



# ФИНАЛЬНЫЙ ПРОЕКТ

по курсу «Аналитик данных»

(группа 050824)

**Исполнитель:**  
Михаил Филимонов

Берлин – Дрезден  
2025

## Оглавление

1. Очистка и подготовка данных .....	5
1.1     CALLS .....	6
1.1.1 Загрузка данных и преобразование в датафрейм .....	6
1.1.2 Предварительная очистка: .....	6
1.1.3 Обработка отсутствующих значений и преобразование типов данных .....	6
1.1.4 Общие выводы по датафрейму Calls: .....	7
1.2     CONTACTS.....	7
1.2.1    Загрузка данных и преобразование в датафрейм .....	7
1.2.2 Предварительная очистка: .....	7
1.2.3 Обработка отсутствующих значений и преобразование типов данных .....	8
1.2.4 Общие выводы по датафрейму Contacts:.....	9
1.3     SPEND .....	9
1.3.1 Загрузка данных и преобразование в датафрейм .....	9
1.3.2 Предварительная очистка: .....	9
1.3.3 Обработка отсутствующих значений и преобразование типов данных .....	9
1.3.4 Общие выводы по датафрейму SPEND .....	10
1.4     DEALS.....	10
1.4.1 Загрузка данных и преобразование в датафрейм¶ .....	10
1.4.2 Предварительная очистка и обработка:¶.....	10
1.4.3 Обработка отсутствующих значений и преобразование типов данных¶ .....	11
1.4.4 Общие выводы по датафрейму DEALS¶.....	17
2. Описательная статистика.....	18
2.1 CALLS.....	18
2.2 CONTACTS .....	22
2.3 SPEND .....	23
2.4 DEALS .....	29
3. Анализ временных рядов .....	42
3.1 Анализ тенденций создания сделок с течением времени и их связь со звонками. ....	42
Динамика создания сделок и их связь со звонками во времени .....	42
Тенденция заключенных сделок и совершенных звонков по месяцам.....	42
Декомпозиция временного ряда .....	42
Корреляция между числом звонков и числом новых сделок .....	44
Анализ времени перехода сделок на финальные этапы после звонков .....	44
Анализ количества звонков по этапам сделок .....	46
Анализ средней продолжительности звонков по этапам сделок .....	47
3.2 Распределение времени закрытия сделок и продолжительность периода от создания до закрытия ..	48
Распределение времени закрытия сделок.....	48
Анализ продолжительности периода от создания до закрытия сделки .....	49
3.3 Общие выводы по проведенному анализу временных рядов сделок и звонков .....	49

4. Анализ эффективности кампаний .....	51
4.1 Оценка эффективность различных маркетинговых источников (Source) в генерировании качественных лидов .....	51
Анализ количества лидов от каждого источника и коэффициента конверсии в первую покупку: .....	51
Визуализация количества лидов и коэффициента конверсии для каждого источника .....	51
Для более глубокого анализа эффективности рекламных источников добавим в анализ число закрытых сделок и среднюю сумму сделки по источникам .....	52
Визуализация сравнения маркетинговых источников (Source) по среднему времени закрытия сделки (Deals_Closed) и средней сумме сделки (Average_Deal_Value) .....	53
4.2 Сравните эффективность различных кампаний с точки зрения генерации лидов и коэффициента конверсии.....	54
Рассчитаем показатели эффективности кампаний .....	54
Интерактивные диаграммы рейтинга кампаний по генерации лидов и корверсии .....	54
Результаты анализа эффективности маркетинговых кампаний с точки зрения генерации лидов и коэффициента конверсии: .....	55
4.3* Сравнение эффективности маркетинговых кампаний с точки зрения затрат на рекламу .....	56
Проверим соответствие кампаний в датафреймах deals и spends .....	56
Анализ совпадений и различий в кампаниях между deals и spend.....	56
Проанализируем связь между затратами на маркетинговые кампании и их основными индикаторами эффективности.....	57
Визуализируем на интерактивной диаграмме зависимость затрат на рекламные кампании с количеством лидов, числом закрытых сделок и коэффициентом конверсии .....	58
Результаты анализа.....	58
5 Анализ эффективности работы отдела продаж¶ .....	60
5.1 Оцените эффективность отдельных владельцев сделок и рекламных кампаний с точки зрения количества обработанных сделок, коэффициента конверсии и общей суммы продаж.¶ .....	60
5.1.1 Анализ эффективности менеджеров продаж (владельцев сделок).....	60
5.1.2 Анализ эффективности рекламных кампаний .....	62
6. Анализ платежей и продуктов .....	69
6.1 Анализ распределения типов оплаты и их влияния на успешность сделок .....	69
6.2 Анализ популярности и успешности различных продуктов и типов обучения .....	69
Анализ популярности и успешности продуктов (Product) и типов обучения (Education_Type) .....	69
Анализ доходности продуктов (Product) и типов обучения (Education_Type) .....	71
7 Географический анализ .....	73
7.1 Анализ географического распределения сделок по городам .....	73
Строим сводную таблицу рейтинга успешных сделок по городам .....	73
Визуализируем топ 100 городов по числу успешных сделок¶ .....	73
Отображаем на карте города с числом сделок, количеством успешных из них и коэффициента успешности сделки .....	73
Выводы из анализа географического распределение сделок по городам .....	74
Фильтрация городов по странам и успешности сделок – определение наиболее перспективных локаций .....	74
Выводы из анализа успешных сделок по странам.....	75

7.2 Анализ влияния уровня знания немецкого языка на успешность сделок в разных городах.....	76
Посчитаем и визуализируем статистику успешности сделок в разрезе уровня владения немецким языком .....	76
Выводы из анализа успешности сделок в зависимости от уровня знания немецкого языка .....	77
В каких городах знание языка оказывает наибольшее влияние на успешность сделки? .....	77
Интерактивная карта успешности сделок по городам в зависимости от уровня немецкого языка .....	78
Проверим гипотезу о наличии статистической зависимости между уровнем знания немецкого языка и успешностью сделок.....	78
8. Продуктовая аналитика.....	80
8.1. Расчет юнит-экономики по продуктам .....	80
Определение количества лидеров(UA), маркетингового бюджета (AC) и стоимости привлечения лидера (LTC) .....	80
Рассчет метрик юнит-экономики про продуктам (предлагаемый курс обучения) .....	80
Рассчет метрик юнит-экономики про продуктам (предлагаемый курс обучения) и типам обучения (утренний курс/вечерний курс) .....	81
Визуализируем маржинальную прибыль по продуктам .....	81
Анализ юнит-экономики образовательных продуктов .....	81
8.2. Определение точек роста бизнеса .....	82
8.3. Дерево метрик для бизнеса.....	84
8.4. Формулирование гипотез для метрик продукта .....	84
8.5. Описание метода проверки гипотез с формулированием условий проверки .....	85
HADI-цикли для проверки гипотез .....	85
Расчет доверительных интервалов и размера выборки для проведения A B тестирования .....	85
Формулирование исходные данных для экспериментов .....	86
Проверим срок проведения тестирования.....	86
Моделирование проведения A B теста .....	87
9. Общие выводы и рекомендации по проекту.....	89
9.1 SWOT-анализ стратегии бизнеса.....	89
9.2 Итоговые выводы и рекомендации для бизнеса .....	89

## 1. Очистка и подготовка данных

Очистка и подготовка данных — важнейший этап анализа данных, который влияет на точность, надежность и интерпретацию результатов.

Что сделано при очистке данных:

- удаление ошибок и аномалий;
- исправление некорректных или отсутствующих значений (NaN);
- устранение дубликатов, несуществующих или ошибочных записей;
- приведение данных к единообразному виду;
- упорядочивание форматов (например, даты YYYY-MM-DD).
- преобразование типов (str → int, object → datetime | category).

Обогащение данных:

- добавление новых полезных переменных (Paid, T, AOV);
- формирование групп (категориальные типы данных);
- удаление шума и выбросов (Initial\_Amount\_Paid, Offer\_Total\_Amount);
- исключение значений, которые выходят за разумные границы (Initial\_Amount\_Paid > Offer\_Total\_Amount).

Важность этапа очистки и подготовки данных:

- ошибки могут исказить выводы → некорректные данные приводят к ложным результатам;
- неправильные типы данных могут «сломать» анализ, например, нельзя построить график, если datetime хранится как str.
- кодирование категорий (deal.Stage) важно для Хи-тестирования и моделей машинного обучения;
- выбросы могут испортить тренды, например, "аномальная" оплата (1 €) изменит среднее значение.

Следуя принципам функционального программирования при выполнении задания мною было разработано ряд функций, которые использовались для выполнения часто повторяющихся задач, например при очистке данных, а также специфических задач обработки данных.

Имя модуля - **My\_Function\_050824\_M\_Filimonov.py**, импортируется с алиасом **mvf**.

**На этапе очистки и обработки данных с целью уменьшения рутинных действий я использовал две функции :**

**df\_pre\_clean()**, которая выполняет следующие операции:

- удаляет дубликаты строк;
- заменяет пробелы в именах столбцов на подчеркивания;
- удаляет столбцы, где все значения NaN;
- удаляет строки, где все значения NaN;
- заменяет пробелы в именах колонок символом подчеркивания, для обеспечения синтаксиса „dot.notation“.

**my\_columns\_describe()** используется для:

анализа столбцов DataFrame на основе их типов данных;  
отображения статистических сводок;  
анализа гомогенности значений столбца (содержание неоднородных типов данных);  
предоставления информации об уникальных значениях и отсутствующих данных.

**На этапе анализа описательной статистики числовых и категориальных данных, с целью расчета расширенной сводной статистики и визуализации распределения данных я использовал функции :**

**num\_fields\_analyze():**

Анализирует числовые поля датафрейма:

- выводит среднее, медиану, моду и диапазон для указанных числовых столбцов;
- логарифмирует указанные поля, если передан список log\_fields.
- строит гистограмму и «ящик с усами» для каждого числового поля (до и после логарифмирования, если применимо).

**plot\_categorical\_distribution():**

Функция для анализа категориальных переменных в DataFrame:

- выводит распределение значений в каждом категориальном столбце с процентами.
- строит график подсчета для каждой категориальной переменной, добавляя числовые метки и проценты.

**На этапе обработки данных**

разработаны и использовались ряд других функций для выполнения специальных задач, они находятся в соответствующих фрагментах кода и будут описаны в контексте.

## 1.1 CALLS

### 1.1.1 Загрузка данных и преобразование в датафрейм

	<b>Id</b>	<b>Call Start Time</b>	<b>Call Owner Name</b>	<b>CONTACTID</b>	<b>Call Type</b>	<b>Call Duration (in seconds)</b>	<b>Call Status</b>	<b>Dialled Number</b>	<b>Outgoing Call Status</b>	<b>Scheduled in CRM</b>	<b>Tag</b>
<b>0</b>	5805028000000805001	30.06.2023 08:43	John Doe	NaN	Inbound	171.0	Received	NaN	NaN	NaN	NaN
<b>1</b>	5805028000000768006	30.06.2023 08:46	John Doe	NaN	Outbound	28.0	Attended Dialled	NaN	Completed	0.0	NaN
<b>2</b>	5805028000000764027	30.06.2023 08:59	John Doe	NaN	Outbound	24.0	Attended Dialled	NaN	Completed	0.0	NaN

### 1.1.2 Предварительная очистка:

DataFrame 'calls' info:

ROWS number: 95874

COLUMNS number: 11

No ROWs, where all values are NaN found

No duplicated ROWs found.

COLUMNS with all NaN values removed:

- Dialled Number
- Tag

Old columns names: ['Id', 'Call Start Time', 'Call Owner Name', 'CONTACTID', 'Call Type', 'Call Duration (in seconds)', 'Call Status', 'Outgoing Call Status', 'Scheduled in CRM']

New columns names: ['Id', 'Call\_Start\_Time', 'Call\_Owner\_Name', 'CONTACTID', 'Call\_Type', 'Call\_Duration\_(in\_seconds)', 'Call\_Status', 'Outgoing\_Call\_Status', 'Scheduled\_in\_CRM']

### Вывод по предварительной очистке датафрейма Calls:

- удалено два столбца (*Dialled Number, Tag*) как неинформативные.
- дубликатов строк не выявлено

### 1.1.3 Обработка отсутствующих значений и преобразование типов данных

#### Предложения по преобразованию типов данных в датафрейме Calls:

```
#   #   Column          Dtype   new Dtype
#   --  --  --
#  0   Id      object  -> object  - тип данных без изменений, идентификатор
#  1 Call_Start_Time  object  -> datetime - для анализа временных рядов
#  2 Call_Owner_Name  object  -> category - имена повторяются и их набор фиксирован *
#  3 CONTACTID       object  -> object  - тип данных без изменений, идентификатор
#  4 Call_Type        object  -> category - значения повторяются и их набор фиксирован *
#  5 Call_Duration    float64 -> int32   - т.к. это целые числа
#  6 Call_Status      object  -> category - значения повторяются и их набор фиксирован *
#  7 Outgoing_Call_Status  object  -> category - значения повторяются и их набор фиксирован *
#  8 Scheduled_in_CRM float64 -> category - значения повторяются и их набор фиксирован
#                                         [1.0,0.0,NaN]->["yes","no","unknown"]*
#
# * ускорит операции фильтрации и группировки
```

#### Предложения по замене пропущенных значений в датафрейме Calls:

- **CONTACTID** т.к это идентификатор, пропущенные значения (4%) **целесообразно заполнить модой**, из групп по Call\_Start\_Time (предварительно выделив из него дату) ,Call\_Owner\_Name и Call\_Type для сохранения целостности данных и корректного объединения датафреймов (таблиц) при дальнейшем анализе.
- **Call\_Duration** пропущенные значения целесообразно **заполнить медианным значением (8 сек)** вместо среднего по столбцу (**165 сек**). Это позволит нивелировать влияние "выбросов" (очень длинных звонков) и сохранить тенденцию типичной длительности звонков.
- **Outgoing\_Call\_Status** пропущенные значения целесообразно заполнить значением "**unknown**", которое обозначит, что статус исходящего звонка не был указан. Это позволит сохранить структуру данных для последующего анализа.
- **Scheduled in CRM** пропущенные значения целесообразно заполнить значением "**unknown**", которое указывает на то, что нам неизвестно, был ли звонок запланирован в CRM.

#### 1.1.4 Общие выводы по датафрейму Calls:

1. Изначально датафрейм имел 95874 строки и 11 колонок.
2. После предварительной очистки было удалено два пустых столбца (Dialled Number, Tag) как неинформативные, дубликатов и пустых строк не выявлено.
3. На основании анализа типов данных и их гомогенности, а также распределения отсутствующих значений приняты решения по изменению типов и категорированию данных. Также для каждого столбца выбрана своя тактика работы с отсутствующими значениями.
4. После очистки данных датафрейм имеет 95588 строк и 9 столбцов.

## 1.2 CONTACTS

#### 1.2.1 Загрузка данных и преобразование в датафрейм

	<b>Id</b>	<b>Contact Owner Name</b>	<b>Created Time</b>	<b>Modified Time</b>
0	5805028000000645014	Rachel White	27.06.2023 11:28	22.12.2023 13:34
1	5805028000000872003	Charlie Davis	03.07.2023 11:31	21.05.2024 10:23
2	5805028000000889001	Bob Brown	02.07.2023 22:37	21.12.2023 13:17
3	5805028000000907006	Bob Brown	03.07.2023 05:44	29.12.2023 15:20
4	5805028000000939010	Nina Scott	04.07.2023 10:11	16.04.2024 16:14

#### 1.2.2 Предварительная очистка:

```
DataFrame 'contacts' info:
ROWS number: 18548
COLUMNS number: 4
No ROWs, where all values are NaN found
No duplicated ROWs found.
No COLUMNS with all NaN values were found.

Old columns names: ['Id', 'Contact Owner Name', 'Created Time', 'Modified Time']

New columns names: ['Id', 'Contact_Owner_Name', 'Created_Time', 'Modified_Time']
```

Вывод по предварительной очистке датафрейма Contacts:

- пустые строки и колонки отсутствуют
- дублирующиеся строки отсутствуют

## 1.2.3 Обработка отсутствующих значений и преобразование типов данных

### Анализ данных по колонкам

```
DataFrame info:  
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 18548 entries, 0 to 18547  
Data columns (total 4 columns):  
 #   Column           Non-Null Count  Dtype     
---  --    
 0   Id               18548 non-null   object    
 1   Contact_Owner_Name 18548 non-null   object    
 2   Created_Time      18548 non-null   object    
 3   Modified_Time     18548 non-null   object    
 dtypes: object(4)  
 memory usage: 579.8+ KB  
...  
Пример выявления неоднородных данных в колонке Call_Owner_Name:
```

```
List of OBJECT-type columns  
Column: 'Id'  
Unique values: 18548  
Missing values (NaN): 0 | 0%  
  
Column: 'Contact_Owner_Name'  
Attention! The data in the column is not homogeneous:  
Contact_Owner_Name  
<class 'str'>    18547  
<class 'bool'>     1  
Name: count, dtype: int64  
Unique values: 28  
Values distribution:  
Contact_Owner_Name  
Charlie Davis      2018  
Ulysses Adams       1816  
Julia Nelson        1769  
Paula Underwood     1487  
Quincy Vincent      1416  
...  
Zachary Foster      8  
Tina Zhang          2  
Wendy Clark          2  
False                1  
Derek James          1  
Name: count, Length: 28, dtype: int64  
Missing values (NaN): 0 | 0%
```

### Исправление ошибки гомогенности данных в столбце Contact\_Owner\_Name

Row with incorrect Contact\_Owner\_Name:

		<b>Id</b>	<b>Contact_Owner_Name</b>	<b>Created_Time</b>	<b>Modified_Time</b>
2197	5805028000008772190			False 2023-09-24 09:01:00	13.10.2023 16:44
Значение Contact_Owner_Name в ошибочной строке заменено на: Jane Smith					

### Предложения по преобразованию типов в датафрейме Contacts:<sup>¶</sup>

```
# # Column           Dtype  New Dtype  
# ---  --    
# 0   Id               object -> object - оставить без изменений  
# 1   Contact_Owner_Name  object -> category - значения повторяются и их набор фиксирован *  
# 2   Created_Time      object -> datetime - для анализа временных рядов  
# 3   Modified_Time     object -> datetime - для анализа временных рядов  
# * ускорит операции фильтрации и группировки
```

#### 1.2.4 Общие выводы по датафрейму Contacts:

- исправили ошибку гомогенности данных Contact\_Owner\_Name;
- пропущенных значений нет;
- типы данных преобразованы в соответствии с выводами из анализа датасета

### 1.3 SPEND

#### 1.3.1 Загрузка данных и преобразование в датафрейм

	Date	Source	Campaign	Impressions	Spend	Clicks	AdGroup	Ad
0	2023-07-03	Google Ads	gen_analyst_DE	6	0.00	0	NaN	NaN
1	2023-07-03	Google Ads	performancemax_eng_DE	4	0.01	1	NaN	NaN
2	2023-07-03	Facebook Ads	NaN	0	0.00	0	NaN	NaN
3	2023-07-03	Google Ads	NaN	0	0.00	0	NaN	NaN

#### 1.3.2 Предварительная очистка:

DataFrame 'spend' info:

```
ROWS number: 20779
COLUMNS number: 8
No ROWs, where all values are NaN found
Number of duplicate rows: 917
All duplicates ROWs have been removed.
No COLUMNS with all NaN values were found.
```

Old columns names: ['Date', 'Source', 'Campaign', 'Impressions', 'Spend', 'Clicks', 'AdGroup', 'Ad']

New columns names: ['Date', 'Source', 'Campaign', 'Impressions', 'Spend', 'Clicks', 'AdGroup', 'Ad']

Вывод по предварительной очистке датафрейма Spend:

- пустые строки и колонки отсутствуют
- выявлено 917 дубликатов строк. Дубликаты удалены.

#### 1.3.3 Обработка отсутствующих значений и преобразование типов данных

##### Анализ данных по колонкам

```
DataFrame info:
<class 'pandas.core.frame.DataFrame'>
Index: 19862 entries, 0 to 20778
Data columns (total 8 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   Date        19862 non-null   datetime64[ns]
 1   Source      19862 non-null   object  
 2   Campaign    14785 non-null   object  
 3   Impressions 19862 non-null   int64   
 4   Spend       19862 non-null   float64 
 5   Clicks      19862 non-null   int64   
 6   AdGroup     13951 non-null   object  
 7   Ad          13951 non-null   object  
dtypes: datetime64[ns](1), float64(1), int64(2), object(4)
memory usage: 1.4+ MB
...
```

##### Предложения по преобразованию типов в датафрейме Spend:

- **Date** datetime64[ns] -> оставить без изменений
- **Source** object -> преобразовать в категориальный тип данных

- **Campaign** object -> преобразовать в категориальный тип данных
- **Impressions** int64 -> преобразовать в Int32
- **Spend** float64 -> оставить без изменений
- **Clicks** int64 -> преобразовать в Int32
- **AdGroup** object -> преобразовать в категориальный тип данных
- **Ad** object -> преобразовать в категориальный тип данных

#### **Предложения по замене пропущенных значений в датафрейме Spend:**

- **Campaign** - пропущенные значения возможно частично заменить на основе подходящего значения из группы по Source (по совпадению с датой указанной в названии Campaign), остальные пропуски заполним значением "unknown".
- **AdGroup** - пропущенные значения возможно частично заменить на основе столбца «Campaign» с использованием предопределенной логики, остальные пропуски заполним значением "unknown".
- **Ad** - пропуски заполним значением "unknown".

#### **Выполняем преобразование типов и обработку отсутствующих значений**

- Number NaN records in 'Campaign' before processing = 5077
- Number NaN records in 'Campaign' after processing = 3230
- Processed NaN are 1847 values, which is 36.38%
- Number of NaN records in 'AdGroup' before processing = 5911
- Number of NaN records in 'AdGroup' after processing = 3167
- Processed NaN are 2744 values, which is 46.42%.

#### **1.3.4 Общие выводы по датафрейму SPEND**

- Всего было 20779 строк и 8 столбцов;
- Удалено 917 дублирующихся строк;
- Проведение сопоставительного анализа в группах данных Source-Campaign и Campaign-AdGroup позволило восстановить 36.38% отсутствующих значений в Campaign и 46.42% значений в AdGroup;
- Типы данных преобразованы в соответствии с выводами из анализа датасета

## **1.4 DEALS**

### **1.4.1 Загрузка данных и преобразование в датафрейм¶**

0	5805028000056864695	Ben Hall	NaN	NaN	New Lead	NaN	/eng/test	03.07.23 women	NaN	v16	...	NaN	NaN	2024-06-21 15:30:00	NaN	NaN	NaN	NaN	5805028000056849495	NaN	NaN		
1	5805028000056859489	Ulysses Adams	NaN	NaN	New Lead	NaN	/at-eng	NaN	NaN	NaN	...	Web Developer	Morning	2024-06-21 15:23:00	6.0	NaN	0	2000	5805028000056834471	NaN	NaN		
2	5805028000056832357	Ulysses Adams	21.06.2024	D - Non Target	Lost	Non target	/at-eng	engwien_AT	00:26:43	b1-at	...	NaN	NaN	2024-06-21 14:45:00	NaN	NaN	NaN	NaN	5805028000056854421	NaN	NaN		

3 rows × 23 columns

### **1.4.2 Предварительная очистка и обработка:¶**

DataFrame 'deals' info:

```
ROWS number: 21593
COLUMNS number: 23
No ROWs, where all values are NaN found
No duplicated ROWs found.
No COLUMNS with all NaN values were found.
```

Old columns names: ['Id', 'Deal Owner Name', 'Closing Date', 'Quality', 'Stage', 'Lost Reason', 'Page', 'Campaign', 'SLA', 'Content', 'Term', 'Source', 'Payment Type', 'Product', 'Education Type', 'Created Time', 'Course duration', 'Months of study', 'Initial Amount Paid', 'Offer Total Amount', 'Contact Name', 'City', 'Level of Deutsch']

New columns names: ['Id', 'Deal\_Owner\_Name', 'Closing\_Date', 'Quality', 'Stage', 'Lost\_Reason', 'Page', 'Campaign', 'SLA', 'Content', 'Term', 'Source', 'Payment\_Type', 'Product', 'Education\_Type', 'Created\_Time', 'Course\_duration', 'Months\_of\_study', 'Initial\_Amount\_Paid', 'Offer\_Total\_Amount', 'Contact\_Name', 'City', 'Level\_of\_Deutsch']

Вывод по предварительной очистке датафрейма Deals:

- пустые колонки отсутствуют;

- пустые строки отсутствуют
- дублирующихся строк нет;

### 1.4.3 Обработка отсутствующих значений и преобразование типов данных¶

#### Анализ данных по колонкам

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21593 entries, 0 to 21592
Data columns (total 23 columns):
 #   Column           Non-Null Count Dtype  
--- 
 0   Id               21593 non-null   object  
 1   Deal Owner Name 21564 non-null   object  
 2   Closing Date    14645 non-null   object  
 3   Quality          19340 non-null   object  
 4   Stage            21593 non-null   object  
 5   Lost Reason     16124 non-null   object  
 6   Page              21593 non-null   object  
 7   Campaign         16067 non-null   object  
 8   SLA              15533 non-null   object  
 9   Content          14147 non-null   object  
 10  Term             12454 non-null   object  
 11  Source            21593 non-null   object  
 12  Payment Type    496 non-null    object  
 13  Product          3592 non-null   object  
 14  Education Type  3299 non-null   object  
 15  Created Time    21593 non-null   datetime64[ns] 
 16  Course duration 3587 non-null   float64  
 17  Months of study 840 non-null    float64  
 18  Initial Amount Paid 4165 non-null   object  
 19  Offer Total Amount 4185 non-null   object  
 20  Contact Name    21532 non-null   object  
 21  City              2511 non-null   object  
 22  Level of Deutsch 1251 non-null   object  
dtypes: datetime64[ns](1), float64(2), object(20)
memory usage: 3.8+ MB
...
...
```

#### Предложения по преобразованию типов в датафрейме DEALS:

```
# '''
# Id               object -> object, так как это идентификатор
# Deal_Owner_Name object -> category - имена повторяются и их набор фиксирован *
# Closing_Date    object -> Datetime - для анализа временных рядов
# Quality          object -> category - имена повторяются и их набор фиксирован *
# Stage            object -> category - имена повторяются и их набор фиксирован *
# Lost_Reason      object -> category - имена повторяются и их набор фиксирован *
# Page              object -> category - имена повторяются и их набор фиксирован *
# Campaign         object -> category - имена повторяются и их набор фиксирован *
# SLA              object -> Timedelta - для дальнейшего анализа (исправляем ошибку гомогенности)
# Content          object -> category - имена повторяются и их набор фиксирован *
# Term              object -> category - имена повторяются и их набор фиксирован *
# Source            object -> category - имена повторяются и их набор фиксирован *
# Payment_Type     object -> category - имена повторяются и их набор фиксирован *
# Product           object -> category - имена повторяются и их набор фиксирован *
# Education_Type   object -> category - имена повторяются и их набор фиксирован *
# Created_Time     DateTime -> Datetime - оставить без изменения
# Course_duration  float64 -> Int8 целое число
# Months_of_study  float64 -> Int8 целое число
# Initial_Amount_Paid float64 -> float64 для корректного преобразования и поддержки NaN (исправляем ошибку гомогенности)
# Offer_Total_Amount float64 -> float64 для корректного преобразования и поддержки NaN (исправляем ошибку гомогенности)
# Contact_Name     object -> object, так как это идентификатор
# City              object -> category - имена повторяются и их набор фиксирован *
# Level_of_Deutsch object-> category - имена повторяются и их набор фиксирован *
# * ускорит операции фильтрации и группировки
# ...
# '''
```

#### Очистка столбцов 'Initial\_Amount\_Paid' и 'Offer\_Total\_Amount' и преобразование к float64

Initial\_Amount\_Paid unique before cleaning:

```
[nan 0 1000 '€ 3.500,00' 500 100 4500 300 200 2000 11000 4000 3000 3500
11500 1200 1500 1 5000 600 700 350 9 400 450]
```

Offer\_Total\_Amount unique before cleaning:

```
[nan 2000 9000 11000 3500 4500 '€ 2.900,00' 6500 4000 3000 10000 2500 5000
11500 1 1000 1200 0 1500 '€ 11398,00' 11111 6000]
```

---

#### Очистка столбца SLA

##### SLA

- <class 'datetime.time'> 13672
- <class 'datetime.timedelta'> 1861
- создаем новый столбец SLA\_sec (приводим все значения в секунды) и затем восстанавливем оригинальный SLA с типом Timedelta
- отсутствующие значения оставляем NaN

	SLA	SLA_sec
<b>21456</b>	311 days 10:34:24	26908464
<b>20045</b>	227 days 16:54:57	19673697
<b>18971</b>	217 days 07:51:48	18777108
<b>17653</b>	207 days 16:31:32	17944292
<b>18611</b>	207 days 04:20:30	17900430
...	...	...
<b>2903</b>	0 days 00:00:00	0
<b>13238</b>	0 days 00:00:00	0
<b>2904</b>	0 days 00:00:00	0
<b>2906</b>	0 days 00:00:00	0
<b>0</b>	0 days 00:00:00	0

21593 rows × 2 columns

#### Предложения по замене пропущенных значений в датафрейме DEALS:

- **Deal\_Owner\_Name** – пропущенные значения заполним значениями из справочника contacts, если есть совпадения.
- **Closing\_Date** – для обеспечения целостности данных использовали при преобразовании типов **errors='coerce'**, чтобы оставить пропуски NaT, т.к. пропущенные значения могут означать, что сделка еще не закрыта или дата закрытия не была зарегистрирована.
- **Quality** – пропущенные значения заполним значением "unknown".
- **Stage** – нет пропущенных значений.
- **Lost Reason** – пропущенные значения заполним значением "unknowp", значения "Duplicate" в колонке игнорируем, ввиду некорректного заполнения данных менеджерами и отсутствия обратной связи из-за специфики проекта.
- **Page** – нет пропущенных значений.
- **Campaign** – пропущенные значения заполним значением "unknown".
- **SLA** – пропущенные значения заполнили нулями при корректном расчете временных интервалов.
- **Content** – пропущенные значения заполним значением "unknown".
- **Term** – пропущенные значения заполним значением "unknown", **предварительно заполнив где возможно данными, взятыми из Campaign**.
- **Source** – нет пропущенных значений.
- **Payment\_Type** – пропущенные значения заполним значением "unknown".
- **Product** – пропущенные значения заполним значением "unknown", **предварительно заполнив где возможно данными, полученными на основе анализа Campaign и Term**.
- **Education\_Type** – пропущенные значения заполним случайнм значением "Morning/Evening", однако **пропорционально** исходному соотношению датафрейма, предварительно сгруппировав по Product.
- **Course\_duration** – предполагалось заменить пропущенные значения **значением моды**, взятым из группы по Product и Education Type. Однако при замене выяснилось, что из-за слишком большого количества пропущенных значений результат только ухудшился, поэтому решение — **оставить колонку как есть**.
- **Months\_of\_study** – пропущенные значения оставляем, чтобы не нарушить целостность данных.
- **Initial\_Amount\_Paid** – пропущенные значения оставляем, чтобы не нарушить целостность данных. Используем значение как признак стадии сделки (начало обсуждения с лицом условий контракта) на этапе до Stage = "Payment Done".
- **Offer\_Total\_Amount** – пропущенные значения **заполним медианным значением** из группировки по Product, Education Type и Course\_duration.
- **Contact\_Name** – пропущенные значения оставляем, чтобы не нарушить целостность данных. Так как это идентификатор, то при анализе для корректного объединения с другими датафреймами по этому полю будем использовать соответствующий тип Join.
- **City** – проверим введенные значения на **ошибки ввода**, используем модуль **geopy** для определения стран и координат, пропущенные значения заполним "unknown".
- **Level\_of\_Deutsch** – проведем **лексическую обработку** и заполним некорректный ввод **таргетными значениями**, пропущенные значения — заполним значением "unknown".

#### Дополнительные колонки для расчета юнит-экономики

Для корректного расчета **юнит-экономики**, а также значений **суммы платежей и среднего платежа** добавим в датафрейм три колонки:

- **Paid** – оплаченная клиентом сумма;
- **T** – число транзакций;
- **AOV** – средний чек.

#### Логика их расчета

- Если Months\_of\_study == 0, то:
  - Paid = 0, T = 0, AOV = 0 – клиент еще не получил услугу и может изменить решение об обучении, соответственно этот платеж не учитывается.

- **Иначе:**
  - **Если** Initial\_Amount\_Paid == Offer\_Total\_Amount, то:
    - Paid = (Offer\_Total\_Amount / Course\_duration) \* Months\_of\_study
    - T = Months\_of\_study
    - AOV = Paid / T
  - **Иначе:**
    - Paid = Initial\_Amount\_Paid + ((Offer\_Total\_Amount - Initial\_Amount\_Paid) / (Course\_duration - 1)) \* Months\_of\_study
    - T = Months\_of\_study
    - AOV = Paid / T

#### Примечания

- "unknown" выбрано для категориальных столбцов, чтобы устраниить пропуски и обеспечить целостность данных.
- Заполнение числовых столбцов **медианным значением** проведено с целью нивелирования смещения данных.

**Выполняем обработку отсутствующих значений DEALS на базе вышеуказанной логики**

#### *Deal\_Owner\_Name*

##### Результат:

Deal\_Owner\_Name missing values before processing: 29  
Deal\_Owner\_Name missing values after processing: 0

#### *Education\_Type*

- Рассчитываем пропорцию Morning/Evening в имеющихся данных
- Создаем копию DataFrame, чтобы избежать SettingWithCopyWarning
- Фильтруем только те строки, где Education\_Type = NaN
- Группируем данные по Product и Course\_duration и применяем замену
- Генерируем случайные значения для замены, но пропорционально рассчитанной пропорции
- Заполняем в оригинальном DataFrame

##### Результат:

Education\_Type missing values before processing: 18294  
Education\_Type missing values after processing: 17992

#### *Course\_duration*

# нет результата из за большого количества NaN -> Course\_duration лучше оставить как есть

#### *Term*

- Используем разработанную функцию **fill\_adgroup()** для заполнения отсутствующих значений Term(=AdGroup) на основе Campaign (значения подготовлены с помощью LMM)
- Входные данные: уникальные списки значений столбцов Campaign и Term

##### Результат:

Number of NaN records in 'Term' before processing = 9139  
Number of NaN records in 'Term' after processing = 4686  
Processed NaN are 4453 values, which is 48.73%.

#### *Product*

**Анализируем подгруппу (Term) в рекламной кампании (Campaign) которая направлена на курс аналитик данных**

Campaign		Term
15639	gen_analyst_DE	аналитик данных обучение
15669	gen_analyst_DE	data analyst курсы
15693	gen_analyst_DE	курсы аналитик данных
15799	gen_analyst_DE	обучение на аналитика данных
15932	gen_analyst_DE	курсы аналитик данных
...	...	...
17863	gen_analyst_DE	data analysis analytics
17965	gen_analyst_DE	data analyst курсы
21522	gen_analyst_DE	data analyst
21532	gen_analyst_DE	курсы аналитика с нуля
21543	gen_analyst_DE	data analyst обучение

35 rows × 2 columns

#### Фильтруем продукт лиды в который пришли по этой группе объявлений в рекламной кампании

	Product	Source	Campaign	Term
15639	UX/UI Design	Google Ads	gen_analyst_DE	аналитик данных обучение
15669		Nan	gen_analyst_DE	data analyst курсы
15693		Nan	gen_analyst_DE	курсы аналитик данных
15799		Nan	gen_analyst_DE	обучение на аналитика данных
15932		Nan	gen_analyst_DE	курсы аналитик данных
...	...	...	...	...
17863	Digital Marketing	Google Ads	gen_analyst_DE	data analysis analytics
17965		Nan	gen_analyst_DE	data analyst курсы
21522		Nan	gen_analyst_DE	data analyst
21532		Nan	gen_analyst_DE	курсы аналитика с нуля
21543		Nan	gen_analyst_DE	data analyst обучение

35 rows × 4 columns

- Делаю предположение (которое при выполнении реального проекта обязательно согласовываю с отделом продаж), что **лиды пришедшие по рекламе курсов датааналитики в атрибуте Product должны иметь пометку 'Data Analytics'** и соответствующим образом восстановливаю пропущенные значения Product и заменяю ошибки ввода данных
- 

#### Результат:

```
Product distribution:
Product
Data Analytics      1
Digital Marketing   1990
Find yourself in IT 4
UX/UI Design       1022
Web Developer       575
Name: Product, dtype: int64
```

	Product	Source	Campaign	Term
15639	Data Analytics	Google Ads	gen_analyst_DE	аналитик данных обучение
15669	Data Analytics	Google Ads	gen_analyst_DE	data analyst курсы
15693	Data Analytics	Google Ads	gen_analyst_DE	курсы аналитик данных
15799	Data Analytics	Google Ads	gen_analyst_DE	обучение на аналитика данных
15932	Data Analytics	Google Ads	gen_analyst_DE	курсы аналитик данных
...	...	...	...	...
17863	Data Analytics	Google Ads	gen_analyst_DE	data analysis analytics
17965	Data Analytics	Google Ads	gen_analyst_DE	data analyst курсы
21522	Data Analytics	Google Ads	gen_analyst_DE	data analyst
21532	Data Analytics	Google Ads	gen_analyst_DE	курсы аналитика с нуля
21543	Data Analytics	Google Ads	gen_analyst_DE	data analyst обучение

35 rows × 4 columns

#### Final Product distribution:

```
Product
Data Analytics      36
Digital Marketing   1978
Find yourself in IT 4
UX/UI Design        1019
Web Developer       575
Name: Product, dtype: int64
```

#### Level\_of\_Deutsch

Применяю разработанную функцию `clean_value()` для очистки `Level_of_Deutsch`, которая

- делает лексический анализ того, что записано в поле `_Level_of_Deutsch`, используя регулярное выражение для сопоставления определенных ключевых слов для каждого уровня владения языком;
- реализует логику замены;
- подсчитываем вхождений уникальных значений после очистки.

#### Результат:

```
Counter({'unknown': 20375, 'B1': 780, 'B2': 162, 'A2': 137, 'A0': 85, 'A1': 26, 'C1': 25, 'C2': 3})
```

#### City

**Для дальнейшего анализа** нам потребуется собственно название города (City), а также страна (Country) и географические координаты (Latitude,Longitude).

- для верификации города и определения дополнительных полей создадим функцию `check_address()` на основе модуля `geopy`

#### Шаги обработки:

- на самом деле в поле City содержится микс от собственно города, до полного адреса с использованием кириллицы и уникальных символов немецкого, румынского, молдавского и алфавитов и других славянских языков;
- выделим уникальные значения City в список city\_list;
- напишем функцию `clean_addresses()`, которая определит некорректно заполненные поля City (с учетом вышесказанных проблем заполнения поля) и запишет их в список wrong\_addresses;
- на основе wrong\_addresses создадим словарь некорректных значений и исправим данные в датафрейме, заполнив отсутствующие значения и "мусор" значением "unknown";
- создадим собственно функцию `check_address()` и проверифицируем город, а также найдем необходимые поля Country, Latitude, Longitude.

**ВНИМАНИЕ!** Процесс обработки данных функцией `check_address()` может занять более 30 минут в зависимости от характеристик компьютера и скорости интернета. Прогресс бар - показывает процент выполнения процесса.

100% | [██████████] | 21593/21593 [36:28<0  
0:00, 9.87it/s]

	City	Country	Latitude	Longitude
56	Crailsheim	Germany	49.136563	10.072019
60	Dortmund	Germany	51.514227	7.465279
65	Stuttgart	Germany	48.778449	9.180013
71	Munich	Germany	48.137108	11.575382
75	Berlin	Germany	52.510885	13.398937
...	...	...	...	...
21313	Frankfurt	Germany	50.110644	8.682092
21314	Wolfsburg	Germany	52.420559	10.786168
21410	Ingolstadt	Germany	48.763016	11.42504
21551	Essen	Germany	51.458224	7.015817
21578	Brake	Germany	53.325611	8.48063

2159 rows × 4 columns

### Обработка аномалий в Created\_Time > Closing\_Date

Подсчитываем аномалии (ошибки ввода), где «Closing\_Date» меньше «Created\_Time», и исправляем их.

Результат:

It's 44 anomaly founded. Fixing it...

Created_Time	Closing_Date
454 2024-06-11 00:06:00	2024-06-16
2083 2024-05-22 21:29:00	2024-05-25
2787 2024-05-07 11:19:00	2024-05-12
3019 2024-05-07 15:31:00	2024-05-08
...	

### Рассчет по каждой сделке количества транзакций (T), среднего чека (AOV) и оплаченной суммы (Paid)

	Product	Education_Type	Payment_Type	Course_duration	Months_of_study	Initial_Amount_Paid	Offer_Total_Amount	Paid	T	AOV
468	Digital Marketing	Evening	Recurring Payments	11	1	500.0	4500.0	900.0	1.0	900.00
478	Digital Marketing	Evening	Recurring Payments	11	1	500.0	4500.0	900.0	1.0	900.00
497	Web Developer	Morning	unknown	6	1	1000.0	9000.0	2600.0	1.0	2600.00
833	Web Developer	Morning	unknown	6	1	1000.0	9000.0	2600.0	1.0	2600.00
1040	Digital Marketing	Morning	unknown	11	1	1000.0	11000.0	2000.0	1.0	2000.00
...	...	...	...	...	...	...	...	...	...	...
21314	Web Developer	Morning	unknown	6	2	1000.0	5000.0	2600.0	2.0	1300.00
21367	Digital Marketing	Morning	unknown	11	11	1000.0	11000.0	12000.0	11.0	1090.91
21410	Digital Marketing	Morning	One Payment	11	11	450.0	4000.0	4355.0	11.0	395.91
21555	Digital Marketing	Evening	unknown	11	8	1000.0	11500.0	9400.0	8.0	1175.00
21586	Digital Marketing	Morning	unknown	11	11	1000.0	11000.0	12000.0	11.0	1090.91

841 rows × 10 columns

### Обработка оставшихся данных согласно принятому решению

Заполняем пропущенные значения 'unknown' в колонках ['Quality', 'Stage', 'Lost\_Reason', 'Page', 'Campaign', 'Content', 'Source', 'Payment\_Type', 'Product', 'Education\_Type']

```
'deals' DataFrame size (row, column): (21593, 30)
DataTypes info:
Id                  object      Id          0
Deal_Owner_Name    category   Deal_Owner_Name  1
Closing_Date       datetime64[ns] Closing_Date  6948
Quality             category   Quality        0
Stage               category   Stage         0
Lost_Reason         category  Lost_Reason   0
Page                category   Page          0
Campaign            category   Campaign       0
SLA                 timedelta64[ns] SLA          0
Content              category  Content        0
Term                object     Term        4686
Source              category   Source        0
Payment_Type        category   Payment_Type  0
Product              category   Product       0
Education_Type      category   Education_Type 0
Created_Time        datetime64[ns] Created_Time  0
Course_duration     Int8      Course_duration 18006
Months_of_study     Int8      Months_of_study 20753
Initial_Amount_Paid float64   Initial_Amount_Paid 17428
Offer_Total_Amount  float64   Offer_Total_Amount 17408
Contact_Name         object    Contact_Name  61
City                category   City          0
Level_of_Deutsch    category   Level_of_Deutsch 0
SLA_sec              int64     SLA_sec       0
Country              object    Country       0
Latitude             object    Latitude      0
Longitude            object    Longitude     0
Paid                float64   Paid         0
T                   float64   T           0
AOV                float64   AOV         0
dtype: int64
Number of missing values:
```

#### 1.4.4 Общие выводы по датафрейму DEALS¶

- Датафрейм Deals содержит информацию по сделкам на обучение;
- До обработки датафрейм имел 21 593 строки и 23 столбца, объем 3.8 Мб
- После обработки он имеет 21 532 строки (удалены 61 строка отсутствующих идентификаторов в Contact\_Name), 27 столбцов (добавлен столбец SLA\_sec - время отклика в секундах и столбцы Country - для анализа охвата регионов откуда студенты ,Latitude Longitude для точной идентификации геолокации), объем 2.6 Мб - уменьшен за счет оптимизации типов данных;
- Проведение сопоставительного анализа в группах Campaign - Term позволило восстановить 48.73% отсутствующих значений в Term и 35 значений в Product;
- Лексический анализ Level\_of\_Deutsch позволил сопоставить по категориям 1218 значений что составило 99% доступных данных в этом столбце;
- Обработка при помощи регуляного выражения позволила очистить 875 из 877 доступных значений данных City, дальнейшая обработка которых средствами библиотеки georou позволило однозначно определить страну и геолокацию объектов, что позволит провести более качественно дальнейший анализ;
- Исправление ошибок гомогенности данных в столбцах SLA, Initial\_Amount\_Paid, Offer\_Total\_Amount и Level\_of\_Deutsch, замена оставшихся NaN в категориальных столбцах на "unknown" и заполнение пропусков медианными значениями в числовых столбцах Course\_duration и Initial\_Amount\_Paid позволило корректно заменить отсутствующие значения.

## 2. Описательная статистика

- Сводная статистика для числовых полей (среднее значение, медиана, стандартное отклонение диапазон)
- Анализ категориальных полей

### 2.1 CALLS

Загрузка датафрейма из файла 01\_calls.pkl

#### CALLS Рассчет сводной статистики

	count	mean	std	min	25%	50%	75%	max
Call_Duration	95588.00	164.86	401.45	0.00	4.00	8.00	97.00	7625.00

- данные по Call\_Duration приведены в секундах

#### CALLS Визуализация сводной статистики

##### Statistical characteristics of number fields:

Call\_Duration:

Max: 7625.00

Min: 0.00

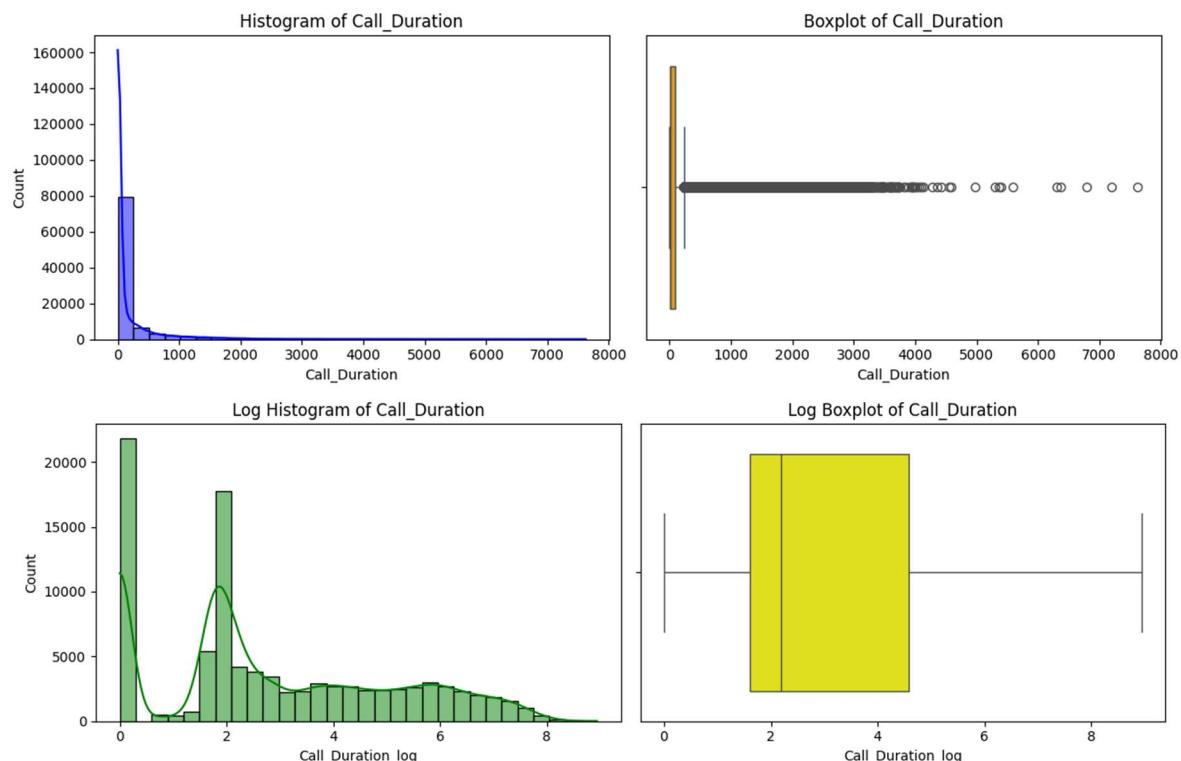
Mean: 164.86

std: 401.45

Median: 8.00

Mode: 0.00

Range: 7625.00



#### CALLS Анализ сводной статистики

Распределение времени звонков Call\_Duration показывает достаточно широкий разброс данных:

- Значения варьируются от 0 до 7625 (2.1 часа), показывая экстремальные выбросы;
- Средняя продолжительность звонка составляет 164,86 сек, или около 3 минут;
- При этом медиана составляет 8 сек, что указывает на сильное смещение данных в сторону малых значений;
- 50% всех звонков лежат в пределах 4 - 97 сек (1,62 мин);
- Мода: 0 секунд, что указывает на большое количество неудавшихся или сброшенных звонков.
- Статистика выбросов:

- Верхняя граница (upper fence) равна 236, что объясняет наличие значительного числа выбросов выше этого значения;
- Большая часть данных сосредоточена в пределах значений от 0 до 100, а остальное — аномально высокие значения.

## CALLS Анализ категориальных полей

	<b>count</b>	<b>unique</b>	<b>top</b>	<b>freq</b>
<b>Call_Owner_Name</b>	95588	32	Yara Edwards	9057
<b>Call_Type</b>	95588	3	Outbound	86835
<b>Id</b>	95588	95588	5805028000056839476	1
<b>CONTACTID</b>	95588	15214	580502800003329100	105
<b>Call_Status</b>	95588	11	Attended Dialled	70673
<b>Outgoing_Call_Status</b>	95588	5	Completed	86759
<b>Scheduled_in_CRM</b>	95588	3	no	86700

### Distribution of categorical variables:

Call\_Owner\_Name (Unique values: 32):  
 Yara Edwards: 9057 records (9.5%)  
 Julia Nelson: 7433 records (7.8%)  
 Charlie Davis: 7180 records (7.5%)  
 Ian Miller: 7177 records (7.5%)  
 Diana Evans: 6846 records (7.2%)  
 Ulysses Adams: 6062 records (6.3%)  
 Amy Green: 5978 records (6.3%)  
 Nina Scott: 5565 records (5.8%)  
 Victor Barnes: 5422 records (5.7%)  
 Kevin Parker: 5394 records (5.6%)  
 Paula Underwood: 4570 records (4.8%)  
 Quincy Vincent: 4367 records (4.6%)  
 Jane Smith: 3746 records (3.9%)  
 Cara Iverson: 3300 records (3.5%)  
 Ben Hall: 2947 records (3.1%)  
 John Doe: 2927 records (3.1%)  
 Alice Johnson: 1250 records (1.3%)  
 Mason Roberts: 1165 records (1.2%)  
 Derek James: 945 records (1.0%)  
 George King: 850 records (0.9%)  
 Zachary Foster: 518 records (0.5%)  
 Eva Kent: 493 records (0.5%)  
 Fiona Jackson: 470 records (0.5%)  
 Sam Young: 456 records (0.5%)  
 Rachel White: 441 records (0.5%)  
 Xander Dean: 304 records (0.3%)  
 Ethan Harris: 280 records (0.3%)  
 Hannah Lee: 175 records (0.2%)  
 Wendy Clark: 162 records (0.2%)  
 Bob Brown: 96 records (0.1%)  
 ... (too many categories, showing first 30)

-----

Call\_Type (Unique values: 3):  
 Outbound: 86835 records (90.8%)  
 Missed: 5794 records (6.1%)  
 Inbound: 2959 records (3.1%)

-----

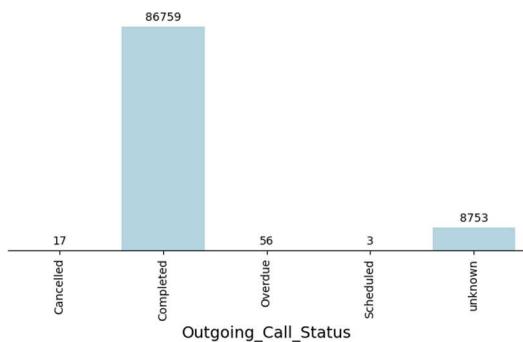
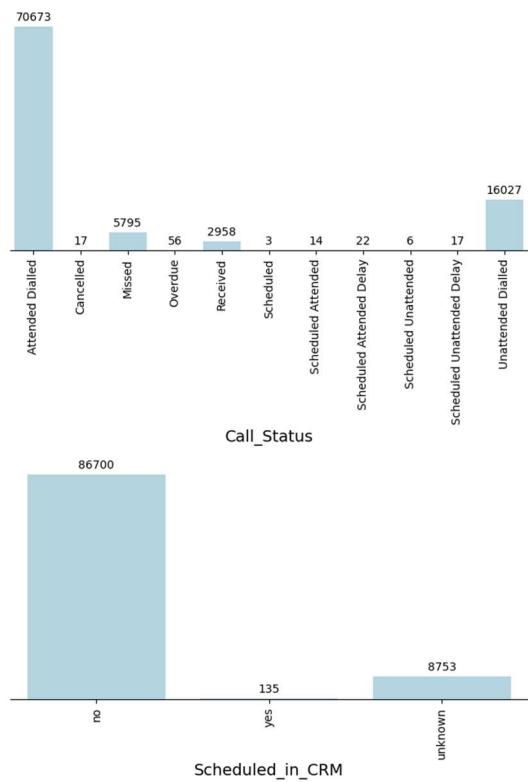
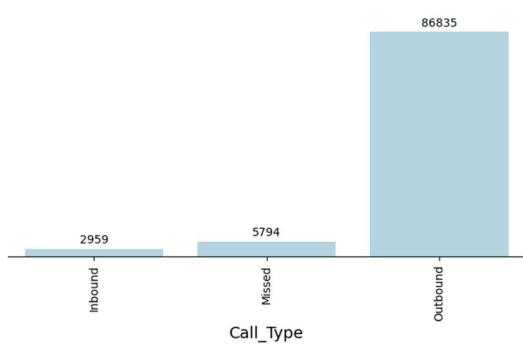
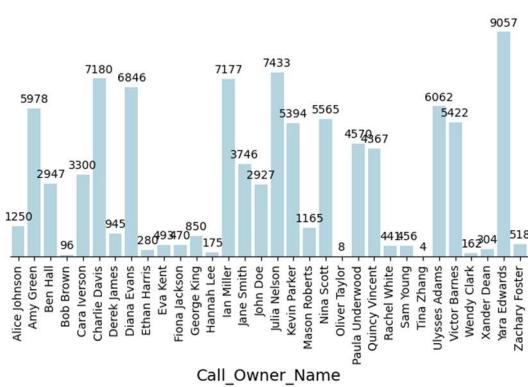
Call\_Status (Unique values: 11):  
 Attended Dialled: 70673 records (73.9%)  
 Unattended Dialled: 16027 records (16.8%)  
 Missed: 5795 records (6.1%)  
 Received: 2958 records (3.1%)  
 Overdue: 56 records (0.1%)  
 Scheduled Attended Delay: 22 records (0.0%)  
 Cancelled: 17 records (0.0%)  
 Scheduled Unattended Delay: 17 records (0.0%)

Scheduled Attended: 14 records (0.0%)  
 Scheduled Unattended: 6 records (0.0%)  
 Scheduled: 3 records (0.0%)

Outgoing\_Call\_Status (Unique values: 5):  
 Completed: 86759 records (90.8%)  
 unknown: 8753 records (9.2%)  
 Overdue: 56 records (0.1%)  
 Cancelled: 17 records (0.0%)  
 Scheduled: 3 records (0.0%)

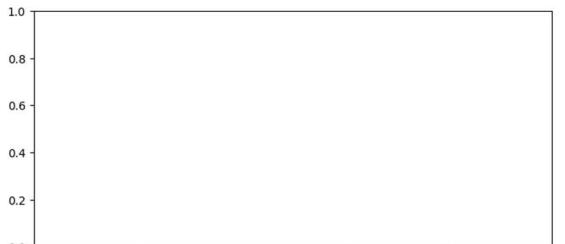
Scheduled\_in\_CRM (Unique values: 3):  
 no: 86700 records (90.7%)  
 unknown: 8753 records (9.2%)  
 yes: 135 records (0.1%)

Distribution of Categorical Variables (Visualisation)



### Типы звонков Call\_Type

- Outbound:** 86,835 (90.8%) - Большинство звонков являются исходящими, что указывает на активную работу отдела продаж;



- **Missed:** 5,794 (6.1%) - Количество пропущенных звонков достаточно велико, свидетельствует о том, что количество сотрудников для обработки звонков недостаточно;
- **Inbound:** 2,959 (3.1%) - доля входящих звонков составляют лишь небольшую часть от общего числа, что может указывать на неэффективную стратегию привлечения клиентов.

#### **Статусы звонков (Call\_Status):**

- **Attended Dialled:** 70673 records (73.9%) - Исходящий звонок, на который ответил адресат. Наибольшее количество звонков было успешно обработано, что говорит о хорошем уровне взаимодействия с клиентами;
- **Unattended Dialled:** 16027 records (16.8%) - Исходящий звонок, на который адресат не ответил (Сотрудник звонит клиенту для подтверждения заказа, но трубка не взята), значение высокое и требует внимания при оценке работы отдела продаж;
- **Missed:** 5795 records (6.1%) - Повторное появление в статусах звонков, что подчеркивает необходимость решения проблемы с пропущенными вызовами;
- **Received:** 2958 records (3.1%) - Входящие звонки, на которые был дан ответ, показывают небольшую долю успешно обработанных входящих звонков. При сравнении с числом пропущенных вызовов показывает значительный потенциал для отдела продаж;
- **Overdue:** 56 records (0.1%) - Просроченный звонок, который должен был быть выполнен в установленный срок (Сотрудник не позвонил клиенту в обозначенное время) можно использовать при персональной оценке сотрудников отдела продаж;
- **Scheduled Attended Delay:** 22 records (0.0%) Запланированный звонок, выполненный с задержкой, клиент ответил. Подчеркивает лояльность и заинтересованность клиента;
- **Cancelled:** 17 records (0.0%) - Звонок, который был отменен до завершения, небольшое количество, свидетельствует о том что ситуация когда клиент завершает звонок до ответа или сотрудник отменяет вызов достаточно редкие для исходящих звонков;
- **Scheduled Unattended Delay:** 17 records (0.0%) - Запланированный звонок, выполненный с задержкой, но клиент не ответил, несущественное количество.
- **Scheduled Attended:** 14 records (0.0%) - Запланированный звонок, выполненный с задержкой, и клиент ответил, несущественное количество.
- **Scheduled Unattended:** 6 records (0.0%) - Запланированный звонок, выполненный с задержкой, но клиент не ответил, несущественное количество.
- **Scheduled:** 3 records (0.0%) - Запланированный звонок, низкое количество, может свидетельствовать о том, что информация в CRM не заносится регулярно.

#### **Статусы исходящих звонков (Outgoing\_Call\_Status):**

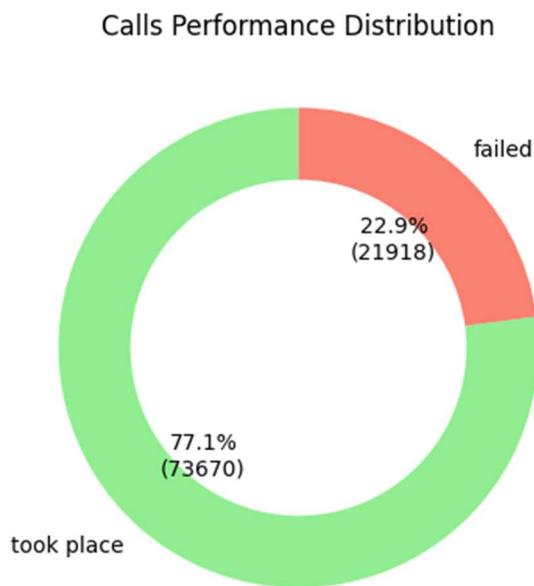
- **Completed:** 86759 records (90.8%) - Высокий процент завершённых звонков. Является суммой Attended Dialled + Unattended Dialled + Overdue + Scheduled из статуса звонков. Дублирование информации с полем статуса звонков ;
- **unknown:** 8753 records (9.2%) - Значительное количество звонков с неопределенным статусом требует внимания, так как это может указывать на необходимость улучшения учёта звонков ;
- **Overdue:** 56 records (0.1%) - Просроченный звонок, небольшое количество. Дублирование со статусом звонков.
- **Cancelled:** 17 records (0.0%) - Звонок, который был отменен до завершения, небольшое количество. Дублирование со статусом звонков.
- **Scheduled:** 3 records (0.0%) - Запланированный звонок, низкое количество, может свидетельствовать о том, что информация в CRM не заносится

#### **Вывод:**

- среди всех звонков преобладают исходящие (90,8%), что с одной стороны говорит о интенсивной работе отдела продаж, с другой стороны может указывать на неэффективную стратегию привлечения клиентов;
- между сотрудниками отдела продаж количество звонков распределено очень неравномерно, только 10 из 32 имеют более 5000 общего количества звонков. Заслуживает отдельного анализа далее;
- достаточно высокий процент пропущенных звонков (6.1%) свидетельствует о недостатках подбора и работы с персоналом в отделе продаж;
- дублирование информации статуса звонков (Call\_Status) со статусом исходящих звонков (Outgoing\_Call\_Status) свидетельствует о нерациональных решениях при проектировании базы данных CRM;
- одинаковое число звонков со значением 'unknown' в поле Outgoing\_Call\_Status ( 9.2%) и поле Scheduled\_in\_CRM ( 9.2%) свидетельствовать о том, что информация в CRM не заносится регулярно.

**Определим продуктивность звонков (performance). Для этого разделим их по группам:**

- звонки, которые имели место и могли стать эффективными; **took\_place**
- звонки статус которых показывает, что они не состоялись **failed**



*Вывод:*

- Из диаграммы видно, что только 77,1% звонков были продуктивными.
- Таким образом определено **существует значительный потенциал** для улучшения работы отдела продаж.

## 2.2 CONTACTS

Загрузка датафрейма из файла 02\_contacts.pkl

**⚠ CONTACTS не содержит числовых полей, следовательно нельзя построить сводную статистику**

**CONTACTS Анализ категориальных полей**

	<b>Id</b>	<b>Contact_Owner_Name</b>
<b>count</b>	18548	18548
<b>unique</b>	18548	27
<b>top</b>	5805028000056907001	Charlie Davis
<b>freq</b>	1	2018

**Id** - идентификатор, следовательно в анализ не включаем

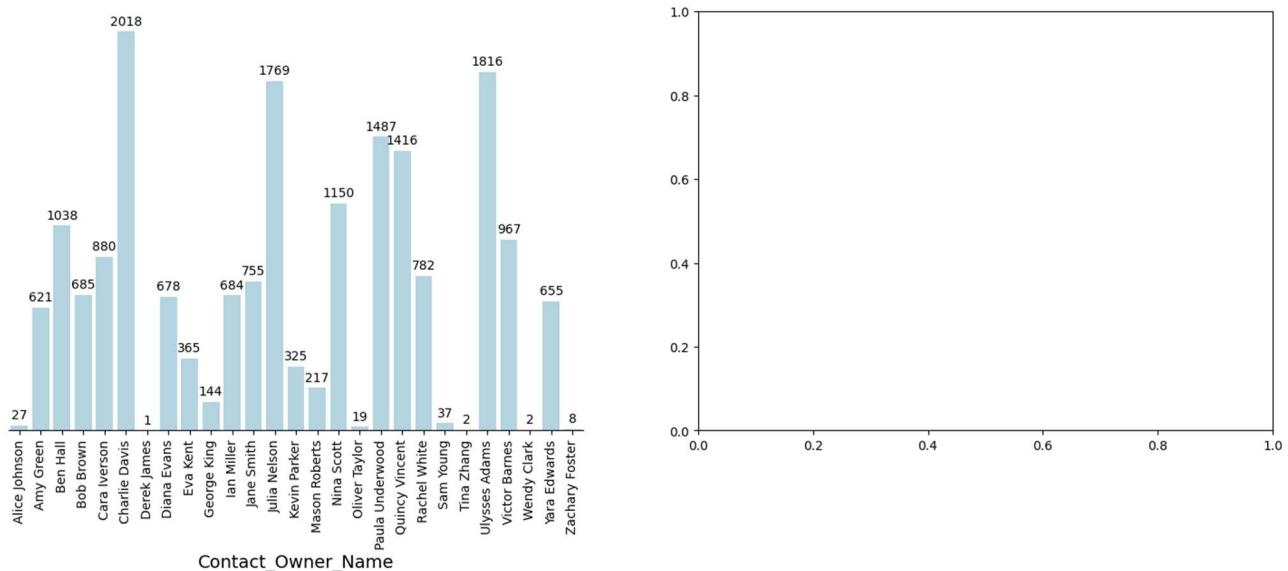
**Distribution of categorical variables:**

Contact\_Owner\_Name (Unique values: 27) :

Charlie Davis: 2018 records (10.9%)  
Ulysses Adams: 1816 records (9.8%)  
Julia Nelson: 1769 records (9.5%)  
Paula Underwood: 1487 records (8.0%)  
Quincy Vincent: 1416 records (7.6%)  
Nina Scott: 1150 records (6.2%)  
Ben Hall: 1038 records (5.6%)  
Victor Barnes: 967 records (5.2%)  
Cara Iverson: 880 records (4.7%)  
Rachel White: 782 records (4.2%)  
Jane Smith: 755 records (4.1%)  
Bob Brown: 685 records (3.7%)  
Ian Miller: 684 records (3.7%)  
Diana Evans: 678 records (3.7%)  
Yara Edwards: 655 records (3.5%)

Amy Green: 621 records (3.3%)  
 Eva Kent: 365 records (2.0%)  
 Kevin Parker: 325 records (1.8%)  
 Mason Roberts: 217 records (1.2%)  
 George King: 144 records (0.8%)  
 Sam Young: 37 records (0.2%)  
 Alice Johnson: 27 records (0.1%)  
 Oliver Taylor: 19 records (0.1%)  
 Zachary Foster: 8 records (0.0%)  
 Tina Zhang: 2 records (0.0%)  
 Wendy Clark: 2 records (0.0%)  
 Derek James: 1 records (0.0%)

Distribution of Categorical Variables (Visualisation)



#### Вывод:

В датафрейме представлено только одно категориальное поле **Contact\_Owner\_Name** которое отражает число контактов у каждого сотрудника отдела продаж. Среди 27 сотрудников наблюдается большой разброс от 1 (Derek James) до 2018/10.9% (Charlie Davis) контактов. Эта ситуация свидетельствует о неравномерной загрузке сотрудников, вероятно в зависимости от их уровня профессиональной подготовки, а так же возможно от продолжительности работы в отделе.

### 2.3 SPEND

Загрузка датафрейма из файла 03\_spend.pkl

#### SPEND Рассчет сводной статистики

	Impressions	Spend	Clicks
<b>count</b>	19862.00	19862.00	19862.00
<b>mean</b>	2571.70	7.53	25.10
<b>std</b>	11691.23	27.33	87.03
<b>min</b>	0.00	0.00	0.00
<b>25%</b>	1.00	0.00	0.00
<b>50%</b>	82.00	0.74	2.00
<b>75%</b>	760.75	6.16	13.00
<b>max</b>	431445.00	774.00	2415.00

#### SPEND Визуализация сводной статистики

##### Statistical characteristics of number fields:

Impressions:

Max: 431445.00

Min: 0.00

Mean: 2571.70

```
std: 11691.23  
Median: 82.00  
Mode: 0.00  
Range: 431445.00
```

---

Spend:

```
Max: 774.00  
Min: 0.00  
Mean: 7.53  
std: 27.33  
Median: 0.74  
Mode: 0.00  
Range: 774.00
```

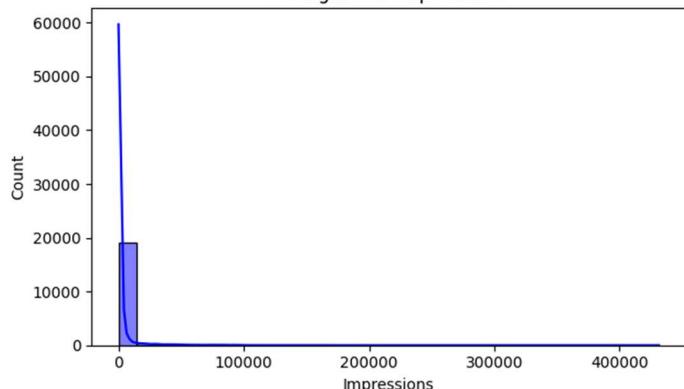
---

Clicks:

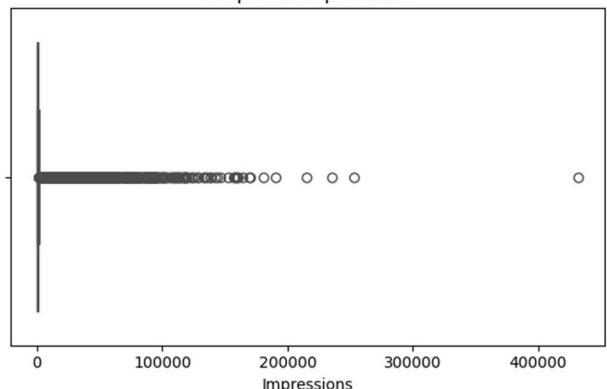
```
Max: 2415.00  
Min: 0.00  
Mean: 25.10  
std: 87.03  
Median: 2.00  
Mode: 0.00  
Range: 2415.00
```

---

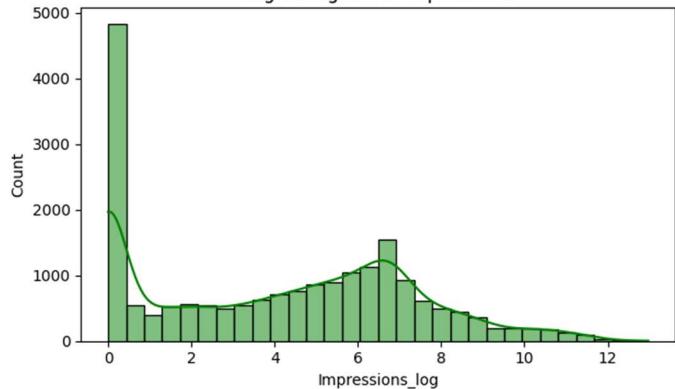
Histogram of Impressions



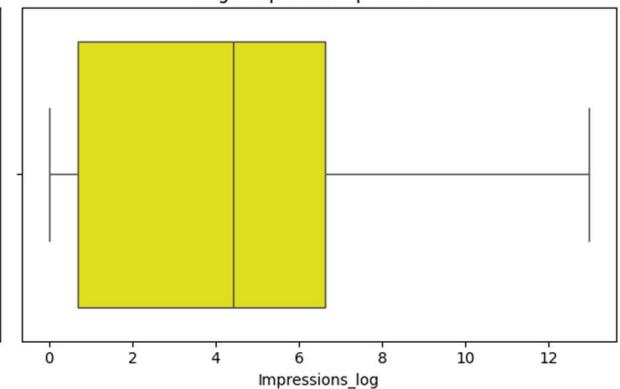
Boxplot of Impressions



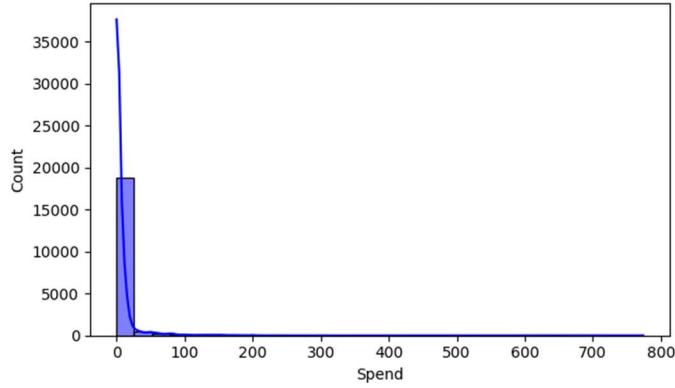
Log Histogram of Impressions



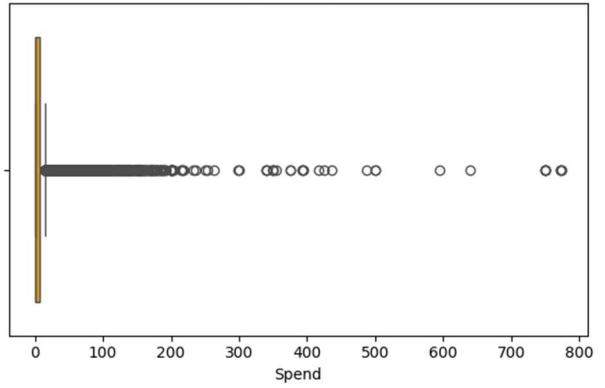
Log Boxplot of Impressions

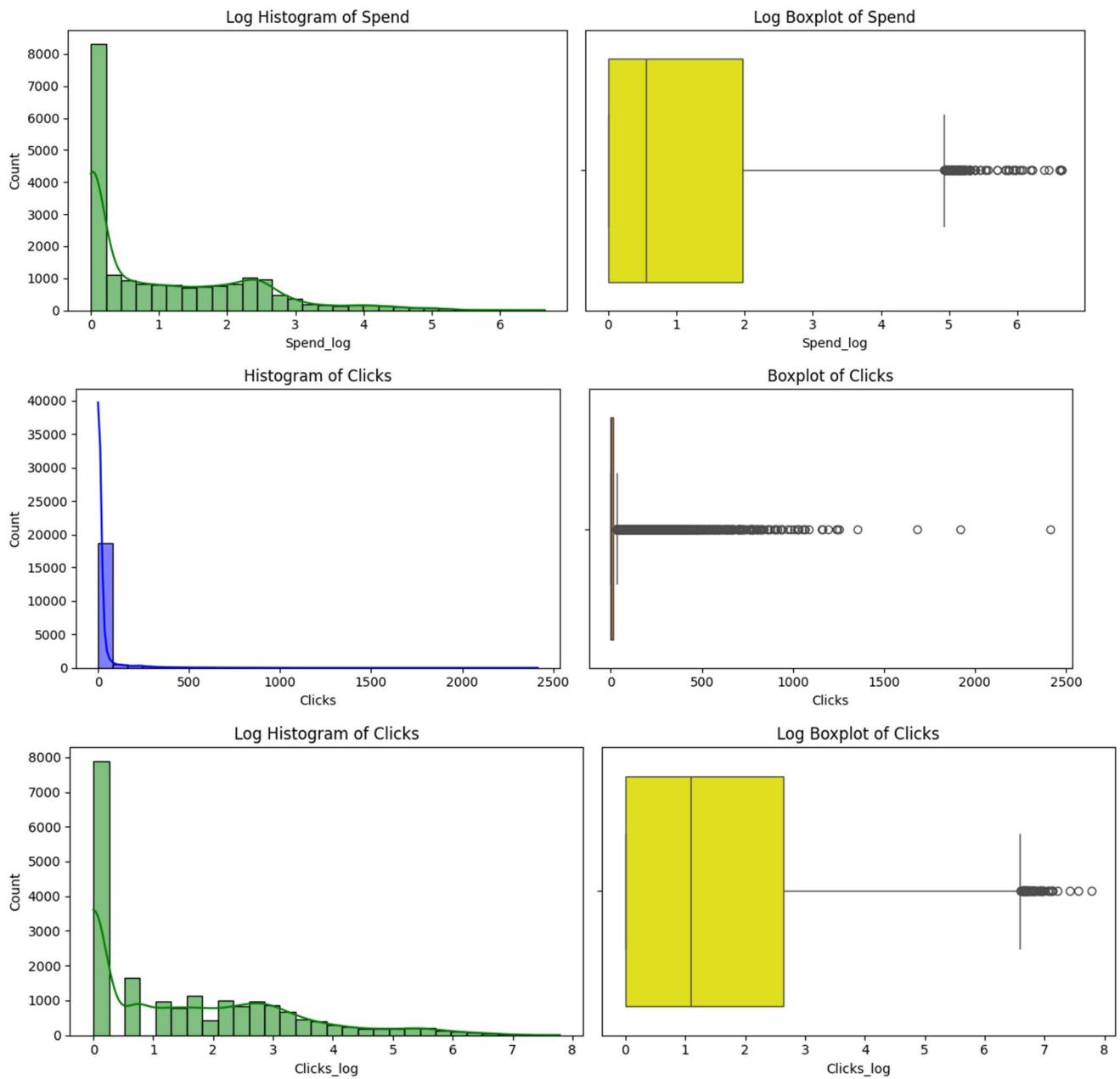


Histogram of Spend



Boxplot of Spend





	Date	Impressions	Spend	Clicks	Impressions_log	Spend_log	Clicks_log
<b>count</b>	19862		19862.00	19862.00	19862.00	19862.00	19862.00
<b>mean</b>	2024-01-10 18:21:55.879568896	2571.70	7.53	25.10	4.08	1.06	1.51
<b>min</b>	2023-07-03 00:00:00	0.00	0.00	0.00	0.00	0.00	0.00
<b>25%</b>	2023-10-09 00:00:00	1.00	0.00	0.00	0.69	0.00	0.00
<b>50%</b>	2024-01-20 00:00:00	82.00	0.74	2.00	4.42	0.55	1.10
<b>75%</b>	2024-04-12 00:00:00	760.75	6.16	13.00	6.64	1.97	2.64
<b>max</b>	2024-06-21 00:00:00	431445.00	774.00	2415.00	12.97	6.65	7.79
<b>std</b>	NaN	11691.23	27.33	87.03	3.18	1.23	1.64

### SPEND Анализ сводной статистики

- **Date:** данные охватывают период с июля 2023 года по июнь 2024 года. Медианная запись приходится на январь 2024 года.
- **Impressions** - Количество показов рекламы пользователям:
  - наблюдается очень большой разброс данных (диапазон - 431445);

- при этом 75% записей имеют относительно небольшое количество показов (до 760 - 3 квантиль), однако у нескольких записей экстремально большие значения (см.соответствующий бохплот) которые значительно увеличивают на среднее значение;
- 25% записей имеют до одного показа;
- среднее число показов составляет 2571, а медианное (50-й перцентиль) — 82. Это говорит о сильной асимметрии данных, так как среднее значение существенно выше медианы.;
- стандартное отклонение (11691) достаточно высокое, что указывает на сильную изменчивость в количестве показов
- **Spend** - Количество денег, потраченных на рекламную кампанию или группу объявлений за указанный период:
  - среднее значение затрат (7.53), медиана (0.74) значительно ниже среднего, что указывает на большое количество очень низких расходов, возможно предположить ставку на бесплатную рекламу;
  - большинство записей имеют низкие или нулевые затраты, но есть несколько записей с высокими затратами, что приводит к высокому среднему и стандартному отклонению (27.33);
- **Clicks** - Количество нажатий пользователей на рекламу:
  - 75 % записей содержат до 13 кликов, что при среднем значении (25.10) и медиане (2) говорит о том, что распределение данных по кликам с перекосом в сторону низких значений и в целом достаточно низкой эффективностью большинства рекламы;
  - максимальное число кликов (2415) указывают на наличие рекламных кампаний с крайне высокой активностью.

#### **Выводы:**

- в датафрейме 19862 записей, что представляет достаточно большой объем данных для анализа рекламной активности;
- все числовые значения характеризуются большим диапазоном данных и перекосом в сторону низких значений;
- минимальные значения во всех столбцах равны 0, что означает, что были кампании без показов, затрат и кликов;
- максимальное число показов 431 445, максимальные затраты 774, максимальное число кликов 2 415 — указывают на рекламные кампании с крайне высокой активностью.
- высокое стандартное отклонение показов (11691.23) и кликов (87.03) говорит о сильной изменчивости данных.

#### **SPEND Анализ категориальных полей**

	Source	Campaign	AdGroup	Ad	AdvCountry
<b>count</b>	19862	19862	19862	19862	19862
<b>unique</b>	14	52	26	177	4
<b>top</b>	Facebook Ads	unknown	wide	unknown	Germany
<b>freq</b>	9569	3230	6002	5911	13636

#### **Distribution of categorical variables:**

Source (Unique values: 15):

- Facebook Ads: 9569 records (48.2%)
- Tiktok Ads: 2985 records (15.0%)
- Youtube Ads: 1784 records (9.0%)
- Google Ads: 1266 records (6.4%)
- Telegram posts: 836 records (4.2%)
- Webinar: 766 records (3.9%)
- Bloggers: 632 records (3.2%)
- SMM: 571 records (2.9%)
- Organic: 514 records (2.6%)
- CRM: 355 records (1.8%)
- Test: 262 records (1.3%)
- Partnership: 234 records (1.2%)
- Offline: 61 records (0.3%)
- Radio: 27 records (0.1%)
- unknown: 0 records (0.0%)

Campaign (Unique values: 52):  
unknown: 3230 records (16.3%)  
12.07.2023wide\_DE: 2423 records (12.2%)  
02.07.23wide\_DE: 1685 records (8.5%)  
youtube\_shorts\_DE: 1578 records (7.9%)  
03.07.23women: 1508 records (7.6%)  
04.07.23recentlymoved\_DE: 1399 records (7.0%)  
07.07.23LAL\_DE: 1182 records (6.0%)  
12.09.23interests\_Uxui\_DE: 1144 records (5.8%)  
15.07.23b\_DE: 530 records (2.7%)  
24.09.23retargeting\_DE: 505 records (2.5%)  
gen\_analyst\_DE: 405 records (2.0%)  
web2408\_DE: 403 records (2.0%)  
performancemax\_eng\_DE: 355 records (1.8%)  
07.12.23test\_DE: 269 records (1.4%)  
20.03.2024wide\_PL: 241 records (1.2%)  
30.11.23wide\_DE: 234 records (1.2%)  
05.07.23interests\_DE: 215 records (1.1%)  
17.03.24wide\_AT: 199 records (1.0%)  
20.03.24interests\_WebDev\_PL: 169 records (0.9%)  
discovery\_DE: 166 records (0.8%)  
brand\_search\_eng\_DE: 146 records (0.7%)  
08.04.24wide\_webinar\_DE: 134 records (0.7%)  
youtube\_shortsin\_AT: 133 records (0.7%)  
20.05.24interests\_DE: 131 records (0.7%)  
20.03.24\_widde\_PL: 130 records (0.7%)  
15.03.2024wide\_AT: 112 records (0.6%)  
15.11.23wide\_webinar\_DE: 108 records (0.5%)  
01.04.23women\_PL: 103 records (0.5%)  
performancemax\_wide\_AT: 93 records (0.5%)  
05.09.2023wide\_DE: 89 records (0.4%)  
... (too many categories, showing first 30)

---

AdGroup (Unique values: 26):  
wide: 6002 records (30.2%)  
unknown: 3167 records (15.9%)  
b: 2246 records (11.3%)  
women: 1611 records (8.1%)  
recentlymoved: 1444 records (7.3%)  
LAL1: 1222 records (6.2%)  
Com\_august: 1073 records (5.4%)  
interest\_work\_WebDev: 733 records (3.7%)  
interest\_programming\_WebDev: 636 records (3.2%)  
retargeting: 505 records (2.5%)  
Com\_march: 206 records (1.0%)  
interest\_work: 181 records (0.9%)  
discovery: 166 records (0.8%)  
Com\_july\_1: 150 records (0.8%)  
interest\_programming: 121 records (0.6%)  
promoposts\_b: 71 records (0.4%)  
wide\_python-developer: 56 records (0.3%)  
wide\_qa-engineer: 50 records (0.3%)  
berlin\_wide: 49 records (0.2%)  
wide\_webdesigner: 42 records (0.2%)  
interest\_all: 40 records (0.2%)  
interest\_python-developer: 28 records (0.1%)  
interest\_dataanalytics: 27 records (0.1%)  
accountant\_wide: 21 records (0.1%)  
interest\_programming - Copy: 8 records (0.0%)  
interest\_dataanalytics - Copy: 7 records (0.0%)

---

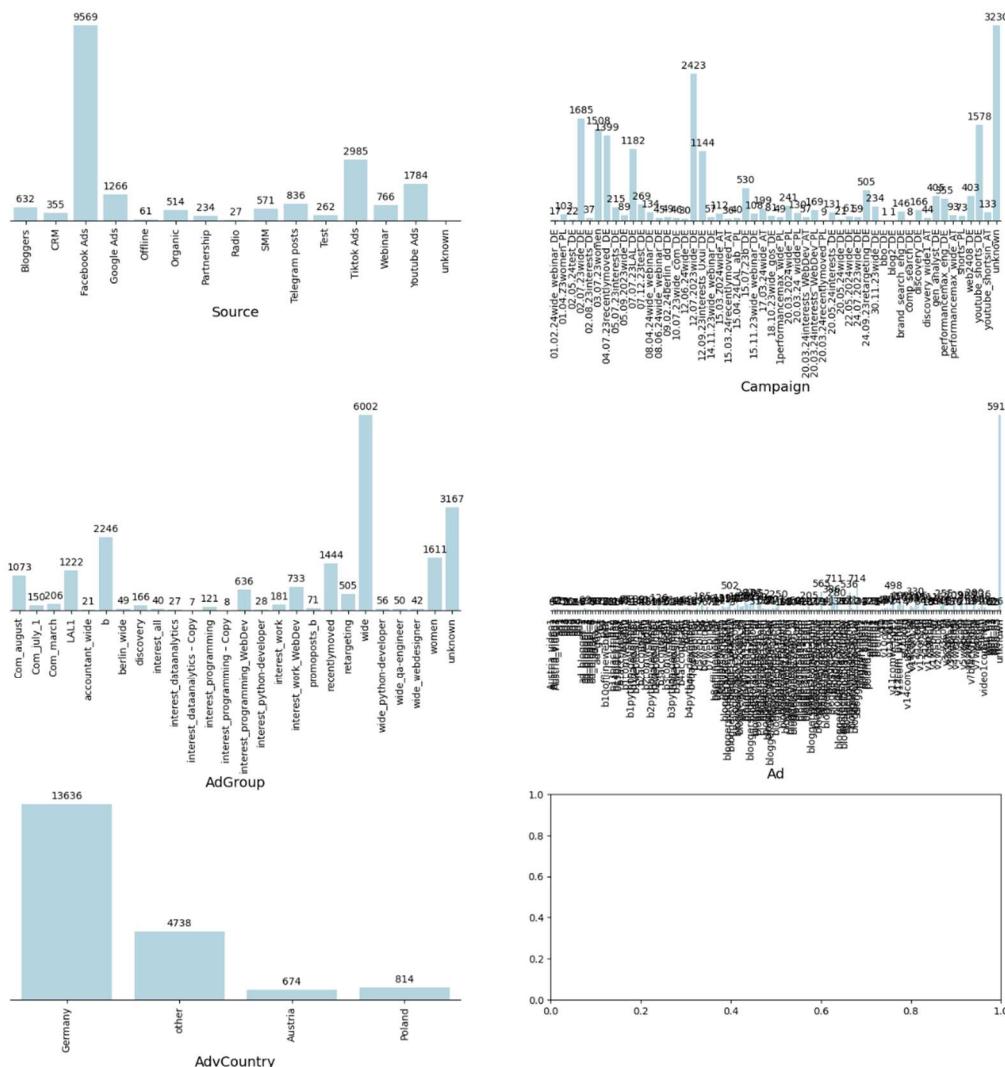
Ad (Unique values: 177):  
unknown: 5911 records (29.8%)  
bloggersvideo9com: 714 records (3.6%)  
bloggersvideo5: 711 records (3.6%)  
bloggersvideo3com: 565 records (2.8%)  
bloggersvideo8com: 536 records (2.7%)  
bloggersvideo11: 502 records (2.5%)  
v11comwebdev: 498 records (2.5%)  
bloggersvideo4com: 386 records (1.9%)  
v15: 330 records (1.7%)  
bloggersvideo14com: 287 records (1.4%)  
bloggersvideo16com: 285 records (1.4%)  
v7com: 280 records (1.4%)  
bloggersvideo5com: 280 records (1.4%)  
v3com: 255 records (1.3%)  
bloggersvideo17com: 252 records (1.3%)  
bloggersvideolcom: 250 records (1.3%)  
v9com: 236 records (1.2%)  
v6com: 228 records (1.1%)

```
v4com: 209 records (1.1%)
v15com: 206 records (1.0%)
bloggersvideo2com: 205 records (1.0%)
bloggersvideo15com: 197 records (1.0%)
v13com_python: 192 records (1.0%)
v5com: 190 records (1.0%)
bloggersvideo12com_at: 185 records (0.9%)
b6: 185 records (0.9%)
bloggersvideo15com_python: 183 records (0.9%)
v14: 170 records (0.9%)
v3: 162 records (0.8%)
v12com_python: 156 records (0.8%)
... (too many categories, showing first 30)
```

AdvCountry (Unique values: 4):

Germany: 13636 records (68.7%)  
other: 4738 records (23.9%)  
Poland: 814 records (4.1%)  
Austria: 674 records (3.4%)

## Distribution of Categorical Variables (Visualisation)



## **Анализ распределения данных**

- **Source** - источники рекламы (Канал, на котором было показано объявление):
    - 15 уникальных значений, из которых один источник **Test** вероятно использовался как тестовый (1,3% записей) но из дальнейшего анализа не будем исключать, т.к. полной уверенности нет. Будем расценивать его как специфический источник;
    - Facebook Ads (48.2%) явный лидер по числу записей;

- в первую пятерку источников рекламы по числу записей входят **Tiktok Ads(15.0%)**, **Youtube Ads(9.0%)**, **Google Ads(6.4%)** и **Telegram posts(4.2%)**
- остальные источники суммарно насчитывают 17,2% записей;
- для проверки эффективности рекламы в дальнейшем потребуется рассчитать и проанализировать:
  - суммарные затраты по каждому источнику **CpS** (Cost per Source);
  - **CpC** (Cost per Click) стоимость одного клика;
  - показатель кликабельности **CTR** (Click-Through Rate) — сколько пользователей нажали на рекламное объявление или ссылку после просмотра;
  - конверсию в первую покупку **C1**.
- **Campaign** - Кампания, в рамках которой было показано объявление (52 уникальных значения). Наличие 16,3% значений с неизвестным статусом и лидером по числу записей **12.07.2023wide\_DE(12.2%)**. Тестовые кампании - '07.12.23test\_DE' и '02.05.24test\_DE'.
- **AdGroup** - Подмножество в кампании, содержащее одно или несколько объявлений с одинаковыми целями или настройками (26 уникальных значений). Наличие 15,9% значений с неизвестным статусом и лидером по числу записей **wide(30.2%)**. Тестовые группы: 'b', 'wide\_webdesigner', 'wide\_python-developer', 'wide\_qa-engineer', 'interest\_python-developer', 'accountant\_wide';
- **Ad** - Конкретная реклама, показываемая пользователям (177 уникальных значений). Наличие 29,8% значений с неизвестным статусом без явного лидера по числу записей от 0 до 3,6%. Тестовые рекламы: *b1accountant, b2accountant, b3accountant, b4accountant*.

#### **Выход**

- Категориальные данные отражают стратегию маркетинга онлайн школы и представляют собой иерархическую структуру Source-Campaign-AdGroup-Ad;
- Наличие достаточного количества отсутствующих значений в полях Campaign и AdGroup могут внести искажения в расчеты и выводы и говорят о существенных недостатках ведения базы данных CRM и необходимости доработки интерфейса ввода данных;
- Для определения эффективности маркетинга потребуется рассчитать ряд дополнительных показателей таких как CpS, CpC, CTR, C1

## **2.4 DEALS**

### **DEALS Рассчет сводной статистики**

```
# загружаем датафрейм
deals = pd.read_pickle("04_deals.pkl")
deals.describe(include=['int', 'int8', 'float']).\n    drop(columns=['Id', 'Contact_Name'], errors='ignore').T
```

Out[20]:

	count	mean	std	min	25%	50%	75%	max
<b>Course_duration</b>	3587.00	10.20	1.83	6.00	11.00	11.00	11.00	11.00
<b>Months_of_study</b>	840.00	5.44	2.92	0.00	3.00	5.00	8.00	11.00
<b>Initial_Amount_Paid</b>	4165.00	929.80	1347.44	0.00	300.00	1000.00	1000.00	11000.00
<b>Offer_Total_Amount</b>	4185.00	7128.72	4636.02	0.00	3000.00	11000.00	11000.00	11500.00
<b>SLA_sec</b>	21593.00	83320.00	627455.11	0.00	0.00	6335.00	46660.00	26908464.00
<b>Paid</b>	21593.00	183.19	1102.21	0.00	0.00	0.00	0.00	12550.00
<b>T</b>	21593.00	0.21	1.20	0.00	0.00	0.00	0.00	11.00
<b>AOV</b>	21593.00	36.91	205.89	0.00	0.00	0.00	0.00	2600.00

### **DEALS Визуализация сводной статистики**

Statistical characteristics of number fields:

Course\_duration:

Max: 11.00  
Min: 6.00  
Mean: 10.20  
std: 1.83  
Median: 11.00  
Mode: 11.00  
Range: 5.00

-----

Months\_of\_study:

Max: 11.00  
Min: 0.00  
Mean: 5.44  
std: 2.92  
Median: 5.00  
Mode: 6.00  
Range: 11.00

-----

Initial\_Amount\_Paid:

Max: 11000.00  
Min: 0.00  
Mean: 929.80  
std: 1347.44  
Median: 1000.00  
Mode: 1000.00  
Range: 11000.00

-----

Offer\_Total\_Amount:

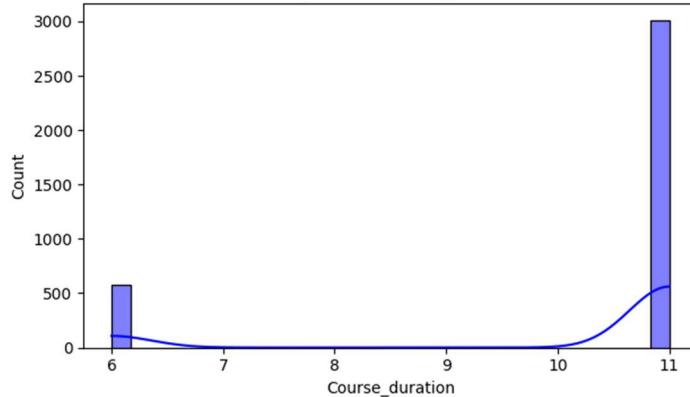
Max: 11500.00  
Min: 0.00  
Mean: 7128.72  
std: 4636.02  
Median: 11000.00  
Mode: 11000.00  
Range: 11500.00

-----

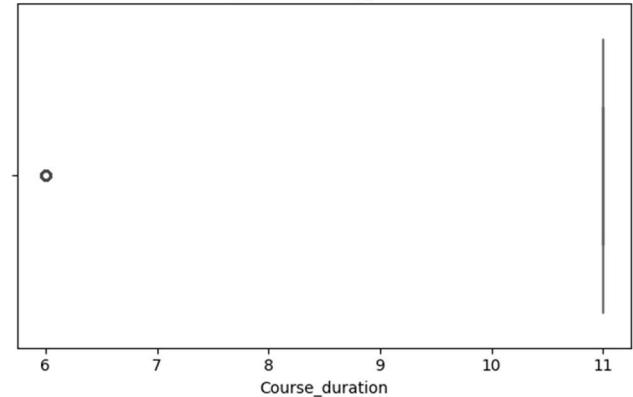
SLA\_sec:

Max: 26908464.00  
Min: 0.00  
Mean: 83320.00  
std: 627455.11  
Median: 6335.00  
Mode: 0.00  
Range: 26908464.00

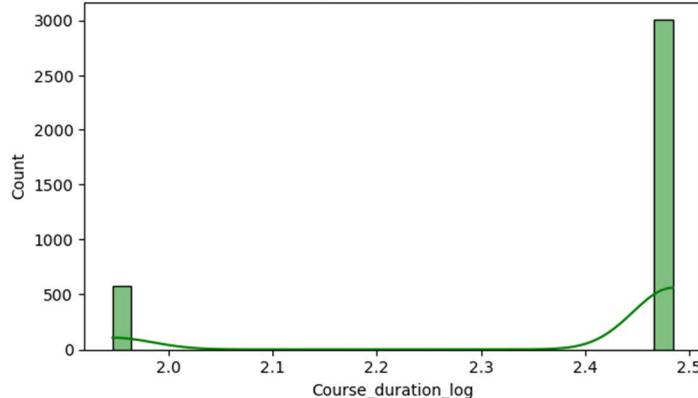
Histogram of Course\_duration



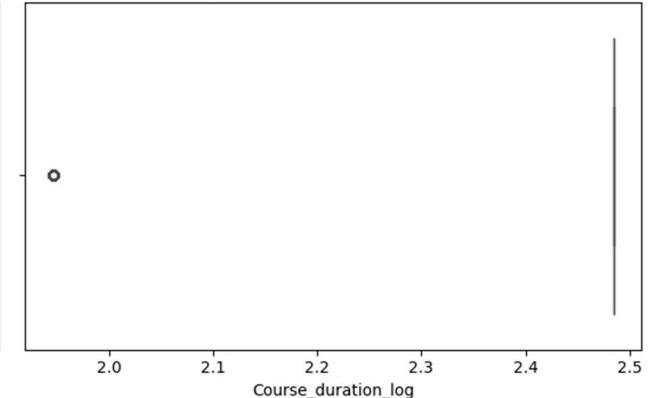
Boxplot of Course\_duration

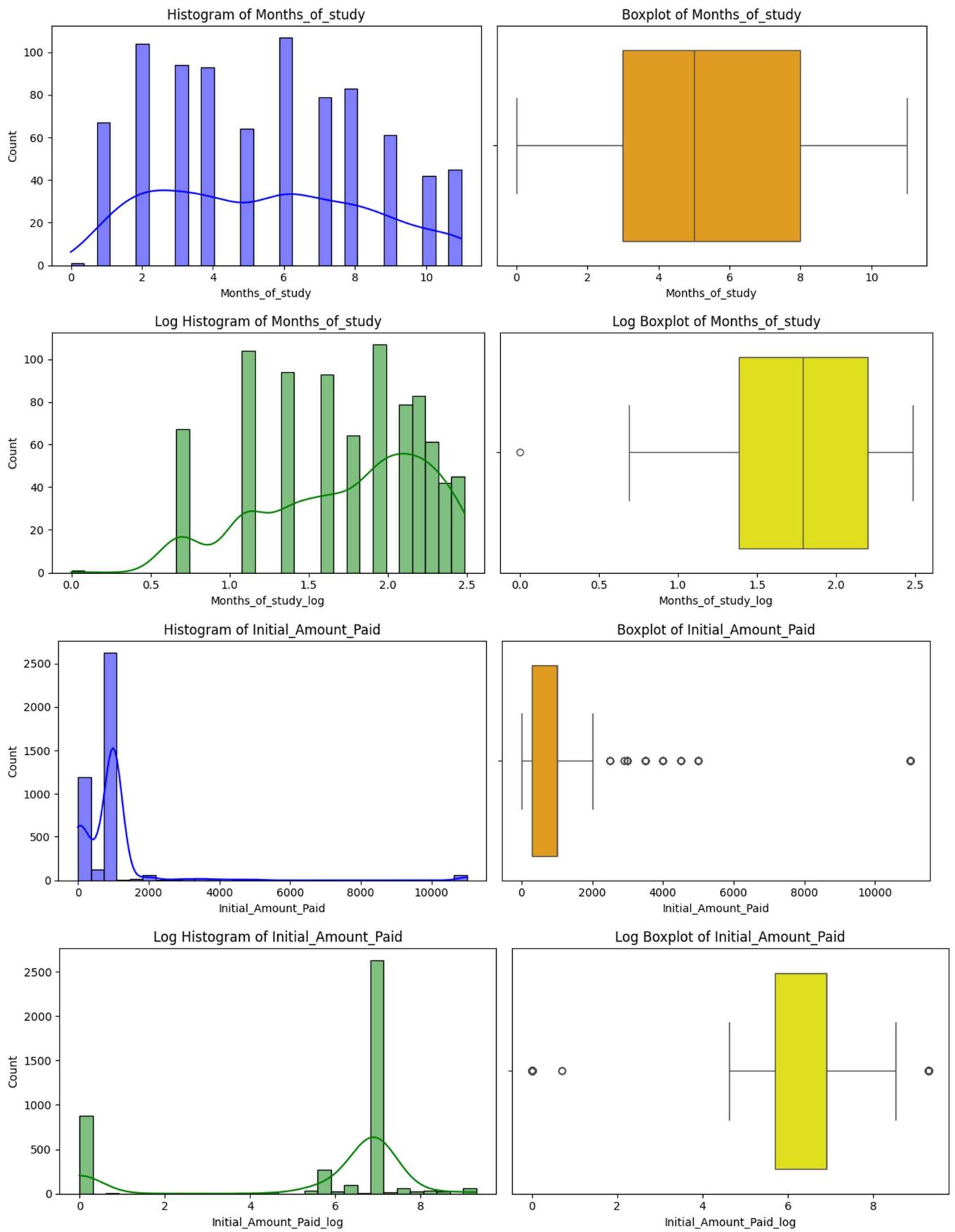


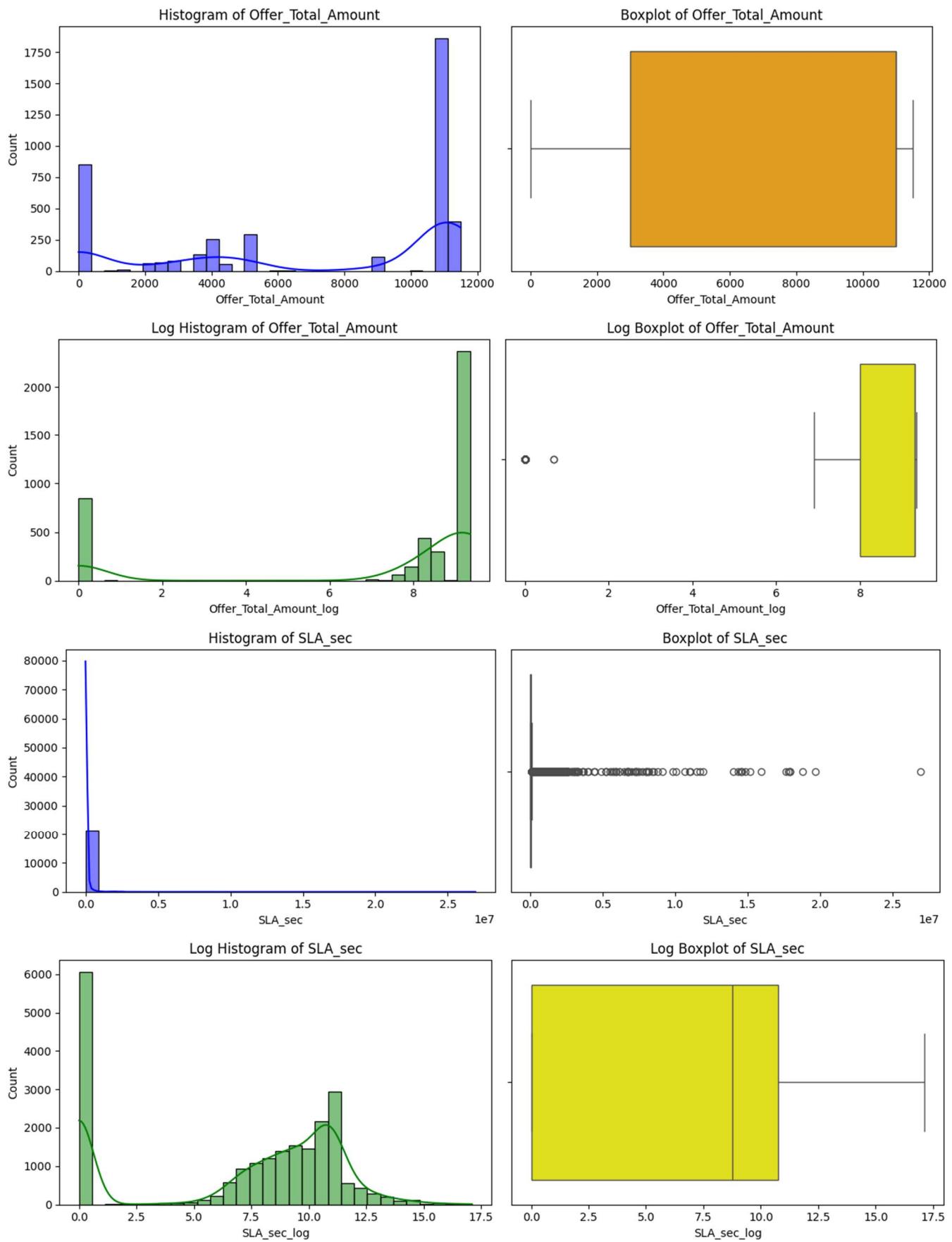
Log Histogram of Course\_duration



Log Boxplot of Course\_duration







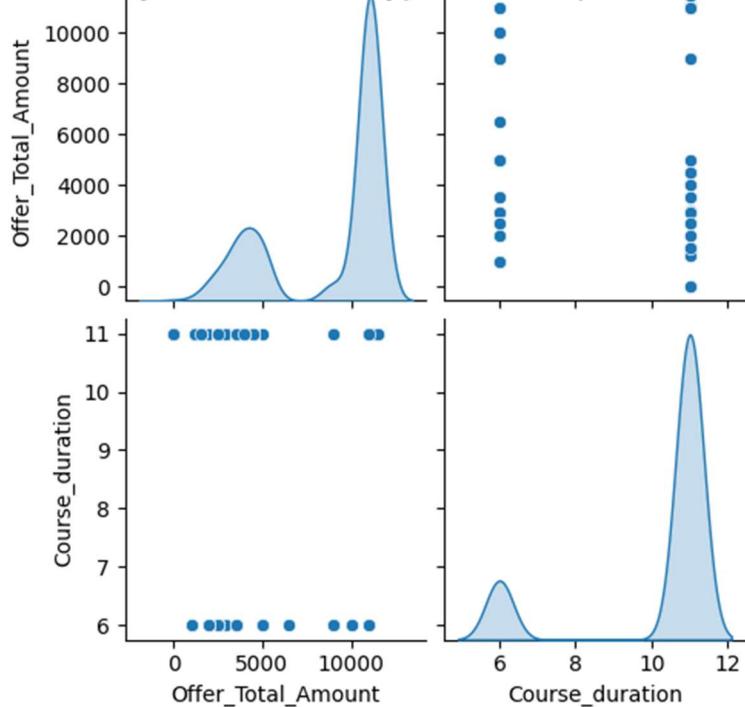
#### Определим степень зависимости переменных *Offer\_Total\_Amount* и *Course\_duration*.

Для этого построим диаграмму рассеяния (`sns.pairplot()`), чтобы визуально оценить тренд и создадим тепловую карту корреляции (`sns.heatmap()`), чтобы наглядно отобразить силу связи.

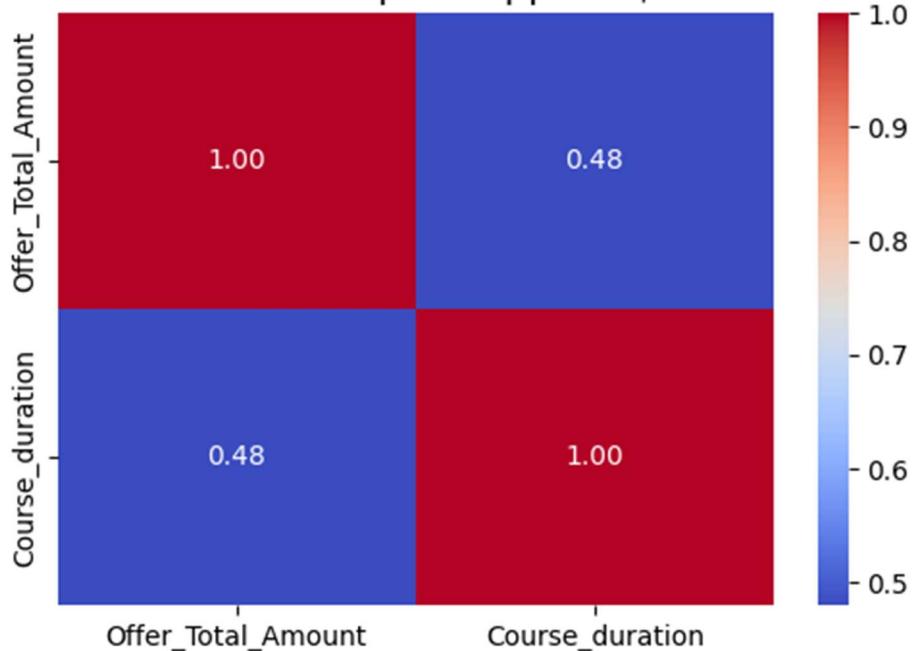
Корреляционная матрица:

	Offer_Total_Amount	Course_duration
Offer_Total_Amount	1.00	0.48
Course_duration	0.48	1.00

Корреляция между стоимостью курса и его продолжительностью



Тепловая карта корреляции

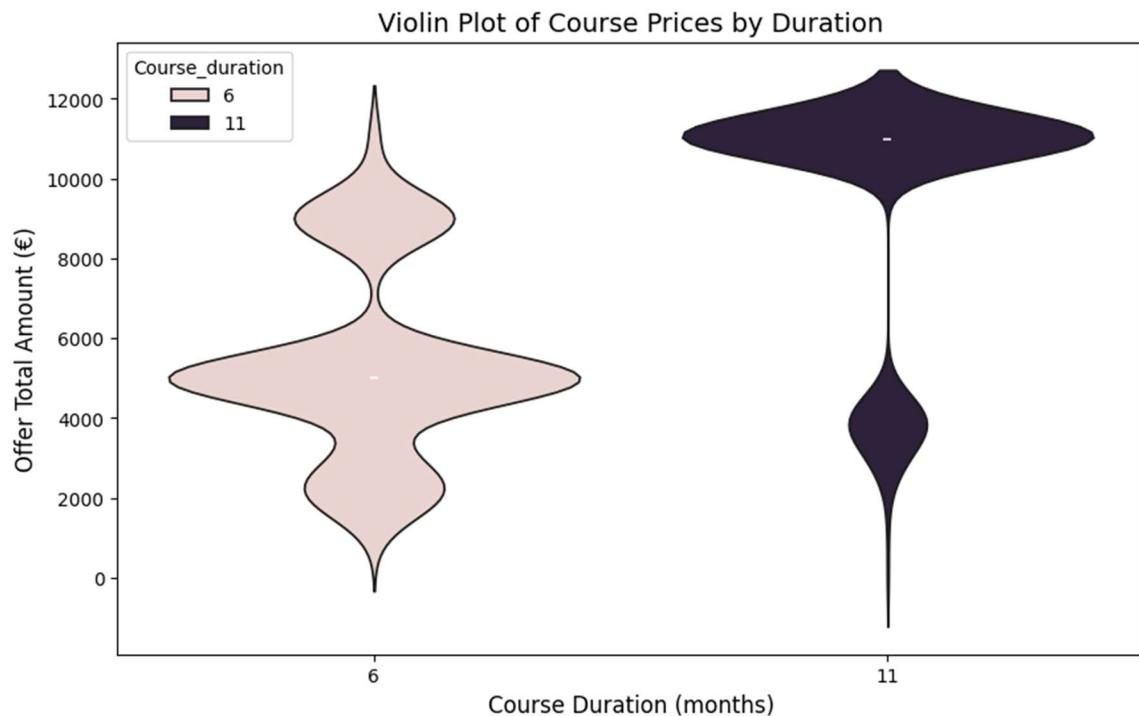


- Корреляция между **Offer\_Total\_Amount** (стоимость курса) и **Course\_duration** (продолжительность курса) составляет 0.48.
- Это означает умеренно положительную корреляцию: чем дороже курс, тем дольше его продолжительность, но взаимосвязь не является строгой, т.е. это означает, что стоимость курса частично влияет на его продолжительность, но есть и другие факторы.

Возможные интерпретации:

- Дорогие курсы чаще всего длиннее, но есть исключения.
- Курсы средней стоимости могут иметь разную продолжительность.
- Дешёвые курсы могут быть как короткими, так и долгими (например, субсидируемые программы).
- Не все дорогие курсы обязательно самые длинные, возможно, есть премиальные, но короткие программы.

*Для дальнейшего анализа построим виолинплоты стоимости курсов по их продолжительности, чтобы понять распределение данных:*



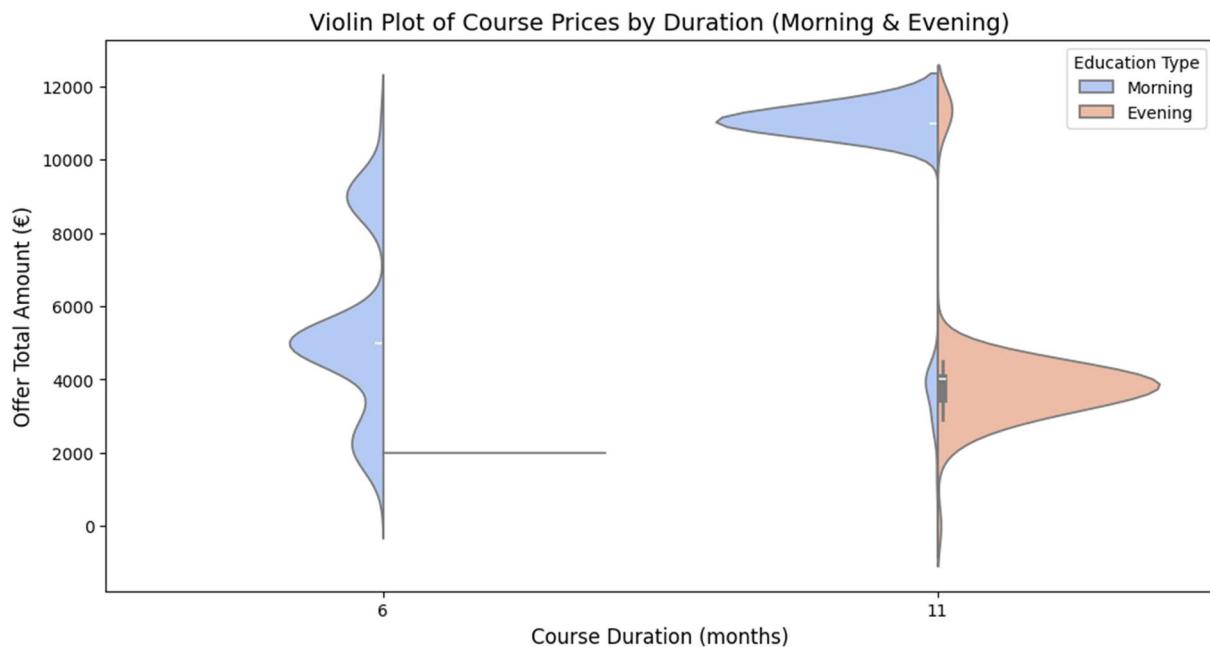
- violinplot показывает не только медиану и квартильные границы, но и размытость данных, то есть где находится больше всего точек.
- широкие зоны показывают скопление курсов с определёнными ценами

#### Основные выводы:

- Разброс цен для 6-месячных курсов:
  - широкое распределение цен с пиками около 2000€, 6000€ и 10000€.
  - это говорит о том, что 6-месячные курсы могут быть как недорогими, так и премиальными.
- Разброс цен для 11-месячных курсов:
  - цены сосредоточены вокруг 8 000€.
  - меньшая вариативность цен — возможно, эти курсы имеют фиксированные стоимости.
- Общая тенденция:
  - длительные курсы (11 месяцев) имеют более стабильные цены.
  - короткие курсы (6 месяцев) варьируются сильнее, что может означать разные уровни сложности или типов обучения.

**Таким образом :** Стоимость длинных курсов (11 месяцев) в основном выше, но короткие (6 месяцев) могут быть как доступными, так и премиальными.

Для более полного анализа добавим колонку *Education\_Type*, которая уточняет тип курсов: утренние или вечерние



- Для 6-месячных курсов:
  - Morning курсы имеют более широкий разброс цен, с основными пиками около 4 000€ и 8 000€.
  - Evening курсы представлены гораздо меньшим числом данных, и их распределение более узкое.
- Для 11-месячных курсов:
  - Morning курсы имеют большую концентрацию около 8 000€.
  - Evening курсы сосредоточены в районе 4 000€, что указывает на возможное разделение ценовых сегментов между утренними и вечерними программами.
- Общие тенденции:
  - Morning курсы чаще дороже, особенно при большой продолжительности (11 месяцев).
  - Evening курсы менее распространены, и их стоимость, как правило, ниже, особенно на длинных курсах.
- **Вывод:** Стоимость курсов зависит от типа обучения.
  - Morning курсы более дорогие и разнообразные по цене.
  - Evening курсы имеют меньший разброс стоимости, особенно на долгосрочных программах.
  - Курсы на 11 месяцев чаще стоят дороже, особенно если они проводятся утром

#### DEALS Анализ сводной статистики

1. Датафрейм содержит 21 593 записи, но не все столбцы полностью заполнены (например, *Course\_duration* имеет 3 580 записей, *Months\_of\_study* — 840). ¶
2. Анализ продолжительности обучения (*Course\_duration*) Средняя продолжительность курса: 10.2 месяца, но медиана и 75-й перцентиль равны 11 месяцам, что говорит о стандартной длительности учебных программ.
3. Время реального обучения (*Months\_of\_study*):
  - Среднее 5.44 месяцев, но широкий разброс значений (стандартное отклонение 2.92). Максимальное значение 11 месяцев
  - Большинство студентов учатся от 3 до 8 месяцев (по квартилям).
  - Минимальное значение 0 месяцев, что может указывать на студентов, которые только начали курс или ждут начала обучения.
4. Финансовые показатели (*Initial\_Amount\_Paid*, *Offer\_Total\_Amount*) Средний размер платежа: 930, но медианное значение 1000, что говорит о смещении распределения. Разброс значений: Стандартное отклонение 1348.83, что указывает на значительную вариативность. Минимальная сумма 0, что возможно связано с полными скидками или отсутствием платежа. Максимальная сумма 11 000, что подтверждает наличие дорогих курсов.
5. Общая сумма предложения (*Offer\_Total\_Amount*):
  - Среднее 7 128, медиана 11 000, что говорит о большом разрыве между фактическими платежами и предложенной ценой.

- Нулевая минимальная стоимость → возможны бесплатные предложения.
  - Максимальная стоимость 11 500, что говорит о стандартном ценовом диапазоне курсов.
  - SLA и время обработки (SLA\_sec)
  - Среднее время ответа: 83 319 секунд ( $\approx$  23 часа), но медианное значение всего 6 335 секунд ( $\approx$  1.76 часа).
  - Сильный разброс значений (стандартное отклонение 627 455 секунд):
  - Минимальное 0, что указывает на мгновенные ответы.
  - Максимальное время ответа — 26 908 464 секунд ( $\approx$  312 дней), что указывает на аномалии или долгие задержки.
7. Логарифмированные данные (log версии столбцов) Логарифмическое преобразование помогло нормализовать распределения:
- Initial\_Amount\_Paid\_log: среднее 5.4, медиана 6.9 → данные стали ближе к нормальному распределению.
  - Offer\_Total\_Amount\_log: среднее 7.16, медиана 9.3 → аналогично скорректировано влияние высоких значений.
  - SLA\_sec\_log: среднее 6.98, медиана 8.75 → снижено влияние выбросов.
  - Course\_duration\_log и Months\_of\_study\_log показали небольшую вариативность, но логарифмическое преобразование привело их к компактному диапазону.

#### **Выводы:**

- Курсы в основном делятся 6 и 11 месяцев.
- Большинство предложений и первоначальных платежей сосредоточены вокруг значений 1000 и 11000 соответственно, что указывает на возможное фиксированное ценообразование.
- В целом, существует зависимость между длительностью и ценой курса, но она не является строгой - стоимость курса оказывает умеренное влияние на его продолжительность, но есть и другие важные факторы, например как тип курса утренний (когда обучение преимущественно является основным видом деятельности для студента на данный момент) или вечерний (в большинстве случаев работающие студенты).
- Время отклика по SLA значительно варьируется, что говорит о различной оперативности обслуживания клиентов. Причины длительных (свыше 10 дней) и экстремальных ( $\approx$  312 дней) задержек в обслуживании требуют обсуждения в отделе продаж.

#### **DEALS Анализ категориальных полей**

##### **Distribution of categorical variables:**

```
Deal_Owner_Name (Unique values: 27):
Charlie Davis: 2963 records (13.7%)
Julia Nelson: 2241 records (10.4%)
Ulysses Adams: 2166 records (10.0%)
Quincy Vincent: 1884 records (8.7%)
Paula Underwood: 1862 records (8.6%)
Ben Hall: 1345 records (6.2%)
Nina Scott: 1283 records (5.9%)
Victor Barnes: 1232 records (5.7%)
Cara Iverson: 1056 records (4.9%)
Diana Evans: 1015 records (4.7%)
Jane Smith: 988 records (4.6%)
Rachel White: 871 records (4.0%)
Kevin Parker: 574 records (2.7%)
Ian Miller: 497 records (2.3%)
Eva Kent: 460 records (2.1%)
Bob Brown: 340 records (1.6%)
Mason Roberts: 268 records (1.2%)
Oliver Taylor: 163 records (0.8%)
George King: 109 records (0.5%)
Yara Edwards: 85 records (0.4%)
Sam Young: 67 records (0.3%)
Amy Green: 66 records (0.3%)
Alice Johnson: 31 records (0.1%)
John Doe: 20 records (0.1%)
Xander Dean: 3 records (0.0%)
Wendy Clark: 2 records (0.0%)
Zachary Foster: 1 records (0.0%)
```

```
-----
Quality (Unique values: 7):
E - Non Qualified: 7634 records (35.4%)
D - Non Target: 6248 records (28.9%)
C - Low: 3459 records (16.0%)
unknown: 2253 records (10.4%)
B - Medium: 1564 records (7.2%)
A - High: 432 records (2.0%)
F: 3 records (0.0%)
```

```
-----
Stage (Unique values: 14):
```

Lost: 15743 records (72.9%)  
Call Delayed: 2248 records (10.4%)  
Registered on Webinar: 2072 records (9.6%)  
Payment Done: 858 records (4.0%)  
Waiting For Payment: 325 records (1.5%)  
Qualificated: 128 records (0.6%)  
Registered on Offline Day: 100 records (0.5%)  
Need to Call - Sales: 33 records (0.2%)  
Need To Call: 31 records (0.1%)  
Test Sent: 25 records (0.1%)  
Need a consultation: 23 records (0.1%)  
New Lead: 6 records (0.0%)  
Free Education: 1 records (0.0%)  
unknown: 0 records (0.0%)

---

Lost\_Reason (Unique values: 22):  
unknown: 5469 records (25.3%)  
Doesn't Answer: 4135 records (19.1%)  
Changed Decision: 2146 records (9.9%)  
Duplicate: 1771 records (8.2%)  
Non target: 1761 records (8.2%)  
Stopped Answering: 1588 records (7.4%)  
Invalid number: 1481 records (6.9%)  
needs time to think: 655 records (3.0%)  
Expensive: 626 records (2.9%)  
Conditions are not suitable: 531 records (2.5%)  
Next stream: 288 records (1.3%)  
Inadequate: 176 records (0.8%)  
Gutstein refusal: 172 records (0.8%)  
Considering a different direction in IT: 148 records (0.7%)  
Not for myself: 145 records (0.7%)  
Does not speak English: 138 records (0.6%)  
Didn't leave an application: 133 records (0.6%)  
Thought for free: 110 records (0.5%)  
Does not know how to use a computer: 50 records (0.2%)  
Went to Rivals: 48 records (0.2%)  
The contract did not fit: 21 records (0.1%)  
Refugee: 1 records (0.0%)

---

Source (Unique values: 14):  
Facebook Ads: 4850 records (22.5%)  
Google Ads: 4226 records (19.6%)  
Organic: 2590 records (12.0%)  
Tiktok Ads: 2051 records (9.5%)  
SMM: 1730 records (8.0%)  
Youtube Ads: 1657 records (7.7%)  
CRM: 1656 records (7.7%)  
Bloggers: 1089 records (5.0%)  
Telegram posts: 1001 records (4.6%)  
Webinar: 379 records (1.8%)  
Partnership: 203 records (0.9%)  
Test: 159 records (0.7%)  
Offline: 2 records (0.0%)  
unknown: 0 records (0.0%)

---

Payment\_Type (Unique values: 4):  
unknown: 21097 records (97.7%)  
Recurring Payments: 350 records (1.6%)  
One Payment: 141 records (0.7%)  
Reservation: 5 records (0.0%)

---

Product (Unique values: 6):  
unknown: 17981 records (83.3%)  
Digital Marketing: 1978 records (9.2%)  
UX/UI Design: 1019 records (4.7%)  
Web Developer: 575 records (2.7%)  
Data Analytics: 36 records (0.2%)  
Find yourself in IT: 4 records (0.0%)

---

Education\_Type (Unique values: 3):  
unknown: 17992 records (83.3%)  
Morning: 3157 records (14.6%)  
Evening: 444 records (2.1%)

---

Level\_of\_Deutsch (Unique values: 8):  
unknown: 20375 records (94.4%)  
B1: 780 records (3.6%)  
B2: 162 records (0.8%)  
A2: 137 records (0.6%)  
A0: 85 records (0.4%)

A1: 26 records (0.1%)

C1: 25 records (0.1%)

C2: 3 records (0.0%)

-----  
Country (Unique values: 29):

unknown: 19434 records (90.0%)

Germany: 2033 records (9.4%)

France: 28 records (0.1%)

Poland: 22 records (0.1%)

Austria: 19 records (0.1%)

Russia: 11 records (0.1%)

Netherlands: 6 records (0.0%)

Ukraine: 5 records (0.0%)

Czechia: 5 records (0.0%)

Kazakhstan: 4 records (0.0%)

Slovakia: 3 records (0.0%)

Serbia: 2 records (0.0%)

Italy: 2 records (0.0%)

United Arab Emirates: 2 records (0.0%)

Hungary: 2 records (0.0%)

Uzbekistan: 2 records (0.0%)

Montenegro: 1 records (0.0%)

Saudi Arabia: 1 records (0.0%)

Tajikistan: 1 records (0.0%)

Azerbaijan: 1 records (0.0%)

United States: 1 records (0.0%)

Belarus: 1 records (0.0%)

Latvia: 1 records (0.0%)

United Kingdom: 1 records (0.0%)

Israel: 1 records (0.0%)

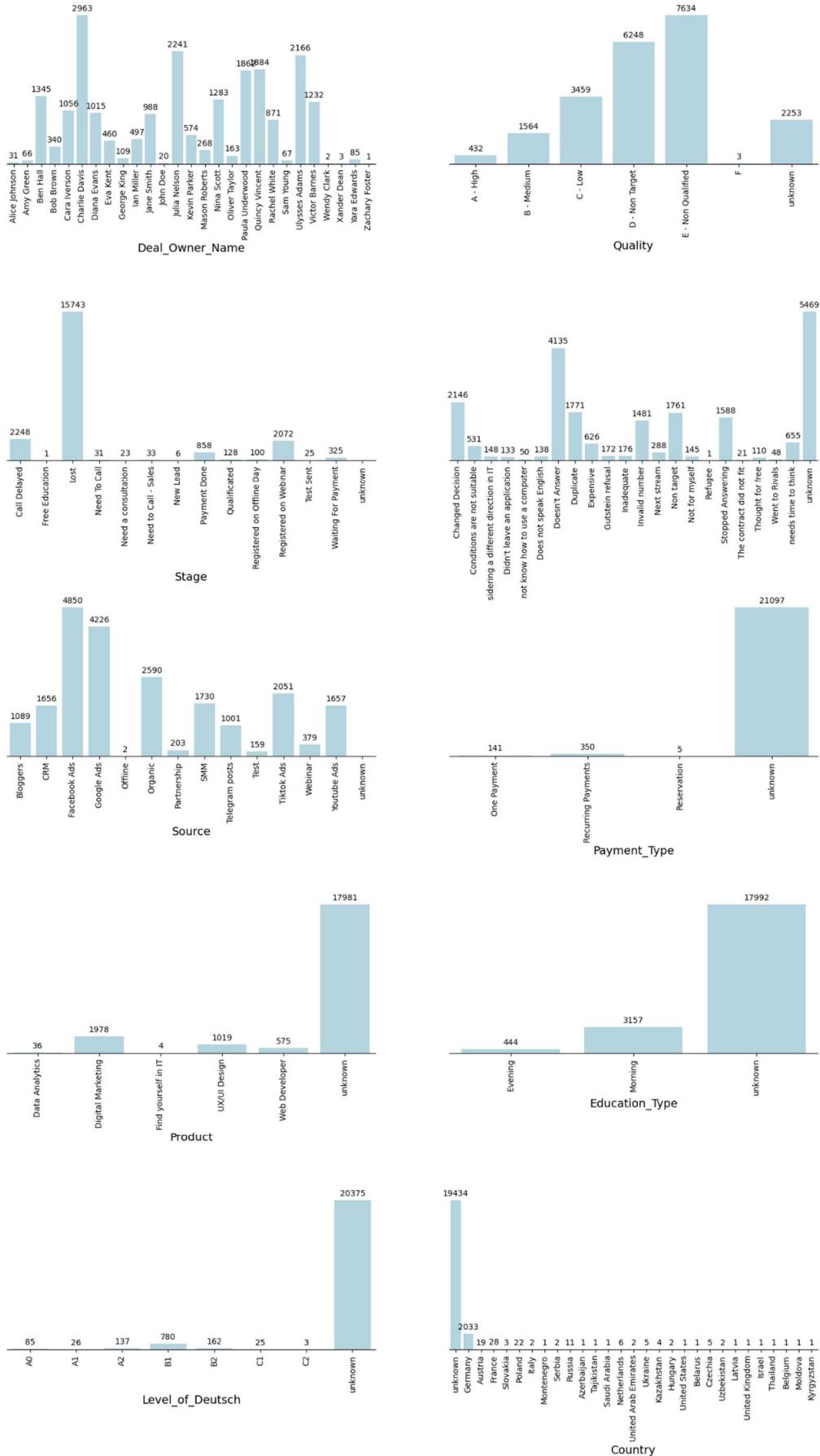
Thailand: 1 records (0.0%)

Belgium: 1 records (0.0%)

Moldova: 1 records (0.0%)

Kyrgyzstan: 1 records (0.0%)

### Distribution of Categorical Variables (Visualisation)



*Предварительные выводы из описательной статистики категориальных данных DataFrame deals:*

1. **Владельцы сделок (Deal\_Owner\_Name)**
  - 13.7% сделок у Charlie Davis, далее идут Julia Nelson (10.4%) и Ulysses Adams (10.1%).
  - в топ-10 владельцев концентрируется более 70% всех сделок, что указывает на сильную асимметрию в распределении.
  - некоторые владельцы имеют < 1% сделок, что говорит о низкой вовлечённости части команды.
2. **Качество лидов (Quality)**
  - 35.4% лидов — "E - Non Qualified" - эти сделки не соответствуют критериям (могут быть нецелевыми или неконвертируемыми).
  - 29% — "D - Non Target" → ещё одна большая категория потенциально неподходящих клиентов.
  - "A - High" составляет всего 2% всех сделок → очень мало лидеров с высокой квалификацией.
  - 10.5% записей имеют "unknown" свидетельствует о низкой дисциплине по заполнению данных в CRM.
3. **Этапы продаж (Stage)**
  - 72.9% сделок находятся на этапе "Lost", что указывает на высокую долю потерь клиентов.
  - "Registered on Webinar" и "Call Delayed" — два больших этапа (~10% каждый), что говорит о том, что многие лиды проходят вебинары и нуждаются в повторных звонках.
  - только 3.9% сделок завершились оплатой, а ещё 1.5% ожидают оплаты.
  - низкий процент сделок достигает стадии "Qualified" (0.6%) → сложности с переводом клиентов на финальные этапы.
4. **Причины потерь (Lost\_Reason)**
  - 25.3% сделок потеряны по "unknown" причине - детальный анализ этого сегмента ввиду отсутствия данных невозможен.
  - чаще всего клиенты просто "Не отвечают" (19.2%) следует рассмотреть резервные каналы связи при регистрации.
  - 9.9% "Changed Decision", возможна необходимость работы с возражениями.
  - 6.9% потерь из-за "Invalid number", возможно, ошибки в сборе контактных данных.
  - 5% клиентов ошибочно думали, что курс бесплатный, важно проверять, как подаётся цена.
5. **Источник лидов (Source)**
  - Главные источники:
    - Facebook Ads (22.5%) и Google Ads (19.6%) → основные рекламные каналы.
    - Tiktok Ads (9.5%) и SMM (8%) также вносят значительный вклад.
    - Блогеры, Telegram и CRM — около 5-7% каждого.
    - Webinar (1.8%) и Offline (0%) → низкая конверсия через вебинары и офлайн-мероприятия.
6. **Способы оплаты (Payment\_Type)**
  - 97.7% данных "unknown" → платежные данные требуют серьёзного уточнения!!!
  - 1.6% сделок через Recurring Payments (рассрочку).
  - 0.6% — One Payment (полная оплата сразу).
  - минимальная категория — Reservation (5 сделок) - почти не используется.
7. **Продукт (Product)**
  - 83.3% записей "unknown" → данные о продукте почти отсутствуют.
  - 9.1% сделок связаны с Digital Marketing, 4.7% — UX/UI Design.
  - Web Developer (2.6%) и Data Analytics (0.2%) имеют меньшее присутствие.
  - низкая диверсификация продуктов — 4 ведущих направления занимают почти весь рынок.
8. **Тип обучения (Education\_Type)**
  - 84.8% записей "unknown", что затрудняет анализ.
  - из известных категорий: "Morning" (13.4%) и "Evening" (1.9%).
  - курсы, проходящие утром, более популярны, чем вечерние.
9. **Уровень владения немецким (Level\_of\_Deutsch)** 94.4% записей "unknown", что делает анализ неполным.
  - среди известных уровней:
    - B1 (3.6%) и B2 (0.8%) встречаются чаще всего.
    - C1 и C2 (0.1%) — крайне редки.
    - большая часть клиентов — либо без знаний языка, либо на начальных уровнях.
10. **Страна (Country)**
  - 90% записей "unknown" - географические данные не заносятся и требуют уточнения.
  - среди известных стран:
    - Германия (9.4%) — единственная значимая категория.
    - Франция, Польша, Австрия, Россия и другие страны имеют меньше 0.1% записей.
  - распределение стран показывает, что курсами интересуются в основном клиенты из Германии.

*Общие выводы по описательной статистике Deals :*

- Много данных с unknown, что затрудняет полноценный анализ.
- Основные переменные, которые требуют уточнения: Payment\_Type, Product, Education\_Type, Country, Level\_of\_Deutsch.
- Высокая доля потерянных сделок (Lost = 72.9%), требуется дополнительный анализ потерь.
- Основные рекламные каналы — Facebook и Google Ads, но другие источники тоже имеют значимый вклад.
- Конверсия в успешные сделки низкая (Payment Done = 3.9%), нужно улучшать стратегию продаж.
- Время обучения влияет на стоимость (Morning курсы более распространены).
- Клиенты чаще не отвечают или изменяют решение, чем уходят к конкурентам.

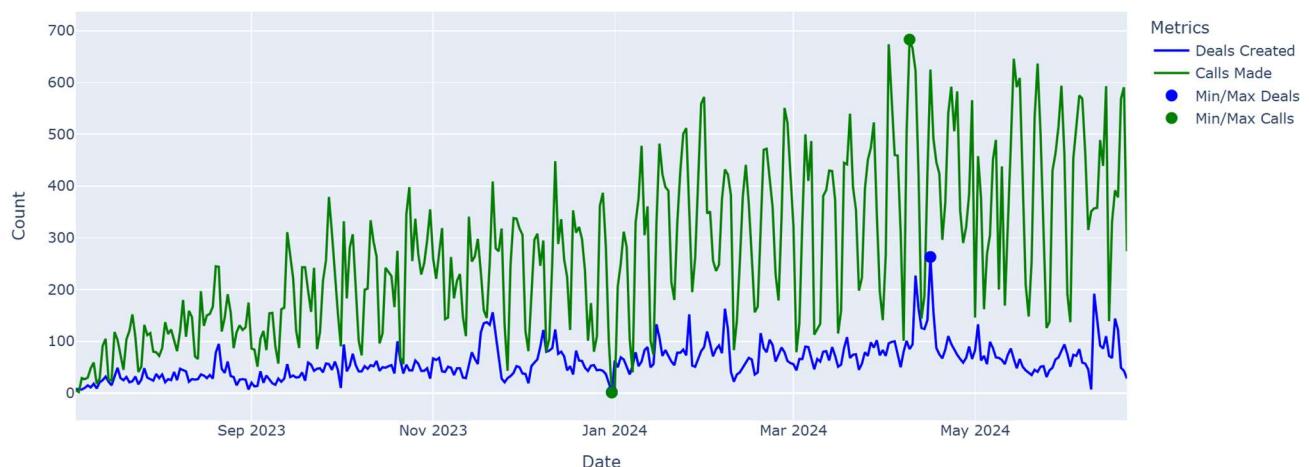
### 3. Анализ временных рядов

(Time Series Analysis)

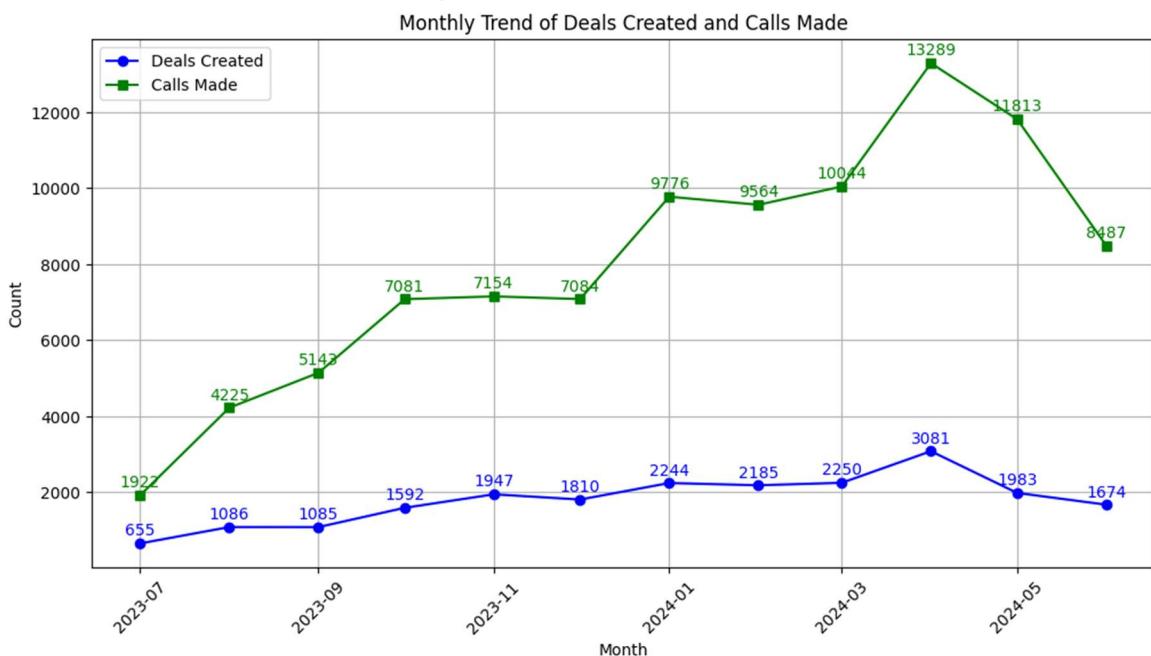
#### 3.1 Анализ тенденций создания сделок с течением времени и их связь со звонками.

##### Динамика создания сделок и их связь со звонками во времени

Dynamics of trade creation and their relationship with calls over time (day interval)



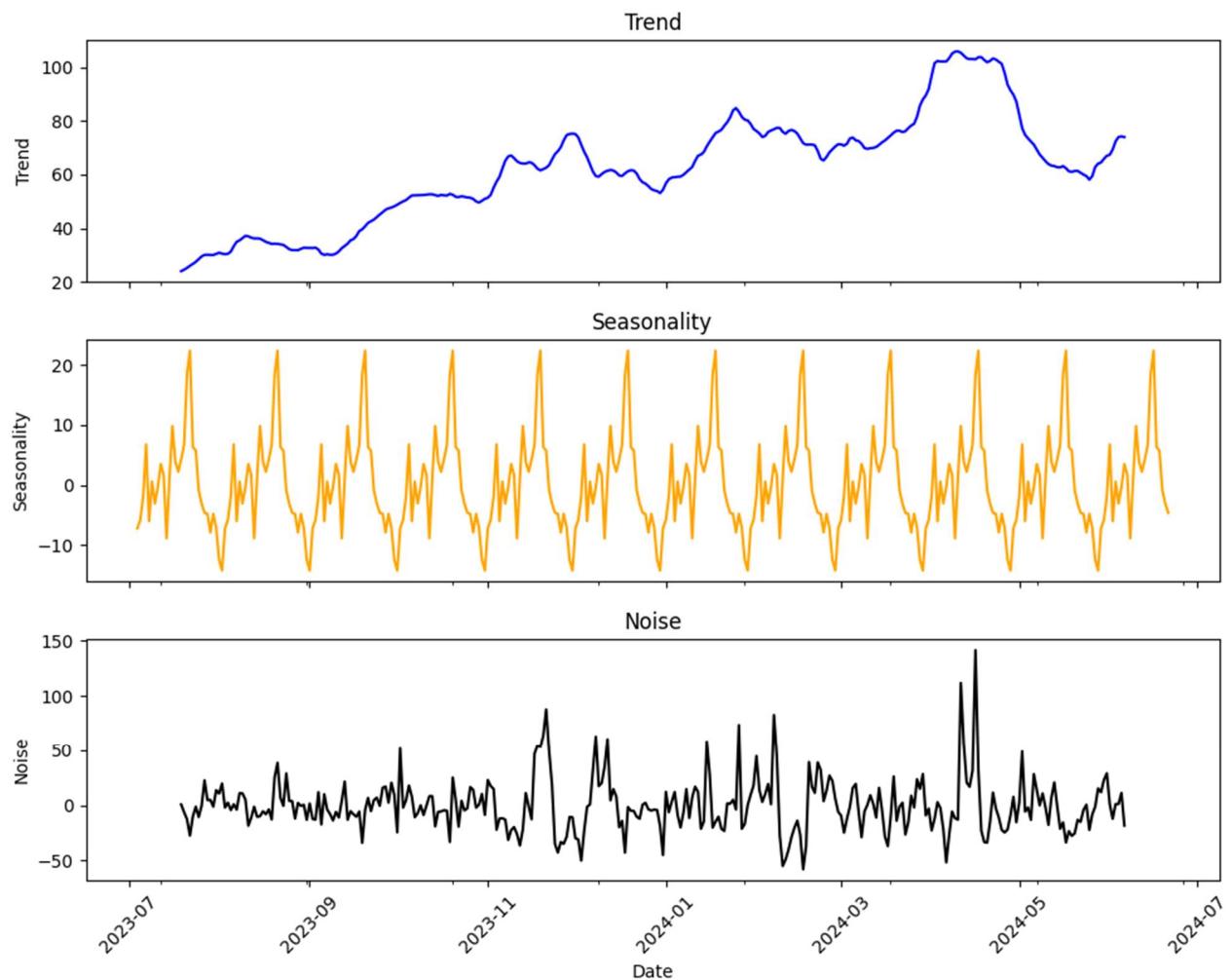
##### Тенденция заключенных сделок и совершенных звонков по месяцам



- На графиках видны колебания в количестве сделок и звонков, что может указывать на сезонность или внешние факторы.
- В некоторые моменты увеличение количества звонков предшествует росту сделок, что может подтверждать эффективность звонков для продаж.
- Различные пики и спады могут соответствовать маркетинговым кампаниям или изменению спроса.
- Максимальное значение по звонкам отмечено 09.04.2024 (683), максимальное значение сделок отмечается 16.04.2024 (263)
- Минимальные значения 1 звонок и 2 сделки отмечены 31.12.2023

##### Декомпозиция временного ряда

- Тренд (долгосрочные изменения)
- Сезонность (периодические колебания)
- Шум (случайные флуктуации)



Average amplitude of seasonality: 7.99

Average noise level: 24.34

Представленные графики представляют разложение временного ряда количества сделок (Deals\_Created) на три компонента:

- **Тренд (Trend)**
  - отражает общую динамику числа сделок с течением времени.
  - наблюдается постепенный рост с некоторыми колебаниями, что может говорить о общем увеличении активности или влиянии маркетинговых стратегий.
- **Сезонность (Seasonality)**
  - Демонстрирует повторяющиеся циклы в данных.
  - Значение сезонности колеблется между -10 и 20, показывая регулярные спады и всплески в активности.
  - Средняя амплитуда сезонных колебаний: 7.99, что указывает на умеренно выраженную сезонность.
- **Шум (Noise)**
  - Это случайные флуктуации, не связанные с трендом или сезонностью.
  - Разброс значений колеблется в диапазоне от -50 до 150, что говорит о значительном количестве нерегулярных колебаний.
  - Средняя величина шума: 24.34, что указывает на довольно высокий уровень случайных изменений.

#### Выводы:

- Тренд показывает рост сделок со временем. Возможно, спрос на продукт увеличивается, либо улучшается работа с лидами.

- Сезонность умеренная. Есть повторяющиеся колебания, но они не экстремальные. Возможны циклы, связанные с рекламными кампаниями или внешними факторами (например, началом учебных периодов).
- Шум достаточно высокий. В данных есть существенные нерегулярные колебания, что может свидетельствовать о спонтанных изменениях в активности, нерегулярных маркетинговых усилиях или неожиданных внешних факторах.

### Корреляция между числом звонков и числом новых сделок

Correlation between the number of calls and the number of deals:

	Deals_Created	Calls_Made
Deals_Created	1.00	0.57
Calls_Made	0.57	1.00



Вывод:

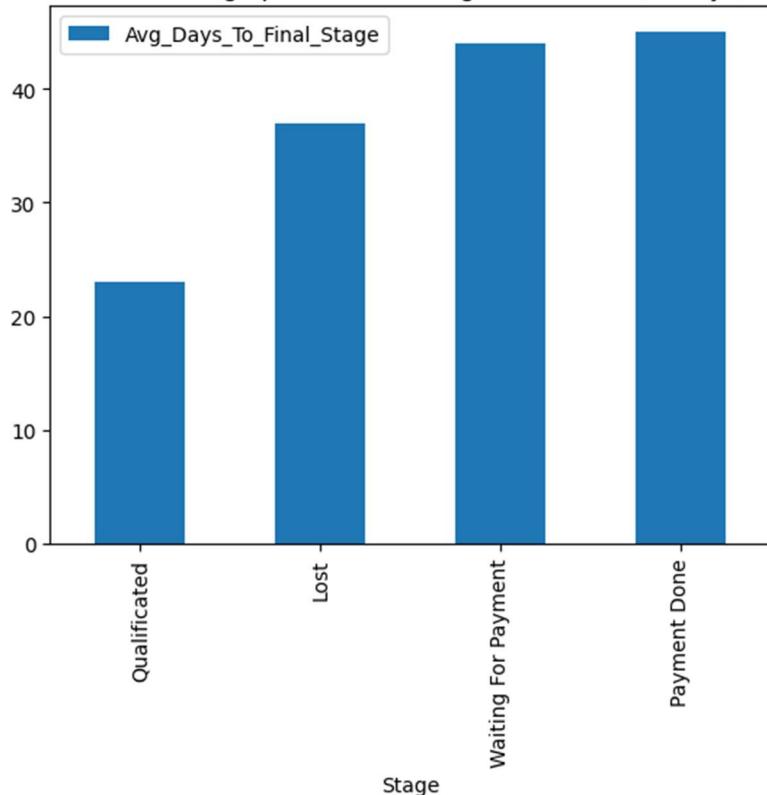
- Кэффициент корреляции 0.57 указывает на **умеренную положительную связь** между количеством звонков (Calls\_Made) и количеством созданных сделок (Deals\_Created).
- Значение коэффициента свидетельствует, что **чем больше звонков совершается, тем больше создаётся сделок**, но связь не является идеальной (т.е. не все звонки приводят к сделкам).
- Корреляция не означает причинно-следственную связь: высокая активность звонков может способствовать увеличению сделок, но на сделки также влияют другие факторы (качество звонков, маркетинг, сезонность).
- **Значение 0.57 говорит о довольно сильной взаимосвязи, но не абсолютной.** Если коэффициент был бы ближе к 1, можно было бы утверждать, что звонки почти всегда ведут к сделкам.
- Таким образом, **стратегия общения с клиентом должна быть несколько скорректирована в сторону увеличения конверсии звонков в сделки.**

### Анализ времени перехода сделок на финальные этапы после звонков

	Stage	Avg_Days_To_Final_Stage
8	Qualificated	23.00
2	Lost	37.00
12	Waiting For Payment	44.00

	Stage	Avg_Days_To_Final_Stage
7	Payment Done	45.00

Deals Moving Speed to final Stages after Calls (in days)



#### Скорость перехода сделок на финальные этапы после звонков:

- "Qualificated" – 27 дней,
  - сделки проходят квалификацию быстрее всего среди всех финальных стадий
  - это может означать, что процесс первичной оценки клиента организован эффективно, но требует дальнейших действий для закрытия сделки.
- "Waiting For Payment" – 34 дня
  - среднее время до стадии ожидания платежа составляет чуть больше месяца.
  - возможны задержки из-за внутренних согласований у клиента или ожидания одобрения финансирования.
- "Lost" – 37 дней
  - потерянные сделки переходят в этот статус не намного медленнее, чем "Waiting For Payment".
  - это может означать, что клиенты долго принимают решения, но в итоге либо отказываются, либо сталкиваются с проблемами при финальном согласовании.
- "Payment Done" – 40 дней -это самая длительная стадия перед окончательным закрытием сделки. - вероятно, процесс оплаты требует дополнительных шагов (согласований, подписания контрактов, банковских переводов).

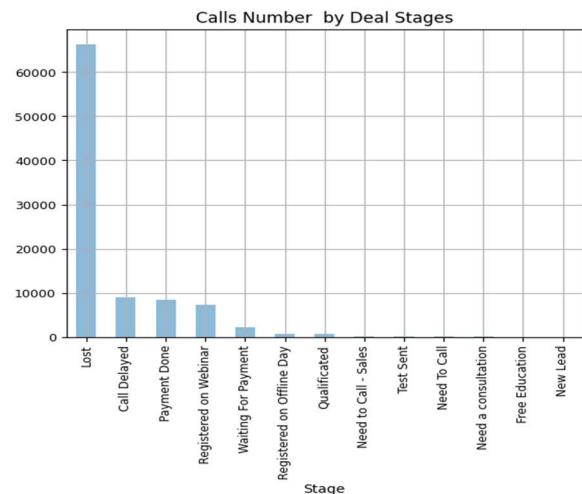
#### Основные выводы:

- Быстрее всего достигается этап квалификации, но до завершения сделки проходит ещё несколько недель.
- Потерянные сделки (Lost) могут быть следствием долгого принятия решения клиентом — возможно, стоит внедрить дополнительные методы ускорения процесса.
- Ожидание платежа занимает почти столько же времени, сколько потерянные сделки — улучшение условий оплаты или финансовой структуры может ускорить процесс.
- Финальное закрытие сделки занимает дольше всего (40 дней), что может указывать на административные барьеры.
-

## Анализ количества звонков по этапам сделок

Количество звонков по этапам сделок:

	Stage	Total_Calls
2	Lost	66337
0	Call Delayed	9044
7	Payment Done	8455
10	Registered on Webinar	7378
12	Waiting For Payment	2285
9	Registered on Offline Day	783
8	Qualified	610
5	Need to Call - Sales	218
11	Test Sent	178
3	Need To Call	115
4	Need a consultation	98
1	Free Education	8
6	New Lead	1



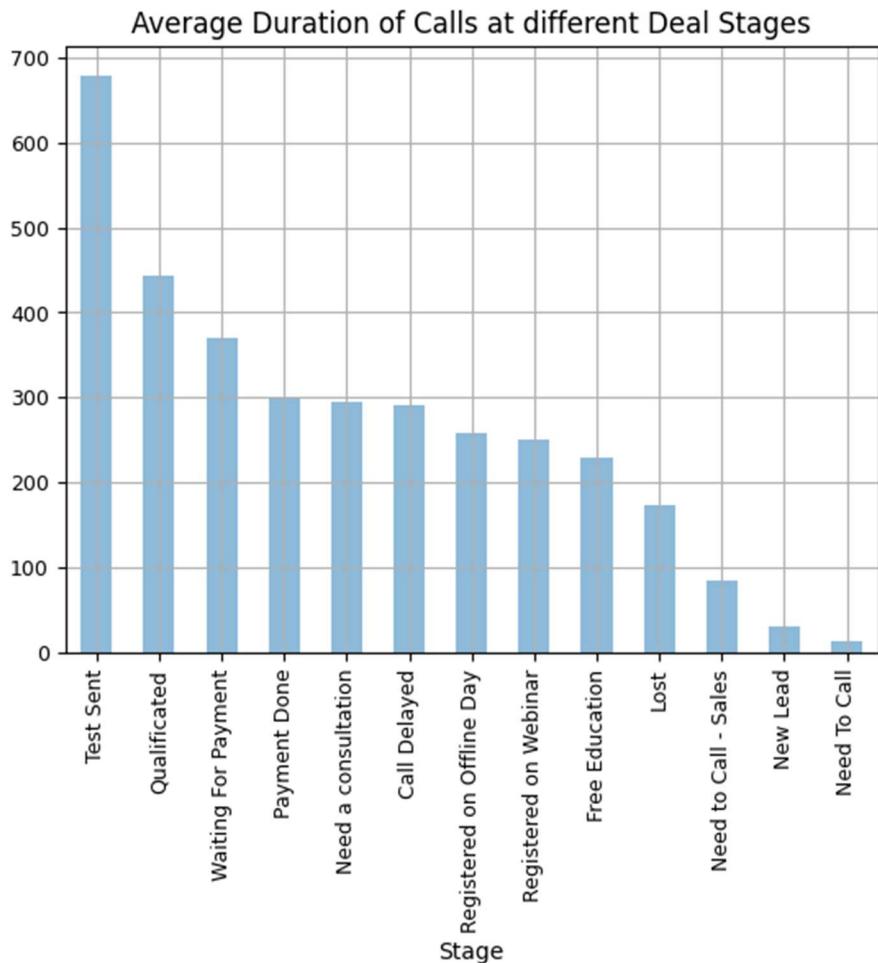
### Основные этапы, где было больше всего звонков:

- **"Lost"** (61 341 звонков) - потерянные сделки обрабатывались интенсивно, но клиенты не были сконвертированы. Это может указывать на неэффективные повторные звонки или сложности в закрытии продаж.
- **"Registered on Webinar"** (14 843 звонков) → вебинары являются важным этапом перед решением клиента, возможно, они помогают повысить конверсию.
- **"Call Delayed"** (9 615 звонков) → частые задержки связи с клиентами, возможно, требуется улучшение процессов.
- **"Payment Done"** (5 560 звонков) → активные звонки после успешного оформления сделок, вероятно, для подтверждения или дальнейшего сопровождения.

### Вывод:

- Большое количество звонков в категории "Lost" говорит о повторных попытках работы с отказавшимися клиентами.
- Этап "Registered on Webinar" получает много звонков, подтверждая важность вебинаров как инструмента продаж.
- Высокая активность на "Payment Done" и "Waiting For Payment" означает, что звонки помогают закрывать сделки.

## Анализ средней продолжительности звонков по этапам сделок



### 1. Этапы с самой длинной средней продолжительностью звонков

- "Test Sent" (634 сек.) → самый продолжительный этап, возможно, из-за детального обсуждения тестов, вопросов по заданиям или разъяснения условий дальнейших шагов.
- "Qualified" (356 сек.) → продленные разговоры, вероятно, касаются подтверждения уровня квалификации или сбора дополнительной информации перед переходом к следующим шагам.
- "Waiting For Payment" (338 сек.) → звонки на этом этапе могут включать объяснение условий оплаты, уточнение деталей транзакций и закрытие возможных вопросов клиента.

#### Вывод:

- Длительность звонков на этих этапах указывает на более сложные или детализированные обсуждения, требующие значительного времени.
- Этап "Test Sent" имеет самые долгие разговоры, возможно, клиенты задают много вопросов перед тестированием.

### 2. Средняя продолжительность звонков на ранних этапах

- "New Lead" (327 сек.) → долгие звонки при первом контакте, вероятно, связаны с представлением продукта и ответами на вопросы потенциальных клиентов.
- "Need a consultation" (236 сек.) → разговоры на этапе консультации требуют объяснения деталей курса.
- "Free Education" (229 сек.) → звонки в рамках бесплатного обучения также занимают время, возможно, клиенты уточняют нюансы программы.

#### Вывод:

- Долгие звонки на первых этапах говорят о высокой вовлеченности клиентов, которые хотят разобраться в деталях перед принятием решения.
- Этап "New Lead" показывает, что первый контакт является ключевым — требуется достаточно времени для убеждения клиента.

### 3. Этапы с самой короткой средней продолжительностью звонков

- "Lost" (166 сек.) → короткие разговоры с отказавшимися клиентами, возможно, отказ фиксируется быстро без детального обсуждения.
- "Need to Call - Sales" (98 сек.) → возможны быстрые проверки интереса клиента, но без долгих бесед.-

- "Need To Call" (56 сек.) → самые краткие звонки — скорее всего, это просто попытки связи или напоминания.

#### Вывод:

- Короткие звонки на "Lost" могут означать, что после отказа не предпринимаются дополнительные усилия для возврата клиента.
- Очень короткие звонки на "Need To Call" — возможно, только напоминания или технические проверки, без углубленного общения.

#### Общие тенденции

- Максимальная продолжительность на этапах проверки и оформления сделки ("Test Sent", "Waiting For Payment").
- Длинные звонки на первых этапах ("New Lead", "Need a consultation") → клиенты активно интересуются перед решением.
- Короткие звонки на "Lost" → отказ фиксируется быстро, возможно, стоит анализировать эффективность работы с этими клиентами

## 3.2 Распределение времени закрытия сделок и продолжительность периода от создания до закрытия

### Распределение времени закрытия сделок

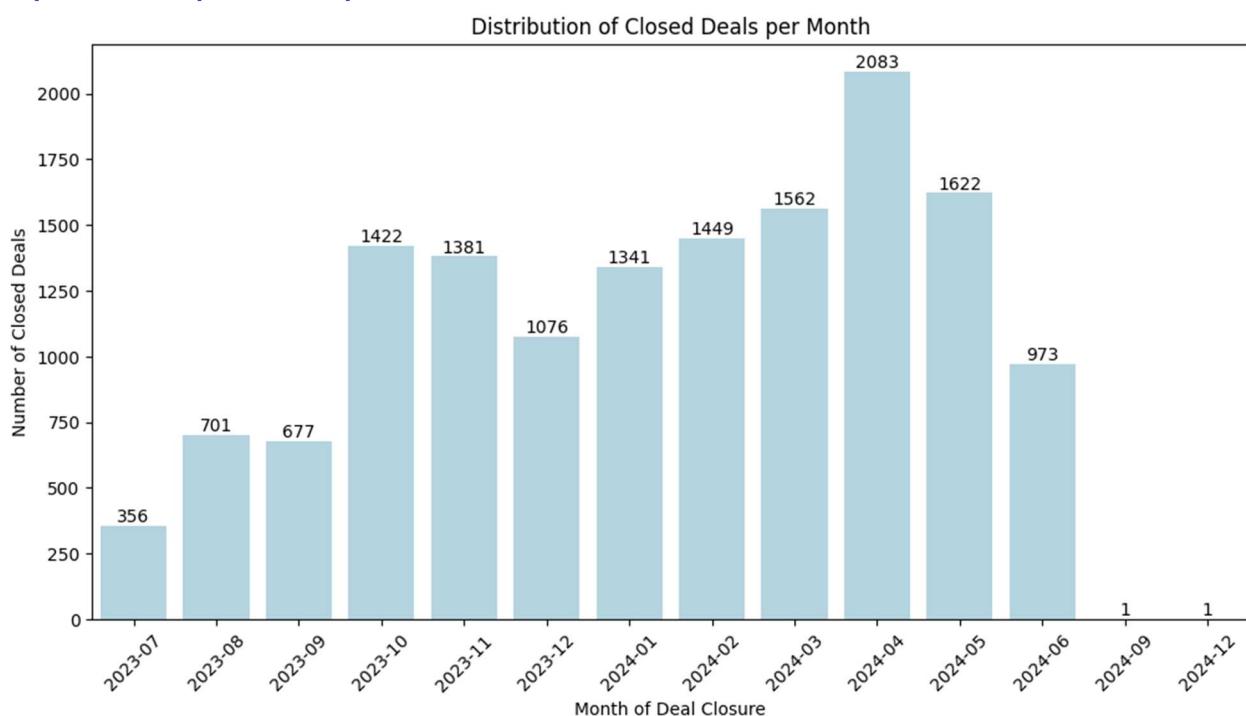


График "Распределение закрытых сделок по месяцам" показывает динамику закрытия сделок с июля 2023 по декабрь 2024

- рост активности в октябре 2023 (1422 сделки), может быть связан с маркетинговыми кампаниями или сезонностью;
- пик в апреле 2024 (2083 сделки) — вероятно, успешный период для продаж или запуск выгодных предложений;
- падение в июне 2024 (973 сделки), может указывать на спад активности или уменьшение новых лидов;
- единичные сделки в сентябре и декабре 2024 (по 1 сделке) — возможно, данные за эти периоды ещё не полностью собраны или были аномалии.

### Выводы и рекомендации

- **Анализ сезонности:** активные периоды (октябрь, февраль, март, апрель) показывают, когда школа закрывает больше всего сделок. Стоит уточнить причины таких всплесков при анализе рекламных кампаний.
- **Управление спадами:** июнь 2024 показал снижение — возможно, нужна корректировка стратегии или дополнительное привлечение клиентов.

- **Дополнительная проверка данных:** единичные сделки в сентябре и декабре 2024 могут быть ошибкой или аномалией.

### Анализ продолжительности периода от создания до закрытия сделки

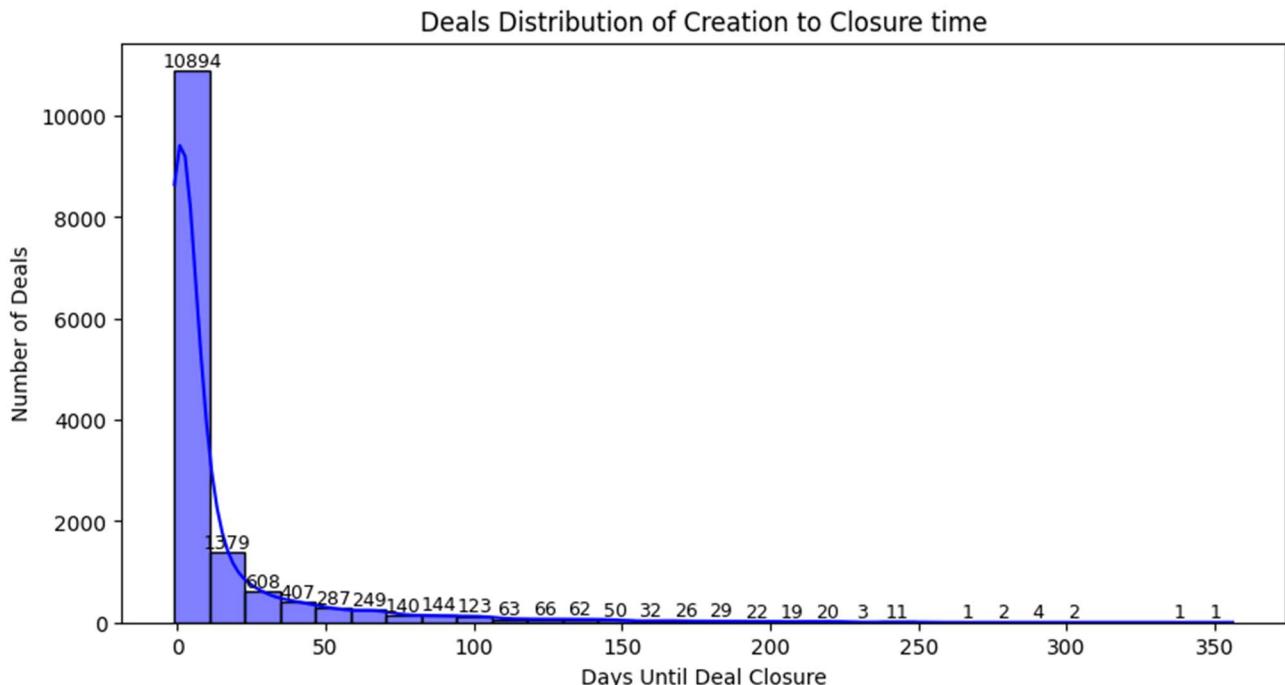


График распределения длительности сделки (от создания до закрытия), измеренное в днях.

- большая часть сделок закрывается в пределах 50 дней, что указывает на относительно короткий цикл продаж.
- есть сделки, которые закрываются **дольше 150 дней**, но их число значительно меньше. Это могут быть более дорогостоящие или более сложные сделки, связанные с неопределенностью клиента.
- наблюдается **пик закрытия сделок** в районе 20–40 дней, что может соответствовать стандартному процессу продаж и принятия решений клиентами.

### Выводы и рекомендации:

- **Оптимизация процессов:** если большинство сделок закрывается в первые 50 дней, стоит проверить, какие факторы ускоряют процесс и как их можно усилить.
- **Анализ долгих сделок:** сделки, требующие более 150 дней, требуют отдельного анализа—что задерживает их закрытие и можно ли сократить этот период?
- **Маркетинговые стратегии:** если пик закрытий приходится на 20–40 дней, стоит сфокусироваться на усилении взаимодействий именно в этом периоде.

## 3.3 Общие выводы по проведенному анализу временных рядов сделок и звонков

### Динамика создания сделок и их связь со звонками

Рост числа звонков часто предшествует увеличению сделок, подтверждая их влияние на продажи.

- **Максимальные значения:** апрель 2024
  - **Минимальные значения:** декабрь 2023
- Это может указывать на сезонность или влияние маркетинговых факторов.

Декомпозиция временного ряда показала:

- **Умеренную сезонность** (амплитуда 7.99), отражающую повторяющиеся колебания.
- **Высокий уровень случайных изменений** (24.34), свидетельствующий о нестабильности спроса.

### Корреляция звонков и сделок

- **Коэффициент корреляции:** 0.57
- Подтверждает умеренную положительную связь, но не гарантирует превращение звонков в сделки.
- Требуется **корректировка стратегии общения** с клиентами для повышения конверсии.

### Время перехода сделок на финансовые этапы

Этап                    Среднее время (дней)

Qualified              27

Waiting For Payment 34

Этап	Среднее время (дней)
Lost	37
Payment Done	40

- **Быстрее всего достигается квалификация клиентов.**
- Финальная оплата занимает **самое большое время**, возможно из-за административных процессов.

#### **Анализ звонков по этапам сделок**

- **Наибольшее число звонков (61 341)** приходится на этап "Lost", что указывает на попытки работы с отказавшимися клиентами.
- **Активность на этапе "Registered on Webinar"** подтверждает значимость вебинаров как инструмента продаж.
- **Средняя продолжительность звонков** наиболее высока на этапах **тестирования, квалификации и ожидания платежа**.
- **Длинные разговоры** на ранних стадиях ("New Lead") говорят о высокой вовлеченности клиентов.

#### **Распределение закрытых сделок по месяцам**

- **Пики активности:** октябрь 2023 и апрель 2024 (вероятно, маркетинговые кампании).
- **Спад в июне 2024** требует дополнительных исследований и корректировки стратегии.

#### **Распределение длительности сделок (от создания до закрытия)**

Анализ показывает, что большая часть сделок закрывается в пределах **50 дней**, что указывает на относительно короткий цикл продаж.

- Сделки, требующие **более 150 дней**, встречаются редко и могут быть более дорогостоящими или сложными.
- **Пик закрытия сделок наблюдается в районе 20-40 дней**, что соответствует стандартному процессу продаж и принятию решений клиентами.

#### **Выводы и рекомендации**

1. **Оптимизировать процессы**, выявляя факторы, ускоряющие закрытие сделок в первые 50 дней, и усиливая их.
2. **Анализировать долгосрочные сделки**, чтобы понять, что задерживает их закрытие и как сократить этот период.
3. **Фокусироваться на маркетинговых стратегиях**, поддерживающих активность клиентов в ключевой период 20-40 дней.
4. **Усилить маркетинг в активные сезоны** (октябрь, февраль, март, апрель) для максимизации продаж.
5. **Пересмотреть работу с отказавшимися клиентами**, улучшив тактику их возврата и ускоряя процесс принятия решений.

## 4. Анализ эффективности кампаний

(Campaigns Effectiveness Analysis)

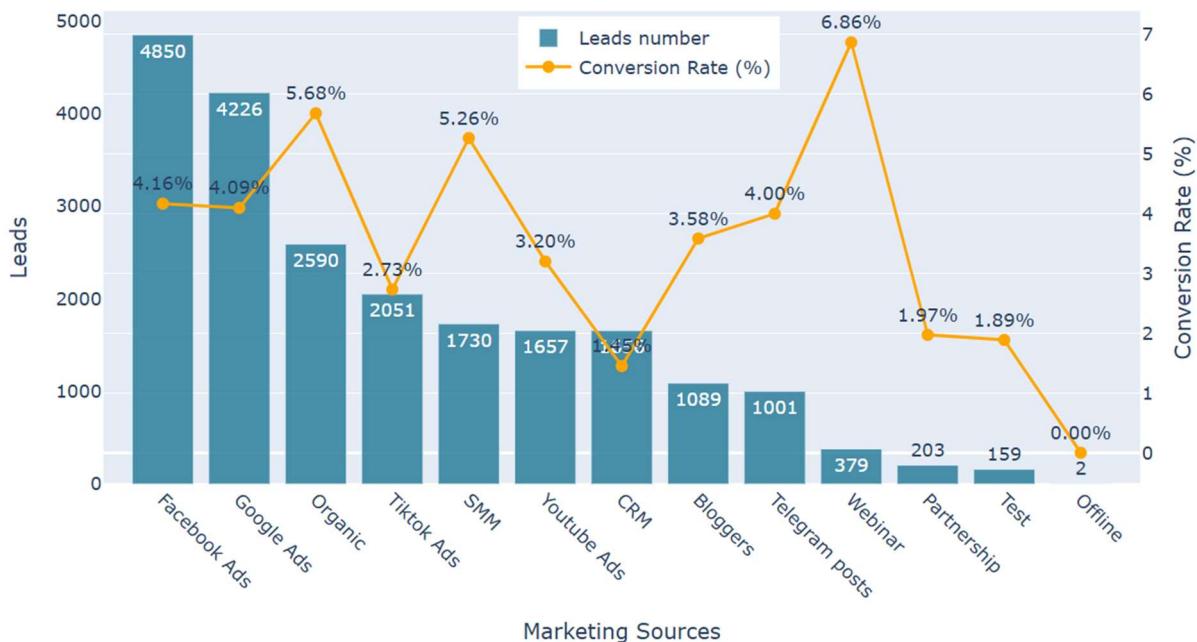
### 4.1 Оценка эффективность различных маркетинговых источников (Source) в генерировании качественных лидов

Анализ количества лидов от каждого источника и коэффициента конверсии в первую покупку:

Source	Leads	Success Deals	Conversion Rate (%)
2 Facebook Ads	4850	202.00	4.16
3 Google Ads	4226	173.00	4.09
5 Organic	2590	147.00	5.68
10 Tiktok Ads	2051	56.00	2.73
7 SMM	1730	91.00	5.26
12 Youtube Ads	1657	53.00	3.20
1 CRM	1656	24.00	1.45
0 Bloggers	1089	39.00	3.58
8 Telegram posts	1001	40.00	4.00
11 Webinar	379	26.00	6.86
6 Partnership	203	4.00	1.97
9 Test	159	3.00	1.89
4 Offline	2	0.00	0.00

Визуализация количества лидов и коэффициента конверсии для каждого источника

Marketing Sources Performance



Для более глубокого анализа эффективности рекламных источников добавим в анализ число закрытых сделок и среднюю сумму сделки по источникам

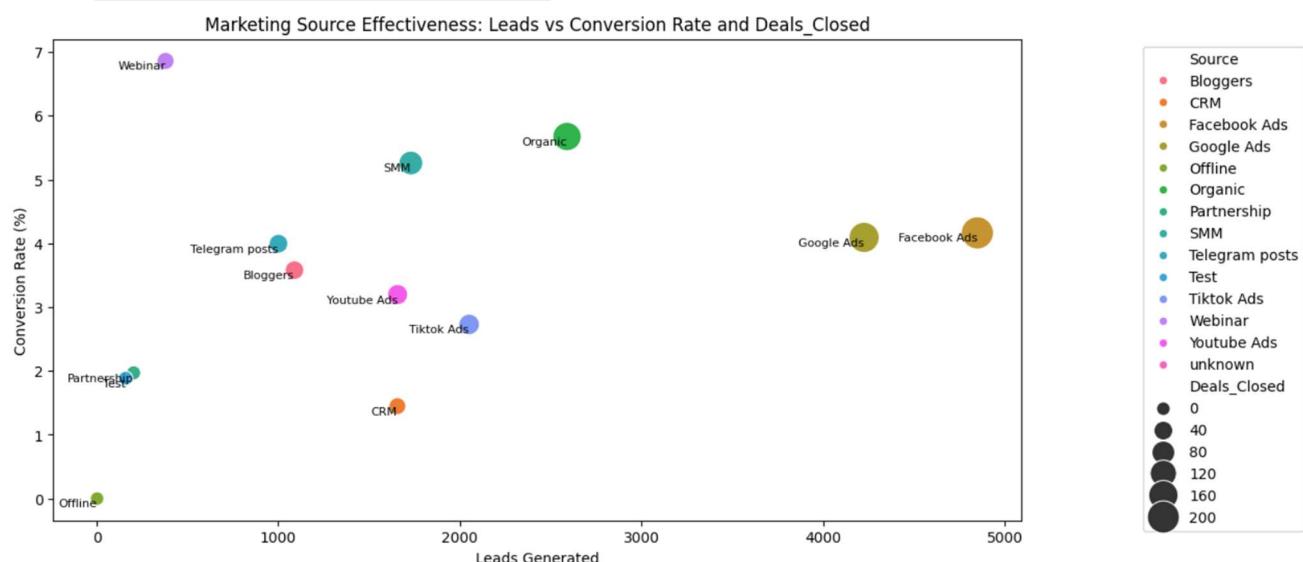
Marketing Source performance indicators - summary table

	Source	Leads_Generated	Deals_Closed	Revenue	Average_Deal_Value	Closing_Duration	Conversion_Rate
2	Facebook Ads	4850	202	1042445.46	927.58	18.67	4.16
3	Google Ads	4226	173	829330.45	917.36	11.78	4.09
5	Organic	2590	147	657147.88	949.86	9.69	5.68
10	Tiktok Ads	2051	56	232650.00	879.50	13.00	2.73
7	SMM	1730	91	343071.96	903.56	15.29	5.26
12	Youtube Ads	1657	53	241093.49	980.86	12.81	3.20
1	CRM	1656	24	90950.00	913.66	11.94	1.45
0	Bloggers	1089	39	190256.82	936.50	17.51	3.58
8	Telegram posts	1001	40	177270.16	1012.21	16.41	4.00
11	Webinar	379	26	122500.91	804.95	21.72	6.86
6	Partnership	203	4	17580.00	1220.30	26.06	1.97
9	Test	159	3	11360.00	1031.11	13.50	1.89
4	Offline	2	0	0.00	0.00	0.00	0.00

Интерактивная диаграмма рассеяния показывает соотношение по **рекламным источникам** (выделены цветом):

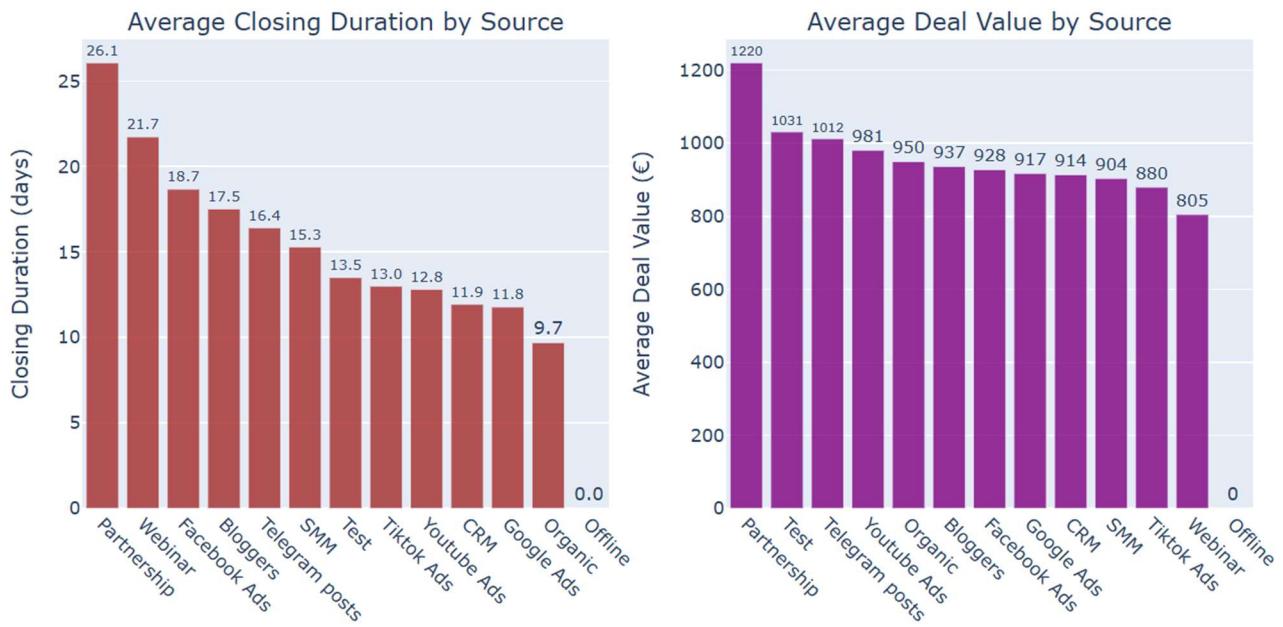
- число лидов (ось X)
- коэффициента конверсии (ось Y)
- число закрытых сделок / средняя сумма сделки - диаметр круга. Выбор метрики из выпадающего списка.

Metric: Deals\_Closed



## Визуализация сравнения маркетинговых источников (Source) по среднему времени закрытия сделки (Deals\_Closed) и средней сумме сделки (Average\_Deal\_Value)

Marketing Source Performance



Основные результаты анализа эффективности маркетинговых источников:

1. **Лиды и конверсия**
  - **наибольшее количество лидов** поступает от **Facebook Ads** (4850 лидов) и **Google Ads** (4226 лидов), что делает их мощными каналами привлечения. Третья позиция у **Organic** (SEO) - 2590 лидов, эффективный канал без рекламных затрат;
  - **высокая конверсия лидов в сделки** наблюдается у **Webinar** (7%) и **Organic** (6%), что говорит о **более качественных лидах**, готовых к оплате.
  - **самая низкая конверсия** у **CRM** (1.45%), **Partnership** (1.97%) и **Test** (1.89%), что может указывать на сложность трансформации льда в реальную сделку;
  - **Offline** - не приносит успешных сделок — его доля в общем маркетинговом миксе незначительна;
  - **Telegram** (3.99%) и **SMM** (5.26%) показывают хорошую конверсию, несмотря на средний объем лидов.
2. **Среднее время закрытия сделки**:
  - **Самые быстрые**: **Organic** (9 дней) и **Google Ads** (12 дней) — клиенты быстро принимают решение.
  - **Самые долгие**: **Partnership** (26 дней) и **Webinar** (22 дня) — требуют большего взаимодействия перед закрытием сделки, **Facebook Ads** (18.51 дней) — высокая генерация, но долгий цикл.
3. **Средняя сумма сделки**:
  - **Самая высокая**: **Test** (2 861€), но этот канал приносит относительно мало лидов (159 льда), 3 закрытых сделки и самой низкой конверсией (1.89).
  - **Стабильные показатели**: **CRM**, **Webinar** и **Partnership** — дорогие сделки, но относительно небольшое число лидов и низкая конверсия.

**Выводы и рекомендации:**

- **Оптимизация Facebook Ads и Google Ads**: у этих каналов огромный поток лидов, но конверсия в сделки не максимальная. Нужно проверить качество лидов и оптимизировать рекламные послания.
- **Расширение Organic и Webinar**: эти источники показывают высокую конверсию — можно увеличить инвестиции в оптимизация для поисковых систем - SEO (Search Engine Optimization) и вебинары для более целевой аудитории.
- **Улучшение CRM**: быстрые сделки, высокая сумма, но очень низкая конверсия (1%). Возможно, нужен пересмотр подхода к работе с клиентами, более точная сегментация и персонализированные сценарии взаимодействия.
- **Telegram и SMM**: средние показатели по конверсии (4-5%), можно протестировать новые стратегии привлечения лидов для масштабирования этих каналов.
- изучить **Offline**, имеет нулевую конверсию и только 2 льда.

## 4.2 Сравните эффективность различных кампаний с точки зрения генерации лидов и коэффициента конверсии.

### Рассчитаем показатели эффективности кампаний

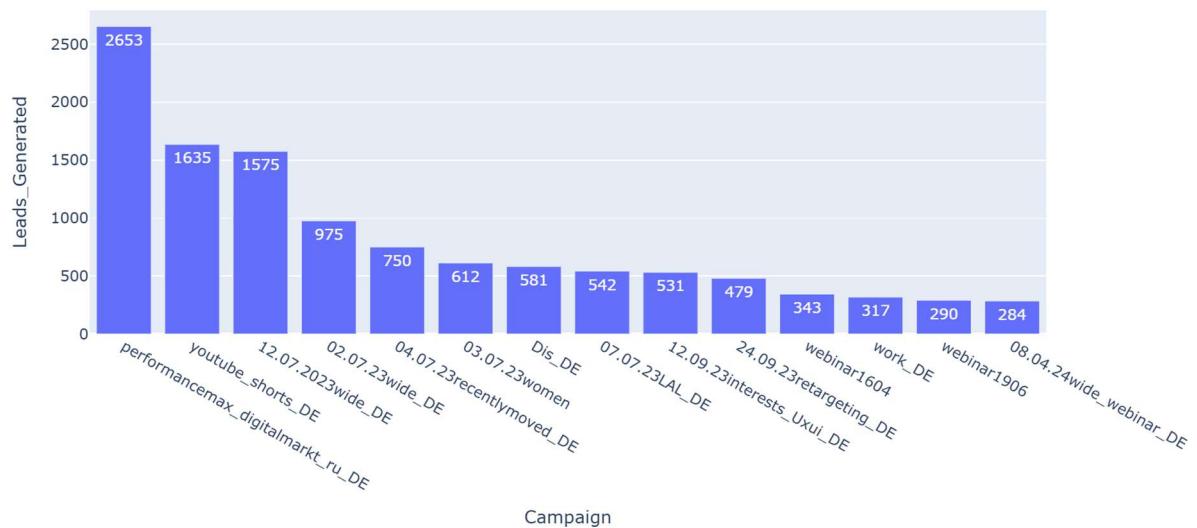
Marketing Campaign performance indicators - summary table

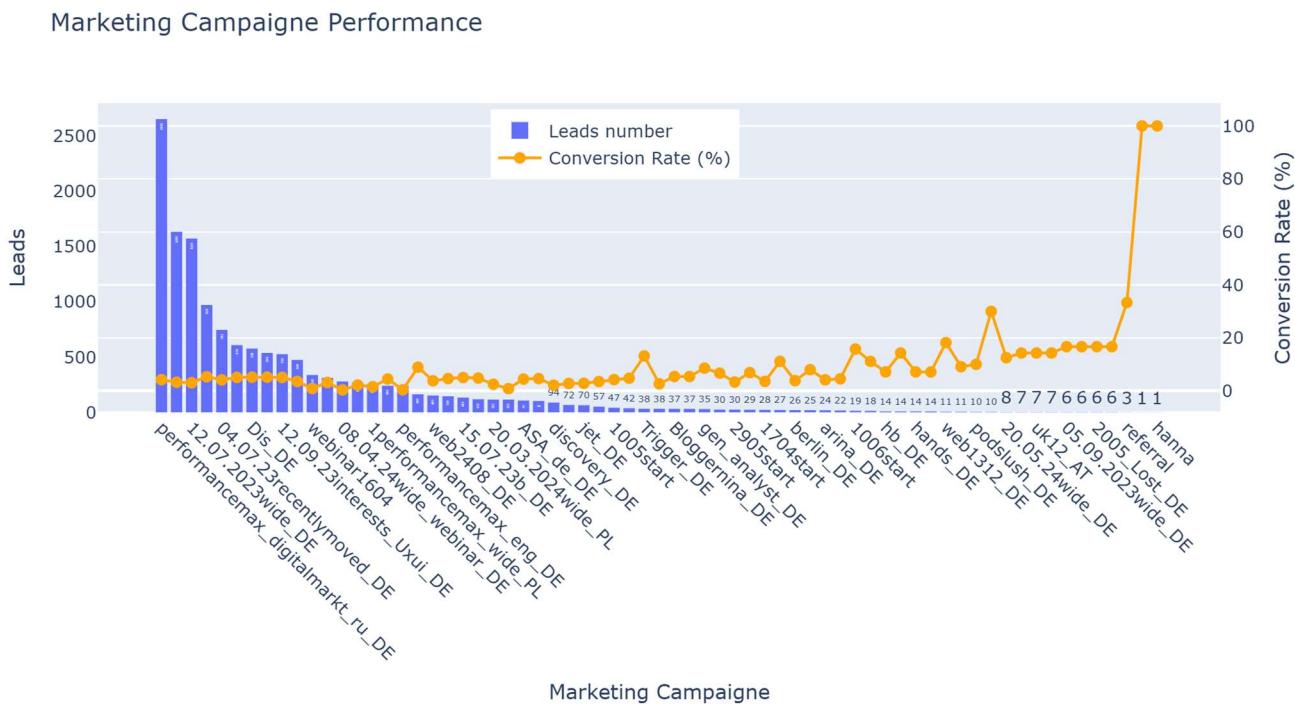
	Campaign	Leads_Generated	Deals_Closed	Average_Deal_Value	Closing_Duration	Conversion_Rate
154	unknown	5526	271	927.12	13.16	4.90
118	performancemax_digitalmarkt_ru_DE	2653	112	905.72	12.48	4.22
152	youtube_shorts_DE	1635	53	980.86	12.88	3.24
18	12.07.2023wide_DE	1575	48	924.71	13.59	3.05
2	02.07.23wide_DE	975	52	889.20	18.73	5.33
...	...	...	...	...	...	...
99	grad_DE	1	0	0.00	1.00	0.00
91	eha_DE	1	0	0.00	0.00	0.00
113	marue_DE	1	0	0.00	0.00	0.00
141	web2410_DE	1	0	0.00	167.00	0.00
147	women_DE	1	0	0.00	69.00	0.00

### Интерактивные диаграммы рейтинга кампаний по генерации лидов и конверсии

Top N:  15

Top 15 (from 155) Campaign by Leads generation





## Результаты анализа эффективности маркетинговых кампаний с точки зрения генерации лидов и коэффициента конверсии:

## **Основные показатели**

- Кампания с наибольшим количеством лидов
    - **performancemax\_digitalmarkt\_ru\_DE** - 2653 льда, 112 закрытых сделок, средняя стоимость сделки 5066.26 €.
    - Лидер по объему трафика, но конверсия средняя (4.22%).
  - Лучшие кампании по конверсии
    - **hanna** → 100% конверсии (1 лид → 1 сделка).
    - **domain** → 100% конверсии (очень редкий результат).
    - **referral** → 33.33% конверсии, что выше среднего, но всего 3 льда.
    - Эти кампании эффективны, но имеют низкий объем лидов
  - Кампании с самой дорогой средней сделкой
    - **2005\_Lost\_DE** → €11,000 за сделку, закрытие 0.33 дня.
    - **hanna** → €11,000, закрытие мгновенное.
    - **hands\_DE** → €11,500, но всего 1 лид.
    - Важно понимать, срабатывают ли эти дорогостоящие сделки в долгосрочной стратегии.
  - Самые быстрые закрытия сделок
    - **2005\_Lost\_DE** → 0.33 дня, практически мгновенное закрытие.
    - **col\_DE** → 0.00 дней, сделка закрывается в день обращения.
    - Высокая скорость закрытия означает легкость в принятии решения клиентом.
  - Кампании с долгим циклом закрытия
    - **web1312\_DE** → 70.66 дней для закрытия сделки.
    - **web2408\_DE** → 42.31 дней.
    - **Akademia** → 40.67 дней.
    - Длительный цикл может указывать на сложность продаж или высокий уровень размышлений перед покупкой.

## **Выводы:**

- **Оптимизировать лид-генерацию**
    - Маркетинговые гиганты (**performancemax\_digitalmarkt\_ru\_DE, youtube\_shorts\_DE**) приносят много лидов, но конверсия средняя.
    - Необходимо работать над улучшением их качество лидов, чтобы закрывать больше сделок.
  - **Использовать кампании с быстрой конверсией**
    - Кампании типа **domain** и **hanna** показывают идеальную конверсию (100%), но они генерируют очень мало лидов.

- Есть смысл масштабировать такие модели.
- **Проанализировать кампании с долгим циклом закрытия**
  - **web1312\_DE** требует 70+ дней для закрытия сделки.
  - Возможно, стоит рассмотреть ускорение обработки лидов и стратегии работы с клиентами.
- **Сегментировать кампании по уровню доходности**
  - **2005\_Lost\_DE** и **hanna** дают самые дорогие сделки, но их объем очень низкий.
  - Возможно, стоит интегрировать их в премиальную стратегию.

#### 4.3\* Сравнение эффективности маркетинговых кампаний с точки зрения затрат на рекламу

**Проверим соответствие кампаний в dataframes deals и spends**

**deals** contains 155 campaigns

**spend** contains 52 campaigns

##### Campaigns in deals but not in spend:109

```
['1004start', '1005start', '1006start', '11.10_prizes_DE', '1406start', '1704start', '1706_DE', '2004start', '2005_Lost_DE', '26.10start', '29.11start', '2905start', '2day_DE', '5555_DE', 'ASA_de_DE', 'Akademia', 'Aussiedler_DE', 'BVRE', 'Berlin_DE', 'Blogff_DE', 'Bloggerlvan', 'BloggerShina_DE', 'Blogger_snezh_DE', 'Bloggerel_DE', 'Bloggernina_DE', 'Bloggerperr_DE', 'Bolgspeak_DE', 'Consult_DE', 'Dis_DE', 'Forum_DE', 'Genie_DE', 'Jobs_germany_DE', 'Kak_PL', 'LOST_de', 'Live_DE', 'Markt_DE', 'PL_life', 'Relocat_DE', 'Trigger_DE', 'adv_DE', 'anastasia_de_DE', 'anuta_DE', 'arina_DE', 'berlin_DE', 'bf_DE', 'blog_DE', 'blog_de_DE', 'bloggerdr_DE', 'bloggerfrai_DE', 'chat_DE', 'clever_DE', 'col_DE', 'domain', 'eha_DE', 'engwien_AT', 'euro_DE', 'forum_DE', 'germania_DE', 'germany_DE', 'ggg10_DE', 'grad_DE', 'hands_DE', 'hanna', 'hb_DE', 'ineng_DE', 'jet_DE', 'karta_DE', 'liveberlin_DE', 'liv_DE', 'magda_DE', 'mail_3', 'mail_DE', 'mailreg_DE', 'maria_DE', 'marue_DE', 'mu_DE', 'nina', 'nochtum_DE', 'of_DE', 'performancemax_digitalmarkt_ru_DE', 'pishet_DE', 'pl_com', 'podrugi_DE', 'podslush_DE', 'potsdam', 'price_DE', 'referral', 'ruslana_DE', 'sushechka_PL', 'talav_DE', 'tyk_DE', 'uk12_AT', 'uk_DE', 'wa_PL', 'warsaw_PL', 'web1312_DE', 'web2211_DE', 'web2311_DE', 'web2410_DE', 'web8_DE', 'webanons', 'webinar1604', 'webinar1906', 'welt_DE', 'women_DE', 'work_DE', 'work_wr', 'workingin_DE', 'yo_DE']
```

##### Campaigns in spend but not in deals:6

```
['01.02.24wide_webinar_DE', '07.12.23test_DE', '15.03.24recentlymoved_AT', '15.11.23wide_webinar_DE', '20.03.24recentlymoved_PL', '30.11.23wide_DE']
```

##### The number of campaigns is the same in both dataframes:46

```
['01.04.23women_PL', '02.05.24test_DE', '02.07.23wide_DE', '02.08.23interests_DE', '03.07.23women', '04.07.23recentlymoved_DE', '05.07.23interests_DE', '05.09.2023wide_DE', '07.07.23LAL_DE', '08.04.24wide_webinar_DE', '08.06.24wide_webinar_DE', '09.02.24berlin_dd_DE', '10.07.23wide_com_DE', '12.06.24wide_DE', '12.07.2023wide_DE', '12.09.23interests_Uxui_DE', '14.11.23wide_webinar_DE', '15.03.2024wide_AT', '15.04.24LAL_ab_PL', '15.07.23b_DE', '17.03.24wide_AT', '18.10.23wide_gos_DE', '1performancemax_wide_PL', '20.03.2024wide_PL', '20.03.24_widde_PL', '20.03.24interests_WebDev_AT', '20.03.24interests_WebDev_PL', '20.05.24interests_DE', '20.05.24wide_DE', '22.05.2024wide_DE', '24.07.2023wide_DE', '24.09.23retargeting_DE', 'bbo_DE', 'blog2_DE', 'brand_search_eng_DE', 'comp_search_DE', 'discovery_DE', 'discovery_wide1_AT', 'gen_analyst_DE', 'performancemax_eng_DE', 'performancemax_wide_AT', 'shorts_PL', 'unknown', 'web2408_DE', 'youtube_shorts_DE', 'youtube_shortsin_AT']
```

#### Анализ совпадений и различий в кампаниях между deals и spend

1. **Общие наблюдения**
  - в **deals** присутствует 155 уникальных кампаний, а в **spend** — 52.
  - **совпадающих кампаний** (тех, что есть в обеих таблицах) — **46**.
  - кампании, которые генерировали сделки, но не имеют рекламных затрат (deals только) → **109**.
  - кампании, которые были профинансираны, но не привели к сделкам (spend только) → **6**.
2. **Разбор кампаний, присутствующих только в deals**
  - 109 кампаний есть в deals, но нет в spend, что означает:
    - эти кампании привели к сделкам, но их расходы не зафиксированы в spend.
    - Возможные причины:
      - это органические или неплатные кампании (например, referral, domain, bloggerdr\_DE);
      - ошибка учета рекламных расходов (некоторые кампании могли быть профинансираны, но данные не были загружены в spend);
      - использование внешних источников привлечения клиентов, которые не отслеживаются как платная реклама.
3. **Разбор кампаний, присутствующих только в spend**
  - 6 кампаний есть в spend, но отсутствуют в deals, что означает:
    - по этим кампаниям были расходы на рекламу, но сделки не зарегистрированы;
    - возможные причины:
      - эти кампании не дали результата (низкая конверсия или некачественные лиды);

- данные по сделкам могли не синхронизироваться в deals;
- кампании были недавними, сделки могут появиться позже.

#### 4. Выводы и рекомендации

- 46 кампаний показали результат (они есть и в deals, и в spend), что говорит о стабильной маркетинговой активности.
- 109 кампаний со сделками, но без затрат → важно проверить, как они генерируют лиды (органика, внешний трафик).
- 6 кампаний без сделок, но с затратами → возможная низкая эффективность, стоит изучить CTR и качество лидов.

**Следующие шаги:**

- провести анализ ROI (возврата инвестиций) по spend для выявления неэффективных трат.
- визуализировать разницу по кампаниям, чтобы увидеть отклонения в эффективности.
- дополнительно проверить, какие кампании генерировали самые дорогие сделки без затрат (возможная бизнес-оптимизация).

### Проанализируем связь между затратами на маркетинговые кампании и их основными индикаторами эффективности

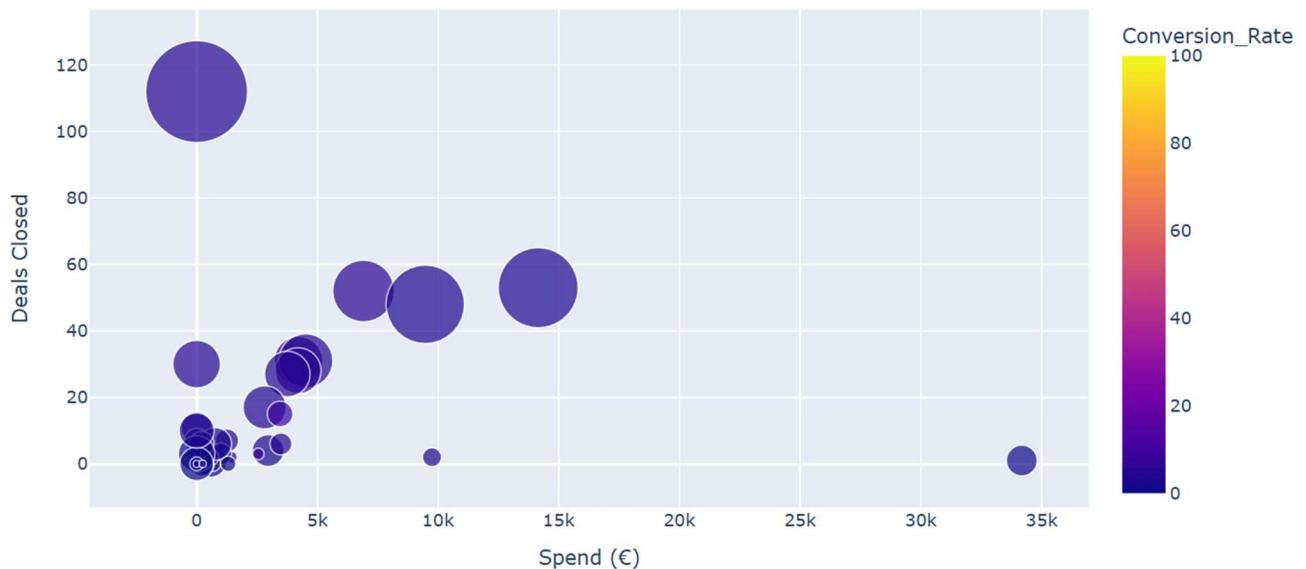
Summary table of campaign costs, leads, closed deals and conversion rate

	Campaign	Spend	Leads_Generated	Deals_Closed	Source	Conversion_Rate
119	performance_max_eng_DE	34183.45	245	1	Google Ads	0.41
152	youtube_shorts_DE	14149.22	1635	53	Youtube Ads	3.24
88	discovery_DE	9750.63	94	2	Google Ads	2.13
18	12.07.2023wide_DE	9471.52	1575	48	Tiktok Ads	3.05
2	02.07.23wide_DE	6913.60	975	52	Facebook Ads	5.33
...	...	...	...	...	...	...
147	women_DE	0.00	1	0	Bloggers	0.00
149	work_wr	0.00	6	0	Telegram posts	0.00
148	work_DE	0.00	317	10	Telegram posts	3.15
151	yo_DE	0.00	13	0	SMM	0.00
150	workingin_DE	0.00	48	0	Telegram posts	0.00

154 rows × 6 columns

Визуализируем на интерактивной диаграмме зависимость затрат на рекламные кампании с количеством лидов, числом закрытых сделок и коэффициентом конверсии

Marketing Campaigns Efficiency: Costs vs. Deals + Leads + Conversion Rate



### Как интерпретировать график?

- Точки справа → кампании с высокими затратами.
- Точки вверху → кампании с большим числом закрытых сделок.
- Чем теплее (красно-оранжевый), тем выше конверсия → показатель успешности кампаний.
- Маленькие точки → мало сделок, большие точки → высокая результативность кампаний.
- Кампания с высокой конверсией, но низкими затратами → идеальный баланс стоимости и качества.
- Кампания с высокими затратами, но низкой конверсией → возможно, неэффективна.

### Результаты анализа

1. **ТОП кампаний по расходам**
  - **performancemax\_eng\_DE** (€34,183.45) → затраты высокие, но всего 1 закрытая сделка. Конверсия 0.41% → крайне низкая эффективность.
  - **youtube\_shorts\_DE** (€14,149.22) → 1635 лидов и 53 сделки, конверсия 3.24% → разумные вложения, но можно улучшить конверсию.
  - **discovery\_DE** (€9,750.63) → 94 лидов, но только 2 сделки, конверсия 2.13% → требует пересмотра рекламной стратегии.

**Вывод:** некоторые дорогостоящие кампании приводят к минимальному числу сделок, что указывает на низкую рентабельность инвестиций (ROI).

2. **ТОП кампаний по конверсии**
  - **05.09.2023wide\_DE** (16.67%) → всего 6 лидов, но 1 сделка, высокая конверсия;
  - **20.05.24wide\_DE** (12.50%) → небольшие вложения (€76.72), но 1 сделка;
  - **gen\_analyst\_DE** (8.57%) → 35 лидов → 3 сделки, хорошая конверсия при разумных затратах (€2,552.93);
  - **brand\_search\_eng\_DE** (8.93%) → 168 лидов → 15 сделок, отличный показатель.

**Вывод:** кампании с высокой конверсией могут быть дешевыми, но они приносят стабильный поток сделок. Стоит обратить внимание на их масштабирование.

3. **Кампании с нулевой конверсией**
  - **discovery\_wide1\_AT** (€1,308.18) → 62 лидов, но ни одной сделки.
  - **youtube\_shortsin\_AT** (€249.28) → 14 лидов, но ни одной сделки.
  - **08.06.24wide\_webinar\_DE** (€345.16) → 112 лидов, но ни одной сделки.

**Вывод:** кампании, которые генерируют лиды, но не дают сделок, требуют глубокого анализа. Возможно, качество лидов низкое, аудитория неправильно сегментирована, или условия сделки недостаточно привлекательны.

### Выводы для бизнеса

- **Пересмотреть стратегию дорогих кампаний** → **performancemax\_eng\_DE** требует оптимизации, так как затраты огромные, а результат минимальный.

- Увеличить бюджет на кампании с хорошей конверсией → **brand\_search\_eng\_DE, gen\_analyst\_DE** показывают высокую отдачу и могут быть масштабированы.
- Проанализировать кампании с нулевой конверсией → возможно, требуется изменение аудитории, формата рекламных сообщений или креативов.
- Сделать A/B тестирование на **discovery\_DE** и **youtube\_shorts\_DE**, чтобы улучшить конверсию и рентабельность.

## 5 Анализ эффективности работы отдела продаж¶

(Sales Dept Efficiency Analys)

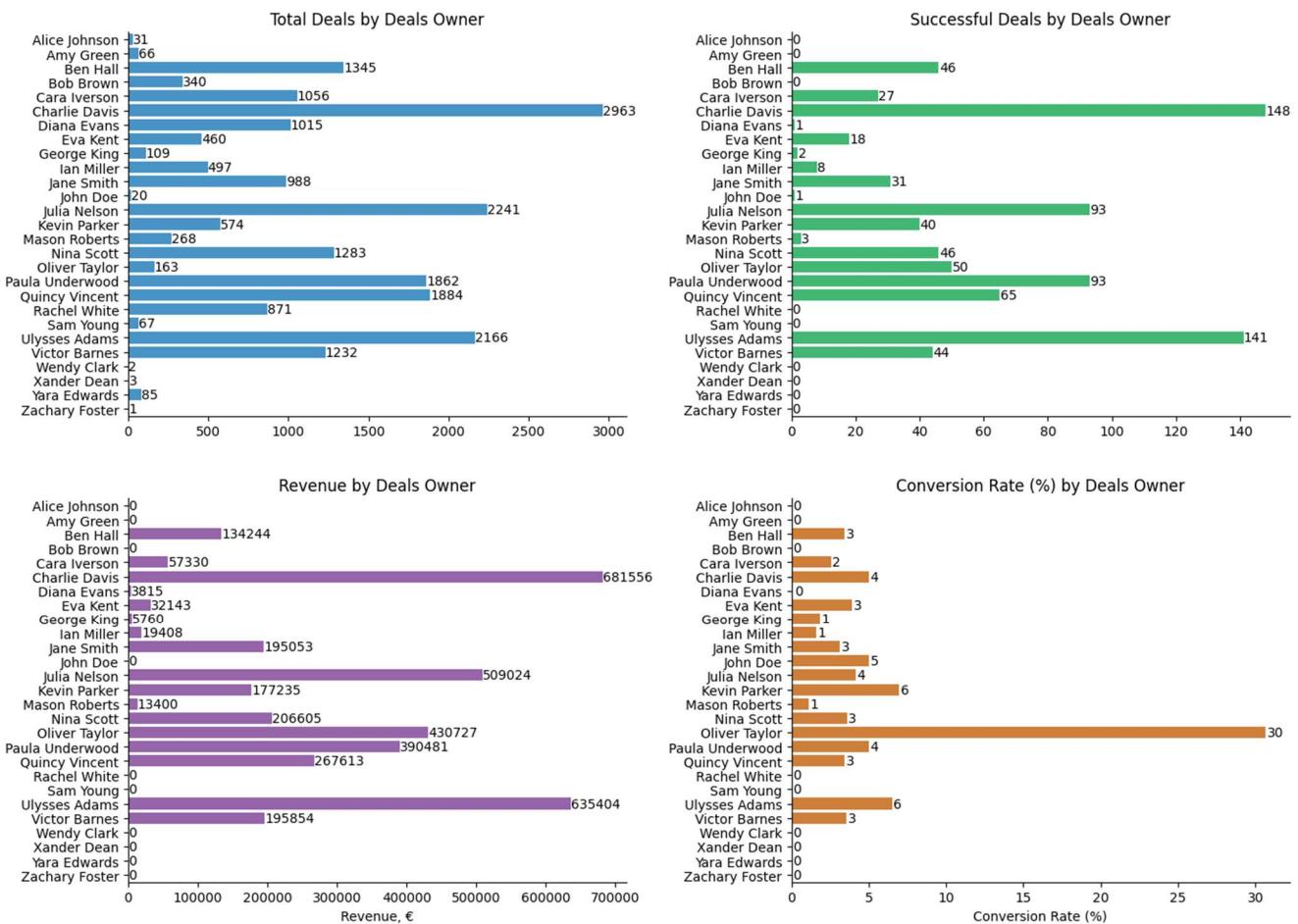
**5.1 Оцените эффективность отдельных владельцев сделок и рекламных кампаний с точки зрения количества обработанных сделок, коэффициента конверсии и общей суммы продаж.¶**

### 5.1.1 Анализ эффективности менеджеров продаж (владельцев сделок)

Summary table of sales manager performance indicators

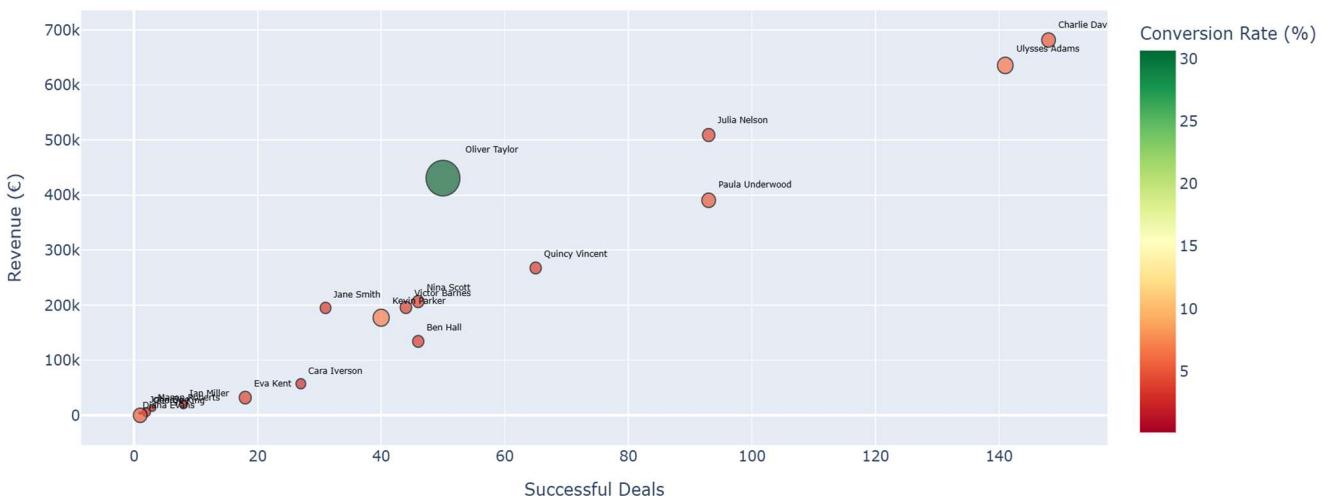
	Total_Deals	Successful_Deals	Revenue	Conversion_Rate (%)
Deal_Owner_Name				
Oliver Taylor	163	50	430,727.27	30.67
Kevin Parker	574	40	177,235.45	6.97
Ulysses Adams	2166	141	635,404.40	6.51
John Doe	20	1	0.00	5.00
Charlie Davis	2963	148	681,556.22	4.99
Paula Underwood	1862	93	390,481.51	4.99
Julia Nelson	2241	93	509,024.54	4.15
Eva Kent	460	18	32,143.02	3.91
Nina Scott	1283	46	206,605.76	3.59
Victor Barnes	1232	44	195,854.55	3.57
Quincy Vincent	1884	65	267,613.04	3.45
Ben Hall	1345	46	134,244.54	3.42
Jane Smith	988	31	195,053.19	3.14
Cara Iverson	1056	27	57,330.00	2.56
George King	109	2	5,760.00	1.83
Ian Miller	497	8	19,408.64	1.61
Mason Roberts	268	3	13,400.00	1.12
Diana Evans	1015	1	3,815.00	0.10
Alice Johnson	31	0	0.00	0.00
Bob Brown	340	0	0.00	0.00
Amy Green	66	0	0.00	0.00
Sam Young	67	0	0.00	0.00
Rachel White	871	0	0.00	0.00
Wendy Clark	2	0	0.00	0.00
Xander Dean	3	0	0.00	0.00
Yara Edwards	85	0	0.00	0.00
Zachary Foster	1	0	0.00	0.00

## Визуализация индикаторов эффективности менеджеров продаж



## Визуализация эффективности владельцев сделок, которые имеют не нулевую конверсию

Deal Owner Efficiency (Conversion Rate > 0%)



**Результаты анализа эффективности менеджеров продаж:  
ТОП-5 менеджеров по успешным сделкам и доходу**

Имя	Успешные сделки	Доход (€)	Конверсия (%)
Charlie Davis	148	681,556.22	4.99
Ulysses Adams	141	635,404.40	6.51
Oliver Taylor	50	430,727.27	30.67
Paula Underwood	93	390,481.51	4.99
Julia Nelson	93	509,024.54	4.15

**Менеджеры с наименьшими показателями**

Имя	Успешные сделки	Доход (€)	Конверсия (%)
Diana Evans	1	3,815.00	0.10
George King	2	5,760.00	1.83
Ian Miller	8	19,408.64	1.61
Alice Johnson	0	0.00	0.00
Bob Brown	0	0.00	0.00

**Основные выводы**

- Оливер Тейлор обладает самым высоким коэффициентом конверсии (30.67%), но у него сравнительно небольшое количество сделок.
- Чарли Дэвис и Улисс Адамс лидируют по количеству успешных сделок и совокупному доходу.
- Менеджеры с нулевой конверсией (например, Алиса Джонсон, Боб Браун) возможно не участвуют в активных продажах или требуют дополнительного обучения.
- Низкие показатели успешности (0-1 сделки) требуют анализа: возможно, есть сложности с закрытием или отсутствует проработанная стратегия.

**Рекомендации по повышению результативности отдела продаж.**

1. **Оптимизировать работу менеджеров с низкой конверсией**, выявить причины отказов и предложить дополнительные инструменты продаж.
2. **Проанализировать эффективность Оливера Тейлора**, изучить его методы работы и распространить среди других менеджеров.
3. **Усилить стратегию по увеличению количества сделок у топ-менеджеров** (Дэвис, Адамс, Тейлор) для масштабирования успешных подходов.
4. **Пересмотреть работу менеджеров с нулевыми продажами**, определить, нужна ли им дополнительная поддержка или перераспределение задач.

### 5.1.2 Анализ эффективности рекламных кампаний

Summary table of Campaign performance indicators

	Campaign	Total_Deals	Successful_Deals	Revenue	Conversion_Rate (%)	Spend	ROI(%)
118	performancemax_digitalmarkt_ru_DE	2653	112	570,485.31	4.22	0.00	0.00
2	02.07.23wide_DE	975	52	291,282.27	5.33	6,913.60	41.13
152	youtube_shorts_DE	1635	53	241,093.49	3.24	14,149.22	16.04
18	12.07.2023wide_DE	1575	48	208,790.00	3.05	9,471.52	21.04
4	03.07.23women	612	31	179,658.18	5.07	4,219.75	41.58
...	...	...	...	...	...	...	...
146	welt_DE	2	0	0.00	0.00	0.00	0.00
149	work_wr	6	0	0.00	0.00	0.00	0.00
150	workingin_DE	48	0	0.00	0.00	0.00	0.00
151	yo_DE	13	0	0.00	0.00	0.00	0.00
153	youtube_shortsin_AT	14	0	0.00	0.00	249.28	-1.00

154 rows x 7 columns

**Результаты анализа эффективности рекламных кампаний:**

**Лидеры по доходу (Revenue)**

Кампания	Сделки	Успешные сделки	Доход (€)	Конверсия (%)	ROI (%)
performancemax_digitalmarkt_ru_DE	2653	112	570,485.31	4.22%	0.00%
02.07.23wide_DE	975	52	291,282.27	5.33%	41.13%
youtube_shorts_DE	1635	53	241,093.49	3.24%	16.04%

Кампания	Сделки	Успешные сделки	Доход (€)	Конверсия (%)	ROI (%)
12.07.2023wide_DE	1575	48	208,790.00	3.05%	21.04%
03.07.23women	612	31	179,658.18	5.07%	41.58%

#### Выводы

- **performancemax\_digitalmarkt\_ru\_DE** лидирует по доходу (570K €), но у нее **отсутствуют данные о расходах**, поэтому невозможно рассчитать ROI.
- **02.07.23wide\_DE** имеет высокий доход (291K €) и ROI 41.13%, что делает ее успешной.
- **youtube\_shorts\_DE** имеет значительный доход (241K €), но ROI ниже (16.04%), что может говорить о высоких затратах.
- **03.07.23women** показывает хорошую конверсию (5.07%) и ROI 41.58%.

#### Лучшие кампании по конверсии (Conversion Rate)

Кампания	Конверсия (%)	Доход (€)	ROI (%)
hanna	100%	8,000.00	0.00%
domain	100%	833.33	0.00%
referral	33.33%	4,000.00	0.00%
web2311_DE	30.00%	12,858.18	0.00%

#### Выводы

- **hanna** и **domain** показывают 100% конверсию, но у них **нет данных о расходах**, поэтому ROI невозможно оценить.
- **referral** (33.33%) и **web2311\_DE** (30%) также демонстрируют высокий уровень успешных сделок **без вложений в рекламу**.
- **Эти кампании могут быть органическими источниками клиентов**, что стоит учитывать при планировании маркетинга.

#### Лучшие кампании по ROI (Return on Investment)

Кампания	ROI (%)	Доход (€)	Конверсия (%)
web2408_DE	153.41%	31,795.00	3.80%
18.10.23wide_gos_DE	54.37%	18,000.00	4.76%
02.07.23wide_DE	41.13%	291,282.27	5.33%

#### Выводы

- **web2408\_DE** демонстрирует невероятно высокий ROI (153.41%) — сверхэффективная реклама с минимальными затратами (205.91 €).
- **18.10.23wide\_gos\_DE** также имеет хороший ROI (54.37%) при сравнительно небольшом доходе.
- **02.07.23wide\_DE** — одна из самых прибыльных кампаний по доходу и ROI.

#### Убыточные кампании (ROI < 0%)

Кампания	ROI (%)	Доход (€)	Расходы (€)
performancemax_eng_DE	-0.94%	2,000.00	34,183.45
discovery_DE	-0.49%	5,000.00	9,750.63
01.04.23women_PL	-1.00%	0.00	357.25

#### Выводы

- **performancemax\_eng\_DE** убыточна: вложенные средства в 34K € не принесли прибыли.
- **discovery\_DE** тоже в минусе, имея низкий доход при значительных расходах (9.75K €).
- Некоторые кампании вообще не получили дохода, но имели расходы (например, **01.04.23women\_PL**).

#### Кампании без успешных сделок (0% конверсия)

Кампания	Сделки	Конверсия (%)	Расходы (€)
blog_DE	3	0%	0.00
forum_DE	5	0%	0.00
jobs_germany_DE	7	0%	0.00
genie_DE	12	0%	0.00

#### Выводы

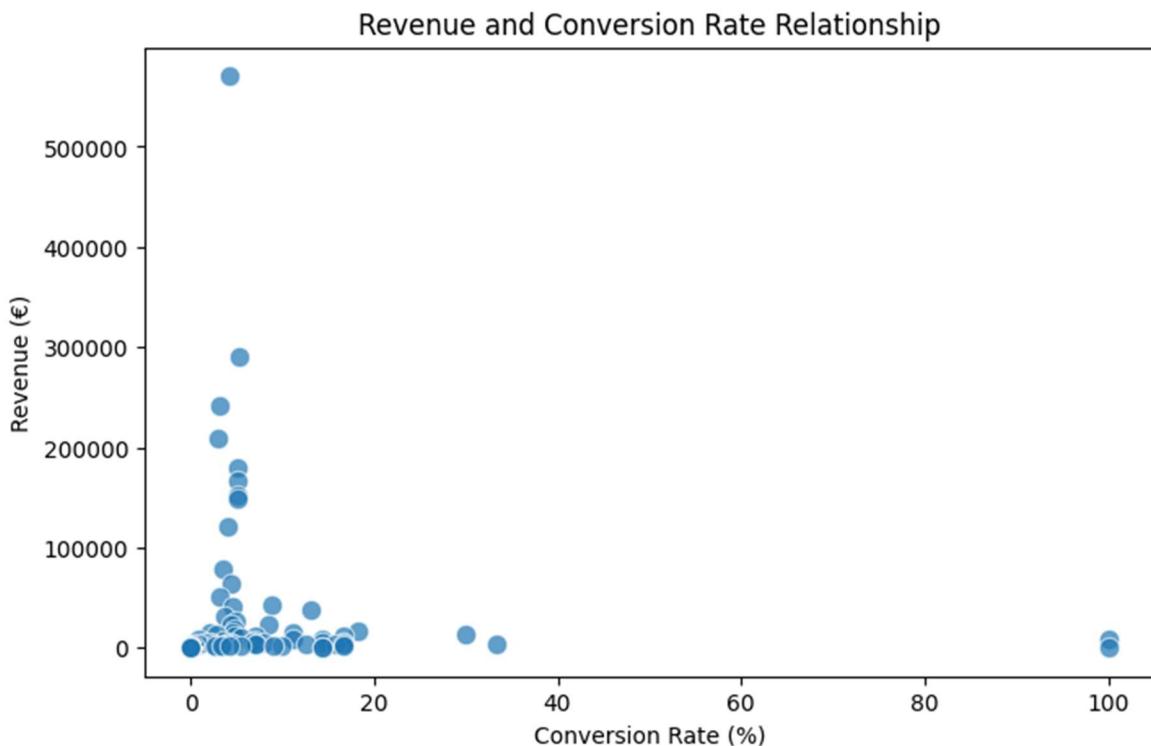
- **blog\_DE, forum\_DE, jobs\_germany\_DE, genie\_DE** — 0% конверсии и 0 € дохода.
- Эти кампании могли быть неправильно настроены или ориентированы не на ту аудиторию.
- Рекомендация: провести аудит и определить причины низкой эффективности.

#### Общие выводы и рекомендации

- Увеличить инвестиции в кампании с высоким ROI (**web2408\_DE, 02.07.23wide\_DE, 18.10.23wide\_gos\_DE**).

- Пересмотреть стратегию убыточных кампаний, возможно, требуется изменить таргетинг или предложение (`performancemax_eng_DE`, `discovery_DE`).
- Проанализировать органические кампании (`hanna`, `referral`, `domain`, `web2311_DE`), чтобы усилить их эффективность без вложений в рекламу.
- Оптимизировать расходы на кампании с высоким доходом, но низким ROI (например, `youtube_shorts_DE`).
- Удалить или полностью изменить кампании, которые не приносят дохода, но требуют затрат (например, `01.04.23women_PL`).
- Аудит кампаний с нулевыми конверсиями (`blog_DE`, `forum_DE`, `jobs_germany_DE`, `genie_DE`) — возможно, они настроены неэффективно.

#### *Связь коэффициента конверсии и общей суммы продаж по кампаниям*



#### **Основные наблюдения:**

- Концентрация данных в нижней части графика: большинство точек расположено в диапазоне 0–20% конверсии и 0–100,000€ продаж. — Это говорит о том, что **большая часть кампаний имеет умеренную конверсию и средний объём продаж**.

#### **Выбросы и аномальные значения**

- Есть один явный выброс с конверсией около 100% и общей суммой продаж ~190,000€.
- Это может быть специфическая кампания, где малый объём сделок даёт высокий процент конверсии.

#### **Общий тренд**

- Если точки сгруппированы вдоль восходящей линии, это означает, что увеличение конверсии приводит к росту продаж.
- Если точки хаотично расположены, значит, конверсия не всегда коррелирует с продажами, возможно, есть другие влияющие факторы.

#### **Выводы и рекомендации:**

- Анализировать выбросы → понять, почему одна кампания значительно превышает другие по показателям.
- Оценить эффективность низкоконверсионных кампаний → улучшить таргетинг, тестировать новые стратегии.
- Построить регрессионную модель → чтобы лучше понять, насколько конверсия влияет на продажи.

#### *Анализ степени влияния коэффициента конверсии (Conversion Rate) на доход (Revenue)*

Для этого построим регрессионную модель, которая поможет выявить, насколько коэффициент конверсии влияет на доход.

**Выберем модель линейной регрессии:**

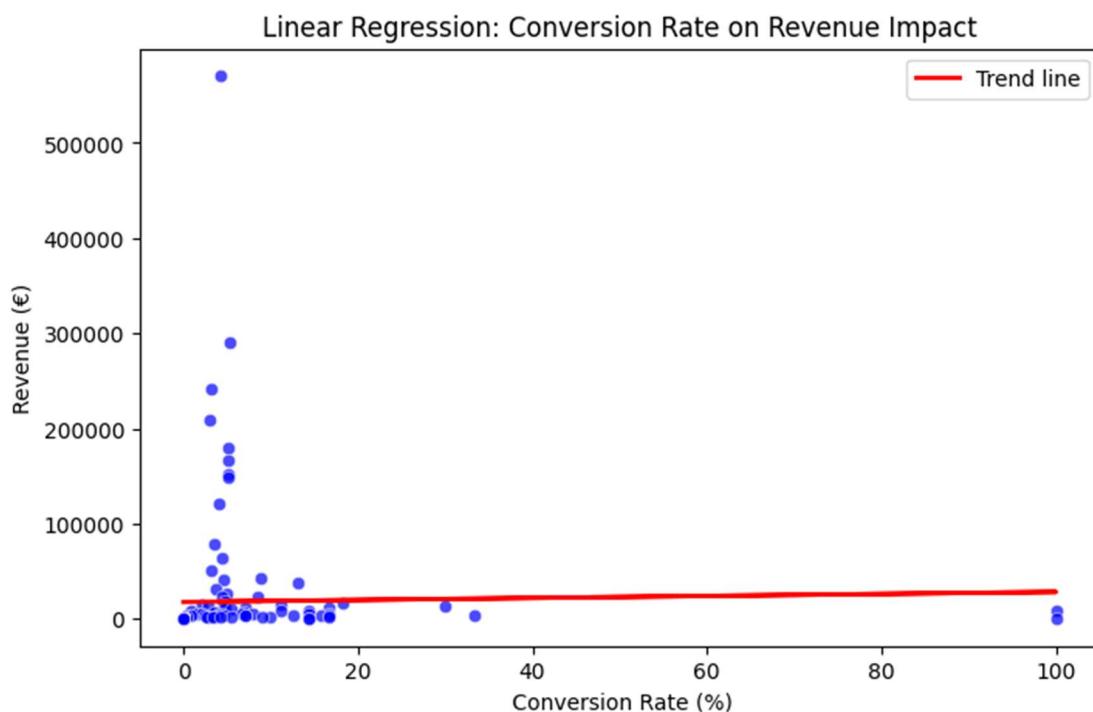
- Conversion\_Rate - предикат
- Revenue - целевая переменная

**Шаги анализа:**

1. Подготовить данные → убрать пропущенные значения.
2. Выполнить регрессионный анализ → использовать LinearRegression из sklearn.
3. Визуализировать зависимость → график с линией тренда.
4. Оценить качество модели → коэффициент детерминации  $R^2$ .

**Model Quality Assesment (coefficient  $R^2$ ):**

Determination Coefficient ( $R^2$ ): 0.0004



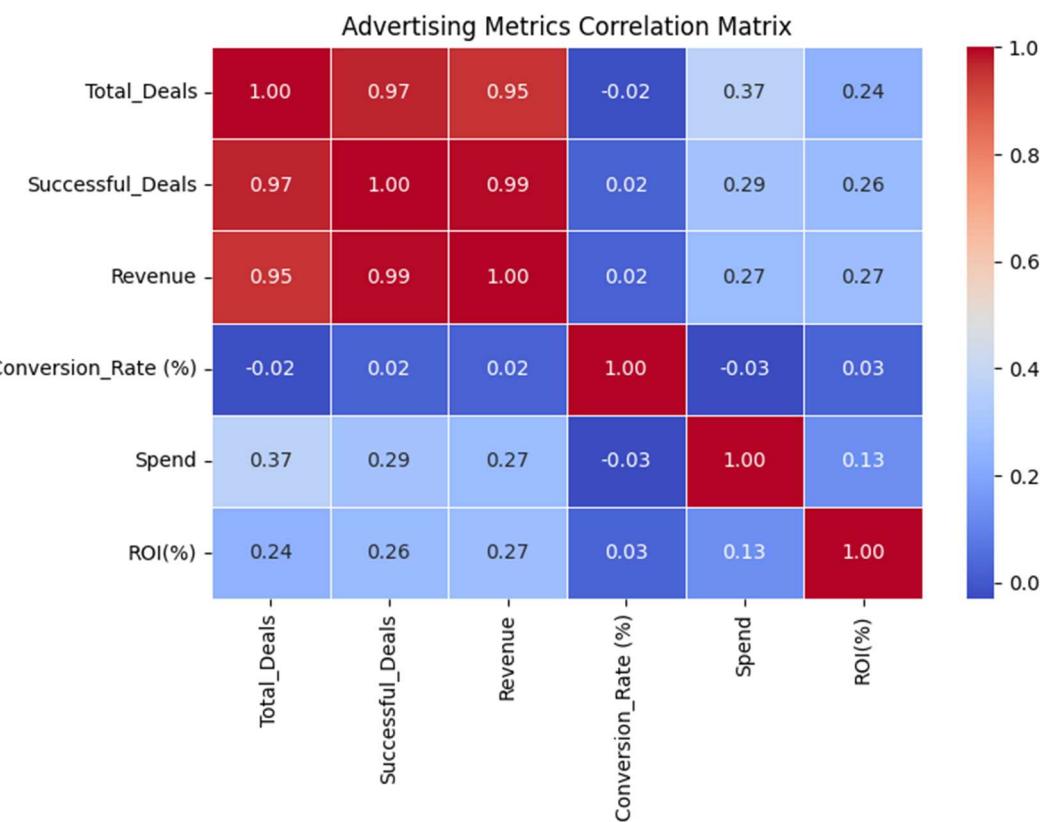
**Вывод по регрессионному анализу**

- Коэффициент детерминации ( $R^2$ ): 0.0009
- Это **крайне низкое значение**, означающее, что **конверсия практически не объясняет изменения в доходе**.
- Линейная регрессия предполагает наличие связи между коэффициентом конверсии и выручкой, но полученное значение  $R^2$  показывает отсутствие сильной зависимости.
- Красная линия тренда на графике почти горизонтальна, что говорит о том, что изменения в коэффициенте конверсии не приводят к значительным изменениям дохода.
- Разброс точек → данные распределены хаотично, без явного линейного тренда, что дополнительно подтверждает слабую связь.

**Что это означает для бизнеса?**

- Конверсия — не единственный фактор влияния на доход → другие переменные (например, средний чек сделки, качество аудитории, маркетинговые расходы) могут играть более значительную роль.
- Следует проанализировать дополнительные факторы → сегментировать сделки по другим параметрам (средний размер сделки, источник клиентов).

**Матрица корреляции рекламных метрик**



На основе представленной матрицы корреляции рекламных метрик можно выделить несколько ключевых наблюдений:

### 1. Общая картина корреляций

- Вариативность значений от -1 (сильная отрицательная связь) до 1 (сильная положительная связь).
- Холодные цвета (синие оттенки) → низкие значения.
- Тёплые цвета (красные оттенки) → высокие значения.
- Промежуточные (белый/светлый тон) → средние значения.
- Более светлые оттенки означают слабую корреляцию, а более насыщенные — сильную взаимосвязь.

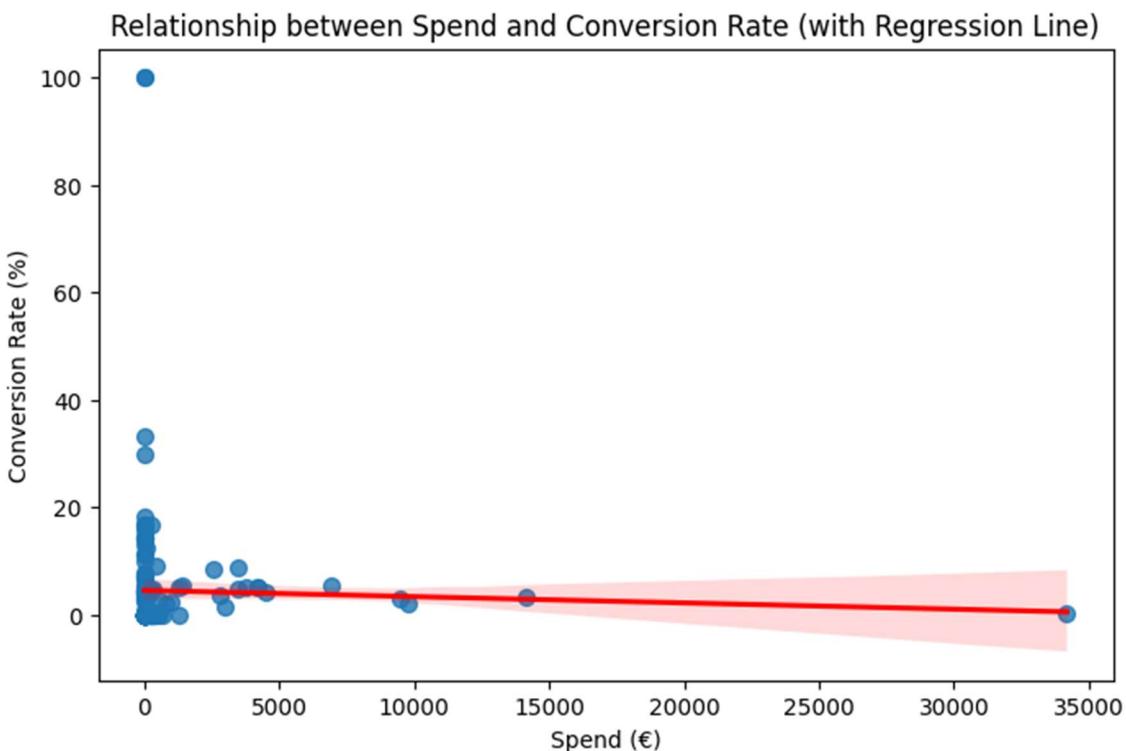
### 2. Основные выводы

- **Сильная положительная корреляция** между доходом (Revenue) и количеством успешных сделок (Successful\_Deals). Это логично: больше успешных сделок приводит к большему доходу.
- **Слабая или отрицательная корреляция** между затратами (Spend), процентом возврата инвестиций (ROI) и конверсией (Conversion\_Rate)
- Это означает, что увеличение затрат не всегда ведёт к росту конверсии (подтверждение вывода регрессионного анализа). Возможно, некоторые кампании неэффективны.
- **ROI** имеет положительную связь с доходом, но слабую с затратами. Это говорит о том, что увеличение бюджета не гарантирует рост ROI.

### 3. Рекомендации

- Провести анализ эффективности затрат – найти кампании, где высокий бюджет не приводит к росту конверсии.
- Усилить фокус на качественных сделках – чем выше процент успешных сделок, тем выше доход.
- Перераспределить ресурсы – инвестировать в кампании с высоким ROI, а не просто увеличивать бюджет на рекламу.

**Анализ взаимосвязи между затратами на маркетинг (Spend) и конверсией (Conversion Rate)**



#### Распределение точек

- Большинство данных сконцентрировано в нижней части графика, где конверсия <5%, даже при значительных расходах.
- Некоторые точки выделяются из общего тренда – они показывают конверсию выше 10–15%, что может говорить о более эффективных кампаниях.

#### Тренд

- общая тенденция**
- линия близка к горизонтальной, что означает **слабую или отсутствующую связь между затратами и конверсией**;
- при этом просматривается **слабый отрицательный наклон** линии тренда, что можно интерпретировать как "**больше расходов не приводит к улучшению конверсии**", возможно, есть проблемы с таргетингом или качеством аудитории.

#### Выбросы и аномалии

- свидетельствуют о наличии кампаний с высокими затратами, но низкой конверсией — это повод пересмотреть стратегию.
- и наоборот, кампании с высокой конверсией при небольших затратах — отличный пример эффективного использования бюджета.

#### Кампании, где высокий бюджет не приводит к росту конверсии

##### Критерии отбора:

- Высокие затраты – кампании с Spend > среднего значения
- Низкая конверсия – кампании с Conversion\_Rate < среднего значения.
- ROI < 1

Campaigns with low Spend budgets and high Conversion Rates  
(sort by Conversion Rate)

	Campaign	Spend	Conversion_Rate (%)	ROI(%)
88	discovery_DE	9,750.63	2.13	-0.49
29	1performancemax_wide_PL	2,961.38	1.53	0.84
119	performancemax_eng_DE	34,183.45	0.41	-0.94
89	discovery_wide1_AT	1,308.18	0.00	-1.00

#### Что делать с этими кампаниями?

- Аудит стратегии: проверить настройки таргетинга, аудитории и рекламных объявлений.

- Перераспределение бюджета: возможно, средства стоит направить в более успешные кампании.
- А/В тестирование: попробовать разные варианты креативов и форматов.

### **Кампании с низкими затратами, но высокой конверсией**

Выявление таких кампаний поможет выявить наиболее эффективные рекламные стратегии

#### **Критерии отбора:**

- Низкие затраты – кампании с Spend < среднего значения
- Высокая конверсия – кампании с Conversion\_Rate > среднего значения
- ROI > 1

Campaigns with low Spend budgets and high Conversion Rates

(sort by Convrtrision Rate)

	Campaign	Spend	Conversion_Rate (%)	ROI(%)
7	05.09.2023wide_DE	244.51	16.67	43.99
35	20.05.24wide_DE	76.72	12.50	38.10
22	15.03.2024wide_AT	425.01	9.09	3.19
28	18.10.23wide_gos_DE	325.10	4.76	54.37
39	24.07.2023wide_DE	280.35	4.55	14.34

#### **Что делать с этими кампаниями?**

- Масштабировать успешные стратегии – увеличить бюджет на кампании с доказанной эффективностью.
- Анализ аудитории – какие сегменты срабатывают лучше при минимальных затратах?
- Применить успешные методы к другимкампаниям – использовать подход из этих кампаний в других рекламных активностях.

## 6. Анализ платежей и продуктов

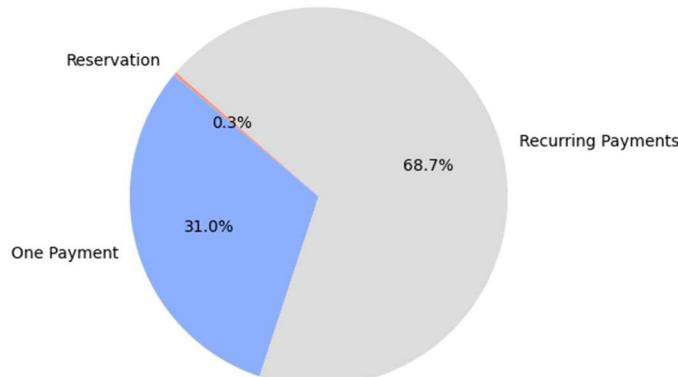
(Payments and Products Analysis)

### 6.1 Анализ распределения типов оплаты и их влияния на успешность сделок

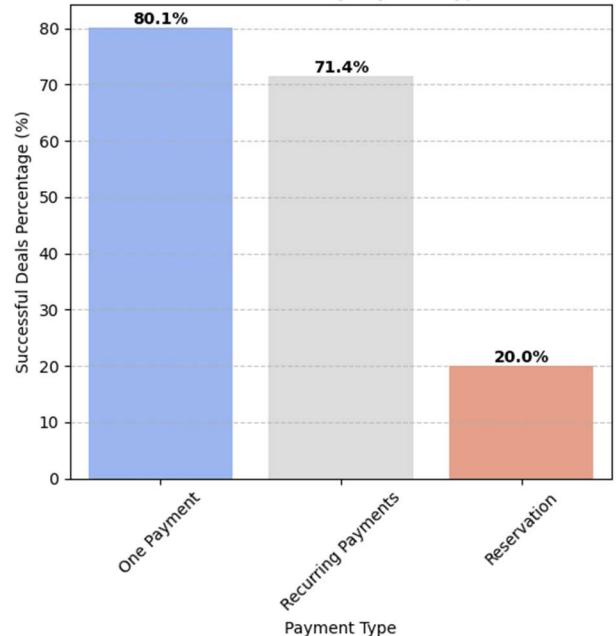
Summary table of Success Deals Rate by Payment Type

	Payment_Type	Total_Deals	Successful_Deals	Success_Rate (%)
0	One Payment	141	113	80.14
1	Recurring Payments	350	250	71.43
2	Reservation	5	1	20.00

Successful Deals Distribution by Payment Type



Success Rate by Payment Type



#### Анализ успешности сделок по типу оплаты

- "One Payment" (80.14%) → самый успешный метод оплаты, обеспечивает высокую конверсию. Это может указывать на предпочтение клиентов платить сразу, без долгосрочных обязательств.
- "Recurring Payments" (71.43%) → также высокие показатели, но чуть ниже. Это логично, потому что подписочные платежи могут быть отменены клиентами до завершения сделки.
- "Reservation" (20.00%) → значительно ниже других методов. Это может означать, что многие клиенты резервируют, но не оплачивают финальную сумму, либо система бронирования требует доработки.

#### Выводы и рекомендации

- Единоразовые платежи — ключевой метод → усиливать удобство оплаты и мотивировать клиентов завершать сделки.
- Продвижение подписочных платежей → возможно, стоит внедрить бонусы или скидки, чтобы увеличить конверсию.
- Оптимизировать систему бронирования → проверить UX (User Experience) процесса (логичность шагов, доступность информации, скорость выполнения, работу в мобильной версии, прозрачность оплаты), рассмотреть оплату депозита за бронирование.

### 6.2 Анализ популярности и успешности различных продуктов и типов обучения

Определим:

- Популярность → какие продукты и типы обучения имеют больше всего сделок
- Успешность → процент успешных сделок (Success\_Rate или Conversion Rate - в юнит-экономике коэффициент конверсии в первую сделку)

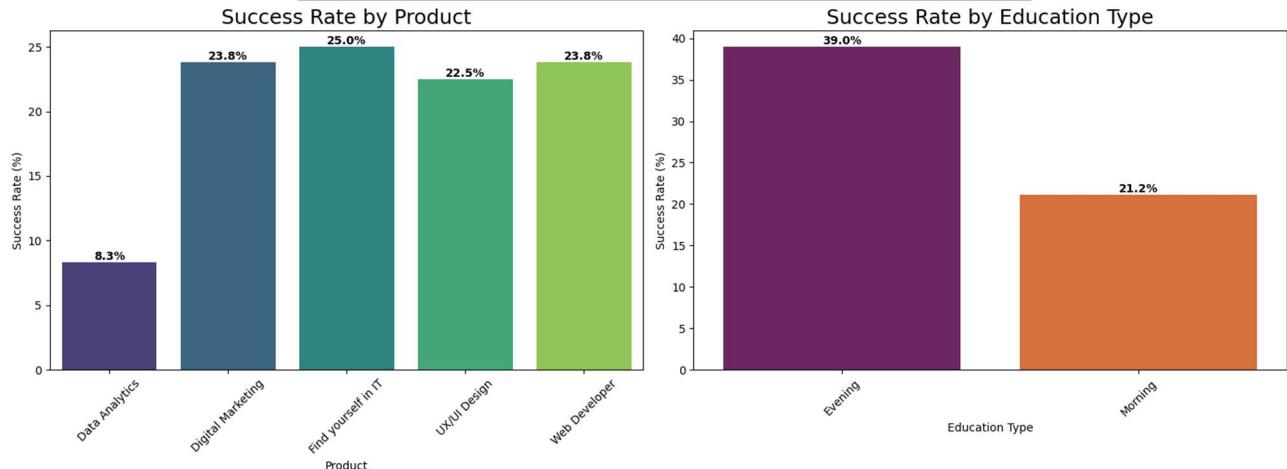
#### Анализ популярности и успешности продуктов (Product) и типов обучения (Education\_Type)

Summary table of Success Rate by Product

	Product	Total_Deals	Successful_Deals	Success_Rate (%)
0	Data Analytics	36	3	8.33
1	Digital Marketing	1978	471	23.81
2	Find yourself in IT	4	1	25.00
3	UX/UI Design	1019	229	22.47
4	Web Developer	575	137	23.83

Summary table of Success Rate by Education Type

	Education_Type	Total_Deals	Successful_Deals	Success_Rate (%)
0	Evening	444	173	38.96
1	Morning	3157	668	21.16



## Популярность и успешность продуктов

### Анализ продуктов

- Общие тренды
- Digital Marketing** (23.81%) → наиболее популярный продукт по количеству сделок (1978), но успех ниже 25%. Это может говорить о высокой вовлеченности, но недостаточной конверсии.
- UX/UI Design** (22.47%) → востребованный курс, но процент завершённых сделок невысокий. Можно протестировать маркетинговые улучшения.
- Web Developer** (23.83%) → схожие показатели с маркетингом и UX/UI, но меньше сделок (575). Возможно, спрос ниже или аудитория требует персонализированного подхода.
- Data Analytics** (8.33%) → самый низкий процент успешных сделок. Возможно, связано с высокой конкуренцией или сложностью курса.
- Find Yourself in IT** (25.00%) → малый объём данных (всего 4 сделки), но успех выше среднего. Для выводов надо больше информации которая к сожалению имеет очень много пропуски в CRM.

### Рекомендации:

- Улучшить доведение сделок до оплаты в Digital Marketing и UX/UI Design.
- Усилить маркетинг для Web Developer и изучить уникальные потребности аудитории.
- Проверить факторы отказов для Data Analytics (цена, сложность, конкуренция).
- 

## Популярность и успешность типов обучения

### Общие тренды

- Вечерние занятия** (42.33%) → имеют значительно более высокий процент успешных сделок. Это говорит о том, что студенты предпочитают обучаться после работы или дневных дел, когда у них больше времени на обучение.
- Утренние занятия** (22.87%) → хоть и имеют гораздо больше сделок (2895), но процент успешности почти в 2 раза ниже. Возможно, клиенты не успевают завершать оплату из-за занятости в первой половине дня.
- Что может влиять на успешность?**
  - Гибкость вечернего обучения → людям удобнее совмещать работу и обучение.
  - Проблемы утренних занятий → не все успевают пройти процесс покупки или выделить время.

- Конверсия в оплату → у вечерних студентов выше мотивация довести курс до конца, что повышает процент оплаты.

### Рекомендации

- Оптимизировать утренние курсы → возможно, стоит предложить гибкие форматы (записи занятий, старт в удобное время).
- Активнее продвигать вечерние курсы → они имеют высокий потенциал успеха и привлекают мотивированную аудиторию.
- Провести A/B-тестирование → сравнить разницу в маркетинговых кампаниях для каждого типа обучения.

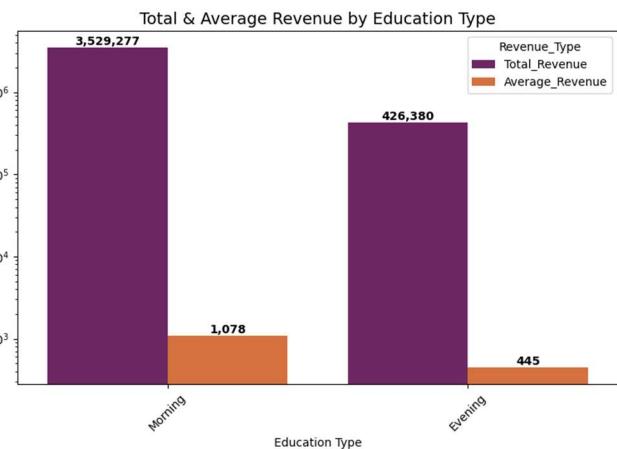
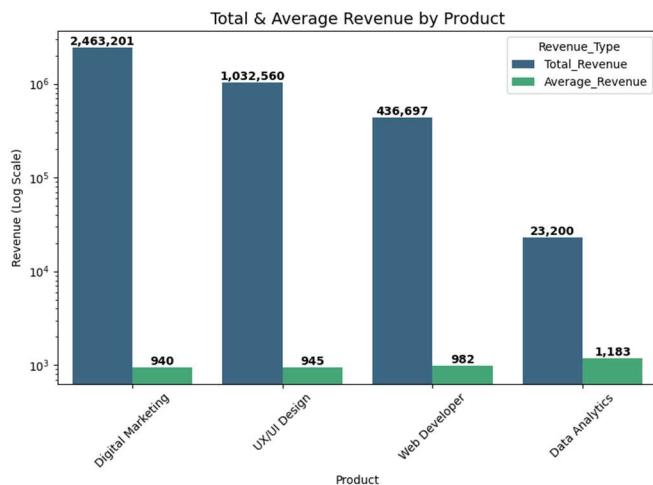
### Анализ доходности продуктов (Product) и типов обучения (Education\_Type)

Summary table of Total Revenue by Product

	Product	Total_Deals	Successful_Deals	Total_Revenue	Average_Revenue
1	Digital Marketing	1978	471	2,463,200.91	939.66
2	UX/UI Design	1019	229	1,032,559.55	944.86
3	Web Developer	575	137	436,696.67	981.79
0	Data Analytics	36	3	23,200.00	1,183.33

Summary table of Total Revenue by Education Type

	Education_Type	Total_Deals	Successful_Deals	Total_Revenue	Average_Revenue
1	Morning	3157	668	3,529,276.68	1,077.92
0	Evening	444	173	426,380.45	444.85



### Доходность продуктов

#### Общие тренды:

- "Digital Marketing" (2,463,200€) → лидер по суммарному доходу, но его средний чек (939.66€) немного ниже, чем у других популярных продуктов. Вероятно, это связано с большим объемом сделок при относительно невысокой стоимости курса.
- "UX/UI Design" (1,032,559€) → второй по доходности продукт, но его средний чек (944.86€) чуть выше, что говорит о потенциальной устойчивости курса.
- "Web Developer" (436,696.67€) → хорошая доходность при среднем чеке (981.79€), но объем сделок заметно ниже.
- "Data Analytics" (23,200€) → крайне низкая доходность при среднем чеке (1,183€), что может указывать на проблемы с продажами или слабый спрос.

#### Что можно улучшить?

- Пересмотреть маркетинг Data Analytics → изучить причины низкой конверсии и протестировать новые подходы.

- Оптимизировать **Web Developer** → возможно, стоит сделать упрощенные тарифы или рассмотреть больше рекламных инвестиций.
- Поддерживать **Digital Marketing** и **UX/UI** → эти продукты успешно приносят доход, но могут быть дополнительные возможности роста через upselling (например базовый курс - upsell на PRO-версию с доступом к закрытым материалам, или групповые занятия → upsell на индивидуальные уроки с наставником) или другие, продвинутые версии курсов.

#### **Вывод:**

- **Digital Marketing** и **UX/UI Design** — самые прибыльные продукты с высоким объемом успешных сделок.
- **Web Developer** тоже приносит доход, но требует дополнительного маркетинга.
- **Data Analytics** — требует внимания к стратегии продаж и спросу.

#### **Доходность типов обучения**

##### **Общие тренды:**

- **Morning** (3,529,276€) → лидер по общему доходу, так как имеет наибольшее количество сделок (3157). Средний доход на успешную сделку — 1,077.92€, что выше вечерних курсов.
- **Evening** (426,380€) → гораздо меньший доход, но процент успешных сделок выше. Средний чек ниже (444.85€), возможно, вечерние программы привлекательны из-за более доступных цен.

##### **Выводы**

- **Утреннее обучение** приносит больше денег, но требует анализа причин отказов (как увеличить успешность сделок).
- **Вечерние курсы** имеют ниже доход, но могут быть перспективны для роста, если предложить дополнительные услуги или повысить цены.

##### **Рекомендации**

- **Проверить маркетинг утреннего сегмента** → изучить аудиторию, готовы ли они платить больше, почему сделки завершаются не всегда.
- **Оптимизировать вечерние курсы** → повысить цены или расширить предложение, сохраняя удобный формат для студентов.

## 7 Географический анализ

(Geographical analysis)

### 7.1 Анализ географического распределения сделок по городам

Строим сводную таблицу рейтинга успешных сделок по городам

Summary table of Deals Statistic by Cities

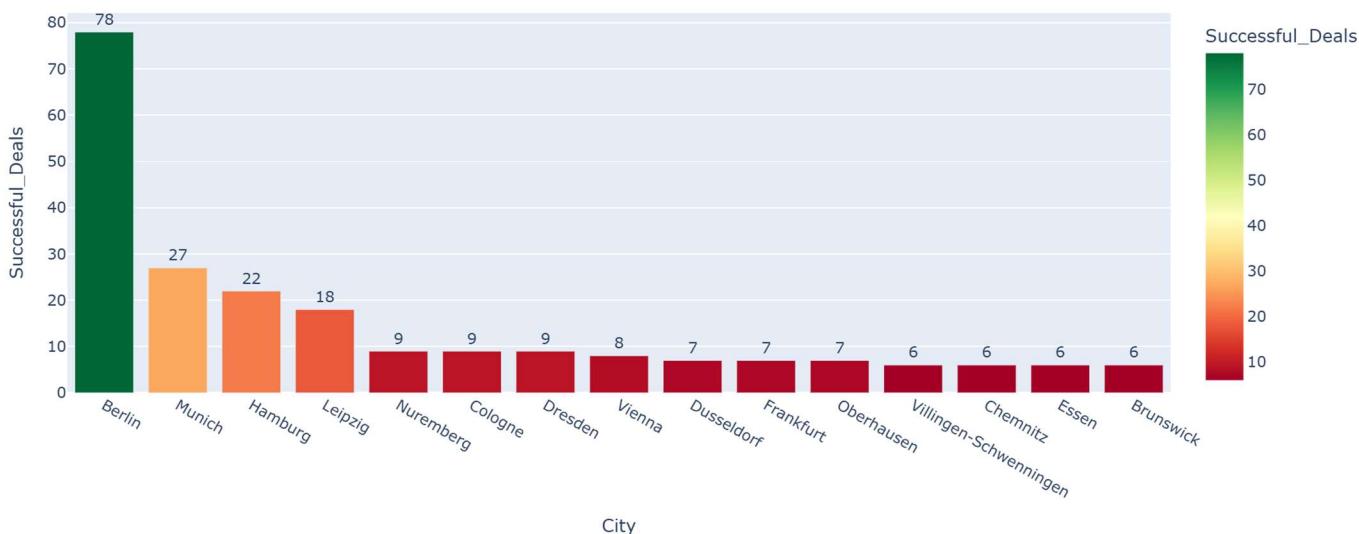
	City	Country	Total_Deals	Successful_Deals	Latitude	Longitude	Success_Rate (%)
86	Berlin	Germany	182	78	52.51	13.40	42.86
517	Munich	Germany	74	27	48.14	11.58	36.49
289	Hamburg	Germany	62	22	53.55	10.00	35.48
435	Leipzig	Germany	45	18	51.34	12.37	40.00
573	Nuremberg	Germany	45	9	49.45	11.08	20.00
...	...	...	...	...	...	...	...
867	Śrem	Poland	1	0	52.09	17.02	0.00
852	Würzburg	Germany	6	0	49.78	9.94	0.00
853	Włodawa	Poland	1	0	51.54	23.53	0.00
843	Wolfhagen	Germany	2	0	51.33	9.20	0.00
848	Wysoka Gryfińska	Poland	1	0	53.29	14.61	0.00

867 rows × 7 columns

Визуализируем топ 100 городов по числу успешных сделок

Top N:  15

Top 15 (from 867) Cities by Successful Deals



Отображаем на карте города с числом сделок, количеством успешных из них и коэффициента успешности сделки

Map of Total Deals, Successful\_Deals and Success\_Rate (%)



## Выводы из анализа географического распределение сделок по городам

Из географии распределения сделок видно:

- Германия - ключевой рынок для школы онлайн программирования.
- Визуально в пределах страны сделки распределены неравномерно, с примерным соотношением:
  - Запад-Восток 60%/40, (наибольшая концентрацией соответственно **Дюссельдорф-Ессен** и **Берлин-Лейпциг-Дрезден**);
  - Север - Юг 30%/70%, (наибольшая концентрацией соответственно **Гамбург и Мюнхен**);
  - центральные земли Германии Тюрингия и Хессен наименее охвачены,
- Конверсия в успешные сделки **в восточных землях** сравнительно выше чем **в западных**,
- Такая же примерно картина наблюдается по отношению конверсии **в южных землях к северным**.
- Наибольшее число сделок отмечается в столицах земель и промышленных центрах.
- **Лидеры по количеству сделок: Берлин (182 сделок), Мюнхен (74 сделок), Гамбург (62 сделок).**
- **Лидеры по успешному проценту сделок:**
  - Города с 100% успешностью (например, Heidenheim an der Brenz, Schwäbisch Hall, Minden, Markgröningen).
  - Высокая успешность также наблюдается в Оберхаузене (77,78%), Наумбурге (75%) и Вене (66,67%).
- **Города с низкой успешностью:** Нюрнберг (20%), Росток (8,33%), Шверин (20%).

Другие географические наблюдения Германия доминирует, но есть сделки в Австрии, Польше, Казахстане, России, Нидерландах, Франции, Таиланде, Словакии и ОАЭ.

- Восточная Германия (Дрезден, Лейпциг, Хемниц) имеет средний показатель успеха (32-60%).
- Западная Германия показывает вариативность, от низких (15% в Дортмунде) до высоких (77% в Оберхаузене).

## Выводы и возможные направления дальнейшего анализа

- Факторы, влияющие на успех – почему в Оберхаузене такой высокий процент сделок, а в Ростоке низкий?
- Более полное сравнение регионов – выявление тенденций между севером и югом, востоком и западом.
- Связь с экономическими показателями – корреляция успешности сделок с экономическим развитием городов.

## Фильтрация городов по странам и успешности сделок – определение наиболее перспективных локаций

### Рассчитаем и визуализируем статистику сделок по странам

Summary table of Deals Statistic by Country

Total Country: 28

Total City: 866

Total Deals: 2159

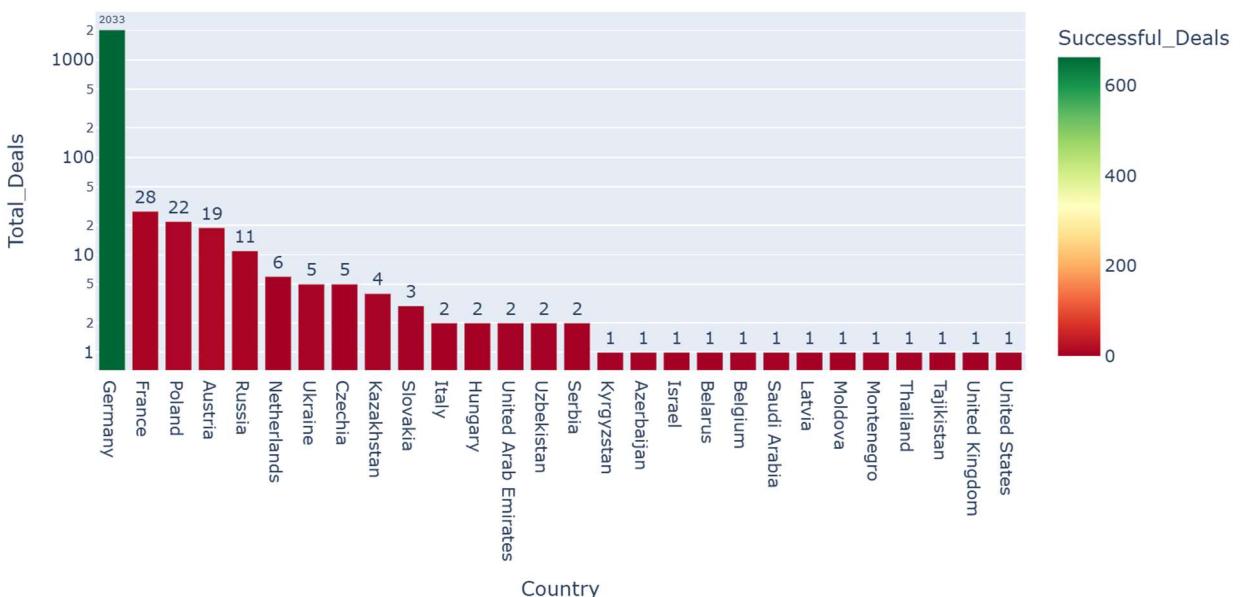
Total Successful Deals: 722

Average Successful Rate: 55.29%

	Country	City_num	Total_Deals	Successful_Deals	Success_Rate (%)
6	Germany	769	2033	664	32.66
5	France	23	28	7	25.00
16	Poland	16	22	8	36.36
0	Austria	8	19	11	57.89
17	Russia	9	11	6	54.55
...	...	...	...	...	...
14	Montenegro	1	1	1	100.00
22	Thailand	1	1	1	100.00
21	Tajikistan	1	1	0	0.00
25	United Kingdom	1	1	0	0.00
26	United States	1	1	1	100.00

28 rows × 5 columns

Tonal Deals and Successful Deals by Country



## Выводы из анализа успешных сделок по странам

### Общий обзор

- Всего стран: 28
- Всего городов: 867
- Всего сделок: 2,159
- Всего успешных сделок: 722
- Средний процент успешных сделок: 55.29%

### Лидеры по количеству сделок

- Германия → 2,033 сделок, 664 успешных (32.66% успеха)
- Франция → 28 сделок, 7 успешных (25.00% успеха)
- Польша → 22 сделки, 8 успешных (36.36% успеха)
- Австрия → 19 сделок, 11 успешных (57.89% успеха)
- Россия → 11 сделок, 6 успешных (54.55% успеха)
  - Германия — лидер по количеству сделок, но процент успешности ниже среднего (32.66%).
  - Австрия и Россия демонстрируют высокий уровень конверсии, несмотря на меньший объем сделок, это **наиболее перспективные рынки** после Германии для масштабирования.
  - Франция и Польша — **перспективные рынки**, но конверсия требует внимания.

### Страны с высокой успешностью

- Италия, Сербия, Венгрия, Кыргызстан, Азербайджан, Бельгия, Саудовская Аравия, Черногория, Таиланд, США → 100% успешных сделок!
- Казахстан → 75.00% успешных сделок (3 из 4)
- Украина → 60.00% успешных сделок (3 из 5)
- Австрия → 57.89% успешных сделок (11 из 19)
- Россия → 54.55% успешных сделок (6 из 11)
- Эти страны показывают высокую конверсию, из-за небольшого количества сделок, однако можно предположить высокую заинтересованности клиентов.
- У остальных стран из перечисленных конверсия - 100%, при одной сделке.

#### **Страны с 0% успешных сделок**

- Израиль, Беларусь, Латвия, Молдова, Таджикистан, Великобритания → сделки были, но ни одной успешной.

#### **Рекомендации:**

- Отделу маркетинга следует повести **анализ причин отказов** → отсутствие спроса, слабый маркетинг или проблемы с процессом оплаты?
- Отделу продаж для масштабирования следует работать над оптимизацией стратегии продаж **на перспективные рынки** → тестировать индивидуальные предложения, рекламные кампании или дополнительные скидки.

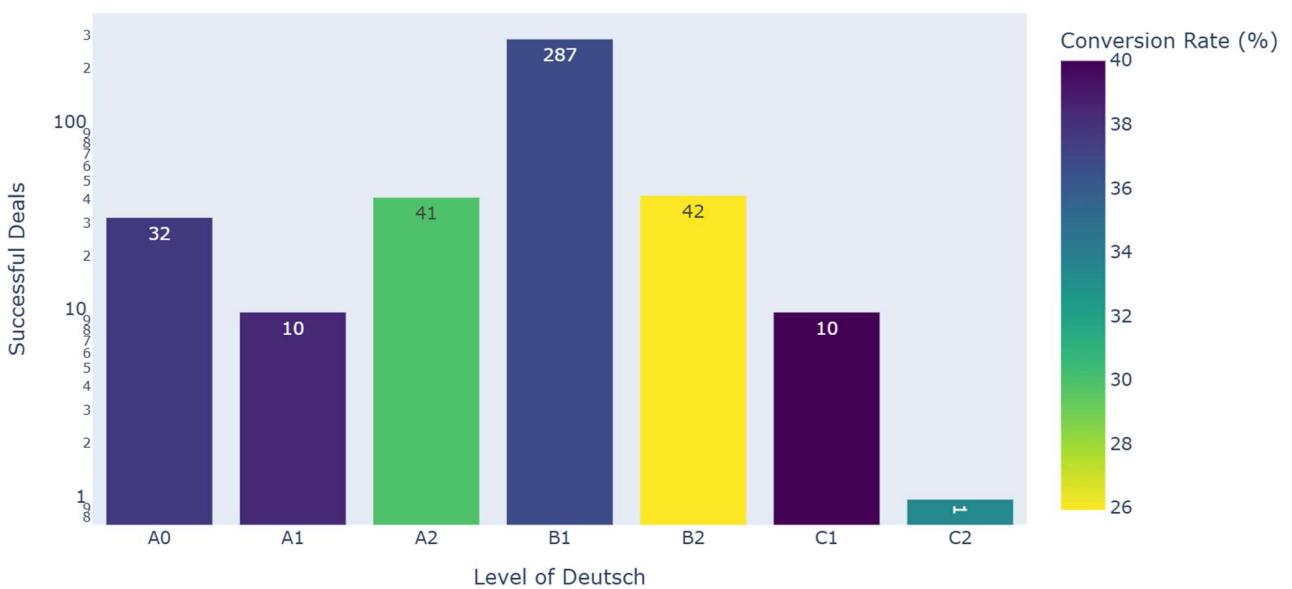
## **7.2 Анализ влияния уровня знания немецкого языка на успешность сделок в разных городах**

**Посчитаем и визуализируем статистику успешности сделок в разрезе уровня владения немецким языком**

Summary table of Deals Statistic by Level of Deutsch

	Level_of_Deutsch	Total_Deals	Successful_Deals	Lost_Deals	Conversion_Rate
0	A0	85	32	39	37.65
1	A1	26	10	11	38.46
2	A2	137	41	74	29.93
3	B1	780	287	353	36.79
4	B2	162	42	90	25.93
5	C1	25	10	11	40.00
6	C2	3	1	0	33.33

## Successful Deals and Conversion Rate by German Language Level (log scale)



### Выходы из анализа успешности сделок в зависимости от уровня знания немецкого языка

#### Общие наблюдения

- Наибольшее количество сделок → уровень B1 (780 сделок)**, что подтверждает, что средний уровень владения немецким языком (B1) является наиболее распространённым среди клиентов.
- Самая высокая конверсия → уровень C1 (40.00%)**, что говорит о том, что высокий уровень владения языком способствует успешности сделок.
- Самая низкая конверсия → уровень B2 (25.93%)**, несмотря на относительно большое количество сделок (162).
- Новички (A0, A1, A2)** → имеют среднюю конверсию (29-38%), что может указывать на сложность коммуникации на низком уровне владения языком.

#### Уточненные выводы

- Уровень B1** → самый распространённый среди клиентов, но конверсия не максимальная (36.79%).
- Уровень C1** → показывает лучшую конверсию (40.00%), что подтверждает важность владения языком на продвинутом уровне.
- Уровень B2** → неожиданно низкая конверсия (25.93%), возможно, из-за недостаточного уровня владения языком для сложных переговоров.
- Начальные уровни (A0, A1, A2)** → сделки происходят, но конверсия средняя, что может указывать на барьеры в коммуникации.

#### Рекомендации

- Оптимизировать работу с клиентами уровня B2 → возможно, требуется дополнительная поддержка или адаптация материалов.
- Усилить маркетинг для клиентов с уровнем C1 → они показывают лучшую конверсию, значит, стоит привлекать больше таких клиентов.
- Разработать стратегии для клиентов A0-A2 → возможно, стоит предложить дополнительные языковые ресурсы или поддержку на английском.

**Вывод:** Чем выше уровень владения немецким языком, тем выше вероятность успешной сделки, но B2 требует дополнительных данных для анализа.

### В каких городах знание языка оказывает наибольшее влияние на успешность сделки?

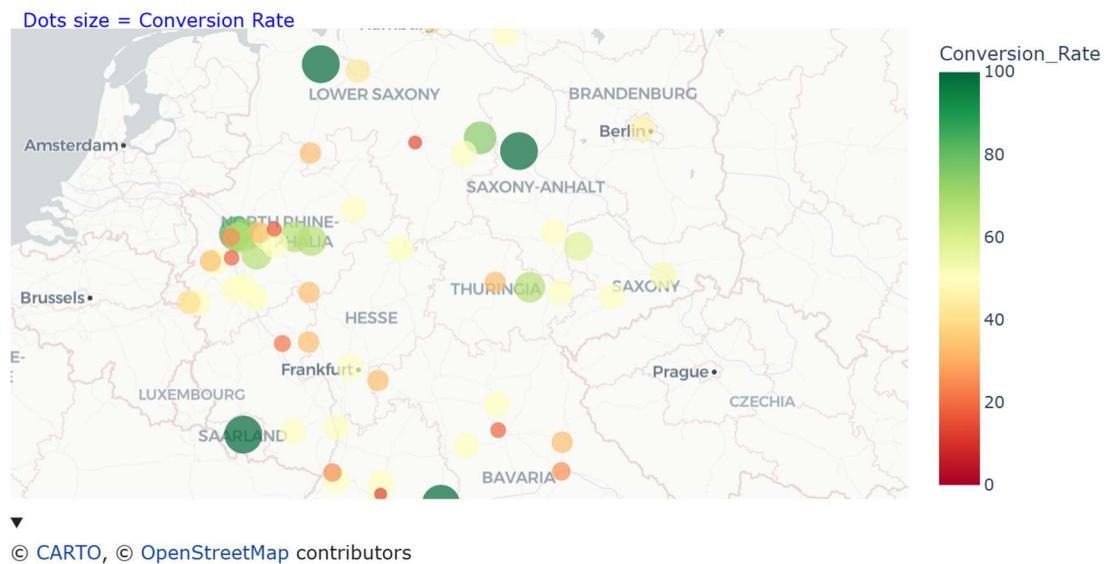
#### Рассчитаем показатели успешности сделок в зависимости от уровня владения немецким

Summary Table of Deals Success Rate depending on Level of German Language by City

	City	Conversion_Rate	Latitude	Longitude
0	Eppelborn	100.00	49.41	6.96
1	Heidenheim an der Brenz	100.00	48.68	10.15
2	Haldensleben	100.00	52.29	11.41
3	Oldenburg	100.00	53.14	8.21
4	Schwarzwalde	100.00	47.98	7.82
5	Oberhausen	83.33	51.47	6.85
6	Wolfsburg	75.00	52.42	10.79
7	Essen	71.43	51.46	7.02
8	Menden (Sauerland)	66.67	51.44	7.80
9	Jena	66.67	50.93	11.59
10	Arnsberg	66.67	51.40	8.06
11	Mülheim an der Ruhr	66.67	51.43	6.88
12	Wuppertal	66.67	51.26	7.18
13	Leipzig	57.78	51.34	12.37
14	Villingen-Schwenningen	55.56	48.06	8.49
15	Munich	53.33	48.14	11.58
16	Dresden	52.22	51.05	13.74
17	Kassel	50.00	51.32	9.50
18	Ludwigsburg	50.00	48.90	9.19
19	Korschenbroich	50.00	51.19	6.51

## Интерактивная карта успешности сделок по городам в зависимости от уровня немецкого языка

Map of Deal Success Rate by City depending on the level of German



Проверим гипотезу о наличии статистической зависимости между уровнем знания немецкого языка и успешностью сделок

Проведем хи-квадрат тест на выборке успешных и потерянных сделок по уровням знания немецкого языка.

Хи-квадрат тест проверяет гипотезу о независимости двух категориальных переменных.

Если  $p\text{-value} < 0.05$ , можно считать, что есть статистически значимая связь между знанием немецкого и успешностью сделки.

Если  $p\text{-value} \geq 0.05$ , значит данные не дают оснований утверждать, что знание языка влияет на финальный статус сделки.

Основные шаги теста:

#### Фильтруем данные

- Исключаются строки, где уровень немецкого "unknown".
- Оставляются только сделки, которые закрылись успешно (Payment Done) или потеряны (Lost).

#### Кодируем категории

- Сделки, которые успешно завершены (Payment Done) получают значение 1.
- Потерянные сделки (Lost) кодируются как 0.
- Новый столбец Stage\_Encoded позволяет анализировать эти категории как числовые данные.

#### Создание таблицы сопряженности (частотная таблица)

- Создается таблица частот, где строки — уровень немецкого, а столбцы — количество успешных и потерянных сделок.

#### Хи-квадрат тест (chi2\_contingency)

- Проверяет зависимость между двумя категориальными переменными (Level\_of\_Deutsch и Stage).
- Возвращает статистику хи-квадрат (chi2), p-value, степени свободы (dof) и ожидаемые частоты (expected).

#### Вывод результатов с цветным форматированием

- Выводятся статистика хи-квадрат и p-value, которые помогут определить, есть ли значимая связь.

**Chi2 Statistic: 11.80**

**P-value: 0.06664**

#### Интерпретация результатов хи-квадрат теста

- **Chi2 Statistic (11.80)** → это значение характеризует степень отклонения наблюдаемых данных от ожидаемых при условии независимости Level\_of\_Deutsch и стадиями Stage. Чем выше значение Chi-square, тем больше различий между группами.
- **P-value (0.06664)** → это вероятность получить такие же или более экстремальные различия случайно, если между Level\_of\_Deutsch и Stage нет реальной связи.

#### Анализ результата:

- Обычно пороговое значение p-value = 0.05 → если  $p < 0.05$ , можно сказать, что есть статистически значимая связь.
- В данном случае **p-value = 0.06664** → немного больше 0.05, но все же близко к значимости.
- Это означает, что есть некоторая зависимость, но она не достигает строгого уровня статистической значимости.

**Вывод:** возможно, уровень немецкого языка влияет на успешность сделок, но эффект не настолько сильный, чтобы быть строго подтвержденным статистикой.

## 8. Продуктовая аналитика

### 8.1. Расчет юнит-экономики по продуктам

#### Определение количества лидов(UA), маркетингового бюджета (AC) и стоимости привлечения льда (LTC)

сравниваем число лидов в Deals и Contacts и берем там где больше логика:

- пользователь не определился, его контакты занесены в Contacts, но сделка не открыта
- контакты пользователя занесены в Contacts, но менеджер не отреагировал
- сделка автоматически создалась при регистрации, но по каким-то причинам данные не занесены в Contacts

Результат:

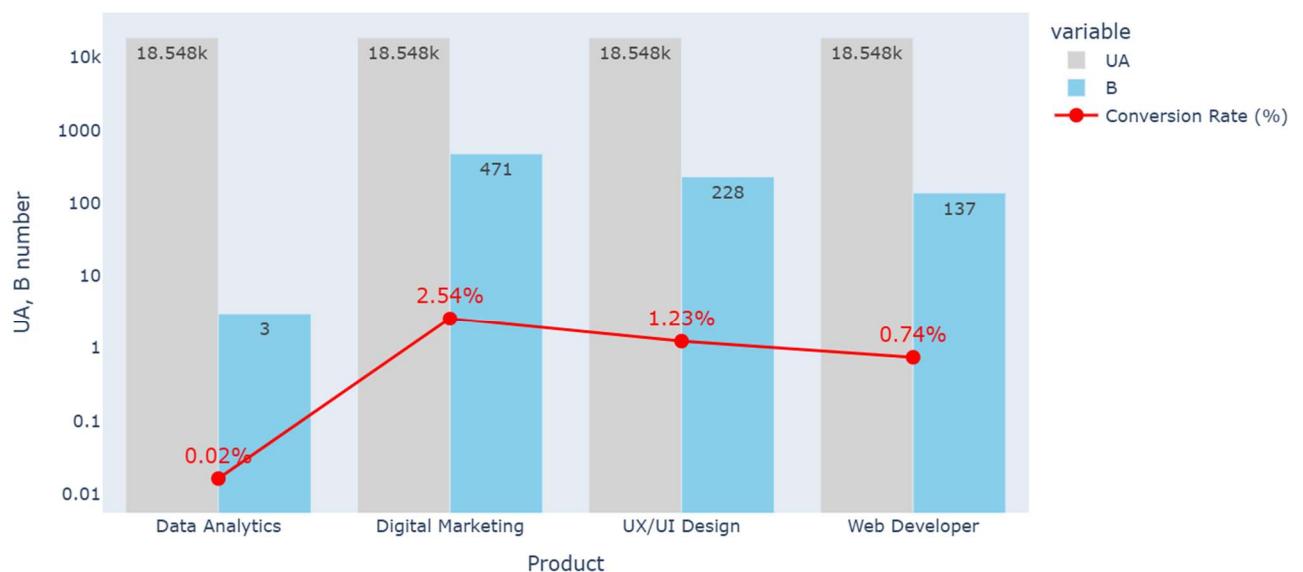
- Количество лидов в таблице Deals: 18089
- Количество лидов в таблице Contacts: 18548
- Решение по количеству лидов: 18548
- Маркетинговый бюджет: 149523.45€
- Стоимость привлечения потенциального клиента: 8.06€

#### Рассчет метрик юнит-экономики про продуктам (предлагаемый курс обучения)

Unit Economics Summary Table (Product)

	Product	B	T	AOV	Revenue	UA	AC	LTC	C1	CLTC	APC	CLTV	LTV	CM
0	Data Analytics	3	20.00	1,183.33	23,200.00	18548	149,523.45	8.06	0.02	498.41	6.67	7,888.89	127.60	22,171.43
1	Digital Marketing	471	2,877.00	939.66	2,463,200.91	18548	149,523.45	8.06	2.54	3.17	6.11	5,739.70	14,575.16	2,701,905.24
2	UX/UI Design	228	1,170.00	944.86	1,032,559.55	18548	149,523.45	8.06	1.23	6.56	5.13	4,848.65	5,960.16	1,103,996.13
3	Web Developer	137	505.00	981.79	436,696.67	18548	149,523.45	8.06	0.74	10.91	3.69	3,619.00	2,673.08	494,307.76
4	TOTAL:	839.00	4,572.00	865.19	3,955,657.13	18,548.00	149,523.45	8.06	0.05	178.22	5.45	4,714.73	213.27	3,806,133.68

Units Acquisition (UA), Buyers (B), Conversion Rate (C1) per Product (log scale)

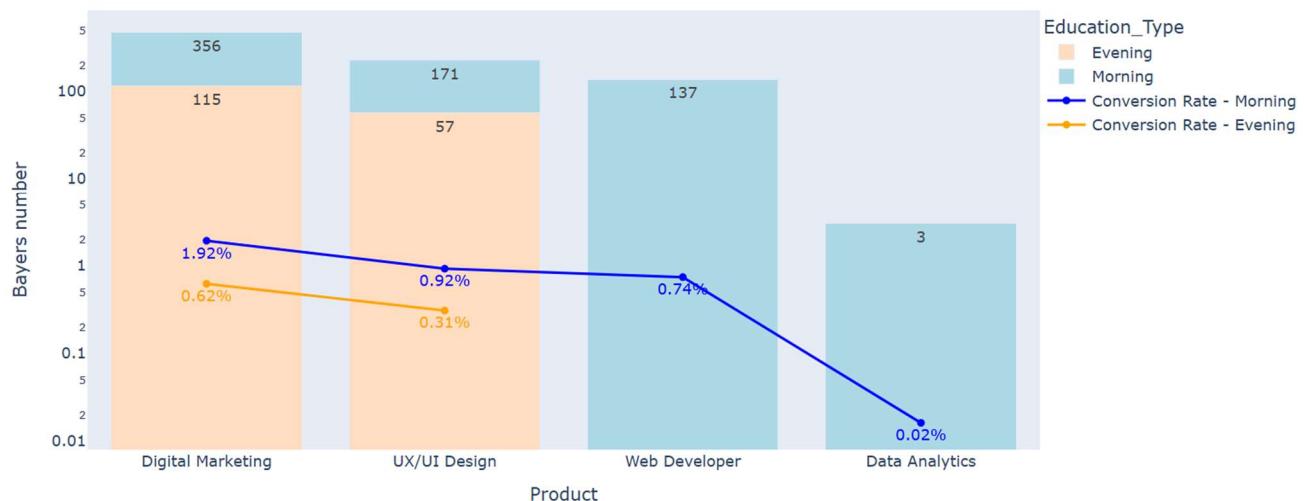


## Рассчет метрик юнит-экономики про продуктам (предлагаемый курс обучения) и типам обучения (утренний курс/вечерний курс)

Unit Economics Summary Table (Product & Education Type)

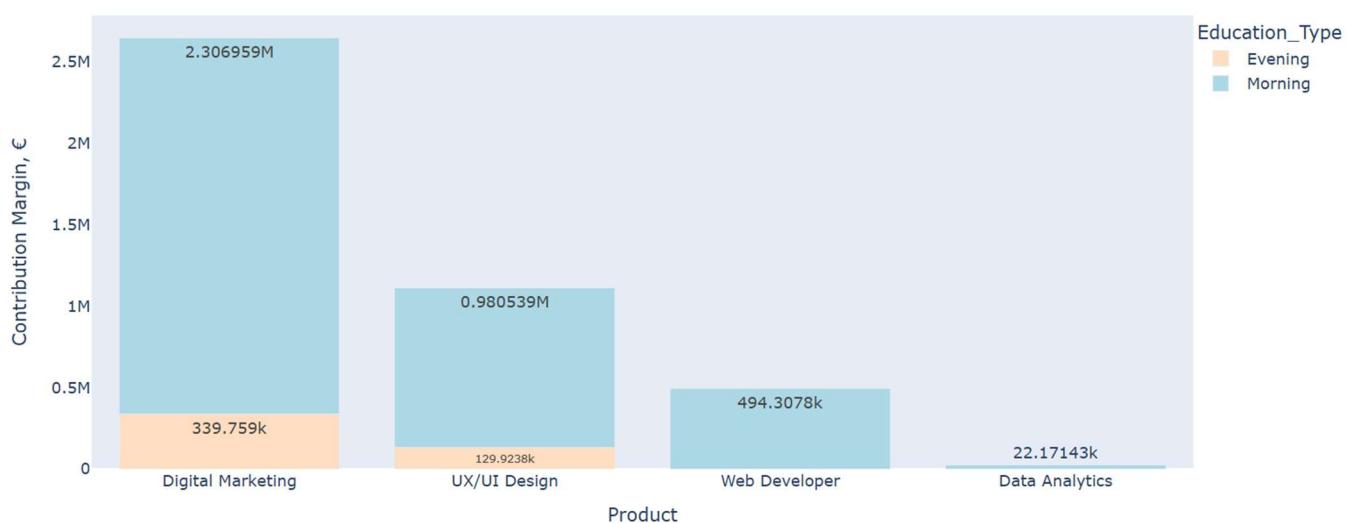
	Product	Education_Type	B	T	AOV	Revenue	UA	AC	LTC	C1	CLTC	APC	CLTV	LTV	CM
0	Data Analytics	Morning	3	20.00	1,183.33	23,200.00	18548	149,523.45	8.06	0.02	498.41	6.67	7,888.89	127.60	22,171.43
1	Digital Marketing	Evening	115	783.00	435.83	309,581.82	18548	149,523.45	8.06	0.62	13.00	6.81	2,967.43	1,839.84	339,758.97
2	Digital Marketing	Morning	356	2,094.00	1,102.41	2,153,619.09	18548	149,523.45	8.06	1.92	4.20	5.88	6,484.42	12,445.84	2,306,959.13
3	UX/UI Design	Evening	57	284.00	462.74	116,798.63	18548	149,523.45	8.06	0.31	26.23	4.98	2,305.60	708.53	129,923.81
4	UX/UI Design	Morning	171	886.00	1,108.39	915,760.92	18548	149,523.45	8.06	0.92	8.74	5.18	5,742.89	5,294.56	980,539.13
5	Web Developer	Morning	137	505.00	981.79	436,696.67	18548	149,523.45	8.06	0.74	10.91	3.69	3,619.00	2,673.08	494,307.76
6	TOTAL:		839.00	4,572.00	865.19	3,955,657.13	18,548.00	149,523.45	8.06	0.05	178.22	5.45	4,714.73	213.27	3,806,133.68

Buyers (B), Conversion Rate (C1) per Product and Education Type (log scale)



## Визуализируем маржинальную прибыль по продуктам

Contribution Margin per Product and Education Type (log scale)



## Анализ юнит-экономики образовательных продуктов

### 1. Ключевые метрики:

Метрика	Значение
Число клиентов (B)	839
Число платежей (T)	4,572
Средний чек (AOV)	865.19 €
Общий оборот	3,955,657.13 €
Средняя конверсия (C1)	0.05 (5%)
Маржинальная прибыль (CM)	3,806,133.68 €

## 2. ТОП-3 самых прибыльных продуктов:

Продукт	Формат	Маржинальная прибыль (CM, €)
Digital Marketing	Morning	2,306,959.13
UX/UI Design	Morning	980,539.13
Web Developer	Morning	494,307.76

Выводы:

- Утренний Digital Marketing приносит наибольшую прибыль с высоким средним чеком.
- UX/UI Design (Morning) и Web Developer (Morning) также высокодоходные, стоит увеличить рекламный бюджет.

## 3. Наименее прибыльные направления:

Продукт	Формат	Маржинальная прибыль (CM, €)
Data Analytics	Morning	22,171.43
UX/UI Design	Evening	129,923.81

Выводы:

- Data Analytics (Morning) показывает низкую прибыльность ( $CM = 22,171.43 €$ ) и слабую конверсию ( $C1 = 0.02$ ).
- UX/UI Design (Evening) имеет низкую маржинальную прибыль, требует пересмотра маркетинговой стратегии.

## 5. Анализ стоимости привлечения клиентов (CLTC)

Продукт	Формат	CLTC (€)
Data Analytics	Morning	498.41
UX/UI Design	Evening	26.23
Digital Marketing	Morning	4.20

Выводы:

- Data Analytics (Morning) — высокая стоимость привлечения ( $CLTC = 498.41 €$ ) при слабой конверсии.
- Digital Marketing (Morning) — низкая стоимость привлечения ( $CLTC = 4.20 €$ ), лучший канал для роста.

## 6. Рекомендации:

- Фокус на утренние курсы Digital Marketing, UX/UI Design, Web Developer, они самые прибыльные.
- Пересмотреть стратегию Data Analytics (Morning), низкая конверсия + высокая стоимость привлечения.
- Оптимизировать UX/UI Design (Evening), низкая прибыльность требует переработки продукта или маркетинга.
- Увеличить маркетинговый бюджет на Digital Marketing (Morning), лучший канал для роста доходов.

## 8.2. Определение точек роста бизнеса

Следуя теории юнит-экономики обозначим шаги для определения точек (драйверов) роста бизнеса:

- Используя Теорию ограничений Голдратта найдем "бутилочное горлышко", а именно, улучшим поочередно метрики UA, LTC, AOV, C1, CLTC на 10% ;
- Определим, какие из метрик оказывают наибольшее влияние на CM, при наименьших затратах;
- Найдем оптимальную конфигурацию метрик с учетом ограничений - затрат на улучшение метрики.
- Генерируем гипотезы, связанные только с этой метрикой и строим HADI циклы для проверки гипотез.

5. Рассчитываем параметры тестов.
6. Моделируем проведение A/B тестирование.

**Impact on Contribution Margin (CM) of 10% improvement in metrics: ['UA', 'LTC', 'AOV', 'C1', 'CLTC']**

	Product	Education_Type	Metric	Tested	CM	CM_Change
1	Data Analytics	Morning	AOV	24,538.10	10.67	
1	Data Analytics	Morning	C1	22,307.36	0.61	
1	Data Analytics	Morning	CLTC	22,171.43	0.00	
1	Data Analytics	Morning	LTC	22,320.96	0.67	
1	Data Analytics	Morning	UA	22,171.43	0.00	
2	Digital Marketing	Evening	AOV	373,884.40	10.04	
2	Digital Marketing	Evening	C1	339,894.91	0.04	
2	Digital Marketing	Evening	CLTC	339,758.97	0.00	
2	Digital Marketing	Evening	LTC	339,908.50	0.04	
2	Digital Marketing	Evening	UA	339,758.97	0.00	
3	Digital Marketing	Morning	AOV	2,537,804.56	10.01	
3	Digital Marketing	Morning	C1	2,307,095.06	0.01	
3	Digital Marketing	Morning	CLTC	2,306,959.13	0.00	
3	Digital Marketing	Morning	LTC	2,307,108.65	0.01	
3	Digital Marketing	Morning	UA	2,306,959.13	0.00	
6	UX/UI Design	Evening	AOV	143,065.71	10.12	
6	UX/UI Design	Evening	C1	130,059.74	0.10	
6	UX/UI Design	Evening	CLTC	129,923.81	0.00	
6	UX/UI Design	Evening	LTC	130,073.33	0.12	
6	UX/UI Design	Evening	UA	129,923.81	0.00	
7	UX/UI Design	Morning	AOV	1,078,742.57	10.02	
7	UX/UI Design	Morning	C1	980,675.06	0.01	
7	UX/UI Design	Morning	CLTC	980,539.13	0.00	
7	UX/UI Design	Morning	LTC	980,688.66	0.02	
7	UX/UI Design	Morning	UA	980,539.13	0.00	
9	Web Developer	Morning	AOV	543,888.06	10.03	
9	Web Developer	Morning	C1	494,443.69	0.03	
9	Web Developer	Morning	CLTC	494,307.76	0.00	
9	Web Developer	Morning	LTC	494,457.28	0.03	
9	Web Developer	Morning	UA	494,307.76	0.00	

**Impact on Contribution Margin (CM) of 10% improvement in metrics: ['UA', 'LTC', 'AOV', 'C1', 'CLTC']**

(Zero values filtered, sorted by improvement percentage)

	Product	Education_Type	Metric	Tested	CM	CM_Change
1	Data Analytics	Morning	AOV	24,538.10	10.67	
6	UX/UI Design	Evening	AOV	143,065.71	10.12	
2	Digital Marketing	Evening	AOV	373,884.40	10.04	
9	Web Developer	Morning	AOV	543,888.06	10.03	
7	UX/UI Design	Morning	AOV	1,078,742.57	10.02	
3	Digital Marketing	Morning	AOV	2,537,804.56	10.01	
1	Data Analytics	Morning	LTC	22,320.96	0.67	
1	Data Analytics	Morning	C1	22,307.36	0.61	
6	UX/UI Design	Evening	LTC	130,073.33	0.12	
6	UX/UI Design	Evening	C1	130,059.74	0.10	
2	Digital Marketing	Evening	LTC	339,908.50	0.04	
2	Digital Marketing	Evening	C1	339,894.91	0.04	
9	Web Developer	Morning	LTC	494,457.28	0.03	
9	Web Developer	Morning	C1	494,443.69	0.03	
7	UX/UI Design	Morning	LTC	980,688.66	0.02	
7	UX/UI Design	Morning	C1	980,675.06	0.01	
3	Digital Marketing	Morning	LTC	2,307,108.65	0.01	
3	Digital Marketing	Morning	C1	2,307,095.06	0.01	

#### Выход:

- **Наилучшим драйвером роста маржинальной прибыли (СМ) при улучшении одного показателя из 5 тестируемых является **увеличение АOV** - среднего чека;**
- **Вторым в рейтинге драйвером роста СМ является **уменьшение LTC** - затраты на юнит масштабирования, или стоимость привлечения потенциального клиента**
- **Третим в рейтинге драйвером роста СМ является **увеличение С1** - конверсия в первую сделку**

### 8.3. Дерево метрик для бизнеса

#### Дерево метрик

Ключевой показатель бизнеса			
<b>Маржинальная прибыль (CM)</b>			
$CM = UA \times (LTV - LTC) = UA \times (CLTV \times C1 - LTC) = B \times (CLTV - CLTC)$			
UA - поток привлекаемых потенциальных клиентов; C1 - конверсия в первую покупку; LTV - средняя валовая прибыль на потенциального клиента;		LTC - стоимость маркетинговых затрат на потенциального клиента; CLTV - валовая прибыль на клиента; CLTC - средние затраты на привлечение одного клиента.	
<b>Метрики принятия решений</b> (показатели юнит-экономики)			
Конверсия в первую покупку(C1): $C1 = B / UA$	Средний чек (AOV): $AOV = Gross Revenue / T$	Средняя стоимость лода в долгосрочной перспективе (LTC): Стоймость перехода на сайт $LTC = CLTC \times C_1$	Финансовые
Среднее число сделок приходящееся на одного клиента (APC): $APC = T / B$	Поток юнитов масштабирования (UA) $UA = COUNT(Id)$		Оборот $Revenue = B \times AOV \times APC$  Прибыль – разница между доходами и расходами.  $Net Profit = Revenue - Costs$  Costs - общие расходы
<b>Продуктовые метрики</b> (это метрики, с которыми работают продуктовые команды)			
Общее число клиентов (B): $B = UA * C1$	Сумма всех платежей, пришедших от клиентов (Gross Revenue) $Gross Revenue = SUM(Paid)$	Количество совершенных транзакций(T): $T = (Course duration   Months of study)$	Платеж, пришедший от клиента (Paid)  $Paid = (Initial Amount Paid   (Offer Total Amount / Course duration) * Months of study)$
Средняя стоимость привлечения одного клиента (CLTC): $CLTC = LTC * C1$	Средняя валовая прибыль на клиента (CLTV): $CLTV = AOV \times APC$ (для этой задачи)	Соотношение числа посетителей решивших оплатить курс (UA_wp) к числу посетителей дошедших до этапа обсуждения контракта (UA_ns):  $C_wp = UA_wp / UA_ns$  ns - negotiation stage wp - waiting for payment stage	
Соотношение числа посетителей оплативших курс (UA_paid)к числу посетителей решивших оплатить курс (UA_wp):  $C_paid = UA_wp / UA_paid$  wp - waiting for payment stage paid - paid stage		Соотношение числа посетителей дошедших до обсуждения контракта (UA_ns) к общему числу посетителей (UA):  $C_ns = UA_ns / UA$  ns - negotiation stage UA - total UA number	
<b>Атомные метрики</b> (это метрики, которые автоматически фиксируются при работе продукта)			
<b>Id (contacts)</b> - Число посетителей сайта (UA)	<b>Created Time(contacts)</b> - Дата внесения контакта в базу	<b>Created Time (deals)</b> - Дата создания сделки	<b>Closing Date (deals)</b> - Дата закрытия сделки
<b>Contact Name (deals)</b> - Идентификатор контактного лица по сделке.	<b>Product (deals)</b> - Конкретный продукт или услуга, связанная со сделкой	<b>Education Type (deals)</b> - Тип образования или обучения	<b>Months of study (deals)</b> - Количество месяцев которые отучился студент
<b>Course duration (deals)</b> - Длительность курса на который поступает студент	<b>Payment Type (deals)</b> - Тип используемого или ожидаемого способа оплаты	<b>Initial Amount Paid (deals)</b> - Первоначальный платеж клиента	<b>Offer Total Amount (deals)</b> - общая стоимость курса

### 8.4. Формулирование гипотез для метрик продукта

- "Бутылочное горлышко" в юнит-экономике это метрика, изменение которой на небольшую величину дает кратный рост маржинальной прибыли при наименьших затратах на изменение метрики.
- Согласно выводу из анализа результатов расчета по определению драйверов роста маржинальной прибыли - это метрики **AOV LTC** и **C1**.
- Установим, что затраты на улучшение метрики LTC выше чем AOV и C1 и исключим ее из дальнейшего анализа (в реальности данные по затратам на улучшение метрик определяются руководством на основе анализа затрат).
- В рамках проекта (в учебных целях) для формулировки и тестирования гипотез выберем два продукта: **лидера Digital Marketing (Morning)** и **аусайдера Data Analytics (Morning)** продаж. Тем более, что необходимость их улучшения прямо связана с выводами и рекомендациями, которые были сделаны по результатам проведенного анализа юнит-экономики образовательных продуктов:

  - Оптимизировать **Data Analytics (Morning)** → доходность (средний чек) и маркетинг (снизить затраты на привлечение, повысить конверсию). По степени затрат эти направления соизмеримы, но бутылочным горлышком здесь есть метрика

среднего чека AOV, которая вносит наибольший среди всех продуктов вклад (10.67%) в маржинальную прибыль. Для этого продукта увеличить AOV (1183,33€) можно, например за счет предложения при покупке курса дополнительных продуктов (продвинутые учебные материалы, книги, сертификаты на онлайн тренинги по углубленному курсу дата-анализа и .т.д.)

Формулировка гипотезы №1 (Продукт: Data Analytics (Morning) по S.M.A.R.T:

- **Предложения дополнительных продуктов при покупке курса (продвинутые учебные материалы, книги, сертификаты на онлайн тренинги по углубленному курсу дата-анализа и .т.д.) увеличит средний чек на 10%. Доверительный интервал AOV 5%:  $1301.66 \pm 32.54\text{€}$**
- 2. Сконцентрироваться на продажах **Digital Marketing (Morning)**, так как он лидер по доходности → стоит делать агрессивные рекламные кампании ( $C1=1.92\%$ ) Здесь протестируем гипотезу увеличения конверсии, хотя ее вклад в увеличение маржинальной прибыли (0.01%), ниже чем у AOV, однако затраты на ее увеличение ниже.

Формулировка гипотезы №2 (Продукт: Digital Marketing (Morning) по S.M.A.R.T:

- **Введение реферальной программы с моделью «5% на 1 месяц обучения за приглашенного клиента» увеличит конверсию на 10%. Доверительный интервал C1 5%:  $2.12 \pm 0.02\%$**

## 8.5. Описание метода проверки гипотез с формулированием условий проверки HADI-циклы для проверки гипотез

### HADI-циклы

для проверки гипотез направленных на улучшение среднего чека и конверсии образовательных продуктов онлайн школы программирования

Этап HADI	Продукт: Data Analytics (Morning)	Продукт: Digital Marketing (Morning)
Hypothesis	Предложения дополнительных продуктов при покупке курса (продвинутые учебные материалы, книги, сертификаты на онлайн тренинги по углубленному курсу дата-анализа и .т.д.) увеличит средний чек на 10%. Доверительный интервал AOV 5%: $1301.66 \pm 32.54\text{€}$	<b>Введение реферальной программы</b> с моделью «5% скидки на 1 месяц обучения за приглашенного клиента» увеличит конверсию на 10%. Доверительный интервал C1 5%: $2.12 \pm 0.05\%$
Null Hypothesis	Предложения дополнительных продуктов при покупке курса не влияет на средний чек.	<b>Введение реферальной программы</b> с моделью «5% на 1 месяц обучения за приглашенного клиента» не влияет на конверсию
Action	Разработать дополнительные продукты. Случайно разделить пользователей на контрольную и тестовую группы. Контрольная группа получает стандартное предложение. Тестовая группа получает предложение с дополнительными продуктами.	Разработать реферальную программу. Случайно разделить существующих клиентов на контрольную и тестовую группы. В обеих группах запускается реферальная программа с уведомлением по e-mail, при этом в контрольной группе предложение 5% скидки на обучение не делается. Тестовая группа получает предложение с указанием условий реферальной программы с 5% скидкой.
Data	Фиксировать данные по стоимости заказов для обеих групп на протяжении 14 дней	Фиксировать продаж клиентам, пришедшим из обеих групп Провести сравнение по истечению 14 дней.
Insight	Если средний чек увеличится на 10% в пределах доверительного интервала или больше, распространить стратегию продаж на весь продукт.	Если конверсия увеличится на 10% в пределах доверительного интервала или больше, продолжить и масштабировать реферальную программу.

S.M.A.R.T	Цикл 1	Цикл 2
Specific	Отделу обучения разработать продвинутые учебные материалы, книги, сертификаты на онлайн тренинги по углубленному курсу дата-анализа	Отделу маркетинга разработать реферальную программу с моделью «5% на 1 месяц обучения за приглашенного клиента»
Measurable	Число клиентов, купивших дополнительный продукт при покупке курса	Количество клиентов по реферальной программе с 5% скидкой на обучение.
Achievable	Цель достижима при предложении курса в пакете с дополнительным продуктом.	Цель достижима при правильной настройке и продвижении программы.
Relevant	Точка роста для бизнеса это средний чек (AOV).	Точка роста для бизнеса это конверсия в первую покупку (C1) Реферальная программа имеет хороший потенциал для этого.
Time-bound	Тестирование и анализ результатов в течение 14 дней.	Тестирование и анализ результатов в течение 14 дней.

### Расчет доверительных интервалов и размера выборки для проведения А|В тестирования

При расчетах примем следующие допущения:

- Вероятность отклонения основной (или нулевой) гипотезы при проверке статистических гипотез в случае, когда конкурирующая (или альтернативная) гипотеза верна (confidence level) - 0.95
- Доверительный интервал - 5%
- Минимально обнаруживаемый эффект - 1%

Для этих допущений **итоговое число экспериментов** в каждой группе рассчитаем по формуле:

$$n = \frac{16 \cdot p \cdot (1-p)}{x^2},$$

где:

- $n$  - итоговое число экспериментов в каждой группе;
- $p$  - базовая конверсия;
- $x$  - минимально обнаруживаемый эффект в %.

Результат:

#### Calculation for the first hypothesis:

Offering additional products when purchasing a course (advanced study materials, books, certificates for online trainings on an advanced data analysis course, etc.) will increase the Average Order Value by 10%. Confidence interval 5%: 1301.66 ± 32.54€  
 Base conversion rate p=0.02%  
 Minimum detectable effect xi=2.00%  
 Base Average Order Value AOV=1183.33%  
 Required sample size for A|B testing: 7 participants  
 Total number of participants required for the experiment: 14  
 New AOV with confidence interval,€: 1301.66 ± 32.54€

#### Calculation for the second hypothesis:

Introducing a referral program with the model "5% off 1 month of training for a referred client" will increase conversion by 10%. Confidence interval 5%: 2.12 ± 0.05%  
 Base conversion rate p=1.92%  
 Minimum detectable effect xi=2.00%  
 Required sample size for A|B testing: 753 participants  
 Total number of participants required for the experiment: 1506 participants  
 New Conversion Rate with confidence interval, %: 2.11 ± 0.05%

## Формулирование исходные данных для экспериментов

Результаты расчетов позволяют сформулировать исходные данные для эксперимента:

### Гипотеза №1

- Общее число участников тестирования - **14**
- Среднее значение чек **AOV** ожидается: **1301.66 ± 32.54€**
- Срок проведения тестирования **14 дней**.

### Гипотеза №2

- Общее число участников тестирования - **1506**
- Среднее значение коэффициента конверсии **C1** ожидается: **2.12 ± 0.05%**
- Срок проведения тестирования **14 дней**.

## Проверим срок проведения тестирования

По условию эксперимента мы планируем тест на 14 дней (2 недели).

- для гипотезы №1 для эксперимента нам необходимо **14 участник**
- для гипотезы №2 для эксперимента нам необходимо **1506 участников**

Сделаем расчет среднего числа лидов для этих интервалов и построим графики по количеству сделок (UA) с течением времени  
 При этом сгруппируем сделки по временным периодам:

- **2 недели**
- **месяц**

### Результат расчета:

Average Deals number per **two weeks** period: **415**

Average Deals number per **month** period: **1799**

### Deals Trends (Selectable Periods)



### Deals Trends (Selectable Periods)



Из расчетов и графика следует:

- среднее число сделок в месяц - 1799, в двухнедельном периоде - 415;
- для проверки Гипотезы №1 - участников достаточно;
- для проверки Гипотезы №2 - участников сделки как в двухнедельном периоде не достаточно, однако если взять срок эксперимента 1 месяц, то участников эксперимента будет достаточно. Вывод:
- Проведение эксперимента по проверке Гипотезы №1 в выбранный срок 2 недели - ВОЗМОЖНО;
- Проведение эксперимента по проверке Гипотезы №2 в срок 2 недели - НЕ ВОЗМОЖНО из-за недостаточной численности участников, поэтому срок эксперимента увеличим и установим один месяц

### Моделирование проведения А|В теста

#### Гипотеза №1 - проверка по t-критерию Стьюдента

Результат расчета:

T-statistic: -6.7227

P-value: 0.0000

Hypothesis NOT rejected: AOV increase by 10% is statistically significant.

Интерпретация результата t-теста Стьюдента для гипотезы №1:

- Т-статистика: -6.7227 — показывает степень различий между двумя выборками.
- P-value: 0.0000 — вероятность того, что наблюдаемые различия могли возникнуть случайно, если нулевая гипотеза верна.
- Уровень значимости (alpha): 0.05 — заданный порог для принятия решения.

#### Вывод

- p-value = 0.0000 < 0.05, что означает, что мы отвергаем нулевую гипотезу.
- Гипотеза НЕ отвергнута, значит, предложение дополнительных продуктов действительно статистически значимо увеличивает средний чек на 10%.
- Внедрение дополнительных предложений при покупке курса может существенно увеличить AOV.

#### Рекомендация:

- Разработать стратегию upsell (дополнительных продуктов), чтобы максимизировать доход от каждой покупки
- Полезно оценить долгосрочное влияние на LTV (Lifetime Value), чтобы понять, насколько устойчив этот рост AOV.

### Гипотеза №2 - проверка по t-критерию Стьюдента

#### Результат расчета:

T-statistic: -79.4407

P-value: 0.0000

Hypothesis NOT rejected: C1 increase by 10% is statistically significant.

Интерпретация результата t-теста Стьюдента для гипотезы №2:

- **Основные показатели**
  - Т-статистика: -79.4407 (очень высокая абсолютная величина)
  - Р-значение: 0.0000 (немного ниже уровня значимости  $\alpha = 0.05$ )
  - Уровень значимости ( $\alpha$ ): 0.05
- **Выводы по тесту:**
  - P-value (0.0000) <  $\alpha$  (0.05), значит, мы НЕ отвергаем гипотезу о том, что внедрение реферальной программы увеличивает конверсию на 10%.
  - Высокая абсолютная величина t-статистики (-79.4407) подтверждает, что разница между контрольной и тестовой группами существенная и не случайная.
  - Следовательно, реферальная программа действительно оказывает положительное влияние на конверсию (C1).
- **Вывод для бизнеса**
  - Реферальная программа работает и статистически значимо увеличивает C1 на 10%.
  - Можно масштабировать программу, предлагая 5% скидку за приглашенного клиента на более широкую аудиторию.
  - Высокая статистическая значимость подтверждает, что этот эффект устойчив, а не случайный.
  - Следующий шаг: протестировать различные размеры скидки (например, 3%, 7%) и проанализировать их влияние.
- **Рекомендации:**
  - Углубить анализ на долгосрочное влияние на LTV (Lifetime Value).
  - Разработать автоматизированный процесс реферальных вознаграждений для удобства клиентов.
  - Провести A/B тестирование с разными скидками и определить оптимальный размер бонуса.

## 9. Общие выводы и рекомендации по проекту

### 9.1 SWOT-анализ стратегии бизнеса

Strengths	Weaknesses	Opportunities	Threats
Успешные рекламные каналы Facebook Ads, Google Ads и Webinar, высокая конверсия	Высокие затраты на некоторые рекламные кампании (Performancemax_Eng_DE – 34 K€, но конверсия 0.41%)	Расширение успешных рекламных каналов → усилить SEO и вебинары	Высокая конкуренция на рынке онлайн-образования
Реферальная программа доказала эффективность, рост конверсии на 10%	Низкая успешность Data Analytics (Morning) (22,17 K€ прибыли, конверсия 0.02%)	Географическая экспансия → усилить маркетинг в восточной Германии, Австрии, Польше, Франции	Экономическая нестабильность, может повлиять на платежеспособность клиентов
Юнит-экономика Digital Marketing (Morning) → 2,3 млн€ маржинальной прибыли	Нерегулярность данных в CRM (unknown значения в Product, Payment Type, Education Type.)	Масштабирование реферальной программы → проверено, увеличивает конверсию	Аномальные долгосрочные сделки, могут указывать на проблемы в процессах продаж
Высокая географическая диверсификация, сделки не только в Германии, но и в Австрии, Польше, Франции	Долгие сделки (150+ дней) → снижают скорость оборота капитала	Оптимизация работы менеджеров → обучение и персонализация работы с клиентами	Низкое качество лидов по некоторым рекламным кампаниям, необходимо тестирование и анализ ROI

### 9.2 Итоговые выводы и рекомендации для бизнеса

Фокус на масштабирование успешных направлений	Устранение слабых мест	Оптимизация юнит-экономики	Географическое развитие	Улучшение отдела продаж
<ul style="list-style-type: none"> <li>• Digital Marketing (Morning);</li> <li>• вебинары;</li> <li>• оптимизация веб-сайта (Search Engine Optimization)</li> </ul>	<ul style="list-style-type: none"> <li>• CRM;</li> <li>• Data Analytics;</li> <li>• нерентабельные рекламные кампании.</li> </ul>	<ul style="list-style-type: none"> <li>• повышение AOV;</li> <li>• увеличение конверсии (C1);</li> <li>• снижение затрат на привлечение клиентов (LTV).</li> </ul>	<ul style="list-style-type: none"> <li>усилить маркетинг в:</li> <li>• Австрии;</li> <li>• Польше;</li> <li>• восточной Германии.</li> </ul>	<ul style="list-style-type: none"> <li>• обучение менеджеров;</li> <li>• персонализация работы с клиентами.</li> </ul>