

Дашборд с аналитикой эффективности рекламной площадки

Целью работы «Дашборд с аналитикой эффективности рекламной площадки» является визуализация с помощью дашборда основных продуктовых метрик рекламы в социальной сети, статистики по дням, конверсии показов в клик и финансовых показателей. Такой дашборд может использоваться менеджером продукта (рекламной площадки социальной сети) для отслеживания динамики показателей рекламной площадки и принятия решений по продукту, таких как отбор наиболее прибыльных и кликабельных рекламных объявлений, выбор платформ для показа, выбор типа оплаты рекламных объявлений за клики или показы и др.

Аналитика эффективности считалась на синтетических данных, похожих на реальные данные рекламной площадки одной из популярных социальных сетей в России. Для построения дашборда использовался Apache Superset — открытое программное обеспечение для исследования и визуализации данных, ориентированное на большие данные. Для ответов на вопросы об эффективности также использовались SQL-запросы на диалекте ClickHouse к базе данных внутри системы Superset.

Работа имеет высокую актуальность и практическую значимость как пример аналитики эффективности в интернет-маркетинге и пример визуализации основных показателей эффективности, на примере рекламной площадки. Данный проект иллюстрирует выполнение типичной рабочей задачи аналитика данных в рекламе.

Постановка задачи

К аналитику отдела рекламы со стороны менеджера рекламной площадки был поставлен ряд вопросов:

1. Получить статистику по дням. Посчитать число всех событий по дням, число показов, число кликов, число уникальных объявлений и уникальных кампаний.
2. Разобраться, почему случился такой скачок 2019-04-05? Каких событий стало больше? У всех объявлений или только у некоторых?
3. Найти топ 10 объявлений по CTR за все время. CTR — это отношение всех кликов объявлений к просмотрам. Например, если у объявления было 100 показов и 2 клика, $CTR = 0.02$. Различается ли средний и медианный CTR объявлений в наших данных?
4. Похоже, в наших логах есть баг, объявления приходят с кликами, но без показов! Сколько таких объявлений, есть ли какие-то закономерности? Эта проблема наблюдается на всех платформах?
5. Есть ли различия в CTR у объявлений с видео и без? Чему равняется 95-перцентиль CTR по всем объявлениям за 2019-04-04?
6. Для финансового отчета нужно рассчитать наш заработок по дням. В какой день мы заработали больше всего? В какой меньше? Мы списываем с клиентов деньги, если произошел клик по CPC объявлению, и мы списываем деньги за каждый показ CPM объявления. Если у CPM объявления цена - 200 рублей, то за один показ мы зарабатываем $200 / 1000$.

7. На какой платформе больше всего показов? Сколько процентов показов приходится на каждую из платформ?
8. А есть ли такие объявления, по которым сначала произошел клик, а только потом показ?

Помимо ответов на вопросы в итоге должен получился дашборд, который поможет менеджерам самостоятельно находить на них ответы и узнавать состояние рекламной системы. В дашборде помимо графиков, которые будут отвечать на поставленные выше вопросы, должны быть выведены основные продуктовые метрики, такие как показы, клики, конверсии в клик, сумма заработанных на площадке денег.

Для учебных целей для анализа использовались синтетические данные, похожие на реальные данные рекламной площадки одной из популярных социальных сетей в России. Размер датасета (базы данных): 1 млн строк.

Ответы на представленные выше вопросы позволят проанализировать эффективность рекламной площадки, выявить причины аномалий в данных (таких как скачок кликов и показов в одну из дат), посчитать кликабельность (CTR) и заработок за размещение объявлений.

Для построения дашборда с ответами на вопросы использовался Apache Superset — открытое программное обеспечение для исследования и визуализации данных, ориентированное на большие данные.

Для удобства чтения визуализации аналитики эффективности рекламной площадки был создан не один, а три дашборда: «Основные продуктовые метрики и статистика по дням», «CTR и финансовый отчет» и «Аномалии и баги рекламной площадки». (Ссылки на дашборды и их скриншоты приведены в Приложении 1).

Визуализация на дашбордах объединена по темам, а не по последовательности вопросов по аналитике рекламной площадки: на дашборде «Основные продуктовые метрики и статистика по дням» даны ответы на вопросы 1 и 7; на дашборде «CTR и финансовый отчет» - ответы на вопросы 3, 5 и 6; на дашборде «Аномалии и баги рекламной площадки» - ответы на вопросы 2, 4 и 8. В некоторых случаях на дашборде представлена дополнительная аналитика, которая не упоминалась в вопросах, но которая дает дополнительную полезную информацию по обозначенным темам (например, дополнительное распределение по дням или распределение по платформам и т.п.)

Для первого дашборда («Основные продуктовые метрики и статистика по дням») предусмотрены фильтры: по дате, типу объявления (CPM и CPC), наличию видео и платформе (Android, iOS и Web). Графики на нем строились на основе основного датасета, что позволило применить фильтры. На двух других дашбордах фильтры не предусмотрены, т.к. графики на них строились на основе кастомных SQL-запросов (в ClickHouse), что ограничивает применение фильтрации, однако потенциально их сделать тоже возможно.

Кроме дашборда, для ответов на вопросы использовались SQL-запросы к базе данных на диалекте ClickHouse внутри системы Superset. Результаты представлены в следующем разделе.

Описание данных¹

date	день, в который происходят события
time	точное время события
event	тип события, может быть или показ, или клик по рекламе
platform	платформа, на которой произошло рекламное событие
ad_id	id рекламного объявления
client_union_id	id рекламного клиента (рекламодателя)
campaign_union_id	id рекламной кампании
ad_cost_type	тип объявления с оплатой за клики (CPC) или за показы (CPM)
ad_cost	стоимость объявления в рублях, для CPC объявлений – это цена за клик, для CPM – цена за 1000 показов
has_video	есть ли у рекламного объявления видео
target_audience_count	размер аудитории, на которую таргетируется объявление

¹ Данные можно скачать по ссылке: <https://disk.yandex.ru/d/6jONLO9mrLmHpA>

Результаты анализа эффективности рекламной площадки (с интерпретацией)

Ниже представлены ответы на вопросы из раздела «Методы»².
Ответы приведены в виде скриншотов либо результатов запросов в ClickHouse, либо с дашбордов.

1. Получить статистику по дням. Посчитать число всех событий по дням, число показов, число кликов, число уникальных объявлений и уникальных кампаний.

Число всех событий по дням

date	events_number
2019-04-01	22073
2019-04-02	47117
2019-04-03	59483
2019-04-04	275735
2019-04-05	519707
2019-04-06	75885

Больше всего событий – 2019-04-05, меньше всего – 2019-04-01.
Общее число событий – 1 000 000.

Число показов по дням

² Код для запросов и для графиков на дашбордах приведен в Приложении 2 (в той же последовательности и с теми же названиями подразделов, что и в данном разделе).

date	event	events_number
2019-04-01	view	21782
2019-04-02	view	46572
2019-04-03	view	59023
2019-04-04	view	275092
2019-04-05	view	427386
2019-04-06	view	60967

Больше всего показов - 2019-04-05, меньше всего – 2019-04-01.

Число кликов по дням

date	event	events_number
2019-04-01	click	291
2019-04-02	click	545
2019-04-03	click	460
2019-04-04	click	643
2019-04-05	click	92321
2019-04-06	click	14918

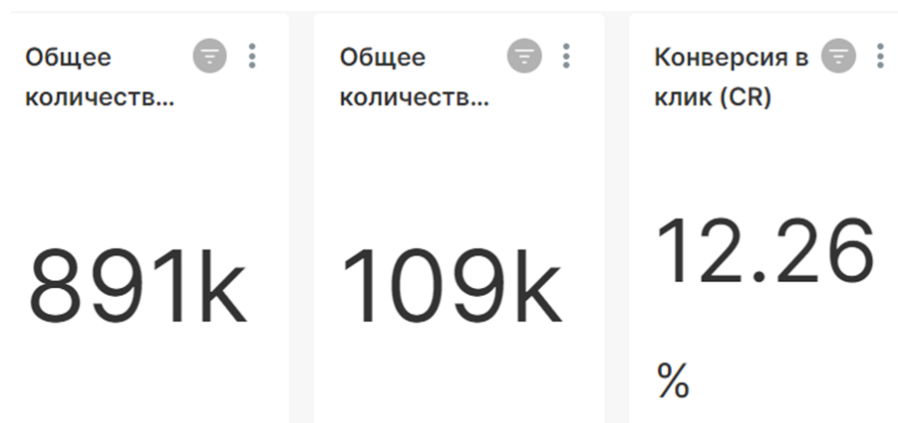
Больше всего кликов – 2019-04-05, меньше всего – 2019-04-01.

Также число показов и кликов по дням отображено с помощью графика на дашборде «Основные продуктовые метрики и статистика по дням»:



В качестве основных метрик на дашборд выведены следующие виджеты:

- Общее количество показов
- Общее количество кликов
- Конверсия в клик (CR)



Общее количество показов (точное значение)

views_number
890822

Общее количество кликов (точно значение)

clicks_number
109178

Число уникальных объявлений по дням

date	uniq_ad_id
2019-04-01	150
2019-04-02	344
2019-04-03	360
2019-04-04	407
2019-04-05	465
2019-04-06	220

Больше всего уникальных объявлений – 2019-04-05, меньше всего – 2019-04-01.

Число уникальных кампаний по дням

date	uniq_campaign
2019-04-01	149
2019-04-02	336
2019-04-03	352
2019-04-04	396
2019-04-05	442
2019-04-06	212

Больше всего уникальных кампаний – 2019-04-05, меньше всего – 2019-04-01.

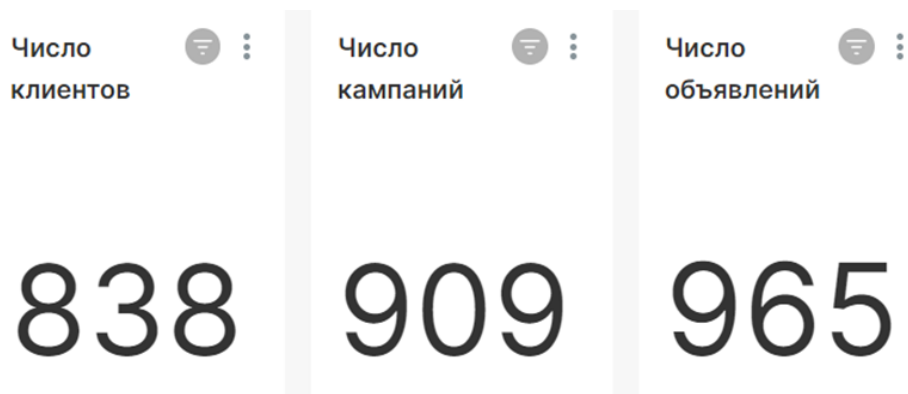
Также число уникальных объявлений и уникальных кампаний по дням отображено с помощью графика на дашборде «Основные продуктовые метрики и статистика по дням»:

Число уникальных объявлений и уникальных кампаний по дням

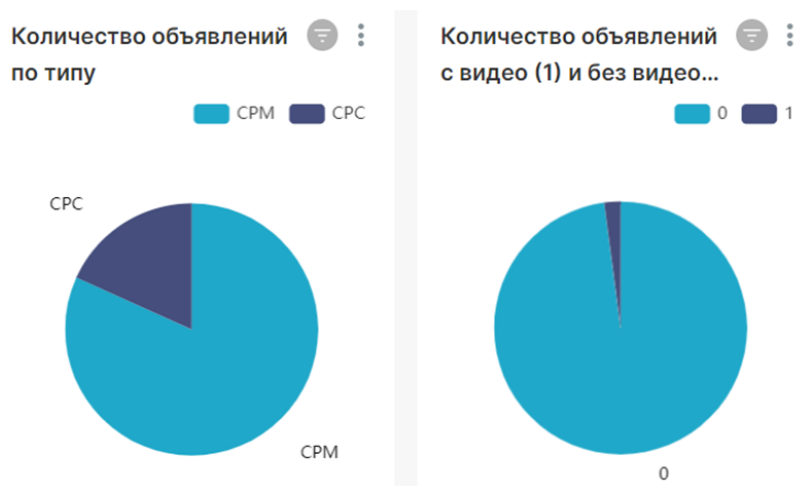


В качестве основных метрик на дашборд выведены следующие виджеты:

- Общее количество показов
- Общее количество кликов
- Конверсия в клик (CR)



Дополнительно на дашборд выведены доли объявлений по типу (CPM и CPC) и доли объявлений с видео и без видео в виде круговых диаграмм:



Количество CPM объявлений превышает количество CPC объявлений: 789 (82%) против 176 (18%).

Количество объявлений без видео значительно перевешивает количество объявлений с видео: 945 (98%) против 20 (2%).

2. Разобраться, почему случился такой скачок 2019-04-05? Каких событий стало больше? У всех объявлений или только у некоторых?

Скачок 2019-04-05: топ 10-объявлений по количеству всех событий

ad_id	events_number
112583	393828
107729	29745
28142	20903
38892	8437
107837	8341
37720	4944
45008	3145
18425	2313
46629	2305
42518	1942

Скачок случился за счет объявления 112583. Количество событий по нему намного больше, чем у прочих (в 13 раз больше, чем у ближайшего по количеству событий объявления).

Отображение на дашборде «Аномалии и баги рекламной площадки»:

Скачок 2019-04-05: топ 10-объявлений по количеству всех событий

ad_id	events_number
112583	393828
107729	29745
28142	20903
38892	8437
107837	8341
37720	4944
45008	3145
18425	2313

Скачок 2019-04-05: количество показов и кликов по топ-3 объявлениям

ad_id	event	events_number
112583	view	302811
112583	click	91017
107729	view	29724
107729	click	21
28142	view	20872
28142	click	31

У объявления 112583 больше как кликов, так и показов, по сравнению с прочими объявлениями (CTR = 0,3).

Также на дашборд «Аномалии и баги рекламной площадки» выведена диаграмма с распределением кликов по платформам:



Диаграмма показывает, что случился заметный пик кликов 5 апреля 2019 г., причем пик наблюдается на всех платформах. Аналогичную диаграмму можно было сделать и для показов (пик случился не только для кликов, но и для показов).

3. Найти топ 10 объявлений по CTR за все время. CTR — это отношение всех кликов объявлений к просмотрам. Например, если у объявления было 100 показов и 2 клика, $CTR = 0.02$. Различается ли средний и медианный CTR объявлений в наших данных?

На дашборде «CTR и финансовый отчет» представлены топ-10 объявлений по CTR (от 500 показов) с их характеристиками:

Характеристики топ-10 объявлений по CTR (объявления от 500 показов)

ad_id	has_video	ad_cost_type	ad_cost	platform_freq	views_count	clicks_count	CTR
112583	0	CPM	210	android	351802	105767	0.301
38892	0	CPM	212	android	9323	516	0.055
4369	0	CPM	195	android	584	27	0.046
10794	0	CPM	190	android	526	22	0.042
20916	0	CPM	208	android	512	20	0.039
14617	0	CPC	40	android	596	22	0.037
18087	0	CPM	204	android	661	23	0.035
46723	0	CPM	187	android	639	21	0.033
111335	0	CPM	193	android	1081	35	0.032
114645	0	CPM	200	android	720	23	0.032

Максимальный CTR – у объявления 112583 – того же, по которому случился пик кликов и показов 5 апреля.

Все объявления с максимальным CTR не содержат видео. Большинство из них – типа CPM (только одно из топ-10 – типа CPC).

Чаще всего объявления с наибольшим CTR показывались на платформе Android.

Средний CTR объявлений

average_ctr
0.0158

Медианный CTR объявлений

median_ctr
0.0029

Медианный CTR меньше среднего примерно в 5,5 раз. Это означает, что выборка смещена в сторону относительно немногочисленных объявлений с высоким CTR, в то время как половина объявлений имеет CTR значительно ниже среднего.

Также средний и медианный CTR объявлений выведены на дашборд «CTR и финансовый отчет» с помощью виджетов:

Средний CTR всех объявлений

0.0158

⋮

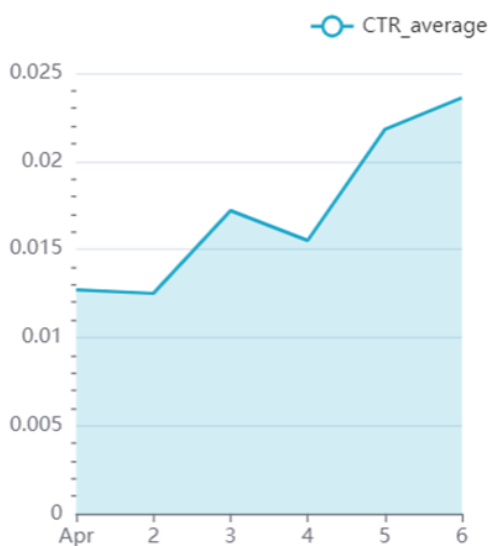
Медианный CTR всех объявлений

0.0029

На дашборд дополнительно выведена аналитика CTR объявлений по дням:

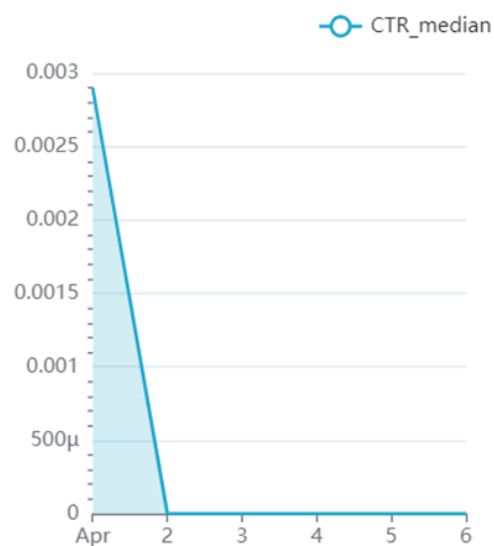
Средний CTR по дням

⋮



Медианный CTR по дням

⋮



По среднему CTR наблюдается поступательный рост. По медианному CTR – наоборот: начиная со 2 апреля он достигает 0. Т.е. как минимум половина объявлений за эти даты – с нулевым CTR.

4. Похоже, в наших логах есть баг, объявления приходят с кликами, но без показов! Сколько таких объявлений, есть ли какие-то закономерности? Эта проблема наблюдается на всех платформах?

Количество объявлений с кликами, но без показов

ad_id_number
9

Характеристики объявлений с кликами, но без показов

ad_id	clicks	views	has_video	ad_cost_type
120431	35	0	0	CPM
41500	20	0	0	CPM
19223	7	0	0	CPM
117364	7	0	0	CPC
26204	6	0	0	CPM
120536	6	0	0	CPC
115825	4	0	0	CPM
45418	3	0	0	CPM
120796	1	0	0	CPM

Во всех объявлениях с кликами, но без показов, нет видео. Преобладают CPM-объявления (7 из 9).

Распределение объявлений с кликами, но без показов, по платформам

platform	ad_id_number
ios	8
android	7
web	5

Проблема наблюдается на всех платформах. Больше всего – 8 из 9 объявлений – было на платформе IOS.

Распределение кликов без показов по платформам

platform	clicks_number
android	47
ios	30
web	12

Больше всего кликов без показов – на платформе Android.

Распределение кликов без показов по датам

date	clicks_number
2019-04-01	72
2019-04-02	17

Клики без показов были 2 дня, больше – 2019-04-01.

Пример отображения той же таблицы на дашборде «Аномалии и баги рекламной площадки»:

Распределение кликов без показов по датам ⋮

date ⌵	clicks_number ⌵
2019-04-01	72
2019-04-02	17

5. Есть ли различия в CTR у объявлений с видео и без? А чему равняется 95 перцентиль CTR по всем объявлениям за 2019-04-04?

Топ-10 объявлений по CTR, у которых есть видео

ad_id	clicks	views	CTR	has_video
43522	6	64	0.09	1
43837	1	15	0.07	1
43565	4	60	0.07	1
37422	2	41	0.05	1
31336	5	163	0.03	1
41059	4	213	0.02	1
39715	1	41	0.02	1
26269	4	763	0.01	1
38376	24	1841	0.01	1
23829	1	82	0.01	1

Топ-10 объявлений по CTR, у которых нет видео

ad_id	clicks	views	CTR	has_video
117164	6	19	0.32	0
112583	105767	351802	0.3	0
42507	3	11	0.27	0
98569	3	16	0.19	0
23599	4	24	0.17	0
46639	44	253	0.17	0
19912	4	25	0.16	0
110414	5	32	0.16	0
45969	2	13	0.15	0
20662	4	26	0.15	0

Отображение на дашборде «CTR и финансовый отчет» топ объявлений с видео от 100 показов:

**Топ объявлений по CTR с видео (объявления от 100 :
показов)**

ad_id ↕	video ↕	views_count ↕	clicks_count ↕	CTR ↕
31336	1	163	5	0.031
41059	1	213	4	0.019
117353	1	288	4	0.014
38376	1	1841	24	0.013
31491	1	273	2	0.007
26269	1	763	4	0.005
44949	1	667	1	0.001

В десятку объявлений с наиболее высоким показателем CTR входят объявления без видео. Если брать наиболее высокий показатель CTR у объявлений с видео, то он достигает 0,03 у объявления, которое показали

хотя бы 100 раз, и 0,09 без такого ограничения. В то же время наиболее высокий показатель CTR у объявлений без видео достигает 0,3 (объявление с хотя бы 500 показами) и 0,302 (у объявления без ограничения количества показов).

Средний CTR объявлений с видео

average_ctr
0.0201

Средний CTR объявлений без видео

average_ctr
0.0157

При этом средний CTR у объявлений с видео немного выше, чем у объявлений без видео.

95-перцентиль CTR по всем объявлениям за 2019-04-04³

quantile_ctr
0.0821

У 5% объявлений за 4 апреля CTR выше 0,08, у остальных – ниже.

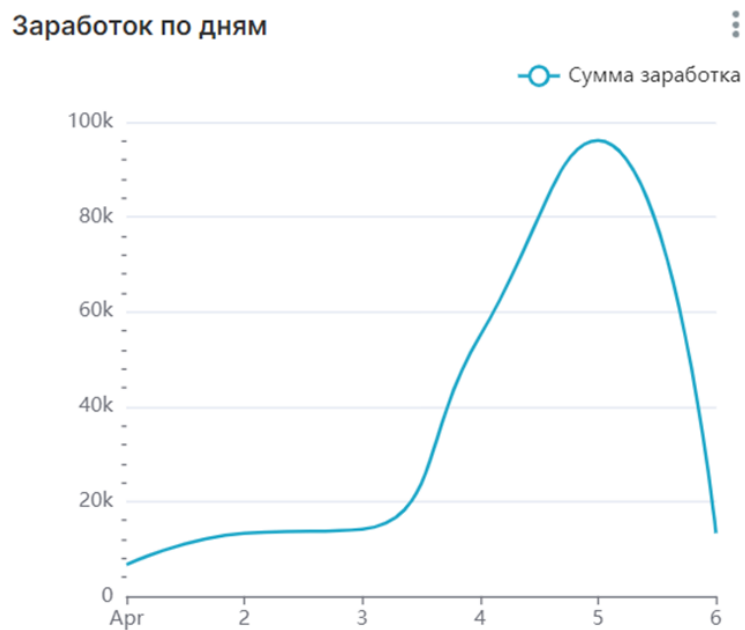
6. Для финансового отчета нужно рассчитать наш заработок по дням. В какой день мы заработали больше всего? В какой меньше? Мы списываем с клиентов деньги, если произошел клик по CPC объявлению, и мы списываем деньги за каждый показ CPM объявления, если у CPM объявления цена - 200 рублей, то за один показ мы зарабатываем $200 / 1000$.

Заработок по дням

date	money_all
2019-04-01	6656
2019-04-02	13286
2019-04-03	14146
2019-04-04	54989
2019-04-05	96123
2019-04-06	13347

Визуализация заработка по дням на дашборде «CTR и финансовый отчет»:

³
Данная информация не была выведена на дашборд, т.к. не является достаточно значимой.



Дни, в которые заработали больше и меньше всего

date_max_money	date_min_money
2019-04-05	2019-04-01

Сумма заработанных на площадке денег

money_all
198546

Общий заработок за все дни также отображен на дашборде:

Сумма заработанных на площадке
денег

199k

На дашборде также есть отображение в виде таблицы топ-10 объявлений, по которым заработали больше всего денег.

Топ-10 объявлений, по которым заработали больше всего денег

ad_id	money_sum
112583	73914
107729	35239
107837	10456
28142	4826
45008	3928
42518	3774
39191	3728
113912	3478
116718	2323
38892	1975

Наиболее выгодным объявлением оказалось № 112583 – то же объявление, которое показало рекорд показов.

Распределение совокупного заработка по типу объявлений

ad_cost_type	money_all
CPM	171235
CPC	27311

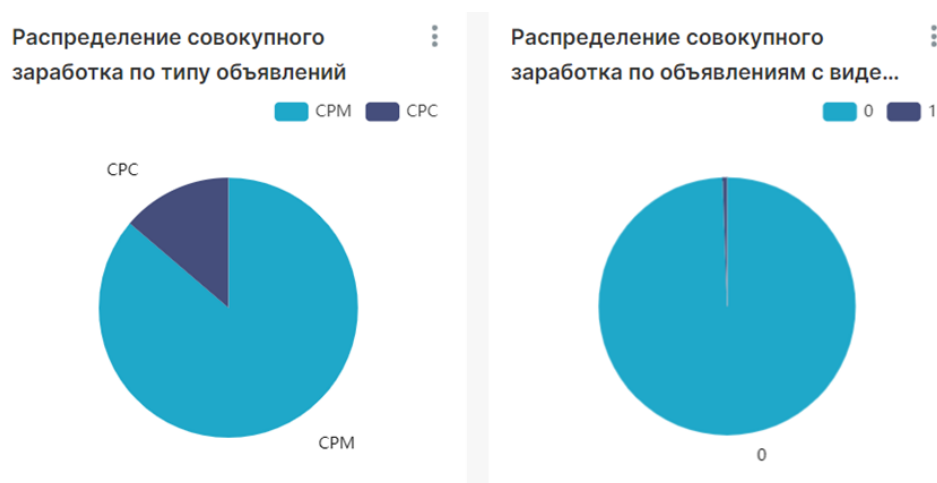
На CPM объявлениях было заработано в несколько раз больше, чем на CPC объявлениях.

Распределение совокупного заработка по объявлениям с видео и без видео

has_video	money_all
0	197454
1	1092

«0» означает, что у объявления нет видео, «1» означает, что есть. Намного больше было заработано на объявлениях без видео.

Также распределение заработка по типам объявлений отображено на дашборде:



На CPM объявлениях заработано 86% денег, на CPC-объявлениях – 14%. И 99% заработка принесли объявления без видео (их большинство).

Распределение совокупного заработка по платформам (отображение на дашборде):



50% заработка дали показы объявлений на платформе Android, 30% - на платформе iOS, 20% - на Web.

Дополнительно на дашборде приведены данные по средней стоимости объявления, минимальной и максимальной стоимости, а также по средней таргетируемой аудитории:

Средняя стоимость объявления

172.46

Минимальная стоимость
объявления

33.2

⋮

Максимальная стоимость
объявления

227.5

⋮

Средний размер аудитории, на
которую таргетируется объявление

1.07M

7. На какой платформе больше всего показов? Сколько процентов показов приходится на каждую из платформ?

Платформа с максимальным количеством показов

platform_max_views
android

Распределение показов по платформам

platform	views_number	percent
android	445722	50.03
ios	267117	29.99
web	177983	19.98

50% показов приходится на платформу Android, 30% показов – на платформу IOS, 20% показов – на платформу Web.

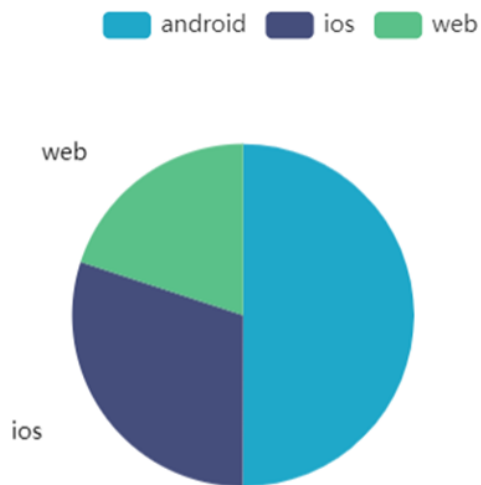
Распределение кликов по платформам

platform	clicks_number	percent
android	54634	50.04
ios	32862	30.1
web	21682	19.86

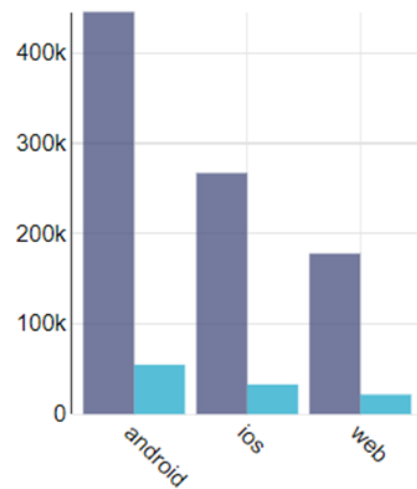
50% кликов также приходится на платформу Android.

Диаграммы на дашборде «Основные продуктовые метрики и статистика по дням» показывают распределение всех событий (показов и кликов) по платформам, а также соотношение кликов и показов по платформам:

Распределение всех
событий (показов и...



Распределение показов
и кликов по платформам



8. А есть ли такие объявления, по которым сначала произошёл клик, а только потом показ?

Объявления, по которым сначала произошёл клик, а только потом показ

ad_id	first_event
18681	click
23599	click
32386	click
33033	click
36758	click
38224	click
44283	click
44766	click
46639	click
98569	click
107798	click
114886	click

Количество объявлений, по которым сначала произошёл клик, а только потом показ

number
12

Такие объявления действительно есть, их 12, – это аномалия рекламной площадки, с которой нужно разбираться дополнительно, потому что так не должно быть (как и объявлений с кликами, но без показов).

Заключение

В представленной работе была проведена аналитика эффективности рекламной площадки социальной сети (на синтезированных данных, похожих на реальные данные). С использованием датасета (1 млн строк) и кастомных SQL-запросов в ClickHouse были построены три дашборда в системе Superset, отвечающих на все 8 поставленных вопросов к аналитику рекламы, а также отражающих основные и дополнительные метрики: «Основные продуктовые метрики и статистика по дням», «CTR и финансовый отчет» и «Аномалии и баги рекламной площадки». На дашбордах были использованы различные типы визуализации: “big number”, “time-area chart”, “pie chart”, “table”, “histogram” и др. Каждый тип визуализации был подобран для оптимального и наиболее информативного отображения тех или иных данных на дашборде. Для дашборда «Основные продуктовые метрики и статистика по дням» были добавлены фильтры: по датам, типам объявлений и платформам.

Выводы, полученные на данных:

1. Конверсия кликов в показ на рекламной площадке в целом составила 12%, тогда как средний CTR по объявлениям – 1,6%.
2. Платформа, по которой было больше всего кликов и показов, а также которая принесла больше всего выручки – Android (50%), следом за ней – iOS (30%) и только потом Web (20%).
3. 5 апреля произошел резкий скачок числа показов и кликов за счет объявления 112 583. По нему же был выявлен наиболее высокий показатель CTR (0,3, или 30%) и оно же принесло наибольшую выручку.
4. CPM объявлений больше по количеству, чем CPC объявлений, и они дали большую выручку.

5. Объявлений без видео больше по количеству и они дали большую выручку (99%). В то же время, средний CTR по объявлениям с видео немного выше, чем по объявлениям без видео.
6. На рекламной площадке есть аномалии и баги: помимо скачка по дате, мы увидели, что по 9 объявлениям не было показов, но были клики, а по 12 объявлениям произошел сначала клик, потом показ.

В качестве идей для дальнейшей аналитики данных можно предложить провести более глубокий анализ причин (например, резкого скачка кликов и показов по объявлению в определенную дату, появления объявлений с кликами, но без показов). Для этого необходимы дополнительные данные.

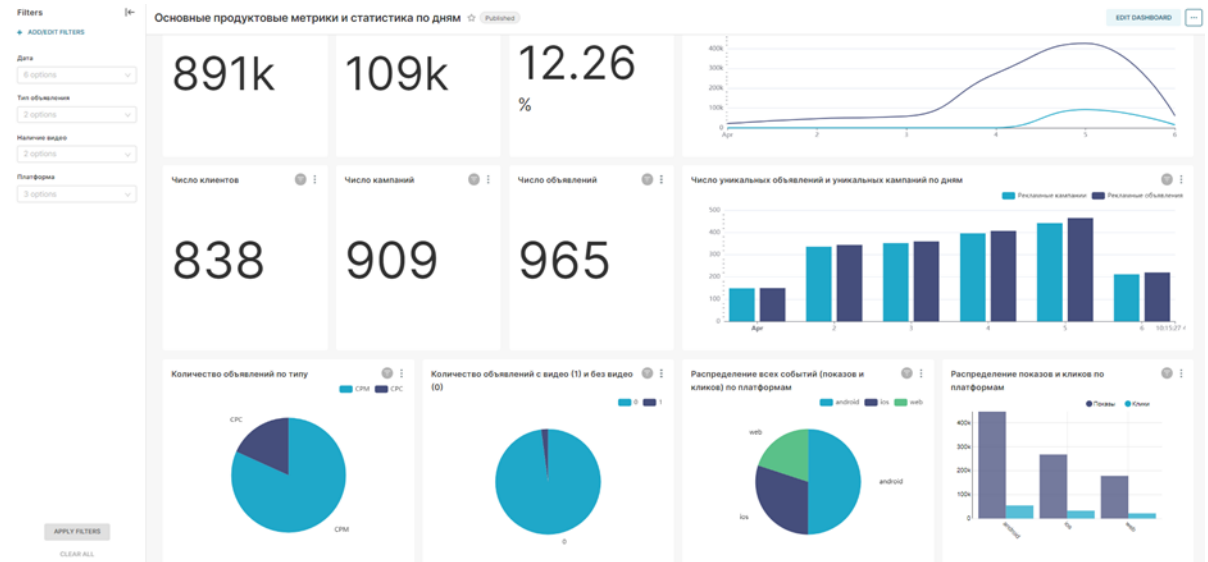
Можно проводить более детальный анализ текущего датасета, рассматривать больше параметров во взаимосвязи, сделать статистический анализ. Например, сделать более детальный анализ по датам и по времени, корреляционный анализ для заработанных денег на рекламной площадке и показателей CTR и др.

Также для расчета дополнительных метрик, рассмотренных в обзоре литературы, можно обогатить датасет дополнительными данными, такими как лиды, заказы, количество заинтересованных пользователей, количество отказов и т.д.

Приложение 1. Дашборды в Superset

1) Основные продуктовые метрики и статистика по дням:

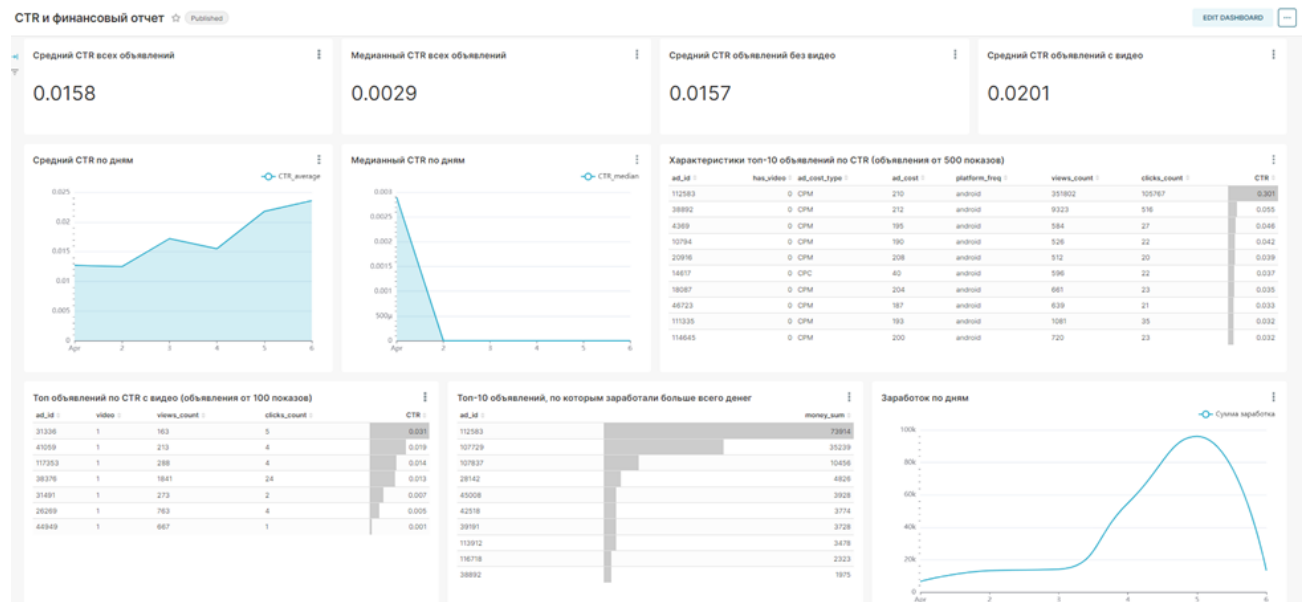
<http://146.185.243.0:8088/superset/dashboard/p/ZkgoNppArqJ/>



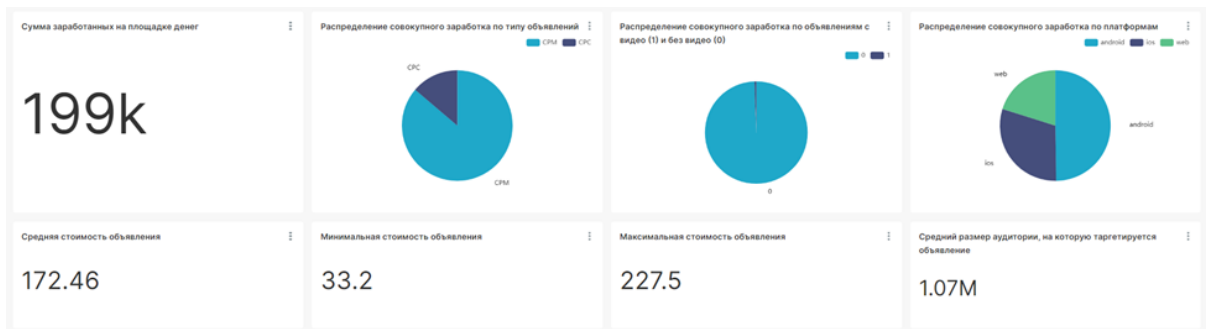
2) CTR и финансовый отчет:

<http://146.185.243.0:8088/superset/dashboard/p/Qlqy2YdoKx2/>

Ч. 1 (верхняя):

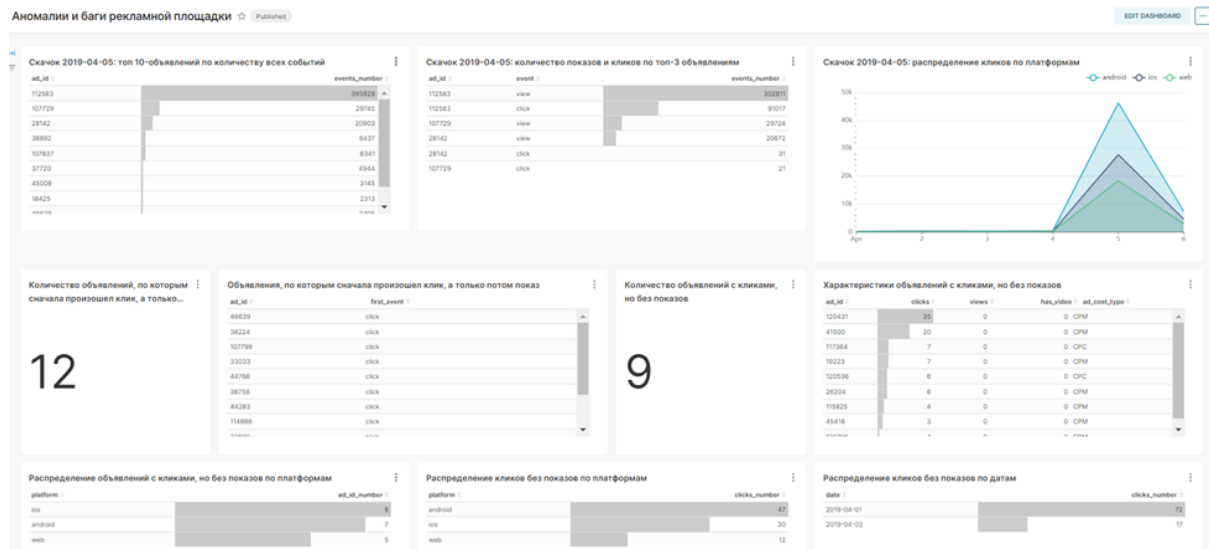


Ч. 2 (нижняя):



3) Аномалии и баги рекламной площадки:

<http://146.185.243.0:8088/superset/dashboard/p/nmMDOM5yBer/>



Приложение 2. Код SQL-запросов

1. Получить статистику по дням. Посчитать число всех событий по дням, число показов, число кликов, число уникальных объявлений и уникальных кампаний.

Число всех событий по дням

```
SELECT date,  
  
COUNT(event) AS events_number  
  
FROM ads_data  
  
GROUP BY date  
  
ORDER BY date
```

Число показов по дням

```
SELECT date, event,  
  
COUNT(event) as events_number  
  
FROM ads_data  
  
WHERE event = 'view'  
  
GROUP BY date, event  
  
ORDER BY date
```

Число кликов по дням

```
SELECT date, event,  
  
COUNT(event) as events_number  
  
FROM ads_data  
  
WHERE event = 'click'  
  
GROUP BY date, event  
  
ORDER BY date
```

Общее количество показов

```
SELECT COUNT(event) views_number  
  
FROM ads_data  
  
WHERE event = 'view'
```

Общее количество кликов

```
SELECT COUNT(event) clicks_number  
  
FROM ads_data  
  
WHERE event = 'click'
```

Конверсия в клик (CR)

```
WITH new_table AS
```

```
(SELECT SUM (event = 'click') as clicks,  
  
SUM (event = 'view') as views,  
  
ROUND(clicks/views, 2) as CTR  
  
FROM ads_data  
  
HAVING views != 0)  
  
SELECT CTR * 100 CR_percent  
  
FROM new_table
```

Число уникальных объявлений по дням

```
SELECT date,  
  
COUNT(DISTINCT ad_id) uniq_ad_id  
  
FROM ads_data  
  
GROUP BY date  
  
ORDER BY date
```

Число уникальных кампаний по дням

```
SELECT date,  
  
uniqExact(campaign_union_id) uniq_campaign  
  
FROM ads_data
```

GROUP BY date

ORDER BY date

2. Разобраться, почему случился такой скачок 2019-04-05? Каких событий стало больше? У всех объявлений или только у некоторых?

Скачок 2019-04-05: топ 10-объявлений по количеству всех событий

```
SELECT ad_id,  
  
COUNT(event) events_number  
  
FROM ads_data  
  
WHERE date = '2019-04-05'  
  
GROUP BY ad_id  
  
ORDER BY events_number DESC  
  
LIMIT 10
```

Скачок 2019-04-05: количество показов и кликов по топ-3 объявлениям

```
SELECT ad_id, event, COUNT(event) events_number  
  
FROM ads_data  
  
WHERE (date = '2019-04-05' AND ad_id = '112583')  
  
OR (date = '2019-04-05' AND ad_id = '107729')
```

```
OR (date = '2019-04-05' AND ad_id = '28142')
```

```
GROUP BY ad_id, event
```

```
ORDER BY ad_id DESC, events_number DESC
```

3. Найти топ 10 объявлений по CTR за все время. CTR — это отношение всех кликов объявлений к просмотрам. Например, если у объявления было 100 показов и 2 клика, $CTR = 0.02$. Различается ли средний и медианный CTR объявлений в наших данных?

Характеристики топ-10 объявлений по CTR (объявления от 500 показов)

```
SELECT DISTINCT toString(ad_id) ad_id,
```

```
has_video, ad_cost_type,
```

```
toString(ROUND(ad_cost)) AS ad_cost,
```

```
anyHeavy(platform) AS platform_freq,
```

```
toString(SUM(event = 'view') AS views) AS views_count,
```

```
toString(SUM(event = 'click') AS clicks) AS clicks_count,
```

```
ROUND(clicks/views, 3) AS CTR
```

```
FROM ads_data
```

```
GROUP BY ad_id, has_video, ad_cost_type, ad_cost
```

```
HAVING views >= 500
```

ORDER BY CTR DESC

LIMIT 10

Средний CTR объявлений

WITH result AS

(SELECT DISTINCT ad_id,

SUM (event = 'click') as clicks,

SUM (event = 'view') as views,

clicks/views as CTR

FROM ads_data

GROUP BY ad_id

HAVING views != 0)

SELECT ROUND(AVG(CTR), 4) average_ctr

FROM result

Медианный CTR объявлений

WITH result AS

(SELECT DISTINCT ad_id,

SUM (event = 'click') as clicks,

```
SUM (event = 'view') as views,  
  
clicks/views as CTR  
  
FROM ads_data  
  
GROUP BY ad_id  
  
HAVING views != 0  
  
ORDER BY CTR DESC)  
  
SELECT ROUND(median(CTR), 4) median_ctr  
  
FROM result
```

Средний CTR объявлений по дням

```
WITH result AS  
  
(SELECT DISTINCT ad_id, date,  
  
SUM (event = 'click') as clicks,  
  
SUM (event = 'view') as views,  
  
clicks/views as CTR  
  
FROM ads_data  
  
GROUP BY ad_id, date  
  
HAVING views != 0)  
  
SELECT date, ROUND(AVG(CTR), 4) CTR_average
```

FROM result

GROUP BY date

Медианный CTR объявлений по дням

WITH result AS

(SELECT DISTINCT ad_id, date,

SUM (event = 'click') as clicks,

SUM (event = 'view') as views,

clicks/views as CTR

FROM ads_data

GROUP BY ad_id, date

HAVING views != 0)

SELECT date, ROUND(median(CTR), 4) CTR_median

FROM result

GROUP BY date

4. Похоже, в наших логах есть баг, объявления приходят с кликами, но без показов! Сколько таких объявлений, есть ли какие-то закономерности? Эта проблема наблюдается на всех платформах?

Количество объявлений с кликами, но без показов


```
WITH new_table AS

(SELECT DISTINCT ad_id,

SUM (event = 'click') as clicks,

SUM (event = 'view') as views

FROM ads_data

GROUP BY ad_id

HAVING views = 0)

SELECT COUNT(ad_id) ad_id_number

FROM new_table
```

Характеристики объявлений с кликами, но без показов

```
SELECT DISTINCT ad_id,

SUM (event = 'click') as clicks,

SUM (event = 'view') as views,

has_video, ad_cost_type

FROM ads_data

GROUP BY ad_id, has_video, ad_cost_type

HAVING views = 0

ORDER BY clicks DESC
```

Распределение объявлений с кликами, но без показов, по платформам

```
WITH new_table AS

(SELECT DISTINCT ad_id,

SUM (event = 'click') as clicks,

SUM (event = 'view') as views,

has_video, ad_cost_type, platform

FROM ads_data

GROUP BY ad_id, has_video, ad_cost_type, platform

HAVING views = 0)

SELECT platform, COUNT(ad_id) ad_id_number

FROM new_table

GROUP BY platform

ORDER BY ad_id_number DESC
```

Распределение кликов без показов по платформам

```
WITH new_table AS

(SELECT DISTINCT ad_id,

SUM (event = 'click') as clicks,
```

```
SUM (event = 'view') as views,  
  
has_video, ad_cost_type, platform  
  
FROM ads_data  
  
GROUP BY ad_id, has_video, ad_cost_type, platform  
  
HAVING views = 0)  
  
SELECT platform, SUM(clicks) clicks_number  
  
FROM new_table  
  
GROUP BY platform  
  
ORDER BY clicks_number DESC
```

Распределение кликов без показов по датам

```
WITH new_table AS  
  
(SELECT DISTINCT ad_id,  
  
SUM (event = 'click') as clicks,  
  
SUM (event = 'view') as views,  
  
has_video, ad_cost_type, date  
  
FROM ads_data  
  
GROUP BY ad_id, has_video, ad_cost_type, date  
  
HAVING views = 0)
```

```
SELECT date, SUM(clicks) clicks_number
```

```
FROM new_table
```

```
GROUP BY date
```

5. Есть ли различия в CTR у объявлений с видео и без? А чему равняется 95 процентиль CTR по всем объявлениям за 2019-04-04?

Топ-10 объявлений по CTR, у которых есть видео

```
SELECT DISTINCT ad_id,
```

```
SUM (event = 'click') as clicks,
```

```
SUM (event = 'view') as views,
```

```
ROUND(clicks/views, 2) as CTR, has_video
```

```
FROM ads_data
```

```
WHERE has_video = 1
```

```
GROUP BY ad_id, has_video
```

```
HAVING views != 0
```

```
ORDER BY CTR DESC
```

```
LIMIT 10
```

Топ-10 объявлений по CTR, у которых нет видео

```
SELECT DISTINCT ad_id,  
  
SUM (event = 'click') as clicks,  
  
SUM (event = 'view') as views,  
  
ROUND(clicks/views, 2) as CTR, has_video  
  
FROM ads_data  
  
WHERE has_video = 0  
  
GROUP BY ad_id, has_video  
  
HAVING views != 0  
  
ORDER BY CTR DESC  
  
LIMIT 10
```

Топ объявлений по CTR с видео (объявления от 100 показов)

```
SELECT DISTINCT toString(ad_id) ad_id,  
  
has_video,  
  
toString(SUM(event = 'view') AS views) AS views_count,  
  
toString(SUM(event = 'click') AS clicks) AS clicks_count,  
  
ROUND(clicks/views, 3) AS CTR  
  
FROM ads_data
```

WHERE has_video = 1

GROUP BY ad_id, has_video

HAVING views >= 100 AND clicks > 0

ORDER BY CTR DESC

LIMIT 10

Средний CTR объявлений с видео

WITH result AS

(SELECT DISTINCT ad_id,

SUM (event = 'click') as clicks,

SUM (event = 'view') as views,

clicks/views as CTR

FROM ads_data

WHERE has_video = 1

GROUP BY ad_id

HAVING views != 0

ORDER BY CTR DESC)

SELECT ROUND(AVG(CTR), 4) average_ctr

FROM result

Средний CTR объявлений без видео

WITH result AS

(SELECT DISTINCT ad_id,

SUM (event = 'click') as clicks,

SUM (event = 'view') as views,

clicks/views as CTR

FROM ads_data

WHERE has_video = 0

GROUP BY ad_id

HAVING views != 0

ORDER BY CTR DESC)

SELECT ROUND(AVG(CTR), 4) average_ctr

FROM result

95 процентиль CTR по всем объявлениям за 2019-04-04

WITH result AS

(SELECT DISTINCT ad_id,

SUM (event = 'click') as clicks,

```
SUM (event = 'view') as views,  
  
clicks/views as CTR  
  
FROM ads_data  
  
WHERE date = '2019-04-04'  
  
GROUP BY ad_id  
  
HAVING views != 0  
  
ORDER BY CTR DESC)  
  
SELECT ROUND(quantile(0.95)(CTR), 4) quantile_ctr  
  
FROM result
```

6. Для финансового отчета нужно рассчитать наш заработок по дням. В какой день мы заработали больше всего? В какой меньше? Мы списываем с клиентов деньги, если произошел клик по CPC объявлению, и мы списываем деньги за каждый показ CPM объявления, если у CPM объявления цена - 200 рублей, то за один показ мы зарабатываем $200 / 1000$.

Заработок по дням

```
SELECT date,  
  
round(  
  
plus(  
  
sumIf(ad_cost, event = 'click' AND ad_cost_type = 'CPC'),
```



```
sumIf(ad_cost/1000, event = 'view' AND ad_cost_type = 'CPM')
```

```
),
```

```
0) AS money_all
```

```
FROM ads_data
```

```
GROUP BY date
```

```
ORDER BY date
```

Дни, в которые заработали больше и меньше всего

```
WITH result AS
```

```
(SELECT date,
```

```
round(
```

```
plus(
```

```
sumIf(ad_cost, event = 'click' AND ad_cost_type = 'CPC'),
```

```
sumIf(ad_cost/1000, event = 'view' AND ad_cost_type = 'CPM')
```

```
),
```

```
0) AS money_all
```

```
FROM ads_data
```

```
GROUP BY date)
```

```
SELECT argMax(date, money_all) date_max_money,
```

```
argMin(date, money_all) date_min_money
```

```
FROM result
```

Сумма заработанных на площадке денег

```
SELECT round(
```

```
plus(
```

```
sumIf(ad_cost, event = 'click' AND ad_cost_type = 'CPC'),
```

```
sumIf(ad_cost/1000, event = 'view' AND ad_cost_type = 'CPM')
```

```
),
```

```
0) AS money_all
```

```
FROM ads_data
```

Топ-10 объявлений, по которым заработали больше всего денег

```
SELECT DISTINCT toString(ad_id) ad_id,
```

```
round(
```

```
plus(
```

```
sumIf(ad_cost, event = 'click' AND ad_cost_type = 'CPC'),
```

```
sumIf(ad_cost/1000, event = 'view' AND ad_cost_type = 'CPM')
```

```
),
```

```
0) AS money_sum  
  
FROM ads_data  
  
GROUP BY ad_id  
  
ORDER BY money_sum DESC  
  
LIMIT 10
```

Распределение совокупного заработка по типу объявлений

```
SELECT ad_cost_type, round(  
  
plus(  
  
sumIf(ad_cost, event = 'click' AND ad_cost_type = 'CPC'),  
  
sumIf(ad_cost/1000, event = 'view' AND ad_cost_type = 'CPM')  
  
),
```

```
0) AS money_all  
  
FROM ads_data  
  
GROUP BY ad_cost_type
```

Распределение совокупного заработка по объявлениям с видео и без видео

```
SELECT has_video, round(  
  
plus(  
  

```

```
sumIf(ad_cost, event = 'click' AND ad_cost_type = 'CPC'),  
sumIf(ad_cost/1000, event = 'view' AND ad_cost_type = 'CPM')  
,  
0) AS money_all  
FROM ads_data  
GROUP BY has_video
```

Распределение совокупного заработка по платформам

```
SELECT platform, round(  
plus(  
sumIf(ad_cost, event = 'click' AND ad_cost_type = 'CPC'),  
sumIf(ad_cost/1000, event = 'view' AND ad_cost_type = 'CPM')  
,  
0) AS money_all  
FROM ads_data  
GROUP BY platform
```

Средний размер аудитории, на которую таргетируется объявление

```
SELECT DISTINCT ad_id, target_audience_count
```

FROM ads_data

GROUP BY ad_id, target_audience_count

7. На какой платформе больше всего показов? Сколько процентов показов приходится на каждую из платформ?

Платформа с максимальным количеством показов

WITH new_table AS

(SELECT platform, COUNT(event) as events_number

FROM ads_data

WHERE event = 'view'

GROUP BY platform)

SELECT argMax(platform, events_number) platform_max_views

FROM new_table

Распределение показов по платформам

WITH

(SELECT COUNT(event)

FROM ads_data

WHERE event = 'view') AS views_all

SELECT platform,

```
COUNT(event) as views_number,  
  
ROUND(views_number / views_all * 100, 2) AS percent  
  
FROM ads_data  
  
WHERE event = 'view'  
  
GROUP BY platform  
  
ORDER BY views_number DESC
```

Распределение кликов по платформам

```
WITH  
  
(SELECT COUNT(event)  
  
FROM ads_data  
  
WHERE event = 'click') AS clicks_all  
  
SELECT platform,  
  
COUNT(event) as clicks_number,  
  
ROUND(clicks_number / clicks_all * 100, 2) AS percent  
  
FROM ads_data  
  
WHERE event = 'click'  
  
GROUP BY platform  
  
ORDER BY clicks_number DESC
```

8. А есть ли такие объявления, по которым сначала произошёл клик, а только потом показ?

Объявления, по которым сначала произошёл клик, а только потом показ

```
WITH new_table AS  
  
(SELECT ad_id,  
  
argMin(event, time) AS first_event,  
  
arrayUniq(groupArray(event)) AS number_of_events  
  
FROM ads_data  
  
GROUP BY ad_id  
  
HAVING first_event = 'click'  
  
AND number_of_events > 1)  
  
SELECT ad_id, first_event  
  
FROM new_table  
  
ORDER BY ad_id
```

Количество объявлений, по которым сначала произошёл клик, а только потом показ

```
WITH new_table AS
```

```
(SELECT ad_id,  
  
argMin(event, time) AS first_event,  
  
arrayUniq(groupArray(event)) AS number_of_events  
  
FROM ads_data  
  
GROUP BY ad_id  
  
HAVING first_event = 'click'  
  
AND number_of_events > 1)  
  
SELECT COUNT(ad_id) number  
  
FROM new_table
```