

Курс: Data Science: Анализ данных и визуализация в R

Длительность: 40 ак. часов

О курсе

Данный курс является необходимым для получения практических навыков работы с языком программирования **R** и **R-Studio**. Вы узнаете, как загружать, сохранять и преобразовывать данные, а также как писать функции, создавать графики и сопоставлять базовые статистические модели с данными. В дополнение к теоретическим занятиям, на которых вы изучите процесс анализа данных, этот курс фокусируется на практических инструментах, необходимых для анализа и визуализации больших данных. По окончании курса вы овладеете основными навыками обработки, манипулирования и анализа данных различных типов, создания отчетов и документирования вашего кода.

Аудитория

Специалисты по работе с большими данными, бизнес аналитики и руководители желающие получить расширенную теоретическую и практическую подготовку по использованию **R-Studio** в проектах анализа больших данных.

Предварительная подготовка

- Понимание основ статистики

1. Основы статистики и простая линейная регрессия

Что такое ваши данные. Статистические выводы. Введение в машинное обучение. Простая линейная регрессия. Диагностика и трансформация. Коэффициент определенности.

2. Базовое программирование с R

Введение в **R**. Что такое **R**? **R-Studio**, пакеты и рабочая область. Основные элементы языка **R**. Типы объектов данных. Локальный импорт / экспорт данных. Введение функций и управляющих операторов. Углубленное изучение объектов данных. Функции. Программирование функций.

3. Базовые элементы данных

Преобразование данных(**Reshape, Split, Combine**). Символы. Обработка строк. Даты и временные метки. Сбор Веб-данных. API источники данных. Подключение к внешней базе данных.

4. Манипулирование данными с помощью "dplyr"

Подмножество, преобразование и переупорядочение наборов данных. Объединение наборов данных. Групповые операции над наборами данных.

5. Графика данных и визуализация данных

Основные подходы к визуализации данных и графики данных. **Base, Grid, Lattice, ggplot2**. Построение графиков больших данных с **ggplot2**.

6. Расширенная визуализация в R

Настраиваемая графика с помощью **ggplot2**. Титулы, системы координат, масштабы, темы, метки, легенда, **scatterplot** с многомерными данными, визуализация временной последовательности, карты. Интерактивная визуализация в **R**.