

```

1 -----
2 -- Курсовой проект. Часть I
3 -----
4 -- В первом задании по проекту вам нужно будет исследовать данные и ответить на несколько аналитических вопросов.
5 --
6 -- Представьте, что вы устроились работать аналитиком в отдел рекламы,
7 -- и ваша первая задача помочь коллегам разобраться с некоторыми вопросами:
8
9 -----
10 -- 1. Получить статистику по дням.
11 -- Просто посчитать число всех событий по дням, число показов, число кликов,
12 -- число уникальных объявлений и уникальных кампаний.
13

```

```

14 SELECT date,
15        count() AS events_all,
16        countIf(event == 'view') AS views,
17        countIf(event == 'click') AS clicks,
18        countIf(ad_id) AS unique_ads,
19        countIf(campaign_union_id) AS unique_campaigns
20 FROM ads_data
21 GROUP BY date
22 ORDER BY date;
23

```

```

24 -- result = [
25 --   {
26 --     "date": "2019-04-01",
27 --     "events_all": 22073,
28 --     "views": 21782,
29 --     "clicks": 291,
30 --     "unique_ads": 150,
31 --     "unique_campaigns": 149
32 --   },
33 --   {
34 --     "date": "2019-04-02",
35 --     "events_all": 47117,
36 --     "views": 46572,
37 --     "clicks": 545,
38 --     "unique_ads": 344,
39 --     "unique_campaigns": 336
40 --   },
41 --   {
42 --     "date": "2019-04-03",
43 --     "events_all": 59483,
44 --     "views": 59023,
45 --     "clicks": 460,
46 --     "unique_ads": 360,
47 --     "unique_campaigns": 352
48 --   },
49 --   {
50 --     "date": "2019-04-04",
51 --     "events_all": 275735,
52 --     "views": 275092,
53 --     "clicks": 643,
54 --     "unique_ads": 407,
55 --     "unique_campaigns": 396
56 --   },
57 --   {
58 --     "date": "2019-04-05",
59 --     "events_all": 519707,
60 --     "views": 427386,
61 --     "clicks": 92321,
62 --     "unique_ads": 465,
63 --     "unique_campaigns": 442
64 --   },
65 --   {
66 --     "date": "2019-04-06",
67 --     "events_all": 75885,
68 --     "views": 60967,
69 --     "clicks": 14918,
70 --     "unique_ads": 220,
71 --     "unique_campaigns": 212
72 --   }
73 -- ]
74

```

```

75 -----
76 -- 2. Разобраться, почему случился такой скачок 2019-04-05?
77 -- Каких событий стало больше?
78 -- У всех объявлений или только у некоторых?
79

```

```

80 SELECT date,
81        ad_id,

```

```

82     countIf(event == 'click') AS clicks,
83     countIf(event == 'view') AS views
84 FROM ads_data
85 GROUP BY date,
86         ad_id
87 ORDER BY clicks DESC
88 LIMIT 10;
89
90
91 WITH 112583 AS anormal_ad_id
92 SELECT date,
93         sum(event == 'click') AS clicks,
94         sum(event == 'click' AND ad_id == anormal_ad_id) AS clicks_anormal_ad,
95         round(clicks_anormal_ad / clicks * 100, 2) AS clicks_anormal_ad_perc,
96         sum(event == 'view') AS views,
97         sum(event == 'view' AND ad_id == anormal_ad_id) AS views_anormal_ad,
98         round(views_anormal_ad / views * 100, 2) AS views_anormal_ad_perc
99 FROM ads_data
100 GROUP BY date
101 -- HAVING clicks_anormal_ad != 0
102 ORDER BY date;
103
104 -- Аномальный скачок по событиям наблюдается 2019-04-05 из-за событий объявления с ad_id = 112583,
105 -- причём, количество кликов по нему ~98.59% от всех кликов за день, а количество просмотров ~70.85%.
106 -- Аномалия продолжается и на 2019-04-06, в меньших абсолютных показателях,
107 -- но сопоставимых относительных (98.87% по кликам и 80.36% просмотрам).
108 -- до 2019-04-05 это объявление в базе не значится.
109
110 -----
111 -- 3. Найти топ 10 объявлений по CTR за все время.
112 -- CTR — это отношение всех кликов объявлений к просмотрам.
113 -- Например, если у объявления было 100 показов и 2 клика, CTR = 0.02.
114 -- Различается ли средний и медианный CTR объявлений в наших данных?
115
116 SELECT ad_id,
117         sum(event == 'click') AS clicks,
118         sum(event == 'view') AS views,
119         round(clicks / views, 3) AS CTR
120 FROM ads_data
121 GROUP BY ad_id
122 HAVING views != 0
123 ORDER BY CTR DESC
124 LIMIT 10;
125
126 -- Без учёта бажных объявлений (с кликами, но без показов) топ 10:
127 -- ad_ids = {117164, 112583, 42507, 98569, 46639, 23599, 19912, 110414, 45969, 20662}
128
129 SELECT avg(CTR),
130        median(CTR)
131 FROM (
132     SELECT sum(event == 'click') AS clicks,
133            sum(event == 'view') AS views,
134            round(clicks / views, 3) AS CTR
135     FROM ads_data
136     GROUP BY ad_id
137     HAVING views != 0
138     ORDER BY CTR DESC
139 );
140
141 -- Без учёта бажных объявлений (с кликами, но без показов)
142 -- среднее: 0.015834728033472772
143 -- медиана: 0.003
144
145 -----
146 -- 4. Похоже, в наших логах есть баг, объявления приходят с кликами, но без показов!
147 -- Сколько таких объявлений, есть ли какие-то закономерности?
148 -- Эта проблема наблюдается на всех платформах?
149
150 SELECT count()
151 FROM (
152     SELECT sum(event == 'click') AS clicks,
153            sum(event == 'view') AS views
154     FROM ads_data
155     GROUP BY ad_id
156     HAVING views = 0
157            AND clicks != 0
158 );
159
160 -- Всего 9 таких объявлений.
161
162 SELECT ad_id,
163        count(ad_id) AS click_count,
164        platform
165 FROM ads_data
166 WHERE ad_id GLOBAL IN (

```

```

167 SELECT ad_id
168 FROM ads_data
169 GROUP BY ad_id
170 HAVING sum(event == 'view') = 0
171 AND sum(event == 'click') != 0
172 )
173 GROUP BY platform,
174 ad_id
175 ORDER BY ad_id;
176
177 -- Этот баг наблюдается на всех платформах.
178 -- На андроиде в среднем таких багов в ~1.5 раза больше, чем на айосе, а на айосе в ~ 2.5 раза больше, чем на вебе.
179 -- Из распределения багов по бажным объявлениям выбивается # 120431 с на порядок превышающим количеством кликов.
180
181 SELECT ad_id
182 FROM ads_data
183 WHERE ad_id GLOBAL IN (
184 SELECT ad_id
185 FROM ads_data
186 GROUP BY ad_id
187 HAVING sum(event == 'view') = 0
188 AND sum(event == 'click') != 0
189 )
190 GROUP BY ad_id;
191
192 -- ad_ids = {115825,120536,117364,45418,120431,120796,41500,19223,26204}
193
194 SELECT *
195 FROM ads_data
196 WHERE ad_id IN (115825,120536,117364,45418,120431,120796,41500,19223,26204)
197 AND has_video == 1
198 ORDER BY time;
199
200 -- Да, все бажные объявления не содержат видео.
201
202 -----
203 -- 5. Есть ли различия в CTR у объявлений с видео и без?
204 -- А чему равняется 95 процентиль CTR по всем объявлениям за 2019-04-04?
205
206 SELECT avg(CTR),
207 median(CTR)
208 FROM (
209 SELECT countIf(event == 'click') AS clicks,
210 countIf(event == 'view') AS views,
211 round(clicks / views, 3) AS CTR
212 FROM ads_data
213 GROUP BY ad_id, has_video
214 HAVING views != 0
215 AND has_video == 1
216 ORDER BY CTR DESC
217 );
218
219 -- [
220 -- {
221 -- "avg(CTR)": 0.020150000000000008,
222 -- "median(CTR)": 0.0095
223 -- }
224 -- ]
225
226 SELECT avg(CTR),
227 median(CTR)
228 FROM (
229 SELECT countIf(event == 'click') AS clicks,
230 countIf(event == 'view') AS views,
231 round(clicks / views, 3) AS CTR
232 FROM ads_data
233 GROUP BY ad_id, has_video
234 HAVING views != 0
235 AND has_video == 0
236 ORDER BY CTR DESC
237 );
238
239 -- [
240 -- {
241 -- "avg(CTR)": 0.015742521367521335,
242 -- "median(CTR)": 0.003
243 -- }
244 -- ]
245
246 -- Различия CTR между объявлениями с видео и без - есть.
247 -- Среднее выше почти на треть, а медиана более чем в три раза.
248
249 SELECT quantile(0.95)(CTR)
250 FROM (
251 SELECT countIf(event == 'click') AS clicks,

```

```

252 countIf(event == 'view') AS views,
253 round(clicks / views, 3) AS CTR
254 FROM ads_data
255 WHERE date == '2019-04-04'
256 GROUP BY ad_id
257 ORDER BY CTR DESC
258 );
259
260 -- 0.08179999999999996
261
262 -----
263 -- 6. Для финансового отчета нужно рассчитать наш заработок по дням.
264 -- В какой день мы заработали больше всего? В какой меньше?
265 -- Мы списываем с клиентов деньги, если произошел клик по CPC объявлению,
266 -- и мы списываем деньги за каждый показ CPM объявления,
267 -- если у CPM объявления цена - 200 рублей, то за один показ мы зарабатываем 200 / 1000.
268
269 SELECT date,
270 round(sumIf(ad_cost, ad_cost_type == 'CPC' AND event == 'click'), 2) AS cpc_sum_cost,
271 round(sumIf(ad_cost / 1000, ad_cost_type == 'CPM' AND event == 'view'), 2) AS cpm_sum_cost,
272 round(cpc_sum_cost + cpm_sum_cost, 2) AS total_cost
273 FROM ads_data
274 GROUP BY date
275 ORDER BY total_cost DESC;
276
277 -- Больше всего заработали 2019-04-05: ~96123.12,
278 -- меньше всего - 2019-04-01: ~6655.71.
279
280 -----
281 -- 7. Какая платформа самая популярная для размещения рекламных объявлений?
282 -- Сколько процентов показов приходится на каждую из платформ (колонка platform)?
283
284 SELECT platform AS most_popular_platform
285 FROM ads_data
286 GROUP BY platform
287 ORDER BY count() DESC
288 LIMIT 1;
289
290 -- самая популярная для размещения рекламных объявлений платформа - android
291
292 SELECT count() AS views_cnt,
293 round(countIf(platform == 'web') / views_cnt * 100, 2) AS web_cnt_perc,
294 round(countIf(platform == 'ios') / views_cnt * 100, 2) AS ios_cnt_perc,
295 round(countIf(platform == 'android') / views_cnt * 100, 2) AS android_cnt_perc
296 FROM ads_data
297 WHERE event == 'view';
298
299 -- На андроид приходится 50.03% всех показов
300
301 -----
302 -- 8. А есть ли такие объявления, по которым сначала произошел клик, а только потом показ?
303
304 SELECT ad_id,
305 minIf(time, event == 'click') AS first_click,
306 minIf(time, event == 'view') AS first_view
307 FROM ads_data
308 GROUP BY ad_id
309 HAVING first_click < first_view
310 AND first_click != '1970-01-01 00:00:00'
311 ORDER BY ad_id;
312
313 -- Да, есть, 12 штук:
314 -- ad_ids = {18681, 23599, 32386, 33033, 36758, 38224, 44283, 44766, 46639, 98569, 107798, 114886}
315 -----
316

```