part 2

Based on your initial evaluation, the received feedback and the strategies of other groups, perform data pretreatment. There will be a max. 2 pages addition to your project PDF.

Breakdown of the points:

- 0.5p - feedback incorporation and adjustment of the initial pretreatment plan
- 0.5p - data centering and scaling techniques
- 0.5p - evaluation of extreme/ missing values and mitigating actions, data synchronization (for time series), sampling
- 0.5p - visualizing pretreated data

Modelling plan

Make a modelling plan for your modelling goal. Create a flowchart of the operations needed to accomplish the task, and assign roles to the team members for each step. The start of each flow chart should be the cleaned data, and the end of it should be the modelling goal achieved. If there is more than one modelling goal, the flowchart can take multiple paths. There will be a max. 2 pages addition to your project PDF. There will be a max. 2 pages addition to your project PDF.

- 0.5p - modelling goal (What is the purpose of your data analysis?);

- 0.5p - model calibration strategy: tools used, methodology, data used;
- 0.5p - model validation strategy: scope of validation, methodology;
- 0.5p - model testing strategy: metrics and data used;
- 0.5p - description of the mathematical methods used in modelling;
- 0.5p - model diagram/ operations flowchart.

Part 3

Points breakdown:

- 1p - feedback incorporation
- 1p - code (correctness, efficiency, usability)
- 1p - reporting (readability, completeness, presentation, mathemathical description of utilised methods)
- 1p - modelling correctness
- 1p - achieving the modelling goal

For A-level teams: 5p - difficulty