Analytics Engineer Discussion Prompts

Prior to our first interview I would look you to take the time to look at the two below prompts and have you get together some thoughts on how you might optimize or simplify the code. You DO NOT need to and SHOULD NOT rewrite this code, but just come up with a few broad ways you might edit it to make it better.

We will discuss these prompts during our first interview. You can either write down some thoughts and email back to prompt the discussion or we can just talk it through in the interview. There is no expectation you submit written responses, but if you are more comfortable replying in written format the option is available.

Prompt 1

An analyst has created the below code to create aggregates of targets by different demo groups. They are complaining that is takes a long time to run. What steps might someone take to make this query run faster

---

```
select
        a.var
        ,a.level
        ,a.n_dist
        ,a.pct_dist
   ,a,concentration
   ,a.hh_count
   ,a.hh_dist
from
(
        select
        '00 topline' as var
        ,'All Voters' as level
        ,sum(univ_flag) as n_dist
        ,sum(univ_flag)/sum(sum(univ_flag)) over()::float as pct_dist
   ,sum(univ_flag)/sum(case when b.sporadic_dem_target =1 or b.swing_target =1 then 1 else 0
end)::float as concentration
   ,count(distinct householdid1) as hh_count
   ,count(distinct householdid1)/sum(count(distinct householdid1)) over()::float as hh_dist

        from (select * from model_scores.univ_base where target_geo2 in
('AZ','CA-47','CA-49','CO','CT-5','GA','IA-3','IL-14','KS-3','MI','NH','NJ-3','NV','OH-9','PA','VA-2','VA-7','WA-8','
WI')) b
   left join (select *, '1' as univ_flag from model_scores.model_base_20220818_final left join
model_scores.household using(id,state)) using(id,state)
   left join voterfile_install.district d using(id,state)
        group by 1,2

        union all

        select
        '01 age bucket' as var
        ,b.demo_age_bucket_full as level
        ,sum(univ_flag) as n_dist
        ,sum(univ_flag)/sum(sum(univ_flag)) over()::float as pct_dist
```

```
    ,sum(univ_flag)/sum(case when b.sporadic_dem_target =1 or b.swing_target =1 then 1 else 0
end)::float as concentration
    ,count(distinct householdid1) as hh_count
    ,count(distinct householdid1)/sum(count(distinct householdid1)) over()::float as hh_dist

            from (select * from model_scores.univ_base where target_geo2 in
('AZ','CA-47','CA-49','CO','CT-5','GA','IA-3','IL-14','KS-3','MI','NH','NJ-3','NV','OH-9','PA','VA-2','VA-7','WA-8','
WI')) b
    left join (select *, '1' as univ_flag from model_scores.model_base_20220818_final left join
model_scores.household using(id,state)) using(id,state)
            left join voterfile_install.district d using(id,state)
            group by 1,2

            union all

            select
            '02 gender' as var
            ,b.gender_female::varchar as level
            ,sum(univ_flag) as n_dist
            ,sum(univ_flag)/sum(sum(univ_flag)) over()::float as pct_dist
    ,sum(univ_flag)/sum(case when b.sporadic_dem_target =1 or b.swing_target =1 then 1 else 0
end)::float as concentration
    ,count(distinct householdid1) as hh_count
    ,count(distinct householdid1)/sum(count(distinct householdid1)) over()::float as hh_dist

            from (select * from model_scores.univ_base where target_geo2 in
('AZ','CA-47','CA-49','CO','CT-5','GA','IA-3','IL-14','KS-3','MI','NH','NJ-3','NV','OH-9','PA','VA-2','VA-7','WA-8','
WI')) b
    left join (select *, '1' as univ_flag from model_scores.model_base_20220818_final left join
model_scores.household using(id,state)) using(id,state)
            left join voterfile_install.district d using(id,state)
            group by 1,2

            union all

            select
            '03 race/ethnicity' as var
            ,b.demo_combined_ethnicity_4way as level
            ,sum(univ_flag) as n_dist
            ,sum(univ_flag)/sum(sum(univ_flag)) over()::float as pct_dist
    ,sum(univ_flag)/sum(case when b.sporadic_dem_target =1 or b.swing_target =1 then 1 else 0
end)::float as concentration
    ,count(distinct householdid1) as hh_count
    ,count(distinct householdid1)/sum(count(distinct householdid1)) over()::float as hh_dist

            from (select * from model_scores.univ_base where target_geo2 in
('AZ','CA-47','CA-49','CO','CT-5','GA','IA-3','IL-14','KS-3','MI','NH','NJ-3','NV','OH-9','PA','VA-2','VA-7','WA-8','
WI')) b
    left join (select *, '1' as univ_flag from model_scores.model_base_20220818_final left join
model_scores.household using(id,state)) using(id,state)
            left join voterfile_install.district d using(id,state)
            group by 1,2

            union all

            select
            '04 income' as var
            ,b.demo_income_bucket_full as level
            ,sum(univ_flag) as n_dist
            ,sum(univ_flag)/sum(sum(univ_flag)) over()::float as pct_dist
    ,sum(univ_flag)/sum(case when b.sporadic_dem_target =1 or b.swing_target =1 then 1 else 0
end)::float as concentration
```

```
    ,count(distinct householdid1) as hh_count
    ,count(distinct householdid1)/sum(count(distinct householdid1)) over()::float as hh_dist

            from (select * from model_scores.univ_base where target_geo2 in
('AZ','CA-47','CA-49','CO','CT-5','GA','IA-3','IL-14','KS-3','MI','NH','NJ-3','NV','OH-9','PA','VA-2','VA-7','WA-8','
WI')) b
    left join (select *, '1' as univ_flag from model_scores.model_base_20220818_final left join
model_scores.household using(id,state)) using(id,state)
            left join voterfile_install.district d using(id,state)
            group by 1,2

            union all

            select
            '05 party' as var
            ,b.demo_combined_party as level
            ,sum(univ_flag) as n_dist
            ,sum(univ_flag)/sum(sum(univ_flag)) over()::float as pct_dist
    ,sum(univ_flag)/sum(case when b.sporadic_dem_target =1 or b.swing_target =1 then 1 else 0
end)::float as concentration
    ,count(distinct householdid1) as hh_count
    ,count(distinct householdid1)/sum(count(distinct householdid1)) over()::float as hh_dist

            from (select * from model_scores.univ_base where target_geo2 in
('AZ','CA-47','CA-49','CO','CT-5','GA','IA-3','IL-14','KS-3','MI','NH','NJ-3','NV','OH-9','PA','VA-2','VA-7','WA-8','
WI')) b
    left join (select *, '1' as univ_flag from model_scores.model_base_20220818_final left join
model_scores.household using(id,state)) using(id,state)
            left join voterfile_install.district d using(id,state)
            where b.demo_combined_party in ('D','I')
            group by 1,2

            union all

            select
            '06 urbanicity' as var
            ,b.catalistsynthetic_urbanity as level
            ,sum(univ_flag) as n_dist
            ,sum(univ_flag)/sum(sum(univ_flag)) over()::float as pct_dist
    ,sum(univ_flag)/sum(case when b.sporadic_dem_target =1 or b.swing_target =1 then 1 else 0
end)::float as concentration
    ,count(distinct householdid1) as hh_count
    ,count(distinct householdid1)/sum(count(distinct householdid1)) over()::float as hh_dist

            from (select * from model_scores.univ_base where target_geo2 in
('AZ','CA-47','CA-49','CO','CT-5','GA','IA-3','IL-14','KS-3','MI','NH','NJ-3','NV','OH-9','PA','VA-2','VA-7','WA-8','
WI')) b
    left join (select *, '1' as univ_flag from model_scores.model_base_20220818_final left join
model_scores.household using(id,state)) using(id,state)
            left join voterfile_install.district d using(id,state)
            group by 1,2) a
```

Prompt 2

Below is some code that gets used when creating a analytics basetable. How might someone use dbt to simplify or avoid duplicative code

```
,case when a.model_a > 80 then 'a80to100'
            when a.model_a > 60 then 'b60to80'
            when a.model_a > 40 then 'c40to60'
            when a.model_a > 20 then 'd20to40'
            when a.model_a > 0 then 'e0to820'
            else 'unknown' end as model_a_buckets

,case when a.model_b > 80 then 'a80to100'
            when a.model_b > 60 then 'b60to80'
            when a.model_b > 40 then 'c40to60'
            when a.model_b > 20 then 'd20to40'
            when a.model_b > 0 then 'e0to820'
            else 'unknown' end as model_b_buckets

,case when a.model_c > 80 then 'a80to100'
            when a.model_c > 60 then 'b60to80'
            when a.model_c > 40 then 'c40to60'
            when a.model_c > 20 then 'd20to40'
            when a.model_c > 0 then 'e0to820'
            else 'unknown' end as model_c_buckets

,case when a.model_d > 80 then 'a80to100'
            when a.model_d > 60 then 'b60to80'
            when a.model_d > 40 then 'c40to60'
            when a.model_d > 20 then 'd20to40'
            when a.model_d > 0 then 'e0to820'
            else 'unknown' end as model_d_buckets
```