Big Data Algorithms - 2023

List 0 for the 3rd lab.

Marek Klonowski

I. Pandas

1. Download CSV dataset about diet and losing weight. You can use this link:
   WeightLoss.csv . Some descriptions can be found here: WeightLoss.html .
   Create a DataFrame structure called **df**. Print its first 20 rows. The rest of the exercises refer to this data structure.
2. Find the number of observations (rows) and the number of attributes (columns). Find the basic statistics for weight loss in the first month (max, min, std, ...). HINT: use `describe`.
3. Find the person with maximal total weight loss.
4. Find the person with minimal average self-esteem.
5. Sort the rows by weight loss at the second month in descending order.
6. Print all the rows such that at least one weight loss was higher than 4.
7. Add a column that contains average self-esteem.
8. List all the rows with the average self-esteem in the range $[12 - 15]$.
9. Print, all the rows such that at least one weight loss was higher than 3 and self-esteem this month, was at least 13.
10. Set index to the group column.
11. Set the index to the self-esteem in the 1st month. List all the rows with the index between 13 and 15 using `loc`.
12. Sort the rows by self-esteem at the 1st month and self-esteem at the 2nd level. List all the rows with self-esteem in the first two months above 4. HINT: You can use indexing.
13. Add a column "Stable"that is `True` if and only if the self-esteem does not change in time (i.e., is the same in all three columns).
14. Find the number of rows in each group wherein the self-esteem in the 1st month is higher than the self-esteem in the 3rd month.
15. Draw a histogram of the total weight loss for each group.
16. Remove all the information about the first month.
17. Remove all the rows related to the Control group.
18. Remove all the rows suspected to represent the same person (with the same weight loss in three months).
19. Assign each row into one of three categories concerning the average self-esteem (below 13, between 13 and 14.5, and above 14.5). Using pivoting, find for each category and each group number of rows and the average weight loss.
20. What is the total weight loss of all the participants? Assume that a magician can turn lost kilograms into high-quality gold. Is it possible to buy each living person at least one donut for this gold?