

ML Basic

Библиотека Pandas

• REC

Проверить, идет ли запись

Меня хорошо видно
&& слышно?



Ставим “+”, если все хорошо “-”,
если есть проблемы

Тема вебинара

Pandas

Игорь Струйко



Руководитель курсов: Reinforcement Learning, ML Professional, ML Basic, MLOps, FinML

Teamlead, главный инженер проекта,
Физический факультет МГУ, PhD теоретическая физика

Опыт:

Более 15 лет занимался прикладной математикой и мат моделированием
(Data Scientist) (Python, C++)

Анализ временных рядов, эволюционные модели, финансовые модели

@stureiko (TG)

LinkedIn: [igor-stureiko](#)

Правила вебинара



Активно
участвуем



Off-topic обсуждаем
в учебной группе



Задаем вопрос
в чат



Вопросы вижу в чате,
могу ответить не сразу

Условные обозначения



Индивидуально



Время, необходимое
на активность



Пишем в чат



Говорим голосом

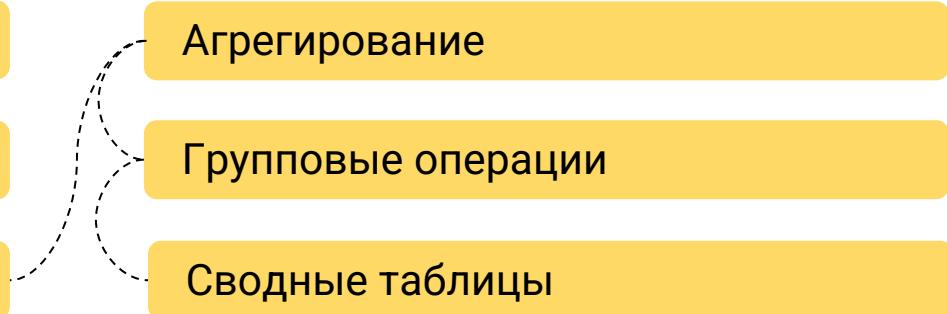


Документ



Ответьте себе или
задайте вопрос

Маршрут вебинара



Цели вебинара

К концу занятия вы сможете

1. Узнать, в чём практическая польза работы с Pandas
2. Назвать различие между Series и Dataframe
3. Загружать данные в Dataframe различными методами
4. Очищать и предобрабатывать данные в Pandas
5. Строить группировки и соединять данные в Pandas

Смысл

Зачем вам это знать

1. Работа с большими наборами данных

2. Работа с большим количеством источников данных (csv, json, excel, html)

3. Масштабируемость под сложность задачи

4. Интуитивная понятность (схожесть методов с SQL)

Библиотека Pandas



pandas Series and DataFrame

Series

| | apples |
|---|--------|
| 0 | 3 |
| 1 | 2 |
| 2 | 0 |
| 3 | 1 |

Series

| | oranges |
|---|---------|
| 0 | 0 |
| 1 | 3 |
| 2 | 7 |
| 3 | 2 |

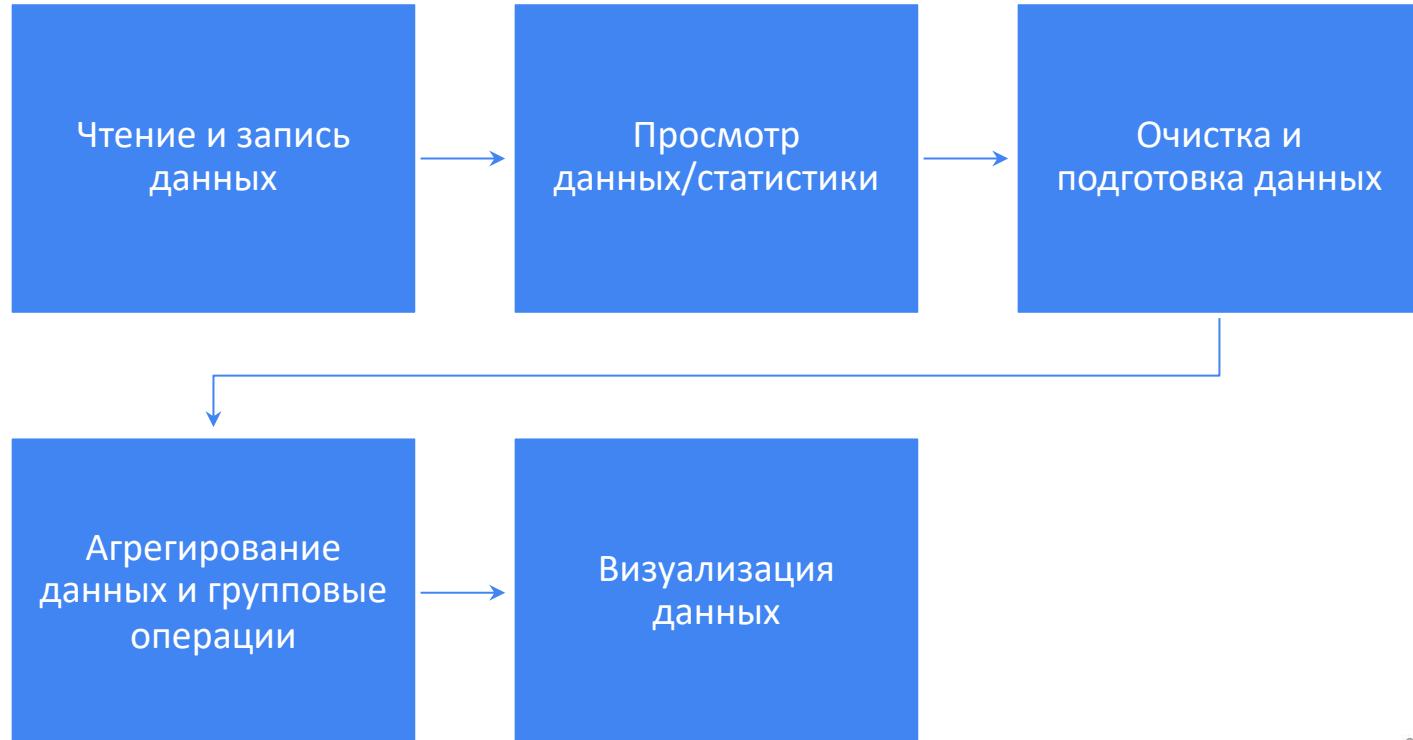
 $+$ $=$

DataFrame

| | apples | oranges |
|---|--------|---------|
| 0 | 3 | 0 |
| 1 | 2 | 3 |
| 2 | 0 | 7 |
| 3 | 1 | 2 |

Разведочный анализ данных

(exploratory data analysis - EDA)



Методы Pandas

- `.read_csv()`
- `.read_excel()`
- `.to_csv()`
- `.to_excel()`

Чтение и запись файлов

- `.head()`
- `.dtypes()`
- `.info()`
- `.describe()`

Просмотр данных/статистики

- `.groupby()`
- `.pivot_table()`
- `.crosstab()`

Группировка и агрегирование

- `.apply()`
- `.map()`
- `.replace()`

Подготовка данных

Вопросы?



Ставим “+”,
если вопросы есть



Ставим “-”,
если вопросов нет

Практика

Вопросы?



Ставим “+”,
если вопросы есть



Ставим “-”,
если вопросов нет

Список материалов для изучения

1. Документация Pandas

- <https://pandas.pydata.org>

2. Работа с Pandas

- <https://habr.com/ru/companies/ruvds/articles/494720/>
- <https://habr.com/ru/articles/196980/>
- <https://habr.com/ru/companies/otus/articles/727222/>
- <https://habr.com/ru/companies/ods/articles/322626/>

3. Учебные задачи на Pandas

- https://grossmend.com/blog/post/pandas_120_part_1/
- https://grossmend.com/blog/post/pandas_120_part_2/

Рефлексия

Рефлексия



С какими впечатлениями уходите с вебинара?



Как будете применять на практике то,
что узнали на вебинаре?

**Заполните, пожалуйста,
опрос о занятии
по ссылке в чате**

Спасибо за внимание!

Приходите на следующие вебинары

Визуализация данных

Игорь Струйко



Руководитель курсов: Reinforcement Learning, ML Professional, ML Basic, MLOps, FinML

Teamlead, главный инженер проекта,
Физический факультет МГУ, PhD теоретическая физика

Опыт:
Более 15 лет занимался прикладной математикой и мат моделированием
(Data Scientist) (Python, C++)

Анализ временных рядов, эволюционные модели, финансовые модели

@stureiko (TG)

LinkedIn: [igor-stureiko](#)