



TOTI Diversidade
Curso: Análise de Dados

Apresentação técnica
Modelagem de dados em SQL server

Aluno: Xavier Alejandro Mata Ibarra

Maio, 2020

Levantamento dos Requisitos

O Análise descritivo que será apresentado foi focado em dados públicos disponíveis no site do Instituto Brasileiro de Geografia e Estatística (IBGE), os microdados usados foram da Pesquisa Nacional por Amostra de Domicílio (PNAD) - PNAD COVID19 no período de julho - outubro 2020.

O Público alvo da pesquisa foram todos os moradores residentes nos domicílios selecionados para responder a pesquisa.

A Forma de coleta da informação foi entrevista exclusivamente por telefone. A pesquisa conta com 5 módulos com 53 perguntas no total.

As exigências do análise foram:

- Escolher 15 perguntas como máximo
- Período mínimo de meses

Para a modelagem de dados foi usado o Microsoft SQL management Studio Server 18 e para apresentação dos resultados foi usado o POWER BI Desktop.

A escolha das perguntas

O interesse do Análise foi focado nas perguntas sobre a internação durante esse período e sua relação com resultado positivo do SWAB COVID para relacionar as comorbidades que os entrevistados referiam ter, para observar a relação entre aqueles que ingressaram na Unidade de Tratamento Intensivo (UTI) ou não.

Na revisão dos formulários foram escolhidas as seguintes perguntas:

- UF: Unidade Federal
- V1013: Mês da pesquisa
- A002: Idade do morador
- A003: Sexo
- B005: Ao procurar o hospital, teve que ficar internado por um dia ou mais?
- B006: Durante a internação, foi sedado, entubado e colocado em respiração artificial com ventilador?
- B009A: Fez o exame coletado com cotonete na boca e/ou nariz (SWAB)?
- B009B: Qual o resultado?
- B0101: Algum médico já lhe deu o diagnóstico de diabetes?
- B0102: Algum médico já lhe deu o diagnóstico de hipertensão?
- B0103: Algum médico já lhe deu o diagnóstico de asma/bronquite/enfisema/doenças respiratória crônica ou doença de pulmão?

Escolhi filtros para obter uma amostra mais específica com critérios de seleção, os critérios foram: por grupo etário de maiores ou igual a 18 anos, se respondeu “sim” na pergunta de internação, e se respondeu “sim” para um resultado positivo de SWAMB COVID 19. Só foi escolhido o resultado do SWAB porque é o exame laboratorial que confirma a infecção ativa pelo COVID 19 com a menor margem de falsos positivos ou falsos negativos.

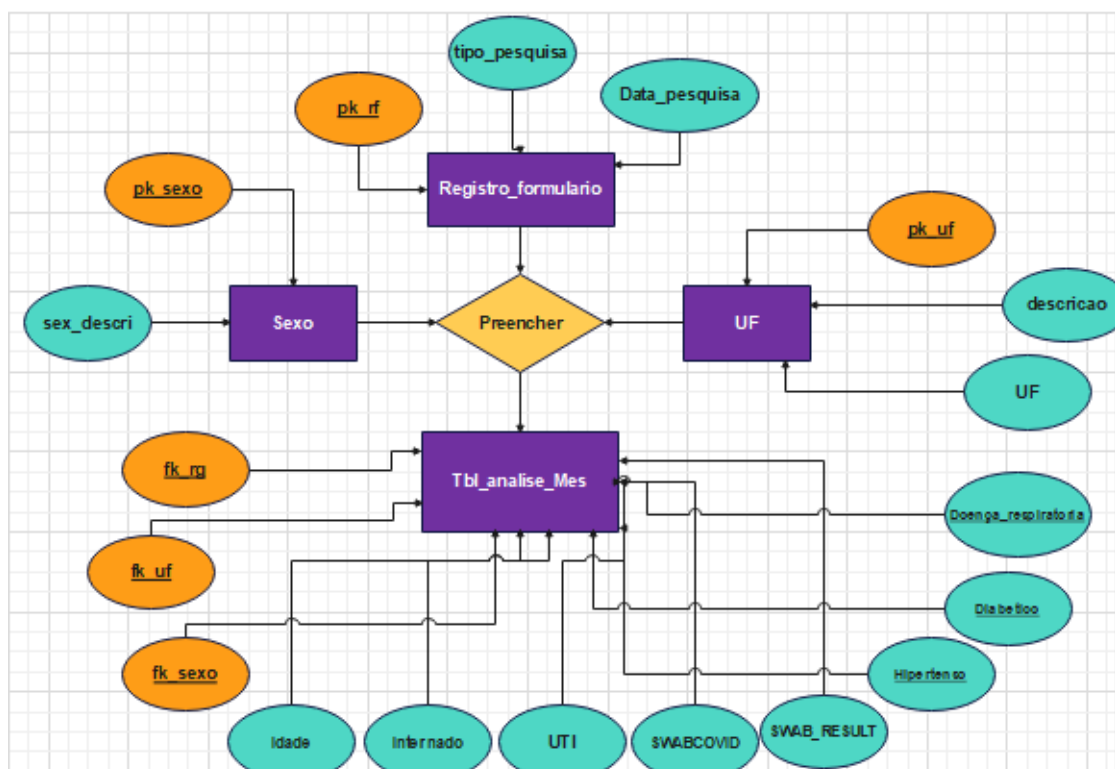
Para a manipulação dos dados foi feito a importação dos dados no SQL, mas antes disso foi criado o modelo entidade relacionamento e modelo lógico para começar a modelagem dentro da base de dados.

Identificação de entidades e relacionamento

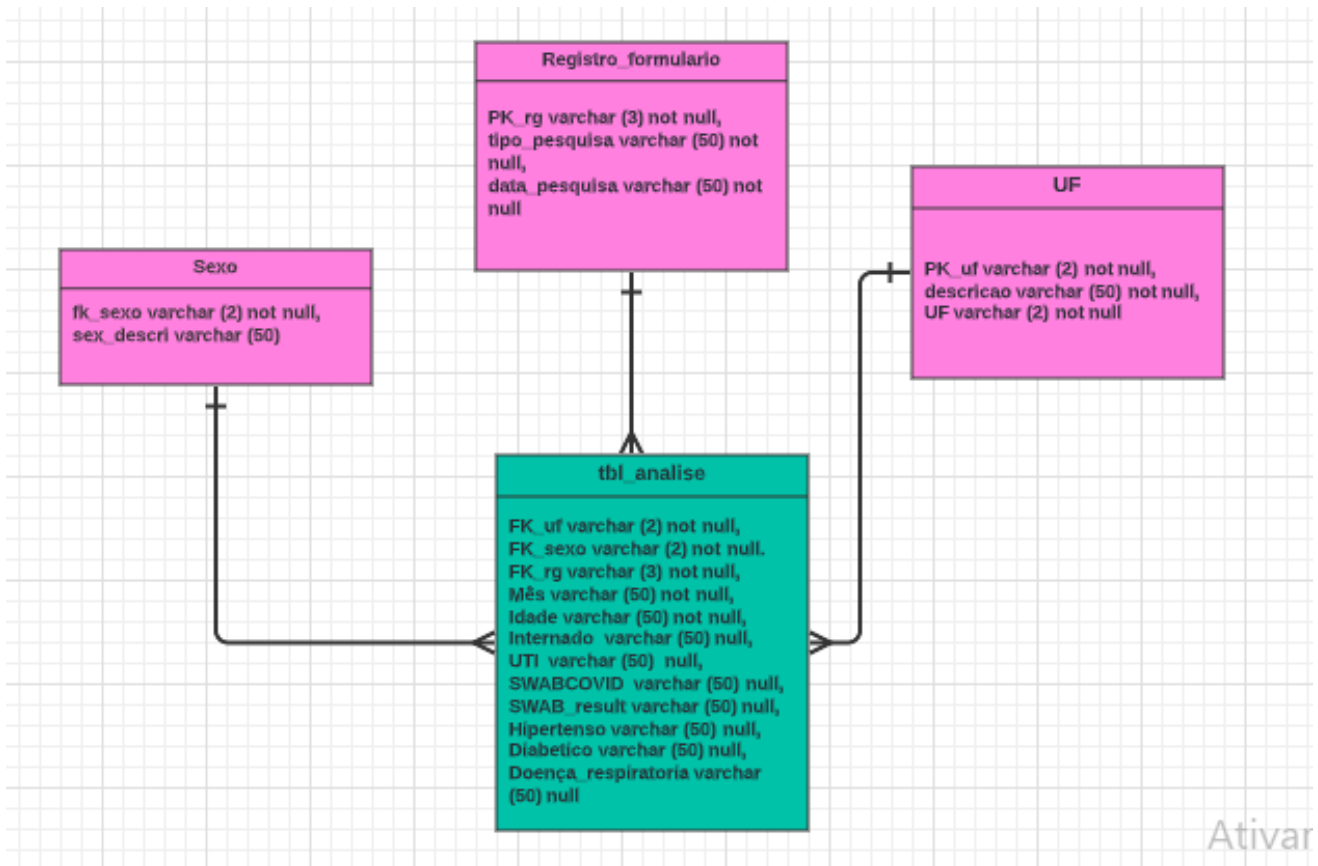
O modelo a seguir foi pensado na manipulação dos dados de forma eficiente criando o modelo em base a formulários que devem ser armazenados e relacionados a traves de uma entidade registro formulário, Unidade federal e pelo Sexo dos entrevistados, que são perguntas padrões que ajudam a esquematizar e entender melhor os dados.

Modelo Entidade relacionamento.

O modelo representa a organização da modelagem e o fluxo de informação, a entidade tbl_analise_mes terá uma tabela por mês, todas as tabelas estarão relacionadas com as entidades UF, Sexo, registro_formulario.



Modelo Lógico



CÓDIGOS SQL COMENTADOS

```
-- Realizei a seleção das perguntas da tabela 'desafio_072020' que apresenta todos os dados do PNAD_COVID de julho.
-- Vou dar uma olhada nos dados aplicando meus criterios de seleção e fazendo uma olhada geral.

SELECT COUNT(B005) FROM desafio_072020 where B005 = 1; -- FORAM 265 ENTREVISTADOS INTERNADO NO MES DE JULHO
SELECT COUNT(B009B) FROM desafio_072020 where B009B = 1; -- 2052 COM SWAB COVID POSITIVO
SELECT * FROM desafio_072020 WHERE B005 = 1 AND B009B = 1; -- 70 ENTREVISTADOS REFEREM SER INTERNADOS COM UM SWAB COVID POSITIVO

-- Com o proximo comando vou criar uma tabela nova com as colunas que preciso, utilizo uma clusula WHERE para filtrar o resultado em função a minha pesquisa

SELECT UF, Ano, V1013, A002, A003, B005, B006, B009A, B009B, B0101, B0102, B0103 --Colunas
INTO tbl_analise -- Nome da Tabale nova
FROM desafio_072020 -- Tabela PNAD_COVID
where B005 = 1 AND B009B = 1 AND A002 >= 18; --Filtros

SELECT * FROM tbl_analise --Confirmo a existencia da minha tabela

-- Agora inicio o modelagem da tabela nova.

--1 - Vamos renomear as colunas para que fique entendivél os dados escolhidos, usando o comando 'sp_rename'

sp_rename 'tbl_analise.UF', 'id UF', 'COLUMN'; -- Depois do sp_rename segue: 'nome_da_tabela.coluna', 'Novo_nome_da_coluna', 'COLUMN';

sp_rename 'tbl_analise.V1013', 'Mes', 'COLUMN';

sp_rename 'tbl_analise.A002', 'Idade', 'COLUMN';

sp_rename 'tbl_analise.B005', 'Internado', 'COLUMN';

-- 2- Com os nomes das colunas trocadas, vamos começar a trocar os dados que estão dentro. As columns contem dados de tipo varchar (50)
-- posso trocar elas por outro valor que seja mais entendivél. A maioria de meus dados contem valores 1 e 2.
-- Sendo 1 = 'sim' e 2 = 'Não'. Podemos trocar isso com a clausula UPDATE (atualizar).

UPDATE tbl_analise -- Nome da minha tabela
SET Sexo = 'F' -- Nome da coluna com o novo valor que quero nela
WHERE Sexo = '2'; -- Nome da coluna com o valor atual que quero substituir

-- Percebi uma inconsistencia na tabela analise2 (Agosto), onde tem um valor '9, definido no dicionario do PNAD como 'ignora',
-- a pergunta é se algum Médico deu diagnostico de alguma das enfermidades descritas, o valor 'ignora'
-- pode ser considerado como Não, sendo uma resposta negativa de não saber se recebeu o diagnostico, por isso sera considerada como 'Não' na minha analise.

UPDATE tbl_analise2
SET Hipertenso = 'Não'
WHERE Hipertenso= '9';
```

```
-- Agora temos um melhor visual da tabela para compreender as colunas e seus dados

-- 3- Vou criar minhas tabelas UF, Sexo e Registro_Formulario para associar e identificar minhas tabelas

CREATE TABLE UF --Nome da tabela
(pk_UF varchar(50) primary key not null, -- Nome da coluna e seu tipo, sendo no nulo.
descricao varchar(100) not null,
UF varchar(3) not null);

-- INSERIMOS OS DADOS

INSERT INTO UF (pk_UF, descricao, UF) -- Nome da tabela e entre aspas as colunas que vou atualizar
VALUES -- Valores que vou atualizar dentro de cada coluna em ordem
('11', 'Município de Porto Velho', 'RO'),
('12', 'Município de Rio Branco', 'AC'),
('13', 'Município de Manaus', 'AM'),
('14', 'Município de Boa Vista', 'RR'),
('15', 'Município de Belém', 'PA'),
('16', 'Município de Macapá', 'AP'),
('17', 'Município de Palmas', 'TO'),
('21', 'Município de São Luís', 'MA'),
('22', 'Município de Teresina', 'PI'),
('23', 'Município de Fortaleza', 'CE'),
('24', 'Município de Natal', 'RN'),

-- CRIAMOS TABELA SEXO

CREATE TABLE Sexo
(pk_sexo varchar(2) primary key,
descricao varchar (50) not null);

-- INSERIMOS OS DADOS

INSERT INTO Sexo (pk_sexo, descricao)
values ('F', 'Femenino'),
('M', 'Masculino');

ALTER TABLE Sexo
ALTER COLUMN pk_sexo varchar (50);
```

```
-- CRIANDO TABLE REGISTRO_FORMULARIO
```

```
CREATE TABLE Registro_formulario
( pk_rf varchar(10) primary key,
  tipo_pesquisa varchar (50) not null,
  data_pesquisa varchar(10) not null );
```

```
-- INSERINDO DADOS
```

```
INSERT INTO Registro_formulario ( pk_rf , tipo_pesquisa, data_pesquisa)
VALUES ('F7', 'PNAD_COVID', '07/2020'),
       ('F8', 'PNAD_COVID', '08/2020'),
       ('F9', 'PNAD_COVID', '09/2020'),
       ('F10', 'PNAD_COVID', '10/2020');
```

```
-- 4- Criamos uma coluna que será chave estrangeira entre minha tabela analise e as nova tabela Registro_formulario.
```

```
ALTER TABLE tbl_analise
ADD fk_registro varchar (50) foreign key
REFERENCES Registro_formulario (pk_rf);
```

```
--Usei o mesmo código para as outras chaves estrangeiras com a tabela Sexo e UF.
```

```
-- Preenchemos os dados dessa coluna nova em tbl_analise
```

```
UPDATE tbl_analise
SET fk_registro = 'F7'
```

```
-- CONFIRMANDO A FUNCIONALIDADE DAS CHAVES COM A SEGUINTE CONSULTA
```

```
SELECT UF, descricao, Genero, Idade, UCI -- AS COLUNAS
FROM tbl_analise AS ta
JOIN Sexo ON Sexo.pk_sexo = ta.fk_sexo
JOIN UF ON UF.pk_UF = ta.fk_UF;

SELECT data_pesquisa, tipo_pesquisa, idade
FROM tbl_analise AS ta
join Registro_formulario AS rf ON rf.pk_rf = ta.fk_registro;
```

```
-- 5- Os proximos passos é utilizar os mesmos codigos nas tabelas analise2 (Agosto), analise3 (setembro) e analise4 (outubro)
```

```
-- 6- Importar em power BI
```