

Atividade Técnica

Continuação...

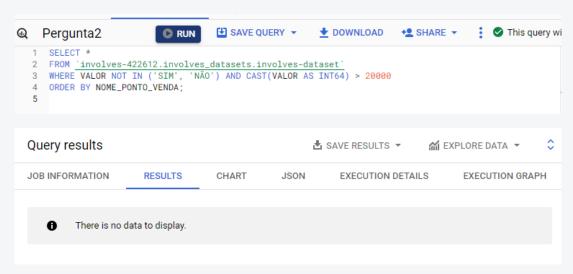
→ Pergunta 2

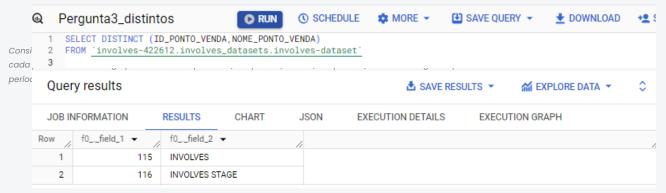
Crie uma query, considerando o SGBD MySQL, para exibir todos os dados de uma tabela de Pontos de Venda (tabela

origer										
(cam										
	ID_DATASET	ID_PONTO_VENDA	NOME_PONTO_VE	PERFIL_PONTO_VE	DATA	ID_LINHA_PRODUT	NOME_LINHA_PRO	MARCA_LINHA_PRO	TIPO_COLETA	VALOR
	1021	115	INVOLVES	ATACADAO	01/09/2020	398	BISCOITOS SORTIE	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

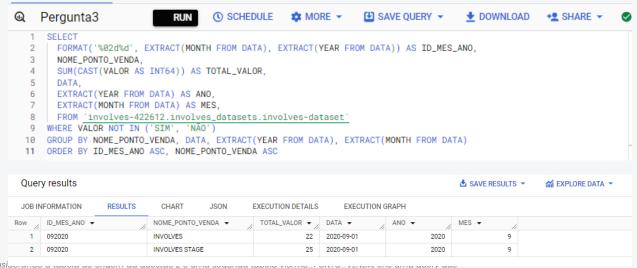
O dataset enviado anexo ao email tem o nome **dataset_teste_de.csv**. Entendo que se trata do mesmo citado acima na descrição da questão, **PONTO_VENDA_UNIDADE**. Conforme o print da imagem acima, não há o campo **SELLIN** na tabela, pode ser que seja a coluna **VALOR**, mas não tenho certeza.

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.





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.



calcule a quantidade de visitas do ponto de venda de nome INVOLVES, sabendo-se que a tabela de visitas possui um canyo perquinta aponto de venda foi visitado ou não chamado FL_VISITADO (Se 1 = Ponto de venda visitado / Se 0 = Ponto de venda não visitado). O campo chave que liga as duas tabelas é ID_PDV (na tabela PONTO_VENDA_UNIDADE) e FK_PDV (na tabela VISITAS_PONTO_VENDA). A query deve mostrar apenas as informações de nome do ponto de venda e quantidade de visitas realizadas.

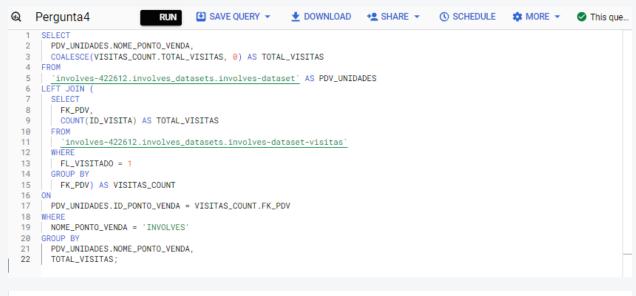
Não foi enviada uma tabela de nome VISITAS_PONTO_VENDA, entao criei com dados aleatórios, para garantir que a consuta funcionaria.

```
SAVE QUERY ▼

    Inserir_Visitas

                                 C RUN
                                                                1 --Criar tabela
  2 CREATE TABLE 'involves-422612.involves_datasets.involves-dataset-visitas' (
       ID_VISITA INT64 NOT NULL,
       FK_PDV INT64 NOT NULL,
       FL_VISITADO INT64 NOT NULL,
       DATA_VISITA DATE NOT NULL
  6
  8
  g
     -- Se necessario limpar a tabela
     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,
         FL_VISITADO,
 16
         DATA_VISITA)
 17
     SELECT
 18
 19
       id AS ID_VISITA,
 20
       CAST(115 + FLOOR(RAND() * 2) AS INT64) AS FK_PDV,
       CAST(FLOOR(RAND() * 2) AS INT64) AS FL_VISITADO,
DATE_ADD('2024-01-01', INTERVAL CAST(FLOOR(RAND() * 365) AS INT64) DAY) AS DATA_VISITA,
 21
 22
 23 FROM
 24 UNNEST(GENERATE_ARRAY(1, 100)) AS id;
```

Resposta:





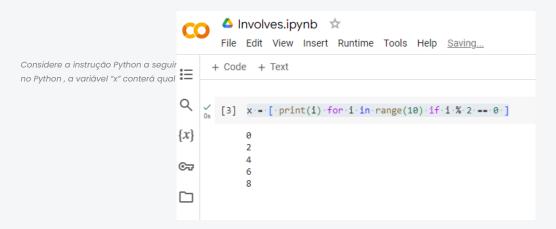
Considerando a query abaixo, a pessoa engenheira de dados identificou que a performance da query está muito

abaixo de esperado. Imaginando que um dos problemas possa estar relacionado aos índices das tabelas do banco de encontro entre dois campos e evita a necessidade de uma consulta de dados, a pessoa esolveu silar os indices nas tabelas. Liste quais possiveis campos devem ser indexados nas tabelas Liste quais possiveis campos de evita a necessidade de uma consulta tabelas liste quais possiveis campos de evita a necessidade de uma consulta tabelas liste quais possiveis campos de evita a necessidade de uma consulta tabelas liste quais possiveis campos de evita a necessidade de uma consulta tabelas liste quais possiveis campos de evita a necessidade de uma consulta tabelas liste quais possiveis campos de evita a necessidade de uma consulta tabelas liste quais possiveis campos de evita a necessidade de uma consulta tabelas liste quais possiveis campos de evita a necessidade de uma consulta tabelas liste quais possiveis campos de evita a necessidade de uma consulta tabelas liste quais possiveis campos de vita a necessidade de uma consulta de de evita a necessidade de uma consulta de la consult

```
Q Untitled query
                         C RUN

■ SAVE ▼

                                                业 DO
  1 SELECT
      FT.CICLO.
  2
     FT.ID_DIM_PDV,
  3
       FT.ID_BLOCO_ITEM,
  4
  5
      SUM(FT.QTD_PONTO_EXTRA),
      SUM(FTPI.TOTAL_NOTA_ITEM)
  6
  7 FROM
     FT_DOMINANCIA_PONTO_EXTRA_COMPLIANCE FT
  9 INNER JOIN
 10 TABREF_PAINEL_LOJAS_LP TPLL
 11 ON
     FT.ID_DIM_PDV = TPLL.ID_DIM_PDV
 12
 13
     AND FT.CICLO = TPLL.CICLO
 14 INNER JOIN
 15 FT_PERFECTSTORE_ITEM FTPI
 16 ON
     FT.CICLO = FTPI.CICLO
 17
 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 7

Faça um script em Python que peça dois números e imprima a soma

```
[11] # @title Default title text
# 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
```

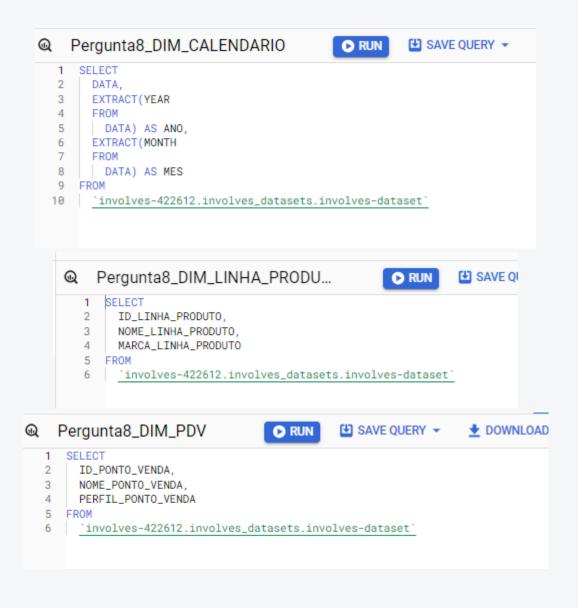
Para responder às questões 8, 9 e 10 A ETL final deve conter um job princi

10. Além disso, que tal ganhar um ponto a mais nessas questões ? Para isso, inclua o projeto criado em um repositório do Githyb **pergunta S**eja público para termos visibilidade, ok ?). Compartilhe por aqui o link para o repositório.

Construa uma transformação que deve usar como datasource o dataset (DATASET_TESTE_DE.csv) que contém informações de coletas de dados nos ponto de vendas. A ETL deve consultar o dataset e inserir, em uma base de

dados (nperguntora) não especifica o formato do año e mês (ex. YY, YYYY-MM, MMM, etc..), deixei a) Dimensão Calendário (DIM_CALENDARIO): Deve conter data, mês e ano da coleta

- b) Dinan-padrão vala (função ev Não rencontreir o poampo perfil da linha de produto, adicionei
- c) Dimensão Linha de Produto (DIM_LINHA_PRODUTO): Deve conter o id, nome e perfil da linha de produto MARCA_LINHA_PRODUTO



Construa uma transformação que deve usar como datasource o dataset (DATASET_TESTE_DE.csv) que contém informações de coletas de dados nos ponto de vendas. A transformação deve consultar o dataset e inserir, em uma base de dados (modelo dimensional), as informações coletadas, conforme as tabelas abaixo:

- a) Fato Disponibilidade (FT_DISPONIBILIDADE): Deve conter os ids de ligação das tabelas de dimensões criadas na questão anterior e a quantidade de presenças de cada linha de produto no mês de Setembro/20.
- b) Fato Disponibilidade Agregada (FT_DISPONIBILIDADE_AGREGADA): Deve conter os ids de ligação das tabelas de

dimerNão agreguei por DATA, porque conforme descrição apenas, pos dados de setembro são por ponto de venda no mês de Setembro/20. nevessarios.

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

Query results							
JOB INFORMATION		RESULTS CHA	ART 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_PRESENCA
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_PRESENCA DESC;
```

Query results

JOB IN	IFORMATION	RESULTS	CHART	
Row	ID_PONTO_VENDA	QTD_PRESENCA	7/	
1	115		7	
2	116		4	