SQL - Case Studies

Alessandro Paiva da Silva

2023

Summary

Introduction

I created this document to highlight my abilities in SQL, since most of the examples are from sensitive data from the Brazilian government, I had to censor some results or table names.

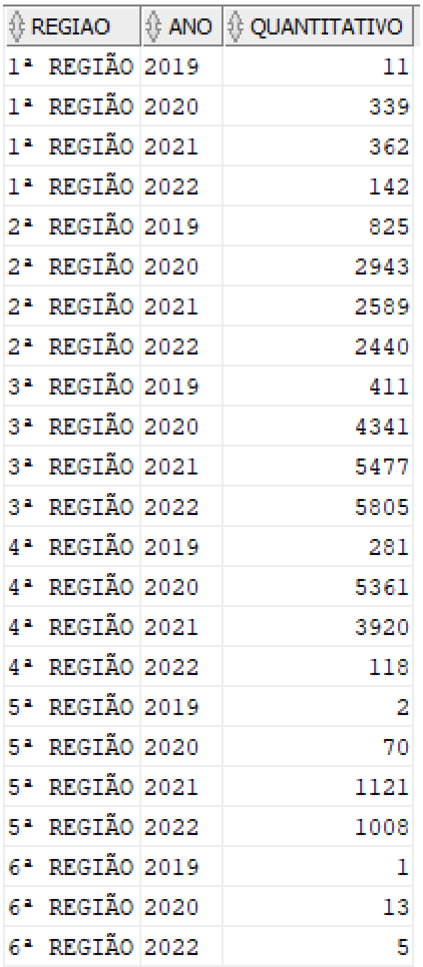
STF

Request from the Brazilian Supreme Federal Court to provide some data to analyze a lawsuit with a fiscal impact of more than **R$470 billion**. It consists of searching for every model ID provided and counting them by region and year.

```sql
/* Each region is composed of a varied number of states currently, there are six regions of
justice in Brazil, as this data is not available within the database and to avoid a very long
GROUP BY clause, and to improve readability, I decided to use a CTE to perform the
transformation of the data. */

-- The chosen field was UF, which is an abbreviation of the state name.


WITH regiao AS(
  SELECT [redacted] AS id,
    CASE
       WHEN v_e.uf IN ('AC', 'AM', 'AP', 'BA', 'DF', 'GO', 'MA',
                       'MT', 'PA', 'PI', 'RO', 'RR', 'TO')
       THEN '1ª REGIÃO'
       WHEN v_e.uf IN ('ES', 'RJ')
       THEN '2ª REGIÃO'
       WHEN v_e.uf IN ('MS', 'SP')
       THEN '3ª REGIÃO'
       WHEN v_e.uf IN ('PR', 'RS', 'SC')
       THEN '4ª REGIÃO'
       WHEN v_e.uf IN ('AL', 'CE', 'PB', 'PE', 'RN', 'SE')
       THEN '5ª REGIÃO'
       WHEN v_e.uf = ('MG')
       THEN '6ª REGIÃO'
    END AS regiao
  FROM [redacted])
```

```sql
-- With the CTE ready, I connected it, counted, and grouped the data.

SELECT TO_CHAR([redacted], 'YYYY') AS ano,
        regiao.regiao,
        COUNT ([redacted]) AS quantitativo
FROM [redacted]
INNER JOIN [redacted] ON [redacted] = [redacted]
INNER JOIN [redacted] ON [redacted] = [redacted]
INNER JOIN [redacted] ON [redacted] = [redacted]
INNER JOIN [redacted] ON [redacted] = [redacted]
INNER JOIN [redacted] ON [redacted] = [redacted]
INNER JOIN [redacted] ON [redacted] = [redacted]
INNER JOIN [redacted] ON [redacted] = [redacted]
WHERE [redacted] IN     (335202, 189308, 189305, 187546,
                        186488, 220289, 189026, 160296, 188292,
                        305101, 207865, 288423, 305179, 333962,
                        399000, 309723, 341956, 292438, 209726,
                        251054, 376162, 209725, 137150, 315401,
                        397960, 450649, 450637, 214401, 311420,
                        188312, 193854, 209774, 224186, 259378,
                        291681, 242480, 300847, 291649, 302280,
                        305642, 310359, 315187, 325131, 327160,
                        322222, 333047, 368186, 395462, 397549,
                        410203, 189249, 445229, 450650, 450684,
                        450683, 450638, 450641, 324160, 324446)
AND [redacted] >= TO_TIMESTAMP ('01/01/2019 00:00:00', 'DD/MM/YYYY HH24:MI:SS')
AND [redacted] < TO_TIMESTAMP ('24/11/2022 00:00:00', 'DD/MM/YYYY HH24:MI:SS')
GROUP BY regiao.regiao, TO_CHAR([redacted], 'YYYY')
ORDER BY regiao.regiao, TO_CHAR([redacted], 'YYYY') ASC
```

Result



REGIAO	ANO	QUANTITATIVO
1ª REGIÃO	2019	11
1ª REGIÃO	2020	339
1ª REGIÃO	2021	362
1ª REGIÃO	2022	142
2ª REGIÃO	2019	825
2ª REGIÃO	2020	2943
2ª REGIÃO	2021	2589
2ª REGIÃO	2022	2440
3ª REGIÃO	2019	411
3ª REGIÃO	2020	4341
3ª REGIÃO	2021	5477
3ª REGIÃO	2022	5805
4ª REGIÃO	2019	281
4ª REGIÃO	2020	5361
4ª REGIÃO	2021	3920
4ª REGIÃO	2022	118
5ª REGIÃO	2019	2
5ª REGIÃO	2020	70
5ª REGIÃO	2021	1121
5ª REGIÃO	2022	1008
6ª REGIÃO	2019	1
6ª REGIÃO	2020	13
6ª REGIÃO	2022	5

Internal Request

This request was sent from one of our internal teams, and they needed some models were labeled based on the id number and according to the court of the lawsuit.

```
/* To avoid a very long GROUP BY clause, and to improve readability, I decided to use a CTE to
perform the labeling process of each model. */

WITH etq AS(
   SELECT [redacted] AS id,
          [redacted] AS modelo,
          CASE
                WHEN v_mod.id = 247154 THEN 'FAV PRES'
                WHEN v_mod.id = 350224 THEN 'DEVOLUÇÃO STF'
                WHEN v_mod.id = 243522 THEN 'Súm 42 TNU'
                WHEN v_mod.id = 243503 THEN 'Súm 43 TNU'
                WHEN v_mod.id = 243488 THEN 'QO 22 TNU'
                WHEN v_mod.id = 249654 THEN 'QO 24 TNU'
                WHEN v_mod.id = 243484 THEN 'QO 10 TNU'
                WHEN v_mod.id = 243454 THEN 'QO 13 TNU'
                WHEN v_mod.id = 243485 THEN 'QO 18 TNU'
                WHEN v_mod.id = 243481 THEN 'REP CONT TNU' END AS etiqueta
   FROM [redacted])

/* That was my chance to do a different approach with the CASE aggregating function to count
the data instead of labeling it. */

SELECT  etq.modelo,
        etq.etiqueta,
        COUNT(CASE WHEN v_tr.id = 1 THEN etq.modelo ELSE NULL END) AS STF,
        COUNT(CASE WHEN v_tr.id = 3 THEN etq.modelo ELSE NULL END) AS STJ,
        COUNT(CASE WHEN v_tr.id = 10 THEN etq.modelo ELSE NULL END) AS TST,
        COUNT(CASE WHEN v_tr.id = 95 THEN etq.modelo ELSE NULL END) AS TNU,
        TO_CHAR([redacted],'MONTH','NLS_DATE_LANGUAGE= ''BRAZILIAN
        PORTUGUESE''') AS mes_modelo,
        [redacted] AS estado
```

```
FROM [redacted]
INNER JOIN [redacted] ON [redacted] = [redacted]
INNER JOIN [redacted] ON [redacted] = [redacted]
INNER JOIN [redacted] ON [redacted] = [redacted]
LEFT JOIN [redacted] ON [redacted] = [redacted]
LEFT JOIN [redacted] ON [redacted] = [redacted]
INNER JOIN [redacted] ON [redacted] = [redacted]
LEFT JOIN [redacted] ON [redacted] = [redacted]
LEFT JOIN [redacted] ON [redacted] = [redacted]
INNER JOIN [redacted] ON [redacted] = [redacted]
LEFT JOIN [redacted] ON [redacted] = [redacted]
LEFT JOIN [redacted] ON [redacted] = [redacted]
LEFT JOIN [redacted] ON etq.id = [redacted]
WHERE v_tr.id IN (1      /*STF*/,
                  3      /*STJ*/,
                  10     /*TST*/,
                  95     /*TNU*/)
AND EXTRACT (YEAR FROM [redacted]) >= '2022'
/* Most of the time I like to add some comments, since most of my filters are done using ids. I
remember that someone told me that this decreases the time of processing, and that's how I
learned. */
AND etq.id IN   (247154/*[redacted]*/,
                 350224 /*[redacted]*/,
                 243522 /*[redacted]*/,
                 243503 /*[redacted]*/,
                 243488 /*[redacted]*/,
                 249654 /*[redacted]*/,
              243484 /*[redacted]*/,
              243454 /*[redacted]*/,
              243485 /*[redacted]*/,
              243481 /*[redacted]*/)
AND v_s.id =  24630  /*[redacted]*/
GROUP BY      etq.modelo,
              etq.etiqueta,
              TO_CHAR([redacted],'MONTH','NLS_DATE_LANGUAGE= ''BRAZILIAN PORTUGUESE''')
              v_e.nome
```

Result