



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

