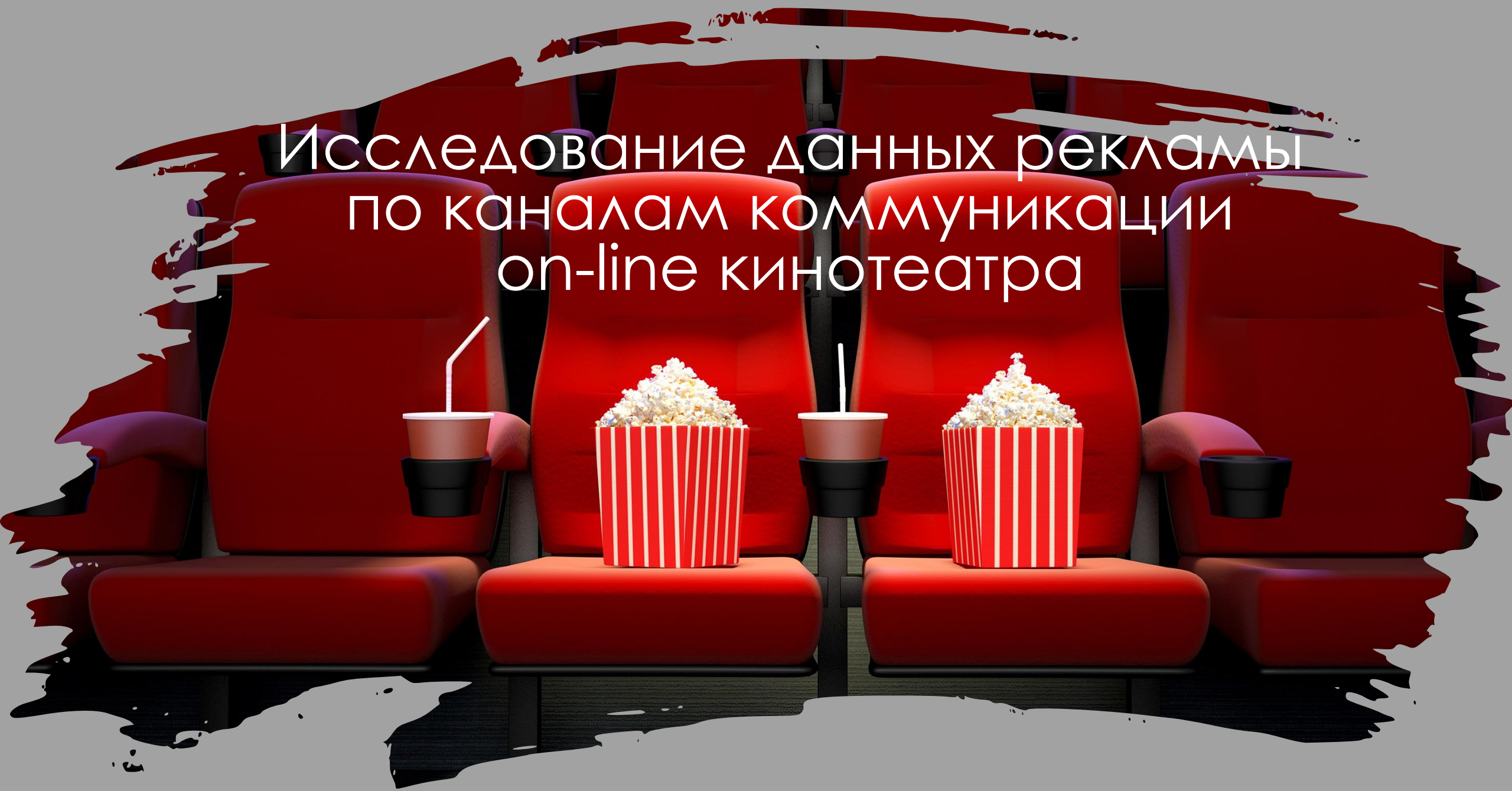


Исследование данных рекламы по каналам коммуникации on-line кинотеатра



Проблема: Исследование данных о каналах коммуникации on-line кинотеатра для использования при проведении рекламных кампаний

Цель:

- оценить какой вклад в итоговый трафик внесли рекламные кампании по разным каналам коммуникации
- выявить проблемные этапы воронки продаж, определить причины и дать рекомендации по изменению подходов

Исходные данные:

Таблицы 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

Таблица с ценами на рекламу:

- ad_channel: название рекламного канала;
- price: стоимость рекламы

Способ реализации:

- ✓ Получить данных из разных источников, сформировать единую витрину данных, провести разведочный анализ
- ✓ Определить степень сфальсифицированности данных и принять решение об их использовании
- ✓ Исследовать –
 - корреляцию признаков, факторный анализ, фродовые данные
 - воронку продаж по каналам коммуникации
 - затраты и прибыль по рекламным каналам
 - операционные системы пользователей по каналам коммуникации
- ✓ Сделать общие выводы и дать бизнес-рекомендации по итогам исследования

Исследование исходных данных, подготовка для аналитики

1.1. Объединение данных, Data cleaning датасета

- Сформирована общая витрина данных — 3 845 707 строк и 7 столбцов
- В целом данные имеют хорошую структуру, аномальности не выявлены, незаполнены только строки в столбце «profit»

3 845 707 строк
7 столбцов

Были сделаны следующие действия:

- замена значений в "funnel_stage" с прописных на строчные, для унификации,
- 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
- удалены дублирования из таблицы (исключено 74,2% данных)

991 739 строк
7 столбцов

| | user_id | funnel_stage | timestamp | profit | os | ad_channel | price |
|---|---------|-----------------|-----------|--------|-----|------------|-------|
| 0 | 387356 | 1_interest | 150000 | 0.0 | iOS | search | 1.5 |
| 1 | 619571 | 2_consideration | 150000 | 0.0 | iOS | search | 1.5 |
| 2 | 230119 | 1_interest | 150000 | 0.0 | iOS | search | 1.5 |
| 3 | 399556 | 1_interest | 150000 | 0.0 | iOS | search | 1.5 |

1.2. Исследование данных - Каналы коммуникаций ("ad_channel") —

| | interest - интерес | | consideration - движение к покупке | | intent - желание купить | | purchase - покупка | | |
|---|-----------------------|----------------|---------------------------------------|-----------------|----------------------------|-----------------|-----------------------|-----------------|-----------------|
| | кол-во | % от общего | кол-во | % от 1 этапа | кол-во | % от 2 этапа | кол-во | % от 3 этапа | % от 1 этапа |
| Итого | 597 381 | | 338 293 | 56,6% | 51 231 | 15,1% | 4 834 | 9,4% | 0,8% |
| search — платный поиск | 228 197 | 67,5% | 134 305 | 58,9% | 27 955 | 20,8% | 3 090 | 11,1% | 1,4% |
| banner — баннерная реклама на сайтах | 93 443 | 27,6% | 47 541 | 50,9% | 7 743 | 16,3% | 688 | 8,9% | 0,7% |
| video — рекламные видеоролики | 99 766 | 29,5% | 60 880 | 61,0% | 60 | 0,1% | 5 | 8,3% | 0,0% |
| organic_traffic — бесплатные переходы на сайт | 79 990 | 23,6% | 53 592 | 67,0% | 9 299 | 17,4% | 542 | 5,8% | 0,7% |
| partner_network — реклама партнёрской сети | 67 212 | 19,9% | 24 721 | 36,8% | 4 504 | 18,2% | 450 | 10,0% | 0,7% |
| native — нативные баннеры | 19 685 | 5,8% | 12 597 | 64,0% | 418 | 3,3% | 37 | 8,9% | 0,2% |
| rich — брендирование страницы, большой баннер | 9 088 | 2,7% | 4 657 | 51,2% | 1 252 | 26,9% | 22 | 1,8% | 0,2% |

986 564 строк
7 столбцов

- после исключения фродовых данных из датасета

2 836 555 строк

- итоговая таблица исключенных данных – 73,8% от числа первоначально количества (дублирующие данные и фроды по одновременности действий)

- Как показывает анализ воронки продаж по каналам коммуникации есть вопросы к возможной фальсификации данных
- Была проведена проверка выполнения >1 действия в один момент времени пользователем. Выявлены данные, которые следует считать фродовыми

5,2 тыс. (0,98%) • фродовых данных

Исследование исходных данных, подготовка для аналитики

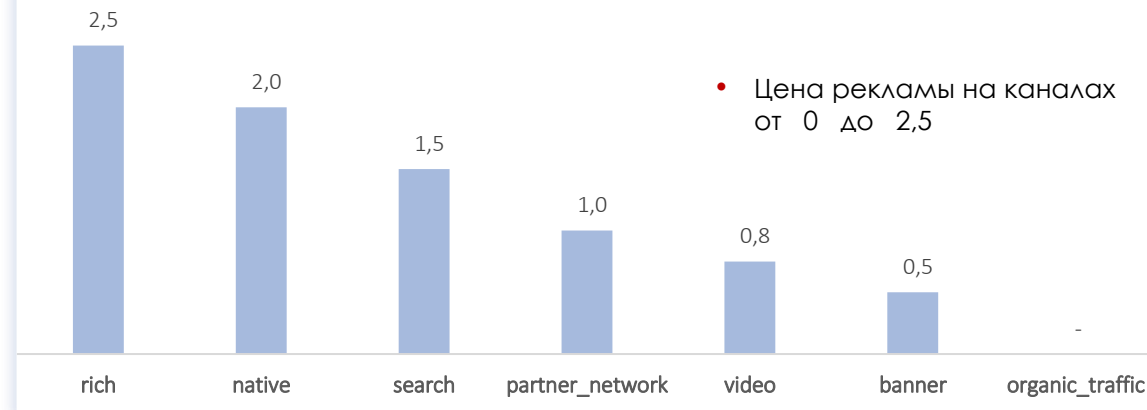
986 564 строк
и 7 столбцов

- итоговая витрина данных

529 359

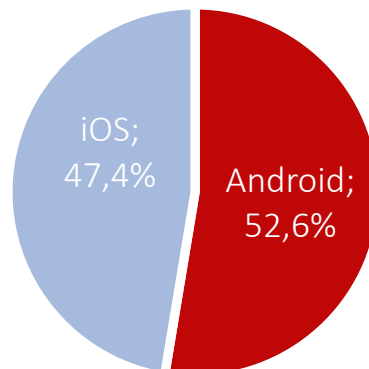
- пользователей (user_id)

Стоимость канала коммуникаций (price)



Операционные системы пользователей (os)

- Примерно равное соотношение использования ОС

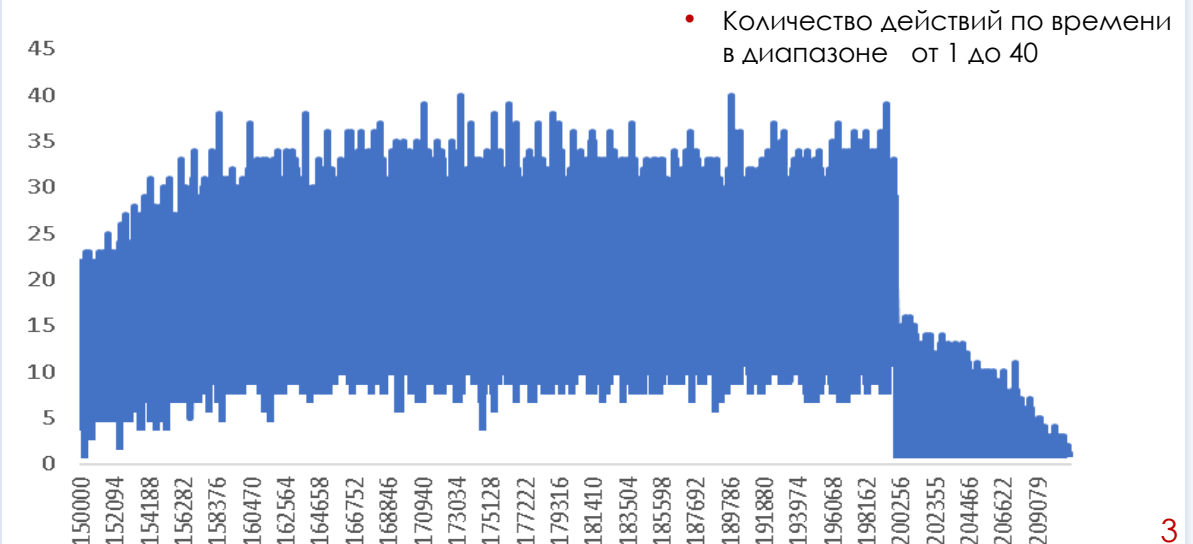


Доход (profit)

| Диапазон суммы покупок | Кол-во покупок | Сумма покупок | % суммы |
|------------------------|----------------|---------------|---------|
| Общий итог | 4 834 | 3 498 027 | |
| <=100 | 647 | 32 350 | 0,9% |
| 100-500 | 1 400 | 271 820 | 7,8% |
| 500-1000 | 843 | 404 750 | 11,6% |
| 1000-2000 | 925 | 914 947 | 26,2% |
| 2000-3000 | 879 | 1 648 608 | 47,1% |
| >3000 | 140 | 225 552 | 6,4% |

- Несмотря на то, что основное количество покупок (30%) имеет средний чек от 100 до 500 руб,
- Основной доход – 54% приносят дорогие покупки со средним чеком свыше 2000 рублей

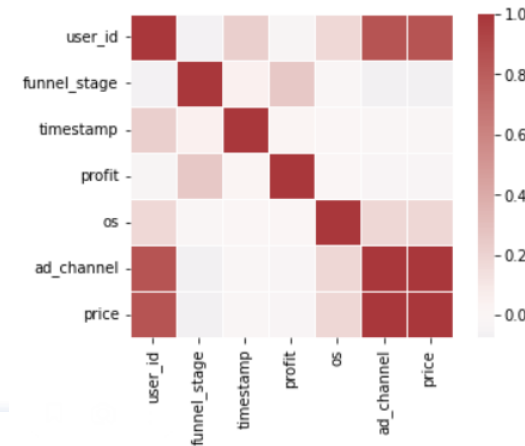
Время действия (timestamp)



2 Исследование корреляции признаков, факторный анализ

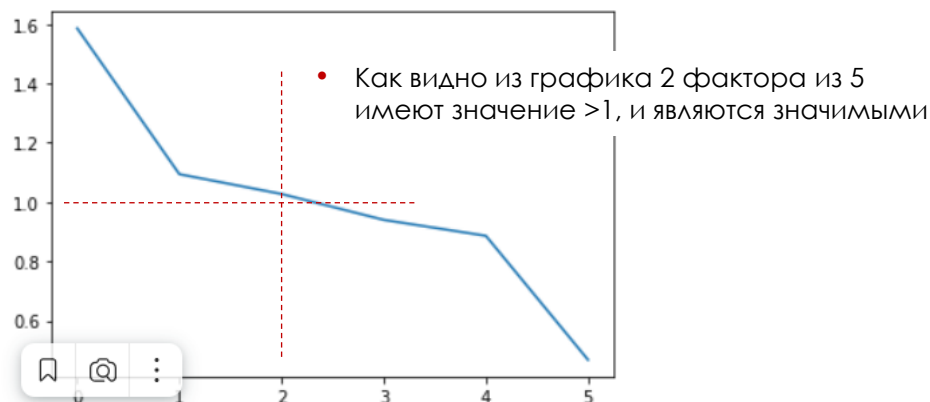
2.1. Исследование корреляции цены

- По тепловой карте видна высокая корреляция между ценой и каналом коммуникации, а также идентификатором пользователя
- Несмотря на то, что коэффициент корреляции между ценой и id очень высокий - 0,85, его достоверность = 0, т.е. судить о наличии зависимости цены от идентификатора пользователя некорректно, в данном случае речь идет скорее всего о предпочтении пользователей при использовании канала коммуникации



2.2. Факторный анализ

- Тест Бартлета показывает возможность выполнения факторного анализа, а также то, что в выборке существуют значимые корреляции между признаками



| | factor_1 | factor_2 |
|--------------|-----------|-----------|
| price | 0.988214 | 0.060782 |
| ad_channel | 0.525504 | -0.186476 |
| funnel_stage | 0.005842 | 0.076615 |
| timestamp | -0.000801 | 0.017767 |
| user_id | -0.025308 | 0.579600 |
| os | -0.218092 | -0.080403 |

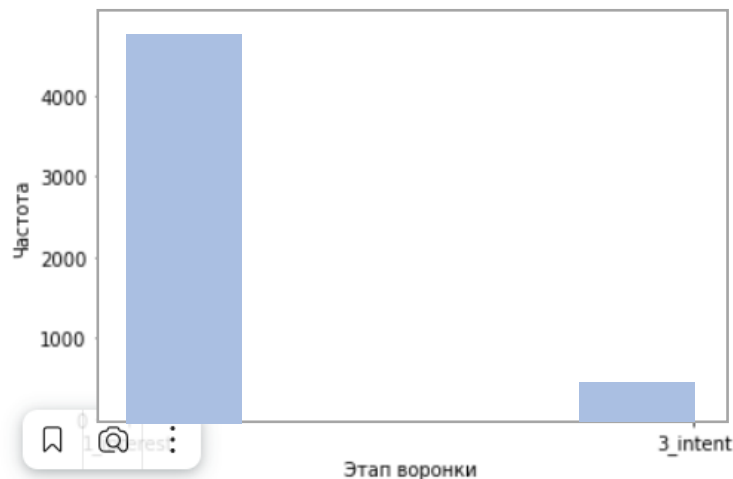
- Преобразовали изначальную матрицу с помощью метода transform() класса FactorAnalyzer и получили итоговое представление исходных данных в виде факторов
- По итогам изучения факторной матрицы видим, что фактор 1 выделяет значительное влияние на доход цены рекламного канала и связанного с ним фактора самого рекламного канала
- Фактор 2 – влияние id пользователя не является смысловым и не рассматриваем далее

2.3. Выводы по итогам факторного анализа

- Факторный анализ выделяет основной влияющий на доход фактор это рекламный канал и связанный с ним признак цены
- 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

3 Исследование фродовых данных

3.1. Распределение фродовых данных по этапам воронки



- Исследуем фродовые значения по > 1 действия в один момент времени – 5,2 тыс. действий (0,98%)
- Фродовые значения установлены только на этапах interest и intent
- Наибольшее количество — почти 92% на этапе интереса
- Причиной такого количества фродовых значений на этих этапах может быть обусловлено —
 - накруткой" откликов со стороны рекламных каналов
 - сбоем в обработки данных со стороны Интернета

- 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

3.2. Выводы исследования фродовых данных

В 3,5 раза больше

- заплатила бы компания за рекламу, при отсутствии простейших проверок данных (н-р, на дублирование данных)

минимум на 0,5% больше

- заплатила бы компания, если бы компания не контролировала данные на фальсификацию

| Итого | | 5175 | 0,80% |
|----------------|-----------------|------|-------|
| ВСЕГО | search | 2002 | 0,78% |
| | banner | 780 | 0,77% |
| | video | 904 | 0,91% |
| | organic_traffic | 710 | 0,80% |
| | partner_network | 534 | 0,74% |
| | native | 155 | 0,77% |
| Bcero interest | | 4758 | 0,80% |
| interest | search | 1765 | 0,77% |
| | banner | 721 | 0,77% |
| | video | 903 | 0,91% |
| | organic_traffic | 642 | 0,80% |
| | partner_network | 492 | 0,73% |
| | native | 152 | 0,77% |
| Bcero intent | | 417 | 0,81% |
| intent | search | 237 | 0,85% |
| | banner | 59 | 0,76% |
| | video | 1 | 1,67% |
| | organic_traffic | 68 | 0,73% |
| | partner_network | 42 | 0,93% |
| | native | 3 | 0,72% |
| | rich | 7 | 0,56% |

4 Исследование воронки продаж по каналам коммуникации

4.1. Анализ воронки продаж по каналам коммуникации в части конверсии в покупку

Высокая конверсия:

search

- канал с самой высокой конверсией в покупку – 1,4% (> средней)
- канал с самым большим использованием – 38%

Средняя конверсия:

banner
organic
partner

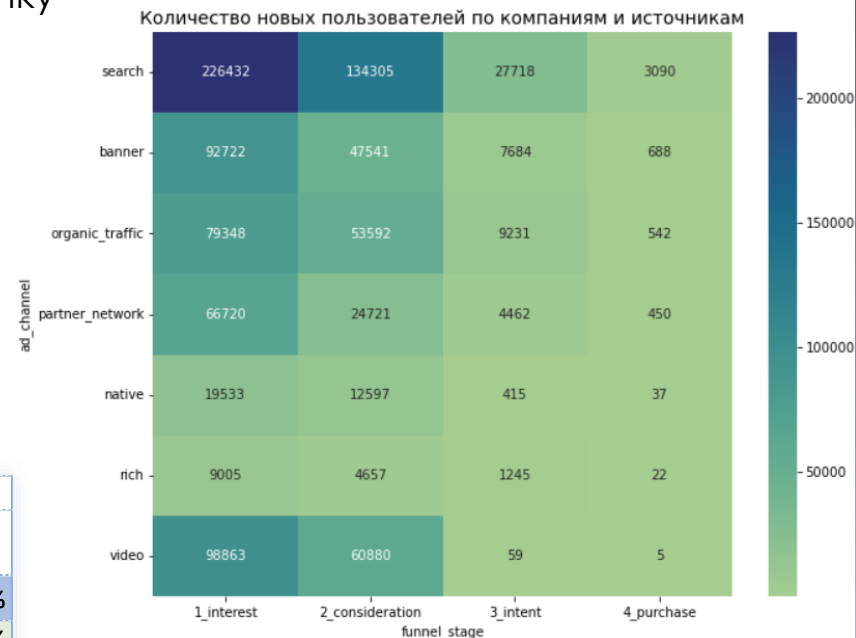
- - каналы со средней конверсией в покупку – 0,7% (около средней)
- - каналы с существенной долей привлечения – 40% в сумме

Низкая конверсия:

video
native
rich

- - канал с самой низкой конверсией в покупку – близко к 0%
- - канал с существенной долей привлечения – 16,7%
- - каналы с низкой конверсией в покупку – 0,2% (ниже средней)
- - каналы с низкой долей привлечения – 3,3 и 1,5%

| | Общий итог | 1_interest | | 2_consideration | | 3_intent | | 4_purchase | | |
|-----------------|------------|------------|-------------|-----------------|--------------|----------|--------------|------------|--------------|--------------|
| | кол-во | кол-во | % от общего | кол-во | % от 1 этапа | кол-во | % от 2 этапа | кол-во | % от 3 этапа | % от 1 этапа |
| Общий итог | 986 570 | 592 623 | | 338 293 | 57,1% | 50 814 | 15,0% | 4 834 | 9,5% | 0,8% |
| search | 391 546 | 226 432 | 38,2% | 134 305 | 59,3% | 27 718 | 20,6% | 3 090 | 11,1% | 1,4% |
| video | 159 808 | 98 863 | 16,7% | 60 880 | 61,6% | 59 | 0,1% | 5 | 8,5% | 0,0% |
| banner | 148 636 | 92 722 | 15,6% | 47 541 | 51,3% | 7 684 | 16,2% | 688 | 9,0% | 0,7% |
| organic_traffic | 142 714 | 79 348 | 13,4% | 53 592 | 67,5% | 9 231 | 17,2% | 542 | 5,9% | 0,7% |
| partner_network | 96 354 | 66 720 | 11,3% | 24 721 | 37,1% | 4 462 | 18,0% | 450 | 10,1% | 0,7% |
| native | 32 583 | 19 533 | 3,3% | 12 597 | 64,5% | 415 | 3,3% | 37 | 8,9% | 0,2% |
| rich | 14 930 | 9 005 | 1,5% | 4 657 | 51,7% | 1 245 | 26,7% | 22 | 1,8% | 0,2% |



- 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

4.2. Выводы исследования

- можно отказаться от канала video – очень низкая конверсия в результат
- следует уделить внимание каналам banner, 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

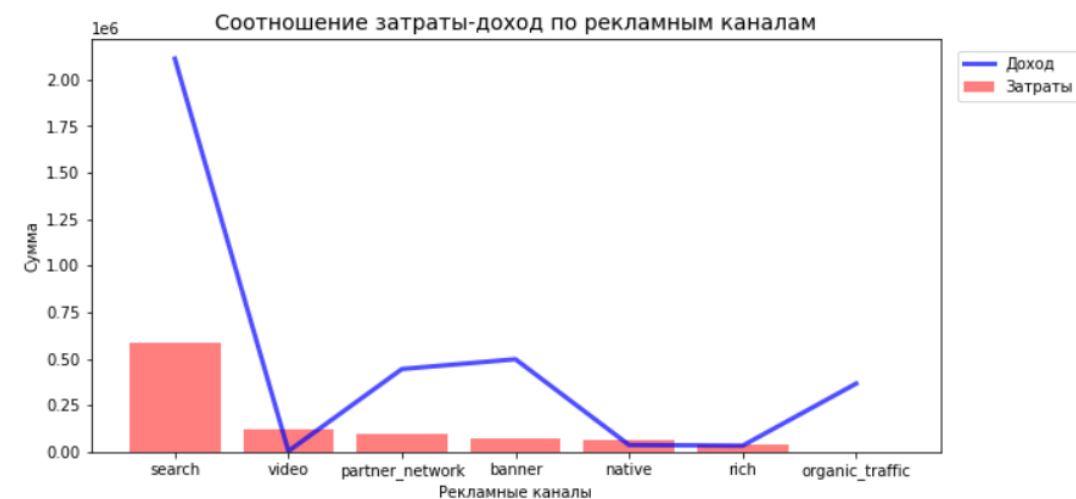
5 Исследование затрат и прибыли по рекламным каналам

5.1. Затраты и прибыль по рекламным каналам на последнем этапе воронки продаж

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

| | Доход | Цена рекламы | Расходы на рекламу | % доходы к расходам |
|-----------------|-----------|--------------|--------------------|---------------------|
| Общий итог | 3 498 027 | 1,0 | 980 330 | 356,8% |
| organic_traffic | 367 094 | - | - | |
| banner | 497 881 | 0,5 | 74 318 | 669,9% |
| partner_network | 445 411 | 1,0 | 96 353 | 462,3% |
| search | 2 112 193 | 1,5 | 587 318 | 359,6% |
| rich | 33 511 | 2,5 | 37 323 | 89,8% |
| native | 37 355 | 2,0 | 65 164 | 57,3% |
| video | 4 582 | 0,8 | 119 855 | 3,8% |

- Самые выгодные — каналы organic_traffic (нулевые издержки), banner, partner и search. На них и следует обращать внимание при разработке рекламной стратегии



- Из графика видно:
 - каналы video, native и rich являются убыточными
 - остальные каналы являются прибыльными

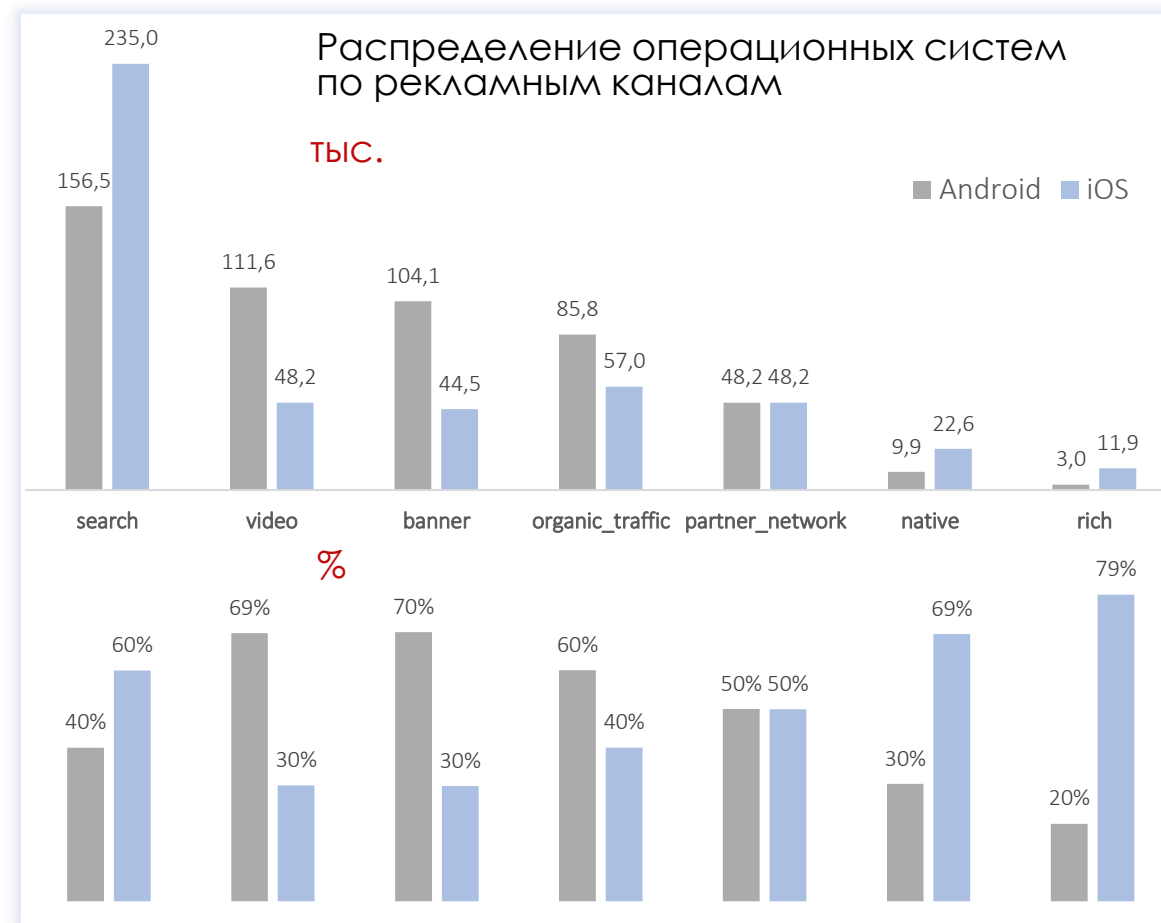
5.2. Выводы исследования

- 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

6 Исследование операционных систем пользователей по каналам коммуникации

6.1. Исследование ОС

- Распределение использования операционных систем по рекламным каналам дает характеристику и тип пользователей на текущий момент и в среднем по исследованию составляет почти 50% : 50%
- Следует учитывать общие тенденции и ее перспективу изменения соотношения использования iOS и Android (предполагаемое изменение в России в сторону Android)
- Предположительно исторически пользователи с iOS - более денежный класс пользователей, пользователи с Android - с меньшим достатком, сейчас ситуация меняется и подобных жестких определений уже проводить нельзя
- Но на момент проведения анализа —
 - преобладающее использование Android пользователями 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
- Проводились исследования особенностей людей, использующих разные ОС, можно их учитывать в планировании рекламных кампаний —
 - пользователи Android - воспринимаются окружающими как более честные, и уступчивые люди, но не более индивидуальны и не любят быть похожими на всех,
 - любители iPhone воспринимаются окружающими как общительные и более открытые, но при этом их отличает высокомерность и заносчивость, некоторые выражают свое социальное положение в обществе, демонстрируя свой новый телефон



6.2. Выводы исследования

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

7 Заключение

7.1. В части пользователей

- 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 (бо́льшие 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
- 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

7.2. В части каналов коммуникации

- 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
- Основным влияющим на доход фактором является рекламный канал (связанный с ним признак цены), остальные переменные (ОС и пр.) не являются влиятельными
- 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. **partner_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 затрат на рекламу по этому каналу
- 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
- Высвободившиеся затраты распределить на каналы 'banner' и 'partner_network' как наиболее выгодные каналы по вложению денег, в размере 70% и 30% соответственно

7.3. В части работы по этапам воронки продаж

| Воронки продаж по доходным каналам | | 1_interest | 2_consideration | 3_intent | 4_purchase | |
|------------------------------------|-----------------|-------------|-----------------|--------------|--------------|--------------|
| | | % от общего | % от 1 этапа | % от 2 этапа | % от 3 этапа | % от 1 этапа |
| | Общий итог | | 57,1% | 15,0% | 9,5% | 0,8% |
| | search | 38,2% | 59,3% | 20,6% | 11,1% | 1,4% |
| | banner | 15,6% | 51,3% | 16,2% | 9,0% | 0,7% |
| | organic_traffic | 13,4% | 67,5% | 17,2% | 5,9% | 0,7% |
| | partner_network | 11,3% | 37,1% | 18,0% | 10,1% | 0,7% |

- 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)

The image features a classic 'The End' title card. The text 'The End' is written in a white, elegant, cursive script font, centered on the screen. The background consists of several concentric circles in varying shades of gray, creating a tunnel-like or target-like effect. The circles are centered around the text, with the darkest circle being the innermost and the lightest being the outermost.

The End