

Финальный аналитический отчёт

по данным CRM онлайн-школы программирования X

Описание проекта

Данный проект выполнен в роли аналитика данных для онлайн-школы программирования X.

Школа ведёт учёт всех клиентов, лидов и сделок в CRM-системе, где хранятся данные о контактах, звонках, рекламных кампаниях, источниках трафика, продажах и платежах.

Задача: провести очистку и анализ этих данных, выявить слабые и сильные стороны бизнеса, оценить эффективность маркетинга и продаж, а также предложить точки роста.

Особенность проекта заключается в том, что он разделён на две части:

1. **Python DA (Data Analysis)** — технический анализ CRM-данных с использованием Python (pandas, matplotlib, seaborn и других библиотек).
Результат: чистый датасет, статистика, визуализации и аналитические выводы.
2. **Продуктовая аналитика** — поиск бизнес-точек роста на основе данных, расчёт юнит-экономики, построение дерева метрик, формулирование гипотез и описание механики их проверки.

Цель проекта

Главная цель: **повысить эффективность работы онлайн-школы программирования через анализ CRM-данных и формулирование продуктовых гипотез.**

Для достижения цели необходимо:

- Очистить данные и подготовить их к анализу.
- Провести исследовательский анализ (EDA).
- Изучить ключевые показатели: количество сделок, конверсии, источники лидов, эффективность кампаний, работу отдела продаж, платежи, продукты и географию.
- Найти слабые места и аномалии (потери на воронке, неэффективные компании, пропуски в данных).
- Рассчитать показатели юнит-экономики.
- Определить метрики и построить их иерархию.
- Сформулировать гипотезы по росту и описать условия их тестирования.

План работы

Работа над проектом разделена на два блока.

Блок 1. Python DA

1. [Подготовка и очистка данных](#)
 - Удаление дубликатов и неактуальных столбцов.
 - Обработка пропусков и ошибок.
 - Приведение типов данных.
2. [Описательная статистика](#)
 - Расчет базовой статистики для числовых полей.
 - Анализ категориальных переменных.
3. [Анализ временных рядов](#)
 - Динамика сделок и звонков во времени.
 - Средняя длительность сделок.
4. [Анализ эффективности кампаний и источников](#)
 - Сравнение кампаний по лидам и конверсии.
 - Эффективность источников (Source).
5. [Анализ работы отдела продаж](#)
 - Эффективность отдельных владельцев сделок.
 - Роль кампаний.
6. [Анализ платежей и продуктов](#)
 - Типы оплаты и их связь с успешностью.
 - Популярные продукты и типы обучения.
7. [Географический анализ](#)
 - Распределение сделок по городам.
 - Роль уровня немецкого языка.

Блок 2. [Продуктовая аналитика](#)

1. [Расчёт юнит-экономики по продуктам.](#)
2. [Определение точки роста бизнеса.](#)
3. [Построение дерева метрик.](#)
4. [Формулировка продуктовых гипотез.](#)
5. [Описание механики тестирования \(с учетом ограничения \$\leq 2\$ недели\).](#)

1. Подготовка и очистка данных

Данные были загружены из CRM-системы онлайн-школы программирования. На этом этапе главная задача заключалась в том, чтобы привести их к корректному и единообразному виду, устранить ошибки и пропуски, чтобы в дальнейшем можно было проводить анализ без искажений.

Очистка выполнялась в GoogleColab с использованием **pandas**.

1.1 Общие преобразования

1. Приведение формата дат

- Все столбцы с датами (*Created Time*, *Closing Date*, *Call Start Time*, *Date* и др.) были приведены к формату `datetime64[ns]`.
- Это позволило:
 - корректно строить временные ряды,
 - рассчитывать длительность сделок

2. Категориальные переменные

- Столбцы с фиксированными наборами значений (*Stage*, *Source*, *Product*, *Education Type*, *Call Type*, *City* и др.) преобразованы в тип `category`.
- Это уменьшило объём памяти и позволило строить более быстрые группировки.

3. Числовые переменные

- Столбцы с количественными показателями (*Impressions*, *Clicks*, *Call Duration (in seconds)*, *Initial Amount Paid*, *Offer Total Amount*) приведены к типам `int` или `float`.
- Это обеспечило возможность расчётов и построения статистики без ошибок.

1.2 Очистка таблицы *Contacts*

- Дубликаты и полностью пустые строки удалены
- *Created Time* и *Modified Time* приведены к типу `datetime64[ns]`
- *Contact Owner Name* очищено от лишних пробелов и приведено к `category`
- Некорректное значение `False` заменено на `"Unknown"`

Структура данных:

Содержит 4 поля: *Id*, *Contact Owner Name*, *Created Time*, *Modified Time*.

1.3 Очистка таблицы *Calls*

- Дубликаты и пустые строки удалены.
- *Call Start Time* приведено к `datetime64[ns]`.

- Все категориальные поля (Call Owner Name, Call Type, Call Status, Outgoing Call Status, Scheduled in CRM) приведены к типу `category`, пропуски заменены на `"Unknown"`
- Scheduled in CRM: нормализация 0/1 → NO/Yes.
- Call Duration (in seconds): числа очищены и приведены к `int64`; пропуски заменены:

-- 0 для статусов, где длительность не имеет смысла (Missed, Cancelled и др.),

-- медиана по остальным звонкам.

- Добавлен флаг `_missing_contact` для звонков без CONTACTID.

Структура данных:

Ключевые поля: Id, CONTACTID, Call Start Time, Call Owner Name, Call Type, Call Status, Call Duration (in seconds), Outgoing Call Status, Scheduled in CRM.

1.4 Очистка таблицы *Spend*

- **Удалены дубликаты**
 - Обнаружено 917 дубликатов, они были удалены и пустые строки удалены
- **Date**
 - Приведена к `datetime64[ns]`.
- **Campaign, Source, AdGroup, Ad**
 - Приведены к `category`.
- **Impressions / Clicks / Spend**
 - Очищены от лишних символов и Приведены к `int64`.

Структура данных

Ключевые поля: Date, Source, Campaign, Impressions, Clicks, Spend, а также AdGroup и Ad

1.5 Очистка таблицы *Deals*

Структура данных

Таблица Deals содержит ключевые бизнес-поля: ID сделки, владельца, стадии (Stage), качество (Quality), суммы (Initial Amount Paid, Offer Total Amount), даты (Created Time, Closing Date), SLA, уровень языка, продукт и пр.

Дубликаты и пустые строки

- Все дубликаты и полностью пустые строки удалены → в выборке каждая сделка представлена единожды.
- **Deal Owner Name**
 - Обнаружены пропуски.
 - Решение: заменить на `"Unknown"`, чтобы не терять данные и сохранить возможность группировки.
- **Даты**
 - Created Time всегда заполнено → корректная база для анализа динамики

- BClosing Date может быть NaT (сделка ещё открыта). Это норма для CRM: такие записи помечены флагом `_open_deal`.
- **Quality**
 - Пустые значения заменены на `"Unknown"`.
 - Тип приведён к `category`.
- **Lost Reason**
 - Для сделок со статусом `Lost`, где причина не указана, значение заменено на `"Unknown"`
 - Это важно для анализа причин отказов.
- **Campaign / Source / Product**
 - Пропуски заменены на `"Unknown"`.
 - Тип данных — `category`.
- **Числовые поля**
 - Course duration, Months of study приведены к `int`
 - Пропуски заменены на 0 (означает отсутствие данных)
 - Суммы (Initial Amount Paid, Offer Total Amount) приведены к `float`, очищены от символов валюты и форматирования.
- **Initial Amount Paid / Offer Total Amount**
 - Проверяет все сделки на стадии **"Payment Done"** (оплата проведена).
 - Если в таких сделках сумма платежа или сумма предложения (Initial Amount Paid, Offer Total Amount) равна **0**, то это считается ошибкой данных (ведь у оплаченной сделки не может быть нуля).
 - В этих случаях мы подставляем **медианное значение** по всем корректным сделкам с оплатой.
 - Медиана выбрана, потому что она устойчива к выбросам и лучше отражает «типичный» платёж, чем среднее.

важно:

- Значения NaN **оставляем** без изменений.
Это даёт возможность в аналитике показать:
 - сколько сделок имеют корректные суммы,
 - сколько записей ошибочно содержат 0,
 - сколько сделок вообще не содержат данных об оплате.

Почему так:

- Если заменить NaN тоже, то потеряем возможность показать проблемы CRM (например, когда сделка помечена как оплаченная, но суммы просто не внесены).

Payment/Offer Categories

- Классификация строится по следующим правилам:
 - `"Unknown"` → значение отсутствует (NaN), данные не внесены в CRM.
 - No Payment / No Offer → значение равно 0, то есть клиент ничего не заплатил или не получил предложение.
 - Demo Access → символическая сумма (меньше 10), например 1, 9 — это демо-доступ или тестовая оплата.

- Regular Payment / Regular Offer → значения больше 10, то есть стандартные платежи и офферы.

Зачем:

- Это позволяет разделить клиентов на разные категории и анализировать:
 - долю сделок без оплат,
 - долю демо-студентов,
 - долю полноценных платёжных клиентов,
 - долю сделок, где в CRM отсутствуют данные.
- Такая категоризация делает данные более интерпретируемыми и полезными для анализа воронки и качества лидов.

Очистка поля *City*

- В поле *City* встречались некорректные записи (полные адреса вместо города, пустые значения, ошибки формата).
- Примеры исправлений:
 - "Karl-Liebknecht str. 24, Hildburghausen, Thüringen" → "Thüringen".
 - "Poland , Gdansk , Al. Grunwaldzka 7, ap. 1a" → "Gdansk".
 - Пустые значения и дефисы → "Unknown".

Очистка поля *Level of Deutsch*

- В исходных данных встречались разные обозначения уровня языка: латиница и кириллица (A1 / A1, B1 / Б1, C1 / C1).
- Решение:
 - Привести всё к верхнему регистру.
 - Кириллицу заменить на латиницу.
 - Разрешённые значения: A1, A2, B1, B2, C1, C2.
 - Всё остальное заменено на "Unknown".

SLA

Поле SLA приведено к формату Timedelta, рассчитаны часы (_SLA_hours). Теперь можно анализировать время реакции менеджеров

Итог подготовленного датасета

- **Calls:** 95,874 строк, 9 столбцов.
- **Spend:** 19,862 строк, 6 столбцов.
- **Deals:** 21,593 строк, 20 столбцов.
- **Contacts:** 18,548 строк, 4 столбца.

Все данные приведены к единому формату, очищены от дубликатов, некорректных и пустых значений.

В итоге структура связей такая:

Spend → Deals → Contacts → Calls

Spend даёт рекламные затраты,

Deals показывает сделки (с привязкой к кампаниям и контактам),

Contacts связывает всё с конкретными людьми,

Calls фиксирует звонки с этими людьми.

Почему не удаляем строки без CONTACTID?

Чтобы не терять данные для операционных метрик по звонкам (нагрузка на менеджеров, распределение статусов). Для сквозной аналитики используем фильтр CONTACTID.notna().

Почему в Closing Date оставляем NaT?

Потому что открытые сделки не закрыты — это норма. Подставлять фиктивные даты нельзя: ломает длительности и тренды.

Зачем делать категории (astype('category'))?

Быстрее группировки/агрегации, меньше память, чёткая типизация категориальных признаков.

Почему медиана, а не среднее для сумм?

Медиана устойчивее к выбросам (одно большое значение не уводит результат).

Почему заполняем пропуски именно 'Unknown'?

****1. Категориальные поля ****(Stage, Quality, Source, Payment Type, City и т.п.)

- Если оставить NaN, то при группировке Pandas/Excel/BI-системы будет создавать отдельную «пустую» категорию, и часто такие строки будут выпадать из расчётов.
- Заполнив 'Unknown', мы сохраняем строку в выборке и явно говорим: «Информация отсутствует».
- Так проще анализировать: можно всегда видеть, какая доля данных не заполнена.

2. Унификация

- Вместо того чтобы часть пропусков была NaN, часть None, часть пустая строка " → всё сведено в одно значение 'Unknown'.
- Это упрощает код (нет необходимости каждый раз проверять разные типы пропусков).

3. Корректность джойнов

- В join-ах NaN не сопоставляются ни с чем → такие строки теряются.
- 'Unknown' сопоставляется с 'Unknown' → в merge строки сохраняются.

4. Прозрачность для бизнеса

- При отчёте, вместо пустого поля написано 'Unknown', это лучше читается.
- Можно сразу объяснить: «Вот процент данных, где не указано».

Важный нюанс

- Для ключевых полей (например, Id, CONTACTID) мы не ставили 'Unknown', а оставляли NaN + добавляли флаг. Потому что Unknown как ID может ввести в заблуждение при джойне.
- 'Unknown' — это стратегия именно для категориальных признаков, которые используются в аналитике (Stage, Source, City и т.п.)

Логика

- В категориальных колонках мы ставили 'Unknown', чтобы:
 - строки не терялись в анализе,
 - группировки работали корректно,
 - было видно долю пропусков.
- В ключевых колонках (Id, ContactID, Contact Name) мы NaN не трогали, чтобы не ломать джойны.

2. Описательная статистика

Цель - провести базовый исследовательский анализ (EDA), чтобы понять общие характеристики данных, выявить распределения, а также заметить аномалии и потенциальные проблемы.

2.1 Числовые переменные

Ход работы:

- Для количественных переменных (Initial Amount Paid, Offer Total Amount, Course duration, Months of study, Impressions, Clicks, Spend, Call Duration (in seconds)) были рассчитаны:
 - Среднее (mean)
 - Медиана (median)
 - Мода (mode)
 - Минимум и максимум
 - Диапазон (max – min)

- Стандартное отклонение (std)

Наблюдения и выводы:

- В большинстве полей среднее оказалось **значительно выше медианы** → распределение скошено вправо (наличие выбросов).
- **Initial Amount Paid** и **Offer Total Amount** содержат нули и редкие крупные значения, которые искажают среднее.
- **Call Duration (in seconds)** показывает, что **медиана звонков очень мала** (несколько секунд), тогда как среднее выше ($\approx 2\text{--}3$ минуты). Это указывает на большое количество коротких или неуспешных звонков и небольшую долю долгих (до 2 часов).
- **Impressions** и **Clicks** распределены крайне неравномерно: большинство компаний имеют низкие значения, а несколько компаний выбиваются с аномально высокими.

Интерпретация для бизнеса:

- Метрики, чувствительные к выбросам, лучше оценивать по медиане.
- Необходим дополнительный анализ аномально долгих звонков: возможно, это либо технические ошибки, либо сложные случаи клиентов.
- Эффективность рекламных кампаний стоит анализировать после фильтрации выбросов, так как единичные крупные кампании сильно искажают общую картину.

2.2 Категориальные переменные

Ход работы:

- Для переменных (**Stage**, **Source**, **Product**, **Education Type**, **Payment Type**, **Quality**, **Lost Reason**, **City**, **Level of Deutsch**) были построены частотные распределения.

Наблюдения:

- **Stage (Стадия сделки):**
 - Большая часть сделок находится в статусе **Lost**.
 - Доля **Payment Done** относительно невелика (успешных сделок мало).
 - Основные потери происходят на ранних этапах (**First Contact**, **Call Delayed**).
- **Source (Источник):**
 - 2–3 источника дают основную часть лидов.
- **Product (Продукт):**

- Основные курсы: *Digital Marketing, UX/UI Design, Web Developer*.
- Много пропусков → это снижает возможность полноценной оценки.
- Есть много «Unknown» → проблема качества данных.
- **Payment Type (Тип оплаты):**
 - Наиболее частые: *Recurring* и *One Payment*.
 - Успешность сделок выше у *One Payment*.
- **Lost Reason (Причина потери):**

Total Lost Deals: 15 743

ТОП-причины

- **Doesn't Answer — 4 074 (~25.9%)** → главная проблема на этапе первичного контакта.
- **Changed Decision — 2 122 (~13.5%).**
- **Duplicate — 1 746 (~11.1%).**
- **Non target — 1 736 (~11.0%).**
- **Stopped Answering — 1 556 (~9.9%).**
- **Invalid number — 1 460 (~9.3%).**
- Остальные причины занимают <4% каждая (Expensive, Conditions not suitable, Needs time to think и др.).

Группы причин

- **Коммуникация (~35.8%):** Doesn't Answer + Stopped Answering → проблема взаимодействия с лидами.
- **Качество лидов (~34%):** Duplicate + Non target + Invalid number → слабый таргетинг и некачественные контакты.
- **Оффер и цена (~11.1%):** Expensive + Conditions not suitable + Needs time to think → необходимость доработки ценностного предложения.
- **Барьеры (~1–2%):** Does not speak English, Does not know how to use a computer → отдельные сегменты требуют иной коммуникации.
- **Level of Deutsch (Уровень немецкого):**
 - Основные уровни: A2, B1, B2.
 - A1 чаще встречается в городах с низкой успешностью сделок.

Интерпретация для бизнеса:

- Основной барьер — **низкая конверсия на ранних стадиях** → стоит улучшать процесс сопровождения после первого контакта.
- Доля *Unknown* в *Product* мешает правильной аналитике → нужно доработать систему учёта.

- Причины потерь показывают явную проблему коммуникации: много лидов теряются, потому что не отвечают → стоит внедрить повторные напоминания, автосообщения.
- Языковой барьер (низкий уровень немецкого) заметно влияет на успешность сделок в отдельных регионах.

Рекомендации

1. **Коммуникация:** ввести многошаговый процесс контакта (звонок → мессенджер → email → повторный звонок), авто-напоминания, контроль скорости ответа (первые 15–30 минут после заявки).
2. **Качество лидов:** включить автоматическую проверку номеров, дедупликацию заявок, более строгий таргетинг и фильтры в форме заявки.
3. **Оффер и цена:** подготовить FAQ по стоимости и рассрочке, показывать примеры успеха выпускников, обучить менеджеров работе с возражениями.
4. **CRM-практика:** оптимизировать список причин «Lost» (убрать дубли, контролировать обязательность заполнения). Основной вывод: почти 2/3 всех потерь вызваны либо отсутствием контакта, либо плохим качеством лида. Улучшение коммуникации и фильтрации контактов может дать значительный прирост конверсии.

2.3 Корреляционный анализ

Ход работы:

- Построена корреляционная матрица по числовым переменным.
- Визуализирована через heatmap (`seaborn.heatmap`).

Основные результаты:

- Сильная положительная корреляция между `Course duration` и `Offer Total Amount` → чем длиннее курс, тем выше стоимость.
- Средняя корреляция между `Initial Amount Paid` и `Offer Total Amount` → чем выше предоплата, тем дороже общий курс.
- Слабая положительная корреляция между `Months of study` и `Initial Amount Paid`.

Интерпретация:

- Студенты, выбравшие более длинные программы, чаще платят выше.
- Однако слабая корреляция между количеством месяцев обучения и предоплатой может указывать, что не все студенты доходят до конца курса.

2.4 Метрики рекламной эффективности (CTR, CPC)

Что считалось:

- **CTR (Click-Through Rate)** = (Clicks / Impressions) × 100
- **CPC (Cost Per Click)** = Spend / Clicks

Результаты:

- Средний CTR ≈ **0.98%** (ниже нормы 1–2%).
- CPC ≈ **0.30 €** (низкий, но кликов мало).
- Распределения сильно скошены: большинство кампаний с нулевыми или минимальными значениями, а несколько аномально крупных «вытягивают» средние.
- **Facebook Ads** — лидер по затратам (58% бюджета), но не по эффективности.
- **Google Ads, YouTube Ads, Organic** генерируют значимую долю кликов и сделок.

Выводы:

- CTR низкий → требуется оптимизация рекламных креативов и сегментации целевой аудитории.
- Большие кампании с сотнями тысяч показов и минимальными кликами неэффективны.
- Сильная зависимость от Facebook Ads — риск для бизнеса
- Есть потенциал в более дешёвых и эффективных каналах (Organic, Webinar)

Итог

Этап описательной статистики показал:

- Данные содержат выбросы и перекосы в распределениях.
- Наиболее проблемные места:
 - низкая конверсия воронки, потери на ранних стадиях
 - высокие доли **Unknown** в продуктах,
 - слабая валидация контактов,
 - низкий CTR кампаний
- Основные точки роста: улучшение качества данных, оптимизация рекламных каналов и работа над коммуникацией с клиентами.

Все эти выводы будут использованы в дальнейшем анализе временных рядов и эффективности кампаний.

3. Анализ временных рядов

Цель – изучить динамику сделок и звонков во времени, а также изучить длительность сделок от момента создания до их закрытия.

3.1 Тенденции создания сделок во времени и их связь со звонками

Ход работы:

- Построены временные ряды по:
 - количеству созданных сделок (**Created Time**),
 - количеству закрытых сделок (**Closing Date**),
 - количеству звонков (**Call Start Time**).
- Для наглядности данные агрегированы по месяцам.
- Для корректности учитывались только завершённые звонки (**Call Duration > 0**).
- Построен линейный график, отображающий динамику сделок и звонков во времени.

Результаты:

- Количество звонков растёт с июля 2023 ($\approx 1\,900$ звонков) до апреля 2024 (пик $\approx 13\,300$ звонков), затем снижается ($\approx 8\,500$ звонков в июне 2024).
- Количество сделок также растёт: от 655 в июле 2023 до пика 3 081 в марте 2024, после чего падает до 1 674 в июне 2024.
- Динамика звонков и сделок схожа → можно говорить о прямой зависимости.

Интерпретация:

- Активность отдела продаж напрямую влияет на количество сделок.
- Спад в звонках в мае–июне 2024 приводит к снижению сделок. Это подтверждает, что звонки являются ключевым драйвером воронки продаж.

3.2 Конверсия сделок (Deals / Calls)

Ход работы:

- Рассчитана метрика:
$$\text{Conversion} = \frac{\text{Количество сделок}}{\text{Количество звонков}}$$
- Построена динамика конверсии по месяцам.

Результаты:

- В начальных месяцах конверсия выше (так как звонков меньше, а доля успешных сделок выше).
- С ростом количества звонков конверсия постепенно снижается.
- В последние месяцы показатель стабилизируется, но остаётся на более низком уровне, чем в начале.

Вывод:

- При высоком объеме звонков менеджеры не успевают качественно обрабатывать все контакты → конверсия падает.
- Необходим баланс: количество звонков должно сопровождаться качественными процессами работы с лидами.

3.3 Продолжительность закрытия сделок

Ход работы:

- Для закрытых сделок рассчитана длительность: $\text{Closing Date} - \text{Created Time}$.
- Использованы два варианта расчёта:

Вариант А – в календарных днях (same-day = 0).

Вариант В (inclusive) – включая день закрытия (same-day = 1).

- Для защиты от ошибок CRM, если $\text{Closing Date} < \text{Created Time}$, значение принудительно установлено в 0.
- Построено распределение длительности сделок (гистограмма) с линиями медианы и среднего.

Результаты:

- Медианная длительность сделки невелика (показывает типичный случай).
- Среднее значение выше медианы → наличие сделок с очень длинным периодом закрытия.
- Распределение скошено вправо: большинство сделок закрывается быстро, но есть выбросы, которые тянутся месяцами.

Интерпретация:

- В целом сделки закрываются за сравнительно короткий срок, но отдельные «долгие» сделки указывают на неэффективность работы с определенными клиентами.

- Для оптимизации необходимо выявить причины задержек и проверить, какие типы клиентов или продуктов чаще «зависают» в системе.

3.4 Тепловая карта

Основные наблюдения:

- Большинство сделок закрываются быстро Основная масса сделок завершается в срок 3–30 дней. Это видно на диагонали карты: чем ближе месяц создания к месяцу закрытия, тем меньше средняя длительность.
- Есть затянутые процессы В ячейках с длительностью 90–150 дней заметны сделки, которые «зависали» на несколько месяцев (особенно для сделок, созданных осенью 2023 и закрытых весной 2024).
- Аномальные хвосты Отдельные сделки тянутся 200–335 дней. Они единичные, но сильно искажают картину. Возможные причины:

-- ошибка при закрытии в CRM (сделка оставалась открытой слишком долго);

-- «замороженные» сделки, возобновлённые позже;

-- дубликаты или некорректные данные.

- Положительная динамика Для сделок, созданных в 2024 году, средняя длительность заметно снижается (часто 2–12 дней). Это может указывать на улучшение процессов работы с клиентами.

Итог анализа временных рядов

1. Количество звонков и сделок движутся синхронно → подтверждает зависимость бизнеса от активности отдела продаж.
2. При большом объёме сделок длительность их закрытия увеличивается → узкое место в ресурсе команды.
3. Конверсия падает при росте нагрузки → нужно искать баланс качества и количества.

Рекомендации:

- Перераспределять ресурсы в пиковые месяцы (март, апрель, ноябрь).
- Рассмотреть возможность увеличения штата/автоматизации в период высокого спроса.
- Поддерживать маркетинговую активность в начале года (январь–февраль), чтобы снизить сезонные провалы.

4. Анализ эффективности кампаний и источников

Цель: оценить результативность маркетинговых кампаний и источников трафика с точки зрения генерации лидов и успешных сделок, а также определить наиболее эффективные каналы.

4.1 Эффективность рекламных кампаний

Ход работы:

- Данные по кампаниям сгруппированы.
- Для каждой кампании рассчитаны:
 - общее количество лидов,
 - количество успешных сделок (**Stage = Payment Done**),
 - коэффициент конверсии (успешные / все лиды).
- Для удобства визуализации построены графики с распределением кампаний по лидерам и аутсайдерам.

Результаты:

- Существует множество кампаний, однако значительную часть лидов и успешных сделок дают только несколько из них.
- В топ-кампаниях одновременно фиксируется высокий объём лидов и более высокий процент конверсии.
- Есть компании с огромным числом лидов, но низкой конверсией → они тянут бюджет, но не дают результата.

Выводы:

- Компания `performancemax_digitalmarkt_ru_DE` — лидер по объёму лидов (>2600), но конверсия низкая (4.2%). Это «лидогенератор», но с большими потерями.
- Кампании `youtube_shorts_DE` и `wide_DE` также показали высокий объём (1600+), но конверсия ещё ниже (3–3.2%).
- Есть кампании с малым объёмом, но более высокой конверсией (>5%), например `brand_search_eng_DE`, что указывает на более точное попадание в аудиторию.

Рекомендации:

- Масштабировать кампании с высокой конверсией даже при меньших объёмах (ниша, точное попадание).
- Оптимизировать лидогенераторы с низкой конверсией: улучшить сегментацию, корректировать креативы и работу с лидами.

- Убрать/заморозить кампании с крайне низкой конверсией (<3%) и низким объёмом.
- Поддерживать баланс: часть бюджета направлять на широкие кампании ради объёма, часть — на точные ради качества.
- Проводить A/B тесты креативов для крупных кампаний с низкой конверсией.

4.2 Эффективность источников (Source)

Ход работы:

- Для каждого источника рассчитаны:
 - общее количество лидов,
 - количество успешных сделок,
 - коэффициент конверсии.
- Построены графики сравнения источников по числу лидов и уровню успешности.

Результаты:

- Наибольший объем лидов приносят источники (Facebook Ads, Google Ads).
- Лидеры по количеству успешных сделок: Facebook Ads, Google Ads.
- Источник Organic показывает более высокую конверсию при меньшем объеме → это качественный канал с низкой стоимостью.
- Webinar также демонстрирует высокий процент конверсии при относительно небольшой выборке.
- SMM — тоже неплохая конверсия (5.3%) при объеме ~1700 лидов

Выводы:

- Школа сильно зависит от Facebook Ads (основной бюджет и поток лидов). Это риск, так как эффективность кампаний нестабильна.
- Organic и Webinar показывают лучшую конверсию → перспективные каналы для инвестиций.
- Источники обозначены как Unknown, мешают полноте анализа → необходимо улучшить CRM-трекинг.
- Инвестировать в SMM, так как у него оптимальный баланс объёма и конверсии.
- Источники с конверсией <3% анализировать и при необходимости отключать.

4.3 Итоговые рекомендации

- Внедрить систему валидации и маркировки источников в CRM, чтобы сократить долю Unknown.
- Проводить A/B тестирование кампаний для выявления наиболее результативных креативов.
-

Кампании:

- Оптимизировать крупные низкоконтверсионные кампании.
- Масштабировать точечные кампании с >5% конверсии.
- Постоянно мониторить и пересматривать стратегию «широкие vs нишевые».

Источники:

- Основные лидогенераторы — Facebook и Google.
- Лидеры по качеству — Organic, Webinar, SMM.
- Оптимизировать бюджет в пользу более эффективных источников.

5. Анализ работы отдела продаж

Цель: оценить эффективность сотрудников (владельцев сделок) и кампаний с точки зрения количества обработанных сделок, коэффициента конверсии и общей суммы продаж.

5.1 Эффективность владельцев сделок (Deal Owner Name)

Ход работы:

- Для каждого владельца сделки рассчитаны:
 - количество обработанных сделок,
 - количество успешных сделок (**Stage = Payment Done**),
 - коэффициент конверсии,
 - общая сумма продаж (по **Offer Total Amount**).
- Построены таблицы и диаграммы для сравнения сотрудников.

Leads volume:

- Charlie Davis (2,963 leads) and Julia Nelson (2,241 leads) лидеры по количеству обработанных сделок.
- Однако их conversion rate сравнительно низкий (4–5%), что снижает общую эффективность.

High conversion:

- Oliver Taylor показывает лучший результат по конверсии (30.7%) при сравнительно небольшом объёме (163 leads).
- Kevin Parker (6.9%) и Ulysses Adams (6.5%) также выше среднего уровня (~4–5%).

Weak performance:

- Несколько сотрудников имеют очень низкую конверсию 1%.

Результаты:

- Распределение неравномерное: небольшая группа сотрудников обрабатывает значительно больше сделок, чем остальные.

- Среди активных сотрудников есть те, кто показывает высокую конверсию и большие суммы продаж, а есть и те, кто при большом объеме лидов даёт низкий результат.
- Часть сотрудников обрабатывают мало сделок и имеют низкий коэффициент конверсии, вплоть до 0%.

Интерпретация:

- Основная нагрузка и результаты бизнеса сосредоточены у ограниченного числа сотрудников.
- Есть менеджеры с высокой эффективностью (высокий % успешных сделок и большая выручка).
- Есть сотрудники с низкой результативностью → требуется дополнительное обучение или пересмотр обязанностей.

5.2 Эффективность рекламных кампаний (в контексте продаж)

Ход работы:

- Для кампаний рассчитаны:
 - количество лидов,
 - количество успешных сделок,
 - коэффициент конверсии,
 - общая сумма продаж.
- Сравнены топ-кампании и слабые кампании.

Результаты:

By leads volume:

- performancemax_digitalmarkt_ru_DE (2,653 leads) – главный генератор лидов, но низкая конверсия (4.2%).
- youtube_shorts_DE (1,635 leads) и wide_DE (1,575 leads) — также лидеры по количеству, но конверсия лишь 3–3.2%.

By conversion rate (≥100 leads):

- brand_search_eng_DE (8.9%), wide_DE (02.07.23) (5.3%), LAL_DE (5.2%).
- Эти кампании привлекают меньше лидов, но качество выше.

Интерпретация:

- Не количество лидов, а их качество определяет ценность кампании.
- Рекламные вложения не всегда напрямую связаны с результатом — важнее эффективность в терминах продаж.

5.3 Общая оценка работы отдела продаж

1. Сильные стороны:

- Несколько менеджеров стабильно обеспечивают высокие результаты.

- Часть кампаний работает эффективно, давая и лиды, и продажи.

2. Слабые стороны:

- Сильная нагрузка на ограниченный круг сотрудников.
- Наличие неэффективных менеджеров и кампаний.
- Отсутствие равномерного распределения сделок внутри отдела.

5.4 Рекомендации

- Провести **внутренний аудит отдела продаж**: выявить лучших сотрудников и перенять их практики.
- **Обучить** менеджеров с низкой результативностью или перераспределить их задачи.
- Внедрить **более равномерное распределение сделок** между сотрудниками, чтобы снизить перегрузку отдельных менеджеров.
- Сосредоточить маркетинговый бюджет на кампаниях, которые показывают высокую конверсию и выручку.
- Использовать дашборды для регулярного мониторинга KPI каждого сотрудника и кампании.
- Увеличить бюджеты для brand_search_eng_DE, LAL_DE.
- Снять нагрузку с Davis/Nelson и передать часть лидов более эффективным сотрудникам.
- Провести обучение для менеджеров с низкой конверсией.
- Использовать успешные кейсы для увеличения среднего чека в других командах.

6. Анализ платежей и продуктов

Цель: изучить распределение типов оплаты, их влияние на успешность сделок, а также проанализировать популярность и результативность различных продуктов и форматов обучения.

6.1 Распределение типов оплаты

Ход работы:

- Взято поле **Payment Type**.
- Для каждой категории рассчитано:
 - количество сделок,
 - количество успешных (**Stage = Payment Done**),
 - коэффициент конверсии.
- Построено распределение по типам оплаты.

Результаты:

- Основные способы оплаты:
 - **Recurring Payments (рассрочка / регулярные платежи)** — самая многочисленная категория.

- **One Payment (разовый платеж)** — менее частая, но с более высоким процентом успешных сделок.
- **Reservation (предоплата/бронь)** встречается редко и имеет низкую конверсию.
- Конверсия у **One Payment** выше, чем у **Recurring**.

Интерпретация:

- Основной спрос приходится на **recurring payments**, что подтверждает доверие клиентов к долгосрочному обучению.
- Высокая конверсия **one payment** говорит о важности наличия этого варианта для гибкости.
- Формат **reservations** неэффективен — низкий объём и слабая конверсия.

Рекомендации

- Сделать упор на продвижение **recurring payments** как основного бизнес-моделя.
- Сохранить **one payment** как дополнительный привлекательный вариант для клиентов.
- Рассмотреть возможность отказа от **reservations** или трансформацию этого формата.

6.2 Популярность и успешность продуктов

Ход работы:

- Для поля **Product** рассчитано:
 - количество сделок,
 - количество успешных,
 - коэффициент конверсии,
 - суммарная выручка (**Offer Total Amount**).
- Сравнены топ-продукты и менее популярные.

Результаты:

- Лидеры по популярности:
 - **Digital Marketing** — наибольшее количество сделок и высокая выручка (1990 сделок, конверсия 23,8%).
 - **UX/UI Design** — устойчиво популярный продукт (1 022 сделки, конверсия 22,4%)
 - **Web Developer** — стабильный, но менее прибыльный по сравнению с Digital Marketing.
- Продукты с низкой эффективностью:
 - **Data Analytics** и **Find yourself in IT** почти не генерируют доход.

Интерпретация:

- Школа зарабатывает на 3 «якорных» продуктах (Digital Marketing, UX/UI, Web Developer).
- Остальные продукты требуют либо переработки, либо снятия с продаж.
- Отсутствие данных по части сделок (**Unknown**) мешает полноценно оценить портфель продуктов.

6.3 Анализ по типам обучения

Ход работы:

- Для каждого продукта рассмотрены 2 формата обучения (**Morning, Evening**).
- Рассчитано количество сделок и выручка по каждому формату.

Результаты:

- **Утренние курсы** доминируют по всем продуктам:
 - Digital Marketing — 1 533 сделки
 - UX/UI Design — 808 сделок
 - Web Development — 545 сделок
- **Вечерние курсы** показывают значительно меньшие результаты (например, Web Development — всего 1 сделка).

Интерпретация:

- Утренние программы являются **основным источником дохода** и успешных сделок.
- Вечерние программы существенно отстают, особенно по направлению **Web Development**.

Рекомендации

- Активнее продвигать **вечерние курсы** как решение для занятых специалистов (гибкость, смена профессии).
- Пересмотреть формат и расписание вечерних занятий для повышения привлекательности.
- Сохранять приоритет на **утренних курсах**, как на ключевом драйвере продаж.

6.4 Итоговые выводы

1. Основной вклад в выручку дают продукты **Digital Marketing, Web Development** и **UX/UI Design**.
2. Утренние курсы работают значительно лучше вечерних.
3. Высокая доля **Unknown** в продуктах и типах оплаты искажает аналитику → нужно доработать CRM.

6.5 Рекомендации

- Дальше продвигать **Digital Marketing** и **UX/UI Design** — они наиболее прибыльные.
- Провести аудит продуктов с низкой выручкой (**Data Analytics, Find yourself in IT**).
- Рассмотреть корректировку вечерних программ (например, удобное время, дополнительные опции).

- Усилить контроль качества данных в CRM, чтобы минимизировать **Unknown**.

7. Географический анализ

Цель: изучить распределение сделок по городам, а также влияние уровня владения немецким языком на успешность сделок в разных регионах.

7.1 Распределение сделок по городам

Ход работы:

- Взято поле **City**.
- После очистки данные сгруппированы по городам.
- Для каждого города рассчитано:
 - количество сделок,
 - количество успешных (**Stage = Payment Done**),
 - коэффициент конверсии.
- Построены таблицы и графики распределения сделок по ключевым городам.

Результаты:

- Топ-10 городов формируют основную часть сделок.
- Берлин лидирует по объему: 182 сделки (43% конверсии).
- Мюнхен, Гамбург, Лейпциг также показывают значительные объемы и конверсии выше среднего уровня.
- Есть города с минимальным числом сделок, где оценивать конверсию статистически некорректно.

Интерпретация:

- Школа сосредоточена в нескольких ключевых городах → это риск зависимости от ограниченного числа регионов.
- Потенциал роста виден в других крупных городах (Франкфурт, Кёльн, Дюссельдорф).

7.2 Влияние уровня немецкого языка (Level of Deutsch)

Ход работы:

- Для каждого города дополнительно рассчитано распределение по полю **Level of Deutsch**.
- Построены графики успешности сделок в зависимости от уровня языка.

Результаты: (рассматриваем топ 10 городов)

- В городах с большим количеством успешных сделок чаще встречаются клиенты с уровнем **A2–B2**.
- Города, где уровень **A1** нулевую конверсию.

Интерпретация:

- Владение немецким языком напрямую влияет на вероятность успешного завершения сделки.
- В городах, где доля слабого уровня языка выше, необходимо адаптировать коммуникации.

7.3 Итоговые выводы

1. Лидеры по количеству сделок — Берлин, Мюнхен, Гамбург.
2. Города с доминированием A1–A2 показывают низкую успешность.
3. B1–B2 — «золотая середина», на которую приходится большинство успешных сделок

7.4 Рекомендации

- Усилить маркетинговое присутствие в **других крупных городах** Германии (Франкфурт, Кёльн, Дюссельдорф).
- Ввести дополнительные инструменты для клиентов с низким уровнем немецкого
- Сосредоточиться на сегменте B1–B2 как на основной целевой аудитории.
- Продумать стратегию выхода за пределы топ-городов, чтобы снизить региональную зависимость.

Блок 2. Продуктовая аналитика

Цель: найти точки роста бизнеса онлайн-школы на основе расчёта юнит-экономики и анализа продуктовых метрик, сформулировать гипотезы для улучшения ключевых показателей и описать механику их тестирования.

1. Расчёт юнит-экономики по продуктам

1. Юнит-Экономика по продуктам:												
	number of potential customers	converting potential customers into paying customers	number of customers received	average order value			average payment count per one customer	gross profit per customer (customer lifetime value)	gross profit per lead	customer acquisition cost (UA acquisition)	acquisition buget	contribution margin
		c=b/ua		aov=revenue/t			apc=t/b	cltv=aov*apc	ltv=cltv*c1	cpa=ac/ua		cm=ua*(ltv-cpa)
Product	UA	C1	B	AOV	Revenue	T	APC	CLTV	LTV	LTC/LCPA	AC	CM
Web Developer	18548	0.74%	137	€660.51	€348,746.67	528	3.85	€2,545.60	€18.80	€8.06	€149,523.45	€199,223.22
Digital Marketing	18548	2.55%	473	€1,180.56	€2,148,610.91	1820	3.85	€4,542.52	€115.84	€8.06	€149,523.45	€1,999,087.46
UX/UI Design	18548	1.23%	229	€944.03	€876,999.55	929	4.06	€3,829.69	€47.28	€8.06	€149,523.45	€727,476.10
Find yourself in IT	18548	0.01%	1	€0.00	€0.00	2	2.00	€0.00	€0.00	€8.06	€149,523.45	-€149,523.45
Data Analytics	18548	0.00%	0	€0.00	€0.00	0	0.00%	€0.00	€0.00	€8.06	€149,523.45	-€149,523.45

Ход работы:

- Для каждого продукта рассчитаны ключевые метрики:
 - **UA (number of acquired potential customers)** — количество привлечённых пользователей (контактов).
 - **C1 (Conversion Rate)** — конверсия в покупку.
 - **B (number of customers received)** — число пользователей, совершивших покупку.
 - **AOV (Average Order Value)** — средний чек.
 - **T (Transactions)** — кол-во транзакций.
 - **Revenue** — выручка.
 - **APC (average payment count per one customer)** — средняя частота покупок.
 - **CLTV (gross profit per customer (customer lifetime value))** — пожизненная ценность клиента.
 - **CPA (Cost Per Acquisition)** — стоимость привлечения клиента.
 - **LTV (gross profit per lead)** — валовая прибыль за лида.
 - **AC (acquisition buget)** — бюджет на меркетинг
 - **CM (Contribution Margin)** — маржинальная прибыль.

Результаты:

- Продукты **Digital Marketing, Web Developer** и **UX/UI Design** показали наибольшую выручку и положительную маржинальность, являются «Локомотивами» бизнеса
- **Data Analytics** и **Find yourself in IT** почти не генерируют выручки и показывают отрицательную маржинальность.

2. Определение точки роста бизнеса

2. Юнит-Экономика по всему бизнесу:												
	UA	C1	B	AOV	Revenue	T	APC	CLTV	LTV	LTC/CPA	AC	CM
	18548	4.53%	840	€1,029.08	€3,374,357.12	3279	3.90	€4,017.09	€181.93	€8.06	€149,523.45	€3,224,833.67
Для поиска точки роста - изменим метрики на 10%												
UA	20403	4.53%	924	€1,029.08	€3,711,792.83	3607	3.90	€4,017.09	€181.93	€8.06	€164,475.80	€3,547,317.04
C1	18548	4.98%	924	€1,029.08	€3,711,792.83	3607	3.90	€4,017.09	€200.12	€8.06	€149,523.45	€3,562,269.38
AOV	18548	4.53%	840	€1,131.99	€3,711,792.83	3279	3.90	€4,418.80	€200.12	€8.06	€149,523.45	€3,562,269.38
APC	18548	4.53%	840	€1,029.08	€3,711,792.83	3607	4.29	€4,418.80	€200.12	€8.06	€149,523.45	€3,562,269.38
CPA	18548	4.53%	840	€1,029.08	€3,374,357.12	3279	3.90	€4,017.09	€181.93	€8.87	€164,475.80	€3,209,881.33
Точка роста - C1 (напрямую влияет на Revenue и CM, без роста AC)												

3. Дерево метрик

Целевая метрика:	
- CM (маржинальная прибыль)	
Метрики принятия решений (юнит-экономика):	
- UA (кол-во лидов)	
- C1 (конверсия в покупку)	
- AOV (средний чек)	
- APC (среднее кол-во сделок на клиента)	
- CPA (стоимость привлечения потенциального клиента)	
Финансовые метрики:	
- Revenue	
Продуктовые метрики:	
- AC	
- T (общее количество транзакций)	
- R_I (выручка на клиента)	
- CLTV (средняя валовая прибыль на клиента)	
- LTV (средняя валовая прибыль на юнит масштабирования)	
- B (клиенты, совершившие покупку)	
Атомные метрики:	
- Offer Category	
- Payment Category	
- Months of study	
- Course duration	
- Offer Total Amount	
- Initial Amount Paid	
- Product	
- Payment Type	
- Id	
- Spend	

4. Формулировка гипотез

HADI-циклы (цель: рост C1 на 1п.п.(процентный пункт) за ≤ 2 недели)				
	HADI #1 Быстрый контакт (Speed-to-Lead ≤15 min)	HADI #2 Welcome-цепочка за 24 часа	HADI #3 Соц-доказательства + гарантии на лендинге/чекхуте	HADI #4 Оффер 48 часов для новичков
H (Гипотеза)	Если менеджер связывается с лидом в первые 15 минут, -> C1 растет.	Серия сообщений лиду - сразу, через 6 ч, через 22 ч - повышает выполнение (CTR) и оплаты. -> рост C1	Если на лендинге и в чекхуте показать отзывы выпускников, истории успеха, кейсы трудоустройства, гарантию возврата, и мини-FAQ - это уменьшает сомнения. Итог - лиды будут чаще платить.	Если новичку дать ограниченное предложение (-5% или бонусный модуль) в первые 48 часов - это ускорит решение об оплате
A (Действие)	Настроить авто-сообщение или звонок ≤15 мин, + напоминание менеджеру, + штилон VIA/email. Итог - SLA (время отклика) ≤15 min	Настроить welcome-цепочку из четкого сообщений "записись на демо/получи пробный урок"	Сделать вариант страницы с блоком доверия	Отправка персонального оффера сразу после регистрации
D (Данные)	Время заявки, время контакта, факт оплаты (Payment Done)	Открытие писем, бронирование демо, факт оплаты	- переходы → заявки → оплаты; - AOV (средний чек)	Активации оффера, оплаты, выручка/маржа
I (Инсайт)	Если конверсия выросла хотя бы на +1 п.п. (с 4.5% до 5.5%), тест успешен	Если бронирование демо выросло хотя бы на 10% относительно, и C1 тоже вырос, гипотеза подтверждается	Если конверсия выросла ≥ +1 п.п. и при этом AOV не упал, гипотеза верна.	Если мы получили больше регистраций за ту же сумму, значит, наш подход работает — CPA стал лучше
Н0 (Нулевая гипотеза)	Связь с лидом в первые 15 мин не дала роста конверсии (конверсия осталась прежней)	Welcome-цепочка не работает, C1 остается прежней	Новые креативы не работают, C1 остается прежней	Если конверсия выросла ≥ +1 п.п. и прибыль (CM) не снизилась, гипотеза подтверждается

5. Условия проведения тестов

Как делаем:				
1. Берём всех новых лидов.				
2. Случайно делим их на 2 группы + следим за репрезентативностью:				
- А (контроль) - все по старому				
- В (тест) - применяем изменения (быстрый контакт, welcome, оффер, соц-доказательства)				
3. Сравниваем C1 = В/А между группами				
Для оценки статистической значимости используется z-тест долей (α = 0.05)				

Условие проведения гипотезы:				
Гипотеза считается подтверждённой, если одновременно выполняются условия:				
1. Рост конверсии (C1) не менее чем на 1 п.п. (MDE = 0.01) (4.5% → 5.5%) в тестовой группе относительно контрольной.				
2. p-value < 0.05 (разница статистически значима)				
3. Дополнительные метрики не ухудшились:				
- CM (маржинальная прибыль) не снизилась				
- CPA (стоимость привлечения) не выросла				
Если хотя бы одно условие не выполняется → гипотеза считается неподтверждённой.				

	А/В тест		Метрика	Условие успеха
	Контрольная группа А	Тестовая группа В		
Быстрый контакт	текущая связь менеджера с лидом	Ответ ≤15 мин	C1	+ 1 п.п., p<0.05
Welcome-цепочка	нет серии	3 письма за 24ч	C1	рост C1
Соц-доказательства	стандартная версия лендинга	Страница с отзывами и FAQ	C1, CTR	+ 1 п.п., рост CTR
Оффер 48ч	нет оффера за 48ч	Купон/бонус новичкам, персональный оффер	C1, CM	+ 1 п.п., CM не снизилась

Расчеты для проведения эксперимента:				
Сколько лидов нужно: $n = (16 \cdot p(1-p))/x^2$				
p = базовая конверсия (4.5% = 0.0453)				
x = минимальный рост, который хотим заметить (+1 п.п. = 0.01)				
$n = (16 \cdot 0.0453 \cdot (1 - 0.0453)) / 0.01^2 \approx 6\,920$ лидов в каждую группу (13.8k всего)				
Итого - Срок: до 14 дней (нужно ~1k лидов/день)				

