Exercise 3

Intelligent data analysis
DV1597

April 6, 2024

This third exercise consists of four questions regarding 4 datasets. The questions are extracted from Steven Skiena's material [1] and Montgomery's material [2].

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.

1. 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.

    a. Analyze the *Manhattan rolling sales* dataset. Where in Manhattan is the most/least expensive real estate located? What is the relationship between the sales price and gross square feet?

    b. 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?

2. What is a time series? Explain the meaning of trends effects, seasonal variations, and random error.

3. Your company manufactures and distributes soft drink beverages, sold in bottles and cans at retail outlets such as grocery stores, restaurants and other eating/drinking establishments, and vending machines in offices, schools, stores, and other outlets. Your product line includes about 25 different products and many of these are produced in different package sizes.

    a. What forecasts do you need in this business to be successful?

b. Is the demand for your product likely to be seasonal? Explain why or why not.

c. Does the shelf life of your product impact the forecasting problem?

d. What data do you think that you would need to be able to produce successful forecasts?

Reference:

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

2. Montgomery, Douglas C., Jennings, Cheryl L., and Kulahci, Murat. Time Series Analysis and Forecasting. Wiley, 2016.