



Программа курса «Data Science. Уровень 1: основы работы с данными»

Номер	Название темы	Количество часов	Описание темы
1	Основы программирования на Python для data science	18	Основы языка: переменные и их типы, арифметические операции, операции над строками, функции. Основы объектно-ориентированного программирования: классы, экземпляр класса, поля и методы, наследование, безопасный доступ. Библиотеки NumPy, SciPy: векторы, матрицы, операции над векторами и матрицами, основы линейной алгебры. Работа в среде Pycharm - создание законченных скриптов. Работа в среде Google Colab - Python notebooks - рабочий инструмент.
2	Основы работы с базами данных	16	Основные типы баз данных для data science: основы хранения информации, таблицы и коллекции, отношения между ними, транзакции, уровни изоляции. Реляционные СУБД как основной источник данных для data science: разбор основных возможностей РСУБД на примере PostgreSQL: получение основных данных о ядре БД, psql - встроенный клиент БД; операции DML: выборка, изменение, удаление, вставка данных; операции DDL: создание и удаление баз данных, таблиц, функций, процедур. NoSQL базы данных на примере MongoDB: коллекции, операции по созданию, изменению, вставке и удалению данных. In-Memory базы данных на примере Redis: хранение информации «ключ - значение» - операции вставки, манипуляции и удаление. Примеры работы с основными типами БД из Python из Pycharm и Python Notebook.
3	Основы анализа данных	22	Основы работы с библиотеками Request, BeautifulSoup - сбор данных с HTML страниц. Библиотеки Pandas и Polars: универсальные табличные процессоры без ограничений. Внутреннее устройство, понятие датафрейма, основные операции, преобразование типов, уменьшение размеров датафреймов. Загрузка данных в датафрейм из разных источников, сохранение данных. Основные аналитические операции: query, groupby, aggregate, sort, pivot_table. Библиотеки Matplotlib и Seaborn: основы визуализации данных. Создание графиков разных типов: linechart, barchart, boxplot, catplot, etc. Тюнинг параметров графиков, ресурсные записи. Exploratory Data Analysis (EDA) - базовые принципы анализа данных с использованием графиков. Основы бизнес-анализа и unit-экономики: анализ данных и расчет бизнес-метрики: LTV, RFM, ARPU, ARPPU, etc.
4	Математика для data science	14	Элементарные функции: основные определения, предел функции. Непрерывность и дифференцируемость: производные простых и сложных функций, оптимизация функции одной переменной. Теория вероятностей: сложение и произведение вероятностей, формула Бернулли, условная вероятность: теорема Байеса. Математическая статистика: проверка гипотез, статистический вывод, центральная предельная теорема, сравнение средних значений, корреляция, регрессия, bootstrap, AB тесты.
7	Итоговая аттестация	2	Зачет.

