

# Мобильное приложение «Ненужные вещи»

Выделение групп пользователей на основе поведения

Выполнил: Горячев Д.О.

# Задание

Выделите группы пользователей, которые различаются по метрикам:

- retention rate,
- время, проведённое в приложении,
- частота действий,
- конверсия в целевое действие — просмотр контактов.

Проведите исследовательский анализ данных

Сегментируйте пользователей на основе действий

Проверьте статистические гипотезы

1. Некоторые пользователи установили приложение по ссылке из yandex, другие — из google. Проверьте гипотезу: две эти группы демонстрируют разную конверсию в просмотры контактов.
2. Сформулируйте собственную гипотезу. Дополните её нулевой и альтернативной гипотезами. Проведите статистический тест.

# Описание данных

Датасет содержит данные о событиях, совершенных в мобильном приложении "Ненужные вещи". В нем пользователи продают свои ненужные вещи, размещая их на доске объявлений.

В датасете содержатся данные пользователей, впервые совершивших действия в приложении после 7 октября 2019 года.

Датасет `mobile_dataset.csv` содержит колонки:

- `event.time` — время совершения
- `event.name` — название события
- `user.id` — идентификатор пользователя

Датасет `mobile_sources.csv` содержит колонки:

- `userId` — идентификатор пользователя
- `source` — источник, с которого пользователь установил приложение

Расшифровки событий:

- `advert_open` — открытие карточки объявления
- `photos_show` — просмотр фотографий в объявлении
- `tips_show` — пользователь увидел рекомендованные объявления
- `tips_click` — пользователь кликнул по рекомендованному объявлению
- `contacts_show` и `show_contacts` — пользователь нажал на кнопку "посмотреть номер телефона" на карточке объявления
- `contacts_call` — пользователь позвонил по номеру телефона на карточке объявления
- `map` — пользователь открыл карту размещенных объявлений
- `search_1` — `search_7` — разные события, связанные с поиском по сайту
- `favorites_add` — добавление объявления в избранное

# Общий вывод

**Итак, в ходе проведения анализа, мы выполнили все запланированные манипуляции с данными:**

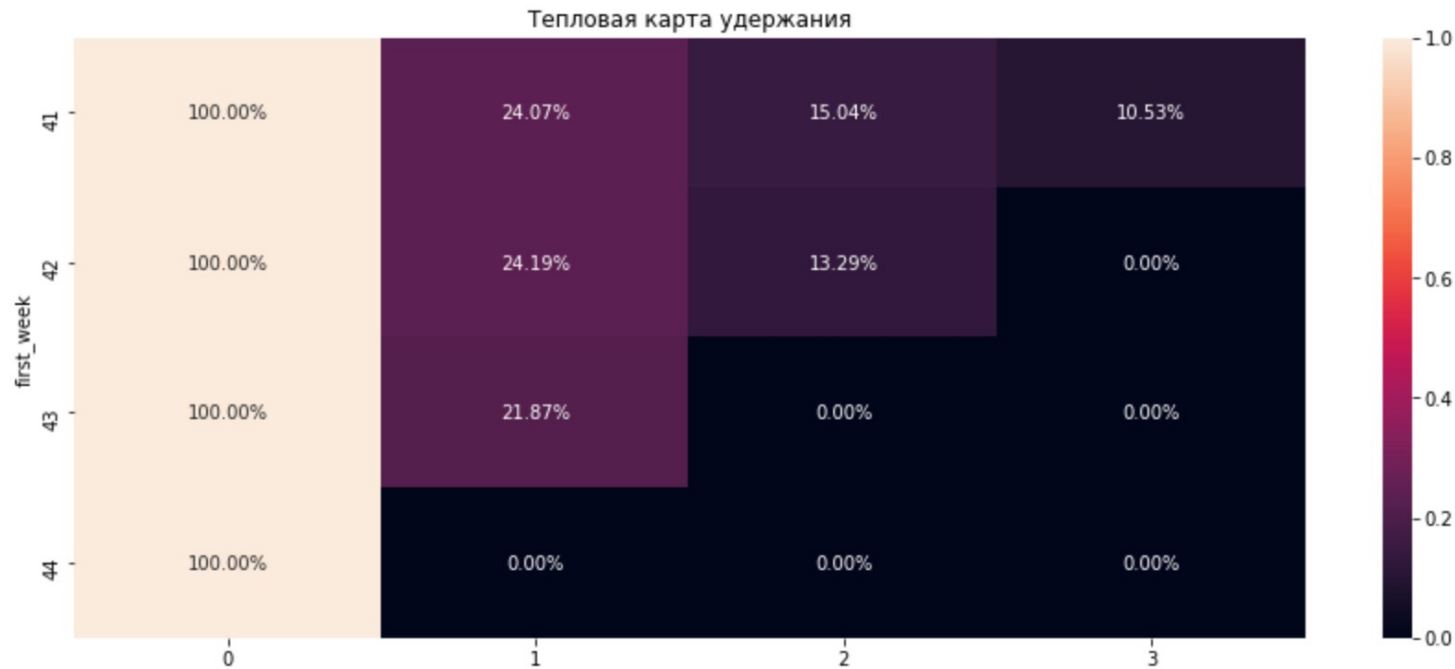
- провели предобработку данных, в результате которой объединили 2 датафрейма в 1, очистили его от дубликатов и пропусков
- провели исследовательский анализ данных, выяснили что:
  - Удержание у пользователей второй недели хуже, чем у пользователей первой недели, а у пользователей 3 недели существенно хуже. Но диапазон исследования больно мал, поэтому что-либо утверждать преждевременно. Пока понятно только то, что каждая когорта ведет себя по-разному. Если у первой когорты между 2 и 3 лайфтаймом падение составило 8,85%, то у второй падение было на 10,21, разница более 2%. У 3 когорты в 1й лайфтайм оказался ниже первых 2х когорт болле, чем на 2%. С 4 когортой вообще ничего не понятно
  - общая конверсия пользователей в целевое действие составила 22,81 %
  - в среднем пользователи проводят в приложении по 2 сессии
  - самое популярное действие после показа рекомендованных объявлений - просмотр фото
- поделили пользователей на группы в зависимости от источника прихода и выяснили, что:
  - пользователей из гугла и из других источников примерно одинаковое количество, а посетителей из яндекса более чем в 1,5 раза больше
  - удержание лучшего всего у пользователей из других источников, между яндексом и гуглом лидирует гугл
  - конверсия пользователей из гугла и яндекса составляет 24,36% и 24,72% соответственно, а конверсия пользователей из других источников 18,39%
- проверили гипотезы о равенстве конверсий между пользователями из яндекс и пользователями из гугл и о равенстве конверсий между теми, кто просматривает фото и теми, кто этого не делает. По результатам проверки гипотез получили следующие результаты:
  - статистически значимой разницы между конверсией групп Яндекс и Гугл не обнаружено
  - разница между конверсией пользователей, просматривающих фото и конверсией пользователей, не делающих этого статистически значима

# Рекомендации

**По результатам проведенного анализа можно дать следующие рекомендации:**

- необходимо поработать над удержанием пользователей, в особенности над удержанием пользователей, пришедших по ссылке из яндекса, потому что оттуда приходит больше всего людей
- привлекать больше пользователей из источника Гугл, так как у них хорошее удержание
- странным показалось, что по сути самым популярным действием является просмотр фото (всплывание рекомендованных объявлений происходит само собой и для этого не требуется активных действий пользователя), а не поиск. Поиск в таком приложении очень важен, т.к. именно с его помощью пользователи могут эффективно найти то, что им требуется. Возможно, на данный момент поиск расположен не очень удобно и стоит поменять его размещение на странице.

# Retention Rate



## Вывод №1

Удержание у пользователей второй недели хуже, чем у пользователей первой недели, а у пользователей 3 недели существенно хуже. Но диапазон исследования больно мал, поэтому что-либо утверждать преждевременно. Пока понятно только то, что каждая когорта ведет себя по-разному. Если у первой когорты между 2 и 3 лайфтаймом падение составило 8,85%, то у второй падение было на 10,21, разница более 2%. У 3 когорты в 1й лайфтайм оказался ниже первых 2х когорт болле, чем на 2%. С 4 когортой вообще ничего не понятно

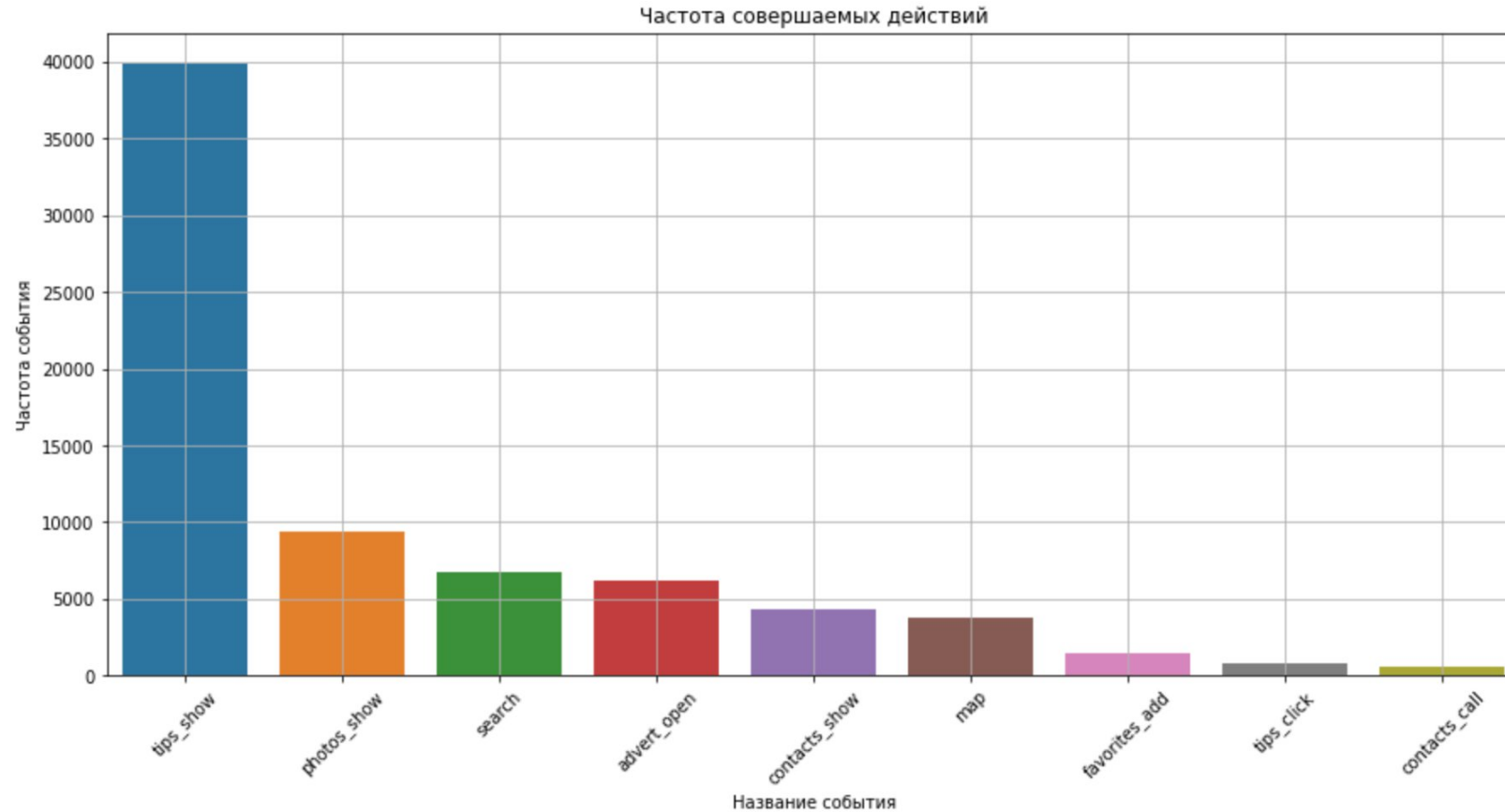
# Время проведенное в приложении

## Вывод №2

Исходя из проведенного анализа, видим, что пользователи в среднем проводят в приложении 2 сессии, то есть около часа. Достаточно для того, чтобы найти подборку нужных товаров, изучить ее и остановиться на чем-то конкретном, после чего совершить целевое действие. По идее конверсия должна быть хорошая, но это мы выясним дальше.

Также, мы выяснили, что средняя продолжительность сессии - 12 минут 52 секунды, а еще в датафрейме 10,5 тысяч строк с нулевыми сессиями, в которые входят все возможные события. Нулевые сессии могут означать нестабильную работу и "вылет" из приложения. Поскольку в нулевых сессиях мы видим наличие всех видов событий, "вылеты" не связаны с совершением какого-либо конкретного действия, а касаются стабильности работы приложения в целом.

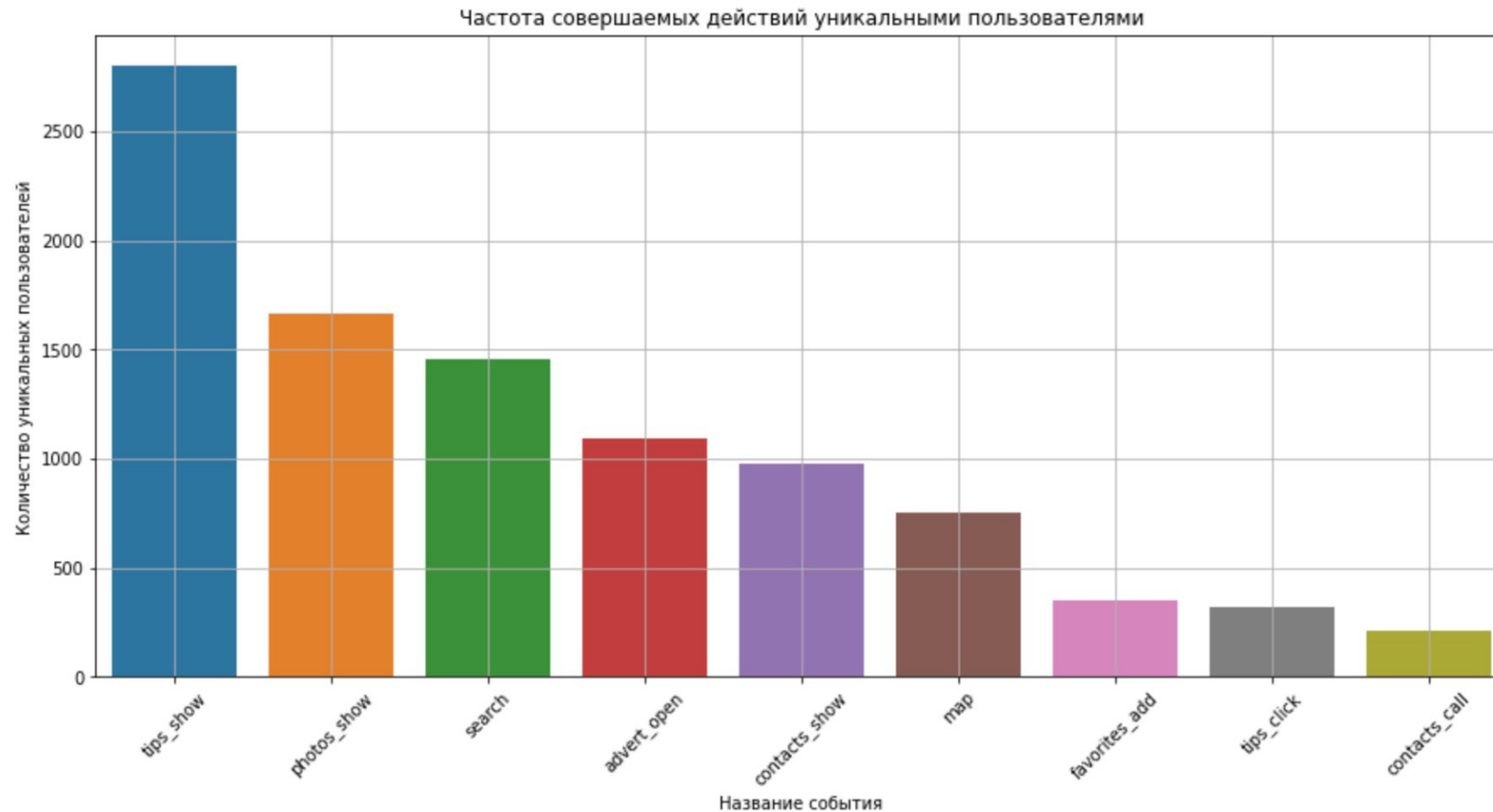
# Частота действий



Самое часто совершаемое действие - просмотр рекомендованных объявлений, целевое действие - просмотр контактов на 5 месте. Конверсия должна быть высокой.



# Частота действий



В пересчете частоты на уникальных пользователей распределение частоты совершаемых действий остается неизменным.

# Частота действий



График динамики событий сильно скачет, пик активности пользователей пришелся на 23.10.2019, а самые большие спады отмечены 12.10.2019 и 02.11.2019. В остальном график показывает стабильность в том плане, что за более активными днями следуют менее активные и в основном пользователи совершают от 2600 до 3000 действий в сутки.

# Частота действий

## Вывод №3

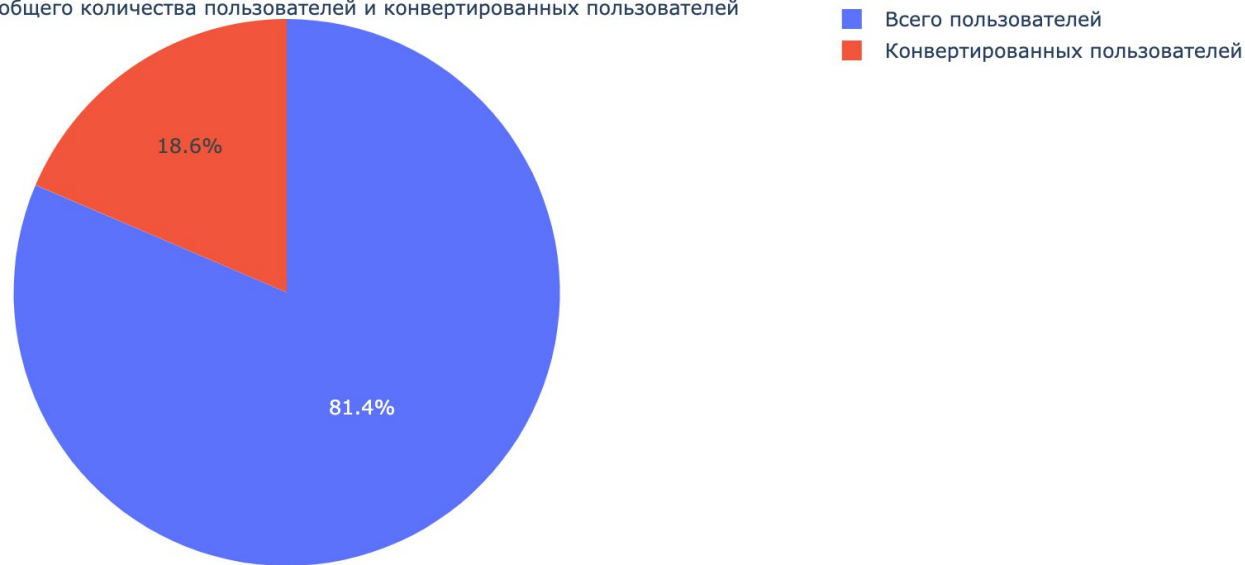
Самым частым действием является всплывание рекомендованных объявлений, далее с большим отрывом просмотр фото, на 3-м месте поиск, на 4-м открытие карточки объявления и на 5-м запрос контактов продавца.

Примечательно, что звонок по контактам из карточки - самое непопулярное действие. Несмотря на то, что рекомендованные объявления у пользователей всплывают очень часто, кликают по ним крайне редко. Почти также редко пользователи добавляют карточки объявлений в избранное

Что касается динамики активности пользователей, то пик активности пользователей пришелся на 23.10.2019, а самые большие спады отмечены 12.10.2019 и 02.11.2019.

# Конверсия в целевое действие

Распределение общего количества пользователей и конвертированных пользователей

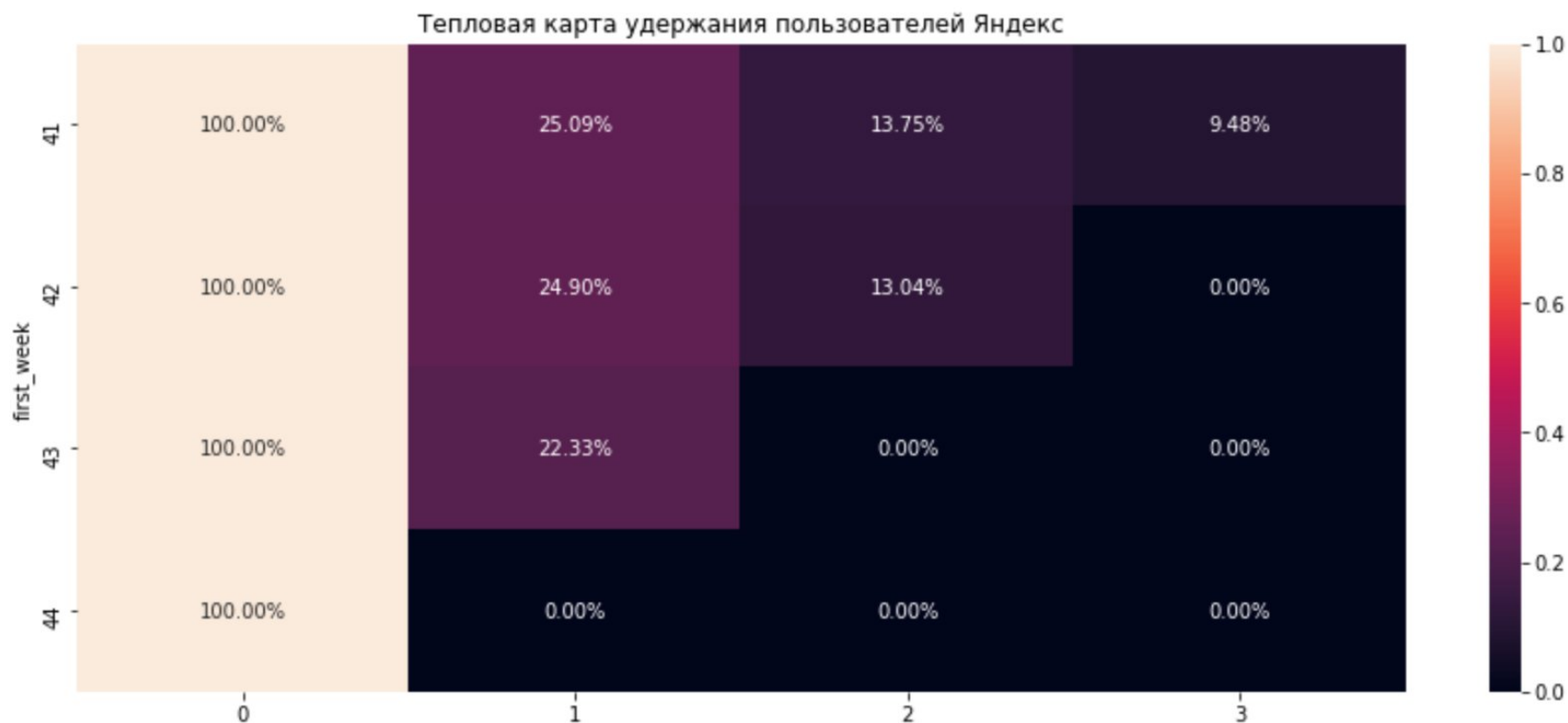


## Вывод №4

Общая конверсия пользователей в целевые действия - 22,8% - это достаточно высокий показатель. Также мы выяснили, что 18,6% пользователей совершают целевые действия.

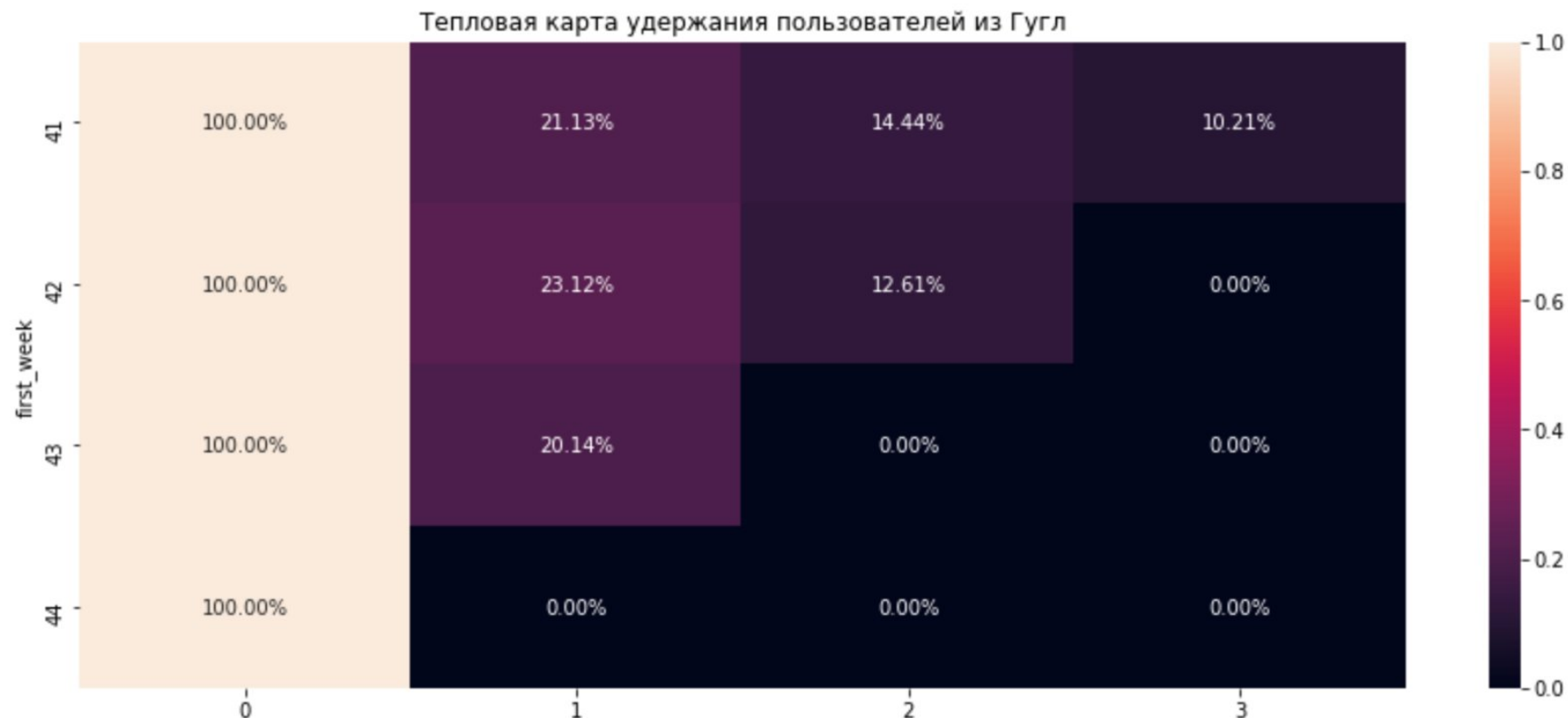
Конверсию пользователей в зависимости от источника прихода изучим в следующем разделе анализа.

# Retention Rate пользователей из Яндекса



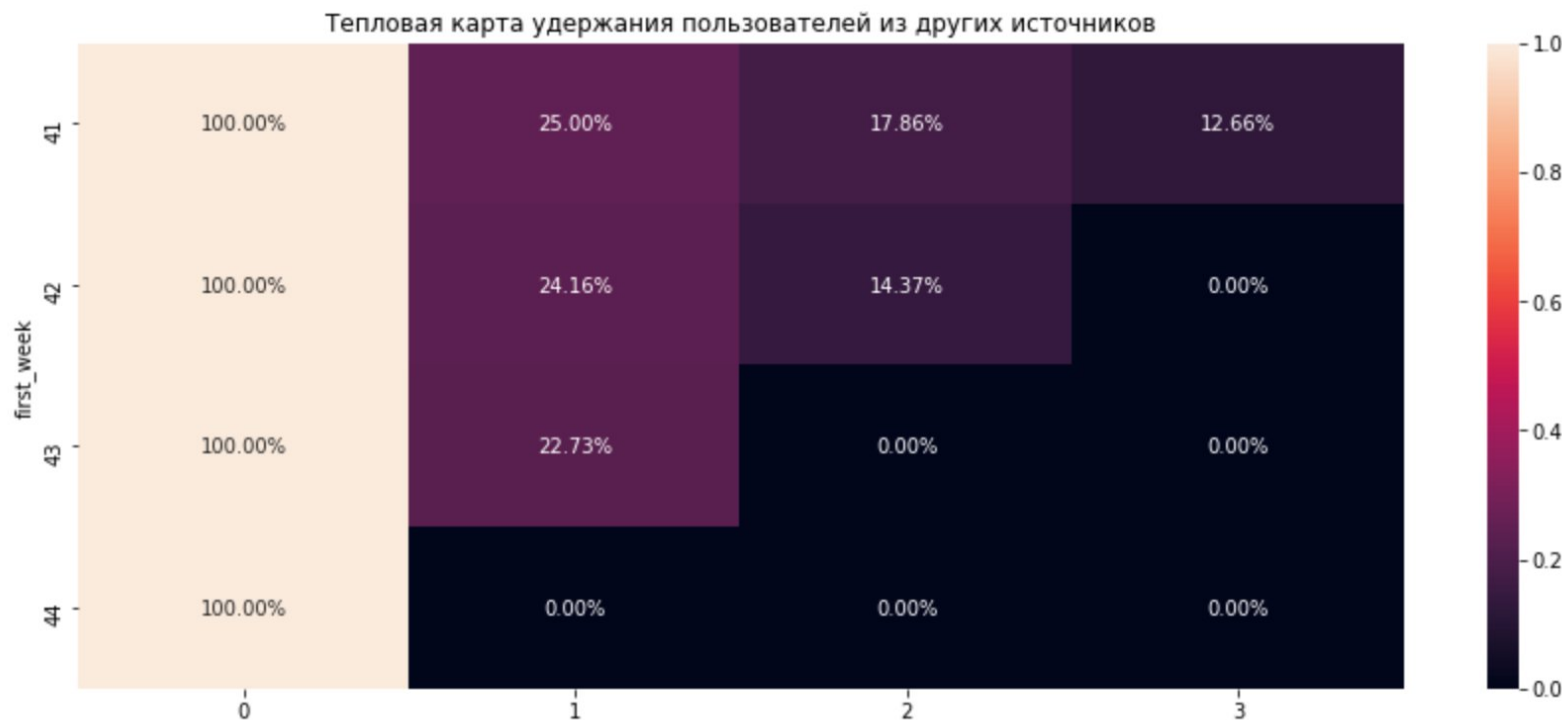
Удержание пользователей, пришедших из Яндекса в 1 лайфтайме с каждой неделей падает, во 2 лайфтайме примерно одинаковые показатели.

# Retention Rate пользователей из Гугл



Удержание пользователей из Гугл на каждой неделе отличается, какой-то единой тенденции не прослеживается. При этом, существенных отличий от группы Яндекс в первом лайфтайме нет, однако во 2 и 3 лайфтаймах удержание группы Гугл выше.

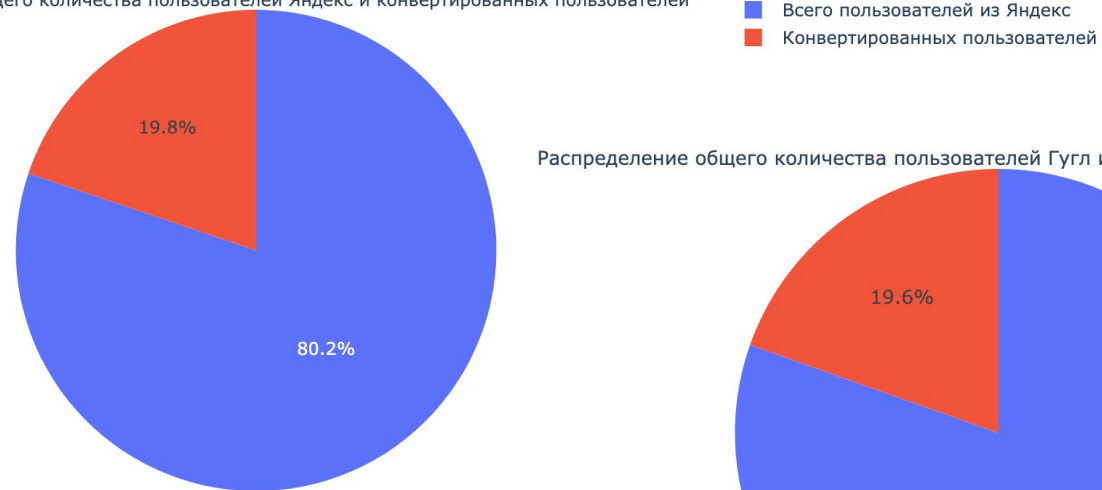
# Retention Rate пользователей из других источников



Удержание пользователей из других источников несколько выше, чем у групп Яндекса и Гугла. Это может быть связано с тем, что люди, пришедшие из других источников скачивают приложение с конкретно целью или по рекомендации, в то время, как пользователи по ссылкам из Яндекса и Гугла могли прийти из любопытства и для ознакомления с приложением. Это бы объясняло отток пользователей, которые не заинтересовались приложением.

# Конверсия для каждой группы

Распределение общего количества пользователей Яндекс и конвертированных пользователей



Распределение общего количества пользователей Гугл и конвертированных пользователей



Распределение общего количества пользователей из других источников и конвертированных пользователей

