

Тестовое задание «Анализ влияния времени суток, дня недели и промежутка между постами на количество лайков под постами ВК»

В качестве страницы для анализа постов была выбрана страница Дмитрия Медведева (ссылка на страницу: <https://vk.com/dm>).

Фрагмент таблицы в виде файла формата .csv, полученной после парсинга (файл posts.csv):

1	id,"likes_count","date"	
2	55089,506,"2025-10-09 18:34:57"	
3	55088,575,"2025-10-01 19:43:42"	
4	55087,459,"2025-09-30 16:16:21"	
5	55086,1770,"2025-09-19 18:53:44"	
6	55085,1123,"2025-09-16 19:39:24"	
7	55084,1341,"2025-09-15 00:47:52"	
8	55083,655,"2025-09-08 18:13:44"	
9	55082,1306,"2025-09-08 13:39:37"	
10	55081,1145,"2025-09-06 17:08:09"	

Таблица содержит следующие поля:

- id – уникальный номер поста
- likes_count – количество лайков под постом
- date – дата и время публикации поста

Таблица содержит информацию о 100 последних постах по состоянию на 10.10.2025, отсортированных по дате публикации от новых к старым.

1. Время суток (likes_per_day_period)

Код SQL-запроса (файл likes_per_day_period.sql):

```
SELECT
  CASE
    WHEN DATE_PART ('hour', date) >= 4 AND DATE_PART('hour', date) < 12 THEN 2
    WHEN DATE_PART ('hour', date) >= 12 AND DATE_PART('hour', date) < 17 THEN 3
    WHEN DATE_PART ('hour', date) >= 17 AND DATE_PART('hour', date) < 23 THEN 4
    ELSE 1
  END AS day_period_number,
  --Задаем номер для каждого времени суток (это нужно только для сортировки таблицы от 0 до 24 часов)
  CASE
    WHEN DATE_PART ('hour', date) >= 4 AND DATE_PART('hour', date) < 12 THEN 'Morning (4-11)'
    WHEN DATE_PART ('hour', date) >= 12 AND DATE_PART('hour', date) < 17 THEN 'Afternoon (12-16)'
    WHEN DATE_PART ('hour', date) >= 17 AND DATE_PART('hour', date) < 23 THEN 'Evening (17-23)'
    ELSE 'Night (0-3)'
  END AS day_period,
  --Делим часы дня на периоды: ночь, утро, день, вечер
  SUM(likes_count) AS likes_count,
  --Считаем количество лайков для последующей группировки по периодам дня
  COUNT(id) AS posts_count,
  --Считаем количество постов для последующей группировки по периодам дня
  ROUND(SUM(likes_count)::DECIMAL / COUNT(id), 2) AS likes_per_post
  --Считаем количество лайков на пост для последующей группировки по периодам дня
FROM posts
GROUP BY day_period, day_period_number
ORDER BY day_period_number
```

Результат выполнения запроса (файл likes_per_day_period.csv):

	day_period_number integer	day_period text	likes_count bigint	posts_count bigint	likes_per_post numeric
1	1	Night (0-3)	11414	3	3804.67
2	2	Morning (4-11)	70690	17	4158.24
3	3	Afternoon (12-1...	107067	39	2745.31
4	4	Evening (17-23)	71151	41	1735.39

Все посты исходной таблицы сгруппированы по часам публикации по схеме:

- 0-3 часа – Ночь
- 4-11 часов – Утро
- 12-16 часов – День
- 17-23 часа – Вечер

Для каждого из этих периодов посчитаны общее число лайков (likes_count) и количество лайков на пост (likes_per_post)

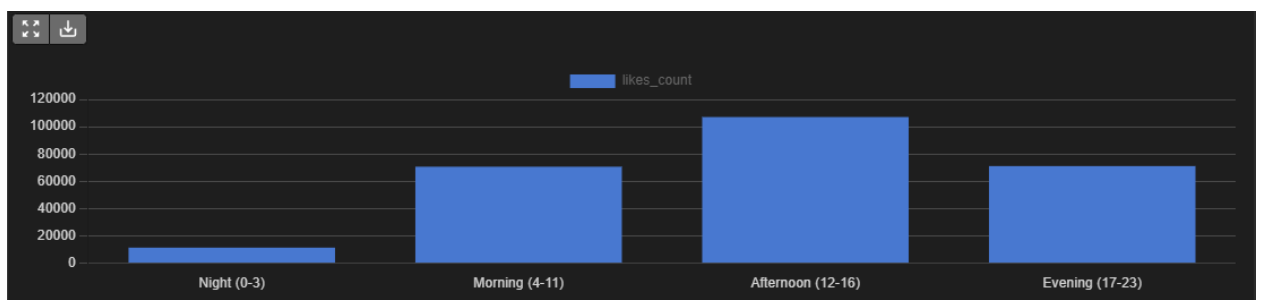


График зависимости количества лайков от времени суток (файл likes_per_day_period.png)

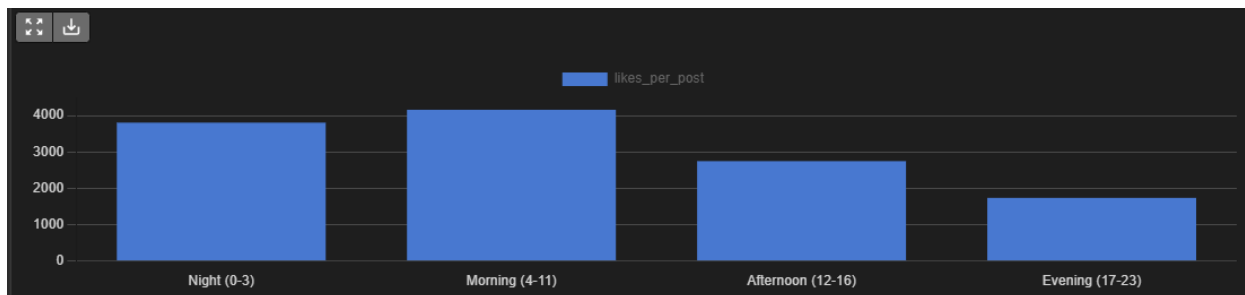


График зависимости количества лайков на пост от времени суток (файл likes_per_post.png)

Вывод: наибольшее количество лайков было поставлено днем (12-16 часов), но по показателю количества лайков на пост более выгодными являются публикации, созданные утром (4-11 часов).

2. День недели (likes_per_day_of_week)

Код SQL-запроса (файл likes_per_day_of_week.sql):

```
SELECT DATE_PART('isodow', date) AS day_of_week_number,
--Получаем номер дня недели (для последующей сортировки)
TO_CHAR(date, 'Day') AS day_of_week,
--Получаем названия дней недели
SUM(likes_count) AS likes_count,
--Считаем количество лайков для группировки по дням недели
COUNT(id) AS posts_count,
--Считаем количество постов для группировки по дням недели
ROUND(SUM(likes_count)::DECIMAL / COUNT(id), 2) AS likes_per_post
--Считаем количество лайков на пост для группировки по дням недели
FROM posts
GROUP BY day_of_week_number, day_of_week
ORDER BY day_of_week_number
```

Результат выполнения запроса (файл likes_per_day_of_week.csv):

	day_of_week_number double precision	day_of_week text	likes_count bigint	posts_count bigint	likes_per_post numeric
1	1	Monday	23820	9	2646.67
2	2	Tuesday	46210	14	3300.71
3	3	Wednesday	34194	17	2011.41
4	4	Thursday	50810	27	1881.85
5	5	Friday	37134	16	2320.88
6	6	Saturday	42759	10	4275.90
7	7	Sunday	25395	7	3627.86

Все посты исходной таблицы сгруппированы по дням недели.

Для каждого дня недели посчитаны число лайков (`likes_count`) и количество лайков на пост (`likes_per_post`).

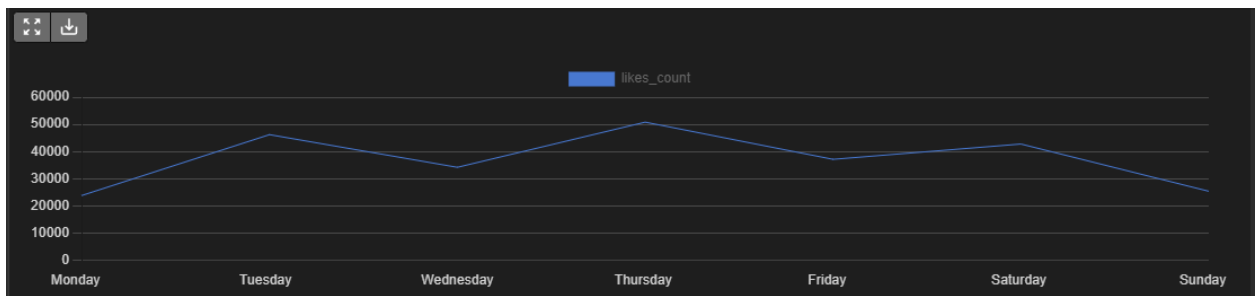


График зависимости количества лайков от дня недели (файл `likes_per_day_of_week.png`)

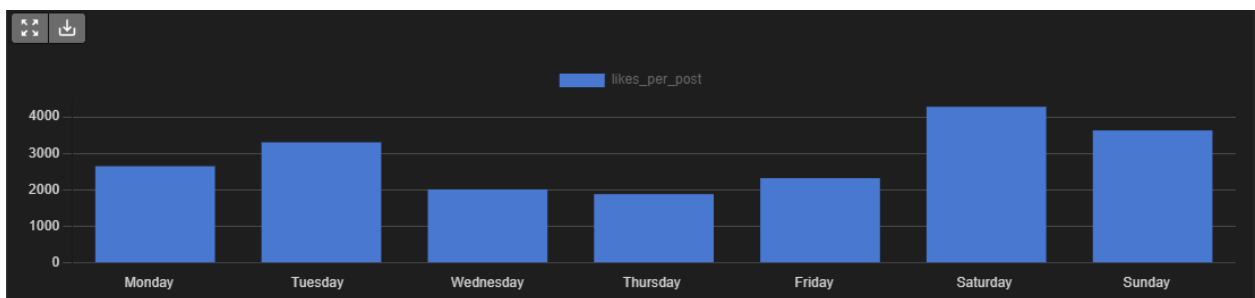


График зависимости количества лайков на пост от дня недели (файл `likes_per_post.png`)

Вывод: наибольшее число лайков было поставлено в четверг, но показатель числа лайков на пост выше в субботу, хотя количество постов в субботу в 2.7 раза меньше, чем в четверг.

3. Промежуток между постами (`likes_per_time_difference`)

Код SQL-запроса (файл `likes_per_days_difference.sql`):

```
SELECT
    days_between_posts,
    SUM(likes_count) AS likes_count,
    COUNT(id) AS posts_count,
    ROUND(SUM(likes_count)::DECIMAL / COUNT(id), 2) AS likes_per_post
FROM
    (SELECT
        COALESCE(DATE_PART('days', AGE (LAG(date) OVER(), date)), 0) AS days_between_posts,
        --Считаем количество дней между постами
        likes_count,
        id
    FROM posts) AS days_interval
GROUP BY days_between_posts
ORDER BY days_between_posts
```

Результат выполнения запроса (файл likes_per_days_difference.csv):

	days_between_posts double precision	likes_count bigint	posts_count bigint	likes_per_post numeric
1	0	71999	29	2482.72
2	1	39438	19	2075.68
3	2	47555	15	3170.33
4	3	21767	9	2418.56
5	4	26499	6	4416.50
6	5	16911	5	3382.20
7	6	15784	9	1753.78
8	7	575	1	575.00
9	8	3099	1	3099.00
10	9	4610	2	2305.00
11	10	4735	2	2367.50
12	11	7350	2	3675.00

Все посты исходной таблицы сгруппированы по количеству дней, прошедших после публикации предыдущего поста.

Для каждого промежутка в днях между постами посчитаны число лайков (likes_count) и количество лайков на пост (likes_per_post).

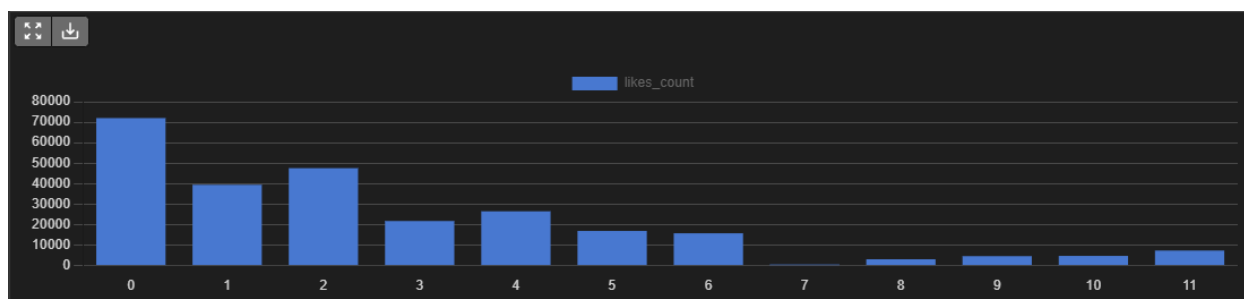


График зависимости количества лайков от промежутка между постами (файл likes_per_days_difference.png)

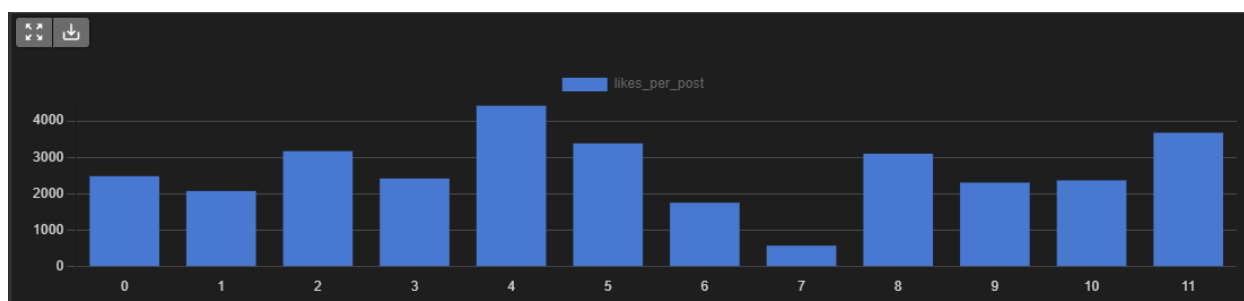


График зависимости количества лайков на пост от промежутка между постами (файл likes_per_post.png)

Вывод: Наибольшее число лайков было поставлено под постами, которые были опубликованы в тот же день, что и предыдущий. По показателю количества лайков на пост лидируют посты, которые были опубликованы через 4 дня после предыдущих.

Заключение (общий вывод)

Из трех рассмотренных критериев для страницы ВК Дмитрия Медведева больше всего на количество лайков влияет промежуток между постами.

Показатель количества лайков на пост для критерия «промежуток между постами» в 4 дня в 1.2 раза больше, чем тот же показатель для промежутка в 11 дней (2 наибольших значения likes_per_post в таблице likes_per_days_difference).

Для критериев «день недели» и «время суток» максимальный и второй по численности показатели likes_per_post отличаются в 1.17 и 1.09 раз соответственно.

Идеальной формулой для получения наибольшего количества лайков на рассматриваемой странице можно считать посты, опубликованные утром субботы после четырехдневного перерыва.

Но наиболее важными критериями, конечно, являются тема и содержание поста.