otus
онлайн-образование

Полная программа

Аналитик данных

Длительность курса: 114 часов

Модуль 1. Принятие решений в бизнесе на основе данных

Тема 1
Введение в профессию "Аналитик данных"

Цель занятия
познакомиться;
узнать, какие темы предстоит изучить и в каком объеме, чтобы освоить профессию аналитика данных.

Краткое содержание
описание профессии "Аналитик данных".

Тема 2
Принятие решений в бизнесе на основе данных

Цель занятия
понять, как обычно используются данные в бизнесе;
изучить, какие стандартные метрики существуют для решения распространенных задач;
разобрать, какие типичные метрики существуют в различных отраслях;
понять, почему важно сделать дашборд максимально информативным и легким для восприятия.

Краткое содержание
принятие решений в бизнесе на основе данных;
метрики и воронки, иерархия метрик.

Тема 3
Сбор требований и прототипирование

Цель занятия
изучить техники сбора требований;
научиться задавать правильные вопросы заказчикам;
научиться правильно фиксировать полученную информацию;
изучить назначения прототипа при создании дашборда и научиться его создавать.

Краткое содержание
сбор требований;
кристаллизация требований и создание прототипа.

Тема 4
Итеративная работа с заказчиком на этапе разработки

Цель занятия
научиться применять итеративный гибкий подход при работе с заказчиками;
научиться проводить качественное демо готового дашборда;
научиться запускать пользовательское тестирование;
научиться собирать и фиксировать обратную связь;
изучить, как не делать разработку бесконечной и доводить работу до релиза.

Краткое содержание
итеративная работа с заказчиком на этапе разработки;
демо готовой версии и этап пользовательского тестирования.

Тема 5
Этап релиза и мониторинг востребованности

Цель занятия
научиться подготавливать набор артефактов для успешного релиза;
научиться представлять результаты разработки наиболее эффектно и получать подтверждение о принятии работы;
научиться отслеживать востребованность дашборда;
изучить полезные техники для регулярной поддержки.

Краткое содержание
релиз и постпродакшен;
мониторинг востребованности и получение обратной связи.

Домашние задания

Карта-шаблон для сбора требований
Цель
В этом ДЗ вы заполните карту-шаблон для сбора требований. Эта карта поможет вам приступить к выполнению итогового проекта по курсу.

Модуль 2. СУБД и SQL

Тема 1
Введение в теорию баз данных. Группы операторов в SQL. Выбор данных

Цель занятия
рассмотреть основные виды баз данных и азы реляционной алгебры;
понять особенность организации хранения данных в реляционной БД;
понять отличия между DDL, DML и DCL;
научиться работать с СУБД.

Краткое содержание
введение в теорию БД, обзор видов БД;
виды ключей и нормальных форм реляционной БД;
создание и редактирование таблиц: DDL. DML, DCL;
создание БД.

Тема 2
Типы объединений и соединений таблиц. Порядок выполнения запроса

Цель занятия
рассмотреть, в каком порядке выполняется SQL - запрос;
научиться писать простые запросы к БД;
научиться получать в результатах запроса данные из нескольких таблиц.

Краткое содержание
выбор данных, условия, срезы данных в SQL;
группировка и сортировка данных;
типы объединений и соединений таблиц.

Домашние задания

SQL part 1
Цель
В результате выполнения ДЗ вы создадите тестовую БД, наполните её данными и отработаете простые и сложные выборки данных.

Тема 3
Функции в SQL. Вложенные запросы и временные таблицы

Цель занятия
научиться применять агрегирующие, аналитические, ранжирующие функции в SQL-запросах;
научиться писать вложенные запросы и Common Table Expressions (CTE).

Краткое содержание
агрегирующие функции;
вложенные запросы и временные таблицы;
встроенные функции на SQL.

Тема 4
Объекты базы данных. Оптимизация производительности запросов

Цель занятия
рассмотреть отличие таблицы от представления;
изучить назначение индексов и партиций;
научиться читать план запроса и оценивать его оптимальность;
научиться оптимизировать производительность запроса.

Краткое содержание
объекты базы данных;
таблицы и представления;
индексы и партиции;
план запроса и оптимизация производительности.

Домашние задания

SQL part 2
Цель
В результате выполнения ДЗ вы разработаете базу данных (БД), представляющую собой модель предметной области в соответствии с выбранным вариантом задания.

Модуль 3. Введение в Business Intelligence и визуальный анализ данных

Тема 1
Введение в Business Intelligence и Tableau

Цель занятия
изучить, что такое Business Intelligence;
изучить виды BI-платформ;
изучить особенности продуктов экосистемы Tableau.

Краткое содержание
введение в Business Intelligence;
обзор экосистемы Tableau Desktop / Public.

Тема 2
Основные типы источников данных в Tableau, типы соединений

Цель занятия
научиться использовать Tableau Desktop / Public для подключения к различным СУБД и файлам с данными;
научиться создавать кастомизированные источники данных;
научиться предобрабатывать данные для целей дальнейшего анализа и визуализации.

Краткое содержание
основные типы источников данных в Tableau, соединения; интерфейс Tableau Desktop и основные рабочие понятия.

Тема 3
Визуализация в Tableau. Виды диаграмм и основные сценарии их использования

Цель занятия
научился использовать Tableau Desktop / Public для создания различных типов визуализаций;
понять, какой тип графика выбрать для решения поставленной задачи;
научиться настраивать внешний вид диаграмм и таблиц для наилучшего восприятия пользователями.

Краткое содержание
визуализация: создание различных типов диаграмм и графиков;
выбор правильного вида визуализации;
основные сценарии использования различных типов визуализаций;
визуализации "из коробки" в Tableau и кастомные лайфхаки;
настройка оптимального внешнего вида визуализаций.

Домашние задания

Визуализация в Tableau. Виды диаграмм и основные сценарии их использования
Цель
Это домашнее задание позволит отработать навык подключения к источникам данных в Tableau и созданий визуализаций для решения различных задач.
Мы создадим набор из основных диаграмм и укажем, в каких случаях их лучше всего использовать. В результате у нас получится собственная шпаргалка с основными типами визуализаций.

Тема 4
Вычисляемые поля и функции в Tableau

Цель занятия
научиться использовать встроенные функции Tableau и вычисляемые поля для расширения аналитических возможностей;
научиться обрабатывать числовые и текстовые поля;
научиться использовать функции для определения длительности события.

Краткое содержание
предустановленные и кастомные калькуляции.

Тема 5
Организация данных в Tableau

Цель занятия
научиться группировать и сортировать данные в Tableau, а также организовывать сеты данных;
научиться добавлять интерактив к визуализации через добавление релевантных фильтров;
изучить виды фильтров в Tableau и порядок, в котором они исполняются.

Краткое содержание
организация данных в Tableau;
порядок выполнения операций в Tableau.

Тема 6
Введение в информационный дизайн

Цель занятия
рассмотреть основы информационного дизайна;
изучить основы психологии восприятия визуальной информации и техники, которые позволяют направлять внимание пользователя ожидаемым образом;
изучить основные ошибки при создании дашбордов и способы, как их не допускать или исправлять.

Краткое содержание
введение в информационный дизайн;
как устроено восприятие пользователя;
основные ошибки при создании дашбордов

Тема 7
Проектирование дашборда с учетом User Experience

Цель занятия
научиться проектировать дашборды для достижения различных бизнес-целей;
научиться создавать оптимальный макет для различных устройств, на которых дашборд будет просматриваться;
научиться учитывать будущее взаимодействие пользователей с дашбордом при разработке макета.

Краткое содержание
проектирование дашборда;
макетирование под различные задачи и девайсы;
планирование взаимодействия пользователя с дашбордом.

Домашние задания

Дашборд в Tableau
Цель
В рамках этого домашнего задания мы создадим свой первый дашборд в Tableau, состоящий из нескольких визуализаций. Настроим фильтрацию и добавим немного интерактива будущему взаимодействию между пользователем и интерфейсом. При работе с дизайном мы будем придерживаться чек-листа по основам инфодизайна в Tableau, и на практике оценим, как информация воспринимается лучше.

Модуль 4. Введение в Python

Тема 1
Введение в языки программирования. Знакомство с синтаксисом и основными понятиями Python

Цель занятия
иметь представление о видах языков программирования;
научиться использовать Jupyter Notebook;
научиться подключать базовые библиотеки;
научиться задавать переменные и производить с ними арифметические операции;
научиться выводить результат на экран.

Краткое содержание
введение в языки программирования;
знакомство с синтаксисом;
Jupyter Notebook;
переменные и типы данных;
вывод данных и арифметические операции.

Тема 2
Основы Python. Структуры данных. Операторы, циклы

Цель занятия
иметь представление о списках, кортежах, словарях в Python;
понимать принцип действия и уметь писать циклы For и While для решения задач, требующих итеративной проверки условия;
научится использовать готовые функции и создавать собственные;
иметь представление о Lambda.

Краткое содержание
структуры данных Python;
строки, списки, кортежи и словари;
операторы, циклы, функции.

Тема 3
Библиотеки (Pandas/NumPy), модули и функции

Цель занятия
понимать, что такое массивы в Python;
подключать базовые библиотеки;
иметь общее представление о существовании библиотек, которые могут пригодиться аналитику при решении различных исследовательских задач и задач преобразования данных - NumPy, Pandas;
иметь детальное представление о библиотеке Pandas;
уметь подключать только нужные модули из библиотек;
уметь загружать файлы с локальной машины и из гугл-шитов;
уметь модифицировать датафреймы, применять основные преобразования с фокусом на предобработку и анализ данных;
уметь формировать выгрузку обработанных данных.

Краткое содержание
массивы;
библиотеки NumPy, Pandas;
датафреймы.

Тема 4
Методы визуализации в Python

Цель занятия
иметь общее представление о существовании библиотек, которые могут пригодиться аналитику при решении задач визуализации - Matplotlib, Seaborn, Plotly;
знать и уметь применять Matplotlib для целей визуализации;
настраивать внешний вид графиков и диаграмм.

Краткое содержание
методы визуализации;
библиотеки Matplotlib, Seaborn, Plotly.

Домашние задания

Python
Цель
Вы научитесь визуализировать данные, выполнять действия с табличными данными для вычисления статистических характеристик и анализа данных.

Тема 5
Работа с базами данных с помощью Python

Цель занятия
уметь подключаться к СУБД из ноутбука;
писать SQL-запросы внутри кода Python;
решать исследовательские задачи на базе живого коннекта к данным (например - задачу сравнения двух файлов между собой).

Краткое содержание
SQL-запросы внутри кода Python.

Модуль 5. Специальные методы и направления в дата-аналитике

Тема 1
Продуктовая аналитика

Цель занятия
изучить, кто такой Аналитик продукта, какими специальными навыками и приемами в аналитике он должен обладать;
рассмотреть, с какими задачами и типами анализа Аналитик продукта сталкивается в работе чаще всего.

Краткое содержание
продуктовая аналитика;
юнит-экономика;
A/B-тесты;
когортный анализ.

Тема 2
BI-аналитика. Дата журналистика и дата-сторителлинг

Цель занятия
изучить, кто такой BI-аналитик (BI-разработчик);
рассмотреть, какими специальными навыками и приемами в аналитике он должен обладать;
изучить, с какими задачами и типами анализа BI-аналитик (BI-разработчик) сталкивается в работе чаще всего;
рассмотреть дата-журналистику как отдельный подтип в области анализа и визуализации данных;
рассмотреть, какие умения позволят попробовать себя в роли дата-журналиста.

Краткое содержание
BI-аналитика;
дата-журналистика.

Тема 3
Дата-арт и дата-дизайн

Цель занятия
изучить, специфику дата-арта и дата-дизайна;
рассмотреть из отличия от дата-анализа.

Краткое содержание
дата-арт;
дата-дизайн.

Модуль 6. Основы статистики

Тема 1
Основы статистики. Генеральная совокупность и выборка. Типы данных

Цель занятия
иметь представление о типах данных - категориальные (номинальные, порядковые), числовые (интервальные, нормативные);
отличать генеральную совокупность от выборки, вычислять базовые статистики с помощью Python.

Краткое содержание
генеральная совокупность и выборка, уровни измерений;
медиана, мода, среднее арифметическое;
типы данных.

Тема 2
Нормальное распределение. Стандартное отклонение. Доверительные интервалы

Цель занятия
знать, что такое нормальное распределение данных;
уметь визуализировать распределение данных и разбираться в его особенностях - нормальное распределение, скошенное распределение;
вычислять стандартное отклонение;
знать, в чем состоит центральная предельная теорема;
строить доверительные интервалы;
определять стандартную ошибку.

Краткое содержание
нормальное распределение;
стандартное отклонение;
центральная предельная теорема;
доверительные интервалы и стандартная ошибка.

Тема 3
Статистические гипотезы. A/B-тесты

Цель занятия
ставить и проверять статистические гипотезы;
иметь представление о методах сравнения данных;
иметь представление об A/B-тестах.

Краткое содержание
уровень значимости, статистические гипотезы;
методы сравнения данных;
сравнение номинальных данных;
методы сравнения средних;
A/B-тесты и их применение на практике.

Домашние задания

A/B-тесты
Цель
В этом ДЗ вы научитесь обоснованно выбирать критерий и оценивать результаты A/B-теста.

Модуль 7. Предобработка данных, исследовательский и статистический методы анализа данных

Тема 1
Исследовательский анализ данных. Предобработка и очистка данных, работа с пропусками и дубликатами

Цель занятия
научиться проводить исследовательский анализ данных;
научиться обнаруживать и обрабатывать выбросы, пропуски и дубликаты;
научиться категоризировать данные;
изучить типы данных и преобразовывать одни типы в другие;
обработка даты и времени.

Краткое содержание
изучение срезов данных;
работа с пропусками и дубликатами;
категоризация данных;
преобразование типов данных;
обработка даты и времени.

Тема 2
Нормирование данных. Анализ временных рядов. Корреляционный анализ

Цель занятия
научиться нормировать данных для целей дальнейшего анализа;
научиться правильно анализировать временные ряды (последовательность значений какого-либо показателя за несколько периодов времени);
иметь представление о корреляционном анализе;
уметь вычислять коэффициент корреляции различными методами и интерпретировать результат.

Краткое содержание
нормирование данных;
анализ временных рядов;
корреляционный анализ.

Тема 3
Взаимосвязь данных. Линейная регрессия

Цель занятия
оценивать взаимосвязь и взаимовлияние данных;
иметь представление о методе линейной регрессии как способе оценивать взаимосвязь данных.

Краткое содержание
взаимосвязь данных;
линейная регрессия.

Домашние задания

Обработка данных и линейная регрессия
Цель
В этом ДЗ вы научитесь выполнять все этапы предобработки данных и делать предсказания при помощи модели линейной регрессии.

Модуль 8. Проектная работа

Тема 1
Выбор темы и организация проектной работы

Цель 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

Проектная работа
Цель
Итоговая проектная работа - полный набор артефактов, описывающих требования заказчика к отчету, сам отчет, исполненный в виде интерактивного дашборда в Табло, а также те выводы и наблюдения, которые получилось с его помощью сделать.

Что нужно подготовить: шаблон, описывающий требования заказчика к отчету, сам отчет, исполненный в виде интерактивного дашборда в Табло (ссылка на Tableau Public), а также краткая презентация с обзором проекта, наблюдениями и рекомендациями по результатам исследования, представленная в ходе защиты.

Тема 2
Консультация по проектам и домашним заданиям

Цель 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.

Тема 3
Защита проектов

Цель занятия
защитить проект и получить рекомендации экспертов;
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
презентация проектов перед комиссией;
вопросы и комментарии по проектам;
организационные вопросы;
рынок вакансий по направлению;
проведение собеседований;
статистика курса и вопросы по курсу.