

## Problem Solving: Data Exploration

1. Revise your Mini Project dataset and explore the following parameters of it:
  - number of columns and rows
  - type of data
  - distribution of numeric data (is it a *normal distribution*?)
  - descriptive parameters of the numeric data (mean, median, mode, min, max, range, count)
  - missing values
2. Prepare the data for further analysis by
  - replacing/removing the cells with damaged/missing data
  - drilling down/rolling up attributes
  - coding the text data with numeral labels
3. Analyze the correlations and/or the similarity of the properties as appropriate
4. Optional: Apply PCA of attributes, if needed.
5. Optional: Use different diagrams/geo-maps to visualize the data in a user-friendly way

### Note

This is a class exercise, a group work, part of Mini-Project 1.

Have fun!  
the instructor