

Анализ данных в Python

Data analysis with Python

1 триместр, обязательный

Лектор: Данила Кочнев

Объем дисциплины: 3 зачетных единицы, всего 108 академических часов, в том числе 52 академических часов составляет контактная работа с преподавателем, 56 академических часов составляет самостоятельная работа магистранта.

Формат занятий и отчетность студентов: Курс состоит из лекционных и практических занятий. Отчетность студентов включает 6 домашних работ (60%), практическую работу (20%) и итоговый экзамен (20%).

Критерии оценки: Для оценки «отлично» необходимо набрать 85% баллов по курсу, «хорошо» – 65% баллов, удовлетворительно – 40% баллов.

Содержание курса:

- 1. Введение в язык Python и его базовый синтаксис.* Числовые типы данных, арифметические действия над числами. Переменные и связывание. Циклы for и while. Подключение модулей. Строки и списки, срезы и стандартные методы строк и списков. Списковые включения. Особенности ссылочной модели данных в Python.
- 2. Функции.* Встроенные и пользовательские функции, пространства имен. Рекурсивное задание функций, концепция динамического программирования. Анонимные функции. Декорирование функций.
- 3. Базовые алгоритмы.* Понятие алгоритма и его сложности, O-нотация. Однопроходные алгоритмы, линейный и двоичный поиск. Классические теоретико-числовые алгоритмы (наивный тест простоты, алгоритм Евклида, решето Эратосфена). Алгоритмы сортировки (пузырьком, вставками, слиянием). Понятие жадного алгоритма.
- 4. Элементы продвинутого синтаксиса.* Функции “map”, “filter”, “reduce”, “enumerate”, “zip”. Концепции итераторов и генераторов. Идея хэш-таблицы и построенные на ней типы данных: множества и словари, их методы и приложения. Элементы объектно-ориентированного программирования: классы, экземпляры классов, инкапсуляция и наследование.
- 5. Основные библиотеки для анализа данных в Python.* Модуль “numpy”, матричная алгебра, массивы “numpy.array” и их отличия от списков. Модуль “pandas”, методы работы с табличными данными, типы “pandas.series” и “pandas.DataFrame”, их методы. Предобработка данных. Модуль “matplotlib” и визуализация данных в Python.
- 6. Базовая статистика и эконометрика в Python.* Разведывательный анализ данных, расчет описательной статистики. Модуль “statsmodels”, оценка и анализ классических эконометрических моделей в Python.

(Программа носит предварительный характер и может незначительно изменяться)

Литература:

Любанович, Б. Простой Python. Современный стиль программирования. СПб.: Питер. 2022. 592 с.

Маккинни, У. Python и анализ данных. М: ДМК Пресс. 2020. 482 с.

Нисчал, Н. Python – это просто. Пошаговое руководство по программированию и анализу данных. БХВ. 2022. 416 с.

Федоров, Д. Ю. Программирование на Python: учебное пособие для вузов. Москва: Издательство Юрайт. 2025. 187 с.

Хайнеман, Дж. Алгоритмы. С примерами на Python. СПб.: Питер. 2023. 304 с.

Cormen, T. Algorithms unlocked. MIT Press. 2013. 237 p.

Heiss, F., Brunner, D. Using Python for Introductory Econometrics. Independently published (2020). 428 p.

Буйначев, С. К. Основы программирования на языке Python: учебное пособие. Екатеринбург: Изд-во Урал. ун-та. 2014. 91 с.

Бхаргава, А. Грокаем Алгоритмы. Иллюстрированное пособие для программистов и любопытствующих. Лидер. 2024. 288 с.

Кнут, Д. Искусство программирования (том 1). Вильямс. 2020. 720 с.

Лутц, М. Изучаем Python (том 1). Вильямс. 2019. 832 с.

Мартин, Р. Чистый код: создание, анализ и рефакторинг. СПб.: Питер. 2022. 464 с.

Мэттиз, Э. Изучаем Python: программирование игр, визуализация данных, веб-приложения. СПб.: Питер. 2020. 512 с.

Чернышев, С. А. Основы программирования на Python: учебное пособие для вузов. Москва: Издательство Юрайт. 2021. 286 с.

Tao, R., Brooks, C. Python Guide to Accompany Introductory Econometrics for Finance. Cambridge University Press (2019). 256 p.

Академическая честность: Любые нарушения норм академической этики, в том числе плагиат, несамостоятельное выполнение домашних и контрольных работ по курсу, использование запрещенных вспомогательных материалов – считаются недопустимыми.