

# Introduction into Big Data analytics

## Assignment 2 - working with big data sets

Janusz Szwabiński

Download the 311 Service Requests from 2010 to present from <https://nycopendata.socrata.com/Social-Services/311-Service-Requests-from-2010-to-Present/erm2-nwe9>

**Task 1** Working with plain csv file, find:

- the most common complaints,
- the most common complaints in each borough,
- the agencies with the most common complaints.

Measure the times needed to complete the queries.

**Task 2** Now convert the data from the csv file into an SQL database, then complete the above tasks again.

- What are the times needed for the queries?
- What is the total time needed to complete the task, i.e. to convert the data and find the queries?
- Compare the execution times for both SQLite and Mysql databases.

Preferably, use a programming language like Python to connect to the database and to report the results instead of working with raw SQL queries and a database client (hint: `sqlite3` library).

**Task 3** How would you improve the efficiency of searching for queries on a standalone computer? Implement your ideas and present the results (even in the case of no performance gains!).