



Обработка, анализ и визуализация больших данных на языке Python. Библиотеки pandas, numpy, matplotlib и seaborn

Код курса: PTN-DA-2

Обработка, анализ и визуализация больших данных на языке Python. Библиотеки pandas, numpy, matplotlib и seaborn

Код курса: PTHN-DA-2

| | |
|--------------------------|--------------------------|
| Длительность | 40 ак. часов |
| Формат | |
| Разработчик курса | Академия АйТи |
| Тип | Учебный курс |
| Способ обучения | Под руководством тренера |

О курсе

Всего за несколько лет средства анализа данных, доступные разработчикам на Python, совершили рывок вперед. Появились мощные пакеты, реализующие алгоритмы машинного обучения, обработку естественных языков, статистический анализ и визуализацию. Инструменты языка Python просты в использовании, при этом имеют широкие возможности применения. Программирование на Python – простой и эффективный вариант для вхождения в популярную сферу Data Science. Уникальная особенность языка – возможность быстрого встраивания анализа данных в веб-приложения. Курс «Data Science. Обработка, анализ и визуализация больших данных на языке Python Библиотеки pandas, numpy, matplotlib и seaborn» предназначен для тех, кто ищет гибкий инструмент обработки, анализа и визуализации данных и планирует применять свои навыки в работе большими массивами информации. Слушатели изучат важнейшие и широко распространенные библиотеки numpy, Matplotlib и Pandas, которые массово применяются в различных областях производственной, финансовой и научной деятельности. Обучение на курсе построено на примерах реальных проектов в области обработки данных. Курс читается на базе операционной системы Linux, однако его материал может быть применен и на Windows. Предварительное требование – прохождение курса «Вводный курс по Python для анализа данных» или знания в объеме этого курса.

Подробная информация

Профиль аудитории:

- программисты
- аналитики
- научные работники

Предварительные требования:

- знание основ программирования на языке Python

По окончании курса слушатели смогут:

уметь:

- объединять и переформировывать данные
- работать с большими массивами информации
- применять библиотеки numpy, Matplotlib и Pandas в различных областях производственной, финансовой и научной деятельности

Программа курса

Модуль 1. Установка и настройка программного обеспечения. Базовый инструментарий работы в Jupyter notebook

- Принципы организации и управления вычислениями
- Используемые языки программирования и взаимосвязь между ними
- Обзор библиотек и инструментов
- Необходимые пакеты Python: numpy, pandas, matplotlib, seaborn, Jupyter и другие
- Дистрибутив Anaconda, его преимущества и недостатки
- Установка пакетов в Linux
- Установка пакетов в Windows. Возможные проблемы и методы их решения
- Методы и функции
- Основные команды и методы работы в Jupyter notebook

Модуль 2. Библиотека numpy. Вычислительные задачи

- Структура библиотеки
- Типы данных библиотеки numpy
- Векторы и массивы, специальные типы данных
- Понятие массива и его основные характеристики
- Принципы вычислений
- Универсальные функции
- Важнейшие стандартные функции
- Работа с массивами и матрицами

Модуль 3. Библиотека pandas и анализ данных

- Объект Series
- Объект DataFrame
- Загрузка данных
- Выгрузка данных
- Чтение и запись данных в формате Excel
- Чтение и запись JSON-файлов
- Загрузка данных из интернета и из базы данных SQL
- Загрузка CSV-файлов из Интернета
- Основы работы с датафреймами
- Организация колонок и строчек
- Создание срезов объекта Series

- Выравнивание данных по меткам индекса
- Выполнение логического отбора
- Переиндексация объекта Series

Модуль 4. Первичная обработка данных

- Сбор данных
- Очистка данных
- Подготовка данных
- Объединение данных
- Сцепление и наложение
- Слияние данных
- Замена значений
- Работа с пропущенными значениями и их заполнение
- Повторы в данных
- Объединение и переформирование данных

Модуль 5. Статистика и анализ

- Фильтрация
- Группировка
- Агрегация
- Описательные статистики
- Получение итоговых описательных статистик
- Измерение центральной тенденции: среднее, медиана и мода
- Вычисление дисперсии и стандартного отклонения
- Вычисление ковариации и корреляции
- Распределения
- Разведывательный анализ
- Постановка гипотезы

Модуль 6. Библиотека matplotlib. Визуализация данных

- Основные элементы диаграммы
- Создание диаграммы
- Виды графиков и диаграмм
- Оформление и кастомизация графиков
- Форматы изображений

Модуль 7. Библиотека seaborn. Наглядная визуализация сложных данных

- Виды графиков, их корректное использование и интерпретация
- Принципы хорошей визуализации, основные ошибки при визуализации данных
- График функции
- Гистограмма
- Визуализации отфильтрованных и сгруппированных данных

Обработка, анализ и визуализация больших данных на языке Python. Библиотеки pandas, numpy, matplotlib и seaborn

Код курса: PTHN-DA-2

[Посмотреть расписание курса и записаться на обучение](#)

Обращайтесь по любым вопросам

к менеджерам Академии АйТи

+7 (495) 150 96 00 | academy@academyit.ru