# **Исследование данных рекламы по каналам коммуникации он-лайн кинотеатра**

Оглавление

[Исследование данных рекламы по каналам коммуникации он-лайн кинотеатра 1](#_Toc154769043)

[1. Ситуация. Исходные данные 1](#_Toc154769044)

[2. Проблема - задача и алгоритм выполнения 2](#_Toc154769045)

[Общая задача: 2](#_Toc154769046)

[Алгоритм выполнения: 2](#_Toc154769047)

[3. Подготовка данных для аналитики 2](#_Toc154769048)

[3.1. Объединение данных, формирование витрины 2](#_Toc154769049)

[3.2. Проверка на полноту и аномальность 2](#_Toc154769050)

[3.3. Проверка на дублирование 3](#_Toc154769051)

[3.4. Оценка фальсификации данных 3](#_Toc154769052)

[4. Исследование корреляции признаков, факторный анализ 3](#_Toc154769053)

[4.1. Исследование корреляций 3](#_Toc154769054)

[4.1.1. Перекодировка значений, матрица корреляций 3](#_Toc154769055)

[4.1.2. Исследование корреляции цены 3](#_Toc154769056)

[4.1.3. Исследование парных корреляций, тест Бартлета 3](#_Toc154769057)

[4.2. Факторный анализ 4](#_Toc154769058)

[4.2.1. Анализ 4](#_Toc154769059)

[4.2.2. Выводы по итогам факторного анализа 4](#_Toc154769060)

[5. Исследование фродовых данных 4](#_Toc154769061)

[6. Исследование воронки продаж по каналам коммуникации 5](#_Toc154769062)

[7. Исследование затрат и прибыли по рекламным каналам 5](#_Toc154769063)

[Графики затрат и прибыли по рекламным каналам на последнем этапе воронки продаж 5](#_Toc154769064)

[8. Исследование операционных систем пользователей по каналам коммуникации 6](#_Toc154769065)

[9. Заключение. Бизнес-рекомендации 6](#_Toc154769066)

[9.1. В части пользователей - 6](#_Toc154769067)

[9.2. В части каналов коммуникации 7](#_Toc154769068)

[9.3. В части работы по этапам воронки продаж 7](#_Toc154769069)

# **Ситуация. Исходные данные**

Обратился известный онлайн-кинотеатр, предоставляющий к просмотру как фильмы и сериалы, так и мультфильмы с целью получения оценки и бизнес-рекомендаций по итогам исследования каналов коммуникаций.

Онлайн-кинотеатр отслеживает различные каналы коммуникации —

* banner — баннерная реклама на сайтах;
* video — рекламные видеоролики;
* native — нативные баннеры;
* partner\_network — реклама партнёрской сети, размещается на агрегаторах контента;
* rich — брендирование страницы, большой баннер на весь экран;
* search — платный поиск;
* organic\_traffic — переходы по прямым ссылкам или переходы с бесплатной поисковой выдачи на сайт.

Данные о реакции людей на различные рекламные каналы собраны на разных этапах воронки продаж. Воронка состоит из пяти последовательных этапов: осведомлённость о бренде (awareness (факт показа рекламы), в данных отсутствует) → интерес (interest - клик по рекламному баннеру и переход на сайт) → рассмотрение бренда к покупке (consideration - выбор планов подписки, хождение по продуктовым страницам) → желание купить (intent - выбор плана, заявка на обратную связь, регистрация) → покупка (purchase).

Данные поступили в двух видах:

*Структура таблиц с данными различных каналов (9 файлов):*

* user\_id: зашифрованный идентификатор пользователя;
* funnel\_stage: этап воронки;
* timestamp: время совершения действия юзером;
* profit: прибыль на этапе покупки, в остальных случаях маркируется как None;
* price: затраты на рекламу;
* os: операционная система, которой пользуется человек.

*Структура таблицы с ценами на рекламу (1 файл с ценами рекламных каналов)* *:*

* ad\_channel: название рекламного канала;
* price: стоимость рекламы.

# **Проблема - задача и алгоритм выполнения**

## **Общая задача:**

В течение месяца клиент провёл серию рекламных кампаний и теперь хочет оценить, какой вклад в итоговый трафик они внесли.

Также нужно выявить проблемные этапы воронки продажи, определить причины и дать рекомендации по изменению подходов к продажам.

## **Алгоритм выполнения:**

1. Получить данных из разных источников, сформировать единую витрину данных, провести разведочный анализ
2. Определить степень сфальсифицированности данных и принять решение об их использовании
3. Исследовать –

3.1. корреляцию признаков, факторный анализ, фродовые данные

3.2. воронку продаж по каналам коммуникации

3.3. затраты и прибыль по рекламным каналам

3.4. операционные системы пользователей по каналам коммуникации

1. Сделать общие выводы и дать бизнес-рекомендации по итогам исследования

# **Подготовка данных для аналитики**

## **Объединение данных, формирование витрины**

## **Проверка на полноту и аномальность**

Изображение выглядит как снимок экрана, Прямоугольник, дизайн

Автоматически созданное описаниеСформирована общая витрина данных, в ней 3 845 707 строк и 7 столбцов.

Изображение выглядит как текст, Шрифт, снимок экрана, белый

Автоматически созданное описаниеИзображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описаниеВ целом данные имеют хорошую структуру, аномальности не выявлены, незаполнены только строки в столбце «profit».

Были сделаны следующие действия:

* замена значений в "funnel\_stage" с прописных на строчные, для унификации значений,
* замена пустых значений в "profit" на нулевые (по смыслу),
* исключение из выборки строк с пустым значением столбца "price" – одновременно был исключен из данных канал «Другое», который не интересовал Заказчика.

## **Проверка на дублирование**

Установлено, что количество полностью дублирующих строк в базе – 2 831 380, т.е. 74,2% от всего числа данных, удаляем их все.

В результате в итоговой таблице остается 991 739 строк (25,8% от первоначального количества данных)

## **Оценка фальсификации данных**

Воронки продаж по каналам коммуникации существенно отличаются между собой, что вызывает вопросы по возможной фальсификации данных. Такое же мнение у Заказчика.

При общем количестве пользователей – 529, 359 тыс., кол-во пользователей, совершавших более 1 действия в один момент времени – 5,175 тыс. (0,98%), это явно фродовые данные, которые требуют исключения из данных для исследования. Итоговая фродовая таблица – это дублирующие данные и фроды по одновременности действий, всего 2 836 555 строк или 73,8% от числа первоначально представленных данных.

# **Исследование корреляции признаков, факторный анализ**

## **Исследование корреляций**

### **Перекодировка значений, матрица корреляций**

Изображение выглядит как текст, Шрифт, рукописный текст, снимок экрана

Автоматически созданное описаниеПри подготовке данных для факторного анализа были перекодированы категориальные признаки –

- данные операционной системы – в бинарный вид

- данные этапов воронки – в нумерованный вид

Изображение выглядит как текст, снимок экрана, прямоугольный, линия

Автоматически созданное описаниеДалее данные были факторизованы и построена тепловая карта корреляционной матрицы -

### **Исследование корреляции цены**

По тепловой карте видна высокая корреляция между ценой и каналом коммуникации, а также идентификатором пользователя.

Изображение выглядит как текст, Шрифт, белый, типография

Автоматически созданное описаниеСчитаем между ценой и id коэффициент Пирсона и значение р-value ---

Несмотря на то, что коэффициент корреляции очень высокий - 0,85, его достоверность = 0, т.е. судить о том, что цена канала зависит от идентификатора пользователя некорректно.

### **Исследование парных корреляций, тест Бартлета**

Методом LabelEncoder() был прокодированы значения рекламных каналов.

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описаниеПо итогам расчета парных корреляций функцией corrwith с признаком «profit» видно, что корреляция факторов достаточно низкая.

Изображение выглядит как текст, Шрифт, снимок экрана, белый

Автоматически созданное описаниеТест Бартлета показывает возможность выполнения факторного анализа, p\_value = 0. Нулевая гипотеза об отсутствии корреляции между переменными отвергается (гипотеза о диагональности при стандартном уровне значимости 0,95), что означает что в выборке существуют значимые корреляции по крайней мере между некоторыми из переменных

## **Факторный анализ**

### **Анализ**

Изображение выглядит как линия, диаграмма, График, скат

Автоматически созданное описаниеСтандартизируем переменные с помощью класса StandardScaler.

Проводим построение, отбор и вращение факторов, на основе класса FactorAnalyzer (модуль factor\_analyzer). Как видно из графика 2 фактора из 5 имеют значение >1, и являются значимыми.

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описаниеПо итогам изучения факторной матрицы видим, что фактор 1 выделяет значительное влияние цены рекламного канала и связанного с ним фактора самого рекламного канала.

Фактор 2 – влияние id пользователя не является смысловым и не рассматриваем далее. Преобразовали изначальную матрицу с помощью метода transform() класса FactorAnalyzer. Получили итоговое представление исходных данных в виде факторов.

### **Выводы по итогам факторного анализа**

Матрица корреляций выделяет основные коррелирующие переменные с результатом - это **рекламный канал и его цена**. При этом корреляция больше 0, т.е. чем выше стоимость рекламного канала, тем выше итоговый результат

Значимая корреляция между ценой канала и идентификатором пользователя не подтверждается и использовать ее некорректно.

Остальные переменные (влияние операционной системы, времени и этапа воронки) имеют крайне низкую корреляцию между собой по модулю.

# **Исследование фродовых данных**

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описаниеИзображение выглядит как текст, снимок экрана, дисплей, число

Автоматически созданное описаниеФродовые значения установлены только на этапах interest и intent, т.е. на начальных этапах взаимодействия пользователя с сайтом.

Наибольшее количество - почти 92% на этапе интереса. Причиной такого количества фродовых значений на этих этапах может быть обусловлено -

1. "Накруткой" откликов со стороны рекламных каналов
2. Сбоем в обработки данных со стороны Интернет сетей

- Следует отметить, что фродовые значения есть по всем каналам коммуникация примерно в одинаковом соотношении – около 0,8% от общего количества данных, в т.ч. и по бесплатному каналу, что свидетельствует о скорее всего техническом сбое

- Чуть выше процент фродовых значений по каналам «rich» и «video» и на них следует обратить особое внимание

Это логично, т.к. рекламные каналы получают свои деньги за количество откликов, т.е. прямой интерес в максимизации данных цифр. И если компания не проверяет их, то заплатит за их услуги значительно больше, чем должна.

В приведенном примере, при отсутствии простейших проверок датасета (в т.ч на дублирование), компания заплатила бы **в три раза больше**, чем необходимо. При этом следует отметить, что **фродовые данные есть по всем каналам** в существенных объемах, что может свидетельствовать о каком-либо действительно сбое обработки данных, либо о том, что это стандартная процедура увеличения своих доходов со стороны любых рекламных каналом.

# Изображение выглядит как текст, снимок экрана, диаграмма, число Автоматически созданное описание**Исследование воронки продаж по каналам коммуникации**

Как видно из тепловой карты, наименее перспективным и ликвидным является канал "video". От него надо отказываться либо разбираться с причинами низкой эффективности более подробно.

Есть также вопросы и к каналам “**native” и ‘rich”** – достаточно низкая доля привлечения и даже при этом очень низкая конверсия

Общие выводы:

* можно отказаться от канала video – очень низкая конверсия в результат
* следует уделить внимание каналам баннер, organic\_traffic и partner network с точки зрения определения путей повышения результативности отклика по ним

Необходимо понаблюдать за каналами native и rich в части оценки целесообразности отказа от них в связи с низкой конверсией в отклик и в покупку

# **Исследование затрат и прибыли по рекламным каналам**

## **Графики затрат и прибыли по рекламным каналам на последнем этапе воронки продаж**

Изображение выглядит как текст, снимок экрана, График, линия

Автоматически созданное описаниеИз соотношения графиков явно следует, что

- убыточными каналами являются каналы video, native и rich

- остальные каналы являются прибыльными для компании.

Самые выгодные - каналы organic\_traffic (нулевые издержки) и search. На них и следует обращать внимание при разработке рекламной стратегии

В 3,5 раза превысил доход расходы на рекламу в целом , но этого недостаточно

Самый высокодоходный канал является бесплатным – это organic\_traffic

Далее за счет низкой цены рекламы высокую доходность показывает канал banner

Partner\_network и search также имеют доходы существенно выше, чем расходы, но следует работать с этими каналами – как в части повышения конверсии в покупку, так и в части снижения затрат на рекламу

Каналы video, native и rich являются убыточными, а учитывая в целом отрицательные показатели по всему исследованию, есть смысл рассмотреть необходимость отказа от них, в особенности от канала video с самыми существенными убытками

# **Исследование операционных систем пользователей по каналам коммуникации**

Распределение использования операционных систем по рекламным каналам дает характеристику и тип пользователей на текущий момент. Следует учитывать общие тенденции и ее перспективу изменения соотношения использования iOS и Android (предполагаемое изменение в России в сторону Android).

Но на момент проведения анализа -

* преобладающее использование Android пользователям по каналам video, banner, organic\_traffic
* примерно равное соотношение iOS и Android на канале partner\_network
* преобладание использования iOS на каналах search, native и rich.

Предположительно исторически пользователи с iOS - более денежный класс пользователей, вип-клиенты и т.п. пользователи с Android - пользователи с меньшим достатком. Но в настоящее время ситуация значительно меняется и подобных жестких определений уже проводить нельзя.

Проводились даже исследования особенностей людей, использующих разные ОС, можно их учитывать при оценке поведенческих действий людей при планировании рекламных кампаний -

Смартфон может многое сказать о характере владельца, если программно анализировать поведение пользователя и "обучить" приложение. Оказывается, у пользователей iPhone и Android разный характер – об этом говорят психологи. Даже самый надежный смартфон требует ремонта или профилактики. Оказывается, пользователи Android воспринимаются окружающими как более честные, и уступчивые люди, нежели владельцы iPhone. В свою очередь, любители iPhone воспринимаются окружающими как общительные и более открытые.

Приверженцы операционной системы Android более индивидуальны и не любят быть похожими на всех – это проявляется даже в выборе одежды, которая приобретается необычного фасона, отличная от того, что носит большинство людей.

Владельцев смартфонов iPhone отличает высокомерность и заносчивость – многие покупатели iPhone признались, что приобретение данной модели было выбрано по принципу его высокого статуса.

Известно, что некоторые владельцы iPhone выражают свое социальное положение в обществе, демонстрируя свой новый телефон.

Считается, что на характер человека оказывает влияние любая вещь, которой он пользуется. Так как телефон для многих людей является частью жизни, то по его модели, манере человека говорить, чистоте и техническому состоянию вполне можно судить о характере.

*Источник:*[*https://wek.ru/uchenye-issledovali-xarakter-polzovatelej-iphone-i-android*](https://wek.ru/uchenye-issledovali-xarakter-polzovatelej-iphone-i-android)

# **Заключение. Бизнес-рекомендации**

## **В части пользователей -**

1. Основное количество покупок (30%) имеет средний чек 100 -500 руб, но основной доход (54%) приносят дорогие покупки со средним чеком свыше 2000 рублей. Для повышения дохода необходимо – ,

— направлять рекламные кампании на покупки по дорогим тарифам (бо′льшие скидки на пакетные услуги, скидки постоянным, VIP клиентам и т.п.)

— направлять адресные кампании на пользователей с меньшим чеком для перехода их в более доходную, дорогую категорию

1. Можно учитывать распределение ОС при планировании акций по приоритетным рекламным каналам. Следует иметь в виду изменение в РФ соотношения использования ОС в сторону Android, а также характеристики любителей данных ОС

## **В части каналов коммуникации**

1. Входящие данные по каналам коммуникации требуют отдельного контроля — их следует проверять на полноту и достоверность, в т.ч. на фальсификацию, как техническую, так и сознательную, в ином случае

— могут существенно необоснованно вырасти издержки на рекламу

— могут быть сделаны неверные выводы по итогам проведенных и планируемых рекламных мероприятий

1. Основным влияющим на доход фактором является рекламный канал (связанный с ним признак цены), остальные переменные (ОС и пр.) не являются влиятельными
2. Необходимо выбирать каналы коммуникации исходя из их доходности, и на эти каналы, по возможности, делать основную ставку при планировании рекламных кампаний:
   1. organic\_traffic – он самый высокодоходный и бесплатный, но составляет всего 13% от всего трафика, необходимо повышать трафик и результативность отклика по нему
   2. banner - второй по доходности канал, высокая доходность за счет низкой цены рекламы, составляет всего 16%, необходимо повышать трафик
   3. рartner\_network – третий по доходности канал, составляет всего 11%, необходимо повышать трафик
   4. search – канал с самой высокой конверсией в покупку – 1,4% (> средней) и с самым большим использованием – 38%, это доходный канал, но для повышения общей прибыли есть необходимость работать как в части повышения конверсии в покупку, так и в части снижения
3. Каналы video, native и rich являются убыточными и, учитывая в целом отрицательные показатели по всему исследованию, есть смысл рассмотреть отказ от них, в особенности от канала video с самыми существенными убытками, самой низкой конверсией в результат и фальсификацией данных

## **В части работы по этапам воронки продаж**

Воронки продаж по доходным каналам:

1. ****Низкая средняя конверсия в покупку – 0,8%. Необходимо работать в данном направлении по всем каналам, включая самый эффективный канал search, по которому текущая конверсия в покупку 1,4%

Низкая средняя конверсия в покупку – 0,8%. Необходимо работать в данном направлении по всем каналам, включая самый эффективный канал search, по которому текущая конверсия в покупку 1,4%

Следует принять меры на этапе покупки, так как именно на нем происходит основное снижение покупателей (т.е. в корзину складывают, но не покупают)

1. Следует принять меры на этапе покупки, так как именно на нем происходит основное снижение покупателей (т.е. в корзину складывают, но не покупают)