

¼ Project (Data Analysis & ClickHouse)

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

```
select distinct(date), count(event) as events_number,  
countIf(event='click') as clicks_number,  
countIf(event='view') as views_number,  
uniqExact(ad_id) as unique_ads,  
uniqExact(campaign_union_id) as  
unique_campaign  
from ads_data  
group by date;
```

Ответ: таблица со статистикой по дням.

distinct(date)	events_number	clicks_number	views_number	unique_ads	unique_campaign
2019-04-01	22073	291	21782	150	149
2019-04-02	47117	545	46572	344	336
2019-04-03	59483	460	59023	360	352
2019-04-04	275735	643	275092	407	396
2019-04-05	519707	92321	427386	465	442
2019-04-06	75885	14918	60967	220	212

2. Разобраться, почему случился такой скачок 2019-04-05? Каких событий стало больше?

```
select countIf(event='click') as clicks_number,  
countIf(event='view') as views_number  
from ads_data  
where date = '2019-04-05';
```

Ответ: 2019-04-05 увеличилось число просмотров и кликов.

Date	clicks_number	views_number
2019-04-05	92321	427386

У всех объявлений или только у некоторых?

```
select distinct(ad_id), countIf(event='click') as
clicks_number, countIf(event='view') as views_number
from ads_data
where date = '2019-04-05'
group by ad_id
order by clicks_number desc, views_number desc
limit 5;
```

distinct(ad_id)	clicks_number	views_number
112583	91017	302811
38892	451	7986
37720	83	4861
42518	57	1885
18425	39	2274

Ответ: не у всех, только у некоторых, например, есть ad_id 112583, у которого 91017 кликов и 302811 просмотров.

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, $CTR = 0.02$.

```
select distinct(ad_id), countIf(event='click') as
clicks_number, countIf(event='view') as views_number,
round(clicks_number/views_number, 2) as CTR
from ads_data
group by ad_id
```

```
having views_number!=0
order by CTR desc
limit 5;
```

Ответ: чтобы CTR не получался infinity, добавлено условие, что число показов не должно быть равным нулю, CTR округлен до второго знака после запятой.

distinct(ad_id)	clicks_number	views_number	CTR
117164	6	19	0.32
112583	105767	351802	0.3
42507	3	11	0.27
98569	3	16	0.19
23599	4	24	0.17

Различается ли средний и медианный CTR объявлений в наших данных?

```
select median(CTR), round(avg(CTR), 2)
from (select distinct(ad_id), countIf(event='click') as
clicks_number, countIf(event='view') as views_number,
round(clicks_number/views_number, 2) as CTR
from ads_data
group by ad_id
having views_number!=0);
```

Ответ: да, различается — медианный равен 0, а средний равен 0,02.

4. Похоже, в наших логах есть баг, объявления приходят с кликами, но без показов! Сколько таких объявлений, есть ли какие-то закономерности

```
select sum(ads_number), sum(clicks_number),
sum(views_number)
from (select count(distinct ad_id) as ads_number,
countIf(event='click') as clicks_number,
countIf(event='view') as views_number
from ads_data
group by ad_id
having views_number=0 and clicks_number!=0);
```

Ответ: всего таких уникальных объявлений 9 (неуникальных — 89) с суммарным количеством кликов 89 и 0 показов.

```
select sum(ios), sum(android), sum(web), sum(ads_number),  
sum(clicks_number), sum(views_number)  
from (select countIf(platform='ios') as ios,  
countIf(platform='android') as android,  
countIf(platform='web') as web,  
count(ad_id) as ads_number,  
countIf(event='click') as clicks_number,  
countIf(event='view') as views_number  
from ads_data  
group by ad_id  
having views_number=0 and clicks_number!=0);
```

sum(ios)	sum(android)	sum(web)	sum(ads_number)	sum(clicks_number)	sum(views_number)
30	47	12	89	89	0

Ответ: число кликов совпадает с общим числом объявлений на всех платформах.

Эта проблема наблюдается на всех платформах?

Ответ: да, на всех, но преобладает на платформе android.

5. Есть ли различия в CTR у объявлений с видео и без?

```
select median(CTR), round(avg(CTR), 2)  
from (select distinct(ad_id), countIf(event='click') as  
clicks_number, countIf(event='view') as views_number,  
round(clicks_number/views_number, 2) as CTR  
from ads_data  
where has_video=1  
group by ad_id  
having views_number!=0);
```

Ответ: да, есть — без видео: медианный CTR равен 0, а средний CTR равен 0,02, с видео: медианный CTR равен 0,01, а средний CTR равен 0,02, не включая объявления с нулем просмотров.

А чему равняется 95 процентиль CTR по всем объявлениям за 2019-04-04?

```
select round(quantile(0.95)(CTR), 2)
from (select distinct(ad_id), countIf(event = 'click') as clicks_number, countIf(event
= 'view') as views_number, round(clicks_number/views_number, 2) as CTR
from ads_data
where date='2019-04-04'
group by ad_id
having views_number!=0);
```

Ответ: 95 процентиль CTR по всем объявлениям за 2019-04-04 равен 0.08.

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.

```
select distinct(date), sum(clicks_number),
sum(views_number), round(sum(cost), 2)
from (select distinct date, countIf(event='click') as
clicks_number,
countIf(event='view') as views_number,
if(ad_cost_type='CPC',
round(clicks_number*ad_cost, 2),
round(views_number*ad_cost/1000, 2)) as cost
from ads_data
group by date, ad_cost_type, ad_cost)
group by date;
```

date	sum(clicks_number)	sum(views_number)	cost
2019-04-01	291	21745	6648.57
2019-04-02	545	46530	13282.47
2019-04-03	460	58883	14138.06
2019-04-04	643	275020	54987.22
2019-04-05	92321	427255	96121.01
2019-04-06	14918	60899	13340.3

Ответ: 2019-04-05 заработок был выше всего — 96121.01, 2019-04-01 — меньше всего — 6648.57.

7. Какая платформа самая популярная для размещения рекламных объявлений?

```
select platform, count(ad_id) as ads, uniqExact(ad_id) as
unique_ads
from ads_data
group by platform
order by ads desc;
```

Ответ: самая популярная платформа для размещения объявлений - android, но с точки зрения уникальных объявлений разница по платформам незначительна.

platform	ads	unique_ads
android	500356	963
ios	299979	964
web	199665	959

Сколько процентов показов приходится на каждую из платформ (колонка platform)?

```
select platform, count(event) as count_ads,  
round(count_ads*100/(select count(*) from ads_data), 2) as  
perc  
from ads_data  
where event='view'  
group by platform  
order by count(event) desc;
```

platform	count_ads	perc
android	445722	44.57
ios	267117	26.71
web	177983	17.8