Computer Science & Information Systems

# Data Management for IoT - Lab sheet 1

Exercise 1 – Data preparation using pandas

# Objective

The objective is to

* + understand the steps involved in the usual data preprocessing step
    - Load the structured data into Pandas Dataframe
    - Getting familiarity with Pandas functions for data preparation
    - Checking and fixing the data quality issues
    - Carry out data preparation activities like filtering, grouping etc.

# Steps to be performed

**Tool** Python3 and Jupyter Notebook

**Libraries required** Pandas, NumPy

**Input**

beach\_water\_quality\_automated\_sensors\_1.csv

**Analytics Technique** Descriptive Analytics

**Implementation** DMIoT - Lab1 - Beach Water Quality Analysis -1.ipynb

**Steps**

* + Import required Python library and convert to as data frame
  + Understand the structure of the dataset
  + Slice out the part of the dataset
  + Explore the summary stats
  + Prepare the data for exploration
  + Sort the data
  + Apply filtering on the dataframe
  + Try out data grouping using the columns of merged dataframe
  + Check data quality issues
  + Rename and modify the content of dataframe column
  + Create the derived columns from the existing columns

# Results

* + Import / Export of datasets performed
  + Data preparation steps carried out
  + Familiarly with data preparation functions available in Pandas

# Observation

* + The usual data preprocessing and simple explorations are carried out with the help of Pandas libary