

Коллега оставил вам код, который он делал по следующему ТЗ:

**Выгрузить статистику по данным и раннего и позднего взыскания. В качестве группировочных признаков использовать**

- **забиненный регион (три значения: отдельно Центр, отдельно Север и отдельно остальные регионы). Исключить из рассмотрения Сибирь.**
- **забиненное количество дней просрочки (произвольное количество бинов). Для всех клиентов позднего взыскания можно считать, что количество дней просрочки равно 720 дней.**
- **Пол клиента (если значения нет, то использовать фиктивную переменную: 0, NULL, -1 и тд)**

**В качестве агрегаций взять:**

- **Количество кредитов**
- **Сумму кредита**

Код не работает, в нём есть несколько ошибок. Плюс, есть большое количество неточностей, лишний подзапрос и странные решения о бинировании. Предложите свои варианты исправления кода и опишите свою логику.

Опишите ваши рекомендации относительно использования среднего: делать ли, как ваш коллега, агрегацию avg в SQL или создать вычисляемое поле в сводной таблице в Excel?

```
select case when current_date-month_paid_inst<'10 month' then 1 when current_date-month_paid_inst<'25 month' then 2 when current_date-
when current_date-month_paid_inst<'200 month' then 3 else -1 end len_bin
, count(*) as cnt,region_bin,sum(amt_credit),avg(amt_credit)
from (
select * from (
select id_client, date_trunc('month',last_paid_inst) month_paid_inst,
case when name_region = 'центр' then 1 when name_region = 'Север' then 2 else -1 region_bin, amt_credit
sum(amt_credit) over (partition by name_region order by last_paid_inst) as sum_all_by_region
from skybank.early_collection_clients a left join skybank.region_dict b on a.id_city = b.id_city
where lower(name_region) not like 'сибирь' or upper(name_region) not like 'СИБИРЬ' union
select id_client, '1900-01-01' month_paid_inst, case when name_region = 'центр' then 1 when name_region = 'Север' then 2 else -1 end ,
amt_loan,
sum(amt_loan) over (partition by name_region) as sum_all_by_region
from skybank.late_collection_clients a join skybank.region_dict b on a.id_city = b.id_city where name_region not like '%сибирь%') tt
group by len_bin,region_bin
```

Исправленный код :

```
select
case when current_date - day_paid_inst < '180 days' then '180 дней'
when current_date - day_paid_inst < '360 days' then '360 дней'
when current_date - day_paid_inst < '540 days' then '540 дней'
when current_date - day_paid_inst < '720 days' then '720 дней'
else 'Больше 720 Дней' end len_bin
```

```
, region_bin
, gender
, count(*) as количество_кредитов
, sum(amt_loan) as сумма_кредитов
```

```
from (
select id_client
, date_trunc('day',last_paid_inst) as day_paid_inst
```

```

, amt_credit as amt_loan
, case when name_region = 'Центр' then 'Центр'
when name_region = 'Север' then 'Север'
else 'Другие регионы' end as region_bin
, sum(amt_credit) over (partition by case when name_region = 'Центр' then 'Центр'
when name_region = 'Север' then 'Север'
else 'Другие регионы' end
order by last_paid_inst) as sum_all_by_region
, null as gender
from skybank.early_collection_clients a
left join skybank.region_dict b
on a.id_city = b.id_city
where lower(name_region) not like 'сибирь'
-- order by random()

```

```

union

select id_client
      , '1900-01-01'::date as day_paid_inst
      , amt_loan as amt_loan
      , case when name_region = 'Центр' then 'Центр'
            when name_region = 'Север' then 'Север'
            else 'Другие регионы' end as region_bin
      , sum(amt_loan) over (partition by case when name_region = 'Центр' then 'Центр'
                                when name_region = 'Север' then 'Север'
                                else 'Другие регионы' end ) as sum_all_by_region
      , gender
from skybank.late_collection_clients a
join skybank.region_dict b
on a.id_city = b.id_city
where lower(name_region) not like 'сибирь'
-- order by random()

) tt

```

```

group by region_bin, len_bin , gender
order by len_bin

```