

Задача:

Возьмите свою страницу Вконтакте, соберите по ней таблицу с датой постов и количеством лайков.

Напишите SQL-запросы, которые позволят ответить на вопрос: что больше всего влияет на количество лайков: время суток публикации, день недели или промежуток между постами.

Описание данных:

Данные взяты со страницы моей соседки

При создании таблицы я не учитывала репосты: на них мало лайков вне зависимости от других факторов, значит, они помешают выделить основные закономерности

id - id записи

date - дата и время поста

likes - количество лайков

id	date	likes
3	2024-08-09 15:53:00	74
4	2024-07-14 10:11:00	55
5	2024-07-13 16:15:00	76
6	2024-07-05 11:17:00	62
7	2024-06-27 23:11:00	66
8	2024-06-14 22:30:00	105
9	2024-05-27 18:42:00	75
10	2024-05-26 21:07:00	86
11	2024-05-23 09:37:00	166
12	2024-05-01 23:39:00	122
13	2024-04-05 17:58:00	57
14	2024-04-02 20:39:00	71
15	2024-03-26 07:03:00	65
16	2024-03-20 17:38:00	55
17	2024-03-05 08:07:00	67
18	2024-01-26 21:56:00	82
19	2023-10-28 18:49:00	123
20	2023-09-24 20:28:00	123
21	2023-09-14 23:37:00	177
22	2023-09-05 21:56:00	89
23	2023-08-21 09:39:00	153
24	2023-08-14 12:51:00	60
25	2023-07-10 08:38:00	171
26	2023-06-14 23:59:00	186
27	2023-04-20 10:10:00	90
28	2023-03-11 18:45:00	21
29	2023-02-12 21:23:00	180
30	2023-01-23 19:50:00	184
31	2022-12-09 15:41:00	100
32	2022-10-13 23:43:00	188
33	2022-07-05 16:30:00	51
34	2022-06-26 14:33:00	289
35	2022-06-17 19:23:00	47
36	2022-06-08 20:05:00	45
37	2022-04-04 04:32:00	20

1. Среднее количество лайков по времени суток

```
SELECT ROUND(AVG(likes)) as avg_likes_by_time_of_day,
```

-- Определим время суток на основе часа публикации поста

```
CASE
```

```

        WHEN HOUR(date) BETWEEN 6 AND 12 THEN 'Morning'
        WHEN HOUR(date) BETWEEN 12 AND 18 THEN 'Afternoon'
        WHEN HOUR(date) BETWEEN 18 AND 24 THEN 'Evening'
        ELSE 'Night'
    END AS time_of_day
FROM posts
GROUP BY time_of_day -- Сгруппируем результаты по времени суток
ORDER BY avg_likes_by_time_of_day DESC

```

avg_likes_by_time_of_day	time_of_day
117	Evening
93	Morning
92	Afternoon
20	Night

Самое лучшее время для выставления постов - ожидаемо вечер.
 Результат в 20 лайков ночью можно не считать значимым, потому что ночной пост был только один. Результаты утренних и дневных постов примерно одинаковы

2. Среднее количество лайков по дням недели

```

SELECT WEEKDAY(date) as day_of_week, ROUND(AVG(likes)) as
avg_likes_per_day_of_week
FROM vk.posts
GROUP BY day_of_week
ORDER BY day_of_week

```

day_of_week	avg_likes_per_day_of_week
0	100
1	69
2	102
3	137
4	78
5	73
6	147

Больше всего лайков собирают воскресные посты. Но мы не можем сделать точный вывод, что лучше выставлять записи на выходных, потому что на субботах постах лайков относительно мало, а в четверг, например, наоборот необъяснимо много. Кажется, что нет точной зависимости между днем недели и количеством лайков

3. Среднее количество лайков в зависимости от временного промежутка с последнего поста

```

WITH t1 AS (
    SELECT
        id,
        date,
        LAG(date) OVER (ORDER BY date) AS last_post_time,
        likes

```

```

FROM posts
) -- добавили столбец с датой прошлого поста
SELECT ROUND(AVG(likes)) as avg_likes_by_period,
CASE -- распределяем по периодам
    WHEN TIMESTAMPDIFF(DAY, last_post_time, date) <= 3 THEN 'до 3 дней'
    WHEN TIMESTAMPDIFF(DAY, last_post_time, date) <= 7 THEN 'до недели'
    WHEN TIMESTAMPDIFF(DAY, last_post_time, date) <= 14 THEN '1-2 недели'
    WHEN TIMESTAMPDIFF(DAY, last_post_time, date) <= 30 THEN '2 недели - месяц'
    ELSE 'больше месяца'
END AS time_period
FROM t1
GROUP BY time_period
ORDER BY avg_likes_by_period DESC

```

avg_likes_by_period	time_period
116	1-2 недели
104	больше месяца
102	2 недели - месяц
88	до недели
68	до 3 дней

Посты набирают меньше лайков, если промежуток между постами был меньше недели. Идеальный промежуток - 1-2 недели

4. Интерпретация результатов

Если смотреть “на глаз”, то два наиболее влияющих фактора - период между публикациями (не стоит выкладывать пост раньше, чем через неделю после предыдущего) и время дня (предпочтителен вечер)

Если посчитать коэффициенты корреляции, то окажется, что корреляция в целом слабая, но time_period влияет сильнее всего

	likes	time_of_day	day_of_week	time_period
likes	1.000000	-0.098961	0.115667	0.175496