

Российская экономическая школа, программы ФИБ, МНФ

Анализ данных в Python

Домашнее задание 1

Срок сдачи 15 сентября 2023 г., пятница, 10:00

Задача 1. (не оценивается) Установите и настройте на своем компьютере рабочее окружение Python, например, одно загрузив Anaconda отсюда: <https://www.anaconda.com/download>.

Задача 2. (не оценивается) Установите и настройте на своем компьютере систему управления версиями Git, например одним из следующих способов:

- Утилита командной строки Git с <https://git-scm.com/downloads>
- Расширение оболочки Windows TortoiseGit с <https://tortoisegit.org/download/>
- Графический клиент SmartGit с <https://www.syntevo.com/smartgit/download/>
- Какой-нибудь графический клиент из перечисленных на <https://git-scm.com/downloads/guis>

Далее, зарегистрируйтесь на <https://github.com>.

Задача 3. Разыщите онлайн какой-нибудь достаточно богатый набор данных (в идеале связанный с экономикой). Сформулируйте пять нетривиальных вопросов к этому набору данных. В качестве хорошего примера набора данных можно рассмотреть MovieLens (<https://grouplens.org/datasets/movielens/>). Очень хорошие вопросы для него (и ответы на них) можно найти здесь:

<http://www.gregreda.com/2013/10/26/using-pandas-on-the-movielens-dataset/>. Наборы данных и вопросы не должны повторяться. Очки за каждый вопрос начисляются с учетом повторений: чем меньше повторов, тем больше оценка за вопрос/набор данных. Если какой-нибудь из наборов данных мне очень понравится, я использую его в одном из следующих домашних заданий, за что нашедшему будут выданы дополнительные баллы.