

## EXERCISE 3

### Intelligent data analysis DV1597

April 18, 2022

This third exercise consists of four questions regarding 4 datasets. The questions are extracted from Steven Skiena's material [1].

The exercise should be performed **individually**, that is, without group cooperation. Please hand in your solutions as a **Jupyter notebook** file via Canvas. The report should be in **English** and include your name and the answers to all the questions below. The exercise is graded with G/Ux/U.

Provide answers to the questions associated with the following data sets, available at <http://www.data-manual.com/data>. For all datasets, download the "**Local Data**". The exercise includes exploratory data analysis, i.e., for each question, you need to analyze the data set, interpret the results, and possibly visualize them.

1. Analyze the *movie* dataset. What is the range of movie gross in the United States? Which type of movies are most likely to succeed in the market? Comedy? PG-13? Drama?
2. Analyze the *Manhattan rolling sales* dataset. Where in Manhattan is the most/least expensive real estate located? What is the relationship between sales price and gross square feet?
3. Analyze the *2012 Olympic* dataset. What can you say about the relationship between a country's population and the number of medals it wins? What can you say about the relationship between the ratio of female and male counts and the GDP of that country?
4. Analyze the *GDP per capita* dataset. How do countries from Europe, Asia, and Africa compare in the rates of growth in GDP? When have countries faced substantial changes in GDP, and what historical events were likely most responsible for it?

## Reference:

1. Skiena, Steven S. "Visualizing Data" in The data science design manual. Springer, 2017.