

АНАЛИЗ ДАННЫХ И DATA SCIENCE ДЛЯ ШКОЛЬНИКОВ



Курс рассчитан для студентов от 15 до 17 лет
Длительностью 9 месяцев, 144 ак. часа с преподавателем,
144 ак. часа самостоятельно, 2 пары в неделю

1. Введение в Data Science



Узнаем про задачи профессии, развеем мифы и расскажем о перспективах работы аналитика данных.

- Темы:**
- Область Data Science и её применений типов аналитических задач
 - Цикл аналитического исследования
 - Формулирование гипотез на основе данных

Инструменты: Среды разработки (Google Colab, Jupyter Notebook)

Навыки Понимание области Data Science и её применений



2. Основы Python

В этом модуле ученики познакомятся с основами программирования на Python — языком №1 для анализа данных. От простых переменных до создания собственных функций — здесь закладывается фундамент для всего дальнейшего обучения.

- Темы:**
- Введение в Python
 - Работа с данными и типами данных
 - Инструкции, выражения и операторы в Python
 - Условные конструкции и логические операторы в Python
 - Циклы и range()
 - Строки
 - Списки и кортежи
 - Множества и словари
 - Функции и параметры функций
 - Основы ООП
 - Исключения и управление потоком выполнения
 - Модули, пакеты



- Инструменты:**
- Python 3
 - Среда разработки (Google Colab, Jupyter Notebook)
 - Основные встроенные библиотеки (math, datetime, random)
 - Работа с файлами различных форматов

- Навыки**
- Написание и отладка программного кода
 - Работа с различными типами данных и структурами
 - Создание функций для многократного использования кода
 - Разработка простых программ для анализа данных



3. Работа с системами версификации

Ученики освоят современные инструменты разработки и коллаборации. Git дает возможность сохранять историю изменений и работать в команде.

Темы:

- Системы контроля версий Git
- Ветки Git
- Синхронизация с облаком в Git

Инструменты:

- Git
- GitHub/GitLab
- Командная строка для работы с Git
- Системы документирования кода и Markdown

Навыки

- Управление историей изменений в коде
- Взаимодействие с удаленными репозиториями
- Работа в команде над общим кодом
- Разрешение конфликтов при объединении кода



4. Базы данных SQL

Этот модуль посвящен языку запросов SQL — универсальному инструменту для работы с данными в реляционных базах данных. Ученики научатся извлекать данные из БД.

Темы:

- Базы данных
- Оператор WHERE (продолжение). Команды ORDER BY и LIMIT
- Соединение таблиц, команда JOIN
- Соединение таблиц. Операции с множествами
- Агрегирование и группировка данных
- Группировка данных (продолжение)
- Оконные функции
- Подзапросы
- Обобщённые табличные выражения (CTE). Views. Временные таблицы
- Основы DDL. Внесение изменений в БД. Основы оптимизации
- Работа с датой и временем. Условные выражения в SELECT

**Инструменты:**

- SQL
- SQLite

Навыки

- Написание запросов различной сложности
- Агрегирование и группировка данных
- Соединение данных из разных таблиц
- Интеграция SQL с Python для анализа данных



5. Библиотеки анализа данных и визуализация

В этом модуле ученики погрузятся в инструменты для эффективной работы с данными. NumPy предоставит возможности быстрых вычислений, а Pandas станет основным инструментом для обработки, очистки и анализа структурированных данных. Также, научившись работать с данными, ученики освоят искусство их визуализации. Модуль фокусируется на превращении чисел в наглядные графики и диаграммы, которые помогают увидеть закономерности и рассказать историю, скрытую в данных.

Темы:

- Введение в работу с NumPy
- Введение в работу с Pandas
- Основные методы и функции библиотеки Pandas. Фильтрация данных
- Загрузка и считывание основных форматов в DataFrame. Загрузка данных из файлов
- Визуализация данных с Seaborn
- Группировка и Seaborn
- Matplotlib

Инструменты:

- NumPy
- Pandas
- Matplotlib и Seaborn

Навыки

- Обработка и трансформация табличных данных
- Статистический анализ наборов данных
- Создание различных типов визуализаций
- Разработка интерактивных информационных панелей



6. Сбор, подготовка и очистка данных

Ученики изучат методы получения данных из различных источников, их предварительную обработку и подготовку для дальнейшего анализа.

Темы:

- Данные как основа аналитики
- Сбор и загрузка данных
- Очистка и подготовка данных для аналитики

Инструменты:

- Pandas для очистки и трансформации данных
- Библиотеки для работы с различными форматами (CSV, JSON, Excel)
- Инструменты для доступа к API
- Методы для валидации и проверки качества данных

Навыки

- Извлечение данных из различных источников
- Обнаружение и обработка пропущенных значений
- Трансформация данных в удобный для анализа формат
- Обработка выбросов и аномалий
- Оценка качества данных и подготовка аналитических отчетов



По окончанию студент получит сертификат Академии ТОП

