

Кто такие аналитики данных

Продолжительность курса — 7 месяцев

01. В чём суть профессии

Аналитики данных помогают компаниям принимать решения на основе данных. Они ищут в данных закономерности, чтобы понять, почему происходят те или иные события, или предугадать, как всё может поменяться в будущем.

02. В каких сферах работают

Анализ данных нужен везде: в маркетинге, финансах, промышленности, разработке любых новых продуктов и технологий.

03. Какие задачи решают

Ритейл-сеть хочет найти районы с высокой плотностью населения и небольшим количеством супермаркетов. Нужно проанализировать данные, которые есть в компании, а также информацию из внешних источников.

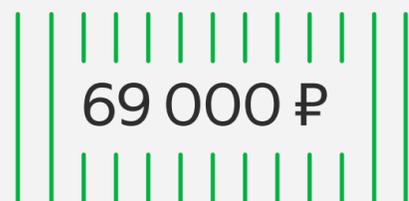
Маркетплейсу нужно сравнить два варианта текста рекламной рассылки и определить, на какой из них лучше реагируют пользователи, чтобы увеличить продажи.

Отделу маркетинга магазина одежды нужно определить причины провала рекламной кампании и оценить факторы, которые к этому привели.

04. Карьерные перспективы

Согласно исследованию «Анкора», около 45% всех компаний в России ищут специалистов по работе с данными.

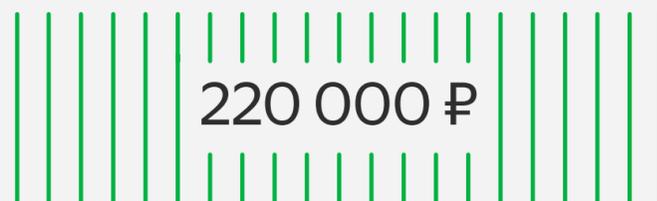
Зарплатная вилка



Джуниор аналитик данных



Мидл аналитик данных



Сеньор аналитик данных

На курсе — всё, что нужно, чтобы начать карьеру аналитика

Диплом о переподготовке

Непрерывная практика

Портфолио из 13+ проектов

Помощь в трудоустройстве

Почему этот курс подойдёт студентам без опыта

Свободное расписание

Читать теорию и практиковаться можно в любое время, главное — соблюдать сроки сдачи проектов. У нас есть мобильное приложение для учёбы, чтобы вы не были привязаны к компьютеру.

Понятная теория

Термины и правила подкреплены примерами из жизни. Сложность и длина программы курса рассчитаны так, чтобы каждую следующую главу вы понимали всё лучше. Мы постоянно обновляем программу: последнее обновление на данный момент — июль 2024 года

Практика в тренажёре

Учитесь читать, визуализировать и интерпретировать данные. Ошибайтесь, быстро получайте обратную связь и исправляйте ошибки

Учёба на реальных задачах

Вас ждут типичные для аналитика задачи из разных сфер. Вы решите их и сможете добавить в своё портфолио.

Команда сопровождения

Мы поддержим, объясним сложные темы, поможем улучшить проекты и не дадим сдать на полпути.

Основы анализа данных
с помощью SQL и BI

Анализ данных
с помощью Python

Расчёт и визуализация бизнес-
метрик и показателей

Визуализация данных с помощью
DataLens. Создание дашбордов

Развитие бизнес-мышления и работы
с метриками продукта и бизнеса

Анализ результатов A/B тестирования
с помощью Python

Чтобы начать учиться, не нужен опыт в IT

Среди наших выпускников, которые стали аналитиками данных, есть спортсмены, экономисты, таксисты и разнорабочие. После наших курсов аналитики находят работу и в стартапах, и в корпорациях вроде Яндекса, «Леруа Мерлен», «Вкусвилла» и Сбера.

Аналитик данных, Аналитик данных расширенный

Основная программа — 7 месяцев

Расширенная программа — 12 месяцев

01

Основы анализа данных
с помощью SQL и BI

12 недель

02

Анализ данных
с помощью Python

8 недель

03

Продвинутый анализ
данных для бизнеса

10 недель

Блоки, входящие в тариф «Аналитик данных расширенный»

Программа специализации «Продуктовый аналитик»

04

Погружение в продукт

2 недели

05

Мониторинг состояния
продукта

5 недель

06

Поддержка принятия
продуктовых решений

5 недель

Программа специализации «BI-аналитик»

04

Понимание бизнес-
требований
и прототипирование

2 недели

05

Визуализация данных
и подготовка
отчётности

2 недели

06

Работа с данными
и разработка витрины
данных

2 недели

07

Инструменты для работы
с большими данными

7 недель

01

Основы анализа данных с помощью SQL и BI

Спринт 1

2 недели

Введение в аналитику. Аналитический отчёт в Google Таблицах

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

Проект

Создание аналитического отчёта в Google Таблицах для небольшого салона для кудрявых

Темы

- 1. Использование данных в бизнесе**
Структурированные и неструктурированные данные. Роль данных в бизнесе.
- 2. Процесс анализа данных и задачи аналитика**
Задачи аналитика данных. Пайплайн работы аналитика. Специализации в аналитике.
- 3. Excel как инструмент аналитика. Основы Google Таблиц**
Табличные редакторы, начало работы в [Google Таблицах](#). Константы и формулы.
- 4. Предобработка данных в Google Таблицах**
Типы данных: числовые, текстовые. Форматирование данных. Очистка данных. Использование панели автоподсчёта. Сортировка и фильтрация данных.
- 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.
- 6. Презентация данных**
Построение простых визуализаций. Как поделиться отчётом.

Инструменты
и технологии

Пайплайн работы
аналитика

Google Таблицы

Формулы

Функции

Отчёты



Спринт 2

2 недели

Основы SQL. Извлечение данных для анализа

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

Инструменты и технологии

БД и СУБД

SQL

PostgreSQL

Типы данных

Группировка данных

Сортировка данных

Проект

Выгрузка данных для проведения статистики прослушиваний музыкального стримингового сервиса Поток.

Темы

- 1. Работа с базами данных**
Откуда аналитик получает данные. Что такое база данных. Как управляют базами данных с помощью СУБД. Первый запрос: выгружаем данные с помощью SQL.
- 2. Типы данных и их преобразования**
Какими бывают данные. Меняем тип данных. Округляем данные. Арифметические операции.
- 3. Фильтрация данных и агрегация**
Когда нужна фильтрация данных. Фильтрация по условиям. Агрегация данных.
- 4. Группировка и сортировка данных**
Группируем данные. Группировка по нескольким полям. Фильтрация после агрегации. Сортировка данных по одному полю. Сортировка данных по нескольким полям.

Спринт 3

2 недели

SQL. Обработка данных

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.

Инструменты и технологии

SQL

Пропуски

Дубликаты

Подзапросы

CTE

Присоединения таблиц (JOIN)

Проект

Проанализируете данные о тарифных планах и активности клиентов федерального оператора сотовой связи "Мегасеть"

Темы

- 1. Связи между таблицами**
Как таблицы хранятся в базе данных. Для чего нужны связи в базе данных. ER-диаграммы
- 2. Работа с пропущенными значениями и дубликатами**
Как работать с пропусками и дубликатами. Работа с пропусками. Заполнение пропусков. Работа с дубликатами.



- 3. Присоединение таблиц**
Присоединение таблиц. Разные типы присоединения таблиц (INNER JOIN, LEFT JOIN и RIGHT JOIN, FULL OUTER JOIN). Особенности присоединения таблиц.
- 4. Операции множеств и подзапросы**
Объединение множеств. Пересечение и вычитание, Подзапросы (в секции WHERE и в секции FROM). CTE (обобщённые табличные выражения).
- 5. Категоризация значений. Создание новых столбцов**
Операции со столбцами (вычитание, сложение, усреднение). Категоризация значений (CASE WHEN THEN END). Обработка неявных дубликатов.
- 6. 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.

Спринт 4

2 недели

SQL. Анализ данных и решение ad hoc задач

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.

Проект

Кейс-проект с ревью: Проанализируете данные о продажах внутри онлайн-игры "Секреты Темнолесья"

Темы

- 1. Знакомство с базой данных**
Как изучать БД самостоятельно. Выводы об устройстве БД.
- 2. Оконные функции. Агрегирующие функции.**
Назначение оконных функций, их классификация. Агрегирующие оконные функции (SUM(), COUNT(), AVG(), MIN(), MAX()). Предложения PARTITION BY и ORDER BY.
- 3. Оконные ранжирующие функции**
Назначение функций ROW_NUMBER(); RANK() и DENSE_RANK(), NTILE(). Особенности ранжирующих оконных функций.
- 4. Оконные функции смещения**
Назначение и особенности функций LEAD(), LAG(), FIRST_VALUE(), и LAST_VALUE().
- 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.

Инструменты и технологии

Решение ad hoc задач

Декомпозиция SQL

Агрегирующие оконные функции

Ранжирующие оконные функции

Оконные функции смещения

Аналитические оконные функции

Мода

Медиана

Среднее

Перцентиль

Размах



6. **Практика решения ad-hoc задач**
Что такое ad hoc запросы. Алгоритм решения ad hoc запроса.
Декомпозиция. Решение ad hoc запросов повышенной сложности.

Спринт 5

2 недели

Визуализация данных с помощью DataLens. Создание дашбордов

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

Инструменты
и технологии

SQL

DataLens

BI-инструменты

Чарт

Визуализация
данных

Дашборд

Проект

Кейс-проект с ревью: Разработка дашборда по чёткому ТЗ на данных конференция TED

Темы

1. **Визуализация в работе аналитика. Знакомство с DataLens**
Визуализация как задача 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.
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.
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.
4. **Дашборды**
Назначение дашбордов. Прототипирование дашбордов. Виджеты. Чарты. Добавление чартов на дашборд. Селекторы. Связи: входящие и исходящие. Настройка селекторов. Тексты. Заголовки. Композиция дашборда.
5. **Параметры**
Использование параметров для дашборда. Использование параметров для чарта. Ситуации, в которых используются параметры. Специальные параметры.
6. **Интерпретация данных из дашбордов**
Задачи визуализации и её аудитория. Принципы хорошей визуализации. Работаем с дашбордом.



1 неделя

Итоговый проект модуля

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

Проект

Проанализируете объявления о продаже жилой недвижимости с помощью SQL-запросов и доработаете существующий дашборд по требованиям заказчика.

Каникулы → 1 неделя



02

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.

Инструменты и технологии

Проект

Решение проверочных заданий на знание синтаксиса Python.

Python

Переменные

Типы данных

Строки

Списки

Циклы

Условный оператор

Функции

Множества

Словари

Темы

- 1. Основы синтаксиса Python**
Что такое Python. Почему аналитики пишут на Python. Что такое переменная. Как выбрать имя переменной. Арифметические операции. Используем переменные.
- 2. Определение данных и их типы**
Типы данных. Преобразование типов. Как работать со строками. Заполнение списков. Операции со списками. Строки и списки.
- 3. Условные выражения**
Логические выражения и операторы сравнения. Сложные логические выражения. Возвращаемые элементы True и False. Условная конструкция. Дополнительные ветки кода.
- 4. Циклы и их организация**
Понятие цикла. Элементы цикла. Перебор элементов цикла. Циклы и строки. Вложенные циклы. Обработка списков в цикле. Функция enumerate() в заголовке цикла. Функция zip() в заголовке цикла. Форма алгоритма. Управление циклом. Операторы break, continue и pass.
- 5. Функции в Python**
Встроенные функции и методы. Пользовательская функция и её синтаксис. Аргументы функции. Тело функции. Вывод результатов с помощью return. Математические функции.
- 6. Словари и множества**
Создание словаря. Поиск в словаре. Изменение словарей. Словари и циклы. Множества.
- 7. Работа с вложенными структурами**
Вложенные структуры: списки. Списковые включения. Работа с вложенными списками. Вложенные словари. Словари с вложениями. Классы, объекты, конструкторы.



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.

Инструменты
и технологии

Python

Pandas

Предобработка
данныхОбработка
пропусковОбработка
дубликатовКатегоризация
данных

Проект

Кейс-проект с ревью: Проведёте предобработку данных в Python онлайн-игры "Секреты Темнолесья" для последующего анализа

Темы

- 1. Основы библиотеки pandas. Обзор данных**
Что такое библиотека в Python. Библиотека pandas. Объекты DataFrame и Series. Работа с файлами в датафрейме. Обзор данных в датафрейме. Сортировка данных.
- 2. Типы данных. Работа с датой и временем**
Типы данных pandas. Тип данных object. Пропуски. Преобразование типов данных. Типы данных для даты и времени. Операции с датой и временем. Интервал времени timedelta64. Среда Jupyter Notebook.
- 3. Индексация в датафреймах. Фильтрация данных**
Индексация и фильтрация в pandas. Индексация с помощью loc и iloc. Индексация с помощью логических операторов.
- 4. Работа с пропущенными значениями**
Методы работы с пропусками. Заполнение пропусков. Удаление пропусков. Явные и неявные дубликаты. Работа со строками.
- 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.
- 6. Категоризация данных**
Категоризация данных. Функция apply() для Series. Функция apply() для датафреймов. Группировка данных. Агрегация с помощью метода agg().



Исследовательский анализ данных и визуализация с помощью Python

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

Инструменты и технологии

Python

Pandas

Matplotlib

Seaborn

Jupyter notebook

Проект

Кейс-проект с ревью: Проведение исследовательского анализа данных рынка общественных заведений Москвы для инвестиционного фонда

Темы

- 1. Объединение датафреймов**
Присоединение датафреймов. Присоединяем датафреймы по столбцам. Присоединяем датафреймы по индексам. Методы `join()` и `merge()`.
- 2. Описательная статистика**
Меры центральной тенденции. Меры разброса. Строим гистограмму. Строим диаграмму размаха. Изучаем меры размаха. Оформляем графики.
- 3. Взаимосвязь переменных**
Знакомство с корреляцией. Виды корреляции. Считаем и оцениваем корреляцию. Визуализируем корреляцию.
- 4. Визуализация для изучения данных**
Столбчатая диаграмма. Линейный график. Тепловая карта. График распределения. Как аналитик исследует графики.
- 5. Сводные таблицы**
Сводные таблицы в `pandas`. Работа с мультииндексами в сводной таблице. Фильтрация и работа с пропусками в сводной таблице.
- 6. Пример исследовательского анализа данных**
Разбор исследовательского анализа данных в среде Jupyter Notebook.



1 неделя

Итоговый проект модуля

Под запрос бизнеса выполните исследовательский анализ данных с последующей визуализацией с помощью инструментов Python. Предоставите рекомендации бизнесу по итогам исследования.

Проект

Исследование данных с помощью Python о стартапах для финансовой компании

Каникулы —————> 1 неделя



03

Продвинутый анализ данных для бизнеса

Спринт 9

2 недели

Расчёт и визуализация бизнес-метрик и показателей

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

Инструменты и технологии

SQL

DataLens

Проект

Кейс-проект с ревью: Расчёт, визуализация и анализ основных метрик сервиса доставки еды.

Темы

- 1. Что такое бизнес-метрики и зачем они нужны**
Различия бизнес-анализа и продуктового анализа. Продуктовый подход к анализу метрик. Классификация метрик. Data-driven-подход и генерация гипотез для анализа метрик.
- 2. Расчёт, визуализация и интерпретация продуктовых метрик**
Расчёт DAU. Визуализация DAU. Расчёт конверсии. Визуализация и интерпретация конверсии. Расчёт среднего чека. Визуализация и интерпретация среднего чека. Расчёт LTV. Визуализация и интерпретация LTV.
- 3. Продуктовая воронка**
Что такое продуктовая воронка. Расчёт воронки. Визуализация воронки и формулировка рекомендаций для бизнеса
- 4. Расчёт метрик конверсии и LTV по когортам**
Что такое когортный анализ. Расчёт Retention Rate по когортам. Визуализация и интерпретация Retention Rate по когортам. Расчёт LTV по когортам. Визуализация и интерпретация LTV по когортам. Юнит-экономика: окупаемость пользователя или продукта.
- 5. Выбор ключевых и кастомных метрик**
Что такое ключевые метрики. Что такое кастомные метрики. Выбор ключевых метрик для разных сфер.



Формулировка и проверка гипотез. Статистический анализ данных

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

Инструменты
и технологии

Проект

Проверка гипотез в бизнесе для сервиса проката самокатов GoFast в тренажёре.

Темы

- 1. Основы теории вероятностей**
Эксперимент, исходы, вероятностное пространство и события. Пересекающиеся и взаимоисключающие события. Диаграмма Эйлера — Венна. Сложение и умножение вероятностей. Вероятность: классическое и геометрическое определения. Закон больших чисел.
- 2. Случайные величины**
Случайные величины и их виды. Вероятность попасть в интервал для дискретной случайной величины. Функция распределения дискретной случайной величины. Математическое ожидание дискретной случайной величины. Дисперсия дискретной случайной величины.
- 3. Распределения**
Эксперимент Бернулли. Биномиальный эксперимент. Биномиальное распределение. Распределение Пуассона. Нормальное распределение и плотность вероятности. Параметры нормального распределения. Функция нормального распределения. Percent Point Function для нормального распределения.
- 4. Основы статистики**
Генеральная совокупность и выборки. Точечные оценки. Выборочное распределение. Центральная предельная теорема. Стандартная ошибка среднего.
- 5. Проверка гипотез**
Гипотезы. Логика проверки гипотез. P-value. Проверка гипотез с одной выборкой в Python. Проверка гипотез с двумя выборками в Python. t-тест Стьюдента. Ошибки I и II рода. Проблема множественной проверки гипотез. Поправки на множественную проверку гипотез.

Python

Pandas

scipy

statsmodels

нулевая гипотеза

альтернативная
гипотеза

распределения
метрик

ошибка первого
рода

ошибка второго
рода

статистический
тест

множественная
проверка гипотез

t-тест

биномиальный
тест



Анализ результатов A/B тестирования с помощью Python

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

Инструменты и технологии

SQL

Python

Pandas

Scipy

Statsmodels

Jupyter Notebook

t-тест

тест Манна-Уитни

z-тест

Проект

Кейс-проект с ревью: Разработка A/B-тестирования и анализ результатов для новой рекомендательной системы развлекательного приложения.

Темы

- 1. Что такое A/B-тесты и зачем они нужны**
Знакомство с A/B-тестированием. Особенности проведения A/B-тестирования. Преимущества и ограничения A/B-тестирования. Процесс проведения A/B-тестирования.
- 2. Выбор метрики для проверки гипотезы**
Классификация метрик эксперимента. Выбор метрик.
- 3. Расчёт размера выборки. Валидация результатов**
Расчёт размера выборки и длительности эксперимента. Минимальный размер эффекта. Расчёт размера аудитории с помощью калькуляторов. Проверка данных. Расчёт метрик по результатам эксперимента.
- 4. Проверка результатов A/B-теста. Тест Манна-Уитни. Z-тест пропорций**
Алгоритм проверки результатов A/B-теста. T-тест и t-тест Уэлча. Тест Манна — Уитни. Z-тест пропорций.

Итоговый проект модуля

Рассчитаете несколько метрик с помощью SQL и интерпретируете полученные результаты. Затем проверите гипотезу с помощью Python и сформулируете выводы о результатах эксперимента для бизнеса.

Проект

На основании данных сервиса Яндекс Книги рассчитаете и интерпретируете метрики о ежедневной активности пользователей в SQL, после чего проверите гипотезу в Python. Проанализируете результаты A/B тестирования новой версии сайта интернет-магазина BitMotion Kit.



2 недели

Дипломный проект курса «Аналитик данных»

Потренируетесь использовать все полученные на курсе навыки на большом объеме реальных данных, используя (SQL (PostgreSQL), Yandex Datalens, Python (pandas, scipy, statsmodels)).

Проект

Изучите данные Яндекс Афиши и рассчитаете метрики с помощью SQL-запросов, подготовите дашборд в Yandex Datalens с описанием основных трендов, проведёте исследовательский анализ данных и проверите гипотезы статистическими методами с помощью Python, сформулируете основные выводы и рекомендации.

3-5 часов
на кейс

Плюс 10 кейсов от реальных работодателей

Мы попросили у наших партнёров реальные задачи, проработали решение и добавили в курс. Вы сможете дополнительно выбрать, сколько задач взять в работу, необязательно решать их все.

Благодаря этим задачам вы с первых дней обучения будете видеть, с чем сталкиваются аналитики на работе.

В начале курса будут более простые задачи, адаптированные под ваш текущий уровень, но с каждым модулем они будут становиться сложнее и сложнее.



Блоки, входящие в тариф «Аналитик данных расширенный»

Программа специализации «Продуктовый аналитик»

04 Погружение в продукт

Спринт 12

Как живёт продукт. Логирование

2 недели

Узнаете, как работает продуктовая команда и какое место в ней занимает продуктовый аналитик. Научитесь самостоятельно анализировать действия пользователей в продукте и составлять аналитические спецификации исходя из ТЗ продакт-менеджера.

Инструменты и технологии

Excel

Логирование

Спецификации

Проект

Разработка аналитических спецификаций для приложения

Темы

- 1. Жизненный цикл продукта**
Как живёт и развивается продукт. Что такое data-driven цикл разработки. Как аналитик участвует в жизненном цикле продукта. Работа в продуктовой команде.
- 2. Анализ работы продукта**
Исследование работы продукта и его экономики. Анализ особенностей продукта глазами пользователя. Выявление ключевых точек взаимодействия. Монетизационные модели продуктов.
- 3. Подготовка аналитических спецификаций**
Аналитические логи и их роль в работе аналитика. Проектирование и разработка аналитических событий для отслеживания взаимодействий пользователей с продуктом. Использование электронных таблиц (Excel) для разработки аналитических спецификаций.

Итоговый проект модуля

Разработаете аналитические спецификации для приложения



05

Мониторинг состояния продукта

Спринт 13

2 недели

Построение системы метрик продукта

Погрузитесь в работу с метриками. Разберётесь, как они связаны и как ухудшение одной метрики может повлиять на другие. Научитесь оценивать состояние продукта с помощью анализа и сравнения метрик. Научитесь обнаруживать аномалии и делать визуализацию метрик.

Проект

Работа с системой метрик продукта для обнаружения проблем

Темы

- 1. Система метрик в продукте**
Основные метрики продукта (DAU, ARPU, удержание и другие). Система метрик продукта, связь между метриками и их взаимное влияние. Классификация метрик. Иерархия метрик, главные метрики продуктов в различных областях бизнеса. Кастомные метрики.
- 2. Расчёт и визуализация пользовательских метрик**
Пользовательские метрики. Расчет пользовательских метрик продукта с помощью Python. Визуализация пользовательских метрик и их анализ.
- 3. Расчёт и визуализация монетизационных метрик**
Монетизационные метрики. Расчет монетизационных метрик продукта с помощью Python. Визуализация монетизационных метрик и их анализ.
- 4. Кейсы и анализ поведения метрик**
Анализ поведения продуктовых метрик. Обнаружение аномалий в поведении метрик и методы поиска причин их возникновения.

Инструменты и технологии

Python

SQL

Метрики

Монетизация

Визуализация

Аномалии



2 недели

Углубитесь в расчёты метрик юнит-экономики и когортного анализа.

Инструменты
и технологии

Проект

Проведение когортного анализа для решения ad-hoc задачи

Python

SQL

Юнит-экономика

Когортный
анализ

Окупаемость

Конверсия

Удержание

LTV

ROI

Темы

- 1. Мониторинг здоровья продукта с точки зрения качества привлекаемых пользователей**
Основы когортного анализа. Методики привлечения пользователей продукта. Методики и метрики исследования качества привлекаемой аудитории (удержание, конверсия, LTV, ROI и другие).
- 2. Расчёт и визуализация удержания для когорт**
Расчёт и визуализация удержания с точки зрения когортного анализа с помощью Python. Анализ удержания для различных когорт.
- 3. Расчёт и визуализация конверсии и LTV для когорт**
Расчёт и визуализация LTV и конверсии с точки зрения когортного анализа с помощью Python. Анализ метрик монетизации для различных когорт. Методики самопроверки.
- 4. Расчёт и визуализация метрик окупаемости для когорт**
Расчёт и визуализация окупаемости и ROI с точки зрения когортного анализа с помощью Python.
- 5. Основы юнит-экономики**
Базовые понятия юнит-экономики. Юнит-экономика продукта и юнит-экономика клиента. Расчёт юнит-экономики в примерах.
- 6. Расчёт метрик для юнит-экономики**
Анализ сходимости юнит-экономики для различных когорт с помощью Python. Разбор кейсов по анализу юнит-экономики. Методы самопроверки.

1 неделя

Итоговый проект модуля

С помощью системы продуктовых метрик и когортного анализа на Python вы сможете продуктовой команде найти/идентифицировать проблемные группы пользователей.

Проект

Работа с системой метрик продукта для обнаружения проблем в нескольких когортах



06

Поддержка принятия продуктовых решений

Спринт 15

2 недели

А/В-тесты в работе продуктового аналитика

Научитесь готовить, проводить и делать выводы по результатам А/В-тестирования.

Инструменты и технологии

Проект

Проектирование, проведение и анализ результатов А/В-теста

Темы

- А/В-тесты в работе продуктового аналитика**
Роль А/В-тестирования в жизненном цикле продукта. Этапы А/В-тестирования. Функции аналитика на каждом из этапов А/В-тестирования.
- Статистические тесты для проверки гипотез**
Основы выбора и применения статистических тестов (t-тест, тест Манна-Уитни, z-тест). Множественное тестирование, выбор теста в ситуации множественных гипотез (ANOVA, тест Краскел-Уоллиса, критерий хи-квадрат). Статистические тесты с помощью Python и библиотек Scipy и Statmodels.
- Подготовка и проведение теста**
Определение назначения и гипотез теста. Выбор метрик, длительности и момента запуска теста. Методики проверки адекватности проведения теста, выбор момента остановки теста.
- Анализ результатов А/В-теста**
Применение статистических тестов для анализа результатов теста. Получение бизнес-инсайтов из результатов теста и презентация результатов теста заказчику.

А/В-тесты

T-тест

Тест Манна-Уитни

Z-тест

ANOVA

Scipy

Statmodels

Спринт 16

2 недели

Поиск инсайтов в работе продукта

Научитесь использовать наиболее популярные методики поиска точек роста для обнаружения потенциальных точек улучшения продукта.

Инструменты и технологии

Проект

Сегментация аудитории продукта с использованием машинного обучения

SQL

Python

User Journey

Сегментация

Воронки



Темы

- 1. Продуктовый аналитик и задача поиска точек роста**
Методики поиска точек роста продукта. Роль аналитика в процессе поиске точек роста.
- 2. Методики поиска точек роста: воронки и UJ**
Применение воронок для поиска точек роста. Исследование user journey для сравнения ожидаемого и реального поведения пользователей. Использование Python для построения воронок и оптимизации user journey.
- 3. Методики поиска точек роста: сегментация**
Методики сегментации. Основы статистики для сегментации пользователей. Кластеризация пользователей с помощью алгоритма K-Means в Python.
- 3. Кейсы поиска точек роста**
Обзор примеров поиска точек роста в различных областях бизнеса.

1 неделя

Итоговый проект модуля

Проведёте сегментацию пользователей. Для обнаруженных сегментов проведёте статистические тесты, чтобы выяснить, есть ли среди них различия в ключевых показателях.

Проект

Проведение АВ-теста для определения оптимального варианта user journey.

2 недели

Дипломный проект

Решите несколько сложных аналитических кейсов на большом объеме реальных данных по запросу бизнеса, используя изученные в курсе инструменты (SQL (PostgreSQL), Yandex Datalens, Python, pandas, scipy, statsmodels) и презентуете бизнесу решение.

Проект

Несколько кейсов из различных областей бизнеса с ревью



04

Понимание бизнес-требований и прототипирование

Спринт 12

2 недели

От выбора инструмента до создания дашборда

Научитесь проводить интервью с заказчиком, выявлять и документировать бизнес-требования, создавать эскизы дашбордов. Узнаете, как приоритизировать требования и выделять критически важные для бизнеса элементы.

Проект

Решение конкретной бизнес-задачи с помощью выбранного BI-инструмента. Заполнение брифа заказчика на основе готового дашборда. Подбор подходящего дизайна и визуализации для представления данных

Темы

- 1. Знакомство с BI-аналитикой**
Применение BI в бизнесе. Основные компоненты BI-систем. Примеры успешного использования BI.
- 2. Сравнение инструментов BI и их реализация**
Обзор популярных BI-инструментов. Ключевые функции и возможности каждого инструмента. Преимущества и недостатки различных BI-решений.
- 3. Типы визуализаций и их применение**
Подготовка к интервью с заказчиком. Методики проведения интервью. Вопросы для выявления бизнес-требований. Анализ и документирование собранной информации.
- 4. Принципы дизайна дашбордов**
Основы дизайна дашбордов. Важность юзабилити и UX. Принципы размещения и группировки элементов. Использование цветов и шрифтов.
- 5. Использование Miro для создания эскизов дашбордов**
Что такое Miro и зачем он нужен BI-аналитику. Обзор возможностей платформы. Работа с фреймами и сеткой. Управление слоями. Создание прототипа дашборда. Создание интерактивных элементов.
- 6. DataLens создание по мокапу из Miro**
Интеграция с Miro. Настройка параметров доступа. Добавление динамических элементов.

Инструменты и технологии

Miro

DataLens

Визуализация данных

Проведение интервью

Сбор требований

Прототипирование дашбордов

Документирование требований



05

Визуализация данных и подготовка отчётности

Спринт 13

2 недели

Проектирование дизайна дашборда и документирование его логики

Изучите принципы дизайна дашбордов, работу с различными типами чартов и инструментами визуализации данных. Освойте создание интерактивных и параметризованных дашбордов, а также документирование и закрепление результатов работы.

Инструменты
и технологии

[DataLens](#)

[Superset](#)

[Markdown](#)

[Miro](#)

[Визуализация
данных](#)

[Администриро-
вание BI-
инструментов](#)

[Документирова-
ние результатов](#)

[Интерактивные
дашборды](#)

Проект

Создание эскиза дашборда в Miro на основании предоставленного брифа заказчика. Сборка макета дашборда в DataLens на основании предоставленных шаблонов графиков (мокапов) и предоставленного интервью заказчика

Темы

- 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.
- 2. Интерактивность и параметризация дашбордов**
Добавление интерактивных элементов на дашборды. Использование фильтров и параметров для персонализации данных. Примеры интерактивных дашбордов.
- 3. Форматирование текста на дашбордах с помощью Markdown**
Основы синтаксиса Markdown. Примеры форматирования текста с помощью Markdown. Добавление заголовков, списков и ссылок на дашбордах.
- 4. Основы администрирования в DataLens**
Управление пользователями и ролями. Настройка доступа к данным и дашбордам. Мониторинг и оптимизация производительности DataLens.
- 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.



Завершение работы над дашбордом и документацией. SUPERSET

Научитесь работать с Superset: создавать чарты, датасеты, интерактивные и параметризованные дашборды. Освойте документирование и администрирование.

Инструменты и технологии

Superset

SQL

Параметризация дашборда/ интерактивность

Администрирование Superset

Проект

Создание эскиза дашборда в Miro на основании предоставленного брифа заказчика. Сборка макета дашборда в Superset на основании предоставленных шаблонов графиков (мокапов) и предоставленного интервью заказчика

Темы

- 1. Введение в Superset: чарты, датасеты и дашборды**
Основные возможности Superset. Создание и настройка чартов. Работа с датасетами. Создание и управление дашбордами.
- 2. Интерактивность и параметризация дашбордов в Superset**
Добавление интерактивных элементов на дашборды. Использование фильтров и параметров для персонализации данных. Примеры интерактивных дашбордов в Superset.
- 3. Улучшение визуализации в Superset с помощью Markdown и CSS**
Обзор возможностей кастомизации в Superset. Зачем нужны Markdown и CSS. Создание кастомных тем. Анимация элементов.
- 4. Основы администрирования в Superset**
Управление пользователями и ролями. Настройка доступа к данным и дашбордам. Мониторинг и оптимизация производительности Superset.
- 5. Тестирование дашбордов**
Виды ошибок в дашбордах. Функциональное и нефункциональное тестирование. UI/UX тестирование. Написание тестов.

Итоговый проект модуля

Соберёте дашборд в Superset на основе данных из PostgreSQL и документируете свой дашборд в Confluence. Создадите уникальный портрет для портфолио по собственной идее.

Проект

Сборка дашборда в Superset на основе предоставленных данных в базе данных PostgreSQL по интервью. Документирование своего дашборда в Confluence. Создание уникального проекта для портфолио по собственной идее



06

Работа с данными и разработка витрины данных

Спринт 15

2 недели

Проектирование витрины данных и предобработка данных

Узнаете, как создать витрину данных. Изучите параметризацию и фильтрацию данных, продвинутые техники SQL и основы администрирования в Superset.

Инструменты
и технологии

DataLens

Superset

SQL

Витрины данных

Параметризация
данных

Оптимизация
запросов

Документиро-
вание данных

Качество данных

Проект

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

Темы

- 1. Введение в базы данных**
Основы управления данными. Модели данных. Базовые концепции. Резервное копирование.
- 2. Устройство базы данных**
Архитектура БД. Витрины данных. Партиционирование. Фрагментация. Денормализация. Настройка партиционирования.
- 3. Оптимизация запросов**
Основы SQL. Анализ производительности. Индексирование. Материализованные представления. Хеширование. Оптимизация JOIN.
- 4. Применение оптимизации в дашбордах**
Особенности дашбордов. Оптимизация витрин. Интеграция с BI. Создание оптимизированной витрины. Настройка ETL.



Углублённая работа с данными и разработка отдельных визуализаций

Углубите свои знания в SQL. Научитесь строить дашборды по готовому макету, оптимизировать запросы и обеспечивать качество данных.

Инструменты и технологии

SQL

Superset

Оптимизация запросов

Качество данных

Визуализация данных

Проект

Практика SQL задач на различных бизнес-кейсах

Темы

- 1. Основы SQL для BI-аналитиков**
Введение в SQL. Основные команды и операторы. Написание простых запросов для извлечения данных.
- 2. Построение дашбордов по макету и готовому датасету**
Чтение и интерпретация макета дашборда. Импорт и настройка готового датасета. Создание дашборда в соответствии с макетом.
- 3. Оптимизация запросов и правила хорошего кода**
Принципы написания оптимальных SQL-запросов. Избежание распространенных ошибок. Примеры оптимизации кода для повышения производительности.
- 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.
- 5. Обеспечение качества данных для дашбордов**
Важность качества данных для аналитики. Методы проверки и очистки данных. Инструменты для обеспечения качества данных.

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

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



Дополнительные модули программы «Аналитик данных. Расширенный»

07

Инструменты для работы с большими данными

Спринт 17

1 неделя

Введение в большие данные

Познакомитесь с концепцией и характеристиками больших данных.

Проект

Решение проверочных заданий

Темы

- 1. Большие данные**
Понятие больших данных. Характеристики больших данных. Как применяют большие данные.
- 2. Хранение и обработка больших данных**
Как работать с большими данными. Хранение данных. Обработка данных. Автоматизация данных.
- 3. Устройство команд в больших компаниях/корпорациях**
Кросс-функциональная команда. Роли в кросс-функциональной команде. ETL процесс.
- 4. Роль аналитика и пайплайн работы с большими данными**
Роль аналитика в продуктовых командах. Взаимодействие аналитика и команды разработки. Взаимодействие аналитика и дата-инженера.

Инструменты
и технологии

Большие данные
ETL
DWH

Спринт 18

2 недели

ClickHouse как аналитическая СУБД

Познакомитесь с концепцией и характеристиками больших данных.

Проект

Кейс-проект с ревью: Решение нескольких SQL-запросов с помощью Clickhouse

Инструменты
и технологии

Clickhouse
Движки таблиц
Массив
Комбинато
Масштабирование
данных



Темы

- 1. Что такое Clickhouse и как он устроен**
Особенности СУБД Clickhouse. Особенности синтаксиса Clickhouse. Движки таблиц (table engines).
- 2. Типы данных в Clickhouse**
Типы данных в Clickhouse. Что такое массив. Работа с массивами в Clickhouse.
- 3. Комбинаторы и аналитические функции в Clickhouse**
Комбинаторы в Clickhouse. Специфичные агрегирующие функции и комбинаторы. Аналитические функции описательной статистики в Clickhouse. Функции - статистические тесты. Оконные функции в Clickhouse.
- 4. Работа с таблицами в Clickhouse**
Создание, переименование и удаление таблиц. Действия со столбцами и строками.
- 5. Масштабирование данных в СУБД**
Для чего нужно масштабирование данных. Механизмы масштабирования данных.

Спринт 19

2 недели

Обработка данных с помощью PySpark

Научитесь извлекать и обрабатывать большие данные с помощью PySpark.

Инструменты
и технологии

PySpark

S3

RDD

Датафрейм

Ленивые
вычисления

Кэширование

Проект

Кейс-проект с ревью: Извлечение и обработка сырых данных из хранилища больших данных с помощью PySpark

Темы

- 1. Хранилище данных S3 и PySpark**
Что такое хранилище данных S3. Знакомство с файлами с помощью S3. Назначение pySpark. Основные форматы данных и компоненты PySpark. Архитектура распределённых вычислений Spark. Создание объектов SparkSession.
- 2. Основные структуры данных в PySpark**
RDD. Датафрейм (DataFrame). Схема датафрейма и типы данных.
- 3. Трансформации датафреймов и действия**
Трансформации датафреймов. Фильтрация и группировка данных в PySpark. Сортировка данных и создание новых столбцов. Удаление полей и удаление строк-дубликатов. Присоединение данных и операции множеств. Действия в PySpark.
- 4. Кэширование и запись данных в другие источники**
Кэширование. Запись датафреймов в Clickhouse.
- 5. Пример обработки данных с помощью PySpark.**
Работа с PySpark в Jupyter Notebook



Спринт 20

Автоматизация (AirFlow)

2 недели

Научитесь извлекать и обрабатывать большие данные с помощью PySpark.

Инструменты
и технологии

Проект

Автоматизация процесса извлечения и переноса данных из S3 в Clickhouse

Airflow

Темы

1. Airflow для аналитических задач
2. Создание DAG в Airflow
3. Операторы и сенсоры в Airflow
4. Автоматизация выполнения задач по расписанию
5. Мониторинг и отладка в Airflow

2 недели

Дипломный проект курса «Аналитик данных расширенный»

Решите несколько сложных аналитических кейсов на большом объеме реальных данных по запросу бизнеса, используя изученные в курсе инструменты (SQL (PostgreSQL), Yandex Datalens, Python, pandas, scipy, statsmodels) и презентуете бизнесу решение.

Проект

Несколько кейсов из различных областей бизнеса с ревью.