

BI-аналитик

Научитесь превращать данные в наглядные графики и отчёты, которые помогают бизнес



Модуль 1 Основы анализа данных с помощью SQL и BI

Спринт 1 Введение в аналитику. Аналитический отчёт в Google Sheets

2 недели

Узнаете, кто такой аналитик данных и какие задачи он решает. Познакомитесь с пайплайном работы аналитика. Создадите свой первый аналитический отчёт в Google Sheets.

Тема 1. Использование данных в бизнесе

- Структурированные и неструктурированные данные
- Роль данных в бизнесе

Тема 2. Процесс анализа данных и задачи аналитика

- Задачи аналитика данных
- Пайплайн работы аналитика
- Специализации в аналитике

Тема 3. Excel как инструмент аналитика. Основы Google Sheets

- Табличные редакторы, начало работы в Google таблицах
- Константы и формулы

Тема 4. Предобработка данных в Google Sheets

- Типы данных: числовые, текстовые
- Форматирование данных
- Очитска данных
- Использование панели автоподсчёта

Тема 5. Использование формул и функций

- Формулы и функции
- Обзор базовых функций, синтаксис, использование
- Математические функции (SUM, COUNT, ROUND, MIN, MAX, AVERAGE)
- Логические функции AND, OR, NOT, функции с условиями (IF, SUMIF и др)
- Абсолютные и относительные ссылки
- Функции даты и времени
- Использование VLOOKUP (ВПР)
- Сводные таблицы

Тема 6. Презентация данных

- Построение простых визуализаций
- Как поделиться отчётом

Финальное задание спринта

Создание аналитического отчёта в Google Sheets.

Пайплайн работы аналитика

Google Sheets

Формулы

Функции

Отчёты

Спринт 2 Основы SQL. Извлечение данных для анализа

2 недели

Узнаете, как могут храниться данные и познакомитесь с языком запросов SQL для работы с базами данных. Напишете первые запросы на SQL и научитесь извлекать данные под задачу с фильтрацией, группировкой, сортировкой.

Тема 1. Работа с базами данных. СУБД

- Сбор данных в пайплайне аналитика
- База данных, ER-диаграмма, модели данных
- Реляционные БД
- Первичный ключ
- СУБД
- SQL, PostgreSQL
- SELECT-запросы

Тема 2. Типы данных и их преобразования

- Основные типы данных (числовые, символьные, логические, для работы с датой и временем)
- Изменение типа данных (CAST, ::)
- Округление данных (ROUND, CEIL и FLOOR)
- Арифметические операции

Тема 3. Фильтрация данных и агрегация

- Фильтрация по условиям (WHERE), логические операции AND, OR, NOT, IN (), NOT IN ()
- Составные условия
- Функции для агрегации (COUNT(), SUM(), MIN(), MAX(), AVG())

Тема 4. Группировка и сортировка данных

- GROUP BY, агрегация значений при группировке
- Группировка по нескольким полям
- Фильтрация после группировки
- Отличия WHERE и HAVING
- Псевдонимы
- Сортировка данных ORDER BY

Финальное задание спринта

Решение задач на извлечение данных с помощью SQL-запросов в тренажёре.

БД и СУБД

SQL

PostgreSQL

Типы данных

Группировка данных

Сортировка данных

Спринт 3 SQL. Обработка данных

2 недели

Продолжите знакомиться с инструментами SQL и научитесь обрабатывать данные для анализа: устранять дубликаты и работать с пропущенными значениями. Сможете извлекать данные из нескольких таблиц, используя JOIN-ы, использовать подзапросы и CTE.

Тема 1. Работа с пропущенными значениями и дубликатами

- Что такое пропуски и дубликаты
- План работы
- Концепция пропуска (NULL)
- Заполнение пропусков (COALESCE)
- Удаление явных дубликатов

Тема 2. Присоединение таблиц

- Хранение таблиц в БД
- Нормализация
- Связи между таблицами (один к одному, один ко многим, многие ко многим)
- Разные типы присоединения таблиц (INNER JOIN, LEFT JOIN и RIGHT JOIN, FULL OUTER JOIN)

Тема 3. Операции множеств и подзапросы

- Объединение множеств
- Пересечение и вычитание
- Подзапросы (в секции WHERE и в секции FROM)
- CTE (обобщённые табличные выражения)

Тема 4. Категоризация значений. Создание новых столбцов

- Операции со столбцами (вычитание, сложение, усреднение)
- Категоризация значений (CASE WHEN THEN END)
- Обработка неявных дубликатов

Тема 5. Работа с датой и временем

- Типы данных для даты и времени (TIMESTAMP, DATE, INTERVAL).
- Функции EXTRACT() и DATE_TRUNC()
- Фильтрация по дате и работа с интервалами

Финальное задание спринта

Решение задач на извлечение и обработку данных с помощью SQL-запросов в тренажёре

Пропуски

SQL

Дубликаты

Подзапросы

CTE

Присоединения таблиц (JOIN)

Спринт 4 SQL. Анализ данных и решение ad-hoc задач

2 недели

Научитесь применять продвинутые инструменты SQL (оконные функции) для решения ad-hoc задач аналитика разной сложности. Познакомитесь с необходимыми для решения таких задач понятиями описательной статистики.

Тема 1. Знакомство с базой данных

- Как изучать БД самостоятельно
- Выводы об устройстве БД

Тема 2. Оконные функции. Агрегирующие функции

- Назначение оконных функций, их классификация
- Агрегирующие оконные функции (SUM(), COUNT(), AVG(), MIN(), MAX())
- Предложения PARTITION BY и ORDER BY

Тема 3. Оконные функции. Ранжирующие функции

- Назначение функций ROW_NUMBER(), RANK() и DENSE_RANK(), NTILE()
- Особенности ранжирующих оконных функций

Тема 4. Оконные функции смещения

- Назначение и особенности функций LEAD(), LAG(), FIRST_VALUE(), и LAST_VALUE()

Тема 5. Описательная статистика. Аналитические функции

- Категориальные и количественные переменные
- Меры центральной тенденции (среднее значение, мода, медиана, различие среднего и медианы, перцентили)
- Меры разброса
- Аналитические функции PERCENTILE_DISC(), PERCENTILE_CONT(), оператор WITHIN, функция STDDEV()
- Назначение и использование

Кейс-проект с ревью

Решение ad-hoc аналитических запросов различной сложности на SQL в рамках задачи исследовательского анализа данных.

Решение ad-hoc задач

Декомпозиция

SQL

Агрегирующие оконные функции

Ранжирующие оконные функции

Оконные функции смещения

Аналитические оконные функции

Мода

Медиана

Среднее

Перцентиль

Размах

Разберётесь с основами визуализации данных в BI-инструменте DataLens. Научитесь подбирать тип визуализации под задачу. Изучите основы создания и настройки дашбордов.

Тема 1. Визуализация в работе аналитика. Знакомство с DataLens

- Визуализация как задачи аналитика
- BI-инструменты. Порядок работы в BI-инструментах
- Интерфейс DataLens. Подключения в DataLens
- Данные (credentials) для подключения к базе данных. Датасеты в DataLens
- Типы данных в DataLens

Тема 2. Основы визуализации. Чарты

- Виды визуализаций
- Типы графиков
- Линейная диаграмма, столбчатая диаграмма, линейчатая диаграмма, кольцевая диаграмма, круговая диаграмма, накопительная диаграмма с областями, таблица, сводная таблица, индикатор
- Элементы визуализации
- Создание чартов в DataLens
- Оформление графиков
- Графики для визуализации сравнения, соотношения части и целого, отображения изменений во времени

Тема 3. Вычисляемые поля

- Вычисляемые поля на уровне датасета и на уровне чарта
- Формулы
- Агрегирующие функции: MIN(), MAX(), AVG(), AVG_IF(), COUNT(), COUNT_IF(), SUM(), SUM_IF(), COUNTD() и другие
- Логические функции: IF(), CASE()
- Функции для работы со строками: REPLACE(), CONCAT(), (I)STARTSWITH / (I)ENDSWITH, (I)CONTAINS и другие
- Функции для работы с датами: DATEADD(), DATETRUNC(), DATEPART(), YEAR(), MONTH() и другие

Тема 4. Дашборды

- Назначение дашбордов
- Прототипирование дашбордов
- Виджеты
- Чарты
- Добавление чартов на дашборд
- Селекторы
- Связи: входящие и исходящие
- Настройка селекторов
- Тексты
- Заголовки
- Композиция дашборда

Тема 5. Параметры

- Использование параметров для дашборда
- Использование параметров для чарта
- Ситуации, в которых используются параметры
- Специальные параметры

Тема 6. Интерпретация данных из дашбордов

- Задачи визуализации и её аудитория
- Работа с готовым дашбордом
- Выбор чарта и настройка фильтрации
- Чтение графиков на дашборде

Кейс-проект с ревью

Предварительный анализ данных с помощью SQL, создание дашборда по ТЗ.

SQL

DataLens

BI-инструменты

Чарт

Визуализация данных

Дашборд

1 неделя

Итоговый проект модуля

Познакомьтесь с БД через SQL и создадите дашборд с использованием связки SQL и BI.

Создание и настройка дашборда, содержащего интерактивные графики и другие визуализации, для автоматизации отчётности.

Каникулы

Модуль 2 Анализ данных с помощью Python

Спринт 6 Основы Python

2 недели

Начнёте знакомство с языком программирования Python. Изучите основы синтаксиса, необходимые для последующего написания кода.

Тема 1. Знакомство с Python

- Язык программирования Python
- Особенности и назначение Python
- Переменные
- Объекты в Python
- Тип данных
- Преобразования типов данных
- Ошибки типов
- Математические операции в Python.

Тема 2. Строки. Списки

- Строки в Python
- Создание строк
- Операции со строками
- Длина строк
- Индексы в строках
- Прямые и обратные индексы
- Срезы строк
- Форматирование строк
- f-строки
- Методы строк
- Списки в Python
- Создание списков
- Длина списков
- Операции со списками
- Индексы в списках
- Срезы в списках
- Методы списков
- Сортировка списков
- Кортежи

Тема 3. Логические выражения. Условный оператор

- Логические выражения в Python
- Логические операции в Python
- Логические значения
- Условный оператор в Python
- Секция elif
- Секция else
- Предикаты

Тема 4. Циклы. Списковые включения

- Циклы в программировании
- Цикл for в Python
- Принцип работы циклов
- Использование циклов
- Функция range()
- Функция enumerate()
- List comprehension
- Использование list comprehension
- Преобразования списков.

Тема 5. Вложенные списки

- Списки, вложенные в списки
- Циклы по вложенным спискам
- Преобразования вложенных списков
- Сортировка вложенных списков

Тема 6. Функции

- Функции
- Создание функций в Python
- Переиспользование кода
- Параметры и аргументы функций
- Значения аргументов по умолчанию
- Позиционные и именованные аргументы
- Возвращение результата из функций
- Возвращение нескольких значений из функции

Тема 7. Множества. Словари

- Множества
- Создание множеств
- Удаление повторяющихся значений
- Пересечение множеств
- Объединение множеств
- Разность множеств
- Преобразование множеств в списки
- Словари
- Создание словарей
- Ключи и значения
- Что может быть ключами и значениями словаря
- Поиск значения по ключу
- Добавления элементов в словари
- Словари и списки
- Словари, содержащие другие словари

Финальное задание спринта

Решение проверочных заданий с помощью Python.

Python

Переменные

Типы данных

Строки

Списки

Циклы

Условный оператор

Функции

Множества

Словари

Спринт 7 Python. Предобработка данных

2 недели

Начнёте работу с библиотекой Pandas. Научитесь предобрабатывать данные с помощью Python: очищать данные от выбросов, пропусков и дубликатов и преобразовывать разные форматы данных.

Тема 1. Основы библиотеки pandas. Обзор данных

- Библиотека pandas
- Датафрейм как структура данных pandas
- Свойства датафреймов
- Создание датафреймов
- CSV и TSV файлы
- Считывание данных из CSV файлов в датафрейм
- Среда Jupyter Notebook

Тема 2. Типы данных. Работа с датой и временем

- Типы данных pandas
- Тип данных object
- Преобразования строковых и числовых значений
- Преобразование значений в тип даты и времени
- Методы astype(), to_numeric(), to_datetime()
- Ошибки при преобразовании типов в pandas и способы их решения

Тема 3. Индексация в датафреймах

- Индексы в pandas
- Выбор нужных строк и столбцов датафрейма
- Атрибут loc[]
- Логическая индексация

Тема 4. Работа с пропущенными значениями

- Пропущенные значения в данных
- Виды пропусков: MAR, MNAR, MCAR
- Нахождение пропусков в датафрейме
- Замена пропусков в датафрейме
- Удаление пропусков
- Подсчёт количества и доли пропусков в данных
- Выбор стратегии обработки пропусков

Тема 5. Обработка дубликатов

- Дубликаты в данных
- Виды дубликатов: явные и неявные дубликаты
- Влияние дубликатов на дальнейший анализ и визуализации
- Нахождение дубликатов в датафрейме
- Удаление дубликатов
- Подсчёт количества дубликатов

Тема 6. Категоризация данных

- Категоризация данных
- Метод apply()
- Параметр axis
- Применение функций к строкам и столбцам датафрейма
- Объединение числовых значений в диапазоны
- Извлечение информации из даты и времени в отдельный столбец

Кейс-проект с ревью

Подготовка с помощью Python "сырых" данных для последующего анализа.

Python

Pandas

Предобработка данных

Обработка пропусков

Обработка дубликатов

Категоризация данных

Спринт 8

2 недели

Исследовательский анализ данных и визуализация с помощью Python

Научитесь использовать Python для исследования и визуализации данных. Разберётесь с основами описательной статистики на примерах.

Тема 1. Присоединения датафреймов

- Присоединения датафреймов
- Виды присоединений
- Методы `join()` и `merge()`
- Использование индексов и столбцов для присоединения

Тема 2. Описательная статистика

- Выборка и генеральная совокупность
- Особенности подсчёта стандартного отклонения и среднего значения в `pandas`
- Степени свободы
- Распределение данных
- Столбчатые диаграммы
- Гистограммы
- Диаграммы размаха

Тема 3. Сводные таблицы

- Отличие сводных таблиц от группировки
- Сводные таблицы
- Составление сводных таблиц
- Анализ среднего, медианы и других показателей с помощью сводных таблиц
- Многоуровневые сводные таблицы
- Метод `cut()`
- Методы `stack()` и `unstack()`
- Визуализация сводной таблицы с помощью `heatmap`
- Библиотека `seaborn`

Тема 4. Анализ данных. Базовая визуализация данных

- Базовая проверка данных
- Взаимосвязь данных
- Корреляция данных для численных значений
- Коэффициент корреляции Пирсона
- Универсальный коэффициент корреляции
- Диаграммы рассеяния
- Матрица диаграмм рассеяния
- Линейный график
- Настройка линейного графика
- Круговая диаграмма

Тема 5. Расширенная визуализация данных

- Размещение нескольких подграфиков на одном графике
- Продвинутая настройка графиков
- Интерактивные визуализации
- Библиотека `plotly`

Тема 6. Пример исследовательского анализа данных

- Разбор исследовательского анализа данных в среде `Jupyter Notebook`

Кейс-проект с ревью

Исследовательский анализ данных для решения бизнес-кейс и подготовка отчёта по исследованию

Python

Pandas

Matplotlib

Seaborn

Jupyter notebook

1 неделя

Итоговый проект модуля

Под запрос бизнеса выполните исследование данных с последующей визуализацией с помощью инструментов Python. Предоставьте рекомендации бизнесу по итогам исследования.

Исследование данных с последующей визуализацией с помощью инструментов Python в Jupyter notebook, с рекомендациями бизнесу по итогам исследования.

Каникулы

Модуль 3 Продвинутый анализ данных для бизнеса

Спринт 9 Расчёт и визуализация бизнес-метрик и показателей

3 недели

Погрузитесь глубже в контекст бизнеса и продукта. Рассчитаете и визуализируете важные для бизнеса показатели с помощью SQL и разберётесь с основами когортного анализа.

Тема 1. Что такое бизнес-метрики

- Метрики бизнеса
- Цель подсчёта и использования бизнес-метрик
- Как бизнес использует аналитические показатели

Тема 2. Основные бизнес-метрики

- Уникальные пользователи
- DAU
- Конверсия
- Средний чек
- CTR
- Retention rate
- Churn rate
- LTV. ROI. CAC
- Расчёт метрик бизнеса
- Визуализация бизнес-метрик

Тема 3. Когортный анализ для расчёта метрик

- Когортный анализ
- Цели использования когортного анализа
- Выбор признака начала когорты
- Момент наблюдения и горизонт анализа
- Расчёт метрик с использованием когортного анализа

Тема 4. Выбор ключевых метрик для кейса

- Использование аналитического мышления
- Выбор ключевых метрик для кейса

Тема 5. Продуктовые воронки

- Продуктовые воронки
- Для чего нужны воронки
- Воронкообразные метрики
- Анализ продуктовой воронки

Кейс-проект с ревью

Решение ad-hoc задач на расчёт и визуализацию бизнес-метрик.

SQL

DataLens

Спринт 10 **Формулировка и проверка гипотез.**

3 недели

Статистический анализ данных

Изучите основы статистического анализа данных и сможете применять статистику для проверки продуктовых гипотез.

Тема 1. Основы теории вероятностей

- Эксперимент
- Исходы
- События
- Вероятность
- Классическое и геометрическое определение вероятности
- Достоверные, случайные и невозможные события
- Совместные и несовместные события
- Умножение и сложение событий
- Вероятностное пространство
- Закон больших чисел.

Тема 2. Случайные величины

- Случайные величины
- Дискретная случайная величина
- Функция распределения дискретной случайной величины
- Математическое ожидание дискретной случайной величины
- Дисперсия математической случайной величины

Тема 3. Распределения

- Распределения
- Эксперимент Бернулли
- Биномиальное распределение
- Нормальное распределение
- Стандартное нормальное распределение
- Распределение Стьюдента
- Распределение Пуассона
- Экспоненциальное распределение.

Тема 4. Проверка гипотез. Т-тест Стьюдента

- Генеральная совокупность
- Случайные выборки
- Репрезентативность выборок
- Оценка параметров генеральной совокупности
- Проверка гипотез
- Статистические тесты
- Ошибки первого и второго рода
- Двусторонние и односторонние гипотезы
- Т-тест Стьюдента
- Т-тест Уэлча

Тема 5. Множественная проверка гипотез

- Множественная проверка гипотез
- Ситуации, в которых возникает множественная проверка гипотез
- Проблемы при множественной проверке гипотез. Поправки на множественную проверку гипотез
- Поправка Бонферрони
- Проблемы, связанные с поправками на множественную проверку гипотез
- Специальные тесты для проверки нескольких групп

Кейс-проект с ревью

Проведение мини-исследования данных и проверка нескольких гипотез с помощью статтестов

Python

Pandas

Scipy

Нулевая гипотеза

Statsmodels

Альтернативная гипотеза

Распределения метрик

Ошибка первого рода

Ошибка второго рода

Статистический тест

Множественная проверка гипотез

T-тест

Тест Манна-Уитни

Z-тест

Биномиальный тест

Спринт 11 Анализ результатов A/B тестирования с помощью Python

2 недели

Познакомьтесь с дорожной картой AB-тестирования и сможете самостоятельно анализировать результаты AB-теста и формулировать выводы для бизнеса.

Тема 1. Что такое A/B-тесты и зачем они нужны

- Продуктовые гипотезы
- Генерация продуктовых гипотез
- Оценка гипотез
- Что такое A/B-тесты
- Зачем и когда нужны A/B-тесты
- A/A-тесты
- Общий план подготовки A/B-теста
- Этапы A/B-теста

Тема 2. Знакомство с этапом «Подготовка A/B теста»

- Основные параметры A/B теста
- Выборка и длительность проведения теста
- Размер выборок в A/B-тесте
- Оценка длительности AB-теста

Тема 3. Проверка результатов A/B-теста

- Выбор статистического теста для проверки результатов A/B-теста
- Непараметрические тесты
- Тест Манна-Уитни
- Z-тест пропорций
- Интерпретация результатов для бизнеса

Тема 4. Анализ результатов A/B-теста: примеры

- Кейсы с анализом результатов проведённых A/B-тестов и их интерпретация

Кейс-проект с ревью

Анализ результатов проведенного A/B-теста с помощью статтестов.

SQL

Python

Pandas

Scipy

Statsmodels

Jupyter Notebook

Множественная проверка гипотез

T-тест

Тест Манна-Уитни

Z-тест

Биномиальный тест

1 неделя

Итоговый проект модуля

Чтобы помочь бизнесу вырастить целевые метрики, проведёте исследование результатов AB-теста и сформулируете выводы о результатах эксперимента для бизнеса.

Исследование результатов AB-теста с поледующей формулировкой выводов о результатах эксперимента и рекомендаций для бизнеса.

Каникулы

Модуль 4 Понимание бизнес-требований, прототипирование и начало визуализации

Спринт 12

2 недели

Анализ бизнес-требований и планирование отчетности

В этом модуле студенты научатся проводить интервью с заказчиком, выявлять и документировать бизнес-требования, создавать эскизы дашбордов. Будет уделено внимание приоритизации требований и выделению критически важных для бизнеса элементов.

Тема 1. Знакомство с BI-аналитикой. Введение в BI-аналитику

- Применение BI в бизнесе
- Основные компоненты BI-систем
- Примеры успешного использования BI

Тема 2. Сравнение инструментов BI и их реализации

- Обзор популярных BI-инструментов
- Ключевые функции и возможности каждого инструмента
- Преимущества и недостатки различных BI-решений

Тема 3. Проведение интервью с заказчиком для сбора требований

- Подготовка к интервью с заказчиком
- Методики проведения интервью
- Вопросы для выявления бизнес-требований
- Анализ и документирование собранной информации

Тема 4. Принципы и методы документирования бизнес-требований

- Основные принципы документирования требований
- Форматы и шаблоны документов
- Инструменты для документирования требований
- Управление изменениями требований

Тема 5. Типы визуализаций

- Обзор основных типов визуализаций данных (графики, диаграммы, карты)
- Когда и какие типы визуализаций использовать
- Примеры эффективных визуализаций
- Ошибки при создании визуализаций и как их избегать

Тема 6. Принципы дизайна дашбордов

- Основы дизайна дашбордов
- Важность юзабилити и UX
- Принципы размещения и группировки элементов
- Использование цветов и шрифтов

Тема 7. Создание эскизов дашбордов

- Инструменты для создания эскизов дашбордов
- Процесс разработки эскиза: от идеи до финального макета
- Примеры удачных эскизов
- Проверка и утверждение эскиза с заказчиком

Проект модуля с автоматической проверкой

Выберите BI-инструмент для решения конкретной бизнес-задачи. Заполните бриф заказчика на основе готового дашборда, выберите подходящий дизайн и визуализацию для представления данных.

Студенту необходимо выбрать инструмент под конкретные задачи, заполнить образец брифа заказчика на основании готового дашборда, а также выбрать дизайн и визуализацию, подходящую под поставленные задачи.

Miro

DataLens

Визуализация данных

Проведение интервью

Сбор требований

Прототипирование дашбордов

Документирование требований

Модуль 5 Визуализация данных и подготовка отчетности

Спринт 13 Проектирование дизайна дашборда и документирование его логики

2 недели

Студенты изучат принципы дизайна дашбордов, работу с различными типами чартов и инструментами визуализации данных, освоят создание интерактивных и параметризованных дашбордов, документирование и закрепление результатов работы.

Тема 1. Дополнительные чарты в DataLens (карты-полигоны, Calendar Heatmap)

- Использование карт-полигонов для визуализации данных
- Создание и настройка Calendar Heatmap
- Примеры использования дополнительных чартов в DataLens

Тема 2. Интерактивность и параметризация дашбордов

- Добавление интерактивных элементов на дашборды
- Использование фильтров и параметров для персонализации данных
- Примеры интерактивных дашбордов

Тема 3. Форматирование текста на дашбордах с помощью Markdown

- Основы синтаксиса Markdown
- Примеры форматирования текста с помощью Markdown
- Добавление заголовков, списков и ссылок на дашбордах

Тема 4. Основы администрирования в DataLens

- Управление пользователями и ролями
- Настройка доступа к данным и дашбордам
- Мониторинг и оптимизация производительности DataLens

Тема 5. Документирование и закрепление результатов созданного дашборда

- Создание документации для дашбордов
- Описание целей и метрик дашборда
- Сбор обратной связи и корректировка дашборда на основе результатов

Кейс-проект спринта 2 с небольшой проверкой ревьюера

Студент создает эскиз дашборда в Miro на основании предоставленного брифа заказчика, сборка макета дашборда в Datalens на основании предоставленных шаблонов графиков (мокапов) и предоставленного интервью заказчика.

DataLens

Superset

Markdown

Miro

Визуализация данных

Администрирование BI-инструментов

Документирование результатов

Интерактивные дашборды

Спринт 14 Завершение работы над дашбордом и документацией. SUPERSET

2 недели

Студенты освоят работу с Superset, включая создание чартов, датасетов и дашбордов, интерактивных и параметризованных дашбордов, а также документирование и администрирование.

Тема 1. Введение в Superset: чарты, датасеты и дашборды

- Основные возможности Superset
- Создание и настройка чартов
- Работа с датасетами
- Создание и управление дашбордами

Тема 2. Интерактивность и параметризация дашбордов в Superset

- Добавление интерактивных элементов на дашборды
- Использование фильтров и параметров для персонализации данных
- Примеры интерактивных дашбордов в Superset

Тема 3. Форматирование текста на дашбордах в Superset

- Основы синтаксиса форматирования текста
- Примеры форматирования текста на дашбордах в Superset
- Использование заголовков, списков и ссылок

Тема 4. Основы администрирования в Superset

- Управление пользователями и ролями
- Настройка доступа к данным и дашбордам
- Мониторинг и оптимизация производительности Superset

Кейс-проект спринта 3 с небольшой проверкой ревьюера

Студент создает эскиз дашборда в Miro на основании предоставленного брифа заказчика, сборка макета дашборда в Superset на основании предоставленных шаблонов графиков (мокапов) и предоставленного интервью заказчика.

SQL

DataLens

Superset

Параметризация данных

Витрины данных

Оптимизация запросов

Администрирование BI-инструментов

Документирование данных

Качество данных

Итоговый проект модуля

Объединяет знания и навыки, полученные в спринтах 2 и 3. Студенты собирают дашборд в Superset на основе данных из PostgreSQL и документируют свой дашборд в Confluence. Тема для проекта выбирается самостоятельно для создания уникального проекта в портфолио.

Проект модуля с проверкой ревьюера

Студент собирает дашборд в Superset на основе предоставленных данных в базе данных PostgreSQL по интервью и документирует свой дашборд в Confluence. Тему для проекта студент выбирает самостоятельно, чтобы создать уникальный проект в своем портфолио.

Superset

PostgreSQL

Документирование

Визуализация данных

Модуль 6 Работа с данными и разработка витрины данных

Спринт 15 Проектирование витрины данных и предобработка данных

2 недели

Студенты изучат создание витрин данных, параметризацию и фильтрацию данных, продвинутые техники SQL и основы администрирования в Superset.

Тема 1. Принципы создания витрин данных

- Основные принципы и этапы создания витрин данных
- Оптимизация данных для аналитики
- Примеры эффективных витрин данных

Тема 2. Параметризация и фильтрация данных в Superset

- Использование параметров для персонализации данных
- Настройка и применение фильтров
- Примеры параметризации и фильтрации данных в Superset

Тема 3. Продвинутые техники SQL в DataLens

- Использование сложных SQL-запросов для анализа данных

Тема 4. Основы администрирования в Superset

- Управление пользователями и ролями
- Настройка доступа к данным и дашбордам
- Мониторинг и оптимизация производительности Superset

Тема 5. Документирование и закрепление знаний

- Создание документации для аналитических решений
- Описание процессов и метрик
- Сбор обратной связи и корректировка на основе полученных данных

Кейс-проект спринта 4 с автопроверкой

Подключение и очистка нескольких датасетов, оптимизация запроса и создание дашборда в DataLens.

DataLens

Superset

SQL

Витрины данных

Параметризация данных

Администрирование BI-инструментов

Документирование результатов

Оптимизация запросов

Спринт 16 Углубленная работа с данными и разработка отдельных визуализаций

2 недели

Студенты углубят свои знания в SQL и научатся строить дашборды по готовому макету, оптимизировать запросы и обеспечивать качество данных.

Тема 1. Основы SQL для BI-аналитиков

- Введение в SQL
- Основные команды и операторы
- Написание простых запросов для извлечения данных

Тема 2. Построение дашбордов по макету и готовому датасету

- Чтение и интерпретация макета дашборда
- Импорт и настройка готового датасета
- Создание дашборда в соответствии с макетом

Тема 3. Оптимизация запросов и правила хорошего кода

- Принципы написания оптимальных SQL-запросов
- Избежание распространенных ошибок
- Примеры оптимизации кода для повышения производительности

Тема 4. Использование сложных SQL в Superset

- Написание и выполнение сложных SQL-запросов в Superset
- Примеры использования подзапросов, объединений и агрегатных функций
- Оптимизация сложных запросов в Superset

Тема 5. Обеспечение качества данных для дашбордов

- Важность качества данных для аналитики
- Методы проверки и очистки данных
- Инструменты для обеспечения качества данных

Проект спринта 5 с автопроверкой

Практика SQL задач на различных бизнес-кейсах.

Superset

Качество данных

SQL

Визуализация данных

Оптимизация запросов

1 неделя

Итоговый проект модуля

Объединяет знания и навыки, полученные в спринтах 4 и 5. Студенты будут работать над проектом, включающим оптимизацию SQL-запросов, создание витрин данных и разработку визуализаций на основе бизнес-кейсов.

Проект модуля с автоматической проверкой (тестирование и проверка в SQL)

Практика SQL задач с интервью для подготовки к собеседованию, включая теоретические вопросы, задачи и чек-лист для самопроверки.

Финальный спринт

2 недели

Финальный проект

У студентов будет 2 недели на подготовку дипломного проекта. Им предстоит решить несколько сложных аналитических кейсов на основе реальных данных, используя изученные в курсе инструменты (SQL, DataLens, Superset). Студенты разработают комплексный дашборд для отслеживания ключевых показателей, задокументируют свои результаты в Confluence и презентуют готовое решение. Ревьюеры оценят проекты по рубриктору и дадут развёрнутую обратную связь.

Дипломный проект с ревьюерской проверкой

Студенту предоставляется набор данных от компании, а также задание на разработку комплексного дашборда для отслеживания ключевых показателей. Задача включает сбор и анализ данных, создание витрины данных, разработку интерактивного дашборда в Superset и документирование результатов в Confluence.

SQL

DataLens

Superset

Документирование

Анализ данных

Дашборды