

Задание 1

Есть следующий набор данных авторов публикаций исследований о России [ranking_dataset.csv](#):

id - идентификатор публикации,
authour_name - имя автора,
title - название публикации,
issn - номер ISSN журнала,
journal - название журнал,
year - год издания,
address - полная аффилиация автора.

Из переменной *address* извлечь название организации и страну, добавив к датасету колонки *org_name* и *country*.

На полученных данных сделать:

- количество публикаций в названии которых встречается Russia или Russian
- список ТОП-20 **организаций** по количеству **авторов** с указанием их количества
- список ТОП-20 **авторов** по количеству **публикаций** с указанием их количества
- список ТОП-20 **организаций** по количеству **публикаций** с указанием их количества
- список ТОП-20 **журналов** по количеству **публикаций** с указанием их количества

Задание 2

На сайте РИНЦ найти в каталоге журналов <https://elibrary.ru/titles.asp> российские журналы по тематике информатика и со страницы анализа публикационной активности для каждого журнала собрать [Число статей в РИНЦ](#) по годам.

В результате должна получиться таблица следующего вида:

id - идентификатор журнала в РИНЦ (https://elibrary.ru/title_profile.asp?id=38061)

name - название журнала

n_papers_risc_year - число статей в РИНЦ, где *year* меняется от 2011 до 2020

Т. о. должен получиться датасет размерностью N строк - количество журналов и M столбцов - описанных выше переменных.

КАТАЛОГ ЖУРНАЛОВ

ПАРАМЕТРЫ

Название журнала, издательства или ISSN: Страна:

Тематика:

Язык публикаций:

Сведения о переводе:

Сведения о включении в Web of Science:

Сведения о включении в Scopus:

Сведения о включении в РИНЦ:

Доступ к полным текстам:

- входит в базу данных RSCI (887) - входит в перечень ВАК (3280)

- входит в ядро РИНЦ (30142) - с полными текстами (8065)

- выходит в настоящее время (56019) - только научные журналы (70809)

Сортировка: Порядок:

ПОИСК

НАВИГАТОР

Задание 3

Представьте, что по вам нужно придумать дизайн и методологию исследования, посвященному публикациям на тему исследований о России. Вдохновитесь этой статьей <https://link.springer.com/article/10.1007/s11192-020-03527-0> на тему Network Science и опишите возможный проект: какой исследовательский вопрос, какие данные нужно собрать, сложности реализации проекта. Объемом до 500 слов.

Результаты заданий нужно сохранить в архиве, где должны быть следующие файлы:

- мотивационное письмо (до 500 слов)
- CV
- документ Word с таблицами и их интерпретацией из задания 1.
- файлы с кодом, который использовался для выгрузки данных из задания 2 (R, Python, или в другом удобном для вас стеке).
- CSV-файл из задания 2.
- документ Word с описанием исследовательского дизайна и данных из задания 3.

Пожалуйста, назовите архив следующим образом: [фамилия]_task_[дата].[расширение], например, [ivanov_task_11.09.2021.zip](#).