



Python для аналитики

Python для аналитики

Длительность курса: 140 академических часов

1 Знакомство с Jupyter Notebook, синтаксис Python

Цели занятия:

познакомиться с Python и средой разработки Jupyter Notebook;
объяснить какие бывают типы данных, виды арифметических операций;
пройдем базовые типы данных, простейшие операции с числами.

Краткое содержание:

python 3, jupyter notebook.

Домашние задания

1 Jupyter и библиотеки

Цель: Установить необходимое для работы ПО.

1. Установить jupyter
 2. Установить оговоренные библиотеки
 3. Проверить, что всё загружается и работает
-

2 Итерируемые типы данных. Методы работы со списками, кортежами, словарями, множествами

Цели занятия:

разобраться с методами работы с данными, хранящимися в списках, словарях, множествах, кортежах.

Краткое содержание:

python 3;
jupyter notebook.

3 **Условные операторы, обработка исключений, циклы. Comprehensions, генераторы**

Цели занятия:

научиться пользоваться условными операторами и обработкой исключений, итерироваться по данным.

Краткое содержание:

python 3;
jupyter notebook.

Домашние задания

1 Скрипт с рандомными данными

Цель: Научиться писать скрипт, который бы выбрал случайные города и числа, соединил бы их в словарь и посчитал бы кол-во пользователей по названию города.

1. Используя библиотеку генерации случайных данных, создать 2 списка: первый - со 100 случайными целыми числами (они будут id клиента), второй - со 100 случайными городами(их надо выбрать из заданного списка городов)
 2. Создать из 2ух списков словарь, в котором key - id клиента, value - город.
 3. С помощью цикла и условного оператора посчитать кол-во клиентов из Москвы
-

4 **Функции, области видимости, глобальные и локальные переменные, лямбда функции**

Цели занятия:

научиться составлять простейшие функции в сочетании с условными операторами и циклами.

Краткое содержание:

python 3;
jupyter notebook.

Домашние задания

1 **Функция "Калькулятор"**

Цель: Написать калькулятор на функциях.

Студент поймет, что есть переменные с разными типами данных, функции, есть циклы и условия, поймет, как писать текст программы, чтобы он выполнялся.

Написать функцию ""Калькулятор"", которая запрашивает у пользователя 2 числа и знак и выполняет операции:

1. сложения
2. вычитания
3. деления
4. умножения

В случае выполнения деления на ноль или ввода некорректного знака, функция должна предупреждать соответствующим образом

5 **Работа со строковыми данными. Регулярные выражения**

Цели занятия:

поработать со строковыми типами данных, разобраться со встроенными в Python методами для работы и преобразования строк; обсудить библиотеку re и принципы построения регулярных выражений для поиска по строкам.

Краткое содержание:

python 3, jupyter notebook.

Домашние задания

1 Проверка госномеров, телефонных номеров

Цель: Научиться писать с помощью регулярных выражений код, выполняющий поиск различных объектов (декораторов, почтовых ящиков, тел. номеров, аббревиатур и т.д.) в строке.

1. Используя регулярные выражения, необходимо выделить в список все адреса принадлежащие домену .com из строки:

```
*"https://habr.com/,  
https://practice.geeksforgeeks.org/, slack.com,  
https://docs.google.com/, https://stackoverflow.com,  
https://flower.dev, https://spb.rr.ru, https://mail.ru/  
https://research.ibm.com/, https://pythonworld.ru/  
https://github.com/"*
```

2. Найдите (в тексте reg_text_1.txt) используя регулярные выражения все декораторы (т. е. конструкции вида начинающихся со знака @, далее латинская буква и произвольное число латинских букв и цифр) и все функции (т. е. все подстроки начинающиеся на "def", далее латинская буква и произвольное число латинских букв и цифр, а после "():").

3. Выберите из текста (reg_text_2.txt из материалов к уроку):

- все дробные числа с разделителем дробной части в виде точки. Разряды целой части могут не выделяться или отделяться пробелом или запятой.
 - Все аббревиатуры (т. е. сокращения вида IP, SQL) и сохраните их в список. Подсчитайте суммарное число найденных аббревиатур и число уникальных аббревиатур.
-

6 Работа с файловой системой и модули

Цели занятия:

научиться читать и записывать файлы, импортировать сторонние библиотеки, пользоваться системой управления пакетами pip.

Краткое содержание:

python 3, jupyter notebook.

Домашние задания

1 Скрипт, читающий из файла

Цель: Научиться писать скрипт, который бы читал из файла и выводил на экран и в другой файл части строк, удовлетворяющие определенным критериям.

1. Нужно открыть файл при помощи Python на чтение
2. Прочитать оттуда строки и проитерироваться по ним
3. Для каждой строки нужно применить паттерн, который бы доставал нужный нам номер телефона в определенном формате
4. Найденные номера телефонов записать построчно в другой файл.

2 Библиотеки по работе с данными и визуализациями

1 Библиотека NumPy, вектора и матрицы

Цели занятия:

научиться пользоваться библиотекой NumPy и её возможностями.

Краткое содержание:

python 3, jupyter notebook.

2 Библиотека Pandas, Series и Dataframe, методы для работы с ними

Цели занятия:

поработать с табличными данными в библиотеке Pandas, узнать разницу между Series и Dataframe.

Краткое содержание:

python 3, jupyter notebook.

3 Основы статистики, статистические отчеты

Цели занятия:

вспомнить основы статистики и узнать какие графики подходят для отображения.

Краткое содержание:

python 3, jupyter notebook.

4 Визуализация в Matplotlib, Seaborn, Plotly

Цели занятия:

научиться строить понятные графики, гистограммы, скаттерплоты и т.д.

Краткое содержание:

python 3, jupyter notebook.

Домашние задания

1 Работа с библиотеками визуализации

Цель: Загрузить датасет, построить визуализации.

Студент ознакомится с библиотекой для построения графиков, научится сам агрегировать данные, чтобы получить нужный ему результат.

Выберите датасет из sklearn https://scikit-learn.org/stable/datasets/toy_dataset.html
Постройте 3 графика, используя matplotlib, 3 – seaborn, 1 – plotly (всего 7. Можно больше).
Сделайте выводы по каждому из графиков (как переменные распределены, особенности описательных статистик, взаимодействия между переменными и т.д.)
Прикрепите ipynb файл для сдачи

**5 Разбор кейса.
Разведочный
анализ данных
(EDA) на
примере**

Цели занятия:

применить полученные в предыдущем модуле знания для подробной аналитики на тестовой датасете.

Краткое содержание:

python 3, jupyter notebook.

Домашние задания

1 Разведочный анализ данных на датасете

Цель: Провести разведочный анализ данных на датасете adult.csv.

1. Нужно будет загрузить классический датасет adult.csv (в материалах лекции)
 2. Провести очистку и предобработку данных
 3. Построить визуализации, позволяющие выявить зависимости между различными показателями
 4. Оформить выявленные инсайты в виде выводов.
-

**6 Q&A (разбор
домашних
заданий)**

Цели занятия:

получить ответы на вопросы по ДЗ;
получить ответы на вопросы по приложениям.

Краткое содержание:

типичные ошибки при выполнении ДЗ;
наставники и преподаватели ответят на ваши вопросы.

3 Практики. Продуктовая и маркетинговая аналитика

1 Введение в продуктовую аналитику на Python. Метрики и методы

Цели занятия:

разобраться, что такое продуктовая и маркетинговая аналитика, для чего она нужна, где применяется и какие метрики и методы для нее используются.

Краткое содержание:

python 3, jupyter notebook.

2 Когортный анализ. Работа с датами и библиотекой dt

Цели занятия:

применить когортный анализ в Python.

Краткое содержание:

python 3, jupyter notebook.

Домашние задания

- 1 Построение когорт по времени регистрации посетителей на сайте и вычисление показателя

Цель: Построить когорт по времени регистрации посетителей на сайте и вычислить показатель Retention Rate по когортам на тестовом датасете.

1. Нужно будет на датасете, показывающим посещаемость условного сайта, разбить пользователей на когорты по месяцу/неделе.
 2. Далее необходимо будет рассчитать lifetime по пользователям и сгруппировать по когортам.
 3. Затем надо рассчитать Retention Rate по когортам и Lifetime.
 4. В конце построить heatmap и сделать соответствующие выводы.
-

3 RFM - анализ

Цели занятия:

применить RFM анализ в Python.

Краткое содержание:

python 3, jupyter notebook.

4 ABC - XYZ анализ

Цели занятия:

применить ABC-XYZ анализ на Python.

Краткое содержание:

python 3, jupyter notebook.

Домашние задания

1 RFM и ABC анализ

Цель: Разбить товары и клиентов на сегменты при помощи RFM и ABC анализа на примере тестового датасета с продажами.

На основе датасета с продажами товара, необходимо разбить покупателей по сегментам на основе RFM анализа.

Продукцию в датасете (<https://www.kaggle.com/datasets/kyanyoga/sample-sales-data>) надо также разбить на группы с помощью ABC-XYZ анализа.

Начальные данные обогатить данными проведённых анализов.

Сделать выводы о том, какие покупатели и товары представляют наибольший интерес/на какие обратить внимание.

5 Проверка гипотез. A/B тестирование

Цели занятия:

научиться правильно проводить A/B тестирование.

Краткое содержание:

python 3, jupyter notebook.

Домашние задания

1 Проверка гипотез в рамках AB теста

Цель: Проверить гипотезы с помощью различных статистических тестов и сравнить результаты.

Используя код и данные с урока, выберите два любых других статистических теста. Сравните их результаты с разобранным тестом. В чем различия между тестами?

4 Работа с базами данных, парсинг данных с сайтов, взаимодействие с API

- 1 **Обзор про базы данных, SQL и теорию множеств. Таблицы, представления, простые выборки, функции**
- Цели занятия:**
- научиться делать простые выборки из реляционной базы.
- Краткое содержание:**
- python 3, jupyter notebook, sqlite.
-

- 2 **Join, exists, вложенные запросы, group by, having. Индексы и план запроса**
- Цели занятия:**
- научиться делать более сложные выборки из реляционных баз, которые включают данные из нескольких таблиц.
- Краткое содержание:**
- python 3, jupyter notebook.
-

3 Оконные функции в SQL, виды и область применения

Цели занятия:

научиться пользоваться оконными функциями для построения сложных запросов.

Краткое содержание:

python 3, jupyter notebook.

Домашние задания

1 Работа с датасетом

Цель: 1. Загрузить датасет, записать его в базу.
2. Сделать выборку, вывести ее на экран.

Студент ознакомится с тем, как записывать данные в базу и как сделать при помощи Python выборку.

1. Нужно скачать датасет
 2. Любыми способами загрузить его в базу (написать цикл на много инсертв, написать цикл делающий один большой инсерт, через командную строку положить файл в базу, если вам это удастся)
 3. Написать простую выборку, которая бы выбирала все данные, вывести их на экран (первые 10 строк).
-

4 HTTP. Обзор по работе с HTTP в Python. Библиотеки, requests

Цели занятия:

научиться делать запросы по HTTP через Python и получать ответ от сервера.

Краткое содержание:

python 3, jupyter notebook.

**5 Парсинг
данных из
HTML**

Цели занятия:

научиться пользоваться библиотеками для разбора XML - eTree, lxml, beautiful soup.

Краткое содержание:

python 3, jupyter notebook.

6 Работа с JSON, работа с CSV и xls

Цели занятия:

выбрать данные из JSON, и электронных таблиц.

Краткое содержание:

python 3, jupyter notebook.

Домашние задания

1 IMDB

Цель: Спарсить с IMDB 50 фильмов с актерами, записать в базу.

Студент ознакомится с тем как делать запросы по сети. Также ознакомится с тем как парсить HTML, с тем как проектировать простейшую базу данных для своих нужд, и записывать туда прочитанные данные.

1. Поставить requests
2. Найти на IMDB любой понравившийся список фильмов - по ссылкам с этой страницы мы будем ходить скриптом
3. Нам нужно спарсить со страницы с фильмом: название, год, жанры, актеров. Запомним же это.
4. Нужно продумать каким способом хранить все эти данные, чтобы избежать дублирования - нам понадобятся отношения многие-ко-многим.
5. Нужно спарсить со странички ссылки фильмов
6. Нужно проитерироваться по этим ссылкам, получить страницы фильмов
7. Со страницы с фильмом взять нужную нам информацию и положить ее в базу данных выбранным нами способом

1 **Введение в машинное обучение. Обучение с учителем/без учителя. Библиотека Scikit-learn**

Цели занятия:

рассмотреть виды машинного обучения, познакомиться с библиотекой scikit-learn.

Краткое содержание:

python 3, jupyter notebook.

2 **Линейная регрессия, Построение модели и метрики**

Цели занятия:

научиться использовать способы построения и метриках простой линейной регрессии.

Краткое содержание:

python 3, jupyter notebook.

Домашние задания

1 Модуль машинного обучения - 1

Цель: В этом ДЗ вы построите модель машинного обучения, предсказывающую цены на жилье в Бостоне.

Нужно будет построить модель линейной регрессии:

- 1) с гиперпараметрами отличными от тех, что "из коробки"
 - 2) удалить один или несколько показателей, имеющих маленькую корреляцию с таргетом или обладающими признаками мультиколлениарности
 - 3) надо будет сравнить получившиеся метрики MAE, MSE RMSE и сделать соответствующие выводы
-

3 **Логистическая регрессия. Построение модели и метрики**

Цели занятия:

научиться использовать способы построения и метриках логистической регрессии.

Краткое содержание:

python 3, jupyter notebook.

Домашние задания

1 Модуль машинного обучения - 2

Цель: В этом ДЗ вы построите модель машинного обучения, предсказывающую, какие пассажиры Титаника выживут.

Нужно будет построить модели логистической регрессии с нормализацией данных и без нее. Сравнить полученные метрики ROC AUG, Accuracy, Recall, F1 и сделать выводы.

4 **Кластеризация. Построение модели и метрики**

Цели занятия:

научиться использовать способы построения и метриках кластеризации.

Краткое содержание:

python 3, jupyter notebook.

Домашние задания

1 Построение модели метрики

Цель: В этом ДЗ вы раобьете ирисы на основе размера цвета чашелистника по кластерам.

Нужно будет построить модель с различным количеством кластеров (от 4 до 8) и сравнить полученные метрики. Сделать выводы.

5 Q&A (разбор домашних заданий)

Цели занятия:

получить ответы на вопросы по ДЗ;
получить ответы на вопросы по приложениям.

Краткое содержание:

типичные ошибки при выполнении ДЗ;
наставники и преподаватели ответят на ваши вопросы.

1 Выбор темы и организация проектной работы

Цели занятия:

выбрать и обсудить тему проектной работы;
спланировать работу над проектом;
ознакомиться с регламентом работы над проектом.

Краткое содержание:

правила работы над проектом и специфика проведения итоговой защиты;
требования к результату проекта и итоговой документации.

Домашние задания

1 Проектная работа

Цель: Выбрать и утвердить в чате по ДЗ темы проекта, разработать и презентовать проект.

1. Выбрать тему
2. Утвердить темы в чате по ДЗ
3. Презентовать проект

Проектная работа будет представлять из себя получение данных, разведочный анализ, обогащение данных с помощью когортного, RFM или ABC анализа, построение модели машинного обучения.

2 Консультация по проектам и домашним заданиям

Цели занятия:

получить ответы на вопросы по проекту, ДЗ и по курсу.

Краткое содержание:

вопросы по улучшению и оптимизации работы над проектом;
затруднения при выполнении ДЗ;
вопросы по программе.

**3 Защита
проектных
работ**

Цели занятия:

защитить проект и получить рекомендации экспертов.

Краткое содержание:

презентация проектов перед комиссией;
вопросы и комментарии по проектам.

**4 Подведение
итогов курса**

Цели занятия:

узнать, как получить сертификат об окончании курса,
как взаимодействовать после окончания курса с OTUS и
преподавателями, какие вакансии и позиции есть для
выпускников (опционально - в России и за рубежом) и
на какие компании стоит обратить внимание.

Краткое содержание:

организационные вопросы;
рынок вакансий по направлению;
статистика курса и вопросы по курсу.