Полная программа

Аналитик данных

Длительность курса: 126 часов

Модуль 1. СУБД и SQL

Тема 1
Введение в профессию «Аналитик данных»

Цель занятия

представлять будущий маршрут курса;
понимать содержание профессии «аналитик данных»;
разбираться в предметных областях и специализациях, которые можно выделить в макро области «аналитика данных»;
понимать особенности каждого карьерного направления с точки зрения ключевых навыков и компетенций и типичных задач.

Краткое содержание

описание профессии «аналитик данных».

Тема 2
Введение в теорию баз данных. Группы операторов в SQL. Выбор данных

Цель занятия

понять концепцию «базы данных» и особенности реляционных баз данных;
оценить качество организации хранения данных в реляционной БД;
создать схему реляционной базы данных;
самостоятельно 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 и DCL;
создание БД.

Тема 3
Типы объединений и соединений таблиц. Порядок выполнения запроса // ДЗ

Цель занятия

написать простой запрос на выборку данных из таблицы;
написать аналитический запрос с преобразованием данных;
понять концепцию соединения множества 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
создать тестовую БД и наполнить её данными, чтобы отработать простые и сложные выборки данных.

Тема 4
Функции в SQL. Вложенные запросы и временные таблицы

Цель занятия

написать аналитический запрос с вложенностью;
понять концепцию временного хранения данных;
написать сложный аналитический запрос с оконной функцией.

Краткое содержание

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

Тема 5
Объекты базы данных. Оптимизация производительности запросов // ДЗ

Цель занятия

создать представление в БД;
написать собственную функцию или процедуру;
использовать оптимизацию кода для работы с БД.

Краткое содержание

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

Домашние задания

SQL part 2
Цель
разработать базу данных (БД), которая представляет модель предметной области в соответствии с выбранным вариантом задания.

Тема 6
Q&A-сессия

Модуль 2. Принятие решений в бизнесе на основе данных

Тема 1
Принятие решений в бизнесе на основе данных

Цель занятия

понять концепцию «принятия решений на основе данных»;
обосновать необходимость этого подхода в своей предметной области;
оценивать полученную информацию «глазами заказчика».

Краткое содержание

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

Тема 2
Сбор требований и прототипирование

Цель занятия

инициализировать проект;
знать техники сбора требований к проекту;
задавать правильные вопросы заказчикам и фиксировать полученную информацию;
знать назначение прототипа при создании дашборда и уметь его создавать.

Краткое содержание

сбор требований;
кристаллизация требований и создание прототипа.

Тема 3
Итеративная работа с заказчиком на этапе разработки

Цель занятия

выбрать наилучший метод управления проектом;
формировать функциональную спецификацию;
проводить качественное демо готового продукта;
проводить пользовательское тестирование;
собирать и фиксировать обратную связь.

Краткое содержание

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

Тема 4
Этап релиза и мониторинг востребованности // ДЗ

Цель занятия

готовить набор артефактов для успешного релиза вашего продукта;
осуществлять контроль и поддержку выпущенного продукта;
представлять результаты проекта заказчику и получать подтверждение о его завершении.

Краткое содержание

релиз и постпродакшен;
мониторинг востребованности и получение обратной связи.

Домашние задания

Карта-шаблон для сбора требований
Цель
заполнить карту-шаблон для сбора требований, которая поможет вам приступить к выполнению итогового проекта по курсу.

Модуль 3. Введение в Business Intelligence и визуальный анализ данных

Тема 1
Введение в визуализацию данных и Business Intelligence

Цель занятия

знать, почему данные важно визуализировать;
понимать, что такое визуальный анализ и зачем он нужен;
понимать, что такое Business Intelligence, и зачем это аналитику;
знать виды BI-платформ и их применимость в условиях импортозамещения.

Краткое содержание

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

Тема 2
Введение в Tableau: установка, настройка, основные типы подключений

Тема 3
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 дальнейшего анализа и визуализации.

Краткое содержание

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

Тема 4
Визуализация в Tableau. Виды диаграмм и основные сценарии их использования // ДЗ

Цель занятия

использовать Tableau Desktop/Public для создания различных типов визуализаций;
знать, какой тип графика выбрать для решения поставленной задачи;
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
создать набор из основных диаграмм и указать, в каких ситуациях лучше всего их использовать (в результате у вас получится собственная шпаргалка с основными типами визуализаций).

Тема 5
Вычисляемые поля и функции в Tableau

Цель занятия

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

Краткое содержание

предустановленные и кастомные калькуляции.

Тема 6
Организация данных в Tableau

Цель занятия

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

Краткое содержание

организация данных в Tableau;
порядок выполнения операций в Tableau.

Тема 7
Введение в информационный дизайн

Цель занятия

иметь представление об информационном дизайне;
знать основы психологии восприятия визуальной информации и техники, которые позволяют направлять внимание пользователя ожидаемым образом;
знать 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 8
Проектирование дашборда с учетом User Experience // ДЗ

Цель занятия

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

Краткое содержание

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

Домашние задания

Дашборд в Tableau
Цель
создать дашборд в Tableau, состоящий из нескольких визуализаций.

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;
использовать Jupyter Notebook;
подключать базовые библиотеки;
задавать переменные и производить с ними арифметические операции;
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;
писать циклы For и While для решения задач, требующих итеративной проверки условия;
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 // ДЗ

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
научиться визуализировать данные и выполнять действия с табличными данными для вычисления статистических характеристик и анализа данных.

Тема 5
Работа с базами данных с помощью Python

Цель занятия

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.

Тема 6
Парсинг данных из HTML-страниц

Модуль 5. Специальные методы и направления в дата-аналитике

Тема 1
Продуктовая аналитика

Цель занятия

отличать направления работы для аналитиков;
знать, что такое продуктовая аналитика;
знать какими специальными навыками обладает продуктовый аналитик;
отличать задачи продуктового аналитика от всех остальных.

Краткое содержание

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

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

Цель занятия

знать, как устроен Business Intelligence;
знать, что такое BI-аналитика;
знать какими специальными навыками обладает BI-аналитик;
определить задачи дата-журналиста, и как они отличаются от всех остальных.

Краткое содержание

BI-аналитика;
дата-журналистика;
дата-сторителлинг.

Тема 3
Дата-арт и дата-дизайн

Цель занятия

понимать разницу между дата-артом и дата-дизайном;
расставлять приоритеты между «понятно» и «красиво»;
в случае необходимости обращаться к дизайн гайдлайнам.

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;
визуализировать распределение данных и разбираться в его особенностях: нормальное распределение, скошенное распределение;
вычислять стандартное отклонение;
знать, в чём 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 // ДЗ

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

проводить исследовательский анализ данных;
обнаруживать и обрабатывать выбросы, пропуски и дубликаты;
категоризировать данные;
различать типы данных и преобразовывать одни типы в другие.

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

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

Краткое содержание

нормирование данных;
анализ временных рядов;
корреляционный анализ.

Тема 3
Взаимосвязь данных. Линейная регрессия // ДЗ

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
научиться выполнять все этапы предварительной обработки данных и делать предсказания при помощи модели линейной регрессии.

Модуль 8. Введение в машинное обучение

Тема 1
Основы машинного обучения

Домашние задания

Основы машинного обучения

Тема 2
Задачи регрессии

Домашние задания

Задачи регрессии

Тема 3
Задачи классификации

Домашние задания

Задачи классификации

Тема 4
Обучение без учителя

Домашние задания

Обучение без учителя

Тема 5
Как ИИ помогает автоматизировать рутину аналитика

Домашние задания

Как ИИ может автоматизировать рутину аналитика

Тема 6
Q&A-сессия

Модуль 9. 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.