

Домашнее задание 10

Файл с выполненным заданием необходимо загрузить на Dropbox до дедлайна, указанного на сайте.

Домашние задания, сданные после срока, оцениваются с использованием понижающих коэффициентов: опоздание в пределах часа – штраф 10% от полученной оценки, в пределах суток – штраф 20%, в пределах недели – штраф 50%. Домашние задания, сданные через неделю после указанного срока и позже, не принимаются и не оцениваются.

Если при проверке работ установлен факт нарушения академической этики, студент получает оценку «0» за данную работу. Работа студента, предоставившего свою работу для списывания, также аннулируется.

Задача 1

Напишите функцию, которая принимает на вход ключевые слова для поиска курсов и специализаций на Coursera, и возвращает датафрейм с результатами первых 5 страниц поиска. В датафрейме должны сохраняться только курсы (COURSE) и специализации (DEGREE). Датафрейм должен содержать следующие столбцы:

- **course**: название курса (тип `character`);
- **university**: название университета (тип `character`);
- **type** : тип (курс или специализация, тип `factor`);
- **rate**: оценка из 5 (тип `numeric`);
- **students**: число слушателей (тип `numeric`);
- **level**: уровень (тип `factor`).

Внимание. Число слушателей, записанных на курс, изначально сохраняется в виде `33k` или `1.4m`, поэтому дополнительно их нужно преобразовать. Например, вместо `33k` должно быть `33000`, а вместо `1.4m` должно быть `1400000`.

Подсказка 1. Пример ссылки, которая содержит запрос по ключевым словам `web design`:

<https://www.coursera.org/search?query=web+design>.

Пример ссылки на вторую страницу поиска по тем же ключевым словам:

https://www.coursera.org/search?query=web+design&page=2&index=prod_all_products_term_optimization.

Подсказка 2. Удобнее всего сначала выгрузить результаты с первой страницы, сохранить в датафрейм, а потом постепенно доклеивать к нему датафреймы с остальных четырех страниц.

Задача 2

Потестируйте написанную функцию на ключевых словах `web design`. По полученному датафрейму посчитайте корреляцию между числом слушателей (`students`) и оценкой курсов (`rate`).