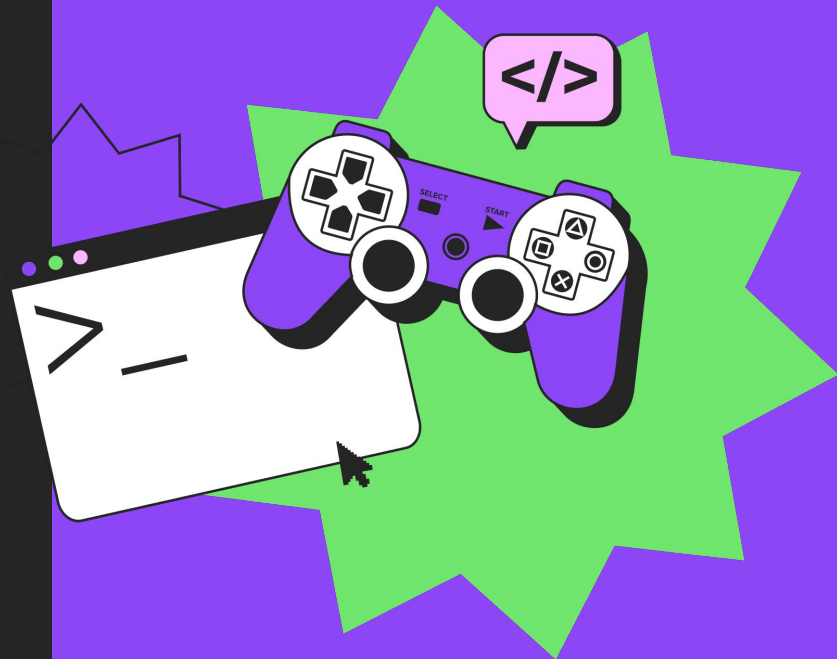





Анализ поведения клиентов на примере сайта GB





Что будет на уроке сегодня

-  Описание проекта
-  Этапы работы над проектом
-  Практическая часть





Вводные

Задача: провести анализ логов с сайта GB, рассчитать некоторые из метрик и выявить инсайты, которые могут быть полезны бизнесу.





Вводные

Задача: провести анализ логов с сайта GB, рассчитать некоторые из метрик и выявить инсайты, которые могут быть полезны бизнесу.

Данные: логи сайта

session_id	идентификатор сессии
claim	признак наличия отправки формы заявки во время текущей сессии (0 – нет заявки, 1 – заявка отправлена)
60sec	признак длительности сессии 60 секунд и более (0 – сессия менее 60 сек, 1 – сессия 60 сек и более)
scroll_90	признак вертикальной прокрутки страницы глубиной 90% (0 – прокрутка менее 90%, 1 – прокрутка 90% и более)
hit_date	дата сессии
referer_url	адрес реферера страницы просмотра
url	адрес страницы просмотра
utm_source	utm-метка источника
utm_medium	utm-метка канала
gender	пол посетителя из куки (-1 – не определен, 0 – мужской, 1 – женский)
age	возраст посетителя из куки (-1 – не определен)
touch_screen	touch экран (0 – не определен, 1 – нет, 2 – есть)
has_vk_id	наличие идентификатора профиля пользователя Вконтакте (0 – нет, 1 – есть)
has_ok_id	наличие идентификатора профиля пользователя Одноклассников (0 – нет, 1 – есть)



Этапы работы над проектом







1. Загрузка данных. Проверка корректности типов данных
2. Исследовательский анализ данных
3. Статистический анализ данных
4. Предобработка данных (обработка пропусков, аномалий)
5. Построение воронки клиентов
6. Выявление инсайтов в данных



Исследовательский анализ данных

Разведочный анализ данных (exploratory data analysis, EDA) — предварительный анализ данных для определения общих закономерностей и свойств. Понятие введено математиком Джоном Тьюки.

Цели анализа:

-  Максимальное «проникновение» в данные
-  Выявление основных структур
-  Выбор наиболее важных переменных
-  Обнаружение отклонений и аномалий
-  Проверка основных гипотез
-  Разработка начальных моделей



Статистический анализ данных



Анализ агрегированных характеристик (среднего значения, моды, медианы и так далее)



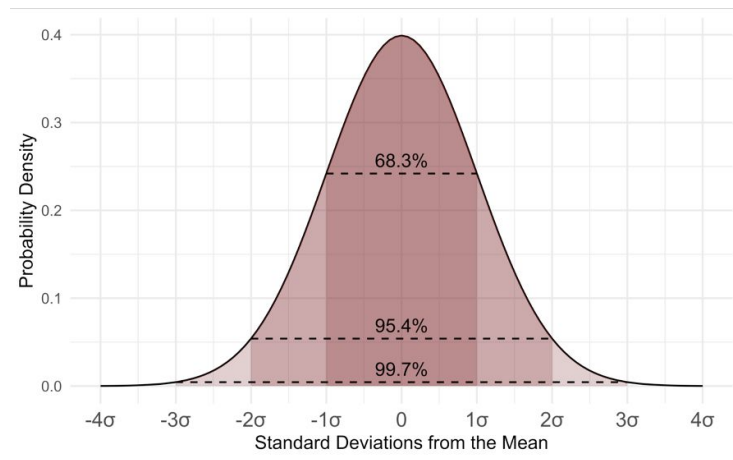
Выявление аномалий с помощью статистических методов (например, IQR)



Анализ типа распределения признаков

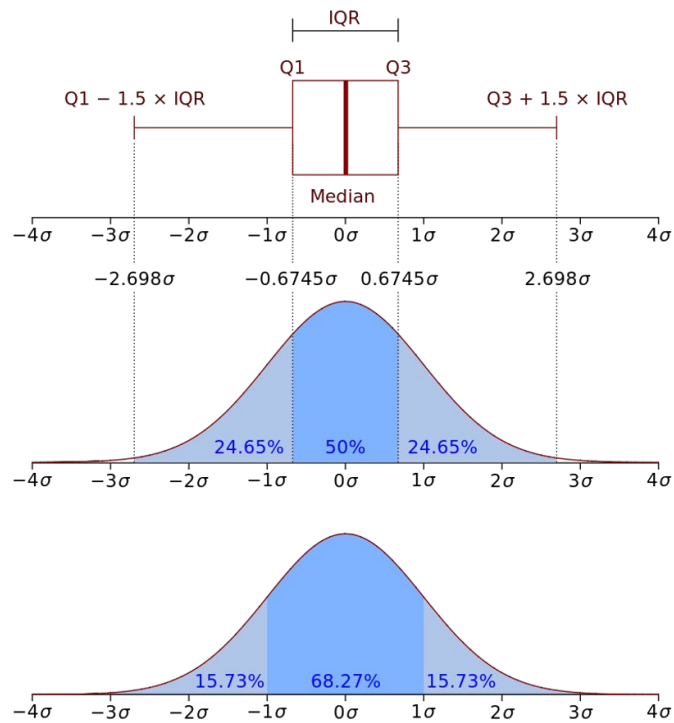


Проверка гипотез (проведение статистических тестов)





IQR (метод межквартильных расстояний)





Воронка продаж

Воронка продаж — это аналитический инструмент, который позволяет понять, как ваш потенциальный клиент приходит к решению о покупке ваших товаров или услуг.

Путь покупателя AIDA:





Воронка продаж

Воронка продаж — это аналитический инструмент, который позволяет понять, как ваш потенциальный клиент приходит к решению о покупке ваших товаров или услуг.

В рамках проекта нужно:

- Провести обработку поля 'referrer_url' и выделить категории (посещение урока, поиск, вход в личный кабинет и так далее).
- Определить, какие события будут показателями того, что клиент доволен сервисом.
- Сформулировать выводы по проведённому анализу.

Путь покупателя AIDA:





На какие вопросы ищем ответы



Какие действия клиенты GB совершают чаще всего?



Есть ли аномалии в поведении пользователей: например, большое число коротких (менее 60 с.) сессий?
С чем эти аномалии могут быть связаны?



На каких страницах прокрутка 90% и более?



В какие даты было больше всего сессий? Рассчитайте DAU, MAU, WAU.



Какие изменения на сайте могут улучшить показатели: число посещений, длительность сессии и другие?



Полезные ссылки/дополнительные материалы



[Understanding Boxplots](#)



[Предварительная обработка данных](#)



[Что такое воронка продаж](#)



[Анализ данных с использованием Python](#)



Итоги урока



Описание проекта



Этапы работы над проектом



Практическая часть



Вопросы?

Вопросы?



Вопросы?

