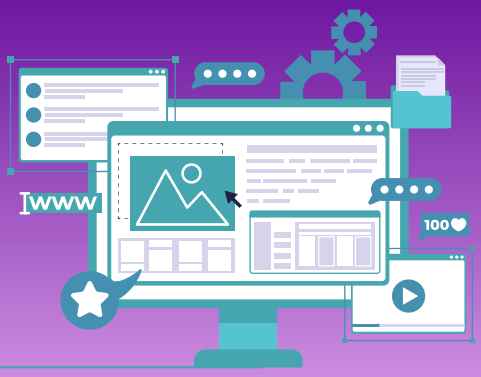
CONSULTA SQL

BIGQUERY - GCP



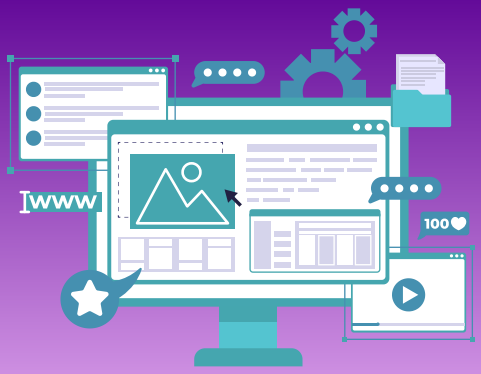
NADJA GUERRA

BIGQUERY

- Vá em algum banco de dados público e selecione aquele de tema da sua escolha
- Crie/selecione um projeto
- Abra BigQuery
- Suba seu conjunto para BigQuery (seja via local ou cloud storage)
- Monte o conjunto de dados e crie a tabela com permissão para letras maiúsculas e minúsculas (opcional);
- Faça alterações caso julgar necessário (eliminação/seleção/renomeação de colunas)
- Faça no mínimo de 5 consultas de negócio;
- Explique as consultas realizadas;
- Documente o que foi feito;
- Entregue no classroom;
- Pode fazer em grupo ou individual;

CONSULTA SQL

BIGQUERY - GCP



BOLSAS E AUXÍLIOS PAGOS PELO CNPQ

O CNPq concede bolsas para a formação de recursos humanos no campo da pesquisa científica e tecnológica, em universidades, institutos de pesquisa, centros tecnológicos e de formação profissional, tanto no Brasil como no exterior. Além de promover a formação de recursos humanos em áreas estratégicas para o desenvolvimento nacional, o CNPq aporta recursos financeiros para a implementação de projetos, programas e redes de Pesquisa e Desenvolvimento (P&D), diretamente ou em parceria com os Estados da Federação. O CNPq investe, ainda, em ações de divulgação científica e tecnológica com apoio financeiro à editoração e publicação de periódicos, à promoção de eventos científicos e à participação de estudantes e pesquisadores nos principais congressos e eventos nacionais e internacionais na área de ciência e tecnologia. Público Alvo: Jovens de ensino médio e superior, em nível de pós-graduação, interessados em atuar na pesquisa cientifica, e especialistas para atuarem em Pesquisa e Desenvolvimento (P&D) nas empresas e centros tecnológicos. Modalidades de bolsas: O CNPq oferece varias modalidades de bolsas de formação e fomento a pesquisa, a alunos de ensino médio, graduação, pós-graduação, recém-doutores e pesquisadores já experientes do País e do exterior. As bolsas são concedidas diretamente pelo CNPq ou por instituições de ensino e pesquisa para as quais o CNPq destina quotas de bolsas.

Organização
Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

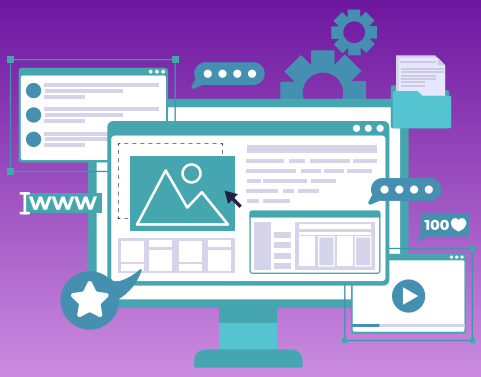
Cobertura temporal

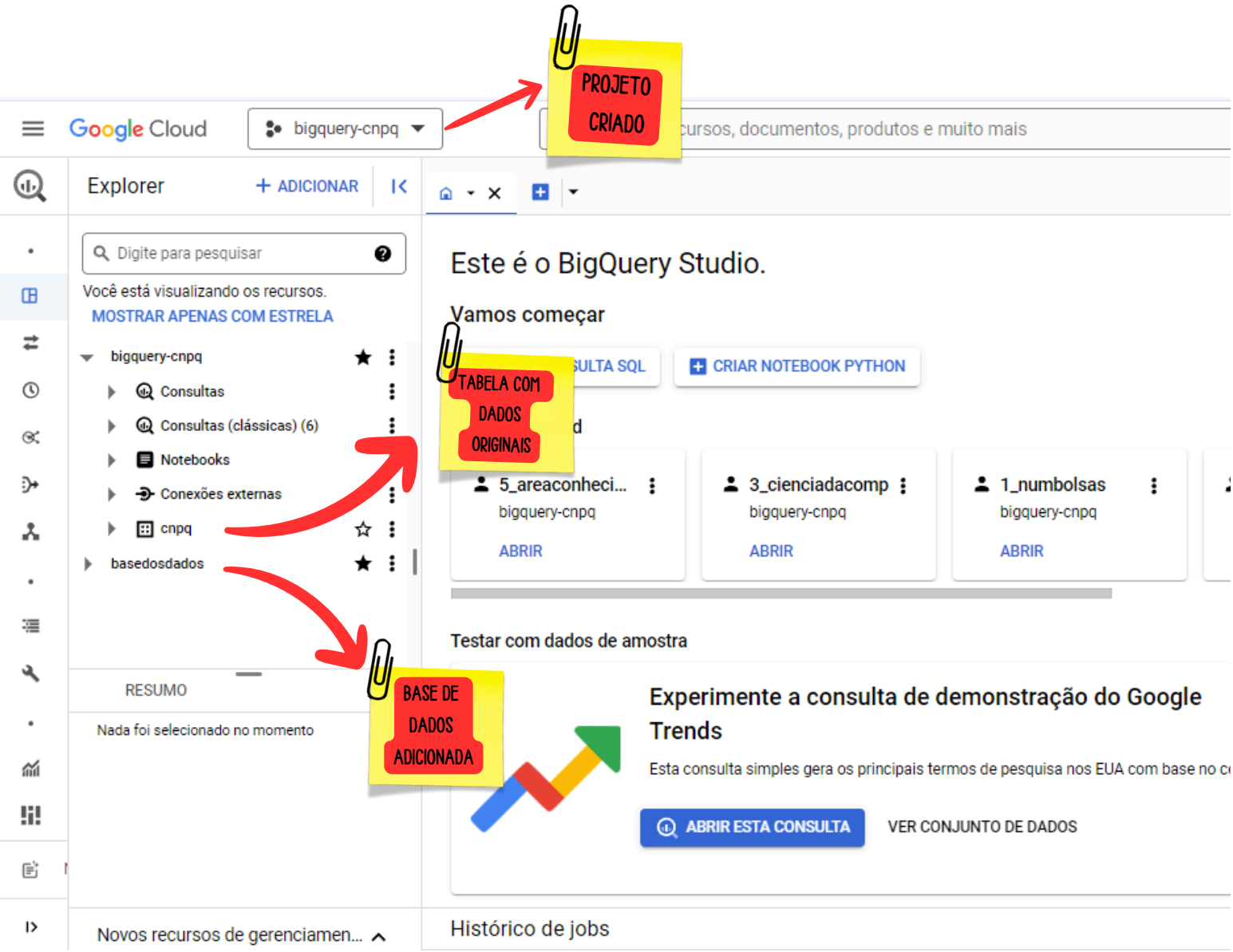
2002 --- 2022

Fonte: Base De Dados

CONSULTA SQL

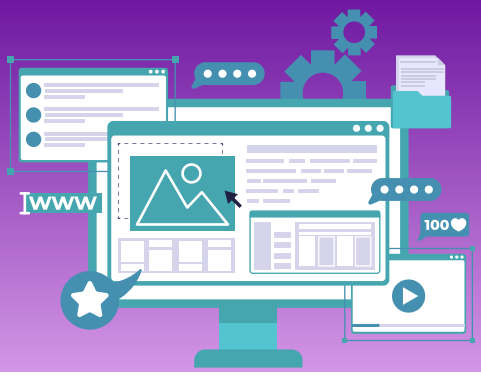
BIGQUERY - GCP





CONSULTA SQL

BIGQUERY - GCP



Consulta para identificar as áreas de conhecimento com maior número de bolsas concedidas na UFPE em 2022.

```
-- Selecionando as colunas de interesse (area_conhecimento) e fazendo a contagem de bolsas para cada área.
SELECT
   area_conhecimento,
   COUNT(*) AS quant_bolsas
-- Tabela que contém os dados de bolsas do CNPq
FROM `bigquery-cnpq.cnpq.dados_originais`
-- Filtrando por UFPE:
WHERE sigla_instituicao_destino LIKE '%UFPE%'
-- Filtrando pelo período todo de  2022
   AND data_inicio_processo BETWEEN '2022-01-01' AND '2022-12-31'
-- Agrupando por área para contabilizar as bolsas
GROUP BY area_conhecimento
-- Ordenando em ordem decrescente de acordo com o número de bolsas
ORDER BY quant_bolsas DESC
-- Limitando o resultado:
LIMIT 10;
```

1_numbolsas

EXECUTAR | SALVAR CONSULTA (MODO CLÁSSICO) | COMPARTILHAR | PROGRAMAÇÃO

```
SELECT
  area_conhecimento,
  COUNT(*) AS quant_bolsas
FROM `bigquery-cnpq.cnpq.dados_originais`
WHERE sigla_instituicao_destino LIKE '%UFPE%'
    AND data_inicio_processo BETWEEN '2022-01-01' AND '2022-12-31'
GROUP BY area_conhecimento
ORDER BY quant_bolsas DESC
LIMIT 10;
```

Pressione Alt+F1 para abrir as opções de acessibil

Resultados da consulta

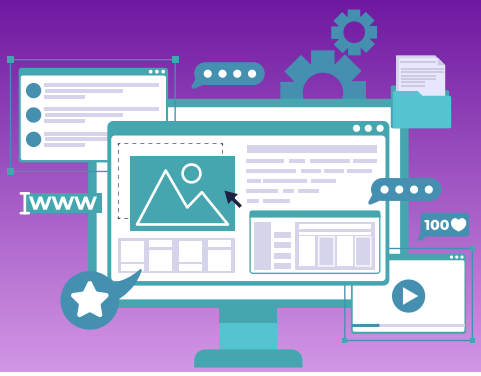
SALVAR RESULTADOS | EXPLORAR DADOS

INFORMAÇÕES DO JOB | RESULTADOS | GRÁFICO | JSON | DETALHES DA EXECUÇÃO | GRÁFICO DE EXECUÇÃO

Linha	area_conhecimento	quant_bolsas
1	BOTÂNICA	330
2	EDUCAÇÃO	157
3	CIÊNCIA DA COMPUTAÇÃO	125
4	ODONTOLOGIA	112
5	QUÍMICA	106
6	NUTRIÇÃO	93
7	ENFERMAGEM	90
8	FARMÁCIA	84
9	MEDICINA	83
10	ENGENHARIA SANITÁRIA	82

CONSULTA SQL

BIGQUERY - GCP



Instituições em **todo país** com maior número de bolsas concedidas de 2019 à 2022

```
-- Selecionando a coluna das Unidades Federativas e fazendo a contagem de bolsas.
SELECT
    sigla_uf_origem,
    COUNT(*) AS quant_bolsas
-- Tabela que contém os dados de bolsas do CNPq
FROM `bigquery-cnpq.cnpq.dados_originais`
-- Filtrando pelo período entre 2019 e 2022
WHERE data_inicio_processo BETWEEN '2019-01-01' AND '2022-12-31'
-- Agrupando por estado para contabilizar as bolsas
GROUP BY sigla_uf_origem
-- Ordenando em ordem decrescente de acordo com o número de bolsas
ORDER BY quant_bolsas DESC
-- Limitando o resultado:
LIMIT 10;
```

2_instituicoes | EXECUTAR | SALVAR CONSULTA (MODO CLÁSSICO) | COMPARTILHAR | PROGRAMAÇÃO | MAIS

```
1  -- Instituições com maior número de bolsas concedidas:
2  -- Selecionando a coluna das Unidades Federativas e fazendo a contagem de bolsas.
3  SELECT
4      sigla_uf_origem,
5      COUNT(*) AS quant_bolsas
6  -- Tabela que contém os dados de bolsas do CNPq
7  FROM `bigquery-cnpq.cnpq.dados_originais`
8  -- Filtrando pelo período entre 2019 e 2022
9  WHERE  data_inicio_processo BETWEEN '2019-01-01' AND '2022-12-31'
10 -- Agrupando por estado para contabilizar as bolsas
11 GROUP BY sigla_uf_origem
12 -- Ordenando em ordem decrescente de acordo com o número de bolsas
13 ORDER BY quant_bolsas DESC
14 -- Limitando o resultado:
15 LIMIT 10;
```

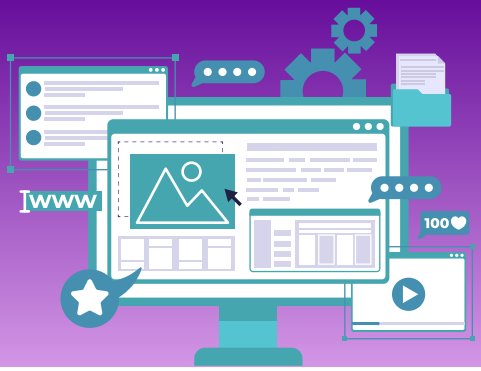
Resultados da consulta

INFORMAÇÕES DO JOB | RESULTADOS | GRÁFICO | JSON | DETALHES DA EXECUÇÃO | GRÁFICO DE EXECUÇÃO

Linha	sigla_uf_origem	quant_bolsas
1	RJ	43567
2	SP	41453
3	DF	20159
4	MG	18640
5	RS	17169
6	PR	11260
7	PE	7729
8	SC	7356
9	BA	7015
10	CE	5752

CONSULTA SQL

BIGQUERY - GCP



CONSULTA 3

Consulta para obter o total de registros de pesquisas na área de Ciência da Computação realizadas nos estados de Pernambuco (PE) e São Paulo (SP) durante o ano de 2022. Apresentando os resultados separadamente para cada estado.

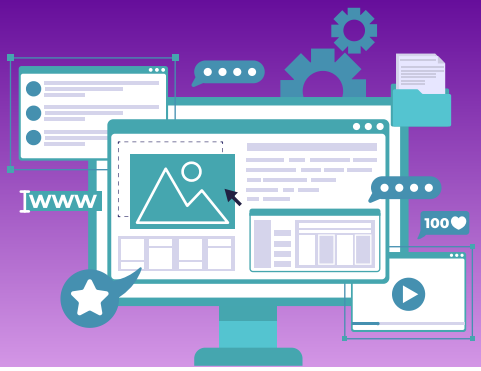
```
-- Consulta para obter o total de registros de Ciência da Computação em Pernambuco
SELECT 'Total de Ciência da Computação em PE' AS estado, COUNT(*) AS total
FROM `bigquery-cnpq.cnpq.dados_originais`
-- Seleciona a tabela de dados originais do CNPq no BigQuery
WHERE data_fim_processo BETWEEN '2022-01-01' AND '2022-12-31'
-- Filtra os registros com data de término do processo dentro do intervalo de tempo especificado
AND ano IS NOT NULL
-- Filtra os registros onde o ano não é nulo
AND sigla_uf_destino = 'PE'
-- Filtra os registros onde a sigla do estado de destino é 'PE' (Pernambuco)
AND area_conhecimento = 'CIÊNCIA DA COMPUTAÇÃO'
-- Filtra os registros onde a área de conhecimento é 'CIÊNCIA DA COMPUTAÇÃO'

UNION ALL

-- Consulta para obter o total de registros de Ciência da Computação em São Paulo
SELECT 'Total de Ciência da Computação em SP' AS estado, COUNT(*) AS total
FROM `bigquery-cnpq.cnpq.dados_originais`
-- Seleciona a tabela de dados originais do CNPq no BigQuery
WHERE data_fim_processo BETWEEN '2022-01-01' AND '2022-12-31'
-- Filtra os registros com data de término do processo dentro do intervalo de tempo especificado
AND ano IS NOT NULL
-- Filtra os registros onde o ano não é nulo
AND sigla_uf_destino = 'SP'
-- Filtra os registros onde a sigla do estado de destino é 'SP' (São Paulo)
AND area_conhecimento = 'CIÊNCIA DA COMPUTAÇÃO';
-- Filtra os registros onde a área de conhecimento é 'CIÊNCIA DA COMPUTAÇÃO'
```

CONSULTA SQL

BIGQUERY - GCP



CONSULTA 3

RESULTADOS

3_cienciadaco... EXECUTAR SALVAR CONSULTA (MODO CLÁSSICO) COMPARTILHAR

```sql
SELECT 'Ciência da Computação em Pernambuco' AS categoria, COUNT(*) AS total
FROM `bigquery-cnpq.cnpq.dados_originais`
WHERE data_fim_processo BETWEEN '2022-01-01' AND '2022-12-31'
AND ano IS NOT NULL
AND sigla_uf_destino = 'PE'
AND area_conhecimento = 'CIÊNCIA DA COMPUTAÇÃO'

UNION ALL

SELECT 'Ciência da Computação em São Paulo' AS estado, COUNT(*) AS total
FROM `bigquery-cnpq.cnpq.dados_originais`
WHERE data_fim_processo BETWEEN '2022-01-01' AND '2022-12-31'
AND ano IS NOT NULL
AND sigla_uf_destino = 'SP'
AND area_conhecimento = 'CIÊNCIA DA COMPUTAÇÃO';
```

Pressione Alt+

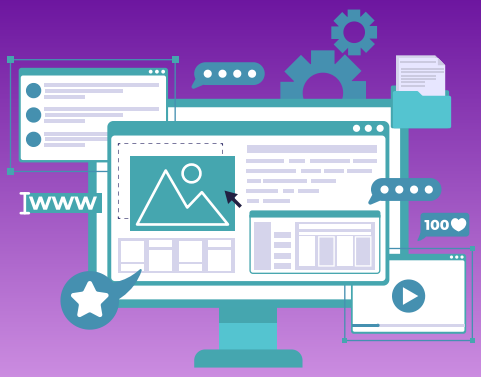
Resultados da consulta SALVAR RESULTADOS

INFORMAÇÕES DO JOB | RESULTADOS | GRÁFICO | JSON | DETALHES DA EXECUÇÃO | GRÁFICO

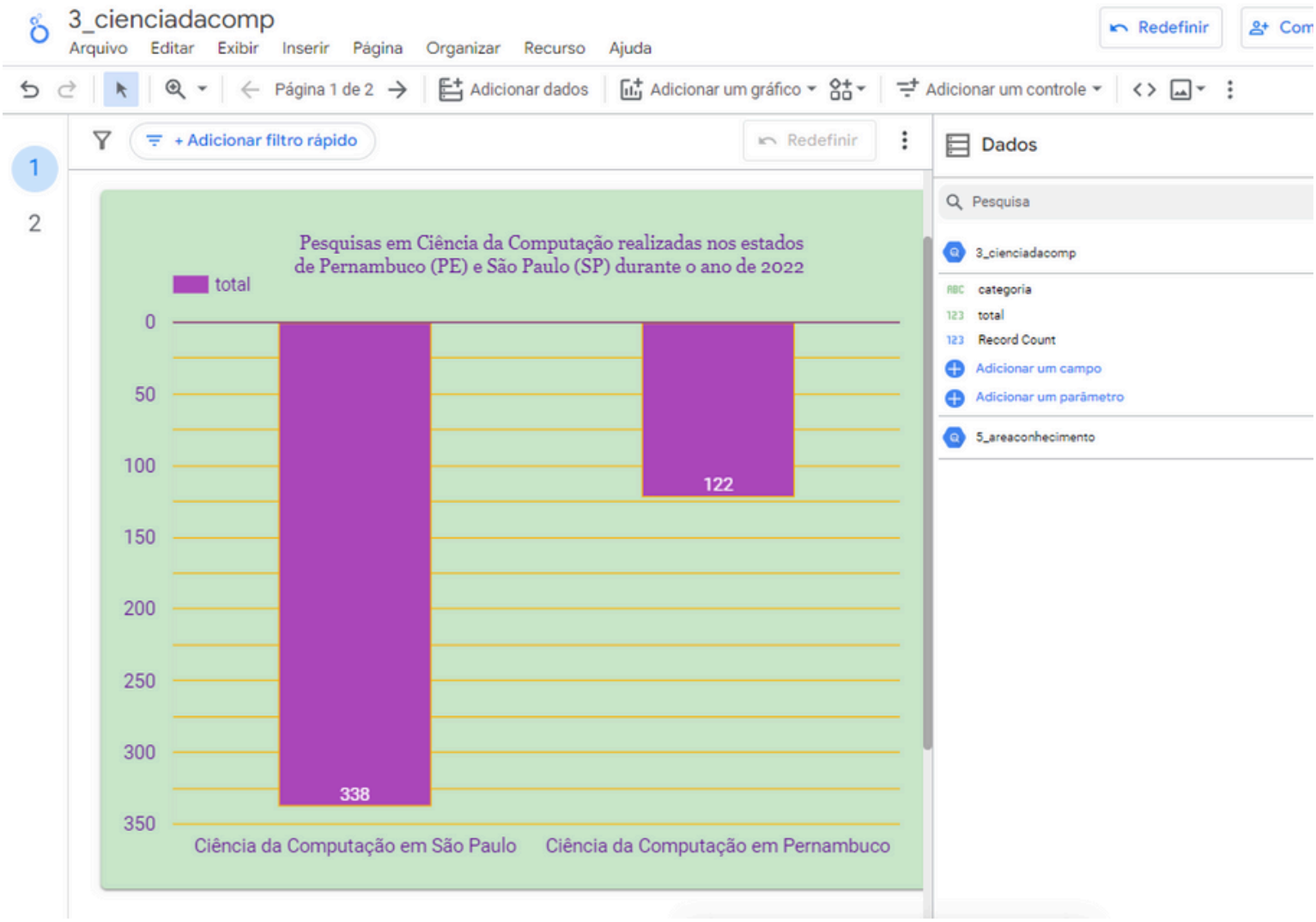
Linha	categoria	total
1	Ciência da Computação em São Paulo	338
2	Ciência da Computação em Pernambuco	122

CONSULTA SQL

BIGQUERY - GCP

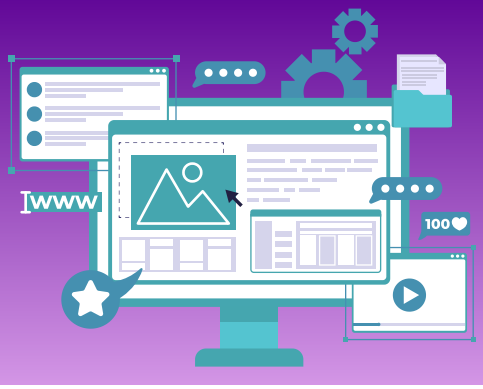


INTEGRAÇÃO COM O LOOKER



CONSULTA SQL

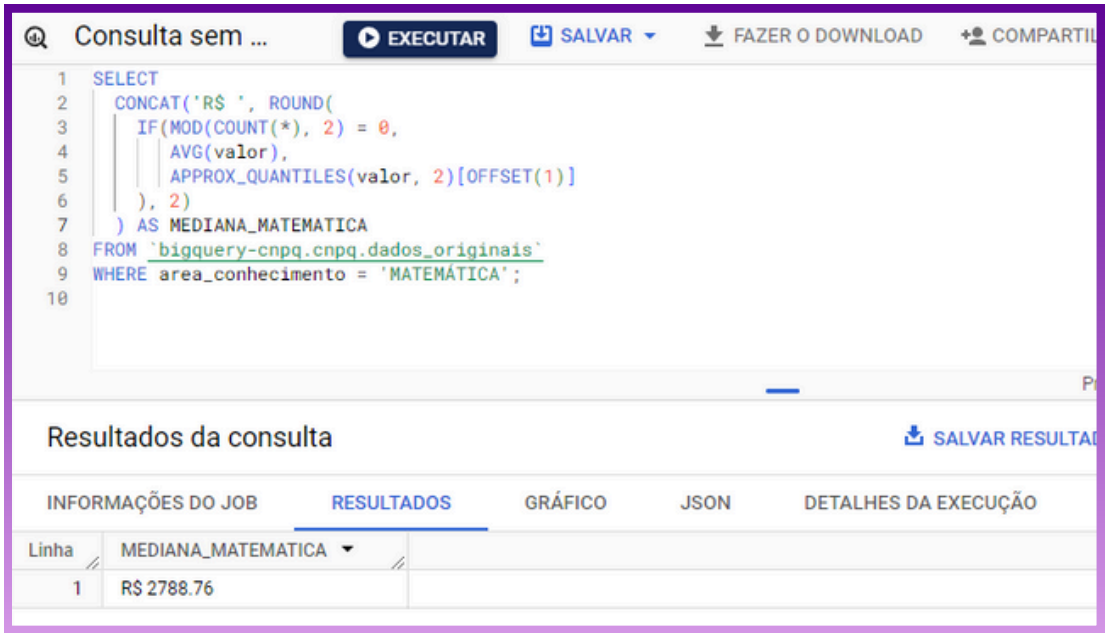
BIGQUERY - GCP



CONSULTA 4

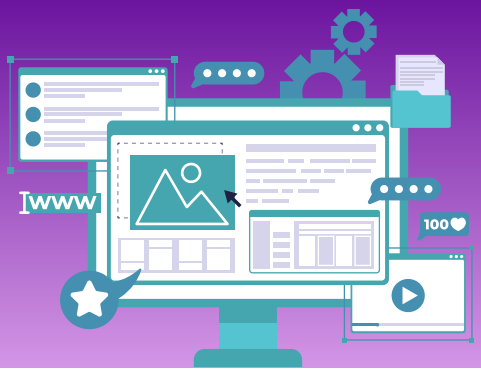
Consulta para calcular a mediana do valor em reais para a área de Matemática em todas as universidades (sem considerar que as outras áreas recebem o mesmo).

```
-- Calcula a mediana do valor em reais para a área de Matemática
SELECT
-- Concatena 'R$ ' ao valor arredondado
 CONCAT('R$ ', ROUND(
  -- Verifica se o número de registros é par
  IF(MOD(COUNT(*), 2) = 0,
    -- Calcula a média dos valores se o número de registros for par
    AVG(valor),
   -- Calcula a média dos valores se o número de registros for par e se for impar
    APPROX_QUANTILES(valor, 2)[OFFSET(1)]
  ), 2) -- Arredonda o valor para duas casas decimais
 ) AS mediana_valor_matematica_em_reais -- representando a mediana dos valores em reais
FROM `bigquery-cnpq.cnpq.dados_originais`
WHERE area_conhecimento = 'MATEMÁTICA'; -- Filtra os registros onde a área de conhecimento é 'MATEMÁTICA'
```



CONSULTA SQL

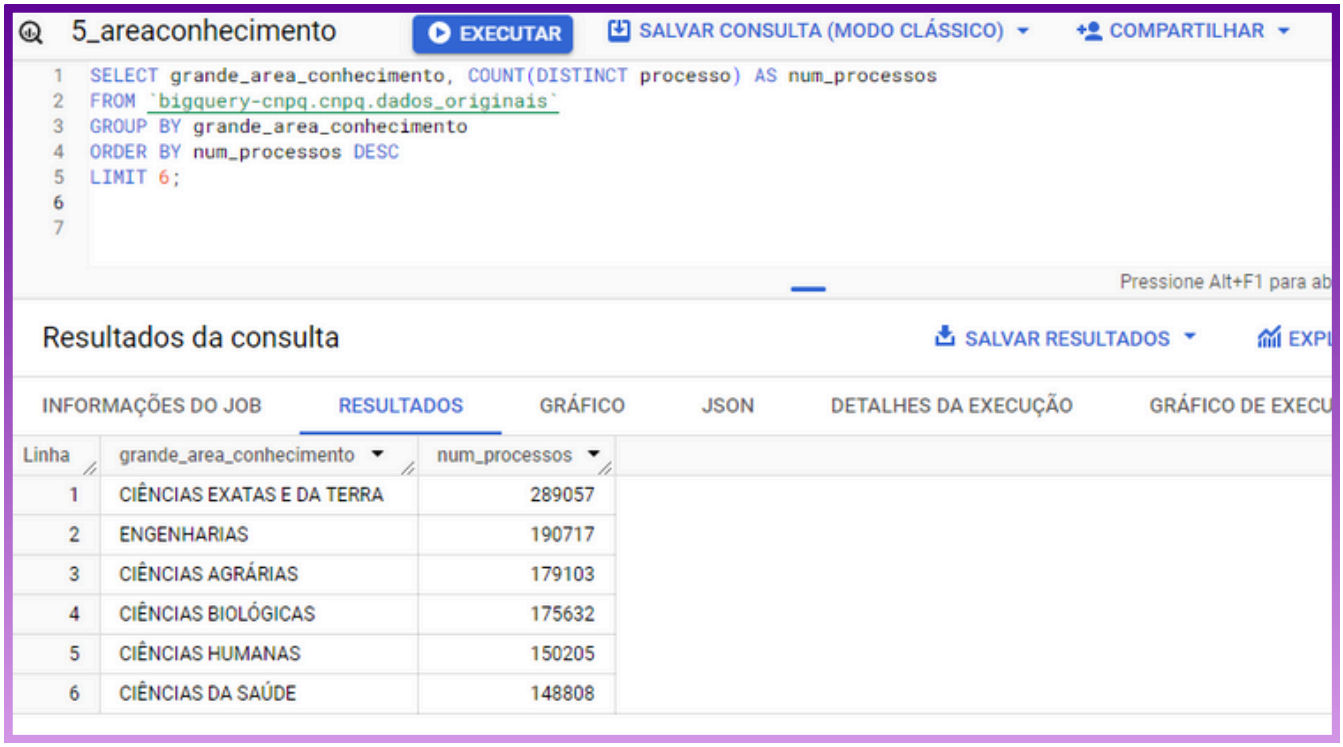
BIGQUERY - GCP



CONSULTA 5

Consulte as áreas de conhecimento com o maior número de processos.

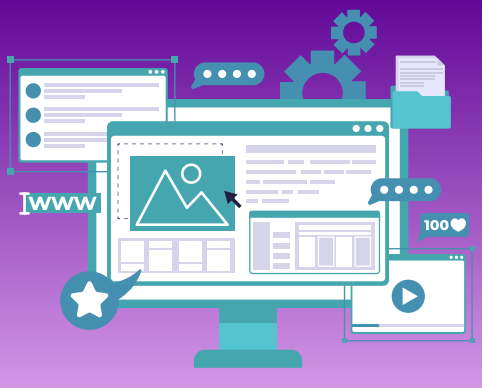
```
-- Seleciona a coluna 'grande_area_conhecimento' e conta o número de processos únicos em cada uma delas
SELECT grande_area_conhecimento, COUNT(DISTINCT processo) AS num_processos
-- Seleciona a tabela de dados originais do CNPq no BigQuery
FROM `bigquery-cnpq.cnpq.dados_originais`
-- Agrupa os dados pela coluna 'grande_area_conhecimento'
GROUP BY grande_area_conhecimento
-- Ordena os resultados em ordem decrescente com base no número de processos
ORDER BY num_processos DESC
-- Limita o resultado
LIMIT 6;
```



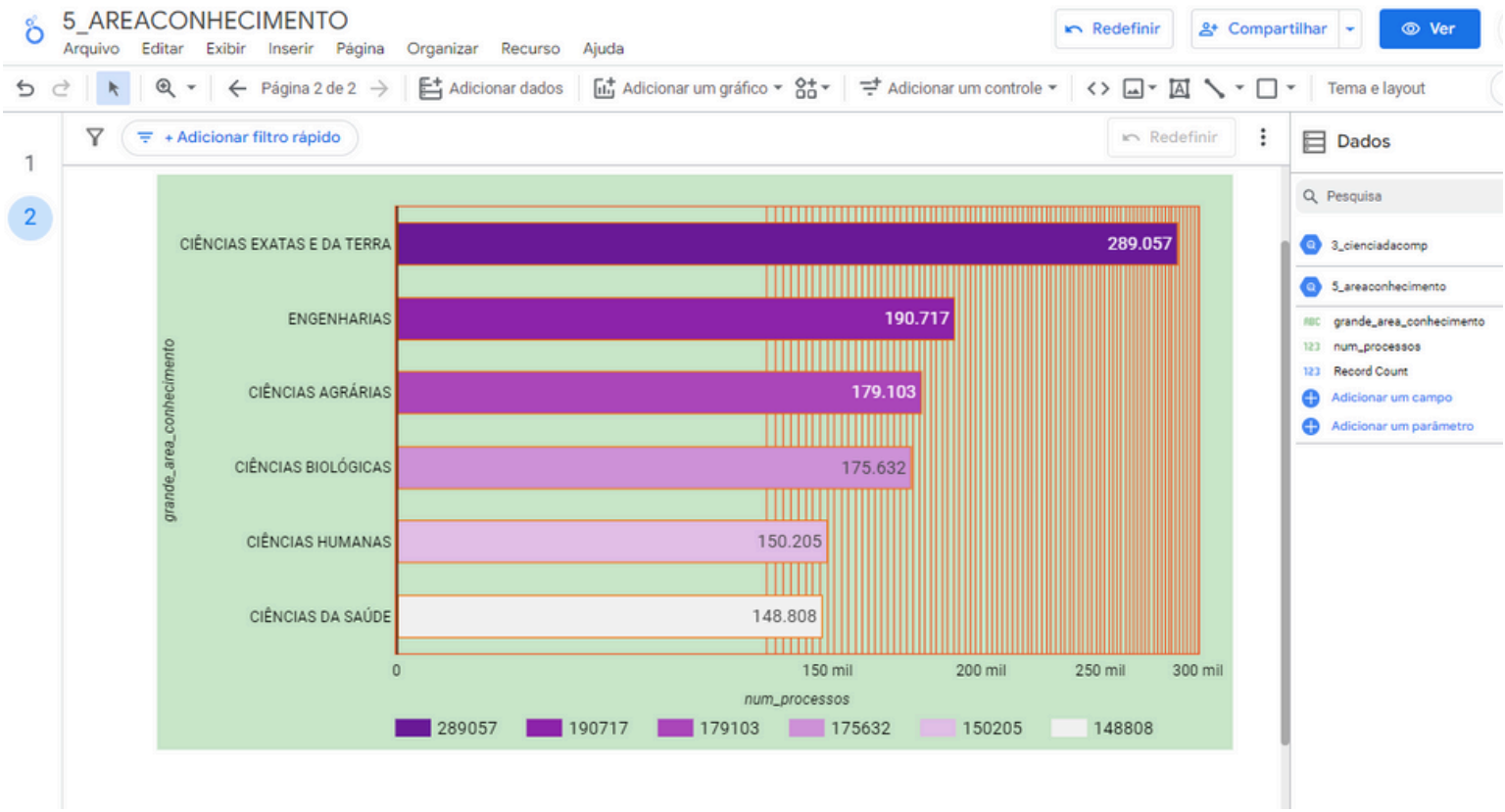
Linha	grande_area_conhecimento	num_processos
1	CIÊNCIAS EXATAS E DA TERRA	289057
2	ENGENHARIAS	190717
3	CIÊNCIAS AGRÁRIAS	179103
4	CIÊNCIAS BIOLÓGICAS	175632
5	CIÊNCIAS HUMANAS	150205
6	CIÊNCIAS DA SAÚDE	148808

CONSULTA SQL

BIGQUERY - GCP



CONSULTA 5

Consulte as áreas de conhecimento com o maior número de processos.

INTEGRAÇÃO COM O LOOKER