Exercise sheet 1 – October 8, 2021

Please submit your solution electronically until October 15, 2021, 23:59
Send your script-file to jerke@soziologie.uzh.ch

Notes:

- **Please sufficiently comment your script and structure it according to the different tasks.**
- **Whenever a task asks for an explicit answer, please write down your answer directly in the script within a comment.**
- **Make sure that you fully document your solution in your script. If you give an answer, but there is no code to clearly reconstruct how the answer was determined, the answer cannot be counted.**
- **The tasks vary in difficulty. For some of them you may have to combine commands in a new way or have to look in the documentation of the respective libraries.**
- **Most exercise sheets will contain bonus questions, providing the possibility to obtain extra points.**
- **The solution of this sheet will be published on OLAT after the submission deadline expires.**

1. Download the data set «Dogs_of_Zurich" and load it with Python.
   [Background on the data set: This dataset provides the most current information on dogs and their owners from the municipal dog register of Zurich. For those keeping dogs, information on age group, sex and district (Quartier) of residence is provided. For each dog, the breed, breed type, sex, year of birth and color are recorded. The dog register is maintained by the Dog Control Department of the Zurich City Police.]

2. How many dogs are currently registered in Zurich?

3. Analyze the missing values in the data set:
   a. There are a few variables in the data set that have many missings (i.e., less than 1000 valid values). Identify these variables and delete them from the data set.
   b. After dropping these variables, how many observations still have at least one missing value? (Make sure not to delete these cases.)

4. Rename the column names in such way that they can be better used afterwards and are more accessible for non-German speaking users of the data set. Therefore, transform them to lower case and choose English names

5. Let's have a look at how the dogs of Zurich are distributed across dog owners:
   a. How many people own dogs in Zurich?
   b. How many people own more than one dog?
   c. What is the highest number of dogs a single dog owner owns?
   d. What types of dog breeds does this person have?

6. We are interested in how old the dogs of Zurich are:
    a. Create a new variable that contains their approximate age in years.
    b. What is the mean age of the dogs? How old is the oldest dog in the data set?
    c. [BONUS] You might have noticed that the age variable contains some implausible values. Replace these values with a missing value (remember: missing values can be expressed with `np.nan` from the *numpy* library). What is the mean age of the dogs after that correction? How old is the oldest dog?

7. We want to know more about the dogs and where they live:
    a. What are the top five of most favorite dog breeds in Zurich?
    b. In what district do the most dogs live?
    c. In what district do the oldest dogs live on average?
    d. In what district do the most wiener dogs (Dachshund) live?
    e. What is the favorite dog in district 7?

8. [BONUS] We also want to know a bit more about the dog owners:
    a. How many female and male dog owners are there in Zurich?
    b. What is the age distribution of dog owners?

9. Export the final data frame to an excel sheet.