Laboratory Work 6

Working with Strings in Pandas

Goal: Learning Pandas features for working with string data.

2. Tasks:

Notes:

1. In this lab you should use Pandas string vectorised functions, str attribute and regular expressions and DO NOT use loops, list comprehensions and other iterations.
2. To calculate the number of variant use the formula $N = (n + 2)\%3 + 1$, where *N* is the variant number, *n* is the number in the group list.

Variant 1

1. Upload data set books.csv.
2. Create a new column "WordsInTitle" with the number of words in the title.
3. Find the longest title in dataset.
4. Find the last name of author of the book with the longest title.
5. Find author with 3 consonant in row in first name.
6. Find books written by author whose name starts with ‘A’
7. Create a new DataFrame with all books, where there are numbers in the title.
8. Find books with "C" as the first letter of the second word in the title

Variant 2

1. Upload data set books.csv.
2. Create a new column in the DataFrame created in task 5 with the numbers from title. If there are several numbers in the title, join them by ‘+’ (for example, ‘39+1’).
3. Find books that published by publisher with 0 or 1 vowel in its name.
4. Find authors whose last name starts and ends with consonants.
5. Find books with the maximum number of vowels in its title.
6. Find books with 2 lettered words in the title
7. Find author with 3 consonant in row in first name.
8. Find the author with 5 letters in the last names

Variant 3

1. Upload data set books.csv.
2. Create the columns with number of vowels in author's first name
3. Find authors with 3 vowels in row in their last names.

4. Find books where there are one letter word in the title
5. Create a column with 1 if there is number in the title and o otherwise
6. Create a new column with last word of books' titles
7. Find the books that were published by publisher with 2 words in its name
8. Find authors with the 3rd ‘y’ in the last name

3. The content of the report

1. Cover page of the report.
2. Topic and goal of the lab.
3. Progress of the work.
4. Link to the created Jupyter Notebook on GitHub, rendered by nbviewer.
5. Conclusions.