|  |  |
| --- | --- |
| Задание 1 | SQL  (Предварительное формирование входного набора данных в Redash) |
| В наших данных использования ленты новостей есть два типа юзеров: те, кто пришел через платный трафик source = 'ads', и те, кто пришел через органические каналы source = 'organic'.  Ваша задача — проанализировать и сравнить Retention этих двух групп пользователей. Решением этой задачи будет ответ на вопрос: отличается ли характер использования приложения у этих групп пользователей. | Для графиков по недельной динамике  SELECT this\_week,  previous\_week,  -uniq(user\_id) as num\_users,  status,  source  FROM  (SELECT user\_id,  max(source) as source,  groupUniqArray(toMonday(toDate(time))) as weeks\_visited,  addWeeks(arrayJoin(weeks\_visited), +1) this\_week,  if(has(weeks\_visited, this\_week) = 1, 'retained', 'gone') as status,  addWeeks(this\_week, -1) as previous\_week  FROM simulator\_20220920.feed\_actions  GROUP BY user\_id  )  WHERE status = 'gone'  GROUP BY this\_week, previous\_week, status, source  HAVING this\_week != addWeeks(toMonday(today()), +1)  UNION ALL  SELECT this\_week,  previous\_week,  toInt64(uniq(user\_id)) as num\_users,  status,  source  FROM  (SELECT user\_id,  max(source) as source,  groupUniqArray(toMonday(toDate(time))) as weeks\_visited,  arrayJoin(weeks\_visited) this\_week,  if(has(weeks\_visited, addWeeks(this\_week, -1)) = 1, 'retained', 'new') as status,  addWeeks(this\_week, -1) as previous\_week  FROM simulator\_20220920.feed\_actions  GROUP BY user\_id  )  GROUP BY this\_week,  previous\_week,  status,  source |
| Для графиков с тепловыми картами  SELECT toString(start\_day) AS start\_day,  toString(day) AS date,  count(user\_id) AS active\_users,  source,  toDate(date)-toDate(start\_day) AS day\_in\_cohorts  FROM  (SELECT \*  FROM  (SELECT user\_id,  source,  min(toDate(time)) AS start\_day  FROM simulator\_20220920.feed\_actions  GROUP BY user\_id,  source  ) t1  JOIN  (SELECT DISTINCT user\_id,  toDate(time) AS day  FROM simulator\_20220920.feed\_actions  ) t2  USING user\_id  WHERE (start\_day >= today()-15) AND (day < (today() -1))  )  GROUP BY start\_day,  day,  source |



Выводы:

По итогам анализа можно сказать, что по пользователям, пришедшим через платный трафик, обычно чуть больше (чем у тех, кто пришел через органические каналы) группа новых пользователей, зато существенно меньше группа остающихся пользователей. Ну, и как следствие, группа оттока (тех, кто ушел), выше у пользователей с source = 'ads'. То есть через платный трафик в приложение приходит людей чуть больше, но остается в дальнейшем гораздо меньше.

Существенных различий по поведению когорт пользователей, различающихся типом подписки, в зависимости от типа дня недели (будни или выходной) обнаружено не было (на данных за последние 2 недели).

Настройки дашборда:

Блок «JSON METADATA» в «Dashboard properties», с изменениями, позволившими задать фиксированные цвета в графиках рядов «Сравнительная недельная динамика» и «Недельная динамика пользователей» для удобства визуального анализа.

{

"label\_colors": {

"organic, retained": "red",

"ads, retained": "blue",

"organic, new": "red",

"ads, new": "blue",

"retained": "#21759B",

"new": "#1AB7EA",

"gone": "orange"

},

"timed\_refresh\_immune\_slices": [],

"expanded\_slices": {},

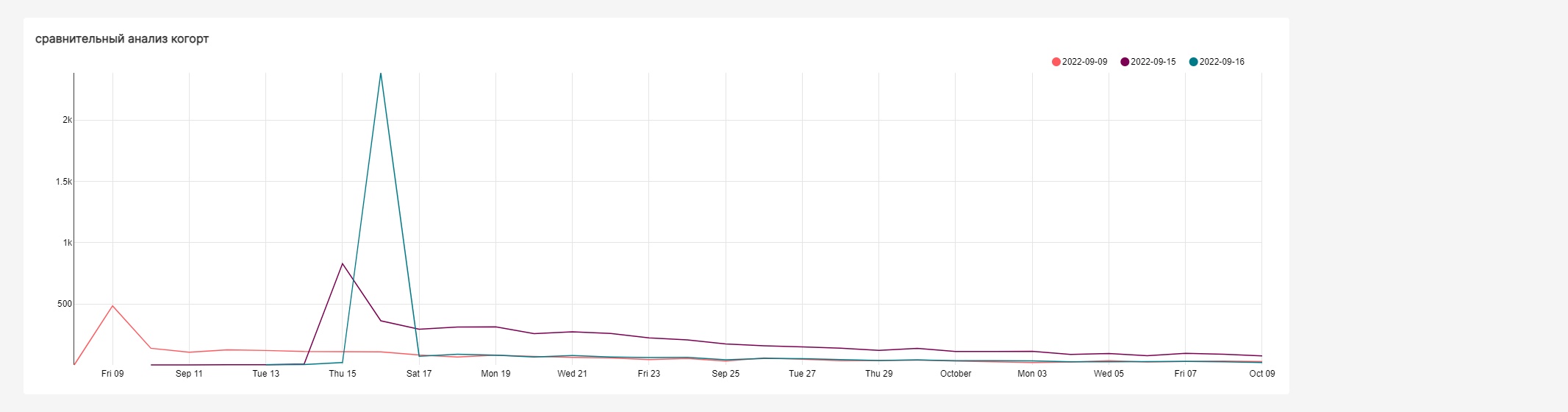
"refresh\_frequency": 0,

"default\_filters": "{}",

"color\_scheme": "SUPERSET\_DEFAULT"

}

|  |  |
| --- | --- |
| Задание 2 | SQL  (Предварительное формирование входного набора данных в Redash) |
| Маркетологи запустили массивную рекламную кампанию, в результате в приложение пришло довольно много новых пользователей, вы можете видеть всплеск на графике активной аудитории.  Однако у нас есть сомнение в качестве трафика, изучите, что стало с рекламными пользователями в дальнейшем, как часто они продолжают пользоваться приложением? | SELECT toString(start\_day) AS start\_day,  toString(day) AS date,  count(DISTINCT user\_id) AS user\_unique,  count(user\_id) AS count\_action,  os,  action,  source  FROM  (SELECT \*  FROM  (SELECT user\_id,  action,  os,  source,  min(toDate(time)) AS start\_day  FROM simulator\_20220920.feed\_actions  GROUP BY user\_id,  os,  action,  source  ) t1  JOIN  (SELECT DISTINCT user\_id,  toDate(time) AS day  FROM simulator\_20220920.feed\_actions  ) t2  USING user\_id  WHERE start\_day in (toDate('2022-09-16') , toDate('2022-09-16') - 1, toDate('2022-09-16') -7))  GROUP BY start\_day,  day,  os,  action,  source |

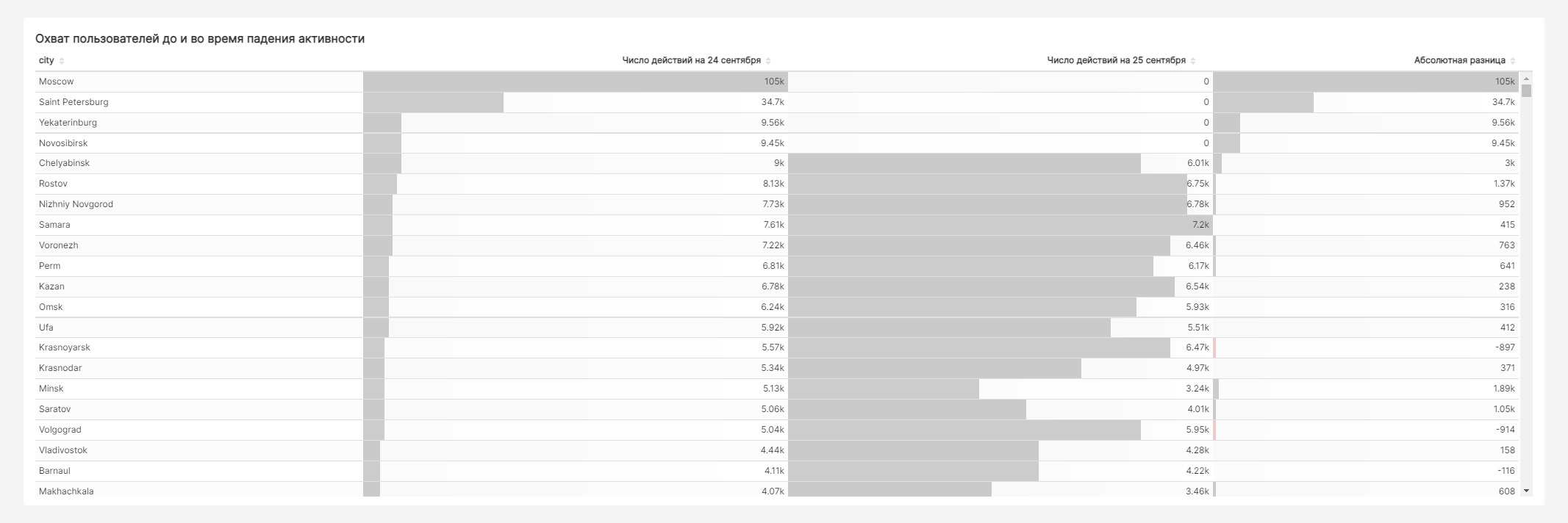


Выводы:

На рисунке на дашборде показана динамика (в абсолютных числах) новых рекламных пользователей, пришедших 16 сентября (испытавшие на себе влияние рекламной компании, помечены как «2022-09-16»), пришедших на день раньше (помечены как «2022-09-15»), и пришедших на неделю раньше (чтобы можно было провести сравнительный анализ с тем же самым днем недели, помечены как «2022-09-09»)). Из графика видно, что в количественном отношении рекламная компания привела к резкому возрастанию (в 5 раз больше, чем на неделю раньше, и почти в 3 раза больше, чем в предыдущий день) числа новых пользователей. Но уже на следующий день от этого числа осталось в приложении столько же пользователей, как и у когорты, сформированной на неделю раньше, и меньше, чем у когорты, сформированной на день раньше. А с учетом общего тренда на увеличение числа пользователей, наблюдаемое соотношение с когортой с недельной разницей говорит о том, что уже на второй день осталось в приложении пользователей даже меньше, чем если бы это были обычные условия, без рекламной компании.

Таким образом, можно сделать вывод о том, что рекламная компания привела к значительному росту новых пользователей, однако это наращение никак не конверсировалось в пользователей, оставшихся в приложении.

|  |  |
| --- | --- |
| Задание 3 | SQL  (Предварительное формирование входного набора данных в Redash) |
| Мы наблюдаем внезапное падение активной аудитории! Нужно разобраться, какие пользователи не смогли зайти в приложение, что их объединяет? | SELECT toStartOfDay(toDateTime(time)) AS timestamp,  count(DISTINCT user\_id) AS user\_unique,  count(user\_id) AS count\_action,  country\_gr,  os,  action,  city  FROM  (SELECT country,  city,  action,  os,  source,  time,  user\_id,  multiIf(country LIKE 'Russia', 'Russia', country LIKE 'Kazakhstan','Kazakhstan',  country LIKE 'Belarus', 'Belarus', country LIKE 'Ukrain' , 'Ukrain', 'Other')  AS country\_gr  FROM simulator\_20220920.feed\_actions  )  GROUP BY timestamp,  country\_gr,  os,  action,  city  HAVING timestamp between toDate('2022-09-25') - 2 AND toDate('2022-09-25') + 2 |



Выводы:

Поскольку на основном дашборде по новостям были подготовлены рисунки, позволяющие сравнить динамику DAU по таким характеристикам, как пол, возраст, os, sourse, и по ним не было обнаружено влияния этих факторов на падение активности, случившееся 25 сентября, было решено сразу начать анализ с влияния географического фактора.

Самой подробной характеристикой пользователей с точки зрения географии является город, поэтому было решено сразу рассматривать группировку по городам.

Но поскольку городов много, а определить хочется в первую очередь тех, кто вносит наибольший вклад, было решено представить результаты в виде таблицы, в которой гистограммы позволяют хорошо разглядеть вклад городов в активность в приложении в тот день, когда случилось падение, и предшествующий ему. Также это позволяет отследить, при надобности, динамику по всем городам, что позволяет выявить потенциальные неожиданности.

По итогам анализа построенного графика: жители четырех крупнейших (по крайней мере с точки зрения предыдущего вклада в активность в приложении) городов: Москвы, Санкт-Петербурга, Екатеринбурга и Новосибирска – не смогли 25 сентября зайти в приложение. Поскольку активность была полностью нулевой, смысла в дальнейшем когортном анализе нет. По всей видимости, причина была чисто техническая, связанная с каналами связи. Поскольку по остальным городам каких-либо видимых аномалий нет, а множество пользователей полностью покрывается объединением их подмножеств по городам, можно считать, что обнаруженное отсутствие трафика по 4 крупнейшим городам полностью объясняет наблюдаемый скачок вниз активности.