



OTUS

Онлайн образование

otus.ru

• REC Проверить, идет ли запись

Меня хорошо видно && слышно?



Python для аналитики



Козлов Роман Владимирович

руководитель курса Python для аналитики

Эл. почта: kozlovrv87@gmail.com



Преподаватель



Козлов Роман

Образование:

МЭСИ (сейчас РЭУ им. Плеханова),
МГУУ Правительства Москвы,
Финансовая Академия Правительства РФ

Работа:

Комплекс городского хозяйства и Социальной
защиты Правительства Москвы
Мониторинг коммунальной техники
Адресная социальной помощь 65+ населения

Среди реализуемых проектов —
информационная система Контакт Центр
«Социальная защита»

- автор/преподаватель курса “Python для аналитики” в
OTUS



Маршрут вебинара

Знакомство

Об OTUS

Основы Pandas. Series и
Dataframe

Работа с данными. Основные
методы

Команда курса

О курсе, программа обучения

Бонус: карьерная информация

Рефлексия



Правила вебинара



Активно участвуем



Задаем вопросы в чат



Вопросы вижу в чате,
могу ответить не сразу



06 OTUS

0 проекте OTUS



Миссия проекта

соединить лучшие компании с самыми
мотивированными специалистами отрасли

Образование в OTUS



Программы курсов

лицензия на образовательную деятельность
разработаны с учетом требований
к специалистам на открытые вакансии
топовых IT-компаний middle и senior уровня

Библиотека Pandas.

Основы и методы работы



Козлов Роман

главный аналитик ДТСЗН г. Москвы

kozlovrv87@gmail.com ~ email
@rokokzk ~ telegram



Цели вебинара

После 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 уметь

Работа с большим
КОЛ-ВОМ ИСТОЧНИКОВ
данных



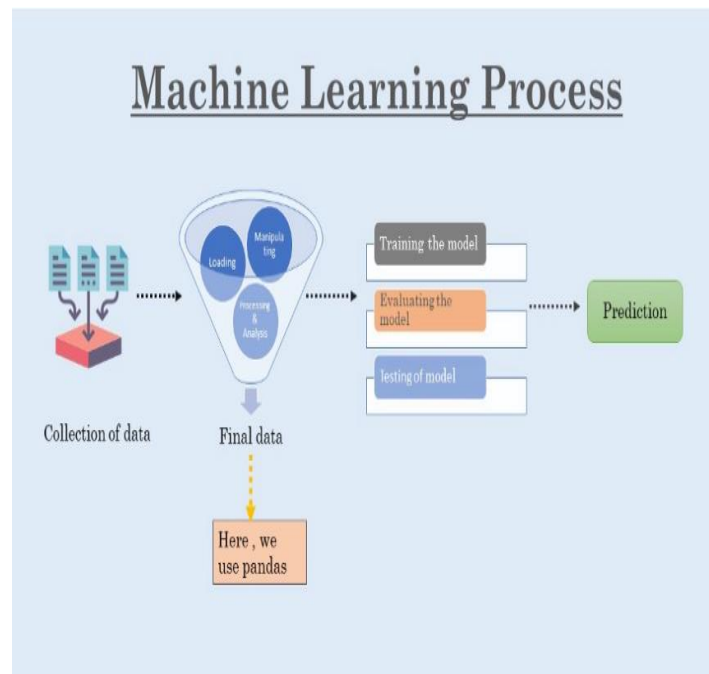
- CSV
- Excel
- HTML-таблицы
(`<table></table>`)
- JSON



Смысл

Зачем вам это уметь

Масштабируемость
под сложность
задачи



Смысл

Зачем вам это уметь

Интуитивная
понятность
(схожесть методов
с SQL)



- JOIN - .join и .merge
- GROUPBY – .groupby
- ORDER BY – .sort_values



Основы Pandas. Series и Dataframe



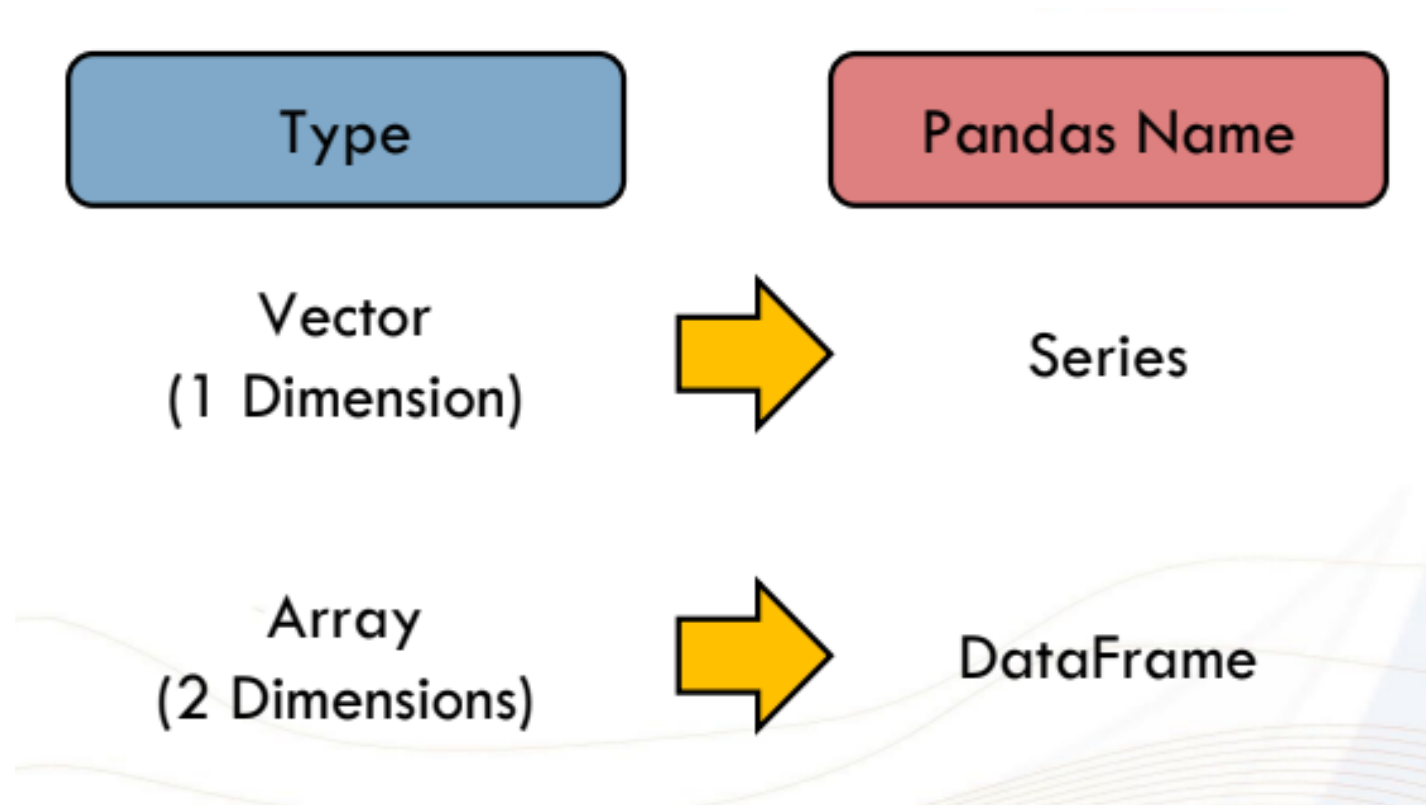
Series

Одномерный
маркированный(индексированный)
объект, содержащий массив данных

Dataframe

Двумерный
маркированный(индексированный)
объект, содержащий массив данных

Связь с Numpy



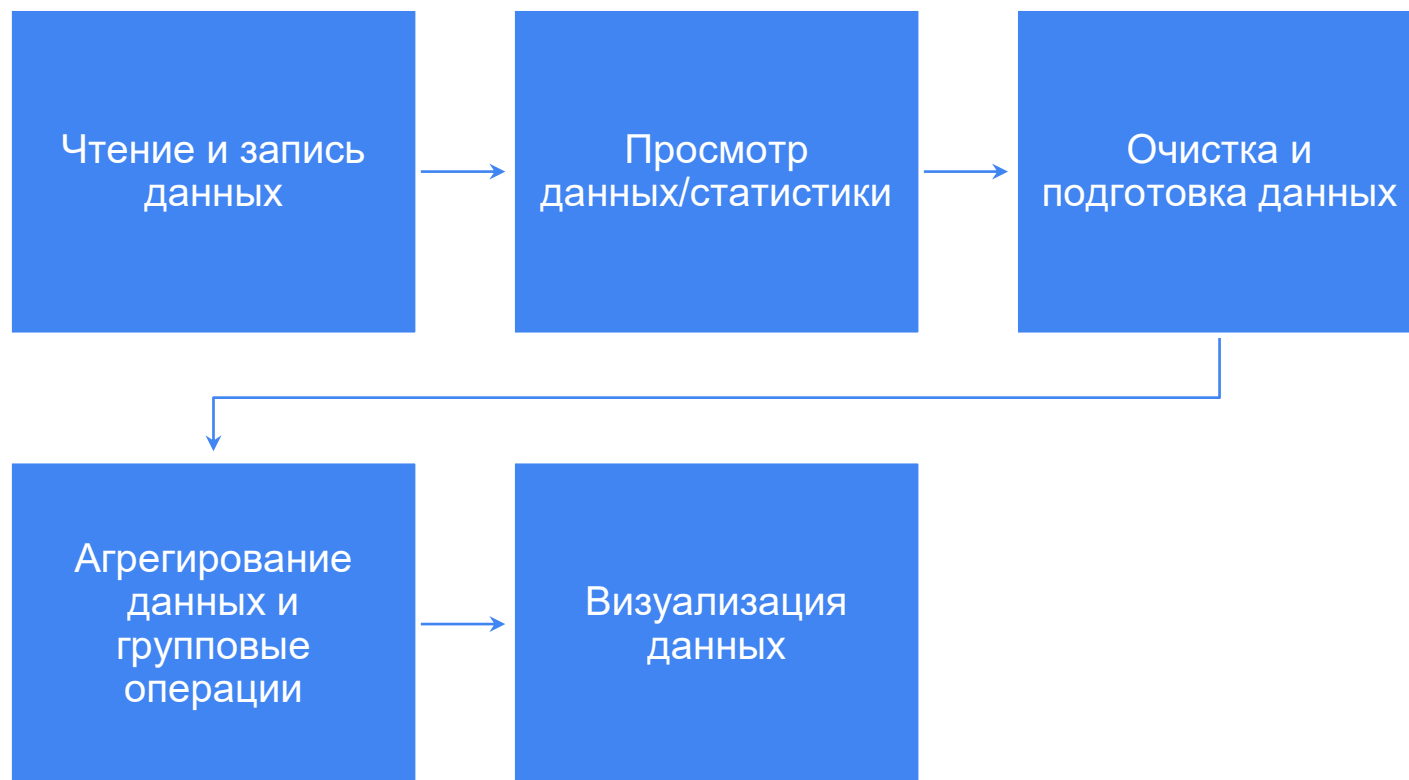
Series и DataFrame

Series 1		Series 2		Series 3		DataFrame			
Mango		Apple		Banana		Mango	Apple	Banana	
0	4	0	5	0	2	0	4	5	2
1	5	1	4	1	3	1	5	4	3
2	6	2	3	2	5	2	6	3	5
3	3	3	0	3	2	3	3	0	2
4	1	4	2	4	7	4	1	2	7

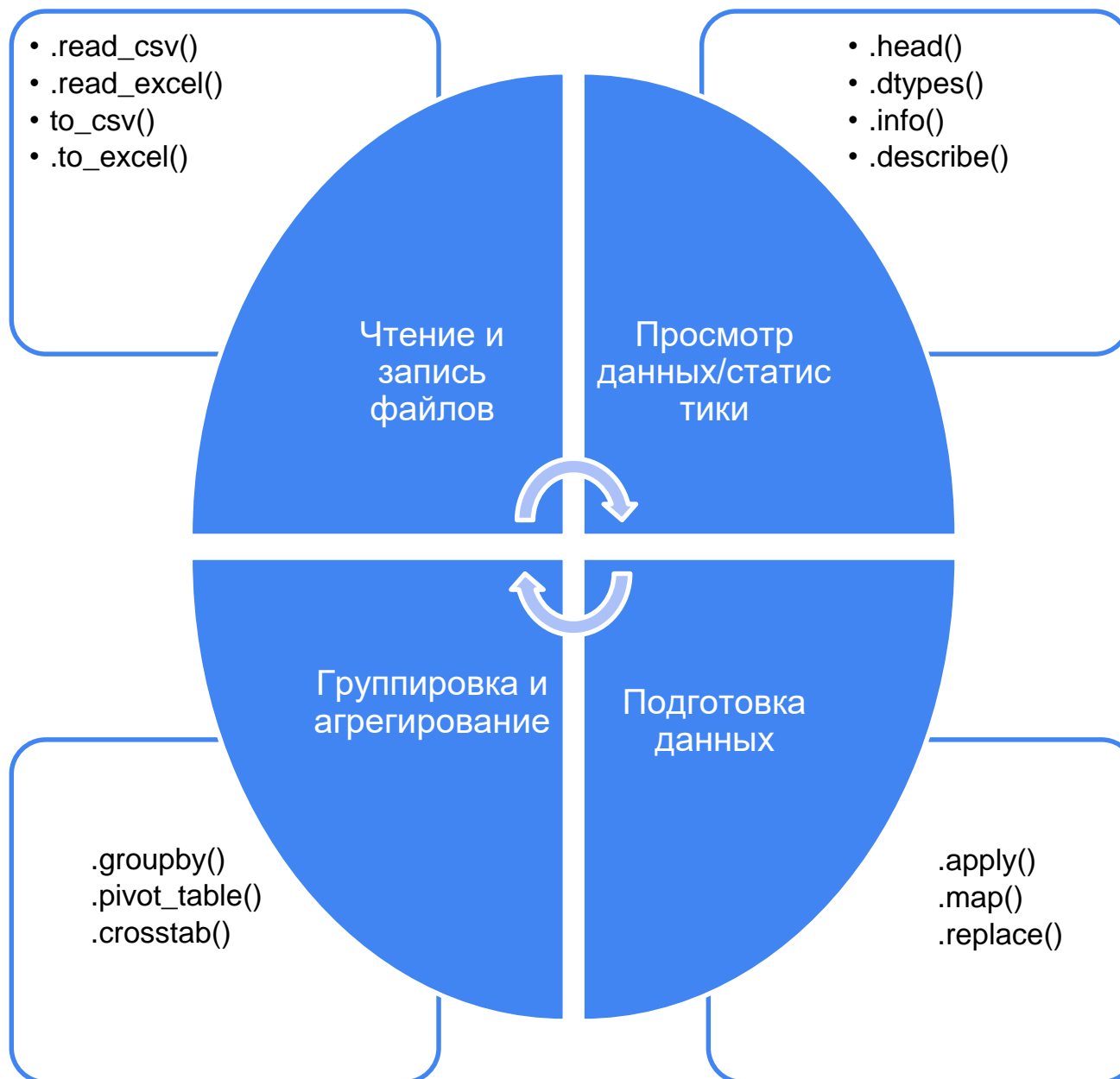
Практика

Работа с данными. Основные методы

Этапы работы с данными (разведочный анализ данных)



Основные методы Pandas



Практика



Знакомство с командой и программой курса



Карьерная информация

otus.ru

Анализ позиции Аналитик данных

811

Вакансий Аналитик данных в феврале
2022 г. для соискателей без опыта.

*вакансии для аудитории с опытом работы до года. Источник — hh

Медиана зарплатных предложений (Junior - Middle)

Аналитик данных
без опыта

60 000

Аналитик данных
1-3 года

150 000

Вакансии из Телеграм каналов

Младший аналитик данных (Junior data analyst)

от 90 000 руб. на руки

ООО Финансово-правовой альянс ✓

ОСНОВНЫЕ ЗАДАЧИ:

- Работа с множеством источников данных: сбор, очищение, изучение, анализ данных в корпоративной информационной среде;
- Визуализация полученных результатов, создание схем, диаграмм, графиков, дашбордов;
- Проведения исследований для отделов продаж, маркетинга, финансового отдела, юридического отдела с целью поиска способов увеличения эффективности компании;
- Прогнозирование: изучение, разработка, тестирование методов анализа и прогнозирования временных рядов;
- Выдвижение гипотез и активное обсуждение их с внутренним заказчиком.

ТРЕБОВАНИЯ К СПЕЦИАЛИСТУ:

- Владение Python (в т.ч. библиотеками NumPy, Pandas, scikit-learn);
- Умение работать с реляционными СУБД, знание SQL;
- Владение базовыми знаниями мат.статистики и теории вероятностей на уровне выпускника ВУЗа;
- Хорошая межотраслевая коммуникация, умение общаться с представителями разных профессий;
- Системное мышление, аналитический склад ума.

БУДЕТ ПЛЮСОМ:

- Владение каким-либо из BI-инструментов (MS Power BI, Tableau, Dataiku и т. п.);

Источник — hh



Вакансии из Телеграм каналов

Аналитик данных

от 150 000 руб. на руки

Ожидаемые результаты на этой позиции:

- Построение системы доступа к данным по извлечению статистических данных для PR материалов.
- Приоритизация задач для разработки на основе данных с помощью воронок, когортного анализа, сегментации.
- Регулярная подготовка статистических данных.

Ожидания от вашего опыта:

- Perfect SQL.
- Опыт работы с Python (или R), Jupyter Notebooks.
- Опыт работы с BigQuery или другими облачными базами данных.
- Высшее математическое/техническое законченное образование.
- Внимание к деталям, трепетное отношение к точности данных.
- Опыт в mobile. Базовое понимание UX и его влияние на метрики. Работа с системами аналитики: Firebase, Amplitude, Mixpanel.
- Опыт визуализации данных и создания dashboards.

Источник — hh

Вакансии из Телеграм каналов

Аналитик данных

Yandex.Cloud

офис

ссылка на отклик



от 150 000 ₽
net по данным careerspace

Что нужно делать:

- извлекать из исходных данных ценную информацию и скрупулезно обрабатывать ее для последующего анализа;
- работать с данными на каждом из этапов их трансформации: от появления информации в конкретном источнике до представления в витринах;
- вместе с командой аналитиков добиваться стабильного расчета бизнес-метрик;
- поддерживать и модифицировать созданное решение, отвечать за корректность данных, предоставляемых для принятия решений.

Мы ждем, что вы:

- программируете на Python;
- хорошо знаете SQL и понимаете устройство современных реляционных и нереляционных баз данных;
- работали с хранилищами данных, понимаете их устройство;
- работаете со специализированными аналитическими базами данных Greenplum, Vertica, ClickHouse, Teradata.

Будет плюсом, если вы:

- работали со стеком технологий Hadoop, Spark, Hive;
- создавали процессы ETL и ELT, знаете инструменты Luigi, Airflow;
- проектировали хранилища данных.

Источник — hh



Вакансии из Телеграм каналов

Компания: Leroy Merlin

Должность: Data Analyst (Поиск и Рекомендации)

Вилка: от 200 000р до 300 000р на руки

Город: удаленно или офис (Москва)

Задачи, которые решает команда:

Поисковые системы для всех публикационных каналов компании;

Создание моделей ранжирования и алгоритмов работы с поисковыми запросами пользователей;

Разработка алгоритмов для систем рекомендации товаров;

Разработка моделей NLP.

Основной стек: Elasticsearch, ClickHouse, Airflow, Apache NiFi, Greenplum, Python, Power BI.

Обязанности:

Анализ различных источников данных с целью выявления новых возможностей для развития продукта;

Умение анализировать и визуализировать данные с помощью Python;

Декомпозиция задач для команды разработчиков.

Требования:

Аналитический склад ума, системное мышление, умение быстро вникать в суть проблемы, широкий кругозор, быстрая обучаемость;

Опыт использования SQL;

Умение обрабатывать и структурировать большие массивы информации;

Умение решать аналитические задачи с использованием Python или других языков программирования.

Рефлексия

Вопросы?



Ставим "+",
если вопросы есть



Ставим "-",
если вопросов нет

Цели вебинара

Проверка достижения целей

1. Узнали, зачем нам изучать Pandas
2. Узнали, из каких структур(объектов) состоит Pandas
3. Научились загружать данные в Dataframe
4. Узнали методы предобработки данных
5. Научились группировать и преобразовывать данные

Вопросы для проверки

1. Чем Dataframe отличается от Series?
2. Что делают методы Pandas drop и append?
3. Как группировать данные в Dataframe методом groupby?

Список материалов для изучения

Аналитика и Python:

1. Уэс Маккини: Python и анализ данных
2. <https://progerlib.ru/python-books>

Pandas:

1. <https://coderlessons.com/tutorials/python-technologies/vyuchit-python-panda/python-pandas-kratkoe-rukovodstvo>
2. <https://pandas.pydata.org/docs>

Практика:

1. <https://www.stratascratch.com/>
2. <https://www.kaggle.com/>



0 курсе

Python для аналитики



старт обучения: 25.02.2022



Заполните, пожалуйста,
опрос о занятии

<https://otus.pw/W98R/>

Спасибо за внимание!