



Atividade Técnica


Continuação...






→ Pergunta 2

A	B	C	D	E	F	G	H	I	J
ID_DATASET	ID_PONTO_VENDA	NOME_PONTO_VENDA	PERFIL_PONTO_VENDA	DATA	ID_LINHA_PRODUTO	NOME_LINHA_PRODUTO	MARCA_LINHA_PRODUTO	TIPO_COLETA	VALOR
1021	115	INVOLVES	ATACADAO	01/09/2020	398	BISCOITOS SORTIDOS	INVOLVES ZERO	DISPONIBILIDADE	SIM
1022	115	INVOLVES	ATACADAO	02/09/2020	407	MARGARINA	INVOLVES ZERO	DISPONIBILIDADE	SIM
1023	115	INVOLVES	ATACADAO	03/09/2020	408	MANTEIGA	INVOLVES ZERO	DISPONIBILIDADE	SIM

O dataset enviado anexo ao email tem o nome **dataset_teste_de.csv**. Entendo que se trata do mesmo citado acima na descrição da questão, **PONTO_VENDA_UNIDADE**. Conforme o print da imagem acima, não há o campo **SELLIN** na tabela, pode ser que seja a coluna **VALOR**, mas não tenho certeza.

O dataset enviado tem apenas 40 linhas/registros, sendo que na coluna VALOR há campos 'SIM/NÃO' do tipo texto (20 linhas) e campos do tipo inteiro (min 0, max 6). O máximo da tabela é bem inferior ao solicitado na questão (20.000); portanto, mesmo que eu limpe e deixe apenas os registros de número inteiro, a tabela retorna vazia.

 Pergunta2



 RUN
  SAVE QUERY
  DOWNLOAD
  SHARE
  This query works

```

1 SELECT *
2 FROM `involves-422612.involves_datasets.involves-dataset`
3 WHERE VALOR NOT IN ('SIM', 'NÃO') AND CAST(VALOR AS INT64) > 20000
4 ORDER BY NOME_PONTO_VENDA;
5

```


Query results

 SAVE RESULTS
  EXPLORE DATA

JOB INFORMATION

RESULTS

 CHART
 JSON
 EXECUTION DETAILS
 EXECUTION GRAPH

 There is no data to display.

→ Pergunta 3

🔍	Pergunta3_distintos	▶ RUN	🕒 SCHEDULE	⚙️ MORE ▾	💾 SAVE QUERY ▾	⬇️ DOWNLOAD	👤
1	SELECT DISTINCT (ID_PONTO_VENDA, NOME_PONTO_VENDA)						
2	FROM `involves-422612.involves_datasets.involves-dataset`						
3							
Query results							📄 SAVE RESULTS ▾ 📊 EXPLORE DATA ▾ ↕
<div>JOB INFORMATIONRESULTSCHARTJSONEXECUTION DETAILSEXECUTION GRAPH</div>							
Row	f0__field_1 ▾	f0__field_2 ▾					
1	115	INVOLVES					
2	116	INVOLVES STAGE					

Conforme consulta acima, o campo NOME_PONTO_VENDA possui **apenas 2 valores distintos**, assim como ID_PONTO_VENDA. Por este motivo a consulta solicitada na questão retorna apenas a soma de VALOR para dois pontos de vendas.

Pergunta3

RUN

SCHEDULE

MORE

SAVE QUERY

DOWNLOAD

SHARE

```
1 SELECT
2   FORMAT('%02d%d', EXTRACT(MONTH FROM DATA), EXTRACT(YEAR FROM DATA)) AS ID_MES_ANO,
3   NOME_PONTO_VENDA,
4   SUM(CAST(VALOR AS INT64)) AS TOTAL_VALOR,
5   DATA,
6   EXTRACT(YEAR FROM DATA) AS ANO,
7   EXTRACT(MONTH FROM DATA) AS MES,
8   FROM `involves-422612.involves_datasets.involves-dataset`
9 WHERE VALOR NOT IN ('SIM', 'NAO')
10 GROUP BY NOME_PONTO_VENDA, DATA, EXTRACT(YEAR FROM DATA), EXTRACT(MONTH FROM DATA)
11 ORDER BY ID_MES_ANO ASC, NOME_PONTO_VENDA ASC
```

Query results

SAVE RESULTS

EXPLORE DATA

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS		EXECUTION GRAPH	
Row	ID_MES_ANO	NOME_PONTO_VENDA	TOTAL_VALOR	DATA	ANO	MES		
1	092020	INVOLVES	22	2020-09-01	2020	9		
2	092020	INVOLVES STAGE	25	2020-09-01	2020	9		

→ Pergunta 4

Não foi enviada uma tabela de nome VISITAS_PONTO_VENDA, então criei com dados aleatórios, para garantir que a consuta funcionaria.

🔍 Inserir_Visitas

RUN

SAVE QUERY

DOWNLOAD

SHARE

SCHEDULE

MORE

```
1 --Criar tabela
2 CREATE TABLE `involves-422612.involves_datasets.involves-dataset-visitas` (
3   ID_VISITA INT64 NOT NULL,
4   FK_PDV INT64 NOT NULL,
5   FL_VISITADO INT64 NOT NULL,
6   DATA_VISITA DATE NOT NULL
7 );
8
9 -- Se necessario limpar a tabela
10 TRUNCATE TABLE `involves-422612.involves_datasets.involves-dataset-visitas`
11
12 -- Inserir dados na tablea
13 INSERT INTO
14   `involves-422612.involves_datasets.involves-dataset-visitas` (ID_VISITA,
15     FK_PDV,
16     FL_VISITADO,
17     DATA_VISITA)
18 SELECT
19   id AS ID_VISITA,
20   CAST(115 + FLOOR(RAND() * 2) AS INT64) AS FK_PDV,
21   CAST(FLOOR(RAND() * 2) AS INT64) AS FL_VISITADO,
22   DATE_ADD('2024-01-01', INTERVAL CAST(FLOOR(RAND() * 365) AS INT64) DAY) AS DATA_VISITA,
23 FROM
24   UNNEST(GENERATE_ARRAY(1, 100)) AS id;
```

Resposta:

🔍 Pergunta4

RUN

SAVE QUERY

DOWNLOAD

SHARE

SCHEDULE

MORE

✔ This que...

```
1 SELECT
2   PDV_UNIDADES.NOME_PONTO_VENDA,
3   COALESCE(VISITAS_COUNT.TOTAL_VISITAS, 0) AS TOTAL_VISITAS
4 FROM
5   `involves-422612.involves_datasets.involves-dataset` AS PDV_UNIDADES
6 LEFT JOIN (
7   SELECT
8     FK_PDV,
9     COUNT(ID_VISITA) AS TOTAL_VISITAS
10  FROM
11    `involves-422612.involves_datasets.involves-dataset-visitas`
12  WHERE
13    FL_VISITADO = 1
14  GROUP BY
15    FK_PDV) AS VISITAS_COUNT
16 ON
17   PDV_UNIDADES.ID_PONTO_VENDA = VISITAS_COUNT.FK_PDV
18 WHERE
19   NOME_PONTO_VENDA = 'INVOLVES'
20 GROUP BY
21   PDV_UNIDADES.NOME_PONTO_VENDA,
22   TOTAL_VISITAS;
```

Query results

SAVE RESULTSEXPLORE DATA

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTION DETAILS	EXECUTION GRAPH
Row	NOME_PONTO_VENDA	TOTAL_VISITAS				
1	INVOLVES	22				

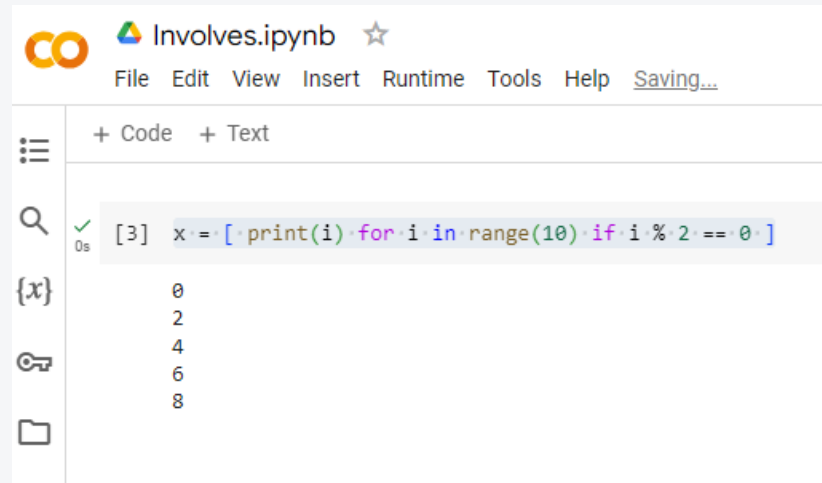
→ Pergunta 5

A indexação agiliza o encontro entre dois campos e evita a necessidade de uma consulta ser realizada linha por linha. Vários campos podem ser indexados, mas a prioridade na consulta apresentada deve ser para os campos que estão nas junções, nos filtros e nos agrupamentos. Abaixo, podemos observar que os campos FT.CICLO e FT.ID_DIM_PDV estão presentes em todas as junções e nos filtros. Portanto, são eles que precisam ser indexados com urgência, inclusive com a possibilidade de ser criado um índice composto.

```
Untitled query [RUN] [SAVE] [DO]

1 SELECT
2   FT.CICLO,
3   FT.ID_DIM_PDV,
4   FT.ID_BLOCO_ITEM,
5   SUM(FT.QTD_PONTO_EXTRA),
6   SUM(FTPI.TOTAL_NOTA_ITEM)
7 FROM
8   FT_DOMINANCIA_PONTO_EXTRA_COMPLIANCE FT
9 INNER JOIN
10  TABREF_PAINEL_LOJAS_LP TPLL
11 ON
12  FT.ID_DIM_PDV = TPLL.ID_DIM_PDV
13  AND FT.CICLO = TPLL.CICLO
14 INNER JOIN
15  FT_PERFECTSTORE_ITEM FTPI
16 ON
17  FT.CICLO = FTPI.CICLO
18  AND FT.ID_DIM_PDV = FTPI.ID_DIM_PDV
19  AND FT.ID_BLOCO_ITEM = FTPI.ID_BLOCO_ITEM
20  AND FT.SEMANA_LP = FTPI.SEMANA_LP
21 WHERE
22  FT.CICLO = 202009
23  AND FT.ID_DIM_PDV = 223459792
24 GROUP BY
25  FT.CICLO,
26  FT.ID_DIM_PDV;
```

→ Pergunta 6



The screenshot shows a Jupyter Notebook window titled "Involves.ipynb". The interface includes a menu bar (File, Edit, View, Insert, Runtime, Tools, Help, Saving...) and a sidebar with icons for file operations. The main area displays a code cell with the following Python code:

```
[3] x = [print(i) for i in range(10) if i % 2 == 0]
```

The output of the code is displayed below the cell:

```
0
2
4
6
8
```

→ Pergunta 7

```
✓ [11] # @title Default title text
5s    # Apresenta
      print ('Olá, vamos fazer a soma de dois números')

      # pede um 1º número
      num1 = float(input("Digite o 1º número: "))

      # pede um 2º número
      num2 = float(input("Digite o 2º número: "))

      # Calcula a soma dos dois números inseridos
      soma = num1 + num2

      # Imprime o resultado
      print("O resultado da soma é:", soma)
```

Olá, vamos fazer a soma de dois números
Digite o 1º número: 50
Digite o 2º número: 150
O resultado da soma é: 200.0

→ Pergunta 8

A pergunta a) não especifica o formato do ano e mês (ex. YY, YYYY-MM, MMM, etc..), deixei o padrão da função. Não encontrei o campo *perfil da linha de produto*, adicionei *MARCA_LINHA_PRODUTO*

```
Pergunta8_DIM_CALENDARIO  RUN SAVE QUERY ▾  
1 SELECT  
2   DATA,  
3   EXTRACT(YEAR  
4   FROM  
5   | DATA) AS ANO,  
6   EXTRACT(MONTH  
7   FROM  
8   | DATA) AS MES  
9 FROM  
10 | `involves-422612.involves_datasets.involves-dataset`
```

```
Pergunta8_DIM_LINHA_PRODU... RUN SAVE QI  
1 SELECT  
2   ID_LINHA_PRODUTO,  
3   NOME_LINHA_PRODUTO,  
4   MARCA_LINHA_PRODUTO  
5 FROM  
6 | `involves-422612.involves_datasets.involves-dataset`
```

```
Pergunta8_DIM_PDV RUN SAVE QUERY ▾ DOWNLOAD  
1 SELECT  
2   ID_PONTO_VENDA,  
3   NOME_PONTO_VENDA,  
4   PERFIL_PONTO_VENDA  
5 FROM  
6 | `involves-422612.involves_datasets.involves-dataset`
```

→ Pergunta 9

Não agreguei por DATA, porque conforme descricao apenas os dados de setembro são neccessarios.

```
Pergunta9_FT_DISPONIBILIDADE RUN
1 SELECT
2   ID_PONTO_VENDA,
3   ID_LINHA_PRODUTO,
4   COUNT(*) AS QTD_PRESENCA
5 FROM
6   `involves-422612.involves_datasets.involves-dataset`
7 WHERE
8   EXTRACT(YEAR
9     FROM
10      | DATA) = 2020
11   AND EXTRACT(MONTH
12     FROM
13      | DATA) = 9
14   AND VALOR = 'SIM'
15 GROUP BY
16   ID_PONTO_VENDA,
17   ID_LINHA_PRODUTO
18 ORDER BY
19   ID_LINHA_PRODUTO,
20   ID_PONTO_VENDA,
21   QTD_PRESENCA DESC;
```

Query results

JOB INFORMATION		RESULTS	CHART	JSON
Row	ID_PONTO_VENDA	ID_LINHA_PRODUTO	QTD_PRESENCA	
1	115	398	2	
2	115	407	1	
3	116	407	2	
4	115	408	1	
5	115	422	1	
6	116	422	2	
7	115	423	2	

Pergunta9_FT_DISPONIBILIDADE_AGREGA...

```
1 SELECT
2   ID_PONTO_VENDA,
3   COUNT(*) AS QTD_PRESENCIA
4 FROM
5   `involves-422612.involves_datasets.involves-dataset`
6 WHERE
7   EXTRACT(YEAR
8     FROM
9       DATA) = 2020
10  AND EXTRACT(MONTH
11    FROM
12      DATA) = 9
13  AND VALOR = 'SIM'
14 GROUP BY
15   ID_PONTO_VENDA
16 ORDER BY
17   ID_PONTO_VENDA,
18   QTD_PRESENCIA DESC;
19
```

Query results

JOB INFORMATION		RESULTS	CHART
Row	ID_PONTO_VENDA	QTD_PRESENCIA	
1	115	7	
2	116	4	