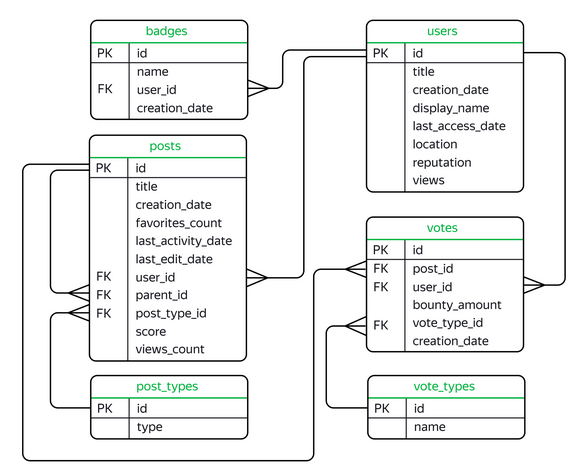
# База данных



## Задача 1

Найдите количество вопросов, которые набрали больше 300 очков или как минимум 100 раз были добавлены в «Закладки».

SELECT count(post\_type\_id)

FROM stackoverflow.posts

where post\_type\_id='1' and score>300 or favorites\_count>=100

Ответ:



## Задача 2

Сколько в среднем в день задавали вопросов с 1 по 18 ноября 2008 включительно? Результат округлите до целого числа.

SELECT round(avg(count\_day)) AS question\_avg

from

(SELECT count(post\_type\_id) AS count\_day,

creation\_date::date as day

FROM stackoverflow.posts

where post\_type\_id='1' and creation\_date between '2008-11-01' and '2008-11-19'

group by 2

order by 2) counts

Ответ:



## Задача 3

Сколько пользователей получили значки сразу в день регистрации? Выведите количество уникальных пользователей.

select count(distinct b.user\_id)

FROM stackoverflow.users u

join stackoverflow.badges b on u.id=b.user\_id

where b.creation\_date::date=u.creation\_date::date

Ответ:



## Задача 4

Сколько уникальных постов пользователя с именем Joel Coehoorn получили хотя бы один голос?

select count(distinct p.id)

FROM stackoverflow.posts p

join stackoverflow.users u on u.id=p.user\_id

join stackoverflow.votes v on v.post\_id=p.id

where u.display\_name='Joel Coehoorn'

Ответ:



## Задача 5.

Выгрузите все поля таблицы vote\_types. Добавьте к таблице поле rank, в которое войдут номера записей в обратном порядке. Таблица должна быть отсортирована по полю id.

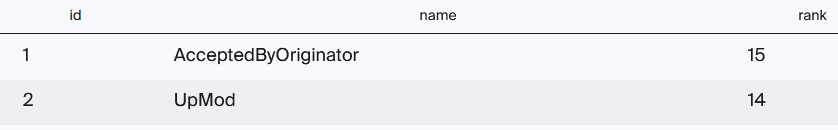
select \*,

row\_number() over (order by id desc) as rank

from stackoverflow.vote\_types

order by id

Ответ:



## Задача 6.

Отберите 10 пользователей, которые поставили больше всего голосов типа Close. Отобразите таблицу из двух полей: идентификатором пользователя и количеством голосов. Отсортируйте данные сначала по убыванию количества голосов, потом по убыванию значения идентификатора пользователя.

select u.id,

COUNT(vt.id)

from stackoverflow.votes as v

join stackoverflow.users u ON v.user\_id = u.id

join stackoverflow.vote\_types vt ON vt.id = v.vote\_type\_id

where vt.name='Close'

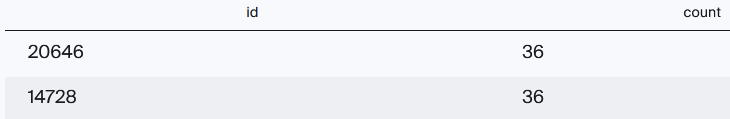
group by 1

order by 2 DESC,

1 DESC

limit 10

Ответ:



## Задача 7.

Отберите 10 пользователей по количеству значков, полученных в период с 15 ноября по 15 декабря 2008 года включительно. Отобразите несколько полей:

* идентификатор пользователя;
* число значков;
* место в рейтинге — чем больше значков, тем выше рейтинг.

Пользователям, которые набрали одинаковое количество значков, присвойте одно и то же место в рейтинге.

Отсортируйте записи по количеству значков по убыванию, а затем по возрастанию значения идентификатора пользователя.

select user\_id,

COUNT(id),

dense\_rank() over (order by count(id) desc)

from stackoverflow.badges

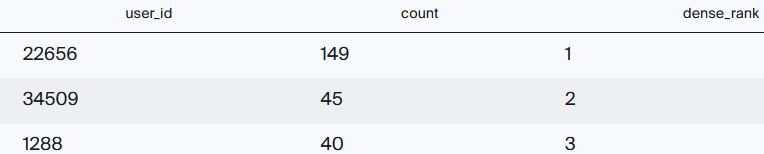
where creation\_date:: date between '2008-11-15' and '2008-12-15'

group by 1

order by 2 DESC, 1

limit 10

Ответ:



## Задача 8.

Сколько в среднем очков получает пост каждого пользователя?

Сформируйте таблицу из следующих полей:

* заголовок поста;
* идентификатор пользователя;
* число очков поста;
* среднее число очков пользователя за пост, округлённое до целого числа.

Не учитывайте посты без заголовка, а также те, что набрали ноль очков.

select title,

user\_id,

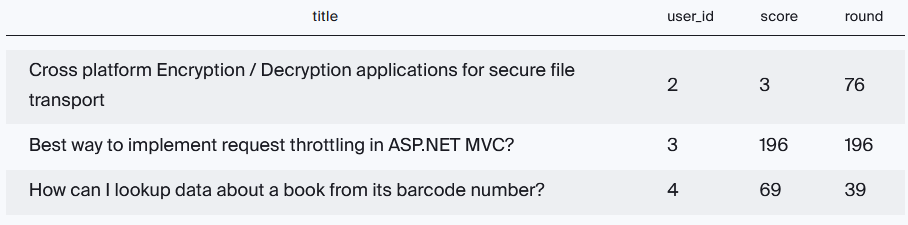
score,

round(avg(score) over (partition by user\_id))

from stackoverflow.posts

where title is not null AND score <> 0

Ответ:



## Задача 9.

Отобразите заголовки постов, которые были написаны пользователями, получившими более 1000 значков. Посты без заголовков не должны попасть в список.

with usr as

(select distinct user\_id,

count(id)

from stackoverflow.badges

group by 1

having count(id)>1000)

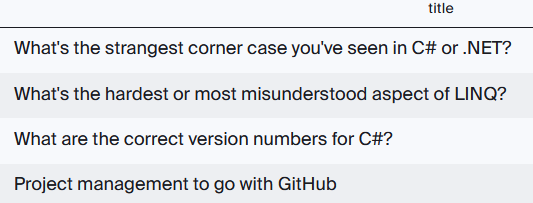
select title

from stackoverflow.posts p

join usr on usr.user\_id = p.user\_id

where title is not null

Ответ:



## Задача 10.

Напишите запрос, который выгрузит данные о пользователях из США (англ. United States). Разделите пользователей на три группы в зависимости от количества просмотров их профилей:

* пользователям с числом просмотров больше либо равным 350 присвойте группу 1;
* пользователям с числом просмотров меньше 350, но больше либо равно 100 — группу 2;
* пользователям с числом просмотров меньше 100 — группу 3.

Отобразите в итоговой таблице идентификатор пользователя, количество просмотров профиля и группу. Пользователи с нулевым количеством просмотров не должны войти в итоговую таблицу.

select id,

views,

CASE

WHEN views >=350 then 1

when views>=100 then 2

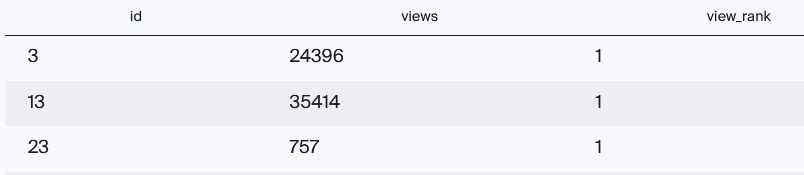
else 3

END AS view\_rank

from stackoverflow.users

where location like '%United States%' and views <> 0;

Ответ:



## Задача 11.

Дополните предыдущий запрос. Отобразите лидеров каждой группы — пользователей, которые набрали максимальное число просмотров в своей группе. Выведите поля с идентификатором пользователя, группой и количеством просмотров. Отсортируйте таблицу по убыванию просмотров, а затем по возрастанию значения идентификатора.

with max\_v as (with cases as (select id,

views,

CASE

WHEN views >=350 then 1

when views>=100 then 2

else 3

END AS view\_rank

from stackoverflow.users

where location like '%United States%' and views <> 0)

select \*,

max(views) over (partition by view\_rank) as max\_views

from cases)

select id,

view\_rank,

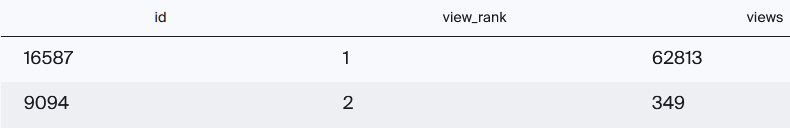
views

from max\_v

where views=max\_views

order by 3 desc, 1

Ответ:



## Задача 12.

Посчитайте ежедневный прирост новых пользователей в ноябре 2008 года. Сформируйте таблицу с полями:

* номер дня;
* число пользователей, зарегистрированных в этот день;
* сумму пользователей с накоплением.

with cnt as (select extract (day from creation\_date) as day,

count(id) over (partition by creation\_date::date) as count,

count(id) over (order BY creation\_date::date)as count\_cum

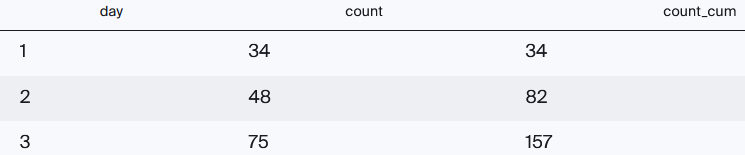
from stackoverflow.users

where date\_trunc('month', creation\_date)::date = '2008-11-01')

select distinct \*

from cnt

order by day



## Задача 13.

Для каждого пользователя, который написал хотя бы один пост, найдите интервал между регистрацией и временем создания первого поста. Отобразите:

* идентификатор пользователя;
* разницу во времени между регистрацией и первым постом.

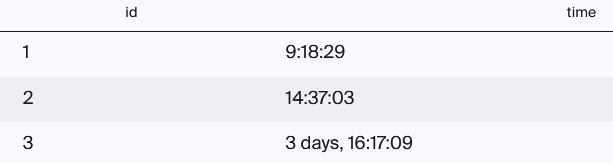
select distinct u.id,

first\_value(p.creation\_date) over (partition by u.id order by p.creation\_date) - u.creation\_date as time

from stackoverflow.users u

join stackoverflow.posts p on p.user\_id = u.id

Ответ:



# Задания (Вторая часть)

## Задача 1.

Выведите общую сумму просмотров постов за каждый месяц 2008 года. Если данных за какой-либо месяц в базе нет, такой месяц можно пропустить. Результат отсортируйте по убыванию общего количества просмотров.

select date\_trunc('month', creation\_date)::date as month,

sum(views\_count)

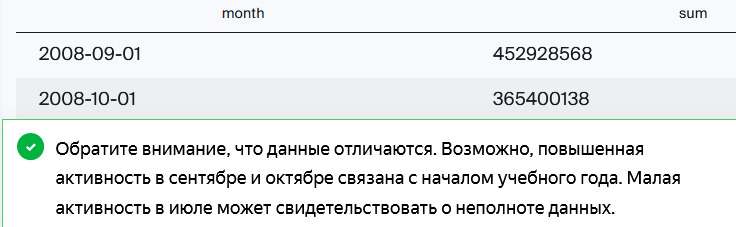
from stackoverflow.posts

where date\_trunc('month', creation\_date)::date between '2008-01-01' and '2008-12-01'

group by 1

order by 2 desc

Ответ:



## Задача 2.

Выведите имена самых активных пользователей, которые в первый месяц после регистрации (включая день регистрации) дали больше 100 ответов. Вопросы, которые задавали пользователи, не учитывайте. Для каждого имени пользователя выведите количество уникальных значений user\_id. Отсортируйте результат по полю с именами в лексикографическом порядке.

SELECT u.display\_name, COUNT(DISTINCT u.id)

FROM stackoverflow.users AS u

JOIN stackoverflow.posts AS p ON u.id = p.user\_id

WHERE p.post\_type\_id = 2

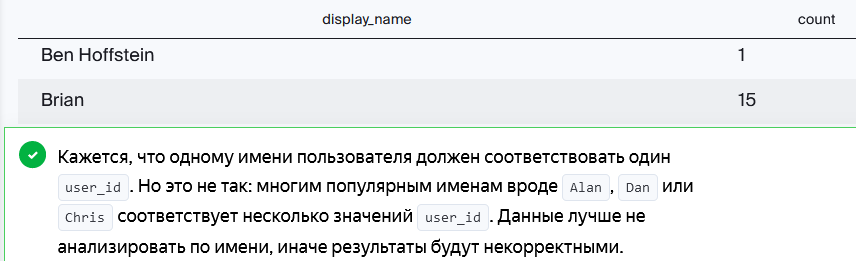
AND p.creation\_date::date <= u.creation\_date::date + INTERVAL '1 month'

GROUP BY u.display\_name

HAVING COUNT(\*) > 100

ORDER BY display\_name

Ответ:



## Задача 3.

Выведите количество постов за 2008 год по месяцам. Отберите посты от пользователей, которые зарегистрировались в сентябре 2008 года и сделали хотя бы один пост в декабре того же года. Отсортируйте таблицу по значению месяца по убыванию.

select distinct date\_trunc('month', creation\_date)::date as month,

count(\*) over (partition by date\_trunc('month', creation\_date)::date)

FROM stackoverflow.posts

where user\_id in (SELECT distinct u.id

FROM stackoverflow.users AS u

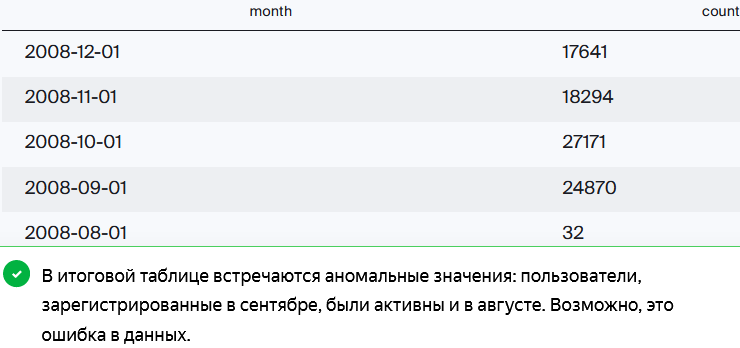
JOIN stackoverflow.posts AS p ON u.id = p.user\_id

WHERE date\_trunc('month', u.creation\_date)::date='2008-09-01'

AND date\_trunc('month', p.creation\_date)::date='2008-12-01')

order by 1 desc

Ответ:



## Задача 4.

Используя данные о постах, выведите несколько полей:

* идентификатор пользователя, который написал пост;
* дата создания поста;
* количество просмотров у текущего поста;
* сумму просмотров постов автора с накоплением.

Данные в таблице должны быть отсортированы по возрастанию идентификаторов пользователей, а данные об одном и том же пользователе — по возрастанию даты создания поста.

select user\_id,

creation\_date,

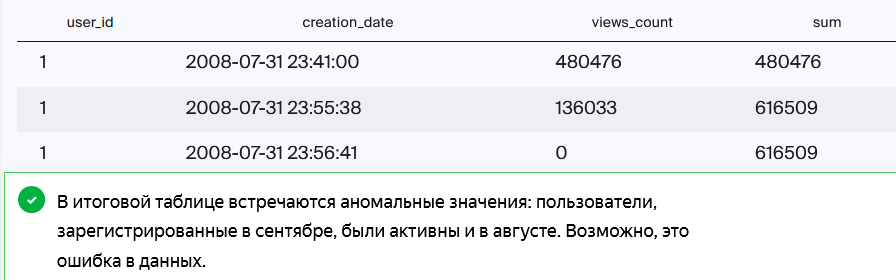
views\_count,

sum(views\_count) over (partition by user\_id order by creation\_date)

FROM stackoverflow.posts

order by 1, 2

Ответ:



## Задача 5.

Сколько в среднем дней в период с 1 по 7 декабря 2008 года включительно пользователи взаимодействовали с платформой? Для каждого пользователя отберите дни, в которые он или она опубликовали хотя бы один пост. Нужно получить одно целое число — не забудьте округлить результат.

with activ as (

select distinct user\_id,

date\_trunc('day', creation\_date)::date,

row\_number() over (partition by user\_id) as count

from stackoverflow.posts

where creation\_date::date between '2008-12-01' and '2008-12-07'

group by 1, 2),

count\_days as (select distinct user\_id,

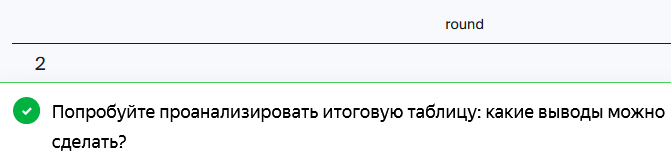
last\_value(count) over(partition by user\_id) as count\_days

from activ)

select round(avg(count\_days))

from count\_days

Ответ:



## Задача 6.

На сколько процентов менялось количество постов ежемесячно с 1 сентября по 31 декабря 2008 года? Отобразите таблицу со следующими полями:

* номер месяца;
* количество постов за месяц;
* процент, который показывает, насколько изменилось количество постов в текущем месяце по сравнению с предыдущим.

Если постов стало меньше, значение процента должно быть отрицательным, если больше — положительным. Округлите значение процента до двух знаков после запятой.

Напомним, что при делении одного целого числа на другое в PostgreSQL в результате получится целое число, округлённое до ближайшего целого вниз. Чтобы этого избежать, переведите делимое в тип numeric.

with cnt as (

select extract(month from creation\_date::date) as month,

count(\*) AS posts\_cnt

from stackoverflow.posts

where creation\_date::date between '2008-09-01' and '2008-12-31'

group by 1

order by 1

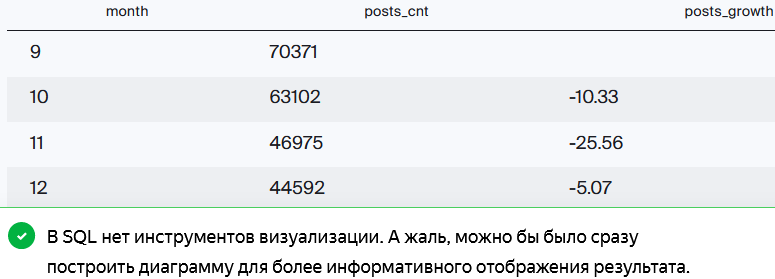
)

select \*,

round(((posts\_cnt::numeric / LAG(posts\_cnt) OVER (ORDER BY month)) - 1) \* 100, 2) AS posts\_growth

from cnt

Ответ:



## Задача 7.

Выгрузите данные активности пользователя, который опубликовал больше всего постов за всё время. Выведите данные за октябрь 2008 года в таком виде:

* номер недели;
* дата и время последнего поста, опубликованного на этой неделе.

with usr as (

select distinct user\_id,

count(\*) AS posts\_cnt

from stackoverflow.posts

group by 1

order by 2 desc

limit 1),

weeks as (

select p.creation\_date,

extract(week from p.creation\_date::date) as week

from stackoverflow.posts p

join usr on usr.user\_id=p.user\_id)

select distinct week,

last\_value(creation\_date) OVER (PARTITION BY week) as last\_post

from weeks

where date\_trunc('month', creation\_date)::date='2008-10-01'

order by 1

Ответ:

