**Public Orders Database Scraping**

We provide you with 3 python files and 2 CSV files. Here is the description of them.

1. **old\_db\_all.py** and **old\_db\_all\_test.py** are basically the same files. They both contain scraper for the old database. **old\_db\_all.py** is a full program which covers the whole time period which was specified in rules of this project. However, because the database for mentioned period is really huge and the program would have to scrape about 400k I decided to create the test program **old\_db\_all\_test.py**. In test program the initial date is set to 01/01/2017. It allows program to incorporate all its features.
2. The file **old\_db\_all\_test.csv** contains the results of running **old\_db\_all\_test.py** script.
3. **new\_db.py** contains a scraper for a new database
4. **new\_db\_test.csv** contains sample data for code for the new database. Again, there was a need to limit the number of data scraped. The code is supposed to work with 5000 records (which is the maximum num of elements limited by new database), however in the example code it scrapes 200 observations. There are two places in code (labeled as num\_of\_el) where the number of elements scrapped can be set up to 5000. In the attached code num\_of\_el = 100.
5. There is a slight difference between column names in csv files. In the old database there were no elements like “podwykonawcy”, “wartość udziałów podwykonawcy”. Therefore they were excluded. Instead of “podwykonawcy” there is a column “wykonawcy” where the program scrapped the info about contractors.